SIGBPS Workshop on Business Processes and Services (BPS'13)

December 15, 2013
Milan, Italy

Table of Contents

Multi-agent Based Cooperative Framework for Managing Information Systems Security Risk

Nan Feng*,Minqiang Li, Fuzan Chen, Jin Tian
College of Management and Economics, Tianjin University, Tianjin 300072, P. R. China
*Corresponding author. Tel.: +86-22-27404796; fax: +86-22-27404796.
E-mail address: fengnan@tju.edu.cn

Abstract: Given the increasing collaboration between organizations, the information sharing across the allied organizations is critical to effectively manage information systems security (ISS) risk. Nevertheless, few previous studies on ISS take the issue of information sharing into account. In this paper, we develop a multi-agent based cooperative framework (MACF) to assess the risk level of each allied organization's IS and support the decision making of proactive security treatment in a distributed environment. In MACF, each analysis agent corresponding to an organization's IS encapsulates a Bayesian network (BN) supporting the flexible information sharing with other analysis agents. Moreover, for an organization's IS, the encapsulated BN is utilized to model its security environment and dynamically predict its security risk level, by which the control agent can choose an optimal action among alternatives to protect the information resources proactively.

Keywords: information systems security; information sharing; multi-agent; Bayesian networks

1. Introduction

Organizations' heavy reliance on information systems (IS)requires them to manage the security issues associated with those systems. Nowadays, risks related to information systems security (ISS) are a major challenge for many organizations, since these risks mayhave dire consequences, including corporate liability, loss of credibility, and monetary damage (Cavusoglu, Cavusoglu, and Raghunathan 2004). Ensuring ISS has become one of the top managerial priorities in many organizations.

The management of ISSrisk is distributed across the allied organizations and requires a great deal of collaborative activity (Bulgurcu et al. 2010).For example, to detect the risk of network-based distributed attacks, cooperation among the allied enterprises is imperative in a supply chain environment. Thus, the development of models for information sharing among the inter-connected IS becomes very important for ISS management. Furthermore, in the scenarios of ISS, the form of sharing information includes not only the hard findings, i.e. the observations, but also the soft findings, i.e. the beliefs or the probability distributions (Gal-Or and Ghose 2009). But, in existing literature, there are few researches taking the information sharing into account in the scenario of ISS management. The models, supposed by Spafford (2000) and Gal-Or (2009), can support the sharing of filtered raw data and binary decisions (e.g., yes/no) among the inter-connected IS, while they are not capable of sharing the security risk beliefs. This limitation often influences their reliability of results.

In this article, we propose a multi-agent based cooperative framework (MACF), in which each component corresponding to one organization's IS is able to process its own data and to integrate local findings with the findings from other components. More specifically, each analysis agent in MACF encapsulates a Bayesian network supporting the sharing in the form of

both the soft findings and the hard findings with other analysis agents. According to an allied organization's risk level yielded by the analysis agent, the control agent is capable to choose an optimal security action that minimizes the expected loss among alternatives.

2. Literature review

In recent years, an emerging research stream, utilizing multi-agent technology, attempts to address the ISS issues under distributed environment. Among these researches, the approach proposed by Boudaoud et al. (2000) provides a flexible integration of a multi-agent system in a classical networked environment to enhance its protection level against inherent attacks. Besides Boudaoud's approach, Helmer et al. (2003) designed and implemented an intrusion detection system prototype that involves lightweight mobile agents. The agents in the system travel between monitored systems in a network of distributed systems, obtain information from data cleaning agents, classify and correlate information, and report the information to a user interface and database via mediators. More recently, adopting the adaptive agent model, Xiao (2009) put forward a security-aware model-driven mechanism by using an extension of the role-based access control model. The major contribution from the approach is a method for building adaptive and secure multi-agent system, following model-driven architecture.

Although these previous studies are highly informative and provide the groundwork for the field of ISS risk management, little attention has been devoted to the security information sharing under distributed environment. Few researches (Spafford and Zamboni, 2000; Gal-Or and Ghose, 2009) achieve the information sharing among the inter-connected IS through the sharing of the hard findings (i.e. the observations), but they are not adequate for supporting the sharing of the soft findings (i.e. the beliefs or the probability distributions). Therefore, our study mainly focuses on addressing the issue that how an allied organization can share the security information in the form of the soft findings as well as the hard findings with other allied members and effectively support the decision making for security practitioners to protect the information resources ina distributed environment.

3. Multi-agent based cooperative framework

Basically, MACFis made up of someinter-connectedIS, which are called as "allied members''. Foreach allied member,there are three kinds of agents:monitor agent, analysis agent and control agent. Suppose that there are *n*allied members and one registry agent. Then Fig. 1 demonstrates the MACF architecture, in which the interactions among the analysis agent and the registry agent of MACF are shown. Specifically, the agent actions and their corresponding messages are given in Table 1.

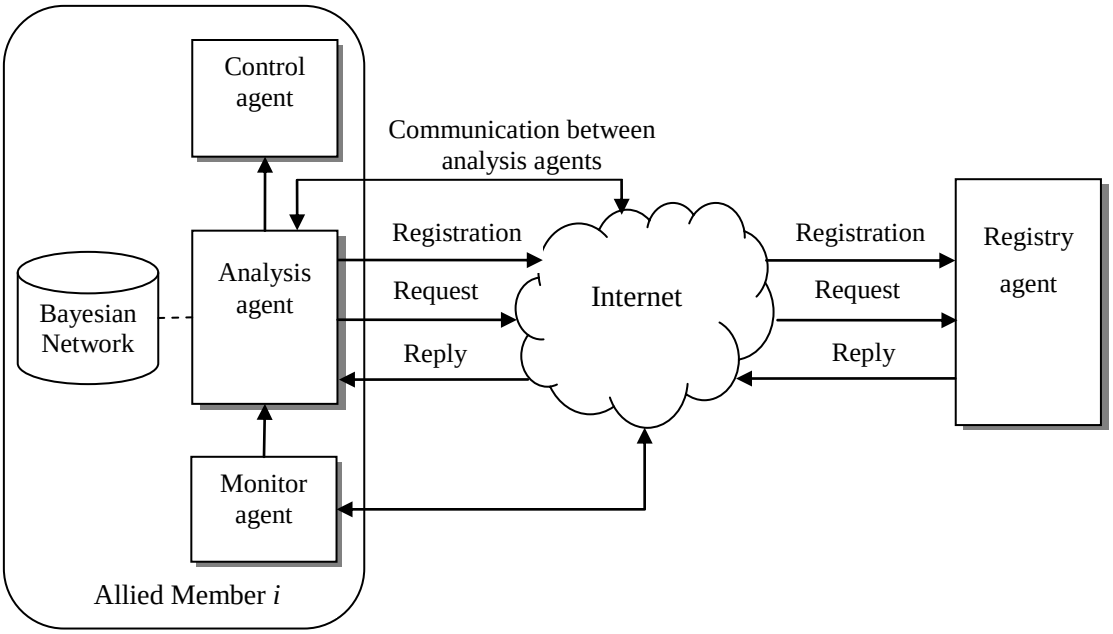


Fig. 1.MACF architecture ($1 \leq i \leq n$).

Table 1. The agent actions and their corresponding messages.

Message	Message content	Agent action
Registration message (from analysis agent to registry agent)	the registering agent's agent-id; IP address;list of published variables (output variables) and their possible states;digital signature and certificate	Each analysis agent must register with the registry agent. The registry agent issues an acknowledgment message upon successfully entering the new agent in its database.
Search request (from analysis agent to registry agent)	the requester's agent-id; IP address; the required input variables	Each analysis agent has a set of input variables. To find agents capable of providing required input data, the analysis agent sends a search request to the registry agent.The message is digitally signed by the requester.
Registry agent's reply (from analysis agent to registry agent)	the requested variable name; the agent-id of the agent publishing the variable; its IP address and status	Upon receiving a search request, the registry agent verifies that the request is legitimate before searching its database to determine which agents can supply the requested variables and the status of these agents.

Belief subscription request (between analysis agents)	the requester's agent-id; requester's IP address; requested input variable name; the duration ofsubscription time; the desired time interval between subsequent updates;the request-id; the timestamp of the request	Upon receiving the list of agents capable of providing the required input from the registry, the subscribing agent sends requests directly to these agents.The message is digitally signed by the requester.
Belief-update message (between analysis agents)	the request-id, the sender's id, the probability distribution of the requested variable	Upon receiving a belief subscription request the publishing agent sends regular updates within the agreed intervals and duration of the subscription. The message is digitally signed by the publisher.

The functionsof each agent are described as follows：

(1) Monitor agent: The monitor agent performs either online or offline processing of log data, communicates with the application systems, and monitors IS resources. The IS profile and deviations generated by the monitor agent can be utilized by the analysis agent as the new evidences (facts and beliefs derived from observations) to update its encapsulated BN.

(2) Analysis agent: The analysis agent encapsulates a BNwhich is used to estimate the probability of each risk level of an allied member based on the operation of belief updating.The nodes of the BN are variablesthat describethe security environment of an allied member. In MACF, the variable can be divided intoinput variable, output variable, and local variable. An analysis agent,to obtain more security information associated with its input variables, can subscribeto the output variables publishedby other analysis agents providing the beliefs of their output variables as the new evidences. Therefore, once the new evidence is obtained through the monitor agents, the analysis agent is able to makeits encapsulated BN modify its own beliefs in real time and import or export beliefs from or to other encapsulated BNs.

(3) Control agent: The control agent belonging to one allied member chooses the action that minimizes the expected loss.Suppose that the control agent has *m* possible security actions and there are six risk levels (from l_0 to l_5) for each allied member. Then the expected loss of each action is defined as follows:

$$Eloss(a_i) = \sum_{j=0}^{5} loss(a_i, l_j) P(l_j), \quad i = 1, \cdots, m \quad j = 0, \cdots 5 \tag{1}$$

where*loss*(a_i, l_j) is the loss function of the *i*th action with respect to l_j risk level; *P* (l_j) is the probability of the allied member that suffers the l_j risk level. According to Eq. (1), the control agent select the action, Min (*Eloss* (a_i)), with the minimum expected loss to protect its information resources.

(4) Registry agent: The registry agent maintains information about the published variables for each analysis agent. It is required that all analysis agents of MACF must register with the registry agent. The registry agent also maintains the location and current status of all theregistered agents. Agent status is a combination of two parameters alive and

reachable. The status of a communication link between any two agents is determined by attempting to achieve a reliable communication between them.

4. Conclusions

In future research, the proposed framework will be applied to a real life supply chain environment, in which there aresix allied members with inter-connected information systems,to illustratehow our model can be employed to manage security risks in distributed information systems.

Acknowledgements

The research was supported by the National Science Fund for Distinguished Young Scholars of China (No.70925005) and the General Program of the National Science Foundation of China (No.71271149, No.71101103, No.61074152, and No.71001076).

Reference

Boudaoud, K., Labiod, H., Boutaba, R., Guessoum, Z. 2000. Network security management with intelligent agents", IEEE Security Management (I), pp. 579–592.

Bulgurcu, B., Cavusoglu, H., Benbasat, I. 2010. "Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness", MIS Quarterly 34(3), pp. 523-548.

Cavusoglu, H., Cavusoglu, H., and Raghunathan, S. 2004. "Economics of IT Security Management: Four Improvements to Current Security Practices", Communications of the Association for Information Systems (14), pp. 65-75.

Gal-Or, E., Ghose, A. 2009."The economic incentives for sharing security information", Information Systems Research 16(2),pp. 186–208.

Helmer, G., Wong, J. S. K., Honavar, V., Miller, L., Wang, Y. 2003."Lightweight agents for intrusion detection", Journal of Systems and Software 67(2), pp. 109–122.

Spafford, E. H., Zamboni, D. 2000. "Intrusion detection using autonomous agents", Computer Networks, 34(4),pp. 547-570.

Xiao, L. 2009. "An adaptive security model using agent-oriented MDA", Information and Software Technology, 51 (5), pp. 933-955.

