LNBIP 108

Michael J. Shaw
Dongsong Zhang
Wei T. Yue (Eds.)

E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life

10th Workshop on E-Business, WEB 2011
Shanghai, China, December 2011
Revised Selected Papers

Springer

Lecture Notes in Business Information Processing 108

Michael J. Shaw
Dongsong Zhang
Wei T. Yue (Eds.)

E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life

10th Workshop on E-Business, WEB 2011
Shanghai, China, December 4, 2011
Revised Selected Papers

Springer

Volume Editors

Michael J. Shaw
The University of Illinois at Urbana-Champaign
Champaign, IL, USA
E-mail: mjshaw@illinois.edu

Dongsong Zhang
The University of Maryland
Baltimore County, MD, USA
E-mail: zhangd@umbc.edu

Wei T. Yue
City University of Hong Kong
HKSAR, China
E-mail: wei.t.yue@cityu.edu.hk

ISSN 1865-1348 e-ISSN 1865-1356
ISBN 978-3-642-29872-1 e-ISBN 978-3-642-29873-8
DOI 10.1007/978-3-642-29873-8
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012936045

ACM Computing Classification (1998): J.1, K.4.4, H.3.5, H.4

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

New advances in information and communication technologies (ICT) continue to rapidly transform how business is done and to change the role of information systems in business. E-business technologies and applications are also stimulating new business models and redefining relations and processes within and across business organizations. The ubiquity of digital devices is fast dismantling the barriers between commercial, work, and social activities. Consequently, shifts in managerial principles and practices are emerging in organizations, as Generation M or the Net Generation joins the workforce. Similarly, businesses are adopting lessons learned from social networks and social computing. This marks a striking contrast to the pre-Web eras where technology emanated from the workplace with limited impact on personal or social activities. Many exciting research questions are emerging.

This book constitutes the refereed proceedings of the Workshop on E-Business, held in Shanghai, China, in December 2011. The workshop provided an open forum for e-business researchers and practitioners worldwide to explore and respond to the challenges of next-generation e-Business systems, share the latest research findings, explore novel ideas, discuss success stories and lessons learned, map out major challenges, and collectively chart future directions of e-business. We received 88 papers from 13 countries. This proceedings volume includes 40 papers on topics that are diverse yet highly relevant to today's e-business problems. The papers are organized in topical sections on social networks, business intelligence and social computing, economics and organizational implications of electronic markets, and e-business systems and applications.

February 2012

Michael J. Shaw
Dongsong Zhang
Wei T. Yue

Organization

Honorary Chair

Andrew B. Whinston	University of Texas at Austin, USA

Conference Chair

Michael J. Shaw	University of Illinois at Urbana-Champaign, USA

Program Co-chairs

Dongsong Zhang	University of Maryland, Baltimore County, USA
Christian Wagner	City University of Hong Kong, SAR China
Wei T. Yue	City University of Hong Kong, SAR China

Local Arrangements Chairs

Zhangxi Lin	Texas Tech University, USA and Tongji University, China
Jie Lin	Tongji University, China
Jia Li	East China University of Science and Technology, China

Track Chairs

Dirk Neumann	University of Freiburg, Germany
Raymond Lau	City University of Hong Kong, SAR China
Zhijun Yan	Beijing Institute of Technology, China
Lina Zhou	University of Maryland, Baltimore County, USA
Mu Xia	Santa Clara University, USA
Ke-Wei Huang	National University of Singapore, Singapore
Kai Lung Hui	Hong Kong University of Science and Technology, SAR China
Michael Chau	The University of Hong Kong, SAR China
Dmitri Roussinov	University of Strathclyde, UK
Philip O'Reilly	University College Cork, Ireland
Aidan Duane	Waterford Institute of Technology, Ireland
Steve Russell	George Washington University, USA
Selwyn Piramuthu	University of Florida, USA

Goel, Sanjay	University at Albany, USA
Heng Xu	Penn State University, USA
Chee Wei Phang	Fudan University, China
John Zhang	University of Connecticut, USA
Arvind Tripathi	University of Auckland, New Zealand
Roman Beck	University of Frankfurt, Germany
Daniel Zeng	University of Arizona, USA
Sinan Aral	New York University, USA
Samuel Fosso Wamba	University of Wollongong, Australia
Ji Yonghua	University of Alberta, Canada
Sidne G. Ward	University of Missouri, Kansas City, USA
Suparna Goswami	Technische Universität München, Germany

Program Committee

Benjamin Shao	Arizona State University, USA
Chen Zhang	The University of Memphis, USA
Chih-Ping Wei	National Taiwan University, Taiwan, ROC
Christian Schaupp	West Virginia University, USA
Christoph Riedl	Harvard Business School, USA
Dan Ma	Singapore Management University, Singapore
Dan Zhu	Iowa State University, USA
Fang Fang	California State University at San Marcos, USA
Gregory Kersten	Concordia University, Canada
H.R. Rao	State University of New York Buffalo, USA
Han Zhang	Georgia Institute of Technology, USA
Hong Guo	University of Notre Dame, USA
Honghui Deng	University of Nevada, Las Vegas, USA
Hsing Kenneth Cheng	University of Florida, USA
Hulisi Öğüt	TOBB University of Economics and Technology, Turkey
Indranil Bose	The University of Hong Kong, SAR China
Jackie Rees Ulmer	Purdue University, USA
Jan Kraemer	Karlsruhe Institute of Technology (KIT), Germany
Joseph Barjis	Delft University of Technology, The Netherlands
Juliana Sutanto	ETH Zurich, Switzerland
Katina Michael	Wollongong University, Australia
Kexin Zhao	University of North Carolina, Charlotte, USA
Kutsal Dogan	Ozyegin University, Istanbul, Turkey
Lemuria Denise Carter	North Carolina A&T State University, USA
Manish Agrawal	University of South Florida, USA
Matt E. Thatcher	University of Louisville, USA
Minder Chen	California State University Channel Islands, USA

Mingfeng Lin	University of Arizona, USA
Nirup Menon	George Mason University, USA
Olivera Marajnovic	University of Sydney, Australia
Patrick Y.K. Chau	The University of Hong Kong, SAR China
Raj Sharman	State University of New York, Buffalo, USA
Ravi Sen	Texas A&M University, USA
Robert P. Biuk-Aghai	University of Macau, SAR China
Rui Chen	Ball State University, USA
Seung Hyun Kim	National University of Singapore, Singapore
T.S. Raghu	Arizona State University, USA
Ulrich Bretschneider	University of Kassel, Germany
Verny Jerome	Rouen Business School, France
Victoria Yoon	Virginia Commonwealth University, USA
Wei Zhou	ESCP Europe, France
Xianjun Geng	University of Texas at Dallas, USA
Xue Yang	Nanjing University, China
Jie Lin	Tongji University, China
Yiching Liou	National Chin-Yi University of Technology, Taiwan, ROC
Zhangxi Lin	Texas Tech University, USA and Tongji University, China
Zhengrui Jiang	Iowa State University, USA
Zhiling Guo	City University of Hong Kong, SAR China

Table of Contents

Part I: Social Networks, Business Intelligence and Social Computing

Social Networks and Social Computing

Enterprise Web Mining, Web Analytics, and Business Intelligence

Part II: Economics and Organizational Implications of Electronic Markets

Online Security Informatics and Privacy Issues

Economics of E-Commerce

Organizational Implications of Electronic Markets

Part III: E-Business Systems and Applications

Cloud Computing and Applications

Collaborative Systems

Supply Chain and Distributed Systems

Part I

Social Networks, Business Intelligence and Social Computing

Mining Implicit Social Network with Context-Aware Technologies

Eun Jung Yoon[1] and Wei Zhou[2]

[1] Information Systems and Operations Management
University of Florida, Gainesville, FL 32611, USA
[2] Information Systems and Technologies
ESCP Europe, 75543 Paris Cedex 11, France

Abstract. Given their growing importance with the fast advance of today's information technologies, social networks have been extensively studied. However, a majority of existing published literature in this area consider only the explicit form of social networks. We consider its complement - implicit social networks. We adapt the social distance model and influence model to a basic implicit social network scenario. We then extend the basic model by incorporating the concept of multiple network paradigms.

Keywords: Implicit Social Network, Context-Aware Technology, RFID.

1 Introduction

With the development of an increasing number of online social networking platforms such as Flickr, Facebook, Youtube, among others, frequent users of these services find them to be powerful means to share, organize and search for content and contacts from such Internet social networks. The structural analysis of social networks has proved to be an important tool for business and sociological study of individual action and collective phenomena. However, a majority of existing literature on social networking are limited to *explicit* social relations/ties such as those that are explicitly created by individuals or groups in those online social sites.

Social relations/ties also exist alongside as *implicit* knowledge and decision-making artifacts in a given situation. These implicit phenomena exist in addition to or, in some cases, even independent of the explicit phenomena. For example, questions such as "Do individuals in a professional team generally see each other as friends?" implies the existence of implicit social networks. While the answer remains unknown from our review of existing literature, we are interested in exploring such unclaimed social ties. We believe that implicit social networks could even have a more profound impact on business practice from the perspective of organizational behavior and structure to marketing design compared to information generated from explicit social networks.

Motivated by the gap in existing literature, we first investigate the possibility to discover implicit social networks by utilizing context-aware technologies. For example, one can use RFID(Radio-Frequency IDentification)-generated data to determine social relationships among consumers of a B&M (Brick&Mortar) retailer. We argue that such implicit social networks discovered in the context of a local retailer could prove to be

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 3–8, 2012.

complementary to that from the Internet social network. As a result, it could generate valuable information for Internet business practitioners who own both physical and virtual channels. We then attempt to model implicit social networks in a general form based on traditional distance model as well as influence model.

The remainder of this paper is organized as follows. We begin with a brief literature review and motivating examples. We then present the basic scenario with single paradigm in Section 3. We then extend the model by incorporating multiple paradigms. Section 4 concludes this research-in-progress paper and highlights possible extensions for future research.

1.1 Literature Review

Although our focus in this study is on implicit social network discovery, the concept of unobserved or hidden relationships in a social network is not new. [7] develops a set of latent structure analysis of social network where the probability of a relationship between individuals depends on the positions of individuals in an unobserved social space, with the results derived under maximum likelihood and Bayesian frameworks.

In some sense, we find implicit social network mining to be very similar to business process mining. [9] classifies business processes from three different perspectives including the process perspective, the organizational perspective and the case perspective. Our focus is on the organizational perspective, which we believe is appropriate for Social Networking Analysis (SNA) that considers the relationship between individual entities in a group. [9] describe the offices of the Dutch national public works department in charge of maintenance of the road and water infrastructure in a city. They analyze the processing of invoices sent by suppliers by using various process mining techniques. As for mining from an organizational perspective, they examine the relationship between workers who perform the different steps and how they are related to one another.

[4] suggests a business process framework to classify social network analysis and mining from a business perspective. This framework introduces tools and techniques for analyzing social networks that could be used in order to improve overall business processes. Several researchers have studied sociograms from email logs in order to analyze the communication structure between organizations. [5] investigate corporate mailing lists to automatically generate approximate corporate social relationships. They found that co-occurrence in mailing lists provided a good predictor to understand relationships among users. [6] show how software can be developed to help support collaboration among individuals by using the Soylent tool. They found particular characteristics from the patterns observed in their study that can be used to explain contacts and communication in this environment.

Although there exists many methodologies and models to study social networks, we find the latent distance model [7] [8] and social influence model [2] to be more appropriate for our study.

2 Implicit Social Network Mining

In contrast to explicit social network, in which individuals claim/develop their relationships with one another, implicit social network deals with implicitly developed and confirmed relationships. The relationship/tie has too be inferred from the activities within the group. Our first example illustrates probably the simplest implicit social network in our daily lives, by observing the paths taken by involved persons in a community through context-aware technologies. We study the hidden (i.e., implicit) relationships among these people from tracking data. The technology considered, RFID, is not uncommon in practice.

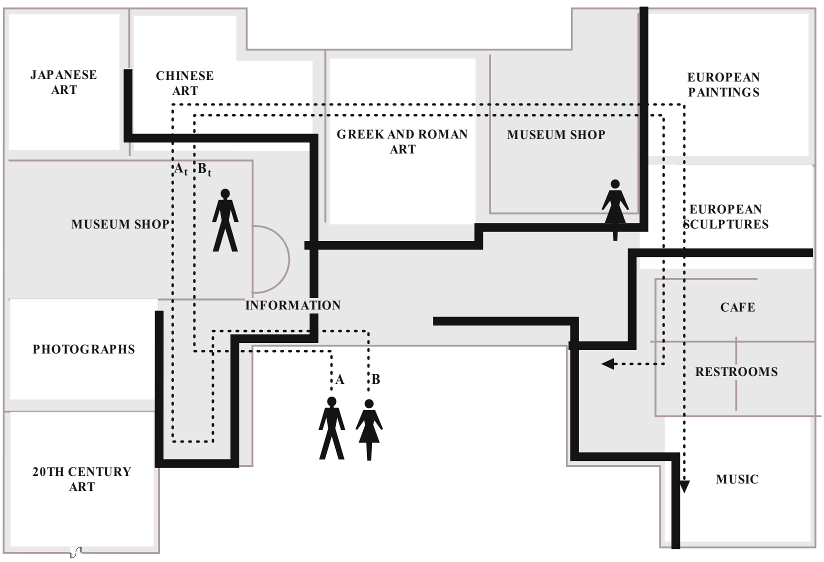


Fig. 1. Context-aware path tracking of museum visitors

Figure 1 illustrates the physical movement of museum visitors. By studying the patterns in the tracking data, based on common-sense, we can observe and “guess" that two individuals may have a relationship/tie if they are observed to be together most of the time while they are both inside the museum. Such implicit social network can also be discovered even on a virtual platform by examining individuals’ cyber-footprints.

2.1 One Paradigm Scenario

Similar to existing literature in social network modeling, our study is based on a setting where a group of individuals (also called agents or users) are represented by nodes in a social network G. Unlike a majority of existing literature in this area, we use the concept of reverse-engineering to mine the implicit (hidden) social network.

$$P(Y|Z,X,\theta) = \prod_{i \neq j} P(y_{i,j}|z_i, z_j, x_{i,j}, \theta) \tag{1}$$

The latent distance model studies the cross-relationship among the individuals, such as earlier studies using the Monk Data (Sampson 1968, Holland and Leinhardt 1981, Reitz 1982, Wang and Wong 1987). The relationship between the monks in the Monk Data are transitive in the sense that "a friend's friend is also a friend" ($i \rightarrow j, j \rightarrow k, \Rightarrow i \rightarrow k$). For example, equation 1 may take the form of parametrization

$$\eta_{i,j} = logodds(y_{i,j} = 1|z_i, z_j, x_{i,j}, \alpha, \beta) = \alpha + \beta' x_(i,j) - |z_i - z_j| \tag{2}$$

which means that $d_{i,j} \leq d_{i,k} + d_{k,j}$, letting $d_{i,j} = |z_i - z_j|$.

In contrast, we are interested in converting the physical/virtual social distance to meaningful implicit social networks. Thus different from the latent distance model, we take a reverse perspective by studying the measurement of distance Z. We argue that the observed implicit social distance at any point in time $d_{i,j}|_t$ is a pure random event and only the accumulated observation reveals the real implicit relationships among the individuals.

$$P(Y|Z, X, \theta) = \int_0^T \prod_{i \neq j} P(y_{i,j}|z_i, z_j, x_{i,j}, \theta)|_t dt \tag{3}$$

The inclusion of time period $[0, T]$ makes it similar to the social influence model. In the influence and correlation model, social correlation is observed when a subset of the group becomes active after an individual performs the action for the first time [2]. A simple social correlation coefficient can be found as $ln\left(\frac{p(a)}{1-p(a)}\right) = \alpha ln(\alpha + 1) + \beta$, where $p(a) = \frac{e^{\alpha ln(\alpha+1)+\beta}}{1+e^{\alpha ln(\alpha+1)+\beta}}$ represents the probability of activation for an individual with α already active agents.

The implicit social network that we consider in this study ignores the assumption of influence that influential social correlation only exists when actions are followed. Instead, we argue that two individual can simultaneously initiate an action, and as time continues, they can choose to keep the action or to deactivate it at any time in the future. The correlation matters only when it's significant after accumulation.

The simplest form of social distance model is to measure the physical distance of individuals. In this setting, Z is a two-dimensional vector that represents one's location in physical space.

$$P(Y|Z, X, \theta)|_{\bar{i}} = \frac{\int_0^T \prod_{i \neq j} P(y_{i,j}|z_i, z_j, x_{i,j}, \theta)|_t dt}{\int_0^T z_i dt} \tag{4}$$

2.2 Extended Scenario

As an extension to the previously considered scenario, we use the concept of paradigm to distinguish the various angles through which social networks could be rendered different weights and values. For example, the observation that two individuals go shopping together every Saturday afternoon may have less of an impact on the professional paradigm of their own jobs than on the personal paradigm which can be used by retailers to discover their shopping behavior (Figure 2). We introduce the $n \times m$ paradigm matrix $|\lambda|_{n \times m}$ to study the inherent inter-correlations, with each cell $\lambda_{i,j}$ representing the weight of the i^{th} paradigm on the j^{th} network setting. This paradigm matrix enables us to distinguish various implicit networks in the same social network setting. For example, the existence of a tie/relationship between the two individuals i and j under the k^{th} paradigm (personal) in the l^{th} network setting (shopping), $y_{i,j,k,l} = 1$, may disappear under another paradigm (professional, or some other special-interest paradigms) in the same network setting, $y_{i,j,\overline{k},l} = 0$.

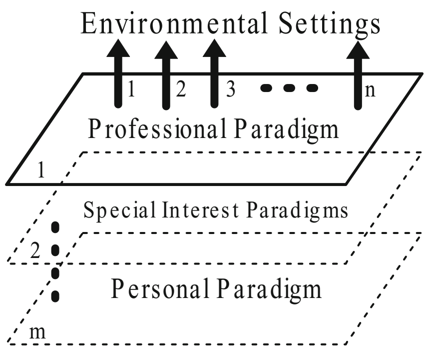


Fig. 2. Paradigmatic view of implicit social networks

In a simple form, $|\lambda|_{n\times m}$ is multiplied with $P(Y|Z, X, \theta)$. We leave the different mechanisms for integrating the paradigm matrix in implicit social network discovery for future research. Furthermore, by incorporating implicit social network in different paradigm layers, we can infer the social relationships in one paradigm from another paradigm.

3 Discussion

In this study, we considered implicit social network with a distance model. We then extended the basic model to incorporate the concept of social paradigm. The primary purpose of this paper is to point out a deficit or gap present in existing research literature on social networks. We, therefore, did not present extensive analyses in this paper. Out next step is to conduct experiments and data analysis to validate the model. We have begun this process and, due to time constraints, we are unable to complete this as of now. We leave this as an exercise for our future studies/publications.

References

1. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, pp. 29–42 (2007)
2. Anagnostopoulos, A., Kumar, R., Mahdian, M.: Influence and correlation in social networks. In: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 7–15 (2008)

3. Blackwell, D.: Equivalent Comparisons of Experiments. The Annals of Mathematical Statistics 24(2), 265–272 (1953)
4. Bonchi, F., Castillo, C., Gionis, A., Jaimes, A.: Social network analysis and mining for business applications. ACM Transactions on Intelligent Systems and Technology (TIST) 2(3), 22 (2011)
5. Farnham, S., Portnoy, W., Turski, A.: Using email mailing lists to approximate and explore corporate social networks. In: Proceedings of the CSCW (2004)
6. Fisher, D., Dourish, P.: Social and temporal structures in everyday collaboration. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2004)
7. Peter, D.H., Adrian, E.R., Mark, S.H.: Latent Space Approaches to Social Network Analysis. Journal of the American Statistical Association 97(460), 1090–1098 (2002)
8. Wang, Y.J., Wong, G.Y.: Stochastic Blockmodels for Directed Graphs. Journal of the American Statistical Association 82(397), 8–19 (1987)
9. Van der Aalst, W.M.P., Reijers, H.A., Weijters, A.J.M.M., van Dongen, B.F., Alves de Medeiros, A.K., Song, M., Verbeek, H.M.W.: Business process mining: An industrial application. Information Systems 32(5), 713–732 (2007)

Using Social Network Classifiers for Predicting E-Commerce Adoption

Thomas Verbraken[1], Frank Goethals[2], Wouter Verbeke[1], and Bart Baesens[1,3]

[1] Katholieke Universiteit Leuven, Department of Decision Sciences and Information Management, Naamsestraat 69, 3000 Leuven, Belgium
{thomas.verbraken,wouter.verbeke,bart.baesens}@econ.kuleuven.be
[2] IESEG School of Management (LEM-CNRS), IS Department, Lille, France
f.goethals@ieseg.fr
[3] School of Management, University of Southampton, Highfield Southampton, SO17 1BJ, United Kingdom

Abstract. This paper indicates that knowledge about a person's social network is valuable to predict the intent to purchase books and computers online. Data was gathered about a network of 681 persons and their intent to buy products online. Results of a range of networked classification techniques are compared with the predictive power of logistic regression. This comparison indicates that information about a person's social network is more valuable to predict a person's intent to buy online than the person's characteristics such as age, gender, his intensity of computer use and his enjoyment when working with the computer.

Keywords: E-commerce adoption, Social network analysis, Classification, Data mining.

1 Introduction

In the last 20 years many different purchasing process models have been studied and presented (Engel et al., 1995; Anderson and Chambers; Hansen, 2005). Nowadays, internet based services can support all steps in the purchasing process: potential customers can use commercial websites to find information about a product or service, to find suppliers, to actually complete a purchase, etc. The question arises which users accept to use the internet for a particular step in the purchasing process. One of the most examined topics in the Information Systems (IS) field concerns IS use and acceptance. User acceptance was defined as the demonstrable willingness within a user group to employ information technology for the tasks it is designed to support (Dillon and Morris, 1996). Several models have been developed that present antecedents of use behavior. Two major antecedents of use behavior are the intent to use the system and facilitating conditions (Venkatesh et al., 2003). Furthermore, several antecedents of the intent to use systems (in general) have been identified as well (Venkatesh et al., 2003). The Technology Acceptance Model (Davis, 1989) identifies two main antecedents of the intention to use a system (behavioral intention):

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 9–21, 2012.

the perceived usefulness and the perceived ease of use of the system. The Unified Theory of Acceptance and Use of Technology (UTAUT), developed by Venkatesh et al. (2003), builds upon the Technology Acceptance Model (TAM) and many other theories (such as the Theory of Planned Behavior (Taylor and Todd, 1995) and the Theory of Reasoned Action (Karahanna and Straub, 1999)), and is considered as a more complete model in modern literature. The UTAUT unifies the general determinants of system appreciation (e.g. performance expectancy, facilitating conditions and effort expectancy) with individual-related variables such as gender, age or experience using the system and social influences. The key dependent variable of behavioral intention is defined as the degree to which a person has formulated conscious plans to perform or not perform some specified future behavior (Warshaw and Davis, 1985).

This paper aims to gain insight into one of the antecedents of behavioral intention: the social influence. Social influence (SI) was defined as the degree to which an individual perceives that important others believe he or she should use the system (Venkatesh et al., 2003). This concept is usually measured by asking users questions such as 'People who influence my behavior think that I should use the system' and 'People who are important to me think that I should use the system'. As suggested by its definition, the social influence construct should thus be regarded as the user's perception of the social influence. One shortcoming of existing research methods is thus that they only consider the perceived social influence and not the latent social influence. A second shortcoming is that the use of surveys to measure perceived social influence – as is usually done – possibly distorts the picture. People may prefer not to reveal that they think they are influenced by others. The perceived social influence construct was found to be significant when system usage was mandatory, but was not significant in voluntary contexts (Venkatesh et al., 2003). The paper at hand intends to gain more insight in the role of social networks in system acceptance in a voluntary context, by avoiding the two shortcomings mentioned above. More specifically, with the goal to adapt the UTAUT in future research, this study will show that knowledge about the social network of a (potential) user is valuable when predicting the intent to use a system in a voluntary context; in casu the intent to buy products online. Moreover, this paper will show that information about the social network is more valuable than the person's characteristics such as age, gender and enjoyment when working with computers. Possible explanations for the existence of social network effects are the word-of-mouth effect and social leader influence. The principle that a contact between similar people occurs at a higher rate than among dissimilar people has been observed in many kinds of social networks and is called homophily (Blau, 1977; McPherson et al., 2001; Macskassy and Provost, 2007), assortativity (Newman, 2010), or relational autocorrelation (Jensen and Neville, 2002).

In this paper, network based classification techniques are applied to predict whether a user is likely to adopt e-commerce, based on social network information. Network based classifiers have received a growing amount of attention during the last decade. This is partly due to the increasing amount of networked data that has become available, i.e. data with entities that are interlinked, such as for instance web pages on the internet and co-authorship in the scientific community. Section 2 introduces a range of network based classification techniques. Section 3 presents the method used to collect data about people's willingness to buy products online and about their social networks. Section 4 discusses the experimental setup followed in this study. In Section 5, the results of the experiments are discussed, and Section 6 summarizes the conclusions and outlines tracks for future research.

2 Classification in Social Networks

Formally, a social network is defined by a graph $\boldsymbol{G}$, consisting of a set of nodes $v \in \boldsymbol{V}$,that are connected by a set of links $e \in \boldsymbol{E}$**;** and $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$. The number of elements in the sets $\boldsymbol{V}$ and $\mathbf{E}$ are denoted respectively by n and k. The nodes in this study represent persons whereas the links represent social ties. The links in a graph can be directed or undirected. Directed links point from an origin to a destination node and incorporate a direction property, whereas undirected links do not. In this study, undirected links are adopted, and maximum one link exists between two persons v_i and v_j. A graph can be represented by an adjacency matrix $\boldsymbol{A} = (a_{ij})$ or a weight matrix $\boldsymbol{W} = (w_{ij})$ of size $n \times n$. An entry a_{ij} or w_{ij} represents the edge between a vertex v_i and a vertex v_j. The value of a_{ij} in an adjacency matrix is equal to one if an edge exists between vertices v_i and v_j, and equal to zero when no connection exists. In principle, the weights $w_{ij} \in \boldsymbol{W}$ can take any value. Typically, a weight expresses a property or characteristic of a link, in this paper the strength of a social tie between two persons. A value of w_{ij} equal to zero means that no relation exists between two vertices.

Besides the social network structure, there is further information about the nodes themselves. Assume a data set $\boldsymbol{D} = (\boldsymbol{x}_i; l_i)$ with $i = 1 \dots n$, x_i an attribute vector representing the information about person i (e.g. age, gender, etc.), and l_i a class label indicating the value for a binary target variable of the customer i (e.g. whether the customer intends to purchase the product online). In the remainder of this study, c will be used to refer to a non-specified class label value, either equal to one or zero. Typically, the class labels are not known for every node, and the set of nodes (persons) with unknown labels is denoted as $\boldsymbol{V_U}$, whereas $\boldsymbol{V_K}$ is the set of nodes with known labels. The aim of a classification method is to predict the unknown labels. In general, a classification technique M maps to each attribute vector $\boldsymbol{x}$ either a class label c or produces a continuous score $s(x)$. Typically, a higher value of the score indicates a higher probability, e.g. to purchase a certain product, and the continuous score function $s(x)$ can be mapped to class labels by setting a cut-off value.

Traditional (i.e. non-relational) data mining techniques do not take into account information contained in the social network structure. In order to incorporate network effects, Macskassy and Provost (2007) introduced a framework for classification in networked data. In this node-centric framework, a full relational classifier comprises a local (non-relational) classifier M_L, a pure relational or network classifier M_R, and a collective inference (CI) procedure. Local classifiers, which solely make use of the attributes that are related to the entity that is to be classified, have been extensively studied by the data mining community. Typical examples of local classifiers are decision tree classifiers, neural networks, support vector machines, statistical classifiers such as logistic regression or Bayesian based methods, and many others.

In this paper, the emphasis lies on the network based models which have been subject of few research so far, i.e. the pure relational classifier and the collective inference procedure. A pure relational classifier, uses information in a node's neighborhood to predict its label or score. However, it is clear that, when two connected nodes have unknown labels, a pure relational classifier will experience problems to predict unknown labels. Collective inference (CI) procedures are developed to assign a label or score to all unknown nodes, by employing an iterative scheme. The relational classifiers and CI procedures applied in this study are briefly discussed in Section 2.1 and Section 2.2 respectively. For detailed information and formulas, one may refer to the paper of Macskassy and Provost (2007).

2.1 Relational Classifiers

In this study, four relational classifiers are applied. The **Weighted-Vote Relational Neighbor classifier (wvrn)** estimates the probability of a customer to buy a product online as a function of the probabilities of his neighbors. More specifically, it will calculate the weighted sum of the neighbor's probabilities, with the linkage strengths w_{ij} serving as weights. This sum is then normalized in order to yield probabilities between zero and one.

The **Class-Distribution Relational Neighbor classifier (cdrn)** learns a model based on the distribution of neighbor class labels, instead of simply using the weighted sum of neighbor class labels as the wvrn classifier does. The class vector $CV(v_i)$ of a node v_i is defined as the vector of summed linkage weights w_{ij} to the various classes. The reference vector $RV(c)$ of class c on the other hand is defined as the average of the class vectors for nodes known to be of class c. Then the probability for a customer to have a label equal to one can be calculated as the normalized vector similarity between v_i's class vector and the class reference vector of class one. Note that any vector similarity measure can be used, such as for instance L_1 or L_2 matrix norms, cosine similarity, etc., normalized to lie in the range $[0;1]$. In the experimental section of this study cosine similarity will be applied.

The **Network-only Link Based classifier (nlb)** applies logistic regression on feature vectors which are constructed for each node by aggregating the labels of neighboring nodes. Various methods exist to aggregate the labels of the neighboring nodes, e.g. existence (binary), the mode, and value counts. The count model is equivalent to the class vector $CV(v_i)$ as used by the class-distribution relational neighbor classifier, and has been shown to perform best by Lu and Getoor (2003). The count vectors are normalized, since this results in better performance, as indicated by Macskassy and Provost (2007). Then, a logistic regression model is fitted to the class vectors of the nodes in V_K, and used to predict the labels of the nodes in V_U.

The **Spreading Activation Relational Classifier (spaRC)** is based on the Spreading Activation technique (SPA) proposed by Dasgupta et al. (2008) and models the propagation of a certain action undertaken by a customer (which in this paper is equivalent to purchasing a product online) through a network as a diffusion process. The SPA technique of Dasgupta can be decomposed in a relational classifier and a collective inference procedure, and therefore incorporated within the modular framework of Macskassy and Provost (2007). The relational classifier part boils down to the energy spreading step, in which activation energy is transferred from the node's neighbors. The amount of energy transferred is proportional to the link strength.

It is noteworthy to mention that the wvrn and spaRC classifier explicitly assume the existence of *homophily*, meaning that linked entities have a propensity to belong to the same class. Furthermore, all relational classifiers discussed above implicitly make a *Markov assumption* by only including the neighbors of order one.

2.2 Collective Inference Procedures

A relational classifier cannot assign labels to all nodes in one single step, since two connected nodes may both have unknown labels. Therefore, collective inference procedures are needed in combination with the relational classifiers discussed above. The four collective inference procedures applied in this study are discussed below.

Gibbs Sampling (GS) (Geman and Geman, 1984) initializes the unknown nodes in $\boldsymbol{V_U}$ using a local classifier model M_L on the data in $\boldsymbol{D}$. A node is initialized by sampling a label value l_i based on the predicted probabilities. If no local classifier is used, the unknown nodes are left unknown. In a next step, an iterative procedure is followed, whereby for each unknown node a score (probability) is predicted based on a relational classifier M_R. Hereafter, a label value is again sampled, based on this new probability. Note that the relational classifier only sees hard labels, i.e. zeros or ones, due to the sampling process. Labels in the neighborhood of the node under consideration which are unknown in a particular iteration step are ignored. During the first 200 iterations, called the burnin period, no statistics are kept. Afterwards, during 2000 iterations, the labels assigned by the sampling process are counted. Normalizing these counts yields final class probability estimates.

Iterative Classification (IC) is based on a procedure described by Lu and Getoor (2003). As with Gibbs sampling, the unknown nodes are initialized based on a local classifier M_L. Then, a relational classifier M_R is applied to all nodes in $\boldsymbol{V_U}$, whereby entities which have not yet been classified are ignored. The label assigned to each node is the one with the highest probability, based on the estimate of the relational classifier. This is repeated for 1000 times, and the estimates from the last iteration are used as the final class probability estimates.

Relaxation Labeling (RL), based on a method introduced by Chakrabarti et al. (1998), adopts the following procedure (Macskassy and Provost, 2007). Again, the nodes element from $\boldsymbol{V_U}$ are assigned a probability by a local model M_L, as in the two previous CI procedures. With relaxation labeling however, no hard labels, i.e. one or zero, are generated based on this probability. Instead, soft probabilities are used as input for the relational classifier M_R, which is iterated 1000 times. The estimates from the last iteration again serve as the final class probability estimates.

The **Spreading Activation Collective Inference (spaCI)** is the CI counterpart of spaRC (both are the result from the decomposition of SPA into a relational classifier and a collective inference procedure). The unknown labels are initialized with a local model, after which the relational model is applied to the probabilities. As all other CI procedures, spaCI applies a relational learner in each iteration, using the result of the previous iteration as input. However, whereas Gibbs sampling, relaxation labeling, and iterative classification consist of a specified number of iterations, the spaCI procedure ends when a stopping criterion is met. The stopping criterion consists of the following conditions. The procedure ends when (1) the set of active nodes is not extended, and (2) the amount of energy that is spread, i.e. the change in the assigned labels, is smaller than a predefined amount. When combining spaCI with spaRC, the stopping criterion will be met since the amount of energy that is passed is reduced in each iteration. However, when combining spaCI with any of the other relational classifiers defined above, it is not guaranteed that the assigned probabilities will converge, and that the stopping criterion will be met. Therefore, a third stopping criterion is added, i.e. (3) the number of iterations is larger than a predefined number T_{max}.

3 Data Collection and Preprocessing

In order to analyze the influence of a person's social network on his e-commerce adoption, two surveys have been conducted in April 2011. A first survey was meant to gather information about the social network of the respondents. The survey was distributed to university-level management students from the third bachelor. Students were asked to list their closest friends (up to seven) at the school. They also had to grade the intensity of the relation (A: we are together most of the day, B: we meet once a day for a short talk, C: we meet a few times every week, D: we meet once a week). This survey yields a social network indicating whether there is a link between two individuals, and how strong this link is. The second survey concerned e-commerce adoption. A subset of the students involved in the first survey was explained the role of the internet in doing business and were introduced to the survey. As homework, students had to fill out the survey themselves. They also had to ask both their parents to fill out the survey. In order to motivate students to collect data rigorously (and of course to have them learn about the issue), they received the additional task to write a paper, formulating findings based on the data they collected.

Table 1. Characteristics of the data set and social network

	Product 2 (Books)	**Product 4 (Computers)**
# Nodes	681	681
# Edges (undirected)	1102	1102
# Known labels	435 (64%)	433 (64%)
# Zero labels	214 (31%)	222 (33%)
# One labels	221 (33%)	211 (31%)
# Unknown labels	246 (36%)	248 (36%)
Prior one (only known labels)	49.2%	51.3%
Prior zero (only known labels)	50.8%	48.7%

More specifically, the second survey asked respondents (including students) to answer questions related to three steps in the acquisition of a product: 1) product information gathering; 2) supplier selection; and 3) purchasing medium, for different products. This paper only examines the last step: how the product would be purchased. Twelve products were covered in the survey, but this paper only examines two products: books (product 2) and computers (product 4); future research will examine the ten other products. The respondents were asked to provide information about how they purchased each of the twelve products, or how they would purchase if they did not purchase it before. No distinction was made in the tests between an actual purchase and the reflection of how a customer intends to make a purchase, following the reasoning of Jackson et al. (1997).

Table 2. Description of the variables used in the logistic regression (the last variable is the target variable)

Variable Description	Type	Range
Age	Continuous	19-73
Gender	Categorical	Male / Female
Education	Categorical	Primary / Secondary / Polytechnic / University
City Size	Categorical	Big / Medium / Small
Nr of persons in family	Integer	0-9
Intensity of Internet use	Continuous	0-10
Intensity of e-mail use	Continuous	0-10
Nationality	Categorical	French / Indian / Other
Occupation	Categorical	Student / Working / Unemployed
PC use at school / work	Categorical	Constantly / Sometimes / Never / na
Level of Internet access at school / work	Categorical	Full / Restricted / LittleOrNone / na
Internet access at school / work since	Continuous	0-19 years
I like working with the PC	Continuous	0-10
If working: company size	Categorical	Big / Medium / Small / na
If I would buy online, I would only buy on site in my own language	Continuous	0-10
Response concerns a purchase I really did	Categorical	Yes / No
Purchasing online is considered appropriate	Binary	0 / 1

For each product, the respondent was asked about his opinion on the appropriateness of some medium for several tasks in the purchasing process. For each product, the respondent had to give a value of one to five to reveal his opinion about the information medium that could be used to achieve the task (e.g., using a website to make the purchase). The value '1' was used to denote a medium that was considered 'very appropriate'; the value '2' indicated 'appropriate, but less important', etc. Since this paper cannot report on all data, only the way the purchase was made is examined, and more specifically whether online media were considered appropriate. Therefore, for each respondent the minimum was taken of his answers on two questions: 'it is appropriate to buy this product via the website of the seller' and 'it is appropriate to buy this product via another website (e.g. E-bay)'. For the test reported below, we recoded the value as '1' (i.e. the medium was considered appropriate) in case the minimum was 1 or 2, whereas a minimum of 3, 4 or 5 was coded as '0' (i.e. the medium was considered inappropriate).

Whereas the first survey uncovered the social network, the second survey provided information about the target variable (also called label), i.e. whether it is considered appropriate by the respondent to purchase the product online, along with other variables. The information about the social network and the labels is summarized in Table 1. The network contains 681 persons and there are 1102 social ties. The number of people with known labels is respectively 435 (64%) and 433 (64%) for the products books and computers. For the people with known labels, further information from the survey is available, displayed in Table 2. These variables would typically be used by a traditional data mining approach (by a local model, as defined in Section 2), whereas the social network information is used by the relational classifiers.

4 Experimental Setup

The aim of this study is to examine whether social network data contains useful information in the context of e-commerce adoption. In order to carry out the analysis, a data set on e-commerce adoption and the corresponding social network information was gathered, as discussed in Section 3. The network classifiers, outlined in Section 2, are then used to predict e-commerce adoption. It is important to mention that the network based classifiers employed in this study only consist of a relational classifier M_R and a collective inference procedure, and do not incorporate a local model. The four relational classifiers introduced in Section 2.1 have been combined with the four CI procedures from Section 2.2, leading to 16 networked classification techniques, which only take into account the network structure and the known labels. Since the main goal is to examine whether network effects are present in e-commerce adoption, no local model is being used to initialize the labels. However, logistic regression (LR), a local classifier, has been included in the study as a benchmark for the networked learners. The networked learners have been implemented in Matlab, whereas for logistic regression the Weka toolbox was used.

To gauge the strength of classification techniques, a variety of performance measures has been proposed (Ali and Smith, 2006). A very commonly used tool in performance measurement of classifiers is receiver operating characteristic (ROC) analysis (Fawcett, 2006). Typically, a classifier assigns a score $s(x)$ to each instance, based on the attribute vector. Classification is then based on the score by defining a cutoff t, whereby instances with scores higher (lower) than t are classified as cases (non-cases). A ROC curve shows the fraction of the correctly classified cases (i.e. instances with label one), versus one minus the fraction of the correctly classified non-cases (i.e. instances with label zero), for a varying cutoff. A classifier which ROC curve lies above the ROC curve of a second classifier is superior, and the point $(0; 1)$ corresponds to perfect classification. Although ROC curves are a powerful tool for comparing classifiers, practitioners prefer a single number indicating the performance over the visual comparison of ROC curves. Therefore, the area under the ROC curve (AUC) has been widely adopted in the data mining community for its simplicity of use. Moreover, the AUC has an interesting statistical interpretation in the sense that it is *the probability that a randomly chosen positive instance will be ranked higher than a randomly chosen negative instance* (Fawcett, 2006). In other words, if the AUC of a classifier is larger than 0.5, then the classifier is able to draw meaningful conclusions with regard to the target variable.

Even though the AUC has been extensively used, it was recently pointed out by Hand that in a business context, i.e. a profit oriented environment, the AUC suffers from the fact that it does not correctly account for the misclassification costs (for details, see (Hand, 2009)). In order to deal with this issue, Hand developed a performance measure, the H-measure, which is based on the expected misclassification loss. Hereby, the misclassification costs are not exactly known but follow a probability distribution, more specifically a beta distribution with parameters α and β. The H-measure is a normalized metric, ranging from zero for a random classifier to one for a perfect classifier.

For classification performance measurement in this study, the data set has been split into a test set and a training set. Two scenarios have been experimented with: a test set consisting of 10% and 30%, drawn as a stratified random sample from the set of students with known labels. By assuming the labels in the test set to be unknown, the percentage of customers (nodes) with unknown labels in the network increases from 35% to respectively 42% and 55%. The reason why the instances with unknown labels, other than those in the test set, have not been removed from the data set is because they may contain relevant information on the social network structure. The above procedure to obtain a random test set and measure classification performance has been carried out five times in order to obtain stable results. Then, the classification performance on the test set is assessed with both measures. The AUC is reported because of its statistical interpretation, whereas the H-measure is included since it correctly measures the potential profits arising from the predictive ability of the model. The H-measure is calculated with default values for the distribution parameters, i.e. $\alpha = 2$ and $\beta = 2$, implying symmetric misclassification costs (i.e. misclassifying a case is as costly as misclassifying a non-case).

5 Discussion of Results

The results of the simulations are shown in Table 3, with the left panel displaying the AUC for each classification method, and the right panel the H-measure. The results pertain to two products (books and computers) and one target variable (i.e. whether the respondent would intend to buy the product online), for which some trends can be noticed. First of all, the network classifiers (the top 16 classifiers in Table 3) almost invariably outperform logistic regression, a local classifier. It is interesting to note that the performance of the network based techniques hardly seems to fluctuate among the two products, whereas the performance of the logistic regression is very dependent upon the product. Among the relational classifiers, wvrn and spaRC generally are in the top performing techniques, for the collective inference procedures there is no specific method outperforming the others.

The network based classifiers also outperform logistic regression in terms of the H-measure, illustrating the potential of social network information for achieving better predictive performance. Furthermore, as one would expect, when the amount of nodes with unknown labels increases, the predictive performance tends to decrease, although it still performs better than a random classifier. This is the case for the product books, where more information invariably leads to better prediction, which is an incentive for companies to collect as much data as possible. However, this is not true for computers, since for this product more data leads to better predictive performance only 9 times out of 16. The reason why this is observed in the data is not

Table 3. Results of the full network learners and logistic regression (LR) as a local model benchmark. The reported performance metric is the AUC in the left panel and the H-measure in the right panel, for two sizes of the test set, i.e. 10% and 30%.

	Books		**Computers**		**Books**		**Computers**	
	10% (42%)	30% (55%)	10% (42%)	30% (55%)	10% (42%)	30% (55%)	10% (42%)	30% (55%)
IC-wvrn	0.630	0.576	0.594	0.588	0.115	0.053	0.083	0.084
IC-cdrn	0.606	0.552	0.575	0.524	0.103	0.032	0.066	0.015
IC-nlb	0.609	0.526	0.609	0.587	0.082	0.012	0.088	0.065
IC-spaRC	0.619	0.565	0.594	0.591	0.094	0.043	0.082	0.079
RL-wvrn	0.623	0.583	0.590	0.605	0.108	0.049	0.090	0.122
RL-cdrn	0.615	0.555	0.578	0.546	0.113	0.036	0.091	0.067
RL-nlb	0.609	0.524	0.609	0.588	0.082	0.010	0.088	0.069
RL-spaRC	0.623	0.583	0.590	0.605	0.108	0.049	0.090	0.122
spaCI-wvrn	0.623	0.583	0.590	0.605	0.108	0.049	0.090	0.122
spaCI-cdrn	0.615	0.555	0.578	0.546	0.113	0.036	0.091	0.067
spaCI-nlb	0.609	0.526	0.614	0.587	0.082	0.010	0.093	0.068
spaCI-spaRC	0.623	0.583	0.590	0.605	0.108	0.049	0.089	0.122
GS-wvrn	0.625	0.582	0.592	0.609	0.108	0.047	0.092	0.120
GS-cdrn	0.593	0.535	0.578	0.565	0.079	0.032	0.074	0.040
GS-nlb	0.611	0.524	0.620	0.589	0.082	0.009	0.094	0.064
GS-spaRC	0.622	0.584	0.591	0.606	0.103	0.050	0.092	0.122
LR	0.580	0.539	0.417	0.507	0.057	0.016	0.002	0.015

Entirely clear, but it suggests that the network effects are more pronounced for books than for computers, since the addition of new data consistently improves classification performance, whereas this consistency is not observed for computers. The weaker network effect for computers is possibly caused by differences in computer literacy, computer requirements, requested after-sales-service, etc. which increase the burden to purchase this product online.

Secondly, since the AUC is (considerably) higher than 0.50, the results suggest there are social influences in e-commerce acceptance, especially since no local model was used to initialize the labels (meaning that no traditional variables were taken into account by the networked learners). Also note that wvrn and spaRC typically perform best. This has an interesting implication, since both wvrn and spaRC explicitly assume the existence of *homophily*. The fact that especially these two relational classifiers outperform the others, indicates that there is a strong resemblance between neighbors in terms of e-commerce adoption. It is less obvious to determine whether this resemblance is a result of social influence, or whether the resemblance just shows up because of the fact that people with similar behavior tend to have more contact. In any case, it opens perspectives for direct marketing applications in an e-commerce context.

Although the respondents were not explicitly questioned about social influences, detecting signals of social influence is in contrast to what would be expected based on the literature on acceptance of systems which are voluntarily used. Prior research on social influence can be considered to study the perceived social influence and would suggest there is no perceived social influence in the adoption of e-commerce. The results of this study then suggest something such as latent social influence, of which the respondent is less aware. Our future research will examine whether the social influence construct in the classic UTAUT should be converted in a combination of a latent social influence construct and a perceived social influence construct. Similarly, it will be examined whether adding the social network as an antecedent of the behavioral intention to use a system in the UTAUT significantly increases the R^2 of the UTAUT. It has often been suggested, based on intuition, that friends would show similar e-commerce behavior. However, this paper is, as far as we are aware, the first to empirically justify this. The fact that friends behave alike when it comes to the acceptance of e-commerce is important for practitioners. Studying social networks of customers (e.g. via Facebook profiles) is valuable to identify potential customers. While marketing campaigns in the past were aimed at specific profiles (people of some age, gender, etc.), companies can now determine the exact names of potential customers. Importantly, these results show that knowledge about the social network is a better predictor of e-commerce behavior than classic factors that describe customer segments, such as age and gender. When setting up a Facebook Adds campaign, companies can choose to only show the add to friends of fans for example. This paper reinforces the idea that such practices really work.

6 Conclusion and Future Research

This paper examines the impact of social network information on the prediction of e-commerce adoption. Even though the number of products and target variables analyzed is limited, some very interesting trends arise. In contrast to what previous research suggests, there seem to be social network effects apparent in e-commerce adoption. Moreover, network based classification techniques are able to outperform traditional models, such as logistic regression.

These results provide some interesting tracks for further research. First of all, more data from the survey is available and will be analyzed. There is information on ten more products and on more target variables (such as the intent to look up product information online and the intent to search the Internet for a supplier). Extending the analysis to these products, and to more randomly drawn test sets, will allow to statistically underpin the indications found in this study. Besides, the networked classification techniques were benchmarked against logistic regression, which is a commonly applied technique. However, more advanced local models exist (e.g. neural networks, support vector machines, etc.), which could be included in a future study. Furthermore, the classification models in this study exclusively use either social network data or traditional data. Future research should examine whether the performance of an e-commerce adoption prediction model can be improved by combining traditional information and the social network structure in a hybrid classification technique. Finally, while prior research showed that the intent to use a system is an antecedent of real system use (Venkatesh et al., 2003), it might be interesting in future research to distinguish between both.

References

1. Ali, S., Smith, K.A.: On learning algorithm selection for classification. Applied Soft Computing 6(2), 119–138 (2006)
2. Anderson, P.F., Chambers, T.M.: A Reward/Measurement model of organizational buying behavior. Journal of Marketing 49, 7–23 (1985)
3. Blau, P.M.: Inequality and Heterogeneity: A Primitive Theory of Social Structure. Free Press, New York (1977)
4. Chakrabarti, S., Dom, B., Indyk, P.: Enhanced hypertext categorization using hyperlinks. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 307–319 (1998)
5. Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A.A., Joshi, A.: Social ties and their relevance to churn in mobile telecom networks. In: Proceedings of the 11th International Conference on Extending Database Technology: Advances in Database Technology, EDBT 2008, pp. 697–711 (2008)
6. Davis, F.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information. MIS Quarterly 3(13), 319–339 (1989)
7. Dillion, A., Morris, M.G.: User acceptance of information technology: Theories and Models. Annual Review of Information Science and Technology, 3–32 (1996)
8. Engel, J.F., Blackwell, R.D., Miniard, P.W.: Consumer Behavior, 8th edn. The Dryden Press, Chicago (1995)
9. Fawcett, T.: An introduction to ROC analysis. Pattern Recognition Letters 27(8), 861–874 (2006)
10. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741 (1984)

11. Hand, D.J.: Measuring classifier performance: a coherent alternative to the area under the ROC curve. Machine Learning 77(1), 103–123 (2009)
12. Hansen, T.: Perspectives on consumer decision making: an integrated approach. Journal of Consumer Behaviour 4(6), 420–437 (2005)
13. Jackson, C.M., Chow, S., Leitch, R.A.: Towards an understanding of the behavioral intention to use an information system. Decision Sciences 28(2), 357–389 (1997)
14. Jensen, D., Neville, J.: Linkage and autocorrelation cause feature selection bias in relational learning. In: Proceedings of the 19th International Conference on Machine Learning, pp. 259–266 (2002)
15. Karahanna, E., Straub, D.W.: The psychological origins of perceived usefulness and ease of use. Information and Management 35(4), 237–250 (1999)
16. Lu, Q., Getoor, L.: Link-based classification. In: Proceedings of the 20th International Conference on Machine Learning (ICML), pp. 496–503 (2003)
17. Macskassy, S.A., Provost, F.: Classification in networked data. Journal of Machine Learning Research 8, 935–983 (2007)
18. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. Annual Review of Sociology 27, 415–444 (2001)
19. Newman, M.E.J.: Networks: An Introduction. Oxford University Press, Oxford (2010)
20. Taylor, S., Todd, P.A.: Understanding information technology usage: a test of competing models. Information Systems Research 6(4), 144–176 (1995)
21. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 27(3), 425–478 (2003)
22. Warshaw, P., Davis, F.: Disentangling Behavioral Intention and Behavioral Expectation. Journal of Experimental Social Psychology 21(3), 213–228 (1985)

Exploring Innovation in the Context of Employee Relationship and IT-Enabled Knowledge Sharing

Jianping Peng[1], Guoying Zhang[2], Zhengping Fu[1], and Yong Tan[3]

[1] Sun Yat-Sen Business School, Sun Yat-Sen University
{mnspjp,mnsfzp}@mail.sysu.edu.cn
[2] Midwestern State University
grace.zhang@mwsu.edu
[3] Foster School of Business, University of Washington
ytan@uw.edu

Abstract. Innovation is one of the critical success factors for organizations. It is essential for business to understand the driving factors of innovation. This study investigates impacts of the following two aspects on innovation: employee relationship and knowledge sharing. 167 samples of firm level data were collected to construct the measurements of innovation, intensity of employee relationship, employee diversity, quality of knowledge sharing, and IT application maturity. It is found that all of these factors have significant impacts on innovation. Furthermore, IT application maturity has a significant moderating effect of enabling knowledge sharing to improve innovation. In order to further refine the characteristics of employee relationship, an individual level study was conducted to construct three consolidated employee social networks. It is shown that centrality of the information social network, which can be perceived as relationship characteristics of an individual employee in the network, is positively related to performance. With analysis at both organizational and individual level, this study empirically illustrates the importance of exploring innovation in the context of employee relationship and IT-enabled knowledge sharing.

Keywords: intensity of employee relationship, employee diversity, quality of knowledge, innovation, social network, contextual performance.

1 Introduction

Innovation is one of the critical success factors for organizations, often referred as "the intentional introduction and application . . . of ideas, processes, products or procedures, new to the relevant unit of adoption, designed to significantly benefit [1]." Because innovation involves new ideas and technologies, it has been used to incorporate creativity, in addition to the implementation process [2]. As Porter (1998) pointed out, innovation can improve product quality and business operation, which are essential for gaining competitive advantages [3]. Innovative organizations often continuously seek for better ways to realize business strategies utilizing cutting edge technologies and new methodologies. Therefore, the ways to improve organization's capability of innovation always draw attention from both researchers and

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 22–36, 2012.

practitioners. This study complements the literature by introducing intensity of employee relationship, employee diversity, and quality of knowledge sharing into the framework of investigating innovation.

Employee relationship in organizations can be investigated from either subjective perception or the characteristics of employee social networks. Relationship among employees evolves in a traditional format of interactions as well as in a digital format of communications. These interactions and communications at both individual and organizational levels are often embedded in a social network structure, influenced by the characteristics of network, and consequently, affect the evolution of network. Here in this study, the approach of subjective perception on the intensity of employee relationship is adopted to construct the measurements at organizational level [4]. On the other hand, the approach of social network characteristics is applied to represent employee relationship at individual level. Prior studies often emphasized on examining whether strong or weak ties of a network are more important to improve innovation and performance [5,6,7]. We use network centrality to characterize an employee's position in the social network.

Organization needs employee diversity to avoid the pitfalls of group thinking and resistance to changes in innovation processes. New cognitive skills, different personalities, and new interactions among employees can be critical to ignite innovation. Dalton and Todor (1979) started the study on this topic and show that team diversity can enhance team flexibility and adaptation [8]. Paulus (2000) agreed that diversity in a team can facilitate generation of new ideas and innovative problem solving [9]. Gilson (2001) showed that individual creativity is higher in a team with a higher diversity composition; furthermore, organizations with highly diversified employees are more effective in knowledge sharing [10].

Beyond explaining innovation from the perspective of employee relationship and employee diversity, it is essential to note that knowledge sharing and disseminating also serve as incubators for most innovations. In fact, innovation is often rooted from an extensive range of knowledge sharing which is helpful to implement new ideas, processes, products, and services. Extant studies verified the direct linkage between knowledge sharing and innovation [11,12,13]. However, with information overload, it becomes a big challenge for employees to be able to identify relevant knowledge. In other words, the sharing quality matters. Yang and Long (2008) proposed a framework to evaluate quality of knowledge sharing and validated that good quality of shared knowledge can greatly motivate employees' contribution to future knowledge sharing [14]. With regards to the implementation of knowledge sharing, it is noticeable that modern organizations exert significant efforts on information technology to implement core business functions, including knowledge sharing. It was observed that organization performance is highly dependent on IT application maturity [15]. Hence, in this study, while incorporating quality of knowledge sharing into the investigation of innovation, the platform for knowledge sharing – IT applications – is also considered.

We collected 167 samples of firm level data based on the constructs used in the literature. The data were processed to generate variable measurements of: innovation, intensity of employee relationship, employee diversity, quality of knowledge sharing, and IT application maturity. Empirical models were then established to investigate the research questions. It is found that intensity of employee relationship, employee diversity, and quality of knowledge sharing all have positive impacts on innovation, while IT application maturity significantly moderates the impact of quality of knowledge sharing. The findings complement the literature on employee relationship theory and organization resource theory, and provide managerial insights to organizations on how to improve innovation capability.

Prior studies also analyzed individual level creativity. We include here an extension to represent employee relationship characteristics using social network analysis at individual level. An individual level study was conducted to construct three consolidated employee social networks. We find that the centrality of the information social network is positively significant in determining employee's contextual performance. Innovation at organization level has to be eventually realized by individual employees, therefore, the validation on the importance of employee relationship at individual level can be aggregated throughout the organization.

2 Theoretical Background

Innovation capability is determined by a myriad of factors in an organization, which include individual creativity, organization learning capability, culture of knowledge sharing among employees, employee relationship, among others. However, not until the late 80's, there was no agreed upon theory on the nature of organization innovation capability, although how to improve team working has been emphasized along the way [17]. As discussed in [18, 19], three factors are identified for creativity, namely, domain-relevant knowledge, creativity-relevant skills, and motivation. This so called "Componential Model" also describes the characteristics of work environment for innovation: organizational motivation to innovate, resources, and management practices. In Ref. [2], the discussion on work environment is combined with individual factors: sense making, motivation, and knowledge and ability. These factors lead to creative actions when interacting with other contextual variables. West and Farr (1990)'s model of team climate for innovation includes four factors: vision, participative safety, task orientation, and support for innovation [1].

Intense interactions among employees usually indicate a dense network of employee relationship, which can have important impacts on business performances. Lots of prior studies endorsed the influence of intense employee relationship on innovation. Vander and Elfring (2002) proposed a mechanism of investigating the effect of relationship network structure on knowledge disseminating [20]. Meagher and Rogers (2004) studied the relationship network features and see their impacts on integrating innovation capabilities [21]. Wang *et al.* (2009) also employs an empirical study to validate that the internal strength of network ties has a positive impact on team creativity [4]. Chen and Guan (2009) established a network among patent innovators, and find that a strong small world characteristic will lead to more innovative outputs [22]. Granovetter (1992) believed that weak link as an information bridge is important to organizations [5]. Burt (1992), however, emphasized more on structure holes in the network, and found no relationship between the strength of relationship and the amount of network resources [6]. Other scholars studied the components of embedded relationship and how embeddness shape organizational outcomes [23]. Bian (1997) utilized the resource flow in the network to demonstrate the importance of weak ties in terms of information distributing, and proposed that strong ties based on trust and obligation have more advantages [24]. Hansen (1999) found out that weak ties help a project team search for useful information, but impede the transfer of complex knowledge, which tends to require a strong tie in between [25]. In Ref. [26] and [27], authors presented similar results in explaining the role of weak ties. Bian and Zhan (2001) also noticed cultural difference in interpreting the role of weak and strong ties, and argue that strong ties can be more effective in job seeking in China [7]. In this study, social networks are not constructed at the firm level analysis, and the intensity of employee relationship is measured directly through respondents' opinions. Instead, at the individual level analysis, three consolidated social networks are constructed and the centrality of each social network is calculated for the empirical model.

With regards to employee diversity, author of Ref. [10] showed that individual creativity is higher in a team with a higher diversity composition, and organization with high diversified employees is more effective in knowledge sharing and disseminating. New cognitive skill, different personality, and new interactions among employees can improve the team creativity. Dalton and Todor (1979) showed that team diversity can enhance team flexibility and adaptation [8]. With diversity, team can take advantage of versatile skill sets and perceptions from various angles. Paulus (2000) agreed that diversity in the team can facilitate the generation of new ideas and improve capability of innovative problem solving [9].

Based on the above discussion, two important factors need to be incorporated in our study toward innovation capability: intensity of employee relationship and employee diversity.

Hypothesis 1: *Intensity of employee relationship and employee diversity have positive impacts on innovation.*

Innovation is often rooted from an extensive range of knowledge sharing to implement new idea, process, product, and service. According to Ref. [11], knowledge sharing among individuals can help learning at both individual and group levels, and learning capability is the foundation of innovation. Ipe (2003) also agreed that interactions among individuals with different domain knowledge can boost innovation much more significant than pure individual efforts [12]. Lin (2007) found a high correlation between voluntary knowledge sharing and organization innovation capability [13]. Further, in the information age, with the information overloading, it becomes a big challenge for employees to be able to share and identify relevant knowledge. In other words, quality of knowledge sharing matters. Yang and Long (2008) proposed a framework to evaluate the quality of knowledge sharing and find that good quality of shared knowledge is important to motivate employees' contribution to knowledge sharing [14]. In general, the higher the quality of knowledge sharing, the more efficient the utilization of knowledge, and consequently, more expectation can be placed on innovation. Therefore, we propose the following hypothesis:

Hypothesis 2: *The quality of knowledge sharing has a positive impact on innovation.*

Finally, most organizations use information technology to facilitate business operations. IT application platform is an essential enabler to implement knowledge sharing among employees. The higher the quality of knowledge sharing, the higher the chance of success in innovation is [28]. IT application is also important to implement knowledge management practice. Hansen (1999) emphasized the importance of IT application platform, which can code and organize a huge amount of documents and information generated during business operations, provide indexing on these documents and information, and store them into database for further retrieval [25]. Similarly, Xia and Cai (2001) pointed out that an essential condition for knowledge sharing is to establish an IT platform for knowledge acquisition, processing, storing, sharing, and reusing, and provide safeguards for knowledge base [29]. Van den Hooff and De Ridder (2004) further studied the impact of computer-mediated-communication on organization commitment, which in turn can influence knowledge sharing and innovation [30]. Based on these studies, the following hypothesis is proposed:

Hypothesis 3: *IT application maturity has a significant moderation effect on the quality of knowledge sharing.*

3 Methodology

3.1 Research Design

Before the measurement constructs are introduced, it is necessary to first define the research objectives discussed above: innovation, intensity of employee relationship, employee diversity, quality of knowledge sharing, and IT application maturity. Established constructs were adopted for this study in order to measure these variables. First, innovation was measured based on the framework proposed by Ref. [4], which is a modification from the model in Ref. [22]. In Ref. [4], the author also provided constructs for intensity of employee relationship and employee diversity. Second, quality of knowledge sharing can be established by the instrument in Ref. [14]. Third, the measurement construct for IT application maturity was adopted from that in Refs. [31] and [32].

Combining the above mentioned measurement instruments (except for IT application maturity), a survey study was first designed and conducted. The questionnaires were distributed to selected MBA and MPA students in authors' university during March 2010 to May 2011. In order to collect more reliable firm level information, these selected respondents were required to have at least three-year working experience as a supervising manager in a company. A total of 195 questionnaires were distributed and 167 effective responses were collected.

Based on the effective responses, about half of them represent companies with state-owned capital. 59 State-owned-enterprises account for 46.5% of the total sample, whilst 21.3% are 27 private companies and 32.3% are 41 joint-venture companies. In terms of industry type, 53.9% of companies are in manufacturing and production and 46.1% in service industry. Major businesses of these companies range from electronics, telecommunications, mechanics, to home appliances, finance, etc. With regards to the number of total employees, 23 companies have less than 100 employees, 27 companies with 100 to 500 employees, 53 with 500 to 1000 employees, 44 with 1000 to 10000 employees, and 20 companies with more than 10000 employees.

3.2 Data Processing

Data collected were processed for reliability and validity analysis. After eliminating the items that cannot fulfill requirement of reliability and validity test, we consolidated items into the following factor variables: innovation, quality of knowledge sharing, intensity of employee relationship, employee diversity. Results are presented in Table 1:

Table 1. Factor loading and reliability scores of first questionnaire items

Item	Innovation	Quality of knowledge sharing	Intensity of employee relationship	Employee diversity		Reliability – Alpha
Originality	0.786	-	-	-	-	
New ideas	0.841	-	-	-	-	
Creative problem solving	0.822	-	-	-	-	
Methods of accomplish tasks	0.802	-	-	-	-	
Learning ability	0.696	-	-	-	-	0.920
Methods of problem solving	0.828	-	-	-	-	
Perspective of defining problems	0.822	-	-	-	-	
Information integration	0.818	-	-	-	-	
In time sharing	-	0.813	-	-	-	
Reliability of knowledge	-	0.833	-	-	-	
Comprehensive sharing	-	0.857	-	-	-	0.869
Relevance of knowledge	-	0.818	-	-	-	
Accuracy of knowledge	-	0.728	-	-	-	
Colleague interactions	-	-	0.863	-	-	
Colleague communications	-	-	0.918	-	-	0.841
Collegiality	-	-	0.775	-	-	

Table 1. (*continued*)

Age difference	-	-	-	0.830	-	0.731
Title difference	-	-	-	0.832	-	
Education background difference	-	-	-	0.711	-	
Team work experience	-	-	-	-	0.882	0.677
Creative thinking capability	-	-	-	-	0.888	
KMO	0.889	0.840	0.650	0.612		
Eigen value	5.158	3.290	2.188	1.909	1.611	
Variance explained (%)	64.480	65.280	72.940	70.389		

In order to further collect corresponding data in the dimension of IT application maturity, a second set of survey was designed on the basis of the measurement framework proposed in Ref. [32]. The questionnaires were then distributed to those companies giving effective responses in the first survey. Similar analysis of reliability and validity on the items in this data set was conducted. All measurement constructs for the five research objects have good reliability and validity, the derived variables (factors) hence were further used for empirical analysis.

4 Model

In order to test Hypotheses 1, 2, and 3, a regression model was proposed to verify the relationship between innovation and intensity of employee relationship, employee diversity, quality of knowledge sharing, and an interaction term of IT application maturity and quality of knowledge sharing:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 IT + \beta_5 x_3 \times IT + \varepsilon \,. \tag{1}$$

In the above equation, y represents innovation capability and is the dependent variable. The independent variables are x_1 representing intensity of employee relationship, x_2 representing employee diversity, x_3 the quality of knowledge sharing, IT representing IT application maturity, and an interaction term of x_3 and IT to capture the moderating effect.

Using SPSS, the model was estimated and the result is presented in Table 2. Due to multi-colinearity, we also dropped the interaction term, $x_3 \times IT$, and estimated the model without this term. Alternatively, the variable of IT was dropped and the model without this term was estimated. The results are presented in Table 2.

Table 2. Parameter estimation for model (1)

	Model (1)		Model (1) without interaction term		Model (1) without *IT* term	
	Coefficient	VIF	Coefficient	VIF	Coefficient	VIF
β_0	.906	-	.172	-	.816**	-
β_1	.170**	1.281	.175**	1.270	.170**	1.276
β_2	.154**	1.088	.139**	1.023	.151**	1.010
β_3	.109	17.303	.338***	1.599	.135	3.910
β_4	-.032	19.743	.213***	1.424	-	-
β_5	.360	53.501	-	-	.316**	3.858
R^2	.353		.350		.353	
F	17.597***		21.839***		22.128***	
***$p < 0.01$, **$p < 0.05$						

It can be observed that innovation capability improves if intensity of employee relationship, employee diversity, or quality of knowledge sharing increases. In other words, effective knowledge sharing is critical to organizations in improving innovation capability. At the same time, an intense network of employee relationship or a diversified employee profile is also important to stimulate innovation. Therefore, Hypotheses 1 and 2 are supported. Further, it can be seen that IT application maturity has a significant moderating effect on quality of knowledge sharing with regards to innovation. β_5 is positively significant in the revised model. In conclusion, mature IT application utilization does help the implementation of knowledge sharing, and consequently, improve the quality of knowledge sharing, which eventually stimulate organization innovation. Therefore, Hypothesis 3 is also supported.

5 Employee Social Networks

In Section 4, it is shown that intensity of employee relationship and employee diversity have significant positive impacts on innovation from the perspective of organizations. It is also interesting to investigate whether employee relationship matters at an individual level. We chose to use performance as the dependent variable here, because employee performance is the evaluation indicator of an employee's competences and accomplishments, including creativity contributing to work.

In order to investigate the relationship between relationship characteristics and performance at individual level, social networks of employees represented by nodes were constructed. A subsidy unit of a company in the first data set was picked as the sample organization. Eight social networks of employees in this subsidy unit were constructed based on a survey study. The centrality measurements of these networks are obtained based on network structure. Using factor analysis, the centrality measurements were further consolidated to represent the centralities of friendship network, advice network, and information network. They were then combined with a variable representing knowledge sharing to form the empirical model for individual performance analysis.

The company under study is a Fortune 500 company specializing in providing high-end healthcare equipment and healthcare information systems. It is one of the top companies in healthcare equipment manufacturing industry worldwide and has top market share in China. The subsidy branch under study is located in Guangzhou, China. It is composed of sales department, customer service department, and administration department. Overseeing the region of five provinces in southeastern China, this branch is in charge of product sales, after sale installation, after sale maintenance and service. There were more than 50 full-time employees in this branch when we conducted the survey, including department managers, administrative assistants, financial analysts, sales, senior engineers, and technicians. Since all of the employees work together in one business office, intensive interactions among them were observed. Social networks based on this group have clear network boundary because only the full-time employees working for this branch were included.

5.1 Network Construct and Variable Measurement

Krackhardt (1992) proposed that employee social networks can be categorized into three different types: friendship network, advice network, and information network [33]. Based on this network categorization, Luo (2010) established a framework of constructs to establish social networks [34]. A survey study was conducted using Luo's framework. Table 3 is a list of question items used to construct social networks among employees. Social networks built upon each item of B1-B3 were representing friendship networks, networks based each item of B4-B6 representing advice networks, and item B7 or B8 formed information networks capturing information exchange relationship among employees.

Table 3. Sample items in the questionnaire for constructing employee social networks

Item No.	Question
B1	In general, which coworker do you hang out with after work?
B2	If you have personal issues, who do you ask for help?
B3	If you get frustrated from work or get criticism from supervisor, who do you ask?
B4	If you have problems in work, who do you ask for help?
B5	When you need to make important decisions in work, who do you ask for advices and recommendations?
B6	In general, which coworker do you choose to discuss with on issues in work?
B7	Who do you often choose to have casual conversations?
B8	In general, who are the people having email communications with you?

The eight items presented in Table 3 generated 8 social networks, and the centrality of each network was calculated to represent the unique position of each employee in the network. The consolidated centrality measurements were further used in the empirical model as independent variables. To maintain consistency with the organization level analysis, knowledge sharing was also considered. The measurement construct established in Ref. [35] was adopted for knowledge sharing at individual level analysis.

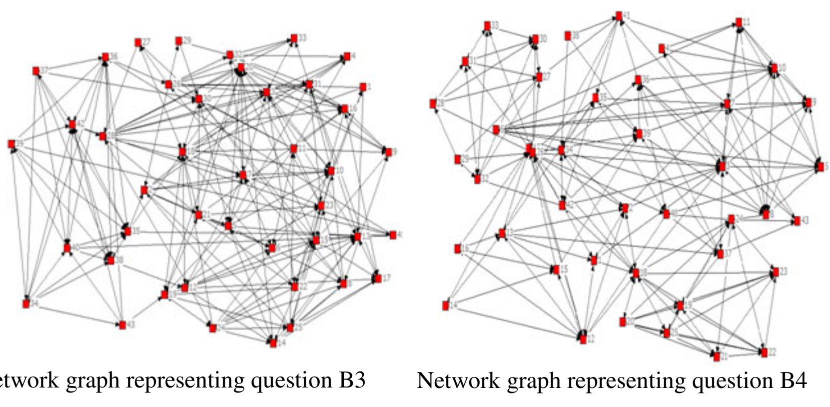
The dependent variable here is individual performance. Borman and Motowidlo (1993) categorized performance into task performance and contextual performance [36]. Contextual performance is defined as activities that contribute to the social and psychological culture of an organization [37]. We used the measurement construct proposed by that in Ref. [38] to generate the variable of contextual performance. This construct has been widely used by other researchers. With regards to task performance, which simply evaluates the completion of assigned tasks, we obtained the performance evaluations from records of Human Resource department of the company.

Although we didn't use creativity for individual level analysis, it is believed that data of individual performance, which are easier to measure, is including creativity characteristics. Bai *et al.* (2008) used a similar approach to take individual performance measurement as the proxy for creativity [39].

5.2 Data Collection and Processing

A survey study on employee social networks was conducted toward employees of Guangzhou branch of the company from December 2010 to January 2011. 54 questionnaires were distributed and 43 effective responses were collected with an effective response rate of 79.6%. Based on the effective answers collected from 8 items listed in Table 3, 8 social networks were constructed using UCINET. The centrality of each network was calculated to represent the unique position of each employee in the network. Figure 1 depicts two sample networks built upon question item B3 and B4.

The data were further processed through reliability and validity analysis. 15 items establish the construct of contextual performance of each employee, and 10 items represent individual knowledge sharing behavior. The centrality measurements calculated for the original 8 networks were consolidated into the three centralities of a friendship network, an advice network, and an information network. With regards to task performance, performance evaluation was obtained from records of Human Resource department of the company. We are now ready to examine the relationship between these network centralities and employee performance.



Network graph representing question B3 Network graph representing question B4

Fig. 1. Two sample social networks

5.3 Model

Now, with all valid constructs in place, the empirical model can be set up to investigate how network centrality, representing employee relationship, influences employee's task and contextual performances:

$$y_i = \beta_0 + \beta_3 x_3 + \sum_{j=1}^{3} \rho_j C_j + \gamma_4 D + \gamma_5 W + \varepsilon_i \,. \tag{2}$$

In the above equation, y, as the dependent variable, represents performance, with subscript i=1 for contextual performance, and i=2 for task performance. Independent variables include x_3 representing knowledge sharing, D representing a control for employee education background, and W representing a control of years of work experiences. Each employee's network centrality is captured through variable C, while subscript j=1 referring to centrality in the friendship network, j=2 the advice network, and j=3 indicating the information network.

Using SPSS, we first tested a preliminary model which includes only the centrality variables from three networks to examine the relationship between network characteristics and performance. Then the full model was examined by incorporating other factors: knowledge sharing, education background, and work experiences. The results are presented in Table 4.

Table 4. Parameter estimation for model (2)

	Preliminary model (2)				Model (2)			
	y_1		y_2		y_1		y_2	
	Coefficient	VIF	Coefficient	VIF	Coefficient	VIF	Coefficient	VIF
β_0	0.000	-	0.000	-	-0.214	-	-0.730	-
ρ_1	0.412***	1.694	0.298**	1.694	0.165	2.332	0.164	2.332
ρ_2	-0.014	2.326	0.060	2.326	-0.011	2.642	-0.013	2.642
ρ_3	0.502***	2.530	0.488***	2.530	0.465***	3.497	0.537***	3.497
β_3	-	-	-	-	0.399***	2.107	0.217	2.107
γ_4	-	-	-	-	0.139	1.217	0.354	1.217
γ_5	-	-	-	-	-0.007	1.318	0.002	1.318
R^2	0.660		0.541		0.757		0.628	
F	25.233***		17.529***		18.731***		10.142***	
***$p < 0.01$, **$p < 0.05$								

In model (2), only the centrality measure of the information network stays positively significant. In other words, a higher centrality of employee in the information network is essential to explain better contextual and task performances. Counter-intuitively, a higher centrality in the advice network doesn't help with performances. This might be explained by the reason that centrality measurement treats incoming and outgoing links equally. In other words, high centrality in the advice network can indicate the employee is a person the others would like to seek for advices, or, at the other extreme, a person that always asks others for advices. In the latter scenario, the employee might not be a prospect of good performance. It is also interesting to note that knowledge sharing is positively significant in improving contextual performance, while it is not playing a role in improving task performance.

Both task and contextual performances were investigated in this section. A higher centrality in the information network often leads to better task and contextual performance. It is conceivable that if an employee has a hub or central position in the information social network, which means other employees often perceive him/her as a reliable source for information, then this information hub person must have a good personality in willingness to help, a sincere care about the organization culture, and a comprehensive skill set of information collection and dissemination. All of these characteristics lead to a higher chance of achieving good performance.

Social network characteristics, representing employee relationship here, have significant impact on employee performance. Although this analysis is on an individual level, it can be argued that social networks influence employee performance, and at an aggregate level, employee performance eventually translates into business performance. Innovation capability, as one of the most important indicators of business performance is thus also influenced by employee relationship characteristics. Here, although the measurement of network characteristics, representing employee relationship, is different from the construct used in the firm level model, the importance of understanding employee relationship is emphasized at both firm and individual levels.

6 Conclusion and Future Directions

This study investigates several factors: intensity of employee relationship, employee diversity, quality of knowledge sharing, in terms of their impacts on organization innovation capability. A survey study was conducted through 2 sets of questionnaires. Mature constructs in prior literature were utilized to measure both dependent and independent variables in the empirical model. The model demonstrates that all of these factors have positive impacts in improving innovations. The result is consistent with the strong tie theory in Ref. [7], and also validates the common assumption that a diversified employee profile facilitates knowledge sharing and disseminating [10], which consequently improves innovation.

Traditional organization resource theory hasn’t taken it into consideration in terms of employee relationship characteristics. In fact, these are important resources for an organization. The formation of an organization implies the formation of employee social networks. These social networks are intangible assets to business, and they can significantly affect organizational culture, innovation, and performance. This study validates the importance of relationship characteristics through demonstrating their impacts on innovation. By incorporating employee relationship into organization resources, it provides an extension to the theory of traditional organization resource.

According to our research results, both the intensity and employee diversity have positive significant impacts on innovation. This result has lots of managerial insights in human resource management and talent retention practice. In fact, how to identify the required education background and work experiences for different job positions has always been challenging for human resource experts. From the perspective of improving innovation capability, a homogenous profile of employees is not advantageous. Also, in order to increase the intensity of employee relationship, organization can consider incorporating more team building and culture building activities.

The quality of knowledge sharing has been a big challenge for organizations. In this study, beyond validating the positive impact of knowledge sharing on innovation, we also consider the implementation platform – IT application. Organizations today cannot survive without information technology. A good utilization of IT application can improve business operations. Therefore, when an IT-based knowledge management system is built, there should be built-in mechanism to encourage knowledge sharing. Organizations differ in the ways of utilizing IT applications [32]; consequently, a higher IT application maturity often provides employees with a digital platform of higher quality contents. Such digital platform is essential for information exchange and knowledge sharing among employees, such that they can effectively obtain the relevant information and knowledge to support their job tasks. The study tested this assumption by empirically validating the moderating effect of IT application maturity on knowledge sharing.

A refined investigation is further carried out at individual level. It is found that the information network centrality does have significant positive impact on both task and contextual performance. Such impact can be aggregated from individual level to organization level. Therefore, characteristics of employee relationship, represented here by social network characteristics, cannot be ignored when analyzing organization behavior. A good social network structure is beneficial not only to each employee but also to the organization as a whole, which can lead to better performance and higher innovation capability.

There are several future avenues for the current study. First, in the first set of questionnaire, intensity of employee relationship is obtained through traditional survey instead of from construction of social networks. Second, only the intensity of employee relationship and the employee diversity were captured in the model, there might be other employee relationship related variables to consider. Finally, it will be very interesting as well as challenging to examine whether there exist an "optimal" network structure of employee relationship. We will extend the research by working on these avenues.

Acknowledgments. This study is based upon work supported by the National Natural Science Foundation of China under Grant Number 71072086 and the Ministry of Education of China, Humanities and Social Sciences Grant: "Impacts of Social Network on Employee Knowledge Sharing and Innovation (No. 10YJA630124)."

References

1. West, M.A., Farr, J.L. (eds.): Innovation and creativity at work: Psychological and organizational strategies, p. xiii, 349. John Wiley & Sons, Oxford (1990)
2. Ford, C.M.: A Theory of Individual Creative Action in Multiple Social Domains. Academy of Management Review 21(4), 1112–1142 (1996)
3. Porter, M.E.: Competitive strategy: techniques for analyzing industries and competitors: with a new introduction. The Free Press, New York (1998)

4. Wang, D., Guo, W., Liu, X.: Study on the Impact of Group Internal Social Network on Group Creativity. Soft Science 23(9), 25–28 (2009)
5. Granovetter, M.S.: Networks and organizations, Structure, form and action. In: Nohria, N., Eceles, R. (eds.) Problem of Explanation in Economic Sociology, pp. 25–56. Harvard Business School Press, Boston (1992)
6. Burt, R.S.: Structural Holes: The Social Structure of Competition. Harvard University Press, Cambridge (1992)
7. Bian, Y., Zhang, W.: Economic Systems, Social Networks and Occupational Mobility. Social Sciences in China (2), 77–90 (2001)
8. Dalton, D.R., Todor, W.D.: Turnover Turned Over: An Expanded and Positive Perspective. Academy of Management Review 4, 225–235 (1979)
9. Paulus, P.B.: Groups, Teams and Creativity: The Creativity Potential Of Ideal Generation Group. Applied Psychology: An International Review 49(2), 237–262 (2000)
10. Gilson, L.: Diversity, Dissimilarity and Creativity: Does Group Composition or Being Different Enhance or Hinder Creative Performance. In: Proceedings of Academy of Management Meetings, Washington D.C. (2001)
11. Andrews, K.M., Delahaye, B.L.: Influences on knowledge processes in organizational learning: the psychological filter. Journal of Management Studies 37(6), 2322–2380 (2000)
12. Ipe, M.: Knowledge sharing on organizations: A conceptual framework. Human Resource Development Review 2(4), 337–359 (2003)
13. Lin, H.F.: Knowledge sharing and firm innovation capability: an empirical study. International Journal of Manpower (28), 315–332 (2007)
14. Yang, Y., Long, J.: A construct to measure knowledge sharing behavior in organizations. Physiological Journal 40(3), 257–350 (2008) (Chinese)
15. Peng, J., Zhang, G., Chen, R., Tan, Y.: Impacts of Essential Elements of Management on IT Application Maturity – A Perspective from Firms in China. Decision Support Systems 51(1), 88–98 (2011)
16. Klein, K.J., Kozlowski, S.W.J.: Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions, pp. 3–90, xxix, 605. Jossey-Bass, San Francisco (2000)
17. Edmondson, C., Nembhard, I.M.: Product Development and Learning in Project Teams: The Challenges Are the Benefits. Journal of Product Innovation Management 26(2), 123–138 (2009)
18. Amabile, T.M.: A model of creativity and innovation in organizations. In: Staw, B.M., Cummings, L.L. (eds.) Research in Organizational Behavior, vol. 10, pp. 123–167. JAI Press, Greenwich (1988)
19. Amabile, T.M.: Motivating creativity in organizations: On doing what you love and loving what you do. California Management Review 40(1), 39–58 (1997)
20. Vander, A.W., Elfring, T.: Realizing innovation in services. Scandinavian Journal of Management 18(2), 155–171 (2002)
21. Meagher, K., Rogers, M.: Network density and R&D spillovers. Journal of Economic Behavior & Organization 53(2), 237–260 (2004)
22. Chen, Z.F., Guan, J.C.: The Impacts of Small Worlds On Innovation. Chinese Journal of Management Science 17(3), 115–120 (2009)
23. Uzzi, B.: Social structure and competition in inter-firm networks: the paradox of embeddedness. Administrative Science Quarterly 42(1), 35–67 (1997)
24. Bian, Y.: Bringing strong ties back in: Indirect ties, network bridges, and job searches in China. American Sociological Review 62(3), 366–385 (1997)

25. Hansen, T.M.: The search-transfer problem: The role of weak ties in sharing knowledge across organization sub-units. Administrative Science Quarterly 44, 82–111 (1999)
26. Dyer, J.H., Nobeoka, K.: Creating and managing a high performance knowledge sharing network: the case of Toyota. Strategic Management Journal 21(3), 345–367 (2000)
27. Gilsing, V., Nooteboom, B.: Density and strength of ties in innovation networks: an analysis of multimedia and biotechnology. European Management Review 2(3), 179–197 (2005)
28. Fu, Z., Zhou, W.: The Influence of Cluster Firm's Social Capital on Its Innovation Performance. Journal of Strategic Management 2(1), 25–36 (2010)
29. Xia, H., Cai, S.: Effective knowledge management systems. Management Information Systems (Chinese Journal) (6), 47–50 (2001)
30. Van den Hooff, B., De Ridder, J.A.: Knowledge sharing in context: The influence of organizational commitment, communication climate and CMC use on knowledge sharing. Journal of Knowledge Management 8(6), 117–130 (2004)
31. Xiao, J., Xie, K.: An Exploratory Study of IT Application Maturity Model in Enterprises. Journal of Sun Yat-sen University 47(5), 110–118 (2007)
32. Peng, J., Zhong, A.: A Comparative Analysis of Key Factors Influencing the Enterprise IT Application Maturity. Chinese Journal of Management 8(4), 606–613 (2011)
33. Krackhardt, D.: The strength of strong ties: The importance of Philos in organization. In: Nohria, N., Eeeles, R.G. (eds.) Networks and Organizations, pp. 216–239. Harvard Business School Press, Boston (1992)
34. Luo, J.: Social Network Analysis. Social and Science Literature Publishing, Beijing (2010)
35. Zhen, R., Li, S.: Fairness, Trust, and Knowledge Sharing. Human Resources Management Journal 1(2), 69–93 (2001)
36. Borman, W., Motowildo, S.: Expanding the criterion domain to include elements of contextual performance. In: Schmitt, N., Bormaned, W.C. (eds.) Personnel Selection in Organizations, pp. 71–98. Jossey-Bass, San Francisco (1993)
37. Borman, W., Motowildo, S.: Task performance and contextual performance: The meaning for personnel selection research. Human Performance 10(2), 99–109 (1997)
38. Motowidlo, S., Van Scotter, J.R.: Evidence that task performance should be distinguished from contextual performance. Journal of Applied Psychology 79(4), 475–480 (1994)
39. Bai, Y., Wang, Y., Xi, Y.: The Mode of Multi-level Leaderships Effects on Employee's Trust in Leadership, Performance and Innovation: An Empirical Study. Journal of Industrial Engineering and Engineering Management 22(3), 24–29 (2008)

The Study of Construction and Analysis Method of Social Network Model Based on Cooperator Relationship

Xiang Chen and Ning Gao

School of Management and Economics, Beijing Institute of Technology,
No.5, Zhongguancun Nandajie, 100081 Beijing, P.R. China
chenxiang@bit.edu.cn, ningloveacf@163.com

Abstract. As the new work mode develops, virtual collaborative environments play an increasingly important role in future business and personal activities. Our research proposes a new method of constructing collaborators social network model which considers both communication relations and cooperative relations between cooperators. This method uses public information on the work platform to discover the cooperator's behaviors, discovers community relationship intensity between collaborators and uses these to construct the coordinators social network model. Finally, we use the test data from the open source community website *www.codeplex.com* to prove the validity of the model and the method. As a result, this method does not only improve the connectivity of social networks, but also reflects each collaborator's role of team better.

Keywords: Social network, Open source community, Cooperation relationship, User behavior.

1 Research Background and Related Work

With the development of information technology, there are many new work modes emerged. A lot of works have been finished through the distribute work platform in different time and places instead of the traditional work mode. Due to lack of strict organization, carrying out work activities in this work mode always rely on social relations between collaborators. The old work-flow analysis methods are no longer meaningful for this type of work. Because of this, how to collaborate in virtue of the virtual network has become a hot issue in theory and industry. More and more researches have been published in Nature, Science, MS and other world-class magazines. Matthew Van Antwerp, who used social network analysis method to analysis the collaborators' relationship in open source community project in SourceForge.net, proved that the pre-existing relationship between developers played a significant impact on the progress of future projects [1]. France, Giulio, Priscilla and Lin had all proposed a social network model constructed method using contact links between coordinators, and confirmed that the model would yield good results in studying the coordinators' individual behavior and role of work[2],[3],[4],[5]. Korba used social network analysis method to develop a fuller understanding of interactions

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 37–45, 2012.

between workers, and described a software prototype that automatically measures as well as analyzed aspects of collaboration developing visualizations of likely social interactions [6].

Most of the existing researches only use information exchange between collaborators to construct social network model. But in most of the work environment, collaborators' cooperation is equally important for measuring the relation of collaborators. So it is necessary to propose a new method of constructing social network model with work information. It will reflect the relations between the collaborators more correctly.

To solve the above problems, our research proposes a new method of constructing social network model which considers the social relation both occurred during the cooperation and the communication. We use contact information on the working platform to discover collaborators' community relationship intensity, and use work information on the working platform to discover collaborators' cooperation relationship intensity, both of the relationship degree are used to construct the coordinators social network model.

2 The Rules of Defining the Contact Relationship between Collaborators

To explain the virtual network collaboration more clearly, it is necessary to define all kinds of relationships intensity. Relationship intensity is "a (probably) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie. [7]" In the virtual world, collaborators established link through social relations and attention, so the collaboration interaction requires further analysis.

2.1 The Definition and Rules of Cooperator Behavior

Considering the frequency and the time range of collaborators' information exchange, our research defines three rules of conducting the information exchange of the collaborators base on the definition of information issue sequence, as: reply rule, quote rule, co-present rule.

Let $C:\{c_1, c_2, \ldots, c_n\}$ and $I:\{(i_1, t_1), (i_2, t_2), \ldots, (i_m, t_m)\}$ be a collaborators collection and its information sequence, in which t_k is the release time of information i_k. Definite that $c_n \rightarrow i_m$ represents that collaborator n published the information m. We define the three rules:

Rule 1. Reply behavior:
If:$c_1, c_2 \in C, i_1, i_2 \in I, c_1 \rightarrow i_1, c_2 \rightarrow i_2$;$t_1 < t_2, i_2$ includes c_1;
Then c_1 and c_2 have reply behavior.
Rule 2. Quote behavior:
If: $c_1, c_2 \in C, i_1, i_2 \in I, c_1 \rightarrow i_1, c_2 \rightarrow i_2$; $t_1 < t_2, i_2$ includes i_1 ;
Then c_1 and c_2 have quote behavior.
Rule 3. Co-present behavior:

If: $c_1, c_2 \in C, i_1, i_2, i_3, i_4 \in I, c_1 \to i_1, c_1 \to i_2, c_2 \to i_3, c_2 \to i_4$;

$(t_3 < t_1 \& t_1 < t_4) | (t_3 < t_2 \& t_2 < t_4) = 1$, or $(t_1 < t_3 \& t_3 < t_2) | (t_1 < t_4 \& t_4 < t_2) = 1$;

Then c_1 and c_2 have co-present behavior.

It can be seen that reply behavior and reference behavior are the direct interactions between the users, the co-present behavior is an indirect interaction, these is different between users reflect the behavior is different.

2.2 Definition and Measurement of Contact Relationship Intensity

It can be learned from the definition of user behavior, contact relationship is composed of two parts: the direct interaction and the indirect interaction. Thus, the relationship intense is composed of two parts:

Definition 1. The contact relationship intense between cooperators $X_{ij,t}$:

$$X_{ij,t} = (\omega_1 X_{d,ij,t} + (1 - \omega_1) X_{u,ij,t})$$

In which:

（1）$X_{d,ij,t}$ represents the relationship intensity which generated by direct interaction during time t between user i,j, $X_{u,ij,t}$ represents the relationship intensity which generated by indirect interaction during time t between user i,j.

(2) $\omega_1$represents the influence weight of contact relationship generated by direct interaction, $(1 - \omega_1)$ represents the influence weight of contact relationship generated by indirect interaction.

(3) the contact relationship intensity generated by users' direct interaction

Haythornthwaite proposed that the relationship intensity in the network is mainly determined by the following factors: frequency of contact between collaborators, the types of collaborators' contact, the Intimacy between the linkage of collaborators and the common interest of collaborators [8]. Based on the characteristics of the Internet information exchange behavior combined the definition and rules of information issue sequence, our research consider that the relationship intensity is related to the frequency of contact, contact type and Intimacy.

Contact Frequency. In unit time t, the amount of Information exchange taken place between two collaborators is the contact frequency. The contact frequency will be got by the amount of Information exchange taken place between two collaborators. In order to avoid the influence by the difference of activity between the collaborators, we use the incidence of user behavior, which substitute of the amount of user behavior, to measure user relationship intensity.

$$N_{ij,t} = T_{ij,t} / (T_{i,t} + T_{j,t}) \tag{1}$$

In which $T_{i,t}$ is the total number of interaction behavior of user c_i during unit time t, $T_{j,t}$ is the total number of interaction behavior of user c_j during unit time t, $T_{ij,t}$ is the total number of interaction behavior of c_i and c_j during unit time.

Respondents and information publisher has a direct contact in reply behavior, which affects the publisher's information weakly, so the Interaction degree is relatively weak. In quote behaviors, the user represents his recognition to the publisher's information, which can be considered as a possible information exchange behavior, the contact degree is bigger than reply behaviors. We use modulus βr,βq to distinguish reply behavior and quote behavior.

Definition 2. The direct relationship intensity generated by users' direct interaction:

$$X_{d,ij,t} = \beta r N_{r,ij,t} + \beta q N_{q,ij,t} \tag{2}$$

In which:

(1) $N_{r,ij,t}$ represent the reply behavior rate between user i,j during time t, $N_{q,ij,t}$ represent the quote behavior rate between user i,j during time t.

(2) βr,βq respectively represent the modulus of reply relationship and quote relationship.

(3) The contact relationship intensity generated by users' direct interaction.

According to Bo, the contact relationship intensity generated by users' direct interaction can be obtained by the frequency of users pay attention to the same content [9].In order to avoid the influence by the difference of activity between the collaborators, we use the incidence of user behavior, which substitute of the amount of user behavior, to measure user co-present relationship intensity.

Definition 3. The contact relationship intensity generated by users' direct interaction is:

$$X_{u,ij,t} = T_{ij,t}/(T_{i,t} + T_{j,t}) \tag{3}$$

In which $T_{i,t}$ is the total number of co-present behavior of user c_i during unit time t, $T_{j,t}$ is the total number of co-present behavior of user c_i during unit time t.

According to Haythornthwaite and Joinson, intimacies of users' interaction take an important part in the contact relationship intensity of users [8], [10]. Intimacy refers to the time factor of user's interaction occurrence. For the users whose interactions occurred closely, they can be considered as having more closely in the frequency of contact; otherwise, their relationship intensity is weaker.

During the measure of the relationship intensity, time factor λ_t is used to represent the time factor of interaction behavior. To ensure the rules that λ_t will decrease during the time range from now being longer.

Definition 4. The contact relationship intensity between cooperator i,j:

$$DS_{ij,t} = \sum_{t=1}^{T} X_{ij,t}\,\lambda_t \tag{4}$$

In which T is the total time, t represents the unit time, $X_{ij,t}$ represent the contact relationship intensity between users i,j during t, λ_t represents the time factor. our research use the decay function proposed by Gu [11], as :

$$\lambda_t = a^{(T-t)} \tag{5}$$

In which $a(0 < a < 1)$ is time-sensitive factor.

3 The Rules of Defining the Cooperative Relationship between Collaborators

In one project, if there are two identical or similar work activities, and the first activity is executed by worker i, the second activity is executed by worker j, i and j are defined as having a collaboration relationship. The cooperative relationship between these two collaborators is closer than others [12].

In many cases, the similarity degree between the work activities is unknown; it needs to be discovered from the descriptions of the work activities. It can be measured by the cosine similarity of the morpheme in the work activities description. Here are the definitions of similar activities and similar activities set [13], [14].

Rule 4. Let A be a work activities set, two work activities $a \in A, b \in A$, Sim_{ab} is similarity degree of a and b obtained by formula 6, $SimT$ is the threshold of activity similarity degree. If:

$W_{ab}: \{(w_1, a_1, b_1), (w_2, a_2, b_2), \dots, (w_n, a_n, b_n)\}$ is a set of morpheme in the description of work activities a and b in which a_k is the numbers of morpheme in the description of Work activities a; b_k is the numbers of morpheme in the description of Work activities b; n is the number of morpheme in W_{ab} and $Sim_{ab} > SimT$.

Then a and b is considered as similar activities, recorded as $a \backsim b$.

$$Sim_{ab} = \frac{\sum_{k=1}^{n} a_k b_k}{\sqrt{(\sum_{k=1}^{n} a_k{}^2)(\sum_{k=1}^{n} b_k{}^2)}} \tag{6}$$

Definition 5. For a work activities set $A: \{a_1, a_2, \dots, a_n\}$, $\exists a_i, a_j \in A, a_i \backsim a_j$, the set A can be defined as a similar work set.

Base on the similar work set definition in definition 5, our research propose a definition of the collaborator relationship degree between the collaborators.

Definition 6. The collaborator relationship degree between the collaborators:

$$DW_{pq} = \sum_{j=1}^{m} (\sum_{k=1}^{Nm} \min\,(c_p \to a_k^j, c_q \to a_k^j)) \tag{7}$$

In which:

(1) $C: \{c_1, c_2, \dots, c_n\}$ is a collaborator set in one project;

(2) $A_1, A_2, \dots, A_m$, is a group of similar work set, in which a_n^m is a working activities of A_m, m is the number of activities in set A_m;

(3) $c_n \to a_n^m$ is the number of times collaborator c_n work in the activity a_n^m .

4 Social Network Model Construction Methods

Social networks model is mainly used to describe the relationship intense between members within the platform, it can be express by social network diagram.

In the social network diagram, the nodes represent coordinators, the edges between nodes represent the social relations between the collaborators and the weight of the edge represents social relations degree between the two members. This social networks model is in order to discover the relationship between the collaborators, so the direction of the edge is not helpful in our study. For the purpose of discovering the relationship between collaborators more easily, we remove the direction of the network to get the no-direction diagram under the premise of not changing other properties of the network [14].

In the dissemination networks of information in the internet, user node can aggregate other user node by information. In other words, aggregating the information carried by the other nodes information dissemination is the Internet's basic social networking features, and the model formed by the polymerization of "core - edge" is the basic mode of communication in Internet social network. For customer relationship-based social networks, the most useful factor to reflect the importance of the nodes in the network is the degree of the node, the definition of node degree is given below:

Definition 7. A node's degree is the total amount of its contacts with all other nodes in its structure.

Many collaborators do not have the close contact, only because the chance to discuss or cooperate few times, they are measured to have social contact by the rules. Therefore, we define the filtering rules of social relations in this situation.

Rule 5. Set the filter degree of social relationship DT. Let C be the collaborators set, $\mathrm{c_p} \in \mathrm{C}, \mathrm{c_q} \in \mathrm{C}, \mathrm{DS_{pq}}$ is the contact relationship degree between $\mathrm{c_p}$ and $\mathrm{c_q}$. $\mathrm{DW_{pq}}$ is the coordinator relationship degree between $\mathrm{c_p}$ and $\mathrm{c_q}$, $\mathrm{DA_{pq}}$ is the total degree of social relations, if one of the following two assumption established:

(1) $\mathrm{DS_{pq}}$=0, $\mathrm{DW_{pq}} < DT$.
(2) $\mathrm{DW_{pq}}$=0, $\mathrm{DS_{pq}} < DT$.

It is defined that the relation is occurred by accidental, $\mathrm{DA_{pq}} = 0$ otherwise $\mathrm{DA_{pq}} = \mathrm{DS_{pq}} + \mathrm{DW_{pq}}$.

Definition 8. The social network model: the social network model is obtained by using methods described above. Model expressed as a five-tuple: (User1, User2, DW, DS, DA). In which:

(1) User1 and User2 is the two collaborators in the work group.
(2) DW is the degree of cooperative relationship between collaborators.
(3) DS is the degree of contacts relationship between collaborators.
(4) DA is the total degree of social relationship between collaborators.

It can get the sub-graph from the social network graph by using the small groups dividing method of social network model, and moderate the core node of each sub-graph node by the degree of each node. The core node has an important influence on other nodes. We give the definition of core nodes in community work:

Definition 9. Let D_i represents the degree of Node i, in work group S ,if $D_i = \max(D_1, D_2, ...D_n)$, node i is the core of this work group, it is expressed as $n_i \odot S$。

By the above method, we can get social network model of collaborators, and the division of corresponding work sub-network.

We can use the key nodes analysis to get the key graph of social network model to reduce the time required in find other nodes in the graph. There are two types of key nodes in the social network: one is the core node of sub-group, these nodes have powerful influence on other nodes in their sub-group, another is the node which degree is biggest, these nodes is the most important node in link sub-group. In our research, we consider the node which its degree is in the top 5% of all as the core node.

5 Experiment

Our research uses the Microsoft's open source community site *www.codeplex.com* as the experimental data sources. Open source community is a new work mode, it allows users to publish and share project code, discuss the progress of work with other users and finish their work in different spaces or different times. *Codeplex* is *Microsoft*'s open source community platform, which has a total of 19,449 projects, and its project organization is relatively strict. To facilitate the experiment, we used 100 open source project which published the Alpha version and have more than 6 cooperate and have tab ‘sliverlight’, ‘C#.NET’, ‘wpf’ in the library of items. We extracted 1357 project-related collaborators, 12705 posts in the discussion forum and 9707 records of code update from website by the crawler tools *nutch*.

In the experiment, we use the thread discussion in the open source forum as the information exchange sequence. We let the quote modulus β as 0.8, the weight of direct interaction ω_1 is 0.6, the weight of direct interaction ω_2 is 0.4。

We will use the existing cooperation relationship in work group to be verifying data. According to the social network model, we respectively use DS, DW, DA to construct the social network graphs, and use G-N Algorithm to divide the sub-group.

During the validation, we propose the following one indicator as the criterion for measuring the effect of the model.

Definition 10. Predicted relationship intensity D_m is the average relationship intensity of users who have real work relationship in the work platform, the formula is:

$$D_m = (\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} * D_{ij}) / (\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}) \quad (8)$$

In which, C_{ij} is the amount of interactions behavior taken by c_i and c_j, D_{ij} is relationship intensity using experiment data by DS, DW, DA. Predicted relationship intensity will reflect the degree that the model measures the user's behavior, the more the value is, the ability that the model reflects the interaction of user is more powerful.

We use the average distance between the regular nodes and the core graph Pm to verity the effective of the core graph. The lower the value is, the effective of the core is more effective. The experiment result is listed in the table:

Table 1. The experiment result of three model

	isolated nodes	work teams	Dm
DA	78	86	0.0382
DS	146	89	0.0309
DW	253	99	0.0237

Experimental results can be obtained from the above: the model constructed by both the contact information and the cooperator information predict the behavior of social network collaborators more accurately, and divide the project team more accuracy. These reflect that the model make the real social relationship and cooperation relationship known more accuracy. It will make the user to get the interesting work group easily and find work partner easily.

6 Conclusion and Outlook

Our research proposes a new method of constructing collaborators social network model. This method obtains the contact relationship coordinator from the contact information and cooperation relationship from the work information, uses these to construct the social network model which can react the social relationship of collaborators more clearly, and uses the open source community data from *www.codeplex.com* to test and verity the feasibility and effectiveness of the method.

It can be inferred from the experiment that, the new model does not only reflect the work role in the work group more accurately, but also makes the cooperator find the work partner they interest easily. This research provides the bases of the following study about the problem of open source work.

Acknowledgments. This research was supported by the Natural Science Fund of China (No. 71102111, No. 70972006).

References

1. Matthew, V.A.: The Importance of Social Network Structure in the Open Source Software Developer Community. In: Proceedings of the 43rd Hawaii International Conference on System Sciences, Hawaii, USA, pp. 1–10. IEEE Computer Society, Washington (2010)
2. France, B.: Communication Patterns in Distributed Work Groups: A Network Analysis. IEEE Transactions on Professional Communication 42(4), 261–275 (1999)

3. Giulio, C., Manuela, L.: Open Source Communities as Social Networks: an analysis of some peculiar characteristics. In: 19th Australian Conference on Software Engineering, Melbourne, Australia, pp. 387–391. IEEE Press, New York (2008)
4. Priscilla, A.: Redefining and measuring virtual work in teams: An application of social network analysis. In: Proceedings of the 40th Annual Hawaii International Conference on System Sciences, Hawaii, USA, pp. 1–10. IEEE Computer Society, Washington (2007)
5. Lin, C.: Mining Web Social Networks. Ph.D. Dissertation. Fudan University, Shanghai, China (2009) (in Chinese)
6. Korba, L., Song, R., Yee, G., Patrick, A.: Automated Social Network Analysis for Collaborative Work. In: Luo, Y. (ed.) CDVE 2006. LNCS, vol. 4101, pp. 1–8. Springer, Heidelberg (2006)
7. Mark, G.: The Strength of Weak Tie. Ameriean Joumal of City 78(6), 1360–1380 (1973)
8. Haythornthwaite, C.: Strong Weak and Latent Ties and the Impact of the New Media. The Information Society 18, 385–401 (2002)
9. Rong, B., Xia, Z.Y.: Study of interactive features of BBS online complex network and its members. Complex Systems and Complexity Science 6(4), 56–65 (2009)
10. Joinson, A.N.: Self Disclosure in CMC: The Role of Self Awareness and Visual Anonymity. European Journal of Social Psychology 31, 177–192 (2001)
11. Gu, C.J., Zhang, S.Y., Feng, H.B., Sun, Y.F.: A Novel Trust Management Model for P2P Network with Reputation and Risk Evaluation. In: 2010 International Conference on E-Business and E-Government, Guangzhou, China, pp. 3544–3547. IEEE Computer Society, Washington (2010)
12. Esslimani, I., Brun, A., Boyer, A.: From Social Networks to Behavioral Networks in Recommender Systems. In: 2009 Advances in Social Network Analysis and Mining, Athens, Greece, pp. 143–148. IEEE Computer Society, Washington (2009)
13. Mutton, P.: Inferring and Visualizing Social Networks on Internet Relay Chat. In: Proc. of the 8th IEEE International Inferences on Information Visualization, pp. 35–43. IEEE Press, New York (2004)
14. Yin, H.Z., Cao, X.B.: Social network mining with similar content and timing. Computer Engineering 34(1), 18–23 (2008) (in Chinese)

Are Amazon.com Online Review Helpfulness Ratings Biased or Not?

Yun Wan[1] and Makoto Nakayama[2]

[1] University of Houston – Victoria
wany@uhv.edu
[2] DePaul University
mnakayama@cdm.depaul.edu

Abstract. The helpfulness rating of Amazon product review, a popular vote feature used by Amazon to rank product reviews and display them to online shoppers, has important implications for online shopping decisions. This research investigates how objective those helpfulness ratings are. The general assumption is that the ratings are "representative" views of the shoppers. However, previous studies on product reviews indicate bias may also exist among helpfulness ratings. Using the survey questionnaire, the study found that there were indeed significant differences between the helpfulness ratings displayed at Amazon.com and those from a simulated online shopper population. The survey results also show that there are evidences of rating differences by gender, age, ethnicity, income and mobile device use for shopping. Thus the "true" ratings on online user reviews may well be quite different from what we see at Amazon.com. Implications and limitations of this research are discussed.

Keywords: online review, bias, amazon.com, B2C ecommerce.

1 Introduction

Online reviews are becoming an essential component of B2C ecommerce and part of the comparison-shopping information provisions [1]. They mainly have two roles. One is to help online shoppers evaluate products and services before making purchase decisions. The other is the informant role to allow consumers become familiar with a product or service even they do not have immediate purchase intent [2].

According to statistics from U.S. Census Bureau in the last 10 years, B2C ecommerce has been growing steadily and consists increasingly percentage of overall revenue of US retail industry. There are more commodities and services previously not available online become available. There are also increasing number of online shoppers. So we expect reviews will become more critical in terms of linking shoppers and products in online environment.

Amazon.com is the leading B2C ecommerce portal. It also revolutionized many online review feature initiatives to further improve shopping experience, including the helpfulness rating of an online review. For example, once an online review on a

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 46–54, 2012.

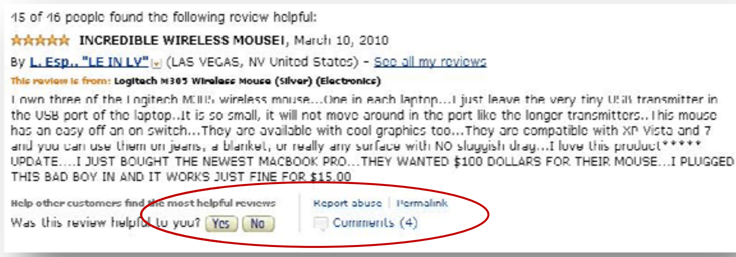


Fig. 1. The helpfulness rating is red-circled

product is being published and read by a registered shopper, the shopper can vote on the helpfulness of the review by simply click the "Yes" or "No" button under the review content (Figure 1).

This simple feature seems very helpful when the number of reviews about a product become too large that a shopper has difficulty to go through even a small portion of them. Thus, based on the overall helpful rating of a review, Amazon.com could use computer algorithms to automatically rank and display those "most helpful" reviews at the top of the review section. So shoppers could spend their valuable time on the most helpful ones and then continue on others until their time allows to avoid overload [3] – an optimized solution for both sides.

This feature is being adopted by many other online retailing sites, becoming a de facto standard. However, we found that there are limitations of this feature and the helpfulness rating is not necessarily the "representative" view of online shoppers.

For the remaining of the paper, we investigated the possible biases of helpfulness rating and using experiment to assess the categorical and extent of such biases.

2 Product Reviews, Their Ratings, and Biases

2.1 Product Reviews

Online reviews refer to those usage experience posted in an online retail website or review site by users about a product and service. Good reviews help online shoppers make informed decisions and mitigate the choice overload effect. For experience and credence products, online reviews become essential to online shoppers to evaluate and make decisions [4]. Because of the positive effect of online reviews for online shopping, in the last five years, we witnessed a tremendous growth of this feature spread among almost all popular online retail sites and the most prominent one is amazon.com. For a popular product, like iPod, at amazon.com, the number of its reviews exceeds hundreds or even thousands.

Product reviews can be biased in many ways [5]. For example, amazon.com provides incentives for reviewers to write good reviews. Product manufacturers want to systematically improve the reputation of their product among consumers through subtle ways like online reviews [6].

2.2 Review Helpfulness Ratings

The abundance of online reviews for a product not only provides convenience to online shoppers to evaluate the product but also pose a challenge – if there are so many reviews for a product with mixed qualities, they could overwhelm online shoppers before the latter even found a helpful one. Thus, online retailers found they have to help online shoppers to identify those helpful reviews effectively. Obviously, established online portals like amazon.com were among those first experienced such problem.

So a filtering mechanism through helpfulness rating was introduced by amazon.com to allow online shopper not only rate a product and write a review but also give a rating about a review's helpfulness. Those reviews receiving the highest proportion of Yes vote against overall vote, or its helpful rating, will be displayed in the prime positions of the product page – the top three positions under the review section. Amazon.com expected such popular voting method to select the most quality reviews thereby helping online shoppers access quality reviews efficiently.

The amazon.com helpfulness rating on each review is a successful feature from user experience perspective. However, the assumption behind this feature, which the reviews selected via this method are among the most helpful because they represent the "voice" of the general population, may not necessarily hold. A "helpful" rating may be biased. Such bias comes from at least two sources: the sequential bias and the self-selection bias.

2.3 Sequential Bias

We use a thought experiment to illustrate the influence from existing ratings, or sequential bias. Suppose there are ten reviews about a product currently available on Amazon.com. They are perfectly equal in helpfulness and all have zero helpfulness rating because they are newly created simultaneously by ten different reviewers. This is actually not unlikely considering the size of product offering and number of reviewers Amazon attracts at any moment.

Now suppose one of the reviews is being displayed at the very beginning and being read by the next 100 online shoppers simultaneously. 30% or 30 of them are busy shoppers. They find the first review helpful, give a "Yes" vote, and then make the purchase before exit. Now this review become most helpful with a 30 out of 30 rating compared with others with zero or less ratings. When the next wave of shopper want to evaluate this product, they will mostly read this review first and probably give the same Yes rating. The Matthew Effect begins to effectuate the cycle and this first review become the most helpful even it has the exact same helpfulness value compared with the others.

We could go a step further by relaxing the assumption slightly and allow those reviews to be not perfectly equal in helpfulness and, say, the first review to be just 2% less helpful than average. It is then not difficult to find that there is still a chance that this not- most-helpful review could stay at the top rank for a while before it is eventually being voted down to a lower rank.

The thought experiment described above is a classic scenario of complexity phenomenon effectuate by sequential bias [7]. Its evolution and eventual outcome – in this case, the current helpful ratings – is path-dependent and unpredictable.

The dynamics of sequential bias and its effect has been systematically studied in several existing literatures, including the online product reviews themselves [8] as well as on artificial music market about how hit songs became hits [9].

Based on above discussion, we suspect that the most helpful reviews for a popular product, which has a highest rating compared with other reviews, may receive a different or probably lower rating if the initial conditions are slightly different. So we have our first hypothesis as following:

Hypothesis 1: The current rating for the most helpful reviews of a product on amazon.com is different from a rating it will receive if all shoppers vote.

2.4 Self-selection Bias

Product reviews is not necessarily objective and suffer from self-selection bias by review authors' gender, ethnicity, income, as well as other social demographic attributes [10]. Product reviews also suffer from homophily effect or its opposite – a review reflect a particular taste and perspective of review authors, which could be very appealing to shoppers that share similar demographic background but not appealing at all to shoppers without such sharing [11]. In another empirical study, Jurca et al. [12] found that groups of users who discuss a certain feature at length are more likely to agree in their rating.

We suspect the self-selection bias exists with helpfulness ratings on product reviews because of the homophily effect. The reviewers are naturally attracted to the product and its reviews because of shared similarities with review authors, not the helpfulness of the review itself.

In addition, not all shoppers write a review or give a vote on a review on amazon.com. Previous research indicates that those who feel imperative about a product tend to leave their comments and ratings about a product [13]. We suspect the same pattern goes to the helpfulness rating. That is only when a voter are particular like the review or hate the review will he or she gives a helpful rating about it.

Thus, we have our second hypothesis:

Hypothesis 2: The most helpful reviews displayed on amazon.com product page reflect the preference of specific online shopper groups with a particular configuration of age, ethnicity, income, as well as other factors.

3 Experiment Design

A survey experiment was designed to test above two hypotheses. We collected some most helpful reviews on most popular products shown in the top page of selected product categories at amazon.com, and masked their helpfulness rating information. Then we asked a group of participants to read those reviews and give their independent helpfulness ratings about these reviews. The results were compared with the original ratings to test hypothesis 1 and the demographic information of participants are being used to test hypothesis 2.

To select the most helpful reviews, we identify top three best selling products from search, experience, and credence product categories respectively [14-16]:

For each product, we chose the top 3 most helpful reviews given by amazon.com. Altogether we collected 27 most helpful reviews. Out of these 27 most helpful reviews, we ranked them based on the total number of votes each review received. And then we select 9 reviews from them to allow there were at least three reviews from each SEC category and the total number of votes each review received were spread out from as small as 0 to 10 range to as high as 90 or higher range.

Table 1. Survey products selection

Search	• T-Mobile G2x 4G Android Phone • Canon PowerShot A30000IS • HP LaserJet Pro P1102w Printer
Experience	• LUNGS (Audio CD) • “Heritage” (Audio CD) • “L.A. Noire” (Video Game)
Credence	• OxyElite Pro • Acidophilus Pearls • Whey

The identified 9 product reviews were screen-captured from the website to retain the original format and style. The helpfulness rating for each review was recorded and then masked. Each review was then being inserted and coded into one web page with a simple “helpful or not” question under it.

About 80 graduate students from four classes in a Midwest university were being invited for the experiment as a convenient sample of randomly selected shoppers.

There is no monetary compensation though an extra credit towards final grade will be given for participation. Each participant was asked to read all 9 reviews and, for each review, they were asked to give a “Yes” or “No” rating for its helpfulness. Demographic information of participants, including gender, age, income, etc. was collected at the beginning of the experiment. A total of 74 valid responses were received.

4 Data Analysis

To test H1, we compared the helpfulness ratings at amazon.com with those obtained from the 74 valid responses. The helpfulness rating is calculated by using following formula:

$$\text{helpfulness rating} = \frac{\text{total number of YES votes received}}{\text{total votes received}}$$

The total votes received including both Yes and No voting. A common rule of thumb to approximate a binomial distribution with the normal distribution is when np and n(1–p) (n: sample size) are both equal to or greater than 10 [17]. In our case, the smaller ones of these values do not exceed 9.62. Therefore, we used the binomial test to compare the amazon.com and sample helpfulness ratings.

In Table 2, "Yes" column and "Total" column under both Ratings categories refer the votes a product review receives. The "Rating" column is the helpfulness rating indicated above. Table 2 lists the p-values in the right-most column regarding the difference between the helpfulness ratings at amazon.com and those obtained from the survey participants. All these p-values are less than .01. Thus, H1 is supported.

We also found that all nine ratings given by participants are lower than those displayed on amazon.com. This is probably because the total votes in our survey is larger than most of their counterparts on amazon.com, which indicates that, all other conditions being equal, ratings given by a larger number of online shoppers will

Table 2. Summary of binomial tests between amazon.com helpfulness ratings and survey respondent ratings

(S: Search Goods, E: Experience Goods, C: Credence Goods)

ID	SEC	Product Name	Amazon.com Ratings			Sample Ratings			Symp. Sig (1-tail)
			Yes	Total	Rating	Yes	Total	Rating	
1	S	T-Mobile G2x 4G Android Phone	13	15	0.87	52	74	0.70	.000
2	S	Canon PowerShot A3000IS	53	57	0.93	54	74	0.73	.000
3	S	HP LaserJet Pro P1102w Printer	61	64	0.95	52	74	0.70	.000
4	E	LUNGS (Audio CD)	36	39	0.92	37	74	0.50	.000
5	E	Heritage (Audio CD)	8	8	1.00	53	74	0.72	.000
6	E	L.A. Noire (Video Game)	23	26	0.88	14	74	0.19	.000
7	C	OxyElite Pro	82	92	0.89	30	74	0.41	.000
8	C	Acidophilus Pearls	47	47	1.00	54	74	0.73	.000
9	C	Whey	64	73	0.88	57	74	0.77	.006

probably lead to lower or more "representative" ratings on the helpfulness of amazon.com reviews.

To test H2, we used chi-square tests to detect any statistical differences on the helpfulness ratings based on 5 profile dimensions of survey participants. These dimensions are: gender, age, ethnicity, income and mobile device use for online shopping. Given the sample size, some cells are expected to contain less than 5. Thus, we applied the Monte Carlo simulation with the sample size of 10,000 for those cases.

Table 3. Demographic influences on helpfulness ratings

(S: Search Goods, E: Experience Goods, C: Credence Goods)

ID	SEC	Product Name	Gender	Age	Ethnicity	Income	Mobile
1	S	T-Mobile G2x 4G Android					
2	S	Canon PowerShot A3000IS	.069		.094*		
3	S	HP LaserJet Pro P1102w	.082			.017*	
4	E	LUNGS (Audio CD)					.028
5	E	Heritage (Audio CD)			.063*		
6	E	L.A. Noire (Video Game)				.060	
7	C	OxyElite Pro					
8	C	Acidophilus Pearls					
9	C	Whey					
		Pearson Chi-Square, df	1	3	3	3	4

*: Based on Monte Carlo simulation (sample size = 10,000)

Table 3 indicates p-values of the chi-square tests regarding the user review helpfulness ratings based on gender, age, ethnicity, income and mobile device use for online shopping of the survey participants (the blank cells mean $p \geq .10$). Provided the small sample size, we use $\alpha = .10$ as the cutoff point.

According to the results, there are at least some significant differences based on one or more profile attributes of online shoppers for the search and experience goods, except for the Android phone.

Gender was a significant factor for two search goods: digital camera and laser printer. The digital camera has some differences in the ratings among different genders and ethnicity. The laser printer saw the different ratings by gender and income levels.

For experience goods, gender and age were not significant factors. Ethnicity was a factor for one audio CD whereas the other audio CD had mobile device use as such a factor.

Finally, income was a factor for the video game. In contrast, we do not see any statistically significant differences for the credence goods. Therefore, the results show that the helpfulness ratings are not the same across at least profile attributes of online shoppers for experience goods and two out of the three search goods we tested.

So H2 were supported by search and experience products.

5 Limitations, Implications and Future Research

One major limitations of this research is the limited number of reviews being used in this study. With a limited number of reviews as sample, we cannot exclude other factors native to these reviews that may lead to the differences we found out in this study. Another limitation is the convenience sample of graduate student we used. There might be inherent bias in the student population even though the graduate students group probably has less such bias compared with undergraduates.

Our finding has important implications to current research related to helpfulness ratings. For example, the helpfulness ratings were being used by mainstream studies as an objective measure about the quality of online review [18]. Our findings indicate that such arrangement maybe problematic in discovering causal relationships.

Online shoppers should also be aware of such bias when they are using helpfulness rating to find product reviews and then use such reviews as buying guide. This is the empirical implication.

Future research could focus on further expanding the review samples and test the extent of such biases in different product categories and domains. Another direction is how to improve the rating process so as to make it more objective and being less influenced by demographic profile of voters. It would also be interesting to use some statistical method to adjust the rating based on the voter and make it more accurate as well as further explore the interaction effect between SEC product categories and homophily effect on ratings.

6 Conclusion

The helpfulness ratings used by Amazon.com is an important feature to help online shoppers identify useful product reviews so as to making shopping decisions. The assumption about the helpfulness ratings is it is representative of online shoppers' opinions and reflects the "true" quality of the review. However, our experiment indicates the helpfulness ratings also suffer sequential and self-selection bias and only represents the opinion of certain group of online shoppers that sharing some common demographic attributes. Online shoppers should be aware of such limitations of helpfulness ratings so as to better utilize them in shopping activities. Online portals like amazon.com should further refine helpfulness rating ranking and voting mechanism so as to make them more representative.

References

1. Wan, Y., Menon, S., Ramaprasad, A.: A Classification of Product Comparison Agents. Communications of the ACM 50, 65–71 (2007)
2. Chen, Y., Xie, J.: Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix. Management Science 54, 477–491 (2008)
3. Maes, P.: Agents that reduce work and information overload. Communications of the ACM 37, 30–40 (1994)
4. Senecal, S., Nantel, J.: The influence of online product recommendations on consumers' online choices. Journal of Retailing 80, 159–169 (2004)
5. Mackiewicz, J.: Reviewer bias and credibility in online reviews. In: 72nd Annual Convention of the Association for Business Communication, Washington, D.C. (2007)
6. Dellarocas, C.: Online reputation systems: how to design one that does what you need. MIT Sloan Management Review 51, 33–38 (2010)
7. Brian Arthur, W., Ermoliev, Y.M., Kaniovski, Y.M.: Path-dependent processes and the emergence of macro-structure. European Journal of Operational Research 30, 294–303 (1987)
8. Kapoor, G., Piramuthu, S.: Sequential bias in online product reviews. Journal of Organizational Computing and Electronic Commerce 19, 85–95 (2009)
9. Salganik, M.J., Dodds, P.S., Watts, D.J.: Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. Science 311, 3 (2006)
10. Li, X., Hitt, L.M.: Self-Selection and Information Role of Online Product Reviews. Information Systems Research 19, 456–474 (2008)
11. Wang, C., Zhang, X., Hann, I.-H.: Social bias in online product ratings: a quasi-experimental analysis. In: WISE 2010, St. Louis, MO (2010)
12. Jurca, R., Garcin, F., Talwar, A., Faltings, B.: Reporting incentives and biases in online review forums. ACM Transactions on the Web 4, 1–27 (2010)
13. Hu, N., Zhang, J., Pavlou, P.A.: Overcoming the J-shaped distribution of product reviews. Communications of the ACM 52, 144–147 (2009)
14. Nelson, P.: Information and consumer behavior. Journal of Political Economy 78, 311–329 (1970)
15. Darby, M.R., Kami, E.: Free Competition and the Optimal Amount of Fraud. Journal of Law and Economics 16, 66–86 (1973)
16. Nelson, P.: Advertising as information. Journal of Political Economy 82, 729–754 (1974)
17. Gravetter, F.J., Wallnau, L.B.: Statistics for the Behavioral Sciences. Wadsworth, Belmont (2009)
18. Mudambi, S.M., Schuff, D.: What makes a helpful online review? a study of customer reviews on Amazon.com. MIS Quarterly 34, 185–200 (2010)

A Social Network Based Analysis of Deceptive Communication in Online Chat

Jinie Pak and Lina Zhou

University of Maryland Baltimore County (UMBC), Department of Information Systems,
1000 Hilltop Circle, Baltimore, MD 21250, USA
{jpak3,zhoul}@umbc.edu

Abstract. The digital information era has allowed increasing online deception opportunities. In addition, the performance of detecting deception has been no better than chance in face-to-face communication, and is reported to be even worse in computer-mediated communication (CMC). Thus, there is a great need to uncover effective cues to deception in CMC. Online interaction weaves an implicit social network. However, Deception in CMC has not been examined from the social network perspective. To fill this gap, this research explores interaction patterns of deceivers in online chat. Based on an analysis of social networks created from online chat messages, this study provides preliminary evidence for the efficacy of social network based metrics in discriminating deceivers from truth-tellers. The findings of this study have significant implications for deception research in social media.

Keywords: Deception Detection, Online Deception, Cues to Deception, Social Network Analysis, Online Chat.

1 Introduction

Despite of the efforts that make Internet more reliable and secure space, there are still sophisticated attacks or acts of deceivers in online communication such as scamming, phishing, impersonating, profile hijacking, cyber stalking, and trust authority compromising [1]. Due to the geographical distance and anonymity, users easily get exposed to such malicious attacks. Furthermore, deception detection in face-to-face communication is no better than chance and some preliminary evidence shows that online deception detection is even more challenging. Therefore, there is a great need to find effective cues to deception in computer-mediated communication (CMC).

Extant online deception studies have analyzed an individual's attributes such as non-verbal and verbal features to understand deception behaviors (e.g., [2, 3, 4]). For instance, deception research on text-based CMC has found that some linguistic features such as negative emotion, language diversity, and cognitive complexity are associated with deceptive behavior of an individual. Although deception is interpersonal and social interaction among people, deceptive behaviors in online communication have not been investigated from the social structure perspective.

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 55–65, 2012.

Social network analysis has been a powerful tool to uncover power and centrality of members and to identify subgroups and social roles such as leaders and gatekeepers in social networks. In recent years, several studies (e.g., [5, 6, 7]) have employed social network analysis techniques for criminal or terrorist network analysis. These studies have examined data collected from crime classification database (e.g., criminal records, financial transaction records) or from online media such as email, discussion forums, and mailing list in order to mine criminal or terrorist social networks. However, few studies have investigated deception in online chat from the social structure perspective. To fill the gap, this study explores structural cues to deception in online chat.

This study potentially contributes to the deception research from the following two perspectives. First and foremost, this study applies social network metrics to the understanding of interaction patterns of deceivers in online chat. Second, this research develops heuristic strategies for the creation of social networks of online communication.

The rest of the paper is organized as follows. Section 2 first provides theoretical background on online chat room discourse and social network analysis, and then proposes hypotheses. Section 3 describes data collection and analysis, and reports results. The last section discusses implications of the findings, limitations of the current research, and further directions.

2 Theoretical Background and Hypothesis Development

This section introduces characteristics of online chat room, social network analysis, and the construction of social network matrix.

2.1 Online Chat Rooms

Online chat is a popular modality in support of computer mediated communication (CMC) that allows users to engage a real time text mediated communication and collaboration with others from geographically different locations. There have been several studies on deception cues in instant messaging [4, 8]. Instant messaging is a client based peer-to-peer chat discussion occurring between small numbers of participants. In contrast to instant messaging, a chat room is a server-based chat discussion occurring on one or more dedicated servers, where participants log on and communicate with several other participants only known by their screen names [9]. Online chat rooms are many-to-many communication. As a result, the messages displayed in a chat room can be read and responded by all connected participants in the chat room. Since media are expected to have influence on cues to deception [4, 10], investgating cues to deception in online chatrooms should be treated separately.

In general, people in an online chat room must be electronically present at the same time, and messages are immediately transmitted through the intermediate servers to all participants, wherever they may be. Thus, compared to asynchronous communication, an online chat room has several distinct features in delivering messages to other people as follows [11]:

- **Shorter and Less Syntactical Messages.** Preparation time for messages is shorter as well as response time to the messages. Thus, messages in online chat rooms are shorter and in smaller chunk in order to contain necessary information.
- **Turn-Taking.** Synchronous online chat involves more rapid exchanges of turns and its messages are transmitted linearly and dynamically.
- **Temporal Properties.** Relevant messages tend to occur temporally adjacent to initiating turns. However, when adjacency is disrupted by delays in message transmission (lag) or differential rates in receiving messages, it is difficult to track sequential exchanges and interaction.

Therefore, online chat rooms are characterized by dynamic and temporal pattern of interactivity among participants. To understand interactions in online chat rooms, it would be helpful to examine not only the individuals within a group but also the relationships among individuals. To this end, social network analysis provides useful methods and metrics.

2.2 Social Network Analysis

All social networks consist of two types of elements: actors (nodes) and relations (links) among actors. The relationships among these actors may indicate communicative interaction, and the nature and strength of relationships may also vary.

Social network analysis (SNA) is defined as a method enabling the patterns of relations among social actors at different levels of analysis such as the node-level and the network-level [12]. SNA explores the structure of social relationships within a network to uncover the informal links between people and to identify how people of the network influence one another [13]. SNA has been used to describe and model a wide range of relations such as information diffusion, terrorist networks, and recently CMC-supported learning networks. Centrality is a popular social network metrics [1], which deals with the roles of individuals in a network. The degree of a particular node (actor) is its number of links that lead into or out of the node; its betweenness is the number of geodesics (shortest paths between any two nodes) passing through it; and its closeness is the sum of all the geodesics between the particular node and every other node in the high betweenness may be a gatekeeper in the network.

2.3 Construction of Social Network Matrix

In order to construct a social network matrix, the first step is to identify actors and the presence of relations in an online chat room. This can be done by monitoring the activity in the chat room and identifying specific user interaction. Inferring actors is relatively simple, since user ID or name can be easily extracted from the messages for matching and consolidation. It is generally assumed that each user has a unique ID and each message is associated with a unique source ID. However, inferring relationships is relatively complicated. In order to create relationships between nodes, heuristic-methods have been developed, which has been used to capture crude relations among users in synchronous CMC such as Internet Relay Chat (IRC)

[13, 14]. Although the accuracy of these approximations is subjective by nature, the results are generally good and afterward should be followed by cross-validation. There are four heuristic methods used for building the set of relationships: direct addressing of users, temporal proximity, temporal density, and monitoring coherence of messages [11, 13, 15, 16].

- **Direct Addressing of User Name/ID.** Occurrences of direct addressing in conversations among participants have been examined to infer relationships. Due to the multi-interaction in a chat room, every initiation and response pair is almost disrupted by intervening material. Nonetheless, the intended recipient of each message can be tracked by the message that contains user name or ID of the recipient explicitly addressed by a sender. This is a very simple yet reliable way of building the set of edges, but it needs to be confirmed by other methods.
- **Temporal Proximity.** Direct addressing is not always used to specify the recipient of a message. A message without explicit direct addressing is either targeted to everybody or, to an individual. Since messages in a chat room are transmitted in a linear order, turn-taking is an important cue to infer relationships among participants. Usually, the sequence of messages: initiation – follow-up – responses, occurs adjacently. For example, a message from one participant is followed up immediately by a message from another participant. Then it is reasonable to imply that those two messages are relevant each other and this sequence can be used for inferring a relationship between the two participants.
- **Temporal Density.** If there are no long delays or overlaps in chat room conversations, it is still possible to derive clues about the structure of the social network by examining other temporal features. If there are *n* messages within a short period time and all *n* messages are sent out by only two participants, then it is reasonable to assume that these two participants are engaged in conversation.
- **Monitoring Coherence of Messages.** This method infers relationship between senders and recipients based on the understandings of topics or subjects discussed by participants in a chat room. Accordingly, conversational and textual coherence in a chat room makes possible to infer relationship between participants. Coherence is the discourse connection of text and conversations such as statements-to-statements, statements-to-people, and people-to-people [17]. Five strategies for creating cohesion in online chatting messages are applied [18]: 1) reference (words used to refer back to previously mentioned subjects), 2) substitution (words used instead of another), 3) ellipsis (connections become clear through the exclusion of certain words), 4) conjunction (segments liked through specific linking words), and 5) lexical cohesion (links created through lexical repetition).

2.4 Hypotheses Development

According to Interpersonal Deception Theory (IDT) [19], deception is not a static process, but an interactive and strategic process. In other words, a deceiver's behaviors and strategic moves are closely influenced by a respondent's behaviors during mutual communication. IDT suggests that there are two types of deception strategies: persuasive strategy and protective strategy. For example, while a deceiver persuades (deceives) the others to have a false belief about a certain topic, he or she also has to pretend to be innocent or truthful in order to protect him or herself from being caught in deception. Therefore, these two dichotomous strategies affect fundamentally on online behaviors of a deceiver [19, 20]. In case of an online group communication, a deceiver has to manage the above two strategies effectively to accomplish deception successfully under the significantly limited time and cognitive resources.

Previous studies show that the amount of messages sent by deceivers in instant messaging communication is significantly less than that sent by truth-tellers [8, 10, 21]. In a real-time online group communication, a deceiver has to deal with multiple receivers. Particularly, a deceiver has to monitor receivers' reactions and guess stance of each different receiver from minute to minute while deceiving. Accordingly, a deceiver's interaction with multiple-receivers becomes increasingly difficult as the group conversation progresses, which can reduce the production of messages by the deceiver. Messages in text-based CMC imply the relationships among participants. Accordingly, the number of social ties in online group chat can be inferred by a production of messages. So the first hypothesis is proposed as follows:

H1: Compared with truth-tellers, deceivers have fewer numbers of ties in online chat.

Influence in a social network is closely related to the concept of power and centrality. Power and centrality can be described in a way that an actor is embedded in a social network as imposing constraints on the actor and offering the actor opportunities [22]. If actors are less dependent on others in a social structure and face fewer constraints and have more opportunities than others, they are in a favorable structural position. In an online group communication, an individual who intends to deceive others should take more influential position in a network structure than other receivers in order to accomplish his or her goal successfully. In addition, a deceiver is more likely to focus on figuring out other receivers' feedbacks and intentions, while a truth-teller doesn't pay as much attention to such information. If a deceiver could effectively apply deception strategies to hide his or her own role, he or she might influence other receivers' decision making processes during online communication. There are three different social network measures to determine an actor's ability in influence and power among actors [23]. The pattern of ties can be described as direct access to actors, brokerage linking together individuals, groups, or even whole network, and information efficiency. Three centrality measures including degree, betweenness, and closeness centralities are employed to look for the relations of a deceiver in comparison to the relations of a truth-teller in a group. Thus, the second hypothesis is proposed as the following:

H2: Deceivers and truth-tellers differ in the network centralities: 1) degree, 2) closeness, and 3) betweenness.

3 Research Methodology

3.1 Data Collection

The data for this study was collected from an online version of the mafia game [24] (www.epicmafia.com). The game is played in an online chat environment. The game was selected to study the deceivers' interaction patterns because game players employ deception strategies in order to win. There are many variants of mafia game, and the simplest version of the game was selected for analysis in this study.

There are three different roles in the selected version of this game: mafia, villager, and third party. Villagers know only their own identity, whereas mafia members know identities of their fellow players. The objective of the mafia game is to eliminate the opponent group. At the end, the group with a large number of survived members would win. Each game consists of two phases (days and nights). During days, all players can discuss and vote who they want to lynch or who the mafia is. During nights, the mafia kills a villager or a third party player. Based on his/her role in the game, each player behaves differently because conversation during a game would affect the final outcome of game: either the mafia wins or the villagers win. Based on game statistics, a homogeneous sample of 44 games were selected, which are all composed of the following roles: one mafia, three villagers, one cop, one hooker, and one watcher. The distribution of game outcomes is 50:50.

3.2 Data Preparation

As discussed earlier, a social network needs to be created before performing social network analysis. The nodes of the network contain individual players of a game, and the relationships are established based on message exchanges in a game chat room, which exclude pre-game and post-game messages and trolling messages such as off-topic messages and interruptive messages. Specifically, the relationships were inferred using a heuristic method, which involves the following rules:

- If a number of messages from two or more users occur in adjacency or a sequential order, a relationship or relationships can be assigned to these users. For example, a directional relationship between TCM and SFS is inferred based on the following messages, since TCM initiates a conversation and SFS responds to TCM.

```
TCM:          msg: at this point
SFS:          msg: kk
TCM:          msg: its stupid
SFS:          msg: I will do that next time
```

- If a user directly addresses another user's ID or name in its message, a relationship can be assigned to these users. For example, a relationship between TCM and GAR is inferred from the following messages.

```
TCM:          msg: gar*
TCM:          msg: claim
GAR:          msg: Blue
```

- If a relationship between two or more users repeatedly occurs, it will be added up to the weight of the relationship. For example, the following back and forth message exchanges between MC and JSL increase the weight of the relationship.

```
MC:           msg: hooker here
MC:           msg: btw
JSL:          msg: I know.
MC:           msg: so theres a cop
MC:           msg: in this setup
JSL:          msg: I`m vanilla.
JSL:          msg: Wanna kill SFS just in case?
MC:           msg: who do u want to kill and who
                   do u think is cop?
```

- If two or more messages are connected with textual cohesion such as reference, substitution, ellipsis, conjunction, and lexical cohesion, a relationship or relationships between these two or more owners of messages can be assigned. For example, adjacent two messages from SFS and MC are connected with conjunction 'and', which indicates textual coherence. Thus, a relationship between SFS and MC is established.

```
SFS:          msg: I`ll be on you
MC:           msg: and then the persons name
SFS:          msg: You be on me
```

- If a message responds to another message including the initiation of conversations such as discussions, questions, or arguments, a relationship between these two owners of messages can be assigned. For example, the following two messages form a question-response relationship, so a relationship is inferred between MC and SFS.

```
MC:           msg: should i claim?
SFS:          msg: You should claim
```

After a social network was created for each game, the actors of the networks are mapped to corresponding game roles for further analysis. NodeXL, an open- source add-in for Excel 2007, was used to derive social network metrics in the current study. NodeXL imports data in Pajek files, other spreadsheets, comma separated value (CVS) files, or incidence matrices [25].

4 Analyses and Results

According to the game rules, a mafia has to successfully deceive other players to survive until the end of a game, whereas a cop and villagers have no intention to deceive other players. Therefore, to discover the patterns of deceptive interaction in online group chatting, we compare the network metrics of the mafia (the deceiver) against those of the cop and villagers (both are truth-tellers) separately. In case of villagers, the average scores of three plays were used in the comparisons.

A paired–sample T-test is performed to test the hypothesis H1. The results are reported in Table1, which did not yield significant results (*p*>.05). Thus, hypothesis H1 is not supported.

Table 1. Paired-sample T-test on the number of social ties

Pair Difference	Mean	Std. Deviation	Std. Error Mean	T	Sig. (2-tailed)
Pair 1 C – M	.93182	19.94924	3.00746	.310	.758
Pair 2 V – M	-.58409	14.46619	2.18086	-1.643	.108

To test hypotheses H2, four popular centrality measures were selected, including in-degree, out-degree, closeness, and betweenness.

- **In-Degree and Out-Degree.** A count of the number of edges that are connected to a node. 1) in-degree is the number of edges that point toward the node of interest and 2) out-degree is the number of edges that the node of interest points toward [26].
- **Closeness.** The number of the shortest paths between nodes [26].
- **Betweenness.** A measure of the average distance from each node to each other nodes [26].

Table 2. Paired-sample T-test on centrality measures and clustering co-efficient

Measure: Mafia vs. Cop			
Social Network Measures	Mean Difference (Cop-Mafia)	Std. Error	Sig.
In-degree	-.909	.334	.009*
Out-degree	-.159	.307	.607
Closeness	-1.317	.391	.002*
Betweenness	-2.595	.316	.000*
Measure: Mafia vs. Villager			
Social Network Measures	Mean Difference (Villager-Mafia)	Std. Error	Sig.
In-degree	-.636	.355	.080
Out-degree	.409	.221	.071
Closeness	-1.324	.392	.002*
Betweenness	-1.527	.574	.011*

Table 2 reports statistical test results of social network measures. The results indicate that most social network metrics significantly differ between the mafia and the cop and between the mafia and villagers. Specifically, the mafia has a higher level of betweenness in comparison with the cop or the villagers. On the other hand, the cop or the villagers have a higher level of closeness (indicated by a lower value) and in-degree. However, the analysis on outdegree did not yield significant results.

5 Discussion

5.1 Major Findings and Implications

The study investigates interaction patterns of deceivers in online chat. The empirical evaluation of the current study yields mixed results. In terms of number of ties in online group chat, there is no significant difference between a mafia and a cop/a villager, which is not consistent with the previous studies [2, 3, 4]. Such a discrepancy could be caused by the different operationalization of productivity, which was defined as the number of messages sent/turns taken by each player and the size of social networks that is relatively small.

The findings on closeness and betweenness centrality metrics confirm our hypotheses. Interestingly, a deceiver has higher betweenness centrality than a truth-teller, while a truth-teller has higher closeness centrality than a deceiver. These two social network measures are particularly revealing of an actor (node)'s advantageous or constrained position in a network. The closeness centrality can be described as the efficiency of information access in a network. The capacity for reaching more people more quickly gives influence in the network. The betweenness centrality can be interpreted as the ability to broker relationships between people that lack other connections. Thus, the node with higher betweenness would allow, withhold, or distort incoming and outgoing resources. For instance, if an individual has high betweenness centrality, he or she is more likely to be the one who has control over information transfer and diffusion in a network. If an individual has high closeness centrality, he or she is likely to receive information more quickly than any others. The results of the current study reveal that a deceiver is likely to play an important role and take an influential position in controlling information flow of other members within a social network. Accordingly, interaction patterns of a deceiver is more likely to function as "a broker", not "a leader or a core" in online group communication.

We can provide the following alternative explanations for the lack of support for our hypotheses on out-degree centrality, despite that there exists difference in in-degree centrality between the mafia and the cop. First, the group size of games selected for the current study ranges between 6 and 7, which is considered as relatively small networks. Since the degree centrality is computed in terms of the number of nodes to which a particular node is adjacent to, so it ignores any indirect connections it may have. As a result, the nodes in the small networks as used in the current study tend to be fully connected, which makes the degree centrality less useful for examining the influence or power of an individual in a network. Secondly, deceptive interactions may have influence on the degree centrality. The degree centrality is used to measure the ability of an individual to directly influence or gain access to other individuals in a network. However, given the intention to deceive, deceivers should be cautious that frequent interactions may arouse suspicion from other receivers. Thus, the deceivers tend to control and maintain appropriate conversational involvement depending on receivers' reactions and other contextual factors. They may go to the other extreme by talking less and becoming nonimmediate and less expressive.

This study has made multifold research contributions. First, this is the first study that applies social network analysis to understand deceiver's interaction behaviors in online chat. Second, this study used real-world data. Third, the study adopts the existing heuristic-based methods in other domains such as asynchronous CMC (discussion forum, emails) and synchronous CMC (IRC, IM) and modifies the methods to apply for identifying relationships among nodes in an online chat environment.

5.2 Limitation and Further Research

Much work remains to be done in this field of online deception and several issues are worthy of future research attention. First, the current study only looks for measures of the connectedness in node-level social network analysis. To understand social structural cues to deception in a network accurately, more in-depth social network analysis should be conducted. Second, deception behavior may change over time, and accordingly dynamic social network analysis would be helpful in revealing temporal structural patterns of deceptive interactions. Third, the construction of social network metrics for online interaction requires cross-validation. Finally, the future research should replicate the current study with larger networks.

References

1. Zejda, D.: From Subjective Trust to Objective Trustworthiness in On-line Social Networks: Overview and Challenges. Journal of Systems Integration, 16–22 (2010)
2. Burgoon, J.K., Blair, J.P., Qin, T., Nunamaker Jr., J.F.: Detecting Deception through Linguistic Analysis. In: Chen, H., Miranda, R., Zeng, D.D., Demchak, C.C., Schroeder, J., Madhusudan, T. (eds.) ISI 2003. LNCS, vol. 2665, pp. 91–101. Springer, Heidelberg (2003)
3. Hancock, J.T., Curry, L., Goorha, S., Woodworth, M.: Automated Linguistic Analysis of Deceptive and Truthful Synchronous Computer-Mediated Communication. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 00(Cmc), 22c–22c. IEEE (2005)
4. Zhou, L.: An Empirical Investigation of Deception Behavior in Instant Messaging. IEEE Transactions on Professional Communication 48(2), 147–160 (2005)
5. Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., Chau, M.: Crime data mining: a general framework and some examples. Computer 37(4), 50–56 (2004)
6. Xu, J., Chen, H.: Criminal Network Analysis and Visualization. Communications of the ACM 48(6) (2005)
7. Yang, C.C., Ng, T.D.: Terrorism and Crime Related Weblog Social Network: Link, Content Analysis and Information Visualization. In: 2007 IEEE Intelligence and Security Informatics, pp. 55–58. IEEE (2007)

8. Burgoon, J.K., Chen, F., Twitchell, D.P.: Deception and its Detection Under Synchronous and Asynchronous Computer-Mediated Communication. Group Decision and Negotiation 19(4), 345–366 (2009)
9. Bengel, J., Gauch, S., Mittur, E., Vijayaraghavan, R.: ChatTrack: Chat Room Topic Detection Using Classification. In: Chen, H., Moore, R., Zeng, D.D., Leavitt, J. (eds.) ISI 2004. LNCS, vol. 3073, pp. 266–277. Springer, Heidelberg (2004)
10. Carlson, J.R., George, J.F.: Media Appropriateness in the Conduct and Discovery of Deceptive Communication: The Relative Influence of Richness and Synchronicity. Group Decision and Negotiation 13(2), 191–210 (2004)
11. Herring, S.C.: Interactional coherence in CMC. In: Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences, HICSS-32. Abstracts and CD-ROM of Full Papers, 00(c), p. 13. IEEE Comput. Soc. (1999)
12. Scott, J.: Social Network Analysis. Sage Publications, London (2000)
13. Mutton, P.: Inferring and visualizing social networks on internet relay chat. In: Proceedings of Eighth International Conference on Information Visualisation, IV 2004, pp. 35–43. IEEE (2004)
14. Tuulos, V.H., Tirri, H.: Combining Topic Models and Social Networks for Chat Data Mining. In: IEEE/WIC/ACM International Conference on Web Intelligence (WI 2004), pp. 206–213 (2004)
15. Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., et al.: Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. Computational Linguistics 26(3), 339–373 (2000)
16. Berglund, T.Ö.: Disrupted Turn Adjacency and Coherence Maintenance in Instant Messaging Conversations. Language@Internet 6, article 2 (2009)
17. Erickson, T., Herring, S., Sack, W.: Discourses Architectures: Designing and Visualizing Computer Mediated Conversation. In: Proceedings of ACM CHI, Minneapolis, MN, April 22 (2002)
18. Halliday, M., Hasan, R.: Cohesion in English. Longman English Languages Series, vol. 9. Longman, London (1976)
19. Buller, D.B., Burgoon, J.K.: Interpersonal Deception Theory. Communication Theory 6(3), 203–242 (1996)
20. Zhou, L., Sung, Y.: Discourse Cues to Online Deception. In: Proceedings of the Credibility Assessment and Information Quality in Government and Business Symposium (2010)
21. Zhou, L., Zhang, D.: A Comparison of Deception Behavior in Dyadic and Triadic Group Decision Making in Synchronous Computer-Mediated Communication. Small Group Research 37(2), 140–164 (2006)
22. Hanneman, R.A., Riddle, M.: Introduction to social network methods. University of California, Riverside (2005), published in digital form `http://faculty.ucr.edu/~hanneman`
23. Hafner-Burton, E.M., Montgomery, A.H.: Centrality in Politics: How Networks Confer Influence. Human Rights, 1–21 (2010)
24. EpicMafia, `http://www.epicmafia.com`
25. Bonsignore, E.M., Dunne, C., Rotman, D., Smith, M., Capone, T., Hansen, D.L., Shneiderman, B.: First Steps to Netviz Nirvana: Evaluating Social Network Analysis with NodeXL. In: International Conference on Computational Science and Engineering, pp. 332–339. IEEE (2009)
26. Smith, M.A., Shneiderman, B., Milic-Frayling, N., Rodrigues, E.M., Barash, V., Dunne, C., Capone, T., et al.: Analyzing (social media) networks with NodeXL. In: Proceedings of the Fourth International Conference on Communities and Technologies CT 2009 (Figure 1), p. 255 (2009)

Research on Financial Super-Network Model Based on Variational Inequalities

Xuan Liu[1], Jia Li[1], Zhigao Chen[1], and Pengzhu Zhang[2]

[1] Department of Management Science and Engineering, Business School, East China University of Science and Technology, Shanghai 200237, China
{xuanliu,jiali,zgchen}@ecust.edu.cn
[2] Department of Management Information System, Antai College of Economics & Management, Shanghai Jiaotong University, Shanghai 200052, China
pzzhang@sjtu.edu.cn

Abstract. How social networks and financial transaction networks interact on each other has drawn more and more interests for supervision agencies and financial institutions in their efforts to combat money laundering. By introducing super-network theory, we proposed a super-network model integrating social network and financial transaction network. Based on this super-network, we presented a multiple objective decision model, and after analyzing the optimal functions of the agents, the equilibrium flows for both social network and financial transition network are found so as for the super-network achieves an equilibrium state. Then we discussed how to analyze the suspicious transaction flows or suspicious financial agents using those equilibrium flows.

Keywords: Super-network, Financial transaction network, Social network, Variational inequalities.

1 Introduction

Money laundering (ML) is attracting more and more attentions from government and scholars all around the world due to its significant illness. For one thing, money laundering is always related with other crimes such as drug abuse, terrorism, and other financial crimes. For another thing, every year, ML related money accounts for large percentages in the world's GDP [1].

In China, our anti-money laundering (AML) process has achieved quite huge progress since the crime of money laundering first issued in "Criminal Law" in 1997. Nowadays, "Administrative Rules for Reporting of Large –value and Suspicious Transactions" was published by People's Bank of China in 2003 [2], he and his second edition issued in 2007 [3] become the major instructive document for related financial institutions in combating ML activities. Today, China has formed a systematic union combining different related agencies. It includes legislation agencies, information agencies, jurisdiction agencies, and administration agencies, within which, information agencies (including financial regulation agencies and

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 66–76, 2012.

various financial institutions) plays important role in collecting, analyzing, delivering and identifying large value and suspicious cases [4]. However, previous researches showed that there exists huge space in enhancing efficiency of detecting SARs for information agencies [4]: Those systems used in financial institutions nowadays are mainly based on fixed rules or given thresholds which can be easily escaped and evaded by real money launders, meanwhile for financial regulation agencies it is lack of tools to trace suspicious clients or transactions based multi-source information, for example, collaborating social information and transaction information .

This paper aims to propose a super-network based model integrating financial transaction network and social network and solve the multi-criteria decision model for each client (agent) using variational inequalities. It is organized as following: the second section introduces those previous researches. The third section proposed a super-network model integrating social networks and financial networks, and we discussed how to use super-network based model to identify suspicious transactions or accounts. In the end we also summarized the paper and elaborate our future direction.

2 Previous Studies

2.1 Suspicious Activity Detection

Previous suspicious activity related detection can be classified into different levels: suspicious transaction detection, suspicious clients/accounts detection, and suspicious group detection. Suspicious transaction detection aims to identify most abnormal transactions; client/account detection aims to separate suspicious clients or accounts that differentiated from those in the same peer groups, while suspicious groups intends to find out criminal gangs or usually hidden organizations. In all, the suspicious activity detection can be viewed as an outlier detection problem, in which, most of the methods are categorized into threshold-based detection and state-based detection. In AML realities, most AML detection systems use human-set thresholds to filter those suspicious transactions according to survey result from China [4]. Among the more elaborate AML systems, the American National Association of Securities Dealers, Inc. uses break detection technology to detect abnormal stock transactions [5] and the U.S. Financial Crimes Enforcement Network AI System (FAIS) uses suspicious scores to flag certain types of transactions and activities, and it also utilizes link analysis technology to detect related crimes [6][7]. Some researchers also focus on using machine learning methods to help detect suspicious activities. For example, [8] used SVM to deal with labeled transaction data in order to find out suspicious activity and [9] introduced decision trees to customer AML risk assessment using manually labeled examples to train the trees. [10] used scan statistic model to identify most unusual transferring fragments (with high frequency or high transaction volumes) within a short time. Also there are papers focusing on finding suspicious accounts whose transaction tendencies were differentiated from other customers in the same industry [11][12]. Compared with transaction detection and account level detection, group detection related research is really limited, representative research includes link analysis methods proposed by [6] and hidden group based model by [13].

Summarizing the previous studies, we found that those information used in AML detection are mainly transaction information includes category data such as flow direction / operation type, as well as numerical data including transaction amount, transferring frequency and so on [14]. While other information such as social relationship levels between different pairs of accounts are not used as meaningful information in detection process. In the next paragraph we would like to summarize interrelationship that does exists between social network and financial networks so as to demonstrate the feasibility of introducing social relationship into suspicious detection.

2.2 Interrelationship between Social Network and Financial Networks

Actually, previous literatures have demonstrated that social networks do play roles on financial networks. And the role that relationships play in financial networks has been studied analytically as well as empirically in several different contexts. [15] described the role of social networks in the context of micro-financing. [16] dealt with the connection between relationships and leading. [17] surveyed on network formation with emphasize on how networks of relationships play an important role in many economic situations. [18] suggested that social relationships and networks affect personal and corporate finance dealings. [19] pointed out that firms are more likely to get loans and to receive lower interest rates on loans if social relationships and network ties exist.

2.3 Super-Network Theory

“Super” networks are networks that are “above and beyond” existing networks, which consist nodes, links, flows, with nodes corresponding to the locations in space, links to connections in forms of roads, cables, ext., and flows to vehicles, data, etc [20]. The super-network framework provides us with tools to study interrelated networks. Tools applied in the network framework includes: optimization theory, game theory, variational inequality theory, projected dynamical system theory, and network visualization tools.

Super-networks so far have been explored into many application areas such as transportation networks, telecommunication networks, economic and financial networks [20]. Also some specific applications of super-networks include: super-networks consisting of social networks interacting with supply chain networks, super-networks consisting of social networks interacting with financial networks, and knowledge super-networks [21].

3 The Super-Network Model Integrating Social Networks and Financial Networks for AML

3.1 Super-Network Model and the Multi-criteria Decision-Making Problem

In this section, we develop the super-network model consisting of the integration of the financial network and the social network.

The depiction of the super-network is given in Fig. 1. As shown in the figure, the super-network is comprised of the social network, which is at the left bottom side, and the financial network, which is the top level network. In financial network, flows are electronic financial transactions from one agent to another (each agent represent a personal customer or corporation), and in the figure are denoted by solid directed line. In the social network, flows are relationship levels between each pair of agents, and are denoted in the figure by solid lines. Subsequently, we describe the inter relationships between the financial and solid networks using dotted arcs.

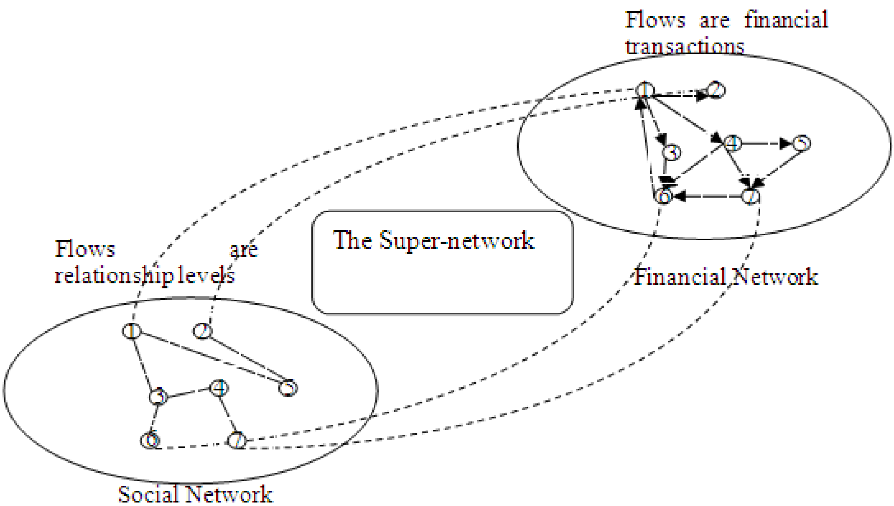


Fig. 1. The Super-network structure of the integrated Financial Network/Social Network System

Then we turn to the description of the behavior of economic decision-makers and discuss their multiple objective decision models.

Consider a cluster containing distinct n customers with a typical agent denoted by i and j. In social networks, we let h_{ij} denote the nonnegative level of the relationship between agent i and j. Relationship levels can take on a value from 0 to 1 where 0 means no relationship while 1 stands for the highest possible relationship. For each social relationship, there exists relationship value which we denoted by v_{ij}, and v_{ij} is assumed to be a function of the relationship level such that $v_{ij} = v_{ij}(h_{ij}), \forall i, j$. We assume that the value functions are continuously differentiable and concave. To establish relationship between pairs of agents, one needs to spend some cost, we further assume the production cost function of relationship level between agents are denoted by b_{ij}, $b_{ij} = b_{ij}(h_{ij}), \forall i, j$. It is convex and continuously differentiable.

In financial networks, we assume the quantity of financial funds transacted between agent i and j is denoted by q_{ij} (transacted from node j to node i); thus the overall volume agent i holds can be denoted by $q_i = q_i(q_{ij}) = q_0 + \sum_j q_{ij}, \forall i, j$; we further assume that the transaction cost between agent i and j is denoted by c_{ij} ,and it depends on the volume of financial transactions between the particular pair and on the relationship level between them, that is $c_{ij} = c_{ij}(h_{ij}, q_{ij}), \forall i, j$; Furthermore, it is also reasonable to assume there exists risk for each agent to deal with another agent, and we let a risk function for agent i dealing with j denoted by r_{ij} , this function depend not only on the quantity of the financial flow transacted between pair of nodes, but also on the corresponding relationship level, that is $r_{ij} = r_{ij}(h_{ij}, q_{ij}), \forall i, j$. Finally, using those financial funds, agent i can get profit denoted by $p_i = p_i(q_i) = p_i(q_{ij}), \forall i, j$.

Then for agent i , he faces 5 different decision-making problems as following:

1) Agent i tries to maximize his total volume of financial funds:

$$\max q_i = \max \sum_j q_i(q_{ij}) \tag{1}$$

2) Agent i tries to maximize the total value of his relationships expressed as:

$$\max \sum_j v_{ij} = \max \sum_j v_{ij}(h_{ij}) \tag{2}$$

3) He also tries to maximize his net revenue he faces the following maximization problem:

$$\max \sum_j p_i = \max \sum_j p_i(q_{ij}) \tag{3}$$

4) Agent i also faces an optimization problem associated with his desire to minimize the total risk:

$$\min \sum_j r_{ij} = \min \sum_j r_{ij}(h_{ij}, q_{ij}) \tag{4}$$

5) Agent i faces to minimize the total cost and corresponding to:

$$\min[\sum_j b_{ij} + \sum_j c_{ij}] = \min[\sum_j b_{ij}(h_{ij}) + \sum_j c_{ij}(h_{ij}, q_{ij})] \tag{5}$$

Then, we can now construct agent $i's$ multicriteria decision-making objective function and is denoted by $U(i)$ .Assume that agent i assigns separately 5 nonnegative weight $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ to the above (1)(2)(3)(4)(5) functions. Therefore, the multicriteria decision-making problem of source i can be expressed as:

$$\max U(i) = \max\{\alpha_1 \sum_j q_i(q_{ij}) + \alpha_2 \sum_j v_{ij}(h_{ij}) + \alpha_3 \sum_j p_i(q_{ij}) - \alpha_4 \sum_j r_{ij}(h_{ij}, q_{ij}) - \alpha_5 [\sum_j b_{ij}(h_{ij}) + \sum_j c_{ij}(h_{ij}, q_{ij})]\}, s.t \begin{cases} 0 \le h_{ij} \le 1 \\ 0 \le q_{ij} < q_i \end{cases} \tag{6}$$

3.2 The Equilibrium Conditions of the Super-Network Integrating Social Networks and Financial Networks

Based on variational inequality theory, the equilibrium state of the super-network is one where the relationship levels, financial flows coincide and satisfy the function (6). We now can establish the following theorem:

Theorem 1 (Variational Inequality Formation)**.** The equilibrium conditions governing the super-network integrating the financial network with the social network are equivalent to the solution of the variational inequality given by: determine $h^* = (h_{ij}^*) \in R_+^n, q^* = (q_{ij}^*) \in R_+^n$, such that $(h^*, q^*) \in K$, satisfying:

$$\begin{Bmatrix} \alpha_5 \sum_{i=1}^n \sum_{j=1}^n \left[\frac{\partial b_{ij}(h_{ij}^*)}{\partial h_{ij}} + \frac{\partial c_{ij}(h_{ij}^*, q_{ij}^*)}{\partial h_{ij}} \right] \\ + \alpha_4 \sum_{i=1}^n \sum_{j=1}^n \frac{\partial r_{ij}(h_{ij}^*, q_{ij}^*)}{\partial h_{ij}} \\ - \alpha_2 \sum_{i=1}^n \sum_{j=1}^n \frac{\partial v_{ij}(h_{ij}^*)}{\partial h_{ij}} \end{Bmatrix} * (h_{ij} - h_{ij}^*)$$
$$+ \begin{Bmatrix} \alpha_5 \sum_{i=1}^n \sum_{j=1}^n \frac{\partial c_{ij}(h_{ij}^*, q_{ij}^*)}{\partial q_{ij}} + \alpha_4 \sum_{i=1}^n \sum_{j=1}^n \frac{\partial r_{ij}(h_{ij}^*, q_{ij}^*)}{\partial q_{ij}} \\ - \alpha_1 \sum_{i=1}^n \sum_{j=1}^n \frac{\partial q_i(q_{ij}^*)}{\partial q_{ij}} - \alpha_3 \sum_{i=1}^n \sum_{j=1}^n \frac{\partial p_i(q_{ij}^*)}{\partial q_{ij}} \end{Bmatrix} * (q_{ij} - q_{ij}^*) \ge 0 \tag{7}$$

$\forall (h, q) \in K$

let

$$F_1(X) = \alpha_5 \sum_{i=1}^{n}\sum_{j=1}^{n}\left[\frac{\partial b_{ij}(h_{ij}^{*})}{\partial h_{ij}} + \frac{\partial c_{ij}(h_{ij}^{*}, q_{ij}^{*})}{\partial h_{ij}}\right] + \alpha_4 \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial r_{ij}(h_{ij}^{*}, q_{ij}^{*})}{\partial h_{ij}} - \alpha_2 \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial v_{ij}(h_{ij}^{*})}{\partial h_{ij}} \quad (8)$$

$$F_2(X) = \alpha_5 \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial c_{ij}(h_{ij}^{*}, q_{ij}^{*})}{\partial q_{ij}} + \alpha_4 \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial r_{ij}(h_{ij}^{*}, q_{ij}^{*})}{\partial q_{ij}} - \alpha_1 \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial q_{i}(q_{ij}^{*})}{\partial q_{ij}} - \alpha_3 \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial p_{i}(q_{ij}^{*})}{\partial q_{ij}} \quad (9)$$

Then the multicriteria decision-making model can be represented as:

$\left\langle F(X^{*}), X - X^{*}\right\rangle \geq 0,$ $X^{*} = (h^{*}, q^{*}), F(X) = (F_1(X), F_2(X))$,where $\langle , \rangle$ presents the inner product of vectors with same dimension.

The proof of Theorem 1 is similarly presented by [22], and because of the paper length limitation, here we omit it.

In Fig. 2, we display the super-network in equilibrium in which the equilibrium financial flows and relationship levels are shown. If the equilibrium values (either financial or relationship levels) on links are identically equal to 0, then those links can be removed from the equilibrium super-network. From Theorem 1 we can also conclude that the effect of relationship levels and transactions flows on cost, risk and revenue determined the equilibrium financial flows and relationships flows (based on (7),(8),(9)). To be more specific, for relationship level, the weighted marginal cost for relationship establishment ($\alpha_5 \sum_{i=1}^{n}\sum_{j=1}^{n}\left[\frac{\partial b_{ij}(h_{ij}^{*})}{\partial h_{ij}} + \frac{\partial c_{ij}(h_{ij}^{*}, q_{ij}^{*})}{\partial h_{ij}}\right]$), the weighted marginal cost for risk reduction ($\alpha_4 \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial r_{ij}(h_{ij}^{*}, q_{ij}^{*})}{\partial h_{ij}}$), and the weighted marginal value of relationship ($\alpha_2 \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial v_{ij}(h_{ij}^{*})}{\partial h_{ij}}$) determined the distribution of social network flows. Similarly in financial transaction network, factors such as the weighted marginal cost for transaction acquirement ($\alpha_5 \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial c_{ij}(h_{ij}^*, q_{ij}^*)}{\partial q_{ij}}$), the weighted marginal cost for risk reduction ($\alpha_4 \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial r_{ij}(h_{ij}^*, q_{ij}^*)}{\partial q_{ij}}$), the weighted marginal transaction flow acquirement ($\alpha_1 \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial q_i(q_{ij}^*)}{\partial q_{ij}}$), the weighted marginal revenue for transaction transferring ($\alpha_3 \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial p_i(q_{ij}^*)}{\partial q_{ij}}$) jointly determined the distribution of equilibrium transaction flows.

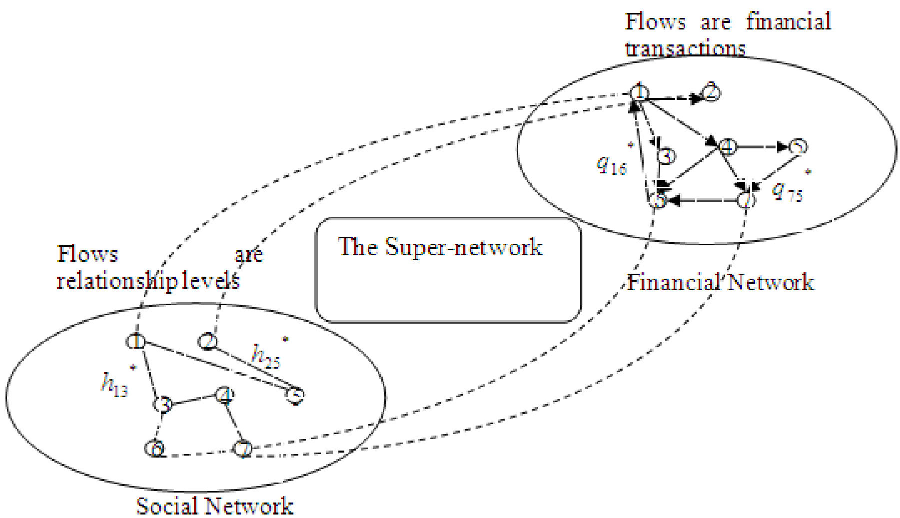


Fig. 2. The Super-network at Equilibirum

3.3 The Solutions to the Variational Inequality Using Projected Dynamical System

In this section, we use Euler method for the computation of solutions to variational inequality (7). It is a special method for the solution of projected dynamic systems and this algorithm not only provides a solution to variational inequality problem, not also yields a time discretization of the continuous-time adjustment process of the projected dynamic system.

Following the Euler method, there are several steps to complete the computation of the solution:

Step 0: Initialization

Set $X^0 \in K$, let Γ denote an iteration counter and set $\Gamma=1$. Set the sequence $\{\alpha_\Gamma\}$ with $\alpha_\Gamma > 0$ for all Γ and $\sum_{\Gamma=1}^{\infty} \alpha_\Gamma = \infty$, as $\Gamma \to \infty$ (which is a requirement for convergence).

Step 1: Computation

Compute $X^\Gamma \in K$ by solving the variational inequality subproblem:

$$\left\langle X^\Gamma + \alpha_\Gamma F(X^{\Gamma-1}) - X^{\Gamma-1}, X - X^\Gamma \right\rangle \geq 0, \forall X = (h,q) \in K$$

Step 2: Convergence Verification

If $\left| X^\Gamma - X^{\Gamma-1} \right| \leq \varepsilon$, with $\varepsilon > 0$, a prespecified tolerance, then stop; otherwise, set $\Gamma = \Gamma + 1$, and then stop; otherwise, go to Step 1.

Convergence results for the Euler method can be found by details in paper written by [20].

4 Suspicious Detection Using Equilibrium State

Using the equilibrium state, we can further think about how to identify suspicious transactions or suspicious accounts. Consider those clusters formed by different persons or corporations, basically, agents with more closer relationship levels, for example, companies in the same supply chain, persons with blood relations and so on, are more likely to have frequent transactions between them, while companies with barely no business deals or persons in different cities would hardly have chance to have any trading. Thus, those agents with high relationship levels in social network, will be also with high homogenous, and vice versa. Thus, when financial transaction networks achieves to be a corresponding pattern compared with social networks, the super-network become at equilibrium. As for suspicious client detection, taking a client denoted as i, first we can extract and analyze its surroundings i agents for both social network and transaction network, explicate relationship levels as ($h_{i1}, h_{i2}, \ldots h_{ij}$) and transaction flow vectors as ($q_{i1}, q_{i2}, \ldots q_{ij}$); second we calculate the super-network equilibrium as represented as ($h_{i1}^*, h_{i2}^*, \ldots h_{ij}^*$) and ($q_{i1}^*, q_{i2}^*, \ldots q_{ij}^*$); thirdly, given a meaningful threshold $r(r \in R_+^n)$, for those $k, k \leq j$,where $\left(\frac{q_{ik} / (\sum_k q_{ik} / j)}{h_{ik} / (\sum_k h_{ik} / j)}\right) / \left(\frac{q_{ik}^* / (\sum_k q_{ik}^* / j)}{h_{ik}^* / (\sum_k h_{ik}^* / j)}\right) > r$, then q_{ik} are highly differentiated from normal patterns. Similarly, we can judge suspicious accounts.

5 Summary and Conclusion

In this paper, we considered the interrelationship between social network and financial networks, and proposed a super-network based model to construct a new model integrating this two kinds of network. By explicating the relationship levels for pairs of clients or agents, we further analyzed the multicriteria decision-making behavior for each clients or agents, including maximizing its total transaction amount, maximizing the total relationship value, maximizing its total net revenue, minimizing its total risk and total cost, using the variational inequality theory, we concludes the super-network in equilibrium, in which the financial flows and the relationship levels coincide and established the variational inequality formulation . Then we discussed how to use the proposed model in suspicious activity detection. In the near future, it would be interesting to apply the theories and models herein to actual AML cases.

Acknowledgment. This research was supported by the National Natural Science Foundation of China under Grant 71101053 and Chinese Ministry of Education Grant 10YJCZH097.

References

1. Camdessus, M.: Money Laundering: The importance of international countermeasures. In: IMF to the Plenary Meeting of the Financial Action Task Force on Money Laundering, Paris, p. 2 (1998)
2. People's Bank of China.: Administrative rules for the reporting of large-value and suspicious activity reports (2003)
3. People's Bank of China.: Administrative rules for the reporting of large-value and suspicious activity reports (2006)
4. Liu, X., Zhang, P.Z.: Research on Constraints in anti-money laundering (AML) business process in China based on theory of constraints. In: IEEE Conference, Proceedings of the 41th Hawaii International Conference on System Sciences (2008)
5. Schneider, S.: Organized crime, money laundering, and the real estate market in Canada. Journal of Property Research 21(2), 99–118 (2004)
6. Goldberg, H., Senator, T.E.: Restructuring databases for knowledge discovery by consolidation and link formation. In: Proceedings of 1998 AAAI Fall Symposium on Artificial Intelligence and Link Analysis (1998)
7. Senator, T.E., et al.: The financial crimes enforcement network AI system(FAIS): identifying potential money laundering from reports of large cash transactions. The AI Magazine 16(4), 21–39 (1995)
8. Tang, J., Yin, J.: Developing an Intelligent Data Discriminating System of Anti-Money Laundering Based on SVM. In: Proceedings of 2005 International Conference on Machine Learning and Cybernetics (June 2005)
9. Wang, S.N., Yang, J.G.: A Money Laundering Risk Evaluation Method Based on Decision Tree. In: 2007 International Conference on Machine Learning and Cybernetics (January 2007)

10. Liu, X., Zhang, P.Z.: A Scan Statistics based suspicious transactions detection model for Anti-Money Laundering (AML) in financial institutions. In: IEEE Conference, 2010 International Conference on Multimedia Communications (Mediacom 2010) (2010)
11. Zhu, T.: Suspicious Financial Transaction Detection Based on Empirical Mode Decomposition Method. In: IEEE Asia-Pacific Conference on Services Computing, APSCC 2006, pp. 300–304 (2006a)
12. Zhu, T.: An Outlier Detection Model Based on Cross Datasets Comparison for Financial Surveillance. In: IEEE Asia-Pacific Conference on Services Computing, APSCC 2006, pp. 2006–601 (2006b)
13. Liu, X., Zhang, P.Z.: Suspicious subgroups recognition in financial networks- A hidden group based model for anti-money laundering. In: IEEE Conference on International Colloquium on Computing, Communication, Control and Management, CCCM 2010 (2010)
14. Xue, Y.W., Zhang, P.Z., Fan, J.: Research on criteria for identifying abnormal capital flows in financial networks. China Soft Science (9), 57–62 (2004) (in Chinese)
15. Ghatak, M.: Exploiting social networks to alternative credit market failures: on the endogenous selection of peer groups in microfinance programs. In: Conference on Credit, Truest and Calculation at the University of California (2002)
16. Berger, A.N., Udell, G.F.: Relationship lending and lines of credit in small firm finance. The Journal of Business 68, 351–381 (1995)
17. Jackson, M.O.: A survey of models of network formation: Stability and efficiency. In: HSS, pp. 228–277. California Institute of Technology, Pasadena (2003), `http://www.grandcoalition.com/papaers/jackson_4.pdf`
18. DiMaggio, P., Louch, H.: Socially embedded consumer transaction: For what kinds of purchases do people most often use networks. American Sociological Review 63, 619–637 (1998)
19. Uzzi, B.: Embeddedness in marketing of financial capital: how social relations and networks benefit firms seeking financing. American Sociological Review 64, 481–505 (1999)
20. Nagurney, A., Dong, J.: Supernetworks: Decision-making for the information Age. Edward Publishing, Elgar (2002) ISBN:1840649682
21. Nagurney, A., Wakolbiinger, T.: Supernetworks: An introduction to the concept and its applications with a specific focus on knowledge supernetworks (2005), `http://supernet.som.umass.edu/articles/uksupernetworks.pdf`
22. Anna, N., Zhang, D.: Projected dynamical systems and variational inequalities with applications. Kluwer Academic Publishers, Boston (1996)

On the Volatility of Online Ratings: An Empirical Study

Christopher S. Leberknight, Soumya Sen, and Mung Chiang

Princeton University, Department of Electrical Engineering, Olden Street,
Princeton, NJ 08544
{csl,soumyas,chiangm}@princeton.edu

Abstract. Many online rating systems represent product quality using metrics such as the mean and the distribution of ratings. However, the mean usually becomes stable as reviews accumulate, and consequently, it does not reflect the trend emerging from the latest user ratings. Additionally, understanding whether any variation in the trend is truly significant requires accounting for the volatility of the product's rating history. Developing better rating aggregation techniques should focus on quantifying the volatility in ratings to appropriately weight or discount older ratings. We present a theoretical model based on stock market metrics, known as the Average Rating Volatility (ARV), which captures the fluctuation present in these ratings. Next, ARV is mapped to the discounting factor for weighting (aging) past ratings and used as the coefficient in Brown's Simple Exponential Smoothing to produce an aggregate mean rating. This proposed method represents the "true" quality of a product more accurately because it accounts for both volatility and trend in the product's rating history. Empirical findings on rating volatility for several product categories using data from Amazon further motivate the need and applicability of the proposed methodology.

Keywords: Consumer confidence, e-commerce, decision support, online ratings, reputation systems.

1 Introduction

The effect of the current downturn of the U.S. Economy on consumer confidence can be observed by a close examination of the new shape in the demand and utility of e-commerce systems. Consumers are, now more than ever, looking for ways to minimize their expenses and e-commerce has moved from novelty and convenience to necessity. This has resulted in the growth of new technologies and revenue from applications such as mobile commerce (m-commerce) and location-based services. The combined revenue of e-commerce and m-commerce accounts for a large volume of the overall online retail market. Forrester Research estimates that both US and European online retail (representing 17 Western European nations) will grow at a 10 percent compound annual growth rate from 2010 to 2015, reaching $279 billion and €134 billion, respectively, in 2015.

Many e-commerce services like eBay and Amazon, as well as some of the newer online shopping applications, such as Yelp and Groupon, all share similar

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 77–86, 2012.

characteristics and advantages over conventional retail stores in providing greater convenience, choice, and customer feedback. Almost all e-commerce systems offer potential buyers with easy access to product ratings from other buyers. The widely held belief is that virtual trust in these systems can be achieved when many consumers provide a similar rating for the product, independent of each other. This concept has been captured in phrases such as "wisdom of the crowd" which suggests that wisdom, and in this particular case quality, can be determined based on how many consumers share the same rating for the product. However, online ratings suffer from subjectivity bias and the arithmetic mean rating is often not the most accurate metric to base purchase decisions. Hu et al., [12] reported that online rating distribution for most products are bimodal (J-shaped) because of the "brag-and-moan" phenomenon among reviewers. Recent research also suggests other aggregators such as the weighted mean, median and mode may portray a more accurate picture of product quality [1].

While pursuing the identification of robust metric is undoubtedly useful, our focus in this paper is to understand and quantify the extent of volatility inherent in the user ratings of current systems. Since product ratings may be one of the strongest predictors of a consumer's decision to purchase a product online, providing such additional information on average volatility and latest rating trends can significantly influence consumer confidence. Using quantitative ideas from the stock market, we introduce a model to estimate rating volatility and use it to analyze an Amazon dataset with about 16,500 ratings across product categories of movies, books, fashion, and electronic goods. We examine the volatility in ratings that potential buyers of such goods are exposed to as they browse through the ratings from other reviewers, and then propose a way to potentially improve consumer's perception of a product's true quality.

2 Related Work

Research on online ratings has primarily focused on understanding various economic and social aspects, such as the impact of online ratings on sales [4],[6],[7],[11] and methods to increase trust and reputation [8],[10].

Besides the economic and social dimensions, researchers have also worked on the developing rating systems based on new quantitative metrics that may provide deeper insight and credibility for consumer ratings. Jøsang and Ismail [9] proposed a new rating metric called the beta reputation system that uses beta probability density functions to combine feedback and derive reputation scores. They also demonstrate an aging mechanism to account for the weighted difference between old and new reviews. Optimization techniques for new rating metrics have also been proposed [3]. In addition to the investigation of new metrics, other researchers have suggested simple transformations for displaying the quality of a given product. While rating feedback on product quality is most commonly aggregated using the arithmetic mean, recent research suggests other aggregators such as the weighted mean, median and mode may portray a more accurate picture of product quality [1]. Utilizing a different aggregator such as the median was further substantiated through the examination of feedback bias [2].

Our paper complements these earlier works by empirically studying the volatility of ratings for several categories of products sold on Amazon. Such an approach can help in enhancing the reliability of existing rating systems as well as improving consumer confidence through better representation of the rating data.

3 Model and Empirical Results

Many online rating systems today use a naïve averaging method for its product ratings (not to be confused with ranking), the simplest of which is an equally weighted average. But this is not necessarily a good metric for products that exhibit some clear trends. For example, consider the diagram shown on the left side of Fig. 1. It shows that there is a clear trend that the new ratings for the product are higher than the older ratings, possibly due to a substantial quality improvement. But averaging out the ratings for this product will fail to inform the buyers about this trend. Consequently, one could argue that discounting the older ratings (i.e. aging) to put greater weightage on the newer ratings would help to account for this latest trend. While that inference is correct, it is not clear whether aggressively discounting older ratings is always the right approach; for example, consider the product rating pattern shown on the right side of Fig. 1. The ratings show a high variability in the user's perception of the product's quality (e.g., because of some undetected flaw in the product). Therefore, in this case the older ratings should be given roughly equal weights in averaging instead of being discounted away. Hence, estimating a product's true rating would require a compromise between a simple averaging and a random walk model.

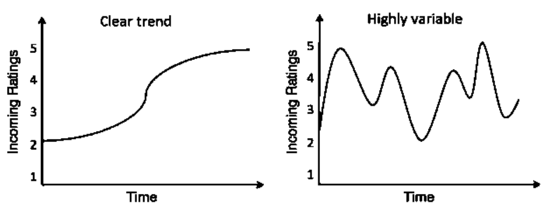


Fig. 1. (a) Left: ratings show a clear trend of improvement, (b) Right: rating is highly variable over time

In order to develop a prediction methodology that balances the need to account for variability in the ratings and the latest trends, we devise a three-stage strategy: First, we propose a metric called Average Rating Volatility (ARV) that captures the extent of fluctuation present in the ratings. Second, we define a map between ARV and the discounting (aging) factor to be used in weighting the past ratings. Third, we use this discounting factor as the coefficient for Brown's Simple Exponential Smoothing to predict the mean rating that we believe is more 'accurate', given that it accounts for both past volatility and the latest trend. These steps are described next followed by an evaluation of the proposed rating strategy on the data gathered from Amazon for four different products categories (movies, books, cameras, and shoes).

3.1 Average Rating Volatility (ARV)

We introduce Average Rating Volatility (ARV) as a measure of the amount of variability inherent in the ratings that a product receives over time. It is calculated based on the mean rating value over non-overlapping intervals (windows) of *N* consecutive ratings. A time series of the product's rating fluctuation is thus obtained as the window moves. The counterpart to the idea of ARV in the stock market [5] is known as the Volatility Index, denoted by VIX or AIV.

In this work, we compute the ARV based on a batch of 10 consecutive ratings, i.e. the window size *N* is set to 10. This size was chosen because most online portals and Amazon in particular, present viewers with a batch of 10 rated reviews on each page. Moreover, site visitors can view these ratings listed according to "newest first" option, in which case, the AVR metric reported here can be interpreted as a direct quantification of the volatility in ratings that a potential buyer is exposed to as he/she moves from one page to the other. However, it should be noted that this approach also accommodates the use of other review window sizes. Choosing the correct window size is a non-trivial task; one direction to explore in future is to choose the window size that gives the highest correlation between the ARV and the standard deviation index of the rating volatility.

Let $R(t) = \{r_t, r_{t-1}, \ldots, r_0\}$ denote the original time series of a product's rating scores, and this series is divided into *M* non-overlapping time windows of size *N*, denoted by $W_1, W_2, \ldots W_M$ (*i.e.* $|W_i| = N$). The Average Rating Volatility is then defined as the fractional change in the average rating values of two consecutive time windows of size *N*:

$$ARV(W_i) = \frac{|\langle R(W_{i+1})\rangle - \langle R(W_i)\rangle|}{\langle R(W_i)\rangle} \;;\; i \in \{1, M\}. \tag{1}$$

where $\langle R(W_i)\rangle$ is the average index value in window W_i, which is given by

$$\langle R(W_i)\rangle = \sum\nolimits_{k \in W_i} \theta_k r_k \quad ; \sum\nolimits_{k \in W_i} \theta_k = 1 \tag{2}$$

θ_k's are the weights assigned to the ratings in window W_i, which in the simplest case is set to $\theta_k = 1/N$. A more sophisticated approach would entail setting the weights in accordance with the review of ratings (*e.g.* "helpful"/ "not helpful" comments left by other reviewers). The absolute value of the difference is used in (1) to capture the true magnitude of shift in public perception, which often shows a synchronized shift in one direction, upward or downward, depending on a sudden discovery of product flaw, introduction of rival products with comparable features etc. The ARV values thus give another time series, the area under which is a measure of the latent variability in the rating time series. But this area depends on the length of the time series data; hence we normalize it with respect to the total number of points in the ARV time series by computing the mean across all ARV windows. This measure, denoted as $\overline{ARV}$, represents the volatility of the product's ratings. For example, the ARV of the 3rd generation iPod constructed from 1907 rating scores on Amazon is shown in Fig. 2. It is interesting to note that the volatility in the ratings have remained almost the same over time.

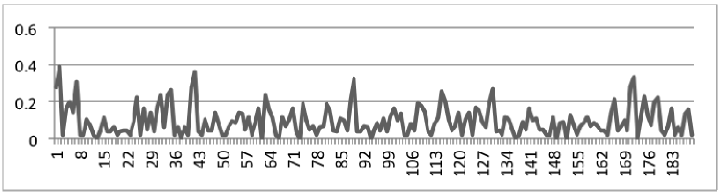


Fig. 2. Average rating volatility (y-axis) of iPod3 (32 GB) for window size of 10 based on 1,908 ratings (x-axis) from Amazon between 9/2009-12/2010

The ARV for the category of commercially successful movies around 1999 is shown in Table 1. The category $\overline{ARV}$, denoted by λ is computed to be 0.128, *i.e.* products of this category show on an average about 12.8% deviation from the mean rating of one page to the next on Amazon. Similar results on the average volatility in books, digital cameras, and women's shoes are reported in Tables 2, 3, and 4 respectively.

Table 1. $\overline{ARV}$ computed from 12,808 ratings across 10 commercially successful movies, with scores received on their Amazon DVDs between the release date and June 2005

Movies (Entertainment)	$\overline{ARV}$	**Number of ratings**
1. Independence Day (1996)	0.203208173	561
2. Titanic (1997)	0.172611210	1816
3. Saving Private Ryan (1998)	0.098727993	1196
4. The Big Lebowski (1998)	0.072623118	516
5. Armageddon (1999)	0.217262387	1171
6. American Beauty (1999)	0.124428681	1061
7. Fight Club (1999)	0.081965368	1276
8. The Matrix (1999)	0.094605664	2926
9. The Mummy (1999)	0.113956289	771
10. Gladiator (2000)	0.099715611	1514
Genre/Category, λ =	0.128 (12.8%)	Total = 12,808

Table 2. $\overline{ARV}$ for 5 critically acclaimed books computed between 1996 and 2005 using Amazon's rating data

Books (Arts & Culture)	$\overline{ARV}$	Number of ratings
1. One hundred years of solitude (G. G. Marquez)	0.095297574	424
2. Beloved (Toni Morrison)	0.146078200	557
3. Blindness (Jose Saramago)	0.098140184	251
4. Lolita (Vladimir Nabokov)	0.090344492	415
5. Satanic Verses (Salman Rushdie)	0.149658322	181
Genre/Category, λ =	0.116 (11.6%)	Total = 1,828

Table 3. $\overline{ARV}$ for a group of digital cameras computed between 2008 and 2011 using Amazon's rating data

Digital Cameras (Electronics)	$\overline{ARV}$	Number of ratings
1. Canon PowerShot SX20IS 12.1MP	0.087368	391
2. Nikon S8000-vibration reduction	0.178723893	552
3. Polaroid CZA-10011P	0.106465356	175
Genre/Category, λ =	0.124 (12.4%)	Total = 1,118

Table 4. $\overline{ARV}$ for popular women's shoe brands based on Amazon's rating data between 2006 and 2011

Women's Shoes (Fashion)	$\overline{ARV}$	Number of ratings
1. Easy Spirit Traveltime	0.056502532	150
2. UGG Classic Footwear	0.047895357	148
3. BearPaw Shearling Boots	0.086294498	201
4. Skechers Shape-Ups	0.12181668	186
5. Tamarac Slippers	0.121079089	120
Genre/Category, λ =	0.087 (8.7%)	Total = 805

3.2 Discounting Factor Function (DFF)

We now introduce a function that maps $\overline{ARV}$ (which provides a measure of the latent volatility in the product's ratings) to the coefficient with which older ratings need to be discounted. This coefficient is also referred to as the discounting (aging) factor, $0 \le \alpha \le 1$. A small value of α puts greater weights on past ratings while $\alpha \approx 1$ puts more the weight on the current rating in making future predictions.

When $\overline{ARV}$ is low (*i.e.,* similar average ratings across time windows), then the discounting factor should be kept large so that more weight is put on the current values, thus any sudden deviation in the trend gets reflected quickly. On the other hand, if $\overline{ARV}$ is high, (*i.e.,* the product's review are volatile), then the discounting factor is kept relatively low to average out any sudden fluctuations, thus responding only when there is a real trend. As shown in Figure 3, we choose a shifted sigmoid function for this mapping between the discounting factor and the average volatility in a product's rating because it satisfies the above properties and makes the discounting factor steadily sensitive to changes around the product category's $\overline{\text{ARV}}$.

$$\alpha = \frac{1}{1 + e^{\eta(\overline{\overline{ARV}} - \lambda)}} \tag{3}$$

where λ is the product's category $\overline{ARV}$ and can be computed as shown in Tables 1-4 of Section 3.1. The parameter η is a scaling constant to make α close to 1 when $\overline{ARV} \simeq 0$ and can be used to also control the sensitivity of the discounting factor to the rating volatility. The exact value of η may be fine-tuned depending on the consumer's price sensitivity of the product category.

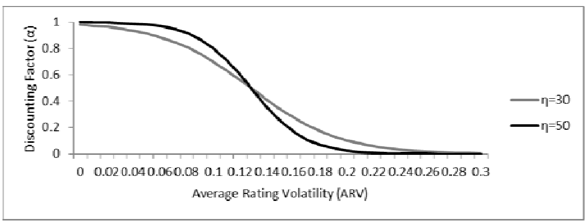


Fig. 3. Plot of aging factor versus (ARV) for λ=0.128

3.3 Simple Exponential Smoothing (SES)

SES is a method that discounts past data in a gradual fashion. For example, it ensures that the most recent rating gets a little more weight than 2nd most recent, and the 2nd most recent gets more weight than the 3rd most recent, and so on. This discounting is done by a "smoothing coefficient", which is the aging factor, α introduced in Section 3.2. The formula used by SES recursively to update the smoothed series as new ratings come in is given by:

$$R_t = \propto r_t + (1 - \propto)\, R_{t-1}; \qquad 0 \leq \propto \leq 1 \tag{4}$$

$$r_{t+1} = R_t;\; R_1 = r_0$$

Thus, the current smoothed value is an interpolation between the previous smoothed value and the current observation, where $\propto$ controls the closeness of the interpolated value to the most recent observation. Notice that if $\propto = 1$ the SES model is equivalent to a random walk model (without growth), while $\propto = 0$ makes it equivalent to a mean model, assuming that the first smoothed value is set equal to the mean. The average

age of the rating data in the SES forecast is $1/\propto$, relative to the period for which the rating prediction is made. For example, when $\propto = 0.2$ the lag is 5 periods etc.

To summarize, our proposed methodology works as follows: for a given rating time series we first compute the mean ARV ($\overline{ARV}$), which is mapped to a discounting factor that is used as the coefficient for the SES method to predict a better aggregate rating. Next this methodology is applied to two online products as shown in Fig. 4 and 5. In Figure 4, the evolution of Amazon's mean rating is depicted (solid black line) for the DVD of Armageddon, which does not reflect the dynamic changes in the real ratings from the users. By accounting for the high volatility in user ratings, a new ratings evolution (dashed black line) is constructed using the proposed approach. It can be observed that this ratings evolution is more responsive to changing user preferences while at the same time avoiding random short-term fluctuation trends that arise from user's subjectivity.

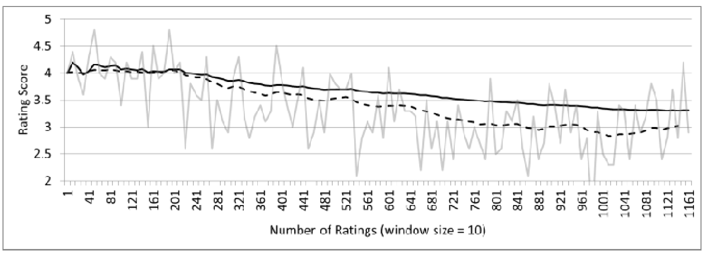


Fig. 4. Armageddon (11/1998-6/2005, 1171 ratings). Amazon's rating score is slow to change as it neglects rating volatility (Solid black: Evolution of the mean rating on Amazon, Grey: Evolution of the actual rating from batches of 10 consecutive ratings, Dashed black: Estimate of true rating that account for volatility using the proposed ARV approach).

In Fig. 5, the ratings for G. G. Marquez's famous novel lie in a much narrower range of values and hence have a low volatility in contrast to the movie Armageddon.

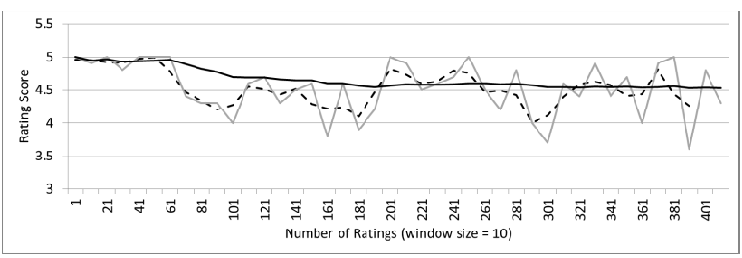


Fig. 5. G. G. Marquez's book One Hundred Years of Solitude (424 ratings). Amazon's rating score does not reveal the user's true score (Solid black: Evolution of the mean rating on Amazon, stabilizes at 4.5, Grey: Evolution of the actual rating from batches of 10 consecutive ratings, Dashed black: Estimate of correct rating while accounting for volatility using our ARV approach; the score goes down to 4 but users never get to see it with simple averaging methods).

Therefore, the prediction for the aggregate rating mostly follows the latest trend and is therefore more responsive to changes in the readership's taste or critical evaluation of the book in the larger social context. Similar results were observed for other categories of products that further strengthen the case for using this AVR-based rating strategy for online rating systems.

4 Conclusion and Future Work

This paper presents a theoretical model for a new rating mechanism which may provide greater insight into the reliability of existing ratings used to convey product quality. The three-tiered strategy of this model attempts to address issues of stabilization, prediction, and discounting. First, we introduce a metric called Average Rating Volatility (ARV) that captures the extent of fluctuation present in the ratings. Second, we define a map between ARV and the discounting (aging) factor to be used in weighting the past ratings. Third, the discounting factor is used as the coefficient for Brown's Simple Exponential Smoothing to predict the mean rating and future values. Lastly, the proposed model is evaluated using data gathered for different types of products from Amazon. A rating mechanism based on this model will provide a more accurate representation of "true" quality of a given product since it will account for both rating volatility and the emergence of long-term trends.

The step-by-step approach presented here towards creating a review window based aggregation metric is only an initial attempt at capturing the potential impact of volatility in the rating aggregation. Extending this study along the dimensions of aggregation granularity, stabilization, and rating prediction can be explored as interesting directions for future work on this study.

Acknowledgments. This work was supported in part by NSF CNS-0905086, CNS-1117126, and ARO W911NF-11-1-0036.

References

1. Garcin, F., Faltings, B., Jurca, R.: Aggregating reputation feedback. In: Proceedings of the International Conference on Reputation: Theory and Technology, vol. 1(1), pp. 62–74 (2009)
2. Jurca, R., Garcin, F., Talwar, A., Faltings, B.: Reporting incentives and biases in online review forums. ACM Transactions on the Web (TWEB) 4(2), 1–27 (2010)
3. De Kerchove, C., Van Dooren, P.: Reputation Systems and Optimization. SIAM News 41(2) (2008)
4. Zhang, X.M., Dellarocas, C.: The lord of the ratings: How a movie's fate is influenced by reviews. In: Proceedings of the 27th International Conference on Information Systems, ICIS (2006)
5. Liu, J., Tse, C.K., He, K.: Fierce stock market fluctuation disrupts scale-free distribution. Quantitative Finance 11(6), 817–823 (2011)
6. Bolton, G.E., Katok, E., Ockenfels, A.: How effective are on-line reputation mechanisms? An experimental investigation. Management Science 50(11), 1587–1602 (2004)

7. Duan, W., Bin, G., Whinston, A.B.: Do online reviews matter? An empirical investigation of panel data. Decision Support Systems (2008)
8. Resnick, P., Zeckhauser, R., Friedman, E., Kuwabara, K.: Reputation systems. Communications of the ACM 43, 45–48 (2000)
9. Jøsang, A., Ismail, R.: The beta reputation system. In: Proceedings of the 15th Bled Conference on Electronic Commerce (2002)
10. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decision Support Systems 43(2), 618–644 (2007)
11. Hu, N., Pavlou, P.A., Zhang, J.: Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of Online word-of-mouth communication. In: Proceedings of the 7th ACM Conference on Electronic Commerce, EC (2006)
12. Hu, N., Zhang, J., Pavlou, P.A.: Overcoming the J-shaped distribution of product reviews. Communications of the ACM 52(10), 144–147 (2009)

Supporting Patent Maintenance Decision: A Data Mining Approach

Chih-Ping Wei, Hung-Chen Chen, Ching-Tun Chang, and Yen-Ming Chu

Dept. of Info. Management, National Taiwan University, Taipei, Taiwan, R.O.C.
cpwei@im.ntu.edu.tw
Dept. of Info. Management, National Taiwan University, Taipei, Taiwan, R.O.C.
JesseHCCHen@gmail.com
Inst. of Service Science, National Tsing Hua University Hsinchu, Taiwan, R.O.C.
luobodun@gmail.com
Inst. of Tech. Management, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.
richardju@mxic.com.tw

Abstract. Nowadays, patents become much more important for companies to protect their rights and intellectual assets under the keen competitive business environments. However, it is not free for a granted patent. In the patent systems of many countries, a patent holder is required to pay a maintenance fee after the initial application to retain patent protection on his/her invention until the expiration of the protection period. Because not all the patents are worth maintaining by patent holders, firms and organizations need to identify "important patents" for maintenance and abandon "unimportant patents" to avoid unnecessary patent maintenance costs. In this paper, we employ the variables suggested by prior studies that would discriminate renewed patents from those abandoned ones and then take the data mining approach to construct a prediction model(s) on the basis of these variables for supporting patent maintenance decisions. Such a data-mining-based patent maintenance decision support system can help firms and organizations improve the effectiveness of their patent maintenance decisions and, at the same time, decrease the cost of their patent maintenance decisions. Our empirical results indicate that the effectiveness of our proposed system is satisfactory and practical for supporting patent maintenance decisions.

Keywords: Patent maintenance, Patent renewal, Data mining, Patent analysis.

1 Introduction

The role of intellectual assets (IA) is becoming more and more significant in the era of knowledge economy. IA includes traditional assets such as patents, copyrights, trademarks, trade secrets, and domain names, as well as certain tacit knowledge like know-how, business practices, and customer preferences [22]. Compared with other IAs, patents are more tangible to measure, and the importance of patents is much more significant in terms of commercial pursuits and technology.

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 87–99, 2012.

The basic function of patents is to provide the granted right to an inventor for a certain period of time (USPTO).[1] Protecting the rights of inventors is not the only purpose of patents. Its significance in technology management, knowledge management, R&D portfolio management, strategic planning, human resource management, and merger and acquisition decision support is also vital [7].

In the patent systems of many countries, patent holders are required to pay periodic renewal (or maintenance) fees until the expiration of the protection period in order to keep their patents in force. For example, in the United States, the term of protection is 20 years from application, and it is necessary for a patent holder to pay a maintenance fee every four years for the first twelve years. Furthermore, the maintenance fee increases over the maintenance period. For example, the charging criterion for a large entity is US$980 in between the 3.5th and the 4th year, $2,480 in between the 7.5th and the 8th year, and $4,110 in the final 11.5th and the 12th year. Table 1 shows the details of maintenance fees in the United States. As shown, a small entity in U.S. needs to pay a total of $3,785 to ensure a full-term legal right and a large entity has to pay a total of $7,570 to secure the same result.

Table 1. Patent Maintenance Fees in U.S

	Due at 3.5 years	**Due at 7.5 years**	**Due at 11.5 years**
Small Entity[2]	$490	$1240	$2055
Large Entity	$980	$2480	$4110

Source: USPTO

Table 2. Patent Renewal Rates

Patent Renewal Rates[3]	**2005**	**2006**	**2007**	**2008**	**2009**
First Stage	83.1%	93.1%	90.1%	83.1%	80.3%
Second Stage	65.4%	69.2%	71.4%	73.7%	65.3%
Third Stage	45.0%	44.4%	48.5%	49.2%	45.4%

Source: USPTO 2009 Performance and Accountability Report

The patent renewal rate from the USPTO 2009 Performance and Accountability Report is shown in Table 2. It shows the percentage of patent renewal rate in the most recent five years. For example, in 2009, 80.3% of the patents issued three years earlier, 65.3% of the patents issued seven years earlier, and 45.4% of the patents issued 11 years earlier were renewed. On the other hand, the maintenance rate could differ from industry to industry. In countries such as Finland and Norway, prior studies show that the maintenance rate is higher in chemical industry than in heavy industry and low-tech

[1] Available at `http://www.uspto.gov/`

[2] Small entity: person, small business concern, nonprofit organization, license to a federal agency, security interest.

[3] The first stage refers to the end of the 3rd year after the initial patent is issued; the second stage refers to the end of the 7th year after the initial patent is issued; and the third stage refers to the end of the 11th year after the initial patent is issued.

industry [16]. In a 1991 U.S. study, patents in the category of "Other Industries" had the lowest renewal rate of 16.34%, while patents in the industry of Computer and Communications had the highest renewal rate of 53.24% [3].

There are several factors that would influence patent maintenance decisions. The context of patent maintenance provides an interesting window through which we could explore how the patent holder weighs "important" or "unimportant" factors in determining whether to maintain a patent or not. In general, the decision of patent renewal might be influenced by the high cost of the maintenance fee [8] or the limited resources of firms. Patent holder thinks some patents might not bring revenues in the future or the expected expenditures might exceed the expected incomes. Abandoning these valueless patents would reduce unnecessary maintenance costs. On the other hand, some companies take patents as their knowledge assets and their proper management will effectively improve their returns. For example, Dow Chemical saved $4 million during the first year of its knowledge management initiative and it is expecting to generate more than $100 million in licensing revenues in the future that might otherwise have forgone. Companies such as IBM (with 3,000% increase in revenues from licensing its intellectual capital), Microsoft, and Texas Instruments demonstrated the power of licensing to improve the financial performance by managing and carefully reviewing the patents [11].

Moreover, companies may decide to maintain their patents to gain the protection from their patent rights, despite that they might not bring revenues in the future. In fact, patents become powerfully competitive tools with strategic functions to litigate their competitors. In response, some patents are therefore maintained to protect the company itself from losing money from litigations and gain advantage in negotiations [5], [15], [19]. Furthermore, companies may need to consider their R&D and knowledge management strategies when making their decisions on patent maintenance. To determine whether a patent is worth to be maintained, the company cannot assess a single patent individually, but needs to take the entire relations of all of its patents into consideration. For example, the renewal probability is higher for those patents that belong to internal sequential innovation than those patents of stand-alone innovation [13].

However, patent maintenance decision making is knowledge-intensive and its process is time-consuming, especially for firms and organizations holding numerous patents. Prior research related to patent maintenance mainly concentrates on comparing the characteristics of renewed patents with those of abandoned ones. In addition, existing studies pay much less research attention on the development of an automated system capable of supporting patent maintenance decisions. In response, in this study, we attempt to exploit the characteristics suggested by prior studies that potentially differentiate renewed patents from abandoned ones as the predictors (i.e., independent variables) and adopt a data mining approach to develop a patent maintenance decision support system in order to help firms and organizations improve the effectiveness and, at the same time, decrease the costs of their patent maintenance decisions.

The remainder of this paper is organized as follows. Section 2 reviews the literature related to this study. In Section 3, we describe our design of the data-mining-based patent maintenance decision support system, including the predictors employed and the detailed process and the design of the proposed system.

Subsequently, we report our empirical evaluation, including data collection, experiment design, and important evaluation results in Section 4. Finally, we conclude in Section 5 with a summary as well as some future research directions.

2 Literature Review

In this section, we review prior literature related to patent maintenance. Specifically, we will first review the aspects that may be considered when making patent maintenance decisions. Subsequently, we will review existing studies that analyze the characteristics of renewed patents as compared to those of abandoned ones.

2.1 Aspects of Patent Maintenance Decisions

Economic Aspect: Prior studies show that economical factors are essential to patent maintenance decisions. These studies assume that patent holder makes a patent maintenance decision on the basis of the expected revenue from the patent and its maintenance fee. In other words, the patent holder decides not to pay the maintenance fee when the expected benefits from the patent are less than its maintenance cost [5], [15], [19].

Technological Aspect: A patent maintenance decision should support the technology portfolio of an organization. A patent is important to subsequent technological innovations of its holding firm or organization, but may not bring any revenues to the assignee. Biases might appear if only the economic aspect is considered; the technological aspect along with the firm's strategies is also crucial in terms of making patent maintenance decisions. In some studies on patent renewal data, the technological aspect may have a greater influence upon renewal decisions than economic concerns regarding the costs of their renewal [23].

2.2 Characteristics of Patent Renewals

Several prior studies compare the characteristics of renewed patents with those of abandoned ones. According to their empirical analyses, renewed patents differ from abandoned patents in the following characteristics:

- *Backward citations*: Patents maintained to the full term cite more patent prior art references than abandoned patents [14].
- *Number of claims*: Patents that were maintained to the full term have more claims than abandoned patents [1], [14]. Patents with more claims are more intimidating to potential infringers because more claims mean higher chances that other companies will infringe these patents and the harder the patents will be to invalidate [14]. Thus, patents with more claims are more likely to be maintained by their assignees than patents with fewer claims.
- *Claim length*: The average claim length of abandoned patents generally is longer than that of patents maintained to the full term [1].

- *Size of patent family*: The size of patent family of patents maintained to the full term is larger than that of abandoned patents [12].
- *Number of forward citations*: Because forward citations are associated with technological value [14], patents that were maintained to the full term receive more citations (i.e., forward citations) than abandoned patents [1], [9], [14]. In addition, the study by Thomas [23] also suggests a strong positive relationship at all three renewal points between patent citations and patent renewals.
- *Length of written specification*: Patents maintained to the full term have longer descriptions than abandoned patents [1].
- *Patent scope*: The theoretical patent literature in economics suggests that the scope of a patent (typically measured by the number of different four-digit IPC classifications) may be an important determinant of the efficacy of patent protection [20], [21]. Thus, patents maintained to the full term are likely to have larger patent scope than abandoned patents.
- *Self citation*: The study by Thomas [23] indicates that there is a strong positive relationship at all three renewal points between self citations and patent renewals.
- *Internal sequential innovations*: Internal sequential innovations refer to a sequence of innovations that build upon the same underlying technological trajectory and therefore share some genealogical connections. Because internal sequential innovations complement each other in co-defining and co-delimiting the underlying technological trajectory, when a patent belongs to a sequence of patented innovations, the firm is in a better position to fully exploit technological potential in this sequence and the underlying technological trajectory. Therefore, patents belonging to internal sequential innovations are more valuable and therefore more likely to be renewed than patents that are stand-alone innovations [13].

In addition to the studies examining the characteristics of renewed patents, some other prior studies investigate the characteristics of valuable and worthless patents. For example, the study by Harhoff et al. [10] suggests that both the number of backward citations to patent prior art references and non-patent literature (i.e., science linkage metric) as well as the citations a patent receives are positively related to a patent's value. Moreover, patents representing large international patent families (i.e., size of patent family) are particularly valuable. From the economic aspect, patent maintenance decisions relate to the expected revenues (i.e., economic values) of patents. Therefore, the abovementioned characteristics between valuable and worthless patents (i.e., including backward citations, science linkage, and size of patent family) also represent important variables to be considered when making patent maintenance decisions.

As mentioned, prior studies mainly examine the characteristics that could differentiate renewed from abandoned patents or valuable from worthless patents; however, each of these prior studies only focuses on certain characteristics rather than a more comprehensive set of characteristics. In addition, existing studies do not attempt to develop an automated system for supporting patent maintenance decisions. Therefore, to address the research gap of the existing research, we will integrate the characteristics suggested by prior studies that potentially differentiate renewed patents from abandoned ones as the predictors (i.e., independent variables) and adopt a data mining approach to develop a decision support system for supporting patent maintenance decisions of firms and organizations.

3 Design of Patent Maintenance Decision Support System

In this study, we formulate the task of patent maintenance decision as a classification problem; i.e., classifying a patent to either the "maintain" or the "abandon" (i.e., not maintain) class. In our formulated classification problem, we integrate all characteristics that could differentiate renewed from abandoned patents or valuable from worthless patents as the independent variables and employ a supervised learning algorithm that learns a prediction model(s) for patent maintenance from a set of training instances (i.e., patents with known maintenance decisions). In the following, we first describe the set of independent variables and their operational definitions. Subsequently, we detail the process and the design of our proposed patent maintenance decision support system.

3.1 Independent Variables Employed

On the basis of the review of existing studies (in Section 2.2), we employ thirteen independent variables in our study, including number of backward citations, science linkage index, number of claims, average length of independent claims, size of patent family (applications), size of patent family (countries), number of forward citations, length of description, patent scope (by IPC), patent scope (by USPC), forward self citation rate, backward self citation rate, and internal sequential innovations. Table 3 provides the definition of each independent variable employed in this study.

Table 3. Summary of Independent Variables Employed

Independent Variable	Definition
• Number of backward citations	Number of patents cited by the focal patent
• Science linkage index	Number of non-patent references cited by the focal patent
• Number of claims	Number of claims in the focal patent
• Average length of independent claims	Average length of independent claims of the focal patent
• Size of patent family (applications)	Number of applications the patent protecting a particular invention
• Size of patent family (countries)	Number of countries in which protection for the focal patent was sought
• Number of forward citations	Number of patents citing the focal patent

Table 3. (*continued*)

• Length of description	Length of the description of the focal patent
• Patent scope (by IPC)	Number of different four-digit IPC classification codes of the focal patent
• Patent scope (by USPC)	Number of different three-digit USPC classification codes of the focal patent
• Forward self citation rate	$\dfrac{\text{\# of forward citations of the focal patent that are from the same assignee as the focal patent}}{\text{\# of forward citations of the focal patent}}$
• Backward self citation rate	$\dfrac{\text{\# of backward citations cited by the focal patent that are from the same assignee as the focal patent}}{\text{\# of backward citations cited by the focal patent}}$
• Internal sequential innovations	= 1 if the focal patent belongs to a sequence of patented innovations (i.e., internal sequential innovations or patents interrelated by divisional and continuation-in-part relationships); otherwise, = 0.

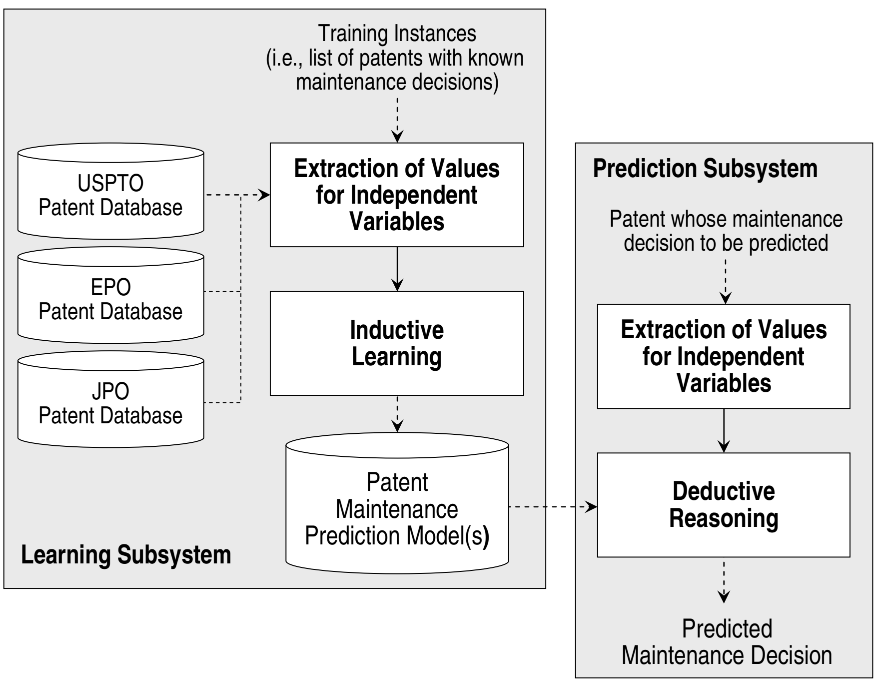


Fig. 1. Overall Process of Our Patent Maintenace Decision Support System

3.2 Process and Detailed Design of Our Patent Maintenance Decision Support System

Because our proposed patent maintenance decision support system takes the data mining approach (specifically, using a supervised learning algorithm), the proposed system can be divided into two subsystems: learning and prediction subsystems. As Figure 1 illustrates, the inputs to the learning subsystem is a set of training instances (i.e., patents with known maintenance decisions, either the "maintain" or "abandon" class). In the "extraction of values for independent variables" phase, the values of the thirteen independent variables (defined in Section 3.1) are extracted for each training patent. The information sources for computing the thirteen independent variables are mainly from the USPTO patent database. Moreover, when calculating the variable of "size of patent family (countries)," we extract related information from EPO (European Patent Office) and JPO (Japan Patent Office) databases.

Following the "extraction of values for independent variables" phase is the "inductive learning" phase. Depending on the supervised learning algorithm employed, the inductive learning phase induces a patent maintenance prediction model or multiple patent maintenance prediction models from the set of training instances. In this study, we choose C4.5 [17], [18], a supervised learning technique, which offers computational efficiency and advantageous interpretability decision tree, as the underlying learning algorithm to construct a patent maintenance prediction model to be used by the prediction subsystem for supporting the maintenance decisions for future patents.

Specifically, C4.5 follows a divide-and-conquer strategy for constructing a decision tree and generally prefers simple trees over complex ones because they are more accurate classifiers of future (new) instances [17], [18]. Given a set of training instances, C4.5 evaluates the information gain (or gain ratio) for each input attribute and selects the one that yields the greatest information gain to branch a target node of the tree. The root node is the initial target node before the decision tree construction process begins (i.e., an empty tree). As the construction proceeds, the decision tree grows by creating child nodes—one for each value of the selected branching attribute emanating from the target node—and then classifying the training instances into appropriate, newly created child nodes. For each child node, the branching attribute selection and tree-construction process continues recursively when the training instances associated with the node do not belong to the same class or the specified termination condition is not satisfied; otherwise, the node becomes a decision node whose class is assigned through the majority class of the training instances associated with the node. Similar to other decision-tree induction techniques, C4.5 is popular partially because of the uncomplicated tree construction process and interpretable and verifiable classification results.

Besides using C4.5 to construct a patent maintenance prediction model, we also adopt a widely employed method, i.e., the bootstrap aggregating (bagging) method [2], [4], to improve the prediction effectiveness of our proposed patent maintenance decision support system. The bagging method, proposed by Breiman [4], generates multiple versions of prediction models and uses these models to arrive at an overall prediction result for a future (new) instance. Figure 2 shows the process of our

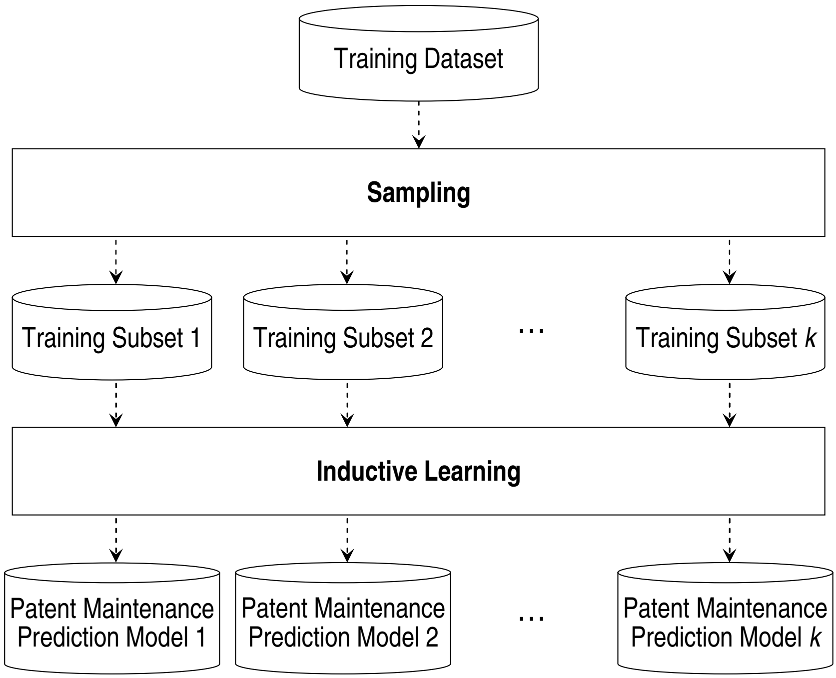


Fig. 2. Process of the Bagging Method

bagging method. In the beginning, we sample with replacements *k* training subsets from the training database (i.e., after the values of all independent variables are extracted for the set of training patents in the "extraction of values for independent variables" phase). In this study, we choose *k* as 15. The size of each training subset is the same as that of the original training dataset. Subsequently, for each training subset, we use C4.5 to train a patent maintenance prediction model. As a result, *k* patent maintenance prediction models are produced and will be used by the prediction subsystem for supporting the maintenance decisions for future patents.

In the prediction subsystem, given a patent whose maintenance decision to be predicted, the extraction of the values for independent variables needs to be performed for this patent. This phase is identical to that in the learning subsystem. Accordingly, if the learning subsystem induces only one patent maintenance prediction model (i.e., without the use of bagging method), the deductive reasoning phase of the prediction subsystem traverses the decision tree (i.e., the prediction model) induced by C4.5, on the basis of the values of the independent variables for the target patent, for arriving at a predicted maintenance decision for the patent. However, if the bagging method is employed by the learning subsystem, the deductive reasoning phase needs to perform multiple deductive reasoning, one for each patent maintenance prediction model. Subsequently, the deductive reasoning phase uses the voting mechanism to combine *k* predictions and assigns the target patent to the class that receives most of the votes from the *k* patent maintenance prediction models.

4 Empirical Evaluation

4.1 Data Collection

The first step of our data collection is to choose industries and companies to be included in our empirical evaluation. In this study, we select industry sectors from Fortune Global 500[4] in 2005. There are fifty industries in total in the list of global 500 companies. However, because some industries (e.g., banks, insurance, and wholesalers) traditionally have fewer patents than other industries, we exclude these industries from our further consideration. Furthermore, for those industries with less than 3 companies listed, we consider them lack of representativeness and thus also exclude them from our consideration. As a result, we select 7 industries for our empirical evaluation purposes. They include *Aerospace and Defense*, *Chemicals*, *Electronics*, *Electrical Equipment*, *Motor Vehicles & Parts*, *Network and Other Communications Equipment*, and *Pharmaceuticals*. In these industries, there are 34 companies and 16,205 patents in the third stage that have faced maintenance decisions during 2004 to 2006. Thus, the 16,205 patents are employed in our empirical evaluation. Table 4 summarizes the number of companies and patents in each industry.

In the second step, we obtain the maintained record of each patent in our dataset from USPTO.[5] There are 121 maintenance event codes. Because we only focus on the maintenance decisions in the third stage, for each patent in our dataset, the status of "payment of maintenance fee, 12th Year" means the patent was maintained and the "Patent Expired for Failure to Pay Maintenance Fees" means the patent was abandoned. Accordingly, we assign the corresponding maintenance status as the known decision to each patent in our dataset.

Table 4. Profiles of Industries and Companies Included in Our Evaluation

Industry	# of Companies	# of Patents
Aerospace and Defense	5	1,365
Chemicals	6	2,144
Electronics, Electrical Equipment	3	5,238
Motor Vehicles & Parts	6	2,131
Network and Other Communications Equipment	4	2,371
Pharmaceuticals	6	966
Telecommunications	4	1,990
Total	**34**	**16,205**

[4] `http://money.cnn.com/magazines/fortune/global500/2005/index.html`
[5] `https://eipweb.uspto.gov/MaintFeeEvents/`

4.2 Evaluation Procedure and Criteria

To evaluate the effectiveness of our proposed patent maintenance decision support system, a ten-fold cross validation is employed. That is, all the instances in our dataset (i.e., 16,205 patents) are randomly partitioned into ten mutually exclusive subsets of nearly equal size. A single subset is retained as the testing set for evaluating the system, and the remaining subsets are used as the training set. The process is then repeated 10 times, with each of the subsets used once as the testing set.

We examine the effectiveness of our proposed patent maintenance decision support system in terms of accuracy, recall, precision, and F1 measures. The accuracy measure is defined as percentage of predictions that are correct. In addition, the recall rate is defined as the portion of patents maintained by respective organizations that are correctly predicted by our proposed patent maintenance decision support system, whereas the precision rate refers to the portion of patents suggested to be maintained by our system that were actually maintained by respective organizations. The F1 measure is the harmonic average of recall and precision rates and is defined as $\frac{2\times Recall\times Precision}{Recall+Precision}$.

4.3 Evaluation Results

Table 5 shows the evaluation results of our proposed patent maintenance decision support system, with or without the use of bagging method (where Single C4.5 denotes the one without the use of bagging method). The accuracy attained by Single C4.5 is 68.23%, which is 1.98% lower than that achieved by the Bagging C4.5 method. However, Single C4.5 slightly outperforms Bagging C4.5 in precision rate (i.e., 74.17% vs. 74.12%), at the cost of its recall rate (i.e., 85.60% for Single C4.5 and 89.93% for Bagging C4.5). The F1 measure indicates that Bagging C4.5 can achieve greater prediction effectiveness (i.e., 81.26%) than Single C4.5 can (i.e., 79.47%). Taken all evaluations together, our empirical evaluation results, on the basis of 16,205 patents, suggest that both methods of our proposed system can achieve satisfactory effectiveness to support patent maintenance decisions and the superiority of the bagging method (i.e., Bagging C4.5) over the non-bagging method (i.e., Single C4.5).

Table 5. Comparative Evaluation Results

Evaluation Criteria	Single C4.5	Bagging C4.5
Accuracy	68.23%	70.21%
Precision (Y)	74.17%	74.12%
Recall (Y)	85.60%	89.93%
F1 Measure (Y)	79.47%	81.26%

5 Conclusion

In this study, we propose and develop a patent maintenance decision support system that takes the data mining approach to construct a patent maintenance prediction model(s) on the basis of the independent variables that we integrate from prior studies. Specifically, we adopt and develop two learning methods: Single C4.5 (i.e., without the use of bagging method) and Bagging C4.5 (with the use of bagging method). Using a set of 16,205 patents as our evaluation dataset, our empirical evaluation results suggest that both methods of our proposed system can achieve satisfactory prediction effectiveness. Furthermore, the bagging method (i.e., Bagging C4.5) outperforms its counterpart (i.e., Single C4.5) in most of the evaluation criteria.

This study contributes to patent maintenance decision support by proposing an effective system and learning methods for constructing patent maintenance prediction model(s). Firms and organizations can employ this system to assist their patent maintenance decision making. In addition, this research also highlights the importance and feasibility of applying data mining techniques to challenging patent analysis research. The integration of these two research areas sheds light on the potential of patent mining for future research.

Our study has several limitations that warrant additional research attention. In this study, we only integrate the variables suggested by prior studies as our independent variables for patent maintenance prediction. To further improve the prediction effectiveness, a comprehensive investigation of other possible factors affecting patent maintenance decisions by firms and organizations should be conducted. For example, we can conduct a Delphi study that involves a panel of senior patent managers and analysts to identify possible patent maintenance decision factors. Subsequently, we can include additional independent variables to extend and enhance our proposed patent maintenance decision support system. In addition, our study only employs a prevalent decision tree induction algorithm (i.e., C4.5) as the underlying supervised learning algorithm. The use and evaluation of other supervised learning algorithms (e.g., backpropagation neural network, support vector machines) for patent maintenance prediction represent another interesting future research direction.

Acknowledgements. This work was partially supported by the National Science Council of the Republic of China under the grant NSC 98-2410-H-002-231-MY3.

References

1. Barney, J.: Comparative patent quality analysis. NACV White Paper (2003)
2. Bauer, E., Kohavi, R.: An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine Learning 36(1-2), 105–142 (1999)
3. Bessen, J.: The value of US patents by owner and patent characteristics. Research Policy 37(5), 932–945 (2008)
4. Breiman, L.: Bagging predictors. Machine Learning 24(2), 123–140 (1996)
5. Crampes, C., Langinier, C.: Information disclosure in the renewal of patents. Annals of Economics and Statistics (49/50), 265–288 (1998)
6. Dueker, K.: Trademark law lost in cyberspace: trademark protection for Internet addresses. Harvard Journal of Law & Technology 9(2), 483–512 (1996)
7. Ernst, H.: Patent information for strategic technology management. World Patent Information 25(3), 233–242 (2003)

8. Griliches, Z.: Patent statistics as economic indicators: a survey. Journal of Economic Literature 28(4), 1661–1707 (1990)
9. Harhoff, D., Narin, F., Scherer, F., Vopel, K.: Citation frequency and the value of patented inventions. Review of Economics and Statistics 81(3), 511–515 (1999)
10. Harhoff, D., Scherer, F., Vopel, K.: Citations, family size, opposition and the value of patent rights. Research Policy 32(8), 1343–1363 (2003)
11. Jones, T., Norris, M., Solomon, I.: Strategies for maximizing value from intellectual capital in a technology-driven business. The Licensing Journal 22(6), 1–7 (2002)
12. Lanjouw, J.O., Pakes, A., Putnam, J.: How to count patents and value intellectual property: The uses of patent renewal and application data. Journal of Industrial Economics 46(4), 405–432 (1998)
13. Liu, K., Arthurs, J., Cullen, J., Alexander, R.: Internal sequential innovations: How does interrelatedness affect patent renewal? Research Policy 37(5), 946–953 (2008)
14. Moore, K.: Worthless patents. Berkeley Technology Law Journal 20, 1521–1552 (2005)
15. Pakes, A.: Patents as options: Some estimates of the value of holding European patent stocks. Econometrica 54(4), 755–784 (1986)
16. Pakes, A., Simpson, M., Judd, K., Mansfield, E.: Patent renewal data. Brookings Papers on Economic Activity. Microeconomics, 331–410 (1989)
17. Quinlan, J.R.: Induction of decision trees. Machine Learning 1(1), 81–106 (1986)
18. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
19. Schankerman, M., Pakes, A.: Estimates of the value of patent rights in European countries during the post-1950 period. The Economic Journal 96(384), 1052–1076 (1986)
20. Scotchmer, S.: Standing on the shoulders of giants: Cumulative research and the patent law. Journal of Economic Perspectives 5(1), 29–41 (1991)
21. Scotchmer, S.: Protecting early innovators: should second-generation products be patentable? RAND Journal of Economics 27(2), 322–331 (1996)
22. Tao, J., Daniele, J., Hummel, E., Goldheim, D., Slowinski, G.: Developing an effective strategy for managing intellectual assets. Research-Technology Management 48(1), 50–58 (2005)
23. Thomas, P.: The effect of technological impact upon patent renewal decisions. Technology Analysis & Strategic Management 11(2), 181–197 (1999)

Neural Network Analysis of Right-Censored Observations for Occurrence Time Prediction

Young U. Ryu[1], Jae Kyeong Kim[2], Kwang Hyuk Im[3], and Hankuk Hong[4]

[1] University of Texas at Dallas, Richardson, Texas, USA
ryoung@utdallas.edu
[2] Kyung Hee University, Seoul, Korea
jaek@khu.ac.kr
[3] Pai Chai University, Daejeon, Korea
khim@pcu.ac.kr
[4] Dong Eui University, Busan Korea
honghk@deu.ac.kr

Abstract. Introduced is a neural network method to build survival time prediction models with censored and completed observations. The proposed method modifies the standard back-propagation neural network process so that the censored data can be used without alteration. On the other hand, existing neural network methods require alteration of censored data and suffer from the problem of scalability on the prediction output domain. Further, the modification of the censored observations distorts the data so that the final prediction outcomes may not be accurate. Preliminary validations show that the proposed neural network method is a viable method.

Keywords: Neural Networks, Survival Time Prediction, Censored Observation, Data Mining.

1 Introduction

The study of estimating timeline of an event occurrence has drawn interests in medicine, finance, and engineering. Often, it involves *censored* (i.e., suspended) observations of expected events as well as completed observations. This form of timeline prediction with censored data does not fit into standard classification or functional form prediction. Survival analysis methods such as life table analysis, Kaplan-Meier estimator, and Cox's proportional hazard model were introduced for study with censored data. They result in a survival function (i.e., the cumulative survival rate on time) or a hazard function (i.e., the failure, or non-survival, probability function on time), instead of survival time estimations on individual cases. Neural network methods have been introduced for survival time estimations [1, 3, 5].

The standard neural network does not allow censored observations to be used when the network is trained. Thus, previous methods make the use of censored observations possible by altering or manipulating the observation data. A major problem with these approaches is in the scalability. That is, if an application has a large domain of output

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 100–109, 2012.

(e.g., prediction in months between 1 and 120 months), either a large number of neural networks or a big neural network must be built. In addition, the actual use of the trained neural networks can be complicated. Further, the modification of the censored observations distorts the data so that the final prediction outcomes may not be accurate.

We propose modification to the standard back-propagation neural network method so that censored data can be used without alteration in network training. The method results in a single neural network whose size does not depend on the output domain. Experiments on synthetic data show that the proposed neural network method performs better than previous neural network approaches. We also provide the details of the modified neural network method and its application to cancer recurrence prediction together with a feature selection technique.

2 Neural Network Models for Right-Censored Observations

A number of neural network methods [1, 3, 5, 8] have been proposed to predict event occurrence time given previous data of censored and completed observations. Using a simple example, we will illustrate data preparation, network structure, and prediction processes of these methods.

Right-censored data, or suspended data, are those without final observation of an expected event, while others are data with completed observation. The lack of final observation of an expected even may be simply due to the fact that the event has not occurred yet or due to termination of the observation before the occurrence of the expected event. For example, as illustrated in Fig 1, suppose we are checking machines' durability after repair services. Machines 1 and 3 failed in 3 and 2 months after the repair; machine 2 is still running after 4 months and under further observation; and machine 4 was still running after 2 months, but the observation was terminated (say, because it was sold at that time). During the observation, we recorded 4 attributes of operating conditions such as average work load, temperature, and others. Then, we have 2 completed observation data points (from machines 1 and 3) and 2 right-censored data points (from machines 2 and 3), as shown in Table 1.

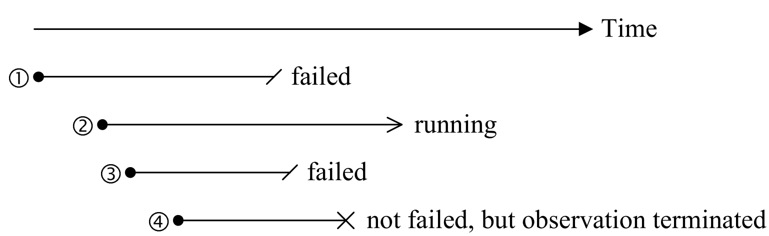


Fig. 1. Censored and Completed Observations (• Time when machine was repaired)

Table 1. Sample Attributes of Observations in Fig 1

Observation	Attributes				Censored?	Time[a]
	1	2	3	4		
1	3	3	4	4	*N*	3
2	2	1	4	3	*Y*	4
3	4	3	5	3	*N*	2
4	3	4	3	3	*Y*	2

[a]censoring time for censored observation; occurrence time for completed observation

The standard neural network processes cannot use censored data for network training. However, there are a number of techniques to deal with censored data in neural network modeling [1, 3, 5, 8]. They are as follows.

2.1 A Single Neural Network with a Single Output

Assuming the machine failure experiment is over a period of 5 months, from the 4 observations in Table 1, we obtain the following data:

(3, 3, 4, 4, 1 | 1), (3, 3, 4, 4, 2 | 1), (3, 3, 4, 4, 3 | 1), (3, 3, 4, 4, 4 | 0), (3, 3, 4, 4, 5 | 0)
(2, 1, 4, 3, 1 | 1), (2, 1, 4, 3, 2 | 1), (2, 1, 4, 3, 3 | 1), (2, 1, 4, 3, 4 | 1)
(4, 3, 5, 3, 1 | 1), (4, 3, 5, 3, 2 | 1), (4, 3, 5, 3, 3 | 0), (4, 3, 5, 3, 4 | 0), (4, 3, 5, 3, 5 | 0)
(3, 4, 3, 3, 1 | 1), (3, 4, 3, 3, 2 | 1),

where first 4 numbers are attribute values, the 5th number is the time of observation, and the 6th output value indicates if the machine was running (1) at the observation time or failed (0) at or before the observation time.

A neural network that consists of 5 input nodes (i.e., 4 nodes for attributes and 1 node for time), a number of hidden nodes, and one output node (indicating that the machine is running or not) is trained using the above data [1]. Once the neural network is trained, an unseen case with estimated or actual attribute values of a_1, a_2, a_3, and a_4 are predicted as follows. The neural network sequentially evaluates data (a_1, a_2, a_3, a_4, 1), (a_1, a_2, a_3, a_4, 2), and so on until the output value becomes less than a certain threshold (e.g., 0.5). If the evaluation stops with (a_1, a_2, a_3, a_4, 2), the case's predicted failure time is 2 months.

2.2 Multiple Neural Networks with a Single Output

For each month under consideration (i.e., 1, 2, 3, or 4 months), we build a neural network that consists of 4 input nodes, a number of hidden nodes, and one output node [3, 5]. For every observation in Table 1, if the machine is still running or has already failed, obtain a data point with the input attributes of the observation and 1 (running) or 0 (failed) as the output. For instance, for the neural network predicting 3rd month failure, the following data are obtained from observations 1, 2, and 3 in Table 1, we have:

(3, 3, 4, 4 | 1), (2, 1, 4, 3 | 1), (4, 3, 5, 3 | 0)

We cannot obtain any data from observation 4, because it was terminated in the 2nd month and the machine's status was not known in the 3rd month. Each neural network is trained with data obtained in this way. An unseen case with estimated or actual attribute values of a_1, a_2, a_3, and a_4 is sequentially evaluated by the neural networks, until the output value becomes less than a certain threshold (e.g., 0.5). If evaluation stops with the 3rd neural network, the case's predicted failure time is 3 months.

2.3 A Single Neural Network with Multiple Outputs

Assuming the machine failure experiment of up to 5 months, we build a neural network that consists of 4 input nodes, a number of hidden nodes, and 5 output nodes (for up to 5 months failure prediction) [8]. For a completed observation with time t, values of 1st to t-th outputs are 1 and others are 0. For instance, from the 1st observation in Table 1, we have

$$(3, 3, 4, 4 \mid 1, 1, 1, 0, 0)$$

For a censored observation with time t, values of 1st to t-th outputs are 1 and those after t-th output are the Kaplan-Meier estimates. For instance, from the 4th observation in Table 1 (assuming that there are many other observations in addition to those in Table 1), we have

$$(3, 4, 3, 3 \mid 1, 1, 0.6, 0.2, 0.1)$$

Once the neural network is trained, the evaluation outputs of an unseen case are sequentially checked from the 1st output until an output value becomes less than a certain threshold (e.g., 0.5). If the 2nd output value is less than the threshold, the case's predicted failure time is 2 months.

3 Modified Neural Network Method for Censored Observations

The previous neural network methods use the standard neural network structures with modified or manipulated data. A practical problem is the neural network scalability on the output domain (i.e., the measure and range of time units). If the time prediction is in months up to 120 months, the multiple neural network method requires the creation and use of 120 neural networks; the single neural network with a single output node requires up to 120 evaluation sessions of a case; the single neural network with 120 output nodes requires a substantially many training cases due to the size of the network. If the output domain is continuous, then discretization is need, which can be another factor that complicates the whole task of learning and prediction. Further, the modification of the censored observations distorts the data so that the final prediction outcomes may not be accurate.

We propose another neural network method for censored and completed observations. Instead of manipulating observation data, we modify the existing back-propagation neural network method so that it can handle both censored and completed observation data. The modified neural network method works as follows.

Given a set of complete observations with n attributes and occurrence time and right-censored observations with n attributes and censoring time, we build a three-layer feed-forward back-propagation neural network with n input nodes, a number of hidden nodes, and one output node. The linear basis function ($u_j = \sum_i w_{ij}\xi_i + b_j$ in Fig 2) and the hyperbolic tangent activation function ($o_j = \tanh(u_j)$ in Fig 2) are used for input aggregation and transformation in the hidden nodes; the linear basis function and the linear activation function ($o_j = u_j$ in Fig 2) are used for the output node.

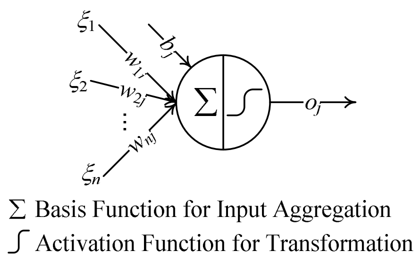


Fig. 2. Data Aggregation and Transformation in Neural Network

Like a typical feed-forward and back-propagation mechanism, training data flow from input to output; the output generated by the neural network is then compared with the known output of training data to determine the error. The critical adjustment in the proposed neural network is in the error determination algorithm. For a *complete training observation* whose actual occurrence time is ζ and the output generated by the neural network is o, we set the error

$$E = \begin{cases} o - \zeta & \text{if } \zeta < o \\ (\zeta - o)/M & \text{if } \zeta \geq o, \end{cases} \tag{1}$$

for some big number $M \gg 1$. That is, over-estimation of occurrence time is penalized much more than under-estimation of occurrence time. For a *censored training observation* whose censoring time is ζ and the output generated by the neural network is o, we set the error

$$E = \begin{cases} 0 & \text{if } \zeta < o \\ \zeta - o & \text{if } \zeta \geq o. \end{cases} \tag{2}$$

That is, only under-estimation of non-occurrence/waiting time is penalized. This information is transferred, or propagated, back to adjust the weights of the arcs using the gradient decent algorithm and the chain rule. That is, the weight w of an arc and a bias term are incremented by

$$\Delta w = -\eta \frac{\partial E}{\partial w} + \alpha w \quad \text{and} \quad \Delta b = -\eta \frac{\partial E}{\partial b} + \alpha b,$$

where η is the learning rate and $0 \leq \alpha < 1$ is the momentum parameter.

4 Experiments with Synthetic Data

To evaluate the prediction accuracies of various neural network methods, we perform experiments on synthetic data sets. We generate them following the method in [2] such that each data point consists of 8 attributes (a_i) having integer values between 1 and 10 and the output value is randomly generated having the binomial distribution of bin(10, p) with

$$p = \left(\frac{\mathrm{sigmoid}(c \cdot a - \overline{a}) - 0.5}{\mathrm{sigmoid}(c \cdot \overline{a}) - 0.5} \right) \cdot 0.5 + 0.5,$$

where $a = \sum_i a_i$. Note, the sigmoid function is defined as $1/(1 + e^{-t})$. We set $\overline{a} = 40$ and $c = 1, 2$, and 3. Generated are six small data sets and three large data sets; a small data set includes 300 data points and a large data set 1200 data points. For each of three small data sets, 50 data points are converted to censored data. For such a data point, its output value o is replaced by censoring time ξ having the discrete uniform distribution of DU(1, o). For each of remaining three data sets, 150 data points are converted to censored data. For each of large data sets, 400 data points are converted to censored data.

A small set does not contain a sufficient number of data to build good neural networks. In such situations, we would like to see various neural network methods' prediction accuracies contingent on the proportion between censored and completed observation data. A large data set contains a sufficient number of data to build good neural networks. We would like to see different neural network methods' accuracies at their best configurations.

For each experiment with a small data set, 200 data points are used for training neural networks and remaining 100 data points are used for validation. As presented in Table 2, the modified neural network performs better than others.

Table 2. Sample Attributes of Observations in Fig 1

Neural Network	Small Data Sets		Large Data Sets
	Less Censored Data	More Censored Data	
Single NN–Single Out.	0.29	0.33	0.27
Mult. NN–Single Out.	0.18	0.27	0.12
Single NN–Mult. Out.	0.22	0.25	0.19
Modified NN	0.11	0.18	0.08

Note: Numeric values indicate prediction errors (of Equation (1) or (2), where $M = 100$) averaged over three data sets.

5 Breast Cancer Recurrence Prediction

5.1 Experiment with All Features

The breast cancer prognosis data set [4] used for our study contains 194 data points. Each data point consists of 32 input features, one class attribute, and one recurrence/survival time attribute. The majority of the features were on fine needle aspirates taken from patients at the time of diagnosis. The first ten features (x_1, x_2, . . . , x_{10}) are the mean values of cell nuclei's radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. The second ten features (x_{11}, x_{12}, . . . , x_{20}) are their standard errors; and the third ten features (x_{21}, x_{22}, . . . , x_{30}) are mean values of their three largest (i.e., worst) values. The last two features (x_{31}, x_{32}) are tumor size (i.e., diameter of the excised tumor) and lymph node status (i.e., number of positive axillary lymph nodes) observed at the time of surgery.

Among 194 data points, 46 are of completed observations with recurrence times varying between 1 month and 125 months, and 148 data points are of censored observations with survival (i.e., censoring) times varying between 1 month and 79 months. Using this data set, we built a modified neural network and perform leave-one-out validation, in which a modified neural network was created using all data points except one and tested on the left-out data point. Experiments were repeated for each of the 194 data points being left out testing.

The modified neural network process was implemented in Matlab/Octave, a numerical computation system. The neural network consisted of 32 input nodes, 20 hidden nodes, and one output node for cancer recurrence time prediction. We set learning rate $\eta = 0.1$ and momentum $\alpha = 0$ for learning. For error measure, we used $M = 10$, 50, and 100. The average prediction errors of cancer recurrence or patient survival time over the 194 leave-one-out validation sessions were 18.08 months (14.46% error rate) for $M = 10$, 16.96 months (13.57% error rate) for $M = 50$, and 18.19 months (14.55% error rate) for $M = 100$. These error rates are low according to doctors' evaluations [8].

5.2 Experiment with Selected Features

Network pruning via input feature selection/reduction is a practical method to minimize the size of the neural network while maintaining good performance. When the training of a neural network is finished, a set of arc weights for the trained neural network is obtained. At this time, we use four feature-weighting measures to check the relative importance of input features: sensitivity, activity, saliency, and relevance [6, 7]. Each of these measures calculates the degree of an input feature's importance by using the connection weights and activation patterns of nodes in the trained neural network. The feature-weighting measures are as follows.

For input feature (i.e. node) i, the *sensitivity* is defined as the impact of i on prediction outcomes of training data. Let L be the set of n training data points, and p_h for $h \in L$ be the prediction outcome of training data point h (using all features) produced by the neural network built with L; let $p_{i,h}$ be the prediction outcome of h with all features except i (i.e., with the i value set to 0). Then, the sensitivity of input feature i is

$$S_i = \frac{\sum_{h \in L} |p_h - p_{i,h}|}{n}.$$

The *activity* of a node measures the variance of activation values of training data. Let $w_{i,j}$ denote the weight of an arc connecting input node (i.e., input feature) i and hidden node j, and w_j the weight of an arc connecting hidden node j and the output node. Then, the activity of hidden node j is

$$A_j = w_j^2 \cdot \underset{h \in L}{\mathrm{var}}\left(g\left(\sum\nolimits_i w_{i,j} x_{i,h} + b_j\right)\right),$$

where $x_{i,h}$ is the i value of data point h, and $g(\cdot)$ is the activation function (i.e., tanh(·)). The activity of input node i is obtained by aggregating the activities of all hidden nodes:

$$A_i = \sum\nolimits_j w_{i,j}^2 A_j \ .$$

For feature selection, we use the activities of input nodes.

The *saliency* of an input node is another parameter measuring the importance of the input node in the derivation of prediction outcome. It is calculated by aggregating arc weights. That is, the salience of input node i is

$$L_i = \sum_j w_{i,j}^2 w_j^2 \ .$$

The *relevance* of a node is a measure of variance of weights in the neural network. The relevance of hidden node j is

$$R_j = w_j^2 \cdot \underset{h \in L}{\mathrm{var}}\left(w_{i,j}^2\right).$$

The relevance of input node i is obtained by aggregating the relevance values of all hidden nodes:

$$R_i = \sum\nolimits_j w_{i,j}^2 R_j \ .$$

We use the relevance values of input nodes for feature selection.

We applied a backward sequential elimination operation on a subset of the data points. The sequential elimination worked as follows. We used all 32 features to train a neural network and calculated all input nodes' sensitivity, activity, saliency, and relevance values. Then we eliminated a node with the average rank of these four feature-weighting values was lowest, which was the mean value of cell nuclei's smoothness (x_5) as shown in Fig 3. Now, we used 31 features to train a neural network, calculated 31 input nodes' feature-weighting values, and eliminated a node with the lowest mean rank of these values. We performed this operation until no improvement was made. The feature selection operation reduced the 32 input attributes into 5 attributes, which were the mean value of cell nuclei's texture, variance of cell nuclei's perimeter values, the extreme value of cell nuclei's concavity, the mean value of tumor size, and the extreme value of cell nuclei's smoothness as shown in Figure 4. When the reduced attributes were used for the training and validation of the neural network, the average prediction errors were 17.58 months (14.06% error rate), 16.04 months (12.83% error rate), and 18.07 months (14.46% error rate) for $M = 10$, 50, and 100, respectively.

var	Sensitivity	Activity	Saliency	Relevance	Sen Rank	Act Rank	Sal Rank	Rel Rank	Average
x1	0.19694	0.33961	0.58281	0.10757	18	31	23	31	25.75
x2	0.63899	2.15034	2.25855	0.65409	2	3	3	3	2.75
x3	0.22391	0.37492	0.60573	0.13869	16	29	22	24	22.75
x4	0.17984	0.3443	0.39085	0.10856	26	30	31	30	29.25
x5	***0.18341***	***0.20763***	***0.36727***	***0.07971***	***25***	***32***	***32***	***32***	***30.25***
x6	0.3199	0.71516	0.78533	0.25203	9	15	18	18	15
x7	0.18644	0.42052	0.45954	0.12884	23	26	30	26	26.25
x8	0.19172	0.40904	0.53568	0.1428	21	27	27	23	24.5
x9	0.50681	1.63075	1.83707	0.58302	3	5	4	4	4
x10	0.43439	0.90172	1.13257	0.30959	5	12	12	13	10.5
x11	0.2477	0.69942	0.7584	0.18764	14	17	19	20	17.5
x12	0.24865	1.23589	1.7922	0.49766	13	6	6	6	7.75
x13	0.36763	0.60041	0.75639	0.16536	6	20	20	22	17
x14	0.17659	0.45355	0.50827	0.12361	27	25	28	28	27
x15	0.17425	0.62555	0.58174	0.18882	28	19	24	19	22.5
x16	0.19502	0.70915	1.01523	0.2736	19	16	14	16	16.25
x17	0.14727	0.48207	0.54582	0.12748	32	23	26	27	27
x18	0.15717	0.45575	0.56437	0.11004	29	24	25	29	26.75
x19	0.19234	1.03019	1.23109	0.35443	20	9	11	10	12.5
x20	0.15218	0.38751	0.48409	0.13608	31	28	29	25	28.25
x21	0.15485	0.83491	1.00129	0.31075	30	13	15	12	17.5
x22	0.48705	1.70244	1.79382	0.54936	4	4	5	5	4.5
x23	0.18893	0.49572	0.62321	0.17631	22	22	21	21	21.5
x24	0.26232	0.66986	1.12944	0.28857	12	18	13	14	14.25
x25	0.88679	2.95362	2.9789	1.01699	1	2	2	2	1.75
x26	0.32153	1.00961	1.31539	0.3604	8	10	9	9	9
x27	0.30449	1.09204	1.25276	0.32907	10	8	10	11	9.75
x28	0.24104	0.76467	0.85941	0.28648	15	14	16	15	15
x29	0.21246	1.15217	1.53539	0.40799	17	7	8	8	10
x30	0.18599	0.58013	0.85202	0.26123	24	21	17	17	19.75
x31	0.34069	3.09557	4.04304	1.22051	7	1	1	1	2.5
x32	0.28521	0.98133	1.56397	0.44835	11	11	7	7	9

Fig. 3. Feature-Weighting Measures

var	Sensitivity	Activity	Saliency	Relevance	Sen Rank	Act Rank	Sal Rank	Rel Rank	Average
x2	2.1751	129.78	171.1686	1.1659	1	1	1	1	1
x13	1.6423	104.301	130.1978	0.6945	5	4	4	4	4.25
x25	1.8974	25.2105	38.9465	0.1942	4	5	5	5	4.75
x27	2.0925	106.886	142.8691	0.7717	2	3	3	3	2.75
x31	1.9751	116.067	153.2707	0.9647	3	2	2	2	2.25

Fig. 4. Feature Selection Result

6 Discussion

In the experiment with breast cancer prognosis data, we successfully predict the cancer recurrence (i.e., patient survival) with the proposed neural network method. Because the domain of output ranged between 1 month and 125 months, we had difficulty in training and validating other types of neural networks. Thus, we could not compare the modified neural network method with previous methods for this data set. However, we generated synthetic data and compared the proposed method with previous approaches. The results indicated the proposed method performed very well.

Further validations and applications of the proposed neural network method for various problems are necessary before we claim the general usability of the method.

The major difficulty is in the collection of data. Nevertheless, we are currently collecting more data sets on which we are testing the proposed method. The results will be presented in a subsequent research report.

References

1. De Laurentiis, M., Ravdin, P.M.: Survival analysis of censored data: Neural network analysis detection of complex interactions between variables. Breast Cancer Research Treatment 32, 113–118 (1994)
2. Jacob, V.S., Krishnan, R., Ryu, Y.U.: Internet content filtering using isotonic separation on content category ratings. ACM Transactions on Internet Technology 1(7), article 1 (2007)
3. Lapuerta, P., Azen, S.P., LaBree, L.: Use of neural networks in predicting the risk of coronary artery disease. Computers and Biomedical Research 28(1), 38–52 (1995)
4. Merz, C.J., Murphy, P.M.: UCI repository of machine learning databases. University of California, Irvine, Department of Information and Computer Sciences (1998)
5. Ohno-Machado, L.: Sequential use of neural networks for survival prediction in AIDS. In: Proceedings of American Medical Informatics Association Annual Fall Symposium, pp. 170–174 (1996)
6. Shin, C.K., Park, S.C.: Memory and neural network based expert system. Expert Systems with Applications 6, 145–155 (1999)
7. Shin, C.K., Yun, U.T., Kim, H.K., Park, S.C.: A hybrid approach of neural network and memory-based learning to data mining. IEEE Transactions on Neural Networks 11(3), 637–646 (2000)
8. Street, W.N.: A neural network model for prognostic prediction. In: Proceedings of the Fifteenth International Conference on Machine Learning, pp. 540–546 (1998)

Impact of Recommendations on Advertising-Based Revenue Models

Philipp Bodenbenner, Markus Hedwig, and Dirk Neumann

University of Freiburg, Germany
{philipp.bodenbenner,markus.hedwig,
dirk.neumann}@is.uni-freiburg.de

Abstract. Online content providers need a loyal user base for achieving a profitable revenue stream. Large number of visits and long clickstreams are essential for business models based on online advertising. In e-commerce settings, personalized recommendations have already been extensively researched on their effect on both user behavior and related economic performance indicators. We transfer this evaluation into the online content realm and show that recommender systems exhibit a positive impact for online content provider as well. Our research hypotheses emphasize on those components of an advertising-based revenue stream, which are manipulable by personalized recommendations. Based on a rich data set from a regional German newspaper the hypotheses are tested and conclusions are derived.

Keywords: E-business, information economics, business model, digital goods, recommendation agent, online customer loyalty, empirical research/study, web metrics.

1 Introduction

The ongoing digital transformation leads to generation of enormous, ever-growing amounts of information that is made available through content-based web pages in the Internet. In contrast to e-commerce web pages, content providers face the challenge of establishing a viable business model. Advertising is being perceived as the primary and dominant source for revenue for the vast majority of online content providers (Evans 2009). For example, the total revenue of newspapers splits up as follows: 81.5% are being earned with advertising; the rest is based on subscription (Clemons et al. 2002). Since advertising is dominating the content providers' revenue streams, it is crucial for them to create heavy traffic on their pages and keep readers to their pages as long as possible. The large article databases of content web-pages demand new techniques to decrease search costs and present items of potential interest quickly and easily to users. Recommender systems particularly serve these requirements. They are a rather new field of academic research, becoming a focus of interest in the mid-1990s (e.g. Resnick and Varian 1997). In the beginning, system design as well as the technical efficiency and predictive accuracy have been the primarily focus of evaluation of recommender systems (e.g. Schafer et al. 1999). Understanding the

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 110–118, 2012.

behavioral effects on users and impact on key economic performance indicators has only been subject of research recently. Examples for such evaluation are two studies conducted on the basis of realworld data from Amazon.com (Kumar and Benbasat 2006) and LeShop (Dias et al. 2008). These studies analyzed the business value of personalized recommendations including the impact on revenue. The largest group of studies focuses on analyzing the effects on sales diversity patterns (e.g. Brynjolfsson et al. 2006; Fleder and Hosangar 2007). Hinz and Eckert (2010) compare multiple search systems (hit lists, recommendations, etc.) and show that recommenders lead to substitution that is beneficial for retailers when niches yield higher margins than substituted top-sellers. As shown, most studies that have hitherto been conducted have dealt with the effect of recommender systems in e-commerce settings. Corresponding impact on content based services has only marginally been evaluated yet. We address this research gap and try to find answers to the overarching question: do recommendations entail an impact on the user behavior in a content page setting, and improve relevant economic performance indicators in advertising-based revenue models? To show these effects we put up a model that inter-relates all components of an online content provider's revenue stream (cf. Bodenbenner and Neumann 2012). We formulate a set of hypotheses on components, which can be influenced by recommender systems, and test them in a study. The study is being conducted based on a real-world data set from a major German newspaper. Core of the study is two-group experiment comparing usage data from a test group with a control group. The control group was provided with random links instead of real, personalized recommendations. The paper is structured as follows: First, we describe our research framework and the related hypotheses. Then, the study environment is presented. Finally, hypotheses are evaluated and preliminary results are derived.

2 Research Framework

For our research framework we have to extract the nucleus of revenue streams of content web pages. The basis for this is a group of performance indicators, so-called web metrics, which have been defined within the realm of web analytics. The topic of web analytics is rather new and scientifically still in its infancy. Recently, a few reference books dealing with web analytics have been published (e.g. Peterson 2006). According to Bhat et al. (2002) exposure, as well as stickiness are core objectives of content web pages. These concepts reflect customer engagement that is a key driver for an effective online customer relationship. In addition, these concepts are supplemented by “content portfolio” that characterizes the depth of usage of a providers’ content base. Together with the external factor “profit” these concepts determine the final outcome, which is total revenue of the content provider.

Concept: Exposure. The first objective, exposure, reflects the number of total visits a web page exhibits. This embraces both the number of unique visitors as well as their average number of visits to the web page (Bhat et al. 2002). The key figure “number of unique visitors” resembles the sum of users that frequent a web page within a certain time frame. For web pages that rely on an advertising revenue model, having a large number of visitors is crucial. New visitors have to be acquired and should view as many different pages as possible in order to generate advertising views. The second factor of the exposure cluster is average number of visits per visitor. This shows visitor retention and measures the number of times a visitor has shown up on the web page within a given period.