Towards a Multi-level Framework to Analyze Mobile Payment Platforms

Avinash Gannamaneni
ESSEC Business School, France;
University of Mannheim, Germany
avinash.gannamaneni@gmail.com

Jan Ondrus
ESSEC Business School, France
ondrus@essec.edu

Abstract:
Despite highly optimistic expectations over the past decade, mobile payments are yet to take off successfully. Repeated failures show that mobile payment platforms are complex to launch. We propose a multi-level framework for analyzing the success and failure factors mobile payment platforms. We use a longitudinal case of mobile payments in South Korea to illustrate the framework. Our results show that three consecutive failures could be explained by essential conditions that were not fulfilled at different levels of the framework. The current attempt displays better potential at each level. Ultimately, the outcome of this fourth attempt is likely to be more positive than for the other previous ones.
Key Words: Mobile payments, multi-sided platforms, multi-level analysis

1. Introduction

Because of the centricity of mobile phones in our lives, the idea of paying for goods and services using mobile phones emerged. The motivation behind mobile payments is to converge the two most indispensable items we carry everyday: our mobile phone and our wallet. Theconcept of mobile payment was received enthusiastically by the corporate world. Over the years, numerous attempts of launching mobile payment services were made by different actors including mobile network operators (MNOs) and financial institutions. In early 2000s, a few optimistic analysts and researchers joined the bandwagon by declaring that mobile payments could be the next killer application in mobile commerce (Rosingh et al., 2001; Zheng and Chen, 2003). However, at the same time, others already started to discuss the challenges for mobile payments to become reality (van der Heijden, 2002; Wrona et al., 2001). The number of early initiatives that failed raised concerns. The last decade of trials and pilots confirmed the difficulties of successfully launching commercial implementations of mobile payments. Academics from different disciplines have been conducting research on various mobile payment issues for the last decade.Despite a number of such studies, this emerging phenomenon still raised numerous questions. The complexity of the questions calls for multi-perspective analyses (Ondrus et al., 2005). Unfortunately, explaining one dimension or level at a time only offers a partial understanding on what the challenges are. Few studies have looked at mobile payments from a more holistic point of view while most of them have focused on only one specific aspect. For example, some researchers tackled the economic (Au and Kauffman, 2008) and strategic (Dahlberg et al., 2008) aspects. In this paper, we propose the use of a holistic multi-level framework to better understand the factors that lead to success or failure. We use a longitudinal case study from South Korea to illustrate how platforms could evolve over time and succeed.

2. Multi-level Framework

Mobile payment platforms are considered to be multi-sided as they bring together more than one group of users of the platform (sides): consumers and merchants. Moreover the platform is created through the collaboration of multiple interdependent actors hailing from different industries. Depending on the type of solution involved, mobile payments call for the involvement

of mobile network operators (MNOs), banks, financial institutions, payment networks, payment service providers, technology providers, mobile handset manufacturers, payment terminal manufacturers and other third parties.

In order to classify the actors of the mobile payment ecosystem, we adopt a multi-level classification adapted from Eisenmannet. al. (2009). We partition the mobile payment platform into three distinct levels: i) sponsor level, ii) platform level, iii) user level. The *sponsor level* encompasses the roles and dynamics involving platform sponsorssuch as MNOs and financial institutions. The *platform level* includes the different points of contact of the users with the platform including the technological solution itself. The*user level* is concerned with the two groups of users of the platform: consumers and merchants. Figure 1 summarizes the different actors who could be involved in mobile payment platforms at each of these three levels. We will use this multi-level framework below to analyze a case of mobile payments.

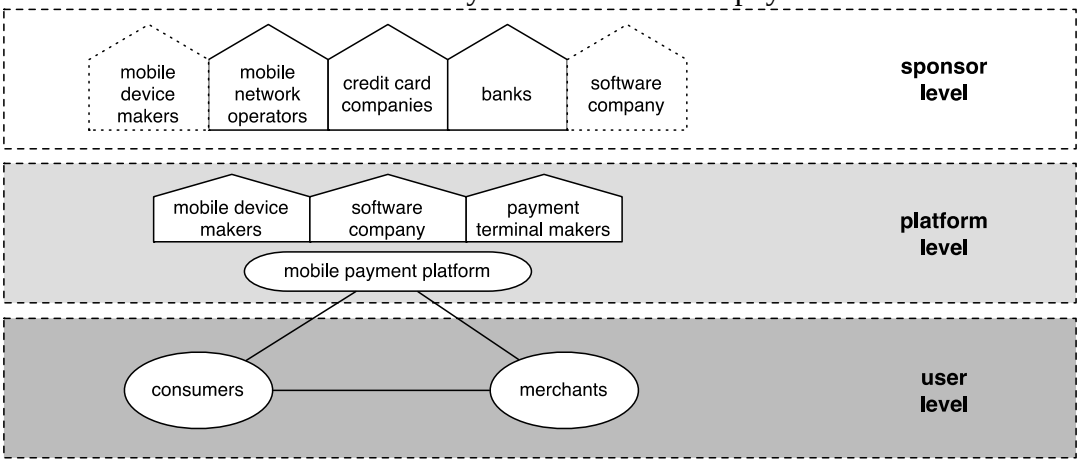


Figure 1: Mobile payments as multi-sided platforms

3. Mobile Payment Case: Moneta

We studied the case of Moneta, a mobile proximity payment solution launched by SK Telecom in South Korea in 2002. Data was collected through interviews with executives from SK Telecom involved in the strategic decision making process behind Moneta. This was coupled with archival data including internal corporate documents, consulting reports and press articles. SKT is persistent with mobile payments. During the last decade, they invested heavily in three successive attempts (2002-2003-2006/7)under the brand name of Moneta. A fourth attempt is on-going.

3.1. Attempt 1 (Feb 2002-2003)

Early 2000s, mobile handsets in South Korea were CDMA-based. No SIM card was required for this technology. In its first attempt, Moneta involved specially designed mobile phones with a full size smartcard reader along with a Moneta credit card. The Moneta card was co-branded by Visa and issued by five domestic credit card companies and banks. Merchants were equipped with a card reader which could read Moneta cards. In order to make a purchase, consumers had to carry both the phone and the card. About 300,000 handsets were sold and 1 million plastic cards were issued. Unfortunately, the users who bought the compatible handsets rarely used the mobile payment service, which made this attempt fail.

Table 1. Multi-level analysis of Attempt 1

Level	**Analysis**
Sponsor	☒ SKT was not able cooperate effectively with financial institutions in order to make Moneta

	compatible with existing the existing credit card system. Financial institutions did not share the same goals with SKT and perceived Moneta to be a threat to their traditional business. SKT was forced to manufacture their own card.
Platform	☒ Multiple proprietary mobile payment solutions were being offered by other players (K-merce by rival MNO KTF, ZOOP by a start-up Harex). It meant that merchants had to install multiple dongles at their point of sale (POS). It became a significant setback to merchants in terms of costs. Both consumers and merchants were wary of adopting any of these proprietary solutions before a winner would emerge as the "*de facto*" standard nationwide.
User	☒ Users had to invest significantly to join Moneta (i.e., consumers: buying an appropriate handset; merchants: installing dongles at payment terminals). In addition, consumers had to carry both their card and mobile handset to make a purchase. Hence there were no added benefits for the users to switch to mobile payments at this stage because of the lack of convenience.

3.2. Attempt 2 (Nov 2003-2005)

In November 2003, SKT introduceda SIM slot for financial chip on their CDMA phones. Six major credit card companies participated in this project. Consumers had to purchase these specially designed phones to be able to use Moneta's services at affiliated stores equipped with Moneta. Merchants had to attach a proprietary dongle supplied by SKT to their existing POS terminal to be able to read Moneta chips. Over the course of this attempt, 400,000 merchants were equipped with the dongle. SKT tried to push the technology to consumers by making most of their handsets compatible with Moneta. By 2005, SKT sold more than 4.9 million Moneta enabled handsets. However, only 300,000 Moneta chips were issued to consumers who registered for the service. This attempt failed as only 21% of these registered users actually made a purchase using their handsets.

Table 2. Multi-level analysis of Attempt 2

Level	**Analysis**
Sponsor	☒ To convince financial institutions to join Moneta, SKT offered them the possibility to manufacture their own SIM sized chips. Therefore, financial institutions would not lose ownership of their customers. Although major credit card companies joined the platform, they did not promote it actively to their customers.
Platform	☒ No change from Attempt 1.
User	☒ Despite the perfect fusion of the phone and the credit card chip, users still did not perceive any added value in using Moneta over existing credit cards. Moreover users still had to invest in appropriate hardware to be able to join the platform (i.e., consumers had to choose phones compatible with Moneta). In addition, the solution was inconvenient for consumers who use multiple credit cards as they had to change the "SIM" chip to use another card.

3.3. Attempt 3 (2006-2007)

Starting in 2006, Visa and Mastercard launched PayWave and Paypass proximity payment platforms based on Near Field Communication (NFC). The two platforms are compatible with both contactless cards as well as mobile payments. The result was a more standardized payment ecosystem in South Korea, where each MNO previously had its own proprietary mobile payment solution. After the mobile telecommunication networks evolved from CDMA to WCDMA in 2006, SKT supplied multi-functional SIMs which could store credit card data. Consumers could download multiple credit cards as well as other applications over the air onto their SIMs. In May 2007, SKT formed an alliance with Visa and started issuing SIM cards with a contactless payment feature. Initially, there was resistance from banks and financial institutions to offer their credit card applications over the air. However, in April 2008, SKT formed another alliance with

Shinhan bank, one of the largest banks in South Korea in order to jointly offer and market Moneta. This attempt raked in 2.6 million subscribers. Despite impressive initial figures, the usage of Moneta's services was too low and the service eventually failed.

Table 3. Multi-level analysis of Attempt 3

Level	Analysis
Sponsor	☒ SKT issued its own SIM cards. As a result, financial institutions were again afraid of losing ownership of their customers. SKT had difficulties convincing all major institutions to join the platform. Moreover, SKT and financial institutions could not agree on revenue sharing. SKT wanted to take 1% of the total 2.5% transaction fees while the financial institutions did not want to give away anything to SKT. Furthermore, there was a complete lack of respect between the MNOs and the financial institutions which manifested in the form of various public bickerings.
Platform	☑ NFC emerged as the standard and all the rival proprietary solutions became interoperable with each other. Merchants no longer needed to install any additional dongles. Consumers had NFC equipped handsets when they upgraded their devices as most of the third generation phones were enabled with NFC.
User	☑ All technology issues were solved as many consumers had NFC-enabled handsets. Merchants automatically received NFC-compatible payment terminals during upgrades. Moneta offered various additional applications, including the extremely popular T-Money (for public transport) and some loyalty schemes. Overall, Moneta offered an additional value over existing payment schemes.

3.4. Post-Moneta: Hana SK Card

In 2010, SKT secured a 49% stake in Hana card, a top-three Korean credit card, and rebranded it as Hana SK card. SKT was able to attract 500,000 subscribers for the Hana SK card's mobile wallet within a year of its launch (i.e., touted as the biggest figure for mobile wallets in the world). As of January 2013, SKT seems to be leading the mobile wallet race in South Korea with a combined annual growth rate of 661.7% since it was launched. This figure points towards an optimistic future for mobile payments in South Korea with a potential for mass adoption.

Table 4. Multi-level Analysis of Hana SK card

Level	Analysis
Sponsor	☑ Since both MNO and financial institution are now owned by the same entity (SKT), the problems related to effective coordination between platform sponsors are solved.
Platform	☑ Solved in Attempt 3.
User	☑ Solved in Attempt 3.

4. Discussion

After analyzing the different attempts of Moneta through the use of a multi-level framework, we uncovered a number of factors responsible for the success or failure of mobile payment platforms. Firstly, at the *sponsor level*, there is a need for effective cooperation between MNOs and financial institutions for mobile payments to succeed. This must be achieved through sharing the same goals, agreeing on value sharing, and respecting each other. Secondly, at the *platform level*, all competing rival platforms must reach a consensus on an industry wide standard. Failure to do so would result in multiple proprietary solutions and it is likely that most of them would fail, especially in a fragmented market with no clear dominant MNO or financial institution. Finally, at the *user level* mobile payments must offer a significant value addition over existing payment schemes to entice users to join the platform. Merely merging phone and credit card is not sufficient to succeed as we have learnt from the case of Moneta.

The contribution of our study is two-fold. First, we demonstrate the relevance of using a multi-

level framework to analyze the success and failure factors of multi-sided platforms, especially mobile payment platforms. Multi-sided platforms are complex by nature. Looking at one dimension at a time does not provide enough depth to explain success or failure of these platforms in a convincing manner. Second, we provide better explanations and justifications why Moneta kept failing in South Korea over the years. We hope that the insights presented in this paper could help practitioners to better design and manage multi-sided platforms. Special efforts must be made to satisfy factors at every level in order to increase their likelihood of success.
In a further research, we aim at testing this framework on more cases/platforms in different contexts in order to validate its relevance in better understanding multi-sided platform design. We expect to extract more success and failure factors from the cases at each level. As a result, we could provide a list of success factors that need to be carefully examined during the design of any multi-sided platform.

5. References

Au, Y.A., Kauffman, R.J., 2008. The economics of mobile payments: Understanding stakeholder issues for an emerging financial technology application. Electronic Commerce Research and Applications 7, 141–164.

Collins, C., Smith, K., 2006. Knowledge exchange and combination: the role of human resource practices in the performance of high-technology firms. Academy of Management Journal, 49(3) 544–560.

Dahlberg, T., Huurros, M., Ainamo, A., 2008. Lost Opportunity–Why Has Dominant Design Failed to Emerge for the Mobile Payment Services Market in Finland? Proceedings of the 41st Annual Hawaii International Conference on System Sciences.

Eisenmann, T.R., Parker, G., Van Alstyne, M., 2009. Opening Platforms: How, When and Why?, in: Gawer, A. (Ed.), Platforms, Markets and Innovation. Edward Elgar Publishing.

Ondrus, J., Camponovo, G., Pigneur, Y., 2005. A Proposal for a Multi-Perspective Analysis of the Mobile Payments Environment.The Fourth International Conference on Mobile Business (ICMB).

Ondrus, J., Lyytinen, K., Pigneur, Y. "Why Mobile Payments Fail? Towards a Dynamic and Multi-perspective Explanation", 42th Annual Hawaii International Conference on System Sciences (HICSS'09), IEEE Computer Society, 5-8 Jan 2009, Hawaii, USA

Rosingh, W., Seale, A., Osborn, D., 2001. Why Banks and Telecoms Must Merge to Surge [WWW Document]. strategy+business. URL http://www.strategy-business.com/article/17163?gko=4cda6 (accessed 6.23.13).

van der Heijden, H., 2002. Factors Affecting the Successful Introduction of Mobile Payment Systems, in:.Presented at the Proceedings of the 15th Bled eCommerce Conference, Bled, Slovenia, June 17-79.

Wrona, K., Schuba, M., Zavagli, G., 2001.Mobile Payments — State of the Art and Open Problems, in: Link.Springer.com, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, 88–100.

Zheng, X., Chen, D., 2003.Study of Mobile Payments System. Proceedings of the IEEE International Conference on Electronic Commerce (CEC), Newport Beach, CA, USA, June 24-27.

Linking IT-Enabled Collaboration and Performance in the SME Context: The Role of Responsiveness Capabilities

Hsin-Lu Chang
hlchang@nccu.edu.tw
National Chengchi University, Taiwan

Daniel Yen Shueh
daniel76308@gmail.com
National Chengchi University, Taiwan

Kai Wang
kwang@nuk.edu.tw
National University of Kaohsiung, Taiwan

Abstract

The service economy has been expanding recently, with small- and medium-sized enterprises (SMEs) playing an important role. Previous research has shown that responsiveness is one of the most important strategic capabilities for SMEs to enhance their performance. How to define responsiveness and how to develop it in the SME context, however, are seldom discussed. A review of the literature leads us to propose the three dimensions of responsiveness in the SME context, which are market sensing, customer linking, and promptness. In addition, we propose that IT-enabled collaboration would facilitate this capability. We develop a research framework to examine the relationshipsamong IT-enabled collaboration, responsiveness, and organizational performance. To verify our research framework, a case study deployed in the Mt. Pillow Leisure Agricultural Area in Yilan County, Taiwan will be carried out. This paper is a work-in-progress and the aim is to contribute to the literature by defining responsiveness capabilities in the SME context and providing a framework for examining the relationships between IT-enabled collaboration, responsiveness capabilities, and organizational performance. The results of this study not only help SMEs in developing their responsiveness capabilities but also guide them to retain competitive advantage in the service economy.

Keywords: SMEs, IT-Enabled Collaboration, responsiveness capabilities, service performance, leisure agriculture.

1. Introduction

Responsiveness is one of the most important strategic capabilities that should be considered for enhancing the performance of service organizations [39, 43]. It contributes to organizations' capability to deal with changes in customer demands [33, 44] and enhances organizational performance [38]. Customers are becoming more sophisticated in their needs and are increasingly demanding a higher standard of service. Therefore, when considering levels of performance as part of setting customer service objectives, service providers must take responsiveness into account as an important capability [29].

Responsiveness is critical for SMEs to remain competitive and sustain high performance. A primary task for resource-limited SMEs in emerging economies has been to develop low-cost and easily implemented measures to improve their sustainability and to increase the chances of success when facing rapid and often unforeseen changes in the external environments.

One strategic response to increasing uncertainty is to establish collaboration between SMEs using information technology (IT) [3, 12, 34]. A well-developed capability to create and sustain fruitful collaborations gives organizations a significant competitive edge [31]. Through IT-enabled collaboration, SMEs can become more responsive by searching and collecting information quickly and efficiently from their partners and customers, thus improving their sustainability [21, 41, 45].

Few existing studies, however, have addressed the role of IT-enabled collaboration among

both SME partners and customers in enhancing SMEs' responsiveness to improve their performance. The current study aims to investigate the relationships between IT-enabled collaboration, responsiveness, and organizational performance. This study proposes that the performance of SMEs could be improved through IT-enabled collaboration, with the relationship between these two constructs mediated by responsiveness. This study aims to answer the following questions: (1)Why is responsiveness so important for SMEs? (2)How will IT-enabled collaboration enhance SMEs' responsiveness? (3)How does responsiveness relate to organizational performance?

2. Conceptual Background and Hypotheses Development

Responsiveness involves three capabilities. *Market-sensing capability* is a process for learning about present and prospective market environments. Market-sensing can be divided into three processes: sensing activities, sense-making activities, and reflection [4, 5]. Sensing activities include the acquisition of information on consumers, competitors, and other channel members. Sense-making activities involve the interpretation of gathered information based on past experiences and knowledge. Reflection means the utilization of the gathered and interpreted information in decision-making.*Customer-linking capability*refers to the ability to develop and manage close customer relationships and is among the most valuable capabilities of any organization [4].An organization's customer-linking capability creates a potential competitive advantage in business [30]. Well-managed customer relationship creates a great opportunity to increase customer value and provides a way to systematically attract, acquire, and retain customers [19].Given the importance of *promptness*, many definitions have emerged [9, 24, 32, 33]. Among all these definitions, Kidd has provided the most comprehensive one: A prompt organization is a fast-moving, adaptable, and robust business. Such a business is founded on processes and structures that not onlysupport speed, adaptation, and robustness buy alsofacilitate a coordinated business capable of achieving competitive performance in a highly dynamic and unpredictable environment to which the current practices are poorly suited.

Responsiveness is especially important for SMEs to maintain customer loyalty because SMEs face an endless stream of competition from larger companies that have the richer resources to be "on call" for their clients constantly. Such competition demonstrates the importance of fast response to and efficient communication withthe market, partners, and customers for SMEs to achieve business success.

Collaboration is an effective way for SMEs to achieve better performance and long-term survival [3, 12, 34]. Collaboration can be conducted either horizontally with SME partners or vertically with customers, using IT to share information more efficiently and effectively to improve coordination and collaboration activities [2, 22].

Appropriate use of information is fundamental to the ability of sensing market requirements because if an organization does not have adequate and accessible resources and information, it stands at a competitive disadvantage [11]. Haeckel and Nolan [10] stressed that information technology is critical to managing conditions that are too turbulent to make sense of. IT-enabled collaboration thus allows resource-limited SMEs to acquire and share information efficiently and effectively, thus strengthening their ability to sense the market. Therefore, we hypothesize the following:

Hypothesis 1-a: IT-enabled collaboration with SME partners has a positive impact on market-sensing capability.

Although customer relationships are viewed as an intangible resource that may be relatively rare and difficult for others to replicate [13, 37], the capability of SMEs to acquire and manage customer information is limited due to their smaller scale. By coordinating information and activities with strategic partners, a SME can develop more ways to attract customers, and

become more responsive to customer requests and build greater customer loyalty and better customer relations [30, 40]. Therefore, we predict that customer-linking capability can be improved by collaboration through information technology:

Hypothesis 1-b: IT-enabled collaboration with SME partners has a positive impact on customer-linking capability.

SMEs usually lack promptness because they lack resources to cultivate such capability. To enhance promptness, it is important to strengthen communication and collaboration and improve decision-making processes [25]. SMEs can acquire necessary resources and capabilities by forming alliances [6]. IšoraItė[36]indicated that organizations involved in alliances are better able to utilize resources to improve their speed to the market and the speed in serving customers. Information technology makes such coordination feasible [18]. Paulraj and Chen [28] stated that IT-enabled collaboration increases information processing speed by providing an intermediary platform for partners to share knowledge, provide timely information, and transcend each firm's boundaries. We thus hypothesize the following:

Hypothesis 1-c: IT-enabled collaboration with SME partners has a positive impact on promptness.

Unlike large organizations, SMEs do not have the resources to engage in formal market research [17]. For SMEs to sense the market precisely and adapt to it, they must collaborate with their customers to collect a significant amount of data and analyze it. This analysis will provide them with better insight into customer requirements and expectations, ultimately resulting in services that are more suited to the market [16]. With the help of information technology, SMEs are able to gather, store, access, and analyze customer data to effectively make strategic business decisions [42]. Therefore, we hypothesize that IT-enabled collaboration with SME customers provides an environment for SME organizations to collect and analyze market data from customers and thus enhances the capability of organizations to sense the market.

Hypothesis 2-a: IT-enabled collaboration with SME customers has a positive impact on market-sensing capability.

One usual but crucial reason for an organization to conduct customer-linking activities iscustomers' low satisfaction with services and products [1, 8]. However, linking to customers is a time-consuming and resource-demanding process for SMEs. It is therefore essential for SMEs to enhance their customer-linking capability by collaborating with customers through a friendly, accessible, and directly interactive channel so that customers feel comfortable to give feedback [16]. Füller et al. [7]indicated that organizations can be able to form new channels to collaborate with customers, effectively share knowledge, and manage relationshipswith the help of IT. As a result, IT-enabled collaboration with customers may reduce the distance between resource-limited SMEs and customers. Thus, we develop the following hypothesis:

Hypothesis 2-b: IT-enabled collaboration with SME customers has a positive impact on customer-linking capability.

Whereas large organizations can employ a number of people to take care of their customers immediately, resource-limited SMEs usually struggle to respond quickly to the demands of their customers. For SMEs, a fast and efficient communication tool to learn about customers is vital if they are to achieve business success. Through collaboration with customers, SMEs are better able to learn about their customers in multiple ways by shortening the response time needed. Moreover, IT enables organizations to reduce the time required to share information and reduce response time to unforeseen events, thereby enhancing promptness [15, 25]. Therefore, we hypothesize that SMEs gain promptness through IT-enabled collaboration with customers.

Hypothesis 2-c: IT-enabled collaboration with SME customers has a positive impact on promptness.

Better market-sensing capability allows an organization to discover underserved market

segments and the segments where the rivals' offerings may not be fulfilling customers' needs [35]. These underserved and unsatisfied segments are good targets for organizations seeking new customers. Hult[14] and Morgan et al. [26] suggested that market-sensing capability provides market insights that enable organizations to reduce their costs through effective use of resources by better matching the organization's resource acquisitions and deployments with customer and prospect opportunities. By doing so, SMEs are better able to forecast the value of different resources accurately, thus enabling them to manage resources in a better way to achieve higher performance [20]. Therefore, we propose hypothesis 3-a as follows:

Hypothesis 3-a: Market-sensing capability has a positive impact on SME performance.

SMEs may sometimes spend their limited resources and time on other tasks at the cost of customer satisfaction. This trade-off may lead SMEs to lose business to their larger competitors. Managing relationships with customers is therefore critical.

In the current study, customer-linking capability is defined as an organization's ability to manage the relationship with its customers through direct contact. Direct customer contacts shorten service cycles and lower service costs. Nielsen [27] and Hooley et al. [13]pointed out that customer-linking capability enables the development and maintenance of strong customer relations and ultimately improves customer satisfaction and loyalty. As a result, we expect the following hypothesis to hold:

Hypothesis 3-b: Customer-linking capability has a positive impact on SME performance.

Due to the smaller scale and limited funds, SMEs need to determine the most efficient as well as effective market strategies for improving their performance. A firm's promptness represents the strength of the interface between the organization and the market [23]. Organizations that are prompt in response to customer requirements demonstrate operational flexibility, which is able to eliminate waste in their operations, better direct their interactions with their customers to improve customer retention, and in general, reduce the costs incurred in servicing the customer base. We therefore argue that promptness can yield better SME performance in two ways, namely gaining profit by quickly adapting to market changes and reducing cost by eliminating waste from operations. Thus, we posit:

Hypothesis 3-c: Promptness has a positive impact on SME performance.