Concept: Stickiness. The second objective is stickiness, which measures the overall visit duration on a web page (Demers and Lev 2001). Moreover, stickiness indirectly measures the relevance of provided information and the associated satisfaction of users. It is derived from the number of page views per visit and the average time spent per page. A visit denotes usage sequences that can be related to a certain user (Peterson 2006). Inactivity between two user activities exceeding a certain threshold implies the end of the old visit and the start of a new visit. The most prevalent measure for comparison of web pages is the number of page views. The number of page views is an excellent indicator of how compelling and easy to navigate visitors find a web page. For content-based web pages, a measure of success is getting visitors to look at a large number of pages. Furthermore, stickiness is determined by the average time spent per page. This describes the timespan between entering and leaving a single web page (Peterson 2006). This metric helps to measure the overall goal of increasing the amount of time users spend on a web page. Multiplied with the average page views per visit, it results in the average time spent on the web page per visit. The duration of a web page visit is important since an increased page exposure time concurrently increases the probability that a user clicks on an advertising banner. This phenomenon has been evidenced in different studies (e.g. Bucklin and Sismeiro 2003).

Concept: Content Portfolio. With the fact that online stores are able to display a much larger number of products than regular stores, consumption moves away from being concentrated on a small number of popular items only (e.g.; Zhu and Zhang 2010). With the help of online search and filtering tools, customers can also search for and discover the lower-selling niche products off the list of top sellers, the so-called "long tail" (Anderson 2006). The long-tail phenomenon is particularly relevant for information goods and online content (e.g. news, music) that can be digitized and distributed at virtually no cost via the Internet. Due to the fact that production and distribution costs for these goods approach zero, the tail can be extremely long, outplaying the traditional 80/20 Pareto principle of offline channels (Elberse 2008). This makes the strategy of extending a top seller portfolio by a substantial long tail promising for online content providers. A long tail portfolio has an increased relevance for a broader audience, since it addresses both mainstream and niche interests. Therefore, it positively affects users' loyalty (exposure) and their visit duration (stickiness).

External Factor: Profit. As described beforehand, most content-based web pages rely on advertising as revenue stream. For online advertising there exist two major classes of pricing models, namely “Cost-per-Mille” (CPM) and “Cost-per-Click” (CPC) pricing (Evans 2009). With CPM-based pricing advertisers pay for exposure of their banners to a specific audience that is measured per thousand impressions.

CPC-advertising, the second type of pricing model, only charges advertisers when the consumer clicks on the advertisement.

Outcome: Total Revenue. The analysis of online customer behavior aims at maximizing the revenue and profitability of the e-business, be it an e-commerce or content-based web page. Based on the previously described performance indicators, the overall revenue of a web page that is based on advertising can be calculated. The resulting revenue is derived from multiplying exposure and stickiness with the profit component.

The five previously described components (i.e. concepts, external factor, and outcome) together form our research framework. This framework creates a relation between the metrics that are relevant to revenue generation of content-based web pages (see figure 1).

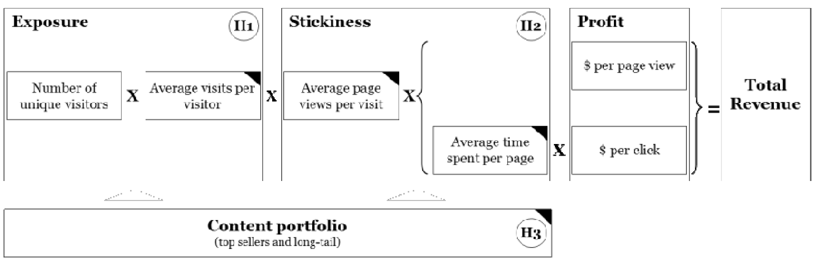


Fig. 1. Research Framework. Triangles indicate components on which an impact by recommendations is assumed. Hypotheses H1-H3 are formulated to verify that.

Personalization has proven to be a key driver for building up user loyalty towards a web page (e.g.; Pathak et al. 2010). There are several ways to achieve personalization, with recommender systems being a very promising. Recommender systems provide added value for the user, create a relationship with the user, and attract users to return to the site (Schafer et al. 1999). The more a user visits a page the more a recommender system learns about his preferences and, consequently, the recommendations become more accurate. This in turn increases the loyalty again and increases switching costs (Schafer et al. 1999; Pathak et al. 2010). In an experiment with Google News, the frequency of website visits significantly increased for the test group compared with the control group (Liu et al. 2010). On average, the frequency of website visits in the test group was 14.1% higher than the control group. We consolidate these findings into the following statement:

H1: *Showing personalized recommendations amplifies exposure of content to customers, i.e. personalized recommendations increase average visits per visitor.*

Our research found no sources relating to the impact of recommender systems on the stickiness of web pages. However, there is research that covers related resp. comparable metrics, such as search costs, loyalty, trust, and number of products sold (for e-commerce settings). These insights can be taken as baseline and valuable input for formulation of our second hypothesis. Recommender systems drive down search costs. Thus, they are only needed when search costs are high (Chen et al. 2004), which particularly applies to online content bases with thousands of accessible content pages. Items with high search costs can particularly profit: showing more recommendations significantly increases their sales figures. For easy-to-find items and small-sized content bases, recommendations could not achieve a major improvement. By decreasing search costs, recommender systems manage to guide users quicker and easier towards information they are looking for and thusly support overcoming the information overflow (Davenport and Beck 2001). Moreover, personalization, which is the core of recommender systems, increases the relevance of shown links on a web page. Most users visit news websites with the attitude of “show me something interesting,” rather than having any specific information goals (Das et al. 2007). Efficient recommenders can help provide users with more relevant content that fits users’ current preferences and thus stick users to websites (Davenport and Beck 2001). The perception of personalization significantly increases customers’ readiness to build up loyalty (Komiak and Benbasat 2006). An experiment conducted with Google News showed that personalization of news recommendations created a more focused reading in the test group (Liu et al. 2010). Users pay more attention and spend more time in the recommended news section. A similar coherence between recommender systems and stickiness is also shown by Hinz and Eckert (2010). In this e-commerce setting it could be observed that recommender systems lead to substitution and not additional consumption; different results were achieved when analyzing other search systems, such as hit lists. In summary, there seems to be coherence between the impact of recommender systems and attractiveness of a web page. This entices users to stay longer on the respective pages and perform more clicks. Thus, we raise the following hypothesis:

H2: *Showing personalized recommendations increases stickiness of a content page. It positively impacts a) the number of page views per visit, and b) leads to more intensely read content pages.*

A focal point of present studies and research related to recommenders’ non-technical impact is on sales concentration respectively diversity. So far, the findings have been derived from e-commerce settings. Two oppositional trends can be found in literature. There is a small group of mainly older research that found evidence for increase in sales concentration when employing recommender systems (e.g.; Mooney and Roy 2000). This means the popularity of already popular products, so-called blockbusters, is further reinforced. The long tail is focus of the second existing group of research. This much stronger and more recently-active group proves recommender systems help consumers discover new products and thus increase sales diversity (e.g. Brynjolfsson et al. 2006; Fleder and Hosangar 2009). That means hitherto niche items are being recommended and adjacently read, respectively sold, more often. This is especially interesting from an up- and cross-selling perspective. We follow the impression of the latter group and deduct our third hypothesis:

H3: *Recommendations lead to diversity in visited content base – from recently top viewed content to long tail: breadth of visited article base, contents’ age as well as articles’ category.*

3 Study Design and Preliminary Results

Our examinations focus on one specific type of content-based web pages, namely online news-papers. Nonetheless, our findings can be easily translated to other content-based services. To evaluate our research framework we are able to conduct a study based on usage data from a major player in the German newspaper market, which employs a YOOCHOOSE recommender system (i.e. a hybrid algorithm, combining stereotype and collaborative filtering) to provide personalization to its visitors. The dataset for our study is derived from a two-group experiment that recorded usage data from a test group and a control group over a period of one week. The full dataset comprises of around 300k records. The groups are constituted randomly, and they are virtually equal in size (test group = 53,828 unique users, control group = 53,604). Any further characteristics about the groups are not known. During the testing period the test group was provided with "real" personalized recommendations, whereas the control group was shown arbitrary links. We have to hazard the consequences of well-known constraints that arise from collecting anonymous measuring data without user login. For example, the refusal of session cookies, multiple persons using one device, and persons using multiple devices lead to inaccuracy in determining user sessions as well as the uniqueness of users. We assume that this effect is balanced out over the two groups and large number of observations that we analyzed.

The evaluation of data is still ongoing. But still, we are able to already present a few noteworthy preliminary results. We have clustered the observations into different segments (see figure 2).

Fig. 2. Segmentation of Study Observation Data

Hypothesis H1. The first hypothesis targets at the impact of personalization on the exposure component. Firstly, we evaluate the test group, and compare users that never clicked on a recommended link with those that have reacted to such a link at least once (UR- vs. UR+). The Wilcoxon test calculates a critical value z that lies outside the boundaries of the 97.5% confidence interval (see test 1). Hence, the null hypothesis (i.e. actual location shift between distributions of visits is greater than 0) can be rejected. Consequently, the average visits per user are stochastically higher for those users that clicked on at least one recommendation in any of their visits during the testing period. H1 is accepted for this part. Secondly, we compare the observed values of the test group with the control group. Results of the Wilcoxon test are inconclusive (see test 2). Thus, a significant advantage of the personalized recommendations, i.e. more visits per user, over the arbitrary links cannot be witnessed. Nonetheless, the observations show that personalized recommendations have a positive impact on "converting" users into loyal visitors (i.e. cluster UR+). The percentage of users that clicked at least one recommendation (UR+) is more than 30% higher for the test group (4.7% of users) compared with the control group (3.5%). This again leads to a higher total average of visits per user. Unfortunately, this effect cannot be fully validated. For interpretation, the position of the recommendations on the web page needs to be taken into account. The recommended links are presented on the lower end of a content page. Thus, they are recognized by a small percentage of visitors only, which results in limited influence of personalized recommendations on the full data set.

Hypothesis 2. H2a sets the focus to the stickiness of a content page: do personalized recommend-dations help to extend a single visit of a user? Again, we start by assessing the test group (UR- vs. UR+). The observed mean of average page views per visit is considerably higher for the cluster using personalized recommendations. A Wilcoxon test confirms this assumption (see test 3); the null hypothesis can be rejected. Hence, personalized recommendations entail a significant increase in page views per visit. H2a is accepted for this part. Secondly, we analyze differences in the impact of using recommended links between the test and control group. Results of the Wilcoxon test are again inconclusive (see test 4). "Personalization", even when arbitrary links can be considered as placebo, has a considerable impact as such, but H2a cannot be accepted when it comes to showing advantages of personalized recommendations over arbitrary links.

Table 1. Overview on results from Wilcoxon-Mann-Whitney tests, $\alpha = 2.5\%$ for all test runs. A chi-square-test showed that the observed values are not normally distributed. The Wilcoxon test can be applied since the observed values are ordinal and observations from the compared groups are independent of each other.

	1) Visits/user		2) Visits/user		3) Clicks/visit		4) Clicks/visit		5) Reads/visit		6) Reads/visit	
	U_{R-}(t)	U_{R+}(t)	U_{R+}(t)	U_{R+}(c)	U_{R-}(t)	U_{R+}(t)	$U_{R+,a}$(t)	$U_{R+,a}$(c)	S_{R-}(t)	S_{R+}(t)	$U_{R+,a}$(t)	$U_{R+,a}$(c)
$\bar{x}$	1.32	2.45	2.45	2.36	1.37	2.59	2.88	2.81	0.94	2.57	1.09	2.05
s / s^2	1.03/1.06	2.16/4.69	2.16/4.69	1.91/3.65	0.90/0.82	2.34/5.49	2.51/6.29	2.22/4.95	0.95/0.91	2.14/4.62	1.08/1.16	1.92/3.68
m / n	54,063	4,948	4,948	3,567	48,757	4,408	3,397	2,529	65,282	2,246	2,529	3,397
interval	]-∞; -4.35e-05]		]-∞; 2.91e-05]		]-∞, -0.999999]		]-∞, 6.83-05]		]-∞, -1.000048]		[0.99996, ∞[	
z	-5.51e-06		-1.40e-05		-0.999994		-9.91e-07		-1.000047		0.99992	

When the user visits a page, a read event is triggered after a certain time. We use the number of these events to measure the average visit duration. For evaluation of H2b we compare distributions of read events for visits with and without clicks on recommended links. The observed means of read events per visits show a clear tendency: it is greater for visits that include clicks on recommendations. Testing confirms the observation; H2b is accepted for this part. For comparing test and control group we again employ a Wilcoxon test, which results in rejection of the null hypothesis. This result is surprising at first sight: the average number of read events per visit is stochastically higher when showing arbitrary recommendation links, and not (as assumed) for personalized recommendations. This result is possibly biased by the way we measure read events. When a user clicks on an arbitrary recommended link he is pushed into a setting that is (most probably) not linked to the prior page. The user requires a certain time to conceive the context and content of the newly called page. Since read events are triggered after a pre-determined time span, the user is still in the state of getting familiar with the new context and not really reading yet. Read events that follow clicks on personalized recommendations consequently exhibit a higher probability that the content has been really consumed.

Hypothesis H3. *For Hypothesis 3 no significant results have been derived yet.*

4 Managerial Implications and Conclusion

Content providers are searching for viable business models in the Internet. Currently, a revenue stream based on advertising is most promising. Therefore, providers need to attract their visitors with relevant, personalized content. As our study in an online newspaper setting demonstrates, recommender systems can contribute to this and significantly improve the revenue stream. The preliminary study results suggest that recommender systems entail a positive influence on key business performance indicators of a content business model, namely exposure and stickiness. Exposure is significantly improved by increasing users' loyalty (i.e. increasing number of recurring visitors and higher average visits per visitors). Stickiness is amended as well: the length of stay both in terms of page views and time spent on the page are positively affected. In the near future, we plan to conduct further evaluations to compare effects of different recommender systems (e.g. popularity). Moreover, we want to extend our research to the actual content to evaluate whether recommendations within or across different categories entail diverse effects.

References

1. Anderson, C.: The Long Tail: Why the Future of Business is Selling Less of More. Hyperion, New York (2006)
2. Bhat, S., Bevans, M., Sengupta, S.: Measuring Users' Web Activity to Evaluate & Enhance Advertising Effectiveness. Journal of Advertising 31(3), 97–106 (2002)
3. Bodenbenner, P., Neumann, D.: Are Personalized Recommendations the Savior for Online Content Providers? In: Proceedings of MKWI 2012, pp. 1933–1945 (2012)
4. Brynjolfsson, E., Hu, Y., Smith, M.: From Niches to Riches: The Anatomy of the Long Tail. Sloan Management Review 47(4), 67–71 (2006)
5. Bucklin, R., Sismeiro, C.: A Model of Web Site Browsing Behavior Estimated on Clickstream Data. Journal of Marketing Research 40(3), 249–267 (2003)
6. Chen, P., Wu, S., Yoon, J.: The Impact of Online Recommendations and Consumer Feedback on Sales. In: ICIS 2004 Proceedings, paper 58 (2004)
7. Clemons, E., Gu, B., Lang, K.: Newly-Vulnerable Markets in an Age of Pure Information Products. In: Proc. of the 35th Hawaii Int. Conf. on System Sciences, p. 218 (2002)

8. Das, A.S., Datar, M., Garg, A., Rajaram, S.: Google News Personalization: Scalable Online Collaborative Filtering. In: Proc. of the 16th Int. Conf. on WWW, pp. 271–280. ACM (2007)
9. Davenport, T., Beck, J.: The Attention Economy: Understanding the New Currency of Business. Harvard Business School Press, Boston (2001)
10. Demers, E., Lev, B.: A Rude Awakening: Internet Shakeout in 2000. Review of Accounting Studies 6(2-3), 331–359 (2001)
11. Dias, M., Locher, D., Li, M., El-Deredy, W., Lisboa, P.: The Value of Personalized Recommender Systems to E-Business. In: Proceedings of the ACM RecSys, pp. 291–294 (2008)
12. Elberse, A.: Should You Invest in Long Tail? Harvard Business Review 86(7-8), 88 (2008)
13. Evans, D.: The Online Advertising Industry: Economics, Evolution, and Privacy. Journal of Economic Perspectives 23(3), 37–60 (2009)
14. Fleder, D., Hosanagar, K.: Blockbuster Culture's Next Rise or Fall: Impact of Recommender Systems on Sales Diversity. Management Science 55(5), 697–712 (2009)
15. Hinz, O., Eckert, J.: The Impact of Search and Recommendation Systems on Sales in Electronic Commerce. Business & Information Systems Engineering 2(2), 67–77 (2010)
16. Komiak, S., Benbasat, I.: The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. MIS Quarterly 30(4), 941–960 (2006)
17. Kumar, N., Benbasat, I.: The Influence of Recommendations and Consumer Reviews on Evaluations of Websites. Information Systems Research 17(4), 425–439 (2006)
18. Liu, J., Dolan, P., Pedersen, E.R.: Personalized News Recommendation Based on Click Behavior. In: Proc. of the 14th Int. Conference on Intelligent User Interfaces, pp. 31–40 (2010)
19. Mooney, R., Roy, L.: Content-based Book Recommending Using Learning for Text Categorization. In: Proc. of 5th ACM Conference on Digital Libraries, pp. 195–204 (2000)
20. Pathak, B., Garfinkel, R., Gopal, R.D., Venkatesan, R., Yin, F.: Empirical Analysis of the Impact of Recommender Systems on Sales. Journal of Management Information Systems 27(2), 159–188 (2010)
21. Peterson, E.: Web Analytics Demystified: The Big Book of Key Performance Indicators. Web Analytics Demystified, Inc. (2006)
22. Resnick, P., Varian, H.R.: Recommender Systems. Communications of the ACM 40(3), 56–58 (1997)
23. Schafer, J.B., Konstan, J., Riedl, J.: Recommender Systems in E-Commerce. In: Proc. of the 1st ACM Conference on Electronic Commerce, pp. 158–166 (1999)
24. Zhu, F., Zhang, X.: Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of Marketing 74(2), 133–148 (2010)

Exploration of a Multi-dimensional Evaluation of Books Based on Online Reviews: A Text Mining Approach

Tianxi Dong[1,*], Matti Hamalainen[2], Zhangxi Lin[1], and Binjie Luo[3]

[1] Department of Information Systems and Quantitative Sciences,
Rawls College of Business, Texas Tech University,
Lubbock, TX 79409-2101, United States
[2] Aalto University, Helsinki, Finland
[2] Southwestern University of Finance and Economics, Chengdu, China
tianxi.dong@ttu.edu,
matti.hamalainen@soberit.hut.fi,
{zhangxi.lin,luobinjie}@gmail.com

Abstract. With advancements made to the Internet, a considerable increase in the number and types of products available online has come. Yet, the large amount of online consumer reviews may present an obstacle to potential buyers. This study proposes a four-dimensional book evaluation system for use by leading online booksellers, thereby enabling potential buyers to form decisions based on differentiated criteria. This book evaluation system was empirically examined by employing a text mining approach and multivariate regression model. The findings here-in may aid in improving the understanding of the construction of online product evaluation systems.

Keywords: Online review, evaluation system, text mining.

1 Introduction

User-generated online reviews of books sold in electronic stores have greatly promoted the eBusiness. Such user-generated feedback has been a critical factor in the success of online business as well as a versatile platform within academic research [1], [2]. At present, the online review of books is an important source of information to consumers that may substitute or complement other forms of word-of-mouth communications regarding book quality [1]. Separately, the sheer volume in addition to the myriad content of online reviews may raise challenges to book customers. Although reading online reviews may help the potential customer make informed decisions, in many cases the large quantity of textual reviews available for a book can be overwhelming thereby impeding the customer's ability to evaluate the product [3]. For example, the number of reviews for any average book on Amazon.com can exceed several hundred. But for the popular books such as The Girl Who Kicked the Hornet's Nest, the number of reviews has reached 1,237.

* Corresponding author.

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 119–129, 2012.

Related to the idea that a large volume of reviews may impede customer decisions, websites have added the helpfulness a helpful voting mechanism that may encourage users to evaluate the helpfulness of user reviews by simply asking anyone who has read the review to vote on the question "Was this review helpful to you?." Yet, the degree to which this mechanism may aid the consumer remains limited. Even for the most popular online booksellers such as Amazon.com and Barnersandnoble.com, a large portion of book reviews do not receive any helpfulness votes. Therefore, if there are not enough helpfulness votes, customer decision-making facilitated by the helpfulness votes will be limited in effect.

In order to assist the user in more easily finding helpful review information, the construction of a numeric multi-dimensional evaluation of books based on online reviews was investigated. The two leading online booksellers, Amazon.com and Barnersandnoble.com, have only a single overall rating outcome for a book; thereby potentially providing limited information about the book. Additionally, potential buyers might employ a set of criteria with which to make a decision, such as: the usefulness of the book, the writing quality, the richness of the text, as they have different purposes for the book use and different knowledge about the book they are seeking. Normally, online reviews of a book express these evaluative concerns but consumers must read the whole review to find useful information. As such, eBay.com provides a four-dimensional evaluative-approach to indicate a seller's performance in addition to the numeral reputation statistics—the latter based on buyers' rating category of positive, negative, or neutral. The four dimensions of evaluation include "Item as described", "Communication", "Shipping time" and "Shipping and handling charges". Another example is Wikipedia.com. This website encourages users to rate each page they have read with four criteria: Trustworthy, Objective, Complete, and Well-written. Under each dimension, users can evaluate the page on a five-point scale. As users' demand of a product becomes diversified, they may come to know the subjects from different angles thus making their decision based on differentiated criteria.

Our objective was to conceive a text-mining method to build a multi-dimensional evaluation system for the books sold on Amazon.com. Based on previous studies [4], [5], regarding the review of books and clustering result from Enterprise Miner, we have constructed the evaluation system for online books from the dimensions of character, plot, completeness, and well-written. The character dimension measures the author's ability to craft meaningful and interesting fictional character. The plot dimension is used to judge the fictional plot's attractiveness. The completeness dimension aims to evaluate the content of the book reviewed. The well-written dimension focuses on the clarity of the book's content. Based on the afore-described four-dimensional evaluation system for books sold online, book reviews were analyzed in a selected category from Amazon.com by employing text-mining techniques and then testing the validity of the four-dimensional evaluation system by scoring the new book reviews in the same category.

The present study contributes to the eBusiness research discussion by proposing a four-dimensional evaluation system for online books. This evaluation system allows users to apply their favored weights of different dimensions to construct their decisions based upon different choices. In this way, it may be easier for potential buyers to find relevant and meaningful book information. This research also contributes to the text-mining research literature by using a text-mining approach in the study of online book reviews. The online book reviews contain a wealth of textual information, traditionally making it difficult to quantify the ambiguous textual information from online reviews [6]. However, some previous studies have employed content analysis and statistical methods to analyze the online reviews [4], [5]. In this paper, a text-mining approach was employed to extract vectors of singular value decomposition (SVD) from reviewers' text and to further use these newly extracted variables in a multivariate regression model to examine what factors determined the rating in each dimension of the four-dimensional evaluation system.

The remainder of this study is presented as such: Section 2 reviews literature relevant to the presented thesis. Data collection techniques and other methodologies are discussed in Section 3. Empirical results are presented in Section 4. Section 5 considers the limitations of current research while proposing meaningful avenues for further exploration.

2 Relevant Work

2.1 Online Book Review

Online book reviews have become an important source of information to consumers, either through substitution or by complement to other forms of word–of-mouth communication regarding the quality of various products [1]. Academic attention has been given to the significance of this community content on the consumer purchasing decision [1], [2]. Chevalier and Mayzlin (2006) examined the effect of voluntarily-supplied customer reviews on the subsequent sale of books at Amazon.com and Barnesandnoble.com. Customer reviews were overwhelmingly positive on both websites but a greater abundance of reviews as well as a comparatively larger word count per review were identified at Amazon.com [1]. Forman et al. (2008) also examined the relationship between online-reviews and product sale. Their results suggested identifying information about a reviewer's background does shape community members' judgment of products and reviews.

Although there is evidence supporting the efficacy of user-generated online review in influencing a consumer's decision to purchase, greater research is needed to address potential improvements to the utilization by buyers regarding the user-generated online reviews. The abundance of information created by the online review system provides a significant obstacle to the consumer's decision towards purchase. This paper investigates the construction of a multi-dimensional evaluation system for online products, potentially helping prospective buyers to more easily identify useful information.

2.2 Analysis of Online Review by Text Mining

There are several emerging studies [4], [7], [12], [13] that examine the detailed textual information contained within the online reviews. Content analysis was utilized to identify the factors affecting the evaluative content of book reviews in sociology

[4]. As this technique is extraordinarily time-consuming when dealing with a large amount of data, such as the online book reviews, text mining is gaining attention in information systems research. Opinion-mining and sentimental-analysis have previously been used to analyze the textual information found in Amazon book reviews [7]. The authors found that the perceived helpfulness of a review depended on its content as well as that the evaluation did relate to other evaluations of the same product. Cao et al. (2011) employed latent semantic analysis to uncover the semantic factors that may influence the number of helpfulness votes given to a particular review.

Several studies have extracted users' opinions from online reviews to predict products sales [2] and examined the effect of semantic characteristics on the helpfulness votes [7], [8]. However, few studies have realized the usefulness and importance of constructing a multi-dimensional evaluation system for online products by automatically extracting the semantic information of reviews, thereby possibly helping consumer decision-making. This research explores the potential ability to utilize a text-mining technique to quantify the semantic information contained in online book reviews.

3 Research Methods

3.1 Data Collection

Data was collected from Amazon.com, a leading online bookseller. Initially established was the training data-set. Three fictional-genre books in a given theme were selected; each book selected received 10 reviews, which included user ratings and user review comments. Regarding the testing data-set was chosen a separate book with selection criteria identical to model training data-set. A total of 30 user reviews for the latter book were collected. This testing data-set also contains user ratings and user reviews.

3.2 Text Mining Method

Latent Semantic Analysis (LSA)-based text mining method was used to explore the textual information in the book reviews. LSA is a statistical approach that is used to analyze relationships between a set of documents and terms contained therein that produces a set of meaningful patterns related to the documents and terms [9]. In the present study, each book review can be understood as a document, and terms are either a single word or a phrase that appears in the document. A concept generated by LSA may refer to the features of some documents. Within the context defined by this research, a concept such as "extremely realistic" could be a key feature of a review with an attractive plot. LSA was utilized to analyze the sentimental information hidden in book reviews and map them into different evaluative dimensions: character, plot, complete and well-written.

Text Miner node in SAS® Enterprise Miner (EM) 6.2 uses a descriptive terms algorithm to describe the contents of both EM clusters and hierarchical clusters. The clustering results are shown in Table 1. Table 1 displays the descriptive terms for each of the four clusters, Cluster ID, that indicates which cluster that a document is grouped into. All documents are grouped into four clusters. Based on the descriptive terms, the clusters could be mapped into the evaluation dimensions in previous studies [4], [5]: character, plot, completeness, and well-written. The character dimension measures if the author has depicted the characters in the fiction well. The plot dimension was used to judge the attractiveness of the fictional plots. The completeness dimension aimed to evaluate the content of the fiction reviewed, while the well-written dimension focused on examining clarity of the fiction. The Frequency refers to how many documents belong to the same cluster.

Table 1. Clustering results

Cluster ID	Example	Frequency	Corresponding Dimension
1	character; man; officers	4	Character
2	war; Jungle; Life; marines	14	Plot
3	Full; richness	9	Completeness
4	Fluent; detail;	3	Well-written

Note: the theme of the fictions was about Vietnam War;

Based on the four-dimensional evaluation system for books, the present empirical study consisted of three steps:

Step 1: Manually coded each review in the training dataset in accordance with the four-dimensional evaluation system: character, plot, completeness and well-written. An expert in the review of fictional works gave the scores for each book review under four dimensions for the different books. Since the rating scale on Amazon is 1-5, it was kept in the four-dimensional evaluation system consistency. One example of this manual coding is shown in Table 1.

Table 2. An Example of Manual Coding Reviews

Review ID	Character	Plot	Completeness	Well-written
Review 1	3	4	5	5
Review 2	4	4	4	5
…	…	…	…	…
Review 30	3	3	2	1

Step 2: Established a multivariate regression model for predicting the four dimensions with which to score books. In order to investigate the relationship between the information contained in the reviews and the score for each dimension of the book, multivariate regression model was developed. Multivariate regression is a

technique that estimates a single regression model with more than one dependent variable. Multivariate regression is very suitable for our study because the dependent variables are the scores of four dimensions for each book. Equation 1 depicts the basics of the multivariate regression:

$$Y = X\beta + \varepsilon \tag{1}$$

Where Y is a $(n \times p)$ data matrix containing the response variables, X is a $(n \times q)$ data matrix containing the factors, β is a $(q \times p)$ data matrix containing the factor coefficients (model parameters), and ε is a $(n \times p)$ data matrix containing the noise terms. In our study, Y is the scores of four dimensions for each book while X is the textual information identified from the review data-set by Singular Value Decomposition (SVD).

Parsing the document collection (here refers to the reviews data-set) may generate a term-document frequency matrix, in which each entry of the matrix represents the number of times that a term appears in a document. If there is a collection of several thousand documents, the term-document frequency matrix can contain hundreds of thousands of words. It requires too much computing time and space to analyze this matrix effectively. To solve this problem, singular value decomposition (SVD) can be implemented to reduce the dimensions of the term-document frequency matrix by transforming the matrix into a lower dimensional, more compact, and informative form. As SVD is a good summary of the document collection, the document collection can be represented by certain number of SVD factors. Each SVD factor summarizes words in the document with similar characteristics [9]. Therefore, SVD factors represent the textual information contained in the reviews. Further details of SVD and more information about SVD can be found at [10].

Step 3: After the parameter estimates of four models were completed, another set of reviews were applied for scoring. To establish the training dataset, 30 user reviews were collected for a new book in the given category. There were two steps for scoring: 1) Use the text-mining model to obtain the input values from the testing data-set for the multivariate regression model. 2) Apply the multivariate regression model to the final scores under each dimension. Based on each review in the training data-set, the predicted scores for the book under each dimension are given in Table 3.

Table 3. Prediction for Scores under Each Dimension

Review ID	Prediction for Character	Prediction for Plot	Prediction for Completeness	Prediction for Well-written
Review 1	0.93491	0.80636	2.99426	2.0893
Review 2	2.03561	1.87524	2.93898	2.69335
…	…	…	…	…
Review 30	2.83699	2.91483	4.47323	3.72766

4 Results and Discussion

Table 4 shows the weights of different terms in each review. “Role” refers to the attributes of the term, “Freq” represents how many times the term appeared in a certain review, “Num of reviews” indicates in what review the term appeared. “Weight” refers to Term Frequency-Inverse Document Frequency (TF-IDF, here document refers to each review) Weighting. Less frequent terms are more useful in distinguishing reviews. TF-IDF weighting [11] was used to place less weight on more frequent terms, and vice versa. For example, the role, frequency, and weight of the term “book” in review 22 is noun, 66, and .13266 respectively. But for the term “story”, its frequency is 33 but its weight is 0.23299 which is higher than that of term “book”. This means that high frequency leads to low weights. Generally, terms with a high frequency contain less useful information and thus have lower weights.

Table 4. Weights for Terms

Term	Role	Freq	Num of reviews	weight
book	Noun	66	22	0.13266
war	Noun	64	19	0.19329
story	Noun	33	17	0.23299
experience	Noun	23	16	0.22224
author	Noun	29	15	0.25677
book	Verb	21	15	0.22132
character	Noun	29	15	0.24545
reader	Noun	27	15	0.25077
novel	Noun	40	14	0.31579
man	Noun	25	13	0.28483

In order to examine which semantic factor had the key impact on the final score in a certain evaluative dimension, the outputs of multivariate regression model were investigated further to include the parameter estimates, standard error, t values and p values. Table 5 through Table 8 individually presents the results of the investigation, respectively.

Table 5 demonstrates that “_SVD_5” had the lowest p value and the parameter estimates for it was -2.9049, indicating that it had a higher negative impact on the final score under character dimension. This affect may be explained by the inclusion in the reviews of negative descriptors such as “boring” and “misleading”. Similar situations can be found in the other three regression models. Based on the following four tables, it is clear which SVD factor had the higher impact on the final scores for the book in certain evaluative dimensions.

Table 5. Parameter Estimates of Regression Model of Character Dimension

Parameter	Estimate	Error	T Value	Pr > |t|
Intercept	5.7182	0.3340	17.12	<.0001
_SVD_1[a]	-4.7778	1.2949	3.69	0.0210
_SVD_11	-0.9135	0.2745	3.33	0.0292
_SVD_14	-0.8735	0.2846	3.07	0.0373
_SVD_16	1.5312	0.4449	3.44	0.0263
_SVD_3	1.4286	0.3723	3.84	0.0185
_SVD_5	-2.9049	0.2893	10.04	0.0006
_SVD_9	-2.5532	0.3982	6.41	0.0030

[a] The "_SVD_" is the prefix of SVD factor loadings. "_SVD_1" presents the loading on the first SVD factors.

Table 6. Parameter estimates of regression model of plot dimension

Parameter	Estimate	Error	T Value	Pr > |t|
Intercept	6.1293	0.3584	17.10	<.0001
_SVD_1[a]	-5.6965	1.3895	4.10	0.0149
_SVD_12	-1.4968	0.4990	-3.00	0.0400
_SVD_15	-1.4036	0.4441	-3.16	0.0342
_SVD_16	2.1991	0.4774	4.61	0.0100
_SVD_3	2.0979	0.3995	5.25	0.0063
_SVD_5	-2.8441	0.3104	9.16	0.0008
_SVD_9	-2.6166	0.4273	6.12	0.0036

[a] The "_SVD_" is the prefix of SVD factor loadings. "_SVD_1" presents the loading on the first SVD factors.

Table 7. Parameter estimates of regression model of completeness dimension

Parameter	Estimate	Error	T Value	Pr > |t|
Intercept	6.1241	0.4354	14.07	0.0001
_SVD_1[a]	-5.9912	1.6881	3.55	0.0238
_SVD_11	-0.8422	0.3579	2.35	0.0782
_SVD_15	-1.3208	0.5395	2.45	0.0706
_SVD_7	1.4142	0.4869	2.90	0.0439
_SVD_3	2.3675	0.4853	4.88	0.0082
_SVD_5	-2.7563	0.3771	7.31	0.0019
_SVD_9	-3.1928	0.5191	6.15	0.0035

[a] The "_SVD_" is the prefix of SVD factor loadings. "_SVD_1" presents the loading on the first SVD factors.

Table 8. Parameter estimates of regression model of well written dimension

Parameter	Estimate	Error	T Value	Pr > |t|
Intercept	5.1964	0.5488	9.47	0.0007
_SVD_21[a]	0.8600	0.4005	2.15	0.0983
_SVD_11	-0.9135	0.2745	3.33	0.0292
_SVD_24	0.8634	0.3906	2.21	0.0916
_SVD_14	-0.6196	0.4676	1.33	0.2558
_SVD_15	-1.0413	0.6801	1.53	0.2005
_SVD_17	1.0701	0.6840	1.56	0.1927
_SVD_19	0.6381	0.4202	1.52	0.2035

[a] The "_SVD_" is the prefix of SVD factor loadings. "_SVD_21" presents the loading on the 21st SVD factors.

To test if the multi-dimensional values were consistent were Amazon's overall rating information, the original overall rating that each review received was compared with the combination of scores under each dimension. To make it more intuitive, this difference was qualified by computing the deviation. The equation 2 for the computation is listed below:

$$\text{Deviation} = \sqrt{(\text{Overall Rating} - \text{Combination Scores})^2} \qquad (2)$$

As shown in Table 9, the average deviation from the overall rating to combined scores is 1.4. But the difference between their average values is 1.3, which is smaller than the deviation. This indicated that the deviation can be reduced while the amount of the reviews increases. Even though the deviation is a little bit larger than expected (within 1), the reasons could be the small sample sizes, ignorance of the correlation between different evaluation dimensions. We directly summed the scores on four dimensions without considering the weights for different dimensions.

Table 9. Deviation between overall rating and predicted combination rating

Review ID	Overall Rating	Combination of scores in four-dimensions	Deviation
Review 1	5	4	1
Review 2	4	4	0
…	…	…	…
Review 30	3	1	2
Average	4.4	3.1	1.4

5 Conclusion

The present study examined an important yet-not-well researched academic issue concerning the multi-dimensional evaluation system for online bookstores. By analyzing readers' review comments, this multi-dimensional evaluation system may be of potential help in solving customer confusion caused by an information-overload for online book buyers. It may allow potential buyers to make a decision based on differentiated criteria. This issue was addressed by examining the multi-dimensional evaluation system for fictional-genre books. Here proposed is a four-dimensional evaluation system, presenting the numerical evaluations of character, plot, completeness, and well-written. Based on the idea of this book evaluation system, the text mining method and a multivariate regression model was used to investigate online reviews of fictional-genre books with the data collected from Amazon.com. Also examined were the semantic factors that may have impacted each dimension using a text mining approach and multivariate regression model. The approach describe herein can be easily applied to the multi-dimensional evaluation for other types of books. Its main contribution is that it can convert previously single-dimension evaluation to a multi-dimensional evaluation system with the existing online review system for the book. The idea potentially might be further generalized to the online review systems for other types of electronic stores.

This exploratory research retains some limitations. Limited by the sample size, the deviation between overall rating and combined scores is not satisfactory. Our training data-set only included 30 observations, thus leading to a large variance from the sample. In future research, the sample size should be expanded to reduce the statistical deviation. Another limitation of this research is that the evaluation system for the book genre of fiction was investigated. When applying this approach to other products or services, there may be new challenges. Even though different products or services require different evaluation criteria, our methodology that combines text mining techniques with multivariate regression modeling can be easily replicated and generalized to analyze other online products or services.

There are a series of interesting directions for future research. Since the deviation between the overall rating and the combined scores is a little larger in this study, how to make the multi-dimensional values consistent with the use of the Amazon overall rating information is worth further research. Separately, some websites, such as Amazon.com and eBay.com provide a helpfulness voting mechanism, how to weigh the inputs with regard to the information of helpfulness of the reviews needs further consideration. Finally, the evaluative dimensions could be automatically identified from the online reviews by employing certain opinion summarization models.

Acknowledgments. We thank the anonymous reviewers for their useful comments. This paper is supported by The National Natural Science Foundation of China (NSFC Program No.71072037).

References

1. Chevalier, J.A., Mayzlin, D.: The effect of word of mouth on sales: online book reviews. Journal of Marketing Research 43(3), 345–354 (2006)
2. Forman, C., Ghose, A., Wiesenfeld, B.: Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets. Information Systems Research 19(3), 291–313 (2008)

3. Liu, Y., Huang, X.J., An, A.J., Yu, X.H.: Modeling and predicting the helpfulness of online reviews. In: Eighth IEEE International Conference on Data Mining, Piscataway, NJ, pp. 443–452 (2008)
4. Snizek, W.E., Fuhrman, E.R.: Some factors affecting the evaluative content of book reviews in sociology. The American Sociologist 14, 108–114 (1979)
5. Nielsen, S.: Reviewing printed and electronic dictionaries: A theoretical and practical framework. In: Nielsen, S., Tarp, S. (eds.) Lexicography in the 21st Century, pp. 23–41. John Benjamins, Amsterdam (2009)
6. Pavlou, P.A., Dimoka, A.: The nature and role of feedback text comments in online marketplaces: implications for trust building, price premiums, and seller differentiation. Information Systems Research 17(4), 392 (2006)
7. Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., Lee, L.: How opinions are received by online communities: A case study on Amazon.com helpfulness votes. In: Proceedings of the 18th International Conference on World Wide Web, pp. 141–150 (2009)
8. Cao, Q., Duan, W.J., Gan, Q.W.: Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. Decision Support Systems 50, 511–552 (2011)
9. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6), 391–407 (1990)
10. GSL Team: Chapter 13.4 Singular Value Decomposition. GNU Scientific Library. Reference Manual (2007)
11. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Communications of the ACM 18(11), 620 (1975)
12. Kim, S.M., Pantel, P., Chklovski, T., Pennacchiotti, M.: Automatically assessing review helpfulness. In: Conference on Empirical Methods in Natural Language Processing, Morristown, NJ, pp. 423–430 (2006)
13. Liu, J., Cao, Y., Lin, C.Y., Huang, Y., Zhou, M.: Low-quality product review detection in opinion summarization. In: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 334–342 (2007)

The Investigation of Online Reviews of Mobile Games

Shu-Chun Ho and Yu-Chung Tu

National Kaohsiung Normal University,
No. 116, Heping 1st Rd., Lingya District, Kaohsiung, Taiwan 802
sch@nknu.edu.tw
Matthieu.tu@gmail.com

Abstract. The increasing computing capacity and popularity of smartphones has stimulated great demand for mobile applications, especially for mobile games. When consumers have a variety of choices of mobile games, online reviews become critical information for consumers in making decisions. Online reviews show the important characteristics of mobile games that consumers care most about. Therefore, the objective of this study is to identify the critical characteristics of mobile games from the online reviews. We conducted content analysis and analyzed 1,485 online reviews of 38 mobile games and identified and classified 2,145 terms into 20 characteristics. We found that the important characteristics of mobile games include fun, information richness, perceived value, after sales services, stableness, and challenge. Our findings provide empirical evidence of the critical determinants of purchasing mobile games which can serve as a precursor of future text mining research. In addition, this study provides practical insights for software developers and for smartphone manufacturers.

Keywords: Consumers' decision, mobile games, online review, software characteristics, word-of-mouth.

1 Introduction

After launching iPhone in 2007, Apple introduced the revolutionary App Store in 2008 which has created a new chapter for the mobile application (mobile app) market. Apple App Store has delivered billions of mobile apps to its users and has established a new business model of mobile application downloads. The hype created by Apple around applications has made iPhone much more personalized and useful for the users. The International Data Corporation (IDC) predicted that downloads of global mobile applications will increase from 38.2 billion in 2011 to 183 billion by the end of 2015 (IDC 2011a). IHS Screen Digest Research predicted that the profit of the four largest mobile app stores (Apple, Google, Nokia and RIM) will account for 77.7% of the total profit of mobile software in 2011, that is, \$3.8 billion US dollars, of which, the Apple App Store dominates with 75% of the market share (IHS Screen Digest research 2011b). According to Gartner (2011a), mobile games account for over 70% of mobile app downloads. Although 60% to 70% of mobile games are "free," the revenue of

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 130–139, 2012.

mobile games worldwide has been predicted to reach 5.6 billion dollars in 2010 and surpass 11.4 billion in 2014 (Gartner 2011a). The hype around mobile app stores has opened this market up and expanded competition among the players in the industry.

Since there are numerous free and trendy mobile games available in app stores, browsing online reviews has become the first step of product information seeking in pre-purchase processes. Consumers usually browse online reviews before they decide whether to download mobile games. The word-of-mouth effect from online reviews has great impact on consumers' purchase decisions because of the non-profit conflict between products and users' online reviews (Scoble and Israel 2006). It is common to see consumers follow earlier buyers' decisions especially when asymmetrical information exists. Although decision making processes are complex and dynamic, precursors' comments and experiences are critical referrals for potential buyers when marking decisions. Therefore, the objective of this study is to identify and classify the key terms in online reviews of the mobile games in the Apple App Store. The classification of characteristics of mobile games serves as the precursor of our further text mining research.

This paper is organized as follows. Section 2 reviews the global market of mobile applications and mobile games, and studies of word-of-mouth. Section 3 addresses our research method, data collection process and research design. Section 4 discusses our data analysis, findings, and the implications of the research results. Section 5 concludes our findings and discusses the contributions, and limitations of this research.

2 Literature Review

This section reviews the development of the mobile platform and global market of mobile applications and mobile games. We further review previous studies of word-of-mouth and online reviews.

2.1 Global Market of Mobile Applications and Mobile Games

IDC predicted that the number of global smartphones sold will reach 472 million in 2011, and rise to 982 million in 2015 (IDC 2011b). Gartner also predicted that the worldwide sales of smartphones will be around 107 million in the second quarter of 2011, accounting for 25% of mobile device sales (Gartner 2011b). The increasing demand and popularity of smartphones indicate the great potential of the mobile application market. The most common activities of smartphone users are to receive short message services (70%), send/receive multi-media services (50.4%), check news/sports/weather (53%), look up direction or maps (50%), send/receive instant messages (e.g. MSN) (41%), and send or receive work email (38%) (Forrester Research 2011). According to Forrester Research, the revenue of mobile applications from customers' downloads in 2015 will exceed $38 billion US dollars

(McCarthy et al. 2011). Around 33% of smartphone users have downloaded mobile applications in the US. The most frequent downloaded categories of mobile applications of smartphone users are games (65%), social networking (46%), news/weather (56%), maps/navigation/search (55%), entertainment/food (38%), banking/finance (31%), sports (30%), shopping/retail (29%), video/movies (25%), and communication (25%) (Nielsenwire 2010). There are four major players worldwide in mobile applications: the Apple App Store, Blackberry App World, Nokia Ovi Store, and Goggle Android Market. Among the four major players, the Apple App store is far ahead in terms of market share and revenue. The sales of the Apple App Store reached $1.78 billion dollars in 2010 with a global market share of 82.7%, with a growth of 132% compared with 2009 (The IHS Screen Digest Research 2011a).

Table 1. The Growth of Global Mobile Applications Store in 2010 and 2009

Store	Share	Revenue (Millions)	Year-over-Year Growth
Apple App Store	92.8 % (2009)	$769	131.9%
	82.7 % (2010)	$1,782	
BlackBerry App World	4.4 % (2009)	$36	360.3 %
	7.7 % (2010)	$165	
NOKIA Ovi Store	1.5 % (2009)	$13	719.4 %
	4.9 % (2010)	$105	
Google Android Market	1.3 % (2009)	$11	861.5%
	4.7 % (2010)	$102	
Total		$828 (2009)	160.2 %
		$2,155 (2010)	

(Source: IHS Screen Digest research 2011a)

Although the revenues of mobile games are predicted to grow significantly year by year, there are only a few studies related to mobile games (Kleijnen et al. 2004; Fin 2005; Schwabe and Goth 2005; Rashid et al. 2006; Salo and Karjaluoto 2007; Demirbilek 2010). Rashiad et al. (2006) developed location-based games via mobile phones. Some researchers studies how to use mobile games in improving learning (Schwabe and Goth 2005; Demirbilek 2010). A few studies investigated the adoption of mobile games and the advertising of mobile games. Solo and Karjaluto (2007) argued that many mobile game developers provide free downloading fees because they can generate revenues from advertising. Based on refined adoption model, Kleijnen et al. (2004) explained the adoption of mobile gaming. However, there is no research studied consumers' review of mobile games.

2.2 Word of Mouth Effects

People usually read online reviews of products or services to make decisions efficiently and effectively (Cheung et al. 2009). Because of information asymmetry, consumers may not have full knowledge of products or services. Thus, word of mouth (WOM) has become an efficient and effective source of getting information. Word of mouth is defined as, “all information communications directed at other consumers about the ownership, usage, or characteristics of particular goods and services or their sellers” (Westbrook 1987; Hennig-Thurau and Walsh 2003). WOM refers to the perspectives from customers’ experiences and evaluations of products or services. Instead of receiving promotional perspectives from companies, consumers can judge their purchase decisions based on customers’ experiences and evaluations. WOM may be affected by customer satisfaction (Richins 1983) and customer satisfaction is influenced by product quality and price. Enterprises can gain benefits and higher evaluation by reducing price, but this strategy only applies for products with moderate quality (Li and Hitt 2010).

The online forum has provided a virtual channel for customers to share comments through the Internet. This new trend of WOM is known as “eWOM” (Cheung et al. 2009). Compared to WOM, eWOM has many distinct characteristics. For example, it is larger scale, available anytime and anywhere, efficient, comparable, and almost costless. However, since there is no limitation for customers to comment online, un-screened information might decrease the credibility of eWOM (Cheung et al. 2009). Previous studies showed that negative eWOM usually have greater impact than positive ones (Cheung et al. 2009). Consumers are usually suspicious of positive eWOM because these comments may be distributed by firms in order to boost sales. By contrast, negative eWOM seems to be more reliable even though it may be posted by competing firms. Therefore, finding a way to ensure authority is a critical issue for eWOM. To establish the reputation of online reviews, Apple App Store has stopped selling thousands of mobile applications from one company because of its purposely manipulative comments. Extremely negative comments are less helpful for consumers than fair or positive comments (Mudambi and Schuff 2010).

3 Research Method

This section addresses our research design, data collection process, coding procedures of content analysis, and the reliability and validity of content analysis.

3.1 Research Design

To identify the key terms of online reviews in mobile games, we collected online reviews from the Apple App Store and further conducted content analysis to analyze our data. Content analysis is a quantitative technique which can systematically and objectively identify specific characteristics from the texts and make inferences (Stone 1966; Krippendorf 1980; Weber 1990; Neuendorf 2002; Berelson 1952).

3.2 Coding Procedures of Content Analysis

The coding and analyzing procedures were as follows. First, the analysis unit was the postings of online review comments. Second, we created a classification guide of the characteristics of mobile games. Third, two coders read through the classification guide and made sure each other understood and agreed upon the classification, operational definitions, and examples. Forth, the two coders read the postings of the online review through their iPhones. Based on the classification guide, the two coders read through the texts of the online review comments and identified the key terms. Overall, there were 2,145 terms identified from the 1,485 online reviews, which were classified into 20 characteristics of mobile games. Their percent agreement was 0.950, while reliability was 0.974, which was higher than the recommended 80% as the acceptable reliability (Gottschalk 1995). We further evaluate the validity of classifications by three researchers to arrive at an agreement on the definitions of all characteristics of mobile games. In the coding process, the two coders had several rounds of discussions in finalizing the classifications of all terms.

4 Findings and Discussion

This section shows the descriptive statistics of our data, results, and discussion.

4.1 Descriptive Statistics

There were 1,485 selected online reviews of mobile games in 38 categories. 2,145 terms were identified by two coders. These online reviews were accessed in April 2011. We summarize the descriptive statistics in Table 2 which includes the names of games in each category, version, price, downloaded ranking, average stars/number of ratings, and number of reviews of the selected mobile games. The descriptive statistics can be discussed as follows. First, the selected mobile games are priced from 0.99 to 15.99 US dollars. Final Fantasy III (Role Playing) is priced the highest, at $15.99, among the sample, but still ranked number 1 in the downloaded ranking. Around half of the mobile games are priced at $0.99. The average price of the selected sample is $2.62. Second, half of our selected mobile games ranked in the top 3 most downloaded in their categories in the Apple App Store Taiwan. The lowest downloaded ranking within our selected sample is 27. Only 6 of the selected games are not ranked in the top 10. The average downloaded ranking is 6.21. Third, 8 selected mobile games have the highest customer ratings, 5 stars. There is one mobile game, AFROian Jewels (Dice), which has the lowest customer rating, 2.5 stars. The averaged numerical customer ratings are 4.12 stars. Forth, Angry birds (Action) has the greatest number of reviews, 870, and Vocabulary Bubble (Word) has the fewest number of reviews, 2. The average number of reviews is 179.

Table 2. Descriptive Statistics of Selected Mobile Games

Categories	Name of Game	Price	Download Ranking	Stars / Rating	No. of Reviews
Action	Angry Birds	0.99	2	5/130	870
	Fruit Ninja	0.99	3	4.5/136	458
Adventure	Infinity Blade	5.99	2	5/325	794
	Devil May Cry 4 refrain	4.99	7	4/84	68
Arcade	iSlash	0.99	10	5/387	581
	STREET FIGHTER IV	4.99	18	4.5/18	303
Board	Bubble Bust!	0.99	2	4.5/63	34
	Gyrotate	0.99	18	3/293	39
Card	UNO™	0.99	2	4/14	20
	i.Game Big2	1.99	5	3/8	25
Casino	Mahjong World	0.99	1	3/55	39
	Fruit MaTsai	0.99	5	3.5/76	45
Dice	Nintaii	0.99	1	3.5/24	3
	AFROian Jewels	0.99	2	2.5/94	4
Education	Memory Planet	0.99	5	3.5/132	41
	iGirlz Alexis	1.99	3	3.5/21	7
Family	Cut the Rope	0.99	1	5/116	657
	Bejeweled2 + Blitz	0.99	3	4/241	81
Kids	Flower Garden	1.99	27	4.5/46	144
	The Creeps!	1.99	16	5/673	164
Music	Tap Tap Revenge 3	1.99	2	4/76	148
	Bongo Touch	0.99	1	4.5/234	155
Puzzle	A Monster Ate My Homework	0.99	16	5/62	92
	Cat Physics	0.99	20	4.5/731	454
Racing	Real Racing 2	6.99	4	4.5/30	53
	Need For Speed™ Undercover	2.99	8	4.5/251	114
Role Playing	Final Fantasy III	15.99	1	5/118	89
	Dungeon Defenders: First Wave	2.99	12	3/81	62
Simulation	Burger Queen	0.99	2	4.5/146	253
	Cooking Mama	6.99	9	3.5/149	29
Sports	HOMERUN BATTLE 3D	4.99	2	4/715	150
	Baseball Superstar®2011 pro	4.99	8	4/70	48
Strategy	Plants vs. Zombies	2.99	2	5/220	650
	Paladog!	1.99	4	4.5/50	91
Trivia	Aha Here	0.99	1	4/31	21
	Are You Still Sober	0*	4	3/81	12
Word	Vocabulary Bubble	2.99	1	4.5/77	2
	Chinese Idiom Bubble	1.99	6	4/74	16
Average		2.62	6.21	4.12/161.37	179.37

Note: *There is one free mobile game because we were unable to find paid mobile games with the online reviews in the Trivia category. “Are you Still Sober” has more online reviews.

4.2 Results and Discussions

We identify the 2,145 terms of mobile games and classify them into the 20 characteristics of mobile game. Our findings are as follows. We found the most frequent terms appearing in the online review of mobile games are fun (29.19%), information richness (12.5%), perceived value (9.49%), after sales service (8.37%), stableness (6.47%), and challenge (4.97%) (See Table 3). First, fun is defined as something that gives us enjoyment and pleasure (Prensky 2001). Fun is the most frequently appearing term which indicates what customers care most about is whether the mobile game is enjoyable, entertaining, and pleasurable. Second, information richness is the second most frequently appearing term. Information richness refers to the ability of information to change understanding within a time interval (Daft and Lengel 1984; Lee 1994). In this study, information richness measures four elements: the contexts of games, music and sound effects, the frame of the game, and complexity and endurance of playing. Third, perceived value is the third most frequently appearing term in the online review. Perceived value is a subjective evaluation which refers to customers' perceived value of a product rather than product quality (Li and Hitt 2010). Forth, customers also care whether mobile games provide good after sales services. Fifth, customers also care about the stableness of mobile games. Stableness refers to the ability of a module to remain largely unchanged when faced with newer requirements and/or changes in the environment (Yu and ramaaswamy 2009). The other appearing terms are expectation (4.35%), stickness (3.85%), ease of use (2.73%), usefulness

Table 3. The Most Frequent Appeared Terms in the Online Reviews (Over 5%)

Game	Fun	Information richness	Value	Service	Stableness
Angry Birds	23	6	4	2	2
Fruit Ninja	16	1	2	0	0
Infinity Blade	10	21	7	4	5
Devil May Cry 4 refrain	6	21	11	6	8
iSlash	27	4	1	3	0
STREET FIGHTER IV	4	3	1	4	1
Bubble Bust!	21	13	6	3	0
Gyrotate	14	3	2	8	2
UNO™	13	0	7	2	0
i.Game Big2	6	6	4	3	1
Mahjong World	8	3	7	11	2
Fruit MaTsai	9	20	9	19	3
Nintaii	2	1	2	1	0
AFROian Jewels™	0	5	0	0	0
Memory Planet	19	3	0	0	0
iGirlz Alexis	2	0	0	1	0
Cut the Rope	31	7	5	0	0
Bejeweled2 + Blitz	5	2	2	15	23
Flower Garden	19	8	0	6	1
The Creeps!	29	10	7	3	0
Tap Tap Revenge 3	12	3	2	4	5
Bongo Touch	28	2	2	0	0
A Monster Ate My Homework	34	5	3	0	1
Cat Physics	32	2	1	0	0
Real Racing 2	12	11	11	1	4
Need For Speed™ Undercover	16	12	10	7	0
Final Fantasy III	5	5	13	2	0
Dungeon Defenders: First Wave	6	5	9	14	22
Burger Queen	30	0	0	5	1
Cooking Mama	9	7	14	7	9
HOMERUN BATTLE 3D	4	1	2	8	13
Baseball Superstar® 2011 pro	14	11	7	0	1
Plants vs. Zombies	13	5	5	2	0
Paladog!	21	6	6	1	3
Aha Here	9	11	2	7	9
Are You Still Sober	7	0	0	0	0
Vocabulary Bubble	1	0	0	0	0
Chinese Idiom Bubble	6	1	6	1	0
Total	523	224	170	150	116
Ratio	29.19%	12.50%	9.49%	8.37%	6.47%

(2.46%), promotion (2.18%), online community (1.51%), accuracy (1.62%), special event (1.51%), style of game (0.89%), innovation (0.78%), and sustainability (0.11%). Customers also care about these characteristics of mobile games. In our analysis, we further analyze the positive and negative feedback in online reviews. In the online reviews, 79.63% of the terms are positive and 20.37% of the terms are negative. It is interesting to notice that positive feedback is much more common than negative feedback.

5 Conclusion

The increasing competitiveness sparks fierce competition among software developers and highlights the importance of understanding the critical characteristics of mobile game. This has motivated us to identify the most frequent terms appearing in the online reviews of mobile games. We selected 2 mobile games in each category of mobile games across 19 categories in the Apple App Store Taiwan which have high representation. Two coders went through 1,485 online reviews and identified 2,145 terms in the online reviews. The most frequently appearing terms are fun, information richness, perceived value, after sales service, stableness, and challenge. This implies that the characteristics that customers care most about in the mobile games are fun, information richness, perceived value, after sales service, stableness, and challenge. The contributions of this research are as follows. First, we define and identify the critical characteristics of mobile games in 2011. Some of the characteristics are from previous studies and some are tailored for mobile games. We applied content analysis and carefully coded the terms. These findings can serve a reference for future text mining research. Second, our findings related to the critical characteristics of mobile games can also serve as a theoretical basis for further quantitative research. Third, our results provide practical insights for mobile game developers and marketers in developing mobile games and developing marketing strategies. However, this study has two limitations. First, we only analyze online reviews from the Apple App Store and do not include online reviews from other platforms. Second, we only analyze the online reviews of mobile games, not other mobile applications. Future research can compare the online reviews in different platforms and among all mobile applications.

Acknowledgment. The authors thank the National Science Council in Taiwan for partially support this research under grant #NSC100-2410-H-017-001-MY2.

References

1. Berelson, B.: Content Analysis in Communications Research. Free Press, Glencoe (1952)
2. Cheung, M.Y., Luo, C., Sia, C.L., Chen, H.: Credibility of Electronic Word-of-Mouth: Informational and Normative Determinants of On-line Consumer Recommendations. International Journal of Electronic Commerce 13, 9–38 (2009)

3. Daft, R.L., Lengel, R.H.: Information Richness: A New Approach to Managerial Beghavior and Organizational Design. In: Cummings, L.L., Staw, B.M. (eds.) Research in Organizational Behavior, pp. 19–233. JAI Press, Homewood (1984)
4. Demirbilek, M.: Investigating Attitudes of Adult Educators towards Educational Mobile Media and Games in Eight European Countries. Journal of Information Technology Education 9, 235–247 (2010)
5. Finn, M.: Gaming goes mobile: issues and implications. Australian Journal of Emerging Technologies and Society 3, 31–42 (2005)
6. Gartner. Gartner Says 428 Million Mobile Communication Devices Sold Worldwide in First Quarter 2011, a 19 Percent Increase Year-on-Year. Gartner Inc. Publishing Gartner Press Info. (2011a), http://www.gartner.com/it/page.jsp?id=1689814 (accessed on May 19, 2011)
7. Gartner. Gartner Says Worldwide Mobile Gaming Revenue to Grow 19 Percent in 2010. Gartner Inc. Publishing Gartner Press Info. (2011b), http://www.gartner.com/it/page.jsp?id=1370213 (accessed on September 5, 2011)
8. Gottschalk, L.A.: Content analysis of verbal behavior: New findings and clinical applications. Lawrence Erlbaum Associates, Inc., Hillside (1995)
9. Hennig-Thurau, T., Walsh, G.: Electronic Word-of-Mouth: Motives for and Consequences of Reading Customer Articulations on the Internet. International Journal of Electronic Commerce 8, 51–74 (2003)
10. IDC. Worldwide and U.S. Mobile Applications, Storefronts, Developer, and In-App Advertising 2011-2015 Forecast: Emergence of Postdownload Business Models (2011a), http://www.idc.com/getdoc.jsp?containerId=228221 (accessed on September 9, 2011)
11. IDC. IDWorldwide Smartphone Market Expected to Grow 55% in 2011 and Approach Shipments of One Billion in 2015 (2011b), http://www.idc.com/getdoc.jsp?containerId=prUS22871611 (accessed on September 9, 2011)
12. IHS Screen Digest research. Apple Maintains Dominance of Mobile Application Store Market in 2010. IHS Publishing isuppli.com. (2011a), http://press.ihs.com/press-release/product-design-supply-chain/apple-maintains-dominance-mobile-application-store-market- (accessed on May 3, 2011)
13. IHS Screen Digest research. Revenue for Major Mobile App Stores to Rise 77.7 Percent in 2011. IHS Publishing isuppli.com. (2011b), http://www.isuppli.com/media-research/news/pages/revenue-for-major-mobile-app-stores-to-rise-77-7-percent-in-2011.aspx (accessed on May 3, 2011)
14. Kleijnen, M., Ruyter, K.D., Wetzels, M.: Consumer adoption of wireless services, discovering the rules, while playing the game. Journal of interactive marketing 18, 51–61 (2004)
15. Krippendorf, K.: Content analysis: An introduction to its methodology. Sage Publications, Beverly Hills (1980)
16. Lee, A.S.: Electronic Mail as a Medium for Rich Communication: An Empirical Investigation Using Hermeneutic Interpretation. MIS Quarterly 18, 143–157 (1994)
17. Li, X.X., Hitt, L.M.: Price Effect in Online Product Reviews: An Analytical Model and Empirical Analysis. MIS Quarterly 34, 809–831 (2010)
18. McCarthy, J.C., Mines, C., Matzke, P., Darashkevich, Y.: Mobile App Internet Recasts the Software and Services Landscape: App Innovation on Phones Will Spread to Cars, Appliances and Entertainment. Forrester Research Publishing Forrester.com. (2011), http://www.forrester.com/rb/Research/mobile_app_internet_recasts_software_and_services/q/id/58179/t/2 (accessed on September 9, 2011)

19. Mudambi, S.M., Schuff, D.: What Makes A Helpful Online Review? A Study of Customer Reviews on Amazon. com. MIS Quarterly 34, 185–200 (2010)
20. Neuendorf, K.A.: The content analysis guidebook. Sage Publications, Thousand Oaks (2002)
21. Nielsenwire. The State of Mobile Apps. The Nielsen Company Publishing Nielsenwire (2010), `http://blog.nielsen.com/nielsenwire/online_mobile/the-state-of-mobile-apps` (accessed on May 3, 2011)
22. Prensky, M.: Digital game-based learning. McGraw-Hill Professional, New York (2001)
23. Rashid, O., Mullins, I., Coulton, P., Edwards, R.: Extending Cyberspace: Location Based Games Using Cellular Phones. ACM Computers in Entertainment 4, 1–18 (2006)
24. Richins, M.L.: Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study. Journal of Marketing 47, 68–78 (1983)
25. Salo, J., Karjaluoto, H.: Mobile games as an advertising medium: towards a new research agenda. Innovative Marketing 3, 72–83 (2007)
26. Schwabe, G., Göth, C.: Mobile learning with a mobile game: design and motivational effects. Journal of Computer Assisted Learning 21, 204–216 (2005)
27. Scoble, R., Israel, S.: Naked Conversations: How Blogs are Changing the Way Businesses Talk with Customers. John Wiley & Sons Inc. (2006)
28. Stone, P.J., Dunphy, D.C., Smith, M.S.: The General Inquirer: A Computer Approach to Content Analysis. MIT Press, Cambridge (1966)
29. Weber, R. P.: Basic content analysis. Sage Publications, Newbury Park (1990)
30. Westbrook, R.A.: Product/ Consumption-Based Affective Responses and Postpurchase Processes. Journal of Marketing Research 24, 258–270 (1987)
31. Yu, L., Ramaswamy, S.: Measuring the evolutionary stability of software systems: case studies of Linux and FreeBSD. IET Software 3, 26–36 (2009)

Part II

Economics and Organizational Implications of Electronic Markets

Do Hacker Forums Contribute to Security Attacks?

Qiu-Hong Wang[1], Wei T. Yue[2], and Kai-Lung Hui[3]

[1] School of Management,
Huazhong University of Science of Technology, Wuhan, China
[2] College of Business, City University of Hong Kong, Hong Kong
[3] School of Business and Management,
Hong Kong University of Science and Technology, Hong Kong
qhwang@mail.hust.edu.cn,
Wei.T.Yue@cityu.edu.hk, klhui@ust.hk

Abstract. There has been an increased amount of discussion about firms needing to account for the hacker's perspective in protecting their information assets. However, we still have very little idea about how attack information is disseminated within the hacker community. In this paper, we study the role of hacker forums in disseminating vulnerability information that leads to attacks. We found that the discussions in online hacker forums correlate significantly with the number of cyber-attacks observed in the real world. Furthermore, hacker forums also play a moderating role in disseminating vulnerability and threat information. As cyber security becomes an increasingly prominent issue for firms, our study indicates that there is a need to study the behaviors of the participants in the hacker forum further in order to better understand the risks that they pose.

1 Introduction

Online communication platforms have made technology diffusion and information sharing more practical in recent years. Although there have been many discussions about how such a platform facilitates the enhancement of productivity (Assimakopoulos & Yan, 2006), less has been said about its negative impact. Indeed, information shared online serves as a double-edged sword as it can be exploited for either a good or a malicious intention. This complicated phenomenon has been observed in online hacker forums, where novice posters seek attack scripts, and veteran attackers solicit help to illegally intrude on online information systems, yet at the same time, there are forum participants seeking computing knowledge that would enhance systems protection. Similarly, we have long recognized that there is a need to disclose vulnerability information to facilitate software patching; on the other hand, some experts have also argued that public disclosure vulnerability could lead to more attacks (Cavusoglu, Cavusoglu, & Raghunathan, 2007). Thus, how do we prejudge whether the information and knowledge disseminated online are used by white hats to improve systems protection or are utilized by black hats to penetrate and destroy systems?

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 143–152, 2012.

Even as more and more information security knowledge is disseminated online, its impact still remains relatively unstudied. Our main objective in this study is to understand whether information shared online can lead to more attacks. Specifically, we address this question by linking vulnerability and threat disclosure events to the events taking place in hacker forums, and studying their combined effect on cyber-attacks. We compiled a unique dataset based on different information sources, and linked them to attack data. Our dataset allows us to capture the complexities of online threat, vulnerability and attack phenomena that have not been discussed before.

The availability of the large amount of security information and malicious codes (e.g., viruses, Trojan horses, works, etc.) has substantially reduced the technological barriers and has shortened the incubation period that allows a novice to commit cyber-attacks. Our findings suggest that the discussions in online hacker forums exhibit a significant relationship to the cyber-attacks observed in the real world. We found that the moderating effects of hacker forums on discovered threats were significant but not consistent across different categories. This observation suggests that the role of an online forum as a medium for attack knowledge dissemination is not only determined by economic and social factors but also is shaped by the nature of the hacking techniques (technology) used by attackers.

2 Background

The hacker's perspective is increasingly being recognized as an important issue in the information security literature. Previous studies have examined the best protection strategies against the presence of strategic hacker(s) (Png & Wang, 2009; Galbreth & Shor, 2010; Cavusoglu, Raghunathan, & Yue, 2008). The hacker's perspective is important because it take into consideration how hackers could react to a firm's protection strategy. While there has historically been a greater emphasis on attacker/hacker motives in protecting a firm's assets, a hacker's activities with respect to information sharing have often been considered to be contained in a black box. Furthermore, many of the studies have assumed the possibility of potential attacks by hackers while the nature of the actual hacker behaviors remains relatively unknown.

Given the abundant amount of data available online, researchers began to recognize the potential of using online data sources as a means to understanding hacker behaviors (Mahmood, Siponen, Straub, Rao, & Raghu, 2010). In the area of online warfare and terrorism, the online medium has been recognized as a means to not only facilitate the exchange of information, but also to enhance the collaboration of various parties' coordinated actions (Baskerville, 2006). To date, there have been relatively few studies which utilize online hacker data. Our study is the first in this area to examine the impact of information being shared in hacker forums on the attacks made on firms. Our work is closely related to that of Ransbotham, Mitra and Ramsey (Forthcoming) that studied which type of vulnerability disclosure mechanism would result in fewer vulnerability exploitations. In their study, their main objective was to link vulnerability disclosure to vulnerability exploitation (attack).

In recent years, there are more and more empirical studies that seek an understanding of the problems related to information. For instance, how does software patching behavior affect the vulnerability disclosure cycle (Arora, Krishnan, Telang, & Yang, 2010), or how does information security enforcement affect attack trends (Png, Wang, & Wang, 2008), or how does vulnerability disclosure of a security attack incident affect a firm's market value (Gordon, Loeb, & Sohail, 2010). Our study used a broader lens to examine how information, which is shared in an online forum, could affect the spread of vulnerability and the threat of disclosure online and also of attacks.

3 Methodology and Hypotheses

Vulnerability comes in many different forms. Although IT security professionals and expert hackers are actively attempting to reveal system vulnerabilities, a vast majority of the system owners and users do not participate in the revealing of system vulnerabilities. Thus, there is a need to disseminate vulnerability information. The purpose is twofold: 1) to allow software firms to develop security patches, and 2) to inform the IT community about the need to apply software patches. Likewise, many hackers are waiting for opportunities to exploit newly-disclosed system vulnerabilities with the intention of exploiting these systems before a patch can be applied. Therefore, vulnerability information is also disseminated through the hacker community.

In a typical attack, a hacker goes through a number of stages in the attack process. The first stage involves probing a system to find an unpatched software application service running (Liu, Zang, & Yu, 2006). Each software application is tied to a specific port. Given that our attack data consists of a series of attacks targeting different ports, we group all of the attacks that target a particular port as being the same type of attack. Although vulnerability disclosures could lead to better and faster software patching (Arora, Krishnan, Telang, & Yang, 2010), it could also result in a broader dissemination of attack knowledge, skill sets, and toolsets within the attack community that in turn leads to a greater attack volume (Ransbotham, Mitra, & Ramsey, Forthcoming).

Hypothesis 1: *Vulnerability disclosure on an attack type generates more attacks for that particular attack type.*

While vulnerability disclosure provides first-hand information to hackers and IT security professionals, the information is further disseminated through other sources. Hacker forums are a perfect medium for participants to discuss the various issues tied to vulnerability. The phenomenon is not too different from other types of discussion forums, where discussions (or so-called buzz) are often generated as a result of new events. Thus, we hypothesize that hacker forum discussions will moderate the impact of vulnerability disclosure on attacks.

Hypothesis 2: *There is a positive moderating effect of discussions in hacker forums on Hypothesis 1.*

In addition to vulnerability disclosure, risk assessments of disclosed vulnerabilities are often conducted by security firms in order to determine the level of risks faced by clients. Not all vulnerabilities are created equal, and some types of vulnerabilities may trigger higher risks than others. Risk and threat information, however, can facilitate the making of risk mitigation decisions. On the other hand, hackers can not only search for viable targets but also they can identify the preferred targets that will incur high payoffs. Studies have found more and more hackers attack because of the possibility of obtaining monetary rewards (Evers, 2005). Hence, a report on a vulnerability with a high risk impact could then be more preferred by hackers. As such, we hypothesize that an online threat report should increase the possibility of more attacks on firms.

Hypothesis 3: *The threat report on a particular attack type generates more attacks for that particular attack type.*

Similar to vulnerability disclosure, threat and risk disclosure also signals a release of information that could be further disseminated in hacker forums. Hence, the hacker forums should have a positive moderating effect on Hypothesis 3.

Hypothesis 4: *There is a positive moderating effect of discussions in hacker forums on* Hypothesis 3.

Hacker forum discussions also allow us to associate forum characteristics with attacks. When an attack type is discussed by more participants, it signals that the attack type has received more attention in the hacker community. Furthermore, for reasons such as resolving technical difficulties and pooling resources, some attacks require a higher degree of collaboration among hackers. For example, DDoS type attacks require substantial resources to carry out an attack. Such a need could encourage hackers to pool resources in launching massive coordinated attacks.

Hypothesis 5: *An attack type with more discussions generates more attacks for a particular attack type.*

Hacker forums are comprised of posters with different backgrounds. Previous studies have found that a hacker community follows a very hierarchical order in which novice hackers work their way up in the community rankings by constantly proving their value to the community (Jordan & Taylor, 1998). Hacker forums provide a perfect medium for hackers to hone and demonstrate their abilities with respect to establishing themselves in the community. We therefore expect that an attack type, which has been discussed by more active participants, should result in more attacks.

Hypothesis 6: *An attack type which has been discussed by authors with a more active status in the forum generates more attacks for the particular attack type.*

Figure 1 presents the attack information dissemination model based on our hypotheses.

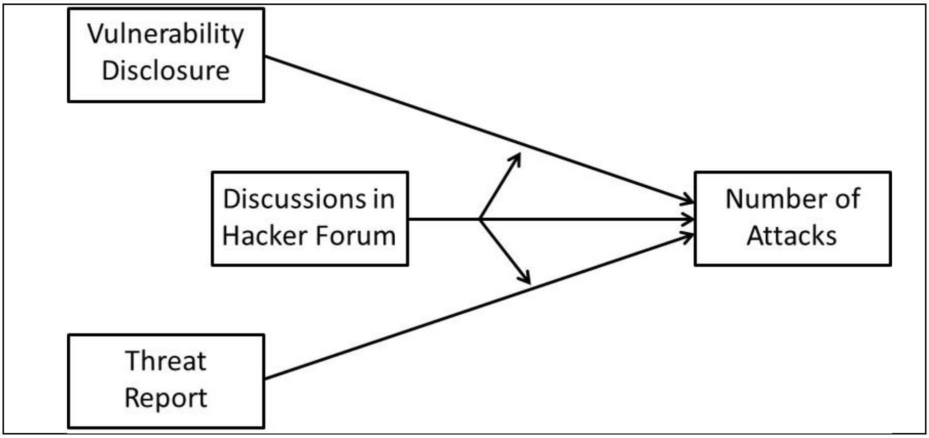


Fig. 1. Attack Information Dissemination Model

3.1 Data Collection

We compiled our dataset from several different sources. The attack data was taken from the SANS Internet Storm Center, where daily attack incidents are captured based on logs. The vulnerability data was taken from the National Vulnerability Database (NVDB) and the Open source vulnerability database (OSVDB). The threat and risk information was taken from the Symantec's enterprise Security Response Unit. The hacker data was taken from two notable online hacker forums. We aggregated the data on an annual basis. Port number was used to link the different sources of data. We generated 9128 observations in our dataset covering the cyber-attacks observed from the years 2002 to 2010. Table 1 provides the descriptive statistics of the main variables used in the subsequent analysis.

4 Analysis

We examined the hypotheses proposed in Section 3 via panel data analysis. All regressions take the double natural logarithmic form and use the robust estimator of variance. Table 2 shows that the interdependent variables are not highly correlated, and thus the model does not have a multicollinearity problem. As a baseline model, we regressed the total volume of attacks in year t on the stock of vulnerabilities and the total number of malwares (categorized as worms, viruses, Trojan horses and others including hoax and hacking tools). Yearly dummy variables were also included to capture the country-constant by year-varied effects.

The results are reported in column "a" of Table 3. It shows that the stock of vulnerabilities and the number of worms is positively and significantly related to the volume of attacks and explains on average more than 27.5% of the variance for each attack type over time. The results provided support for Hypothesis 1 (Hypothesis 3), which states that vulnerability disclosure (a malware report) of a particular attack type

Table 1. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Total volume of cyber attacks	9128	2464399	33500000	0	1.45E+09
Stock of vulnerabilities	9128	0.356376	0.8154924	0	15
No. of worms	9128	0.0606924	1.328253	0	73
No. of Trojans	9128	0.0317704	0.2217461	0	8
No .of viruses	9128	0.0008764	0.0295931	0	1
No. of attack-related threads and their replies	9128	25.15261	356.7223	0	15476
No. of posts made by attack thread authors	9128	94.46166	1202.483	0	44352

Table 2. Correlation

		V1	V2	V3	V4	V5	V6
V1	Stock of vulnerabilities	1					
V2	No. of worms	0.0834	1				
V3	No. of Trojans	-0.0141	0.0511	1			
V4	No. of viruses	0.0552	0.0042	0.0625	1		
V5	No. of attack-related threads and their replies	0.3288	0.1598	0.0491	0.1585	1	
V6	No. of posts made by attack thread authors	0.3988	0.0493	0.0167	0.0627	0.4769	1

generates more attacks for that particular attack type. This observation remains true when we consider the possible impact from hacker forums as being measured by the number of threads (including their replies) which have been discussed in relation to other attacks. As shown in column "b" of Table 3, the coefficient of this hacker forum variable is positive and significant, while the coefficients of the other variables remain similar to the ones observed in column "a". This outcome is also consistent with Hypothesis 5 which states that there is a positive relationship between hacker forum discussions and cyber-attacks.

To investigate the double-edged roles of hacker forums in facilitating user precautions and hacking, we next examined the moderating effects of hacker forums on attack information dissemination and cyber-attacks. The interaction terms of a hacker forum discussion with vulnerability disclosure and various types of malware reports are incorporated into the regression. Column "c" in Table 3 reflects the outcome. Because there is a high correlation with the interaction terms, the coefficient of the hacker forum discussion variable becomes insignificant. However, its interaction terms are all significant. Specifically, the moderating effects are positive for vulnerability disclosure and reports associated with Trojan malware but negative for reports associated with worms and viruses. These results are robust when year dummy variables or the hacker forum discussion term is excluded from regression. As shown in column "d", including the thread author's experience (represented by the number of posts made by individual thread authors in the forum) in the regression does not significantly change the estimated coefficient. Furthermore, the results show that the impact of an author's experience is not significant. In summary, the estimation results presented in columns "c" and "d" support Hypothesis 2, and partially support Hypothesis 4 but do not support Hypothesis 6.

The mixed moderating effects suggest that the question of hacker forums facilitating user precautions or promoting the hacker could very well depend on the nature of the threats. This result is revealing in light of the different classes of threats, that could affect marginal hacking costs and marginal precaution costs differently. For instance, a vulnerability associated with a Trojan horse may require a hacker to find a target-specific means to carry out the attacks, whereas worms and viruses could typically propagate by themselves in order to infect other targets. This may very well explain why hackers incur a higher marginal cost with Trojan horse attacks as compared to attacks that involve worms and viruses. Similarly, compared to viruses and worms, users may incur a higher marginal precaution cost in order to protect Trojan horses as their primary attacking trajectory may be more complicated and subject to changes at any time.

In fact, according to Png and Wang (2009), facilitating a user's use of a precaution is an effective instrument in deterring cyber-attacks when both the marginal hacking costs and the precaution costs are relatively low. Following their proposition, in the scenarios of cyber-attacks related to worms and viruses, the reduced number of attacks could be due to the hacker forum's facilitation in promoting user precaution. Consequently, we observe the negative moderating effects of hacker forums on worms and virus reports.

Table 3. Categorize threats by worm, virus, Trojan horse and others

	a	b	c	d
VARIABLES	Baseline	Hacker forum	Hacker forums' moderating effects	Authors' activeness
Stock of vulnerabilities	1.484*** (0.166)	1.290*** (0.145)	1.248*** (0.144)	1.246*** (0.144)
No. of worms	1.690*** (0.144)	1.639*** (0.179)	1.860*** (0.158)	1.859*** (0.158)
No. of Trojans	0.120 (0.175)	0.101 (0.179)	0.071 (0.184)	0.071 (0.184)
No. of viruses	0.544 (0.677)	0.390 (0.746)	0.759 (0.790)	0.757 (0.790)
No. of attack- related threads and their replies		0.311*** (0.032)	0.077 (0.061)	0.057 (0.073)
No. of attack- related threads and their replies × vulnerability stock			0.255*** (0.068)	0.251*** (0.068)
No. of attack- related threads and their replies × No. of worms			-0.247*** (0.051)	-0.242*** (0.051)
No. of attack- related threads and their replies × No. of viruses			-0.358*** (0.123)	-0.363*** (0.123)
No. of attack threads and their replies × No. of Trojans			0.293** (0.114)	0.298*** (0.114)
Total no. of posts by attack thread authors				0.020 (0.033)
Time-varied effects	included	included	included	included
Constant	-0.006 (0.051)	-0.005 (0.048)	-0.005 (0.048)	-0.005 (0.048)
Observations	9,128	9,128	9,128	9,128
R-squared	0.920	0.922	0.923	0.923
Number of ports	652	652	652	652

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Port and year fixed effects; robust standard errors clustered by port

5 Conclusion

The sharing and dissemination of hacking techniques via online forums brings a significant challenge to how we approach technology governance. Specifically, should policy-makers encourage or limit the phenomenon in which the Internet is used as a platform for sharing security information? Should law enforcement be more vigilant in observing users who disseminate information that threatens the general public? To answer these practical questions, we need to first identify and differentiate between the possible impacts on cyber-attacks from the online diffusion of information security technology among technology enthusiasts and potential security violators. Our study found that there was a significant relationship between the discussions in online hacker forums and the number of cyber-attacks observed in the real world. Furthermore, hacker forums also play a moderating role in disseminating vulnerability and threat information. The moderating effects differ in their security threats.

More hypotheses can be tested in a future version by further differentiating between attack techniques. We can also explore the issues regarding whether or not a participant in a community could be encouraged to carry out deviance behavior (Sherman 1993). We can also further consider whether or not and how the support level of an attack-related topic in the forum affects the frequency of real world cyber-attacks.

Acknowledgements. This research was partially supported by Research Grants 7002631 and 7200173 from City University of Hong Kong, and National Natural Science Foundation of China (71001042).

References

1. Arora, A., Krishnan, R., Telang, R., Yang, Y.: An Empirical Analysis of Software Vendors' Patch Release Behavior: Impact of Vulnerability Disclosure. Information Systems Research 21(1), 115–132 (2010)
2. Assimakopoulos, D., Yan, J.: Sources of Knowledge Acquisition for Chinese Software Enginners. R&D Management 36(1), 97–106 (2006)
3. Baskerville, R.: Hacker Wars: E-Collaboration by Vandals and Warriors. International Journal of e-Collaboration 2(1), 1–16 (2006)
4. Cavusoglu, H., Cavusoglu, H., Raghunathan, S.: Efficiency of V7ulnerability Disclosure Mechanisms to Disseminate Vulnerability Knowledge. IEEE Transactions on Software Engineering 33(3), 171–185 (2007)
5. Cavusoglu, H., Raghunathan, S., Yue, W.T.: Decision-Theoretic and Game-Theoretic Approaches to IT Security Investment. Journal of Management Information Systems 25(2), 281–304 (2008)
6. Evers, J.: Hacking for Dollars (July 6, 2005), `http://CNETNEWS.com`
7. Galbreth, M.R., Shor, M.: The Impact of Malicious Agents on the Enterprise Software Industry. MIS Quarterly 34(3), 595–612 (2010)
8. Gordon, L.A., Loeb, M.P., Sohail, T.: Market Value of Voluntary Disclosures Concerning Information Security. MIS Quarterly 34(3), 567–594 (2010)

9. Jordan, T., Taylor, P.: A Sociology of Hackers. The Sociological Review 46(4), 757–780 (1998)
10. Liu, P., Zang, W., Yu, M.: Incentive-Based Modeling and Inference of Attacker Intent, Objectives, and Strategies. ACM Transactions on Information and System Security 8(1), 78–118 (2006)
11. Mahmood, M.A., Siponen, M., Straub, D., Rao, H.R., Raghu, T.S.: Moving Toward Black Hat Research in Information Systems Security: An Editorial Introduction to the Special Issue. MIS Quarterly 34(3), 431–433 (2010)
12. Png, I., Wang, Q.-H.: Information Security: Facilitating User Precautions Vis-a-Vis Enforcement Against Attackers. Journal of Management Information Systems 26(2), 97–121 (2009)
13. Png, I., Wang, C.-Y., Wang, Q.-H.: The Deterrent and Displacement Effects of Information Security Enforcement: International Evidence. Journal of Management Information Systems 25(2), 125–144 (2008)
14. Ransbotham, S., Mitra, S., Ramsey, J.: Are Markets for Vulnerabilities Effective?. MIS Quarterly (forthcoming)
15. Sherman, L.W.: Defiance, deterrence, and irrelevance: A theory of the criminal sanction. Journal of Research in Crime and Delinquency 30, 445–473 (1993)

A Trust Perspective to Study the Intentions of Consumers to the Group Buying

Deng-Neng Chen[1], Yi-Shan Yang[1], and Yi-Cheng Ku[2]

[1] National Pingtung University of Science and Technology
No. 1, Shuefu Rd., Neipu, Pingtung 912, Taiwan
{dnchen,M9656018}@mail.npust.edu.tw
[2] Providence University
No. 200 Chung-Chi Rd., Salu Dist., Taichung City 43301, Taiwan
ycku@pu.edu.tw

Abstract. With the rapid development of Internet and electronic commerce, more and more innovative online business models have been proposed. Internet group buying, one of the most interesting model, has attracted much attentions. Due to the properties of Internet and group behavior, trust on the Internet and among the group are both important factors that would affect the success of Internet group buying. In this research, we constructed a research model based on trust theory and conducted an experiment to investigate the impacts of trust on Internet group buying. Under different trust scenarios in our experiment, the subjects are invited to participate in a group buying activity. The results show that, the consumers have different intention to attend a group buying activity in different trust scenarios. It also means that different trust perspectives have different impacts on the consumers, and trust is a critical factor for Internet group buying.

Keywords: Internet Group Buying, Trust, Intention of Consumer.

1 Introduction

With the rapidly development of information technology (IT), Internet and electronic commerce (E-commerce, EC) have play important roles in the life of people. More and more consumers spend their time on online shopping, and lots of innovative EC business models have been proposed on the Internet. Transaction on the Internet can cross the limitations of geographical boundaries. As such, the Internet platform can gather more and more consumers with similar interests from different regions. Therefore, the group buying on the Internet is getting popular, such as Groupon.com has been an important company in the world nowadays. The buying costs on the group buying sites could be driven down further by at least 20 percent to 40 percent; therefore, group buying has been brought to the attention of the general public and still remains one of the heated topics within E-Commerce.

The greatest target groups for the traditional group buying are students and white-collar workers in the office. In comparison to traditional group buying, Internet group buying crosses boundaries related to time and distance. In addition, it makes it

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 153–166, 2012.

relatively faster to gather more customers to share the transport costs, obtained from bulk buying to achieve price savings. Other than the foregoing advantages, Internet retailers save overheads over the physical stores and benefit from substantial publicity out of established group buying websites to increase sales volumes. In a nutshell, Internet group buying is beneficial both for buyers and sellers [16].

Although Internet could bring in much assistance to grouping buying, with the inherent properties of Internet, it might also cause many obstacles to group buying. For example, due to the virtualization properties of Internet, the consumers cannot touch the merchandizes and meet with the other group members before they make the decision. This might cause some risk and uncertainty during the group buying process. Current research on consumer behavior has demonstrated that the price is the prime factor to affect group buying behavior. Aside from pricing, the consumer decision making process might be affected by different situational contexts.

Kauffman et al. [13] highlighted the financial risks, psychological risks, and trust in the auction initiator have influences on the consumers' willingness to attend an Internet group buying. Lee and Turban [18] has proposed that trust is an important element in the scenario of perceived risks in E-commerce. Urban et al. [32] further mentioned that trust is a critical component of the consumer decision-making process on the Internet.

Due to the limitation of Internet, given the risks and uncertainty, trust is essential in a virtual collaboration, such as group buying. In addition, several critical factors for group buying have been identified in previous literature, yet no comprehensive empirical study to date has investigated the effects of trust on consumers' willingness in group buying. In this research, we developed an experimental website for Internet group buying to study the impacts of different trust factors on the consumer's intention to participate in a group buying activity. The consumers were invited to a group buying activity under different trust scenarios, and the consumers' perspectives were collected to analyze and discuss the role of trust in group buying.

2 Literature Review

This study focuses on the role of trust in group buying in the context of E-commerce. The relevant literatures on group buying in E-commerce, and trust theory are discussed below.

2.1 Group Buying in E-Commerce

Kauffman and Wang [14] characterized group buying in E-commerce as a group of potential customers, with similar needs arising from one specific good or service over computer networks, exerting their collective power to fight for a lower price. According to Kaufmann and Wang [14], as the size of the group and the purchasing quantities are increasing, the price of the product is decreasing. Consumers can see an array of information over computer networks, such as: latest price, quantities etc. By means of coalition, they can get their products at a lower retail price. Consumers used to bid their products on auction sites where the highest price offered at the end of the auction would get the product. In contrast, in the group buying model consumers can negotiate the selling price, due to their collective power.

In the group buying, as time lapses more buyers are employing the collective bargaining approach. Gradually, a mass number of buyers is built up to obtain a lower price from sellers. Different buyer organizations might be scattered across e-markets, thus the bargaining power of each buyer organization is very limited. Consolidating different buyer organizations based on group similarity, such as products, interests, aggregates their bargaining power. In turn, buyers can obtain a better price to reach another form of group buying.

The buyers wish to get substantial discounts and aggregate their bargaining power; therefore, they are making repetitive purchases from the group with higher bargaining power. To some extent, they are willing to sacrifice some of their own interests and desires to get better deals from the sellers. As such, the success or failure of each transaction lies in bargaining which is the core essence of group buying [4]. The collective bargaining approach is much more effective than the price ceiling in terms of strategic interactions between bidders. An early perspective on the study of bargaining has assumed that all the participants gain the maximization of benefits by attempting to reach acceptable results between buyers and sellers. For buyers, their preferences might fall short of their anticipations, except price. The buyers continue to negotiate the price with the sellers who in turn adjust their price accordingly. In the continuous price negotiation, the final price falls in between the seller's lowest accepting price and the buyer's highest asking price [1]. Apart from the price incentive as a quantity discounts scheme, there would be another factors should be concerned in group buying. In this research, we focus on the trust factors to evaluate the intentions of consumers to group buying.

2.2 Trust

The application of trust theory is very broad and each discipline exhibited a variety of insights into the nature of trust, its methodologies, and the evolving process which demonstrates some degree of ambiguity, multi-dimensional complexity and contradictions [3], [19], [25], [28].

The definition of trust proposed by the literature review on social psychology and marketing [17] exhibits two explicit components: credibility and benevolence [9], [15]. Lindskold [20] further argued that trust "based on a partner's expertise and reliability focuses on the objective credibility of another partner: an expectation held by an individual that the partner can be relied on." Benevolence is defined as the care for a partner's welfare and by extensions, as two partners actively looking for mutual benefits [7].

Rempel et al. [28] found that trust encompasses three essential elements: predictability, dependability and faith. The first element of trust, predictability, is the extent to which the other partner's behaviors can be predicted and the partner continues to act in a trustworthy manner. Thus, the other partner is reliable. The second element of trust, dependability is the extent to which the other partner is dependable in any circumstances: in particular when you are immersed in a harmful situation, you definitely reckon that the other partner is reliable. The final element of trust, faith is the extent to which some signs are unfavorable to you, but you still believe that the other partner is reliable. On the grounds of a review of the relevant literature, how to build trust in an online environment for the first-time consumer has been an open question; hence, the first-time consumer initially only believes in the seller [10]. Trust, a by-product of numerous transactions in online environments, can only be enhanced steadily.

Trust is a phenomenon that carries a social and psychology meaning. Previous research on trust has suggested that trust can be constructed in four levels: individual, interpersonal, relational and social [5]. Researchers have focused on different aspects of trust classifications covering how the nature of trust has developed and its development process. McAllister [24] has asserted that interpersonal trust encompassed two principal forms: "affected-based trust, grounded in reciprocated interpersonal care and concern" and "cognition-based trust, grounded in individual beliefs about peer reliability and dependability."

The above-mentioned concept of trust is multi-dimensional. Based on the theoretical streams developed by Gefen et al. [10], "trust is a product of many antecedents", which in turn foster the formation of trust. These antecedents include personality-based trust, cognition-based trust, knowledge-based trust, calculative-based trust and institutional-based trust. Doney and Cannon [7] provide a comprehensive overview of five evident processes through which trust has developed over years: calculative process, prediction process, capability process and intentionality process as well as transference process.

Drawing from the above literature, Paul and Daniel [26] observed that interpersonal trust comprises of four elements: self-interest, ability, empathy and integrated, which strongly affect virtual collaborative relationship. From the standpoint of operational activities, calculative trust can be understood as self-interest: "the individual in one party perceives that the benefits from directly participating in the collaborative relationship are greater than the costs" [26]. Several other researchers have discussed trust similarly. Mayer et al. [23] identified three factors associated with trust: ability, benevolence and integrity. Amongst them, the factor of ability and the capability process have much in common. Trust is intertwined in the intentionality process involved with the motives hidden in between two parties. Benevolent intention which is consistent to the factor of benevolence allows one party to understand the other party's aims and goals better [7]. Integrity means trustworthiness and honesty. Some past evidence demonstrated that personal integrity affects the levels of trust. Based on this overview of theories proposed by a variety of scholars, this study will define group trust as interpersonal trust comprised of self-interest, ability, empathy and integrated.

Concerning the trust to the seller on the Internet, Jarvenpaa et al. [12] addressed that the size and the reputation of the seller directly affect their levels of trust. Also, the size and the reputation of the selling firm have impacts on the consumers' trust towards them [7]. Furthermore, the verification mechanism is essential to increase the security of the Internet, commonly known as mobile phone certification and credit card certification. Mobile phone certification enables the seller to further contact the buyer to prevent Internet crimes from happening or to reduce the rates of Internet frauds. Credit card certification is used to ensure that the cardholder indeed exists. Meanwhile, each cardholder can accumulate his/her credit profiles to undertake other financial activities such as increase in credit line, overdraft. Of which theory is consistent with the institutional-based trust pointed by Gefen et al. [10]. Thus, we are going to measure the seller's trust as size, reputation and verification mechanism in the research design.

With regard to trust towards the website, trust towards the information system means that the interactions between specific systems and the trustee is deemed as individual computer system. According to Chopra and Wallace [5], four determinants are characterized by the trusted/reliable information system – capability, positive intentions, ethics and predictability. First, capability can be defined as the efficiency with which the system maintains its accuracy and feasibility. Second, positive intentions can be perceived as the absence of malicious codes from the computer system to alleviate users' worries. Third, ethics can be regarded as measures of the copy protection in the computer system to deter hackers' attacks. Last, predictability means that the systems should be sustainable and fit for operations. Therefore, we are going to measure three variables (positive intentions, ethics and predictability) in relation to trust towards the websites.

3 Research Model

The aim of this study concentrates on measuring the effects of trust factors on group buying. We construct one group buying platform from the consumer's viewpoint. Using a trust scale, we are going to observe the factors on which consumers focus. As such, these factors will become reference points for E-commerce developers and group buying participants. The research constructs and the research model are illustrated in Figure 1.

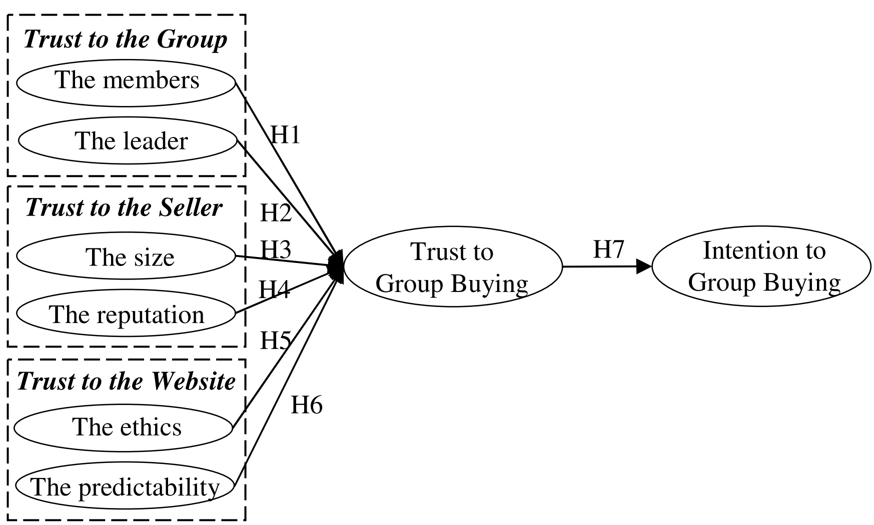


Fig. 1. Research Model

Group buying involves not only buyers and sellers but the platforms. In our research model, we try to investigate the relationship between consumer's trust and intention to group buying. The foregoing factors of trust to group buying are analyzed, including trust to the group, seller, and website.

(1) Trust to group buying

Given the risks and uncertainty inherent in the environments, trust is very crucial, especially in the virtual collaborative relationship [18]. This study considers the trust of consumer to group buying would be the critical factor to the consumer's willingness to participate in group buying.

(2) Trust to the group

Before the consumer would consider joining a group buying activity, he/she must believe the group is worthy to be trusted. The consumer's trust to the group can be divided into the members and the leader in the group [23], [26].

a. Reliability of the Members in the Group: Consumers evaluate costs, profits, the ability to pay installments, familiarity and intentions fostered by the interactions between the group buying members.
b. Reliability of the Leader in the Group: Consumers evaluate the credit certification of the group leader, the ability to handle a number of transactions, the ability to seek massive profits and intentions possessed by the leader of the group [13].

(3) Trust to the seller

The size and reputation of the seller are also critical in a transaction [7], [10], [12].

a. Size: Consumers' perception of the size of the seller in the group buying context.
b. Reputation: Consumers' feelings of feedback ratings of sellers in the group buying context based on comments by other buyers in anticipation of good reputation to enhance the level of trust.

(4) Trust to the website

Chopra and Wallace [5] noted that ethics and predictability should be taken into account. So, the present study has used these criteria to assess the relationship between trust towards the website and the variables. Accordingly:

a. Ethics: Consumer perceptions of the degree of group buying website security have a direct and positive influence on the process of trust development and therefore, consumers expect effective and high security mechanisms in group buying website to be in place.
b. Predictability: Consumers predict that group buying website should perform stably and effectively.

(5) Intention to group buying

Buying intentions imply the possibility of consumers purchasing products/services [11]. Likewise, buying intentions are likely to express intent to purchase products/services [33]. According to Sheth et al. [29], consumer buying intentions would be varying due to product category, home shopping experience, promotions, brand image, product warranty, shop image, and past online shopping experience as well as trust. The preceding factors will influence the willingness to do online shopping [6], [8], [9], [21], [27], [31].

Based on the previous research model and constructs, seven hypotheses are proposed in figure 1. The details of the hypotheses are described as follows:

First, we discuss the influences of interpersonal trust on group buying trust. The hypotheses are listed below:

Hypothesis 1: In a group buying context, the reliability of the member in the group has positive impacts on the trust to group buying.

Hypothesis 2: In a group buying context, the reliability of the leader of the group has positive impacts on the trust to group buying.

Second, we would like to study the relationship between trust to the seller and group buying trust. The hypotheses are listed below:

Hypothesis 3: In a group buying context, the size of the seller has positive impacts on the trust to group buying.

Hypothesis 4: In a group buying context, the reputation of the seller has positive impacts on the trust to group buying

Third, we are going to discuss how group buying trust is determined by trust to the website. As mentioned previously, trust to the website consists of ethics and predictability. Thus, we propose the following hypotheses.

Hypothesis 5: In a group buying context, the ethics of the website has positive impacts on the trust to group buying.

Hypothesis 6: In a group buying context, the predictability of the website has positive impacts on the trust to group buying.

In the aforementioned discussion, the definition of group buying trust is made of the trust to group, the seller and the website. In advance, we propose the final hypothesis:

Hypothesis 7: In a group buying context, the higher the trust to group buying, the consumer will have higher intention to participate in a group buying activity.

4 Research Design

4.1 Experimental System Architecture

As shown in Figure 2, we build a group buying experimental system, in which a randomization process creates a set of parameters to implement a quantitative analysis on trust.

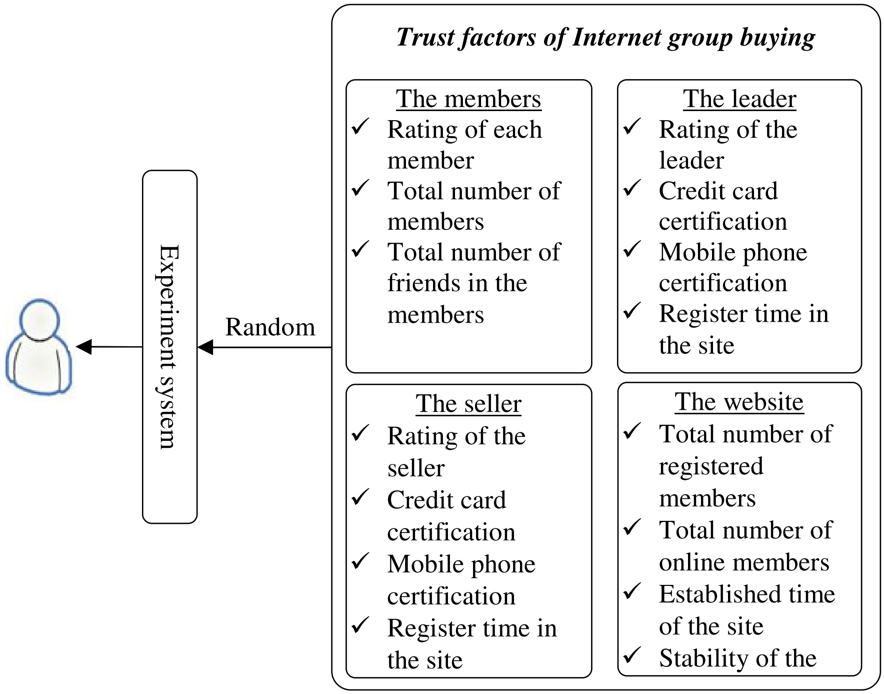


Fig. 2. The trust factors of experimental system

Based on a randomization of sets of variables like the leader, the member, the seller and the website to generate relevant parameters, the system creates mock up websites in different trust scenarios. When the research subject logs in the system, parameters are set by a randomization process and all these data are quantified. Each parameter corresponds to a test score and we subsequently add up each score under each category, i.e. the group, the seller and the website. By doing so, it is much easier to categorize different treatment groups. The main purpose of this system aims to assess how different types of interpersonal trusts (i.e. group, seller, or website) influence consumers. The online questionnaire is implemented in the system, enabling each respondent to answer each questionnaire item, of which parameter specifically retrieved from the system to be included in the questionnaire. The experimental process briefly includes the following four steps.

Step 1: After the subject logs in this experimental website, the system creates different levels (varying scores) of trust parameters by a randomization process.

Step 2: The varying scores of trust parameters will be stored in the database of group buying which in turn generates different trust scenarios.

Step 3: Evaluate his/her feelings towards this trust scenario after browsing over the website

Step 4: Repeat the second and third steps until finishing off the whole process of group buying

4.2 Experimental Design

In this study, we use a factorial design to assign participants to different groups (scenarios). In this design, we have three contextual factors: trust to the group, the seller and the website respectively. Each factor varies across two different modes (high/low trust). In this study, we have a 2 by 2 by 2 factorial design. The subjects would be assigned randomly to each of the 8 possible scenarios to observe the extent to which the subjects are influenced by the different scenarios. The final valid samples are allocated as shown in Table 1.

Table 1. Experiment Design and the Number of Subjects

Trust to website		Trust to group				Total
		High		Low		
Trust to seller	High	High	82	High	70	290
		Low	70	Low	68	
	Low	High	71	High	74	324
		Low	93	Low	86	
Total		316		298		614

The aim of this study focuses on how the group buying environment affects consumers' willingness to participate in group buying, and therefore the selection of experimental products involves the values of the independent variable. In the real world, food products dominate most group buying websites. As a consequence, this study chooses "cake" as the tested product. Some fundamental parameters are set up based on a wide range of parameters available in ihergo.com, which takes a massive share of the domestic group buying market in Taiwan. Furthermore, these fundamental parameters are adapted to the definition of each variable, which we later quantify into a test score within a range of 1 to 10.

4.3 Scale Design

The scale design is divided into six sections based on the research constructs. The first section attempts to understand the disparity between the quality of group buying participants and the environment from the subject's viewpoint on the scenario he/she is in. The second and third sections cover the trust towards the seller and the website, respectively. The subjects should determine whether the questionnaire items match the scenario variables in the preceding three sections. If the questionnaire item matches the scenario variable, this results in high scores, whereby unmatched questionnaire items lead to low scores. After thorough grasp of the above-mentioned three types of trust, subjects should be asked about their trust towards the scenario they are in. Section five covers the questionnaire items about online shopping experiences in order to relate subjects' distribution to their online shopping experience. Subjects should demonstrate varying degrees of acceptability of the particular scenario in which subjects have been allocated in section six. A 5-point Likert Scale is applied to design the questionnaires ranging from “strongly disagree” to “strongly agree.”

5 Data Analysis

5.1 Reliability and Validity Analysis

This study uses factor analysis (principle component analysis) to evaluate questionnaires’ validity. The Varimax method of orthogonal rotation is used for extracting factors with Eigenvalues greater than 1 (as part of total variance explained). The factor loading of each question within each construct should be eliminated if a factor loading is greater than 0.5 to ensure that each construct maximizes its convergent validity. Concerning the justification of questionnaires’ reliability, with Cronbach’s alpha (α), the value of α should be above 0.6. The value of Cronbach α for these six constructs falls between 0.6840-0.9357 which is above 0.6. In the other, the value of Composite Reliability (CR) falls between 0.8175-0.9514 which is higher than the recommended one, 0.7.

Two indicators are applied to measure the reliability of the construct: the factor loading and Average Variance Extracted (AVE). AVE can be defined as the measurement of potential variance within the construct with reference to the amount of variance due to measurement error. If the average variance extracted is higher, it implies that the measurement of potential variance within the construct represents higher reliability and convergent reliability. In this study, the average variance extracted for these six potential variances falls between 0.5408-0.9074 which is greater than the cut-off value of 0.5. Overall, it demonstrates that the questionnaire scaling has considerable convergent reliability. The square root of the AVE for each construct is greater than the absolute value of each construct which indicates that this study demonstrates significant discriminant validity [2].

5.2 Hypotheses Validation

We applied PLS to analyze the data. Coupled with the bootstrap resampling technique, 500 times in this study, we can calculate the path coefficient significance in the model. In this research, we performed eight experimental scenarios to accomplish the aims of this study. The path coefficient analysis and the whole model analysis results are presented in Table 2. The number in the parenthesis is t value, and above the parenthesis is beta coefficient.

Based on the findings of scenarios 1 and 2, with high levels of trust towards both group and seller but low levels of trust towards the website, the reliability of leader has a significantly positive effect on group buying trust. In other words, when there is low trust towards the website, the reliability of the leader is the key factor in promoting group buying trust. From the findings of scenarios 1 and 3, it can be inferred that with high levels of trust towards group and website, and low levels of

Table 2. Results for Hypotheses Testing

	Scenario 1			Scenario 2			Scenario 3			Scenario 4			Scenario 5			Scenario 6			Scenario 7			Scenario 8		
	G	S	W	G	S	W	G	S	W	G	S	W	G	S	W	G	S	W	G	S	W	G	S	W
	H	H	H	H	H	L	H	L	H	H	L	L	L	H	H	L	H	L	L	L	H	L	L	L
H1	0.211 (2.240)			ns			ns			ns			0.314 (2.745)			0.439 (3.342)			0.407 (3.395)			0.338 (2.873)		
H2	ns			0.387 (3.218)			0.405 (2.898)			ns			ns			ns			ns			ns		
H3	ns			ns			ns			ns			ns			ns			ns			ns		
H4	0.333 (2.937)			ns			0.241 (2.010)			0.214 (1.980)			ns			ns			ns			0.354 (2.697)		
H5	0.238 (2.377)			ns			ns			ns			ns			ns			ns			0.284 (2.323)		
H6	ns			ns			0.286 (1.984)			0.226 (1.971)			ns			0.351 (2.574)			0.420 (2.249)			ns		
H7	0.533 (5.807)			0.614 (7.193)			0.608 (8.605)			0.503 (5.523)			0.566 (4.935)			0.588 (6.094)			0.643 (8.272)			0.631 (10.345)		

(G: Group, S: Seller, W: Website, H: High, L: Low, ns: non-significant)

trust towards the seller, the reliability of the leader and the reputation of the seller as well as the predictability of the website have significantly positive effects on group buying trust. Thus, with low levels of trust towards the seller, the reliability of the leader and the reputation of the seller along with the predictability of the website may trigger consumers to take up the activity of group buying.

The findings of scenarios 1 and 5 could reflect that group self-interest has a significantly positive effect on the group buying trust on the condition of high levels of trust towards both seller and website but low levels of trust towards group. Clearly, when the consumers are exposed to with low levels of trust towards the group, assessment of self-interest plays a critical role in gaining trust to group buying.

Results from the scenarios 2 and 4 indicate that the reputation of the seller and the predictability of the website have a significantly positive impact on group buying trust within the context of low level trust towards both website and seller, yet in spite of high level of trust towards group. Group buying trust arises when essential ingredients of the seller's reputation and the website's predictability are in place. Further, the results of scenarios 2 and 6 indicate that when consumers are exposed to high levels of trust towards the seller but low levels of trust towards both group and website, we observe a significantly positive association between group buying trust on one hand and the self-interest for the group and the predictability of the website on the other. Such information suggests that the underlying factors that affect group buying trust are evaluations of self-interest and website predictability.

Observation of the results of scenarios 3 and 7 demonstrates that in the context of high levels of trust towards the website and low levels of trust towards the seller, the self-interest for group members and the size of the seller as well as the predictability of the website have a significantly positive impact on group buying trust. Thus, these findings are best understood as promoting group buying trust.

Our results of scenarios 4 and 8 also indicate that with low levels of trust towards the seller, website and group, three factors including the self-interest for the group, the reputation of the seller and website ethics have significantly positive effects on group buying trust. Combining insights from the results of scenarios 5 and 6 suggests that on the condition of high levels of trust towards the buyer but low levels of trust towards both seller and website, we observe a significantly positive association between group buying trust and the self-interest for the group and website predictability. A couple of noteworthy findings are drawn from scenarios 5 and 7. With high levels of trust towards the website but low level of trust towards both group and seller, a cluster of factors, namely self-interest for the group, size of the seller and predictability of the website, has a significantly positive effect on group buying trust. With low levels of trust towards group, seller and website, the findings of scenario 8 also indicate that group buying trust is strongly determined either by self-interest for the group, seller reputation or website ethics.

6 Conclusions

The findings from five experimental groups demonstrate that the reliability of the members has a significantly positive effect on group buying trust. In particular, when the context of the group environment is set up as low trust, consumers demand high levels of trust towards the group (scenarios 5 to 8). When consumers realize that they have offered little room for profit, the trust to group buying will diminish. On the contrary, consumers' group buying trust strengthens after they realize that they could gain more profits. In addition, results of two experimental groups suggest that the impact of the reliability of the leader on group buying trust is significant. The reliability of the leader is one essential element for consumers to get involved in group buying. The higher the leader's reliability, and the higher the trust to group buying.

The results of all eight experimental groups strongly indicate that the size of the seller has no effect on group buying trust. Therefore, it appears that within the group buying context, consumers do not pay much attention to the size of the seller. However, the findings from four experimental groups suggest that the reputation of the seller has a significantly positive effect on group buying trust. Consumers tend to focus on reputation when they do online shopping. For more than half of the respondents, the higher the reputation of the seller, the greater the trust to group buying.

An overview of the findings from eight experimental groups indicates that group buying trust has a significantly positive impact on the willingness to participate. When the consumers are in risky online environments, they might object to the idea of participating in group buying. On the contrary, if the online environment is favorable and secure to the consumers, consumers would choose to participate in group buying.

Overall, when consumers face up with a variety of group buying environments, they focus on their own interests and assess potential profits and risks. Then, they decide whether to participate in the group buying.

In this research, we make contributions to the understanding of different levels of trust and its association with group buying trust through assessment of different group buying scenarios, either with high or low levels of trust. As a consequence, we could clearly understand what triggers the consumers to participate in a group buying activity. Findings establish a correlation between group buying trust and members, the leader and the website in group buying contexts that are central to real-life organizations and their daily operations. Thus, on the basis of these findings, group buying website designers can tailor their websites to meet potential consumers' needs.

References

1. Anand, K.S., Aron, R.: Group Buying on the Web: A Comparison of Price-Discovery Mechanisms. Management Science 49, 1546–1552 (2003)
2. Anderson, J.C., Gerbing, D.W.: Structural Equation Modeling in Practice: A Review and Recommended Two-step Approach. Psychological Bulletin 103, 411–423 (1988)
3. Butler, J.K.: Toward Understanding and Measuring Conditions of Trust: Evolution of a Conditions of Trust Inventory. Journal of Management 17, 643–663 (1991)
4. Chen, D.N., Jeng, B., Lee, W.P., Chuang, C.H.: An Agent-Based Model for Consumer-to-Business Electronic Commerce. Expert Systems with Applications 34, 469–481 (2007)
5. Chopra, K., Wallace, W.A.: Trust in electronic environments. In: Proceedings of the 36th Hawaii International Conference on Systems Sciences, Hawaii (2003)
6. Crosby, L.A., Evans, K.R., Cowles, D.: Relationship Quality in Services Selling:An Interpersonal Influence Perspective. Journal of Marketing 54, 68–81 (1990)
7. Doney, P.M., Cannon, J.P.: An Examination of the Nature of Trust in Buyer-Seller Relationships. Journal of Marketing 61, 35–51 (1997)
8. Eastlick, M.A.: Consumer Intention to Adopt Interactive Teleshopping. Marketing Science Institute, MA (1996)
9. Ganesan, S.: Determinants of long-term orientation in buyer - seller relationships. Journal of Marketing 58, 1–19 (1994)
10. Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in online shopping: an integrated model. MIS Quarterly 27, 51–90 (2003)
11. Grewal, D., Krishnan, R., Baker, J., Borin, N.: The Effect of Store Name, Brand Name and Price Discounts on Consumers' Evaluations and Purchase Intentions. Journal of Retailing 74, 331–352 (1998)
12. Jarvenpaa, S.L., Tractinsky, N., Vitale, M.: Consumer Trust in an Internet Store. Information Technology and Management 1, 45–71 (2000)
13. Kauffman, R.J., Lai, H., Lin, H.C.: Consumer Adoption of Group-buying Auctions: An Experimental Study. Information Technology and Management 11, 191–211 (2010)
14. Kauffman, R.J., Wang, B.: Bid Together, Buy Together: On the Efficacy of Group buying Business Models in Internet-Based Selling. In: Proceedings of 5th Annual University of Minnesota Electronic Commerce Conference (2001)
15. Kumar, N., Scheer, L.K., Steenkamp, J.E.M.: The effects of supplier fairness on vulnerable resellers. Journal of Marketing Research 32, 54–65 (1995)

16. Lai, H.: Challenging Negotiations to an Electronic Shopping Broker of Virtual Consumer Coalition on The Internet. In: Proceedings of International Conference of the Decision sciences Institute, Athens (1999)
17. Larzelere, R.E., Huston, T.L.: The Dyadic Trust Scale: Toward Understanding Interpersonal Trust in Close Relationship. Journal of Marriage and the Family 42, 595–604 (1980)
18. Lee, M.K.O., Turban, E.: A Trust Model for Consumer Internet Shopping. International Journal of Electronic Commerce 6, 75–91 (2001)
19. Lewis, J.D., Weigert, A.: Trust as a Social Reality. Social Forces 63, 967–985 (1985)
20. Lindskold, S.: Trust Development, the GRIT Proposal and the Effects of Conciliatory Acts on Conflict and Cooperation. Psychological Bulletin 85, 772–793 (1978)
21. Macintosh, G., Lockshin, L.S.: Retail relationships and store loyalty: A multi-level perspective. International Journal of Research in Marketing 12, 498–597 (1997)
22. Matsuo, T., Ito, T., Shintani, T.: A Buyers Integration Support System in Group Buying. In: Proceedings of IEEE International Conference on e-Commerce Technology (2004)
23. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An Integrative Model of Organizational Trust. Academy of Management Review 20, 18–31 (1995)
24. McAllister, D.J.: Affect and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. Academy of Management Journal 38, 24–59 (1995)
25. Muir, B.M.: Trust in automation part I: Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics 37, 1905–1922 (1994)
26. Paul, D.L., McDaniel, R.R.J.: A field study of the effect of interpersonal trust on virtual collaborative relationship performance. MIS Quarterly 28, 183–227 (2004)
27. Raghubir, P., Corfman, K.: When Do Promotion Affect Pretial Brand Evaluations? Journal of Consumer Research 36, 211–222 (1999)
28. Rempel, J.K., Holmes, J.G., Zanna, M.P.: Trust in Close Relationships. Journal of Personality and Social Psychology 49, 95–112 (1985)
29. Sheth, J.N., Newman, B.I., Gross, B.L.: Why We Buy: A Theory of Consumption Values. Journal of Business Research 22, 159–170 (1991)
30. Shim, S., Drake, M.F.: Consumer intention to utilize electronic shopping. Journal of Direct Marketing 4, 22–33 (1990)
31. Tan, S.J.: Strategies for Reducing Consumers' Risk Aversion in Internet Shopping. The Journal of Consumer Marketing 16, 163–180 (1999)
32. Urban, G.L., Sultan, F., Qualls, W.J.: Placing Trust at the Center of Your Internet Strategy. Sloan Management Review 42, 39–48 (2000)
33. Zeithaml, V.A.: Consumer perceptions of price: A means-end model and synthesis of evidence. Journal of Marketing 52, 2–21 (1988)

Effects of Borrower-Defined Conditions in the Online Peer-to-Peer Lending Market

Jiaxian Qiu[1], Zhangxi Lin[1,2], and Binjie Luo[1,2]

[1] Key Lab of Financial Intelligence and Financial Engineering,
Southwestern University of Finance and Economics, Chengdu, 611130, China
jessieqjx@gmail.com
[2] Center for Advanced Analytics and Business Intelligence,
Texas Tech University, Lubbock, TX 79409-2101, USA
{Zhangxi.lin,luobinjie}@gmail.com

Abstract. In online Peer-to-Peer lending market, the borrower-defined conditions of loan requests predetermine the successfulness to receive loans. We analyze the transaction data of PPDai, a leading Peer-to-Peer lending market provider in China. By using the multinomial logit model to investigate the importance of borrowers' decisions and their effects on funding results, we reveal that loan amount, acceptable maximum interest rate, and loan period decided by borrowers significantly influence the loan outcomes. For the unsuccessful listings, the requested loan amount has much more importance than other factors, while for the listings attracting more supply than the requested amount, the borrower's acceptable maximum interest rate are more dominant than other factors to the outcomes. Besides, consistent to prior literature's findings, PPDai borrower's personal information and social capital also play major role in the transactions.

Keywords: Lending, P2P lending, Borrower Decisions, Funding Results.

1 Introduction

As one latest new form of E-commerce, online Peer-to-Peer lending (OP2PL), also called online social lending, allows individuals to lend and borrow money directly among each other without the mediation of a creditor bank institution [1]. It provides versatile online services for reallocating small funds among people, particularly for those with poor credit, which could utilize the large number private idle capital efficiently. Harvard Business Review announced it to be one of the 20 breakthrough ideas for 2009. This market has experienced significant growth since 2005 [2]. Today, there are more than 30 OP2PL marketplaces in more than 10 countries worldwide [3]. Prosper (www.prosper.com), the largest profit market in US, helped its one million members receive over $218 million loans by early 2011; Kiva(www.kiva.org), the largest non-profit OP2PL market, serviced its 876,654 member with the total value of loans more than $192 million by the same time. Different from prior literature, which mainly focuses on the context of Propsper.com in last three years [e.g., 4, 5], this

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 167–179, 2012.

paper is to report our recent study in the PPDai.com, founded in 2007, the largest OP2PL market in china and more comparable to Prosper. PPDai has a pool of subscribers of 400,000 in early 2011, doubled from that of two years ago, and more than 7 million loans contracted each month. PPDai, rooted in a booming economy and inhered with a special oriental culture, demonstrated quite unique features that may help us understand the OP2PL market with a broader view and enhance our knowledge mainly gained from the case of the west.

Although OP2PL market has developed fast during last several years, the successful rate of loans in this kind of market is still very low right now. Take PPDai for example, only about 30% listings on PPDai could successfully fund the loan. So, how to improve the successful rate of the transactions is important for the development of this market, which is also the main motivation of this study.

During the transaction process, borrower's decisions play very important roles on the funding results. In this marketplace, borrower initiates the request of a loan (named as listing) and multiple lenders make bids on it. The borrower in OP2PL market usually makes systematic decisions in loan seeking strategies, such as the loan amount, the acceptable maximum interest rate, and the descriptions of the loan. All the information provided by the borrower should encourage lenders to bid on the listing. Therefore, it is important for borrower to know how her decisions influence on the funding results and how to make efficient decisions to fund the loan with minimum cost under their current situation. While for the OP2PL marketplaces, they should identify the status of the borrower's decision-making in the market and understand the borrowers' behavior to finds proper ways to improve borrower's decisions. Based on the phenomena, this study focuses on the borrower's decisions, exploring the borrower's decision-making in PPDai. The specific objectives are as follows: (1) identify the most important decision variables influencing the funding outcomes; (2) examine the effect of changes in the significant decision variables on the elasticity of funding outcomes; (3) investigate the efficiency of borrower's decisions into different groups according to the listing's demand-supply relationship. The works done by this paper could shed light on optimizing borrowers decisions-making to improve the successful rate of the transactions in the market.

We claim that this study makes several contributions to both academia and practice. First, while OP2PL is a large and growing aspect of commerce on the Internet, relatively little research pays attention to the decision-making from borrowers' perspective except for a few [4, 6], which try to build borrower decision support system from different ways. To our knowledge, the works done in this study becomes unique from previous researches. Second, according to the prior researches, the decisions made by borrowers (i.e., loan amount, interest rate, loan period) are the main factors influencing the funding probability [e.g., 2, 4]. Hence, it is important to find out how these decisions affect the funding results, and how to optimize these decisions, which could be a contribution the practical implementation of such decision support services. Third, previous OP2PL literature has largely been limited to the Western context by utilizing Prosper's public transaction data [7], except for some works done by Qui et al. [3, 8]. While the culture constitutes the broadest effect on many dimensions of human behaviors [9], by doing empirical analysis of Chinese OP2PL market, this paper also presents diversified research outcomes about the same type of market but in a different country with an enriched and unique culture.

The rest of the paper is organized as follows. Section 2 reviews relevant literature. Section 3 is the methodology section, including research scheme, data collection and variable selection. Section 4 covers the empirical model specification and empirical results analyses. Section 5 concludes the results and presents future research tasks.

2 Literature Review

While social lending is a large and growing aspect of commerce on the Internet, it has received only limited attention in the research literature, most of which focusing on empirical studies [5]. Like other online businesses, the fundamental problem in OP2PL is information asymmetry between the lenders and the borrowers [1, 10]. This consequently causes problem of trust, risk control, incentive [11]. How to mitigate the information asymmetric thus becomes a key issue for the OP2PL market. In order to solve this problem, most OP2PL marketplaces build the credit systems which contain the “hard information” credit system and “social capital” credit system. Most OP2PL empirical studies focus on how borrowers’ “hard information” and “social capital information” influence on loan performance, especially the last one.

"Hard information" is a kind of quantity information that can be stored easily and spread objectively [2]. In OP2PL marketplaces, "hard information" credit system provides user’s information, such as credit profile, education, age. When borrowers apply for loans, their personal information is typically requested by the marketplace to provide to lenders to evaluate the credit and riskiness of the borrowers, and then decide whether to bid as well as the interest rate [12]. Iyer [13] finds that lenders infer the most from standard banking “hard” information. Besides, a large number of studies have found that hard information could affect the funding probability, interest rate, default rate, etc. For instance, Lin et al.[2] discover that the lower credit grade the borrower has, the less chance the loan is successful with higher borrowing rates and higher loan default rate; Puro et al. [4] find that loan amount have an negative impact on success rate and interest rate; Collier [12] concludes that the borrower's financial situation (loan/ income ratio) and auction format have impacts on interest rate.

In microfinance literature, asymmetric information risk is mitigated by two principal factors: joint financial liability and personal relationships [14]. Most OP2PL marketplaces build “social capital” credit system to solve the asymmetric information problem. For example, Prosper and PPDai allow their members to build group or friend relationships with others; members of Lending Club can share their backgrounds; Smava and Zopa UK build forums for their members. Much prior literature of OP2PL examines how borrower’s social capital influence P2P loan performance [e.g., 1, 14] based on the Granovetter’s [15] “Embeddedness” social capital theory, and proves that social capital has positive effect on loan performance. Everett [14] finds evidence that higher bidding by borrowers’ social network are associated with lower default rates, and lower interest rates. Lin et al. [2] focus on the relational aspect of networks, and show that borrower’s relational network is significant predictors of lending outcomes (funding probability, default time and interest rate). Lopez et al. [16] conclude that inviting friends and group members to bid on borrower’s listings can increase her chance of getting fund.

However, there are very few studies about people's decisions in this market except several ones, like Kumar [17], Iyer et al. [13], Puro et al. [18], Puro et al. [4] and Wu and Xu [6]. The first three papers analyze the effectiveness of lender's behavior, while the last two focus on the role of borrower. Kumar [17] reports that lenders mostly behave rationally and charge appropriate risk premiums for antecedents of loan default. Iyer et al. [13] find that lenders are able to use available information to infer a third of the variation in creditworthiness that can be captured by a borrower's personal information. While Puro et al. [18] study lender's bidding strategy on Prosper, and show that bidding behavior is not homogeneous among bidders. From the perspective of borrowers, Wu and Xu [6] propose a decision support system (DSS) model which used to recommend loan design to borrower in P2P lending (P2PL-BCS) which uses ontology to set up the knowledge base, and uses intelligent agents to recommend the most optimum lenders to borrower by their expectations. Puro et al. [4] focus on the trade-off between the success rate of loan listings and the interest rate, and try to help borrowers to improve their decision quality. Through calculating the pairwised correlations among a group of decision variables, they identified several key variables that are highly related to the success rate of listings, such as starting rate, amount requested, credit grade, and debt-to-income. Finally, they build a DSS that can provide borrowers with three combinations of predicated interest rates and success rates based on borrowers' own settings for their listings.

In summary, previous research is insufficient, especially on loans' efficiency [6]. However, much of prior empirical literature considers the borrower's decision-making information as the control variables when it examines other problems, such as Lin et al. [2], Collier et al. [12], Everett [14], which could provide some guidance to this study.

3 Methodology

3.1 The Research Scheme

The transaction in this market is conceptualized as a sequential process. First, borrower who wants to borrow money in the market initiates a listing with systematic decisions. Then, lenders decide whether bid on the listing, and the bid amount as well as interest rate when they decide to bid. Finally, the borrower could get the loan if and only if lenders' total bid amount (Supply) is equal to or more than borrower's requested loan amount (Demand), and the bidders with lower bidding interest rate finally fund the loan. From the funding outcomes, based on the relationship of listings' demand-supply relationship, the listings could be categorized into four groups, represented by the variable of ListType in the study. We denote the listings that do not have any bid at all as the first group (ListType=1). The second group of listings has less supply amount than demand amount, and the value of ListType is 2. The supply amount of listing with ListType=3 is exactly equal to the demand amount requested by the borrower. And the fourth group are those listings with more supply amount than demand amount (ListType=4). For the first two groups, listings are failed to fund the loan, while the last two groups of listings successfully fund the loan.

A borrower's decisions in the OP2PL market are complex because she needs to make trade-off among several loan conditions, e.g., the interest rate (cost of the loan) she is willing to pay for lenders, and loan amount. Presumptively, if the borrower delineates a higher interest rate or smaller loan amount, she could have a higher funding probability with a higher cost. On the contrary, if the borrower delineates a lower interest rate or larger loan amount, she could obtain the loan with a lower cost but in the risk of lower funding probability. For borrower, efficient decisions could make the lenders' supply amount exactly equal to her requested amount. Less supply amount than request amount means she could not fund the loan, while more supply amount means she could fund more loan amount with current interest rate or fund the given loan amount with less interest rate. So, based on this point of view, only borrowers of listings belonging to ListType 3 make efficient decisions.

This study is to identify the major borrower decision determinants of loan results, examine the effects of borrowers' decisions, and investigate the market responses to these decisions with regard to these four different types of listings. Based on the empirical results, our study is also to provide suggestions to the borrowers of listings in different groups about how to improve the efficiency of their decisions, and what decisions they should pay more attention on.

3.2 Data Collection and Variable Selection

As a new form of e-commerce in latest few years, the platforms of OP2PL have quit similar business model with each other. Being the leading OP2PL marketplace, PPDai has more than 400,000 members with 7 millions loans each month in early 2011, which could represent the OP2PL market in China. In this research, we randomly collect data from PPDai's website by crawler program. The raw dataset contains 17,211 loan request listings with 55,727 bidding transaction records, and 17,188 subscribers' credit profiles. We remove invalid listing records by comparing the bid count in the listing page to the number of bid which is calculated from bid records, and also removed those records with missing values in key variables. Finally, we obtained 16,748 cleansed listing records with 4,221 listings with bid amount and 12,527 listings with no bid at all. Among the 4,221 listings with bids, there are 1,349 successful ones and 2,872 unsuccessful ones. Considering the available information we could get, we choose the variables mainly guide by the prior literature. Table 1 gives short description of the variables used in this study.

Different from previous studies which largely focus on borrower's loan performance, we analyze the efficient of borrower's decisions from the perspective of listing's demand and supply relationship. Much prior literature utilizes the Prosper transaction data, and some variables used in this study like Repayment are not appeared in prior literature for the different information provided by Prosper.com and PPDai.com. Similar to the credit grade on Prosper.com used in Lin et al. [2], Puro et al., [4] etc., in PPDai, borrower's credit profile concludes two kinds of scores, BCreditS and LCreditS, Besides borrower's decision information and personal information, we also add her social capital information (FrdNbr, FrdbidAmt, Frd2bidAmt) into empirical model as the control variables according to prior literature [1, 2].

Table 1. Variables descriptions

Variables	Description	Reference
Dependent variables		
ListingType	According to the value of DS (Supply amount/ Demand amount), listings are classified into four groups: 1 (DS=0), 2 (DS<1), 3 (DS=1), 4 (DS>1).	
Independent variables (Borrower's Decision information)		
LoanAmt	The loan amount requested by a borrower.	[4, 14]
IR	The acceptable maximum interest rate set by borrowers.	[19]
LoanPeriod	The length of the loan in months.	[1]
LDescription	The length of descriptions of the loan.	[1, 2]
Repayment	Two repayment ways: monthly repayment (1) and pay on due date (0).	
BidType	Three bid modes: Bidding (1), Friendship (2) and offline (3).	[3, 8]
Independent variables (Personal information)		
BCreditS	Borrower's credit score as a borrower.	[3, 8]
LCreditS	Borrower's credit score as a lender.	[3, 8]
SucNbr	Number of previously successful listings.	[1, 3, 8]
FailNbr	Number of previously unsuccessful listings.	[3, 8]
Income	Borrower's monthly income.	[14]
Independent variables (Social capital)		
FrdNbr	Number of borrower's friends.	[2]
FrdbidAmt	The total amount that a borrower's friends bid.	[1, 2]
Frd2bidAmt	The total amount that a borrower's friends' friends bid.	[3, 8]

Based on data exploration outcomes, we log-transform four input variables (LoanAmt, Income, FrdbidAmt and Frd2bidAmt) according to their distribution and relevancy to the output variable. The transformed variables are named with a prefix "Ln" with their original variable names. Table 2 shows the descriptive statistics of all variables.

Table 2. Variables descriptions

Variable	Min	Max	Mean	Std. D	Variable	Min	Max	Mean	Std. D
ListType	1	4	1.38	.769	LCreditS	0	5823	35.97	218.073
LnLoanAmt	3.477	5.041	3.818	.365	SucNbr	0	21	.29	1.270
IR	.0001	.270	.187	.067	FailNbr	0	27	2.76	2.789
LoanPeriod	1	36	7.59	6.694	LnIncome	0	6	2.77	1.495
LDescription	0	5347	33.88	123.080	Lender	0	1	.52	.500
Repayment	1	2	1.03	.174	FrdNbr	0	680	8.22	23.209
BidType	1	3	1.10	.360	LnFrdBidA	0	4.989	.305	.912
BCreditS	0	104	33.62	15.261	LnFrd2BidA	0	5	.350	.963

4 Analysis of Borrower's Decision Issues

We identify the determinants that significantly affect the success rate of listings by logit regression. The Chi-Square of Test of Parallel Line is 531.75, significant under the 5% level, confirming that the slope coefficients of response categories are significantly different from 0. The nature of this categorization calls for the use of multinomial logit model.

4.1 Model Specification

As we discussed in the research scheme part, borrowers of the listings without equal supply amount and demand amount could improve their benefits by adjusting their decisions. Only the borrowers of listings in group of ListType 3 make efficient decisions, for they have equal demand and supply amount under their defined situation. So we set the listings of ListType 3 as the reference category to examine the effects and efficiency of borrowers' decisions. The model is shown as equation(1).

$$Logit\left(\frac{P(ListType=j)}{P(ListType=3)}\right)=\beta_{j0}+\beta_{j1}DV_{ji}+\beta_{j2}HV_{ji}+\beta_{j3}SCV_{ji}=Z_{ji} \tag{1}$$

Where i is the subscription of listings, j is one of the other three categories $(j=1,2,4)$ the listing belonging to relative to the reference one, β_{j0} is the intercept, DV_{ji} are the borrower's decision variables of the listing i, HV_{ji} represent borrower's "hard information" variables of the listing i while SCV_{ji} represent borrower's "social capital information" variables of the listing i, β_{j1}, β_{j2} and β_{j3} are the coefficients vectors for DV_{ji}, HV_{ji} ,and SCV_{ji}, respectively.

There will be 3 equations for each category relative to the reference category, ListType=3. For $j=1,2,4$,

$$P(ListType=j)=\frac{\exp(Z_{ji})}{1+\sum_{j=1}^{4}\exp(Z_{ji})} \tag{2}$$

While, for the reference category, where $ListType=3$,

$$P(ListType=3)=\frac{1}{1+\sum_{j=1}^{4}\exp(Z_{ji})} \tag{3}$$

$P(ListType=j)$ is the probability of being in each of the categories 1, 2, and 4, while $P(ListType=3)$ is the probability of being in the reference category. In the multinomial logit model, the coefficients of the reference group are normalized to 0. Hence, for 4 kinds of listings only 3 distinct sets of parameters can be identified and estimated.

4.2 Results and Analysis

The likelihood Ratio Tests of all dependent variables are statistically significant under 5% level, indicating that neither some predictor variables should be excluded nor some categories should be merged. Table 3 gives the Model-Fitting Information. The value of Chi-Square is 12,690 with P-value less than 0.0001, which means that the explanatory variables are collectively significant in explaining the classification of the listings grouping by the listing's demand-supply relationship. The Pseudo-R Squares and Cox and Snell are 0.532, 0.674, respectively, showing the good-fit for the estimated model.

Table 3. Fitting Statistics of Multinomial Logit Regression

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	DF	Sig.
Intercept Only	25,930	25,960	25,930			
Final	13,340	13,740	13,240	12,690	48	.000

Table 4. The Estimated Results of Multinomial Logit Regression

ListType[a]	Variables	DF	Estimate	Wald	P	OR	Variables	DF	Estimate	Wald	P	OR
1	Intercept	1	1.402	.436	.509		BCreditS	1	-.118	403.331	.000	.889
	LnLoanAmt	1	3.521	133.425	.000	33.824	LCreditS	1	.001	1.710	.191	1.001
	IR	1	-3.754	5.509	.019	.023	SucNbr	1	-.326	10.211	.001	1.386
	LoanPeriod	1	.063	11.761	.001	1.065	FailNbr	1	.155	43.449	.000	1.167
	LDescription	1	-.001	10.210	.001	.999	LnIncome	1	.247	34.193	.000	1.280
	[Repayment=1]	1	-2.643	3.764	.052	.071	[lender=0]	1	.209	1.055	.304	1.233
	[Repayment=2]	0	0[b]	.	.	.	[lender=1]	0	0[b]	.	.	.
	[BidType=1]	1	-2.533	2.278	.131	.079	FrdNbr	1	.043	84.636	.000	1.044
	[BidType=2]	1	-.422	.062	.804	.656	LnFrdBidA	1	-36.122	.	.	2.052E-16
	[BidType=3]	0	0[b]	.	.	.	LnFrd2BidA	1	-6.361	15.758	.000	.002
2	Intercept	1	-5.961	14.417	.000		BCreditS	1	-.031	40.062	.000	.969
	LnLoanAmt	1	3.337	132.905	.000	28.140	LCreditS	1	-.001	16.269	.000	.999
	IR	1	-2.916	3.511	.061	.054	SucNbr	1	-.137	12.333	.000	.872
	LoanPeriod	1	.047	7.277	.007	1.049	FailNbr	1	.147	50.060	.000	1.159
	LDescription	1	.000	2.032	.154	1.000	LnIncome	1	.117	8.939	.003	1.124
	[Repayment=1]	1	-.918	2.048	.152	.399	[lender=0]	1	-.036	.032	.858	.965
	[Repayment=2]	0	0[b]	.	.	.	[lender=1]	0	0[b]	.	.	.
	[BidType=1]	1	-.347	.071	.790	.707	FrdNbr	1	.012	28.048	.000	1.012
	[BidType=2]	1	-1.355	1.098	.295	.258	LnFrdBidA	1	-.764	171.169	.000	.466
	[BidType=3]	0	0[b]	.	.	.	LnFrd2BidA	1	-.796	193.844	.000	.451
4	Intercept	1	1.582	1.187	.276		BCreditS	1	-.012	7.116	.008	.988
	LnLoanAmt	1	-.833	9.116	.003	.435	LCreditS	1	.000	11.501	.001	1.000
	IR	1	7.669	25.184	.000	2,141.36	SucNbr	1	-.002	.006	.937	.998
	LoanPeriod	1	-.033	3.143	.076	.968	FailNbr	1	-.057	7.540	.006	.944
	LDescription	1	.000	1.748	.186	1.000	LnIncome	1	.085	5.728	.017	1.089
	[Repayment=1]	1	1.950	4.372	.037	7.030	[lender=0]	1	.312	2.268	.132	1.366
	[Repayment=2]	0	0[b]	.	.	.	[lender=1]	0	0[b]	.	.	.
	[BidType=1]	1	-1.961	2.056	.152	.141	FrdNbr	1	.006	10.654	.001	1.006
	[BidType=2]	1	-2.582	3.611	.057	.076	LnFrdBidA	1	-.125	4.306	.038	.883
	[BidType=3]	0	0[b]	.	.	.	LnFrd2BidA	1	.446	45.929	.000	1.561

a. The reference category is: 3; b. This parameter is set to zero because it is redundant.

Most of the coefficients in the regression are not equal to 0 statistically significant at the 5% level, except Repayment, BidType and Lender. From the Wald Chi-Square statistics in Table 4, we can compare the weight of effects of those variables on funding results.

Importance of Borrower's Decision Variables

Referring to the borrowers' decisions, the coefficients of LnLoanAmt and IR are statistically significant under the 5% level for all the three equations. Similar to most prior studies find by using Prosper's data, on PPDai, the requested loan amount and acceptable maximum interest rate set by borrower significantly influence the likelihood of getting fund [e.g., 1, 4]. The coefficients of LoanPeriod are statistically significant under 5% level for the ListType 1 and ListType 2, and significant under 10% level for the ListType 4, indicating that time period of the loan is also one important factor affecting the funding results. Besides, the LDescription and Repayment are also significant under 5% level for the first and last equations, respectively.

Ordering the values of Wald Chi-Squares, we get the sequential variables sorted by the weight of effects on funding results among the independent variables. For ListType 1, the importance of LnLoanAmt is only less than the BcreditS. For ListType 2, LnLoanAmt is the third critical factor, less important than borrower's two social capital variables, LnFrd2BidA and LnFrdBidA. However, among the independent variables, IR becomes the second important factors while LnLoanAmt is the fifth one, influencing on the funding results for the ListType 4. Based on these results, we have the following finding.

Finding 1: On PPDai, borrowers' decisions about requested loan amount, acceptable maximum interest rate and loan period, are the most important decision factors influencing on the funding outcomes. Especially, the requested loan amount and acceptable maximum interest rate set by borrower play very critical important roles in the transactions.

Effects of Decisions on Funding Outcomes

In multinomial logit model, the Odds Ratio Estimation gives the marginal effects or partial derivatives are obtained by differentiating equations with respect to the particular independent variable. For ListType 1, the coefficients of LnLoanAmt and LoanPeriod are significant positive, while the coefficients of IR and LDescription are significant negative, under the 5% level. The results imply that the probability of the listing belongs to the group of ListType 3 increasing as decreasing of LnLoanAmt and LoanPeriod, and increasing of IR and LDescription. The odds ratio of LnLoanAmt is 33.824, suggesting that a one percent decreasing of LnLoanAmt leads to about 33.824 times increasing in the probability of classification into the ListType 3 relative to the ListType 1. Contrary to the LnLoanAmt, the odds ratio of IR is 0.023 indicates that a one percent increasing of IR leads to about 0.023 times increasing the successful probability of listing. For ListType 2, the coefficients of decision variables are similar to the ListType 1, except LDescription. The coefficients of LnLoanAmt and

LoanPeriod are significant positive, and the coefficient of IR is significant negative, under the 5% level. The values of OR of LnLoanAmt, LoanPeriod and IR are 28.14, 1.049 and 0.054, respectively. All these results indicating that, decreasing the LnLoanAmt, LoanPeriod and increasing IR will increasing the probability of the listing being classified to the ListType 3. Among these decisions, the funding outcomes are more sensitive of the changing in LnLoanAmt for both ListType 1 and ListType 2. The successful funding probability will be multiple increased by decreasing the value of LnLoanAmt. According to the results, we have the following findings.

Finding 2: On PPDai, the borrowers of unsuccessful listings could increase the successful probability by setting smaller requested loan amount, higher maximum acceptable interest rate or shorter loan period. Among these decisions, the requested loan amount becomes the most important factor for the marginal effect of LnLoanAmt on the funding outcomes is much larger than others.

Comparing to the listings of ListType 3, the listings of ListType 4 have more bid amount than requested amount. Different from the ListType 1 and 2, the coefficients of LnLoanAmt and LoanPeriod are significant negative, while the coefficient of IR is significant positive, under 5% level. The odds ratios of LnLoanAmt, LoanPeriod and IR are 0.435, 2141.36 and 0.968, respectively, which suggest that borrower of ListType 4 could set a larger loan amount, lower interest rate and longer loan period to fund the loan. However, the OR value of IR is 2141.36, indicating that the loan results is very sensitive for the changing of interest rate set by borrowers. Borrower should be very careful when she decides to low down the interest rate.

Finding 3: Comparing to the ListType 3, the borrowers of ListType 4 could set a larger loan amount, lower acceptable maximum interest rate in a longer loan period to fund the loan. However, the latent space for borrower to low the interest rate is very limited for the funding result is very sensitive to the acceptable maximum interest rate set by borrower.

Effects of Borrower's Personal Information and Social Capital

Besides the borrower's decisions, borrower's personal information and social capital also take very important parts on the funding results. The personal information, such as credit profile, income, provided to lenders to evaluate the credit of borrowers, which will influence lenders bidding strategies [2, 12]. Besides, as many prior studies proved, borrower's social capital also influence the funding probability a lot [e.g., 14, 16]. If it is said that borrower's decisions for the requesting loan determines she could fund the loan or not, her personal information and social capital will determine her potential funding capability. The empirical results of our study also proved the findings of prior research.

For ListType 1, BcreditS is the most important factor in affecting the classification for its Chi-Square statistic is the largest one with value of 403.331, while for ListType 2 and 4, borrower's social capital variables are most important factors for the Chi-Square statistics of LnFrd2BidA are 193.844 and 45.929, respectively. The coefficients of social capital and good personal information variables (such as LnFrd2BidA, BCreditS, LcreditS, and SucNbr) are significant negative, and the coefficients of FailNbr are significant positive, for ListType 1 and 2, implying that borrowers of these two types of listings could increase the funding successful rate by providing better personal information and building more social capital. However, the coefficients of ListType 4 are reversed from the last two types, which indicating that the borrowers of the listings of ListType 4 have better personal information and social capital comparing to the listings of ListType 3. Based on these results, we conclude following findings.

Finding 4: On PPDai, borrower's personal information and social capital are the most important factors influencing on the funding outcomes. For unsuccessful listings, in order to increasing the funding probabilities, the first thing borrowers should do is improving their personal information and social capital to lift their funding capacity, and then is making appropriate decisions.

5 Discussion and Conclusions

In this study, by using the transaction data of loans on PPDai.com, we classify the listings into four groups from the perspective of listing's demand-supply relationship: listing without any supply (ListType=1); listing with less supply amount than loan amount (ListType=2); listing with equal supply amount and loan amount (ListType=3); and listing with more supply amount than loan amount (ListType=4).
The empirical results of Multinomial Logit Model shows that, although borrower's personal information and social capital are the most important factors influencing on the funding probability, borrower's decisions also play very important roles. Borrower's decisions of requested loan amount, the acceptable maximum interest rate and loan period, especially the first two, are the most critical decisions affects the loan results.

From the perspective of market demand-supply relationships in PPDai, our analyses reveal the role of borrower's decisions on funding results for different kinds of listings. The results suggest that borrowers of listings with supply not equal to demand could increase their benefit by adjusting their decisions. However, for borrowers of different groups of listings, different aspects they should focus on.

For borrowers of unsuccessful listings (ListType 1 and 2), the most important thing they should do is lifting their funding capacity by improving their credit scores and social capital. Besides, the decision of loan amount requested is also very important according to the Wald Chi-Square value. And the funding probability will increase a lot by decreasing the loan amount, for the marginal effect of LnLoanAmt is around 30. Furthermore, the funding probability of the unsuccessful listings also could be increased by setting lower interest rate and shorter loan period.

For the listings with more supply amount than demand amount (ListType 4), borrowers could improve their benefit by setting a larger loan amount with a lower interest rate and a longer loan period. Borrower's acceptable maximum interest rate becomes most important decisions for the listings in this category. The OR of IR is

2141.36, indicating that the loan result is very sensitive of the interest rate set by borrower, and latent space for borrower to low the interest rate is very limited.

In summary, the results of this study do not only reveal the importance of borrower's decisions and the effects of these decisions on funding results, but also give suggestions on how to improve the efficiency of borrower's decisions. At present, most of the Chinese OP2PL markets do not offer decision support system services for their users. So, based on the findings of this study, designing and building a comprehensive borrower decision support system for OP2PL market is the further work this study will do.

References

1. Greiner, M.E., Wang, H.: Building Consumer-to-Consumer Trust in E-Finance Marketplaces: An Empirical Analysis. International J. of E. Comm. 15, 105–136 (2011)
2. Lin, M., Prabhala, N.R., Viswanathan, S.: Judging borrowers by the company they keep: Social networks and adverse selection in online peer-to-peer lending. In: Western Finance Association 2009 Annual Meeting Paper: Smith School of Business, University of Maryland (2009)
3. Qiu, J., Xu, Y., Chen, D., Zhang, G.: The Roles of Social Capital in Online P2P Lending Markets under Different Cultures: A Comparison of China and America. In: The Tenth International Conference on Electronic Business, Shanghai (2010)
4. Puro, L., Teich, J.E., Wallenius, H., Wallenius, J.: Borrower decision aid for people-to-people lending. Sup. Syst. 49, 52–60 (2010)
5. Chen, N., Ghosh, A., Lambert, N.S.: Auctions for Social Lending: A Theoretical Analysis in Research Paper No. 2078 Stanford Graduate School of Business (2011)
6. Wu, J., Xu, Y.: A Decision Support System for Borrower's Loan in P2P Lending. J. of Comp. 6, 1183–1190 (2011)
7. Xu, Y., Qiu, J., Chen, D.: Profit vs. Non-Profit Business Based on P2P Lending: A Cross-Country Multiple Case Study. In: The Tenth International Conference on Electronic Business, Shanghai (2010)
8. Qiu, J., Xu, Y., Chen, D., Zhang, G.: The effects of Social Capital In Chinese Online P2P Lending Market. In: 2011 International Conference on Innovation and Information Management, Chengdu (2011)
9. Soares, A.M., Farhangmehr, M., Shoham, A.: Hofstede's dimensions of culture in international marketing studies. J. of Busi. Res. 60, 277–284 (2007)
10. Pavlou, P.A., Liang, H.: Understanding and mitigating uncertainty in online exchange relationships: A principle agent perspective. MIS Quarterly 31, 1–32 (2007)
11. Stewart, K.: Trust transfer on the World Wide Web. Org. Sci. 14, 5–17 (2003)
12. Collier, B., Hampshire, R.: Sending mixed signals: Multilevel reputation effects in Peer-to-Peer lending markets. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, Savannah, Georhia, USA (2010)
13. Iyer, R., Khwaja, A.I., Luttmer, E.F.P.: Screening in new credit markets: Can individual lenders infer borrower creditworthiness in peer-to-peer lending? National Bereau of Economic Research: NBER Working Paper No. 15242 (2009)
14. Everett, G.R.: Group Membership, Relationship Banking and Loan Default Risk: The Case of Online Social Lending (2008)

15. Granovetter, M.S.: Economic Action and Social Structure: The Problem of Embeddedness. Amer. J. of Soci. 91, 481–510 (1985)
16. Lopez, S.H.: Social interactions in P2P lending. In: Proceedings of the 3rd Workshop on Social Network Mining and Analysis, New York (2009)
17. Kumar, S.: Bank of One: Empirical Analysis of Peer-to-Peer Financial Marketplaces. In: Proceedings of Thirteenth Americas Conference on Information System, Atlanta (2007)
18. Puro, L., Teich, J.E., Wallenius, H., Wallenius, J.: Bidding strategies for real-life small loan auctions. Dec. Sup. Syst. 51, 31–41 (2010)
19. Berger, S.C., Gleisner, F.: Emergence of Financial Intermediaries on Electronic Markets: The Case of Online P2P Lending. Busi. Res. 2, 39–65 (2009)

Cognitive Elaboration on Potential Outcomes and Its Effects on Employees' Information Security Policy Compliance Intention–Exploring the Key Antecedents

Xue Yang[1], Wei T. Yue[2], and Choon Lin Sia[2]

[1] School of Business, Nanjing University, Nanjing, China
[2] College of Business, City University of Hong Kong, Hong Kong
yangxue@nju.edu.cn, {Wei.T.Yue,iscl}@cityu.edu.hk

Abstract. IS security policy is one of the essential tools to ensure the secure use of information systems and technological assets. To enhance the effectiveness of policy implementation, organizations rely on security training, education and awareness (STEA) programs to help employees understand the IS security issues of the organization. However, different levels of STEA informativeness may have conflicting effects on employees' compliance decisions. In addition, the urgency of a task may also lead employees to abandon the compliance decision occasionally. The existing corporate information security policy (ISP) could also serve as a deterrence message that would influence compliance decisions. An experimental survey was conducted to examine this phenomenon and test the related hypotheses. The results of this study can be used to inform and guide researchers and practitioners as to how to better enforce an IS security policy through better implementation of STEA programs and improved design of ISP in different task scenarios.

Keywords: Security Training, Education and Awareness, Informativeness, Deterrence, Task Urgency, Cognitive Elaboration, Compliance Intention.

1 Introduction

As the use of information systems (ISs) becomes essential in business, protecting the safety of corporate informational assets has introduced many challenges for contemporary organizations. Many of the security challenges are difficult to address because many of the IS usage violations are often carried out by internal employees (Gordon et al., 2006; Warkentin and Willison, 2009). Internal users are often entrusted with the privilege to access some of the most sensitive information of the organization (Shaw et al., 1998), and their detrimental behaviors can result in severe negative impacts on the organization (e.g., corporate liability, loss of credibility, monetary damage) (Cavusoglu et al. 2004). Although most organizations have made progress in using technology-based solutions to eliminate IS security risks (Ernst and Young 2008; PricewaterhouseCoopers 2008), there is still a need to continuously improve information security policies (ISPs) to prevent, detect and react to employees' unauthorized usage (Stoneburner et al., 2002; Whitman et al., 2001).

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 180–194, 2012.

In an increasingly fast-paced and flexible workplace, security policies serve as guidelines for behavior but there is no guarantee that employees will comply (Boss and Kirsch, 2007; Siponen et al., 2007). When giving employees enough privilege to perform their tasks, there is also a need to understand the measures that focus on the behavioral aspect of these users. Thus, an additional layer of protection mechanisms must be implemented to ensure best practices in IS usage. Previous literature often focused on the economic trade-offs in applying controls to prevent breach of information systems (e.g. Yue and Cakanyildirim, 2007) and little attention was directed toward the understanding of the different manifestations of corporate ISP enforcement measures. Our aim in this study is to fill this research gap by examining the extent to which the design of corporate security training, education and awareness (STEA) programs and work-related tasks can influence compliance decisions of the employees.

The implementation of various STEA programs are often aimed at improving ISP compliance behaviors and reducing IS misuse (D'Arcy et al. 2009). These programs may deliver different levels of informativeness in their IS security messages, which may sometimes cause unexpected counterproductive effects. Employees may also occasionally decide not to comply with ISPs when compliance creates inconvenience in work-related tasks (Hui and Hu, 2008). It is not clear if too much exposure to actual security breach incidents would encourage non-compliance behaviors, especially in dire situations incurred by task urgency. Deterrence measures have also been established to prevent employees from committing security abuse (Straub 1990). Hence, in this paper, we focus on STEA program informativeness, task urgency and ISP deterrence and study how they would affect compliance decision. We applied the theoretical lens of cognitive elaboration to uncover the underlying mechanisms of how certain factors affect compliance decisions (Perse 1990; Nenkov et al. 2008).

The survey approach was employed to test the research model. This study contributes to the literature and practice by addressing the joint impacts of the different STEA program features, namely ISP informativeness and deterrence, and work-related task urgency on compliance behavior. Our novel approach builds on extant literature in the area that examined individuals' awareness of the security issues raised by STEA programs (e.g. Bulgurcu et al., 2010) or the effects with the presence of STEA programs (e.g. D'Arcy et al., 2009). Theoretical and practical contributions will be elaborated upon based on the data analysis results.

2 Literature Review and Theoretical Background

Although it has been found that many information system (IS) usage violations are linked to malicious intent, very often work-related factors could also induce such violations. In this paper, we study the circumstances under which an individual may engage in inappropriate IS usage that is deemed to be violating the organizational information security policy (ISP) while fulfilling particular organizational tasks. It is assumed that employees generally having a benign intent may be induced by external factors that motivate them to circumvent or breach the ISP.

IS related security issues are complex and often have a direct impact on organizational welfare. It has long been recognised that resolving security quandaries can improve the operational efficiencies of an organization (Smith and Hasnas, 1999; Harrington, 1996). Extant literature has established that identifying the characteristics of people by applying an ethical lens can be an effective way to deter the misuse of IS (Banerjee, Cronan and Jones, 1998). In particular, this study introduces the possibility that STEA programs, and task and security policy characteristics are critical factors influencing corporate users' ISP compliance decisions through an underlying mechanism of individual cognitive elaboration on potential outcomes. The selection of these key variables of question is generally based on the likelihood that they may effectively raise individuals' cognitive elaborations on various aspects.

2.1 The Security Training, Education and Awareness (STEA) Program

Organizations today promote appropriate IS usage through various STEA programs designed to improve policy awareness and operational knowledge of the employees. Typically, these programs can be designed to serve the informative function. Informative STEA programs provide employees with the technical and procedural knowledge of best practices for IS usage. Technical knowledge provides overall coverage on how a breach of a code of ethics is detected and the likelihood of detection, whereas procedural knowledge is focused on the code of itself (Harrington, 1996).

However, informative STEA programs can be a double-edged sword in inducing unexpected IS usage behaviors given that knowledge is the prerequisite in preventing or executing IS abuse. On one hand, such practices could condition employees to be more aware of the possible ways to breach security and their respective detrimental impacts (D'Arcy et al., 2009). On the other hand, traditionally, excessive individual liberties (e.g. information accessibility) have been traditionally deemed to be harmful to security protection (e.g. national security) (e.g. Sagar, 2009; Sarikakis, 2008; Strickland 2005). Thus, STEA practices offering security information that may not be obtained otherwise could alter an employees' perceived difficulty of executing security breach behavior. This is because they are now exposed to the weaknesses of ISs, and could even elicit IS abuse behavior. In this sense, providing informative STEA programs that can educate those non-IT-savvy employees may not be an ideal means to enforce ISP compliance. Thus, it is thus unclear whether employees may be better off not being aware of the typical methods of circumventing the technical barriers from committing information security breaches.

2.2 Task Characteristics of Workplace

Though STEA has been touted as being effective and practical to deploy in corporate IS security protection practices (D'Arcy et al., 2009), users may still perform noncompliance behaviors mostly due to the cost of the compliance as a work impediment (Bulgurcu et al., 2010). This noncompliance is mostly due to the fact that individuals may be morally open to changes in behavior when facing special situations (Myyry et al., 2009). When employees are required to make a decision on whether to follow ISP, they may also consider the corporate task that is being carried out at the time. In fact, since IS security compliance requirements introduce the need for additional time and effort in completing a task (as compared to there being no ISP), an employee may perceive the compliance behavior as a barrier to productivity (Siponen and Vance, 2010; Warkentin et al., 2004) that would immediately lead to perceptible or actual negative consequences for the employee (Bulgurcu et al., 2010). In some cases, complying with security requirements may even conflict with the employee's primary tasks or be detrimental to an employee's daily job-related tasks and activities (Pahnila et al., 2007). However, if the task at hand is not urgent enough to elicit such negative reactions toward the ISP, an employee may have less difficulty in complying with ISP. Hence, a task characteristic such as its urgency could have a significant impact on an employee ethical model and subsequent security compliance behaviors.

2.3 Deterrence Effects of ISPs

Because some individuals are more sensitive to sanction-based threats, the punishment-as-deterrence doctrine has been widely accepted as a viable approach to preventing policy violation behaviors (Liska and Steven 1999). From the behavioral perspective, general deterrence theory has been adapted from the criminology literature in order to understand the effect of deterrent factors on security policy compliance. Deterrence theory posits that unwanted behaviors can be deterred through the threat of certain, swift, and/or severe punishment (Akers, 1990; Williams and Hawkins, 1986). As the level of punishment certainty and punishment severity is increased, the level of unacceptable behavior decreases (Akers, 1990; Williams and Hawkins, 1986). Previous literature has illustrated the effects of deterrence on illegal computing activities in organizations such as reducing computer abuse (Straub 1990). Furthermore, non-adherence to security policies can be deterred only when certain penalties can be enforced on individuals if any security breach behavior was detected by the organization (Herath and Rao 2009). This further emphasizes the importance of increasing an organization's ability to detect security violation behavior and communicating the information to employees. If employees are aware of the high level of certainty of a penalty for IS misbehavior, their decisions with respect to a security violation would likely to be affected (Herath and Rao 2009). In other words, how to execute sanctions on rules violations is another critical dimension that ensures the effectiveness of deterrence measures (Tyler and Blader 2005).

2.4 Cognitive Elaboration Theory

Beyond identifying the critical factors that influence an organization employees' ISP compliance decisions, we are also interested in uncovering the inner schema of the individual processing of the information security related factors, which can provide a theoretical explanation of the effects of these critical external factors. In particular, the extent to which individuals purposefully spend cognitive elaboration effort on ethical issues has been suggested to influence ethical behaviors in organizational contexts (Street et al. 2001).

Cognitive elaboration in this study is defined as the amount of cognitive effort one puts forth with respect to a target attitude's attributes, merits, and drawbacks (Petty et al., 1995). The target attitude objects can cover many categories of decisions, situations, people, physical objects, or social issues (Fazio et al., 1986), including information security issues in organizations. When under different situations, individuals can engage in different levels of issue-relevant thinking with little (or no) to substantial cognitive effort (Petty, 1995). For example, when the attitude object or issue has greater informational intensity, personal relevance or personal accountability with respect to values, career, relatives, or other considerations (Malaviya 2007), the decision maker is more likely to be motivated to carefully examine the relevant information concerning the attitude object or issue (Petty, 1995). Similarly, in the psychology and marketing literature, the language and the type of metaphor contained in the delivered messages could influence recipients' cognitive elaboration on how information is perceived (e.g., Wyer 2002).

When cognitive elaboration can virtually encompass various aspects, the elaboration on potential outcomes related to an issue is especially important because it can make people conscious of how their decisions could lead to desirable or undesirable outcomes. Such thoughts would then activate self-regulation initiatives (Nenkov et al. 2008). Moreover, research shows that when cognitive elaboration on potential outcomes increases, the strength and certainty of an individual's attitude toward a particular object or action also increases (Smith et al. 2008). For example, if individuals expend a higher level of cognitive elaboration on the potential outcomes of a particular decision, they are more likely to be in the central information processing mode (as suggested by the elaboration likelihood model). Such a development could lead to a higher probability of recognizing the ethical issue, establishing an ethical intent and ultimately engaging in ethical behavior (Street et al. 2001).

Table 1. The Definitions of All the Key Constructs

Key Construct	Definition	References
STEA Informativeness	The extent to which STEA programs provide employees with rich technical and procedural knowledge of best practices for IS usage.	Harrington, 1996
Task Urgency	The immediacy of an employee's daily job-related tasks and activities.	e.g., Bulgurcu et al. 2010
ISP Deterrence	The threatening degree of the ISPs messages of certain, swift, and/or severe punishment on ISP breach behaviour.	Akers, 1990; Williams and Hawkins, 1986
Security-related Cognitive Elaboration	The amount of cognitive effort spent on comprehending the possible outcomes associated with complying or not complying with corporate information security policies	Petty et al., 1995
Task-related Cognitive Elaboration	The amount of cognitive effort toward thinking about the outcomes of specific corporate task accomplishments.	Petty et al., 1995
ISP Compliance intention	An employee's intention to protect the information and technology resources of the organization by following corporate ISP requirements.	Bulgurcu et al. 2010

3 Research Hypotheses

Based on the literature on STEA programs, task urgency characteristics and ISP deterrence as well as the theoretical backbone of cognitive elaboration, in this study we propose a series of theory-based hypotheses related to ISP compliance intention. In general, as suggested by cognitive elaboration theory, people would expend different levels of cognitive effort when facing different situational factors reflecting different levels of informational intensity, personal relevance or personal accountability with respect to values, career, relatives, or other considerations (Malaviya 2007). Hence, the three most influential factors, which are STEA program informativeness, task urgency characteristics and ISP deterrence, will be viewed as being the situational factors leading to varied levels of cognitive elaboration efforts being spent on different aspects. These aspects comprise security-related and task-related outcomes. Generally, we propose that more security-related cognitive elaboration and task-related cognitive elaboration could enhance employees' ISP compliance tendencies, even when tasks completion conflicts with security compliance decisions.

First, as highlighted by cognitive elaboration theory and the dual considerations of both security-related and task-related outcomes, we decompose the construct of cognitive elaboration on potential outcomes into two dimensions: security-related and task-related, accordingly. Specifically, on the one hand, cognitive elaboration on security-related potential outcomes centres on the cognitive effort spent on comprehending the possible outcomes associated with complying or not complying with corporate information security policies. Such elaboration is directly linked to an individual employee's decision-making with respect to ISP compliance. According to cognitive elaboration theory, more elaboration effort being spent on the target decision could lead to a greater possibility of ethical conduct as a result of self-regulation (Nenkov et al. 2008). On the other hand, task-relevant cognitive elaboration directs more cognitive effort toward thinking about the outcomes of specific corporate task accomplishments. It emphasizes task completion as being the core problem at stake. We propose that more cognitive elaboration on task-related potential outcomes could help an individual evaluate the task in a rational way and form an objective judgment with respect to the severity of the consequences of not finishing the task. In other words, the more individuals think about on the potential outcomes of task processing, the more ways they will devise to deal with the negative consequences of not completing the task. Hence, it will be less likely that they would take the risk of being punished for breaking ISP rules in order to finish the task. Hence, we hypothesize that:

Hypothesis 1: *More cognitive elaboration on security-related potential outcomes has a positive effect on corporate employees' ISP compliance intentions.*

Hypothesis 2: *More cognitive elaboration on task-related potential outcomes has a positive effect on corporate employees' ISP compliance intentions.*

Second, STEA programs are designed to deliver security information to employees to enhance the level of corporate information security. However, the extent to which STEA messages are perceived as being informative by employees could result in opposite outcomes. Specifically, due to limited cognitive effort, as the STEA contents become increasingly informative, employees may be less motivated to consider security-related potential outcomes than when they were exposed to less information. Instead, they are more likely to link the security knowledge to other issues (e.g., how to finish the tasks more efficiently with the security knowledge). Such inferential use of security knowledge may not only distract employees' attention from the security compliance, but may also lead them to other unintended uses of that knowledge. For example, highly informative STEA programs may reveal to employees with possible ways of dealing with the security measures that help them complete specific work assignments. In such a situation, employees are encouraged to spend more cognitive elaboration effort on task-related potential outcomes but less effort on security-related potential outcomes. Therefore, it is hypothesized that:

Hypothesis 3: *The informativeness of STEA programs has no significant effect on corporate employees' cognitive elaboration with respect to security-related potential outcomes.*

Hypothesis 4: *The informativeness of STEA programs has a positive effect on corporate employees' cognitive elaboration on task-related potential outcomes.*

Third, employees who are overwhelmed by the need to quickly finish a task are more concerned with task-related potential outcomes. When task urgency becomes primed to be a critical situational factor, the decision-maker places more weight on the importance of the task and the task-related cognitive elaboration will be enhanced. Meanwhile, the cognitive processing of other outcomes, i.e., the security-related outcomes, could be diminished due to their reduced priority. In other words, when employees are carrying out daily routines that are not expected to be completed immediately, it will be more likely for them to think carefully about the task and more carefully consider the detailed requirements of the IS security policy. However, if the task at hand is urgent or is required to generate quick outputs, the IS security policy will become a hurdle in accomplishing the task in terms of time and effort. Hence, when a task is urgent, an employee is more likely to focus more on the cognitive effort required to complete the task rather than on the counterproductive security measures. Hence, we hypothesize that:

Hypothesis 5: *The task urgency has a negative effect on corporate employees' cognitive elaboration with respect to security-related potential outcomes.*

Hypothesis 6: *The task urgency has a positive effect on corporate employees' cognitive elaboration with respect to task-related potential outcomes.*

Fourth, the deterrence information contained in corporate ISPs could make future punishment or sanctions salient to corporate employees. The punishment strength and penalty certainty can increase the level of the receivers' attention and thus motivate them to consider the consequences of security breach outcomes. Under such circumstances, more cognitive elaboration will be expended on the deterrence messages in order to gauge their meaning and scope. Meanwhile, other closely related matters will also seem more important and more effort will also be spent on thinking about them, such as the tasks to be completed. For example, when the deterrence level increases, individuals can easily relate it to the difficulty of finishing the tasks at hand because increasing the level of ISP deterrence means more restrictions on the work. Therefore, we propose that the deterrence information in corporate ISPs motivates employees to focus more on the potential security-related outcomes as well as the potential task-related outcomes.

Hypothesis 7: *The deterrence of ISP has a positive effect on corporate employees' cognitive elaboration on security-related potential outcomes.*

Hypothesis 8: *The deterrence of ISP has a positive effect on corporate employees' cognitive elaboration on task-related potential outcomes.*

4 Research Methodology and Data Analysis

To test the hypotheses, we conducted an experimental survey in a public university. We manipulated the STEA messages, task characteristics and ISP deterrence message, and investigated how these factors may affect ISP compliance intentions.

In the experiment, participants were asked to process relevant information and to answer survey questions. At the beginning of the experiment, participants were provided with STEA material introducing details of password cracking issues in organizations, followed by ISP content specifically directed toward the password usage or cracking issue. The STEA material was designed to reflect the different levels of informativeness. Less informative STEA material only taught participants the basic concepts of password cracking but highly informative STEA material included additional information about how to crack system passwords and the low probability of it being detected (i.e., low vs. high STEA informativeness) (e.g. Kuo and Hsu 2001). The fictional task was retrieving certain required information from a corporate system when the person holding the system password was not available. This situation may induce participants to crack the password for the sake of task accomplishment. The urgency of the task was manipulated by altering the allowed time to finish it (i.e., two days vs. a few hours for task processing as low vs. high task urgency). The task requirement was thus in conflict with the corporate policy of password protection as indicated in the ISP material and the pros and cons of either following the policy or breaking the rules to finish the task were specified. The likelihood of ISP deterrence was manipulated by varying the severity of punishment and the certainty of security breach detection from a low to a high level (i.e., soft tone vs. harsh tone about ISP breach punishment as low vs. high ISP deterrence). Following the instruction materials, participants were required to indicate their ISP compliance decision and answer related survey questions.

The measurements for the constructs of cognitive elaboration and ISP compliance intention were adapted from previous research. Specifically, cognitive elaboration measurements were adapted from items examining the thoughts directed toward the potential outcomes of the issue (see Table 2). Our research design was reviewed internally by the ethics committee at the public university to ensure its appropriateness.

Table 2. Operationalization of Constructs

Constructs	Indicators	Sources
Security-related Cognitive Elaboration (SECURITY ELABORATION)	1. I tried to anticipate as many consequences of my actions related to security policy breach as I could. 2. Before I made a decision, I considered all possible outcomes of a security policy breach. 3. I tried to assess how important the potential consequences of my decision to incur a security policy breach might be. 4. I tried to predict how likely the different consequences of a security policy breach were. 5. Usually, I carefully estimate the risk of the various outcomes of a security policy breach occurring.	Nenkov et al. (2008)
Task-related Cognitive Elaboration (TASK ELABORATION)	1. I tried to anticipate as many consequences of my actions on task execution as I could. 2. Before I made a decision, I considered all the possible outcomes of task execution. 3. I tried to assess how important the potential consequences of my decisions for task execution might be. 4. I tried to predict how likely the different consequences of task execution were. 5. Usually, I carefully estimate the probability of various outcomes of task execution occurring.	Nenkov et al. (2008)
ISP Compliance Intention (COMPLIANCE)	1. I intend to comply with the requirements of the information security policy of my organization in the future. 2. I intend to protect information and technology resources according to the requirements of the information security policy of my organization in the future. 3. I intend to carry out my responsibilities as prescribed in the information security policy of my organization when I use information and technology in the future.	Bulgurcu, et al. (2010)

From a total of 160 participants who completed the survey and each participant was randomly assigned to one of the eight treatments (either high or low in STEA informativeness, task urgency and ISP deterrence). After removing the incomplete responses, we kept 147 data sets for further analysis while the number of data sets for each treatment ranged from 15 to 20. The demographic variables of the sample and descriptive measures of the constructs are illustrated in Table 3 and Table 4.

The convergent and discriminant validity of the measurement model was verified based on standard criteria (see Table 5 and Table 6).

Table 3. Demographics (n = 147)

Demographic Variables	Category	Frequency (%)	Demographic Variables	Category	Frequency (%)
Gender	Male Female	76 (51.7%) 71 (48.3%)	***Age***	18 – 20 21 – 24 25 – 29 30 – 34 35 – 39	1 (0.7%) 115 (78.2%) 29 (19.7%) 1 (0.7%) 1 (0.7%)
Highest Level of Education/ Highest Degree	High School or Below Pre-U Bachelor Master Ph. D	3 (2%) 2 (1.4%) 68 (46.3%) 69 (46.9%) 5 (3.4%)	***Years of Computer Usage***	1 – 4 5 – 8 9 – 12 13 – 16 >16	6 (16.3%) 74 (50.3%) 44 (29.9%) 4 (2.8%) 1 (0.7%)
Computer Expertise	Very Poor Poor Modest Expert Absolutely Expert	2 (1.4%) 4 (2.8%) 91 (61.8%) 40 (27.2%) 10 (6.8%)	***Computer Security Expertise***	Very Poor Poor Modest Expert Absolutely Expert	7 (4.7%) 22 (14.9%) 98 (66.7%) 16 (10.9%) 4 (2.8%)

Table 4. Descriptive Statistics

Construct	Mean	S.D.
Security-related Cognitive Elaboration (SECURITY ELABORATION)	5.3	0.934
Task-related Cognitive Elaboration (TASK ELABORATION)	5.53	1.067
ISP Compliance Intention (COMPLIANCE)	5.62	0.926

We further applied SmartPLS (version 2.0 M3) to analyze the structural model. Results indicated that five out of eight hypotheses were supported (Hypothesis 2, Hypothesis 3, Hypothesis 4, Hypothesis 6 and Hypothesis 8) (see Figure 1). We also tested the effects of control variables and the results showed no significant influence of factors including gender, age, highest education, computer usage experience, Internet usage experience, computer expertise and information security expertise on the ISP compliance intention.

Table 5. Results of Convergent Validity and Discriminant Tests

Constructs and Indicators		Reliability of Indicators	Composite Reliability	Cronbach's Alpha	Average Variance Extracted	Factors		
						1	2	3
SECURITY ELABORATION	Item1	0.691	0.915	0.862	0.699	.690	.243	.023
	Item2	0.771				.786	.297	-.049
	Item3	0.875				.822	.162	.262
	Item4	0.812				.751	.139	.158
	Item5	0.859				.753	.317	.171
TASK ELABORATION	Item1	0.823	0.893	0.848	0.649	.371	.612	.323
	Item2	0.696				.237	.843	-.123
	Item3	0.815				.225	.692	.280
	Item4	0.786				.214	.592	.425
	Item5	0.825				.275	.785	.178
COMPLIANCE	Item1	0.915	0.954	0.895	0.822	.070	.221	.842
	Item2	0.879				.132	.027	.891
	Item3	0.933				.130	.203	.858

Table 6. Inter-Construct Correlations (The diagonal values are the square roots of AVE)

	SECURITY ELABORATION	TASK ELABORATION	COMPLIANCE
SECURITY ELABORATION	**0.836**		
TASK ELABORATION	0.288**	**0.806**	
COMPLIANCE	0.408**	0.612**	**0.906**

** Significant at 0.01 level (two-tailed).

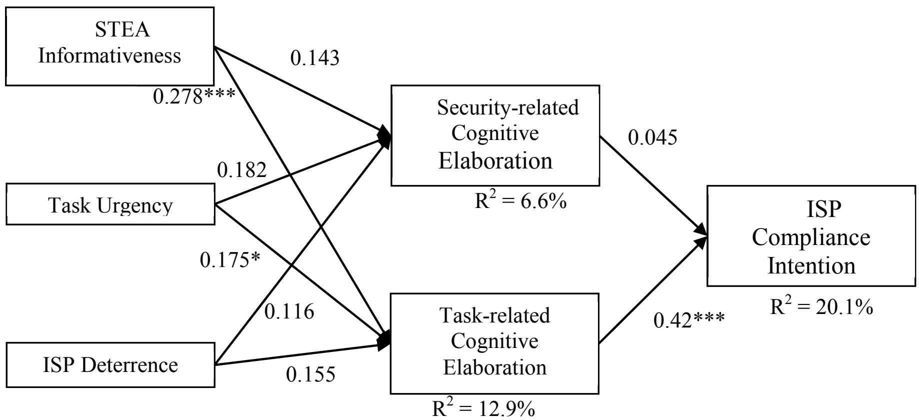


Fig. 1. PLS Analysis Results

* or *** significant at 0.05 or 0.001 level (two-tailed), respectively.

5 Discussion, Implications and Conclusion

This study represents one of the earliest attempts by researchers to empirically investigate how STEA programs and ISP policies can be designed and how task characteristics may affect employees' ISP compliance decisions.

The data analysis of the study showed various unexpected results. First, security-related cognitive elaboration had no significant relationship to ISP compliance intention (Hypothesis 1 not supported). This result was in some ways contrary to intuition, but may imply that an individual's ISP compliance was not simply determined by the security-related issues, but that it was a more complex decision. In contrast, task-related cognitive elaboration had a positive relationship to ISP compliance intention (Hypothesis 2 supported). This implies that when participants expended considerable cognitive effort on the task at hand, it might make them feel that the task was less critical than indicated, especially when they considered the possible outcomes of the task processing in the broader or long-term sense. It is also possible that participants took security-related knowledge or policy into account when elaborating on the task outcomes. The importance of security compliance outweighed that of task completion, thus motivating them to comply with the ISP. In turn, STEA informativeness was seen as having no significant relationship to security-related cognitive elaboration (Hypothesis 3 supported). Task urgency was found to be positively related to security-related cognitive elaboration (Hypothesis 5 not supported) whereas ISP deterrence had no significant effect on security-related cognitive elaboration (Hypothesis 7 not supported). A possible explanation for this result is that ISP deterrence issues are considered to be distant from the individual's immediate concerns such as the urgency of the task, thereby making them less focused on the security issues. However, as the tasks were not entirely separate from the security issue, placing more emphasis on task urgency unavoidably caused participants to think more about the security issue that was closely associated with to the task. Task-related cognitive elaboration was positively related to STEA informativeness (Hypothesis 4 supported) and task urgency (Hypothesis 6 supported). In addition, task-related cognitive elaboration was also positively related to ISP deterrence (Hypothesis 8 supported).

Overall, the unsupported results showed the interplay between STEA, ISP and task information when individuals had to make a decision that was closely related to each of them. Although the cognitive elaboration was directed toward each aspect, that seemingly irrelevant information still managed to increase the level of an individual's attention with respect to other aspects. Furthermore, the results indicated that only when the individuals' cognitive elaboration on task-related issues was enhanced, was their ISP compliance tendency increased, because the task-related cognitive elaboration was more closely related to their daily decision-making. When employees evaluate the pros and cons of any possible method of carrying out a task (including how to breach security rules), they are more likely to follow the existing ISP because failing to finish the task could be less harmful than breaking the security rules.

The results of this study will benefit both academics and practitioners in several aspects. The benefits to academic research in IS will include the conceptualization of STEA informativeness function and the incorporation of task characteristics and ISP deterrence as predictors of compliance intention. More importantly, we propose the critical role of cognitive elaboration and decompose the construct into different dimensions to reflect the unique research background. This approach enabled us to gain a better understanding of the complex effects of various external measures on an individual's thoughts and decision-making. It also reflected the contradictory thoughts that individuals may generate when facing informational messages that are in conflict. Eventually, based on the original theoretical insights into cognitive elaboration, we extended the theory's scope by proposing possible determinants and effects of cognitive elaboration on potential outcomes, especially when they are in conflict with each other. The methodological contribution was demonstrated by the use of experimental techniques in simulating the effects of measures that may result in a threat to IS security arising from trusted users.

The findings from this study will also have significant management implications with respect to the design and implementation of STEA programs and ISP deterrence messages within an organization based on specific task characteristics. Specifically, security practitioners need to discern the potential conflicting effects of informativeness in STEA programs by taking ISP deterrence and task urgency into account. According to the survey results, practitioners should try their best to elicit as much cognitive elaboration on either security- or task-related issues from as many corporate employees as possible, thus further enhancing their knowledge of ISP compliance intention. To achieve this objective, the appropriate establishment setting of external measures should be emphasized.

To summarize, this study recognizes the impact of different levels of informativeness in STEA programs on employees' compliance behaviors and considers the importance of task urgency and ISP deterrence level. We propose a theoretical model based on the cognitive elaboration literature and conducted an experimental survey to examine this phenomenon and test the hypotheses. The results of this study will be suitable to guide researchers and practitioners to better enforce IS security policy. In future endeavours, we will overcome the limitations of the current study by designing a more realistic experiment and soliciting responses from corporate employees as well as public university employees.

Acknowledgements. The authors gratefully acknowledge a research grant from the City University of Hong Kong (SRG Project No. 7002527) in support for this study.

References

1. Akers, R.: Rational choice, deterrence, and social learning theory in criminology: the path not taken. The Journal of Criminal Law and Criminology 81(3), 653–676 (1990)
2. Boss, S.R., Kirsch, L.J.: The last line of defense: motivating employees to follow corporate security guidelines. In: Proceedings of the 28th International Conference on Information Systems, Montreal, December 9-12 (2007)
3. Bulgurcu, B., Cavusoglu, H., Benbasat, I.: Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness. MIS Quarterly 34(3), 523–548 (2010)
4. Cavusoglu, H., Mishra, B., Raghunathan, S.: The effect of internet security breach announcements on market value: capital market reactions for breached firms and internet security developers. International Journal of Electronic Commerce 9(1), 69–104 (2004)
5. D'Arcy, J., Hovav, A., Galletta, D.: User awareness of security counter-measures and its impact on information systems misuse: A deterrence approach. Information Systems Research 20(1), 79–98 (2009)
6. Ernst & Young, Moving beyond compliance: Ernst & Young's, global information security survey (2008)
7. Fazio, R.H.: Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In: Petty, R.E., Krosnick, J.A. (eds.) Attitude Strength: Antecedents and Consequences, pp. 247–282. Erlbaum, Mahwah (1995)
8. Gordon, L.A., Loeb, M.P., Lucyshyn, W., Richardson, R.: CSI/FBI computer crime and security survey. Computer Security Institute (2006)
9. Harrington, S.: The effect of codes of ethics and personal denial of responsibility on computer abuse judgments and intentions. MIS Quarterly 20(3), 257–277 (1996)
10. Herath, T., Rao, H.R.: Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. Decision Support Systems 47, 154–165 (2009)
11. Hui, W., Hu, P.: Examining end-user information security policy compliance: An exploratory study. In: Proceedings of the Workshop on e-Business (WeB), Paris, France, December 13 (2008)
12. Kuo, F., Hsu, M.: Development and validation of ethical computer self-efficacy measure: The case of softlifting. Journal of Business Ethics 32, 299–315 (2001)
13. Liska, A.E., Messner, S.F.: Perspectives on Crime and Deviance, 3rd edn. Prentice Hall, Upper Saddle River (1999)
14. Malaviya, P.: The moderating influence of advertising context on ad repetition effects: The role of amount and type of elaboration. Journal of Consumer Research 34(1), 32–40 (2007)
15. Myyry, L., Siponen, M., Pahnila, S., Vartiainen, T., Vance, A.: What levels of moral reasoning and values explain adherence to information security rules? An empirical study. European Journal of Information Systems 18(2), 126–139 (2009)

16. Nenkov, G.Y., Inman, J.J., Hulland, J.: Considering the Future: The Conceptualization and Measurement of Elaboration on Potential Outcomes. Journal of Consumer Research 35, 126–141 (2008)
17. Pahnila, S., Siponen, M., Mahmood, A.: Employees' behavior towards is security policy compliance. In: Proceedings of the 40th Hawaii International Conference on System Sciences, pp. 156–166. IEEE Computer Society Press, Los Alamitos (2007)
18. Petty, R.E.: Attitude change. In: Tesser, A. (ed.) Advances in Social Psychology, pp. 194–255. McGraw–Hill, New York (1995)
19. PricewaterhouseCoopers. Employee behavior key to improving information security, new survey finds, June 23 (2008)
20. Sagar, R.: Who holds the balance? A missing detail in the debate over balancing security and liberty. Polity 41(2), 166–188 (2009)
21. Shaw, E., Ruby, K., Post, J.: The insider threat to information systems: The psychology of the dangerous insider. Security Awareness Bulletin 2-98, 1–10 (1998)
22. Siponen, M.T., Vance, A.: Neutralization: new insight into the problem of employee information systems security policy violations. MIS Quarterly 34(3), 487–502 (2010)
23. Siponen, M.T., Pahnila, S., Mahmood, A.: Employees' adherence to information security policies: An empirical study. In: Venter, H., Eloff, M., Labuschagne, L., Eloff, J., von Solms, R. (eds.) New Approaches for Security, Privacy and Trust in Complex Environments, pp. 133–144. Springer, Boston (2007)
24. Smith, S.M., Fabrigar, L.R., Macdougall, B.L., Wiesenthal, N.L.: The role of amount, cognitive elaboration, and structural consistency of attitude-relevant knowledge in the formation of attitude certainty. European Journal of Social Psychology 38(2), 280–295 (2008)
25. Stoneburner, G., Goguen, A., Feringa, A.: Risk management guide for information technology systems. NIST Special Publications 800-30, White Paper, United States Department of Commerce, Gaithersburg, MD (2002)
26. Straub, D.W.: Effective is security: an empirical study. Information Systems Research 1(3), 255–276 (1990)
27. Street, M.D., Douglas, S.C., Geiger, S.W., Martinko, M.J.: The impact of cognitive expenditure on the ethical decision-making process: The cognitive elaboration model. Organizational Behavior and Human Decision Processes 86(2), 256–277 (2001)
28. Tyler, T.R., Blader, S.L.: Can Businesses Effectively Regulate Employee Conduct? The Antecedents of Rule Following in Work Settings. Academy of Management Journal 48(6), 1143–1158 (2005)
29. Warkentin, M., Willison, R.: Behavioral and policy issues in information systems security: The insider threat. European Journal of Information Systems 18(2), 101–105 (2009)
30. Warkentin, M., Davis, K., Bekkering, E.: Introducing the check-off password system (cops): An advancement in user authentication methods and information security. Journal of Organizational and End User Computing 16(3), 41–58 (2004)
31. Williams, K., Hawkins, R.: Perceptual research on general deterrence: a critical review. Law and Society Review 20(4), 545–572 (1986)
32. Wyer, R.S.: Language and advertising effectiveness: Mediating influences of comprehension and cognitive elaboration. Psychology & Marketing 19(7-8), 693–712 (2002)
33. Yue, W., Çakanyildirim, M.: Intrusion prevention in information systems: Reactive and proactive response. Journal of Management Information Systems 24(1), 329–353 (2007)
34. Yue, W., Çakanyildirim, M., Ryu, Y., Liu, D.: Network externalities, layered protection and it security risk management. Decision Support Systems 44(1), 1–16 (2007)

Analyzing Monetization Models for the Digital Content Services: Channel Ownership and Royalty Contracts

Yung-Ming Li[1], Yuan Fang[1], and Bih-Huang Jin[2,*]

[1] Institute of Information Management, National Chiao Tung University, Hsinchu, 300, Taiwan
yml@mail.nctu.edu.tw, m1118@ms18.hinet.net
[2] Department of Business Administration, Tunghai University, Taichung 407, Taiwan
bihuang@thu.edu.tw

Abstract. The formats and access models of digital content are increasingly rich and diverse as the advancement of internet and mobile technologies. In this paper, utilizing a game theoretic model, we analyze the adoption and monetization models of heterogeneous content channels (mobile and website content). The pricing and advertising strategies of pay and advertisement-supported content under different channel and content ownership structures are examined. Furthermore, under asymmetric content ownership (only one of two channels own the content), popular monetary transfer contracts are realized to license the opponent channel. The impacts of market parameters (such as quality differentiation and online advertisement factors) on the development of business strategies are presented.

Keywords: Digital content, Mobile service, Monetization and contract models, Channel competition, Content ownership, Revenue sharing.

1 Introduction

With the advance of Internet and mobile technologies, the formats and access models of digital content are increasingly rich and diverse. Various types of digital content are integrated and incorporated to become new products and services. With the addition of the mobile and television/radio broadcasting dimension, various forms of digital services are not just transforming existing content to mobile or TV/Radio platforms. Digital content can integrate the characteristics of "anywhere and personal" in consumption of content in social interaction or new interactions with user's environment [1]. The same content source can have different types of services and business models after going mobile or specialized targeting environment.

Convergence makes now possible for a content provider, initially specialized in a specific platform, to deliver digital services and contents through several different access platforms. The integration among different platforms (internet, mobile, and television/radio broadcasting) allows content providers maximize content value through different access channels. In spite of generalized consensus about the huge

* Corresponding author.

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 195–201, 2012.

potential of FMC, there are only limited cases to develop a multi-access approach to digital content distribution. This phenomenon is caused by channel cannibalization issue, privacy issue, or lack of competence in the new digital convergence era. In a competitive market, content differentiation is essential. A move to go mobile and multi-channel content provider will be the necessary. Thus, our research questions are: (1) what is the effectiveness of market segmentation to the content provider? (2) What happens to the channel adoption decisions to a monopolistic firm? (3) What is the best revenue sourcing contract when some of players may not have the content source? In this research, we analyze the monetization and contract models of heterogeneous digital content services (mobile and Web content) under various channel and content ownership settings and examine the impact of the service quality and online advertisement factors on channel adoption decision, market share, profit level as well as resulting contract design

2 Related Literature

2.1 Versioning/Product Differentiation/ Channel Cannibalization

Digital content service provider design versioning and product differentiation to maximize their profit [2]. Users acquire the digital content they need via various channels. In this circumstance, cannibalization effect may be generated. Simon & Kadiyalib [3] examined the cannibalization effect of digital and print magazine and found that print circulation declined. Feng, Guo, & Chiang [4] investigated the optimal digital content channel strategy . Gallaugher, Auger, & BarNir [5] provide an empirical exploration of different revenue streams and related performance of digital content providers. Diverse ownerships of digital content would design optimal pricing contract to generate the maximized profit and avoid or reduce cannibalization effect.

2.2 Contracts

Different types of contract under different channel may generate different beneficial for digital content providers. Digital content providers choose an optimal contract to sell their products. Luo & Çakanyilirim [6] compared revenue-sharing contract and wholesale price contract. They found revenue-sharing contracts can be Pareto improving for the supplier and retailer. Veen & Venugopal [7] showed that revenue-sharing contracts can optimize both supplier and retail a win-win situation. Yue & Liu [8] compared the performance with and without direct channel and found that manufacturer and the supply chain may better off when some condition was satisfied. Yao, Leung, & Lai [9] found provision of revenue-sharing contract improve the performance of supply chain. Linh & Hong [10] found that wholesale price are set lower than retail price and optimal revenue-sharing ratio is linearly increasing in wholesale price for the retailer. Pan, Lai, Leung, & Xiao [11] analyzed revenue-sharing contract and wholesale contract under manufacturer-dominated and retailed dominated scenarios. They found that revenue-sharing contract is beneficial for either one or both manufacturers under the manufacturer-dominated scenario.

3 The Model

We consider a digital content market in which two differentiated access types (mobile and website channels) of content service are provided. The access service quality of mobile content and Web content is denoted as q_M and q_I respectively, where the subscript M stands for the mobile channel and *I* represents the Web (Internet) channel. As a mobile channel provides more convenient access mode to access content instantly than a website channel, we assume $q_M > q_I$. Denote η_0 as the potential market size; η_M and η_I are the total number of the mobile-channel and the Web-channel customers respectively. Customers have different values on the content access channel chosen. The value of accessed content attached to a typical customer i is defined as $\theta_i q_k$ for $k \in \{M, I\}$, where the variable θ_i stands for the individual valuation on the chosen content access channel, and is uniformly distributed with an interval $[0,1]$. A higher value of θ_i indicates that a higher value on the accessed content. While a few revenue models have been proposed to monetize digital content service, subscription and advertisement are two of the most popular and major revenue streams practiced in the digital market. For a target channel $k \in \{M, I\}$, we denote the access fee as p_k and the advertising level (amount) delivered a_k, The parameters used in the model are listed in Table 1.

Table 1. Model parameters

Parameters	Description
η_0	Total number of potential users (potential market size)
$\eta_M ; \eta_I$	Demand of the pay (mobile) channel; Demand of the ad-supported (Web)channel
$q_M ; q_I$	Service quality of the ad-supported (Web)channel; Service quality of the ad-supported (Web) channel
θ_i	Individual valuation (preference) on the chosen content service ($\theta_i \sim U[0,1]$)
p_M	Access fee of the pay (mobile)channel
a_I	Amount of advertisements exposed to users
λ	Price of advertisement
φ	Revenue sharing ratio
f	Fixed loyalty fee
δ	Disutility coefficient to advertisement
$\pi_M ; \pi_I$	Profit of the pay (mobile) channel; Profit of the ad-supported (Web)

Customer utility function. The utility function of a typical customer i is formulated as

$$U_i = \theta_i q_k - p_k \ - \delta a_k \ , k \in \{I, M\}, \tag{1}$$

where $\delta > 0$ stands for the disutility incurred when an advertisement is exposed to a customer. In the research, we focus on the scenario that the mobile content access is a pay service but the Web content access is advertisement-supported and free to the users (p_I =0 and $a_M = 0$).

Channel demand function. According to the content valuation function θ_i, the mobile content service is more preferable to the users who have a higher valuation on the content. Let $\hat{\theta}_1$ denote a customer type who is indifferent between choosing from a Web channel and not accessing any content. Similarly, let $\hat{\theta}_2$ denote a customer type who is indifferent between choosing from the mobile content and the Web content. The utility function implies that:

$$\hat{\theta}_1 = \frac{\delta a_I}{q_I} \text{ and } \hat{\theta}_2 = \frac{p_M - \delta a_I}{q_M - q_I}. \tag{2}$$

Therefore, all customer types indexed by $\theta_i \in [\hat{\theta}_1, \hat{\theta}_2]$ choose the Web content service and all customers indexed by $\theta_i \in [\hat{\theta}_2, 1]$ choose mobile content service. The demand functions are written as:

$$\eta_M = \left(1 - \hat{\theta}_2\right)\eta_0 = \left(1 - \frac{p_M - \delta a_I}{q_M - q_I}\right)\eta_0, \ \ \eta_I = \left(\hat{\theta}_2 - \hat{\theta}_1\right)\eta_0 = \left(\frac{p_M - \delta a_I}{q_M - q_I} - \frac{\delta a_I}{q_I}\right)\eta_0 \tag{3}$$

4 Monopolistic Ownership of Heterogeneous Channels

We first examine the scenario that a monopolistic firm owns content source and has the capability (in both technology and marketing aspects) to offer both mobile and Web content services. Denote λ as the price of online advertisement per measure unit.

The profit-maximization problem of the firm is formulated as:

$$\max_{p_M, a_I} \pi_{I+M} = \eta_M p_M + \eta_I \lambda a_I \ \text{ s.t. } 0 \le \eta_M, \eta_I \le \eta_0 \tag{4}$$

Solving the profit maximization problem, we obtain the price of mobile content and advertising amount of Web content as:

$$\left(p_M^*, a_I^*\right) = \begin{cases} \left(\frac{q_M}{2}, 0\right) & \text{if } \lambda \le \delta \\ \left(\frac{2\delta\lambda q_M (q_M - q_I)}{4\delta\lambda q_M - (\varphi + \delta)^2 q_I}, \frac{(\lambda + \delta) q_I (q_M - q_I)}{4\delta\lambda q_M - (\lambda + \delta)^2 q_I}\right) & \text{if } \delta < \lambda \le (2q_M / q_I - 1)\delta \\ \left(0, \frac{q_I}{2\delta}\right) & \text{if } \lambda > (2q_M / q_I - 1)\delta \end{cases}. \tag{5}$$

The demands of the two types of content service can be derived as:

$$\left(\eta_M^*,\eta_I^*\right)=\begin{cases}\left(\dfrac{\eta_0}{2},0\right) & if\,\lambda\le\delta\\ \left(\dfrac{\left(2\delta q_M-(\lambda+\delta)q_I\right)\lambda\eta_0}{4\delta\lambda q_M-(\lambda+\delta)^2 q_I},\dfrac{\delta(\lambda-\delta)q_M}{4\delta\lambda q_M-(\lambda+\delta)^2 q_I}\right) & if\;\;\delta<\lambda\le(2q_M/q_I-1)\delta\\ \left(0,\dfrac{\eta_0}{2}\right) & if\,\lambda>(2q_M/q_I-1)\delta\end{cases}\tag{6}$$

The resulting profit of the provider can be obtained as:

$$\pi_{I+M}^{\;\;*}=\begin{cases}\dfrac{q_M\eta_0}{4} & if\,\lambda\le\delta\\ \dfrac{\delta\lambda q_M\left(q_M-q_I\right)\eta_0}{4\delta\lambda q_M-(\lambda+\delta)^2 q_I} & if\;\;\delta<\lambda\le(2q_M/q_I-1)\delta\\ \dfrac{\lambda q_I\eta_0}{4\delta} & if\,\lambda>(2q_M/q_I-1)\delta\end{cases}\tag{7}$$

Therefore, we have the following the observation:

Proposition 1. A monopolistic content provider will (1) Only provide pay mobile content when $\lambda\le\delta$, (2) Provide pay mobile content as well as advertisement-supported Web content when $\delta<\lambda\le\left(2q_M/q_I-1\right)\delta$, (3) Only provide advertisement supported web content when $\lambda>\left(2q_M/q_I-1\right)\delta$.

Proposition 1 quantities the market environments suggesting whether pay mobile content or free Web content or both should be provided. Obviously, to successfully collect more revenue from dual heterogeneous content channels, besides the quality differentiation should be sufficiently large, the market advertisement price is not too small or high. Otherwise, Channel cannibalization effect significantly diminishes the effectiveness of market segmentation.

5 Competing Heterogeneous Channels

5.1 Both Channels Has Self-own Content Ownership

We next analyze the market scenario in which the mobile and Web channels are operated by independent firms. In addition to channel service, they own content source. The profit-maximizing problems the competing firms face can be reformulated as:

$$\max_{p_M}\pi_M=\eta_M p_M \quad\text{and}\quad \max_{a_I}\pi_I=\eta_I\lambda a_I\,.\tag{8}$$

Solving $\dfrac{\partial\pi_M}{\partial p_M}=0$ and $\dfrac{\partial\pi_I}{\partial a_I}=0$ simultaneously, we have the price of mobile content and advertising level of Web content as:

$$p_M^{\;*}=\frac{2q_M\left(q_M-q_I\right)}{4q_M-q_I}\;;\quad a_I^{\;*}=\frac{q_I\left(q_M-q_I\right)}{\delta\left(4q_M-q_I\right)}\,.\tag{9}$$

The demands for the two channels are:

$$\eta_M^* = \frac{2q_M\eta_0}{4q_M - q_I};\ \eta_I^* = \frac{q_M\eta_0}{4q_M - q_I}. \tag{10}$$

The profits for the two channels are:

$$\pi_M^* = \frac{4q_M^2(q_M - q_I)\eta_0}{(4q_M - q_I)^2};\ \pi_I^* = \frac{\lambda q_M q_I (q_M - q_I)\eta_0}{\delta(4q_M - q_I)^2}. \tag{11}$$

5.2 The Pay Channel Has Content Ownership

We further consider the scenario that the mobile channel provider owns the content source. Instead of launching the Web content service, the mobile content provider licenses a business partner providing the website content service and receives content loyalty fee. In the following, we develop the monetary transfer contract of revenue sharing.

Assume the Web channel provider needs to transfer $0 < \varphi_M < 1$ proportion of its collected revenue to the mobile channel provider. The profit-maximizing problems for the two competing firms can be reformulated as:

$$\max_{p_M} \pi_M = \eta_M p_M + \eta_I \varphi_M \lambda a_I \text{ and } \max_{a_I} \pi_I = (1-\varphi_M)\eta_I \lambda a_I. \tag{12}$$

Solving $\partial\pi_M / \partial p_M = 0$ and $\partial\pi_I / \partial a_I = 0$ or $p_M^* = \frac{q_M - q_I + (\delta - \varphi_M\lambda)a_I^*}{2}$ and $a_I^* = \frac{q_I p_M^*}{2\delta q_M}$ simultaneously, we have the price of mobile content and advertisement amount of Web content as:

$$p_M^* = \frac{2\delta q_M (q_M - q_I)}{4\delta q_M - (\delta - \varphi_M\lambda)q_I},\ a_I^* = \frac{q_I(q_M - q_I)}{4\delta q_M - (\delta - \varphi_M\lambda)q_I}. \tag{13}$$

The demands of the two channels are:

$$\eta_M^* = \frac{(2\delta q_M + \varphi_M\lambda q_I)\eta_0}{4\delta q_M - (\delta - \varphi_M\lambda)q_I}, \eta_I^* = \frac{\delta q_M \eta_0}{4\delta q_M - (\delta - \varphi_M\lambda)q_I}. \tag{14}$$

The profits for the two channels are:

$$\pi_M^* = \frac{(4\delta^2 q_M^2 + 3\lambda\delta\varphi_M q_M q_I)(q_M - q_I)\eta_0}{(4\delta q_M - (\delta - \varphi_M\lambda)q_I)^2}, \pi_I^* = \frac{(1-\varphi_M)\delta\lambda q_M q_I (q_M - q_I)\eta_0}{(4\delta q_M - (\delta - \varphi_M\lambda)q_I)^2}. \tag{15}$$

Finally, the mobile content channel decide the optimal revenue sharing ratio by solving $\partial\pi_M^* / \partial\varphi_M = 0$. We have $\varphi_M^* = \frac{\delta(4q_M - 3q_I)}{3\lambda q_I}$. (16)

6 Conclusion

In this research, we analyze the monetization and contract models of heterogeneous digital content services (mobile and Web content) under various channel and content ownership settings. When the entire market is dominated by a monopolistic firm, we present the conditions prescribing whether pay mobile content and advertisement-supported content services should be offered. When the delivery channels are operated independently but the content source is owned by one of the competing channels, we further develop the monetary transfer contracts such as revenue sharing ratio and fixed loyalty fee. The impact of the service quality and online advertisement factors on channel adoption decision, market share, profit level as well as resulting contract design are examined and discussed. Besides the theoretic aspects, our results can further provide useful practical business strategy for the digital content industry.

This study can be further extended in several directions. First, in this paper, the content service quality is mainly measured at some focused dimensions and a macro and aggregate level (e.g. the perspective of instant and convenient access). It will be interesting to incorporate more factors in other dimensions (such as content richness and user interface). These factors may be conflictable. Second, besides the subscription and advertising models, other advanced pricing schemes (nonlinear usage fee) can be further examined. Finally, while we consider two popular loyalty contract formats (revenue sharing and fixed fee) in the model, a potential avenue for future extension is to develop other appropriate contract types.

References

1. FeijUo, C., Maghiros, I., Abadie, F., GUmez-Barroso, J.E.L.: Exploring a heterogeneous and fragmented digital ecosystem: Mobile content. Telematics Informatics 26, 282–292 (2009)
2. Shapiro, C., Varian, H.R.: Information rules: a strategic guide to the network economy. Harvard Business Press, Boston (1999)
3. Simon, D., Kadiyali, V.: The effect of a magazine's free digital content on its print circulation: Cannibalization or complementarity? Inform Econ. Policy 19, 344–361 (2007)
4. Feng, Y., Guo, Z., Chiang, W.K.: Optimal Digital Content Distribution Strategy in the Presence of the Consumer-to-Consumer Channel. J. Manage. Inform. Syst. 25, 241–270 (2009)
5. Gallaugher, J.M., Auger, P., BarNir, A.: Revenue streams and digital content providers: an empirical investigation. Inform. Manage. 38, 473–485 (2001)
6. Luo, S., Cakanyilirim, M.: Pricing and production games under revenue sharing and information updates. J. Revenue Pricing Manage. 4, 270–301 (2005)
7. van der Veen, J.A.A., Venugopal, V.: Using revenue sharing to create win–win in the video rental supply chain. J. Oper. Res. Society 56, 757–762 (2005)
8. Yue, X., Liu, J.: Demand forecast sharing in a dual-channel supply chain. Eur. J. Oper. Res. 174, 646–667 (2006)
9. Yao, Z., Leung, S.C.H., Lai, K.: The effectiveness of revenue-sharing contract to coordinate the price-setting newsvendor products' supply chain. Int. J. Supply Chain Manage. 13, 263–271 (2008)
10. Linh, C.T., Hong, Y.: Channel coordination through a revenue sharing contract in a two-period newsboy problem. Eur. J. Oper. Res. 198, 822–829 (2009)
11. Pan, K., Lai, K., Leung, S.C.H., Xiao, D.: Revenue-sharing versus wholesale price mechanisms under different channel power structures. Eur. J. Oper. Res. 203, 532–538 (2010)

Pricing Centralized and Decentralized Wireless Service: A Mechanism Design Approach

Jhih-Hua Jhang-Li

No. 101, Sec.1, Fenliao Rd., LinKou District, New Taipei City 24452
Hsing Wu Institute of Technology, Taiwan
jhangli@gmail.com

Abstract. This study concentrates on how to price wireless access service and compare the two different operation models: centralized and decentralized service architecture. With the classical model of the mechanism design, some interesting results are discovered. In the mechanism design, a telecommunication service provider offers two service plans, and assumes that consumers select their service plans according to individual type. By involving queuing delay and service availability in our study, we find that the results in centralized service architecture are consistent with those in prior studies in the mechanism design; however, the results in decentralized service architecture are contrary to those. The phenomenon is caused by the factor that the level of service benefit in decentralized service architecture is positively associated with the number of contributors. Also, we examine social welfare in centralized service architecture and indicate that the government has to understand the real benefits received by different group of consumers if a subsidy is provided.

Keywords: Wireless Service, Service Availability, Queuing Delay, Mechanism Design.

1 Introduction

Wireless access service has become one of the most important approaches to connect to the Internet and World Wide Web since wireless technologies, such as Wi-Fi and WiMAX, were widely applied to daily life, industry, entertainment, and so on. From the perspective of operation, wireless access service can be divided into centralized and decentralized service architectures. For example, Bharat Sanchar Nigam Ltd., world's 7th largest telecommunications company in India, offers centralized wireless service at difference prices according to specific usage rates. Currently, the three usage rates, 400MB, 1.0GB, and 4.0GB, are sold at 220, 350, and 750 dollars, respectively. On the other hand, FON is a successful decentralized wireless service which integrates the bandwidth and hardware resources at peers' computers and serves as a manager for the pooling resources. The contributors who set up FON hotspots (that is, a type of wireless routers) and share their bandwidth can free access other FON hotspots in the world and receive compensation when someone connects to their hotspots. Others without installing FON hotspots have to pay subscription fees to access FON network.

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 202–215, 2012.

Both architectures rely on the economies of scale to offer service, but centralized service architecture can offer better control than decentralized service architecture because bandwidth resources and access activity are directly provided and handled by an organization. Although both architectures count on distributed access points around the world to provide access capability, decentralized service architecture can offer more scalability and flexibility than centralized service architecture because their access points are contributed by consumers, rather than service providers themselves.

Price discrimination, which is also known as mechanism design, is a marketing strategy that provides different versions of a service which sells at different fares. Therefore, motivated by the perspective of wireless service architecture and marketing strategies, we study the link between centralized and decentralized service architecture under which the mechanism design are being made. In addition, we consider both usage rate and service rate to model the impact of average queuing delay on the mechanism design. The number of access points is also considered because it is an important indicator of availability in wireless access service. Since few studies analyze wireless access service from the perspectives of service architecture and marketing strategy concurrently, our research questions in this study are: (1) What is the difference in the mechanism design between centralized and decentralized service architectures? (2) With centralized service architecture, how to decide service quality? Furthermore, how much should a government subsidize a telecommunication service provider if wireless access service benefits the society? (3) With decentralized service architecture, who should a telecommunication service provider organize and entice to contribute access capability?

2 Literature Review

Recently, optimal pricing in a wireless network has been studied at many different aspects. In a mobile decentralized network, Heikkinen [1] considers optimal pricing with transmit power control and proposes a linear congestion pricing scheme for the optimal distributed control of a wireless network. In essence, pricing schemes for wireless service can be recognized according to the SLA (Service Level Agreement), the subscription type, the negotiation capabilities between wireless service providers and their subscribers, the network capacity, the available bandwidth and frequency spectrum, the Wi-Fi hotspots, and the WIMAX base stations [2]. The service coverage supplied by wireless nodes is also an important research issues because an insufficient number of active nodes will lead to disconnected components in wireless mesh networks [3]. Moreover, most of telecommunication service providers also utilize bundling strategy to sell wireless access service and voice service together, or encourage consumers to buy cellular phones at a discount price if they subscribe to a service plan with the price above a threshold [4,5]. All prior studies and evidences indicate the importance of pricing programs in telecommunication industry.

In two papers related to our work, Masuda and Whang [6] study a simple tariff structure which is widely used in telecommunication industry and Bandyopadhyay et al. [7] explore the issue that network service providers are trying to propose different pricing strategies centered on usage. Masuda and Whang [6] consider a monopolistic carrier in a telecommunication market which offers a menu of "fixed-up-to" plans and show that a simple FUT menu structure to the monopolist is better than any nonlinear pricing schedule. Our study extends their work by comparing centralized service architecture with decentralized service architecture under an identical service menu structure. Our results indicate that the pricing and usage schedules may be different when capacity cost in not too expensive. Bandyopadhyay et al. [7] verify that a network service provider should charges two-part tariff rather than a uniform fixed fee or a differential fixed fee. However, no matter which pricing strategy is adopted, their social welfares are the same. Although considering a differential subscription fee, our analysis can be straightforward extended to a two-part tariff setup. In addition, we compare the level of social welfare under centralized and decentralized service architectures so a social planner can base our exploration of the service architectures in wireless access service to make subsidies or tax breaks to wireless service industries.

Our major contribution is extending the existing literature by comparing centralized service architecture with decentralized service architecture under the mechanism design. Queuing delay is closely associated with the service quality provided in wireless access services, while service availability depends on the number of access points in centralized service architecture and the number of contributors in decentralized service architecture. In order to make the implication of the mechanism design in wireless access service more insightful, we define the benefit function which is composed of the number of access points and the amount of usage rates in our model so we can analyze the influence of the number of consumers contributing resource in decentralized service architecture and study the influence of the number of access points in centralized service architecture. Moreover, by comparing a differential subscription fee with a uniform subscription fee, we find a counterintuitive finding contrary to prior studies in the mechanism design, which is caused by the feature of decentralized service architecture.

3 The Centralized Wireless Service Model

In this section, we consider that a telecommunication service provider sells its wireless service under centralized service architecture. Consumers are heterogeneous in the level of benefit and QoS (Quality of Service) derived from the wireless service. For the sake of simplicity, we assume that there are two types of consumers and all consumers with the same type are symmetric. That is, we segment the consumers into H-type (for heavy usage) and L-type (for light usage) consumers, both of whom are characterized by their valuations for the consumption of the service and their usage levels. The number of H-type and L-type consumers are denoted as η_H and η_L; thus, we can express the total number of consumers as $\eta_0 = \eta_H + \eta_L$, and define $\beta = \eta_H / \eta_0$ and $\rho = \eta_H / \eta_L$. A summary of all the variables used in the present paper appears in Table 1, which is given as follows.

Table 1. Notation

Notation	Description
η_0	The number of total consumers
η_H, η_L	The number of *H*-type (*L*-type) consumers
β	The fraction of *H*-type consumers
ρ	The ratio of *H*-type consumers to *L*-type consumers
μ	The service rate
$\lambda_H (\lambda_L)$	The usage rate for *H*-type (*L*-type) consumers
$p_H (p_L)$	The subscription fee for *H*-type (*L*-type) consumers
δ	The threshold for the average queuing delay
Λ	The total usage rate
$\theta_H (\theta_L)$	*H*-type (*L*-type) consumers' valuation for the wireless access service
c	The value of time
$\upsilon(n,\lambda)$	The benefit of the wireless access service
$W(n,\Lambda)$	The average queuing delay
$F_c(n)$	The management cost in a centralized service architecture (A convex function)
$F_d(n)$	The management cost in a decentralized service architecture (A concave function)
$K(\mu)$	The capacity cost
κ	The marginal capacity cost

In practice, the telecommunication service provider can design different service plans to maximize its profit. The service plans are composed a subscription fee and a usage rate, which can be expressed as (p_H, λ_H) and (p_L, λ_L) for H-type and L-type consumers, respectively. Moreover, in order to construct a ubiquitous wireless service, the telecommunication service provider has to deploy thousands of Wi-Fi hotspots or WiMAX base stations to overcome the transmission-distance limitations presented by wireless technology. For convenience, we use the term "access points" to refer to these wireless devices in the whole study. Most prior studies adopt M/M/1 queue to formulate the average queuing delay for the consumer subscribing telecommunication service, which can be represented as $delay(\mu,\alpha) = 1/(\mu-\alpha)$ where μ is the service rate and α is consumer's total usage rates [6,8,9]. As the number of access points increases, the availability of wireless service becomes higher and the queuing delay is reduced. Therefore, the average queuing delay in a centralized wireless service with multiple access points can be formulated as $W(n,\Lambda) = 1/(\mu - \Lambda/n)$, where n is the number of access points and $\Lambda = \eta_L\lambda_L + \eta_H\lambda_H$ is the total usage rate.

Since wireless service range is associated with the distribution of access points and work throughput is affected by the usage rate, we use the notation $\upsilon(n, \lambda_i)$ where $i \in \{L, H\}$ to represent the benefit for the consumption of the wireless service. We consider $\upsilon(n, \lambda_i)$ the concave function which increases in the number of access points and the usage rate; in addition, we assume that $\partial^2 \upsilon(n, \lambda_i) / \partial n \, \partial \lambda_i > 0$, which is known as the single-crossing condition and standard in the mechanism design literature [10].

From a consumer's perspective, the value of the wireless service can be measured by service availability, the usage rate, and the average queuing delay. According to consumer's type, we denote θ_i as consumer's valuation for wireless access service, where $\theta_H > \theta_L > 0$. As a result, for an *i*-type consumer subscribing a *j*-type plan, his/her value derived from the wireless service can be expressed as

$$\theta_i \left(\upsilon(n, \lambda_j) - c\lambda_j W(n, \Lambda) \right), \text{ where } c \text{ is the value of time.} \tag{1}$$

The telecommunication service provider can rely on the revelation principle to maximize its profit by assuming that *i*-type consumers select (p_i, λ_i) under incentive rationality and incentive compatibility constraints. The capacity cost of the telecommunication service provider can be expressed as $K(\mu) = \kappa \cdot \mu$ since each access point needs individual bandwidth and access capability to receive and respond messages (that is, data packets). In addition, the service provider has to integrate multiple access points to offer wireless service; therefore, the management cost derived from handling these access points is denoted as $F_c(n)$. Accordingly, its optimization problem can be expressed as follows:

$$\begin{aligned}
&\underset{\Theta}{Max}\ \pi_c = \eta_L p_L + \eta_H p_H - n \cdot K(\mu) - F_c(n) \\
&s.t. \\
&\theta_H \left(\upsilon(n, \lambda_H) - c\lambda_H W(n, \Lambda) \right) - p_H \ge \theta_H \left(\upsilon(n, \lambda_L) - c\lambda_L W(n, \Lambda) \right) - p_L \\
&\theta_L \left(\upsilon(n, \lambda_L) - c\lambda_L W(n, \Lambda) \right) - p_L \ge \theta_L \left(\upsilon(n, \lambda_H) - c\lambda_H W(n, \Lambda) \right) - p_H \\
&\theta_H \left(\upsilon(n, \lambda_H) - c\lambda_H W(n, \Lambda) \right) - p_H \ge 0 \\
&\theta_L \left(\upsilon(n, \lambda_L) - c\lambda_L W(n, \Lambda) \right) - p_L \ge 0
\end{aligned}, \tag{2}$$

where $\Theta = \{ p_L, p_H, \lambda_L, \lambda_H \}$.

Notice that the above model follows the approach proposed by Masuda and Whang [6], in which each consumer has measure zero so a single consumer's increase in the usage rate alone imposes no externalities to other jobs. By adopting the approach, the computing complexity can be eased so we can examine the analytical results existing in the model. Moreover, in order to entice consumers to subscribe the wireless service in place of fixed lines, most telecommunication service providers guarantee a certain service quality to enhance consumers' trusts. Queuing delay is a common indicator of service quality, so we denote δ as the indicator that the average queuing delay cannot be higher than. Based on the mechanism design literature [10], we can have optimal subscription fees as follows.

$$\begin{aligned} p_H^* &= \theta_{\mathrm{H}}\left(\upsilon(n,\lambda_H) - c\lambda_H\delta\right) - \theta_{\mathrm{H}}\left(\upsilon(n,\lambda_L) - c\lambda_L\delta\right) + p_L^* \\ p_L^* &= \theta_{\mathrm{L}}\left(\upsilon(n,\lambda_{\mathrm{L}}) - c\lambda_L\delta\right) \end{aligned} \tag{3}$$

For convenience, we use the notation f_x to indicate $\partial f / \partial x$. Thus, if the interior solution exists, we can derive implicit solutions for the optimal usage rates as follows.

$$\upsilon_\lambda\left(n,\lambda_H^*\right) = c\cdot\delta + \kappa/\theta_{\mathrm{H}}, \quad \upsilon_\lambda\left(n,\lambda_L^*\right) = c\cdot\delta + \kappa/\left(\theta_{\mathrm{L}} - \rho\left(\theta_{\mathrm{H}} - \theta_{\mathrm{L}}\right)\right) \tag{4}$$

Lemma 1

The more the number of access points, the higher the usage rate. Formally, $\partial\lambda_H^*/\partial n > 0$ and $\partial\lambda_L^*/\partial n > 0$. The usage rate for L-type consumers decreases with the fraction of H-type consumers. Formally, $\partial\lambda_L^*/\partial\beta < 0$.

When these consumers select their service plans according to individual type, we can utilize the envelop theorem to examine the relation between the number of access points and other factors in centralized service architecture, which is given by solving the following equation.

$$\begin{aligned} &\partial F_c\left(n^*\right)/\partial n = \\ &\eta_L\left(\theta_{\mathrm{L}}\upsilon_n\left(n^*,\lambda_L^*\left(n^*\right)\right)\right) + \eta_H\left(\theta_{\mathrm{H}}\upsilon_n\left(n^*,\lambda_H^*\left(n^*\right)\right) - \left(\theta_{\mathrm{H}} - \theta_{\mathrm{L}}\right)\upsilon_n\left(n,\lambda_L^*\left(n^*\right)\right)\right) - \frac{\kappa}{\delta} \end{aligned} \tag{5}$$

Subsequently, we use the general form $\upsilon(n,\lambda) = n^{\sigma_n}\lambda^{\sigma_\lambda}$ to examine how other factors affect the level of service availability when the telecommunication service provider makes an attempt at enhancing its profit.

Proposition 1 (The Special Case with $\sigma_n = \sigma_\lambda = 1/2$)

1. The higher level of benefit H-type (L-type) consumers can receive, the higher level of service availability. Formally, $\partial n^*/\partial\theta_L > 0$ and $\partial n^*/\partial\theta_H > 0$.
2. The level of service availability increases in the fraction of H-type consumers when the marginal capacity cost is sufficiently small. On the other hand, the opposite holds true when the marginal capacity cost is sufficiently large. Formally, $\partial n^*/\partial\beta > 0$ for a small κ, but $\partial n^*/\partial\beta < 0$ for a large κ.
3. The level of service availability increases in the average queuing delay when capacity cost is sufficiently large. On the other hand, the opposite holds true when the marginal capacity cost is sufficiently small. Formally, $\partial n^*/\partial\delta > 0$ for a large κ, but $\partial n^*/\partial\,\delta < 0$ for a small κ.
4. The higher the value of time, the lower level of service availability.

In the present study, we only consider the special case where $\sigma_n = \sigma_\lambda = 1/2$; however, given a general concave function (that is, $\sigma_n \neq \sigma_\lambda$), we believe the above results still hold, and a further rigid examination can be conducted in the future. The telecommunication service provider can deploy more access points when the benefit of wireless service increases. However, we also indicate that it should downsize the operation of wireless service under some certain conditions. First, if the time value gets important, reducing the workload is better than increasing the number of access points if the average queuing delay is difficult to be altered. Second, if the marginal capacity cost is expensive, the telecommunication service provider can consider decreasing the number of access points when the fraction of the consumers with high demand (or high willingness-to-pay) increases. Even though they can pay more than other consumers, high usage rate under a high marginal capacity cost structure will be a disaster. Lowering their workloads is more profitable and easier than increasing the number of access points. Likewise, when the average queuing delay is altered, the optimal number of access points will depend on the amount of capacity cost. For instance, if the capacity cost is expensive, the telecommunication service provider shouldn't increase the number of access points when the requirement of the average queuing gets rigid. That is, the telecommunication service provider can reduce the number of data packets instead.

In recent years, some scholars have examined whether the government should subsidize telecommunication service providers to expand wireless service [11,12,13]. The primary function of subsidy is to increase the availability of wireless services in rural area. Accordingly, we examine the subsidy the government offers from the perspective of social welfare. The social welfare can be expressed as follows:

$$SW_c = \pi_c + CS_c, \tag{6}$$

where CS is consumer surplus, which is defined as the sum of the utility of all consumers.

Proposition 2

Based on the number of access points, the government can offer a subsidy to enhance social welfare. Formally, the subsidy is given by $\psi(n) = (\theta_{\mathrm{H}} - \theta_{\mathrm{L}})\eta_H \upsilon(n, \lambda_L)$.

We find that the subsidy increases with the number of H-type consumers and the government has to raise the subsidy when the difference in the level of benefit perceived between H-type consumers and L-type consumers becomes large. That is, if H-type consumers have higher information rent, the government should compensate telecommunication service providers more. Otherwise, telecommunication service providers have no incentive to deploy more access points since the information rent held by H-type consumers decreases its profit. Therefore, in order to successfully raise social welfare, the government should play the role of the third party to investigate the level of benefit received by business consumers and home consumers since the subsidy is positively associated with the difference in the level of benefit.

4 The Decentralized Wireless Service Model

Although most telecommunication service providers, like Boingo and CLEAR, utilize centralized service architecture to offer wireless services, some other service providers like FON operate under decentralized service architecture. In a decentralized service environment, service availability is provided by a part of consumers, rather than telecommunication service providers themselves. The fee charged from the consumers providing access capability may be positive or negative. A negative fee represents a compensation a telecommunication service provider pays to consumers playing the role of access points. Accordingly, if a telecommunication service provider decides to organize H-type consumers and entice them to contribute their bandwidth, we can express its profit optimization problem as follows:

$$\max_{\Theta} \pi_d(H) = \eta_L p_L + \eta_H p_H - F_d(\eta_H)$$

$$s.t.$$

$$\begin{aligned} &\theta_H(\upsilon(\eta_H,\lambda_H) - c\lambda_H W(\eta_H,\Lambda)) - p_H - K(\mu) - F_c(1) \geq \theta_H(\upsilon(\eta_H,\lambda_L) - c\lambda_L W(\eta_H,\Lambda)) - p_L \\ &\theta_L(\upsilon(\eta_H,\lambda_L) - c\lambda_L W(\eta_H,\Lambda)) - p_L \geq \theta_L(\upsilon(\eta_H,\lambda_H) - c\lambda_H W(\eta_H,\Lambda)) - p_H - K(\mu) - F_c(1) \\ &\theta_H(\upsilon(\eta_H,\lambda_H) - c\lambda_H W(\eta_H,\Lambda)) - p_H - K(\mu) - F_c(1) \geq 0 \\ &\theta_L(\upsilon(\eta_H,\lambda_L) - c\lambda_L W(\eta_H,\Lambda)) - p_L \geq 0 \end{aligned} \tag{7}$$

Likewise, we can derive a mirror model if the telecommunication service provider organizes L-type consumers to serve. Notice that H-type consumers in (7) have to pay the capacity cost $K(\mu)$ and the management cost $F_c(1)$, just like what a telecommunication service provider pays $n \cdot K(\mu)$ and $F_c(n)$ in centralized service architecture. However, it is possible that H-type consumers pay a smaller subscription fee or receive compensation but turn off their services to save operation cost. In this case, the telecommunication service provider has to audit consumers' behavior and fine consumers if they fail to offer services. Accordingly, the telecommunication service provider can develop an auditing technology to ensure that consumers contribute their bandwidth. In practice, the fines cannot exceed the consumers' wealth, and the policy involves no change in quantities relative the analysis in the section unless their wealth is affected by the subscription fee or the benefit received from the wireless access service. Thus, we treat this concern as a future research. The management cost for integrating and auditing these consumers are denoted as $F_d(\eta_H)$. The optimal subscription fee is given as follows.

$$\begin{aligned} p_H^* &= \theta_H(\upsilon(\eta_H,\lambda_H) - c\lambda_H W(\eta_H,\Lambda)) - \theta_H(\upsilon(\eta_H,\lambda_L) - c\lambda_L W(\eta_H,\Lambda)) \\ &\quad - K(\mu) - F_c(1) + p_L^* \\ p_L^* &= \theta_L(\upsilon(\eta_H,\lambda_L) - c\lambda_L W(\eta_H,\Lambda)) \end{aligned} \tag{8}$$

Lemma 2

In centralized service architecture, H-type consumers pay more than L-type consumers; however, the opposite may hold true in decentralized service architecture when capacity cost is expensive. In addition, the capacity cost in decentralized service architecture decreases with the fraction of H-type consumers. Formally, $p_H^* > p_L^* > 0$ in centralized service architecture and $p_H^* < 0 < p_L^*$ in decentralized service architecture when κ is sufficiently large. Moreover, $\partial K(\mu^*)/\partial\beta < 0$.

Next, we can derive the implicit optimal usage rates as follows.

$$\upsilon_\lambda(\eta_H,\lambda_H^*) = c\cdot\delta + \kappa/\theta_{\mathrm{H}}, \quad \upsilon_\lambda(\eta_H,\lambda_L^*) = c\cdot\delta + (1-\beta)\kappa/(\theta_{\mathrm{L}} - \beta\theta_{\mathrm{H}}) \tag{9}$$

To compare the service plans between centralized and decentralized service architectures, we consider the number of access points is fixed in centralized service architecture and derive the following findings.

Proposition 3

1. In centralized service architecture, as the fraction of H-type consumers increases, the usage rate for H-type consumers remains unchanged but the usage rate for L-type consumers decreases. Formally, $\partial\lambda_H^*/\partial\beta = 0$ and $\partial\lambda_L^*/\partial\beta < 0$.
2. In decentralized service architecture, as the fraction of H-type consumers increases, the usage rate for H-type consumers increases, but the usage rate for L-type consumers may increase or decrease, which is conditional on capacity cost. Formally, $\partial\lambda_H^*/\partial\beta > 0$, and $\partial\lambda_L^*/\partial\beta > 0$ when κ is sufficiently small.
3. In either service architecture, the subscription fee for H-type consumers increases in the fraction of H-type consumers. Formally, $\partial p_H^*/\partial\beta > 0$.
4. In centralized service architecture, the subscription fee for L-type consumers decreases in the fraction of H-type consumers, but the opposite may hold true in decentralized service architecture when capacity cost is inexpensive. Formally, $\partial p_L^*/\partial\beta < 0$ in centralized service architecture, but $\partial p_L^*/\partial\beta > 0$ in decentralized service architecture when κ is sufficiently small.

These findings exhibit the impact of the fraction of H-type consumers on the two service plans in different service architectures. We consider the case where the fraction of H-type consumers increases. In centralized service architecture, we find that the telecommunication service provider intends to raise the subscription fee for H-type consumers to enhance its profit. Consequently, it has to lower the subscription fee and the system usage for L-type consumers to prevent H-type consumers from selecting the service plan for L-type consumers. On the other hand, we find the provider in decentralized service architecture may enhance its profit by raising the subscription fee for H-type consumers; however, it may raise the subscription fee and the system usage for L-type consumers when marginal capacity cost is sufficiently low. The difference is caused by the advantage of decentralized service architecture; that is, the more H-type consumers, the more accessing points.

A numerical example is shown in Figure 1 and 2 to demonstrate the difference in usage rates and subscription fees between centralized and decentralized service architecture. Figure 1 shows that the usage rates in centralized access service cannot be raised with the fraction of H-type consumers, but the opposite holds true in decentralized service architecture. Figure 2 shows that the subscription fees in centralized service architecture can be raised or decreased with the fraction of H-type consumers, but both can be raised in decentralized service architecture. So far, we only discuss the case in which H-type consumers contribute their access capability. In the following we examine the case in which access capability is provided by L-type consumers.

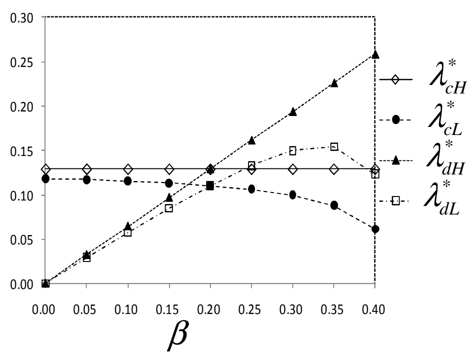


Fig. 1. The optimal usages in centralized and decentralized access service

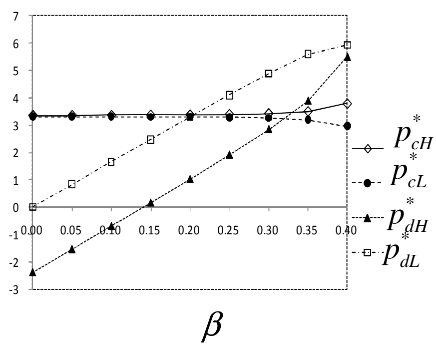


Fig. 2. The optimal subscription fees in centralized and decentralized access service

$v(n,\lambda)=\sqrt{n\lambda}$, $\eta_0 = 100$, $n = 20$, $\kappa = 2$, $c = 3$, $\delta = 2$, $F_c(n) = n^2/5$, $F_d(n) = \sqrt{n}$, $\theta_H = 9$, $\theta_L = 4$

When the fraction of H-type consumers is equal to the fraction of L-type consumers, regardless of the type of consumers contributing access capability, the telecommunication service provider gains the same profit in decentralized service architecture. Comparing the two scenarios in which access capability is provided by H-type or L-type consumers, we find that the telecommunication service provider gains the same profit when the number of H-type consumers is equal to the number of L-type consumers. The result helps us figure out how a telecommunication service provider in decentralized service architecture organizes its distributed access points. Accordingly, if $\beta < 1/2$ ($\beta > 1/2$), the telecommunication service provider implementing two service plans in decentralized service architecture should organize L-type (H-type) consumers to contribute their access capability.

That is, in decentralized service architecture, if the fraction of H-type consumers is higher than that of L-type consumers, the service provider should entice H-type consumers to contribute their access capability; however, if the fraction of L-type consumers is higher than that of H-type consumers, it should entice L-type consumers to contribute their access capability although the result sounds counterintuitive. In most prior studies, resources or efforts are exerted by an efficient individuals because their effort cost is less than others; however, in decentralized service architecture, we show that this may not be true forever because access resource increases with the number of contributors when auditing technology works well, which is more important than contributors' perceived benefit or cost.

Next, we examine the scenario in which the telecommunication service provider only offers a single service plan. As Masuda and Whang [6] stated, the service provider can design a single service plan to either attract both types of consumers (is also known as pooling strategy) or attract only H-type consumers. Other cases are impossible or gain a lower profit. Therefore, offering a single service plan to attract both types of consumers, the service provider's profit in decentralized service architecture can be expressed as:

$$\begin{aligned} &\underset{p_L,\lambda_L}{Max}\ \pi_{dL} = \eta_0 p_L - F_d\left(\eta_0\right) \\ &s.t.\quad \theta_L\left(\upsilon\left(\eta_0,\lambda_L\right) - c\lambda_L W\left(\eta_0,\Lambda\right)\right) - p_L - K\left(\mu\right) - F_c\left(1\right) \geq 0 \end{aligned} \tag{10}$$

Likewise, when offering a single service plan to attract only H-type consumers, the service provider's profit can be expressed as:

$$\begin{aligned} &\underset{p_H,\lambda_H}{Max}\ \pi_{dH} = \eta_H p_H - F_d\left(\eta_H\right) \\ &s.t.\quad \theta_H\left(\upsilon\left(\eta_H,\lambda_H\right) - c\lambda_H W\left(\eta_H,\Lambda\right)\right) - p_H - K\left(\mu\right) - F_c\left(1\right) \geq 0 \end{aligned} \tag{11}$$

Lemma 3

1. For the single service plan, when the fraction of H-type consumers is higher than a specific threshold, attracting only H-type consumers can gain more profit than attracting both types of consumers. Formally, there exists $\tilde{\beta}$ such that $\pi^*_{dH} \geq \pi^*_{dL}$ when $\beta \geq \tilde{\beta}$ but $\pi^*_{dH} \leq \pi^*_{dL}$ when $\beta \leq \tilde{\beta}$.
2. If the single service plan is attractive to both types, the profit of the two service plans is higher (lower) than that of the single service plan when the profit of the two service plans increases (decreases) with the fraction of H-type consumers. Formally, $\pi^*_d\left(L\right) \leq \pi^*_{dL}$ if $\partial\pi^*_d\left(L\right)\big/\partial\beta < 0$ and $\pi^*_d\left(L\right) \geq \pi^*_{dL}$ if $\partial\pi^*_d\left(L\right)\big/\partial\beta > 0$.
3. If the single service plan is attractive to only H-type consumers, the profit of the two service plans is always higher than that of the single service plan. Formally, $\pi^*_d\left(H\right) \geq \pi^*_{dH}$.

When a single service plan is attractive to both types, traditional mechanism design shows that offering two service plans is better than offering a single service plan because the model of two service plans can degenerate to the model of a single service plan by letting H-type consumers pay the same subscription fee and receive the same usage rate as L-type consumers. However, our result indicates that this may be violated in decentralized service architecture because all consumers contribute may be better than a portion of consumers contributes.

Proposition 4 (The Optimal Service Plans in Decentralized Service Architecture)

1. If the fraction of H-type consumers is sufficiently large, the telecommunication service provider should offer the two service plans in which access points are composed of H-type consumers.
2. If the fraction of L-type consumers is sufficiently large, the telecommunication service provider has to compare the two service plans in which access points are composed of L-type consumers with the single service plan in which access points are composed of all consumers.

Our findings help the telecommunication service provider reduces its decision space and facilitate reaching the optimal profit in decentralized service architecture. In the end, we compare the social welfare of the two service plans between centralized and decentralized service architectures. The notation of social welfare in decentralized service architecture is given as follows.

$$SW_d = \pi_d + CS_d \tag{12}$$

Proposition 5 (The Comparison in Social Welfare between Different Service Architectures)

When the two service plans are adopted in both service architecture, if the number of access points in centralized service architecture is equal to the number of contributors in decentralized service architecture, whether the social welfare in centralized service architecture is higher or lower than that in decentralized service architecture is conditional on the management costs in the two service architectures. Formally, $SW_d \geq SW_c$ if and only if $F_c(n) \geq \eta_i \cdot F_c(1) + F_d(\eta_i)$, where *i*-type consumers contribute their access capability.

Comparing the social welfare in centralized and decentralized service architectures, we find the importance of the management cost. For decentralized service architecture, if the telecommunication service provider has to spend more to integrate and audit decentralized access points, centralized service architecture may be better than the other from the perspective of social welfare. Therefore, we claim that a non-profit institution adopting decentralized service architecture shouldn't concentrate on auditing technology too much because this may reduce social welfare. Otherwise, centralized service architecture will be a better choice for non-profit organizations.

5 Conclusion

In this study, we compare two common used architectures in wireless access service: centralized and decentralized service architectures. We summarize the interesting findings as follows. First, examining the two service plans in decentralized service architecture, with low capacity cost, we find that the subscription fees and usage rates for consumers with low willingness-to-pay may increase in the fraction of consumers with high willingness-to-pay. However, this outcome wouldn't appear in centralized service architecture. Second, we point out the optimal number of access points in centralized service architecture is associated with the level of capacity cost. That is, deploying more access points can enhance service quality so as to raise subscription fees; however, reducing the number of access points can save capacity cost. In addition, if the government intends to subsidize a telecommunication service provider, our results indicate that the government has to observe the benefit which different types of consumers receive.

Third, we suggest that a telecommunication service provider in decentralized service architecture entice the biggest group of participants to contribute their access resources because the benefit of wireless access is positively associated with the number of access points. Moreover, we also compare a single service plan (uniform subscription fee) with two service plans (differential subscription fees) in decentralized service architecture. If access points in two service plans are composed of consumers with low willingness-to-pay, we find that the profit of the pooling strategy can be better than that of two service plans under certain condition, which is contrary to prior studies in mechanism design. The counterintuitive finding is caused by the fact that the benefit of wireless access service under decentralized service architecture is positively associated with the number of contributors. As a result, the strategy of uniform subscription fee may recruit more number of contributors than the strategy of differential subscription fees, leading higher revenue for telecommunication service provider.

In order to prevent contributors from turning off their services in decentralized service architecture, auditing technology is substantial. Obviously, if consumer's wealth may be affected by service quality and availability, our results can be refined in a future research, which discusses the impact of the auditing cost on mechanism design in decentralized service architecture. In addition, we only consider a simple setup in which there are two types of consumers in the market. In the future, this study can be extended to analyze the case with multiple types.

Acknowledgments. The author would like to thank the two anonymous reviewers for their insightful comments and helpful suggestions. This research was supported by the National Science Council of Taiwan (Republic of China) under the grant NSC 100-2410-H-266 -005.

References

1. Heikkinen, T.M.: On congestion pricing in a wireless network. Wireless Networks 8, 347–354 (2002)
2. Gizelis, C.A., Vergados, D.D.: A Survey of Pricing Schemes in Wireless Networks. Communications Surveys & Tutorials 13, 126–145 (2011)
3. Allen, S.M., Whitaker, R.M., Hurley, S.: Personalised subscription pricing for optimised wireless mesh network deployment. Computer Networks 52, 2172–2188 (2008)

4. Lahiri, A., Dewan, R.M., Freimer, M.: The disruptive effect of open platforms on markets for wireless services. Journal of management information systems 27, 81–110 (2010)
5. Yang, B., Ng, C.T.: Pricing problem in wireless telecommunication product and service bundling. European Journal of Operational Research 207, 473–480 (2010)
6. Masuda, Y., Whang, S.: On the optimality of fixed-up-to tariff for telecommunications service. Information Systems Research 17, 247–253 (2006)
7. Bandyopadhyay, S., Cheng, H.K., Guo, H.: Usage-based Pricing and Broadband users differentiation. Working Paper for Public Utility Research Center (2009)
8. Tan, Y., Mookerjee, V.S.: Allocating spending between advertising and information technology in electronic retailing. Management Science 51, 1236–1249 (2005)
9. Fan, M., Kumar, S., Whinston, A.B.: Short-term and long-term competition between providers of shrink-wrap software and software as a service. European Journal of Operational Research 196, 661–671 (2009)
10. Salanié, B.: The Economics of Contracts: primer. Massachusetts Institute of Technology, Cambridge Center, MA 02142 (2005)
11. Proenza, F.J.: The Road to Broadband Development in Developing Countries Is through Competition Driven by Wireless and Internet Telephony. Information Technologies and International Development 3, 21–39 (2007)
12. Yuguchi, K.: The digital divide problem: An economic interpretation of the Japanese experience. Telecommunications Policy 32, 340–348 (2008)
13. Abolhasan, M., Wright, A.: Survey of the potential of emerging wireless technologies to improve telecommunication services in remote Australian settlements. The Rangeland Journal 30, 157–165 (2008)

Parallel Importation: An Empirical Investigation of Online Unauthorized Distribution Channels for Luxury Fashion Goods

Kexin Zhao[1], Xia Zhao[2], and Jing Deng[2]

[1] University of North Carolina at Charlotte
kzhao2@uncc.edu
[2] University of North Carolina at Greensboro
{x_zhao3,jing.deng}@uncg.edu

Abstract. Parallel importation is prevalent as billions of dollars worth of genuine products are sold by unauthorized distributors across countries. In this research, we offer one of the first empirical investigations of online parallel importation. We find that for luxury handbags, less expensive products, products with greater market interests and products available via the authorized channel have more parallel importation activities. In addition, there are fewer parallel importation activities for the more prestigious brand.

Keywords: Parallel Importation, Gray Market, Electronic commerce, Brand prestige.

1 Introduction

Parallel importation occurs when the branded products are priced lower in the home countries and the individuals and enterprises can profit by sourcing products in the home countries and selling them in the export countries without the brand owners' authorization [5]. Together with "reimportation" and "lateral importation" [1], it is often referred to as "gray market" [6]. Parallel importation is endemic across a wide variety of industries, ranging from lumber, electronic components, broadcast signals, IPOs, automobiles, heavy construction equipment, watches, cosmetics, bags, health, beauty aids and prescription drugs [4]. For instance, CNN Money reported that half of the iPads sold through Apple's Fifth Avenue store on a busy March Friday were purchased for resale overseas [9]. Uneven purchasing power, tariff charges, exchange rate changes and competition all contribute to the rise of price gaps and parallel importation [6].

The emergency of electronic commerce alters the landscape of parallel importation. Electronic markets expand unauthorized distribution channels into the

---

[1] Reimportation occurs when then branded products are priced lower in the export countries than the home countries and the firms can profit by purchasing the products in the export countries and selling in the home countries. Lateral importation occurs when there is price difference between two export countries and the firms can profit by purchasing the products in the low-priced countries and selling in the high-priced countries.

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 216–224, 2012.

cyberspace, breaking the geographical barriers and reaching out to a wide range of consumers. The ubiquitous accessibility provided by the Internet also enables manufacturers and brand owners to use emerging technologies, such as web crawlers, to search billions of webpages on foreign e-commerce websites and monitor parallel importation activities.

While parallel importation draws a great deal of attention from practitioners, research on parallel importation is scarce [1, 2]. Ahmadi and Yang [1] and Ahmadi et al. [2] use analytical models to study pricing issues in both the authorized channels and parallel importation channels. Surveys of US manufacturers or exporters have been administered to examine factors driving or deterring parallel importation activities [4, 16]. However, little work (of which we are aware) has empirically examined the market dynamics in parallel importation channels due to the sensitive nature of the phenomenon and limited data availability. To fill the gap, we intend to empirically analyze characteristics of online parallel importation channels. To the best of our knowledge, this is the first study to directly observe and analyze market activities in online parallel importation channels. Our research helps unveil the dynamics of gray market.

We focus on the parallel importation activities on Taobao.com, the dominant C2C trading platform in China [18]. Taobao.com is the largest Internet retail and trading website in China, servicing more than 800,000 sellers and nearly 170 million registered users. Annual transaction volume on Taobao.com reached 208 billion CNY in 2009, accounting for approximately 80% of China's e-commerce market.

We choose the category of luxury handbag as our research subject. Parallel importation incidents are more likely to happen when a product has a prominent brand name [8] and large price differential [17]. Luxury handbags in the Chinese market satisfy both conditions. In addition, “handbags are the engine that drives luxury brands today” [19].

We collect the listing and transaction data of 183 unique Coach handbag styles and 246 unique LV handbag styles in June 2011 for our preliminary analysis. LV and Coach are both ranked as the top 12 designer bag brands [13]. These two brands represent two distinct classes of brand prestige as LV always holds the top position among most powerful luxury brand in various lists and Coach, renowned as the “affordable luxury” brand, introduces the accessible luxury to the masses [12]. This enables us to examine the impact of brand prestige on the parallel importation activities.

Our preliminary data analysis suggests that for the luxury handbags, less expensive products, products with greater market interests and products available via the authorized channels have more parallel importation activities. In addition, there are fewer parallel importation activities for the more prestigious brand. Across the brands, the negative impact of product prices on parallel importation activities does not differ, but the positive impact of potential market interests is weakened for the more prestigious brand. We also find that product characteristics, such as new arrival, size, and material, matter in parallel importation channels.

The remainder of the paper is organized as follows. In Section 2, we review the literature and develop our hypotheses. We describe our data in Section 3 and present our preliminary empirical model and results in Section 4. Section 5 discusses the current status and plan for future research.

2 Theory and Hypotheses

To have the first grasp of online parallel importation, we focus on the degree of parallel importation activities in online markets. In the previous literature, due to data availability, parallel importation activity is measured as the likelihood that it occurred during the previous two years [5, pp. 95] or whether it was a significant problem to the exporter [16]. Since we are able to directly observe products listed in and sold via unauthorized channels, we use the number of sellers and transaction quantity to assess the degree of online parallel importation activities. In the case of luxury handbags, the majority of buyers only buy one handbag in a single transaction. Therefore, transaction quantity approximately represents the number of buyers in the market.

The major driver of parallel importation is the difference of the product prices in the authorized channels between home countries and in the export countries [21]. Due to the brand owners' coordinated pricing strategy, a product's price differential—price difference between the home country and the export country—is highly correlated to its base price[2] (i.e., price in the home country). A higher base price leads to a higher price differential, which yields higher profit margins in the unauthorized channels and hence attracts more sellers to pursue the arbitrage opportunities. Large price differential also motivates buyers to switch to unauthorized channels to obtain the product at a discounted price. However, the base prices, even though positively correlated to the price differentials, have an opposite impact on the supply and demand in the gray market. When the base price is higher, sellers incur higher cost to acquire the product via the authorized channels from the original country and face higher risk of lost sales. Buyers may also become reluctant to shop more expensive products in unauthorized channels. For products with higher prices, consumers tend to value the relative savings less [15]. Furthermore, only consumers with higher willingness to pay can afford more expensive products. Those consumers tend to prefer authorized channels due to high values from associated benefits such as warranty and services [1]. Thus, we expect to observe fewer parallel importation activities with more expensive products and we hypothesize that:

H1a: The number of sellers in parallel importation channel is negatively associated with the product's base price.

H1b: Transaction quantity in parallel importation channel is negatively associated with the product's base price.

The ownership of luxury branded goods signals social status. Simply use or display a branded product reflects the owner's association with or dissociation from a specific social group [12]. Consumers in the social group seeking high prestige have a higher willingness to pay. They tend to value more the associated services and prefer the authorized channels. On the other hand, the brands signaling higher prestige are also priced higher on average, yielding a higher risk of lost sales for the parallel importers. Therefore, we expect the more prestigious brand has fewer parallel importation activities.

---

[2] The Pearson Correlation between Coach's US prices and price differential between authorized Chinese distribution channels and US distribution channels is 0.952 ($p<0.01$). The Pearson Correlation between LV's France prices and price differential between authorized Chinese distribution channels and French distribution channels is also 0.952 ($p<0.01$).

H2a: The number of sellers in parallel importation channel is negatively associated with the brand prestige.

H2b: Transaction quantity in parallel importation channel is negatively associated with the brand prestige.

Market demand for branded products is another driver of gray market activities [5]. Parallels importation arises to meet consumers' needs for the branded products in the export countries. Products with higher market interests have a larger market size and higher profit potentials, attracting more sellers to enter the parallel importation channels. On the buyer side, more transactions can happen if more shoppers indicate potential interests in a product [1]. Therefore, we expect to observe positive relationships between parallel importation activities and potential market interests.

H3a: The number of sellers in parallel importation channel is positively associated with the product's potential market interests.

H3b: Transaction quantity in parallel importation channel is positively associated with the product's potential market interests.

A product is not always available in different geographic regions. For instance, in our case, only 52 out of 183 styles are available from the official Coach distribution channels in China. Parallel importers gain more from products available in the authorized channel by free-riding off advertising and marketing efforts from the authorized distributors [3]. The products may also acquire a better awareness among potential buyers if they are displayed or promoted by the brand owner. Thus, we hypothesize that:

H4a: The number of sellers in gray market is positively associated with the product's availability in the authorized channel.

H4b: Transaction quantity in gray market is positively associated with the product's availability in the authorized channel.

3 Research Method

As we have discussed before, we focus on the case of luxury handbags due to its premium positioning and large price disparity between the home countries and the Chinese market. We selected both the most prestigious luxury brand, LV, and the "affordable luxury" brand, Coach. LV has been often top ranked among the most powerful luxury brands whereas Coach is renowned for introducing the accessible luxury to the masses [12]. They are both popular luxury brands among Asian consumers [11, 14] but represent different classes of brand prestige. We include all 183 Coach handbag styles available in US as of April 2011 and obtain authorized channel prices via official US and China Coach websites. For LV, we include 246 LV handbag styles and obtain authorized channel prices via official France LV websites and the offline distribution channel in China[3].

---

[3] Prices are not listed on the official LV China website.

Parallel importation data were collected from www.taobao.com. We fed the Coach and LV style numbers into a Linux shell script to search for all matching items that were listed on Taobao.com. On Taobao, sellers create a webpage with a unique listing ID for each single product style they sell. Consequently, for each style, the number of listings equals the number of sellers. For each listing, we collected all related information such as price, shipping cost, seller reputation scores, transaction history, etc. Data collection was performed once per week. In this preliminary analysis, we have collected data from May 30th to June 28th, about 4 weeks.

The main tools used in the information retrieval included perl, wget, python, as well as other standard tools such as grep, sed, wc, etc. The main webpage retrieval engine was the urllib2 library for python, which was capable of loading dynamic data such number of views and the times being bookmarked. It also addressed the Chinese encoding problem on some webpages that used GBK encoding.

To exclude possible counterfeit products, we only included oversea sellers, who can easily obtain genuine handbags in low-priced regions. They are referred to as DaiGou—people who purchase products for other consumers in a different geographic location. We also exclude listings with suspiciously low prices[4]. Table 1 lists the summary statistics for our preliminary sample.

Table 1. Summary Statistics

		Observations	**Mean**	**St. Deviation**	**Min**	**Max**
Posted Price[5] (CNY)	Coach	26010	2596.33	1189.87	652.00	16042.00
	LV	21143	8055.73	7390.25	1486.00	444444.00
Bookmark	Coach	26010	10.75	78.03	0	992
	LV	21143	4.38	25.99	0	1222
Transaction Price (CNY)	Coach	696	1964.79	762.22	749.00	6888.00
	LV	225	6438.22	2749.95	2024.00	20500.00
Transaction quantity	Coach	696	1.81	2.67	1	41
	LV	225	3.08	4.36	1	27
Official Price in Low-Priced Country (CNY[6])	Coach	732	2367.07	1310.57	439.64	9073.72
	LV	984	10395.86	5978.55	1596.19	31184.67
Official Chinese Price (CNY)	Coach	52	5107.69	2041.85	1900.00	11900.00
	LV	246	13204.88	7571.05	2100.00	39000.00

4 We delete sellers whose posted prices are 30% lower than the official prices.

5 The post prices are calculated by totaling the listing prices and the minimum shipping charges.

6 A handbag corresponds to four price observations taking fluctuation of exchange rates into account.

4 Preliminary Econometric Model and Results

We conduct a product-week level analysis by aggregating the number of listings and transaction quantity for each product style in a week. The following specifications are used to test our hypotheses. Table 2 provides illustrative description of variables. We measure the potential market interests using the times of a product's listing pages being bookmarked. Since a larger number of listing or sellers leads to a higher chance of being bookmarked, our analysis uses the average counts of a product being bookmarked per listing. We capture the brand effect using a dummy variable with 1 representing LV and 0 representing Coach.

$$SELLER_{it} = \alpha_{10} + \alpha_{11}PRICE_i + \alpha_{12}AVAILABILITY_i + \alpha_{13}BMARK_{it} + \alpha_{14}NEW_i + \alpha_{15}SIZE_i + \alpha_{16}MATERIAL_i + \alpha_{7}BRAND_i + \varepsilon_{1it}$$

$$TRANSACTION_{it} = \alpha_{20} + \alpha_{21}PRICE_i + \alpha_{22}AVAILABILITY_i + \alpha_{23}BMARK_{it} + \alpha_{24}NEW_i + \alpha_{25}SIZE_i + \alpha_{26}MATERIAL_i + \alpha_{7}BRAND_i + \varepsilon_{2it}$$

Table 2. Description of Variables

Notation	Variables
i	Product i
t	Time period t
SELLER	Number of sellers per week per style
TRANSACTION	Transaction quantity per week per style
PRICE	Price in the lower-priced country
AVAILABILITY	Whether or not the product is available via the Official China website. (Yes: 1; No: 0)
BMARK	Average times of a product being bookmarked per listing
NEW	Whether or not the product is a new arrival (Yes: 1; No: 0)
SIZE	Size of the product (Small: 1; Medium: 2; Large: 3; Extra Large: 4)
MATERIAL	Material of the product (Leather: 1; Other materials: 0)
BRAND	Brand of the product (LV: 1; Coach: 0)

Our preliminary analysis incorporates the listing and transaction data from two brands in only four weeks, thus we do not expect significant temporal effects. Since the individual equations pertain to the same product style, the error terms may be correlated. Seemingly unrelated regressions (SUR) can handle potentially correlated error terms [20]. Small VIF indicators suggest no significant multicolinearity concern. Results are presented in Table 3.

The results indicate that more parallel importation incidents happen for less expensive products, products with higher market interests and available via the authorized channel in the export country. Thus, H1, H3 and H4 are supported. In addition, fewer sellers offer LV handbags for sale and fewer buyers purchase LV handbags via unauthorized channels compared with Coach. Overall, there are fewer parallel importation activities for the brand with high prestige. Thus, H2 is also supported.

Table 3. Result of Brand Effect

	Number of Sellers	Transaction Quantity
Price	-0.253***	-0.066 **
Availability	0.433***	0.275 ***
Bmark	0.279***	0.296 ***
New	-0.113***	-0.031 (n.s.)
Size	0.064**	-0.008 (n.s.)
Material	-0.093***	-0.054 *
Brand	-0.368***	-0.230 ***
N	1716	1716
R square	0.288***	0.146***

Note: $p<0.001$, ***; $p<0.01$, **; $p<0.05$, *.

The product characteristics also matter in the unauthorized channels. Leather bags are not appealing to either sellers or buyers as fewer sellers offer leather bags and fewer buyers purchase leather bags. The sizing has different implication on the seller and buyer sides. While more sellers offer larger bags, buyers show no difference between different sizes. It takes time for sellers to acquire and distribute new arrivals in the gray market, thus their offering of new arrivals is lagged. However, buyers are ready to purchase those products soon after their release.

To give a closer look at the impact of brand prestige on parallel importation activities, we compare the regression results across two luxury brands. The regression results are presented in Table 4 and Smith-Satterthwaite tests were used to determine whether the difference between the regression coefficients is significant.