The research model is shown in Figure 1.

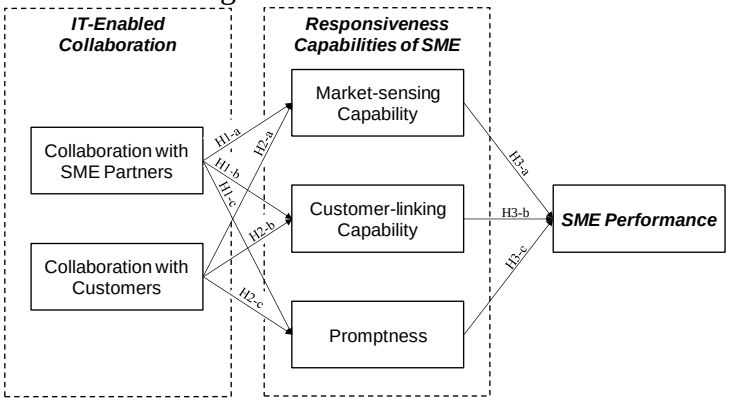


Figure 1. Research Model

3. Research Methodology

Leisure agriculture is a new trend of agricultural operations that combines local industry, cultural characteristics, leisure, natural ecology, and accommodations. With its unique geography and diverse cultural and natural resources, Taiwan possesses significant potential to develop leisure agriculture. However, SMEs involved in leisure agriculture are seldom able to collaborate to enhance their competitiveness. Operating independently, they continue to suffer from low

productivity, lack of innovation, and slow growth due to the lack of necessary resources to manage and fulfill customer needs efficiently.

Based on the discussion above, we believe that IT-enabled collaboration will enable SMEs to enhance their responsiveness capabilities to achieve a higher level of service performance.To better examine the relationships between IT-enabled collaboration, responsiveness capabilities, and performance, we plan to adopt the case study approach andperform in-depth analysis.

For the purpose of the current study, we selected eight SMEs in the Mt. Pillow Leisure Agriculture Area, Yilan County, Taiwan. These SMEs include farms, orchards, gardens, restaurants, natural landscapes, natural ecological areas, and accommodations. Due to space limitation, the detailed descroption of case companies is available on request from the first author.

Because the SMEs in the Mt. Pillow Leisure Agriculture Area do not yet have a unified IT-enabled collaborative platform, we extend the targeted platform to include tools that can provide a channel for SMEs to obtain and share information, directly communicate and interact, and engage in collaborative projects with customers and other SME partners. Therefore, the IT-enabled collaborative platforms we take into consideration include, among others, blogs, guestbooks, and Facebook. We will conduct two interviews for each case, each lasting approximately one to two hours. All interviews will be tape recorded and transcribed before the data analysis. To enhance the validity of answers, summaries of the major finding in each interview will be verified by the interviewees after each interview session. Moreover, to ensure the construct validity, internal validity, external validity, and reliability of the case study, Yin's[85] case study tactics will befollowed.

4. Expected Contribution

The service economy has grown significantly in the last decade, and SMEs are an important part contributing to the growth in this sector. To serve customers in the turbulent environment, SMEs must enhance their responsiveness to retain long-term competitiveness.

Few published articles have addressed the issue of how to enhance SMEs' responsiveness through IT-enabled collaboration. The current study aims to contribute to the literature by defining responsiveness capabilities in the SME context and providing a framework for examining the relationships between IT-enabled collaboration, responsiveness capabilities, and organizational performance. We believe that the results will not only help SMEs to develop their responsiveness capabilities but also offer SMEs a guide to retaining their competitive advantage in the service economy through IT-enabled collaboration.

References

[1] K. Atuahene-Gima, "The Influence of New Product Factors on Export Propensity and Performance: An Empirical Analysis", Journal of International Marketing, 3(2), 1995, 11-28.

[2] A. Barua, P. Konana, A.B.Whinston, and F. Yin, "An Empirical Investigation of Net-Enabled Business Value", MIS Quarterly, 28(4), 2004, 585-620.

[3] P. Bastos, "Inter-Firm Collaboration and Learning: The Case of the Japanese Automobile Industry", Asia Pacific Journal of Management, (4), 2001, 423-441.

[4] G.S. Day, "The Capabilities of Market-Driven Organizations", Journal of Marketing, 58(4), 1994, 37-52.

[5] G.S. Day, "Managing the Market Learning Process", Journal of Business & Industrial Marketing, 17(4), 2002, 240-252.

[6] Z. Fernández and M.J. Nieto, "Internationalization Strategy of Small and Medium-Sized Family Businesses: Some Influential Factors", Family Business Review, 18(1), 2005, 77-89.

[7] J. Füller, H.Mühlbacher, K.Matzler, and G.Jawecki,"Consumer Empowerment through Internet-Based Co-Creation", Journal of Management Information Systems, 26(3), 2009, 71-102.

[8] J. Goldenberg, D.R. Lehmann, andD. Mazursky,"The Idea Itself and the Circumstances of its Emergence as Predictors of New Product Success", Management Science, 47(1), 2001, 23-26.

[9] A. Gunasekaran,"Agile Manufacturing: AFramework for Research and Development", International Journal of Production Economics, 62(1/2), 1999, 87-105.

[10] S.H. Haeckel, and R.L. Nolan,"Managing by Wire", Harvard Business Review, 5(71), 1993, 122-132.
[11] P. Herbig and A.T. Shao,"American Keiretsu: Fad or Future", Journal of Business-to-Business Marketing, 1(4), 1993, 3-30.
[12] M. Hitt, D. Ahlstrom, M. Dain, E.Levitas, andL. Svobodina,"The Institutional Effects on Strategic Alliance Partner Selection in Transition Economies: China vs. Russia", Organization Science, 15(2), 2004, 173-185.
[13] G.J. Hooley, G.E.Greenley, J.W.Cadogan, andJ. Fahy,"The Performance Impact of Marketing Resources," Journal of Business Research, 58(1), 2005, 18-27.
[14] G.T. Hult,"Managing the International Strategic Sourcing Process as a Market-Driven Organizational Learning System", Decision Sciences, 29(1), 1998, 193-216.
[15] H. Katayama andD. Bennett,"Agility, Adaptability and Leanness: A Comparison of Concepts and a Study of Practice", International Journal of Production Economics, 60-61(20), 1999, 43-51.
[16] C. Kausch,ARisk-Benefit Perspective on Early Customer Integration,Springer, Berlin, 2007.
[17] H.T. Keh, T.T.M. Nguyen, and H.P. Ng,"The Effects of Entrepreneurial Orientation and Marketing Information on the Performance of SMEs", Journal of Business Venturing, 22(4), 2007, 592-611.
[18] K. Kumar and H.G.V.Dissel,"Sustainable Collaboration: Managing Conflict and Cooperation in Interorganizational Systems", MIS Quarterly, 20(3), 1996, 279-300.
[19] Y.C. Lin andH.Y. Su,"Strategic Analysis of Customer Relationship Management—AField Study on Hotel Enterprises", Total Quality Management, 14(6), 2003, 715-731.
[20] R. Makadok,"Toward a Synthesis of the Resource-Based and Dynamic-Capability Views of Rent Creation", Strategic Management Journal, 22(5), 2001, 387-401.
[21] E. Malecki and D. Tootle,"The Role of Networks in Small-Firm Competitiveness", International Journal Technology Management, 11(1/2), 1996, 43-57.
[22] C. Martinez-Fernandez,Networks for Regional Development: Case Studies from Australia and Spain, PhD Thesis, University of New South Wales, Sydney, 2001.
[23] R. Mason-Jones and D.R.Towill,"Total Cycle Time Compression and the Agile Supply Chain", International Journal of Production Economics, 62, 1999, 61-73.
[24] R.E. McGaughey,"Internet Technology: Contribution to Agility in the Twenty-First Century", International Journal of Agile Management Systems, 1(1), 1999, 7-13.
[25] A.E.C. Mondragon, A.C. Lyons, andD.F. Kehoe,"Assessing the Value of Information Systems in Supporting Agility in High-Tech Manufacturing Enterprises", International Journal of Operations and Production Management, 24(12), 2004, 1219-1246.
[26] N.A. Morgan, D.W.Vorhies, andC. Mason,"Market Orientation, Marketing Capabilities, and Firm Performance", Emerald Management Reviews, 30(8), 2009, 909-920.
[27] J.F. Nielsen,"Internet Technology and Customer Linking in Nordic Banking", International Journal of Service Industry Management, 13(5), 2002, 475-495.
[28] A. Paulraj andI.J. Chen,"Strategic Buyer-Supplier Relationships, Information Technology and External Logistics Integration", Journal of Supply Chain Management, 43(2), 2007, 2-14.
[29] A. Payne,The Essence of Service Marketing, Prentice-Hall, UK, 1995.
[30] A. Rapp, K.J.Trainor, andR. Agnihotri,"Performance Implications of Customer-Linking Capabilities: Examining the Complementary Role of Customer Orientation and CRM Technology", Journal of Business Research, 63(11), 2010, 1229-1236.
[31] M. Rosabeth,"Collaborative Advantage: The Art of Alliances", Harvard Business Review, 72(4), 1994, 96-108.
[32] V. Sambamurthy, A. Bharadwaj, andV. Grover,"Shaping Agility through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms", MIS Quarterly, 27(3), 2003, 237-263.
[33] H. Sharifi andZ. Zhang,"A Methodology for Achieving Agility in Manufacturing Organizations: An Intruduction", International Journal of Production Economics, 62(1/2), 1999, 7-22.
[34] E. Sivadas and F.R. Dwyer,"An Examination of Organizational Factors Influencing New Product Success in Internal and Alliance-based Processes", Journal of Marketing, 64(1), 2000, 31-49.
[35] S.F. Slater and J.C.Narver,"Intelligence Generation and Superior Customer Value", Journal of the Academy of Marketing Science, 28(1), 2000, 120-127.
[36] M. IšoraItė, "Importance of Strategic Alliances in Company's Activity", Intellectual Economics, 1(5), 2009, 39-46.
[37] R.K. Srivastava, T.A.Shervani, and L. Fahey,"Market-Based Assets and Shareholder Value: A Framework for Analysis",The Journal of Marketing, 62(1), 1998, 2-18.
[38] G. Stalk,"Time: The Next Source of Competitive Advantage", Harvard Business Review, 66(4), 1988, 41-51.
[39] G. Stalk and T.Hout,Competing against Time, Free Press, New York, NY, 1990.
[40] J.R. Stock,"Managing Computer, Communication and Information Technology Strategically: Opportunities and Challenges for Warehousing",The Logistics and Transportation Review, 26(2), 1990, 133-148.

[41] L. Suarez-Villa,"The Structures of Cooperation: Downscaling, Outsourcing, and the Networked Alliances", Small Business Economics, 10(1), 1998, 5-16.
[42] P. Swafford, S. Ghosh, and N. Murthy,"A Framework for Assessing Value Chain Agility", International Journal of Operations and Production Management, 26(2), 2006, 170-188.
[43] R. Teare,"Hospitality Operations: Patterns in Management, Service Improvement and Business Performance", International Journal of Contemporary Hospitality Management, 8(7), 1996, 63-74.
[44] A.S. Tsui,"Reputational Effectiveness: Toward a Mutual Responsiveness Framework",in L.L. Cummings and B.M. Staw (Eds.), Research in Organizational Behavior, 16, 1994, 257-307.
[45] P. Varadarajan and M. Cunningham,"Strategic Alliances: A Synthesis of Conceptual Foundations", Journal of the Academy of Marketing Science, 23(4), 1995, 282-296.

Fostering a Seamless Customer Experience in Cross-Channel Electronic Commerce

Jaeki Song
Texas Tech University
jaeki.song@ttu.edu

Miri Kim
Sogang University
mirikim@sogang.ac.kr

Jeff Baker
American Univ. of Sharjah
jbaker@aus.edu

Junghwan Kim
Texas Tech University
Jun.kim@ttu.edu

ABSTRACT

Many companies have launched mobile channels as an extension of existing websites in order to enhance communication with customers. Despite employing multiple technologies, retailers have difficulty integrating web and mobile channels to deliver a consistent and seamless customer experience. Our initial empirical results suggest that customers' perceived quality of channel integration is determined by the functional configuration of the mobile applications, by the perceived similarity across channels, and by the perceived quality of the website.