Table 4. Group Comparison Results

	No of Seller			Transaction Quantity		
	Coach	**LV**	**Diff-t value**	**Coach**	**LV**	**Diff-t value**
Price	-0.155***	-0.185***	0.59(n.s.)	-0.092*	-0.056(n.s.)	-0.70(n.s.)
Availability	0.303***			0.171***		
Bmark	0.455***	0.129***	8.10***	0.557***	0.078*	10.98***
New	-0.063**	-0.157***	2.82**	0.019(n.s.)	-0.033(n.s.)	1.26(n.s.)
Size	0.182***	0(n.s.)	4.14***	0.148***	-0.091**	5.09***
Material	0.108***	-0.392***	11.89***	-0.03(n.s.)	-0.086*	1.22(n.s.)
N	732	984		732	984	
R Square	0.388***	0.32**		0.391***	0.033***	

Note: $p<0.001$, ***; $p<0.01$, **; $p<0.05$, *.

The negative impact of product base price on parallel importation activities does not differ significantly across the brands. The positive impact of market potential interests on the parallel importation activities is weaker for LV. Many Taobao consumers might follow LV bags only for its prestigious brand name, and their interests might not be readily converted to real transactions.

The negative impact of new arrival on the number of sellers is stronger for LV than for Coach. This is probably because the brand with high prestige tends to have limited supply especially for new arrivals. Therefore it is more difficult to obtain a LV new arrival. But the buyer shows no difference. While larger Coach bags draw more sellers and buyers, sizing does not matter for the LV sellers and larger LV bags are even less appealing to the buyers. The leather material shows countervailing impact on the number of sellers across the brands in the parallel importation channel as the leather Coach bags attract more sellers whereas the LV leather bags attract fewer sellers. On the buyer side, the impact of material does not show difference across the brands.

5 Current Status and Future Research

Currently, we are in the process of collecting more data, which may allow us to observe market dynamics and explore the temporary effects in the parallel importation channels. In addition, many prior studies have examined price dispersion on the electronic markets [7, 10]. We plan to explore the level of price dispersion in online unauthorized channels and identify its drivers.

References

1. Ahmadi, R., Yang, B.R.: Parallel imports: Challenges from unauthorized distribution channels. Marketing Science 19(3), 279–294 (2000)
2. Ahmadi, R., Carr, S.M., Dasu, S.: Gray Markets, Demand Uncertainty and Excess Inventory. Marshall Research Paper Series Working Paper FBE 11-10 (2010)
3. Anderson, E., Weitz, B.A.: Make-or-Buy Decisions: Vertical Integration and Marketing Productivity. Sloan Management Review 27, 3–19 (1986)
4. Antia, K.D., Bergen, M., Dutta, S.: Competing with gray markets. Sloan Management Review 46(1), 62–69 (2004)
5. Antia, K.D., Bergen, M.E., Dutta, S., Fisher, R.J.: How does enforcement deter gray market incidence? Journal of Marketing 70, 92–106 (2006)
6. Assmus, G., Wiese, C.: How to Address the Gray Market Threat Using Price Coordination. Sloan Management Review, 31–41 (Spring 1995)
7. Brynjolfsson, E., Smith, M.D.: Frictionless Commerce? A Comparison of In-ternet and Conventional Retailers. Management Science 46(4), 563–585 (2000)
8. Bucklin, L.: Modeling the International Gray Market for Public Policy Deci-sions. International Journal of Research in Marketing 10(4), 387–405 (1993)
9. Elmer-DeWitt, P.: Inside the iPad 2 Gray Market. CNN Money (March 14, 2011), `http://tech.fortune.cnn.com/2011/03/14/inside-the-ipad-2-gray-market/`
10. Ghose, A., Yao, Y.: Using transaction prices to re-examine price cispersion in electronic markets. Information Systems Research, published online ahead of print, February 1 (2010)

11. Han, H.T.: Taiwanese Consumers' Perceptions of Luxury Handbags: The Country-of-Origin Effect. Dissertation, Lynn University (2005)
12. Han, Y.L., Nunes, J.C., Dreze, X.: Signaling status with luxury goods: the role of brand prominence. Journal of Marketing 74, 15–30 (2010)
13. Hart, G.: Top 12 Handbag brands for (January 22, 2008), http://gill-hart.suite101.com/top-12-handbag-brands-for-2008-a42449 (accessed on November 02, 2011)
14. Lee, J., Harp, S.S., Horridge, P.E., Russ, R.R.: Targeting Multicultural Pur-chase and Consumption Segments in the Leather Handbag Markets: Product Development and Merchandising Implications. Family and Consumer Science Research Journal 31(3), 297–330 (2003)
15. Lindsey-Mullikin, J., Grewal, D.: Imperfect Information: The Persistence of Price Dispersion on the Web. Journal of the Academy of Marketing Science 34(2), 236–243 (2006)
16. Myers, M.B.: Incidents of gray market activity among U.S. exported: occur-rences, characteristics, and consequences. Journal of International Business Studies 30(1), 105–126 (1999)
17. Onkvisit, S., Shaw, J.J.: International Marketing: Analysis and Strategy. Merrill Publishing Company, Columbus OH (1989)
18. Ou, C.X., Davison, R.M.: Why eBay lost to TaoBao in China: The global advantage. Communications of the ACM 52(1), 145–148 (2009)
19. Thomas, D.: Deluxe: How Luxury Lost Its Luster. The Penguin Press, New York (2007)
20. Zellner, A.: An Efficient Method of Estimating Seemingly Unrelated Re-gression and Tests for Aggregation Bias. Journal of the American Statistical Association 57, 348–368 (1962)
21. Zhang, Z., Feng, J.: Price Coordination in Multi-Markets with Unautho-rized Distribution Channel. In: Proceedings of the 9th Workshop on e-Business, St. Louis, MO (2010)

Consumer Segmentation and the Information Rule of Online Reviews in Horizontally Differentiated Product Markets

Qingliang Wang and Khim Yong Goh

15 Computing Drive, Singapore, Republic of Singapore
National University of Singapore
wangqingliang@nus.edu.sg, gohky@comp.nus.edu.sg

Abstract. Previous studies have explored the impact of online reviews on product sales at the aggregate level. This study contributes to the literature by investigating how online reviews take effect at the individual consumer level in a horizontally differentiated product market. We empirically test our hypotheses using data from a popular review website in China and consumers' actual dining records. We find evidence that the information role of online reviews is moderated by consumers' geographical locations. Our results have implications for consumer segmentation and targeting of consumers through local market advertising.

Keywords: consumer segmentation, online reviews, economics of information, horizontal differentiation, geographical location.

1 Introduction

In recent years, user-generated online product reviews have grown in popularity rapidly on the Internet. This large-scale sharing of consumption experiences can potentially reduce consumers' uncertainty about product or service quality prior to actual purchases, especially for experience goods whose quality cannot be inspected before purchase. Recent evidence suggests that consumer reviews have become an integral part of consumer purchase decisions [1]. Previous studies have also explored the relationship between online reviews and product sales [2-6]. However, the online reviews literature generally focuses on the average effect of online reviews on product sales at the aggregate level, which cannot take consumers' heterogeneous characteristics into consideration and suffers from the aggregation bias. An individual level analysis can help researchers and marketers gain a deeper understanding of how individual consumers engage, perceive, and use online information channels.

At the individual consumer level, it is still unclear how and to what extent online reviews can influence consumer choice. Because of consumers' heterogeneity, online reviews website may play different information roles for different consumers. For example, in consumers' dining choices, where restaurants are horizontally differentiated by geographical locations, consumers' information needs may depend

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 225–233, 2012.

on their familiarity with different restaurants and their locations. It has been long recognized in retail marketing that geographical factors play an important role in impacting consumers' purchase behavior [7-9]. Specific online businesses, such as search engines, are increasingly using spatial data to provide local content and advertisement [9]. In this study, we posit that the influence of online reviews will be constrained by the geographical locations of where the consumer is and where the product or service is actually available at. To be specific, we empirically explore the following two questions:

(1) Do online reviews influence consumer choice of horizontally differentiated products?
(2) How and to what extent do consumers' geographical locations moderate the information role of online reviews?

Based on the economics of information, we propose: (1) two competing hypotheses about the influence of online reviews on consumers' choice in a horizontally differentiated product market; (2) the information role of online reviews on consumer choice will be moderated by consumers' experiences across different geographical locations. Using reviews data from a popular restaurant review websites in China and consumers' actual dining records, we empirically test our hypotheses at the individual consumer level. Our results shed light on how geographical locations moderate the influence of online reviews and have implications on how firms should target and advertise to different consumer segments in horizontally differentiated products markets.

2 Literature Review

Previous studies have established a link between online reviews and product sales in diverse product categories. Two dimensions of online reviews have been mostly considered: volume and valence of online reviews. Volume refers to the number of online reviews, while valence refers to the average rating of these reviews. Chevalier and Mayzlin [2] find that the differences between consumer reviews posted on Barnes & Noble and Amazon.com were positively related to the differences in book sales in these two retail sites. Forman et al. [5] examined the moderating effect of reviewer identity disclosure on the relationship between reviews and sales. Duan et al. [4] documented the importance of the number of online reviews in influencing movie box office sales. Liu [10] found that while the volume of online word-of-mouth is positively associated with product sales, the relationship between the reviews' valence and sales is often mixed. Dellarocas et al. [3] found that the valence of online consumer reviews is a better predictor of future movie revenues than other measures they considered. Godes and Mayzlin [11] showed that dispersion of conversations across online consumer communities is the main factor that influences sales performance. Clemons et al. [12] found that the variance of ratings and the strength of the most positive quartile of reviews have a significant impact on the growth of craft beers.

This study contributes to the literature in two dimensions. First, instead of modeling the relationship between online reviews and product sales in an aggregate ad-hoc reduced-form manner, our individual consumer level analysis follows the utility-maximization framework. Second, focusing on the horizontally differentiated products market, we segment consumers into different geographical locations and explore how geographical location moderates the influence of online reviews for different consumer segments.

3 Theory and Hypotheses

We adopt the economics of information as our theoretical foundation to address the role of online reviews in consumer choice processes. Consumers usually make choices with incomplete information and it is costly to seek full information. Since the choice uncertainty pegged to a consumer's decision making cannot be totally eliminated, a consumer typically needs to make a choice under risk, and there are trade-offs between the perceived costs and benefits of additional search of information or choice alternatives [13]. The economics of information literature has made a close connection between information and uncertainty [14]. Online reviews are valuable for consumers because they signal product quality which cannot be inspected before purchase (especially for experience products). Therefore, the influence of online reviews will depend on how much uncertainty it can reduce with regard to the true quality of a product or service under consideration.

3.1 Influence of Online Review on Horizontally Differentiated Products

In markets where products are horizontally differentiated by geographical locations, consumers from different locations may have different information needs [9]. From consumers' point of view, products can be either local (i.e., products are in the same location with consumers) or non-local (i.e., products are in different locations with consumers). Previous studies have identified the positive impact of online reviews' volume and valence on product sales [2-6]. The underlying reason is that the volume of reviews indicates the popularity (or awareness) of a product, while the valence of reviews signals the quality of a product. Although review volume cannot signal product quality directly, previous studies have shown strong linkages between a product's popularity and its perceived quality [15,16]. In other words, being popular in itself signals higher quality. Thus, higher volume or valence of online reviews may lead to higher product sales. We follow the same logic and posit that:

In a horizontally differentiated product market,

H1a: the valence of online reviews has a positive relationship with consumers' choice of a local product.

H1b: the volume of online reviews has a positive relationship with consumers' choice of a local product.

However, this relationship between online reviews and product sales may not be true for horizontally differentiated products due to the following two reasons. First, for horizontally differentiated product markets, such as the choice of a restaurant for dining, consumers usually have different cost structures for different restaurants in terms of search cost, transportation cost, and opportunity cost. For example, consumers generally have a lower overall cost to patronize a local restaurant compared to a non-local one. Thus, consumers may choose a local restaurant even when the local restaurant is lower rated than a non-local restaurant in online reviews. Second, consumers usually have more information of local products than non-local products and it is easy for them to get information for local products. For example, due to the lower cost of patronizing a local restaurant, consumers can learn of the true quality of local restaurants through their own dining experience. In addition, they can easily get information of a local restaurant from offline word of mouth sources such as their family members and friends. Therefore, in a horizontally differentiated product market, it seems that there is less of a need for consumers to seek local products' quality information from online reviews. We thus give the following competing hypothesis:

In a horizontally differentiated product market,

H2a: the valence of online reviews has no significant explanatory power for consumers' choice of a local product.

H2b: the volume of online reviews has no significant explanatory power for consumers' choice of a local product.

3.2 The Moderating Effects of Geographical Location

As we argued before, in horizontally differentiated products market, from the consumers' point of view, non-local products differ from local products in two dimensions. First, comparing with local products, consumers incur higher overall cost and lower potential option value when patronizing non-local products. Second, consumers usually have more information and thus less quality uncertainty of local products compared to those of non-local ones. Consumers maximize their utility by choosing a high quality product and lowering their overall cost in their choice decisions.

Consumers face more quality uncertainty when choosing a non-local product. Therefore, they are more likely to seek for signals of product quality compared to the case when choosing a local product [14]. At the same time, because consumers incur more cost (search cost, transportation cost, and opportunity cost), they will pay more emphasis on the reliability of product quality signals in order to reduce risk. If the quality is not high enough or the quality signal is not convincing enough, there may not be enough incentives for them to choose a non-local product [13]. Because the average rating signals restaurant quality directly, we posit that consumers are more likely to be influenced by the valence of online review when choosing a non-local product.

H3: There is an interaction effect of the valence of a product's online reviews and the location of a horizontally differentiated product, such that the effect of the valence of a product's online reviews on consumers' choice probability is stronger for non-local products compared to local products.

Volume of online reviews generally indicates the popularity of a product. Although popularity signals quality to some extent [16], the volume of online reviews is often not a reliable quality signal. Dellarocas and Narayan [17] examined what motivates consumers to post reviews for different kinds of movies. They found that most consumers rate movies very high or very low, resulting in a bimodal, U-shaped histogram. Andersion [18] developed a utility-based model for the relationship between consumer satisfaction and their engagement in word of mouth activities, and suggest a U-curve distribution function of consumer opinions. These studies suggest that high volume of online reviews indicates the existence of a large number of consumers who extremely like or extremely dislike this product. In other words, volume of online reviews is not convincing enough as a reliable quality signal. Therefore, consumers are less likely to be influenced by the volume of online reviews when choosing a non-local product than when choosing a local product.

H4: There is an interaction effect of the volume of a product's online reviews and the location of a horizontally differentiated product, such that the effect of the volume of a product's online reviews on consumers' choice probability is weaker for non-local products compared to local products.

4 Data

We test our hypotheses in the context of consumers' restaurant choice for two reasons. First, restaurant dining is a kind of experience product (or service) and consumers usually do not have full information about the restaurant's quality before the first patronage. Second, the restaurant dining industry is horizontally differentiated and consumers have different overall transaction costs and levels of information for different alternatives. Our original data sets consist of consumers' actual dining records, online reviews for restaurants, and restaurant attributes such as location and price promotion events. Consumers' transaction records and restaurant review information spans from May 2005 to March 2008. Restaurant promotion information spans from January 2006 to March 2008. Restaurant review data is from a popular review website in China. Review information includes consumers' ratings in terms of taste, ambience, and service. Rating scale ranges from 1 to 5, 1 being very bad and 5 being very good. We use the weighted average of these three rating in our analysis. The website distributes loyalty member cards to their registered customers. When patronizing a restaurant with a joint membership program with the review website, consumers can get a discount by using their member cards. Therefore, data of consumers' actual dining records were collected from customers who were members of the loyalty program.

We use the following criteria to generate our final data set. First, based on the original transaction data, we generate consumers' "home" location as the district in which consumers have the largest number of transactions. From consumers' point of view, we differentiate restaurants into two categories: local restaurants which are in the same location as the consumer's "home" location, and non-local restaurants which are in a different location with the consumer's "home" location. We use this categorization as a qualitative indicator to model the products differentiation in horizontal level. Second, in order to control the influence of consumers' cuisine type preference on their restaurant choice, we limited our econometric analysis to restaurants selling the same cuisine type. Third, we choose restaurants which are geographically adjacent to one another so that all these restaurants are likely to fall under a consumer's concurrent alternatives in a consideration set. We choose two adjacent districts rather than one because we want to test the moderating effect of consumers learning experiences across different locations. We finally identify 9 restaurants as alternatives in consumers' choice set and have 3093 transactions from 1827 consumers. Consumers patronized local restaurants in 2045 transactions and visited non-local restaurants in 1048 transactions.

Table 1. Descriptive Statistics

Variables	Mean	Std Dev	Min	Max
Volume (,000)	0.46	0.31	0.001	1.09
Valence	2.78	0.14	2.45	3.05
Promotion	0.12	0.33	0	1
Non-local	0.61	0.49	0	1
Loyalty	0.11	0.15	0	1
N	3093			

5 Model Specification

Our empirical model is specified in the logit framework. We assume consumer *i*'s utility from restaurant *j* on purchase occasion *t* is a linear function of restaurant *j*'s online review rating and price. Then,

$$\begin{aligned} U_{ijt} &= \beta_1 Vol_{ijt} + \beta_2 Val_{ijt} + \beta_3 Pro_{ijt} + \beta_4 NL_{ij} + \beta_5 Loyalty_{ijt} \\ &\quad + \gamma_1 NL_{ij} {*} Vol_{ijt} + \gamma_2 NL_{ij} {*} Val_{ijt} + \gamma_3 NL_{ij} {*} Pro_{ijt} + \varepsilon_{ijt} \\ &= v_{ijt} + \varepsilon_{ijt} \end{aligned} \tag{1}$$

Vol_{ijt} refers to the volume of online reviews of restaurant *j* before consumer *i*'s purchase occasion *t*; Val_{ijt} refers to the valence of online reviews of restaurant *j* before consumer *i*'s purchase occasion *t*; Pro_{ijt} indicates whether restaurant *j* has coupon promotion at consumer *i*'s purchase occasion *t*. It equals one when the restaurants is on promotion and zero otherwise; NL_{ij} indicates whether consumer *i* and restaurant *j* are in different geographical locations. It equals one when restaurants are in the different locations with the consumer and zero otherwise; $Loyalty_{ij}$ refers to consumer *i*'s loyalty for restaurant *j* at purchase occasion *t*. It is constructed by following Guadagni and Little [19]. The variable measures consumers' state dependence in their choices. We use this variable to capture consumers' heterogeneity in terms of offline experience learning; ε_{ijt} is the error term which varies with consumer *i*, restaurant *j* and purchase occasion *t*. Table 1 shows the descriptive statistics of these variables.

Consumers' choice set *J* consists of the 9 restaurants which we identified from our data. Given the utility function specified above, consume *i* chooses the restaurant with the highest utility on each purchase occasion. We further assume the stochastic components ε_{ijt} are independent, identically distributed with a double exponential distribution. Then, the conditional probability (P_{ijt}) that individual *i* chooses restaurant *j* at time *t* is given by the multinomial logit model [19].

$$P_{ijt} = \exp(v_{ijt}) \,/\, \textstyle\sum_{\mathrm{j}} \exp(v_{ijt}) \tag{2}$$

6 Estimation Results

Table 2 shows our preliminary estimation results. The coefficient for Loyalty is positive and highly significant, which suggests that consumers are more likely to choose restaurants they patronized before. After controlling consumers' state dependence, both the volume and the valence of online reviews still have positive and significant influence on consumers' restaurant choices. Thus, H1a and H1b are supported (and therefore the competing hypotheses H2a and H2b are rejected). The coefficient for the interaction term of non-local dummy and valence is positive and significant. This result suggests consumers are more likely to be influenced by the valence of online reviews when they patronize a non-local restaurant. H3 is supported. The coefficient for the interaction term of non-local dummy and Volume is negative, which suggests that consumers are less likely to be influenced by the volume of online reviews when patronizing a non-local restaurant than when patronizing a local one. However, this result is statistically insignificant. H4 is not supported by our data.

In addition, the coefficient for promotion is positive and significant, which means consumers are more likely to patronize a restaurant in promotion. But we do not find the moderating effect of non-local dummy on promotion. The coefficient for non-local dummy has negative sign and is statistically significant, which means consumers are less likely to choose a non-local restaurant.

Table 2. Estimation Results

Variables	Coefficients	p-value
Volume	1.58***	0.001
Non-local*Volume	-0.38	0.12
Valence	6.09 ***	< 0.001
Non-local*Valence	1.80 ***	< 0.001
Promotion	0.82 ***	< 0.001
Non-local*Promotion	0.01	0.98
Non-local	-4.58 ***	< 0.001
Loyalty	6.31***	< 0.001
N	3093	

Note: constant for each restaurant is not reported here due to the page limitation.

7 Conclusion

This study contributes to the literature of online reviews by exploring the information role of online review for different consumer segments in horizontally differentiated product market. Our preliminary analysis shows that consumers' restaurant choice is influenced by volume and valence of online reviews after we control for consumers' state dependence (or loyalty). At the same time, we find evidence that consumers from different geographical locations are influenced differently by online reviews.

Our results suggest that online reviews increase the market competition in dining industry. Although the dining industry is horizontally differentiated, restaurants compete not only with local competitors but also non-local restaurants with high volume and valence of online reviews. Meanwhile, the fact that non-local consumers are more likely to be influenced by the valence of online reviews suggests that consumers from different segmentations have diverse information needs. Restaurants managers should adjust their online marketing strategy accordingly when they advertise to different consumer segmentations.

References

1. ComScore: Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior (2007)
2. Chevalier, J.A., Mayzlin, D.: The Effect of Word of Mouth on Sales: Online Book Reviews. Journal of Marketing Research (JMR) 43(3), 345–354 (2006)
3. Dellarocas, C., Zhang, X., Awad, N.F.: Exploring the value of online product reviews in forecasting sales: The case of motion pictures. Journal of Interactive Marketing 21(4), 23–45 (2007)
4. Duan, W., Gu, B., Whinston, A.B.: Do online reviews matter? – An empirical investigation of panel data. Decision Support Systems 45(4), 1007–1016 (2008)
5. Forman, C., Ghose, A., Wiesenfeld, B.: Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. Information Systems Research 19(3), 291–313 (2008)
6. Zhu, F., Zhang, X.: Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. Journal of Marketing 74(2), 133–148 (2010)
7. Bronnenberg, B.J., Mahajan, V.: Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables. Marketing Science 20(3), 284–299 (2001)
8. Hofstede, F.T., Wedel, M., Steenkamp, J.-B.E.M.: Identifying Spatial Segments in International Markets. Marketing Science 21(2), 160–177 (2002)
9. Jank, W., Kannan, P.K.: Understanding Geographical Markets of Online Firms Using Spatial Models of Customer Choice. Marketing Science 24(4), 623–634 (2005)
10. Liu, Y.: Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. Journal of Marketing 70(3), 74–89 (2006)
11. Godes, D., Mayzlin, D.: Using Online Conversations to Study Word-of-Mouth Communication. Marketing Science 23(4), 545–560 (2004)
12. Clemons, E.K., Gao, G.G., Hitt, L.M.: When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry. Journal of Management Information Systems 23(2), 149–171 (2006)
13. Stigler, G.J.: The Economics of Information. The Journal of Political Economy 69(3), 213–225 (1961)
14. Nelson, P.: Information and Consumer Behavior. The Journal of Political Economy 78(2), 311–329 (1970)
15. Ramon, C., Vives, X.: Why Market Shares Matter: An Information-Based Theory. The RAND Journal of Economics 27(2), 221–239 (1996)

16. Hellofs, L.L., Jacobson, R.: Market Share and Customers' Perceptions of Quality: When Can Firms Grow Their Way to Higher Versus Lower Quality? Journal of Marketing 63(1), 16–25 (1999)
17. Dellarocas, C., Narayan, R.: A Statistical Measure of a Population's Propensity to Engage in Post-Purchase Online Word-of-Mouth. Statistical Science 21(2), 277–285 (2006)
18. Anderson, E.W.: Customer Satisfaction and Word of Mouth. Journal of Service Research 1(1), 5–17 (1998)
19. Guadagni, P.M., Little, J.D.C.: A Logit Model of Brand Choice Calibrated on Scanner Data. Marketing Science 2(3), 203–238 (1983)

Comparing the Quality of Customer Service in 3D Virtual Worlds to Web-Based Service

Sulin Ba, Dan Ke, Jan Stallaert, and Zhongju Zhang

2100 Hillside Road, Storrs, CT 06269, USA
{sulin.ba,dan.ke,jan.stallaert,john.zhang}@business.uconn.edu

Abstract. In the Internet era, web-based services have become a convenient alternative to physical customer service interactions. However, lack of face-to-face interaction makes web service communication inefficient. The 3D virtual worlds provide a new platform that offers customer service, where users can communicate "face to face" via their representative avatars. We propose a conceptual model to compare the quality of customer service and users' satisfaction in 3D virtual worlds to that of web-based services. Theories of computer display technology, communication, and psychology are applied to address how a 3D virtual world impacts users' sense of presence, and their perception of customer service quality. We design an experiment in Second Life and set up a mock-up website to collect data in a post-study questionnaire. Structural equation model is adopted as the main methodology to conduct the multiple group analysis.

Keywords: 3D virtual worlds, Online customer service, Service quality, User satisfaction, Multiple group analysis.

1 Background

In the Internet era, the web-based service becomes an alternative of traditional customer center. However, due to lack of direct communication through Internet, web-based service is not competent enough if the consumers were not familiar with the product or service. Thus a new channel of online service is needed to meet customers' requirement.

3D virtual worlds provide a new platform for offering customer service through the Internet, where users can communicate "face to face" by means of their representative avatars as if they were acting in the real world. According to media richness theory (Daft and Lengel 1986), the richer format of media is, the more effective the communication is to complete a task. 3D virtual worlds provide a 3D computer interfaces where users can communicate with others as if they are acting in-world. Thus, a 3D virtual world becomes a richer media and thus leads to higher efficiency of business communication.

Current research on online service has covered the following main types of web services. Researchers have examined the quality of online services in different kinds of websites (Yang et al. 2004, Yoo and Donthu 2001). Thus, companies shall take the

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 234–247, 2012.

opportunity of improving their customer service quality in the 3D virtual worlds. In this paper, we propose a conceptual model of online customer service quality and investigate the advantages and disadvantages of virtual service based on user's evaluation. We aim to compare the service quality and user satisfaction in 3D virtual worlds to web-based service.

2 Theory Foundation and Model Development

Figure 1 is the proposed research model of online service quality. We contribute to the research on the evaluation of users' satisfaction of virtual service by examining customers' evaluation of service quality online.

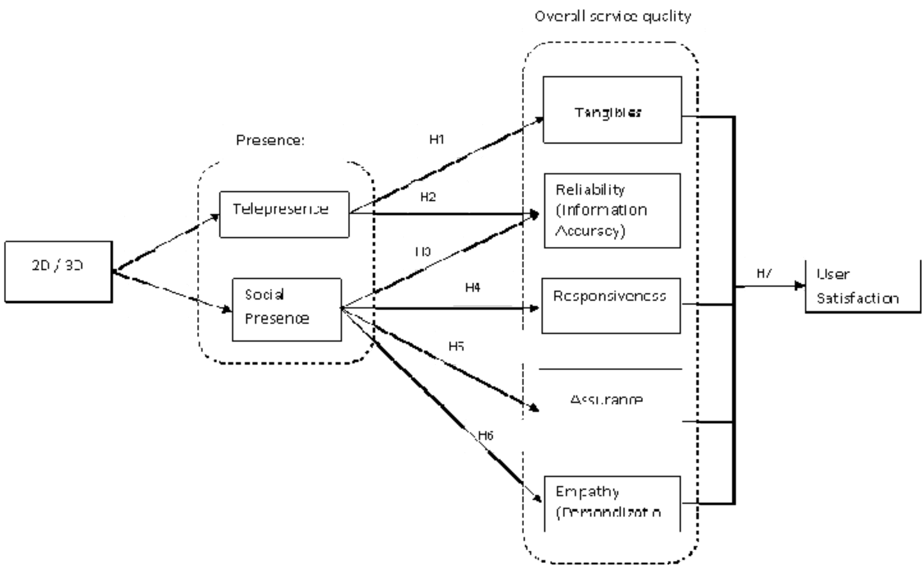


Fig. 1. Research Model of Online Service Quality

2.1 Immersion and Presence

Presence is a psychological consciousness, sense of being in the virtual environment (Slater and Wilbur 1995; Steuer 1992). Immersion in a virtual environment is "a quantifiable aspect of display technology, primarily determined by the extent to which displays are inclusive, extensive, surrounding and vivid" (Slater and Wilbur 1995). Prior studies have contributed to the literature by exploring the relationship between immersion and sense of presence. In Slater's model, they indicate that the higher level of immersion, the higher level of presence (Slater and Wilbur 1995, Slater et al. 1996). According to Slater and Wilbur's (1995) definition of immersion, Users' avatars can "see" or even "touch" the avatars of their partners in the simulated environment, which is not realized with a 2D website (Biocca and Levy 1995).

Presence consists of two interrelated concepts: telepresence (spatial presence or physical presence) and social presence (Heeter 1992, Biocca 1997). Telepresence is defined as the phenomenal sense of “being there” including automatic responses to spatial cues and the mental models of mediated spaces that create the illusion of place. Social presence is defined as the sense of “being together with another” in a virtual environment, including primitive responses to social cues, simulations of “other minds,” and automatically-generated models of the intentionality of others (people, animals, agents, gods, and so on). Based on above statements of the relationship between immersion level and the sense of presence, we find that the higher level of immersion in a 3D virtual world leads to users’ higher sense of telepresence than that in a 2D website. Similarly, the higher level of immersion in a 3D virtual world leads to users’ higher sense of social presence than that in a 2D website.

2.2 Service Quality

Parasuraman et al. (1988) has developed a scale of five core factors named SERVQUAL: reliability, tangibles, responsiveness, assurance, and empathy. Their refined five-core-element scale has become the most widely used scale for measuring service quality. In the Internet era, the importance of online customer service has been recognized by researchers (Janda et al. 2002; Liu and Arnett 2000; Yoo and Donthu 2001; Parasuraman et al. 2005). We propose our scale for measuring service quality in online environment originating from the five core dimensions of SERVQUAL from Parasuraman et al. (1988)’s work. We apply the framework of SERVQUAL and adopt the measurement items for measuring online service quality in prior literature. The reason we follow the 5-dimension framework of the original version of SERVQUAL lies in that virtual worlds are simulation of the real world and the original SERVQUAL was developed directly in the real world setting of customer service.

In SERVQUAL scale (Parasuraman et al. 1988), tangibles originally refer to physical facilities, equipment, and appearance of personnel. Reliability is originally defined as the ability to perform the promised service dependably and accurately. Responsiveness is the willingness to help and provide prompt service. Assurance is the knowledge and courtesy of employees and their ability to inspire trust and confidence. Empathy is the service person’s ability to understand and care individualized attention provided to the customers. Among the five dimensions, tangibles measure the features of customer service based on users’ sensory experiences such as by visual and aural sense; Responsiveness, assurance and empathy measures the quality of human-based features based on users’ psychological response; Reliability measures both the information based on customers’ personal observation and conveyed by the service associate via communication.

2.3 Presence and Service Quality

Prior research has divided the sense of presence into telepresence and social presence. By definition, telepresence describes users’ perception of spatial or physical cues in the environment (Steuer 1995), whereas social presence emphasizes their social cues that come from being with another in a virtual environment (Biocca 2003). In other words, telepresence focuses on users' sensory fidelity in the mediated place and social presence depends on the existence of others and communication with them in the environment. Biocca (1997) discusses the effect of embodiment in virtual environment on the sensation of physical presence, social presence, and self presence.

Accordingly, the dimensions of service quality emphasize on physical features and human interaction respectively. Since tangibles are the physical features, we only explore the impact of telepresence on them. Reliability contains information both from the physical display and via human communication, thus we explore the impact of both telepresence and social presence on it. As for responsiveness, assurance and empathy, these are the users' perceptions based on their interaction with the customer service person. Thus we explore the impact of social presence on them as they are human-related factors and result from users' psychological distance in the virtual environment.

Prior literature has shown that the interactivity and vividness are the two determinants of telepresence that realizes virtual reality (Steuer 1992). The sense of telepresence in a virtual world implies the property of representational richness of its virtual display and interaction with the environment, users of virtual service in such an environment will find the virtual product vivid and their representative avatar easy to navigate around. Furthermore, the sense of telepresence enhances users' psychological state of being involved in an activity (Novak et al. 2000). Thus sense of telepresence enables the users to play with the virtual environments (including the background landscape, virtual objects, etc.) and get absorbed in the virtual world. Thus, we have our proposed hypothesis

Hypothesis 1: A user's sense of telepresence has a positive effect on his/her perception of visual aesthetics and playfulness of the customer service in an online environment.

The reliability of the service requires clear and concise information about a product or service. The sense of telepresence increases users' evaluation on the accuracy of the information from the point of view that the virtual environment delivers clearer information of the product and service visually to the users. Having the sense of telepresence, the user is able to concentrate on the current task and clearly get concrete and complete information of real-world product and service, which in turn improves his decision quality (Novak et al. 2000). Therefore, we propose the following hypothesis by operationalizing service reliability as information accuracy.

Hypothesis 2: A user's sense of telepresence in has a positive effect on his/her perception of the reliability (information accuracy) of the customer service in an online environment.

Short et al. (1976) found the social presence theory about the social effects of communication technology. It argues that the social impact of a communication medium depends on the social presence it allows communicators to have. Social presence is a communicator's sense of awareness of the presence of an interaction partner, and the degree of social presence is equated to the degree of awareness of the other person during the course of a communication interaction (Sallnas 2000). In terms of a particular service task performed in virtual worlds, the sense of social presence in the environment let the user feel about the server's responsiveness, courtesy, considerateness, and such constructs of people's communication properties and key elements to evaluate the service itself. Thus:

Hypothesis 3: A user's sense of social presence has a positive effect on his/her perception of the reliability (information accuracy) of the customer service in an online environment.

Media richness theory (Daft and Lengel 1986) shares some common views with social presence theory. It addresses that the amount of information delivered through communication differs with respect to a medium's richness. This theory assumes that the main goals of communication are to resolve ambiguity and reduce uncertainty, and the more restricted the medium's capacity, the less uncertainty and equivocality it is able to manage. Since it argues that the richer the media the more efficiency the communication is, media richness theory supports that communication in more social media is more able to resolve ambiguity to convey accurate information, and thus more efficient and competitive. Thus we propose the hypothesize that:

Hypothesis 4: A user's sense of social presence has a positive effect on his/her perception of the responsiveness of the customer service in an online environment.

The construal level theory supports that the closer psychological distance positively affects communicators' construal process (Trope et al. 2007). Studies have also demonstrated that mood affects the construal that individuals adopt. In addition, social presence is predicted to increase trust via online communication (Gefen and Straub 2004), which is one of the key constructs of assurance of the service. Hence, based on the ground of social presence theory, media richness theory, and construal level theory, the sense of social presence in a virtual world is associated with assurance.

Hypothesis 5: A user's sense of social presence has a positive effect on his/her perceptions of assurance of the customer service in an online environment.

The social identity model of deindividuation effects (Postmes et al. 1998; Reicher et al. 1995; Spears and Lea, 1994) was developed as a response to the idea that anonymity and reduced presence (or "deindividuated") made communication technology socially impoverished. Instead, it provides an alternative explanation for these "deindividuation effects" based on theories of social identity (Turner et al., 1987). In SIDE model, cognitive effects occur when communication technologies make "salient" particular aspects of personal or social identity. According to the implications of SIDE model, we hypothesize that:

Hypothesis 6: A user's sense of social presence has a positive effect on his/her perception of the empathy (personalization) of the customer service in an online environment.

2.4 Overall Service Quality and User Satisfaction

Prior research has exploring the relationship between service quality and users' satisfaction, behavioral intention and profitability (Gupta and Zeithaml 2006, Iacobucci et al. 1995, Zeithaml et al. 1996), etc. It is accepted that the perceived overall service quality leads to users' satisfaction (Parasuraman et al. 1988, 2005). Thus we propose that each dimension for measuring the service quality has a significant role in users' satisfaction on service:

Hypothesis 7a: A user's perception of service tangibles has a positive effect on his/her satisfaction on the customer service offered in online environments.

Hypothesis 7b: A user's perception of service reliability has a positive effect on his/her satisfaction on the customer service offered in online environments.

Hypothesis 7c: A user's perception of service responsiveness has a positive effect on his/her overall evaluation of service quality in an online environment.

Hypothesis 7d: A user's perception of service assurance has a positive effect on his/her satisfaction on the customer service offered in online environments.

Hypothesis 7e: A user's perception of service empathy has a positive effect on his/her satisfaction on the customer service offered in online environment.

3 Research Methodology

An online experiment was conducted to test the hypotheses. We invited university students and university staff as our subjects in the experiment because they are proficient Internet users and mostly have online purchasing experience. Three extra points of final grade was the incentive to the students who are participating in the experiment. These subjects used would not affect the validity of the findings (McKnight et al. 2002) in that: (a) University students and IT professionals are most proficient Internet users, and conduct more online business; (b) Students and IT professional are younger and better educated than conventional consumers, which closely resemble the online customer population; and (c) Using a homogeneous population can decrease the effect of variance when not exposed to all factors of real world environment.

3.1 Two Experiment Scenarios

We design two customer service scenarios in 2D and 3D online environments respectively. In the 3D version, we choose Second Life as the platform for establishing a 3D virtual customer center. We set up a virtual customer center in Second Life. We then created a virtual cruise in the water. The design of cruise and its surroundings is based on the image of real-world cruises. There is a customer service associate within the subject's field of view, to whom the subject can talk instantly with the local chatting function in Second Life.

In the 2D version, we designed a mock-up website of cruise service, with pictures, texts, and videos. We embed a pop-up chat window to the website that allows the subject to enquire any information about the cruise service from the active customer service associate.

The subjects will participate in the experiment individually. To mediate the practice effects and carryover effects, we use the counterbalancing conditions by letting some of the subjects experience 2D version first and then 3D version, and others in opposite sequence. Thus, we can capture each subject's perception of service quality and their satisfaction for 2D and 3D versions. They experience the 2D and 3D scenarios separately in a random sequence. Before they start each scenario, they took a few minutes for warm-up until they felt familiar with the functions and ready for the service task.

In the experiment, they are assigned to accomplish a service task of figuring out the information of the cruise tour as complete as they can. The information includes the cruise route, service provided, room distribution, price, based on which they could make a decision to reserve a cruise trip in the real world. After finishing their experience in each scenario, the subjects are asked to fill in an administered survey about customer service quality in online environment.

3.2 Scale Development

The variables in our conceptual model – telepresence, social presence, tangibles, reliability, responsiveness, assurance, empathy, overall service quality and satisfaction – are captured in the post-study questionnaire. All items are assessed on a seven-point Likert scale with 1 being strongly disagree and 7 being strongly agree.

Telepresence is original measured as the participant's perception of the immersion level of the virtual environment (Steuer's 1992). We adopt the scale developed by Qiu and Benbasat (2005). Social presence is measured by the subject's perception of how personable, sociable, sensitive, and warm a computer mediate environment is (Short et al. 1976). The dimensions for measuring service quality are justified in the previous section. We measure visual aesthetics using Yoo and Donthu's (2001) scale with arranged wording to be assessable in both the 2D and 3D versions. Liu and Arnett's (2000) scale for assessing playfulness is applied in our paper. We combine Janda et al. (2002)'s measurement of information and Parasuraman et al. (2005)'s measurement of reliability to assess information accuracy in our model. We use the original measurement of SERVQUAL (Parasuraman et al. 1998) to adjust it to assess responsiveness in the online environment in both versions. Since we particularly want to compare the human-related factors in the two scenarios, we refer to some measurement of competence and curtesy (e.g., Cai and Jun 2003; Yang et al. 2004, Parasuraman et al. 1998) to assess assurance in our paper. We take the original measurement of SERVQUAL (Parasuraman et al. 1998) to assess and empathy in the model. We use Janda (2002)'s measurement and reword it to assess user satisfaction in both scenarios.

4 Data Analysis

In this paper, structural equation modeling (SEM) was adopted as the main data analysis method. We applied Amos 18.0 to perform the multiple group confirmatory factor analysis and the path analysis based on the measurement model.

The participants in our experiments involve 189 university students, who experienced the 2D website based online service and 3D virtual world customer center subsequently according to our designed participation procedure. Thus we have collected 378 questionnaires altogether for data analysis. The participants' age ranges from 18 to 40. Among them, 118 are male students and 71 female, 80 are undergraduate students and 109 graduates, and 32 had experiences with user-generated virtual worlds such as Second Life.

Q-Q plots were drawn first to check the normality of measurement items. US5 (user satisfaction measurement item) was observed with heteroscedastic residuals; this problem was solved when this negative item was recoded into a positive item (Drasgo and Hulin 1990).

4.1 Factor Extraction with Exploratory Factor Analysis

We applied each statistical analysis to two groups (each 189 records) including the structural consistency test, confirmatory factor analysis and the structural equation modeling. Our sample size meets the minimum of 100 subjects (MacCallum et al. 1999) and the subjects-to-variables (STV) ratio reaches 5:1 in each group , which is widely accepted as a rule of the sample size for conducting factor analysis (e.g., Bryant and Yarnold 1995; Garson 2008).

We first sent our data set for exploratory factor analysis to extract the dimensions for measuring service quality in online scenarios. We applied the principal axis factoring method using the promax (oblique) rotation method with Kaiser Normalization (Cureton and Mulaik 1975). Since the scale development of service quality has been maturely examined in prior literature and we proposed our scale starting from the five core-dimension SERVQAUL (Parasuraman 1998), we extracted five factors when applying the principal axis factoring method to get the dimensions for measuring service quality online.

It output five factors whose initial eigenvalues were over 1, and the KMO and Barlett's test showed the model was significant and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy was .933 for our data, which was high above the threshold .80 (Kaiser 1970). However, there are several problematic items found according to the rotated factor pattern matrix. We deleted above problematic items and ran the exploratory factor analysis again, and got very clean factor loadings for the five dimension scale.

We then ran exploratory factor analysis for the five dimensions of service quality and other questionnaire items for measuring presence and user satisfaction in the research model. The factor loadings in the rotated pattern matrix are all above .60 (Fabrigar et al. 1999), which indicate significance of the measurement model.

4.2 Measurement Model Comparison and Validation

To validate the revised measurement model in the exploratory factor analysis, we applied the multiple group confirmatory factor analysis to the data set. With this method, we can test the measurement model in two groups and check whether the same metric could be shared across groups, and subsequently whether the factor variances / covariances are equal across groups. We selected the maximum likelihood method to perform confirmatory factor analysis and output various fit indices.

In AMOS 18.0, we established the baseline measurement model (without any constraints across groups), metric model (assuming factor loadings are equal across groups), factor model (assuming factor variances and covariances are equal across groups if metric model is correct), and the residual model (assuming item residuals are equal across groups), for all the factors. The overall fit and estimates for each of the model were calculated by the maximum likelihood method. Table 1 showed the model fit indices of the four measurement models.

Table 1. Model fit indices of multiple group factor confirmatory analysis

Model	Chi-square	df	p	χ2/df	CFI	GFI	NFI	IFI	RMSEA
Unconstrained	1193.831	592	.000	2.017	.923	.822	.860	.924	.052
Metric model	1214.851	611	.000	1.988	.923	.819	.858	.924	.051
Residual model	1415.177	647	.000	2.187	.902	.797	.834	.903	.056
Factor model	1728.804	674	.000	2.565	.865	.764	.797	.866	.065

In Table 1, the unconstrained model and metric model were acceptable measurement models according to the model fit indices. We compared the model fit to test the invariance of factor loadings, given the unconstrained model is correct. We used the chi-square difference test (Anderson and Gerbing 1988) to determine the significance of the difference between the unconstrained model and the metric model. $\Delta\chi^2 = \chi^2_{\text{metric}} - \chi^2_{\text{unconstrained}} = 1214.851\text{-}1193.831 = 21.02$, and $\Delta df = df_{\text{metric}} - df_{unconstrained} = 19$. The corresponding P-value of $\Delta\chi^2$ and Δdf is .336, which means the difference between the metric model and the baseline model is not significant.

The normed Chi-square ($=\chi^2/df$) 1.988, less than the recommended threshold 2, indicated a good model fit with equal factor loadings across groups. Moreover, the standardized factor loadings of the metric invariance measurement model all higher than .6 (Bagozzi and Yi 1988) at $p<.001$ in both 2D and 3D groups reported that it was well fitted.

Similarly, we used the chi-square difference test to compare the model fit of factor model to the metric model. The significant p-value of this comparison meant that the metric model fitted better than the factor invariance assumption. Neither was the residual invariance model acceptable. Further, we did nested comparisons of model by testing each factor variance or each factor covariance individually. With the LaGrange multiplier for constraints in each model to be compared, we found most of the tests were significant except for testing the variance of telepresence, variance of social presence, and their covariance.

Necessarily, the reliability and validity analyses were performed to validate the metric invariance measurement model for each of the group separately. For each group measurements extracted from the metric invariance model, we tested the internal consistency for all the constructs according to the Cronbach's alpha and calculated the composite reliability (CR) for reliability assessment. The values of Cronbach's alpha and CR suggested reasonable construct reliability. Bagozzi and Yi (1988) recommended .60 or above as the acceptable range for the composite reliability. Our results in the two groups met these criteria.

Further, we examined the average variance extracted (AVE) based on the factor loadings of the indicators associated with each constructs in the confirmatory factor analysis. AVE is a common adopted index to demonstrate the convergent validity of a measurement model in factor analysis. The AVE values of our measurement model were well above the threshold of .5. Table 2 is the results of model estimates and validation.

Combined the model comparison results and the significance of the statistics for measurement model validation in the two groups, we used the metric invariance model established in the multiple group confirmatory factor analysis and constraint the variances of telepresence and social presence as well as their covariance across two groups, for the multiple group analysis.

Table 2. Measurement Model Estimates and Validation for 2D Group

Factors	2D webistes			3D virtual worlds		
	CR	AVE	Cronbach's α	CR	AVE	Cronbach's α
Tangibles	0.66	0.60	0.86	0.66	0.60	0.86
Assurance	0.73	0.69	0.90	0.73	0.69	0.90
Responsivness	0.78	0.75	0.89	0.78	0.75	0.89
Ampathy	0.62	0.55	0.77	0.62	0.55	0.77
Reliability	0.77	0.74	0.89	0.77	0.74	0.89
User Satisfaction	0.75	0.72	0.80	0.75	0.72	0.80
Tele Presence	0.85	0.84	0.94	0.85	0.84	0.94
Social Presence	0.78	0.76	0.91	0.78	0.76	0.91

4.3 Multiple Group Analysis for the Structural Model

Maximum likelihood estimator was still adopted as the discrepancy function in the structural equation modeling. We first ran the structural model without constraining that the path coefficients are equal across groups. Three hypotheses were not supported in either of the group. Thus we deleted the three paths (social presence → responsiveness; social presence → empathy; empathy → user satisfaction). Empathy was not related to any factor in the structural model then. However, as one of the key dimensions of overall service quality, empathy, as well as responsiveness, has strong causal relationship with assurance, according to the model modification indices. Thus we added theses two paths to test their impacts on assurance in the model. Similarly, there indicates a relationship between tangibles and reliability, so we added another path (tangibles → reliability).

We constrained the coefficient invariance for each path individually for this step to compare with the unconstrained revised structural model. The same method as we did for multiple group confirmatory factor analysis was applied to compare the significance of chi-square difference. However, for every single test, we got a significant p-value, indicating that the unconstrained model fitted better than with any constraint added to the model. This means the coefficients for each path differs across groups.

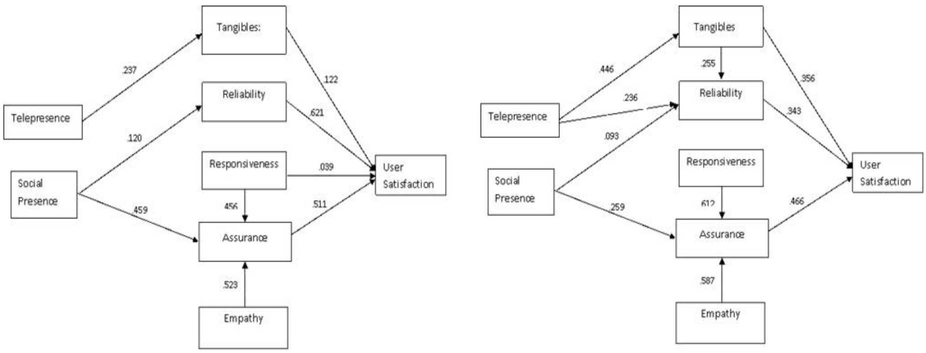


Fig. 2. Structural Model for 2D and 3D Group respectively (at .05 significance level)

Figure 2 respectively illustrated the path coefficients and significance for 2D and 3D groups. The overall fit of the multiple group model indices showed acceptable indicants, with the chi-square=1255.637, degree of freedom=630, thus the normed chi-square = 1.993 indicated good fit of the structure model. Other indices such as CFI=.928 and RMSEA=.051 also indicated a satisfactory structure model.

5 Discussion and Conclusion

Our research model emphasized the effect of environmental features on online users' sense of presence. We focus on the service quality and user satisfaction in the 3D virtual worlds and compare these factors to those in traditional 2D website based service. We aim to demonstrate it is a wise choice to conduct customer service in 3D virtual world, which would be a useful marketing tool to improve customer service quality. We justified our research model and interpreted the underlying factors that are associated with user satisfaction. We demonstrated the advantages of customer service in 3D virtual worlds, that is, the communication efficacy and vivid virtual design in the user-generated environment.

The multiple group analysis indicates that the advantages of customer service offered in 3D virtual world mainly lie in two parts: the vivid presentation of virtual objects as their real world counterpart, and human-related factor (including responsiveness, communication efficacy, and personalized service) due to the increased social presence in the immersive environment. The virtual environment could be designed as exactly same as the real world environment with detailed distribution in the location based world. According to the media richness theory, communication efficiency is increased in the 3D virtual world as if avatars interact face to face. The voice chat setting in Second Life ensures the high quality customer service if the users want detailed explanation about the product of service.

The revised model indicated that one of the dimensions of service quality, empathy, is not directly associated with the user satisfaction in either 2D or 3D scenarios. However, there is a causal relationship between empathy and users' assurance in the online environment. This means, the customer service associate's personalized service in the online platform positively help the users establish online trust with other users or the customer service associate (Ramnath et al. 2005). Therefore, the underlying relationship between empathy and assurance as a new path brings better model fit than our proposed research model. Responsiveness is only associated with user satisfaction in 2D website. However, there is also a causal relationship between responsiveness and empathy. Similarly as empathy, the customer service associate's responsiveness to consumers is helpful to 2D websites explorers, whereas the effect is moderated by the direct interaction with the customer service associate in 3D virtual worlds.

When we propose our scale for measuring service quality, we redefined the reliability as information accuracy, which include the information conveyed by the customer service associate as well as from users' perception online objects in the respective environment. Telepresence in our research model was defined to describe users' perception on the environmental design in the contrast to social presence. The causal relationship between telepresence and information accuracy is only significant in the 3D environment, where the presentation of virtual product and environment brings to consumers a clear idea on the product or service provided.

We failed to find support for the hypotheses (H4, H6) that social presence is associated with responsiveness and empathy, because these two dimensions are closely correlated with assurance. Thus only three out of five dimensions of service quality are positively related to user's sense of presence in the online environment. Even though users in 3D virtual worlds have higher sense of presence than 2D website, we cannot draw the conclusion that the overall service quality in 3D virtual world customer center is significantly higher than 2D website service quality.

In addition to test our research model, we test the validity of scale for measuring service quality by running the causable relationship between each of the five dimensions and overall service quality from the questionnaire items. Expectedly we obtained good model fit (CFI=.936), which indicate the SERVQUAL scale (Parasuraman et al. 1988) is a good source for studying scale development of service quality in the online environment especially in 3D virtual worlds. Moreover, overall service quality is significantly related to user satisfaction.

Our results of multiple group structural equation modeling indicate that users of 3D virtual worlds obtained better sense of being in the online environment and using the 3D platform for customer service can increase user satisfaction by providing higher customer service quality in the immersive environment. To be specific, businesses may focus on virtual product and environment design and increase the customer service associates' communication skills to satisfy potential customers. From a practical point of view, our research brings a new idea for businesses to improve their customer service and thus increase the business value.

References

1. Anderson, J.C., Gerbing, D.W.: Structural Equation Modeling in Practice: A Review and Recommended Two-step Approach. Psychological Bulletin 103(3), 411–423 (1988)
2. Bagozzi, R.P., Yi, Y.: On the Evaluation of Structural Equation Models. Journal of the Academy of Marketing Science 16(1), 74–94 (1988)

3. Biocca, F.: The Cyborg's dilemma: Progressive embodiment in virtual environments. Journal of Computer-Mediated Communication 3(2), 12–26 (1997)
4. Biocca, F., Levy, M.R.: Virtual reality as a communication system. In: Biocca, F., Levy, M.R. (eds.) Communication in the Age of Virtual Reality. Lea'S Communication Series, pp. 15–31. L. Erlbaum Associates, Hillsdale (1995)
5. Biocca, F., Harms, D., Burgoon, J.K.: Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. Presence 12(5), 456–480 (2003)
6. Bryant, F.B., Yarnold, P.R.: Principal components analysis and exploratory and confirmatory factor analysis. In: Grimm, L.G., Yarnold, R.R. (eds.) Reading and Understanding Multivariale Statistics, pp. 99–136. American Psychological Association, Washington, DC (1995)
7. Cai, S., Jun, M.: Internet users' perceptions of online service quality: A comparison of online buyers and information searchers. Managing Service Quality 13(6), 504–519 (2003)
8. Cureton, E.E., Mulaik, S.A.: The weighted varimax rotation and the promax rotation. Psychometrika 40(2), 183–195 (1975)
9. Daft, R.L., Lengel, R.H.: Organizational Information Requirements, Media Richness and Structural Design. Management Science 32(5), 554–571 (1986)
10. Drasgow, F., Hulin, C.L.: Item Response Theory. In: Dunnette, M.D., Hough, L.M. (eds.) Handbook of Industrial and Organizational Psychology, 2nd edn., vol. 1, pp. 577–636. Consulting Psychologists Press, Palo (1990)
11. Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., Strahan, E.J.: Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods 4(3), 272–299 (1999)
12. Gefen, D., Straub, D.W.: Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Service. Omega 32(6), 407–424 (2004)
13. Gupta, S., Zeithaml, V.: Customer Metrics and their Impact on Financial Performance. Marketing Science 25(6), 687–717 (2006)
14. Heeter, C.: Being there: The subjective experience of presence. Presence: T + VE 1, 262, 271 (1992)
15. Iacobucci, D., Ostrom, A., Grayson, K.: Distinguishing Service Quality and Customer Satisfaction: The Voice of the Consumer. Journal of Consumer Psychology 4(3), 277–303 (1995)
16. Janda, S., Trocchia, P.J., Gwinner, K.P.: Consumer perceptions of internet retail service quality. International Journal of Service Industry Management 13(5), 412–431 (2002)
17. Kaiser, H.F.: A Second-Generation Little Jiffy. Psychometrika 35, 401–415 (1970)
18. Liu, C., Arnett, K.P.: Exploring the factors associated with Web site success in the context of electronic commerce. Information and Management 38(1), 23–33 (2000)
19. MacCallum, R.C., Widaman, K.F., Zhang, S., Hong, S.: Sample Size in Factor Analysis. Psychological Methods 4, 84–99 (1999)
20. McKnight, D., Choudhury, V., Kacmar, C.: Developing and Validating Trust Measures for E-Commerce: An Integrative Typology. Information Systems Research 13(3), 334–359 (2002)
21. Novak, T.P., Hoffman, D.L., Yung, Y.F.: Measuring the Customer Experience in Online Environments: A Structural Modeling Approach. Marketing Science 19(1), 22–44 (2000)
22. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing 64(1), 12–40 (1988)
23. Parasuraman, A., Zeithaml, V.A., Malhotra, A.: E-S-QUAL: A multiple-item scale for assessing electronic service quality. Journal of Service Research 7(3), 213–233 (2005)

24. Postmes, T., Spears, R., Lea, M.: Breaching or building social boundaries? SIDE-effects of computer-mediated communication. Communication Research 25, 689–715 (1998)
25. Qiu, L., Benbasat, I.: Online Consumer Trust and Live Help Interfaces: The Effects of Text-to-Speech Voice and Three-Dimensional Avatars. International Journal of Human-Computer Interaction 19(1), 75–94 (2005)
26. Chellappa, R.K., Sin, R.G.: Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. Information Technology and Management 6(2-3), 181–202 (2005)
27. Reicher, S., Spears, R., Postmes, T.: A social identity model of deindividuation phenomena. European Review of Social Psychology 6, 161–198 (1995)
28. Sallnas, E.L., Rassmus-Grohn, K., Sjostrom, C.: Supporting presence in collaborative environments by haptic force feedback. ACM Transactions on Computer-Human Interaction 7(4), 461–476 (2000)
29. Short, J., Williams, E., Christie, B.: The social psychology of telecommunications, London (1976)
30. Slater, M., Linakis, V., Usoh, M., Kooper, R.: Immersion, presence, and performance in virtual environments: An experiment in tri-dimensional chess. In: Proceedings of VRST 1996, Hong Kong (1996)
31. Slater, M., Wilbur, S.: Through the looking glass world of presence: A framework for immersive virtual environments. In: Slater, M. (ed.) FIVE 1995: Framework for Immersive Virtual Environments. QMW University of London (1995)
32. Spears, R., Lea, M.: Panacea or panopticon? The hidden power in computer-mediated communication. Communication Research 21(4), 427–459 (1994)
33. Steuer, J.: Defining virtual reality: Dimensions determining telepresence. Journal of Communication 42(4), 73–93 (1992)
34. Trope, Y., Liberman, N., Wakslak, C.: Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior. Journal of Consumer Psychology 17(2), 83–95 (2007)
35. Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S., Wetherell, M.S.: Rediscovering the social group: A self-categorization theory. Basil Blackwell, Oxford (1987)
36. Yang, Z., Jun, M., Peterson, R.T.: Measuring customer perceived online service quality. International Journal of Operations and Production Management 24(11), 1149–1174 (2004)
37. Yoo, B., Donthu, N.: Developing a Scale to Measure the Perceived Quality of an Internet Shopping Site (SITEQUAL). Quarterly Journal of Electronic Commerce 2(1), 31–47 (2001)
38. Zeithaml, V.A., Parasuraman, A., Malhotra, A.: Service Quality Delivery through Web Sites: A Critical Review of Extant Knowledge. Journal of the Academy of Marketing Science 30(4), 362–375 (2002)

The Impact of Query Suggestion in E-Commerce Websites

Alice Lee and Michael Chau

School of Business, The University of Hong Kong, Pokfulam Road, Hong Kong
alice226@hku.hk, mchau@business.hku.hk

Abstract. In this paper we propose a research agenda for studying the impact of query suggestion features on cognitive load and customer satisfaction during online shopping in e-commerce websites. Despite the popular use of query suggestion features in search engines and large e-commerce websites such as Amazon.com and eBay, there is little research in this area. Based on a review on prior literature in query suggestion and online shopping, a research model and five hypotheses are posed. A lab experiment is proposed to test the hypotheses and potential implications of the research are discussed.

Keywords: query suggestion, e-commerce websites, cognitive load, customer satisfaction, purchase intention.

1 Introduction

There is a rapid growth in the number of people shopping via the Internet. Online shopping provides opportunities for customers to reach the businesses globally and directly. Accordingly, it is important for the online stores to understand the customers' behaviors and purchase intentions in order to sell their products effectively and satisfy their customers. The perceived ease-of-use of the online shopping websites directly influence customers' attitude towards online purchase, and hence, their intention to purchase online [22]. Therefore, it is critical that the working environment of the website is user-friendly and can help users to obtain desired results easily [10].

Product search function is always an important feature of online store navigation. Customers often reach a product they want to purchase through product search engines [16]. Thus, improving the search engine of online shopping websites is significant in increasing the efficiency of getting the results that can better satisfy consumers' needs. Some e-commerce websites, e.g. Amazon.com and eBay, have already added useful features in their search engines. The main idea of these features is to help the users form suitable queries by providing them with queries made by other users before. There are two ways to display these queries. One of the approaches is called real-time query suggestion, in which the recommended queries are shown interactively when users are entering the words in the query entry box. This is often implemented using AJAX technologies like the one used in Google Suggest

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 248–254, 2012.

[19]. Another approach is retrospective suggestion, where the suggested queries are shown after retrieval has been executed [23]. However, the effectiveness of the use of such query suggestion features in online shopping website is little known.

The objective of this paper is to examine the effectiveness of query suggestion in online shopping websites. We propose a research model and an experiment design to test the impact of query suggestion on customers' cognitive load (e.g., number of reformulations, decision time) and satisfaction when using these systems. Two settings will be used in the experiment – a baseline setting that does not provide any query suggestion and a real-time query suggestion system. Knowing the features of the systems help online stores in developing search engines in order to benefit from the market.

2 Literature Review

It has been pointed out that customer satisfaction occupied a central position in the outcome of marketing activity and it is directly in relation to the post-purchase phenomena such as attitude change, repeat purchase and brand loyalty [8]. The increasing usage of e-commerce websites has drawn great research interests in analyzing the factors affecting user satisfaction. It was found that poor usability of websites leads to poor company image [3]. The service quality of online stores has significant effects on user satisfaction. It has been shown that since customers who have experienced high level of service will visit that e-commerce website again, financial success of the stores will be enhanced by high service quality [12]. In order to make a successful shopping website, online stores should actively seek ways to improve their websites [14]. It was also shown that the factors affecting user satisfaction of websites, including control, efficiency and helpfulness, which indicated the ease for user to navigate through the website, to find the information they needed at a reasonable speed, and to offer help when finding information and navigating [13].

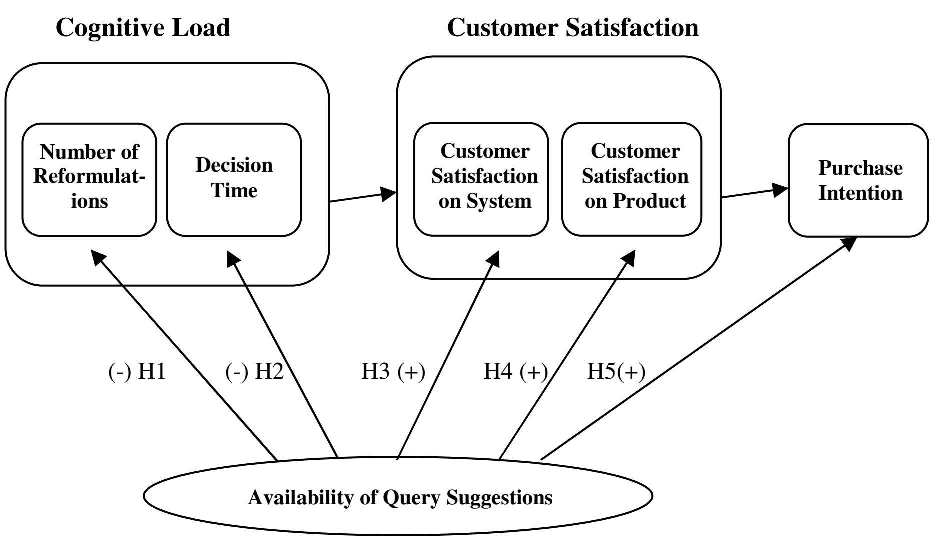
In order to increase customers' satisfaction, some e-commerce websites such as Amazon.com and eBay have implemented query suggestion function in their product search engines. The suggested queries are the queries made by other users before or similar terms available in the databases of those e-commerce websites. Many researches focused on the design of query suggestion method to help users formulate better queries. One possibility is to analyze user queries submitted to search engines. A new query suggestion method has been proposed by using the relationships between query terms and document terms from user logs to aid the query suggestion [9]. A query recommendation technique was also proposed based on reusing previous search histories [2]. This research showed the way in which the queries are scored and ranked using relevance and coverage factors in order to prioritize the results. Semantic nets and ontologies have also been used to support query suggestion and found that the suggested queries provide more relevant results [21].

However, there are only a few studies on the effectiveness of this technique for real searchers. Some researchers have conducted experiments on three search systems: a baseline system which does not offer any query suggestion, a real-time query suggestion which displays the suggested queries when users are entering words in the query entry box, and a retrospective system which displays the suggested queries after the retrieval has been executed [23]. The results show that real-time query suggestion has the best performance in the quality of initial queries, engagement in search and the uptake of query suggestion. However, their study was on general search and may not be applicable to e-commerce applications. We have not been able to identify any study on the effectiveness of query suggestion in e-commerce websites.

3 Research Model and Hypotheses Development

The research model proposed in this study is illustrated in Figure 1. Cognitive load is one of the variable assessing the efficiency and effectiveness of the three systems. Cognitive load can be considered to be a multidimensional construct that represents the load that performing a particular task imposes on the cognitive system of a particular learner [18]. The cognitive load of using the systems will be measured by the number of reformulations [11] and decision time. It is also found that when the cognitive load decreases, user satisfaction of searching on that system increases [20]. Moreover, customer satisfaction leads to higher repurchase intention from the same online stores [1].

Hypothesis 1: When query suggestions are provided, the number of reformulations will be reduced.



(+) denotes positive relation; (-) denotes negative relation between variables.

Fig. 1. The research model

One of the reasons why query reformulation occurs is that users may have a quite specific information need in mind but is uncertain how to express that need in query language [4]. Without the recommendation, users will doubt the correctness and effectiveness of the queries they entered. They cannot obtain assistance when they have difficulties expressing their intention of search until they read through the result displaying pages, and have limited adjustment ability to reformulate their search queries. With assistance provided, the cognitive load of searchers can be significantly reduced. When they are not sure of the spelling of words, the query suggestion function can generate help. When they type a very general term, the function can offer suggestions to make the queries more specific. When they are facing difficulties in presenting their queries, the function can provide suggestions based on the overall popularity of search strings by other users, which can be extracted from search logs [6]. Hence the suggestions displayed by the system can enable customers to express the products they want to find more effectively. The assistance by the query suggestion function will help customers figure out the best query and arrive at the products they intend to search in a minimal number of reformulations.

Hypothesis 2: When suggestions are displayed, the average time to complete the search on products is shorter than that without suggestions.

The product queries suggested can enable customers to express more keywords than they would otherwise have thought of so as to arrive at a solution more quickly. Customers can adjust their queries from the queries made by other users before and form a more effective query by relying on the suggestions from the query suggestion function. They do not have to look through the result displaying pages. Without the assistance of query suggestions, customers have to find a more effective query on their own by reading the result displaying pages and evaluating the products, which takes more time in adjusting the queries to find a suitable product than that when recommendations of queries are displayed.

Hypothesis 3: Customers will have higher satisfaction with the online stores when query suggestions are provided.

Since the query suggestion function can offer help for customers to express their intention of search, the products they found can satisfy their needs better. The customers will appreciate more on the website, and hence, improve the images of the company. The customers will also be more satisfied with the usability features and technical sophistication of the online website. Therefore, the level of satisfaction of customers will be increased.

Hypothesis 4: Customers will have higher satisfaction with the product when query suggestions are provided.

If the query suggestion function is available, when customers are entering their queries in the search box, they can see the suggestions of similar items. They may try to explore these items due to curiosity. The more they explore, the more products they will discover. Since these products may be similar or are accessories to the item that the customers intend to find originally, they may be interested in these products after exploring them. Hence the customers will be more satisfied with the products that they find.

Hypothesis 5: Customers will have higher intention to purchase products when query suggestion function exists.

In e-commerce websites, information quality is considered to be one of the key design factors contributing to the success of online shopping websites [15]. If query suggestions are provided, customers would be able to express the products they intend to search in more specific ways. The availability of query suggestion function can increase the level of user-friendliness of the webpage. It allows customers to more easily determine the products they want to search and purchase. Therefore, customers' purchase intention will increase when query suggestion function is available.

4 Discussion and Future Work

To test the hypotheses, we plan to conduct a lab experiment on using the real-time query suggestion function in an e-commerce website such as Amazon.com. Lab experiment is a widely used method for assessing web search functionality [7]. About 200 subjects will be recruited to participate in this study. The experiment will be conducted using two settings, one with query suggestion and one without. Subjects will be required to perform two shopping tasks on the e-commerce website under each setting. The two tasks include one simple search and one complex search, which will help us reduce the effect of task complexity [5]. For simple search, subjects will be required to find some details of a particular product, such as the screen size and dimensions of a specific camera model. For complex search, subjects have to make purchasing decisions based on the situation provided. An example of the complex search task will be as follows:

> "Next Friday will be your friend's birthday. You want to give him a birthday gift. He is very interested in reading, and you would like to give him a book as present. You have limited budget of $100. As you can choose only one gift in this case, you have to make your decision very carefully."

Subjects will do this experiment independently. Upon the arrival of subjects, they will be given the instructions and overviews of the tasks they need to accomplish in the experiment and will complete a demographics questionnaire focusing on their search and online shopping experience. Before each task, the descriptions of the shopping items will be provided to the subjects. The order of the systems used and the search tasks will be randomized. Number of reformulations and decision time will be automatically recorded. After using each system, subjects will have to answer the questions regarding their cognitive load and satisfactions on the systems used and products searched, and their purchase intention. Other control variables will be measured in the questionnaire as well. They can also give any other open opinions on this experiment.

We expect this research to have several implications. From an academic perspective, it will help us better understand the role of query suggestion functions in the online shopping process in e-commerce websites. From a practical point of view, the study is expected to provide some findings on how customers utilize and react to the query suggestion function and help decide whether query suggestion should be used in these sites. We also plan to study the impact of incorporating advertisements, such as sponsored links [17], in the query suggestion function.

References

1. Anderson, E.W., Sullivan, M.W.: The Antecedents and Consequences of Customer Satisfaction for Firms. Marketing Science 12(2), 125–143 (1993)
2. Balfe, E., Smyth, B.: Improving Web Search Through Collaborative Query Recommendation. In: Proceedings of the 16th European Conference on Artificial Intelligence, pp. 268–272 (2004)
3. Bouch, A., Kuchinsky, A., Bhatti, N.: Quality is in the Eye of the Beholder: Meeting Users' Requirements for Internet Quality of Service. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, vol. 2(1), pp. 297–301 (2000)
4. Bruza, P.D., Dennis, S.: Query Reformulation on the Internet: Empirical Data and the Hyperindex Search Engine. RIAO 9, 488–499 (1997)
5. Chau, M.: Visualizing Web Search Results Using Glyphs: Design and Evaluation of a Flower Metaphor. ACM Transactions on Management Information Systems 2(1), 1–27 (2011)
6. Chau, M., Fang, X., Yang, C.C.: Web Searching in Chinese: A Study of a Search Engine in Hong Kong. Journal of the American Society for Information Science and Technology 58(7), 1044–1054 (2007a)
7. Chau, M., Shiu, B., Chan, I., Chen, H.: Redips: Backlink Search and Analysis on the Web for Business Intelligence Analysis. Journal of the American Society for Information Science and Technology 58(3), 351–365 (2007b)
8. Churchill, G.A.: A Paradigm for Developing Better Measures of Marketing Constructs. Journal of Marketing Research, 64–73 (1979)
9. Cui, H., Wen, J.R., Nie, J.Y., Ma, W.Y.: Probabilistic Query Expansion Using Query Logs. In: Proceedings of the 11th International Conference on World Wide Web, pp. 325–332 (2002)
10. Despotopoulos, I., Korinthios, G., Nasios, I., Reisis, D.: Developing an efficient model for evaluating WWW search engines. In: Proceedings of the Seventeenth IASTED International Conference, pp. 87–89 (1999)
11. Gwizdka, J.: Distribution of cognitive load in web search. Journal of the American Society for Information Science and Technology 61(11), 2167–2187 (2010)
12. Kim, J., Lee, J.: Critical design factors for successful e-commerce systems. Behaviour & Information Technology 21, 185–199 (2002)
13. Kirakowski, J., Claridge, N., Whitehand, R.: Human Centred Measures of Success in Web Site Design. In: Proceedings of the 4th Conference on Human Factors & the Web (1998), `http://www.research.att.com/conf/hfweb/proceedings/Kirakowski/index.html`
14. Liu, C., Arnett, K.L.: Exploring the factors associated with web site success in the context of electronic commerce. Information and Management 38, 23–33 (2000)
15. Liu, C., Arnett, K.P., Litecky, C.: Design quality of websites for electronic commerce: fortune 1000 webmasters' evaluations. Electronic Markets 10(2), 120–129 (2000)
16. Lohse, G.L., Spiller, P.: Electronic shopping. Communications of the ACM 41(7), 81–86 (1998)
17. Lu, Y., Chau, M., Chau, P.: The Impacts of Trust on the Effectiveness of Search Engine Advertising. In: Proceedings of the Seventh Workshop on E-Business (WEB 2008), Paris, France, December 13 (2008)
18. Paas, F., van Merriënboer, J.J.G.: Variability of worked examples and transfer of geometrical problem solving skills: A cognitive-load approach. Journal of Educational Psychology 86, 122–133 (1994)

19. Paulson, L.D.: Building Rich Web Applications with Ajax. IEEE Computer (October 14-17, 2005)
20. Schaefer, A., Jordan, M., Klas, C.-P., Fuhr, N.: Active Support for Query Formulation in Virtual Digital Libraries: A Case Study with DAFFODIL. In: Rauber, A., Christodoulakis, S., Tjoa, A.M. (eds.) ECDL 2005. LNCS, vol. 3652, pp. 414–425. Springer, Heidelberg (2005)
21. Storey, V.C., Burton-Jones, A., Sugumaran, V., Purao, S.: CONQUER: A Methodology for Context-Aware Query Processing on the World Wide Web. Information Systems Research 19(1), 3–25 (2008)
22. van der Heijden, H., Verhagen, T., Creemers, M.: Understanding online purchase intentions: contributions from technology and trust perspectives. European Journal of Information Systems 12, 41–48 (2003)
23. White, R., Marchionini, G.: Examining the effectiveness of real-time query expansion. Information Processing and Management 43(3), 685–704 (2006)

Is Localization Advisable for E-Commerce Websites?

Muller Y.M. Cheung and James Y.L. Thong

Department of Information Systems, Business Statistics and Operations Management,
School of Business and Management, Hong Kong University of Science and Technology
{imcym,jthong}@ust.hk

Abstract. To address the possible impacts of cultural differences, companies tend to set up localized websites for different countries. These localized websites usually feature local content, local language, and local cultural elements. However, in an increasingly globalized economy, individuals around the world are now more exposed to multiple cultures and may even have internalized multiple cultures. Psychology theories suggest that these individuals' perceptions and decision making may vary in response to different cultural cues. As such, localized websites could serve as cultural cues that impact online consumers' building of trust and subsequent decision making. In this study, we focus on one basic consideration of localization (i.e., incorporating cultural elements into websites) and investigate whether the effects of social presence on trust building will be contingent upon it. This study intends to contribute to understanding the effectiveness of these localization strategies in building online consumers' trust in e-commerce websites.

Keywords: Localization strategies, biculturals, social presence, trusting beliefs.

1 Introduction

With rapid development of information technology and extensive adoption of e-commerce websites, the potential impacts of cultural values have been capturing the attention of both industry practitioners and researchers [3, 21]. Companies acknowledge the importance of cultural differences and set up localized websites for different countries and even for different cities. These localized websites usually feature local content, local language, and local cultural elements. Prior research found that individuals across cultures had distinctive preferences for website design [4]. Operators of e-commerce websites implement localization strategies to address varying preferences across cultures in promoting their websites. While localization strategies are believed to be an effective way to address the needs of local individuals, psychology theories suggest that these individuals' perceptions and decisions may vary in response to different cultural cues. Drawing from psychology theories [14], we will extend prior research on localization strategies of websites by studying how the building of trust in e-commerce websites would be contingent on localization strategies.

In today's globally connected world, many individuals have internalized two cultures and speak the languages associated with each of those cultures. It was estimated that 20% of individuals in the United States have lived in another country before moving to the United States [1]. These potential bicultural individuals

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 255–260, 2012.

(a.k.a. biculturals) are familiar with two different cultures and have internalized both cultures. The two internalized cultures may be associated with two distinct culture-specific cognitive structures or mental frames with their respective values, norms and behaviors [2] and with competing theories, such as collectivist versus individualist orientations. Hong et al. (2000) suggest that the competing culture-specific mental frames cannot simultaneously guide cognition. Instead, in response to contextual and environmental cues, a particular mental frame will be more accessible and guide the cognition at any one time. The process of shifting mental frames is called "cultural frame switching". When the shift of mental frame is activated by cultural cues, biculturals could shift their values and change their decision making strategies [2]. The emerging salience of biculturals raises an important question. Is it always advisable for the operators of e-commerce websites to localize their websites for different regions? Cultural knowledge is "like a contact lens that affects the individuals' perceptions of visual stimuli" [14]. It is possible for the biculturals to respond unexpectedly, in favorable or unfavorable manner, when the localized websites activate the specific mental frames.

Individuals in different regions espouse varying values and norms, and could be dimensionalized for systematic comparisons [13]. The varying cultural values of individuals influence their perceptions of web design elements and their judgments. Previous researchers mostly devoted their efforts to investigate the impacts of cultural values and norms at the country level. For example, Cyr (2008) found that the visual design of localized web sites resulted in website trust for Chinese individuals but not for Canadian and German individuals. In another study, cultural differences were found to influence the effectiveness of trust building strategies in the context of Internet shopping [21].

Even though several studies have suggested that individuals could espouse national cultural values to different degrees and exhibit differences in their perceptions of information technology [22], relatively few studies have investigated the cultural impacts at the individual level. This study intends to complement prior studies by investigating how individuals in the same country are affected by their internalized cultural values. It investigates the interplays between localization strategies and biculturals at the individual level. Specifically, we focus on the basic consideration of localization, i.e., incorporating cultural elements into websites. Our research objective is to investigate how the building of trust in e-commerce websites would be contingent on the localization of websites.

2 Theoretical Background and Hypotheses

2.1 Social Presence and Trusting Beliefs

Social presence can be defined as the extent to which a medium allows users to experience others as being psychologically present [8]. Social presence theory views social presence as a quality inherent in a communication medium and the social presence of a medium could be characterized as its capacity to transmit information about non-verbal cues [20]. From a psychological perspective, social presence of a medium is concerned with the medium's ability to transmit the feeling of "warmth" and sociability. Being an emerging medium in the last century, a website's social presence is concerned with its ability to transmit the feeling of human touch and sociability [4]. Social presence has been found to be an important determinant of trust in e-commerce websites [10, 11]. Trust in an e-commerce website can be defined as an individual's willingness to rely on the e-commerce websites and willingness to be vulnerable to the actions taken by the e-commerce websites [16]. Gefen and Straub (2004) have conceptualized trust as a "set of beliefs about other people that serve to justify an interaction with these persons, based on the belief that they will behave socially as expected and in doing so will provide the trusting party with its expected outcomes".

2.2 Biculturals and Cultural Frame Switching

Hong et al. (2000) proposed that biculturals who internalized two cultures have access to two cultural meaning systems. Biculturals could apply either one of the cultural meaning systems in response to contextual cues, such as language and national flags, through a process called cultural frame switching (CFS). If a particular set of norms and values has been learnt in a particular context, the factors that are associated with this context is likely to increase the accessibility of these norms and values, and increase the likelihood of applying these norms in appropriate situations. Hong et al. (2000) showed that when Hong Kong biculturals are primed with Chinese cues, they made more external attribution, i.e., attributed the responsibility to situational or contextual factors. In contrast, when primed with American cues, Hong Kong bicultural individuals made more internal attributions and attributed the responsibility to themselves. By manipulating the contextual cues that would activate different cultural meaning for biculturals, Hong et al. (2000) managed to reproduce the patterns that were observed in cross-national comparisons [18].

Being associated with the largest ethnic groups with distinctive cultural values and norms, biculturals who internalized Chinese and Western cultures were chosen by previous researchers to investigate perceptions and behaviors of biculturals. Examples include Chinese Canadian biculturals [19], Chinese American biculturals [1], and Hong Kong Chinese [2]. Consistent with these studies, we chose the biculturals who internalized both Chinese and Western cultures for investigation and used Chinese and Western cultural cues to activate their respective cultural values in this study. Figure 1 depicts our research model.

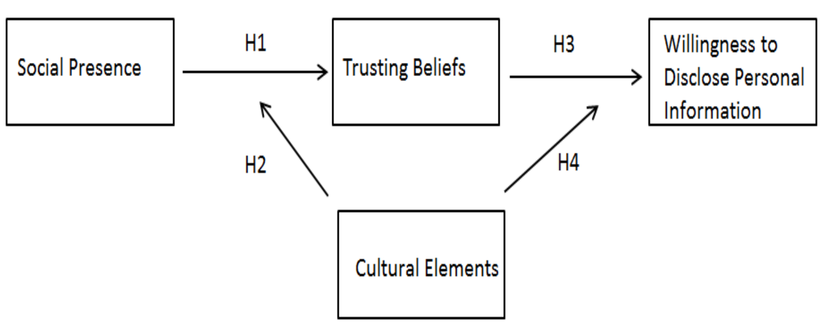


Fig. 1. Research Model

2.3 Hypotheses

Gefen and Straub (2003) suggested that social presence builds trust based on the conveyance of social cues in the online context. Higher levels of social presence help online consumers to build trust in e-commerce websites by making it more difficult for operators of e-commerce websites to hide information and engage in unexpected behavior. With a higher level of social presence, online consumers can gain more confidence that they have greater ability to predict online vendors' behavior and detect untrustworthy behavior. Doney et al. (1998) suggested that trust is built up when the trusting party gains confidence in his/her ability to predict the trusted party's future behavior with accuracy. Compared with individuals in individualist cultures who accept variations in behavior, individuals in collectivist cultures embrace the value of behavioral conformity. Norms and values in collectivist cultures support behavioral conformity and curb variations in individuals' behavior. Also, the costs of deviant behavior are higher in collectivist cultures than in individualistic cultures. Therefore, relative to individuals with individualist cultural values, individuals with collectivist cultural values are more likely to form trust by embracing the value of behavioral conformity among individuals and making it easier for one to predict the trusted party's behavior with accuracy [7].

When biculturals are primed with cues of different cultures, their respective cultural values and norms will be activated. For individuals who internalized both Chinese and Western cultures, the cultural elements incorporated in a particular website could serve as the contextual cues that activate individuals' respective cultural values and norms. For example, the choice of Chinese as the presentation language in web pages activates Chinese culture [2] that is characterized by collectivism [13]. The use of English language activates Western cultures that are characterized by individualism. With Chinese cultural values activated, it is more likely for them to form trusting beliefs toward the website with a high level of social presence. A high level of social presence provides individuals with more social cues and gives individuals more confidence in detecting untrustworthy behavior. As proposed by Doney et al. (1998), relative to individuals with individualist cultural values, individuals with collectivist cultural values are more likely to form trust based on their confidence in predicting the trusted party's behavior. We therefore propose that the positive impact of social presence on trusting beliefs will be more prominent when biculturals' Chinese cultural values are activated by Chinese cultural cues than when their Western cultural values are activated by Western cultural cues.

H1: *Social presence has a positive impact on biculturals' trusting beliefs.*

H2: *The positive impact of social presence on biculturals' trusting beliefs will be stronger when Chinese cultural elements are incorporated into a website than when Western cultural elements are incorporated.*

In the context of e-commerce, disclosing personal information, such as name and contact information, are often required as part of the purchase transaction. This disclosure of personal information may expose online consumers to the risk of privacy invasion. Trusting beliefs have been shown to ease online consumers' negative perceptions and enhance their willingness to disclose personal information [6]. There are two board frames of decision making, i.e., promotion focus and prevention focus [12]. Promotion focus is generally characterized by attention to positive outcomes of one's behavior, while prevention focus is characterized by attention to negative outcomes of one's behavior. Chinese socialization appears to induce a prevention orientation [17]. When biculturals' Chinese cultural values are activated by the Chinese cultural elements presented on the websites, it will induce a prevention orientation and these biculturals will focus on the potential negative outcomes of their decisions. When biculturals are shown with Chinese cultural elements and are induced with a prevention focus, they will have the tendency to avoid loss and choose the loss-minimizing options. Under such circumstances, we hypothesize that trusting beliefs will be less effective in easing biculturals' negative perceptions and in enhancing their willingness to disclose personal information.

H3: *Trusting beliefs are positively related to biculturals' willingness to disclose personal information.*

H4: *The positive impact of trusting beliefs on biculturals' willingness to disclose personal information will be weaker when Chinese cultural elements are incorporated into a website than when Western cultural elements are incorporated.*

3 Future Plans and Expected Contributions

An online experiment will be conducted to test the hypotheses. We will use a between-subject design with two between-subject factors: social presence (high vs. low) and cultural elements (Chinese vs. Western). Monetary incentive will be provided to recruit college students as participants for this study. Participants will be randomly assigned to an experiment condition.

This study intends to make several contributions. First, by drawing from psychology theories we develop hypotheses to test the effects of localization strategies of e-commerce websites on trust building among the biculturals. These biculturals have internalized two cultures and distinct cultural values and norms could be activated by localized web design features. Through this research, we will contribute to understanding the effectiveness of these localization strategies on building online consumers' trust in an e-commerce website. Second, this study investigates how the impact of biculturals' trusting beliefs on decision making would be contingent on the localization of websites. Cultural elements could induce individuals' tendency to focus on the gains or loss when making a decision. This study intends to contribute to a better understanding of the decision making process. Third, this study addresses an important business question that captures the attention of industry practitioners. Companies invest considerable amount of resources in setting up and maintaining localized websites. However, there is no systematic study on whether these localized websites are effective in addressing the distinctive preferences of individuals in different regions. Our study will fill these research gaps and provide practical implications about these localization strategies.

References

1. Benet-Martínez, V., Lee, F., Leu, J.: Biculturalism and cognitive complexity: Expertise in cultural representations. Cross-Cultural Psychology 37, 386–407 (2006)
2. Briley, D.A., Morris, M., Simonson, I.: Cultural chameleons: Biculturals, conformity motives and decision making. Consumer Psychology 15, 351–362 (2005)
3. Cyr, D.: Modeling website design across cultures: Relationships to trust, satisfaction and e-loyalty. Management Information Systems 24, 47–72 (2008)
4. Cyr, D., Hassanein, K., Head, M., Ivanov, A.: The role of social presence in establishing loyalty in e-service environments. Interacting with Computers 19, 43–56 (2007)
5. Cyr, D., Trevor-Smith, H.: Localization of web design: An empirical comparison of German, Japanese, and U.S. website characteristics. American Society for Information Science and Technology 55, 1–10 (2004)
6. Dinev, T., Hart, P.: An extended privacy calculus model for e-commerce transactions. Information Systems Research 17, 61–80 (2006)
7. Doney, P.M., Cannon, J.P., Mullen, M.R.: Understanding the influence of national culture on the development of trust. The Academy of Management Review 23, 601–620 (1998)
8. Fulk, J., Schmitz, J., Power, G.J.: A social information processing model of media use in organizations. Communication Research 14, 520–552 (1987)
9. Gefen, D., Straub, D.W.: Managing user trust in B2C e-Services. e-Service J. 2, 7–24 (2003)
10. Gefen, D., Straub, D.W.: Consumer trust in B2C e-commerce and the importance of social presence: Experiments in e-products and e-services. Omega 32, 407–424 (2004)
11. Hassanein, K., Head, M.: Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. International J. Human-Computer Studies 65, 689–708 (2007)
12. Higgins, E.T.: Beyond pleasure and pain. American Psychologist 55, 1217–1233 (1997)
13. Hofstede, G.H.: Culture consequences: International differences in work-related values. Sage Publications, London (1980)
14. Hong, Y., Morris, M.W., Chiu, C., Benet-Martínez, V.: Multicultural minds: A dynamic constructivist approach to culture and cognition. American Psychologist 55, 709–720 (2000)
15. Luna, D., Ringberg, T., Peracchio, L.A.: One individual, two identities: Frame switching among biculturals. Consumer Research 35, 279–293 (2008)
16. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. The Academy of Management Review 20, 709–734 (1995)
17. Miller, P.J.: Narrative practices: Their role in socialization and self-construction. In: Neisser, U., Fivush, R. (eds.) The Remembering Self: Construction and Accuracy in the Self-Narrative, pp. 158–179. Cambridge University Press, New York (1994)
18. Morris, M.W., Peng, K.: Culture and cause: American and Chinese attributions for social physical events. Personality and Social Psychology 67, 949–971 (1994)
19. Ross, M., Xun, W.Q.E., Wilson, A.E.: Language and the bicultural self. Personality and Social Psychology Bulletin 28, 1040–1050 (2002)
20. Short, J., Williams, E., Christie, B.: The Social Psychology of Telecommunications. Wiley, London (1976)
21. Sia, C.L., Lim, K.H., Leung, K.L., Matthew, K.O., Huang, W.W., Benbasat, I.: Web strategies to promote Internet shopping: Is cultural-customization needed? MIS Quarterly 33, 491–512 (2009)
22. Srite, M., Karahanna, E.: The role of espoused national cultural values in technology acceptance. MIS Quarterly 30, 679–704 (2006)

The Impact of E-Commerce on Organizational Performance: The Role of Absorptive Capacity and Integrative Capability

Qing Hu[1], Jianzheng Yang[2], and Lifan Yang[3]

[1] Iowa State University, Ames, Iowa, USA
qinghu@iastate.edu
[2] Shanghai University for Science and Technology, Shanghai, China
cnyangjz@163.com
[3] East China University of Political Science and Law, Shanghai, China
cnyanglifan@163.com

Abstract. This study examines how e-commerce creates value for firms from the perspective of dynamic capability theory. A theoretical model is proposed and tested using structural equation modeling techniques based on survey data collected from firms that have been using e-commerce in their operations for an average of 4 years and have more than 25% of sales or procurement via e-commerce channels. We find that top management participation is a key contributor to the development of a firm's potential and realized absorptive capacities. These two forms of absorptive capacity in turn contribute to the firm's integrative capability, theorized as a form of dynamic capability, which then impacts the firm performance indicators. Different contributions of the two absorptive capacities are delineated, so are the effects of top management on the absorptive capacities. Theoretical and practical contributions of these findings are discussed.

Keywords: E-Commerce, Absorptive Capacity, Integration, Firm Performance.

1 Introduction

Research literature on e-commerce has been diverse in theory, methodology, focus, and findings [23]. Studies that focus on the question of how e-commerce impacts firm performance offered recommendations and prescriptions based on different theoretical and practical perspectives. For example, Saeed et al. [17] examined the impact of e-commerce on firm performance from the perspective of customer psychology and web site features, while Zhu and Kreamer [27][28] focused on how a firm's e-commerce capabilities and IT infrastructure work together to create positive impact on firm performance. Other studies focused their attentions on process integrations with external business partners in B2B environment [13][15].

In this study, we attempt to provide an integrated view of how e-commerce capabilities impact firm performance without differentiating B2C and B2B operations. We argue that in the current business environment it is rare that a firm

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 261–273, 2012.

only focuses on either B2C or B2B side of e-commerce. Sustainable competitive advantage is more likely to be acquired if a firm fully integrates its B2B and B2C strategies and operations into comprehensive e-commerce strategy and operations that permeate strategic planning, internal and external business processes, supply chain, and customer service operations, i.e., the entire value chain of the firm. Therefore, the core research questions of this study are: 1) where can firms develop their e-commerce capabilities? and 2) how do these capabilities impact the firms' performance? To address these questions, we draw on the dynamic capability theory as advanced by Teece et al. [21] and Eisenhardt and Martin [6] and posit that a firm's integrative capability is the core source of its competitiveness in the global e-market. We further argue that this integrative capability is driven by top management participation in e-commerce activities and initiatives, and derived from the firm's potential and realized absorptive capacities. In the sections that follow, we develop this theoretical thesis and test the hypotheses with survey data collected from 145 firms that have e-commerce as a substantial part of their business operations.

2 Theoretical Development

The dynamic capabilities theory [6][12] addresses the criticisms of the widely used resource based view (RBV) [2][16][[24] in the strategic management literature by shifting the focus from the firm's endowment of resources to the firm's capability to reconfigure and deploy the resources in response to competitive conditions. Teece et al. [21] defines dynamic capability as "the firm's ability to integrate, build and reconfigure internal and external competences to address rapidly changing environments" (p. 516). Instead of positing that it is the possession of rare, valuable, immobile, and non-substitutable resources that gives a firm sustainable competitive advantage as in RBV, the dynamic capabilities view emphasizes that it is the ability to use (configure, combine, deploy, and exploit) these resources in response to changing environment that generates sustainable competitive advantage for the firm. As Teece et al. [21] put it: "In short, identifying new opportunities and organizing effectively and efficiently to embrace them are generally more fundamental to private wealth creation than is strategizing, if by strategizing one means engaging in business conduct that keeps competitors off balance, raises rival's costs, and excludes new entrants." (p. 509).

The theory of dynamic capability is especially salient to the understanding of how e-commerce creates competitive advantages to firms. This is because the most commonly used e-commerce technologies and applications, such as those in B2C transactions and B2B exchanges, are readily available to and easily copied by all participants in the competitive landscape. However, what differentiates the competitors is their ability to integrate e-commerce technologies, applications, and processes with internal and external business processes and infrastructure efficiently and effectively. This argument leads to the core thesis of this study: the integrative capability of a firm, defined as the capability that allows the firm to build, accumulate, deploy, and exploit its e-commerce infrastructure and applications among business units and across business partners, is a dynamic capability. Therefore, integrative capability is a key component for value creation by e-commence in firms. We further argue that the integrative capabilities are developed primarily through the firm's absorptive capacity [5] related to e-commerce, and directly and indirectly enhanced by top management participation in e-commerce related activities and initiatives. The integrative capability of the firm leads to higher levels of internal integration between traditional business processes and e-commerce business processes, and external integration between the internal business processes and those of the upstream and downstream partners (suppliers, distributors, and customers). These integrated business processes constitute another set of valuable, rare, immobile, and causal ambiguous resources, thus lead to better firm performance (operational and financial). This thesis is summarized in the research model as shown in Figure 1, and the definitions of the key constructs in this model are presented in Table 1.

2.1 Integrative Capability and Firm Performance

While not widely discussed in the context of e-commerce, research literature has long recognized that integrations between different systems and processes that make up a firm's operation have significant impacts on the bottom line (cost) and the top line (revenue) of a firm [18][27]. This is because that system level and process level integrations eliminate human errors, reduce operational cost due to decreased inventory level and increased inventory turnover, reduce cycle time from order to delivery, and improve quality of operational decisions due to timely data and shared information among different parties within and outside the firm. In addition to the direct operational impacts, due to the complexity, path dependency, and opaque nature of systems and process integrations, they also have the characteristics of strategic resources.

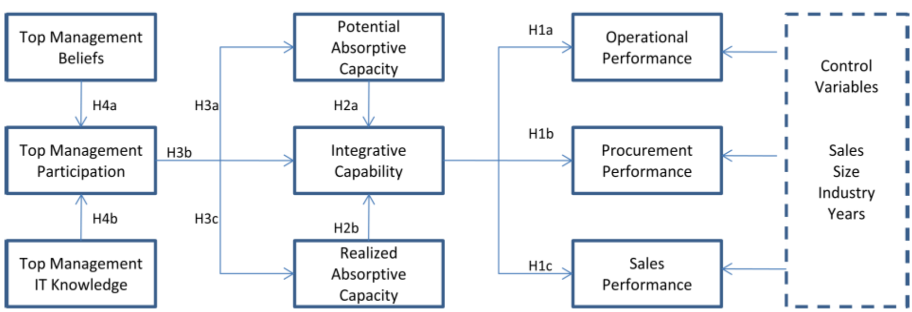


Fig. 1. Research Model

Table 1. Definitions Construct and Reference Sources of Measurement

Construct	Symbol	Definition	Sources of Measurement
Top Management Beliefs	MB	The degree to which top management of the firm believes in the role of electronic commerce in their business.	[14]
Top Management Participation	MP	The degree to which top management of the firm participates in the electronic commerce in their business.	[14]
Top Management IT Knowledge	MT	The degree to which top management of the firm understands IT and the role of IT in their business.	[1]
Potential Absorptive Capacity	PAC	The organizational capacity to learn and understand new technologies and processes for generating solutions to new challenges.	[26]
Integrative Capability	IC	The organizational capability to combine different resources and capacities to support organizational transformation and renewal.	[22][25]
Realized Absorptive Capacity	RAC	The organizational capacity to exploit and transform existing technology and processes for generating solutions to new challenges.	[26]
Impact on Operations	OPR	The impact of ecommerce on a firm's operational performance such as productivity, cycle-time, and customer satisfaction.	[28]
Impact on Procurement	PUR	The impact of ecommerce on a firm's procurement performance such as inventory cost, procurement cost, and cash flow.	[28]
Impact on Sales	SAL	The impact of ecommerce on a firm's sales, such as total revenue, profit margin, and customer satisfaction.	[28]

Integrating heterogeneous systems and disparate business processes as a result of introducing e-commerce into the strategic and operational mix not only requires technique skills, but also demands a firm to be able to coordinate different tasks, teams, and stakeholders, resolve resource, power, and political conflicts and contentions, and overcome unforeseen technical complications related to incompatibility in data, standards, and procedures. To further delineate a firm's ability to integrate its e-commerce systems and processes internally and externally, in this study, we use a single construct called integrative capability to capture a firm's capabilities to combine different resources and processes internal and external to create integrated systems and processes. In this sense, integrative capability is the reincarnation and an operationalization of the dynamic capability concept. The concept of integrative capability was originally proposed by Verona [22] as "an adhesive by absorbing critical knowledge from external sources and by blending the different technical competencies developed in various company departments." (p. 134). Verona defined two types of integrative capabilities: external integrative capabilities and internal integrative capability, with similar dimensions (managerial process, integrative structures, and culture and value for integration) but different orientations (external and internal). On the other hand, Yeoh and Roth [25] distinguished two organizational capabilities: component capability – defined as local abilities that are fundamental to day-to-day problem solving, such as economies of experience, knowledge and skills embedded within the firm, or organizational routines, and integrative capabilities – defined as the ability to deploy or use both resources and component capabilities in new or flexible ways to support organizational renewal. In this study, for theoretical parsimony, we use the single construct – integrative capability – to capture these dynamic capabilities of a firm. Thus, we posit that:

Hypothesis 1a: The level of a firm's integrative capability is positively associated with the operational performance of the firm related to e-commerce.

Hypothesis 1b: The level of a firm's integrative capability is positively associated with the procurement performance of the firm related to e-commerce.

Hypothesis 1c: The level of a firm's integrative capability is positively associated with the sales performance of the firm related to e-commerce.

2.2 Potential and Realized Absorptive Capacity

In the strategic management literature, a closely related concept to dynamic capability is absorptive capacity. Cohen and Levinthal [5] defined the absorptive capacity as a firm's ability to recognize the value of new and external information, assimilate it, and apply it to commercial ends, and argued that it is critical to the innovativeness of the firm. This idea has been widely adopted by scholars and used to explain a broad range of phenomena. In the subsequent literature since Cohen and Levinthal [5], the concept of absorptive capacity has been extended by numerous scholars with new dimensions that go well beyond the original specification, such as personal skills, ability in learning and training, and prior experiences. One of the most critical reconceptualization of absorptive capacity was proposed by Zahra and George [26] in which four fundamental dimensions of absorptive capacity were explicated: acquisition, assimilation, transformation, and exploitation. Zahra and George [26] then proposed two types of absorptive capacity: potential absorptive capacity (PACAP) which consists of the acquisition and assimilation dimensions, and realized absorptive capacity (RACAP) which consists of the transformation and exploitation dimensions. In essence, the PACAP captures the original Cohen and Levinthal [5] specification about absorptive capacity, and the RACAP captures the extensions added by scholars in the later studies.

In this study, for theoretical clarity, we use the concept of potential absorptive capacity (PACAP) and realized absorptive capacity (RACAP) as defined and operationalized in Zahara and George [26]. PACAP emphasizes a firm ability to learn and acquire new knowledge, understand, interpret, and assimilate the acquired knowledge in the new work and organizational environment. On the other hand, RACAP emphasizes the firm's ability to exploit and transform existing knowledge, in the forms of internalization, recodification, and conversion, for generating solutions to new challenges and creating new core competences. By definition, both RACAP and PACAP of a firm will be related to its integrative capabilities which require the firm to learn new technologies and processes, understand the potential of its existing technologies and processes, identify the synergy, and generate new knowledge and solutions. Verona [22] argues that a firm's experimentation and prototyping, learning by doing and learning before doing, which are essentially the characteristics of RACAP and PACAP, are antecedents to its external and internal integrative capabilities. Thus, it is logical that we posit:

Hypothesis 2a: The level of a firm's potential absorptive capacity is positively associated with the firm's integrative capability in the context of e-commerce.

Hypothesis 2b: The level of a firm's realized absorptive capacity is positively associated with the firm's integrative capability in the context of e-commerce.

2.3 Top Management, IT Knowledge, and Participation

Research suggests that top management plays a critical role in the success of innovative and transformative organizational changes [10]. Top management is expected to champion transformational initiatives such as process integration with internal units and external partners by articulating a clear vision and strategy and setting the goals and measures about the initiatives [10]. Top management is expected to convey a strong commitment to the established goals and objectives, and to follow up with the execution by holding lower level managers and employees accountable [19]. Top management can champion new initiatives by participating in meetings, allocating needed resources, and helping resolve conflicts which are usually inevitable with change initiatives. Top management participation in these initiatives also sends a strong signal to other managers and employees about how much they value the initiatives. In other words, top management's active participation and engagement in e-commerce initiative could be felt by managers and employees at all levels. The strongly positive impact of top management participation on the success of many IT related initiatives has been widely supported in the IS literature [11][14]. In the context of ERP implementation, Ke and Wei [11] argued that top management participation contributes to the learning culture of an organization, and organizational learning is directly related to the absorptive capacity [5]. In addition, in a direct test of organizational antecedents to absorptive capacity, Jansen et al. [9] found that participative decision making in organizations has significantly positive impact on both the potential and realized absorptive capacities, especially on the dimensions of acquisition of PACAP and transformation of RACAP, while the impact on the other two dimensions are not statistically significant. Though top management participation is not directly equal to participative decision making in organization, it is certainly an indication of a participative organizational culture which leads to participative decision making processes [11]. Thus, we posit:

Hypothesis 3a: The level of top management participation in e-commerce initiatives is positively associated with the firm's potential absorptive capability (PACAP) related to e-commerce.

Hypothesis 3b: The level of top management participation in e-commerce initiatives is positively associated with the firm's integrative capability related to e-commerce.

Hypothesis 3c: The level of top management participation in e-commerce initiative is positively associated with the firm's realized absorptive capacity (RACAP) related to e-commerce.

However, top management participation in e-commerce initiatives is not necessarily an automatic phenomenon. In a study about ERP assimilation in organizations, Liang et al. [14] found a strong relationship between top management's belief about the role of ERP systems in their business and their support and participation in ERP assimilation activities. On the other hand, Kerns and Sabherwal [12] found that top management's IT knowledge has a significant impact on business managers' participation in IT strategic planning and IT managers' participation in business strategic planning. A general understanding of IT and knowledge about the role of e-commerce will enable top management to determine the priorities of initiatives, direct specific resources for the initiatives, and be more creditable and confident when making the decisions related to the initiatives. Thus, we posit that:

Hypothesis 4a: The level of top management beliefs in e-commerce is positively associated with the level of top management participation in e-commerce initiatives.

Hypothesis 4b: The level of top management knowledge of e-commerce is positively associated with the level of top management participation in e-commerce initiative.

3 Data and Method

The data collection for this project was carried out as part a research grant that calls for conducting seminars to managers in companies across China on the measurement of electronic commerce usage in their companies for a national survey. The seminars were conducted from March 2010 to March 2011. The participants represented companies in more than 10 provinces and cities mostly located in the east coast and central part of China, including Shanghai, Shenzhen, Jiansu, Zhejiang, and Guangdong, where the lion's share of China's GDP is produced. In each of the seminars, the survey instrument was distributed right after the seminar and the participants were asked to complete the survey before they left. A total of 426 questionnaires were distributed and collected, 320 were deemed usable, and 145 were used in the final data set after eliminating those with significant amount of missing data, with total online sales and purchasing less than 5%, with less than one year of using e-commerce as of 2010, or with fewer than 5 employees. This resulted in an effective response rate of 34%. The respondents represented companies range from small (less than 200 employees, with 40 million RMB in sales) to very large (with over 55,000 employees and 55 billion RMB in sales). About half of the companies involved in manufacturing, and over a quarter in services, together they constitute more than 76% of all companies in the final sample. More importantly, on average, about 25% of the sales and procurement of these companies were conducted through e-commerce activities, with about 5 years experience with running e-commerce operations, with the longest being 15 years.

4 Results and Analyses

4.1 Measurement Model Quality

The reliability of measurement addresses the concern of how well the items for one construct correlate or move together [20]. Reliability is usually assessed by two indicators–Cronbach's apha and composite reliability. Cronbach's alpha is a measure of internal consistency among all items used for one construct. Composite reliability addresses similar concept but is considered as a more rigorous reliability measure in the context of structural equation modeling [3]. The reliability indicators of the constructs in this study are shown in Table 2. The lowest composite reliability is 0.813 and the lowest Cronbach's alpha is 0.695, close to or higher than the recommended minimum value of 0.7 [7], indicating good reliability of the measurement for each construct.

Construct validity can be assessed using convergent validity and discriminant validity. Convergent validity is defined as the degree to which the measurement items are related to the construct they are theoretically predicted to be. Convergent validity is shown when the t-values of the outer model item loadings are statistically significant. As it can be seen from Table 2, all item loadings for each construct are significant at $p < 0.01$ ($t > 2.576$), indicating good convergent validity. Hulland [8] recommends that items with loading below 0.5 be dropped. All item loadings in our measurement model are greater than this threshold. All these indicators suggest an acceptable convergent validity in the measurement model.

There are a number of techniques that have been used for testing discriminant validity in the literature [20]. In this study we assess the discriminant validity by comparing the correlations between constructs and the AVE of each construct. This is a widely used technique in the IS literature when component based SEM methods such as PLS is used. Discriminant validity is supported if the square root of construct AVE is greater than the correlations of the construct with all other constructs [8]. In our case, the diagonal values in Table 3 are the square root of AVEs of constructs, which show good discriminant validity for all constructs in the measurement model.

Table 2. Quality Indicators of the Measurement Model

Construct	Direction	Item #	Item Loading	t-stats	AVE	Composite Reliability	Cronbach's Alpha
PAC	Reflective	1	0.812	26.982	0.588	0.850	0.766
		2	0.806	22.757			
		3	0.827	31.214			
		4	0.794	20.546			
IC	Reflective	5	0.802	24.984	0.649	0.881	0.820
		6	0.798	25.577			
		7	0.823	23.663			
		8	0.799	23.845			
RAC	Reflective	9	0.761	15.463	0.656	0.884	0.825
		10	0.806	20.328			
		11	0.799	23.465			
		12	0.697	9.054			
MB	Reflective	25	0.742	13.083	0.621	0.868	0.796
		26	0.801	20.685			
		27	0.764	17.584			
		28	0.843	35.830			
MP	Reflective	29	0.749	11.370	0.611	0.862	0.787
		30	0.830	32.378			
		31	0.803	25.629			
		32	0.740	10.835			
MT	Reflective	33	0.711	11.374	0.611	0.862	0.788
		34	0.829	23.246			
		35	0.759	16.087			
		36	0.821	23.674			
SAL	Reflective	57	0.693	7.443	0.521	0.813	0.695
		58	0.710	9.754			
		59	0.724	10.624			
		60	0.757	12.319			
PUR	Reflective	61	0.895	37.526	0.650	0.880	0.823
		62	0.849	19.397			
		63	0.755	10.334			
		64	0.712	7.652			
OPR	Reflective	65	0.786	14.888	0.591	0.852	0.768
		66	0.712	8.911			
		67	0.755	13.035			
		68	0.818	17.984			

Table 3. Correlations of Constructs

Construct	RAC	IC	PAC	TMB	TMP	TMT	SAL	PUR	OPR
RAC	**0.767**								
IC	0.476	**0.806**							
PAC	0.391	0.717	**0.810**						
TMB	0.463	0.559	0.496	**0.788**					
TMP	0.420	0.602	0.590	0.747	**0.781**				
TMT	0.356	0.493	0.496	0.537	0.591	**0.782**			
SAL	0.316	0.459	0.488	0.578	0.588	0.431	**0.721**		
PUR	0.276	0.434	0.455	0.532	0.507	0.360	0.465	**0.806**	
OPR	0.374	0.513	0.585	0.702	0.649	0.395	0.640	0.639	**0.769**

4.2 Structural Model Analysis

The structural analysis revealed many interesting results, as shown in Figure 2. The overall result supports our core thesis for this research: the dynamic capabilities of a firm, captured by the construct of integrative capability, lead to better organizational performance as indicated by operational, procurement, and sales performance indicators. More specifically, integrative capabilities (IC) in a firm have a significantly positive influence on operational, procurement, and sales performances (H1a, β = 0.538, p<0.01; H1b, β = 0.486, p<0.01; H2b, β = 0.463, p<0.01).

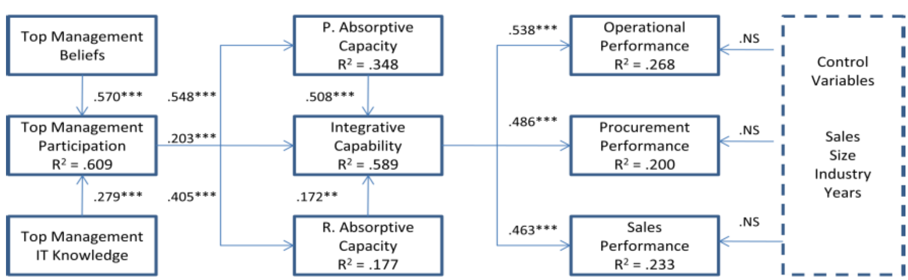


Fig. 2. Results of Structural Analysis (*** 0.01, ** 0.05, *0.1 significant levels)

The second is about how companies can develop its integrative capability – the dynamic capability that is at the core of the sustainable performance in the context of e-commerce. We hypothesized that two distinct but related organizational resources are the essential building blocks – the potential absorptive capacity that focuses on the learning and assimilating aspects of a firm and the realized absorptive capacity that focuses on the exploitative and transformative aspects of the firm. The testing results strongly support this thesis (H2a, β = 0.348, p<0.01; H2b, β = 0.172, p<0.05). Equally important is the fact that 59% of the variances in the integrative capability construct are explained by the three exogenous constructs (PACAP, RACAP, and TMP), a high amount for a latent construct in structural models [4], and thus a strong indication that this model captured the most significant antecedents of the integrative capability construct.

Finally, our results confirm the critical role of top management in the success of e-commerce. Top management participation in e-commerce initiatives significantly impacts directly the integrative capability (H3b, $\beta = 0.203$, $p<0.01$) and indirectly via potential absorptive capacity (H3a, $\beta = 0.548$, $p<0.01$) and realized absorptive capacity (H3c, $\beta = 0.405$, $p<0.01$). An interesting observation from this result is that the effect of top management participation on the integrative capability seems to be mediated by the absorptive capacity, given the significantly larger magnitude of the two indirect paths. However, this issue needs further investigation in future studies.

In addition, our results also show what facilitates top management participation: top management belief (H4a, $\beta = 0.570$, $p<0.01$) and top management IT knowledge (H4b, $\beta = 0.279$, $p<0.01$). Together these two exogenous constructs explain more than 60% of the variance in top management participation. Once again, this is a high percentage by the standard in social science research [4], and an indication of the strong effects of these constructs on the endogenous construct. It is also interesting to note that the magnitude of the two path coefficients: the effect of top management belief construct is about twice as strong as the effect of the top management's IT knowledge, indicating that top management participation in e-commerce initiatives is primarily motivated by the top managers' beliefs in electronic commerce; and that the knowledge related to IT is certainly helpful, but not necessarily a strong determinant.

Among the control variables, none of them is found to have any significant impact on the three dependent variables. Overall the theoretical model is strongly supported by the data. All but one of the hypothesized relationships were found to be significant at $p<0.01$ level, and the R^2 values for the critical endogenous constructs are reasonably high and very high, especially for the three dependent constructs: operational performance (.268), procurement performance (.200), and sales performance (.233), indicating the constructs in the model have captured a significant amount of variances in the phenomenon of interest: how and where electronic commerce activities impact firm performance.

5 Conclusions

We developed and tested a dynamic capability model that explains how e-commerce operations may impact the operational, procurement, and sales performances of a firm. Our results show that the integrative capability of a firm can be developed via organizational absorptive capacity. Moreover, our results show the role of top management in creating and enhancing the integrative capability – directly and indirectly via the two dimensions of absorptive capacity – potential and realized.

This study contributes to the theories of e-commerce research in at least three areas. First, we introduced the construct of integrative capability into electronic commerce research. The notion of integrative capability not only supports the dynamic capabilities view of strategic management literature, more importantly it focuses the attention of scholars on a higher level construct above the commonly studied constructs of system integration and process integration. Second, we tested the re-conceptualization of absorptive capacity by Zahra and George [26] and refined the understanding of how each contributes to the performance and how each is affected by organizational antecedents such as top management participation. By using the new conceptualization of absorptive capacity, we are able to capture the extensions to the original absorptive capacity construct in the recent literature and still preserve the original Cohen and Levinthal [5] concept in our model, thus providing theoretical continuity and clarity to this critically important concept. Finally, this study examines the question of how and where e-commerce generates value for firms by anchoring on the widely accepted and more advanced dynamic capability framework in the strategic management literature, thus providing strong theoretical footing for the proposed model, and a solid foundation for future e-commerce research.

Our findings also offer some prescriptive insights for managing e-commerce initiatives. First, by showing the critical role of the integrative capability in the success of e-commerce operations in firms, we remind managers where to focus their attention and resources. While web site design and customer services are still critical to the success of e-commerce, they are rarely the differentiating factors that can sustain competitive advantage in today's environment. Our study suggests that strengthening integrative capability should be the locus of e-commerce strategy after the infrastructure has been operationalized. The question then is how and where? Our model suggests that two factors are at the core of building the integrative capability: absorptive capacity and top management participation. On the other hand, our model shows that top management can play a critical role by actively participating in e-commerce initiatives. The significant factors that facilitate top management participation are top management beliefs and top management IT knowledge, especially the former. Therefore, firms should consider top management IT training, unless their top management team already has a strong IT background, as one of the most important components of their overall e-commerce strategy.

Acknowledgment. This research is partially supported by a grant from the National Natural Science Foundation of China (70973079) and a grant from Shanghai Leading Academic Discipline Project (S30504).

References

1. Armstrong, C., Sambamurthy, V.: Information technology assimilation in firms: The influence of senior leadership and IT infrastructures. Information Systems Research 10, 304–327 (1999)
2. Barney, J.B.: Firm resources and sustained competitive advantage. Journal of Management 17, 99–120 (1991)
3. Chin, W.W.: The Partial Least Squares Approach to Structural Equation Modeling. In: Marcoulides, G.A. (ed.) Modern Methods for Business Research, pp. 295–336. Lawrence Erlbaum Associates, Hillsdale (1998)
4. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Lawrence Erlbaum Associates, Hillsdale (1988)
5. Cohen, W.M., Levinthal, D.A.: Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly 35, 128–142 (1990)
6. Eisenhardt, K.M., Martin, J.A.: Dynamic capabilities: What are they? Strategic Management Journal 21, 1105–1121 (2000)
7. Gefen, D., Straub, D.W., Boudreau, M.C.: Structural Equation Modeling And Regression: Guidelines for Research Practice. Communications of AIS 4, Article 7 (2000)
8. Hulland, J.: Use of Partial Least Squares (PLS) in Strategic Management Research: A Review of Four Recent Studies. Strategic Management Journal 20, 195–204 (1999)

9. Jansen, J.J.P., Bosch, F.A.J.V.D., Volberda, H.W.: Managing Potential and Realized Absorptive Capacity: How Do Organizational Antecedents Matter? Academy of Management Journal 48, 999–1015 (2005)
10. Jarvenpaa, S.L., Ives, B.: Executive Involvement and Participation in Management Information Technology. MIS Quarterly 15, 205–227 (1991)
11. Ke, W., Wei, K.K.: Organizational culture and leadership in ERP implementation. Decision Support Systems 45, 208–218 (2008)
12. Kearns, G.S., Sabherwal, R.: Strategic Alignment between Business and Information Technology: A Knowledge-Based View of Behaviors, Outcome, and Consequences. Journal of Management Information Systems 23, 129–162 (2007)
13. Lee, S.C., Pak, B.Y., Lee, H.G.: Business value of B2B electronic commerce: the critical role of inter-firm collaboration. Electronic Commerce Research and Applications 2, 350–361 (2003)
14. Liang, H., Saraf, N., Hu, Q., Xue, Y.: Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management. MIS Quarterly 31, 59–87 (2007)
15. Mukhopadhyay, T., Kekre, S.: Strategic and Operational Benefits of Electronic Integration in B2B Procurement Processes. Management Science 48, 1301–1313 (2002)
16. Peteraf, M.A.: The Cornerstones of Competitive Advantage: A Resource-Based View. Strategic Management Journal 14, 179–191 (1993)
17. Saeed, K.A., Hwang, Y., Grover, V.: Investigating the Impact of Web Site Value and Advertising on Firm Performance in Electronic Commerce. International Journal of Electronic Commerce 7, 119–141 (2002)
18. Saraf, N., Langdon, C.S., Gosain, S.: IS Application Capabilities and Relational Value in Interfirm Partnerships. Information Systems Research 18, 320–339 (2007)
19. Sharma, R., Yetton, P.: The Contingent Effects of Management Support and Task Interdependence on Successful Information Systems Implementation. MIS Quarterly 27, 533–555 (2003)
20. Straub, D.W., Boudreau, M.C., Gefen, D.: Validation Guidelines for IS Positivist Research. Communications of the AIS 13, 380–427 (2004)
21. Teece, D., Pisano, G., Shuen, A.: Dynamic Capabilities and Strategic Management. Strategic Management Journal 18, 509–533 (1997)
22. Verona, G.: A Resource-Based View of Product Development. The Academy of Management Review 24, 132–142 (1999)
23. Wareham, J., Zheng, J.G., Straub, D.: Critical themes in electronic commerce research: a meta-analysis. Journal of Information Technology 20, 1–19 (2005)
24. Wernerfelt, B.: A Resource-Based View of the Firm. Strategic Management Journal 5, 171–180 (1984)
25. Yeoh, P.L., Roth, K.: An Empirical Analysis of Sustained Advantage in the U.S. Pharmaceutical Industry: Impact of Firm Resources and Capabilities. Strategic Management Journal 20, 637–653 (1999)
26. Zahra, S.A., George, G.: Absorptive Capacity: A Review, Reconceptualization, and Extension. The Academy of Management Review 27, 185–203 (2002)
27. Zhu, K., Kraemer, K.L.: E-Commerce Metrics for Net-Enhanced Organizations: Assessing the Value of e-Commerce to Firm Performance in the Manufacturing Sector. Information Systems Research 13, 275–295 (2002)
28. Zhu, K., Kraemer, K.L.: Post-Adoption Variations in Usage and Value of E-Business by Organizations: Cross-Country Evidence from the Retail Industry. Information Systems Research 16, 61–84 (2005)

IT Governance: The Key Factor of E-Government Implementation in China

Tianmei Wang[1], Baowen Sun[1], and Zhijun Yan[2]

[1] School of Information, Central University of Finance and Economics, Beijing, P.R. China
[2] School of Management and Economics, Beijing Institute of Technology, Beijing, P.R. China
wangtianmei@cufe.edu.cn, sunbaowen@263.net, yanzhijun@bit.edu.cn

Abstract. According to E-government maturity model, Leavitt's Diamond model, Fountain's technology enactment framework and IT governance theory, this paper discusses the important impact that IT governance bring in the effectiveness of e-government implementation. Furthermore, the critical influence factors model of e-government implementation effectiveness is proposed. Data are collected from a survey and analyzed by following a psychometric procedure. All data from government official of China's public sector across 29 provinces. The research result shows that the E-government governance capability is positively related with e-government implementation effectiveness, Moreover, the environmental readiness and organizational support are also positively related with e-government implementation effectiveness.

Keywords: E-Government, Implementation Effectiveness, Influence Factors.

1 Introduction

Compared with the informationization constructions of governments around the world before the 1990s, E-government has three basic features that can not be ignored: First, E-government develops based on the Internet, and second, E-government emphasizes to structure public services that take the public as the center, and the third the informationization constructions of governments obviously have been from the data processing stage into the information management and business processing stage, the constructions of the large-scale cross-departments E-government application systems have been gradually carrying out. It is thus clear that the implementation of E-government presents a more complex features, its implementation effect will inevitably be affected by a variety of factors. It is necessary for E-government practitioners and managers to observe, analyze, verify, manage and continuous control critical success factors and effectively achieve the implementation objectives of a project[1]. Therefore, to study the critical factors that affect the implementation effects of E-government has important theoretical and practical significance.

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 274–285, 2012.

2 Literature Review

In the field of the theoretical research of information systems, influencing factors research is an important research branch, which aims to monitor the influencing of the key influencing factors in the implementation process through analyzing the key influencing factors in the implementation process of an information system, and to ensure the efficient allocation of resources and the implementation effectiveness of the project[1]. As early as 1987, Kwon summarized the past research literatures about the key factors and identify 22 key factors for the successful implementation of information systems, and divided them into five categories, including Structural Factors, Technological Factors, Task-related Factors, Individual Factors and Environmental Factors[2]. On this basis, Larsen systemized the research literatures about the key factors in the top five international journals on information systems, listed of 69 critical success factors, and divided into seven categories, in addition to the five factors put forward by Kwon and Zmud, added Process Factors and Interorganizatioal Factors[3]. Based on the analysis and summary to the literatures, Santhanam divided 59 critical success factors of the information systems into four categories, including Organizational Factors, Managerial Factors, Technological Factors and Information System-related Task factors [4]. These studies define several major categories for critical success factors of information systems with a more macroscopic view, what is noticeable is that all these studies mentioned the task factors, showing that the task characteristics of an information system to a large extent affected the effectiveness of the implementation of an information system, the degrees of influencing of the same key factors for different types of information systems implementation effectiveness are quite different, so it is necessary to in-depth study of the key factors affecting the implementation effectiveness of information systems in specific situations[4].

As an information system project of the government department, E-government has general rules of the information system project implementations, but also has great differences in the task features. For the study of the key factors of the E-government implementation effectiveness, most of the literatures base on these studies, but have some differences in the study perspective. Some scholars summarized the key success factors of the E-government implementation from a more abstract overall level, for example, through systemizing the literatures about the E-government key success factors in top five journals in public management field (Public Administration Review, Journal of Public Administration Research and Theory, American Review of Public Administration, Administration and Society, and Public Performance and Management) from 1999 to 2003, Gil-Garcia considered that the key factors of E-government includes five categories: Information and Data, Information Technology, Organizational and Managerial, Legal and Regulatory, Institutional and Environmental[5]; Yoon analyzed many critical success factors of E-government in national level, such as ICT infrastructure, financial resources, public education, system, privacy and security, leadership, involvement of private organizations, and through statistical analysis, discussed the significant differences in the priority on the critical success factors between the developed countries, developing countries and underdeveloped countries[6].Other scholars refined the influencing of a certain factor to the E-government implementation effectiveness based on the specific situations, of them Ke summarized the key success factors that Singapore government to implement E-government through reviewing and analyzing the Singapore's E-government implementation process, including the Government's leadership, bridging the digital divide, coordination mechanisms between government departments[7];Chu adopted the empirical research methods and in-depth discussed the user factors that affect the E-government implementation effectiveness such as user action and user attitudes based on the Technology Acceptance Model and the practice of the Taiwan government's ETS(Electronic Tendering System) project[8]. Gichoya summed up a conceptual framework of the success factors and failure factors of E-government from a microcosmic view of a government department's project implementation[9]; Based on the case studies of the implementation of the E-government projects of the Taiwan's Bureau of Foreign Trade from 1998 to 2003, Tseng refined the management factors affecting the E-government implementation effectivenessfrom several views, such as strategic alliances, project implementation and design, management changing and innovation, stakeholder management, MIS development capacity[10]; Prybutok adopted empirical studies method to discuss the relationship between the leadership, quality of information and the interests of the network, explained the important role of leadership[11].

Above studies provides rich references to the key factors researches of the E-government implementation effectiveness, but because of the differences in the study levels, resulting a variety of study findings, because the research perspective of this paper focuses on the analysis of the management issues of the micro-perspective-governmental organizations in the process of implementing E-government projects, therefore, from the perspective of project implementation, we found that three study variables have garnered particular attentions: the degree of readiness of the external environment and the support level of the internal environment, as important variables affecting the E-government implementation effectiveness, become the focuses of discussion in most of the literatures[5, 9, 12-15]; the complex features of a project, as a situational factor of an information system implementing task, also attracts extensive attentions in the key success factors researches in the information system field[5, 16-17]。

However, as a key IT investment project of a government, E-government has great potentials in improving the public administration capacity and public service capacity of government departments. But as the trans-department IT project that touches the interests of multiple stakeholders, the implementation of E-government also contains various risks, therefore, it becomes very important to govern the E-government implementation process[18].

As the extension of the concept of IT governance in the government departments, E-government governance means that the governmental organizations supervise the performance of the E-government construction and managers, ensure the rights of stakeholders, and promote the maximization of the public management and public service capacity through arranging and balancing the critical E-government decision-making authorities. E-government governance specifies the models, structures, processes and mechanisms of the E-government project decision-making for the governmental organizations from a system level, to ensure both rules compliance and the consistency between the information technology and business strategy, to help government departments to improve the decision-making efficiency of the

E-government projects, achieve strategic objectives and avoid risks. Although there is no a lot of empirical researches about the E-government governance, but recently some E-government research literatures[17, 19-20]proposed certain research results about the institutional arrangements influencing the effectiveness of E-government implementation at different levels.

According to Fountain's technology enactment framework[20], this paper argues that E-government implementation is a typical process that the objective information technology is translated into the implemented information technology. The implementation of E-government begins with the expected goals according to the objective information technology, in the technical implementation process, due to hitting the interests of multiple stakeholders, will have a lot of organization and management issues. Along with the continuous generation and solutions of these problems, E-government implementation process will demonstrate a significant dynamic complexity, resulting in an unexpected deviation between E-government implementation effectiveness and the expected implementation goals.

In fact, based on the existing literatures, the dynamic complexity of E-government implementation is decided by three levels, as figure shown: first is the research results based on the E-government maturity model[21-24], this paper argues that E-government as a newborn things, has its own law of development, with continuous deepening of the information work within an organization, E-government projects will increase and lead to the technical and organizational complexities is growing. Second is based on the Leavitt's Diamond model[25], this paper considers that E-government, which as information technology is introduced into governmental organizations, will lead to three key factors in the organization interact, and then influence the accepting degree of the governmental organizations to E-government. Third is based on the Fountain's technology enactment framework[20], This paper argues that organization's institutional arrangements, as a key element affecting the behaviors and expectations of the actors of a government organization, also affect the implementation effectiveness of E-government. Meanwhile, from the perspective of organizational behavior, a major cause of the dynamic complexity of the E-government implementation is that the E-government implementation will affect the new structure of information resources – a kind of organization's critical resources. Thus, the implementation will result the conflicts of interests between stakeholders of the organizations participate in the implementation (also known as the organizational behaviors). It is clear that the implementation of E-government in fact is the result of the "micro-politics" competition among interest groups or between stakeholders in the government organization to strive for the influences on the organizational policies, procedures and resources[26]. The degree of deviation between the E-government implementation effectiveness and expected goals depends on the game between the "micro-politics" forces.

So this paper considers that the E-government governance, as the extension of the concept of IT governance in the E-government implementation, is to study the decision-making institutional arrangement for the E-government implementation, is designed to rationalize the relationships of the responsibility, the right, the obligation between stakeholders in the E-government implementation. Based on the analyzing to the key decision-making items during the process of the E-government implementation and the identification of decision-making bodies and stakeholders in the E-government implementation, to carry out the governance arrangements of the decision-making powers and to design appropriate governance structures and governance mechanisms, to try to control the E-government implementation path and to ensure the E-government implementation effectiveness from the root causes that may result E-government implementation effectiveness deviates the expected results. Therefore, this paper argues that the E-government governance capacity, as a key influencing factor to the E-government implementation effectiveness, has important positive effects the E-government implementation effectiveness.

3 Research Model

Based on the above theoretical researches, mainly from the perspective of micro-government's organization and management of the E-government implementation, this paper considers that the readiness of the external environment and the degree of support of the internal environment, E-government governance capacity and the complexity characteristics of the project affect the E-government implementation effectiveness in different degrees, therefore, this paper proposed a conceptual model of this study as shown in Figure 1, and defined the specific meaning of the main study variables as shown in Table 1.

The E-government implementation effectiveness refers to the system quality and efficiency in the use of the governmental organizations' E-government implementation, most scholars, such as Gichoya, Gil-Garcia, Seddon, etc., believe that the E-government implementation effectiveness is affected by a variety of factors both inside and outside the organization[5, 9, 15].

The readiness of the external environment refers to the readiness of the political, social, economic, technological and other factors outside governmental organizations. Seddon, Gil-Garcia and others all think that the external factors constituted by policy orientations of higher authorities, the regional economic level, development of technical standards and others have impacts on the implementation effectiveness of information technology in government organizations[5, 9, 15, 27-28]. Therefore, We hypothesize the following:

H1: The readiness of the external environment has positive influences on the E-government implementation effectiveness.

The degree of support of the internal environment refers to the degrees of support of the executive heads, information management environment, information resources, and a variety of management personnel within governmental organizations to the E-government implementation. Kamal, Svara all deem that the internal organizational environment such as the supports of all kinds of roles within the organization and the possession statuses of information resources within the organization have important influences on the E-government implementation effectiveness[12-14]. Therefore, We hypothesize the following:

H2: The degree of support of the internal environment has positive influences on the E-government implementation effectiveness.

E-government governance capacity is the general effect of institutional arrangements of the E-government decision-makings related to the governance model, governance structure, and governance mechanism in the E-government implementation process. Fountain, Kim, Liang all consider that the institutional arrangement will have important influences on the E-government implementation effectiveness[20, 29-31]. Therefore, We hypothesize the following:

H3: The E-government governance capacity has positive correlativity on the E-government implementation effectiveness.

The complexity characteristic of an E-government project refers to the degrees of the complexities such as the size of a specific E-government project, scope, the impact on the flow within the governmental organization. Larsen and Santhanam highlighted the organizational information technology task characteristics have important influences on the implementation effectiveness of the information systems, this factor can be regarded as the situational factor influencing the system implementation effectiveness[3-4]. Azad, Gil-Garcia, Tan deem that the complexity of an E-government project is due to its trans-department feature and touching multi-stakeholders, which will largely affect the E-government implementation effectiveness[5, 16-17, 32]. Therefore, We hypothesize the following:

H4: The complexity characteristic of an E-government project has passive influences on the implementation effectiveness.

4 Methodology

4.1 Instrument Development

Based on some researches[21-24], from a microcosmic view of that the government department organizing the E-government implementation, this paper chooses six indexes to measure the E-government implementation effectiveness from two point of views of the system quality and efficiency. Of which the system quality refers to the basis of the E-government implementation effectiveness, since the system quality is a larger concept, which not only includes the stability and response time of the system itself, but the reasonableness and usability of the system functions, also includes the effectiveness, timeliness, accuracy of the information generated by the system, because this study is based on the point of view of the "government affairs" and ignores the point of view of the "technology", so we only choose the system function and the quality of information as the evaluation indexes related to the system quality; There exist some differences in utilization efficiency for different E-government projects, but all include employees' utilization efficiency, organizations' utilization efficiency and communities' utilization efficiency, of which, as the direct users of E-government systems, their utilization efficiency can be measured by the improving of the work efficiency; the measurement for organizations' utilization efficiency is more complicated, due to the goal of the construction of E-government is to improve the quality of public services through enhancing collaboration between organizations, so this paper selects the level of information sharing and the quality of public services as the indexes to measure the organizations' use benefit, of them the level of information sharing can be regarded as the utilization efficiency between departments within the organization, and the quality of public services can be viewed as the utilization efficiency of the organization as a whole; Since the communities' utilization efficiency will differ for different types of E-government project, this paper only selects an universal index, that is to take the customer satisfaction as the evaluation index to measure the utilization efficiency of social users (public, corporate users), in this study, only when all six indexes got positive evaluation, we consider that the E-government has good implementation effectiveness.

Similarly, based on research literatures mentioned on the third part of this paper, to establish measurement indexes for four endogenous latent variables, namely the readiness of the external environment, the degree of support of the internal organization, the E-government governance capacity and the complexity of the project.

4.2 Data Collection

The investigation took more than four months, in the way of combination of the key-point investigation for typical users and the blanket investigation for ordinary users; respondents are mainly government officials in charge of information technology from the province to district and county levels of local government. Over the past 4 months, we commissioned information technology journal to issue 200 copies of paper questionnaires and 300 copies of electronic questionnaires through E-government meetings and lectures, total of 201 questionnaires were returned and rate of return is 40.2%, of which valid questionnaires are 171 copies, the valid rate of return is 34%. The invalid questionnaire mainly refer to questionnaires that data missing rate is over 50%.

5 Main Study and Data Analysis

5.1 Measurement Model

Of which the convergence validity test of the measurement model is as shown in Table 1, normalized factor loadings of all measured variables all are greater than 0.5 and reach significant level, the composite reliability (CR) values of latent variables all are greater than 0.7, average variance extracted (AVE) all are greater than 0.5, fundamentally reach the recommended values of Fornell and Larcker (1981). Discriminant validity test is as shown in Table 2; the related coefficient below the diagonal is far less than the square root of AVE above the diagonal, showing that this questionnaire has good discriminant validity. It is thus clear that the revised model has good convergence validity and discriminant validity; we can accept this measurement model and carry out the equation of structure test in the next step.

5.2 Structural Model

On the basis of the confirmatory factor analysis, we used LISREL8.80 to test entire structural equation model. The fit indexes of the model are as shown in Fig.1.

Table 1. The convergent validity of the measurement model

Construct	Item	Std.Loadin g	*T* value	AVE	CR	Cronbach's α
Effectiveness	Y1	0.82	0.82	0.55	0.88	0.87
	Y2	0.69	0.69			
	Y3	0.72	0.72			
	Y4	0.60	0.60			
	Y5	0.81	0.81			
	Y6	0.79	0.79			
Environment	X1	0.73	0.73	0.54	0.70	0.70
	X2	0.74	0.74			
Organization	X4	0.86	0.86	0.64	0.84	0.82
	X5	0.87	0.87			
	X7	0.65	0.65			
Governance	X7	0.76	0.76	0.50	0.80	0.80
	X8	0.63	0.63			
	X9	0.69	0.69			
	X10	0.82	0.82			
Complexity	X13	0.76	0.76	0.53	0.77	0.75
	X14	0.61	0.61			
	X15	0.80	0.80			

Table 2. Descriptive statistics and factor correlation

$\sqrt{AVE}$	Effectiveness	Environment	Organization	Governance	Complexity
Effectiveness	0.7424				
Environment	0.6012	0.7350			
Organization	0.5653	0.3474	0.7997		
Governance	0.6792	0.6176	0.6319	0.7183	
Complexity	0.4519	0.4609	0.4794	0.7151	0.7279

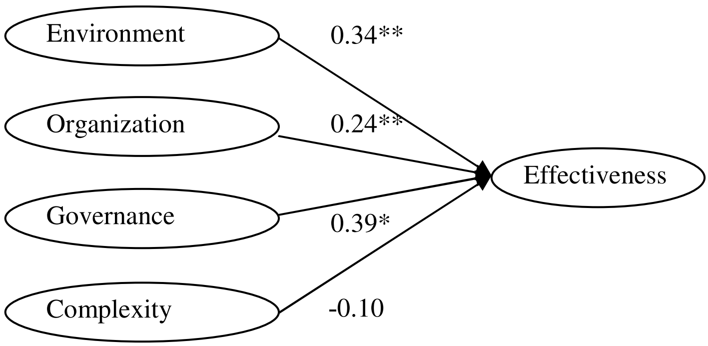


Fig. 1. Standardized LISREL solution for hypothesis testing (Chi-Square=282.99, df=125, P-value=0.00000, RMSEA=0.086, *p<0.05 **p<0.01 ***p<0.001)

Results of the empirical researches confirmed the hypothesis H1, H2, H3 respectively, finding that the readiness of the external environment, the degree of support of the internal organization and E-government governance capacity are three key factors influencing the E-government implementation effectiveness. However, for hypothesis H4, although the empirical results show that there exists negative correlativity between the two, but the correlation is not significant, can not support this hypothesis.

6 Discussion and Implications

Combining the research literatures about influencing factors of the information systems theory and literatures about key influencing factors of the E-government implementation, this paper established a theoretical model of key influencing factors for the E-government implementation effectiveness, and proposes four research hypotheses; then, adopting 171 valid samples from the Chinese government's department to carry out the measurement model test, according to the proposal of the revised indexes, combining with the theoretical analysis, we revised the measurement model, convergent validity and discriminant validity of the revised model and model fit indexes all are up to empirical values, confirmed the effectiveness of the measurement tool; on this basis, we further verified the structure model, three hypotheses of four research hypotheses were confirmed in different significance levels, results of the quantitative analysis confirmed the effectiveness and credibility of the theoretical model presented in this paper.

The readiness of the environment and the degree of support of the organization have significant positive influences on the E-government implementation effectiveness, which is same with research conclusions of many Western scholars. Although there are many differences in administrative system, that is to say, the majority of Western governments adopt Zone Management system, state or local governments have a strong executive authority, with stronger mandatory leadership to functional departments within its region, while Chinese government departments adopt typical matrix management mode, and base on vertical management, there are many barriers between higher and lower levels or between different departments and regions, but in the E-government implementation process, the readiness of the external environment and the degree of support of the internal organization have important influences on the implementation effectiveness, which is also consistent with the general laws of information system implementation.

E-government governance capability is the kernel variable studied in this paper, also is a new latent variable that has important influences on the E-government implementation effectiveness formed by summarizing related literatures, empirical studies have confirmed that, as the readiness of the environment and the degree of support of the organization, the E-government governance capability has a significant positive effect on the E-government implementation effectiveness, which is consistent with the research conclusions of Fountain, Kim, Liang that the institutional arrangement should be paid attention to in the E-government implementation, also is consistent with the practices of the Western governments that relying on systems to promote E-government. Although as early as 10 years ago Chinese government has begin to practice the E-government institutional arrangement in the macro level, such as the emergence of the Information Work Leading Group and the Joint Conference System, but the decision-making institutional arrangement in the field of the E-government implementation in the micro-level can not be ignored, especially with the development of the E-government gradually deepening into the integration phase and there are a large number of trans-department E-government implementation projects, It's of great urgency that to improve the E-government governance capability of the local government departments.

Of course, for studying which factor has the greatest influences on the E-government implementation effectiveness, this paper tried to revise the structural model and found that the revised results are not ideal, meanwhile, for lacking of theoretical basis, empirical research data merely confirmed the E-government governance capability, the readiness of the environment and the degree of support of the internal environment have significant positive effects of the E-government implementation effectiveness.

In this paper, the negative influencing of the complexity characteristics of the project on the E-government implementation effectiveness is not significant, which is not consistent with the research conclusions of some Western scholars. This paper argues that this is determined by the research background, compared to the general information system, in addition to the more complex technological structure, E-government project also involves flow optimization, organization restructuring and other activities, its implementation process is a bit complicated. In this survey, 63.08% of the government departments have launched E-government over five years, have consensus to the "complexity" of the E-government, therefore, this paper thinks that the primary cause resulting in the negative influencing of the complexity characteristics of the project on the E-government implementation effectiveness is not significant is that the sensitivities of the research departments to the complexity characteristics of the project are not high.

Acknowledgement. This research was supported by the National Natural Science Foundation of China(Grant No. 70872119), the Special Funds for Construction

Project of Ministry of Education of Beijing, Science Foundation of Ministry of Education for Young Scholars of China (Grant No. 09YJC630240), Beijing Natural Science Foundation of China (Grant No. 9092014), 211 Project for Central University of Finance and Economics(the 3rd phase), Discipline Construction Foundation of Central University of Finance and Economics, Program for Innovation Research in Central University of Finance and Economics.

References

1. Shank, M., Boynton, A., Zmud, R.: Critical success factor analysis as a methodology for MIS planning. Mis Quarterly, 121–129 (1985)
2. Kwon, T., Zmud, R.: Unifying the fragmented models of information systems implementation. Wiley Series In Information Systems, pp. 227–251 (1987)
3. Larsen, K.: A taxonomy of antecedents of information systems success: Variable analysis studies. Journal of Management Information Systems 20(2), 169–246 (2003)
4. Santhanam, R., Hartono, E.: Common Factors among Management Support Systems' Success. In: AMCIS 2003 (2003)
5. Gil-Garcia, J.R., Pardo, T.A.: E-government success factors: Mapping practical tools to theoretical foundations. Government Information Quarterly 22(2), 187–216 (2005)
6. Yoon, J., Chae, M.: Varying criticality of key success factors of national e-Strategy along the status of economic development of nations. Government Information Quarterly 26(1), 25–34 (2009)
7. Ke, W.L., Wei, K.K.: Successful e-government in Singapore - How did Singapore manage to get most of its public services deliverable online? Communications of the ACM 47(6), 95–99 (2004)
8. Chu, P.Y., et al.: Exploring success factors for Taiwan's government electronic tendering system: behavioral perspectives from end users. Government Information Quarterly 21(2), 219–234 (2004)
9. Gichoya, D.: Factors Affecting the Successful Implementation of ICT Projects in Government. Electronic Journal of e-Government 3(4), 175–184 (2005)
10. Tseng, P.T.Y., et al.: To explore managerial issues and their implications on e-Government deployment in the public sector: Lessons from Taiwan's Bureau of Foreign Trade. Government Information Quarterly 25(4), 734–756 (2008)
11. Prybutok, V.R., Zhang, X.N., Ryan, S.D.: Evaluating leadership, IT quality, and net benefits in an e-government environment. Information & Management 45(3), 143–152 (2008)
12. Ho, A.T.K.: Reinventing local governments and the e-government initiative. Public Administration Review 62(4), 434–444 (2002)
13. Svara, J.H.: Strengthening Local Government Leadership and Performance: Reexamining and Updating the Winter Commission Goals. Public Administration Review 68, 537–549 (2008)
14. Kamal, M.: IT innovation adoption in the government sector: identifying the critical success factors. Journal of Enterprise Information Management 19(2), 192–222 (2006)
15. Seddon, P.: A respecification and extension of the DeLone and McLean model of IS success. Information Systems Research 8(3), 240–253 (1997)
16. Tan, C.W.: Managing stakeholder interests in e-government implementation: Lessons learned from a Singapore e-government project. Journal of Global Information Management 13(1), 31–53 (2005)

17. Gil-Garcia, J.R., Martinez-Moyano, I.J.: Understanding the evolution of e-government: The influence of systems of rules on public sector dynamics (2007)
18. Rocheleau, B.: Whither E-goverment? Public Administration Review 67(3), 584–588 (2007)
19. Azad, B., Faraj, S.: E-Government institutionalizing practices of a land registration mapping system. Government Information Quarterly 26(1), 5–14 (2009)
20. Fountain, J.: Building the Virtual State: Information Technology and Institutional Change. Brookings Institution Press (2001)
21. Layne, K., Lee, J.W.: Developing fully functional E-government: A four stage model. Government Information Quarterly 18(2), 122–136 (2001)
22. Hiller, J., Bélanger, F.: Privacy Strategies for Electronic Government. E-Govemment Series. PricewaterhouseCoopers Endowment for the Business of Govemment, Arlington (2001)
23. Moon, M.J.: The evolution of e-government among municipalities: Rhetoric or reality? Public Administration Review 62(4), 424–433 (2002)
24. Andersen, K.V., Henriksen, H.Z.: E-government maturity models: Extension of the Layne and Lee model. Government Information Quarterly 23(2), 236–248 (2006)
25. Leavitt, H., James, G.: Applied Organizational Change in Industry: Structural, Technological and Humanistic Approaches. Carnegie Institute of Technology, Graduate School of Industrial Administration (1962)
26. Dovifat, A., Brüggemeier, M., Lenk, K.: "The model of micropolitical arenas''–A framework to understand the innovation process of e-government-projects. Information Polity 12(3), 127–138 (2007)
27. Kraemer, K., Perry, J.: The federal push to bring computer applications to local governments. Public Administration Review, 260–270 (1979)
28. Perry, J., Danziger, J.: The Adoptability of Inno Va Tions: An Empirical Assessment of Computer Applications in Local Governments. Administration & Society 11(4), 461 (1980)
29. Kim, S., Kim, H.J., Lee, H.: An institutional analysis of an e-government system for anti-corruption: The case of OPEN. Government Information Quarterly 26(1), 42–50 (2009)
30. Liang, H., et al.: Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. Mis Quarterly 31(1), 59–87 (2007)
31. Norris, D.F., Moon, M.J.: Advancing e-government at the grassroots: Tortoise or hare? Public Administration Review 65(1), 64–75 (2005)
32. Azad, B., Faraj, S.: Making e-Government systems workable: Exploring the evolution of frames. Journal of Strategic Information Systems 17(2), 75–98 (2008)

Part III

E-Business Systems and Applications

The Impact of Cloud Services on Independent Software Vendors: Should We Step into Cloud?

Jhih-Hua Jhang-Li and Chih-Yao Lee

No. 101, Sec.1, Fenliao Rd., LinKou District, New Taipei City 24452
Hsing Wu Institute of Technology, Taiwan
jhangli@gmail.com

Abstract. Cloud services have become one of the most popular industry terms since Google and IBM invested to build large data centers that users can program and research over the Internet. In practice, IT corporations often annually pay independent software vendors considerable license fees to save the cost of software upgrade and technology support. Although cloud services are considered a cost-down solution for small or medium IT corporations, there are some limitations making IT corporations hesitate to adopt it. In this research, we examine the investment strategy and profit for cloud service providers to better understand the business model of cloud services. We find that a cloud service provider with R&D capability will prefer vertical competition rather than partnering with an independent software vendor. In addition, for independent software vendors, maintaining a loose partnership is better than a tight one. Finally, we suggest that an independent software vendor shouldn't support its cloud partner to enhance service security and compatibility, both of which may reduce its overall profit.

Keywords: Durable Goods, Cloud Services, Independent Software Vendor, Software Renting.

1 Introduction

In the market of office software, Google offers its cloud-based tools, known as Google Apps, which include Google Docs and Gmail free to its subscribers for just $50 per user per year. Recently, Microsoft announced Office 365 for cloud services, a subscription web-based cloud service which costs $6 per user per month. From the perspectives of features, ease-of-use, administration, setup, and value, a comparison between both products have been made by a business report, which indicates that the service quality of Office 365 is better than that of Google Apps. Although Office 365 is Microsoft's latest attempt to do battle with Google in the cloud, it is not clear whether the strategy of selling both desktop-based and web-based software can make Microsoft gain more. Some effects of Office 365 in the competitive market have been exposed by the report. For example, Office 2010 isn't outrageously expensive since Microsoft has priced Office 365 [1,2].

Software can be considered a special type of durable goods because it lasts forever with the support of virtual machines, a type of applications can simulate a suitable

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 289–301, 2012.

operating environment in which a variety of software run without the concern of operation systems. Prior studies emphasized a question that manufacturers of durable goods decide how long their products should last. Based on Cournot model [3], a monopoly that sells a durable good in many periods earns less than one that only rents. However, the result cannot be directly applied to study the competition between web-based and desktop-based software because both are different in service quality due to the nature of the Internet and Web-based Interface. Moreover, the uncertainty of future sales also affects the comparison in selling strategy between web-based and desktop-based software.

Prior studies rarely analyzed the impact of cloud services on the selling strategy of an independent software vendor from the perspective of durable goods, the uncertainty in future sales, and service quality (that is, software quality), our research questions in this study are: (1) For a monopolistic independent software vendor, should it sell web-based software before cloud services emerge? (2) After a cloud service provider enters the market, what is the best selling strategy of an independent software vendor if both compete for the same consumer base? Moreover, can the software vendor gain more by simultaneously announcing its desktop-based and web-based software in the market? (3) Should a software vendor convert itself into a cloud service provider when another cloud service provider has entered the market?

2 Literature Review

Prior studies examining Microsoft's behavior is relevant for understanding durable goods markets since software applications are durable. A monopolistic software vendor cannot commit to the event that it will introduce new products that make used applications obsolete too frequently. In addition, a monopolistic software vendor may price low and sell a large quantity in order to deter future entry [4]. Ellison and Fudenberg [5] consider a two-period model to explore the reason for a monopolistic software vendor's incentive to introduce more upgrades than is socially optimal when the upgrade is backward but not forward compatible. Their analysis indicates that consumer heterogeneity leads to this outcome. In addition, the monopolist will want to sell the upgraded product to new consumers. Economides [6] analyzes a model of multi-period monopoly in software markets. Taking into consideration that a monopolistic software vendor stops selling older software when they introduce a replacement model, he shows that software prices may increase over time when new-version software are introduced by the software vendor which are only partially compatible with old-version software.

However, improving software applications so rapidly, a software vendor will find that high-end consumers are tempted to wait for a future new-and-improved version [7]. Indeed, when an improvable software package saturates the market, a software vendor has incentive to release new versions, which may hurt her profit if this is done too frequently. Sankaranarayanan [8] proposes a novel contractual device, which is coined as a Free New Version Rights warranty (free NVR warranty), can help the seller overcome this temptation. Kornish [9] indicates that offering upgrade pricing is not a necessary condition to form an equilibrium pricing strategy for a monopolistic software vendor. In addition, prior results indicate that selling is better than leasing for a competitive market when suppliers of durable goods interact directly with consumers; however, Bhaskaran and Gilbert [10] find a contrary result at the supply chain in which the manufacturer sells her product and then the dealers can either sell or lease it to the final consumer.

Cheng and Koehler [11] model the economic dynamics between an application service provider and its potential customers and show that there exists a unique rational expectation equilibrium under a realistic economies-of-scale assumption. Wu and Banker [12] examine flat fee pricing, pure usage-based pricing, and two-part tariff pricing and suggests that whether consumers are homogeneous or heterogeneous will significantly affect the choice of the best pricing strategy. Chien and Chu [13] indicate network effects may cause that profits from selling software might be higher than from leasing when the products exhibit network effects, which is contrary to the existing literature. Choudhary [14] compare the commonly used perpetual licensing model with a relatively new licensing model called SaaS (Software as a Service) and find that SaaS licensing model leads to greater investment in product development under most conditions. However, most prior studies either consider homogeneous consumers or regard consumer heterogeneity at a monopolistic market. Consequently, we cannot further know the impacts of durable goods (that is, desktop-version software) when cloud services emerge. Although many factors not in our model could have an important influence on our findings, the competition between software vendors and cloud service providers is perhaps the most notable.

3 The Model

Before introducing the model, we first clarify the terms used in this study. For a software application to achieve a certain purpose, there are two different types of a software application: a desktop-version and an online-version. For example, Microsoft Office and CommitCRM are representative instances for desktop-version applications, and consumers pay for them only once and can use them forever as long as they like. Google Apps and Salesforce CRM are representative instances for online-version applications and consumers pay rents to use them on the Internet during a certain period. Based on the two types of the software application, firms selling the software application can be classified into software vendors or cloud service providers. In essence, a software vendor develops and sells its desktop-version application, while a cloud service provider develops and rents its online-version application, which is also known as a cloud service.

We consider the software market in which a software vendor sells a software application to consumers. Due to the nature of durable goods, we model the software market as a two-period game in which the software vendor may sell its product at the two periods with respect to different prices, which are denoted as $p_{s,1}$ and $p_{s,2}$. Obviously, the consumers buying the software application at period 1 take the advantage of using free it at period 2. Accordingly, they can receive more benefit than other consumers purchasing the software application at period 2. We characterize each consumer's preference by θ, which represents its valuation for the software application. Thus, if buying the software application at period 1 and 2, a consumer's utility is given by $U_{s,1}(\theta) = 2\theta - p_{s,1}$ and $U_{s,2}(\theta) = \theta - p_{s,2}$, respectively. Indeed, the software vendor can upgrade its product to raise consumer's valuation for the software application at period 2. In this research, we omit the concern of upgrade pricing and leave it as a future study.

The other approach to sell the software application is to transform it into a cloud service, in which consumers pay a rent to use the online-version application. That is, a software vendor can offer cloud services in which software applications are deployed at its servers and consumers use the service provided by the software applications through the Internet. In fact, the concept of cloud services is not a new idea, which has many parallels with time sharing systems. Instead of purchasing the software application, consumers can share the software application on the Internet without suffering from the expensive cost of ownership. Due to the limitation of bandwidth and security concern, the service quality of the online-version application is lower than that of the desktop-version application, so we denote δ as the difference in service quality between the two versions of the application. Given the rent $p_{c,1}$ and $p_{c,2}$ paid by consumers at different periods, consumer's utilities are given by $U_{c,1}(\theta) = \delta\theta - p_{c,1}$ and $U_{c,2}(\theta) = \delta\theta - p_{c,2}$, respectively. At each period, a consumer can pay a rent to use the online-version application, or pay a rent at period 1 to use the online-version application but buy the desktop-version application at period 2. In the latter case, his/her utility is given by $U_{c,1}(\theta) + U_{s,2}(\theta)$.

In order to understand why software vendors, such as Microsoft, didn't sell its online-version application before cloud services emerge, we examine the scenario in which the software vendor can transform its software application into a cloud service and then offer the online-version and desktop-version applications simultaneously. At period 1, all consumers may choose to pay $p_{s,1}$ to buy the desktop-version application, pay $p_{c,1}$ to use the online-version application, or buy nothing. Likewise, all consumers at period 2 may make the same purchasing decisions except the consumers buying the desktop-version application at period 1. Consequently, solving $U_{s,2}(\theta) = U_{c,2}(\theta)$ can yield $\hat{\theta}_{sc,2} = (p_{s,2} - p_{c,2})/(1-\delta)$, which represents that a consumer with preference $\hat{\theta}_{sc,2}$ receives the same utility between using the desktop-version and online-version applications at period 1. Similarly, we can derive $\hat{\theta}_{sc,1} = (p_{s,1} - p_{c,1} - p_{s,2})/(1-\delta)$ by solving $U_{s,1}(\theta) = U_{c,1}(\theta) + U_{s,2}(\theta)$, which represents the same result at period 2. Moreover, by solving $U_{c,1}(\theta) = 0$ and $U_{c,2}(\theta) = 0$ separately, we can derive $\hat{\theta}_{cn,1} = p_{c,1}/\delta$ and $\hat{\theta}_{cn,2} = p_{c,2}/\delta$, which represents consumers with preference $\hat{\theta}_{cn,1}$ and $\hat{\theta}_{cn,2}$ are indifferent between paying a rent to use the online-version application or not paying at period 1 and 2, respectively.

Due to the uncertainty in future sales, the revenue at period 2 is discounted by the exogenous variable γ . Consequently, the software vendor's pricing decision can be derived by solving the following optimization problem:

$$\underset{p_{s,1},p_{s,2},p_{c,1},p_{c,2}}{Max} \quad \pi_m = p_{s,1}\left(1-\hat{\theta}_{sc,1}\right)+\gamma\cdot p_{s,2}\left(\hat{\theta}_{sc,1}-\hat{\theta}_{sc,2}\right)+p_{c,1}\left(\hat{\theta}_{sc,1}-\hat{\theta}_{cn,1}\right) \\ +\gamma\cdot p_{c,2}\left(\hat{\theta}_{sc,2}-\hat{\theta}_{cn,2}\right) \tag{1}$$

Proposition 1
Before cloud services emerge, a monopolistic software vendor has no incentive to rent an online-version application with less functionality or lower security than its desktop-version application.

Our results recognize software vendor's selling strategy in a monopolistic market, in which its revenue cannot be enhanced by selling desktop-version and online-version applications simultaneously. Moreover, if online-version application's marginal operation cost is higher than desktop-version application's, the result remains although our analysis omits the concern of marginal operation cost. Before Google Apps enters the market of word processing software, Microsoft only sells its desktop-version application because an online-version application is a substitute for a desktop-version application and bears a higher operation cost due to servers and bandwidth. A monopolistic software vendor doesn't want to see the competition between its desktop-version and online-version applications. In addition, the nature of durable goods enables it sell a high-price software at period 1 and a low-price software at period 2, which is better than selling an online-version application with higher operation cost at both periods.

4 The Impact of Cloud Services

Now, we consider the scenario in which a cloud service provider enters the market and rents an online-version application, and a software vendor still sells its desktop-version application. For example, Microsoft sells its Office 2007, whereas Google rents its Google Apps. By backward induction, we solve best response prices at period 2 and then the equilibrium prices at period 1. At period 2, their revenue functions are given by $\pi_{s,2} = p_{s,2}\left(\hat{\theta}_{sc,1}-\hat{\theta}_{sc,2}\right)$ and $\pi_{c,2} = p_{c,2}\left(\hat{\theta}_{sc,2}-\hat{\theta}_{cn,2}\right)$, in which the symbol s and c represents a software vendor and a cloud service provider, respectively. By solving FONC for both firms simultaneously, we derive their best response prices as follows.

$$p_{s,2} = 2\left(p_{s,2}-p_{c,2}\right)/\left(8-\delta\right),\;\; p_{c,2} = \delta\left(p_{s,1}-p_{c,1}\right)/\left(8-\delta\right) \tag{2}$$

At period 1, in addition to the revenue at this period, both firms have to consider their future sales at period 2, which are given by $\pi_s = \pi_{s,1}+\gamma\cdot\pi_{s,2}$ and $\pi_c = \pi_{c,1}+\gamma\cdot\pi_{c,2}$, where $\pi_{s,1} = p_{s,1}\left(1-\hat{\theta}_{sc,1}\right)$ and $\pi_{c,1} = p_{c,1}\left(\hat{\theta}_{sc,1}-\hat{\theta}_{cn,1}\right)$. By utilizing (2), we can have

$$p^{*}_{c,1}=\frac{\Delta(8-\delta)^{2}(1-\delta)}{(6-\delta)(8-\delta)+(1-\Delta)((6-\delta)(8-\delta)-16\gamma)} \tag{3}$$

$$p^{*}_{s,1}=\frac{(8-\delta)^{2}(1-\delta)}{(6-\delta)(8-\delta)+(1-\Delta)((6-\delta)(8-\delta)-16\gamma)}\ , \tag{4}$$

where $\Delta\equiv\dfrac{\delta(8-\delta)(6-\delta)-2\gamma\cdot\delta^{2}}{2(8-\delta)^{2}(1-\delta)+2\delta(8-\delta)(6-\delta)-2\gamma\cdot\delta^{2}}$.

Based on equilibrium prices at period 2, which are derived by plugging (3) and (4) into (2), we can have their revenues at each period as follows:

$$\pi_{s,1}=(8-\delta)^{2}(1-\delta)\frac{(6-\delta)(8-\delta)-16(1-\Delta)\gamma}{\{(6-\delta)(8-\delta)+(1-\Delta)((6-\delta)(8-\delta)-16\gamma)\}^{2}} \tag{5}$$

$$\pi_{c,1}=\frac{\Delta(8-\delta)^{2}(1-\delta)\{(6-\delta)(8-\delta)-\Delta(8-\delta)\{(6-\delta)+(8-\delta)(1-\delta)/\delta\}\}}{\{(6-\delta)(8-\delta)+(1-\Delta)((6-\delta)(8-\delta)-16\gamma)\}^{2}} \tag{6}$$

$$\pi_{s,2}=\frac{8(1-\delta)(8-\delta)^{2}(1-\Delta)^{2}}{\{(6-\delta)(8-\delta)+(1-\Delta)((6-\delta)(8-\delta)-16\gamma)\}^{2}} \tag{7}$$

$$\pi_{c,2}=\frac{\delta(1-\delta)(8-\delta)^{2}(1-\Delta)^{2}}{\{(6-\delta)(8-\delta)+(1-\Delta)((6-\delta)(8-\delta)-16\gamma)\}^{2}} \tag{8}$$

Lemma 1

The demand of the desktop-version application at period 1 (at period 2) increases (decreases) in the uncertainty in sales at period 2, respectively. Formally, $\partial D_{s,1}/\partial\gamma<0$ and $\partial D_{s,2}/\partial\gamma>0$.

The demand of the online-version application at each period decreases in the uncertainty in sales at period 2. Formally, $\partial D_{c,1}/\partial\gamma>0$ and $\partial D_{c,2}/\partial\gamma>0$.

Based on the observation of the demands, which is demonstrated in Figure 1, we can find that the software vendor will sell its product according to the future sales. That is, the software vendor can sell more at the beginning if it is optimistic about future sales. However, the same results cannot be found for the cloud service provider. This indicates that the cloud service provider cannot give up its revenue at period 1 but gain more at period 2 because the software vendor utilizes the advantage of high-quality and durable goods and forces the cloud service provider to adopt a conservative selling strategy.

Proposition 2

For the desktop-version application, no matter how the cloud service provider enhances its service quality, the sales at period 1 is more than that at period 2.

The software vendor's revenue at period 1 may increase or decrease with the uncertainty in sales. Moreover, the software vendor' revenue at period 2 and the cloud service provider's revenue at each period decrease in the uncertainty in sales at period 2.

If the difference in service quality between the online-version and desktop-version applications is significant, the consumers leasing the online-version application at period 1 may purchase the desktop-version application at period 2. However, if the difference is slight, the consumers adopting the online-version application at period 1 may continue to lease it at period 2.

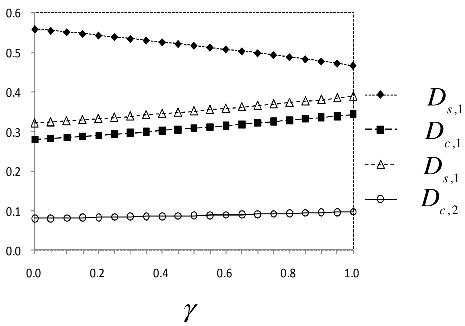


Fig. 1. The demand of the desktop-version and online-version applications

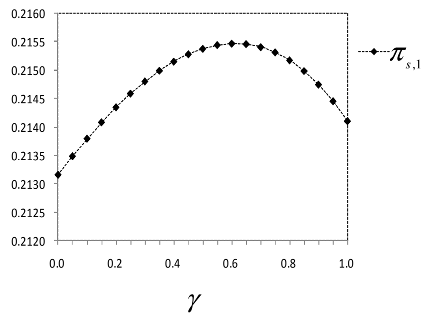


Fig. 2. The software vendor's revenue at period 1

Because of the effect of the uncertainty in sales at period 2, the software vendor will sell more at period 1 than period 2. In addition, the portion of sales at period 1 will increase in the uncertainty in sales at period 2. Furthermore, the software vendor can gain higher revenue by adjusting its pricing strategy as follows. Because of the features of durable goods, although its demand at period 1 decreases if charging a higher price, the software vendor can sell more at period 2. However, the pricing strategy works only when the uncertainty in sales at period 2 slackens off. Figure 2 shows that the impacts of the uncertainty in sales at period 2 on its revenue at period 1 will be most serious when $\gamma = 0$ and $\gamma = 1$. When $\gamma = 0$, both firms are pessimistic for future sales so they face an intensive competition at period 1. When the expectation of future sales becomes less pessimistic, the cloud service provider can relax its pricing decision at period 1 due to the future sales at period 2. In this case, the software vendor's revenue at period 1 can increase in the expectation of future sales. However, when both firms are optimistic for future sales, the software vendor will largely raise its price at period 1 because the sales at period 2 is so optimistic that it can gain more by raising its price at period 2 to offset its decline revenue at period 1.

Moreover, we also recognize that consumers may continue to use cloud services between the two periods if they can perceive that the service quality of the online-version application approaches that of the desktop-version application. That is, with the progress of online-version software, more and more consumers may stay at cloud services. However, if the service quality of the online-version application is far from the service quality of the desktop-version application, consumers using cloud services may be attracted by low-price desktop-version software in the future. This finding indicates that a software vendor has to understand the difference between its product and emerging cloud services. As long as the difference of service quality can be appreciated by consumers, the threat that cloud services change market structure is not convincible; however, if the service qualities of cloud services approach its desktop products, product differentiation or changing selling strategy is a must.

4.1 Is It Time to Change?

Recently, Microsoft, the dominator of word processing software, has rented its Office products to invade cloud services market. The solution of its online-version word processing software is known as Office 356. Thus, we examine the scenario in which a software vendor sells the desktop-version and online-version applications simultaneously when a cloud service provider enters the software market. For convenience, we assume that the quality of the online-version application sold by the software vendor is between the online-version application sold by the cloud service provider and the desktop-version application sold by the software vendor. Accordingly, we denote $\delta \le \overline{\delta} \le 1$ which captures the difference between these software applications.

From the perspective of software quality, if the software vendor makes its online-version application look like its desktop-version application, this will lead to an internal competition that hurts its profit. If the software vendor downgrades its online-version application so that consumers are indifferent between adopting the two online-version applications, the cloud service provider lowers its rent in response that decreases its and software vendor's revenue. Obviously, the software vendor is not willing to incur the outcomes. Although we can examine a general case where $\delta < \overline{\delta} < 1$, this may lead to intractable models if no other assumption is attached. So, we only study the special case where $\overline{\delta} = (1+\delta)/2$. The analytical result derived in the special case and our intensive simulations for the general case support the outlook that the software vendor should give up selling both online-version and desktop-version applications. The reason for that is as follows. If the software vendor sells an online-version application with the same functionality as its desktop-version application, this leads to internal competition, in which hurts its revenue of the desktop-version application; in addition, the income of its online-version application cannot compensate the great loss. On the other hand, if the software vendor sells its online-version application with the same functionality as its competitor's online-version application, this leads to external competition, in which both firms lower their rents so as to decrease both firms' revenues. If the service quality of the online-version application sold by the software vendor is between the other two products, both the influence of internal and external competition becomes weak but the impacts of competition still exists, which causes that the software vendor cannot gain more by selling the additional cloud product.

Trajectory 1
When cloud services emerge, a software vendor hardly gains more by selling desktop-version and online-version applications simultaneously.

Although our result indicates that an online-version application sold by a software vendor has poor prospects, it is possible that the software vendor gains more if it gives up its revenue of the desktop-version application. Subsequently, we examine the scenario in which both software vendor and cloud service provider rent their online-version applications, but the sale of the desktop-version application is stopped. Because the impacts of the features of durable goods disappear in the scenario, we can study a one-period model instead of the two-period model. Consequently, we have the new indifference points, $\hat{\theta}_{sc} = \left(p_s - p_c\right)/\left(\overline{\delta} - \delta\right)$ and $\hat{\theta}_{cn} = p_c/\delta$, which are given by solving $\overline{\delta}\theta - p_s = \delta\theta - p_c$ and $\delta\theta - p_c$, respectively. Consequently, their revenues at each period are given by $\pi_{s,i} = 4\overline{\delta}^2\left(\overline{\delta} - \delta\right)/\left(4\overline{\delta} - \delta\right)^2$ and $\pi_{c,i} = \delta\overline{\delta}\left(\overline{\delta} - \delta\right)/\left(4\overline{\delta} - \delta\right)^2$.

Proposition 3
When the uncertainty in future sales is not too high, the software vendor can sell cloud services but give up its income of the desktop-version application; however, if the uncertainty in future sales is significant, the software vendor should continue to sell its desktop-version application and pass up the chance of stepping into cloud services.

We use Figure 3 and 4 to demonstrate the above findings. Both figures show that both firms' revenues can increase when the uncertainty in future sales goes down. Moreover, when the uncertainty in future sales is significantly low, the software vendor should transform its desktop-version application into an online-version application and give up the business of the desktop-version application. If the uncertainty in future sales is significantly high, software vendor should sell its desktop-version application only. In addition, software vendor's best selling strategy at the two extreme cases is beneficial to the cloud service provider. In order to clarify the impacts of the sales uncertainty and service quality on the selling strategy of the software vendor, we use a numerical example to demonstrate the best selling strategy among all possible cases, which is shown in Figure 5.

Figure 5 exhibits an interesting result that software vendor's best selling strategy is affected by its competitor's service quality. When the value of γ is between a specific range (for example, 0.6 and 0.65), selling its desktop-version application is better than renting its online-version application when its competitor's service quality is poor or extremely high just like the quality of its desktop-version application; otherwise, the software vendor can only rent its online-version application to gain a higher revenue. In this range, if its competitor's service quality is inferior, selling the desktop-version application is the best selling strategy for the software vendor because the impact of competition on the software vendor's revenue is so slight that it can constitute the sales of the desktop-version application. If its competitor's service quality approaches the quality of its desktop-version application, this leads to intensive competition; consequently, the software vendor should still sell its desktop-version application because the nature of durable goods can achieve product differentiation to a certain extent, which is beneficial to its revue.

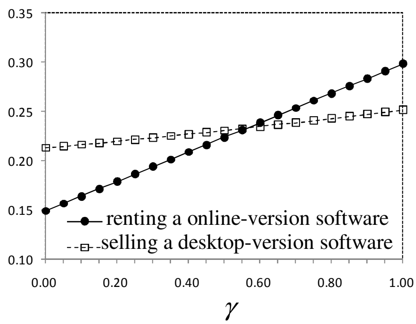


Fig. 3. The profit for the software vendor π_s ($\delta = 0.5$, $\overline{\delta} = 0.95$)

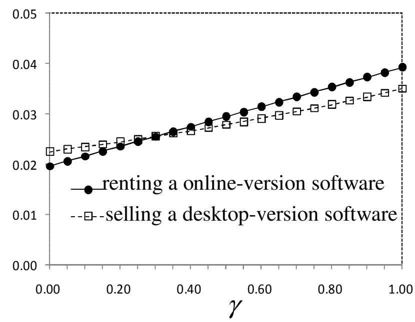


Fig. 4. The profit for the cloud service provider π_c ($\delta = 0.5$, $\overline{\delta} = 0.95$)

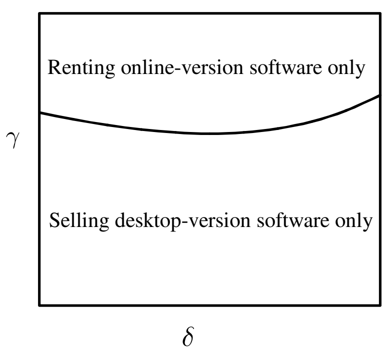


Fig. 5. The best selling strategy for the software vendor ($\overline{\delta} = 1$)

5 Conclusion

In this research, we study the selling strategy of an independent software vendor under different market structures. Based on the nature of durable goods, we utilize a two-period model to examine the impact of cloud services on the software vendor's pricing decisions at different periods. We find that a monopolistic software vendor has no incentive to rent an online-version application with lower service quality because of the competition between its desktop-version and online-version applications plus the impacts of durable goods. As cloud services emerge, a software vendor may set a high price at period 1 to attract those consumers focusing on software quality and a low price at period 2 to attract consumers adopting cloud services at period 1. One of the practical evidences for a low price can be found at recent price change history for the Adobe Photoshop Elements 8.0 for Windows[1] (A Image editing application for Windows). However, the software vendor may have trouble if it sells desktop-version application and rents online-version application simultaneously, because its competitor, a cloud service provider, may decrease its price in response that decreases both firms' revenues.

Subsequently, we examine the scenario in which a software vendor addresses it as a cloud service provider and competes with the other cloud service provider. In this case, we find that both uncertainty of future sales and its competitor's service quality affects its reforming decision. If the uncertainty of future sales is sufficiently high, selling a desktop-version application is better than renting an online-version application. The opposite may hold true when the uncertainty of future sales is sufficiently low. Moreover, given the uncertainty of future sales, the software vendor's reforming decision may be affected by its competitor's service quality. Under certain conditions, the software vendor should sell its desktop-version application only when its competitor's service quality is poor or extremely high; however, renting an online-version application is better than selling its desktop-version application in the other cases.

There are some limitations in this study, which may affect our analytical results. First, we don't involve cloud service provider's marginal operation cost in this study. In the case that both firms have operation costs but the marginal operation cost of selling a desktop-version application is lower than the marginal operation cost of renting an online-version application, we may normalize the software vendor's marginal operation cost as zero; however, cloud service provider's marginal operation cost still exists. Fortunately, although cloud service provider's marginal operation cost will significantly affect its rent, it only affects the quantitative results rather than qualitative results of the software vendor; moreover, this concern will make our model intractable. Thus, we only emphasize the impact of the emerging of cloud services on the software vendor's selling strategy and reforming decisions in this study. Second, the number of consumers in the two periods can be different. That is, some incomers may appear at the second period. In addition, the number of consumers purchasing the desktop-version application will form a positive network effect so as to affect other consumers' willingness-to-pay at the second period. Finally, we don't consider the issue that software publisher's incentive to invest in software quality. All of the concerns can be studied in the future.

Acknowledgments. The authors would like to thank the two anonymous reviewers for their insightful comments and helpful suggestions. This research was supported by the National Science Council of Taiwan (Republic of China) under the grant NSC 100-2410-H-266 -005.

---

[1] http://www.photoprice.ca/product/03197/Adobe-Photoshop-Elements-8.0-for-Windows-price.html

References

1. Office 365 vs. Google Apps: The InfoWorld review, http://www.infoworld.com/d/cloud-computing/office-365-vs-google-apps-the-infoworld-review-447
2. Microsoft launching Office 365 on June 28, http://news.cnet.com/8301-10805_3-20072888-75/microsoft-launching-office-365-on-june-28
3. Carlton, D.W., Perloff, J.M.: Decision Making over Time: Durability. Modern Industrial Organization, 4th edn. Addison-Wiley, New York (2005)
4. Waldman, M.: Durable Goods Theory for Real World Markets. The Journal of Economic Perspectives 17, 131–154 (2003)
5. Ellison, G., Fudenberg, D.: The Neo-Luddite's Lament: Excessive Upgrades in the Software Industry. The RAND Journal of Economics 31, 253–272 (2000)
6. Economides, N.: Durable Goods Monopoly with Network Externalities with Application to the PC Operating Systems Market. Quarterly Journal of Electronic Commerce 1, 1–13 (2000)
7. Dhebar, A.: Durable-Goods Monopolists, Rational Consumers, and Improving Products. Marketing Science 13, 100–120 (1994)
8. Sankaranarayanan, R.: Innovation and the Durable Goods Monopolist: The Optimality of Frequent New-Version Releases. Marketing Science 26, 774–791 (2007)
9. Kornish, L.J.: Pricing for a Durable-Goods Monopolist under Rapid Sequential Innovation. Management Science 47, 1552–1561 (2001)
10. Bhaskaran, S.R., Gilbert, S.M.: Implications of Channel Structure for Leasing or Selling Durable Goods. Marketing Science 28, 918–934 (2009)
11. Cheng, H.K., Koehler, G.J.: Optimal pricing policies of web-enabled application services. Decision Support System 35, 259–272 (2003)
12. Wu, S., Banker, R.D.: Best Pricing Strategy for Information Services. Journal of Association for Information Systems 11, 339–366 (2010)
13. Chien, H.-K., Chu, C.Y.C.: Sale or Lease? Durable-Goods Monopoly with Network Effects. Marketing Science 27, 1012–1019 (2008)
14. Choudhary, V.: Comparison of Software Quality Under Perpetual Licensing and Software as a Service. Journal of Management Information Systems 24, 141–165 (2007)

Appendix

Proof of Proposition 1

Solving $\partial\pi_m / \partial p_{c,2} = 0$ yields $\delta p_{s,2} = p_{c,2}$, which implies that the monopolistic software vendor doesn't rent an online-version application at period 2. So, its optimization problem can be rewritten as follows:

$$\pi_m = p_{s,1}\left(1-\hat{\theta}_{sc,1}\right) + \gamma \cdot p_{s,2}\left(\hat{\theta}_{sc,1} - \hat{\theta}_{sc,2}\right) + p_{c,1}\left(\hat{\theta}_{sc,1} - \hat{\theta}_{cn,1}\right)$$

Then, solving $\partial\pi_m / \partial\hat{\theta}_{sc,1} = 0$, $\partial\pi_m / \partial\hat{\theta}_{sc,2} = 0$, and $\partial\pi_m / \partial\hat{\theta}_{cn,1} = 0$ yields $\hat{\theta}_{cn,1} = 1/2$, $\hat{\theta}_{sc,1} = \left((1-\delta)-(1-\gamma)\hat{\theta}_{sc,2}\right) / \left(2(1-\delta)\right)$, and $\hat{\theta}_{sc,2} = \left(1-(1-\gamma)\hat{\theta}_{sc,1}\right) / (2\gamma)$.

Because of the boundary condition $\hat{\theta}_{sc,1} \geq \hat{\theta}_{cn,1}$, we can have $\hat{\theta}_{sc,1} \geq 1/2$. Therefore, $\hat{\theta}_{sc,1} = \hat{\theta}_{sc,2} = 1/2$ when $\gamma \leq 1$.

Proof of Proposition 2

1. Because $\Delta = \dfrac{\delta(8-\delta)(6-\delta) - 2\gamma\cdot\delta^2}{2(8-\delta)^2(1-\delta) + 2\delta(8-\delta)(6-\delta) - 2\gamma\cdot\delta^2}$, we can observe $0 \le \Delta \le 0.5$. Moreover, $\partial\Delta/\partial\delta > 0$ and $\Delta = 0$ when $\delta = 0$ holds. Notice that $D_{s,1} > D_{s,2}$ if and only if $\gamma < \dfrac{(8-\delta)((6-\delta)-4(1-\Delta))}{16(1-\Delta)}$. Because $\dfrac{\partial}{\partial\Delta}\dfrac{(8-\delta)((6-\delta)-4(1-\Delta))}{16(1-\Delta)} > 0$ and $\left.\dfrac{(8-\delta)((6-\delta)-4(1-\Delta))}{16(1-\Delta)}\right|_{\delta=0} = 1$, we can infer that $\gamma < \dfrac{(8-\delta)((6-\delta)-4(1-\Delta))}{16(1-\Delta)}$ for any δ.
2. We observe $\partial\pi_{s,2}/\partial\gamma > 0$, $\partial\pi_{c,1}/\partial\gamma > 0$, and $\partial\pi_{c,2}/\partial\gamma > 0$ directly. Moreover, the sign of $\partial\pi_{s,1}/\partial\gamma$ is equivalent to the sign of $\dfrac{\Delta\Theta - 16(1-\Delta)\gamma}{\{\Theta + (1-\Delta)(\Theta - 16\gamma)\}^3}$. Consequently, we have $\partial\pi_{s,1}/\partial\gamma > 0$ when $\gamma < \hat{\gamma}$ and $\partial\pi_{s,1}/\partial\gamma < 0$ when $\gamma > \hat{\gamma}$, where $\hat{\gamma} = \Delta\Theta/(6(1-\Delta))$.
3. Focusing on δ, we have the result by comparing $\hat{\theta}_{sc,2}$ and $\hat{\theta}_{cn,1}$ directly.

Proof of Proposition 3

When cloud services emerge, if the software vendor rents an online-version application, its revenue at the two periods is given by $\pi_s = \pi_{s,1} + \gamma\pi_{s,2}$. If the software vendor sells an desktop-version application, its revenue at the two periods is given by $\pi_s = \dfrac{(1-\delta)(8-\delta)^2\{(6-\delta)(8-\delta) - (1-\Delta^2)\gamma\}}{\{(6-\delta)(8-\delta) + (1-\Delta)((6-\delta)(8-\delta) - 16\gamma)\}^2}$. We complete the proof by comparing both revenues.

SLA Based Dynamic Provisioning of Cloud Resource in OLTP Systems

Xiaoqiu Qiu, Markus Hedwig, and Dirk Neumann

Information Systems Research,
Albert-Ludwigs-Universität Freiburg im Breisgau
Kollegiengebäude II, Platz der Alten Synagoge
79085 Freiburg, Germany
{xiaoqiu.qiu,markus.hedwig,dirk.neumann}@is.uni-freiburg.de

Abstract. In the era of cloud computing, an increasing amount of services are moving online. Also increasing is the amount of cloud resource to power these services. Among these modern online transactions, many belong to the category of Online Transaction Processing (OLTP), which can be processed with predictable time and resource. However, with a large user base and fluctuated usage patterns, providing OLTP services efficiently remains a major challenge. In this paper we present an online algorithm that solves for a cost-minimizing provision scheme under fluctuated user requests, constrained by a tail-distribution-based Service Level Agreement (SLA), and incorporated with Neural Network prediction. Experiment shows that the algorithm delivers significant savings in provision, and outperforms a simple look-forward provision plan with the same SLA compliance.

Keywords: Service Level Agreement, SLA, OLTP, Cloud Computing.

1 Introduction

With the emerging trend of Cloud Computing and its increasing application in corporate IT, the perception of computing has changed. Nowadays, enterprise information systems are transformed to service systems, which are detached from the physical hardware and operated in virtual environments. In line with this trend, more and more services are provided by third party providers [1] and provided online to the customer. Though this enables the service consumer to simplify his own computing environments, service provider have to expand their IT continuously. On one hand they have to fulfill the increasing service demand; on the other hand they need to contribute to the increasing service complexity. As a consequence, more and more energy is consumed to power these services.

In this paper, we focus on a certain class of service systems: Online Transaction Processing (OLTP) systems, where the workload is relatively simple with predictable processing time [2]. However, users are impatient and may walk away if their requests are not answered in time. Famous representatives of this class are online banking and e-commerce. A single OLTP transaction may cost negligible

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 302–310, 2012.

computational power to handle, however, with an explosive amount of user and user requests, simple transactions can add up to a giant computational task that requires new data centers to process [3]. Besides, many of these datacenter-workloads display a cyclical pattern [4], resulting a server utilization level as low as 20-30% [5], and 10% servers unused around the world [6]. Considering that energy-related costs remain a significant account of overall data center expenditures and the fastest-growing part [7], low utilization of servers creates a high yet unnecessary cost both economically and environmentally.

The high volatility in user demand makes it ideal to provide cloud resource dynamically. Recent technology advancement in virtualization enables service providers to purchase computing capacity from commercial cloud providers (e.g. Amazon, IBM) with ease. Without the hassle related to physical servers, these virtual servers can be operated in a much leaner and agiler manner without compromising in SLA performance. Such possibility of reducing server usages yields a potential reduce in power and cooling costs, aligns the profit maximization motive with the environmental awareness.

In this paper we propose an online algorithm for dynamic provisioning of cloud resources based on past SLA performances in OLTP systems. We use a simple Neural Network to predict upcoming user demands, and analyze the prediction errors. We show that as the SLA contract extends, our solution yields an asymptotically optimal provision scheme, with increasing tolerance of prediction errors. The algorithm also outperforms simple look-forward dynamic provision algorithm that based on the same Neural Network predicted series.

The remainder of the paper is structured as follows. In the second section we briefly review the current literature on SLA and dynamic provision of cloud resources. Section 3 states a mathematical formation of the problem. Section 4 illustrates our solutions. The computational results based on Wikipedia data [4] are analyzed in Section 5. Section 6 summarizes the paper and concludes with some remarks.

2 Related Works

For a service provider, lower level of service means lower cost to operate, but this often leads to a bad user experience. With SLA that explicitly states service levels and the corresponding prices or penalties, service providers can better optimize their operation strategy without the risk of dampening their reputation in the business.

As many recent studies indicate that energy-related costs constitute a significant portion of overall data center expenditures [7][8], the potential to save such costs brings both economical and environmental benefits. The measures to tackle this cost-saving problem take approaches from the economic side, to improve SLA, as well as from the information science side, to integrate better dynamics into the system.

On designing a better SLA, mathematical approximation and simulations are employed for consistency checks and run-time optimization [9]; new type of revenue-penalty function, which has a linear transition between revenue and penalty states instead of a discontinued jump [10], is expected to give service providers an incentive to provide decent service even when the ideal SLA requirements cannot be fulfilled. Finally, SLA contract itself can also embrace more dynamics, e.g., moving towards electronic contracts and autonomous negotiation and brokerage [11].

As a service system usually consists of accepting tasks, getting resources and processing jobs (allocating resources), the dynamics of the system can be improved through Admission Control, Dynamic Procurement of cloud resource, and Job Scheduling.

From Admission Control aspect, researches usually utilize prediction of user requests, queuing theory as well as different Quality of Service constraints for prioritizing incoming tasks and picking an optimized subset to accept [12]. Novel and improved queuing models are developed to better handle the evolving and unpredictable user requests [13]. New mechanisms are designed to handle emerging situations, e.g. a sudden rush of user demand during major events [14].When the model contains a cluster of servers, admission control is also incorporated with server resource allocation [15].

The idea of dynamic purchase of external cloud resource gains more attention in recent years due to increasing popularity of computation outsourcing. Depending on price differentiation of various virtual resources, finding an optimized procurement portfolio is of crucial importance [16]. Other insights come from earlier research on grid computing, which attempts to utilized unreliable public resources to boost system performance [17].

Finally the aspect that draws most attention is dynamic job scheduling. The idea is to construct a task allocator that assigns suitable computing resources to specific user requests. With highly fluctuated workflows, the allocator is also able to power on/off servers dynamically [18][22]. User requests considered in job scheduling are usually tasks with variable or even unpredictable processing time [19]. On the other hand, multiple constraints are incorporated into analysis, ranging from system configuration [20] to data centers' locations [21]. With simulation, scheduling policies are refined dynamically [9].

Our work differentiates from previous studies that we limit our focus in the simpler OLTP workflows with predictable processing time, so that we can analyze them with a stock model, instead of a flow model where tasks require intensive computational power and unpredictable processing time. Since in a stock model user requests do not carry onto the next period, it captures the time constraint aspect of OLTP workflows well.

3 Model for the SLA Constrained Dynamic Provisioning Problem

3.1 Assumptions

Our model deals with OLTP workloads that require predictable processing resource and time. Whenever available resource falls short of incoming requests, the excess requests are counted as loss. We adopt a tail-distribution-based Service Level Agreement [22], which checks the number of unanswered requests (loss) across the SLA lifecycle. If the total loss exceeds the threshold percentage (e.g. 5%), SLA is violated, and the service provider has to pay a penalty to its customers.

The objective of the model is to minimize resource provision without violating SLA at the end of its lifecycle. Due to the high volatility in incoming requests and the risk of prediction error, losses can be greater than the threshold level in some periods, ideally, the algorithm would over-provide resource in the coming periods to compensate previous under-provisions, and try to achieve the SLA target eventually.

Considering that cloud resources are normally traded in fixed quantity (e.g. a virtual instance [1]), cloud resources in the model are measured in Resource Blocks, which are comparably larger than the resource to satisfy one user demand. The size of a Resource Block states how many user requests can be handled within one period. Then user demands are normalized by the size of Resource Blocks, with the level of provision being discrete integers. Besides, the feature of a Time-Block is also considered: the level of provision stays constant for certain periods and can only be changed when Time-Block alters. The size of a Time-Block can be one, the case where provision level is flexible anytime.

Finally we assume that over-provision does not earn higher revenue from users. Under this assumption, intuitively the best operation mode would be exactly satisfying the user demand with boundary SLA compliance, a lower provision triggers the penalty while a higher provision brings additional cost.

3.2 Model Setup

Consider a total of n periods, the time series of incoming user requests Q, provision P, and unsatisfied user requests counted as loss L. Since resource is provided in blocks, provision P should be a non-negative integer series:

$$\text{User requests} \qquad Q = \{q_1, q_2, \cdots, q_n\}, q_i \in R^+ \tag{1}$$

$$\text{Provision} \qquad P = \{p_1, p_2, \cdots, p_n\}, p_i \in N^0 \tag{2}$$

$$\text{Loss} \qquad L = \{l_1, l_2, \cdots, l_n\}, l_i \in R^+, l_i = \max\{0, q_i - p_i\} \tag{3}$$

$$\text{Time Block} \qquad p_{i+1} = p_i, \forall i \in [1, n), (i+1) \bmod k \neq 0 \tag{4}$$

In our model, the Service Level Agreement states the percentage a that service provider should answer among all user requests. The unanswered requests are counted as loss, and if the loss rate of all requests exceeds the threshold rate 1-a, SLA is violated.

$$\text{SLA Constraint} \qquad \frac{\sum_{i=1}^{t} l_i}{\sum_{i=1}^{t} q_i} = \frac{\sum_{i=1}^{t-1} l_i + l_t}{\sum_{i=1}^{t-1} q_i + q_t} \leq 1 - a \tag{5}$$

$$\text{Objective Function} \qquad \min \sum_{i=1}^{n} p_i \tag{6}$$

Denote the threshold loss that is about to violate SLA as l_t^*, we have

$$\text{Threshold Loss} \qquad l_t^* = (1-a)\cdot(\sum_{i=1}^{t-1} q_i + q_t) - \sum_{i=1}^{t-1} l_i \tag{7}$$

In the baseline model we take current period user requests qt as known, there is no risk involved and hence it is a deterministic model. Assume a unit cost model where cost of provision linearly depends on the amount of service provided. Intuitively the profit maximizing operation point would be to answer exactly the threshold level of requests. The only constraint that prevents the provider to reach this exact SLA threshold point would be resource blocks. Thus the best operation point lies between the threshold level α·100% and the level that an overprovision of exactly one unit of cloud resource would lead to.

Then we consider a more realistic setup where current period demand q_t is not known, but predicted from past periods. The rest of the model remains the same as (1)-(7).

4 Solutions to the SLA Constrained Dynamic Provision Problem

4.1 Online Deterministic Solution

In the deterministic solution all period's user requests are taken as known. The algorithm uses binary search to find the lowest possible provision for every period, without violating SLA. The starting interval is the lowest and highest demand in each period, times the SLA threshold respectively, as these yield the lower and upper bound of the threshold loss rate. During the binary search, the algorithm checks the actual loss rate to ensure the target SLA compliance.

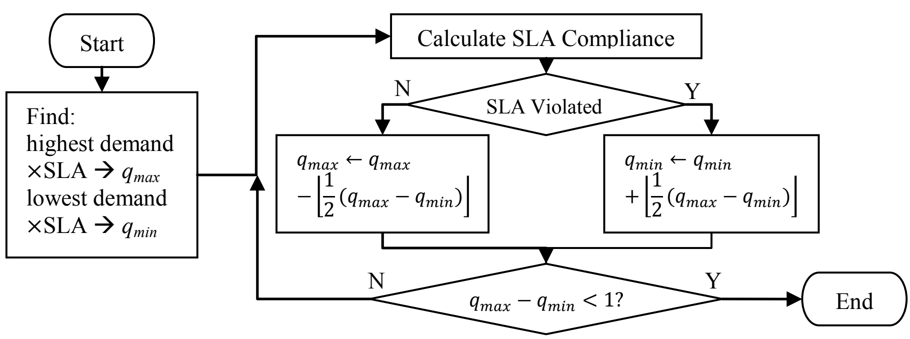


Fig. 1. Process of the Deterministic Algorithm in one Time-Block

At period t, the difference between achieved and target rate is bounded by $1/\sum_{i=1}^{t} q_i$. Because the actual loss rate is calculated by (5), if the difference between the threshold loss l_t* and the achieved loss l_t is larger than 1 (one unit of resource block), the service provider can be better off by decreasing 1 unit of resource block provided without violating the SLA requirement. Clearly, as the contract extends, the difference between achieved SLA compliance and the target rates would reduce. As the number of periods goes to infinity, the bound approaches zero: a service provider can deliver precisely the threshold SLA performance with a long contract time-span. The deterministic solution is asymptotically optimal.

4.2 Online Solution with Demand Prediction

In reality, a service provider has to provide the service p_t before knowing the current period demand q_t. Therefore we need to predict the incoming user requests and make provision decisions based on the prediction and previous SLA compliance.

Here a neural network is employed for prediction. In the solution we just replace the current period request with the predicted one, therefore, other prediction methods can be applied as well. Consider that the current period user requests often depend on previous periods with clear pattern, we chose a non-linear Auto-regressive type Neural Network from MatLab Time Series Tool [23], with 20 neurons and 20 period delays. The data is divided into 50% for network training, 15% for validating the fitted network, and the rest 35% for testing. The rest setup is MatLab default: a random data division, Levenberg-Marquardt training rule, and using Mean Squared Error (MSE) as performance measure. The generalization of the network stops when MSE of the validating set stops converging.

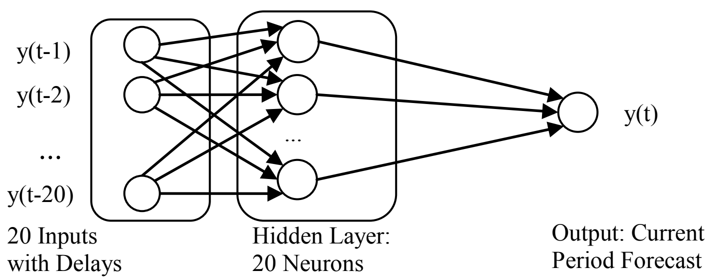


Fig. 2. Neural Network Structure

In the deterministic model, the SLA constraint is calculated by (5), now that the current period demand is unknown, it is replaced by the predicted demand q_t*. When there is more than one period in the time block, all unknown user requests are replaced by predicted demands. The rest of the algorithm (binary search for the threshold provision) remains the same as in 4.1.

5 Computational Result

5.1 Our Solutions Compared to a Simple Full Compliance Algorithm

Both algorithms described in 4.1 and 4.2 are tested with Wikipedia data [4] of hourly page counts from January to April 2011 (German, English, Japanese and Chinese pages). With resource block set to 5000, time block 5, and an SLA target of 95% compliance. The total provisions from the solutions are compared to a simple online algorithm that calculates provision at full compliance. (Numbers after normalization by Resource Block)

Table 1. Provision Level Compare to 100% SLA Compliance

Workload	Prediction R Value	Simple Algorithm 100% SLA	Deterministic 95% SLA	With Prediction 95% SLA
WikiDE	99.241%	1,133,505	999,195 (-11.85%)	1,003,085 (-11.51%)
WikiEN	98.858%	7,148,810	6,364,935 (-10.97%)	6,393,710 (-10.56%)
WikiJA	99.337%	1,257,245	1,097,795 (-12.68%)	1,101,750 (-12.37%)
WikiZH	99.307%	98,325	84,165 (-14.40%)	84,630 (-13.93%)
Average	99.186%	-	-12.47%	-12.09%

Compared to the 100% compliance case, targeting 95% SLA compliance reduce provision by more than 10% on average. The Online Algorithm with Prediction achieves slightly less saving thanks to the high accuracy prediction of the Neural Network.

Further, the simple online algorithm calculates provision with the predicted user demand, which can be seen as a simple look-forward solution. Due to prediction error the resulted provisions cannot reach full compliance but have an average of 0.5% loss. We then run our two solutions targeting the same SLA compliance to compare the resulting provision.

Table 2. Provision Level Compared to Simple Algorithm with Prediction

Workload	Simple Prediction	SLA Compliance	Deterministic	With Prediction
WikiDE	1,130,925	99.34%	1,105,215 (-2.27%)	1,130,690 (-0.020%)
WikiEN	7,077,295	99.44%	6,991,930 (-1.21%)	7,074,695 (-0.037%)
WikiJA	1,253,845	99.66%	1,238,615 (-1.21%)	1,253,535 (-0.025%)
WikiZH	98,095	99.65%	95,540 (-2.60%)	98,085 (-0.010%)
Average	-	99.52%	-1.82%	-0.023%

The computational result from the Deterministic solution shows that even with a high SLA compliance requirement, there is still room for improvement. While the Online Algorithm with Prediction shows that, even facing prediction error, incorporating past provision history as well as SLA fulfillment into the optimization process enables further savings without reducing SLA quality.

5.2 Prediction Error Analysis

We normalize the prediction errors by the corresponding number of user requests so that they can be compared across periods.

Using the distribution fitting tool of MatLab [23], we found that a *t* location-scale distribution, which is a variation from normal distribution, fits the error sample with the highest likelihood. From its cumulative distribution function, we can calculate the probability corresponding to different magnitude of errors. e.g. the chance of a 20% under-provision is 1.19%. Now the service provider faces a classical risk-profit trade off where he can choose to target a higher level of SLA compliance in order to reduce the risk of a serious under-prediction occurred in later periods.

As mentioned in 4.1 that a longer SLA contract enables the provider to better approach its SLA goal, a longer contract also makes him more robust towards prediction error at final periods, because of the diminishing weight on one period's prediction error.

6 Conclusion

In this paper we present an online deterministic algorithm as well as an algorithm with user requests prediction based on SLA for OLTP workflows. Experiments with Wikipedia data show that the algorithm with prediction delivers similar performance as the benchmark deterministic algorithm, while both yield improvements upon the simple look-forward online algorithm. By targeting 95% SLA compliance, provision can be reduced by 12% on average.

As mentioned in section 5.2 that the service provider could choose to insure for later periods by a slight over-provision in early periods. Our future work involves to analysis this risk-profit tradeoff quantitatively, and to incorporate risk-attitudes to find the optimal provision scheme and SLA lifetime.

References

1. Amazon Elastic Compute Cloud (Amazon EC2), `http://aws.amazon.com/ec2/`
2. Claybrook, B.G.: OLTP, online transaction processing systems. J. Wiley (1992)
3. Cozzatti, J.P.: Room to grow: a Twitter data center. Twitter Engineering Blog (2010), `http://engineering.twitter.com/2010/07/room-to-grow-twitter-data-center.html`
4. Wikipedia Pagecount Statistics, `http://dammit.lt/wikistats/`
5. Meisner, D.: PowerNap: Eliminating Server Idle Power. Analysis, 205–216 (2009)
6. Hawkins, A.: Unused Servers survey Results Analysis. The Green Grid (2010)
7. Gartner. Gartner Says Energy-Related Costs Account for Approximately 12 Percent of Overall Data Center Expenditures (2010), `http://www.gartner.com/it/page.jsp?id=1442113`
8. Scaramella, J.: Worldwide Server Power and Cooling Expense 2006-2010 Forecast. IDC Research (September 2006)

9. Aib, I., Boutaba, R.: On Leveraging Policy-Based Management for Maximizing Business Profit. IEEE Transactions on Network and Service Management 4(3), 25–39 (2007)
10. Yeo, C.S., Buyya, R.: Integrated Risk Analysis for a Commercial Computing Service. In: 2007 IEEE International Parallel and Distributed Processing Symposium, pp. 1–10 (2007)
11. Hasselmeyer, P., Qu, C., Schubert, L., Koller, B.: Towards Autonomous Brokered SLA Negotiation. In: Performance Computing (2006)
12. Verma, A., Ghosal, S.: On admission control for profit maximization of networked service providers. In: Proceedings of the Twelfth International Conference on World Wide Web, WWW 2003, p. 128 (2003)
13. Puschel, T., Lang, F., Bodenstein, C., Neumann, D.: A Service Request Acceptance Model for Revenue Optimization-Evaluating Policies Using a Web Based Resource Management Game. In: 2010 43rd Hawaii International Conference on System Sciences, HICSS, pp. 1–10 (2010)
14. Zhang, Q., Smirni, E., Ciardo, G.: Profit-driven service differentiation in transient environments. In: 11th IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer Telecommunications Systems, MASCOTS 2003, pp. 230–233 (2003)
15. Mitrani, I.: Management of Server Farms for Performance and Profit. The Computer Journal 53(7), 1038–1044 (2009)
16. Lackermair, G., Strahringer, S., Mandl, P.: Dynamically Scalable Architectures for E-Commerce. Multikonferenz Wirtschaftsinformatik 2010, 255 (2010)
17. Asaduzzaman, S., Maheswaran, M.: Utilizing unreliable public resources for higher profit and better SLA compliance in computing utilities. Journal of Parallel and Distributed Computing 66(6), 796–806 (2006)
18. Zhang, L., Ardagna, D.: SLA based profit optimization in autonomic computing systems. In: Proceedings of the 2nd International Conference on Service Oriented Computing, ICSOC 2004, p. 173 (2004)
19. Yu, Y., Ren, S., Chen, N., Wang, X.: Profit and penalty aware (PP-aware) scheduling for tasks with variable task execution time. In: Proceedings of the 2010 ACM Symposium on Applied Computing, SAC 2010, p. 334 (2010)
20. Kimbrel, T., Steinder, M., Sviridenko, M., Tantawi, A.: Dynamic application placement under service and memory constraints. Experimental and Efficient Algorithms, 391–402 (2005)
21. Boloor, K., Chirkova, R., Viniotis, Y., Salo, T.: Dynamic Request Allocation and Scheduling for Context Aware Applications Subject to a Percentile Response Time SLA in a Distributed Cloud. In: 2010 IEEE Second International Conference on Cloud Computing Technology and Science, pp. 464–472 (2010)
22. Moon, H.J., Chi, Y., Hacigümüs, H.: SLA-Aware Profit Optimization in Cloud Services via Resource Scheduling. In: 2010 6th World Congress on Services, pp. 152–153 (2010)
23. MATLAB version R2010b. Natick. The MathWorks Inc., Massachusetts (2010)

Integrating Heterogeneous Prediction Models in the Cloud

Hung-Chen Chen, Chih-Ping Wei, Yu-Cheng Chen, and Ci-Wei Lan

Dept. of Info. Management, National Taiwan University, Taipei, Taiwan, R.O.C.
JesseHCCHen@gmail.com
Dept. of Info. Management, National Taiwan University, Taipei, Taiwan, R.O.C.
cpwei@im.ntu.edu.tw
Inst. of Service Science, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.
shamen.chens@gmail.com
IBM Research Collaboratory, Taipei, Taiwan, R.O.C.
ciweilan@tw.ibm.com

Abstract. As the emergence and rapid growth of cloud computing, business intelligence service providers will host platforms for model providers to share prediction models for other users to employ. Because there might be more than one prediction models built for the same prediction task, one important issue is to integrate decisions made by all relevant models rather than adopting the decision from a single model. Unfortunately, the model integration methods proposed by prior studies are developed based on one single complete training dataset. Such restriction is not tenable in the cloud environment because most of model providers may be unwilling to share their valuable and private datasets. Even if all the datasets are available, the datasets from different sources may consist of different attributes and hard to train a single model. Moreover, a user is usually unable to provide all required attributes for a testing instance due to the lack of resources or capabilities. To address this challenge, a novel model integration method is therefore necessary. In this work, we aim to provide the integrated prediction result by consulting the opinions of prediction models involving heterogeneous sets of attributes, i.e., heterogeneous models. Specifically, we propose a model integration method to deal with the models under a given level of information disclosure by adopting a corresponding measure for determining the weight of each involved model. A series of experiments are performed to demonstrate that our proposed model integration method can outperform the benchmark, i.e., the model selection method. Our experimental results suggest that the accuracy of the integrated predictions can be improved when model providers release more information about their prediction models. The generalizability and applicability of our proposed method is also demonstrated.

Keywords: heterogeneous classification models, classifier integration, machine learning, cloud computing.

1 Introduction

Nowadays, companies and individuals are facing the challenge of information overload. In response, information systems for different domains are created to help

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 311–322, 2012.

companies and individuals make better decisions effectively and efficiently from an enormous amount of information. In these information systems, machine learning techniques have been widely adopted to find hidden patterns or support decision making, especially in the business domains, such as company bankruptcy prediction [7], [14], corporate dividend prediction [10], credit scoring [9], [22], prediction of market trends [9], and churn prediction [19]. These machine learning techniques have already proved their capabilities of promoting the productivity, and saving time and money for companies and individuals.

However, there might be many prediction models trained by machine learning techniques for the same problem. We take the prediction of credit card default risk as an example. It is essential for banks to control the risks by ranking the credit score of each customer based on his/her profile. Nevertheless, due to different perspectives, capabilities, or resources of the banks, the prediction models of credit card default risk from different banks may contain different attributes. As shown in Table 1, we collect from the literature five models for credit scoring and there are 31 different attributes involved in the five models [2], [5], [12], [18], [22]. The corresponding statistics show that the average attribute overlapping between any two models is only 36.12%. Specifically, there are only few common attributes in the five models. As also shown in Table 1, two other prediction tasks, i.e., length of hospital stays of burn patients (namely LOS) [21] and company bankruptcy [7], [14], also share similar phenomenon. Since the prediction models for the same task are different in essence, it is difficult to determine which one is the best. Moreover, each prediction model may be effective for different cases. One might consider integrating the underlying datasets used to construct these prediction models and then building a super prediction model from the integrated dataset. However, companies or organizations possessing these underlying datasets often cannot or will not share their data, due to privacy or business competition concerns. Therefore, it is desirable to integrate multiple prediction models (rather than their underlying datasets) pertaining to the same prediction task to improve prediction effectiveness. Note that there are prior studies that focus on the problem of classifier integration. The existing approaches are based on the assumption that the complete data source is available for training. But, in the scenario we concern, such complete data source does not exist. Accordingly, it is still necessary to solve the problem of integrating prediction models trained from different data sources with different attributes.

Table 1. Statistics of Models for Three Prediction Tasks

	# of models	Total # of attributes	Average # of attribute per model	Average attribute overlapping between any 2 models
Credit scoring	5	31	6.2	36.12%
LOS of burn patients	14	27	5.57	20.63%
Company bankruptcy	18	171	19.94	11.66%

With the development of cloud computing, companies and individuals can develop, deploy, and deliver applications in a more productive and efficient way. Specifically, cloud computing can bring the business lower cost, multiple redundant sites, scalable provisioning of resources, on-demand performance improvements, usage billings, fast deployment, and easy maintenance [16]. With respect to business intelligence services in the cloud, Google launches its Google Prediction API to provide Google's machine learning algorithms and prediction model hosting service on the cloud [3]. The Google Prediction API is now working to enable organizations or individuals to expose their models for paid use by other users. It can be foreseen that more and more prediction models will be available on this platform and attract users to search for relevant services for specific prediction requests. However, it is not easy for users to determine which prediction model is most effective among a number of models for their target prediction request. Moreover, using the prediction result from a single model is always too reckless to make the final decision. Therefore, the described business intelligence service environment on the cloud motivates us to develop a model integration method to integrate prediction models to achieve better performance for users.

As mentioned above, our challenge is to integrate prediction models that involve different attribute sets, i.e., *heterogeneous prediction models*, to enhance the overall accuracy for a target prediction request. Besides, we notice that model providers can disclose some useful model information, such as model accuracy or attribute ranking, for reference when users employ their models. How to utilize the given model information for improving the effectiveness of the integrated prediction result is also interesting. In this study, we define that the models for integration are all at a certain level of information disclosure. That is, each prediction model is associated with the same model information. Furthermore, we have to consider the real situation that a user's prediction request may not have all the attributes required by a prediction model.

In summary, there are three research questions in this study. First, we attempt to develop an appropriate method for integrating heterogeneous prediction models pertaining to the same prediction task for improving the prediction effectiveness. Second, we want to explore the relationship between the level of information disclosure and the effectiveness of model integration. Finally, we aim to discover guidelines for business intelligence service providers to manage prediction models provided in the cloud to reach higher service quality. In response, we propose a model integration method to deal with heterogeneous prediction models under a certain level of information disclosure. For simplification, we only consider the models built by the decision-tree induction technique, i.e., C4.5 [13], for integration. In our proposed method, we adopt the concept of weighted voting to combine the results from different models. Specifically, a specific measure based on the given model information is developed to determine the weight of each involved model. Subsequently, the prediction with the highest accumulated weight from all models is selected as the final result. Our experimental results demonstrate that our proposed model integration methods for different information disclosures are all superior to the benchmark, i.e., the model selection method. Moreover, the experimental results also reveal that the integrated prediction accuracy can be improved when model providers disclose more information about their models. This finding suggests that business intelligence service providers should develop incentive schemes to encourage model providers to supply more information about their models for enhancing the overall prediction accuracy.

The remainder of this paper is organized as follows. We first define the problem concerned in this study in Section 2. In Section 3, we review the literature related to this study. In Section 4, the details of our proposed model integration method for different information disclosure are presented. Section 5 reports our experimental evaluation and discusses some important evaluation results. Finally, we conclude the paper with some future directions in Section 6.

2 Problem Definition

As we mentioned in Section 1, our concerned problem is much complicated than that of the related works. Therefore, we formally define our focal problem as follows:

Suppose there are N prediction models M_1, M_2, …, and M_N trained for a specific prediction task under an certain model information with the corresponding attribute set MA_1, MA_2, …, and MA_N, respectively. Let $AUnion = \bigcup MA_i$ for i=1, 2, …N and $AInter = \bigcap MA_i$ for i=1, 2, …N. Then, the corresponding attribute set MA_1, MA_2, …, and MA_N must satisfy the following restriction: $AUnion \neq AInter$ (i.e., there exists at least one MA_i that is different from other MA_j). Moreover, suppose an active user registers a testing instance TI with attribute set UA and $UA \subset AUnion$. Our attempt is to estimate the corresponding W_i for each M_i based on UA and the given model information, such that we can integrate the weighted predictions from all M_i to get the final prediction result.

3 Literature Review

The idea of model integration is to utilize multiple learning methods to solve a machine learning problem [15]. Specifically, the prediction function based on several models is expected to provide better predictions than that based on a single model. Prior studies have demonstrated that integrating several prediction models into an overall prediction model or solution can improve the prediction accuracy of a single model because each single model has its own superiority in different cases. Alpaydin [1] classifies the existing model integration methods into two categories, i.e., *multiexpert combination* and *multistage combination*. Specifically, multiexpert combination methods, such as *voting*, *stacking*, *bagging*, and *boosting*, have base models that work in parallel. On the contrary, multistage combination methods, such as *cascading*, use a serial approach, where the next base model is trained with or tested on only the instances where the previous base models are not accurate enough. The details of these related model integration methods are summarized in the following.

The simplest way to combine multiple models is by voting, which corresponds to take a linear combination of the prediction results of these models. The equal voting method is to given equal weight to the models and then takes the prediction with the highest weight as the result. There are also other combination rules, such as weighted sum, median, minimum, maximum, and product [11]. Among these rules, sum rule is the most intuitive and most widely used one in practice [1]. However, the weighted voting scheme is only applicable when the user can supply additional information of how much s/he votes for each model. This is difficult for a user to supply such information because s/he may not have enough expertise or knowledge background of the models. Unlike voting, stacking tries to learn the weight for each model through a combiner system, which is another learner [20]. The input attributes of this combiner system are the prediction outcomes of the involved models. Therefore, stacking is a means of estimating and correcting for the biases of the multiple prediction models. However, the combiner system should be trained on data unused in training the models.

Bagging is also a kind of voting method whereby the models are trained based on the training datasets with some differences [4]. Specifically, for a training dataset with n instances, we first randomly select n instances with replacement from the training dataset to construct a re-sampled training dataset. If we need L models for integration, L corresponding re-sampled training datasets are required. After L models are trained, we take an average of the predictions made by these models as the final result. In bagging, generating complementary models is left to chance and to the unstability of the learning algorithm. On the contrary, boosting is to generate complementary models by training the next model that focuses more on the mistakes made by the previous models [1]. The original boosting method combines three weak models to generate a strong one [17]. Although the original boosting method is successful in reducing prediction error rate, its disadvantages is that a very large training dataset is required. To overcome this problem, Freund and Schapire [6] proposed a variant, i.e., Adaboost, to use the same training dataset over and over again. However, the prediction models need to be simple to avoid overfitting. Finally, the idea of cascading is to have a sequence of base models sorted in terms of their space or time complexity, or the cost of the representation they use [8]. Unlike boosting, cascading not only uses the misclassified instances of the previous model, but also the instances for which the previous based learner is not confident.

Unfortunately, the model integration methods proposed in the literature all need one single complete dataset for base models training and the corresponding iteration combination process. Such restriction is not tenable in the cloud environment because most of model providers are not willing to share their valuable and private datasets. Even if all the datasets are available, the datasets from different sources may consist of different attributes and hard to train a single model. As a result, these integration methods are unable to deal with our concerned problem. In this study, our proposed model integration method is designed based on the concept of weighted voting. Specifically, we utilize the given model information to automatically decide the weight of each model for each prediction request to achieve better accuracy. Therefore, it is convenient to companies and individuals to adopt our proposed method to get the integrated prediction results from multiple prediction models.

4 Classifier Integration Method

As we mentioned before, we aim to integrate prediction models with different involved attribute sets under a certain level of information disclosure. Since we deal with the models built by C4.5 in this study, four possible levels of information disclosure are considered as shown in Table 2. The simplest way for a model provider is to only provide the involved attributes, so that users can input correct attribute values (if available) into the focal prediction model. Moreover, a model provider can also release the prediction accuracy of the model. Note that the prediction accuracy of a model is evaluated based on the provider's own dataset. Furthermore, releasing the information of the ranking of the attributes involved in a prediction model could be helpful for users to determine the capability of this model. In this study, the ranking of all attributes in a model are decided based on the information gain measure (namely IG), which is widely adopted to determine the power of an attribute for classification [13]. Finally, the information gain of each attribute in a prediction model can also be released for more details. In summary, we define Level 1 as only disclosing the attributes of a model and the Level 4 as disclosing the most information including attributes, model accuracy, attribute ranking, and the corresponding IG. For each level of information disclosure, we propose a corresponding model weight measure for our proposed model integration method. The details of our proposed model integration method are depicted in the following.

Table 2. Model Information Included in Different Information Disclosure Levels

	Attributes of model	Model accuracy	Ranking of attributes	IG of attributes
Level 1	Y			
Level 2	Y	Y		
Level 3	Y	Y	Y	
Level 4	Y	Y	Y	Y

At Level 1, we suppose that users only know the attributes of each model. The ideal situation is that a user's prediction request (i.e., testing instance) has all attributes involved in each prediction model, and then we can just combine the prediction results from all models as the final decision. However, in the real situation, a user is usually unable to provide all attributes required by each model due to the lack of resources or capabilities. Therefore, our proposed model integration method for Level 1 is designed based on the percentage of all attributes in a model that are provided by the target user. Intuitively, more attributes the user can provide, higher prediction accuracy a model can reach. Following this assumption, the measure to calculate the percentage AP_i of all attributes in a model M_i that the target user can provide is defined as: $AP_i = (UA \cap MA_i) / MA_i$, where UA is the attribute set provided by the target user and MA_i is the attribute set required by the target model M_i. As a result, AP_i is used as the weight W_i of M_i for the target prediction request. Accordingly, the score S_j of each class C_j for the target testing instance is defined as: $S_j = \sum_{i=1}^{N} W_i \times P_{i,j}$, where N is the number of prediction models adopted for the prediction request, $P_{i,j}$ is the possibility of the testing instance belonging to class C_j predicted by M_i. Finally, the class with the highest score is then selected as the final result for the target prediction request.

At Level 2, we suppose that each model not only provides its own attributes, but also its prediction accuracy. Although the accuracy of each model is evaluated based on different datasets, the model accuracy is still an important clue for model integration. Intuitively, we may have more confidence on models with higher prediction accuracy. Consequently, the weight W_i of M_i for all its predictions is modified as: $W_i = AP_i \times AC_i$, where AP_i is the percentage of attributes in model M_i that the target user provides and AC_i is the prediction accuracy of M_i. Similarly, we calculate the score S_j of each class C_j for the testing instance and then select the class with the highest score as the final result.

At Level 3, we suppose that the attribute ranking in each model is also available. Since the ranking of all attributes in a model is decided based on the information gain, we can know whether the testing instance has important attributes for the focal model. Specifically, if the testing instance only has attributes with lower ranks in the focal model, the prediction accuracy of this model for this instance could be seriously decreased. Therefore, by using the information of attribute ranking, we can assess the prediction accuracy of each model for a testing instance more precisely. Since attributes with higher ranks should be more important than attributes with lower ranks, we first define the importance IM_k of an attribute A_k with the rank of k in model M_i as: $IM_k = (n_i - k + 1) / n_i$, where n_i is the number of attributes involved in M_i. Consequently, the weight W_i of M_i for all its predictions is further modified as: $W_i = AP_i \times AC_i \times \sum_{A_k \in UA \cap MA_i} IM_k$. Again, we calculate the score S_j of each class C_j for the testing instance and then select the class with the highest score as the final result.

At Level 4, we suppose that each model releases the following information: the attributes involved, model accuracy, attribute ranking, and information gain of each attribute. Actually, the information gain of an attribute is more representative than the ranking of this attribute, because attribute ranking only tells the difference in ranks of two attributes but not the exact difference between the classification power of the two attributes. Because attributes with higher information gain should be more important than those with lower information gain, we first set the importance IM_k of the attribute A_k in model M_i as its information gain IG_k. As with Level 3, the weight W_i of M_i for all its predictions is defined as: $W_i = AP_i \times AC_i \times \sum_{A_k \in UA \cap MA_i} IM_k$. Finally, we calculate the score S_j of each class C_j for the testing instance and then select the class with the highest score as the final result.

5 Empirical Evaluation

In this section, we describe our evaluation design, including evaluation datasets, evaluation procedure, and performance benchmark, and then discuss important experimental results.

5.1 Data Collection

For performance evaluation, we collect the three datasets (i.e., Chess, Dermatology, and Zoo) from the UCI Repository of Machine Learning Databases. A summary of the three datasets is shown in Table 3. The characteristics of these datasets are quite different so that we can demonstrate the generalizability of our proposed model integration method.

Table 3. Summary of Our Datasets

	# of instances	# of classes	# of attributes (discrete/continuous)
Chess	3196	2	36 (36/0)
Dermatology	366	6	34 (0/34)
Zoo	101	7	16 (16/0)

5.2 Evaluation Procedure

To evaluate the performance of our proposed model integration method, each original dataset in Table 3 is first partitioned randomly into two sub-datasets in each evaluation round, i.e., training dataset and testing dataset. The corresponding training dataset consists of 66% of the instances randomly selected from the original dataset and the corresponding testing dataset consists of the remaining 34% of the instances. The training dataset provides a basis for generating training instances to construct each prediction model for the target prediction task, and the testing dataset is then used to evaluate the performance of our proposed model integration method.

To simulate the situation in which heterogeneous prediction models are involved, we build five heterogeneous prediction models based on the target training dataset. Specifically, for a training dataset, each prediction model is not built based on the same training instances and does not use the same attributes as its inputs. Therefore, we have to split the training dataset into five sub-datasets. Because two of our collected datasets (i.e., Dermatology and Zoo) do not have a large number of instances, we allow some instances to appear in more than one sub-dataset. Specifically, we generate a random number for each training instance. If the number is smaller than the pre-defined threshold t (i.e., 0.75 in this study), this instance is assigned to one of the five sub-datasets. If the number is equal to or greater than t, we duplicate this training instance into more than one sub-datasets. Accordingly, we randomly select the number of copies c (between 2 and 5) that specifies the number of sub-datasets this training instance will appear. Thus, each instance in the training dataset will has a 6.25% chance of appearing in 2, 3, 4, or 5 sub-datasets, respectively.

Subsequently, we try to keep different attributes for each sub-dataset to train the corresponding model, such that the prediction accuracy of the five models will be different. Moreover, we also want to reduce the impact of the number of attributes involved in a prediction model. Therefore, we pre-define the composition of the attributes in each of the five models as shown in Table 4. Specifically, we first split the attributes of the original dataset into three groups, i.e., HIG (for high information gain), MIG (for medium information gain), and LIG (for medium information gain), according to the information gains of the attributes in the original dataset.

For example, the attributes of the chess dataset are partitioned into three groups and the top-12 attributes with the highest information gain are assigned to the HIG group. On the contrary, the 12 attributes with the lowest information gain are assigned to the LIG group. Afterward, according to Table 4, we randomly select 50% to 82% of the attributes from HIG, 18% to 50% of the attributes from MIG for the first sub-dataset, and then use these selected attributes to train the first model based on the first sub-dataset. The other four models are built based on the corresponding rules. Note that we employ C4.5 to build all these prediction models.

Table 4. Attribute Composition of the Five Prediction Models

	HIG	MIG	LIG
Model 1	(66±16)%	(34±16)%	0%
Model 2	(34±16)%	(66±16)%	0%
Model 3	(33±16)%	(34±16)%	(34±16)%
Model 4	0%	(66±16)%	(34±16)%
Model 5	0%	(34±16)%	(66±16)%

Finally, as we mentioned before, for a prediction request, the user is usually unable to provide all attributes required by each prediction model due to the lack of resources or capabilities. To simulate this scenario in our evaluation, we discard some attributes for each instance in the testing dataset before inputting to our model integration method. Suppose we decide to discard *x%* attributes for a testing instance. In this study, we randomly discard *x%* of the attributes for the target testing instance.

5.3 Performance Benchmark and Evaluation Criteria

For comparison purpose, we implement a model selection method (namely MS) as our performance benchmark. Specifically, for each testing instance, the prediction result of the model with the highest model weight is selected as the final result. The weight W_i of a target model M_i in a certain level of information disclosure in MS is the same as the weight of this model in the same level of information disclosure in our proposed model integration method. Moreover, in each of the following experiments, we adopt the average prediction accuracy of the 300 evaluation rounds as our evaluation criteria. The prediction accuracy of each evaluation round is defined as:

$$\text{accuracy} = \frac{\text{number of testing instances correctly predicted by the investigated method}}{\text{number of testing instances}}$$

5.4 Evaluation Results

We examine the performance of our proposed model integration method (namely MI) and the corresponding performance benchmark under information disclosure Level 1, 2, 3, and 4 based on the three datasets by following the evaluation procedure discussed in Section 5.2. In these experiments, we range the percentage of attributes in a testing instance to be discarded, i.e., *x* from 20 to 80 in increments of 20. Table 5 shows the corresponding evaluation results. First, the performance of both MS and MI decreases when x increases. It is reasonable that most of the prediction models fail if a user cannot provide enough attributes as required by each prediction model. Second, our proposed integration method outperforms the benchmark across all information disclosure levels investigated. Such results demonstrate that our proposed model integration method can leverage multiple prediction models to achieve greater prediction accuracy. Compared with the benchmark, our proposed method not only considers the condition of the target testing instance to choose the most suitable model by giving a higher weight, but also consults the opinions from other models. Moreover, in most cases, the prediction accuracy of our proposed method is higher when the level of information disclosure is higher. Such results suggest that if model providers can release more model information, the quality of prediction service will be greater. Finally, all the experimental results shown in Table 5 indicate that our proposed method is applicable for all datasets examined.

Table 5. Comparative Evaluation Results of MS and MI

(a) Chess Dataset

x%	Level 1		Level 2		Level 3		Level 4	
	MI	MS	MI	MS	MI	MS	MI	MS
20%	76.59%	69.63%	76.59%	68.83%	77.19%	69.19%	**80.44%**	68.61%
40%	70.36%	65.17%	70.36%	64.41%	71.18%	65.02%	**73.05%**	64.80%
60%	65.73%	61.51%	65.73%	61.32%	66.37%	60.98%	**67.35%**	61.25%
80%	59.03%	56.88%	59.03%	56.74%	59.39%	56.53%	**59.68%**	56.74%

Note: The boldfaced number in each row indicates the highest value across the different levels of information disclosure and different methods investigated.

(b) Dermatology Dataset

x%	Level 1		Level 2		Level 3		Level 4	
	MI	MS	MI	MS	MI	MS	MI	MS
20%	79.57%	66.37%	80.39%	65.65%	**80.69%**	65.49%	80.12%	65.24%
40%	65.25%	54.13%	65.98%	54.36%	**67.13%**	53.08%	66.29%	52.93%
60%	46.27%	42.28%	46.53%	42.48%	**49.16%**	40.90%	48.76%	41.42%
80%	30.78%	32.63%	31.20%	32.52%	**34.43%**	32.10%	34.20%	31.84%

(c) Zoo Dataset

x%	Level 1		Level 2		Level 3		Level 4	
	MI	MS	MI	MS	MI	MS	MI	MS
20%	82.54%	71.67%	83.22%	72.70%	**83.89%**	70.66%	83.75%	71.84%
40%	70.79%	62.37%	70.91%	64.07%	**73.57%**	61.91%	73.53%	61.99%
60%	60.15%	55.76%	60.34%	56.15%	**63.65%**	55.11%	63.56%	55.49%
80%	49.23%	48.04%	49.23%	49.05%	**52.20%**	47.61%	51.74%	48.14%

6 Conclusion and Future Research Directions

In the cloud, business intelligence service providers (e.g., Google) host platforms for model providers to share prediction models for other users to employ. Since there might be more than one prediction models built for the same prediction task, one important issue is to provide the integrated results from the involved models to avoid arbitrary decisions from a single model. Although prior studies have focused on the problem of model integration, existing approaches are basically based on the assumption that a complete dataset is available for training. Such assumption is not tenable because the model providers on the cloud may not be willing to share their own valuable and private datasets to others. Moreover, a user is usually unable to provide all required attributes for a testing instance due to the lack of resources or capabilities. Therefore, it is necessary to develop a model integration method to address this challenge. In this study, we notice that the models for the same prediction task usually have different concerned attributes and, given a prediction request, the information about a model can help assess possible prediction accuracy of the model. Accordingly, we propose a weighted-voting-based model integration method that combines the prediction results of all involved heterogeneous models to obtain the final prediction. Specifically, a specific measure based on the given model information is developed to determine the weight of each model involved. Subsequently, the class with the highest accumulated weight from all models is selected as the final result. Our experimental results show that our proposed integration method outperforms the benchmark, i.e., the model selection method, in all the scenarios investigated. The results also suggest that business intelligence service providers should develop incentive schemes to encourage model providers to release more information about their prediction models so that the service providers can provide more effective prediction services.

Some ongoing and future research directions are discussed as follows. First, evaluating the proposed integration method with more datasets will help further extend the generalizability of the evaluation results reported in this study. Second, in this study, we only consider underlying prediction models with the same level of information disclosure. Such restriction may be too strict. Therefore, the extension of our proposed model integration method to deal with prediction models with different levels of information disclosure is currently underway. Finally, our proposed method is now developed for prediction models trained by C4.5. One future research direction is to evaluate the proposed integration method when underlying prediction models are constructed by different learning algorithms.

Acknowledgement. This work was supported by the National Science Council of the Republic of China under the grant NSC 99-2410-H-002-191-MY2.

References

1. Alpaydin, E.: Introduction to Machine Learning. The MIT Press (2010)
2. Bellotti, T., Crook, J.: Loss given default models incorporating macroeconomic variables for credit cards. International Journal of Forecasting (forthcoming)

3. Boyd, R.: Make your business app intelligent with the google prediction API (2011), http://googleappsdeveloper.blogspot.com/2011/06/make-your-business-app-intelligent-with.html
4. Breiman, L.: Bagging predictors. Machine Learning 24(2), 123–140 (1996)
5. Dunn, L.F., Kim, T.H.: An empirical investigation of credit card default. Working Paper, Department of Economics. The Ohio State University, Columbus, Ohio (1999)
6. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: Proceedings of International Conference on Machine Learning, pp. 148–156 (1996)
7. Hsieh, W.K., Liu, S.M., Hsieh, S.Y.: Hybrid neural network bankruptcy prediction: an integration of financial ratios, intellectual capital ratios, MDA and neural network learning. In: Proceedings of International Conference on Computational Intelligence in Economics and Finance (2006)
8. Kaynak, C., Alpaydin, E.: Multistage cascading of multiple classifiers: one man's noise is another man's data. In: Proceedings of International Conference on Maching Learning, pp. 455–462 (2000)
9. Ketter, W., Collins, J., Gini, M., Gupta, A., Schrater, P.: Detecting and forecasting economic regimes in multi-agent automated exchanges. Decision Support Systems 47(4), 275–560 (2009)
10. Kim, J., Won, C., Bae, J.K.: A knowledge integration model for the prediction of corporate dividends. Expert Systems with Applications 37(2), 1344–1350 (2010)
11. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(3), 226–239 (1998)
12. Koh, H.C., Tan, W.C., Goh, C.P.: A two-step method to construct credit scoring models with data mining techniques. International Journal of Business and Information 1(1), 96–118 (2006)
13. Quinlan, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc. (1993)
14. Ravi Kumar, P., Ravi, V.: Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. European Journal of Operational Research 180(1), 1–28 (2007)
15. Re, M., Valentini, G.: Ensemble methods: a review. In Data Mining and Machine Learning for Astronomical Applications. Chapman & Hall (2011)
16. Reyes, E. P.: A systems thinking approach to business intelligence solutions based on cloud computing (2010), http://dspace.mit.edu/handle/1721.1/59267
17. Schapire, R.E.: The strength of weak learnability. Machine Learning 5(2), 197–227 (1990)
18. Steenackers, M.: A credit scoring model for personal loans. Insurance: Mathematics and Economics 8(1), 31–34 (1989)
19. Wei, C., Chiu, I.: Turning telecommunications call details to churn prediction: a data mining approach. Expert System with Applications 23(2), 103–112 (2002)
20. Wolpert, D.H.: Stacked generalization. Neural Networks 5(2), 241–259 (1992)
21. Yang, C.S., Wei, C., Yuan, C.C., Schoung, J.Y.: Predicting the length of hospital stay of burn patients: Comparisons of prediction accuracy among different clinical stages. Decision Support Systems 50(1), 325–335 (2010)
22. Yeh, I.: The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications 36(2), 2473–2480 (2009)

Software Licensing in the Case of No Network Effects

Shengli Li and Hsing Kenneth Cheng

Department of Information Systems and
Operations Management
Warrington College of Business Administration,
University of Florida
{lsl1217,hkcheng}@ufl.edu

Abstract. Traditionally, consumers purchase software by paying the price upfront and install the software on their computers. However, allowing consumers "pay as you go" by subscribing to the software has become increasingly popular. The two licensing models are referred to as "on-premises model" and "Software as a Service (SaaS) model" respectively. This paper studies the software vendor's choice of the two models considering consumers' uncertainty on the software quality and the software upgrading issue.

Keywords: Software as a Service, Backward compatibility, Consumers' uncertainty, Network Effects.

1 Introduction

Traditionally, consumers purchase software by paying the price upfront and install the software at their computers. This type of software is referred to as on-premises software and we refer to this licensing model as on-premises model[1]. However, with the development of the Internet in the late 90s, Software as a Service (SaaS) has emerged as a new software licensing model and gained significant popularity. In this licensing model, instead of paying the price upfront, consumers pay a subscription fee on a recurring basis [5]. They have access to the software as long as they pay the subscription fee. The SaaS model has attracted such well-established software firms as Oracle, SAP, and Microsoft into offering some of their applications on a subscription basis.

It has been estimated that SaaS sales have reached US$9B in 2010, up 15.7% from 2009, and will increase to US$10.7b in 2011. SaaS applications account for approximately 10% of the total enterprise software market in 2010, and may achieve 16% of worldwide software sales by 2014 [4]. Given that the SaaS model is gaining more and more popularity, a critical question facing the software firms is whether they will be better off by converting to the SaaS model from the on-premises model. A major objective of this paper is to provide guidelines for the software firms on this critical issue.

[1] See `http://en.wikipedia.org/wiki/On-premises_software`

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 323–333, 2012.

The SaaS model in essence amounts to a software leasing model. In this model, the consumers rent the software from the firm and pay rent for the period of usage. The on-premises model is a selling model, in which consumers purchase the software and own the license perpetually. Although the problem of whether to sell or lease the durable goods has been studied extensively, software as a durable good has several distinct features that have not been considered in prior literature. First, software exhibits positive network effects. When there exist positive network effects, the consumer utility from the product increases with the size of the consumers adopting the same product. This study focuses on the case where the network effects are negligible.

Another feature of software is that a newer version becomes available not long after the previous one. Consumers who did not purchase the previous version need to pay the full price of the newer version, while those who did can pay a smaller fee to upgrade. In the SaaS model, consumers pay the same subscription fee, presumably smaller than the full price of the on-premises model, for both versions. To complicate the matter, there is a compatibility issue with the software upgrade. For example, a DOC file created by Microsoft Word 2003 can be opened by Word 2007, which is not the case the other way around. This compatibility issue is commonly referred to as "backward compatible"; that is, the new version is compatible to the old version but not vice versa. In this case, backward compatibility problem can reduce the benefit of network effects for the users of the previous version [3].

Considering the distinct features of software discussed above, this paper aims at providing guidelines for the software firm on which licensing model – SaaS or on-premises – will generate a greater profit. The remainder of this paper is organized as follows. Section 2 describes the SaaS and on-premises licensing models in the absence of network effects, and compares their performances. Section 3 concludes the paper and identifies possible directions for future research.

2 The Baseline Model: No Network Effects

In this section, we derive the software firm's optimal profit for both SaaS and on-premises licensing models in the absence of network effects. The baseline model of no network effects provides us a benchmark case. Comparing the results of the benchmark case with those of the case with network effects enables the firm to discern the impact of network effects. We first analyze the SaaS licensing model in section 2.1, then the on-premises licensing model in section 2.2 and compare the two in section 2.3.

For both SaaS and on-premises licensing models, the software firm under consideration offers the first version of the software at the beginning of the first period. Consumers have a level of uncertainty about the quality of the software before adopting it. Consumers believe that the quality of software will be s_1^H (high quality) with a probability α and s_1^L (low quality) with a probability $1-\alpha$. The software firm offers the second version of the software at the beginning of the second period. Similar to the first period, consumers believe that the quality will be s_2^H with a probability of α, and s_2^L with a probability of $1-\alpha$. The second version of the software will have a higher quality than that of the first version. Then, $s_2^H > s_1^H, s_2^L > s_1^L$.

2.1 SaaS Model in the Absence of Network Effects

Under the SaaS model, the software firm charges a subscription fee p_L for each period of usage of the software. Note that this fee remains the same for both periods to reflect the common practice of the industry. Consumers make the decision whether to subscribe the software or not at the beginning of each period. At the beginning of the first period, consumers decide to subscribe the software or remain inactive; Likewise, consumers makes the decision again for the second period no matter what choices they have made in the previous period. Thus, consumers have four possible strategies under the SaaS licensing model - subscribe in both periods, subscribe in the first period and remain inactive in the second period, remain inactive in the first period and subscribe in the second period, and remain inactive in both periods. Table 1 summarizes the notation of our model.

Table 1. Model Notation

s_1^H	The possible high quality of the first version
s_1^L	The possible low quality of the first version
s_2^H	The possible high quality of the second version
s_2^L	The possible low quality of the second version
α	The probability that consumers believe the software is high quality
θ	Consumers' valuation of the software
c	Usage cost
p_L	The price charged for subscription of one period in the SaaS model
p_S	The price to purchase the software in the on-premises model
p_U	The price to upgrade in the second period in the on-premises model

The consumers have heterogeneous valuations of the software [2, 3]. Following Conner [2], we designate the size of the total population in the market as K and the number of customers who have a positive valuation of the software as N such that $K = a \cdot N, a > 1$. Then the number of people not interested in the software will be $K - N = a(N-1)$. The total population can thus be modeled by the interval $[-a(N-1), N]$. For simplicity, we normalize N to 1. Then the interval $[-(a-1),\ 1]$ corresponds to the total population in the market. Let a customer's valuation type be denoted by θ, where θ is uniformly distributed over $[-(a-1),\ 1]$. Hence, consumers with a valuation type between 0 and 1 have a positive valuation about the software, and therefore correspond to the potential market size of the software firm. Further, let c denote the usage cost incurred by consumers when adopting the software. This usage cost includes time and efforts cost for the consumers to learn and use the software [1].

The consumer's expected net utility by subscribing to the software in the first period equals

$$U = \theta\left(\alpha s_1^H + (1-\alpha)s_1^L\right) - p_L - c. \tag{1}$$

For simplicity, we use Es_1 to designate $\alpha s_1^H + (1-\alpha)s_1^L$ and Es_2 to designate $\alpha s_2^H + (1-\alpha)s_2^L$ in the rest of this paper. Therefore, the net utility for those who remain inactive in both periods (*II* for Inactive, Inactive) equals

$$U_{II} = 0. \tag{2}$$

The net utility for those who remain inactive in the first period and subscribe to the software in the second period (*IS* for Inactive, Subscribe) equals

$$U_{IS} = \theta Es_2 - p_L - c. \tag{3}$$

Likewise, the net utility for those who subscribe to the software in the first period and remain inactive in the second period (*SI* for Subscribe, *I*nactive) equals

$$U_{SI} = \theta Es_1 - p_L - c. \tag{4}$$

Finally, the net utility for those who subscribe to the software in both periods (*SS* for Subscribe, Subscribe) equals

$$U_{SS} = \theta Es_1 - p_L - c + \theta Es_2 - p_L - c. \tag{5}$$

The consumers choose one of the four possible strategies under the SaaS licensing model that maximizes their net utilities.

We find that for a given consumer the expected net utility of adopting *SI* strategy is always smaller than that of *IS* strategy. This implies that the *SI* strategy is always dominated by the *IS* strategy. Furthermore, it can be shown that the consumer who chooses the *IS* strategy has a higher valuation parameter θ than those who choose *II* strategy, and the consumer who chooses the *SS* strategy must have a higher valuation parameter θ than those who choose *IS* (see Appendix for proof). Figure 1 shows the market segments under the SaaS licensing model.

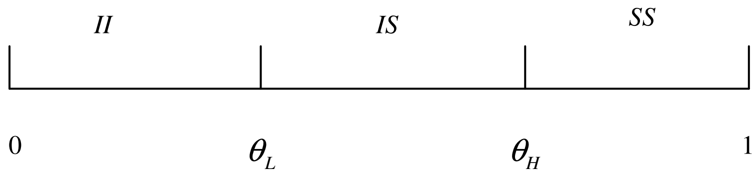


Fig. 1. Market Segments for the SaaS model

The marginal consumer θ_L in Figure 1 has the same expected net utility between choosing *II* and *IS*, which implies $E[\theta_L s_2 - c - p_L] = 0$ and $\theta_L = \dfrac{c + p_L}{Es_2}$. The marginal consumer θ_H has the same expected net utility between choosing *IS* and *SS*, which implies $E[\theta_H s_2 - c - p_L] + E[\theta_H s_1 - c - p_L] = E[\theta_H s_2 - c - p_L]$. Thus, $\theta_H = \dfrac{2(c + p_L)}{Es_1 + Es_2}$. The interval $\theta_H - \theta_L$ corresponds to the demand for the software from the consumers who choose the *IS* strategy, while the interval $1 - \theta_H$ corresponds to the demand for the software from those who choose the *SS* strategy. Thus, the profit for the software firm adopting the SaaS licensing model equals

$$\pi = 2p_L(1 - \theta_H) + p_L(\theta_H - \theta_L) = p_L\left[2 - \frac{(c + p_L)(Es_1 + 3Es_2)}{Es_2 \cdot (Es_1 + Es_2)}\right]. \qquad (6)$$

To guarantee non-negative demands, it requires that $\theta_L \geq 0$, $\theta_L \leq \theta_H$, $\theta_H \leq 1$. Since c, p_S, and Es_2 are non-negative, the condition $\theta_L \geq 0$ is always satisfied. The condition $\theta_L \leq \theta_H$ requires $Es_1 \leq Es_2$, which is implied in our model. The condition $\theta_H \leq 1$ requires $p_L \leq \dfrac{Es_1 + Es_2}{2} - c$. This implies there is an upper bound on the subscription fee p_L the software firm can charge. Therefore, the software firm faces the following maximization problem to achieve the optimal profit.

$$\begin{aligned} &\max_{p_L} \ \pi = p_L\left[2 - \frac{(c + p_L)(Es_1 + 3Es_2)}{Es_2 \cdot (Es_1 + Es_2)}\right]. \\ &s.t. \ 0 \leq p_L \leq \frac{Es_1 + Es_2}{2} - c \end{aligned} \qquad (7)$$

The optimal solution to the above problem is (see the proof in the Appendix)

$$p_L^* = \begin{cases} \dfrac{2Es_2(Es_1+Es_2)-c(Es_1+3Es_2)}{2(Es_1+3Es_2)}, \text{ when } c \le \dfrac{(Es_1+Es_2)^2}{Es_1+3Es_2} \\ \dfrac{Es_1+Es_2}{2}-c, \text{ when c} > \dfrac{(Es_1+Es_2)^2}{Es_1+3Es_2} \end{cases}. \quad (8)$$

The firm realizes the following optimal profit

$$\pi^* = \begin{cases} \dfrac{(c(Es_1+3Es_2)-2Es_2(Es_1+Es_2))^2}{4Es_2(Es_1+Es_2)(Es_1+3Es_2)}, \text{when } c \le \dfrac{(Es_1+Es_2)^2}{Es_1+3Es_2} \\ \dfrac{(Es_1+Es_2-2c)(Es_2-Es_1)}{4Es_2}, \text{when c} > \dfrac{(Es_1+Es_2)^2}{Es_1+3Es_2} \end{cases}. \quad (9)$$

From Eqs. (8) and (9), there are two sets of optimal solutions, which correspond to two market segmentation structures. The market structures depend on the relationship between the usage cost and software quality as follows:

1) When the usage cost c is less than or equal to the threshold value of $\dfrac{(Es_1+Es_2)^2}{Es_1+3Es_2}$, the optimal market structure includes all three segments of *II*, *IS* and *SS*. This shows that when the usage cost is relatively small, every one of the three strategies will be chosen by some consumers.

2) When the usage cost c is greater than the threshold value of $\dfrac{(Es_1+Es_2)^2}{Es_1+3Es_2}$, the optimal market structure includes only *II* and *IS*. This shows that when the usage cost is relatively high, no consumers will choose the *SS* strategy, which implies that no consumers will subscribe to the software for both periods.

It can be shown that when $Es_2 \ge Es_1$ and $c \le Es_1$, the condition $c \le \dfrac{(Es_1+Es_2)^2}{Es_1+3Es_2}$ always holds. The expected quality of the second version is greater than that of the first version ($Es_2 \ge Es_1$) and the usage cost is less than the expected quality of the software ($c \le Es_1$); otherwise, no consumers will purchase the software. In other words, the only feasible market structure in reality is the one that includes all three segments of *II*, *IS* and *SS*. We thus have the following proposition.

Proposition 1: *In the absence of network effects, the software firm implementing the SaaS licensing model should set the optimal subscription price as*

$p_L^* = \dfrac{2Es_2(Es_1 + Es_2) - c(Es_1 + 3Es_2)}{2(Es_1 + 3Es_2)}$ *and realizes the optimal profit of* $\pi^* = \dfrac{(c(Es_1 + 3Es_2) - 2Es_2(Es_1 + Es_2))^2}{4Es_2(Es_1 + Es_2)(Es_1 + 3Es_2)}$ *in the market where all three segments of II, IS and SS consumers coexist.*

2.2 On-Premises Model in the Absence of Network Effects

Under the on-premises licensing model, the software firm charges a price p_S for consumers who purchase the software at the beginning of the first period or the second period. Consumers have access to the software in both periods if they purchase it in the first period. For those who have purchased the software in the first period, the software firm charges them an upgrading fee p_U to upgrade the first version to the second version. The upgrading fee is lower than the purchase price to reflect the reality, i.e., $p_U < p_S$.

The consumers make decisions in the beginning of the first period based on their total expected utilities received in the both periods. They have four strategies to consider remain inactive in both periods (*II*), inactive in the first period and buy in the second period (*IB*), buy in first period and hold it in the second period (*BH*), and buy in the first period and upgrade in the second period (*BU*).

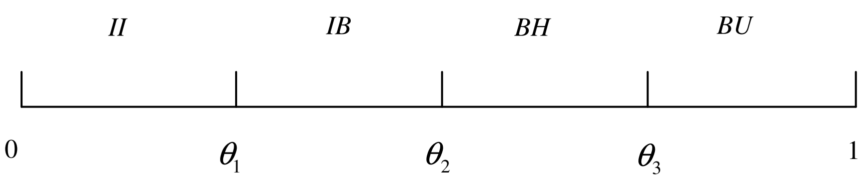


Fig. 2. Market Segments of the On-premises Model

The expected utility of those consumers who choose strategy *II* is $U_{II} = 0$. For consumers who choose strategy *IB*, they derive expected utility of $U_{IB} = \theta Es_2 - p_S - c$. The consumers who choose strategy *BH* realizes expected utility of $U_{BH} = \theta Es_1 - p_S - c + \theta Es_1 - c$. Finally, the expected utilities of those consumers who choose strategy *BU* equals $U_{BU} = \theta Es_1 - p_S - c + \theta Es_2 - p_U - c$. Thus, the market under the on-premises licensing model is described in Figure 2 (see Appendix for proof).

The marginal consumer θ_1 has the same net utility between choosing the strategy *II* and *IB*, which implies $E[\theta_1 s_2 - c - p_S] = 0$ and $\theta_1 = \dfrac{c + p_S}{Es_2}$. The marginal consumer θ_2 has the same net utility between choosing the strategy *IB* and *BH*, which implies $E[\theta_2 s_1 - c - p_S] + E[\theta_2 s_1 - c] = E[\theta_2 s_2 - c - p_S]$ and $\theta_2 = \dfrac{c}{2Es_1 - Es_2}$. Finally, the marginal consumer θ_3 has the same net utility between choosing the strategy BH and BU, which implies $E[\theta_3 s_1 - c - p_S] + E[\theta_3 s_2 - c - p_u] = E[\theta_3 s_1 - c - p_S] + E[\theta_3 s_1 - c]$. Thus, $\theta_3 = \dfrac{p_U}{Es_2 - Es_1}$.

To ensure that each market segment is nonnegative requires $\theta_1 \geq 0, \theta_1 \leq \theta_2, \theta_2 \leq \theta_3, \theta_3 \leq 1$. Since c, p_S, and Es_2 are nonnegative, the condition $\theta_1 \geq 0$ is always satisfied. The condition $\theta_1 \leq \theta_2$ requires that $p_S \leq 2c\dfrac{Es_2 - Es_1}{2Es_1 - Es_2}$. This implies that the selling price of the software must be lower than the upper bound of $2c\dfrac{Es_2 - Es_1}{2Es_1 - Es_2}$. The condition $\theta_2 \leq \theta_3$ requires that $p_U \geq \dfrac{c(Es_2 - Es_1)}{2Es_1 - Es_2}$, suggesting that there is a minimum upgrade price. Finally, the condition $\theta_3 \leq 1$ requires that $p_U \leq Es_2 - Es_1$. It follows that the upgrade price cannot exceed the upper bound of $Es_2 - Es_1$.

The software firm can achieve the following profit by adopting the on-premises model.

$$\pi = p_S(1-\theta_2) + p_U(1-\theta_3) + p_S(\theta_2 - \theta_1) = p_S + p_U + \frac{p_U^{\ 2}}{Es_2 - Es_1} - \frac{p_S(c + p_S)}{Es_2} \tag{10}$$

The problem facing the software firm becomes

$$\begin{aligned} &\max_{p_S, p_U} \ \pi = p_S + p_U + \frac{p_u^{\ 2}}{Es_2 - Es_1} - \frac{p_S(c + p_S)}{Es_2} \\ &s.t.\ 0 \leq p_S \leq 2c\frac{Es_2 - Es_1}{2Es_1 - Es_2} \\ &\quad \frac{c(Es_2 - Es_1)}{2Es_1 - Es_2} \leq p_U \leq Es_2 - Es_1 \end{aligned} \quad . \tag{11}$$

The optimal price of the software and the optimal upgrading fee are (see the Appendix):

$$p_S^* = \begin{cases} 2c\dfrac{Es_2 - Es_1}{2Es_1 - Es_2}, \text{ when } c \le \dfrac{2Es_1 - Es_2}{2} \\ 2c\dfrac{Es_2 - Es_1}{2Es_1 - Es_2}, \text{ when } \dfrac{2Es_1 - Es_2}{2} < c < \dfrac{Es_2(2Es_1 - Es_2)}{3Es_2 - 2Es_1}, \\ \dfrac{(Es_2 - c)}{2}, \text{ when } \dfrac{Es_2(2Es_1 - Es_2)}{3Es_2 - 2Es_1} \le c \le 2Es_1 - Es_2 \end{cases} \quad (12)$$

$$p_U^* = \begin{cases} \dfrac{(Es_2 - Es_1)}{2}, \text{ when } c \le \dfrac{2Es_1 - Es_2}{2} \\ \dfrac{c(Es_2 - Es_1)}{2Es_1 - Es_2}, \text{ when } \dfrac{2Es_1 - Es_2}{2} < c < \dfrac{Es_2(2Es_1 - Es_2)}{3Es_2 - 2Es_1}, \text{ and} \\ \dfrac{c(Es_2 - Es_1)}{2Es_1 - Es_2}, \text{ when } \dfrac{Es_2(2Es_1 - Es_2)}{3Es_2 - 2Es_1} \le c \le 2Es_1 - Es_2 \end{cases} \quad (13)$$

the optimal profit equals

$$\pi^* = \begin{cases} \dfrac{(Es_2 - Es_1)\left((2Es_1 - Es_2)^2 + 8c(2Es_1 - Es_2) - 8c^2\right)}{4(2Es_1 - Es_2)^2}, \text{ when } c \le \dfrac{2Es_1 - Es_2}{2} \\ \dfrac{3c(Es_2 - Es_1)(2Es_1 - Es_2 - c)}{(2Es_1 - Es_2)^2}, \text{ when } \dfrac{2Es_1 - Es_2}{2} < c < \dfrac{Es_2(2Es_1 - Es_2)}{3Es_2 - 2Es_1} \\ \dfrac{1}{4}\left(Es_2 + \dfrac{4c(Es_1 + c)}{2Es_1 - Es_2} - \dfrac{4c^2 Es_1}{(2Es_1 - Es_2)^2} + \dfrac{c^2}{Es_2} - 6c\right), \text{ when } \dfrac{Es_2(2Es_1 - Es_2)}{3Es_2 - 2Es_1} \le c \le 2Es_1 - Es_2 \end{cases} . \quad (14)$$

Eqs. (12) - (14) show that there are three sets of optimal solutions depend on the relationship between usage cost and software quality. Each solution corresponds to a particular market structure. When the usage cost c and software quality satisfy the condition $c \le \dfrac{2Es_1 - Es_2}{2}$, the market structure includes three segments˜ *II*, *BH* and *BU*. This implies that when the usage cost is relatively small, the software firm sets the optimal sales price of the on-premises software at the upper bound of Eqs. (12), leading to $\theta_1 = \theta_2$, i.e., the disappearance of *IB* segment. When the usage cost c and software quality satisfy the condition $\dfrac{2Es_1 - Es_2}{2} < c < \dfrac{Es_2(2Es_1 - Es_2)}{3Es_2 - 2Es_1}$, the market structure includes two segments $\tilde{}$ *II* and *BU*. This implies that when the usage cost is mediate, the software firm sets the optimal sales price at the upper bound and the upgrading fee at the lower bound, which leads to $\theta_1 = \theta_2 = \theta_3$, i.e., the disappearance of segments of *IB* and *BH*. Finally, when the usage cost c and software quality the condition $\frac{Es_2(2Es_1 - Es_2)}{3Es_2 - 2Es_1} \le c \le 2Es_1 - Es_2$, the market structure includes three segments $\tilde{}$ *II*, *IB* and *BU*, leading to $\theta_2 = \theta_3$, i.e., the disappearance of *BH* segment.

2.3 SaaS versus On-Premises Model without Network Effects

Comparing the optimal profit of the SaaS model in Eq. (9) with that of the on-premises model in Eq. (14) shows that it is too complex to determine which licensing model generates a higher profit. Therefore, we resort to computational analysis for managerial insights. To this end, we set s_1^H at the baseline value of 100. Then, we let s_1^L equal 90% of s_1^H, s_2^H equal 20% higher than s_1^H, and s_2^L equal 90% of s_2^H. The probability that the consumers believe the software quality is high is set at 0.7. Figure 3 plots the optimal profits of SaaS and on-premises licensing models with respect to the software usage cost c where c is between 0 and the highest upper bound of $2Es_1 - Es_2$. In Figure 3, the value of c is expressed in terms of the percentage of the baseline parameter s_1^H.

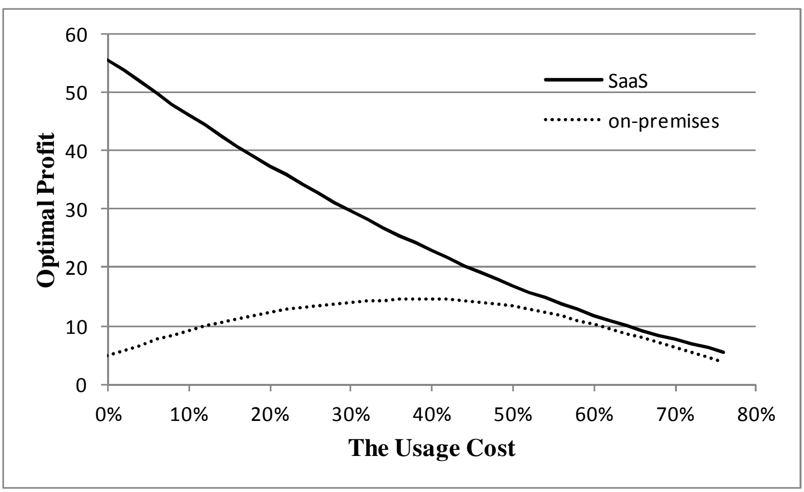


Fig. 3. The Comparison of the Optimal Profits of SaaS vs. On-Premises Model with respect to the Usage Cost

Figure 3 clearly demonstrates that the SaaS model dominates the on-premises model for all values of software usage cost c. This implies that the software firm should implement the SaaS model in the absence of network effects.

3 Concluding Remarks

Software as a Service (SaaS) has become an increasingly popular software licensing model since the term was first coined in February 2001 by the Software & Information Industry's (SIIA) eBusiness Division[2], although most software is still offered through the traditional on-premises model. The objective of this research is to examine the software firm's choice between SaaS and on-premises licensing models in the absence of network effects. We show that SaaS model dominates the on-premises model in generating a higher profit when network effects are not existent or weak. Our future research will examine the impact of network effects on the software firm's optimal choice of licensing model.