Keywords*: e-commerce; m-commerce; mobile applications; cross-channel commerce; seamless customer experience*

1. INTRODUCTION

Mobile commerce is rapidly becoming a popular and valuable retail channel, with both consumers as well as producers realizing numerous benefits (Oracle, 2011). Retailers are finding that a mobile sales channel can function as an extension of their traditional web channel, creating additional value for stakeholders. More than 75% of consumers use two or more channels to research and complete transactions when they purchase a product or service, suggesting that "retailers need not necessarily serve up the identical experience in each channel, but rather they can optimize and connect channel interactions to deliver consistent user experiences" (Oracle 2011, p.4). Consequently, many retailers are investing in mobile services as a valuable channel for e-commerce, and plan their strategies to include websites, mobile sites, and mobile applications(Adobe 2010; Lamont 2012).

At the same that retailers are learning to add mobile channels to their existing web channels, it has been noted that many companies find it difficult to deliver a consistent experience across the various electronic channels customers use (Fodor, 2012). For instance,57% of mobile web users would not recommend a business with a bad mobile site, and 40% of users have turned to a competitor's site after a bad mobile experience (Compuware 2011). A bad experience on a mobile site or application also leaves consumers less likely to utilize or recommend a related website. Consistency in functionality, information, and visualization is important to customers. The research question we therefore seek to investigate is, "What are the factors that influence

customers' perceptions regarding consistency across retailers' various electronic channels?"

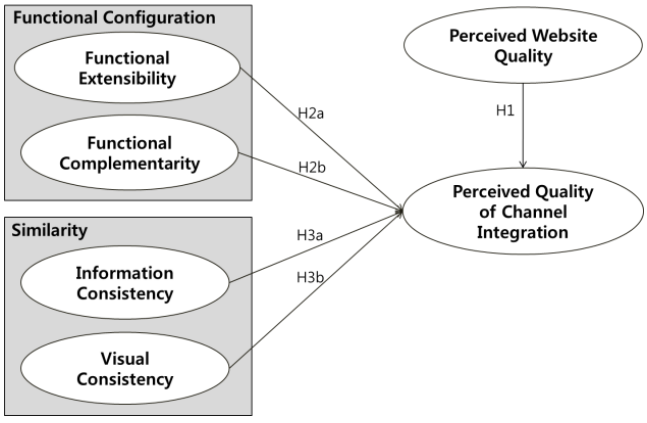
2. THEORETICAL BACKGROUND AND HYPOTHESIS DEVELOPMENT

A Seamless Customer Experience: Perceived Quality of Channel Integration

Cross-channel commerce is a growing trend in retailing (Neilson, et al., 2006; Kwon and Lennon, 2009; Madaleno, et al, 2007). Customers are increasingly using multiple channels, including computers and mobile devices, to obtain information about products and services, or to complete a transaction (Oracle, 2011). Given the growing reality of cross-channel electronic commerce, researchers have begun to investigate the effects of channel integration on a variety of outcomes such as customer satisfaction and retention, as well as its effects on customer relationship management (CRM) (Bendoly, et. al., 2005; Yang, et al., 2013). Researchers have shown that firms with well-integrated channels are more successful than single-channel firms or than those with multiple, but poorly-integrated channels (Sousa et al. 2006). Channel integration can enrich customers' experiences with retailers, and can strengthen customers' overall perceptions regarding the image of a retailer (Kwon and Lennon, 2009). The development of multiple well-integrated electronic channels is thus a goal and a challenge for retailers.

Retailers realize that customer experience occurs during all moments of contact with the firm through multiple channel settings (Rosenbloom 2007; Sousa et al. 2006). Offering a seamless and consistent experience across channels is important because each interaction with a customer plays an essential role in enhancing (or degrading) the quality of the customer relationship (Madaleno et al. 2007; Payne et al. 2004). A seamless customer experience provides value not only to customers as they gain information and make transactions, but also to firms as they cultivate relationships with customers and potential customers. Therefore, this study aims to provide an understanding of how retailers offer a seamless customer experience across channels.

Our research model shows how we plan to investigate customers' perceptions of consistency and channel integration (see Figure 1). We suggest that customers' perceived quality of channel integration is determined by the functional configuration of the mobile applications, by the perceived similarity across channels, and by the perceived website quality. We now describe in greater detail each of the relationships depicted in our model.



[Figure 1] Research Model

The Influence of Website Quality

Companies both large and small have online, virtual means for reaching customers (Rosenbloom 2007). Website quality has been identified as a key factor that influences the success of an e-commerce website, with website quality being defined as "the extent to which a website facilitate the efficient and effective shopping, purchasing, and delivery products and services" (Gounaris et al., 2005, p. 673) We suggest in this study that quality is important not only in traditional websites, but in all electronic channels. Customers' experiences with existing channels necessarily influence how they cognitively or affectively perceive new channels (Zeithmal 2002). Therefore, we hypothesize:

H1.In cross-channel electronic commerce, website quality is positively associated with the perceived quality of channel integration.

Mobile applications lead to new opportunities and challenges for retailers. The limitations of mobile devices, including screen size, processing power, and limited functionality relative to laptop or desktop computers, are critical challenges for retailers. At the same time, portability, ubiquity, and familiarity with the device present several advantages (Buellingen et al. 2004; Steele 2003). With these opportunities and challenges in mind, we focus on two additional factors that may influence the perceived quality of channel integration: the functional configuration of the mobile application and similarity of the mobile application to other channels.

Functional Configuration of the Mobile Application

The *functional configuration* of the mobile application includes two elements as potential indicators: functional extensibility and functional complementarity. *Functional extensibility* refers to the extent to which a mobile channel extends the web channel's functionality by adding new functions (Yang et al. forthcoming). That is, functional extensibility implies that mobile channels deliver different functions for achieving similar goals, which can be utilized in the website, with new or mobile-customized technologies. For example, the application for Amazon.com uses location management technology to determine where a shipper should deliver an item – this is an example of functional extensibility. Mobile applications are designed to provide many of the core functions used in websites in addition to different mobile networking technologies (Varshney et al. 2002). These technologies contribute to a function of the mobile application and result in better customer service by creating synergies (Rosenbloom 2007). Hence, regarding functional extensibility, we hypothesize:

H2a.In cross-channel electronic commerce, the functional extensibility of the mobile application is positively associated with perceived quality of channel integration.

Functional complementarity refers to the extent to which the basic functionality and essential features of a mobile application support or complement the e-commerce activities customers can perform on a website. Providing complementary functions for customers to complete their intended activities is important to ensure superior customer experience (Mithas et al. 2007). For instance, a UPS customer can search for a particular shipping product (such as "Next-Day Air"), schedule a pick-up, and complete payment all online. Then, the customer can use the package tracking feature on the UPS mobile application. The use of this feature on the mobile application

complements the customer's activities on the traditional website. Similarly, online ticketing, for air travel can be completed on a website, with the complementary activity of checking a flight's arrival time performed via a mobile application(Buellingen et al. 2004). By providing customers with the features on a mobile application that complement those of the website, service providers enhance customer experience. Therefore, we posit that

H2b.In cross-channel electronic commerce, the functional complementarity of the mobile application is positively associated with the perceived quality of channel integration.

Perceived Similarity of the Mobile Application to Other Channels

Perceived similarity captures an individual's beliefs about how similar elements are to each other (Brown and Inouye, 1978). In this study, the similarity we are investigating is the similarity between e-commerce sales channels, particularly between mobile applications and websites. Perceived similarity has two indicators: information consistency and visual consistency.

Information consistency refers to the consistency between information exchanged with the customer through different channels, including both outgoing and incoming information (Sousa et al. 2006). If inconsistency or conflict of the information across channels exists, this inconsistency will confuse the customer and may reduce the likelihood of purchase. Firms must provide quality across all channels, ensuring a coherent message with all information conveyed by different channels (Payne et al. 2004). Thus, we hypothesize:

H3a. In cross-channel electronic commerce, the information consistency of the mobile application is positively associated with the perceived quality of channel integration.

Visual consistency refers to the consistency of the relevant and comparable visual attributes, including, visual aesthetics, image, font, order, and complexity, in mobile applications associated with retailers' websites. The similarity of a mobile application to an e-commerce website may increase a customer's uniform and comprehensive experience by enhancing the integrated interactions across channels (Ganesh 2004; Sousa et al. 2006). From a design process perspective, providing the user with consistency of the visual design is the most important factor in mobile application development (Buellingen et al. 2004; Kangas et al. 2005). When the design of an application is simple and operated by in a manner similar to a familiar website, the customer feels that the application is alsofamiliar, and is easy to use. Therefore, we hypothesize:

H3b.In cross-channel electronic commerce, visual consistency of the mobile application is positively associated with perceived quality of channel integration.

3. RESEARCH DESIGN

Instrument Development

For the survey instrument, we identified existing measures that had been repeatedly tested and that possess strong content validity. For example, perceived quality of channel integration is measured by a consistent impression between mobile application and website, integration of the mobile application and website, and combination of mobile application and website as service across channels (adapted for our research domain from Payne et al., 2004 and Sousa et al., 2006)

[Survey instrument and full references available upon request].New items were developed based on the results of our literature review. Constructs are measuredwith a 7-point Likert scale.

Pilot Study Data and Preliminary Data Analysis

We pilot tested our survey to identify any potential problems with the instrument. Of the 29 voluntary participants, 52.7% were male and 47.3% were female. To assess initial reliability and validity, we conducted exploratory factor analysis. All items except two loaded on the expected theoretical construct. Two items exhibit slight cross loading (values of 0.43 and 0.44). Based on the results of the pilot study, those two items are being modified to make them clearer and more straightforward. Reliability and correlations are shown in Table 1.

[Table 1] Reliability and Correlations

	Alpha	Composite Reliability	1	2	3	4	5	6
1. Perceived functional extensibility	.84	.89	(.82)					
2. Perceived functional complementarity	.88	.93	.49	(.90)				
3. Perceived information consistency	.93	.95	.25	.55	(.91)			
4. Perceived visual consistency	.96	.97	.31	.34	.34	(.94)		
5. Perceived website quality	.75	.79	.03	.08	.23	.30	(.76)	
6. Perceived quality of channel integration	.88	.88	.35	.46	.40	.57	.17	(.78)

*The diagonal elements (in bold) represent the square root of the AVE.

Reliability is assessed using internal consistency scores. All values for Cronbach'salpha exceed 0.70, signifying acceptable reliability. In addition, all constructs share more variance with their indicators than with other constructs since all AVEs are well above 0.50.Data collection and analysis is ongoing. We will finalize the measures for constructs based on the results of the pilot study, thenwe will conduct the main survey and analyze the data using SEM.

4. CONCLUSION

This study aims to develop a framework for effectively constructing cross-channel electronic commerce environments by fostering a seamless experience for customers. In this study, we propose that perceived quality of channel integration can be determined by (1) the functional configuration of the mobile applications, (2) the perceived similarity across channels, and (3) the perceived quality of the websites. Our intended theoretical contribution is to develop a research framework extending findings regardingwebsite-based e-commerce into in cross-channel, web and mobile e-commerce. We also intend to provide practitioners with insights for effectively executing integrated, multi-channel e-commerce strategies for their companies.

REFERENCES

Brown, I., and Inouye, D.K., Learned helplessness through modeling: The role of perceived similarity in competence, *Journal of Personality and Social Psychology*, 36(8), 1978, pp.

Buellingen, F. and M. Woerter., Development perspectives, firm strategies and applications in mobile

commerce, *Journal of Business Research*, 57(12), 2004, pp.1402-1408.

Madaleno, R., Wilson, H., & Palmer, R., Determinants of customer satisfaction in a multi-channel B2B environment, *Total Quality Management & Business Excellence*, 18(8), 2007, pp. 915-925.

Mithas, S., et al., Designing web sites for customer loyalty across business domains: a multilevel analysis, *Journal of management information systems*, 23(3), 2007, pp. 97-127.

Oracle, Cross-channel commerce: A consumer research study, in: Oracle Corporation, 2011.

Payne, A., &Frow, P., A strategic framework for customer relationship management, *Journal of marketing*, 2005, pp.167-176.

Rosenbloom, B., Multi-channel strategy in business-to-business markets: prospects and problems, *Industrial Marketing Management*, 36(1), 2007, pp. 4-9.

Sousa, R., & Voss, C. A., Service quality in multichannel services employing virtual channels, *Journal of Service Research*, 8(4), 2006, pp. 356-371.