References

1. Cheng, H., Liu, Y.: Optimal software free trial strategy: The impact of network effects and consumer uncertainty. Information Systems Research (2011)
2. Conner, K.R.: Obtaining strategic advantage from being imitated: When can encouraging "clones" pay? Management Science 41(2), 209–225 (1995)
3. Ellison, G., Fudenberg, D.: The neo-luddite's lament: Excessive upgrades in the software industry. The RAND Journal of Economics 31(2), 253–272 (2000)
4. Hall, K.G.: SaaS sales will grow 16.2% to $10.7bn (2011), `http://www.Computerweekly.com/Articles/2010/12/14/244489/Gartner-SaaS-sales-will-grow-16.2-to-10.7bn-in-2011.htm` (retrieved June 27, 2011)
5. Industry, S.&.I. Software as A service: Strategic backgrounder (retrieved June 27, 2011)

[2] See, `http://en.wikipedia.org/wiki/SaaS`

Human Capital and Information Technology Capital Investments for Firm Innovation: Curvilinear Explanations

John Qi Dong*, Jinyu He, and Prasanna Karhade

School of Business and Management, Hong Kong University of Science and Technology, Cleart Water Bay, Kowloon, Hong Kong
{dongqi,mnjhe,karhade}@ust.hk

Abstract. We investigate the relationship between a firm's investments in knowledge-related assets and innovation. We propose curvilinear (inverted-U shaped) effects of human capital investments and IT capital investments on firm innovation respectively. Based on the analysis of 349 German firms across industries, corroborating support for the proposed relationships is presented. We contribute to the existing work on the resource-based view of the firm (RBV) by identifying and testing the possible downsides associated with the excessive accumulation of strategic resources.

Keywords: Human Capital, Information Technology Capital, Innovation, Resource-based View.

1 Introduction

The resource-based view of the firm (RBV) maintains that firms that are able to accumulate and deploy valuable, rare, costly to imitate, and non-substitutable resources are positioned to generate and sustain competitive advantage (Barney, 1986; 1991). Among various resources that a firm possesses, its knowledge assets are often regarded as having such characteristics prescribed by RBV (Grant, 1996; Kogut and Zander, 1992). Also, scholars have recognized that a firm's knowledge that enables it to generate sustainable competitive advantage is often embedded in its employees (Nonaka and Takeuchi, 1995; Wright et al., 2001). And, as Kogut and Zander point out, "the knowledge of the firm must be understood as socially structured, or, more simply stated, as resting in the organizing of human resources" (1992: p. 385; emphasis added). In this regard, the role of the firm is to create an environment that facilitates the sharing and recombination of individually possessed knowledge (Argote et al., 2003). A key component of this environment is a firm's information technology (IT) assets, which facilitate the exploration, sharing, and organizing of knowledge (Alavi and Leidner, 2001). So, firms are investing both in assembling a capable team of knowledge workers and in acquiring superior IT resources (Sambamurthy and Zmud, 2000).

* Corresponding author.

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 334–346, 2012.

The application of these knowledge-related resources is expected to result in process and product innovations, enabling a firm to gain performance advantages. Scholars often propose innovation as an immediate outcome of knowledge-intensive activities within the firm (Katila and Ahuja, 2002). Accordingly, when examining the return on knowledge-related investments, it is necessary to recognize that a firm's performance in innovation may serve as a more proper unit of analysis than overall financial performance (Ray et al., 2004). On the basis of the RBV logic, one may expect a linear and positive relationship between a firm's human and IT capital investments and innovation. We, however, argue that the relationship between these investments and innovation is more likely to be curvilinear (i.e., following an inverted-U shape). That is, while human capital and IT capital investments may initially increase a firm's performance in innovation, excessive accumulation of these resources likely results in inefficiencies due to increased dynamic adjustment costs resulted from managerial complexity and information overload (e.g., Penrose, 1959; Tan and Mahoney, 2005). As a consequence, further investments in human and IT capitals may hinder a firm's capacity to innovate. We test our predictions using data from 349 German firms in 2002-2006. Our sample has good representativeness of the German economy, which is distributed across nearly all industries. We analyze the effects on performance-based measure of product innovation and find consistent results corroborating our theoretical arguments. In addition, we also test the interactive effect between these two knowledge-related resources as theoretically they may play either complementary or substitutive roles in the process of generating innovation. Our inability to find significant empirical results tends to suggest that the way these two resources interact with each other may be complex and depend on various contingencies. We contribute to the body of work on RBV by recognizing the possible downsides of accumulating strategic resources. Such recognition may constitute a valid explanation for why RBV has only received marginal support in the empirical literature (See Newbert [2007] for a recent review). Our findings of curvilinear effects of knowledge-related resources on firm innovation remind both researchers and practitioners the danger of an oversimplified interpretation of RBV (Barney, 1986; 1991).

2 Theory and Hypotheses

For the assets possessed by a firm, a distinction is often made between invisible and physical assets (Itami, 1987) with the former being those knowledge-based resources that are accumulated and produced by the firm's employees. It has been well recognized that "physical (visible) assets must be present for business operations to take place but invisible assets are necessary for competitive success" (Barney and Arikan, 2001, p. 136). Firms are now making extensive investments in both human capital and IT capital to develop their knowledge bases to compete in today's economy (Krishnan et al., 2007), because the former is the ultimate generator of new knowledge (Davenport and Prusak, 1998; Kogut and Zander, 1992) and the latter facilitates the exploration, sharing, and organization of knowledge (Alavi and

Leidner, 2001). Knowledge assets serve as vital inputs to a firm's innovation activity (Cohen and Levinthal, 1990; Miller et al., 2007), which is often regarded as the driver of a firm's competitiveness (Schumpeter, 1934). Accordingly, we study a firm's accumulation of knowledge assets in terms of both human capital and IT capital investments and maintain that these investments serve as key antecedents to a firm's innovation activity.

2.1 The Inverted-U Relationship between Human Capital and Innovation

Knowledge embedded in a firm's human capital is regarded as possessing valuable, rare, and non-substitutable characteristics (Wright et al., 2001). And, firms often consider investing in accumulating human capital an essential strategy to effectively compete with other firms (Ployhart et al., 2009). Empirical studies have found some support for a positive relationship between the accumulation of human capital and firm performance (Aaronson et al., 2007; Damanpour, 1991; Ployhart et al., 2009), or a U shaped curvilinear relationship between human capital embodied in partners and performance of professional service firms which is negative early in the partners' tenure but becomes positive (Hitt et al., 2001). Yet, we propose that an inverted-U shaped curvilinear relationship between the accumulation of human capital and firm innovation is more likely for two reasons. First, the accumulation of human capital necessarily increases the complexity of interactions among the firm's employees—knowledge workers in particular. As employees need to pool their individually possessed knowledge together, an effective control system is inevitably required to facilitate continuous knowledge sharing and generation. To the extent that individual actors are bounded in their rationality (Simon, 1947), increased complexity is likely to frequently render the existing control systems ineffective. As a consequence, when the accumulation of human capital exceeds a certain point, the control system as a whole may not be able to function effectively to ensure that the firm can continue to generate innovative outputs (Penrose, 1959). Particularly, with an over accumulation of human capital, the ease with which management can productively govern its internal human assets decreases. Therefore, as the level of accumulated human capital increases, the corresponding control systems also need to grow, leading to increased dynamic adjustment costs (Hay and Morris, 1991). The dynamic adjustment costs that management incurs systematically increase in contexts that rely on tacit processes, because explicit organizational structures and policies can only play a little role in coordinating socially complex processes such as knowledge transferring and sharing (Tan and Mahoney, 2005). Innovation activity, which is largely driven by such tacit processes of knowledge recombination, is thus likely to suffer as the size of the internal human resource pool exceeds an optimal level. Second, accumulating more human capital may also increase the diversity among employees. While diversity can bring about more creative ideas, the scarcity of the firm's attention (Simon, 1947; Mahoney, 2005) renders it difficult for the firm to choose the right focus. Moreover, lengthy debates over an excessive number of options generated by a diverse group of knowledge workers may further encumber decision-making processes and eventually weaken the firm's capacity to implement promising R&D projects. Based on the two theoretical arguments presented above, we propose that, after an optimal level, additional investments in human capital are likely to negatively impact the firm innovation activity.

H1: Accumulation of human capital and innovation performance are likely to have an inverted-U shaped relationship.

2.2 The Inverted-U Relationship between IT Capital and Innovation

IT facilitates access to knowledge by creating electronic links not only within a focal firm but also across its entire supply chain (Rai et al., 2006; Krishnan et al., 2007). IT assets are vital for tapping into information that typically resides outside the firm and for supporting business processes that span firm boundaries (Saraf et al., 2007). IT resources support sophisticated business activities through process adaptation and information sharing. Also, as organizations grow in size, they typically rely on IT to systematically organize knowledge dispersed within their boundaries (Gattiker and Goodhue, 2005). IT can help a firm to acquire, assimilate, transform and explore knowledge, which in turn breeds various innovations (Joshi et al., 2010). Empirical studies have shown some support for the positive relationship between the accumulation of IT capital and various firm performance outcomes including innovation. However, we propose that this relationship is more likely to be curvilinear for the following three reasons. First, higher IT capital accumulation and excessive deployments of IT systems can be associated with information overload (Ackoff, 1967) which is likely to reduce the productivity of various parties engaged in IT-enabled knowledge work. This is because the increased amount of information generated by the IT systems intensifies the competition for managers' and employees' attention, likely compromising the firm's decision-making process and quality (Ocasio, 1997). Such an information overload could arguably also explain the presence of an IT productivity paradox where increasing IT capital is associated with reduced employee productivity (Brynjolfsson and Hitt, 1996; Grise and Gallupe, 2000). Due to the challenges associated with managing this information overload (Langley et al., 1995), the productivity of innovation activity is also likely to suffer in the presence of excessive investments in IT capital. Second, after a certain optimal limit, the value of additional investments in IT can be constrained by the firm's lack of investments in complementary capabilities (Teece, 1986; Sambamurthy et al., 2003; Ray et al., 2005). IT resources are not utilized in a vacuum. Complementary capabilities such as business process redesigns (Broadbent et al., 1999) help IT resources realize their value in the actual productive, knowledge-intensive activities of the firm. With more investments in IT assets, a firm likely experiences a greater need to develop additional complementary capabilities. As innovation processes consume IT assets, the efficacy of firm innovation activity also critically depends on investments in complementary capabilities. Thus, with a higher level of IT capital investment, the relative shortage of complementary capabilities is likely to be detrimental for innovation. Third, as a firm continues to invest in IT, it uses a larger number of IT systems to reorganize its business system. The additional IT systems — which are deployed as a result of the incremental investments in IT — need to be systematically integrated into an ecosystem of existing IT applications and infrastructures. As a result, the firm typically incurs high dynamic adjustments (Penrose, 1959) and integration costs (Goodhue et al., 1992). If a firm does not exert additional effort or make the necessary investments to integrate its IT systems (Weill and Ross, 2004), the efficacy of firm innovation activity, which depends on an IT foundation, is likely to suffer. Based on these theoretical arguments, we propose that, after an optimal level, additional investments in IT are likely to negatively impact the efficacy of firm innovation activity.

H2: Accumulation of IT capital and innovation performance are likely to have an inverted-U shaped relationship.

2.3 The Substitutability of Human and IT Capitals

Human capital and knowledge resources are the most frequently examined resources while IT capabilities are the most frequently examined capabilities pertinent for innovation (Newbert, 2007). A resource-based analysis of IT assets reveals that competitive advantages can only be achieved when IT capitals are combined with complementary knowledge and skills (Mata et al., 1995; Ray et al., 2004; 2005). As IT capitals facilitate organizational learning and augment organizational absorptive capacity, they yield a positive influence on firm financial and innovation performance (Joshi et al., 2010; Tippins and Sohi, 2003). Recent studies have deepened our understanding and uncovered the conditions under which IT investments and capabilities lead to better firm performance (Drnevich and Kriauciunas, 2011; Ho et al., 2011). Additionally, Kleis et al. (2011) further provided supportive evidence that IT capital in addition to R&D investment is important inputs to innovation output. The early usage of IT was to reduce personnel costs in labor-intensive operations such as accounting, purchasing and payroll. Then IT has affected all functions of modern enterprises and gradually displaced traditional labor and capital inputs through electronic transactions, electronic data interchange, office automation, sale-of-point applications, enterprise resource planning (ERP) systems, among others. In recent years, IT has also led to the restructuring and reengineering of business processes (Barua et al., 2004; Rai et al., 2006). Prior research has found a net substitute of IT capital for labor, reporting an Allen Elasticity of Substitution (AES) of 1.063 ($p < 0.05$) in 1988-1992, based on the constant elasticity of substitution translog (CES-Translog) production function (Dewan and Min, 1997). Similarly, Chwelos et al. (2010) confirmed that through labor substitution, the increasing factor share of IT comes at the expense of labor. Based on the CES-Translog production function, they reported a higher estimation of AES of 3.012 ($p < 0.01$) among a sample of over 800 Fortune 1000 firms in 1987-1998. While it lacks recent evidence of IT's substitution for labor, we expect a negative interaction effect between IT capital and human capital in knowledge intensive contexts as well. Modern IT applications can extensively facilitate knowledge management and innovation activity (Joshi et al., 2010), releasing the demand for labor to a large extent.

H3: Human capital and IT capital have a negative interaction effect on innovation performance, such that more human (IT) capital will reduce the main effect of IT (human) capital.

3 Methodology

3.1 Data

The data used to examine our hypotheses came from the Mannheimer Innovation Panel database provided by the Center for European Economic Research. This database includes data of innovation activities of firms in a variety of industries representing the entire economy of Germany. The data were collected in an annual base by surveying managers, which generate an unbalanced panel with unique identifier for each firm. However, IT investment data for all sectors are only available for 2003. Thus, the unbalanced panel allows us to design a causal test of hypotheses with time lags. We matched the data from five waves of survey in 2002-2006 by firm identifiers and sorted out variables to be used in our design. It results a sample of 349 firms distributed across all industries covered by the survey. Below we discuss the measures we chose to operationalize each variable and the temporal relationships among variables.

3.2 Measures

Innovation Performance: We selected the most widely used performance-based innovation measure, as the revenue from new or substantially improved products or services scaled by total sales (Hitt et al., 1996; Leiponen and Helfat, 2011; Woerter, 2009). In light the literature suggested that innovation processes typically last across years[1], we used three-year lagged innovation performance in 2006 as the dependent variable[2]. A larger lag can mitigate the concerns of common method bias of survey data and reverse causation of relationships between human and IT capitals to innovation performance. In addition, it can reduce endogeneity bias as it is less likely that any omitted variables simultaneously influence both capital investments and innovation performance three years later. A long lag also provides stronger evidence on the effects of human and IT capitals if any, which are not diminishing rapidly in a short run.

Human Capital: We took account of two components of human capital embedded in a firm's human resources. Education level of employees was extensively used in past research to represent accumulating intelligence of employment from recruitment (Bresnahan et al., 2002). In addition, the development of human resources is also important to accumulate human capital in a firm, which is typically achieved by investing in a variety of training programs for employees (Booth and Bryan, 2005;

[1] Joshi et al. (2010) used a three-year period to capture the innovation processes from staring commercialization to introduction to the market.

[2] We also tried two-year and one-year lagged innovation performance as the dependent variables, and found interesting lagged effect patterns of human and IT capitals. We discuss the findings in Section 5.

Gooderham, 1994). In order to thoroughly reflect the stock of human capital in a firm, we calculated the measure of human capital as the sum of standardized ratio of employees with college degree and standardized ratio of training investment over total sales in 2003. This measure incorporates the two sources of human capital from recruitment and training after recruitment.

IT Capital: We measured IT capital by total IT spending scaled by total sales in 2003. This is an overall indicator of a firm's IT intensity, similar to what has been used in past research (Hitt, 1999; Mata et al., 1995; Ray et al., 2004; 2005).

Control Variables: To control for potential confounds and rule out rival explanations, we controlled a number of variables related to innovation performance. The longitudinal data allow us to control for the past innovation performance reflecting a firm's innovation capabilities. Past innovation performance can serve as the proxy of firm heterogeneities in innovation processes influenced by unknown or omitted factors. Thus, we controlled innovation performance in 2002 (one year before measuring human and IT capitals) in our analysis. Mergers and acquisitions (M&A) may dramatically change a firm's innovation performance (Hitt et al., 1990; Marianna et al. 2010), which was controlled by a dummy variable indicating whether there was M&A occurred with the firm in 2000-2002. We also control for a few variables in 2003 impacting firm performance, such as market share, firm size by logarithm of sales, and age by a dummy variable indicating whether the firm is established for about or more than 3 years (value = 1), or is a new entrant less than 3 years old (value = 0). These variables can reflect the overall firm performance, richness of resources, and inertia along with time. Because our sample covers all sectors of Germany, the number of observations for each industry is relatively small. Adding industry dummies generated serious multicollinearity, which results in unstable estimators. Thus, we created two dummy variables to control for the fixed effects of high-technology vs. low-technology industries[3] and manufacturing vs. services sectors[4]. High-technology industries typically perform better than low-technology industries in innovation (Colombo et al., 2006). Manufacturing sectors are suggested to adopt distinct innovation processes compared to services sectors (Malerba, 2002). In addition, national or regional systems were suggested to affect innovation performance at the firm level (Freeman, 1995; Sharif, 2006), which was controlled by a dummy variable indicating the location of a firm in either East Germany (value = 1) or West Germany (value = 0). Table 1 shows descriptive statistics and correlation matrix of the variables.

[3] We created *Hi/LoTech* dummy by matching firms' NACE Rev 1.1 two-digit codes (similar to SEC industry classification systems in US) to the official classification of high-technology/low-technology industries by EuroStat.

[4] We created *Manu/Serv* dummy according to firms' NACE Rev 1.1 two-digit codes.

Table 1. Descriptive Statistics and Correlation Metrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) *InnoPerf*$_{t+3}$											
(2) *HumanCap*$_t$	0.238***										
(3) *ITCap*$_t$	-0.064	0.057									
(4) *InnoPerf*$_{t-1}$	0.639***	0.214***	-0.076								
(5) *M&A*$_{t-3\sim t-1}$	0.089	0.018	-0.028	0.117*							
(6) *MktShare*$_t$	-0.139**	-0.057	0.041	-0.164**	-0.032						
(7) *Size*$_t$	0.238***	-0.214***	-0.161**	0.192***	-0.031	-0.225***					
(8) *Age*$_t$	-0.024	-0.017	0.340***	0.048	-0.014	-0.049	-0.045				
(9) *Hi/LoTech*	0.301***	0.474***	0.024	0.298***	-0.034	-0.119*	-0.137*	-0.057			
(10) *Manu/Serv*	0.196***	-0.228***	-0.044	0.186***	0.041	-0.049	0.306***	-0.004	-0.019		
(11) *East/West*	0.069	0.246***	-0.049	0.131*	0.069	0.087	-0.110*	0.008	0.123*	-0.079	
Mean	0.018	-0.269	0.018	0.018	0.032	0.794	2.148	0.006	0.364	0.524	0.447
SD	0.025	1.446	0.087	0.026	0.175	0.223	1.643	0.076	0.482	0.500	0.498
Min	0	-2.019	0	0	0	0.200	-1.661	0	0	0	0
Max	0.080	6.946	0.800	0.080	1	1	6.384	1	1	1	1

Note: N = 349. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

4 Results

We used OLS regression to examine the inverted-U hypotheses by the square terms of human and IT capitals (Katila and Ahuja, 2002; Lahiri, 2010). The interaction of human and IT capitals was detected by the product term of their mean-centered values. Table 2 reports the regression results for hypothesis testing and full model. We took a step-wise procedure and estimated a control model first. To examine H1, we added human capital to the control model, and then additionally added human

Table 2. Regression Results for Innovation Performance

	Controls Only	Partial Model (H1)		Partial Model (H2)		Partial Model (H3)	Full Model
$HumanCap_t$		0.002*	0.003***			0.003**	0.002*
		(0.001)	(0.001)			(0.001)	(0.001)
$HumanCap_t^2$			-0.001**				-0.001**
			(0.0003)				(0.0004)
$ITCap_t$				0.003	0.132*	-0.004	0.219*
				(0.012)	(0.065)	(0.013)	(0.090)
$ITCap_t^2$					-0.166*		-0.277*
					(0.082)		(0.117)
$HumanCap_t$						-0.001	-0.008
$\times$ $ITCap_t$						(0.007)	(0.009)
$InnoPerf_{t-1}$	0.521***	0.506***	0.491***	0.522***	0.518***	0.503***	0.492***
	(0.042)	(0.042)	(0.042)	(0.043)	(0.042)	(0.043)	(0.042)
$M\&A_{t-3\sim t-1}$	0.005	0.004	0.004	0.005	0.005	0.004	0.005
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
$MktShare_t$	0.001	0.001	0.003	0.001	0.0003	0.002	0.003
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$Size_t$	0.002**	0.002***	0.002**	0.002**	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age_t	-0.011	-0.011	-0.011	-0.012	-0.011	-0.013	-0.013
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)
Hi/LoTech	0.009***	0.006*	0.006*	0.009***	0.008***	0.006*	0.005*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Manu/Serv	0.003	0.004	0.004*	0.003	0.003	0.004	0.004*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
East/West	-0.0003	-0.001	-0.002	-0.003	0.00001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.002)
Adjusted R^2	0.438	0.447	0.456	0.436	0.441	0.446	0.461
F	34.85***	32.21***	30.18***	30.90***	28.47***	26.47***	23.87***

Note: N = 349. * $p< 0.05$; ** $p < 0.01$; *** $p < 0.001$. Standard errors are in parentheses.

capital squared. It was found that human capital had significantly positive effect on innovation performance, while human capital squared had significantly negative effect on innovation performance. Thus, H1 was supported. Similarly, we added IT capital and then IT capital and its square term to the control model to test H2. Consistent with prior findings (e.g., Mata et al., 1995; Ray et al., 2004; 2005), we did not observe significant effect of IT capital on innovation performance. After additionally adding IT capital squared to the model, however, IT capital became positively significant and IT capital squared was negatively significant. Thus, a curvilinear relationship between IT capital and innovation performance better fits the data and demonstrates significant result. Thus, H2 was supported. For H3, we added human capital, IT capital, and their mean-centered product term to the control model but failed to observe significant interaction effect. The last column of Table 2 reports the result for full model, consistent with previous tests to support H1 and H2, but not to support H3. The model demonstrates good fit being able to explain 46% variation in innovation performance.

5 Conclusion

Our empirical results lend support to the proposed curvilinear relationships between human and IT capitals to firm innovation. After an optimal level, excessive human capital investment is likely to be detrimental to firm innovation outcomes. Similarly, excessive accumulation of IT capital is likely to increase complexity and heterogeneity in organizational communication, which are counterproductive to innovation as well. However, we failed to observe significant interaction of human and IT capitals, suggesting a lack of resource substitutability in German firms. To our best knowledge, this is the first study to explicitly theorize and test the curvilinear effects of accumulating strategic resources. In addition, our focus on innovation rather than firm financial performance enables us to examine the theoretical predictions based on RBV more closely (Newbert, 2007). In the meanwhile, this study has several limitations, which suggest room for future research. The results presented are based on the analysis of data collected by survey, which is largely self-reported data. Also, while we leveraged causal design with time lags, we are unable to adopt a longitudinal research design. Future studies may examine more closely the mechanisms underlying our proposed relationships. For instance, researchers may choose to study how the accumulation of IT and human resource affects the rate of increase in the dynamic adjustment and integration costs over time. Such an investigation could improve our understanding of the dynamics of the tacit processes underlying innovations.

References

1. Aaronson, D., Barrow, L., Sander, W.: Teachers and student achievement in the Chicago public high schools. Journal of Labor Economics 25, 95–135 (2007)
2. Ackoff, R.L.: Management misinformation systems. Management Science 14, 147–156 (1967)

3. Alavi, M., Leidner, D.: Knowledge management and knowledge management systems: Conceptual foundations and research issues. MIS Quarterly 25, 107–136 (2001)
4. Argote, L., McEvil, W., Reagans, R.: Managing knowledge in organizations: An integrative framework and review of emerging themes. Management Science 49, 571–582 (2003)
5. Barney, J.B.: Strategic factor markets: Expectations, luck and business strategy. Management Science 32, 1231–1241 (1986)
6. Barney, J.B.: Firm resources and sustained competitive advantage. Journal of Management 17, 99–120 (1991)
7. Barney, J.B., Arikan, A.M.: The resource-based view: Origins and implications. Blackwell Publishers Inc., Malden (2001)
8. Barua, A., Konana, P., Whinston, A.B., Yin, F.: An empirical investigation of net-enabled business value. MIS Quarterly 28, 585–620 (2004)
9. Booth, A.L., Bryan, M.L.: Testing some predictions of human capital theory: New training evidence from Britain. Review of Economics and Statistics 87, 391–394 (2005)
10. Bresnahan, T.F., Brynjolfsson, E., Hitt, L.M.: Information technology, workplace organization, and remand for skilled labor: Firm-level evidence. Quarterly Journal of Economics 117, 339–376 (2002)
11. Broadbent, M., Weill, P., Clair, D.S., Kearney, A.T.: The implications of information technology infrastructure for business process redesign. MIS Quarterly 23, 159–182 (1999)
12. Brynjolfsson, E., Hitt, L.M.: Paradox lost? Firm-level evidence on the returns to information systems spending. Management Science 42, 541–558 (1996)
13. Chwelos, P., Ramirez, R., Kraemer, K.L., Melville, N.P.: Does technological progress alter the nature of information technology as a production input? New evidence and new results. Information Systems Research 21, 392–408 (2010)
14. Cohen, W.M., Levinthal, D.A.: Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly 35, 128–152 (1990)
15. Colombo, M.G., Grilli, L., Piva, E.: In search of complementary assets: The determinants of alliance formation of high-tech start-ups. Research Policy 35, 1166–1199 (2006)
16. Damanpour, F.: Organizational innovation: A meta-analysis of effects of determinants and moderators. Academy of Management Journal 34, 555–590 (1991)
17. Davenport, T.H., Prusak, L.: Working knowledge: How organizations management what they know. Harvard Business School Press, Cambridge (1998)
18. Dewan, S., Min, C.K.: The substitution of information technology for other factors of production: A firm level analysis. Management Science 43, 1660–1675 (1997)
19. Drnevich, P.L., Kriauciunas, A.P.: Clarifying the conditions and limits of the contributions of ordinary and dynamic capabilities to relative firm performance. Strategic Management Journal 32, 254–279 (2011)
20. Freeman, C.: The "national system of innovation" in historical perspective. Cambridge Journal of Economics 19, 5–24 (1995)
21. Gattiker, T.F., Goodhue, D.L.: What happens after ERP implementation: Understanding the impact of interdependence and differentiation on plant-level outcomes. MIS Quarterly 29, 559–585 (2005)
22. Gooderham, P.: Human capital in organizations: Competence, training and learning. Oxford University Press, New York (1994)
23. Goodhue, D.L., Wybo, M.D., Kirsch, L.J.: The impact of data integration on the costs and benefits of information systems. MIS Quarterly 16, 293–311 (1992)

24. Grant, R.M.: Toward a knowledge-based theory of the firm. Strategic Management Journal 17, 109–122 (1996)
25. Grise, M.L., Gallupe, R.B.: Information overload: Addressing the productivity paradox in face-to-face electronic meetings. Journal of Management Information Systems 16, 157–185 (2000)
26. Hay, D.A., Morris, D.J.: Industrial economics and organization: Theory and evidence. Oxford University Press, New York (1991)
27. Hitt, L.M.: Information technology and firm boundaries: Evidence from panel data. Information Systems Research 10, 134–149 (1999)
28. Hitt, M.A., Bierman, L., Shimizu, K., Kochhar, R.: Direct and moderating effects of human capital on strategy and performance in professional service firms: A resource-based perspective. Academy of Management Journal 44, 13–28 (2001)
29. Hitt, M.A., Hoskisson, R.E., Ireland, R.D.: Mergers and acquisitions and managerial commitment to innovation in m-form firms. Strategic Management Journal 11, 29–47 (1990)
30. Hitt, M.A., Hoskisson, R.E., Johnson, R.A., Moesel, D.D.: The market for corporate control and firm innovation. Academy of Management Journal 39, 1084–1119 (1996)
31. Ho, J.L.Y., Wu, A., Xu, S.X.: Corporate governance and returns on information technology investment: Evidence from an emerging market. Strategic Management Journal 32, 595–623 (2011)
32. Itami, H.: Mobilizing invisible assets. Harvard University Press, Cambridge (1987); Joshi, K. D., Chi, L., Datta, A., Han, S.: Changing the competitive landscape: Continuous innovation through IT-enabled knowledge capabilities. Information Systems Research 21, 472–495 (2010)
33. Katila, R., Ahuja, G.: Something old, something new: A longitudinal study of search behavior and new product introduction. Academy of Management Journal 45, 1183–1194 (2002)
34. Kleis, L., Chwelos, P., Ramirez, R.V., Cockburn, I.: Information technology and intangible output: The impact of IT investment on innovation productivity. Information Systems Research (2011) (forthcoming)
35. Kogut, B., Zander, U.: Knowledge of the firm, combinative capabilities, and the replication of technology. Organization Science 3, 383–397 (1992)
36. Krishnan, M.S., Rai, A., Zmud, R.: The digitally enabled extended enterprise in a global economy. Information Systems Research 18, 233–236 (2007)
37. Langley, A., Mintzberg, H., Pitcher, P., Posada, E., Saint-Macary, J.: Opening up decision making: The view from the black stool. Organization Science 6, 260–279 (1995)
38. Lahiri, N.: Geographic distribution of R&D activity: How does it affect innovation quality? Academy of Management Journal 53, 1194–1209 (2010)
39. Leiponen, A., Helfat, C.E.: Location, decentralization, and knowledge sources for innovation. Organization Science 22, 641–658 (2011)
40. Mahoney, J.T.: Economic foundations of strategy. Sage, Thousand Oaks (2005)
41. Malerba, F.: Sectoral systems of innovation and production. Research Policy 31, 247–264 (2002)
42. Marianna, M., Hitt, M.A., Lane, P.J.: Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions. Strategic Management Journal 31, 602–628 (2010)
43. Mata, F.J., Fuerst, W.L., Barney, J.B.: Information technology and sustained competitive advantage: A resource-based analysis. MIS Quarterly 19, 487–505 (1995)

44. Miller, D.J., Fern, M.J., Cardinal, L.B.: The use of knowledge for technological innovation within diversified firms. Academy of Management Journal 50(2), 308–326 (2007)
45. Newbert, S.L.: Empirical research on the resource-based view of the firm: An assessment and suggestions for future research. Strategic Management Journal 28, 121–146 (2007)
46. Nonaka, I., Takeuchi, H.: The knowledge-creating company. Oxford University Press, New York (1995)
47. Ocasio, W.: Towards an attention-based view of the firm. Strategic Management Journal 18, 187–206 (1997)
48. Penrose, E.T.: The theory of the growth of the firm. Wiley, New York (1959)
49. Ployhart, R.E., Weekley, J.A., Ramsey, J.: The consequences of human resource stocks and flows: A longitudinal examination of unit service orientation and unit effectiveness. Academy of Management Journal 52, 996–1015 (2009)
50. Rai, A., Patnayakuni, R., Seth, N.: Firm performance impacts of digitally enabled supply chain integration capabilities. MIS Quarterly 30, 225–246 (2006)
51. Ray, G., Barney, J.B., Nuhanna, W.A.: Capabilities, business processes, and competitive advantage: Choosing the dependent variable in empirical tests of the resources-based view. Strategic Management Journal 25, 23–37 (2004)
52. Ray, G., Muhanna, W.A., Barney, J.B.: Information technology and the performance of the customer service process: A resource-based analysis. MIS Quarterly 29, 625–652 (2005)
53. Sambamurthy, V., Bharadwaj, A., Grover, V.: Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. MIS Quarterly 27, 237–264 (2003)
54. Sambamurthy, V., Zmud, R.W.: The organizing logic for an enterprise's IT activities in the digital era: A prognosis of practice and a call for research. Information Systems Research 11, 105–114 (2000)
55. Saraf, N., Langdon, C.S., Gosain, S.: IS application capabilities and relational value of interfirmpartnerships. Information Systems Research 18, 320–339 (2007)
56. Schumpeter, J.A.: The theory of economic development. Harvard University Press, Cambridge (1934)
57. Sharif, N.: Emergence and development of the national innovation systems concept. Research Policy 35, 745–766 (2006)
58. Simon, H.A.: Administrative behavior: A study of decision-making processes in administrative organizations. Macmillan, New York (1947)
59. Tan, D., Mahoney, J.T.: Examining the Penrose effect in an international business context: The dynamics of Japanese firm growth in U.S. industries. Managerial and Decision Economics 26, 113–127 (2005)
60. Teece, D.J.: Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. Research Policy 15, 285–305 (1986)
61. Tippins, M.J., Sohi, R.S.: IT competency and firm performance: Is organizational learning a missing link? Strategic Management Journal 24, 745–761 (2003)
62. Weill, P., Ross, J.W.: IT governance: How top performers manage IT decision rights for superior results. Harvard Business School Press, Cambridge (2004)
63. Woerter, M.: Industry diversity and its impact on the innovation performance of firms: An empirical analysis based on panel data (firm-level). Journal of Evolutionary Economics 19, 675–700 (2009)
64. Wright, P.M., Dunford, B.B., Snell, S.A.: Human resources and the resource based view of the firm. Journal of Management 27, 701–721 (2001)

Attention-Aware Collaboration Modeling

Shaokun Fan[1,2] and J. Leon Zhao[1]

[1] Department of Information Systems, City Universtiy of Hong Hong,
Kowloon Tong, Hong Kong
[2] Department of Management Information Systems, The University of Arizona,
Tuson, AZ, USA
fsk@email.arziona.edu, jlzhao@cityu.edu.hk

Abstract. Recently, a great variety of web-based collaboration support technologies (CSTs) have become available for people to collaborate for various purposes. On the other hand, CSTs are leading to more attention stress — more and more people are becoming overwhelmed by many simultaneous projects and the associated tasks. However, little research has been done on how to design collaboration management mechanisms that can help managers control collaboration activities for better collective efficiency. We lay the foundation of research in this regard by developing a model of team collaboration while emphasizing the attention aspects of collaboration, which we refer to as Attention-Aware Collaboration Modeling (AACM). In this paper, we present core concepts and basic principles of attention-aware collaboration management based on Attention Economy Theory.

Keywords: Attention-aware, Collaboration modeling, Attention stress.

1 Introduction

Collaboration management has been proposed to enhance collaboration efficiency (Schuster et al, 2000). However, most existing collaboration management and support mechanisms are based on workflow and groupware systems. Collaboration activities, which are ad hoc and dynamic in nature, cannot be managed very well in such systems. Further, advanced collaboration support technologies have opened a vast set of opportunities for people to collaborate with each other.

Meanwhile, advanced collaboration support technologies such as online meeting systems also created more attention stress: people are being overwhelmed by many tasks and human attention is the scarcest resource. Unfortunately, attention management has not been emphasized in commercial collaborative support systems, rendering them ineffective in helping managers control collaborative attention. We believe that research is needed in this direction. In this paper, we develop an AACM approach for attention management.

The contributions of this paper are threefold: First, we indentify the problem of attention-aware collaboration management and specify requirements for attention-aware modeling. Second, we propose an attention-aware modeling approach to support collaboration management. Third, principles of efficient attention-aware

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 347–355, 2012.

collaboration management are investigated. To the best of our knowledge, this is the first practical endeavor to embed attention-aware management as a key component in collaboration management.

2 Research Background

Traditional workflow management mechanisms and groupware are widely used by collaborative teams for collaboration support. Each of these mechanisms focuses on various aspects of collaboration and different types of collaboration models are embedded in these systems. For example, Collaboration Management Model (Schuster et al., 2000) draws existing primitives from workflow and groupware models and introduces new primitives for previously unsupported collaboration process requirements. However, to support effective collaboration management, a model of collaboration needs to be defined at a level of granularity, which covers all the major components and mechanisms of team collaboration.

As information in working place has grown increasingly abundant and immediately available, attention becomes the limiting factor in the consumption of information. Based on research in cognitive science and psychology, Attention Economics Theory (Davenport, 2001) defines attention as “focused mental engagement on a particular object of interest”. The key argument for attention economy theory is that attention is a limited resource and attention management has a direct impact of company revenue. Further, human attention, as an important resource in organizations, can be scheduled, shifted and protected. Attention models have also been studied in IS research (Horvitz et al., 2003), where attention was described by Bayesian models and take as inputs sensors that provide streams of evidence about attention and provide a means for computing probability distributions over a user’s attention and intentions.

Similar to other collaboration management approaches, the model proposed in this paper also focuses on collaboration activity and process management. The unique feature of our approach is that human attention is used as a key factor to schedule, monitor and improve collaboration.

3 Problem Specification and Requirement Analysis

In this section, we use collaborative software design as an example to illustrate the problems of collaboration management under attention stress. Consider that a group of seven people from a software company are assigned to the task of designing an intelligent campus system. Out of the ten people, one person is designated as the project manager. The rest six people are divided into three teams (with 1,2,3 participants respectively) and each team has a team leader. The process of software design has four main steps, i.e. requirement analysis, architecture design, individual module design, and budgeting. These main tasks can be further divided into more detailed sub-tasks. For example, the design of each module (e.g. financial system, student management, and so on) can be a sub task of the individual module design task. Figure 1 describes the collaborative software design case.

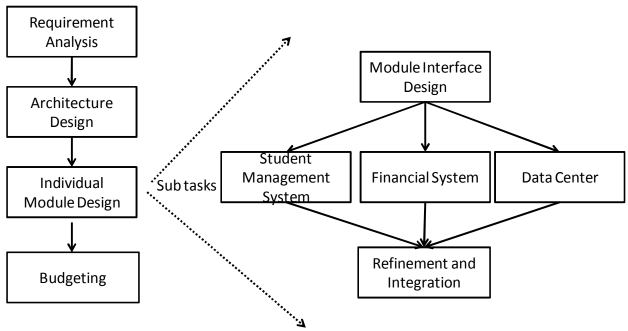


Fig. 1. The Process of Intelligent Campus System Design

Each subtask is assigned to a team. But, design of a software component may depend on another component. People working on different components must communicate with one another and provide support. Otherwise, the efficiency or quality of collaboration will be compromised due to task dependency. The task assignment relationships are fuzzy here, i.e. those who are not yet assigned to a task may support the task implicitly.

If we use existing project or workflow management approaches to manage this collaboration process, people are either assigned or non-assigned to a particular task. On one hand, if we assign a task to all members who should contribute to it, we will end up with an assignment plan that the ten people are assigned to all sub tasks. On the other hand, if our assignment plan only considers the people who are directly responsible for the tasks, the collaboration relationship among people will not be captured. In either way mentioned above, workload of individual collaboration participants cannot be explicitly represented, controlled and scheduled. Further, during the execution of the collaboration process, managers may find that every member in the group are participating many tasks but the degree of participation (different levels of workers, different levels of knowledge supporters,) cannot be managed. This could cause inappropriate attention allocation, i.e. some members participate in many tasks heavily while other members only slightly participate in the tasks.

Because software design is knowledge intensive work and people must process considerable quantities of information in order to get their jobs done. If attention stress is not managed very well, collaboration efficiency will be reduced. Although they are generally only working on a single task at any instant in time, the combination of cognitively intensive processing, considerable quantities of information and multitasking make knowledge work extremely challenging. From the previous discussion, the following requirements for attention-aware collaboration management emerge:

1) Attention, as a key resource in collaboration, should be scheduled for tasks.
2) Attention status should be tracked at different levels of granularities (individual, team and organizational) in collaboration processes.
3) Guidelines for efficient attention scheduling in collaboration should be provided.

To address the attention stress problems in collaboration, we propose an attention-aware collaboration model that enables collaboration management processes model with specialized attention-aware mechanism. These are discussed in detail in the remaining of the paper.

4 Attention-Aware Collaboration Modeling

In this section, we design an attention aware modeling mechanism, in which attention is treated as a four-dimensional resource in collaboration processes. In Fig. 2, three aspects are included in the model: organization structure, collaboration tasks and attention.

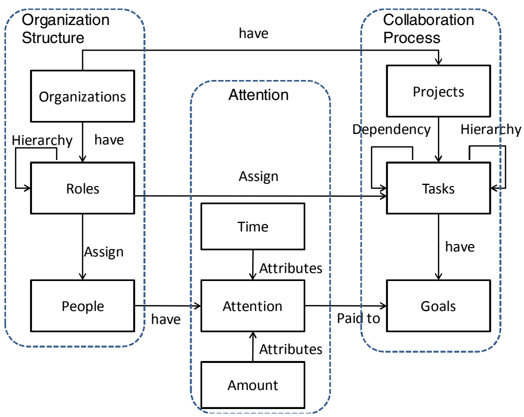


Fig. 2. An Attention-aware collaboration model

A collaboration organization is a group of participants who perform activities together in order to achieve a particular goal. For instance, a software development group includes a group of people who perform software development works. For information systems development, organization is a fundamental concept for analysis of business functions and activities. The concept of role is well-established in the literature (Sandhu, 1996), it can be considered as a job function within the organization that describes the responsibility to the role. Most formal organizations describe their structure by means of an organization chart, which describes a strict partially ordered set of named positions. An example of organization chart is shown in Figure 3. It is in the form of a rooted tree, with the root at the top of the organization.

Definition 1. An **organization structure** (*OS*) is defined as a four-tuple $OS=(P, R, PR, RH)$,

- $P=\{p_1, p_2, \ldots\}$ is a set of people; $R=\{r_1, r_2, \ldots\}$ is a set of roles;
- $PR \subseteq P \times R$: is a binary relation, describes the people role assignment relationships;
- $RH \subseteq R \times R$: Role hierarchy is a strict partial ordered tree structure, where $\forall r \in R$, has only one direct parent.

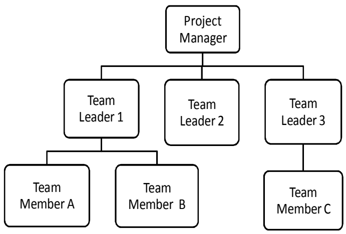


Fig. 3. Role Hierarchy

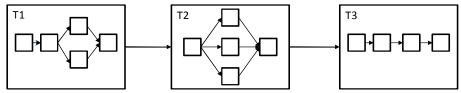


Fig. 4. Task Hierarchy and Task Dependency

Collaboration processes are a fundamental part of collaborative work and require thorough management. In traditional workflow management systems, a process can be defined by a directed graph of connected tasks using process modeling languages such as Petri-Nets or UML-Activity diagrams. Tasks are logical sub steps of a process. They may be either manual or automated in nature. Tasks can be performed by one or many users. In this paper our focus is on attention-aware collaboration modeling. Therefore, task modeling is an important aspect of collaboration modeling. In order to reach business goals, tasks need to be executed.

Definition 2. A task has a goal set that needs to be accomplished within a defined period of time and is represented by a three-tuple task= (tid, time, GOAL), where

- tid is the task id and each task has an unique id;
- time=(start, end), is the planed time period for this task;
- GOAL={$goal_1$, $goal_2$ ..} is a set of goals of tasks.

Definition 3. A project is a set of tasks with hierarchies and dependencies, which is represented by a three tuple project= (TASK, TD, TH), where

- TASK={t_1, t_2,…}is a set of tasks;
- Task Dependency TD$\subseteq$T×T, is a binary relation, describes dependency relationships between tasks;
- Task Hierarchy TH$\subseteq$T×T, is a binary relation, describes the relationship of subtasks. It is a strict partial ordered tree structure, where $\forall t \in T$, has only one direct parent.

Unfortunately, in most WfMSs the user is assumed to work on one single task at a time. In this paper we propose a far less rigid approach where allows users to be partially assigned to a task by dividing users' attention. Attention is best described as the sustained focus of cognitive resources on information while filtering or ignoring extraneous information. Attention is a very basic function that often is a precursor to all other neurological/cognitive functions. Attention has been referred to as the allocation of processing resources (Anderson 2004). One of the most used attention models are proposed by Sohlberg and Mateer (1989) where five different types of attention are defined: focused attention, sustained attention, selective attention, alternating attention, and divided attention. In a recent review, Knudsen (2007) describes a general model which identifies four core processes of attention, with working memory at the center. We first define attention and attention stress as the following. Three assumptions based on theoretical research in the literature are first listed below:

1. **Divided Attention Assumption:** Individual's attention can be divided into different portions and assigned to different tasks at the same time. E.g., people can talk on a cell phone conversation while driving a car.
2. **Limited Attention Assumption:** In a certain time period, one individual's attention is limited. E.g., one person cannot talk to two people on the phone at the same time.
3. **Attention Performance Assumption:** the amount of attention put into a task is proportional to the performance of the tasks if we assume everything else keeps the same. E.g., People can drive the car better if they do not talk on the phone.

Definition 4. Attention is one's mental engagement on certain task. It is represented by a four-tuple: Attention= (p, amount, time, goal), where

- $p \in P$, is the person whose attentional resource is.
- $1 \geq amount \geq 0$, is the portion of one's attention that is paid to a specific goal of a task .
- time=(start, end), is a time period represented by a start time and a end time
- $goal \in GOAL$, is the sub-goal of a task.

Here, we model attention as a four-detention resource that connects essential components in collaboration such as people, tasks, time and role. In our model, attention is the key resource to manage in collaboration. It belongs to a person in collaboration and is a key precondition for any collaboration tasks. The unique characteristic of attention is that it can be divided into portions so that multiple tasks/goals may be achieved at the same time.

5 Principles for Efficient Collaboration Management

In order to achieve better collaboration efficiency, we need to explicitly represent and measure attentional resources for individuals, teams and tasks. Attention can be measured by different approaches. For example, Tarzia el al. (2009) proposed a Sonar-based measurement to determine the presence and attention levels of computer users. In Davenport (2001), he proposed that the best way to measure attention is to ask employees themselves. Attention as a key resource in collaboration can be measured at different levels. Four levels of attention measurement are defined as the following:

Definition 5. Individual attention is defined as the amount of attention spent with a time period. Formally, it is represented by $IndividualAttention(p_i, time) = \sum_{\forall \text{ Attention, Attention.p=pi\&\&Attention.time} \subseteq \text{time}} (\text{Attention.amount* Attention.time})$.

Definition 6. Team level attention is defined as the portion of attention spent with a time period. Formally, it is represented by $TeamAttention\ (r_i, time) = \sum_{\forall \text{ Attention, Attention.p=pi.\&\& (pi, ri)} \in \text{PR\&\&Attention.time} \subseteq \text{time}} (\text{Attention.amount* Attention.time})$.

Definition 7. Task level attention is defined as the portion of attention spent on the task. Formally, it is represented by $TaskAttention\ (t_i, time) = \sum_{\forall \text{ Attention, Attention.goal} \in \text{Task.Goals \&\&Attention.time} \subseteq \text{time}} (\text{Attention.amount* Attention.time})$.

Definition 8. Measurement of attention stress is defined by the number of tasks that one individual needs to pay attention to at the same time. Formally, it is represented by *Attention Stress* $(p_i,$ *time)* *=Count* ($\forall$ *Attention*, *attention.p*$=p_i$*&&attention.time*$\subseteq$*time), where Count is the function that counts the numbers of occurrences that satisfy the conditions.*

Based on these measurements, we propose three principles for attention-aware collaboration management.

Principle 1 (Appropriate attention stress level): In any time period, all collaboration participants should be kept at an appropriate attention stress level. Formally, $\forall p_i$, *time, Attention Stress* $(p_i$, *time)* $\leq a$, *where a is the appropriate level.*

Although collaboration efficiency can enhance through parallelism, people cannot work on too many tasks based on working memory theory. If collaboration participants join too many tasks at the same time period, they may suffer from information overload and reduce collaboration efficiency. Further, interruptions from different tasks my lead to waste of working time. Therefore, the appropriate level of attention stress relies on factors such as individual ability, collaboration tool support, and complexity of tasks and needs to be empirically accessed.

Principle 2 (Focus attention on key tasks): Key tasks in the project should get more attentional resources. Formally, *Key*(t_i) $\rightarrow$ *TaskAttention* $(t_i$, *time)*$>$ $\sum_{\forall tj\ \in TASK}$ *TaskAttention* $(t_i$, *time)/number of tasks, where Key() is a function to determine whether the task is a key task.*

Not all tasks in a project are equally important. Some tasks may require more innovative efforts and are more important than others. The performance of such key tasks has a direct impact to the project result. Thus, attentional resources allocated to key tasks should be more than the average level.

Principle 3 (Schedule attention based on tasks dependency): Attention allocated to tasks should be consistent with tasks dependencies. Formally, $(t_i, t_j) \in TD$*&&* *Attention*$_i$*.goal*$\in t_i$*.GOAL* *&&* *Attention*$_j$*.goal*$\in t_j$*.GOAL* $\rightarrow$ *Attention*$_i$*.time.start*$\leq$ *Attention*$_j$*.time.start* *&&* *Attention*$_i$*.time.end*$\leq$ *Attention*$_j$*.time.end*

Coordination theory (Malone and Crawson, 1994) claims that resource scheduling should consider dependencies that constrain how tasks can be performed. Attention scheduling is an important aspect of collaboration management. Efficient attention scheduling will minimize waiting time and enhance collaboration efficiency. Attention scheduling should be consistent with time scheduling so that the waste of attentional resources can be reduced. Attention-based measurement defined in this section can be used for collaboration management in the following ways:

- Collaboration planning: when project manager starts to plan and schedule task, three principles can be used as guidelines for managers to enhance collaboration efficiency.
- Collaboration monitoring: During the process of collaboration, managers can have an overview of attention allocation within the group and keep track of attention-based performance of individuals, groups, and tasks.

- Collaboration evolution: When unexpected events happen in collaboration processes, managers can reallocate attentional resources based on these collaboration efficiency principles.

Next, we apply our approach into the case of collaborative intelligent campus system design to illustrate the usefulness of the proposed attention-aware collaboration model. Due to the space limit, we only use part of the case to show how the model can be used. Part of the personnel hierarchy, attention, tasks hierarchy and dependency are first modeled as follows:

Suppose the organization structure is similar to the one in Figure 3. People are represented as p_1, $p_2 \ldots p_7$; p_1 is the project manager; p_2 -p_4 are team leaders, and the rest are the team members.

P= $\{p_1, p_2 \ldots p_7\}$

R={manager, leader1, leader2, leader3,member A, member B, member C}

PR= {(p1, manager), (p2, leader1), (p3, leader2), (p4 leader3), (p5, member A), (p6, member B) (p7, member C) }

RH={(manager, leader1), (manager, leader2), (manager, leader3), (leader1, member A),(leader1, member B),(leader3, member C)}

We further assume goals of each task as the following:

TASK={(1, (1,10), {g11,g12,…}), (2, (1,3), {g21,g22}), (3, (3,7), {g31}), (4, (3,7), {g41,g42}), (5, (3,8), {g51}), (6, (9,10), {g61,g62}), }

TD={(2,3),(2,4),(2,5), (3,6), (4,6) (5,6)}

TH= {(1,2), (1,3) (1,4) (1,5) (1,6)}

If the project manager designs an attention plan as is shown in Table 1:

Table 1. An Example of Attention Scheduling Plan

	P_1	P_2	P_3	P_4	P_5	P_6	P_7
g21	(0.5,(1,5))	(0.8,(1,3))	(0.8,(1,3))	(0.8,(1,3))	(0.3,(1,3))	(0.2,(1,3))	(0.3,(1,3))
g22	(0.5,(1,3))	(0.2,(1,3))	(0.4,(1,3))	(0.1,(1,3))	(0.2,(1,3))		
g31	(0.2,(3,7))	(0.7,(3,7))	(0.1,(3,7))	(0.1,(3,7))	(0.8,(3,7))	(0.8,(3,7))	(0.2,(3,7))
g41	(0.4,(3,7))	(0.1,(3,7))	(0.4,(3,7))	(0.2,(3,7))		(0.7,(3,7))	(0.2,(3,7))
g42	(0.2,(3,7))	(0.1,(3,7))	(0.3,(3,7))	(0.2,(3,7))	(0.1,(3,7))		
g51	(0.1,(3,8))	(0.1,(3,8))	(0.1,(3,8))	(0.4,(3,8))	(0.1,(3,8))	(0.1,(3,8))	(0.5,(3,8))
g61	(0.2,(9,10))	(0.2,(9,10))	(0.4,(9,10))	(0.2,(9,10))			(0.8,(9,10))
g62	(0.2,(9,10))	(0.2,(9,10))	(0.4,(9,10))	(0.4,(9,10))	(0.6,(9,10))	(0.7,(9,10))	

Based on the above formal model, attention can be measured at different levels and the assignment plan can be calculated according to our definitions. For example,

$IndividualAttentention(p_5)=(0.2+0.3)*3+(0.8+0.1)*5+0.6*2=7.2$

$AttentionStress(p_5, (3,7))=3$

$TaskAttention\ (t_4, (3,7))=(0.4+0.2+0.1+0.1+0.4+0.3+0.2+0.2+0.1+0.7+0.2)*5=14.5$

Some initial conclusions can be made based on calculating different levels of attention.

- If t_4 is the most important tasks in this project, then this attention scheduling plan is efficient according to Principle 2. Otherwise, it is inefficient.
- During the time period of (3,7), members have the highest attention stress. If the appropriate level of attention stress is 2, then this plan is not efficient according to Principle 1.
- Attention (P_1, 0.5,(1,5), g21) is not efficient because it is not consistent with Principle 3.

6 Conclusions

In this paper, we propose an Attention-Aware Collaboration Modeling (AACM) approach where attention is represented as a four-dimension resource. Principles for efficient collaboration management are also investigated based on AACM. To the best of our knowledge, this is the first attempt to operationalize attention management in collaboration processes for better team efficiency. Our follow-up studies include (1) development of algorithms for enforcing attention-aware principles towards computer-supported collaboration management and (2) implementation of a prototype system that can facilitate attention-aware collaboration management.

Acknowledgement. This work was supported by NSF under grant 07-505 and Hong Kong Public Policy Research under grant cityU1015-PPR-10.

References

1. Anderson, J.: Cognitive psychology and its implications, 6th edn., p. 519. Worth Publishers (2004)
2. Davenport, T., Beck, J.: The Attention economy. Harvard Business School Press (2001)
3. Horvitz, E., Kadie, C.M., Paek, T., Hovel, D.: Models of Attention in Computing and Communications: From Principles to applications. Communications of the ACM 46(3) (2003)
4. Knudsen, E.I.: Fundamental Components of Attention. Annual Rev. of Neuroscience 30(1) (2007)
5. Malone, T.W., Crowston, K.: The interdisciplinary study of coordination. Computing Surveys 26(1), 87–119 (1994)
6. Sandhu, R.S., et al.: Role-Based Access Control Models. IEEE Computer 29(2), 38–48 (1996)
7. Schuster, H., Baker, D., Cichocki, A., Georgakopoulos, D., Rusinkiewicz, M.: The collaboration management infrastructure. In: International Conference on Data Engineering, San Diego (2000)
8. Sohlberg, M., Mateer, C.: Introduction to cognitive rehabilitation: theory and practice. Guilford Press, New York (1989) ISBN 0-89862-738-9
9. Tarzia, S., et al.: Sonar-based measurement of user presence and attention. In: The 11th International Conference on Ubiquitous Computing (2009)

Human–Software Agent Negotiations: An Experimental Study

Rustam Vahidov, Gregory E. Kersten, and Raafat Saade

InterNeg Research Centre, Concordia University,
Montreal, Canada
{rvahidov,gregory,rsaade}@jmsb.concordia.ca

Abstract. Negotiation is a powerful mechanism for facilitating effective economic exchanges. Electronic negotiations allow participants to negotiate online and use analytical support tools in making their decisions. Software agents offer the possibility of automating negotiation process using these tools. This paper aims at investigating the prospects of agent-to-human negotiations in B2C contexts using experiments with human subjects. Various types of agents have been configured and paired up with human counterparts for negotiating product sale. The paper discusses the results obtained both in terms of objective, as well as subjective measures.

Keywords: Electronic negotiations, Software agents, Automated negotiations, Experiments.

1 Introduction

Negotiation is an important mechanism for facilitating economic transactions. In the course of negotiations parties exchange offers in order to jointly explore the possibilities of finding acceptable solutions. Negotiations involving more than issue allow for more degrees of freedom in search for agreements that would be beneficial to the negotiators due to the asymmetry of their preference structures.

Online negotiations supported by electronic negotiation systems allow the parties exchange offers over the internet [1]. In addition to enabling anytime/anywhere mode of interactions, they may also incorporate analytical facilities for supporting negotiators in their preparation and conduct of negotiations. This support can range from such tools as those for capturing and modeling negotiator's preferences, to providing active advice and critique, and all the way to complete automation of the negotiation conduct.

Despite early optimistic expectations of the growth of negotiations as one of the primary mechanisms of conducting online transactions, in reality only few commercial sites offer such capabilities to their customers. One such website that allows customers to make (a limited number of) offers is Priceline.com. One possible

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 356–367, 2012.

explanation to the scarcity of negotiating websites is that negotiations imply a relatively high cognitive load, especially if multiple issues are involved (e.g. price, warranty, product attributes, shipment, etc.). This load may translate into a prohibitive cost when day-to-day transactions involving people who are not negotiation experts are concerned. Software agents may circumvent this problem by automating negotiation process while working with customers towards an acceptable deal. Moreover, they can also ensure consistency in reaching negotiation outcomes according to the set policies.

Software agents can be configured to behave in a competitive or collaborative fashion, depending on the context and the needs of a business. For example, if demand for company products or services is high, the agents could follow competitive strategies. On the other hand if customer loyalty and retention are the priority, the agents may be configured as collaborators.

However, up to date little experimental work has been done in assessing the potential of human customer vs. software agent negotiations in terms of objective and subjective variables.

The purpose of this work is to investigate the prospects of human – software agent negotiations in experimental settings. To this end an electronic negotiation system incorporating software agents has been built. The system was used in experiments with human subjects to measure such outcomes as utility of agreements and number of agreements. Additionally, such subjective variables as satisfaction and perceived usefulness were also measured.

2 Related Work

Research on automated negotiations involving software agents has been extensive [2,3]. While thorough coverage of the past work in the area is well beyond the scope of this paper, we will review the representative publications in the context of business exchanges. One could categorize these in accordance with the context of interactions (i.e. C2C, B2B, B2C), and the extent of automation.

One well-known early work in this direction was the construction of the Kasbah electronic marketplace [4,5]. Targeting primarily the C2C domain the marketplace allowed human users to configure agents, which would then be sent to the marketplace to negotiate with each other. Three types of agents ranging from competitive to the conceding ones were provided. Negotiations included a single issue, i.e. price.

In B2B applications software agents have been proposed for automating various aspects of supply chain management. For example, in [6] an agent-based architecture has been proposed for dynamic supply chain formation. The agents acting as brokers representing various entities within supply chain negotiated agreements with each other in building up the chain.

There has also been work targeting the B2C transactions. In [7] the authors proposed an agent-based architecture for automated negotiations between businesses and consumers. The buyer agents incorporated such components as searcher and negotiator, while seller agents featured negotiator module whose strategy was set by the sales department.

In [8] the authors have proposed an intelligent sales agent with the capabilities for negotiation and persuasion. The agent employed reinforcement learning in the process. In their experiments with human subjects they found that the agent using persuasion capability has increased buyer's product valuation and willingness to pay.

It has been argued by many that complete automation of real-life negotiations, in particular in business contexts does not seem to be a viable solution (e.g. [9]). Automation in general is applicable only when tasks concerned are well-structured, which is rarely the case in many business situations. However, since efficient policies can be set for multiple daily interactions with the customers regarding the sales of products and services, it seems that a relatively high level of automation may be feasible.

While the work reviewed above concerns fully automated negotiations, there has been some research into sharing responsibilities between human negotiators and negotiation agents. In [10] a system has been proposed where agents actively supported human decision making in the negotiation process. An agent advised the human user on the acceptability of the received offer, helped with the preparation of the counter-offer, and critiqued offers composed by the users when it deemed necessary to intervene.

In [11] an agent-based architecture was proposed for managing multiple negotiations. In this architecture a fleet of agents negotiated deals with customers. These negotiations were monitored by a coordinating agent, which, based on analysis of the situation, instructed the negotiating agents to adjust their strategies and reservation levels within the limits of its authority. The overall process was monitored by a human user who could intervene to make changes if necessary.

There has been some experimental work comparing in assessing human-to-agent negotiations. In [8] the authors have described an agent representing a salesperson that employed persuasion and negotiation techniques while interacting with a customer. Persuasion was based on the customer – agent dialogue with the involvement of pre-defined arguments organized into a tree. First, the agent would try to convince a customer to accept an offer. If this did not work, the agent would go into bargaining mode and determine what concession should be made. Price was the single issue in the negotiations. Using the case of a used car sale, the authors conducted both lab and online experiments. Their findings suggested that persuasion increased buyers' product valuation and willingness to pay. Negotiation increased the seller's surplus.

Another related experimental work looked to investigate the effects of framing on the subjective variables when employed by agents using persuasion/ argumentation tactics [12]. Namely, the impacts of gain vs. loss frames adopted in arguments by an agent were studied. In this study subjects were assigned the role of a buyer who had to negotiate purchase of laptops. The issues included unit price, quantity, service level, and delivery terms. Seller was a software agent, and subjects were unaware of it. The authors did not find significant differences in buyer satisfaction with the settlement or with the counter-part when compared across different frames.

The current work is aimed at investigating how software agents perform in agent-to-human dyads as compared to human-human dyads while in multi-issue negotiations. Various types of agents following different strategies have been configured for the comparison of their performance. Subjective measures have also been employed to measure the perceptions on the human side.

3 Negotiation Case and Configurations of Agents

The negotiation case developed for the experimental study concerned the sale of a desktop computer. There were five issues including the price, type of monitor, hard drive, service plan, and software loaded. Each option for each issue had a corresponding level of utility (attractiveness), these levels being different for the buyers vs. sellers. In order to calculate the total utility of the offer the issues were assigned different weights. These were then used in an additive utility function to estimate the level of attractiveness of an offer. Agents used this information in order to decide on the acceptability of the received offers and generate offers.

All agents acted on the seller side, and they were not aware of the buyers' preference structures. The weights were slightly different for sellers than buyers to facilitate tradeoffs, which have been considered one of the key integrative negotiation characteristics [13]. Thus, agents would decide on the utility of the next offer first, according to their concession schedules, and then generate the corresponding offer.

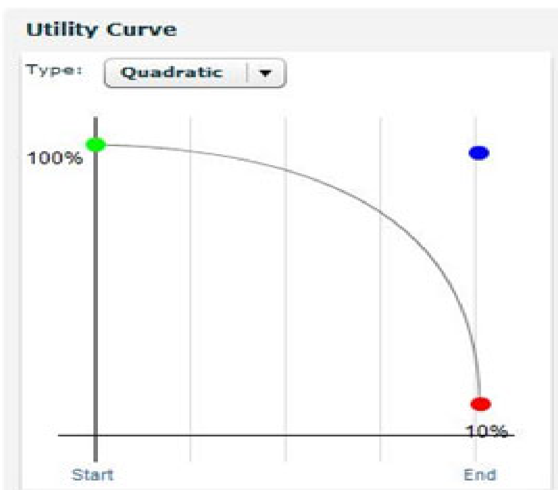


Fig. 1. Competitive Schedule

We have chosen to use five different concession schedules, three of which were similar to those used in Kasbah experiments. These included: competitive, neutral, collaborative, competitive-then-collaborative, and tit-for-tat strategies. The competitive agents (CM) tend to make smaller concessions in terms of utility of generated offers in the beginning of the negotiation period. However, as they approach the end of the period, they would start making larger concessions in search of an agreement (figure 1).

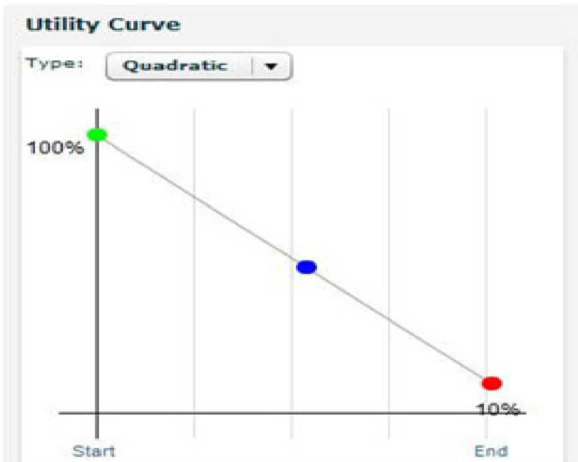


Fig. 2. Neutral Schedule

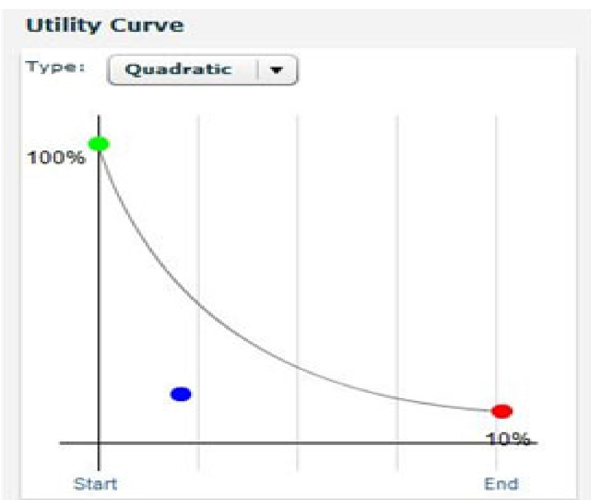


Fig. 3. Collaborative schedule

Neutral strategy (NT) dictates that an agent concedes the constant amount of utility regardless of the time period, i.e. the concession schedule is linear (figure 2). Collaborative schedule (CL) implies making large concessions in the very beginning of the negotiation period in search of a quick agreement. This represents the case where an agent is anxious to sell the product. However, as the agent quickly drops the utility close to the reservation levels, it cannot make large concessions later in the process (figure 3).

Competitive-then-collaborative schedule (CC) models more complex behavior of the agents. In the beginning of the process an agent behaves competitively, however, in the middle of the negotiation period it changes its profile to a collaborative one. Thus, there is an inflexion point in an agent's schedule (figure 4).

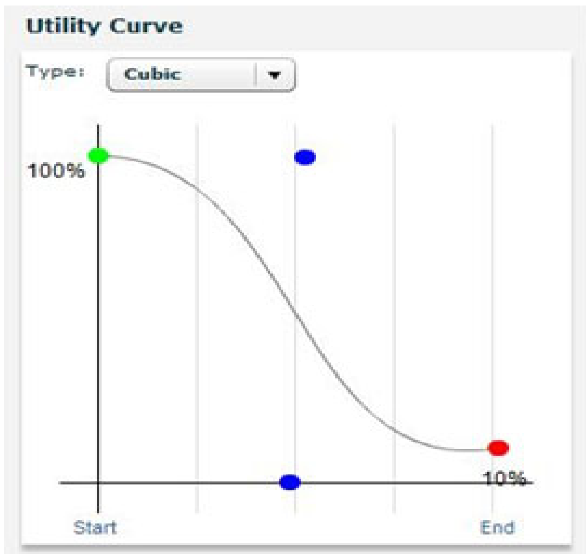


Fig. 4. Competitive-then-collaborative schedule

The reason for introducing this strategy is to imitate the situation when an agent's behavior adjusts due to the overall situation in the market (e.g. the product is not selling well). Moreover, the CC schedule allows introducing less predictable non-obvious behavior, which may be characteristic of human negotiators. (Little circles appearing on the screenshots are used to graphically define the shapes of the curves.)

The final strategy used is tit-for-tat. These agents do not rely on utility calculations. Rather, they watch the opponent moves and simply mirror them in composing counter-offers. In other words, when an opponent makes a new offer an agent determines the difference between this offer and the previous one made by the opponent, and applies the same difference to its own offer. If, say an opponent made a large change to a price, the agent would do the same.

The agent follows the following algorithm. In the beginning of the process it makes an offer that has highest utility to an agent. It then waits for the opponent to respond. If an opponent agrees, the process terminates. If an opponent makes a counter-offer the agent calculates its acceptable utility level according to the concession schedule employed. If the opponent's offer is equal or higher than the acceptable utility, the agent accepts the offer. Otherwise, the agent generates a new offer according to the acceptable utility level. It takes the opponent's offer as a starting point, and employing hill-climbing algorithm changes it to get close to the set utility level. This heuristic method is used instead of analytical one, since most of the issues are not continuous variables. It then sends this offer to the opponent.

4 Variables and Experimental Setup

In the current work we were interested in the objective outcomes of agent – human negotiations, as well as subjective variables capturing human perceptions of the process, outcomes and system. The objective variables included the utility of the agreements, and the proportion of agreements achieved. These relate to the economic benefits of agent-human negotiations. The subjective variables included satisfaction with the outcomes, satisfaction with the process, ease of use, and perceived usefulness of the system. These are important indicators from the information systems literature, especially relating to the acceptance and use of the system by human users.

The subjects in the study were university students enrolled in the introductory course on information technology. Thus, the negotiation case was well in line with the learning objectives of the course. The treatments included pairing up the subjects with various types of agents described in an earlier section. We also paired up humans with humans in a control group.

The experiment was conducted via the web, whereby subjects could perform their tasks from any location in an asynchronous mode during a two-day period. The subjects were invited to join the negotiations via email containing the link to the system. Negotiations began by sellers making the first offer. The agent sellers then checked for the status of negotiations at fixed intervals of time (every 3 hours). At those points of time, if they have not received new offers, they would wait until the next period of time elapsed. If an offer was received they would evaluate it and would either accept it, or would make a counter-offer.

Human subjects were free to terminate the negotiation at any time without reaching an agreement with their counter-parts. After either reaching an agreement, or terminating the negotiations the human subjects were asked to complete a questionnaire measuring their perceptions of the outcome, process, and the system. One final question read: "I was negotiating with: 1) a human; 2) a computer: 3) not sure."

5 Results

For the analysis of the results we have selected only those negotiation instances, which featured at least four offers in total. The rationale for this decision was to include only those cases where the subjects took the task seriously. Thus, we ended up having 436 usable negotiation instances. Of these, 65% ended up in an agreement, while in 35% of cases the agreement was not reached.

Figure 5 shows the results of the question related to whether the participants guessed correctly if they were negotiating with humans or computers. The left side shows the results from human-agent dyads, and the right side shows human-human ones. The leftmost bar in each group indicates the number of responses that read "human", the middle one relates to "computer" responses, and the last one shows "not sure" responses.

As one can see, the majority of subjects in the agent-human dyads were not sure if they were interacting with the humans or agents (183 responses). This was followed by the group of subjects who had thought they were negotiating with other humans (114). The smallest group consisted of those who guessed correctly that they were interacting with agents (65). It is interesting to note that some subjects in the human-to-human dyads thought they were interacting with a computer (2 out of 30).

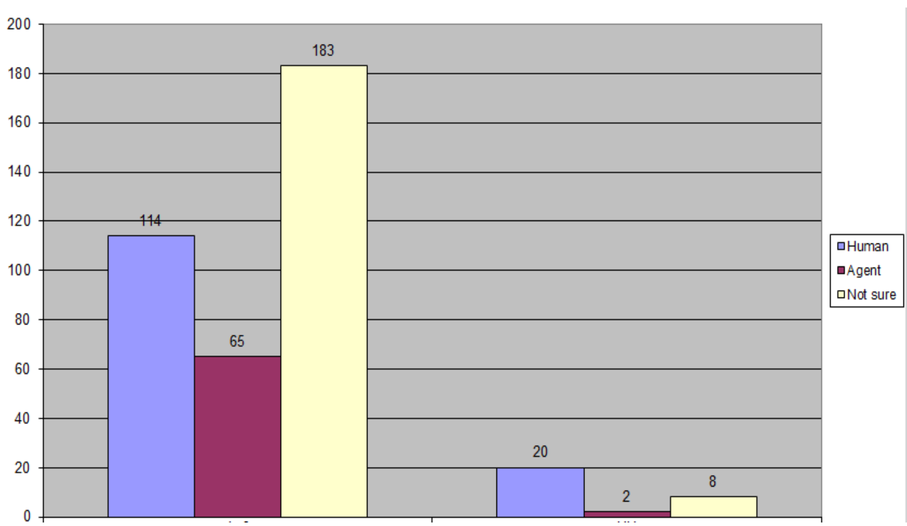


Fig. 5. "I was negotiating with…" agent - human dyads vs. human – hyman dyads

The distribution of answers depended on the type of the agent strategy employed. For example, in competitive-then-collaborative category much larger proportion of subjects thought they were negotiating with a human counter-part as compared to those who had an impression they were dealing with a machine (25 vs. 8). This can be explained by the fact that CC concession schedule results in more complex behavior, less obvious behavior that could be more readily ascribed to humans, rather than machines. Similar, though less prominent results were obtained in competitive agent category (33 vs. 15). On the other hand, the collaborative category was the only one where the number of "human" vs. "machine" responses was equal (21 each). Perhaps, the subjects expected their human counterparts to be more competitive, rather than conceding.

Table 1 shows the proportions of agreements for different compositions of dyads. The largest proportion of agreements was reached in the collaborative agent category. This is an intuitive result, since collaborative agents make large concessions early in the negotiations process, and thus they have a higher chance of making a deal with the human counterparts. It is interesting to see that human-to-human dyads have a second-lowest record in terms of proportion of agreements made. Thus, the majority of agent-involved dyads have reached more agreements than purely human dyads.

Competitive agents were able to reach an agreement in 53% of cases. Competitive-then-collaborative agents have made agreements in 75% of cases, falling between the CL and CM categories, but higher than neutral category. The lowest number of agreements was achieved in tit-for-tat category. This is the only agent strategy that does not employ utility function, and, thus it does not necessarily drop its utility level to the minimum towards the end of the period. Overall, agent-human pairs achieved agreements in 66% of cases vs. 50% exhibited by HH dyads.

Table 1. Proportions of Agreements

Category	**Agreements, %**
All agent categories	66
Competitive	53
Neutral	70
Collaborative	82
Competitive-collaborative	75
Tit-for-tat	43
Human-human	50

Table 2 compares the utilities of reached agreements for sellers and buyers across different categories. In human-human dyads the sellers achieved much lower utility levels than buyers. This could be explained by the adopted reference frames. Since both sellers and buyers in this category were undergraduate student subjects, they tended to shift the price levels downwards to what they consider to be acceptable regions. Nonetheless, as it can be seen from the table, the human sellers had reached the lowest levels of utility.

The highest average utility was achieved by tit-for-tat agents (72.4). However, as already mentioned, they performed worst in terms of proportion of agreements reached. In terms of proportion of agreements the competitive agents performed slightly better than human sellers. However, utility-wise these agents have considerably outperformed their human “colleagues” (63.2 vs. 35.9). Collaborative agents did only slightly better than humans, reaching 36.5 utility. However, they had much higher proportion of agreements. Competitive-then-collaborative agents have reached the average utility level of 40.4, and the neutral ones had a slightly higher value of 43.8. Overall, agents did better than human negotiators (46.8 vs. 35.9).

Table 2. Utilities of Agreements

Category	Seller utility	Buyer utility
All agent categories	46.8	65.6
Competitive	63.2	44.9
Neutral	43.8	69.7
Collaborative	36.5	79.0
Competitive-collaborative	40.4	71.9
Tit-for-tat	72.4	36
Human-human	35.9	73.0

In order to compare the subjects' perceptions a questionnaire was used with three items per construct measuring perceived usefulness, perceived ease of use and satisfaction with the outcome, and four items measuring satisfaction with the process. Factor analysis resulted in an acceptable pattern of loadings (Table 3). We have then used item averages for factors to compare across different categories. Results are shown in table 4.

Table 3. Factor analysis results

	Factor			
	PU	PEU	SP	SO
SO1	.146	.088	.402	**.532**
SO2	.207	.078	.258	**.789**
SO3	.235	.063	.345	**.721**
SP1	-.127	-.010	**-.404**	-.218
SP2	.112	.114	**.604**	.102
SP3	-.131	.002	**-.426**	-.175
SP4	.207	.067	**.712**	.263
PU1	**.684**	.194	.192	.172
PU2	**.816**	.130	.190	.177
PU3	**.809**	.164	.268	.204
PEU1	-.022	**-.635**	-.102	-.042
PEU2	.207	**.845**	.078	.062
PEU3	.191	**.804**	-.026	.059

There were no significant differences among the categories in terms of satisfaction with the process, perceived usefulness, and ease of use. There were, however, some significant differences regarding satisfaction with the outcome, which is understandable. In particular, tit-for-tat and competitive strategies yielded lower satisfaction levels than some other strategies, such as collaborative. As human subjects had lower utility values of their agreements they also felt less satisfies with the outcomes. None of the categories yielded significantly different results as compared with human-human interactions.

Table 4. Comparison of item averages

	Factor			
	SO	SP	PU	PEU
All agents	4.35	4.06	3.48	3.34
CM	3.89	4.19	3.38	3.38
NT	4.60	4.12	3.48	3.31
CL	4.83	4.02	3.44	3.31
CC	4.73	4.04	3.79	3.35
TT	3.35	3.79	3.24	3.33
HH	4.03	3.97	3.36	3.34

6 Conclusions

The purpose of this study was to experimentally investigate the promises of agent-human negotiations in B2C context. To this end various types of agents were configured to conduct negotiations with human subjects. The question of whether humans were able to tell if they were negotiating with machine has important implications, since if they did they would be, in principle, able to predict the opponents moves. Findings indicate that, in most cases, the subjects were not able to make a correct guess. This is especially true when agents employed a complex concession pattern, i.e. compete-then-collaborate.

In regards with the objective outcomes the results show that human negotiators performed worst as compared to agents in terms of utility of agreements. They were also second worst in terms of number of agreements.

When it comes to selling products and services or retaining customers, human representatives of companies do sometimes negotiate with their customers. Some of these negotiations nowadays occur through electronic media, using such facilities as e-mail and chat. Thus, in this study we also looked at perceptive measures related to system acceptance and usage. We found no significant differences between agent-human vs. human-human dyads.

An exciting possibility for future work could be conducting experimental studies where agent and human negotiators could add issues in the course of negotiations.

References

1. Beam, C., Segev, A.: Automated Negotiations: A Survey of the State of the Art. Wirtschaftsinformatik 39(3), 263–268 (1997)
2. Chavez, A., Dreilinger, D., Guttman, R., Maes, P.: A Real-life Experiment in Creating an Agent Marketplace. In: Nwana, H.S., Azarmi, N. (eds.) Software Agents and Soft Computing: Towards Enhancing Machine Intelligence. LNCS, vol. 1198, pp. 160–179. Springer, Heidelberg (1997)
3. Chen, E., Vahidov, R., Kersten, G.E.: Agent-supported negotiations in the e-marketplace. International Journal of Electronic Business 3(1), 28–49 (2005)
4. Huang, C.-C., Liang, W.-Y., Lai, Y.-H., Lin, Y.-C.: The agent-based negotiation process for B2C e-commerce. Expert Systems with Applications 37(1), 348–359 (2010)
5. Huang, S.-L., Lin, F.-R.: The design and evaluation of an intelligent sales agent for online persuasion and negotiation. Electronic Commerce Research and Applications 6(3), 285–296 (2007)
6. Jennings, N.R., Faratin, P., Lomuscio, A.R., Parsons, S., Wooldridge, M.J., Sierra, C.: Automated Negotiation: Prospects, Methods and Challenges. Group Decision and Negotiation 10(2), 199–215 (2001)
7. Kersten, G., Noronha, S.J.: WWW-based Negotiation Support: Design, Implementation, and Use. Decision Support Systems 25, 135–154 (1999)
8. Lin, R., Kraus, S.: Can automated agents proficiently negotiate with humans? Communications of the ACM 53(1), 78–88 (2010)
9. Maes, P., Guttman, R.H., Moukas, A.G.: Agents that Buy and Sell. Communications of the ACM 42(3), 81–87 (1999)
10. Raiffa, H., Richardson, J., Metcalfe, D.: Negotiation Analysis. The Science and Art of Collaborative Decision Making. Harvard University Press, Cambridge (2003)
11. Vahidov, R.: Situated Decision Support Approach for Managing Multiple Negotiations. In: Gimpel, H., Jennings, N.R., Kersten, G.E., Ockenfels, A., Weinhardt, C. (eds.) Negotiation, Auctions, and Market Engineering. LNBIP, vol. 2, pp. 179–189. Springer, Heidelberg (2008)
12. Wang, M., Wang, H., Vogel, D., Kumar, K., Chiu, D.K.W.: Agent-based negotiation and decision making for dynamic supply chain formation. Engineering Applications of Artificial Intelligence 22(7), 1046–1055 (2009)
13. Yang, Y., See, Y.H.M., Ortony, A., Tan, J.J.X.: Subjective Effectiveness in Agent-to-Human Negotiation: A Frame x Personality Account. In: McBurney, P., Rahwan, I., Parsons, S., Maudet, N. (eds.) ArgMAS 2009. LNCS, vol. 6057, pp. 134–149. Springer, Heidelberg (2010)

An Approach for Multiple Attribute Group Decision Making Based on Information Axiom

Jie Lin[1] and Houxing You[1,2]

[1] School of Economics and Management, Tongji University, Shanghai, 200092, China
[2] Department of Electronic Commerce, Jiangxi University of Finance and Economics, Nanchang, 330013, China
youhouxing@126.com

Abstract. In this paper, a new method was presented for multiple attribute group decision making. In the proposed method, the comprehensive evaluation value of each attribute was obtained by set-valued statistics model, and the ranking of all alternatives was judged by the information axiom. Finally, a software product evaluation case was given, and the case was calculated by the proposed method and traditional method. The result illustrates the feasibility and applicability of the proposed method.

Keywords: group decision making, information axiom, information content, experiment compare.

1 Introduction

The increasing complexity of the socio-economic environment makes it less and less possible for a single decision maker to take into account all relevant aspects of a problem [1]. Therefore, most of decision-making processes need to be supported by a group of decision makers, namely multiple attribute group decision making (MAGDM), which need to determine not only the weights of attributes, like multiple attribute decision making (MADM), but also the weights of decision makers. Obviously, group decision making is able to provide a better decision mechanism, which is more democratic and scientific than MADM, but decision-making process of MAGDM is more complicated [2].

The information axiom provides an effective approach for decision making, in which the priority of alternatives is determined by the size of information content, and the information content is measured via both design range and system range of alternative's attributes. According to their own needs, decision makers can also obtain the different preference result by adjusting the design range, and this can avoid determining the weights of attributes. Meanwhile, owing to calculating information content by the form of probability ratio, the process of normalizing attributes values is omitted [3].

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 368–378, 2012.

In view of the above mentioned characteristics, the information axiom is widely used in decision-making field [4]. Babic[5] proposed a method which provided an effective decision support system for flexible manufacturing system designers to determine the appropriate FMS configuration at the design stage. The developed model illustrated the selection of alternatives among machines that produced manufacturing components with respect to the design specifications. The selection procedure was implemented by using the information axiom. Kulak [6] developed a decision support system considering both technical and economic criteria in material handling equipment selection problem, and the final decision for the best equipment selection among the alternatives was given by using the information axiom. Kulak and Kahraman [7] introduced the information axiom under fuzzy environment, and the proposed approach was applied to multiple attribute comparison of advanced manufacturing systems. Coelho [8] used axiomatic design principles as a decision making tool to determine one of the manufacturing technologies, and the information axiom was employed to select the appropriate technology at a high decision level for the subsequent detail design of a mechanical component. Kahraman and Cebi [9] developed a method based on the information axiom for multiple attribute decision making problem, and an application was given by a teaching assistant selection problem to show the usability of the method. Celik, Kahraman, et al.[10] used the fuzzy information axiom to investigate a systematic evaluation model on the docking facilities of shipyards. The information axiom was used for the selection of the best alternatives among shipyards since the information axiom gave the opportunity to decision makers to define the design interval for each criterion.

According to the above, we learn that work done in previous research mainly belong to the domain of multiple attribute decision making. The research work with regard to the information axiom in the field of multiple attribute group decision making is still needed for us. Therefore, through the set-valued statistics model and the information axiom, an integrated methodology is proposed for MAGDM in this paper, and the feasibility and applicability of the proposed method is demonstrated by using a concrete example.

Considering the complexity and uncertainty of decision environment, it is generally difficult for decision makers to express decision information by accurate numerical values. However, if decision makers use interval numbers to express the information, it is convenient and corresponding with people’s thinking patterns. Therefore, in this paper, we are to discuss the situation that the decision information of decision makers is interval numbers.

2 The Proposed Method for MAGDM

2.1 Set-Valued Statistics Model

Set-valued statistics model can synthesize viewpoint of different decision makers, which is the generalizations of classical statistics [11]. A certain point of phase space is gotten by classical statistics in every test, but set-valued statistics gets a subset of phase space. In this paper, this subset of phase space is an interval estimation, namely, $\tilde{a}_{sj}$. A set-valued statistics series is formed which is the distribution of decision information interval overlying. This kind of distribution map can be described as equation (1).

$$\overline{u_j}(a) = \frac{1}{p}\sum_{s=1}^{p}\tilde{a}_{sj}(a) \tag{1}$$

Where $\overline{u_j}(a)$ is fuzzy covering frequency, $\tilde{a}_{sj}(a)$ is shadow-fallen function, $\tilde{a}_{sj}(a)$ can be expressed as equation (2).

$$\tilde{a}_{sj}(a) = \begin{cases} 1, & if\ \underline{a_{sj}} \le a \le \overline{a_{sj}} \\ 0, & else \end{cases} \tag{2}$$

By analyzing $\overline{u_j}(a)$, we can obtain decision makers' viewpoint unified degree on u_j. If decision makers' viewpoint is a unanimous trend, the graphics distribution is relatively concentrated, and $\overline{u_j}(a)$ appears similar normal distribution. On the contrary, the distribution is relatively discrete, and $\overline{u_j}(a)$ appears flat. According to set-valued statistics, we can obtain decision makers' comprehensive evaluation value $\overline{a_j}$ on u_j, which can be expressed as equation (3).

$$\overline{a_j} = \frac{\int_{\min}^{\max} a\times\overline{u_j}(a)d_a}{\int_{\min}^{\max}\overline{u_j}(a)d_a} = \frac{\frac{1}{2p}\sum_{s=1}^{p}[(\overline{a_{sj}})^2-(\underline{a_{sj}})^2]}{\frac{1}{p}\sum_{s=1}^{p}(\overline{a_{sj}}-\underline{a_{sj}})} = \frac{1}{2}\frac{\sum_{s=1}^{p}[(\overline{a_{sj}})^2-(\underline{a_{sj}})^2]}{\sum_{s=1}^{p}(\overline{a_{sj}}-\underline{a_{sj}})} \tag{3}$$

Where min and max are minimum($\underline{a_{sj}}$) and maximum($\overline{a_{sj}}$) respectively.

Definition. Frequency coverage is formed by all decision makers' decision information interval overlying on u_j, the shadow discrete degree of which is named decision makers' viewpoint divergence degree, as g_j. According to the literature [11], we can obtain equation (4).

$$g_j \triangleq \frac{\int_{\min}^{\max} (a-\overline{a_j})^2 \times \overline{u_j}(a) d_a}{\int_{\min}^{\max} \overline{u_j}(a) d_a} = \frac{\frac{1}{3p}\sum_{s=1}^{p}[(\overline{a_{sj}}-\overline{a_j})^3-(\underline{a_{sj}}-\overline{a_j})^3]}{\frac{1}{p}\sum_{s=1}^{p}(\overline{a_{sj}}-\underline{a_{sj}})}$$
$$= \frac{1}{3}\frac{\sum_{s=1}^{p}[(\overline{a_{sj}}-\overline{a_j})^3-(\underline{a_{sj}}-\overline{a_j})^3]}{\sum_{s=1}^{p}(\overline{a_{sj}}-\underline{a_{sj}})} \tag{4}$$

The greater value of g_j indicates viewpoint divergence degree bigger, and the credibility of comprehensive evaluation value $\overline{a_j}$ is lower. Conversely, its credibility is higher. Assume that the coefficient of viewpoint divergence degree is constant γ (normally, the value of γ is 0.1), If $g_j \le \gamma$, this indicates that $\overline{a_j}$ is desirable. Otherwise, $\overline{a_j}$ need to be recalculated. By set-valued statistics model, we can obtain effective comprehensive evaluation value of alternatives' each attribute in no $\overline{w_i}$ situation.

2.2 Principles of the Information Axiom

After obtaining the comprehensive evaluation value, we are to calculate the information content of each alternative by information axiom. The information axiom states that among those designs that satisfy the independence axiom, the design that has the smallest information content is the best design [3]. Information is defined in terms of the information content, I, that is related in its simplest form with the probability of satisfying the given FRs. I determines that the design with the highest probability of success is the best design. In practice, in any design situation, the probability of success is given by what designer wishes to achieve in terms of tolerance and what the system is capable of delivering. The overlap between the designer-specified "design range" and the system capability range "system range" is the region where the acceptable solution exists. Therefore, in the case of uniform probability distribution function P may be expressed as equation (5).

$$I = -lb\frac{C_r}{S_r} = lb\frac{S_r}{C_r} \tag{5}$$

Where S_r is system range, and C_r is common range. If FR is a continuous random variable, the probability of achieving FR in the design range may be expressed as equation (6).

$$P = \int_{dr}^{du} f(FR) d_{FR} \tag{6}$$

Where d_r and d_u are the lower bound of design range and the upper bound of design range respectively. $f(FR)$ is system probability density function for FR. So the information content is equal to equation (7).

$$I = -lbP = -lb\int_{dr}^{du} f(FR) d_{FR} \tag{7}$$

After calculating information content of all attributes of each alternative, and accumulating those information contents, we may obtain total information content of each alternative. According to the information axiom, the best alternative is the smallest total information content. The priority of alternatives is determined by comparing the size of information content, and we avoid both determining the weights of alternative's attributes and normalizing attributes values. At the same time, according to their own needs, decision makers can also obtain the different preference alternative by adjusting the design range.

2.3 The Decision Making Procedure Based on Set-Valued Statistics Model and Information Axiom

Based on subchapters 2.1 and 2.2, the decision-making procedure based on set-valued statistics model and the information axiom is concluded as follows:

Step 1. Interval number $\tilde{a}_{sj}$ of decision-making attribute set $U = \{u_1, u_2, \cdots, u_n\}$ is obtained from decision-making group $D = \{d_1, d_2, \cdots, d_p\}$, as $\widetilde{A}_{pn}$.

Step 2.Calculate the columns vector of $\widetilde{A}_{pn}$ by using set-valued statistics model, and comprehensive evaluation value $\overline{a_j}$ is obtained, as $\widetilde{A}_n$.

Step 3.Calculate viewpoint divergence degree g_j of $\overline{a_j}$, If $g_j \le \gamma$, it indicates that $\overline{a_j}$ is desirable. Otherwise, reacquire $\widetilde{A}_{pn}$ and $\overline{a_j}$ need to be recalculated.

Step 4. According to $\overline{a_j}$, calculate information content of all attributes of each alternative.

Step 5. Calculate the total information content of each alternative.

Step 6. Rank all the alternatives by the information axiom and obtain the best alternative.

The decision-making procedure can be illustrated by Fig.1.

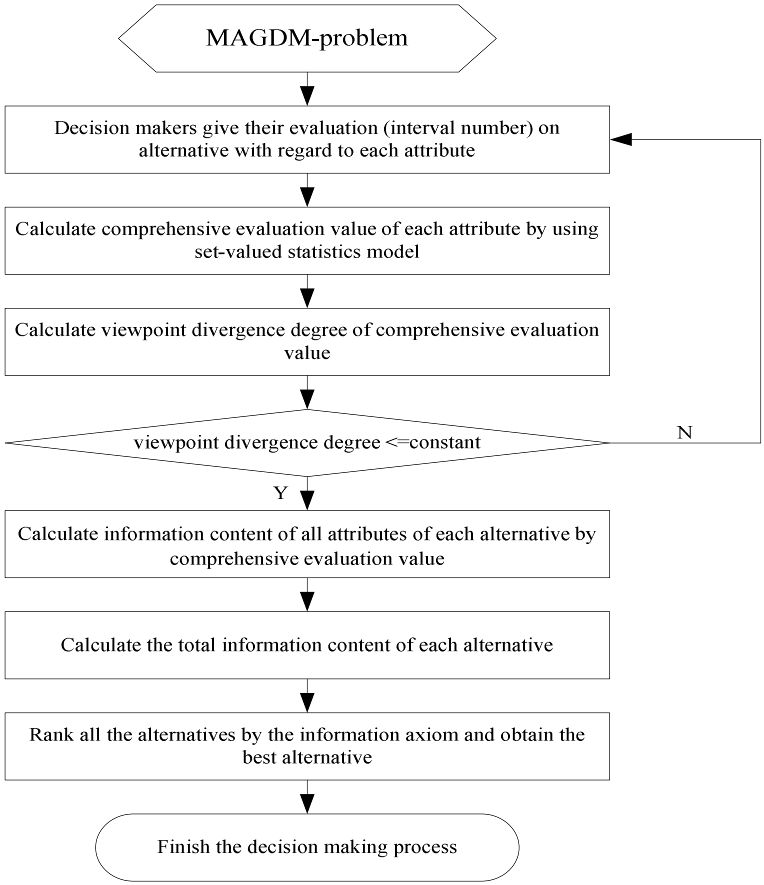


Fig. 1. The decision-making process flowchart

3 An Application Case Study

3.1 An Equipment Software Product Evaluation Case

Many software quality evaluation models have been proposed in the previous research. There were four famous models: CMM [12], Boehm [13], McCall [14] and ISO/IEC9126 [15]. The CMM model belongs to procedure-oriented quality evaluation, and the rest models belong to product-oriented quality evaluation. The ISO/IEC9126 model was widely adopted, which was expressed by a hierarchical structure, including quality factor, quality standard and quality measurement. The model was divided into six quality factors, including functionality, efficiency, reliability, maintainability, portability and usability [15]. Later, many evolution models derive from the ISO/IEC9126.

In this paper, considering the characteristics of equipment software and software testing data provided by five equipment software product developers, we introduce a software quality model on the basis of the ISO/IEC9126. The model consists of completeness (u1), accuracy (u2), efficiency (u3), reliability (u4), maintainability (u5), portability (u6) and usability (u7).

The relevant data of five equipment software products are given in Table 1, and Table 2 shows ten experts' decision-making information from ten different departments, and the information is expressed with interval numbers.

Table 1. Software testing data of five equipment software products

SP	u1	u2	u3	u4	u5	u6	u7
1	0.736	0.945	0.985	0.643	0.470	0.380	0.873
2	0.843	0.790	0.865	0.698	0.660	0.423	0.852
3	0.921	0.932	0.942	0.532	0.553	0.350	0.578
4	0.892	0.671	0.886	0.900	0.762	0.510	0.905
5	0.654	0.996	0.924	0.840	0.740	0.570	0.940

*Note: Data from *acta armamentarii*

Table 2. Experts' decision-making information

Expert	u1	u2	u3	u4	u5	u6	u7
1	[0.65,0.70]	[0.95,1.00]	[0.95,0.98]	[0.98,1.00]	[0.45,0.50]	[0.45,0.65]	[0.80,0.85]
2	[0.65,0.75]	[0.96,1.00]	[0.90,1.00]	[0.90,0.99]	[0.55,0.65]	[0.40,0.55]	[0.85,0.95]
3	[0.60,0.65]	[0.90,1.00]	[0.95,1.00]	[0.99,1.00]	[0.56,0.60]	[0.45,0.65]	[0.80,0.90]
4	[0.65,0.70]	[0.95,1.00]	[0.95,0.99]	[0.95,0.99]	[0.50,0.60]	[0.45,0.55]	[0.85,0.90]
5	[0.75,0.80]	[0.98,1.00]	[0.90,1.00]	[0.95,1.00]	[0.46,0.55]	[0.45,0.55]	[0.75,0.85]
6	[0.65,0.70]	[0.95,1.00]	[0.95,0.98]	[0.90,1.00]	[0.40,0.45]	[0.45,0.70]	[0.80,0.85]
7	[0.75,0.80]	[0.85,1.00]	[0.96,1.00]	[0.90,1.00]	[0.45,0.75]	[0.45,0.68]	[0.60,0.80]
8	[0.70,0.75]	[0.98,1.00]	[0.90,0.98]	[0.95,0.98]	[0.45,0.58]	[0.40,0.55]	[0.90,0.98]
9	[0.65,0.70]	[0.95,1.00]	[0.98,1.00]	[0.92,0.98]	[0.45,0.60]	[0.50,0.65]	[0.85,0.96]
10	[0.70,0.75]	[0.80,1.00]	[0.90,0.96]	[0.95,0.97]	[0.50,0.60]	[0.35,0.60]	[0.85,0.92]

3.2 The Evaluation Based on the Proposed Method

In this subsection, the software quality of five equipment software products is evaluated by using the proposed method. Firstly, according to Table 2, we can calculate $\overline{a_j}$ and g_j, as shown in Table 3.

Table 3. Comprehensive evaluation value and viewpoint divergence degree value

SP	u1	u2	u3	u4	u5	u6	u7
$\overline{a_j}$	0.702	0.942	0.955	0.957	0.549	0.529	0.833
g_j	0.02	0.02	0.01	0.06	0.05	0.05	0.08

Then, according to Table 3, because of $g_j \leq \gamma$, It demonstrates that $\overline{a_j}$ is effective. We can continue Step 4.

After that, according to Table 1, because attributes' value belongs to single point value and all attributes (u1~u7) are benefit type, it indicates the bigger attribute value, and the evaluation result is the better. If the attributes' value is greater than $\overline{a_j}$, it indicates that "system range" completely meet "design range", then information content of the attribute is zero by formula (6). Here, take u6 as an example, where $\overline{a_6}$ =0.529, only u6 of SP5 completely meets the design range of $\overline{a_6}$, as I_{SP5-u6} =0. $I_{SP1-u6} = lb\exp(0.529 - 0.380)$ =0.215.

Using the same method for other attributes, the calculating result shows in Table 4.

Table 4. Information content of SP

SP	u1	u2	u3	u4	u5	u6	u7	$\sum$
1	0	0	0	0.453	0.114	0.215	0	0.782
2	0	0.219	0.130	0.374	0	0.153	0	0.876
3	0	0.014	0.019	0.614	0	0.258	0.367	1.272
4	0	0.390	0.099	0.083	0	0.027	0	0.599
5	0.069	0	0.044	0.169	0	0	0	0.282

Finally, according to the information axiom, we acquire the ranking of SPS: SP5 $\succ$ SP4 $\succ$ SP1 $\succ$ SP2 $\succ$ SP3, namely, SP5 is the best equipment software product.

4 Comparing the Proposed Method with Other Methods

In this section, the proposed method in this paper is compared with two different methods, the traditional AHP and the AHP-ELECTRE [16]. Analytic hierarchy process (AHP) is also a decision-making method for determining the ranking of alternatives. The method has been extensively applied, especially in large-scale problems where many attributes must be considered and where the evaluation of alternatives is mostly subjective. The purpose of AHP provides vector of weights expressing the relative importance of the alternatives for each attribute [17]. AHP requires assessing the decision makers' evaluations by pairwise comparisons and using the eigenvector method to yield priorities for attributes. The scale of importance is defined according to

Saaty 1-9 scale for pairwise comparison, and readers can refer to Saaty's literature [17] for detailed discussion of AHP modeling and solution methodology. As will be shown below, five equipment software products are evaluated by using AHP method in Table 1.

Step 1. We consult experts' opinion, and pairwise comparisons judgment matrix is constructed for attributes (u1~u7) by using 9-scale method, shown as follow.

$$A = \begin{bmatrix} 1 & 1/8 & 1/9 & 1/4 & 3 & 3 & 1 \\ 8 & 1 & 1 & 1/5 & 4 & 8 & 2 \\ 9 & 1 & 1 & 1 & 3 & 9 & 7 \\ 4 & 5 & 1 & 1 & 6 & 9 & 2 \\ 1/3 & 1/4 & 1/3 & 1/6 & 1 & 1 & 1 \\ 1/3 & 1/8 & 1/9 & 1/9 & 1 & 1 & 1/3 \\ 1 & 1/2 & 1/7 & 1/2 & 1 & 3 & 1 \end{bmatrix}$$

Step 2. Calculate maximum eigenvalue of A, as $\lambda_{\max}$ =7.8725.

Step 3. The corresponding eigenvectors of $\lambda_{\max}$ are normalized, and weight vectors W_A is obtained.

$$W_A = (0.061, 0.199, 0.281, 0.316, 0.046, 0.026, 0.070)$$

Step 4. Calculate Consistency Index, as CI.

$$CI = \frac{\lambda_{\max} - n}{n-1} = 0.1454$$

Where n is the matrix order of judgment matrix A, n =7. RI denotes the Mean Random Consistency Index of A, RI =1.36.

Step 5. Calculate Random Consistency Ratio, as CR.

$$CR = \frac{CI}{RI} = 0.1069$$

Owing to CR >0.1, judgment matrix A fails to pass Consistency Check. Therefore, judgment matrix A needs to be adjusted.

Step 6. Calculate quasi-optimal transfer matrix of A, as A'.

$$A'=\begin{bmatrix} 1 & 0.07 & 0.028 & 0.026 & 1.874 & 5.542 & 0.667 \\ 14.334 & 1 & 0.397 & 0.367 & 26.864 & 79.433 & 9.558 \\ 36.126 & 2.52 & 1 & 0.924 & 67.703 & 200.189 & 24.087 \\ 39.079 & 2.726 & 1.082 & 1 & 73.238 & 216.555 & 26.057 \\ 0.534 & 0.037 & 0.015 & 0.014 & 1 & 2.967 & 0.356 \\ 0.181 & 0.013 & 0.005 & 0.005 & 0.338 & 1 & 0.12 \\ 1.5 & 0.105 & 0.042 & 0.038 & 2.81 & 8.31 & 1 \end{bmatrix}$$

Step 7. Calculate maximum eigenvalue of A', as $\lambda_{\max}'$=7.

Step 8. Calculate Consistency Index, as CI'.

$$CI' = \frac{\lambda_{\max}' - n}{n-1} = 0, \text{ then } CR = \frac{CI}{RI} = 0$$

So A' passes Consistency Check.

Step 9. The corresponding eigenvectors of $\lambda_{\max}'$ are normalized, and weight vectors W_A' is obtained.

$$W_A' = (0.011, 0.155, 0.39, 0.421, 0.006, 0.002, 0.016)$$

Step 10. According to W_A', we can calculate comprehensive rating value of the five equipment software products. Their values are (0.83, 0.78, 0.76, 0.85, 0.90), the ranking of SPS:

$$\text{SP5} \succ \text{SP4} \succ \text{SP1} \succ \text{SP2} \succ \text{SP3}$$

Chen [16] used an improved AHP method which is AHP-ELECTRE, obtaining the same result in the same data in Table 1. And Chen also illustrated the advantages of the AHP-ELECTRE method. Consistency of the results demonstrates feasibility and applicability of the proposed method.

5 Conclusions

In this paper, we presented a new method for multiple attribute group decision making problems. The proposed method was based on set-valued statistics model and the information axiom. In the proposed method, comprehensive evaluation value of each attribute was obtained by set-valued statistics model, and the effectiveness of comprehensive evaluation value was judged via viewpoint divergence degree. After that, the ranking of alternatives was obtained by the information axiom. An application case demonstrates the feasibility and applicability of the proposed method. Compared with other two methods, the proposed method has noticeable characteristics, which has stronger practicability. It will be popularized in other MAGDM field, such as supplier selection, process selection, performance evaluation and so on.

References

1. Kim, S.H., Choi, S.H., Kim, J.K.: An interactive procedure for multiple attribute group decision making with incomplete information: Range-based approach. European Journal of Operation Research 118, 139–152 (1999)
2. Hwang, C.L., Lin, M.J.: Group Decision Making under Multiple Criteria: Methods and Applications. Springer, New York (1987)
3. Suh, N.P.: Axiomatic Design: Advances and Applications. Oxford University Press, New York (2001)
4. Kulak, O., Cebi, S., Kahraman, C.: Applications of axiomatic design principles: A literature review. Expert Systems with Applications 37, 6705–6717 (2010)
5. Babic, B.: Axiomatic design of flexible manufacturing systems. International Journal of Production Research 5, 1159–1173 (1999)
6. Kulak, O.: A decision support system for fuzzy multi-attribute selection of material handling equipments. Expert Systems with Applications 29, 310–319 (2005)
7. Kulak, O., Kahraman, C.: Multi-attribute Comparison of Advanced Manufacturing Systems using Fuzzy vs. Crisp Axiomatic Design Approach. International Journal of Production Economics 95, 415–424 (2005)
8. Coelho, A.M.G.: Axiomatic design as support for decision-making in a design for manufacturing context: A case study. International Journal of Production Economics 109, 81–89 (2007)
9. Kahraman, C., Cebi, S.: A new multi-attribute decision making method: Hierarchical fuzzy axiomatic design. Expert Systems with Applications 36, 4848–4861 (2009)
10. Celik, M., Kahraman, C., Cebi, S., Er, I.D.: Fuzzy axiomatic design-based performance evaluation model for docking facilities in shipbuilding industry: The case of Turkish shipyards. Expert Systems with Applications 36, 599–615 (2009)
11. Sun, Y.F., Chen, S.Q., Wu, J.P.: Set-Valued Statistics based fuzzy neural expert evaluation system and its application. Fuzzy Systems and Mathematics 15, 97–101 (2001)
12. CMU/SEI, CMM (2001)
13. Boehm, B.W., Brown, J.R., Lipow, M.: Quantitative evaluation of software quality. In: Proc. of the 2nd Int'l Conf. on Software Engineering. IEEE Computer Society, Long Beach (1976)
14. McCall, J.A., Richards, P.K., Walters, G.F.: Factors in software quality, Vol. I, II, III, Final Technical Report, RADC-TR-77-369.Rome Air Development Center, Air Force Systems Command, Griffiss Air Force Base (1977)
15. Software Product Evaluation-Quality Characteristics and Guideline for Their Use. ISO/IEC Standard ISO-9126 (1991)
16. Chen, C., Guo, J.W., Zhao, C.X.: A method of equipment software quality evaluation based on AHP-ELECTRE. Acta Armamentarii 31, 1481–1486 (2010)
17. Saaty, T.L.: The Analytic Hierarchy Process. McGraw-Hill Book Company, New York (1980)

Banking Event Modeling and Simulation in Scenario-Oriented Stress Testing

Daning Hu[1], J. Leon Zhao[2], and Zhimin Hua[2]

[1] Department of Informatics, University of Zurich,
Binzmühlestrasse 14, CH-8050 Zurich, Switzerland
hdaning@ifi.uzh.ch

[2] Department of Information Systems, City University of Hong Kong,
83 Tat Chee Avenue, Kowloon Tong, Hong Kong
jlzhao@cityu.edu.hk, zmhua2@student.cityu.edu.hk

Abstract. The recent 2008 financial tsunami has made the financial regulators realize the importance of stress testing in banking systems. One of the major challenges in stress testing is to model and calibrate "exceptional but plausible" scenarios in which macroeconomic shocks may cause contagious bank failures that may lead to the breakdown of a banking system. Presently, existing stress testing methods mainly focus on modeling single or multiple risk factors through a "static snapshot" of the banking systems. However, real-world bank crisis scenarios are much more dynamic such that different event occurrence sequences may have different impacts on individual banks and banking systems. For purposes of predicting contagious bank failures in stress testing, we propose the use of event-driven process chains in modeling bank failure scenarios. We refer to this approach as Banking Event-driven Scenario-oriented Stress Testing (or simply the BESST approach). We compare the pros and cons of the BESST approach with two existing approaches in an example scenario. In addition, we conducted a financial simulation based on this example scenario to demonstrate the validity of the BESST approach.

Keywords: Stress Testing, Event Modeling, Scenario, Process Modeling.

1 Introduction

The recent 2008 global financial tsunami has been considered as the worst financial crisis since the Great Depression. It was triggered by the decline of U.S. housing prices and resulted in a liquidity shortfall in the U.S. banking system, pushing the banking system to the brink of a system-wide collapse. One of the major causes of this crisis is that the financial stakeholders, including the major banks and regulators, failed to model and calibrate the "exceptional but plausible" scenarios in bank stress testing in which the macroeconomic shocks may cause contagious bank failures and lead to the breakdown of a banking system [1].

Such scenarios contain events of large magnitude and impacts on banking systems that are often very rare such as the bankruptcy of Lehman Brothers. Such high-impact

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 379–389, 2012.

and rare events that are beyond the realm of normal expectations are referred as "Black Swan" events [2]. The rarity of such "Black Swan" events made it very difficult to model the bank stress testing scenarios since most existing methods rely on historical financial data such as the Value at Risk measure [3].

Moreover, existing stress testing methods focusing on evaluating the vulnerability of the banking system to single risk factors, or just combining the analysis of multiple risk factors into a single estimation of the probability distribution of a bank's aggregate losses. However, real-world financial crisis scenarios are much more complex in which different occurrence sequences of the same set of macroeconomic shocks (events) may have quite different impacts on both individual banks and banking systems. For instance, injecting capitals to selected banks before a set of financial shocks and interbank transaction settlements may largely prevent the breakdown of a banking system. However, injecting capital to the same banks after the shocks and settlements may have little effect. This is because that the shocks may already cause contagious bank failures through the network of interbank exposures and affected far more banks than the selected injection banks. Therefore, a process-oriented perspective is needed to model the full dynamics both within and between banks in stress testing scenarios.

Third, constructing a stress testing scenarios requires modeling different types of risk events or factors such as market risk events and credit risk events. These events are often dependent on or correlated with each other. Current stress testing practices in a bank often requires inputs from different bank departments to model a complex scenario. Oftentimes there will be contradictions among the inputs from different departments. Assuming in a stress testing scenario, a bank's credit risk manager designs an event that U.S. Federal Reserve Committee has cut the interest rate by 0.5%. In the same scenario, the exchange rate risk manager may design a following event that the U.S. dollar rose against other major currencies. However, in reality, the decreasing U.S. interest rate will actually drive the funds away from U.S. dollar and cause it devaluates. To address this problem, we need to develop a modeling approach that can automatically analyze and detect such inconsistencies in stress scenarios.

To address these three major challenges in modeling bank stress testing scenarios, we developed a process-driven modeling approach called Banking Event-driven Scenario-oriented Stress Testing (or simply the BESST approach), largely based on the event-driven process chains (EPC) modeling method. This approach provides banking stakeholders (i.e., bankers and regulators) an effective tool in modeling "exceptional but plausible" scenarios in banking systems. We developed a bank stress testing simulation system which is based on BESST, aiming to evaluate the effectiveness of various risk mitigation strategies. Moreover, to the best of our knowledge, our research is the first to study how to effectively model and simulate bank stress testing scenarios from a technological perspective.

The remainder of this paper is structured as follows. Section 2 reviews the pros and cons of the existing macro stress testing methodologies. In Section 3, we review the existing computer-based algorithms and systems that are used for predicting bank failures. In Section 4, we propose to use the EPC-based approach to model the contagious bank failure scenario and compare its effects with the existing stress testing modeling methods. Section 5 describes a stress testing system we developed to simulate various bank risk scenarios based on the proposed EPC approach. In Section 6, we conclude the paper and discuss the future research directions.

2 Macro Stress Testing Methodologies

Sorge and Virolainen [4] have reviewed current macro stress testing methodologies in finance literature. They proposed a schematic classification (Table 1) of existing stress testing approaches, mainly including two types: 1) the piecewise approach, and 2) the integrated approach. The "piecewise approach" mainly focuses on modeling banks' vulnerabilities to single risk factors by forecasting several financial indicators such as capital asset ratio and exposure to exchange rate risks under different economic environment. The "integrated approach" takes a further step to integrate the analysis of banks' vulnerabilities to multiple risk factors into a single estimate of the probability distribution of banks' losses under a stress scenario.

Table 1. Existing Macro Stress-Testing Methodologies in Finance

	Piecewise Approach	Integrated Approach
Main Modeling Approach	♦ Models a scenario as a set of macro fundamental variables ♦ Linear functional forms	♦ Combining analysis of multiple risk factors into a single distribution ♦ Macro-econometric risk models
Pros	♦ Simple and intuitive models ♦ Low computational costs	♦ Integrating market and credit risks ♦ Models nonlinear effects of macro shocks on credit risk
Cons	♦ Lack of empirical proofs for the validity of the linear relationships ♦ No feedback effects	♦ Non-additivity of VaR measures across institutions ♦ No feedback effects

The piecewise approach mainly focuses on modeling the direct relationships between macro fundamental variables (independent variables) and certain financial risk indicators (dependent variables) (e.g., capital adequacy ratio and return on equity). The estimated coefficients are used to simulate the impacts of possible adverse economic scenarios on the banks' financial risk indicators. Thus, the piecewise approach actually models an individual stress scenario as a combination of a specific set of macro fundamental variables. For instance, Kalirai and Scheicher [5] models the aggregate loan loss provisions in the Austrian banking system as a function of set of macroeconomic variables which include general economic indicators such as GDP, CPI inflation, and income, consumption and investment in the household and corporate sectors. Hoggarth et al. [6] focused on the relationship between banks' loan write-offs and the UK output gap, retail and house price inflation, and the nominal short-term interest rate. Moreover, Saurina and Delgado [7] studied the relationship between loan loss provisions and a set of macroeconomic indicators which includes unemployment rate, interest rates and indebtedness. This piecewise approach is very intuitive and its computational cost is usually low since these models are often in linear functional forms.

The integrated approach differs from the piecewise approach mainly from two perspectives: 1) it focuses on integrating the analysis of banks' market and credit risk factors rather than several single financial risk indicators; 2) it enables researchers to model the relationships between the macroeconomic factors, as opposed to just modeling the direct linear relationship between the financial risk indicators and the macroeconomic factors. In this way, the integrated approach allows the risk managers to model and analyze non-linear relationships between macroeconomic shocks and possible bank losses.

However, both piecewise and integrated approaches are limited in terms of their fundamental assumptions. First, both approaches assumed that all risk events (e.g., changes in macro fundamental variables) will not change during the course of study. But in reality the banks often response to various economical events to reduce its market and credit risk exposures. Thus these two approaches lack the ability to model such "dynamic" behavior responses and the impacts of such responses to other banks in the banking system.

Second, it was assumed that different occurrence sequences of risk events have the same effects on banks' liquidity status. As mentioned earlier in the introduction section, this assumption may underestimate the "domino" effects of interbank exposures on contagious bank failures (i.e., the network of interbank obligations may enable the default of one bank to have contagion effects on other banks, thereby causing more banks to be insolvent).

3 Bank Risk Management from a Technological Perspective

Besides the above macro stress testing approaches, there is another stream of research that aims to better manage bank risk from a technological perspective. This stream of research focuses on predicting bank failures by applying various data mining algorithms on historical bank financial data, aiming to classifying healthy and unhealthy banks. We briefly review these studies in bank failure predictions in terms of techniques they used. Discriminant analysis (DA) was one of the first techniques adopted [8] to predict firm failures across different industries. Later Sinkey [9] applied it to bank failure predictions. However, recent studies [10] have shown that the back propagation neural network (BPNN) algorithm outperformed it in predicting firm bankruptcy. NN-related techniques were first adopted by Odom and Sharda [11] to predict firm failures. Then Tam and Kiang [12] applied NN and demonstrated its effectiveness in predicting bank failures. Later Tam [13] adopted a variation of NN - the back propagation neural network (BPNN) algorithm - for predicting bank bankruptcy. Another set of more recent data mining methods termed as Support Vector Machines (SVMs) were also used in bank failure prediction research. Like the two previous techniques, SVM was initially used for predicting corporate bankruptcy [14]. Then Wang et al. [15] developed and employed a fuzzy support vector machine method for credit risk assessment for the banking industry.

In recent years, there is a set of studies that combines different types of data mining techniques for predicting bank failures. Min et al. [16] proposed to use genetic algorithms (GA) to optimize the feature subsets and parameters of SVM in order to improve its performance in bankruptcy prediction. A similar study done by Wu et al. [17] also used GA to optimize the parameters of SVMs for predicting bankruptcy. They empirically examined the performances of their combined GA-SVM model with other methods, such as DA, NN and standard SVMs, to predict financial crises in Taiwan. Their results showed that the GA-SVM model outperforms other methods, implying the integration of different BI techniques may be an effective approach for improving the prediction performances of single techniques. In summary, these studies adopted various computer-based techniques on real-world business data to analyze the relationships between the financial data items and bank failures. Based on the patterns learned from the data, these techniques aim to measure the risk of future bank failures.

However, these computer-based techniques still only focuses on the "static" information of bank status, but largely ignoring the dynamic processes in bank crisis scenarios. Therefore, information systems that can model, process and analyze real-time large-scale process information in bank crisis scenarios are greatly needed for stress testing purposes. To this end, we also developed a bank stress testing platform which integrates the BESST modeling approach and simulation techniques, aiming to simulate various bank risk scenarios and evaluate possible risk mitigation strategies such as capital injections.

4 A Process-Oriented Approach for Stress Testing Scenario Modeling

To address the two above limitations of the existing stress testing methodologies, we developed process-oriented approach can effectively model both the event occurrence sequences and the interbank exposures in stress testing scenarios. More specifically, we adopted event-driven process chain (EPC) as the main modeling tool to model stress testing scenarios. EPC is a type of flowchart and widely used for modeling business processes [18]. It mainly consists of three types of elements: events, functions, and logical connectors. A function is a process step which needs to be executed and are linked by events. An event describes the situation before or after a function is executed. The logical connectors such as XOR node can be used to connect events and functions, thereby specifying the control flow. Existing EPC modeling paradigms is good at modeling activity sequencing which made it suitable for modeling the occurrence sequences of risk events in stress testing scenarios.

Thus we adopt the event-driven process chains proposed in to formally represent of risk event processes in stress testing scenarios. In this research, the event node is used to represent various economic risk events (e.g., U.S. Fed cut interest rate). The function node is used to represent various information processing activities (e.g., calculating a bank's losses). In addition, there are mainly three kinds of logical connector nodes, AND, OR and XOR node.

4.1 Modeling a Bank Stress Testing Scenario

In this section, we model a typical bank stress testing scenario from a single bank i's perspective using all three aforementioned approaches including BESST, aiming to demonstrate the advantages of the BESST approach in modeling event sequences and "contagion" effects.

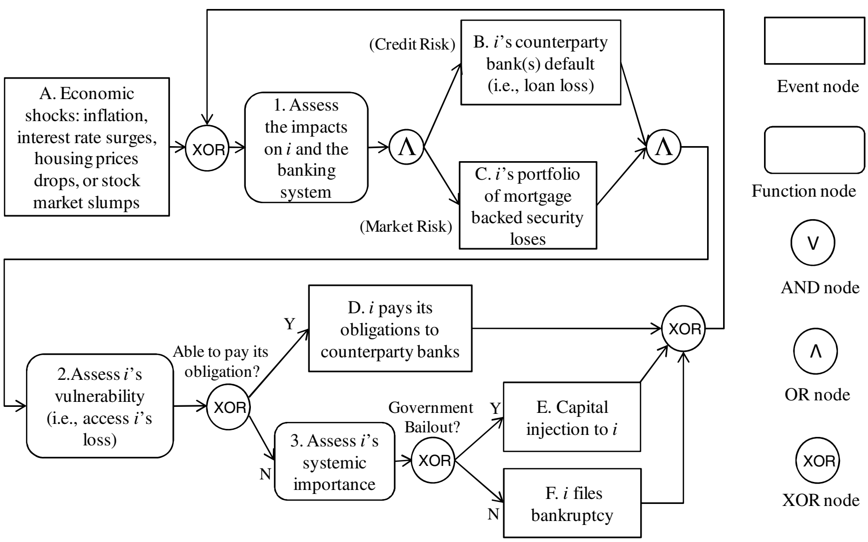


Fig. 1. A typical stress testing scenario for banking systems

We first describe the events of this stress testing scenario as Figure 1 shows. (A) There is one or more adverse economic shocks happened such as the burst of housing bubbles. These shock(s) caused drastic changes in major fundamental economic indicators such as increasing national interest rate, foreclosures, and declining stock market indices.

(B) From an individual bank i's perspective, the negative impacts of these shocks may cause one or more i's counterparty banks (i.e., banks that have payment obligations to i) to suffer great losses and thereby default their obligations to i.

(C) Meanwhile, the declining housing prices also cause great losses in i's investment portfolio on mortgage backed securities. Together with credit risks, this type of market-related risk will reduce i's ability to pay its obligations to others (Event D).

If the loss is greater than i's capital reserve, i will be not able to honor its obligations to other banks (Event F). Then i's survival depends on if the central bank will inject capital to it. If the central bank thinks i's default will significantly contribute to contagious bank failures in the banking system, it may inject capital to i to prevent a system-wide meltdown (Event E).

4.2 Modeling a Bank Stress Testing Scenario

To model the above scenario, the piecewise approach will analyze the market (Event A->B) and credit risk (Event A->C) separately. As reviewed in [4], This approach can be represented as

$$E(\tilde{Y}_{i,t+1} \,/\, \tilde{X}_{t+1} \geq \bar{X}) = f\{X^t\} \quad . \tag{1}$$

where for bank *i*'s expected loan loss (investment portfolio loss) Y at time t+1 is estimated as a linear function of past realizations of a vector X of macroeconomic variables (e.g., interest rates).

Integrated Approach

On the other hand, the integrated approach will focus on analyzing both types of risks (Event A->B,C) by incorporating the selected macro fundamental variables into value at risk (VaR) measures as follows:

$$VaR_{i,t}(\tilde{Y}_{i,t+1} \,/\, \tilde{X}_{t+1} \geq \bar{X}) = f\{E_{i,t}(X_t)\,;P_t(X_t)\,;PD_t(X_t)LGD_t(X_t)\,;\textstyle\sum_t(X_t)\}$$

The VaR measure for the portfolio of the banking system is represented by a vector E of both credit exposures and market positions at time t, and is calculated as a vector of security prices P, bank default probabilities PD, loss given default LGD and a matrix of default volatilities and correlations $\sum$. All these parameters are functions of the vector of macroeconomic variables X.

Event-Driven Process Chains

Both piecewise and integrated approaches only focuses on modeling "static" and isolated events but failed to capture the dynamic processes in the scenario described in Section 3.1. For instance, after the initial economic shocks, if bank i fails to pay its debt obligation(s) to its counterparty banks, such failures may cause contagious bank failures through domino effects. But both modeling approaches fail to capture such transaction level risk events and the event occurrence sequences, which largely determine if contagious bank failures will happen.

The proposed EPC approach provides a process perspective on modeling such dynamic event processes as shown in Figure 2. Unlike the other two approaches, the EPC models three different event processes depending on i's losses in the stress scenario. The event chain (A-> (B,C)->E) indicates that bank i's failure may cause contagious bank failures. Following Eisenberg and Noe [19], i's payment ability can be calculated as a payment clearing vector:

$$p_{i,t}^{*} = \begin{cases} d_{i,t} & if \quad \sum_{j=1}^{N} l_{ji,t} + e_{i,t} - Y_{i,t} \geq d_{i,t} \\ \sum_{j=1}^{N} l_{ji,t} p + e_{i,t} - Y_{i,t} & if \quad d_{i,t} > \sum_{j=1}^{N} l_{ji,t} + e_{i,t} - Y_{i,t} \geq 0 \text{z} \\ 0 & if \quad \sum_{j=1}^{N} l_{ji,t} + e_{i,t} - Y_{i,t} < 0 \end{cases} . \quad (2)$$

where $\sum_{j=1}^{N} l_{ji,t}$ represents all the payment obligations i receives from its counterparty banks at time t. d_{it} is the total amount of i's obligations to others, while e_{it} is i's capital reserve. Y_{it} is i's estimated loss due to the economic shocks.

5 A Bank Stress Testing Simulation System Based on BESST Approach

We also developed a bank stress testing simulation system based on the Banking Event-driven Scenario-oriented Stress Testing (BESST) approach. As Figure 2 shows, this system mainly consists three components. First, the event-drive process modeling module provide users tools for modeling complex bank risk scenarios like the one in Figure 1. This module also automatically checks the consistencies among the modeled risk events and processes. Second, the risk scenario generation module will generate

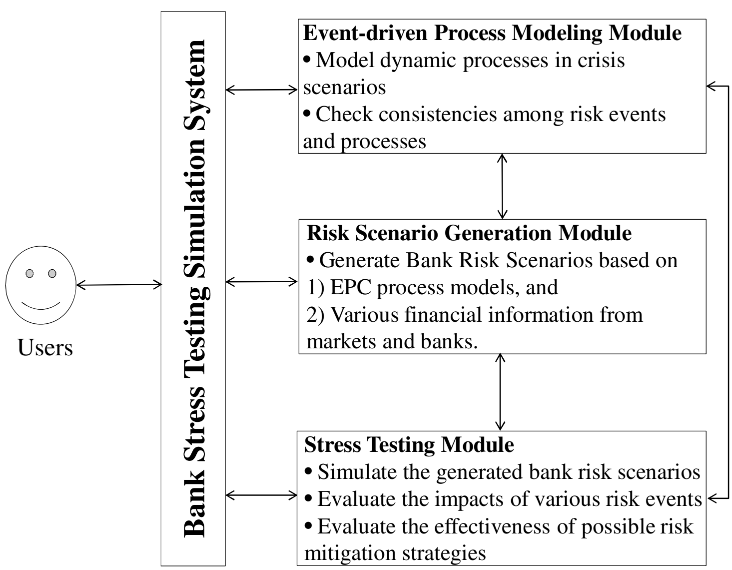


Fig. 2. System Architecture of the Bank Stress Testing Simulation System

various risk scenarios modeled by the first module. This module can implement real-world historical financial data such as security prices and forex exchange rates into the modeled scenarios. The simulation results based on such historical data may provide insights for experts about the cause and dynamics why bank crisis happen how it spread through various interbank relationships. Third, the stress testing module simulates various risk scenarios modeled and implemented by the first two modules. By configuring the value of a financial risk indicator such as interest rates, this module can simulate the impacts of the change in that indicator on banks' solvency. Therefore, we can then evaluate the impacts of various risk events on the stability of the banking systems. In this way, we also can examine the effectiveness of various bank risk mitigation strategies such as capital injections to banks.

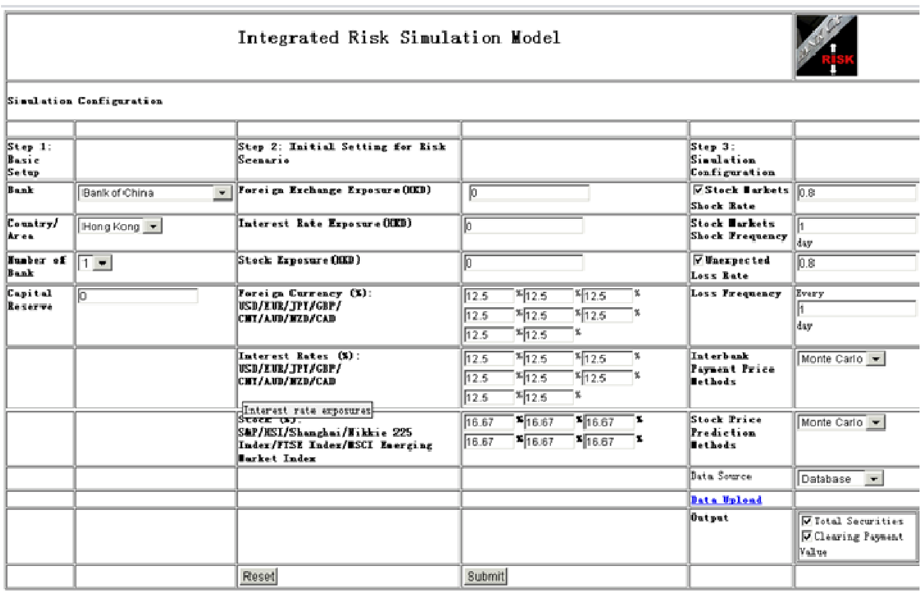


Fig. 3. User Interface for the Bank Stress Testing Simulation System

Figure 3 shows interface for configuring the stress testing scenarios. The user (e.g., risk manager) in a bank can set up their exposures in forex risk, credit risk and market risk. The user can also simulate various types of stock market shocks. Currently, we have only developed the simulation system based on the proposed BESST approach. In our future research, we will also conduct simulation experiment with the same setting using piecewise and integrated approaches to compare the effectiveness of these three approaches. We suggest BESST approach will outperform the other two in term of effectiveness.