Varshney, U. and R. Vetter., Mobile commerce: framework, applications and networking support, *Mobile networks and Applications*, 7(3), 2002, pp. 185-198.

Yang, S., et al., Why do consumers adopt online channel? An empirical investigation of two channel extension, *Decision Support Systems*, forthcoming.

Full reference available upon request.

The Effects of firms' resources and value co-creation activities on mobile service innovation

Miri Kim
Sogang University
mirikim@sogang.ac.kr

Jaeki Song
Texas Tech University
jaeki.song@ttu.edu

Jason Triche
Texas Tech University
jason.triche@ttu.edu

ABSTRACT

The current trend shows that the world economy has moved from a product oriented economy to a service oriented economy. Therefore, understanding service innovation provides opportunities for firms to continuously obtain or sustain competitive advantages by offering innovating services. This study aims to contribute to service innovation in mobile industry through a theoretical formulation of a research framework in which we posit how firms implement better services from a dynamic capabilities framework. Our framework has potential to provide new insights by suggesting the basis for new discourse in IS for service innovation.

Keywords: Value co-creation, dynamic capabilities, resource-based view, service innovation

1. INTRODUCTION

Today's organizations consider the ability to develop innovations as one of their top priorities (Thomson Reuters, 2011). Especially, services operate in a complex and dynamic environment with respect to open characteristics of the markets where emphasized linkages between service providers and recipients are important. These actors perform at designing, producing, and delivering processes together for supporting the high quality products and services to the market. Such involvement outside the firm that leads to service innovations are new developments in the core offerings of service companies that tend to create new revenue streams (Oke 2007). In some cases, these new developments can fundamentally change the way a company does business. As innovation plays a critical role in highly competitive markets, the innovation performance has received great attention to both practitioners as well as researchers.

Predominant studies in management literature have demonstrated how outcomes of innovations can be improved. However, we focus on firm's resource and capabilities, which are distinguished by tangible and intangible characteristics in the service sector and dynamic capabilities as a black box of the co-creation process in order to enhance the service offerings. Such market-driven knowledge becomes one of the most critical factors in creating service innovations by integrating firms' heterogeneous resources with their capabilities (Eisenthardt and Martin, 2000). In this study, we aim to provide a framework that explains the effects of firms' resources and market-driven interactivity on their capabilities and, in turn, the impact of capabilities on their outcomes in mobile commerce service innovation.

2. Theoretical Backgrounds

Dynamic Markets in Mobile Industry

According to the 16th Mobile Competition Report, mobile wireless industry is sufficiently complex and competes in dynamic markets in various aspects, such as types of services, geographic areas of the mobile network, etc. (Federal Communications Commission, 2013). To service firms in the mobile market, firms should formulate strategic approaches on how they can provide a better service than their competitors. The highly competitive businesses in the dynamic-technology industries such as telecommunications, information services, and software, challenges the assumption of the resource-based view (RBV) of the firm. To address this issue, we draw on dynamic capabilities by synthesizing resource-based view and value co-creation.

Dynamic Capabilities

The mobile industry is extremely active therefore any static theory, like resource-based view which is suited to stable environments, is not sufficient to explain dynamic phenomenon. Many previous studies have looked at only a single side of determinates for an innovative organization, or else have proposed a hidden aspect of innovation process, such as Kanter'snewstream (Lawson et al. 2001). The Dynamic Capabilities Framework addresses that internal technological, organizational, and managerial processes enable firms to gain economic rents in settings of rapid change (Baker et al., 1997). They have been defined by as "the firm's ability to integrate, build, and reconfigure internal and external competences to address rapidly changing environments" (Teece et al., 1997). Dynamic capabilities have emphasized the critical role of managerial capabilities rather than firm resources (as in the RBV). Dynamic capabilities enable a firm to adjust its strategy and resources to maintain and sustain competitive advantages (Wade and Hulland, 2004). Without such enduring dynamic capabilities, competitive advantage could erode quickly. Thus, proven organizational capabilities are valuable outcomes, such as creating effective innovation.

Resource-based View (RBV)

The resource-based view explains that competing firms possess heterogeneous sets of resources and capabilities (Wernerfelt, 1984). As conceptualized in this theory, valuable, rare, inimitable, and nonsubstitutable resources are the basis of a firm's competitive advantage (Barney 1986). The RBV applies very broad definition of resources, including such items as physical capital, human capital, and organizational capital (Barney, 1991). Capabilities are defined as competencies that are built by combining resources (Grant, 1991). Its resources create distinct strategic choices that could exploit markets in order to improve market position (Song and Zahedi, 2006). The RBV has been applied to explain that shared domain knowledge between business and IT managers help improve the quality of different outcomes (Kearns and Sabherwal, 2006-7), to explain how the strategy of a firm influences its productive interactions with other firms (Madhok, 2002), how applied to firms' capability to address the rapid changes in markets in which firms have operated (Eisenhardt and Martin, 2000), and how information systems competencies affect process innovations within an organization (Tarafdar and Gordon 2007).

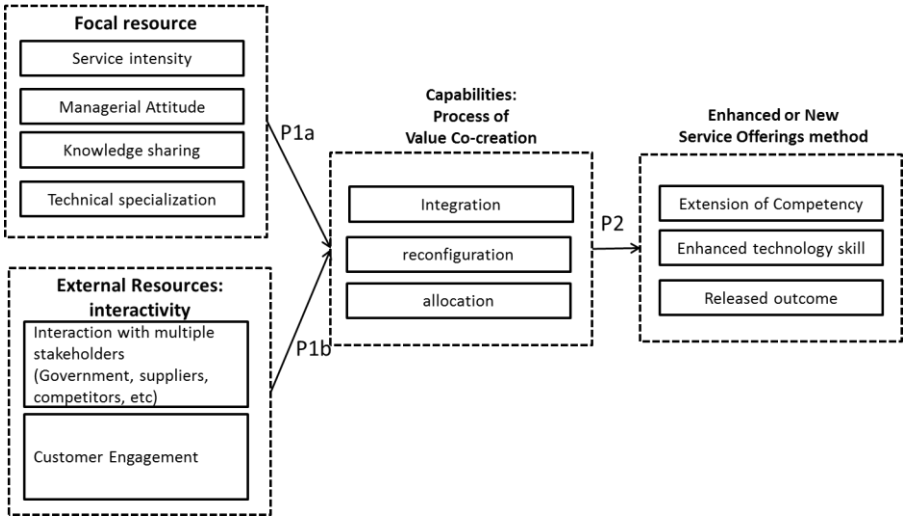
Value Co-creation

Dynamic capabilities emphasize the role of capabilities that react to dynamic external context which is directly related to value co-creation activities with external resources (Ramaswamy and

Gouillart, 2010). In business relationships, value can be examined with customer, suppliers, and other entities including government and even competitors. Especially in the case of customer involvement, the literature suggests that increase of value co-creation activities between a firm and customers provides unique and inimitability of the firm's knowledge and they become new sources of firms' rents (Dierickx and Cool, 1989). In a complex and dynamic environment, ability to see the value creation relationships can lead to obtaining business opportunities because value is dynamic and changes over time (Forsström, 2005). Therefore, a firm's network creates value by enhancing the creation of capabilities through combining the existing capabilities of firms in new value configurations (Ngugi et al., 2010). Furthermore, such value co-creation possibly produces high innovation, effectiveness, and creating more innovative products (Gilbert et al., 2000).

3. Research Framework

In this section, we develop propositions about the relationships between resources and capabilities, and in turn, capabilities and outcomes of service innovation. Our proposition and the links among the critical components are outlined in Figure 1.



[Figure 1] Research Framework

Focal resource and Interactivity

An organizational routine, as defined in the dynamic capabilities literature, is a set of activities that are routinized such that the ability to continuously manipulate key capabilities and resources of their firmsinto new products, processes and systems for successful innovation (Lawson and Samson 2001, Agarwal and Selen 2009). More specifically, innovation capability is about synthesising these two operating paradigms: 1) focal resource and interactivity, and 2) process of value co-creation.

Although there are many heterogeneous resources across an organization, we suggest that they all can be categorized into focal resources and interactivity for co-creation. Focal resources must carry specific characteristics: valuable, rare, inimitable and non-substitutable, and must be

appropriate to the firm (Arend 2006). Within service firms, innovative firms are able to link their core resources such as technical specializations, service intensity, managerial attitude, and knowledge sharing to their capabilities in processing value co-creations (Agarwal and Seldon, 2009; Lawson and Samson 2001; Wu 2010).

According to studies in dynamic capabilities, maintenance of relationships with partners such as suppliers, customers, and other key stakeholders is the fundamental logic of leveraging resources (Agarwal and Selen 2009). Especially, a firms' relationship with its customers through its engagement becomes a co-creation process in the development, design, and delivery of innovative products and services (Agarwal and Selen 2009, Wu 2010). Also, service companies increasingly emphasize the partnerships with suppliers and other business partners (Howells 2000). Therefore, acquiring knowledge and information from their external resources becomes critical (Foss et al., 2011). Given the relationship between focal resources, external resources, and dynamic capabilities, we posit the following:

Proposition 1a: In the context of service innovation, increases in focal resources increase capabilities for producing value co-creation process.

Proposition 1b: In the context of service innovation, increases in external resources increase capabilities for producing value co-creation process.

Capabilities of Service Innovation: Process of Value Co-creation and Their Outputs

Firms that possess dynamic capabilities as integrating, re-configuring, and allocating by linking different levels of resources of firms for service innovation. There are three dynamic capabilities that link different levels of resources for firms that wish to innovate (Eisenhardt and Martin 2000). *Integrative capabilities* allow firms to combine operant sources (i.e., skills and knowledge) from different resources developed by various firms' departments and external resources (Lawson and Samson 2001). *Reconfigurative capabilities* focus on transfer processes including replication and brokering the source within the firm. *Allocative capabilities* are related to distribution and patching the resources for co-evolving to create synergies. Firms build such capabilities to capture market opportunities and enhance or create new service innovation process (Agarwal and Selen 2009).

Firms may pursue a complex of service innovations that may target different type of outcomes. We propose service innovation as implying new or enhanced service offerings and blended the different characteristics through the process of value co-creation (Gallouj and Weinstein 1997). To take the specific characteristics of service, we leverage competency, technology skills, and outcome. *Competency* refers to the basis of unique value-creating in service innovation systems and direct mobilization of own competency (Gallouj and Weinstein 1997, Eisenhardt and Martin 2000). One of the major activities in service is the *technologies* embedded in the form of values (i.e., knowledge and skills) including tangible and intangible technical characteristics processes which directly provide a service innovation. *Outcome* refers to a prior product or service represented by a set of combination of competences and technology characteristics by firm or without non-combination of the characteristics in service innovation system. In terms of a process of dynamic capabilities, various modes of service innovation are highlighted by how

three characteristics formulate in service innovation system (Gallouj and Weinstein 1997). Given the relationship between dynamic capabilities and service offerings, we posit the following:

Proposition 2: In the context of service innovation, increases in dynamic capabilities enhance to offer better services (or new services).

4. Research Design

The context of this study is the mobile service industry. The mobile telecommunication industry is sufficiently complex and competes in dynamic markets in various aspects, such as types of services, and geographic areas of the mobile network. Moreover, mobile services providers supply customers with the ability to connect and communicate over their network, which at its core is the definition of service. We plan to analyze our model using data from the leading providers of mobile communication services in South Korea. In order to explore the data and have a more in-depth analysis, we employ content analysis as a research method. Content analysis is one of the empirical research methods that use existing data analysis methods in order to ensure objectivity and reliability. In addition, using content analysis allows us to explore the data, posit deeper insights, and provide a systematic procedure to discover richer, more textual information in the service innovation variables.

5. Conclusion

Our study provides potential benefits to academics in addition to businesses. Our framework sheds light on innovation literature by synthesizing dynamic capabilities with resource based view and value co-creation activities. Also, our framework is a testable model so that it provides solid foundation of theory in the literature. Our study also has some implications to managers. Mangers could understand how they offer better services in terms of determining how to integrate and allocate their capabilities that are associated with their internal resources. In addition, firms can learn how to reconfigure external resources to maximize value creation from external resources.

REFERENCES

Agarwal, R. and W. Selen., Dynamic capability building in service value networks for achieving service innovation, *Decision sciences*, 40(3), 2009, pp.431-475.

Baker, J., Jones, D., Cao, Q., and Song, J., Dynamic Strategic Alignment Competency: A Theoretical Framework and An Operationalization,*Journal of the Association for Information Systems*, (12:4), 2011, 229-328.