6 Conclusions

We claim that our Banking Event-driven Scenario-oriented Stress Testing (BESST) approach offers three advantages in modeling stress testing scenarios. First, it introduces a process perspective that allows risk managers to model risk event processes, rather than just modeling a "static" snapshot of the banking system in the two conventional approaches (so called piecewise and integrated approaches). The BESST approach also enables risk managers to model sequential interactions among risk events such as contagious bank failures. Moreover, event-driven process chains can be mapped to Petri nets that have formal semantics and provide a wide range of analytical techniques [18]. This feature allows risk managers to check the correctness and consistencies of given stress testing scenarios.

In summary, this paper presents a novel event modeling approach for scenario-oriented stress testing in banks that have both practical importance and academic value in the area of bank risk management, demonstrated through a simple yet realistic case. We also developed a prototype of bank stress testing simulation system based on the proposed BESST approach. Our next step is to further validate the modeling framework and simulation system, and apply it in a banking environment. Specifically, we will use the BESST approach with the extracted real-world bank data to model more complex stress scenarios which include events such as capital injections. The simulation technique beneath the BESST approach can be used to evaluate the effectiveness of bank risk mitigation strategies. In our future research, we will also conduct simulation experiments to compare the effectiveness of the piecewise, integrated and BESST approaches.

Acknowledgement. This work was partially supported by a GRF grant from Hong Kong Research Grant Council (No. 9041582).

References

1. Hu, D., Zhao, J.L., Hua, Z.: ICIS 2010 Proceedings (2010)
2. Taleb, N.N.: The Black Swan: The Impact of the Highly Improbable (Random House Trade Paperbacks) (2011)
3. Jorion, P.: Value at risk: the new benchmark for controlling market risk. Irwin Professional Publishing, Chicago (1997)
4. Sorge, M., Virolainen, K.: Journal of Financial Stability 2(2), 113 (2006)
5. Kalirai, H., Scheicher, M.: Financial Stability Report 3, 58 (2002)
6. Hoggarth, G., Logan, A., Zicchino, L.: Bank of England, paper presented at the Workshop on Financial Stability in Frankfurt (2004)
7. Delgado, J., Saurina, J.: Directorate General Banking Regulation, Bank of Spain (2004)
8. Altman, E.I.: The Journal of Finance 23(4), 589 (1968)
9. Sinkey, J.F.: The Journal of Finance 30(1), 21 (1975)
10. Lee, K., Booth, D., Alam, P.: Expert Systems with Applications 29(1), 1 (2005); Tsukuda, J., Baba, S.-I.: Computers & Industrial Engineering 27 (1-4), 445 (1994)
11. Odom, M.D., Sharda, R.: Presented at the IJCNN International Joint Conference on Neural Networks (1990) (unpublished)

12. Tam, K.Y., Kiang, M.: Applied Artificial Intelligence 4(4), 265 (1990)
13. Tam, K.Y.: Omega 19(5), 429 (1991)
14. Shin, K.-S., Lee, T.S., Kim, H.-J.: Expert Systems with Applications 28(1), 127 (2005)
15. Wang, Y., Wang, S., Lai, K.K.: IEEE Transactions on Fuzzy Systems 13(6), 820 (2005)
16. Min, S.-H., Lee, J., Han, I.: Expert Systems with Applications 31(3), 652 (2006)
17. Wu, C.-H., Tzeng, G.-H., Goo, Y.-J., Fang, W.-C.: Expert Systems with Applications 32(2), 397 (2007)
18. van der Aalst, W.M.P.: Information and Software Technology 41(10), 639 (1999)
19. Eisenberg, L., Noe, T.H.: Management Science, 236 (2001)

Hierarchical RFID Tag Ownership and Transfer in Supply Chains

Wei Zhou[1,3], Eun Jung Yoon[2], and Selwyn Piramuthu[2,3]

[1] Information & Operations Management, ESCP Europe, Paris, France
[2] Information Systems and Operations Management, University of Florida
Gainesville, Florida 32611-7169, USA
[3] RFID European Lab, Paris, France
wzhou@escpeurope.eu, {ejyoons,selwyn}@ufl.edu

Abstract. A majority of RFID tag applications are related to tracking and tracing of the RFID-tagged items throughout supply chains. Tracking and tracing require communication with the tag to identify and authenticate the tag. While identification can be readily accomplished in the absence of adversaries, identification and authentication in the presence of adversaries is not a trivial task. This process is exacerbated when the tagged item transfers ownership as it passes through a supply chain. Recent developments in cryptography facilitates accomplishing these in a seamless manner. We consider a specific scenario in supply chains where ownership of RFID-tagged items follow a hierarchical relationship. We present an ownership transfer protocol for this scenario and briefly consider its security properties. The proposed protocol can be used to seamlessly manage hierarchical ownership transfers in supply chains.

Keywords: RFID, supply chain, ownership transfer.

1 Introduction

Radio Frequency IDentification (RFID) tags have been successfully used in a wide variety of applications over the past several decades. Although RFID tags supplant bar codes in some applications, their memory and processing power allow for applications (e.g., sensor) where bar codes cannot be effectively used. Clearly, automated processing of items in a high-speed batch processed supply chain environment calls for auto-ID technologies such as RFID. Moreover, in distributed environments where continuous access to a database cannot be guaranteed, it helps to store and retrieve important information on the item itself as it passes through a supply chain. RFID tags are ideal candidates in such environments. Not surprisingly, the demand for RFID tag use is on the rise.

Recent years have witnessed a continual decrease in RFID tag unit cost with advances in technology and volume of use. Moreover, the need to optimize processes and to gain competitive advantage, firms are interested in the use of instantaneous fine-granular item-level information that can be generated through

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 390–398, 2012.

RFID tags. RFID tags are very widely used in supply chains at the pallet level. However, item-level use is not far behind as evidenced by the incorporation of item-level RFID tags in Wrangler jeans sold at Wal-Mart stores in the U.S. since August 2010. The importance of RFID tags in tracking and tracing pallets throughout supply chains cannot be overstated.

Given that RFID tags are used in supply chains for tracking, tracing, and continually or even continuously recording ambient conditions, there is a need to communicate with these tags. Communication is necessary to uniquely identify and authenticate these tags and to transfer information to and/or from them. These can be readily accomplished in a secure environment. However, in general, a supply chain is not secure. Assuming the presence of adversaries in a supply chain environment is not far from the truth since direct competitors as well as others have the incentive to communicate with the RFID tags as they pass through a supply chain. This calls for the need to secure communication with RFID tags. This is not trivial since these RFID tags communicate with any reader without too many restrictions. Therefore, there is a need to restrict communication of tags only with authentic and honest readers. A common means to accomplish this is through protocols using cryptography.

RFID cryptography has been a very active area for both researchers and practitioners over the past decade. While a majority of RFID protocols that have been developed thus far address the more common single-tag/single-owner scenario (e.g., [9]), there are other scenarios with other characteristics that warrant development of specialized protocols. For example, the Yoking Proof ([4]) protocol and its variants are appropriate for scenarios that require simultaneous presence of two tags in the field of the reader. Ownership transfer is another such scenario with its own set of characteristics and therefore protocols. The literature to-date has about a dozen ownership transfer protocols. Almost all of these address the one-owner/one-tag scenario, which is more common than the others. In addition to such one-owner/one-tag scenario, supply chains also possess multiple-owner/multi-tag (e.g., [5], [6]) scenarios.

In a supply chain, there are situations that dictate the simultaneous presence of multiple owners for any given tag. For example, in a vendor-managed inventory (VMI) scenario, the vendor owns and manages the inventory of items at (possibly) a third-party retail store. Ownership of an item that is handled in a VMI scenario could have different configurations and privileges - i.e., the vendor and the retail store have different levels of 'ownership' of the item of interest.

We consider a scenario where the simultaneous multi-level ownerships have decreasing levels of authority and privilege at lower levels in the hierarchy. An example of this scenario is where an RFID-tagged item has one primary owner. This owner rents the item to a rental-owner. This rental-owner again rents the item to another rental-owner. The complexity of this scenario arises from the fact that an end to the ownership of the tag at any level above the lowest level would automatically nullify the ownership at every level below. An example of this structure in health care applications is presented in Meiller et al. ([8]). The scenario involves ancillaries that are used for implantation and extraction of prosthetic devices in humans. These ancillaries are shared (rented) among hospitals - in a sense, these are rented from one hospital to another beginning with their primary owner. The different hospitals renting such an ancillary is a node in this hierarchy of ownership.

We develop an ownership transfer protocol for this scenario. The proposed ownership transfer protocol addresses this scenario comprising simultaneous multi-level RFID tag ownerships with varying levels of access to the functionalities of the RFID tag. We first present the scenario and then present the proposed protocol. We provide a brief security analysis to check for some common vulnerabilities that are present in related protocols.

This paper is organized as follows: The considered scenario, the sequence of steps for ownership transfer in this scenario, as well as the proposed protocol following these steps are presented in the next section. Section 3 concludes the paper with a brief discussion.

2 Hierarchical Ownership Levels

We first present the considered scenario, followed by the sequence of events in the modeled scenario, and then the set of notations used in the proposed protocol. We then present and discuss the proposed protocol.

2.1 The Scenario

This scenario comprises a hierarchy of ownerships of the same RFID-tagged item (Figure 1). For example, consider an item X with one primary owner. This primary owner rents the item to a renter. The renter in turn could (sub-)rent the item to another renter. Eventually, the entities that rent a given item at any given point in time could form a hierarchy that is tree-structured with a single (i.e., one sub-owner) branch or several (i.e., several sub-owners) branches originating at each node. The peculiarity of this scenario is that when an intermediate node (i.e., renter) loses ownership (we define ownership of a renter loosely here as having access to that item) of an item, every node below this node in the "ownership tree" lose their ownership as well.

The following are some of the characteristics of this system:

1. The primary (main) key for the owner never changes unless there is a change in the primary owner of the tag. We do not consider this scenario in this paper since any extant ownership protocol can be used to accomplish this.
2. Whenever ownership transfer occurs (ownership transfer in this scenario is said to occur when the item has a new renter or when a renter ceases to have access to the item), a sub-key is updated.
3. We model the keys such that an RFID-tagged item's main key (say, K_i) is a composite of the main key for the owner or renter one level up in the hierarchy (say, K_{i_1} and a sub-key (k_i).
4. The main key at any level is known only to the renter at that level (or owner, if there's no renter) and the TTP

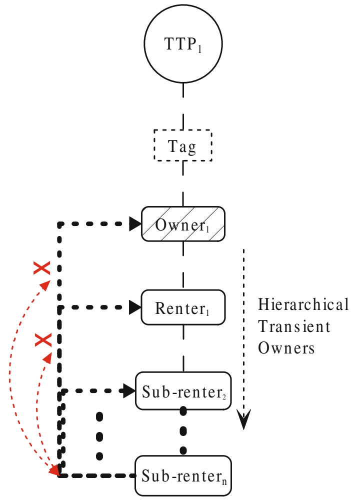


Fig. 1. Hierarchical ownership levels

5. The sub-key is known only to the TTP.
6. The main key of an item at any level (below the root-node in the hierarchy) is, therefore, a composite of the main key of the (primary) owner and the sub-keys of renters until (and including) the level of interest when no cancelation of rental agreement has occurred for this item.

2.2 The Event Sequence

At time zero, we assume that an item is owned by only one entity (i.e., the primary owner). Multiple ownership is handled similarly whereby the same main key (K) is shared among all the owners of this RFID-tagged item. When this RFID-tagged item is rented, the TTP generates a sub-key (k_1) and XORs the main key of the owner with this sub-key to generate the main key (K_1) for the renter (i.e., $K_1 \leftarrow K \oplus k_1$). This main key is then communicated to the renter and the tag. The (primary) owner is now unable to communicate with the tag.

Similarly, when this renter (sub-)rents this RFID-tagged item to another renter, the TTP generates yet another sub-key (k_2) and XORs the main key of the first-level renter (i.e., K_1) to generate the new main key (K_2) for the second-level renter (i.e., $K_2 \leftarrow K_1 \oplus k_2$). The TTP communicates this new key to the second-level renter and the tag. Again, neither the primary owner nor the first-level renter would be able to communicate with the tag. This process continues as long as none of the renters at a higher level cancels its rental agreement. When that happens (say, the $(i+1)^{th}$-level renter cancels the rental agreement), the TTP changes the sub-key (k_i) and communicates the new main key ($K_i \leftarrow K_{i-1} \oplus k_i$) to the tag and the i^{th}-level renter. Now, the i^{th}-level renter is able to communicate with the tag.

2.3 Notations

The following notations are used for this protocol:

- p, t, o, r: m-bit nonce
- K, K_i: tag's shared secret main key with the primary owner and i^{th} renter respectively
- k_i: tag's secret temporary sub-key for the i^{th}-level renter
- s: shared secret key between TTP & primary owner
- s_i: shared secret key between TTP & renter (at level) i
- k_T: shared secret key between tag and TTP
- f_k: keyed (with key k) encryption function

2.4 The Protocol

The proposed protocol (Figure 2) can be used for generation and transfer of the main key whenever ownership transfer (granting or revoking) occurs. The main key is associated with the owner or renter whereas the sub-key is associated with the renter's level in the hierarchy. The messages between any two entities as presented in Figure 2 occur through over-the-air channel that is not assumed to be secure. The entities participating in the protocol are assumed to be in close proximity to one another - at the least, any pair of communicating entities are expected to be in close physical proximity to each other.

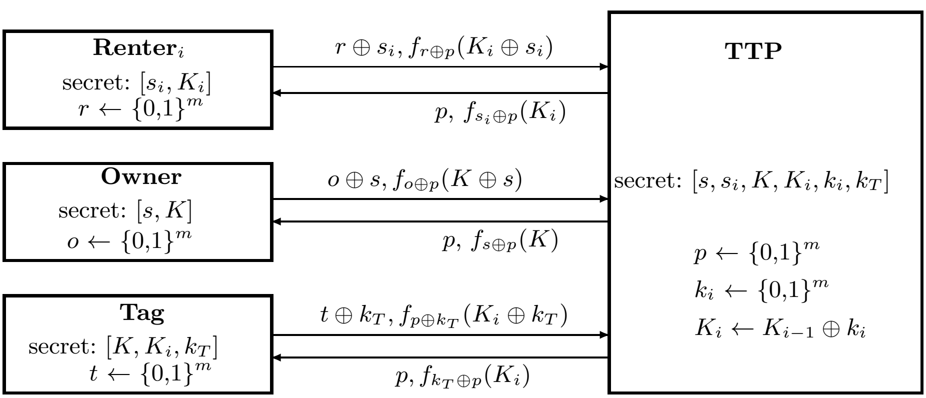


Fig. 2. The proposed hierarchical ownership transfer protocol

This protocol includes several shared secret keys between pairs of entities. The shared secret keys between the main owner and TTP is s and the shared secret key between the TTP and a level$_i$ renter is s_i. The shared secret key between the TTP and the tag is k_T. The shared key between the primary owner and the tag is K. The shared key between a level$_i$ renter and the tag is K_i. Each of the entities generate fresh nonce when initiating a protocol 'loop.' The set of nonce generated and used in the protocol include r, o, t, and p respectively for the renter, owner, tag, and TTP.

The messages in the protocol form three loops: the loop between owner and TTP is used by the TTP to inform the owner of either (1) renting of the RFID-tagged item by a level_1 renter or (2) termination of rental by a level_1 renter. The loop between the TTP and Renter_i is (1) for the TTP to generate a new main key and inform the new renter of the same when this level_i renter rents the RFID-tagged item or (2) for the TTP to generate a new main key and inform the renter at level_i when renter at level_{i+1} terminates its rental contract. The remaining loop is for the TTP to inform the tag when a new renter rents this RFID-tagged item or an existing renter terminates its rental contract on this tagged item.

The protocol begins when a renter rents the RFID-tagged item. The TTP generates a fresh nonce (p) and a fresh sub-key (k_1) and encrypts its message to the new renter using its shared key (s_i) and sends the message to the renter $(p, f_{s_i \oplus p}(K_i))$. The purpose of this loop is to inform the renter of the tag's main key. In response, the renter acknowledges by generating a fresh nonce (r) and the message $(r \oplus s_i, f_{r \oplus p}(K_i \oplus s_i)\ )$ to the TTP. To prevent trivial DoS attacks, the TTP waits for a pre-determined amount of time to receive acknowledgement from the renter. When this does not happen, TTP generates a fresh nonce (say, p') and re-starts the process. This is repeated until successful completion of the loop.

The purpose of the next loop is similar to the preceding one except here the tag is notified of the change in key. The TTP generates a fresh nonce (p) and sends $(p, f_{k_T \oplus p}(K_i))$ to the tag. The tag retrieves the new key (K_i) and acknowledges the message to the tag with $(t \oplus k_T, f_{p \oplus k_T}(K_i \oplus k_T))$. Again, the TTP waits for a pre-determined amount of time to receive a response from the tag. When the response does not materialize, the TTP generates a fresh nonce and repeats its message to the tag after incorporating this nonce.

The remaining loop is used to inform the owner of a change in key. Here, since the main key (for the owner) does not change, this loop is instantiated only when there is a change in access for the owner (i.e., when renter_1 rents the tag from the owner or when renter_1 terminates ownership of the item). The message structure between the TTP and owner are very similar to those between the TTP and tag or TTP and renter. As with the other two loops, the TTP waits for acknowledgement from the owner and repeats the loops with a fresh nonce when acknowledgement is not received within a pre-specified amount of time.

Every loop in the protocol is timed to ensure that all initiating messages reach their destination and the acknowledgement messages reach their destination.

2.5 Security Analysis

We briefly consider a few security violations that can arise in the presented context and evaluate the proposed protocol.

- *Denial of Service (DoS) / desynchronization attack*: Denial of Service (DoS) attacks can arise due to several reasons including desynchronization between tag and reader and can lead to disconnect between tag and reader. When this occurs, the tag (or, reader) will not be able to communicate with the reader (or, tag). In the proposed protocol, there is very little chance for desynchronization to occur since when the shared keys are updated, the entity updating the key (e.g., TTP) ensures that all parties involved (e.g., Owner, renter, tag) are aware of this update through timed message loops.
- *Forward Security*: Forward security is necessary to maintain the integrity of the system and is especially critical in systems where messages are exchanged over the air. Forward security implies that even if an adversary copies every message that has been exchanged between two entities, the adversary cannot decipher any of them even though the secret keys at a later point in time are known to the adversary. This protocol does not guarantee forward security since none of the messages are encrypted using one-way hash functions. It is possible to address this issue through the use of one-way hash functions to encrypt messages throughout the protocol. However, that could lead to increased computational load on the entities, especially the tag, which is bound by extreme resource constraints. The messages sent between the TTP & Owner and TTP & renter can be encrypted using one-way hash functions (instead of the encryption function used) to alleviate effects related to this problem.
- *Replay attack*: Replay attack occurs when an adversary passively observes communication among entities and copies those messages for later use. At some later point in time, the adversary strategically replays some of these messages, which are accepted as valid by the receiving entity. Since most messages that could be replayed in the proposed protocol include freshly generated nonce, replaying any of these will not work (i.e., will not be validated by the receiving entity).
- *Impersonation attack*: Impersonation attack occurs when an adversary is able to completely impersonate an entity to all other entities that participate in the protocol. In the proposed protocol, an adversary cannot impersonate any of the entities to another due to the use of freshly generated nonce in each of the three loops. It is also difficult for an adversary to obtain the shared secrets (i.e., s, s_i, k_{i-1}, k_i, K, K, K_i, k_T, k_i).

3 Discussion

We considered the scenario where ownership gets transferred in a hierarchical fashion. The requirement in this scenario is that cancel of ownership at an intermediate level dictates cascade cancel of ownership at all levels below in the tree-based hierarchy. The use of a new sub-key every time a new renter gains or loses access to the RFID tag ensures maintenance of a certain degree of privacy and security of the RFID-tagged item. This also prevents (both physical and RF) unauthorized access of the RFID-tagged item by anyone, including adversaries and those (renter as well as owner) at higher levels in the tree. This is done to ensure that only the current renter (or, owner if there are no renters of this item) and the TTP can communicate with the RFID tag at any given point in time. Using a composite comprising the key from one level up and a local key prevents unauthorized access to a tag when the current key is compromised. An adversary with such a compromised key would be unable to communicate with the tag once the sub-key at a higher level is modified by the TTP. Moreover, this lends itself to an easy extension - ownership *sharing* - when renters at two different levels in the hierarchy *share* ownership of an item, the 'higher level' renter knows its key and the 'lower level' renter knows the composite key.

While it is possible, in principle, to develop secure authentication protocols, it is difficult to ensure protection from relay attacks. Relay attacks occur when adversaries simply relay messages between a honest reader and a honest tag with or without the knowledge of either party (e.g., [1], [2], [3], [7]). The difficulty with relay attacks lies in the absence of cryptographic manipulations by the adversary. The proposed protocols, therefore, are not immune to relay attacks. This is not a major drawback given that a majority, if not all, of extant authentication protocols are not immune to relay attacks.

We proposed and evaluated an ownership transfer protocol for a hierarchical ownership scenario in a supply chain context. The use of sub- (or local) keys facilitates the assignment of different privileges to different levels in the hierarchy. We intend to model this scenario with different privileges as an extension to this paper. We also intend to extend this work to a scenario where an RFID-tagged entity is simultaneously jointly owned/rented from two different levels in the hierarchy. As RFID authentication and ownership transfer protocols are continually developed and improved, there is a need to identify scenarios that have not yet been studied for protocol development purposes. Although a majority of extant published protocols target one-owner/one-tag/one-way-authentication scenarios, other scenarios are no less important in terms of significance and these scenarios demand attention from researchers. Given the explosion of interest in RFID-tagging entities in a wide variety of applications, there is an urgency in this endeavor.

References

1. Brands, S., Chaum, D.: Distance Bounding Protocols. In: Helleseth, T. (ed.) EUROCRYPT 1993. LNCS, vol. 765, pp. 344–359. Springer, Heidelberg (1994)
2. Desmedt, Y.: Major Security Problems with the 'Unforgeable' (Feige)-Fiat-Shamir Proofs of Identity and How to Overcome Them. In: Proceedings of the Securicom 1988, 6th Worldwide Congress on Computer and Communications Security and Protection, pp. 147–159 (1988)
3. Hering, J.: The BlueSniper 'rifle'. Presented at 12th DEFCON, Las Vegas (2004)
4. Juels, A.: Yoking-Proofs for RFID Tags. In: Proceedings of the First International Workshop on Pervasive Computing and Communication Security (PerSec), pp. 138–143. IEEE Press (2004)
5. Kapoor, G., Piramuthu, S.: Single RFID Tag Ownership Transfer Protocols. IEEE Transactions on Systems, Man, and Cybernetics - Part C (2011)

6. Kapoor, G., Zhou, W., Piramuthu, S.: Multi-item and Multi-owner RFID Ownership Transfer in Supply Chains. Decision Support Systems (2011)
7. Kfir, Z., Wool, A.: Picking Virtual Pockets using Relay Attacks on Contactless Smartcard Systems. In: Proceedings of the 1st International Conference on Security and Privacy for Emerging Areas in Communication Networks (SecureComm), pp. 47–58 (2005)
8. Meiller, Y., Bureau, S., Zhou, W., Piramuthu, S.: Adaptive Knowledge-based System for Health Care Applications with RFID-generated Information. Decision Support Systems 51(1), 198–207 (2011)
9. Piramuthu, S.: RFID Mutual Authentication Protocols. Decision Support Systems 50(2), 387–393 (2011)

Negotiation and Auction Mechanisms: Two Systems and Two Experiments

Gregory E. Kersten[1], Pierpaolo Pontrandolfo[2], Rustam Vahidov[1], and Dmitry Gimon[1]

[1] InterNeg Research Centre, Concordia University, Montreal, Canada
[2] Dep. of Environmental Eng. & Sustainable Development, Politecnico di Bari, Bari, Italy
pontrandolfo@poliba.it,
{gregory,dgimon,r_vahidov}@jmsb.concordia.ca

Abstract. Auction and negotiation are mechanisms used in market exchanges. Behavioral economics experiments focused on the mechanism efficiency which required highly simplified problems and contexts. This paper discusses an on-going project involving an experimental comparison of auction and negotiation mechanisms embedded in software which we have developed. Both reverse multi-attribute auctions and multi-bilateral negotiations are used in a transportation service procurement scenario. The potential contribution includes the verification of theoretical claims that auctions are more profitable for auction givers than negotiations. It also includes formulation of guidelines for appropriate design of multiattribute market mechanisms and their selection.

Keywords: Auction and negotiation, multiattribute auctions, online auctions, e-negotiations, decision support systems, experimental study, e-procurement.

1 Markets, Transactions, and Mechanisms

Exchange mechanisms are sets of rules, which specify the functioning of the market and permissible behavior of its participants. The three standard mechanisms are: (1) catalogues, where requests and offers are posted; (2) auctions, where one side automates the process during which participants from the other side compete against each other; and (3) negotiations, where the participants bargain over the conditions of an exchange. One or more of these mechanisms are implemented in every e-marketplace.

Auctions can be either single-sided, where one seller auctions off goods to a number of bidders, or double-sided, where competition is employed on both sides of a market. They may differ in the process (e.g., ascending and descending), bidder acceptance, winner determination and other rules.

Negotiation is a rich and ill-defined family of processes used for exchanging goods or services among buyers and sellers, and for resolving inter-personal and inter-organizational conflicts. It involves an exchange of information comprised of offers, counter-offers and arguments with the purpose of reaching a consensus [1].

The economic view is the dominant one in auction mechanisms. In fact, the field of market design is almost exclusively focused on the mechanism design and applied

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 399–412, 2012.

auction theory. This focus has a major ramifications for practical market: *the decision regarding which exchange mechanism to use favors auctions* [2].

Auctions are well-structured and can be described completely and unequivocally using a set of rules and formulae. This led the computer science community involved in the design of early e-market mechanism, including negotiations, to propose replacing auctions with negotiations or claim that "negotiations are auctions". Sandholm [3], in an article that is entirely devoted to auctions, makes an opening statement saying that "Negotiation is a key component of e-commerce". Similarly, other authors who write about electronic business negotiations discuss solely auctions [4-6].

In practice, various market mechanisms are needed. In some situations flexibility and adaptability are sought over efficiency and speed, while in others it may be the ease of use and speed that are required. The requirements may depend on the nature of exchange (e.g., complex or simple, one-shot or repetitive, involving standard or unique goods, and between anonymous or well-known participants). Furthermore, market participants may have different needs which can be satisfied by different mechanisms. These requirements and needs motivate this research.

In procurement, for example, activities undertaken by business organizations differ in terms of the importance and the impact of the exchanged goods on profit as well as the inherent complexity and risk of the supply market. This led [7-9] to the identification of four types of buyer-supplier relationships: acquisition, strategic, noncritical and leverage. Recent studies [10, 11] confirmed that business organizations implement their procurement strategies based on these relationships.

About 70% of corporate revenue is spent on purchasing; savings of 5% translate into very significant amounts of money for companies of every size [12, 13]. Reverse auctions have been shown to achieve an average gross savings of 15-20 percent [14]. Most of these auctions are single attribute. However, a survey by Ferrin and Plank [15] found that over 90% of purchasing managers based their decisions on both price and non-price variables (e.g., durability, service, lead-time, and trust). Therefore, many organizations tend to modify the pure single-attribute auctions. The modifications include [16-18]:

1. Pre-selection of bidders so that only bidders who are known to meet the additional criteria are included;
2. Giving incumbents an advantage because their qualifications are known; and
3. The use of disclaimers such as "the lowest bid may not be awarded the contract".

The results of such auction modifications are mixed because of collusion and selection of inferior offers [19, 20]. In some situations the process becomes an auction in name only, as is the case with an auction in which neither the winner nor any other bidder is awarded the contract.

Most procurement decision problems are multi-attribute; therefore, there is a need to develop multi-attribute auctions that can be used in e-procurement. The first step in this direction is the design of an auction, however, the most recent survey of experimental auction research [21] does not include any multi-attribute auction experiment.

This paper reports an ongoing study in which we address three specific issues:

1. Design of a multi-attribute auction procedure and its implementation in a system which can be used by buyers and sellers.
2. Design and implementation of decision support aids which can support bidders and negotiators.
3. Behavioral comparison of multi-attribute auctions and multi-bilateral negotiations in which the participants use the same business case.

We use a multi-attribute auction procedure which requires no information about the bid-taker's preferences and allows the bidders to make progressive bids [22 , 23].

The paper is organized as follows. In Section 2 we summarize earlier studies on multiattribute auctions and negotiations. Review of the existing experimental studies is given in Section 3. Two systems which we have designed and implemented are briefly discussed in Section 4. We used these systems to conduct experiments discussed in Section 5. Conclusions and future work are given in Section 6.

2 Auctions and Negotiations

Auctions are defined by an explicit set of rules which determine resource allocation and prices on the basis of the bids made by the market participants [24]. The following four characteristics differentiate auctions from other exchange mechanisms [2]:

1. Auction rules are *explicit* and known to bidder, prior to the auction. Therefore, rules cannot be modified during the auction.
2. The rules describe the mechanisms completely thus allowing for the determination of one or more winners solely based on the bids. Auctioneers, or any other parties have no discretion in the winner choice.
3. Rules typically include:
   a. *bidding rules* stating how bids can be formulated and when they can be submitted;
   b. *allocation rules* describing who gets what on the basis of submitted bids; and
   c. *pricing rules* stating the corresponding prices the bidders have to pay; and
4. Auction rules allow mapping of bids on a single dimension – the price. Auctions focus on prices to achieve an efficient allocation or revenue maximization.

2.1 Multi-attribute Auctions

Che [25] and Branco [26] initiated studies on buyer's payoffs in two-attribute (i.e., price and quality) reverse auctions. The private information of buyers determining the utility can be represented in one dimension; this shortcut allows to apply the auction design apparatus to these problems. More recently, Beil and Wein [27] analyzed the problem of designing the multi-attribute auction. They were in particular concerned with finding a scoring rule to maximize buyer's utility.

The highly stylized information exchange in auctions makes it impossible for the sellers (buyers) to learn the preferences (needs, limitations) of the buyer (sellers). Therefore, much effort in multi-attribute auctions experiments has been devoted to the role and scope of preference revelation schemes. Bichler [28] conducted several experiments in which the bidders (sellers) were given the utility (value) function of the buyer. The results show that multi-attribute auctions do not provide substantial benefits over comparable single-attribute auctions. In other words, even with fully-revealed utilities the additional complexity outweighs the possible gains.

Koppius and van Heck [29] conducted experimental studies of the impact of information availability on the mechanism efficiency. The information availability specifies the type of information that is given as well as when, how and to whom it becomes available during the auction. They studied two types of multi-attribute English auctions:

1. Auctions with unrestricted information availability, in which suppliers are provided with the standing of the highest bid and the corresponding bidder as well as scores or bid ranking of the most current losing bids; and
2. Auctions with restricted information availability, in which the bidders are informed only about the standing of the highest bid.

The experiments indicated that auctions with unrestricted information availability yield higher efficiency than auctions with restricted information availability.

Strecker [30] analyzed the impact of preference revelation schemes on the efficiency of multi-attribute English and Vickrey auctions. He concluded that English auctions with revealed preference structure of the buyer are more efficient than both Vickrey auctions and English auctions with hidden preferences. Chen-Ritzo, Harrison at al. [31] introduced a multi-attribute English auction, where only partial information about the buyer's utility function was revealed. They showed that this variant performs better in terms of efficiency than a single attribute (price-only) auction. The multi-attribute auction outperformed over the single attribute auctions even though the bids in the former were far away from those predicted by theory. Notably, complexity in the auction mechanism consumes some of the efficiency gains over price-only auctions. This observation contradicts the findings reported by Bichler [28].

2.2 Multi-attribute Negotiations

One of the main questions of the research in multi-issue negotiation is how the representation of multi-issue preferences affects the negotiation outcomes.

Davey and Olson [32] compared a value-based negotiation system that used AHP with its pairwise comparing of criteria and alternatives with a goal-based NEGO system that asked users to set aspiration levels for criteria. This research confirmed the suggestion that conventional decision-making is goal-oriented and negotiators preferred to use goal-oriented method.

Lim [33] conducted an experiment involving executives and managers in Singapore and found that the acceptance of negotiation support systems mainly depends on the subjective norm and perceived behavioral control.

Several experiments were conducted using the e-negotiation system Inspire. The experimental research [34] conducted using Inspire confirmed the theoretical assumption that knowledge about counterpart's preferences contributes to the achievement of better outcomes. Negotiation Assistant [35] was used for the research on the effect of negotiation support on the results of negotiations. Experiments showed that using NSS in structured negotiation settings yields better outcomes for the negotiators as compared to face-to-face or email negotiations.

Experiments with the negotiation support system called Negoisst [36] led the authors to formulate five main challenges for computer-aided negotiations. They are the general limitations of preference elicitation due to problem complexity, the dynamics of preferences, the dynamics of the problem structure itself and its understanding, and the necessity for integrated decision support systems to deal with issue-by-issue negotiations.

3 Experimental Studies and Mechanism Comparison

Theoretical comparisons of auctions and negotiations are difficult because of the significant differences in the assumptions underlying each mechanism, as well as differences in participants' knowledge and behavior. Auctions assume that bidders know the buyer's valuation (price) of the good and follow strict and fixed protocol. Negotiation mechanisms have significantly weaker assumptions; the key assumption is that the parties negotiate in good faith and that the parties have preferences allowing each to compare the alternatives. There are no limitations on communication and no assumptions about the sellers' knowledge of the buyer's valuation.

Bulow and Klemperer [37] have shown in one of the first formal comparative studies that simple English auction with N+1 participating bidders (buyers) always yields higher revenue than a scheme they call "negotiation with N participants".

Bulow and Klemperer did not compare English auctions with anything that resembles negotiation as discussed in social science literature. The basis for their comparison was an exchange mechanism designed so as to maximize revenue of the seller. This mechanism is a type of an auction inasmuch as it does not allow for free interaction among the parties and requires the sellers to compete among themselves.

Kirkegaard revised Bulow and Klemperer's theory and included non-cooperative bargaining but still with very limited communication protocol. Manelli and Vincent [38] showed that the effects of auctions and negotiations would vary according to the situations; it is difficult to judge the effect of these two mechanisms on a given transaction without the consideration of the overall context, including the goods, participants, market, and so on. An important conclusion in this study was that auction mechanisms are frequently inefficient in a procurement environment, which contradicts the two previous studies.

In addition to theoretical comparisons, several experimental studies were conducted to compare auctions with negotiations. Thomas and Wilson [39, 40] conducted two studies in a laboratory settings. The first study [39] compared first-price auctions to multi-bilateral negotiations in a procurement scenario. They found that with more sellers (four sellers) the transaction prices in multi-bilateral negotiations were not significantly different from those in first price auctions. The transaction prices in multi-bilateral negotiations were higher than in first-price auctions when the number of sellers was reduced from four to two. Moreover, these two mechanisms were equal in terms of efficiency.

In their second study, Thomas and Wilson [40] compared second-price auctions to multi-bilateral negotiations with verifiable offers. They found that prices were lower in verifiable multi-bilateral negotiations than in second-price auctions. However, the efficiency of these two mechanisms was found to be statistically equivalent. By comparing these results to the first study, they ordered the four mechanisms (in terms of yielded transaction prices) from highest to lowest: second-price auctions, verifiable negotiations, non-verifiable negotiations, and first-price auctions.

Bajari, McMillan and Tadelis [11] conducted empirical analysis of auctions and negotiations in the construction industry. They observed that the use of the exchange mechanism depends on the knowledge and complexity of the context, task, and good. Negotiations have advantages, if the specifications of the product to be traded are not well-defined a priori, which is often the case in this industry. Negotiations, unlike auctions, allow for the discussion and clarification of the specifications.

We know of no experimental research in which multi-attribute auctions were compared with multi-issue multi-bilateral negotiations when no preference information of one side is disclosed to the other side.

4 Overview of the Imaras and Imbins Systems

Earlier experimental comparative studies of exchange mechanisms dealt with single issue auctions and negotiations (price in most cases) [40, 41]. If multiple issues are involved, the weighing of different issues should reflect the preferences of the business. However, the requirement that the auctioning or negotiating sides inform each other about their respective preferences seems unrealistic. The procedure that is embedded in the proposed auction system does not require disclosure of preference information.

We used two systems: (1) Imaras (InterNeg multi attribute reverse auction system) to conduct auctions; and (2) Imbins (InterNeg multi-bilateral negotiation system) to conduct negotiations. Both systems have been developed in the InterNeg virtual integrated transaction environment (Invite), discussed in detail by Strecker at al. [42].

Imaras supports several types of auction settings, including:

- Disclosure of bids to bidders: only the bidder's own bid is displayed, or both own and winning bids are displayed, or all bids are displayed;
- Bidding process: continuous (asynchronous bidding) or round-based (synchronous bidding); with rounds being defined by time (e.g., number of minutes or hours) or defined by a rule (e.g., number of submitted bids).

Imbins supports multi-bilateral negotiations in which the parties negotiate on the same or similar subsets of issues by exchanging offers consisting of one or more issues (attributes) and free text messages.

Imaras's main screen is shown in Figure 1; it is the bidding screen of a round-based auction in which the bidder can see his or her own bids as well as the winning bids. Imbins's main screen is shown in Figure 2; it is the message and offer submission screen in which the negotiator can see his or her own offers and messages as well as those of the counterpart. Both interfaces have four main components.

The clock (A) shows time from the beginning of the auction and the time left to the deadline. The systems' navigation bars are located on the right-hand side (B) where links to active pages are listed. For auctions the round number and clock are also given. The clock is reset at the beginning of every round.

Section C of both bidding and offer screens contains the most recent winning bids and offers made by the seller (who sees this screen) and by the buyer. In auction, only winning bids and bids made by the bidder who sees this screen are shown. In negotiation, only offers made by the buyer to all sellers or to the seller who sees the screen are shown.

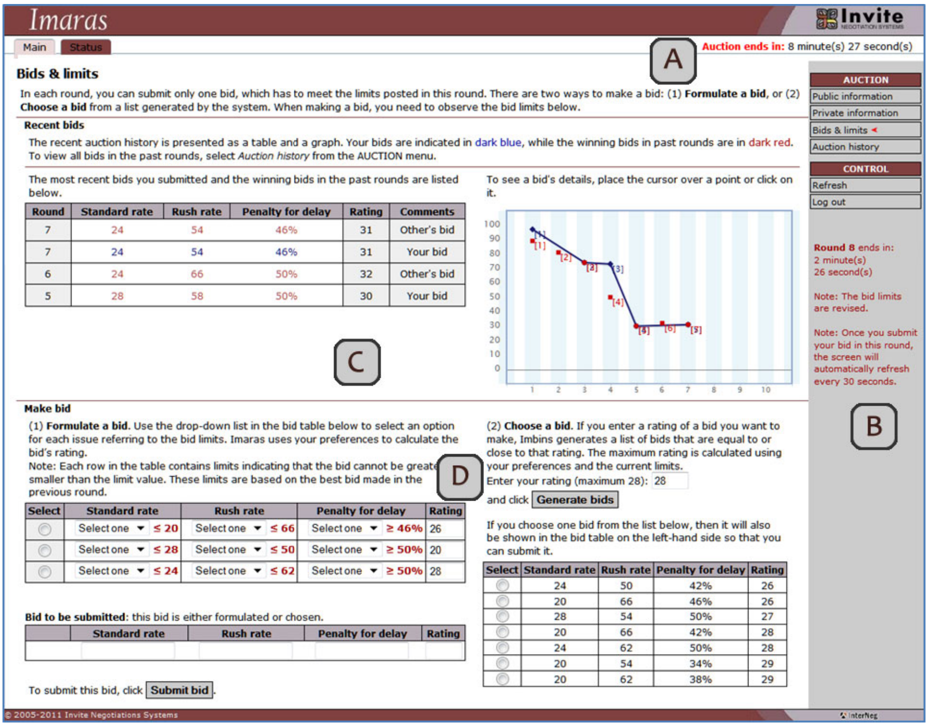


Fig. 1. Imaras auction: Bid construction and submission screen

The most recent bids and offers are shown in both tabular and graphical forms. In this section there is a difference between the auction and negotiation pages. In the auction page, the auction round and the winning bid are listed. In the negotiation page, messages sent by a counterpart can be accessed (they are expandable). Bids and offers are constructed and submitted in Section D. The two main differences between auction and negotiation are in the limit sets for the former and messaging facility for the latter.

The limit sets are determined by the multi-attribute auction procedure [23, 43]. They are bounds imposed on the attribute values and they assure that the bid in one round is not worse for the bid-taker than the winning bid in the previous round. Because there may be several limit sets (three are shown in Figure 1), the bidder can select one set (table row) and then select admissible attribute values for the given set.

Then in the next table the selected values appear. This table shows the auction bid or negotiation offer.

There are two ways to input the values into the bid and offer tables. In addition to the one described above, the user may generate bids (offers). In order to do so, he or she has to enter the preferred value of rating (utility) and click on "Generate bids" ("Generate offers") button. Subsequently, a table (not shown) will appear below the button with up to ten bids (offers) with a rating which is equal to or close to the selected rating value.

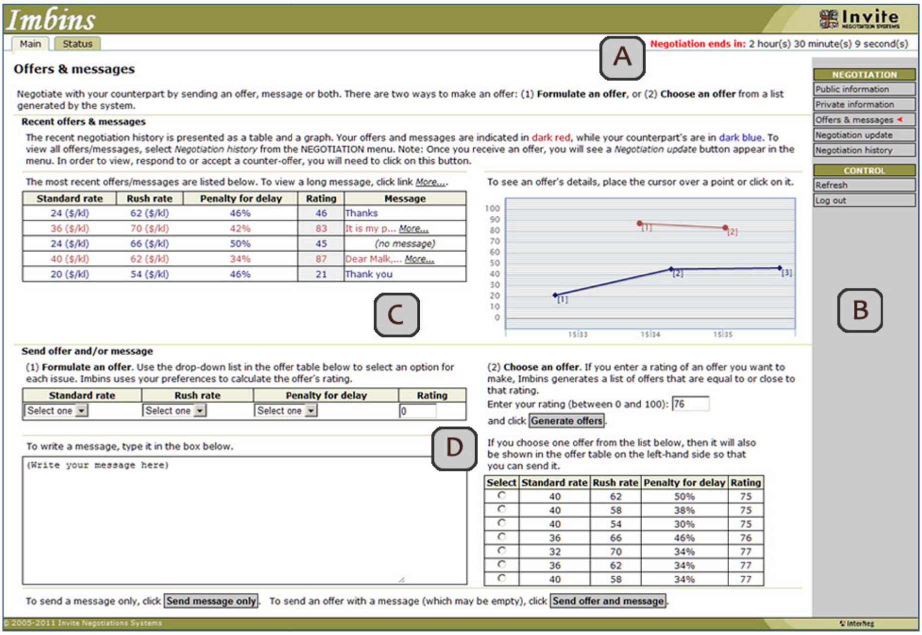


Fig. 2. Imbins negotiation: Offer with/w-out message construction and submission screen

5 Auction and Negotiation Comparison

In this section we describe results obtained from the experiments involving the two systems described above.

5.1 E-Procurement Case

We used the same procurement case in both auctions and negotiations.

A producer of perishable goods (the buyer) is seeking a logistics service provider who would transport goods from a single depot to a large number of customers. The buyer wants to sign a contract with a single provider for one year with a possibility of renewal. The buyer assures the minimum quantity of goods to be transported. There are five attributes: (1) standard rate of transportation; (2) the amount transported above the minimum quantity without additional charges; (3) rush rate for unexpected delivery; (4) penalty for delay in providing customers with the requested goods on time; and (5) penalty for the non-delivery or delivery of spoiled goods. The possible ranges for each attribute are known to every participant.

There are six providers with a proven record who are invited to the auction or negotiation.

Participants are told that the company they represent estimated a revenue function based on the problem attributes. For each configuration of attribute values, revenue value can easily be calculated using a simple calculator which is embedded in the case description. In order to simplify comparison of different offers or bids, the revenue is represented as ratings (secret) between 0 and 100 interval. Participants are also given breakeven ratings and are told that they should not accept contracts below this value. Such contracts bring forth losses for the firms their represent.

5.2 Experiment 1

The first experiment involved students from a Canadian university. The auctions and negotiations were conducted in the lab as well as online. One of the key differences between auctions and negotiations is the buyer's involvement. The buyer may follow different strategies and tactics, making comparison of the two processes difficult. Therefore, we selected buyers from graduate and senior undergraduate students and gave them detailed instructions regarding their behavior. Some of the buyers were asked to follow an integrative strategy, while others—a cooperative strategy.

Selected information about the results is given in Table 1.

Table 1. Experiment 1 results (sellers only)

	Auction		Negotiation*			
	Lab	Online	Lab-In	Lab-Cp	Online-In	Online-Cp
No. of instances	21	15	31	32	7	8
Agreement (%)	—	—	93.5	93.7	100	100
No. of offers (w & w-out msg.)/bids	5.6	3.2	6.3	6.5	2.8	2.9
Seller's profit	-9.5	16.3	13.9	7.7	9.1	4.7
Buyer's profit	80.5	53.3	63.6	69.8	60.4	65.3
No. of dominating alternatives	0.1	2.9	1.7	1.0	3.6	3.4
Overall satisfaction (1-7)	4.7	4.8	5.2	5.3	4.6	4.7
Time (hrs.)	0.24	61.4	0.35	0.36	37.4	45.0

* In – integrative; Cp – competitive.

In the lab setting, negotiations took longer on average than auctions (35 and 36 min vs. 24 min.). This is understandable because the negotiators-sellers were interacting with the buyers and they needed time to write and read messages. This can be contrasted with much longer time used in online auctions than in negotiations. In this case, however, the likely reason is the auction protocol bidder had to follow: the time allocated to each round was fixed and equal to one day.

The number of bids (offers) made significantly differs between the lab and online conditions: for online settings it is significantly lower. The difference is much smaller when auctions are compared with negotiations. This may suggest that making offers is not much more difficult than constructing bids.

In our experimental settings the outcome of every auction is an agreement. This is because the initial auction reservation levels are very favorable for the sellers. It is not the case for the negotiations. Therefore, the percent of agreements is generally lower in negotiations. Interestingly, the sellers reached worse agreements in lab auctions (-9.5) than in negotiations (13.9 and 7.7). In the negotiations, the sellers who negotiated with integrative buyers reached better deals than those who negotiated with competitive ones. For the buyers the results were somewhat opposite: lab auctions yielded the best deals, followed by the deals when buyers were competitive, while the worst deals were achieved by cooperative buyers.

The situation was very different in online settings. Online buyers negotiated better deals through auctions than through negotiations. (Again, this may be due to the way the auctions were set up.) However, the deal-making ability of integrative and competitive negotiators did not change: competitors achieved more. Overall satisfaction levels were similar across all cases.

5.3 Experiment 2

The second experiment involved students from an Italian University. The auctions and negotiations were conducted in the lab and, together with the preparation time they lasted two hours. The participants were third year undergraduate students; they were playing the role of the sellers. In negotiations, the buyers were junior researchers who were trained to follow integrative or cooperative strategies.

The main difference between Experiment 2 and the lab portion of Experiment 1 is that in this experiment two versions of the case were used. The three-attribute case used here was the same as the case used in Experiment 1. In addition, we also used two-attribute case (one attribute was dropped from the initial case).

Selected information about the results is given in Table 2.

Table 2. Experiment 2 results (sellers only)

	Two attributes			Three attributes		
	Auction	Integrate	Compete	Auction	Integrate	Compete
No. of instances	11	7	7	11	7	7
Agreement (%)	—	100	100	—	100	100
Offers (w. & w-out msg.)/bids	2.6	3.5	3.5	4.9	3.1	3.1
Seller's profit	-4.91	21.8	3.3	-9.2	9	2.3
Buyer's profit	76.7	55.0	65.4	77.3	60.3	67.8
No. of dominating alternatives	0	0	0	0.8	7.1	3.6
Overall satisfaction (1-7)	5.4	5.7	5.2	5.3	5.3	5.4
Time (hrs.)	0.9	0.15	0.12	0.22	0.15	0.11

As seen from the table, in this experiment the agreement rate was 100%. In a two-attribute case the average number of offers was higher for negotiation instances than the number of bids in an auction. In a three-attribute case, however, the situation is reversed; perhaps, due to the increase of the negotiation task complexity and the required cognitive effort. Sellers' profits were lower in auctions as compared to negotiations. Sellers made more profit in integrative, vs. competitive settings. Again, for the buyers the opposite was true: the highest profit they made in auctions, followed by competitive negotiations, and then integrative negotiations. In a two attribute case non-dominated alternatives were selected while in a three-attribute case dominated alternatives were agreed upon. This especially applied to negotiation cases, with integrative settings being the worst in this respect.

6 Conclusions

In this paper we have described the exploratory experiments aimed at investigating the differences between multi-attribute auctions and negotiations. The results indicate that there may be important differences between the two types of mechanisms in terms of process and outcome variables. One of the major lessons learned from the experiments is that participants need more time to learn and understand the procedure, system and case.

Observation of the negotiation and bidding processes and comments from the experiment participants led us to treat the experiments as extensive testing rather than a research experiment. The earlier system and usability testing did not show the participants' lack of good understanding of the problem and process. During the two experiments, we observed that some participants were lost and/or uninterested.

We received both positive comments ("positive overall experience", "fun to use", "enjoy the challenge", and "good learning experience") as well as negative ones ("not clear process", "difficult construction of bids", "no guidance"). The latter comments and the results of the experiments, in particular the losses that the winners "brought in" to the firms they represented, led us to realize that multi-attribute auctions are difficult and that we need to provide more and better tools for learning about the system, its use and the specifics of the bidding process. To this end, we have prepared several training materials.

Participants in our experiments are students who differ in their motivation and interest to learn the system and the case. In order to provide a more even field so that every participant knows the basics of the system and the bidding process, we are developing a demo followed by a short quiz which will test students' understanding of the system and its use. Students will watch the demo and will be asked to take the quiz about one week before the experiments. The next step aiming at increasing students' understanding of the process is breaking up the process into two separate phases. One phase will be preparation which will take between three days during which students will log in to the system and learn about the case. Before moving to the next phase, which involves bidding, students will have to pass a test.

We expect that the changes in the system and the experimental process will allow us to formulate concrete postulates for the researchers and practitioners. These will focus on the relationships between the context of market exchanges, the market and characteristics of goods exchanged, and the market participants. There are also accomplishments which this study has already achieved. We have shown that multiattribute reverse auctions can be conducted without unveiling the buyer's preferences. We have also demonstrated that it is possible to conduct feature-rich experiments in which both the systems and cases are similar to these available in real-life. Other implications of this study include the need to consider multiple exchange characteristics; the sole focus on substantive outcomes (e.g., deal values) is not sufficient. This is because in real-life such aspects as reputation, trust, empathy, and fairness often play an important role in interactions between buyers and sellers.

Acknowledgments. We thank Norma Paradis and Shikui Wu for their contributions to the development of the Invite systems and conduct of the experiments. This work has been supported by Regione Puglia (APQ PS025), and grants from the Social Sciences and Humanities Research Council of Canada (SSHRC), the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Bichler, M., Kersten, G.E., Strecker, S.: Towards the Structured Design of Electronic Negotiation Media. Group Decision and Negotiation 12, 311–335 (2003)
2. Kersten, G.E., Chen, E., Neumann, D., Vahidov, R., Weinhardt, C.: On Comparison of Mechanisms of Economic and Social Exchanges: The Times Model. In: Gimpel, H., Jennings, N., Kersten, G.E., Ockenfels, A., Weinhardt, C. (eds.) Negotiation, Auction and Market Engineering, pp. 16–43. Springer, Heidelberg (2008)
3. Sandholm, T.: Automated Negotiation. The Best for All Concerned. Communication of the ACM 42, 84–85 (1999)
4. Beam, C., Segev A., Shanthikumar, J.G.: Electronic Negotiation through Internet-based Auctions. CMIT Working Paper 96-WP-1019, Haas School of Business, University of California (1996), http://haas.berkeley.edu/~citm/WP-1019.PDF
5. Kumar, M., Feldman, S.I.: Business Negotiation on the Internet. IBM Institute for Advanced Commerce (1998)
6. Ströbel, M.: On Auctions as the Negotiation Paradigm of Electronic Markets. Electronic Markets 10, 39–44 (2000)
7. Kraljic, P.: Purchasing Must Become Supply Management. Harvard Business Review 83, 109–117 (1983)
8. Larson, P.D., Carr, P., Dharwall, K.S.: SCM Involving Small versus Large Suppliers: Relational Exchange and Electronic Communication Media. The Journal of Supply Chain Management 41, 18–29 (2005)
9. Handfield, R.B., Straight, S.L.: What Sourcing Channel is Right for You? Supply Chain Management Review 7, 63–68 (2003)
10. Subramanian, G., Zeckhauser, R.: For Sale, but How? Auctions Vs. Negotiations. Negotiation, 3–5 (2004)
11. Bajari, P., McMillan, R., Tadelis, S.: Auctions versus Negotiations in Procurement: An Empirical Analysis. Journal of Law, Economics, and Organization 25, 372–399 (2009)
12. Peleg, B.: The Value of Procurement via Online Bidding. Whitepaper, p. 3. CRMTODAY, Athens (2003)
13. Wagner, S.M., Schwab, A.P.: Setting the Stage for Successful Electronic Reverse Auctions. Journal of Purchasing and Supply Management 10, 11–26 (2004)
14. Cohn, L.: B2B: The Hottest Net Bet Yet? Business Week (2000)

15. Ferrin, B.G., Plank, R.E.: Total Cost of Ownership Models: An Exploratory Study. Journal of Supply Chain Management 38, 18–29 (2002)
16. Bichler, M., Kalagnanam, J.: Configurable Offers and Winner Determination in Multi-attribute Auctions. European Journal of Operational Research 160, 380–394 (2005)
17. Engelbrecht-Wiggans, R., Haruvy, E., Katok, E.: A Comparison of Buyer-determined and Price-based Multi-attribute Mechanisms. Marketing Science 26, 629–641 (2007)
18. Schoenherr, T., Mabert, V.A.: Online Reverse Auctions: Common Myths versus Evolving Reality. Business Horizons 50, 373–384 (2007)
19. Elmaghraby, W.: Auctions and Pricing in E-marketplaces. In: Simchi-Levi, D., Wu, D., Shen, M. (eds.) Handbook of Quantitative Supply Chain Analysis: Modeling in the E-business Era. Kulwer, Norwell (2004)
20. Katok, E., Wambach, A.: Collusion in Dynamic Buyer-Determined Reverse Auctions. Management Science (2011) (forthcoming)
21. Kagel, J.H., Levin, D.: Auctions: A Survey of Experimental Research, 1995-2010. Handbook of Experimental Economics (2012)
22. Pontrandolfo, P., Wu, S., Moramarco, R., Kersten, G.E.: Auctions and Negotiations in Transportation Service Procurement. In: Proceedings of Group Decision and Negotiations Conference, pp. 241–252 (2010)
23. Kersten, G.E., Pontrandolfo, P., Wu, S.: A Multiattribute Auction Procedure and Its Implementation. In: Proceedings of HICSS 45 (2012)
24. McAfee, R.P., McMillan, J.: Auctions and Bidding. Journal of Economic Literature 25, 699–738 (1987)
25. Che, Y.-K.: Design Competition through Multidimensional Auctions. RAND Journal of Economics 24, 668–680 (1993)
26. Branco, F.: The Design of Multidimensional Auctions. Rand Journal of Economics 28, 63–81 (1997)
27. Beil, D.R., Wein, L.: An Inverse-optimization-based Auction Mechanisms to Support a Multiattribute RFQ Process. Management Science 49, 1529–1545 (2003)
28. Bichler, M.: An Experimental Analysis of Multi-attribute Auctions. Decision Support Systems 29, 249–268 (2000)
29. Anderson, E.T.: Sharing the Wealth: When Should Firms Treat Customers as Partners? Management Science 48, 955–971 (2002)
30. Strecker, S., Seifert, S.: Electronic Sourcing with Multi-attribute Auctions. In: HICSS 37 (2004)
31. Chen-Ritzo, C.H., Harrison, T.P., Kwasnica, A.M., Thomas, D.J.: Better, Faster, Cheaper: An Experimental Analysis of a Multiattribute Reverse Auction Mechanism with Restricted Information Feedback. Management Science 51, 1753–1762 (2005)
32. Davey, A., Olson, D.: Multiple Criteria Decision Making Models in Group Decision Support. Group Decision and Negotiation 7, 55–75 (1998)
33. Lim, J.: A Conceptual Framework on the Adoption of Negotiation Support Systems. Information and Software Technology 45, 469–477 (2003)
34. Vetschera, R.: Preference Structures and Negotiator Behavior in Electronic Negotiations. Decision Support Systems 44, 135–146 (2007)
35. Rangaswamy, A., Shell, G.R.: Using Computers to Realize Joint Gains in Negotiations: Toward an "Electronic Bargaining Table". Management Science 43, 1147–1163 (1997)
36. Köhne, F., Schoop, M., Staskiewicz, D.: Decision Support in Electronic Negotiation Systems-New challenges. In: DSS Conference (2004)
37. Bulow, J., Klemperer, P.: Auctions versus Negotiations. American Economic Review 86, 80–194 (1996)

38. Firth, A.E.: The Discourse of Negotiation. Studies of Language in the Workplace. Elsevier, New York (1995)
39. Thomas, C.J., Wilson, B.J.: Verifiable Offers and the Relationship Between Auctions and Multilateral Negotiations. The Economic Journal 115, 1016–1031 (2005)
40. Thomas, C.J., Wilson, B.J.: A Comparison of Auctions and Multilateral Negotiations. The RAND Journal of Economics 33, 140–155 (2002)
41. Thomas, C.J., Wilson, B.J.: Verifiable Offers and the Relationship Between Auctions and Multilateral Negotiations. Economic Journal 115, 1016–1031 (2005)
42. Strecker, S., Kersten, G.E., Kim, J.B., Law, K.P.: Electronic Negotiation Systems: The Invite Prototype. In: Proceedings of the Collaborative Business MKWI 2006, pp. 315–331. GITO (2006)
43. Kersten, G.E., Wu, S., Szapiro, T.: A Procedure for Multiattribute Reverse Auctions with two Strategic Parameters. InterNeg Report INR3/11. Concordia University (2011)

Research on the Hybrid Push/Pull Production System for Mass Customization Production

Jie Lin[1], Xing Shi[1,*], and Yu Wang[2]

[1] School of Economics and Management, Tongji University, No. 1239 Siping Road, 200092, Shanghai, China
linjie@tongji.edu.cn, shixinghoney@yahoo.cn
[2] School of Information Science, Guangdong University of Business Studies, 510320, Guangzhou, Guangdong, China

Abstract. To start with, literatures of the CODP theory and the researches on the hybrid push/pull production system are reviewed. CODP is the point where push production integrate with the push production mode, therefore based on the research of manufacturing strategy of the CODP, the production planning model of the push/pull production of the single-CODP mass customization system is put forward. Afterwards the model is extended to the multi-CODP mass customization production system and the production planning of the working points where the push and pull production coexist is emphatically discussed.

Keywords: CODP(customer order decoupling point), mass customization, production planning.

1 Introduction and Literature Review

The manufacturing mode has experienced a series of transformation from the earliest form of customization production, to mass production mode characterized by the high automation level as well as high production efficiency but low flexibility. In recent decades, the acceleration in production efficiency has shortened the delivery time tremendously, meanwhile the globalization as well as the pursuit of the individualism leads to the diversification among the customer demands. In order to provide a wider range of products, or customized products and even customer-specific one, customers might need to be involved in the activities of delivery, manufacturing, engineering and even designing, the level of which are determined by the practical customization level of the products. Thus inevitably the delivery time of the product is to become longer. So, there emerges the conflict between the delivery time and the customization level.

To solve the conflict, nowadays, many manufacturing enterprises have turned to mass customization production and managed to reduce customer's waiting time and provide products of a higher customization level at the same time. The idea of mass customization was first brought about by Alvin Toffler(1970) and was illustrated and

---

[*] Corresponding author.

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 413–420, 2012.

discussed systematically by B.Joseph Pine II (1993). The strategy of conducting mass customization production involves two key factors: the modularization in the designing process and the postponement strategy during the manufacturing stage (Shao 2001). The key concept in the postponement strategy is customer order decoupling point (CODP).The CODP is normally defined as the point in the flow of goods where forecast-driven production and customer-order-driven production are separated (Giesberts and van den Tang 1992, Wortmann et al. 1997).

A single fixed CODP in the production process is most widely and thoroughly discussed. A single CODP structure connects with the postpone strategy by Bucklin (1965), and also the position of CODP corresponds to distinguished manufacturing strategy: engineer-to-order (ETO); make-to-order (MTO); and assemble-to-order (ATO) (Hoekstra and Romme 1992, Browne et al. 1996, Higgins et al. 1996, Sackett et al. 1997, Wortmann et al. 1997, Mather 1999).

From the production perspective, it is a fact that upstream the CODP, the material flow is forecast driven and the push production is applicable, at the same time downstream the customer order triggers the production and the pull production is applied. Thus the CODP, considered as a major element in the hybrid push- and pull-type manufacturing strategy, is the very point where the push production integrates with the pull manufacturing, in which the inventory of half-finished products are held. Scrutinizing the literatures, many scholars have done researches on the hybrid push/pull manufacturing system (Olhager and Östlund 1990, Omar Ghrayeb et al. 2009, Yasuhiro Hirakawa 1996, Cochran and Kim 1998).

Most scholars accept that all products of enterprise (or supply chain) have one, or several, CODP which is immovable based on the process. However Lin et al.(2004) proposed the mass dynamic customization system(MDCS) in which several CODPs can be set corresponding to different customization levels and they are movable when the customer demand trend changes. Scrutinizing the literatures, at present the majority of the researches on the production planning of the CODP where the push- and pull-type production integrates are based on the condition of one single fixed CODP, in this paper we will discuss the production planning of several CODPs within the mass customization production system.

2 The Production Planning of Single-CODP Mass Customization Production System

2.1 The Single-CODP Mass Customization Production System

The key point of designing single-CODP mass customization production system lies in the positioning of the CODP in the whole production process. The more forward the CODP is located, the higher level of customization the product is and the wider product range is provided to customers, meanwhile the longer lead time is required. The reverse is also true. When the CODP is set relatively closer to customers, manufacturers carry out predicted production of the semi-finished products in the upstream process of the CODP, and directly assemble and deliver products once orders from customers are received. In this situation, the lead time is comparatively short, but at the same time customers can only enjoy the products of little customization. Therefore, currently, the mass customization production system generally weigh and balance the personalization range provided, the customer's bearable delivery time, product characteristics, the modularization and standardization level, and manufacturers' managerial competence so as to determine the appropriate position of the CODP in the production process.

Here the manufacturing process is abstracted to a sequence of working points, and when the raw materials go through this sequence, they are processed in each working point and eventually turned into finished products, as shown in Fig 1., in which the square represents working points and the triangle shows the position where the CODP is located. In Fig 1.we can see that CODP is located between working point i and working point i+1, and according to the processing flow direction, the working point 1 to i are called working points upstream CODP and i+1 to n are the ones downstream CODP. The working points upstream CODP carries out push production (shown as the broken line), while the ones downstream the CODP carry out the pull production (shown as the solid line).

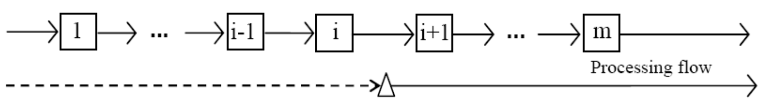


Fig. 1. The single-CODP mass customization production system model

2.2 The Production Planning Modeling of the Single-CODP Production System

Based on the above analysis, there is certain amount of half-finished products in the every place where CODP is located in the whole production process. The followings discuss how to arrange push mode production and the inventory strategy of semi-finished products so as to gain economic benefit of the large-scale production considering inventory cost, production start cost and production cost. The economic batch production model is proposed with the goal of minimizing average cost and by solving the model we get optimal production batch amount and optimal production period.

2.2.1 Assumptions

Firstly, here we assume that the demand of each product conforms to continuous uniform distribution. Since any product can be decomposed to semi-finished products or parts according to BOM (Bill of Material), we assume the demand of semi-finished products is in proportion to that of each product, and the ratio is demand coefficient. Secondly, the (s, S) inventory strategy is adopted, and the out-of-stock situation is not allowed in the following model. Thirdly, at the beginning of each production cycle, the production start cost is required. The total cost TC consists of Stock Cost SC and Production Cost QC which includes production start cost, raw material cost, and unit production cost but transportation cost is excluded.

2.2.2 Modeling

Denote the demand rate of customized product i (i=1,2,…,n) as di, that of semi-finished products is d, the demand ratio of customized product i and semi-finished product i is α_i, so the average demand $\bar{\alpha} = \frac{\sum_{i=1}^{n} \alpha_i d_i}{\sum_{i=1}^{n} d_i}$, and $d = \bar{\alpha} \sum_{i=1}^{n} d_i$.

The demand amount of customized product within the production period T is $Z = \sum_{i=1}^{n} Z_i = \sum_{i=1}^{n} d_i T$.

Denote the production batch by Q, unit raw material cost and processing cost by c, production start cost by K, therefore the total production cost in a production period $QC = K + cQ$.

Because of (s, S) inventory strategy and the instant stock renewal, the stock of semi-finished products at the beginning of each production period is $S = s + Q$.

$I(t)$ denotes the stock of semi-finished products at t, then $I(t) = S - dt$.

The average stock of semi-finished products

$$\bar{S} = \frac{1}{T}\int_0^T I(t)dt = S - \frac{1}{2}dT = s + Q - \frac{Q}{2} = s + \frac{Q}{2}. \tag{1}$$

And within T the stock cost $SC = h\bar{S}T = h(s + \frac{Q}{2})T$, where h is the unit stock cost in unit time.

In summary, the average cost can be expressed in the function below

$$F(s, Q) = \frac{TC}{T} = c + hs + \frac{dK}{Q} + \frac{h}{2}Q. \tag{2}$$

By minimizing $F(s, Q)$, we get the optimal production batch $Q^* = \sqrt{\frac{2dK}{h}}$. And the optimal production period $T^* = \frac{Q^*}{d} = \sqrt{\frac{2K}{hd}}$.

Assume applying the above economic batch production strategy, for each working point upstream the CODP in the stage of push production, the production plan is that the optimal production batch Q^* needs to be manufactured in the optimal production period T^*. However, for the working points of pull production downstream the CODP, the production schedule should be arranged by the actual demand in customer orders.

3 The Multi-CODP Mass Customization Production System Production Planning

3.1 The Multi-CODP Mass Customization Production System

However, one deficiency of the single-CODP mass customization production system is that customers' requirement for customization level is not discriminated. Once the position of the CODP is set, for all customers the delivery time is equal. Therefore, for certain customers who just demand products with less personalization, they still need to endure the same delivery time, which result in reducing customer satisfaction. Thus, in the mass production system with only one fixed CODP, the contradiction between the production personalization and the delivery time always exists. The proposed multi-CODP mass customization production system manages to solve this problem and the followings mainly discuss the production planning model of it.

Similarly the multi-CODP mass customization production system is abstracted as a series of working points with several CODPs. As shown in Fig 2., the square represents working point, and the triangles label where CODPs are positioned, with the number on the up right of it denoting the serial number of CODP. Within the mass customization system with n CODPs and m working points, processing conducted by working points 1 to j are push mode production labelled as Stage I; processing by working points j+1 to k are the hybrid push and pull mode production labeled as Stage II; and the processing by working point k +1 to m is pull mode labeled as Stage III. Here we adopt the concept of the product mix in the positioning of multiple CODPs (Wang et al.2010). Multiple end-products may share some common components processes and have their own customization attributes. Products sharing common components processes correspond to the same CODP, and are categorized to a kind of product mix, thus CODP i corresponds to product mix i (PM i). Therefore n CODPs in Stage II correspond to n product mixes separately and n types of delivery time, which solves the contradiction between customized characteristics and waiting time.

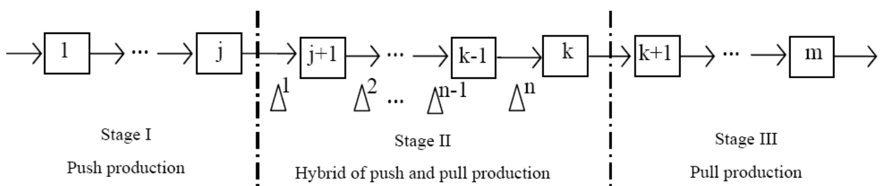


Fig. 2. The multi-CODP mass customization production system model

3.2 The Production Planning Modeling of the Multi-CODP Production System

Based on the multi-CODP mass customization production system model, the followings illustrate how to calculate the production volume of the working point of each stage. In Fig 3**.**, the broken line represents the processing of push mode, while the solid line is that of pull mode. For each single working point, whether it carries out push production or pull production determines how it arranges the production plan. For Stage I and Stage III, the working points within them conduct a single mode of production separately, that is pure push production in Stage I and pure pull production in Stage III. However, the push production and pull production mode coexists in the working points *j*+1 to *k* in Stage II. There are *n* CODP totally in this processing flow, each of which corresponds with a kind of product mix and they are distributed among working points, where the corresponding semi-finished products are kept.

For each working point in Stage II, according to its relative position to n CODPs separately, the processing within it is the combination of push and pull mode production. Thus we can decompose the production in this working point into the pull production part and push production part according to the CODPs. Take the working point m2 for example. In Fig 3., we can see the vertical line L through m2 intersects with the horizontal lines through CODPs. For CODP 1, the line L intersects at the solid line part across CODP 1, representing the processing corresponding to CODP 1 by m2 is pull production and the production volume of this part of processing is determined by the actual demand of the orders for PM 1. It is the similar case for CODP 1 to M-1. However for CODP M, Line L intersects with the broken line part of the line across CODP M. Therefore the processing of the corresponding PM M by m2 is push production and the production volume of this part is calculated based on the prediction of the demand of PM M. This case also suits CODP M to n.

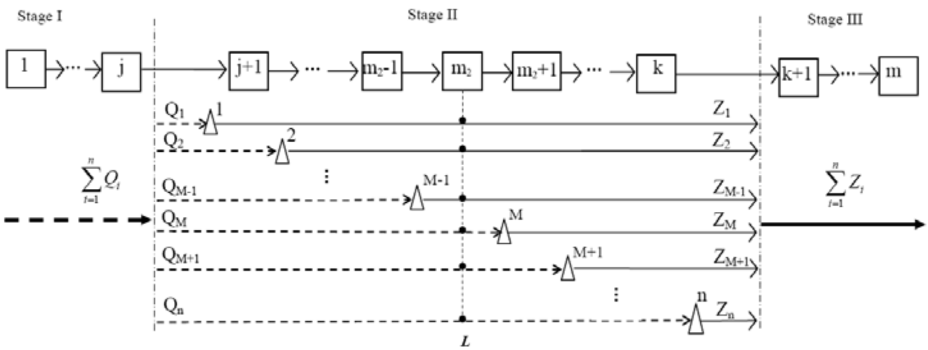


Fig. 3. The production planning model of multi-CODP mass customization production system-decomposing CODPs

Based on the above illustration and the economic lot model of the semi-finished products, the followings give the general model of the production volume calculation of the working points in the multi-CODP mass customization production system.

(1)Denote the node in the Stage I (including node 1 to j) as m_1 (m_1 =1,2,…,j), thus P_{m_1}, the productive task of m_1, is

$$P_{m_1} = \sum_{i=1}^{n} \frac{Q_i^* T}{T_i^*}, \tag{3}$$

where Q_i^* and T_i^* represent the optimal production batch and optimal production cycle respectively.

From the economic lot production model we can get $Q_i^* = \sqrt{\frac{2D_i K_i}{h_i}}$, where K_i and h_i is the production start cost and unit inventory cost of semi-finished product i corresponding to PM i respectively, and D_i is its demand rate.

(2) Denote the node in the Stage III (including node k+1 to m) as m_3 (m_3=k+1,…,m), thus P_{m_3}, the productive task of m_3, is

$$P_{m_3} = Z_1 + Z_2 + ... + Z_n , \tag{4}$$

where Z_i is the actual demand for customized product i in the orders within T.

(3) Denote the node in the Stage II (including node j+1 to k) as m_2 (m_2=j+1,…,k), and assume m_2 is positioned just before CODP M, thus P_{m_2}, the production task of m_2 is

$$P_{m_2} = \sum_{i=1}^{M-1} Z_i + \sum_{i=M}^{n} \frac{Q_i^* T}{T^*} = \sum_{i=1}^{M-1} Z_i + \sum_{i=M}^{n} \frac{T}{T^*}\sqrt{\frac{2D_i K_i}{h_i}}. \tag{5}$$

4 Conclusion

This paper illustrates the concept of multi-CODP mass customization production system based on the research of the mass customization and the CODP, in which multiple CODPs are set in the production system according to different customization characteristics and delivery time. Then within the single-CODP mass production system, during the push production phase, manufacturers produce the standard or modular parts based on the prediction of the market trend and demand, which has the scale economic advantage. In contrast, during the pull production phase, the parts are assembled, packaged and transported specifically according to the customized requirements in the orders. So by solving economic lot production model of semi-finished products with the goal of minimization of average cost, we get the optimal production batch quantity and corresponding optimal production period of semi-finished products. And then the model is extended to the multi-CODP production system, emphasizing on the analysis of the production planning of the working points within the phase where push and pull production coexists. And finally, this paper concludes that when the manufacturers confront with production incompetence, they should consider and weigh factors such as the demand amount in the order, customer loyalty and historical sales information to set an order priority level and how to rearrange the production planning of the hybrid push and pull production.

Acknowledgements. The paper is supported by National Natural Science Foundation, China (No. 71071114), Doctoral Fund of Ministry of Education of China (No. 200802470009) and Shanghai Leading Academic Discipline Project, Project (No.B310).

References

1. Browne, J., Harhen, J., Shivnan, J.: Production management systems: An integrated perspective. Addison-Wesley, Harlow (1996)
2. Buchlin, L.: Postponement, speculation and the structure of distribution channels. Journal of Marketing Research 2, 26–31 (1965)

3. Cochran, J.K., Kim, S.-S.: Optimum junction point location and inventory levels in serial hybrid push/pull production systems. International Journal of Production Research 36(4), 1141–1155 (1998)
4. Ghrayeb, O., Phojanamongkolkij, N., Tan, B.A.: A hybrid push/pull system in assemble-to-order manufacturing environment. J. Intell. Manuf. 20, 379–387 (2009)
5. Giesberts, P.M.J., van den Tang, L.: Dynamics of the customer order decoupling point:impact on information systems for production control. Production Planning & Control 3(3), 300–313 (1992)
6. Higgins, P., Le Roy, P., Tierney, L.: Manufacturing planning and control: Beyond MRPII. Chapman & Hall, London (1996)
7. Hill, T.J.: Manufacturing strategy: Text and cases. Macmillan, Basingstoke (1995)
8. Hirakawa, Y.: Performance of a multistage hybrid push/pull production control system. International Journal of Production Economics 44, 129–135 (1996)
9. Hoekstra, S., Romme, J.: Integrated logistics structures: Developing customer oriented goods flow. McGraw-Hill, London (1992)
10. Lin, J., Jiang, J.J., Xu, H.: The supply chain management system for mass dynamic customisation based on GDSS. Industrial Engineering and Management 1, 63–67 (2004)
11. Mather, H.: How to profitably delight your customers. Woodhead Publishing Ltd., Cambridge (1999)
12. Olhager, J.: The role of the customer order decoupling point in production and supply chain management. Computers in Industry 61, 863–868 (2010)
13. Olhager, J., Ostlund, B.: Integrated push-pull manufacturing strategy. European Journal of Operational Research 45(2/3), 135–142 (1990)
14. Pine II, B.J.: Mass Customization-the New Frontier in Business Competition. Harvard Business Press, Boston (1993)
15. Sackett, P.J., Maxwell, D.J., Lowenthal, P.A.: Customising manufacturing strategy. Integrated Manufacturing Systems 8(6), 359–364 (1997)
16. Shao, X.F., Ji, J.H., Huang, P.Q.: An analysis of Mass Customization. Industrial Engineering and Management 1, 13–17 (2001)
17. Wang, F., Lin, J., Liu, X.: Three-dimensional model of customer order decoupling point position in mass customisation. International Journal of Production Research 48(13), 3741–3757 (2010)
18. Wortmann, J.C., Munstlag, D.R., Timmermans, P.J.M.: Customer-driven manufacturing. Chapman & Hall, London (1997)

A Study of Users' Intention to Voluntarily Contribute Real-Time Traffic Information through Mobile Devices

Chen Zhu[1], Kai Kwong Wat[2], Chao Ren[3], and Stephen Shaoyi Liao[4]

[1] Department of Information Systems City University of Hong Kong
chenzhu2@student.cityu.edu.hk
[2] Hsin Chong Construction Group Limited Hong Kong
richardwat2003@gmail.com
[3] Department of Information Systems City University of Hong Kong
st03rc@gmail.com
[4] Department of Information Systems City University of Hong Kong
issliao@cityu.edu.hk

Abstract. Advanced traffic information systems are increasingly available to provide users with real-time traffic information services. Nowadays, this kind of public information collection relies extensively on contributions made by end users on a voluntary basis. Although the literature has looked into the motivations that may lead to free online contributions, it has largely overlooked the actual use of the same systems based on mobile devices. Considering the issue from the perspectives of complexity, social effects and practical considerations, this paper explores and examines factors influencing mobile device users to freely contribute to a traffic information system. The methodology is planned to be a combination of qualitative and quantitative research methods. From the first stage of this research, which involves three face-to-face interviews and a focus group discussion among mobile application developers and users, we gleaned interesting results which are instructive for future empirical work.

Keywords: Advanced traffic information systems (ATIS), Mobile devices, Free user contribution, Theory of planned behavior (TPB).

1 Introduction

Providing drivers and passengers with relevant information about traffic conditions is generally acknowledged as having the potential to change their behavior when driving and traveling. Nowadays, advanced traffic information systems (ATIS) providing such information on platforms such as on navigation systems, online services, and especially mobile devices are developed in metropolises. For many years, real-time traffic information in Hong Kong has been provided primarily by the Government's Transport Department, from video camera systems, toll-stations, automatic vehicle detectors, and loop-coils. Of the above mentioned methods, the most important data source currently is from the video captured by closed-circuit TV cameras. In April 2010, there were about 400 closed-circuit TV cameras all around Hong Kong.

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 421–428, 2012.

Unfortunately, real-time traffic data collection only covers a few major roads and tunnels. Mobile, location-based applications help collect real-time traffic information and are poised to make a leap in the quality of information provided, which rely on mobile devices for collecting, processing, and disseminating real-time traffic information. As mobile devices are carried by drivers and passengers all the time, mobile device users only need to install the appropriate software on their mobile devices, choose to open the software, and admit to enable the real-time traffic information uploading function. In the near future, public real-time traffic information might rely entirely on voluntary user contributions, so called free user contributions. The greater the extent to which real-time traffic information can be collected, so the more accurate and better the prediction of short future traffic condition services can be provided.

However, the users could directly use the information services analyzed and provided by the application without voluntarily contributing to the public. What differentiate contribution of real-time traffic information behavior using mobile devices seems to be that 1) the users are using mobile devices and encountering constraints (Clarke I 2001), 2) real-time information means time-sensitive, and 3) the third is they are contributing location-sensitive information, which relates to the privacy concern. These concerns become obstacles to voluntary contribution. Impure altruism theory proposes that a contributor receives not only utility from provision of the public good, but also a rather selfish, private benefit or warm glow, such as moral satisfaction and joy of giving (Steinberg 1987; Andreoni 1989). Many studies support the idea that contributors to public goods receive private benefits as a result of "social effects" through demonstration from empirical observations (Andreoni 2007).

As IS researchers, we look into factors and processes that intervene between information technology investments and the realization of the economic value. The objective of this study is to deepen our understanding of the factors that increase or diminish mobile devices users' tendencies to engage in voluntarily contributing behaviors. Since the voluntarily contributing behaviors are likely to be influenced not only by personal motivations, but also by contextual forces. We apply a theoretical frame in which the extrinsic forces and social-psychological forces are integrated with the theory of planned behavior (TPB).

The remainder of the paper proceeds as follows. Section 2 builds the related works for the proposed research model. In Section 3, we present the background for a preliminary study that tested this model in the context of one-day trial usage of the prototype intelligent traffic mobile application. The results of an empirical test are presented in Section 4. Section 5 is a discussion of the future work.

2 Theoretical Development

2.1 Theoretical Framing on TPB

To develop an integrative view of the forces influencing individuals' willingness to share real-time location-based traffic information, we adopted TPB (Ajzen 1991) as an initial theoretical frame. Hence, an individual user's decision to engage in a specified behavior is determined by their intention to perform the behavior, which in turn is determined jointly be their attitude toward (reflecting their salient behavioral beliefs), the subjective norms (regarding their normative beliefs and motivation to comply with these beliefs), and the perceived control (reflecting their beliefs about the presence of factors that may facilitate or impede performance of the behavior). Since TPB can be applied to virtually any behavior, the nature of the beliefs operative for a particular behavior are left unspecified. In a mature field of study where the beliefs that underlie a focal behavior are well specified, prior literature is usually a sufficient source for identifying the relevant beliefs (as well as their motivational drivers). However, in our study, the existing understanding of the factors that shape individuals' intentions to engage in knowledge sharing is anything but mature. Consequently, we interviewed executives leading the voluntary contribution design initiatives and a few first-time users to validate and supplement, if necessary, the motivational drivers identified from the existing literature. While viewed as a subset of e-commerce (Kwon & Sadeh 2004), mobile commerce/technology has its own characteristics (Clarke III 2001). On the other hand, users encounter with special constraints: mobile data quota, mobile device RAM and battery capacity. Our identification hinges on the booming state of free mobile applications, especially those which rely partly or entirely on free user contributions.

2.2 Studies into the Free Contributions

As voluntary contribution does not come without participant costs, personal beliefs that expected benefits will outweigh these costs are likely to be an important determinant of knowledge sharing behaviors. The concept of free contributor is an extension to the so-called "free-rider" hypothesis in resource allocation literature for the production of public goods. Briefly, the economic premise is the private provision of public goods. The free-rider hypothesis is that as group size grows, individual contribution levels will decline (Isaac and Walker 1988). At the very start, researchers built theoretical models from pure altruism, showing supports to the hypothesis that contributors receive utility while providing public goods, as well as private consumptions (Andreoni 1988; Palfrey and Howard 1984). However, empirical observations showed that as the participant number grows, the average contribution level falls down, leaving only those individuals with the lowest costs of contributing or the highest rewards will keep contributing. More studies came out to build models on impure altruism (Steinberg 1987). They indicated that contributors receive not only utility, but also social effects, metaphorized as warm glow (Zhang and Zhu 2011), which represent prestige, reciprocity among peers, social image, and social welfare (Harbaugh 1998; Ellingsen and Johannesson 2008). It is noted that when participant number grows to some extent, giving rise to free-riding behavior, and private benefit dominate the motivation to contribute. Ultimately, individual contribution levels could increase after the crossover point of the particular participant number. Our work will contribute to consider the mobile application scenario in a relatively complete perspective.

3 Methods

Since travel behavior is complex, a deep understanding of users' perceptions, attitudes and behaviors is needed. Qualitative methods are powerful tools to let people express what is really important to them in their own words. Most of the research conducted on mode choices employs researcher-selected variables, which tends to limit to a few attributes. Regarding this, qualitative methods can be a valuable way to allow respondents to specify which factors are important to them (Clifon and Handy 2001). On the other hand, quantitative approaches would be used to measure the reactions of subjects to a set of questions allowing the comparison and statistical aggregation of the data.

A triangulation of in-depth interviews and focus groups was believed appropriate for understanding the underlying reasons and specific circumstances. In the first phase, we conducted a pilot study, including three face-to-face in-depth interviews and one focus group study in Hong Kong. In the second phase, based on the results from pilot study, we will modify and test the model. Laboratory experiments and empirical user data collection will be utilized then. By writing this paper, we have just finished the pilot study.

4 Pilot Study

Three face-to-face in-depth interviews (average 30 minutes) and one focus group (40 minutes) were held in Hong Kong, involving programmers, a project manager and a few users. A prototype traffic information system that relies on mobile devices for collecting, processing, and disseminating real-time traffic information has been developed, by the Intelligent Transportation System R&D group in City University of Hong Kong, and the implementations are substantial supported by the Hong Kong Transport Department and Innovation & Technology Commission. As shown is Figure 1, for mobile network operators, location-based services represent utilizing their fixed infrastructures or even an additional stream of revenue that can be generated by extra data transfer or peripheral revenue like advertisement. For the end user, the application enables mobile phone users with the functions:

- ✓ *Wireless communication technology enabled mobile to vehicle (M2V) and mobile to mobile (M2M) traffic information exchanges;*
- ✓ *Identification of mobile user motion status;*
- ✓ *Identification and prediction of traffic patterns;*
- ✓ *Location-based services, e.g. travelling guide, people search, shopping mall search, etc.*

At the start of the focus group, a broad scope was adopted, for identifying the varying contexts of travel information use.

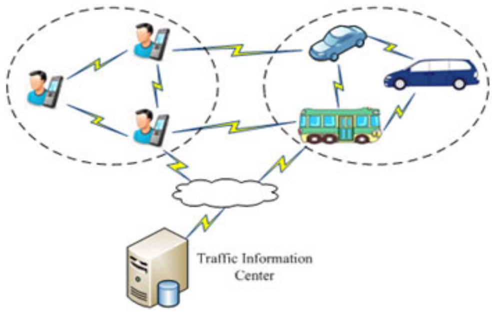


Fig. 1. Integration intelligent Transportation Systems

The target users of this system include private car owners, travelers who need real-time information with smart mobile phones. It offers us an ideal empirical setting to study the users' behavior on mobile applications. Firstly, we could get to know the developers of this system in understanding the initial intention and design solutions. Second, the center could keep the tracking and operating history of our experiment participants and each action can be traced to an ID that uniquely identified the contributor. As a result, we can accurately measure the contributions from each user over time.

4.1 Description of Participants

Different age groups were studied in the focus group. The two programmers are mainly in charge of the design of the prototype application for several platforms (e.g., Windows mobile, Android and iPhone). The project manager works in a Hong Kong local IT company, who is also a senior software engineer and experienced for more than seven years in both development and understanding clients' requirements. They were invited to our semi-structured interviews one by one. In the focus group, six first-time users are recruited via university e-mail lists of students and staff. They have never used this kind of traffic information mobile application before. Each of them was given a smart phone deployed with the indicated application. Among them, two were drivers with private cars and four were public transportation passengers.

The two programmers were interviewed firstly; because we hope that the project manager could tell us his understanding based on his experiences communicate with both programmers and clients. All interviews and focus group activities were tape recorded and transcribed.

4.2 Results

Level of Complexity by Using the Real-Time Traffic Information Application
The level of complexity did not seem to vary much among the users studied. Most indicated that the application was easy to get started and operate. It seems people are open to accept a new mobile application while it has the potential to improve their

knowledge. This finding might be not sufficient to support a solid conclusion, so more participants should be investigated in the next phase. Frankly, there is inconsistence between the users' perceived complexity and the developers' expectation. When asked about the potential barriers for users to operate and contribute the real-time traffic information, the programmers answered in one voice: do not exist. From their point of view, the application is

Simple, easy, and useful

They could tell a lot, for instance: the application is Google Map style, users do not need to learn how to use; wireless communication technology enabled mobile to vehicle and mobile to mobile traffic information exchanges; the contribute function is so easy that they need to do no more than start it while driving or travelling. It indicated that programmers are very confident with their work. However, the project manager said when we mentioned the programmers' answers:

They (programmers) believe that their program is perfect!!

Information needs vary for different users with different career, age, and education. The project manager explained that software and mobile applications should be designed to be intelligent, innovative, of great functions or even expertise ability. But the clients are probably not aware of it. 95% above users started with the easiest function. Half of them throw it over after a period of time. Only few users go deeper into the core function of a well-organized system. It takes even shorter time for clients/users to get bored when they use or play a mobile application. In the focus group, one of the respondents also indicated that

I felt it not always smooth when I use my fingers to move the map. (female, 38)

There might be reasons to explain the control problem. However, it illustrates that there are existing inconsistency between programmers' intuition and users' reaction, corresponding to the theory of structural complexity and users' perceived complexity (Nadkarni and Gupta 2007). The complexity of operations has a good chance to be underestimated by the programmers, while users may feel difficult to get start, nor making full use of it.

Level of Participation into the Free Contribution Function

After one-day trial of the mobile application, six first-time users reflected to be of great interests to continue use and expected to see the real-time traffic information around them during the trip. But the level of contribution function was relatively low. In focus group, while discussing considerations about contribution, respondents showed different starting points in their cognitive process of choice. They would

1. Consider battery remaining (inadequate leading them to not contribute);
2. Consider mobile data quota (inadequate also leading them to not contribute);
3. Have a slight preference for either contribute or not (other appealing functions might leading them to contribute);
4. Consider incentives (leading them to contribute); or
5. Consider reveal of their location (therefore would not contribute).

One driver pointed out that he forgot to start the function of contributing real-time traffic information at the very beginning. For the reason of limited participants in the pilot study, there were few users in the network using the application. Consequently, the prediction function was no guarantee accurate based on the historical data collected by traditional approaches. It requires large number of participants and their continuous contributions. The project manager also pointed out that they always devise some easy and attractive functions to let clients keep using, so that there will be more time for the core functions to be exposed. Besides, they sometimes divide users into several groups to customize the products. Meanwhile, the users reported some feelings while using.

> *The GPS is power consuming. In the afternoon, because my phone left less than 1/3 power, I decide to close the function of contribution.* (male, 49)

The app seems okay…If my friends use it, I will continue. (female, 23)
Summarized from their group talking, we found that: 1) users with different backgrounds might perceive the application to be not easy to control. The so-called "simple, easy, free" experiences imagined by programmers did not cover all the users, especially those with various backgrounds. 2) The special constraints of mobile phones, such as mobile data usage quota and battery, may influence the function use of an application. So its potency should be tested and considered in the design section. 3) Social network is truly influencing, especially for the young users, and car owner associations (such as BMW association, etc.). This could be utilized to enlarge the crowd for real-time traffic collection.

5 Limitations and Future Work

Limitations exist in the pilot study. First, it was held within very few people. As this is a pilot study, we aimed to get senses and understandings of the users and developers attitude and behavior to guide our further study. Second, the participants were selected in a relatively small scope. But the reflections may be instructive. The users' trial investigation could be extent to a larger number for the next step research work. In the second phase of this study, laboratory experiments and empirical user data collection will be utilized to examine the model, in order to reflect common experience of using and preference to freely contribute.

References

1. Ajzen, J.: The theory of planned behavior. Organizational Behavior & Human Decision Processes 50(2), 179–211 (1991)
2. Andreoni, J.: Why Free Ride? Strategies and Learning in Public Goods Experiments. Journal of Public Economics 37(3), 291–304 (1988)
3. Chorus, G., Arentze, T., Molin, E., Timmermans, H., Wee, B.: The Value of Travel Information: Decision Strategy-Specific Conceptualizations and Numerical Examples. Transportation Research B, 504–519 (2006)

4. Clarke, I.: Emerging value propositions for m-commerce. Journal of Business Strategies 18(2), 133–148 (2001)
5. Clifton, K.J., Handy, S.L.: Qualitative methods in travel behavior research. In: The International Conference on Transport Survey Quality and Innovation, South Africa (August 2001)
6. Ellingsen, T.l., Magnus, J.: Pride and Prejudice: The Human Side of Incentive Theory. American Economic Review 98(3), 990–1008 (2008)
7. Harbaugh, W.T.: The Prestige Motive for Making Charitable Transfers. American Economic Review 88(2), 277–282 (1998)
8. Huang, M.H.: Designing website attributes to induce experiential encounters. Computers in Human Behavior 19, 425–442 (2003)
9. Isaac, R.M., James, M.W.: Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism. Quarterly Journal of Economics 103(1), 179–199 (1988)
10. Isaac, R.M., James, M.W., Arlington, W.W.: Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups. Journal of Public Economics 54(1), 1–36 (1994)
11. Kwon, O.B., Sadeh, N.: Applying case-based reasoning and multi-agent intelligent system to context-aware comparative shopping. Decision Support Systems 37(2), 199–213 (2004)
12. Lakhani, K.R., Hippel, E.: How Open Source Software Works: 'Free' User-to-User Assistance. Research Policy 32(6), 923–943 (2003)
13. Loomes, G., Sugden, R.: A Rationale for Preference Reversal. The American Economic Review 73, 428–432 (1983)
14. Miller, D.W., Star, M.K.: The Structure of Human Decisions. Prentice Hall, Englewood Cliffs (1967); 92 Transportation Research Record 2069
15. Nadkarni, S., Gupta, R.: A task-based model of perceived website complexity. MIS Quarterly 31, 501–524 (2007)
16. Palfrey, T.R., Howard, R.: Participation and the Provision of Discrete Public Goods: A Strategic Analysis. Journal of Public Economics 24(2), 171–193 (1984)
17. Sendy, F., Lyons, G.: What affects use of pre-trip public transport information Empirical Results of a qualitative study. Transportation Research Record: Journal of the Transportation Research Board, 85–92 (2008)
18. Steinberg, R.S.: Voluntary Donations and Public Expenditures in a Federal System. American Economic Review 77(1), 24–36 (1987)
19. Sunitiyoso, Y., Avineri, E., Chatterjee, K.: The Role of Minority Influence on the Diffusion of Compliance with a Demand Management Measure. In: 11th International Conference on Travel Behaviour Research (2006)
20. Tanaka, Y., Morita, S., Suzuki, K.: Improvement of Microscopic Traffic Simulation with Driver's Parameter Variation using Travel Time Estimation. In: ITS World Congress 2010, Seoul (2010)
21. Todd, P.M.: How Much Information Do We Need? European Journal of Operational Research 177, 1317–1332 (2007)
22. Xu, T., Sun, L., Hao, Y.: Empirical Analysis and Modeling of Drivers' Response to Variable Message Signs in Shanghai. In: ITS World Congress 2010, Seoul (2010)
23. Zhang, X., Zhu, F.: Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia. American Economic Review 101(4), 1603–1620 (2011)

The Effects of Application Discoverability on User Benefits in Mobile Application Stores

Jaeki Song, Junghwan Kim, and Donald R. Jones

ISQS Area, Rawls College of Business Administration, Texas Tech University,
Box 42101 Lubbock, Texas 79409-2101, United States
{jaeki.song,jung.kim,donald.jones}@ttu.edu

Abstract. This document is in the required format. Mobile applications and mobile application stores are becoming people's commodities in everyday life, offering unprecedented mobile services. In mobile application stores with numerous applications finding the right applications is painstaking for users. Therefore, this study aims to explicate the effect of application discoverability on user benefits in mobile application stores by identifying the relationships of need specificity, application discoverability, and application quantity. Using a survey methodology, we found that app users' need specificity has an impact on application discoverability and quantity-sufficiency of applications, but not quantity-overload of applications. Our findings also show that application discoverability plays a substantial role in enriching users' utilitarian and hedonic benefits in mobile application stores.

Keywords: We would like to encourage you to list your keywords in this section.

1 Introduction

Until recently, people's use of mobile service has been limited to traditional communication services such as voice services and text messaging [1, 2]. New kinds of mobile devices, which are called Smartphones (such as iPhone), are fundamentally changing the way people use mobile services [3]. Such mobile devices provide people with various services beyond routine calls and text messages. In particular, a mobile application (mobile app), which refers to a software application that runs in Smartphones or other portable devices, plays an integral role in new kinds of mobile services. With the adoption of mobile apps for hand-held devices, mobile application stores, which allow people to browse and download mobile apps for user on the mobile device, are also becoming widely used in everyday life. Thus, the popularity of mobile apps and mobile app stores indicates that mobile apps provide value beyond that provided by the mobile device itself.

For example, the Apple's App Store is the most successful and best-known app store, having over 350,000 applications and delivering 10 billion downloads in

M.J. Shaw, D. Zhang, and W.T. Yue (Eds.): WEB 2011, LNBIP 108, pp. 429–441, 2012.

January 2011. Steve Jobs, Chief Executive Officer in Apple Inc., has said that about 15,000 new applications are submitted to the App Store each week. In addition, demand for mobile apps is expected to continue growing [4], with the number of global downloads anticipated to reach 76.9 billion in 2014, with a value of $35 billion [5]. Thus, increasing numbers of apps serve to attract new Smartphone users to the mobile app store while the expanding user pool motivates app developers to offer higher quality apps at lower prices in hopes of tapping this growing market [6]. However, increasing numbers of apps make it difficult for users to find the app they need. As the number of apps drastically increases, the need to sift through a high number of apps can cause app users to spend more time and effort for finding the app they want [7]. As a result, finding the right apps to download can be painstaking for app users. Thus, app discoverability is becoming a critical issue for app users in various mobile platforms.

With regard to the mobile service phenomenon, there are various theoretical approaches including IS theories (e.g., technology acceptance model (TAM), theory of reasoned action (TRA), theory of planned behavior (TPB)) and social psychological theories (e.g., uses and gratification theories, domestication research) [1, 8]. Such theoretical approaches have investigated user benefits in mobile service, including both utilitarian and hedonic perspectives. Particularly in the IS field, many studies have been concerned with utilitarian benefits in which the main purposes for using the systems are effectiveness, efficiency, and utility [1]. In addition, a hedonic perspective reflecting deeper basic human needs and values, has been regarded as a salient benefit in the mobile service phenomenon, in that user needs in the mobile app store are much broader than information needs for task performance in organizations [9]. Therefore, there is a critical need to understand the substantial role of app discoverability on mobile app users' benefits.

In summary, the objective of this study is twofold. First, we identify the relationships of app users' need specificity, app discoverability and app quantity in mobile app stores. Emphasizing the role of need specificity in manipulating the effects of app discoverability and app quantity, we draw on the concept of goal-setting theory, which explains that individuals' cognition and behaviors vary depending on the level of their need specificity [10]. Second, we examine the effect of app discoverability on user benefits in mobile app stores. In detail, we address that mobile app users are more likely to gain utilitarian and hedonic benefits when mobile app stores are more discoverable.

Next, we first describe the background for this research, and the present our research model and hypotheses. In the section that follows we discuss our research methodology and results, and then finally conclude with theoretical and practical implications and limitations of this research.

2 Background

2.1 The Conflicting Goals of Mobile Application Stores

A mobile application (mobile app) is a service for users of mobile devices such as iPhones, BlackBerrys, and Android phones, allowing users to browse and download software applications for use on their mobile devices. Offering new kinds of mobile service, mobile app stores strive to achieve two seemingly-conflicting goals for success: (1) providing a large number of apps that will appeal to the largest possible number of consumers and (2) making the apps that users are searching for easy to find within the store. For instance, the higher number of distinct apps the Apple's App Store accumulates, the greater is the likelihood that the App Store will be attractive to each of the 40 million iPhone, iPad, and iPod Touch users. However, with so many apps to sort through, many app users have difficulty finding the right apps they need in mobile app stores. Therefore, effectively managing this challenging dilemma of app discoverability and app quantity can be critical in sustaining the app stores' competitive advantage in the mobile application market [6].

2.2 Application Discoverability and Application Quantity

The seemingly-conflicting goals of mobile app stores can be explained by the relationships between app discoverability and app quantity. First of all, we define application discoverability as the level of ease of searching for and accessing applications in the mobile application store. App discoverability may have been explained by prior research regarding navigability, search mechanism, website design, etc. [11-13]. Moreover, app discoverability includes the various ways in which an app store facilitates users' search for desired apps, beyond design aspects including graphics, layout, and actual content. Therefore, we address that app discoverability plays an important role in sustaining the success of mobile app store.

Second, perceived quantity of applications refers to the application users' perception of application quantity in mobile app stores [14-16]. Regarding information quantity as a metaphor of this construct, some prior research has addressed that perceived quantity of apps refers to a unitary construct, representing the app users' perception of app quantity in mobile app stores [14-16]. Similarly, some prior studies to date have viewed perceived quantity of information as coming from a continuum, treating it as a single dimension [15, 17]. However, we surmise that attitudinal impacts of the app quantity are not so straightforward because people may have different perception of the same phenomenon. For example, some people may perceive the quantity of apps in an app store to be adequate, and this perception of adequate quantity would not be reduced by the addition of more apps to the store. The same person, whoever, may (or not) be distracted by what they perceive to be an inordinately high number of apps. Indeed, these two different perceptions (e.g., quantity-sufficiency and quantity-overload) of the app quantity can also coexist in mobile app stores. Hence, we regard the perceived quantity of applications as having two different dimensions: quantity-sufficiency and quantity-overload. First, quantity-sufficiency of applications refers to the extent to which the amount of apps provided by a mobile app store is perceived as sufficient to enable or support the finding of a desired app without additional cognitive effort [18]. Second, quantity-overload of applications refers to the extent to which a large number of apps in an app store is chaotic or overwhelming to consumers, making it confusing to find the appropriate apps [15, 19].

2.3 User Benefits in Mobile Application Stores: Utilitarian and Hedonic Benefits

The advent of mobile app store has triggered some changes in everyday life. These changes impact people's daily routines and behaviors, because they tap into a deep-rooted desire to lead more optimized and productive lives. For example, Gravitytank's research [20] reveals that app and app-enabled mobile devices have become digital Swiss Army knives for modern living, providing consumers with near-instant access to the information and services they need.

According to Hoffman and Novak's (1996), mobile services have several different characteristics. For example, text messaging services and contact services represent person-to-person interactive services, whereas gaming services and payment services are machine interactive. Furthermore, text messaging and payment can be characterized as goal-directed services, whereas gaming services and contact services are more experiential. Because mobile app stores provide a variety of services from productivity or business to games or entertainment, app users are motivated from different service categories. In line with this, users' benefits in using mobile services could be studied across service categories [1], and thereby, various theoretical perspectives may provide a richer understanding of the mobile services phenomenon [8].

According to Batra and Ahtola [21], people use goods and services for two basic reasons: (1) consummatory affective gratification and (2) instrumental, utilitarian reasons. The former is called the hedonic perspective, and it results from sensations derived from the experience of using products and services. Hedonic benefits refer to users' aesthetic, experiential, and enjoyment-related benefits [21-24]. On the other hand, the latter results from some type of conscious pursuit of an intended consequence, and it can be task-related, rational, and may be thought of as work [25]. Thus, utilitarian benefits refer to the functional, instrumental, and practical benefits offered by using the services.

3 Research Model and Hypotheses

Figure 1 shows the proposed research model, including the major theoretical constructs and their hypothesized relationships. The posited constructs and their relationships shown in Figure 1 have been identified because of their theoretical relevance and managerial importance, as subsequently described.

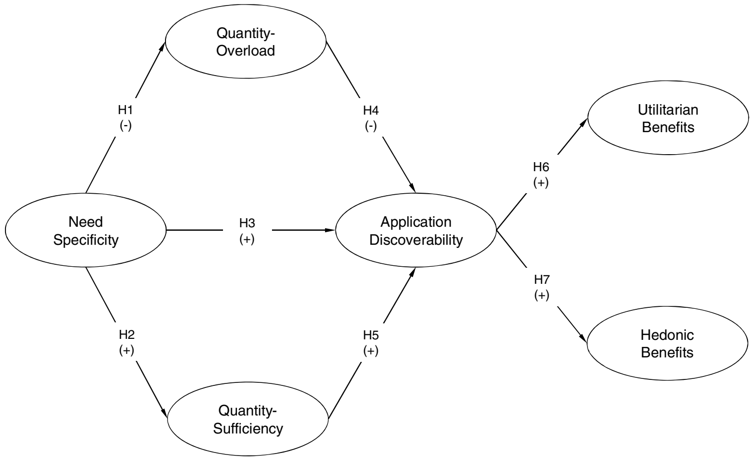


Fig. 1. Research Model

3.1 The Effects of Need Specificity on Application Discoverability and Quantity

According to Locke and Latham [10], much of human action is purposeful and is directed by conscious goals. Therefore, individuals' motivation and subsequent performance increase when they set specific, difficult goals with high valence. In addition, individuals react or change their behaviors when a discrepancy is perceived between their goal and the current state of the environment. Based on goal-setting theory [10, 26], we consider the specificity of the app user's needs as a salient individual factor in mobile app stores.

App user's *need specificity* is defined as the extent to how well individual users know what they want when they visit a mobile app store: In other words, the extent to which the information about a sought app is specific and clear [27, 28]. Users' need is very specific when they know the exact title of app, whereas their need is vaguer when they don't know the exact title of the app satisfying them. For example, if you know the app "Angry Birds", you can easily find that app in the mobile app store. However, when you only know that you seek some game or puzzle that involves destruction and something about eggs and pigs, you may spend more time and effort to search and get the app in the store.

According to Kourfaris, Kambil, and Labarbera [28], users' need specificity influences how to search and access the information in the website (e.g., apps in mobile app stores). More specifically, users with high need specificity will easily find the information they want without any help of search mechanism, whereas users with low need specificity will be more likely to get some help from search mechanism to better define their needs [28]. In addition, users may perceive the app quantity to be sufficient when they know exactly what they want, even though there are so many apps in mobile app stores. However, if users are unsure of what they are looking for in mobile app stores, huge numbers of apps likely cause them to feel overwhelmed in the store. Therefore, we posit that the dynamics of app discoverability and app quantity can be determined by the degree of app users' need specificity.

Hypothesis 1: The users' need specificity will be negatively related to their perceived quantity-overload of applications in a mobile app store.

Hypothesis 2: The users' need specificity will be positively related to their perceived quantity-sufficiency of applications in a mobile app store.

Hypothesis 3: The users' need specificity will be positively related to application discoverability in a mobile app store.

3.2 The Effects of Application Quantity on Application Discoverability

According to O'Reilly [15], decision makers' tendencies are to seek information up to the point of possessing a sufficient amount of information to make a decision. When they perceive that the quantity of information is sufficient, they make higher-quality decisions with less cognitive effort and time. Thus, in the quantity-sufficient situation, app store users are more likely to perceive higher app discoverability.

At the same time, an increase in the quantity of distinct apps makes it harder for users to find the app they seek. Although most mobile app stores provide guides for searching and accessing the apps users want (e.g., categorized features, top-rated applications, etc.), it is not always easy to find the appropriate app [29]. Having huge numbers of apps has the potential to reduce the user benefits of an app store by imposing additional processing costs on users of the mobile app store [30, 31]. As a result, all else equal, increasing the number of apps beyond users' cognitive capacity contributes to reducing the discoverability of the apps. Thus, increases in app quantity may necessitate a higher level of discoverability, or additional functionality to enable discoverability, in order to reduce cognitive overload, and as a result, quantity-overload is likely to impede users' ability to identify the apps they desire [14, 15].

Hypothesis 4: In a mobile application store, the users' perceived quantity-overload of applications will be negatively related to application discoverability.

Hypothesis 5: In a mobile application store, the users' perceived quantity-sufficiency of applications will be positively related to application discoverability.

3.3 The Effects of Application Discoverability on User Benefits

In mobile app stores, some mobile apps are used for goal-directed purposes, and others are used for experiential purposes. Specifically, goal-directed processes are characterized by instrumental orientation with a focus on utilitarian benefits [1, 32]. Utilitarian benefits emphasize task-related and rational benefits with a conscious goal pursuit. In mobile app stores, 50% of users view apps as essential tools for getting things done and staying organized, considering apps as an indispensable tool to manage information, tasks, work, and relationships in their lives [20].

On the other hand, hedonic benefit concerns users' gratifications in media and technology use, based on their individual "needs" or "motivations" [33]. Complementing the utilitarian perspective, hedonic benefits, as non-utilitarian gratifications, are related to (1) enjoyment, fun seeking, and entertainment and (2) fashion, status, and sociability [1]. In addition, as the salient motivator of mobile app users, the hedonic benefits have revealed to be critical antecedents of attitude toward using technology-based services [34] and willingness to recommend services [35]. In the sense, 63% of mobile app store users think that mobile apps keep them entertained, enhancing their hedonic benefits [20].

In terms of user benefits such as utilitarian and hedonic benefits, mobile app stores that are more difficult to find the right apps will demand more mental resources, raising the search cost which most users strive to minimize [14, 36]. The more discoverable the app store is, the more the perceived utilitarian (e.g., informativeness) and hedonic (e.g., entertainment) benefits are [14]. Therefore, we posit that app discoverability enhances utilitarian and hedonic benefits which app users perceive in mobile app store.

Hypothesis 6: In a mobile application store, application discoverability will be <u>positively</u> related to the users' utilitarian benefits.

Hypothesis 7: In a mobile application store, application discoverability will be <u>positively</u> related to the users' hedonic benefits.

4 Methodology

4.1 Measures and Participants

For the survey instrument, we identified existing literature where they had been repeatedly tested and strong content validity exhibited, and adapted them to our research domain. All items (except for demographic questions) used a 7-point Likert scale. Subjects consisted of 174 undergraduate and graduate students in a large university in South Korea. Subjects are currently using the Smartphone such as iPhone or Android Phone. They have experiences of downloading one or more applications from a mobile application store.

4.2 Measurement Model

Data analysis was performed using Partial Least Squares (PLS). Unlike covariance-based approaches, PLS requires minimal demands on measurement scales, sample size, and distributional assumptions [37]. In addition, PLS is well-suited to highly-complex predictive models [38]. We used Smart PLS version 2.0 for our analysis. Smart PLS is a software application for the design of structural equation models (SEM) on a graphical user interface (GUI). We conducted our analysis in two stages. First, we tested the measurement model to ensure that the constructs had sufficient psychometric validity and then we addressed the structural model in which the hypotheses were tested. Table 1 exhibits the results of confirmatory factor analysis.

Measurement reliability was assessed using internal consistency scores, calculated by the composite reliability scores [39]. Internal consistencies of all variables are considered acceptable since they exceed 0.70, signifying acceptable reliability [40]. As can be seen in Table 2, the composite reliability for all constructs is greater than 0.90. In addition, all items exhibit high loadings on their respective constructs. Thus, all constructs in the model exhibit good internal consistency.

Convergent and discriminant validity is supported when the PLS indicators (1) load much higher on their hypothesized factor than on other factors (own-loadings are higher than cross-loadings), and (2) when the square root of each construct's average variance extracted (AVE) is larger than its correlations with other constructs [37]. As shown by comparing inter-construct correlations and AVE in Table 2, all constructs share more variance with their indicators than with other constructs since all AVEs are well above 0.50 [41]. These statistics for the reliability of our measures and analysis are summarized in Table 2.

Table 1. Summary of Factor Analysis Results

Construct	Item	NS	QO	QS	AD	UT	HD
Need Specificity (NS)	NS1	**0.774**	0.005	0.246	0.385	0.367	0.149
	NS2	**0.754**	0.131	0.034	0.322	0.258	0.071
	NS3	**0.825**	0.051	0.089	0.355	0.293	0.065
	NS4	**0.774**	0.015	0.354	0.471	0.431	0.079
Quantity-Overload (QO)	QO1	0.048	**0.877**	0.336	-0.122	0.035	0.026
	QO2	0.077	**0.881**	0.224	-0.119	-0.010	0.021
	QO3	0.075	**0.934**	0.166	-0.183	-0.080	0.023
	QO4	0.008	**0.959**	0.144	-0.229	-0.130	-0.016
Quantity-Sufficiency (QS)	QS1	0.256	0.193	**0.952**	0.438	0.502	0.299
	QS2	0.278	0.142	**0.892**	0.394	0.480	0.303
	QS3	0.257	0.183	**0.951**	0.403	0.454	0.222
	QS4	0.203	0.319	**0.912**	0.347	0.425	0.162
Application Discoverability (AD)	AD1	0.498	-0.074	0.513	**0.877**	0.657	0.301
	AD2	0.410	-0.148	0.291	**0.873**	0.549	0.215
	AD3	0.475	-0.226	0.378	**0.923**	0.628	0.315
	AD4	0.433	-0.233	0.338	**0.921**	0.621	0.321
Utilitarian Benefits (UT)	UT1	0.447	-0.039	0.501	0.660	**0.926**	0.395
	UT2	0.459	0.033	0.422	0.542	**0.882**	0.305
	UT3	0.330	-0.127	0.461	0.650	**0.897**	0.471
	UT4	0.406	-0.091	0.431	0.615	**0.908**	0.457
Hedonic Benefits (HD)	HD1	0.097	0.035	0.259	0.306	0.433	**0.941**
	HD2	0.157	0.057	0.281	0.344	0.477	**0.926**
	HD3	0.094	-0.051	0.222	0.294	0.404	**0.939**
	HD4	0.065	-0.017	0.205	0.196	0.289	**0.804**

Table 2. Reliability Measures for Model Constructs and Construct Correlation

Construct	Cronbach's α	Composite Reliability	Inter-Construct Correlations*					
			NS	QO	QS	AD	UT	HD
Need Specificity (NS)	0.798	0.863	**0.612**					
Quantity-Overload (QO)	0.935	0.952	0.052	**0.834**				
Quantity-Sufficiency (QS)	0.945	0.961	0.270	0.220	**0.859**			
App Discoverability (AD)	0.921	0.944	0.508	-0.189	0.429	**0.808**		
Utilitarian Benefits (UT)	0.925	0.947	0.452	-0.066	0.504	0.686	**0.816**	
Hedonic Benefits (HD)	0.926	0.947	0.119	0.011	0.270	0.324	0.454	**0.818**

*The diagonal elements (in bold) represent the square root of the AVE.

4.3 The Structural Model

In a PLS structural model, loadings of measures of each construct can be interpreted as loadings in a principal components factor analysis. Paths are interpreted as standardized beta weights in a regression analysis. The PLS path coefficients are shown in Figure 2.

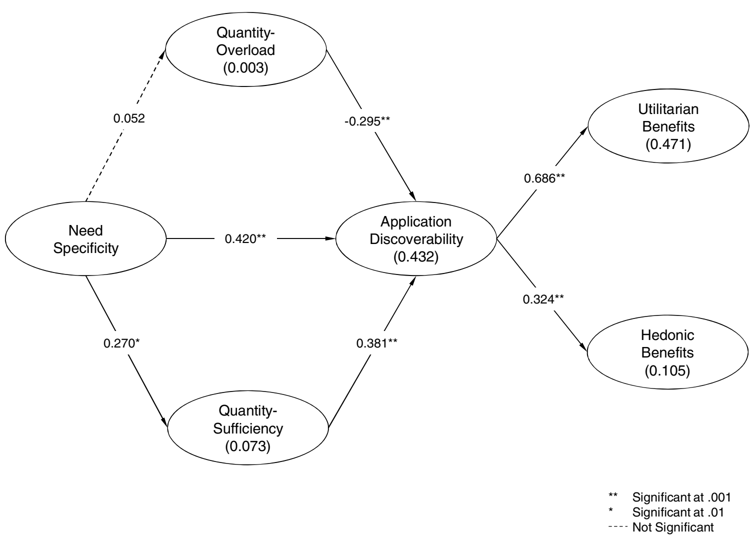


Fig. 2. The Estimated Model: PLS Results

First of all, in the relationship between need specificity and perceived app quantity, PLS results show that H2 is supported, but not H1. In other words, the users' need specificity positively influences perceived quantity-sufficiency of apps (b=0.270, p<0.01), whereas the relationships between need specificity and perceived quantity-overload of apps in this data is not significant (b=0.052, *n.s.*). Moreover, the need specificity positively affects app discoverability (b=0.420, p<0.001), as we predicted.

Second, PLS results provide support for H4 and H5, which assert that two dimensions of perceived quantity of apps have different impacts on app discoverability. Quantity-overload has a negative impact on app discoverability (b=-0.295, p<0.001), whereas quantity-sufficiency positively influences app discoverability (b=0.381, p<0.001). Moreover, high R^2 values show that this model can be used to predict the effect of perceived quantity of apps on app discoverability with the users' need specificity (R^2=0.432).

Third, PLS results address that app discoverability has a positive impact on both utilitarian (b=686, p<0.001) and hedonic (b=324, p<0.001) benefits, supporting H6 and H7. App discoverability accounts for 47.1% of the variance in users' utilitarian benefits, but only 10.5% of the variance in their hedonic benefits.

5 Discussion

The mobile service and mobile app store phenomena are becoming people's commodities in everyday life. Most people who have the new kinds of mobile devices such as iPhone and Android phone usually access to the mobile app store, and download and use the apps with their hedonic and utilitarian benefits. Therefore, we set out to investigate the effects of app discoverability on user benefits in mobile app stores, highlighting the relationships of app discoverability with need specificity and app quantity.

5.1 Theoretical Implications

From a theoretical perspective, our proposed model provides several insights into the effect of app discoverability on user benefits in mobile app stores. First, our results indicate that app users' need specificity significantly influence quantity-sufficiency of app and app discoverability, but not quantity-overload of app. In detail, need specificity had a positive and significant impact on quantity-sufficiency of app and app discoverability. However, quantity-overload of app is not significantly influenced by need specificity. Hence, if app users exactly know what they want in mobile app stores (that is, the need specificity is high), they are more likely to perceive that the number of apps may be appropriate, which in turn mobile app store is discoverable. In sum, app users' need specificity can be considered as an influential factors for explaining the relationships of app discoverability and app quantity in mobile app store.

Second, our findings showed that app discoverability played a considerable role in improving user benefits, both utilitarian and hedonic benefits. This finding is in consistent with the concept of cognitive-overload research, which explains that providing too much information creates the impediment to engaging in the environment with eliciting unpleasant emotions. In line with this, app discoverability can reduce app users' cognitive load in mobile app stores. Essentially, we argue that users' utilitarian and hedonic benefits can be increased when mobile app stores make it more discoverable while simultaneously maintaining or increasing the number of apps.

5.2 Practical Implications

From a practical perspective, the proposed model can help explain the importance of app discoverability in mobile app stores. App discoverability is critical in enriching user benefits such as utilitarian and hedonic benefits, because it enables mobile app users easily access and find the app they need with minimal cognitive efforts. Hence, in order to maximize user benefits, platform owners which run the mobile app stores need to improve app discoverability by utilizing such facilitators as application store coherence, in-marketplace search optimization, user-generated reviews, and reputations from outside or third-parties.

5.3 Limitations and Future Research

It is worthwhile to note some of the limitations of this study. First, we made use of perceptual measures for quantity of applications. The subjective perceptions of application quantity may be different, and as a result, we could not say how many applications are appropriate or are required for the success of the mobile application store. Moreover, it is difficult to set the threshold of differentiating quantity-overload from quantity-sufficiency of applications. Future research can be conducted by using objective measures for quantity of applications, perhaps in addition to the perceptual measures used here.

Second, some have criticized the use of student participants in this study, even though we believe that students are an appropriate population from which to draw a sample. Recent studies supported that most students using Smartphones and other similar mobile devices utilize the apps downloaded from app stores. In addition, college students make up a large part of mobile app store users, with young people ages 15-24 more immersed in mobile technology such as Smartphones and mobile app store than any previous generations [42]. Thus, the selection of our sample from this group can be appropriate. However, in spite of this rationale, we suggest that future research needs to include evidence from other demographic groups to better-represent the total population.

6 Conclusion

Since mobile apps and mobile app stores are permeating people's everyday life, understanding the mobile app store phenomenon is very critical to both theory and practice in a wide variety of domains. Most people who have the new kinds of mobile devices such as Smartphones usually access to the mobile app store, and download and use the apps with their hedonic and utilitarian purposes. Therefore, this study set out to investigate how hedonic and utilitarian benefits can be improved in mobile app store, emphasizing the integral role of app discoverability. Particularly, app discoverability enriches user benefits in mobile app stores, with the emphasis of relationships between app discoverability, app quantity, and need specificity. This study not only contributes to the literature by presenting that user benefits can be enriched when mobile app stores are more discoverable, but also motivates mobile platform owners to develop the facilitators for app discoverability.

References

1. Nysveen, H., Pedersen, P., Thorbjørnsen, H.: Intentions to use mobile services: antecedents and cross-service comparisons. Journal of the Academy of Marketing Science 33, 330–346 (2005)
2. Ling, R.: New tech, new ties: How mobile communication is reshaping social cohesion. The MIT Press (2008)
3. Zittrain, J.: The Future of the Internet, Allen Lane, London (2008)
4. MobiThinking, Global mobile statistics 2011: all quality mobile marketing research, mobile Web stats, subscribers, ad revenue, usage, trends.., in (2011)
5. IDC, IDC Forecasts Worldwide Mobile Applications Revenues to Experience More Than 60% Compound Annual Growth Through 2014, in (2010)
6. Chen, B.X.: For the iPhone's App Store, Quantity Really Does Matter. Wired (2009)
7. Shapiro, A.: The Great App Bubble. Fastcompany.com (2010)
8. Konana, P., Balasubramanian, S.: The social–economic–psychological model of technology adoption and usage: an application to online investing. Decision Support Systems 39, 505–524 (2005)
9. Kim, J., Song, J., Jones, D.R., Lin, Z.: The influences of hedonic and utilitarian benefits in mobile app stores. In: Big XII+ MIS Research Symposium, Manhattan, KS (2011)
10. Locke, E.A., Latham, G.P.: Goal setting theory. In: O'Neil Jr., H.F., Drillings, M. (eds.) Motivation: Theory and Research, pp. 13–29. Lawrence Erlbaum Hillsdale, NJ (1994)
11. Cyr, D.: Modeling Web site design across cultures: Relationships to trust, satisfaction, and e-loyalty. Journal of Management Information Systems 24, 47–72 (2008)
12. Palmer, J.: Web site usability, design, and performance metrics. Information Systems Research 13, 151–167 (2002)
13. Song, J., Zahedi, F.: A theoretical approach to web design in e-commerce: a belief reinforcement model. Management Science 51, 1219–1235 (2005)
14. Kang, Y.S., Kim, Y.J.: Do visitors' interest level and perceived quantity of web page content matter in shaping the attitude toward a web site? Decision Support Systems 42, 1187–1202 (2006)
15. O'Reilly, C.A.: Individuals and information overload in organizations: Is more necessarily better? Academy of Management Journal 23, 684–696 (1980)
16. Grisé, M.L., Gallupe, R.B.: Information overload: addressing the productivity paradox in face-to-face electronic meetings. Journal of Management Information Systems 16, 157–185 (1999)
17. Gu, B., Konana, P., Rajagopalan, B., Chen, H.: Competition among virtual communities and user valuation: The case of investing-related communities. Information Systems Research 18, 68–85 (2007)
18. Gibson, J.J.: The ecological approach to perception. Haughton Mifflin, Boston (1979)
19. Kaplan, S., Kaplan, R.: Cognition and environment, Praeger (1982)
20. Gravitytank, Apps get real: Perspective on the phenomenon, Chicago, IL (2009)
21. Batra, R., Ahtola, O.: Measuring the hedonic and utilitarian sources of consumer attitudes. Marketing Letters 2, 159–170 (1990)
22. Chitturi, R., Raghunathan, R., Mahajan, V.: Form versus function: How the intensities of specific emotions evoked in functional versus hedonic trade-offs mediate product preferences. Journal of Marketing Research 44, 702–714 (2007)
23. Chitturi, R., Raghunathan, R., Mahajan, V.: Delight by design: The role of hedonic versus utilitarian benefits. Journal of Marketing 72, 48–63 (2008)

24. Dhar, R., Wertenbroch, K.: Consumer choice between hedonic and utilitarian goods. Journal of Marketing Research 37, 60–71 (2000)
25. Bridges, E., Florsheim, R.: Hedonic and utilitarian shopping goals: The online experience. Journal of Business Research 61, 309–314 (2008)
26. Pinder, C.C.: Work motivation: Theory, issues, and applications, Scott, Glenville, IL (1984)
27. Winters, D., Latham, G.P.: The effect of learning versus outcome goals on a simple versus a complex task. Group & Organization Management 21, 236–250 (1996)
28. Koufaris, M., Kambil, A., LaBarbera, P.A.: Consumer behavior in web-based commerce: an empirical study. International Journal of Electronic Commerce 6, 115–138 (2001)
29. Chen, B.X.: Mac App Store Provokes Developer Interest, Concern. Wired (2010)
30. Butler, B.S.: Membership size, communication activity, and sustainability: A resource-based model of online social structures. Information Systems Research 12, 346–362 (2001)
31. Moreland, R.L., Argote, L., Krishnan, R.: Socially shared cognition at work: Transactive memory and group performance. In: Nye, J.L., Brower, A.M. (eds.) What's Social about Social Cognition? Research on Socially Shared Cognition in Small Groups, pp. 57–84. Sage, Thousand Oaks (1996)
32. Hoffman, D., Novak, T.: Marketing in hypermedia computer-mediated environments: conceptual foundations. Journal of Marketing 60, 50–68 (1996)
33. Lin, C.: Standpoint: Looking back: The contribution of Blumler and Katz's uses of mass communication to communication research. Journal of Broadcasting & Electronic Media 40, 574–581 (1996)
34. Dabholkar, P., Bagozzi, R.: An attitudinal model of technology-based self-service: moderating effects of consumer traits and situational factors. Journal of the Academy of Marketing Science 30, 184–201 (2002)
35. Johnson, M., Zinkhan, G., Ayala, G.: The impact of outcome, competency and affect on service referral. Journal of Services Marketing 12, 397–415 (1998)
36. Hoque, A.Y., Lohse, G.L.: An information search cost perspective for designing interfaces for electronic commerce. Journal of Marketing Research 36, 387–394 (1999)
37. Chin, W.W.: The partial least squares approach to structural modeling. In: Marcoulides, G.A. (ed.) Modern Methods for Business Research, pp. 295–336. Erlbau Associates, London (1998)
38. Barclay, D., Higgins, C., Thompson, R.: The Partial Least Squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. Technology Studies 2, 285–309 (1995)
39. Werts, C.E., Linn, R.L., Jöreskog, K.G.: Intraclass reliability estimates: Testing structural assumptions. Educational and Psychological Measurement 34, 25 (1974)
40. Nunally, J.: Pshycometric theory, 2nd edn. McGraw-Hill, New York (1978)
41. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 18, 39–50 (1981)
42. TheNielsonCompany, Mobile Youth Around the World, The Nielson Company (2010)

Author Index