Barney, J.B., Types of competition and the theory of strategy: toward an integrative framework, *Academy of Management Review*, 11(4), 1986, pp. 791–800.

Eisenhardt, K.M., and Martin, J.A., Dynamic capabilities: What are they?,*Strategic Management Journal*, 21, 2000, pp. 1105-1121.

Foss, N. J., et al., Linking customer interaction and innovation: The mediating role of new organizational practices, *Organization Science*, 22(4), 2011, pp. 980-999.

Gallouj, F. and O. Weinstein, Innovation in services, Research policy, 26(4), 1997, pp. 537-556.

Ramaswamy, V., and Gouillart, F., Building the Co-Creative Enterprise, *Harvard Business Review*, October

2010, pp. 100-109.

Song, J., and Zahedi, F.M., Internet Market Strategy: Determinants and Implications," *Information & Management*, (43:2), 2006, pp.222-238.

Teece, D.J., Pisano, G., and Shuen, A., Dynamic Capabilities and Strategic Management, *Strategic Management Journal*, 18 (7), 1997, pp.509–533.

Wade, M., and Hulland, J., Review: The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research, *MIS Quarterly*, (28) 1, 2004. pp. 107-142.

Wernerfelt, B., A Resource-Based View of the Firm, *Strategic Management Journal*, (5) 2, 1984, pp. 171-180.

Wu, L.-Y., Applicability of the resource-based and dynamic-capability views under environmental volatility, *Journal of Business Research*, 63(1), 2010, pp. 27-31.

Full reference available upon request.

Exploring the Impact of Social Capital Investment on Doctor's Returns on eHealth Website

ShanShan Guo
School of Management, Harbin Institute of Technology, P .R. China, 150001
E-mail: hit.guoshanshan@gmail.com;
Xitong Guo
School of Management, Harbin Institute of Technology, P .R. China, 150001
E-mail: xitongguo@gmail.com

Abstract: Noteworthy is that, although the use of eHealth websites has been perceived to be valuable for patients, there is limited research that has focused on the social behavior of doctors who are willing to use eHealth websites. Our research examines how the doctors' online efforts and offline attributes impacts the benefits brought from eHealth websites to doctors. As social capital investment will bring social capital returns, the target of this paper is identifying which domains of doctors' social capital investments in eHealth websites, referred to as online efforts and offline attributes domains, lead to which dimension of social capital returns, social returns and economic returns dimensions. Regression results show that the two domains of social capital investments did have significant different effects on the two dimensions of social capital returns. A reasonable explanation about the four types of doctors, defined with differences betweentwo dimensions of social capital returns, will concluded in this study.

1. INTRODUCTION

Nowadays, there are numerous medical demands compared with the dearth of health resourcesto be deal with them, which has become a global problem. In order to solve this problem, information technology has been brought into the health care industry. Research has started to promote the benefits of eHealth websites. Yang and Tan(Yang & Tan, 2010)claim that an online health community provides various kinds of social support,especially for helping patients to improve their health conditions. In fact, only when doctors and patients, the two stakeholders of eHealth websites, are in a win-win situation, will the developed eHealth websites come to their full potential. Noteworthy is that, although the use of eHealth websites has been perceived to be valuable for patients, there is limited research that has focused on the social behavior of doctors who are willing to use eHealth websites. As we all know, doctors in the healthcare context belong to the donors. For this reason, it is essential to explorehoweHealth websites can add to the benefit of doctors.

Our research examines how the doctors' online efforts and offline attributes impacts the benefits brought from eHealth websites to doctors. Similar to physical retailers who are willing to open up their online business, such as Barnes & Noble, doctors who are willing to do online consulting are also like running e-commerce on eHealth websites. Though physical retailers and doctors look the same, they are completely different. First, from the perspective of forms of platform, besides e-commerce websites, eHealth websites have the characteristics of social media(Hawe & Shiell, 2000), which contains communication between doctors and patients, patients and patients.In consequence,research abouteHealthwebsites should fusion both in the field of e-commerce and social media； Second, from the perspective of participants, Doctors areindividuals and entities shopsare organizations composed of more than one person. Hence, doctors are different from physical

shops in behavioral pattern. The doctorsobey individual behavior principles, such as a self-fulfilling intend, etc. Above all, the motivation of doctors usingeHealthwebsitesis not completely the same to the entities stores e-commerce driven by interests. So we will study the individual doctors'online behavior and offline attributes in eHealth websites.

2. CONCEPTUAL ARGUMENTS

Social behavior motivation theories point out that if people want to benefit from a relationship, they have to invest in it (Lin, 1999). As to the form of investment, apart from the physical and mental form, it also includes attempts at the opportunity of communication (Verbrugge, 1977). Similar to the real world, there is capital in the social network called social capital, which isdefined as resources embedded in a social structure which are accessed and mobilized in purposive action (Comet, 2007; Lin, 1999). The social capital theory pointed out that social relationshipamong people could be productive resources (Coleman, 1988). In this study, we verify the speculation that social capital could be the reason for the motivation of doctors'knowledge sharing through eHealth websites(Yli‐Renko, Autio, & Sapienza, 2001).

Furthermore, Lin (1999)declares that the processes of social capital are investment in social capital, access to and mobilization of social capital and returns of social capital. In addition, the returns following investments in social capital can be classified as economic, political and social returns. Particularly, on the eHealth websites, there are social capital returns of economic and social returns for doctors. However, doctors rarely obtain political returns. In this study， we examine the two dimensions of social capital returns based on doctors' social capital investment in eHealth websites, which are economic returns and social returns.

The first dimension of social capital returns, economic returns, refers to the range of forms which can bring direct reality economic benefit(Chumbler, Mann, Wu, Schmid, & Kobb, 2004). Here we only consider the direct economic returns for the doctors from eHealth websites, and do not contain the indirect economic returns such as benefit brought by the increase quantity of offline clinic patients. These economic benefits mainly include the doctors' diagnosis and treatment online incomes and telephone therapeutic benefits(Rumberger & Dansky, 2006). That is to say, on eHealthwebsites, the doctor's economic returns are in the form of revenue from websites. The second dimension refers to social returns, which belongs to virtual returns. Social returns vary in aspects such as virtual thanks letter, online votes from patients, and virtual gifts. These social returns for the doctor from eHealthwebsites is a reflection of a kind of glory. Economic and social returns signify two different dimensions of social capital returns. As Table 1 illustrates, these two dimensions capture four different scenarios of doctors in eHealth websites.

Table 1 Doctors' Performance Across Social Capital Returns

		Social Returns	
		Low	High
Economic Returns	Low	Doctors who can't get both enough wealth and glory from eHealth websites	Doctors who prefer glory to wealth brought by eHealth websites
	High	Doctors who prefer wealth to glory brought by eHealth websites	Doctors who can get both enough wealth and glory from eHealth websites

3. RESEARCH MODEL AND HYPOTHESES

The impact factors of doctors' social capital investment span two domains, referred to as online efforts and offline attributes domains. While the online efforts domain entails the actions and efforts of doctors on eHealth websites which is dynamic investment. The offline attributes entails the originalattributes owned by

doctors. Though both the domains are important, the nature of the factors is different and revolves around professionals with different efforts. Therefore, the benefits brought by these two factors are also distinct and different between these two domains. In this research, we assess the impact of the online effort and offline attributes of doctors on the two dimensions of doctors' online social capital returns.

The target of this paper is identifyingwhichdomains of doctors' social capital investments in eHealth websites lead to which dimension of social capital returns. As mentioned before, the two dimensions of social capital returns capture four different kinds of doctors in eHealth websites. Different kinds of doctors are likely to have different kinds of social capital investment structures. To illustrate, we define Group 1 represents high social returns & high economic returns doctors; Group 2 represents high social returns & low economic returns doctors; Group 3 represents low social returns & low economic returns doctors; Group 4 represents low social returns & high economic returns doctors. This leads us to propose our first two hypotheses:

H1a. For Group 1, both online efforts and offline attributes will be positively associated with social returns.

H1b. For Group 1, only offline efforts will be positively associated with economic returns.

Similarly, for Group 2, we argue that, online efforts are more important for social capital returns gaining.

H2a. For Group 2, only online efforts will be positively associated with social returns.

H2b. For Group 2, only online efforts will be positively associated with economic returns.

For Group 3, the economic and social returns are less. We put forward our third two hypotheses:

H3a. For Group 3,both online efforts and offline attributes will be positively associated with social returns.

H3b. For Group 3,only online efforts will be positively associated with economic returns.

For Group 4, our last two hypotheses are shown below:

H4a. For Group 4,only offline attributes will be positively associated with social returns.

H4b. For Group 4,only offline attributes will be positively associated with economic returns.

4. DATA AND METHODS

For the data, we will use the Good Doctor(www.haodf.com) as our research object. In China，the Good Doctor(www.haodf.com) is the largest eHealthwebsite which includes 3,233 regular hospitals and 303,367 doctors. This study crawled public data of the website through the network spiders on Sep 6, 2013,and the data consisting of numerical information (e.g., fees of telemedicine, the number of papers doctors share, the number of replies to patients, the number of gifts, and recent online time). The observations are all the online doctors belong to coronary heart disease, and the total number of doctors is 1788. To verify the four hypotheses we raised above, quantitative research method, multiple linear regression between groups,will beused.

4.1 Dependent Variables

The dependent variables in this study are social capital returns of doctors. For Good Doctor, the social capital returns are telephone consulting incomes, virtual gifts, thanks letters and votes. K-means cluster was used to make sure which dimension of social capital returns that variables belong to. According to the principle of separabilityand interpretability, final number of clusters is four, which also match our speculation of four kinds of doctors. It was presented in Two-dimensional quadrant method and shown in Figure 2. The front size roughly indicates the amount of observations.

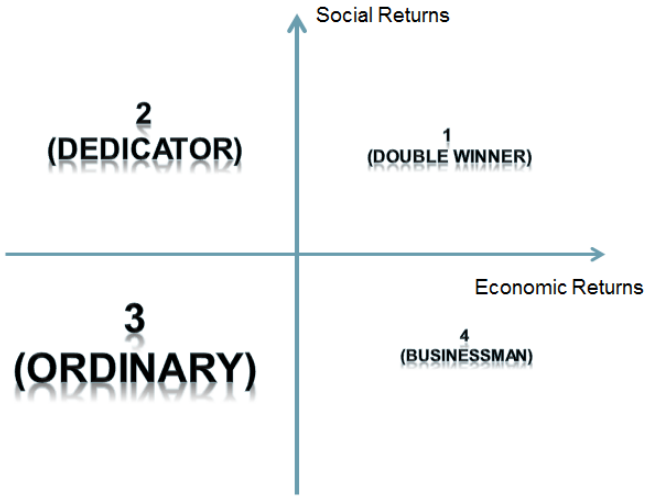


Figure 1.Two-dimensional quadrant of doctors

4.2 Empirical Model

We estimated the following regression model between four groups of doctors. In addition, imaginary lines indicate the moderating effect between online efforts and offline attributes which have not hypothesized yet. However, we will consider the moderating effects in when we do further data analysis (There are speculations that *Online Efforts* significantly impact social returns and *Offline Attributes* act as moderator, on the contrary,*Offline Efforts* significantly impact economic returns and *Online Attributes* act as moderator).

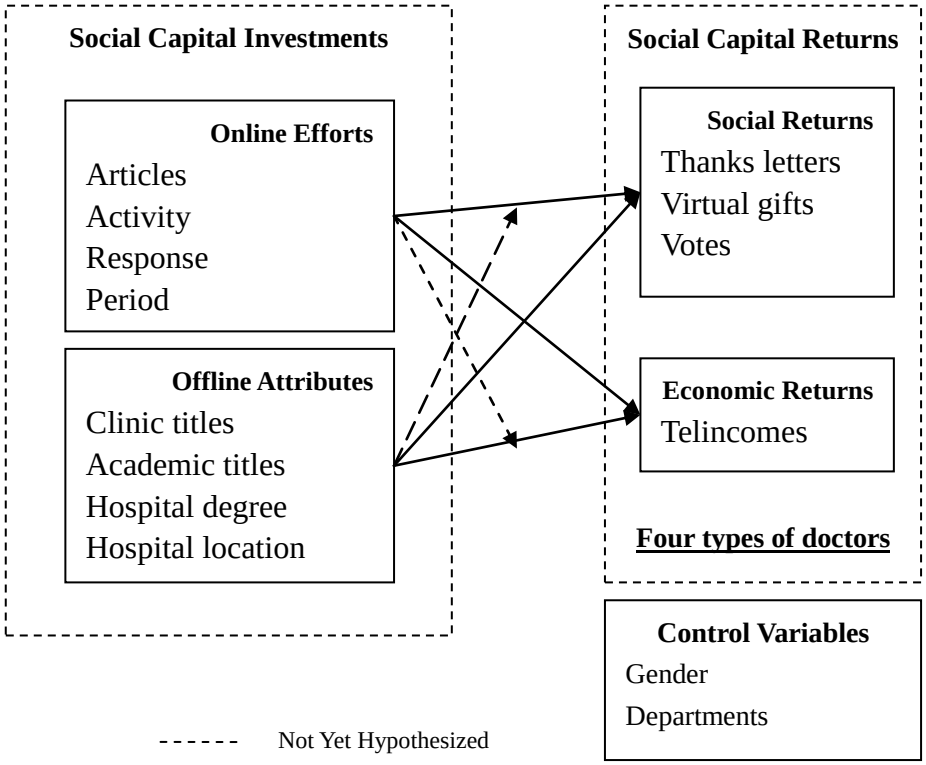


Figure 2. Empirical Model

The expression of multiple liner regression is shown in equation 1. For *SR* and *ER* it represented for *Social Returns* and *Economic Returns*, respectively.

$$\begin{cases} SR_g = \beta_{s,g} + \sum_{i=1}^{I} \beta_{s,i} Online_{i,g} + \sum_{j=1}^{J} \beta_{s,j} Offline_{j,g} + \varepsilon_{s,g} \\ ER_g = \beta_{e,g} + \sum_{i=1}^{I} \beta_{e,i} Online_{i,g} + \sum_{j=1}^{J} \beta_{e,j} Offline_{j,g} + \varepsilon_{e,g} \end{cases}, g \in G(1,2,3,4) \quad (1)$$

In the expression, g comes from sets G, which contents four types of doctors;i comes from sets I, which contents four independent variables belong to online efforts;j comes from sets J, which contents four independent variables belong to offline attributes.

5. CONCLUSION AND DISCUSSION

Regression results show that the two domains of social capital investments did have significant different effects on the two dimensions of social capital returns. According to the hypotheses, it can be indicated that the effect of the different structuresof social capital investments between different groups of doctors should be different in social capital returns.

The purpose of this paper is to discuss the model of doctor'ssocial capital investment and return on eHealth websites. Furthermore, behavioral suggestions to doctors fully usable of eHealth websites andthe operational suggestions to online eHealth websitesare given. More importantly, this study discover the particular eHealth websites which integratee-commerce sites and social media sites together, which provides a new research target for the network and social capital research.

References

Chiu, Chao-Min, Hsu, Meng-Hsiang, & Wang, Eric TG. (2006). Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories. Decision support systems, 42(3), 1872-1888.

Chumbler, Neale R, Mann, William C, Wu, Samuel, Schmid, Arlene, &Kobb, Rita. (2004). The association of home-telehealth use and care coordination with improvement of functional and cognitive functioning in frail elderly men. Telemedicine Journal & E-Health, 10(2), 129-137.

Coleman, James S. (1988). Social capital in the creation of human capital.American journal of sociology, S95-S120.

Comet, Catherine. (2007). Ronald S. Burt: Brokerage and Closure: An Introduction to Social Capital. European Sociological Review, 23(5), 666-667.

Deutsch, Morton, & Gerard, Harold B. (1955).A study of normative and informational social influences upon individual judgment.The journal of abnormal and social psychology, 51(3), 629.

Hawe, Penelope, &Shiell, Alan. (2000). Social capital and health promotion: a review. Social science & medicine, 51(6), 871-885.

Lin, Nan. (1999). Building a network theory of social capital. Connections, 22(1), 28-51.

Rumberger, Jill Schumann, &Dansky, Kathryn. (2006). Is there a business case for telehealth in home health agencies? Telemedicine Journal & e-Health, 12(2), 122-127.

Verbrugge, Lois M. (1977). The structure of adult friendship choices. Social Forces, 56(2), 576-597.

Yang, L, & Tan, Yong. (2010). An empirical study of online supports among patients.

Yli‐Renko, Helena, Autio, Erkko, &Sapienza, Harry J. (2001). Social capital, knowledge acquisition, and knowledge exploitation in young technology‐based firms. Strategic management journal, 22(6‐7), 587-613.

Ambidexterity in Software Product Development: An Empirical Investigation

Kannan Mohan, Baruch College
Balasubramaniam Ramesh, Georgia State University

Abstract: Software product development organizations that are involved in the development of a large number of related (family of) products face considerable challenges in balancing the need for product variety with schedule, cost, and quality constraints. While the organizations that focus on rapid development of individual software products sacrifice the ability to expand their product portfolio, those that follow a platform-based approach sacrifice a short development cycle, the ability to meet unique needs of customers, and cost efficiencies for the initial set of products. In this paper, we examine how organizations can simultaneously pursue such conflicting demands for schedule, cost, quality, and product variety. Based on a case study, we identify practices that help software development organizations gain ambidexterity and enable them to pursue conflicting demands. The findings are summarized in a conceptual framework, which highlights the contexts and conditions under which organizations face conflicting demands and the practices followed to achieve ambidexterity.

1. Introduction

Software product development organizations are increasingly cognizant of the tradeoffs involved in achieving often conflicting needs for quick time-to-market, quality, costs, and providing sufficient product variety. Organizations attempt to address this conflict by following Product Family Development (PFD) approaches which involve the development of a flexible product platform that can be used to derive numerous, customized end products [1]. While upfront investment required to produce the platform in terms of development time and resources is likely to delay the delivery of the initial set of finished products, the derivation of product variants from the platform requires considerably shorter cycles and minimal costs [6]. However, the need to develop software products rapidly [2] is not consistent with long term investments in the development of a product platform. Therefore, these organizations face a critical dilemma – Should they focus on rapidly developing an initial set of software products to gain first or early mover advantage or invest significantly into developing a product platform which will provide the ability to derive numerous products that meet diverse customer needs? In this paper, we examine how organizations can pursue these conflicting demands simultaneously.

Organizational capability to simultaneously pursue conflicting demands such as exploitation and exploration has been conceptualized by past research as organizational ambidexterity [4, 5, 7]. In this research, we identify ambidextrous practices that enable software development organizations to pursue conflicting demands along multiple dimensions, viz. product variety, cost, quality, and schedule. The key research question answered in this paper is: "*How can software product development organizations become ambidextrous by balancing conflicting demands on product variety, cost, quality, and schedule?"*

2. Software Product Family Development

A software product family is a set of software-intensive systems that share a common, managed set of features [1]. The development of software product families demands a well-

disciplined process that facilitates understanding and controlling common and distinguishing characteristics of products in the family [8]. The choice between the single systems development approach (i.e., developing each product separately) and the product family development approach is quite complex because each approach has both significant strengths as well as shortcomings as highlighted in Table 1.

		Single system development approach		**Product family development approach**	
		Early set of products	*Subsequent set of products*	*Early set of products*	*Subsequent set of products*
Schedule		Short	Longest	Long	Short
Cost		Low	Highest	High	Low
Quality	Meeting common customer needs	Low	Low	High	High
	Meeting unique customer needs	High	High	Low	Low
	Extensibility	Low	Low	High	High
	Maintainability	Low	Low	High	High

Table 1: Problems in balancing single system and product family development approaches

3. Organizational Ambidexterity

Organizational ambidexterity refers to the pursuit of conflicting demands such as exploration and exploitation [5]. The development of appropriate processes and systems that can be used to strike a balance between opposing demands is increasingly recognized as critical for organizational success [4]. We conceptualize ambidexterity in software product development as the ability to achieve a balance among conflicting demands on product variety, schedule, cost and quality by using a mix of single-system and product family-based approaches.

4. Research Methodology

Since our approach is exploratory rather than confirmatory, we use a case study design [9]. Our study seeks to develop a conceptual framework from the case study [3] to explain how software development organizations achieve ambidexterity in managing conflicting demands. We selected two software product development organizations (hereafter referred to as SCCo and ECCo) that deal with the phenomenon of focus in this research. SCCo developed warehouse management systems and ECCo was focused on developing electronic control units. Both were using a product family approach. Qualitative data were collected from the two sites progressively via semi-structured interviews and a review of documentation and development processes. We first conducted within-case analysis for each of the cases and then did cross-case analysis to compare them. The theoretical concepts and relationships discovered are included in the framework described in the next section.

5. Findings

Figure 1 depicts our research framework depicting the context, conditions, and practices that are explained earlier in this section.

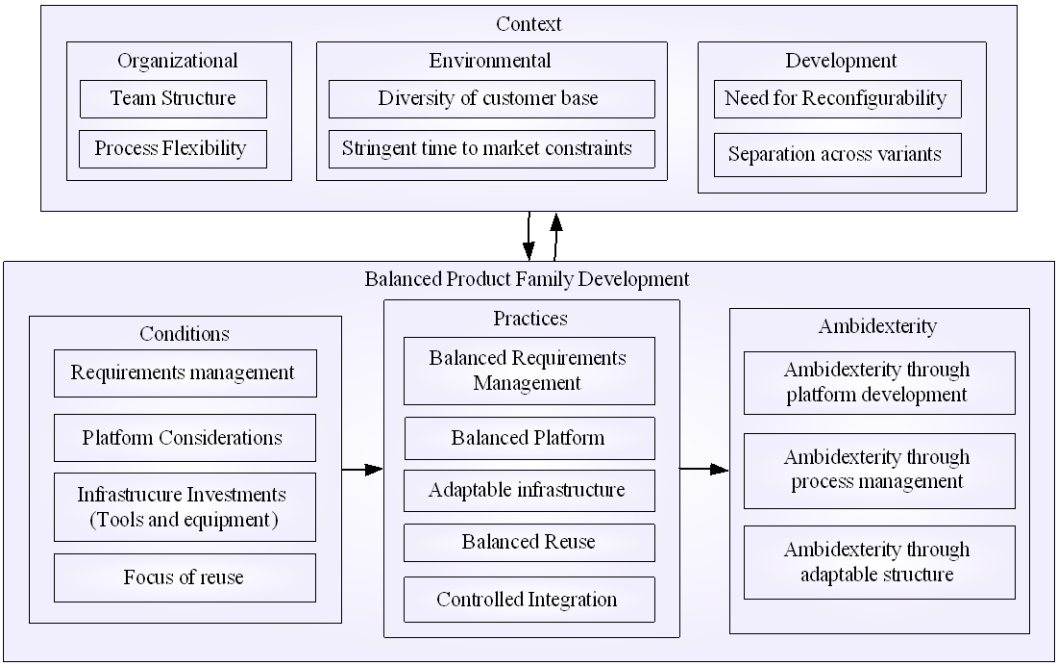


Figure 1: Research Framework

5.1 Context

Both organizations had a team structure that separate domain engineering activities from application engineering. The former, which involved the development and maintenance of the was considered critical because the quality of the design of the platform has tremendous impact on the health of the product family. While both organizations recognized the need for flexibility in the processes used, ECCo followed a more rigid approach in handling new requirements and in incorporating changes to the platform.

5.2 Conditions

The critical dilemma faced by both organizations in requirements management relates to the extent to which they should focus on customer-specific requirements vs. common requirements in the domain. While both organizations recognized the importance of the upfront development of a comprehensive platform architecture that can handle the needs of most customers, they also had to deliver products under stringent time to market constraints. Both sites recognized the need for planned reuse while also realizing the importance of opportunistic reuse of some of the functionalities that are not included in the product platform but may be common among some customers.

5.3 Practices

Both study sites had adapted their practices to strive for a balance between architecting a stable and adaptable platform that can cater to the needs of a variety of customers while simultaneously developing the ability to develop product variants quickly. The motivations for seeking such a balance was associated with organizational, environmental and development conditions that dictated the need for a different approach to developing the family of products.

5.3.1 Balanced Requirements Management Practices

Recognizing that a comprehensive domain analysis involves significant investment in terms of both time and other resources which may delay time-to-market and increase the cost of initial set of products, SCCo decided to carry out this activity in an incremental fashion. It initially restricted domain analysis to a single industry. Only after the development of the initial products for this industry had progressed significantly, the scope of the analysis was expanded to include other industries. While ECCo also followed an incremental process for understanding its domain, it focused on incrementally developing a hierarchy of product platforms that catered to a variety of customer segments.

Recognizing the difficulty in meeting all the requirements of a diverse set of customer segments, SCCo decided to focus on the core functionalities desired by most of its customers. Therefore, domain analysis was geared towards encompassing functionality that was required by a majority (and not necessarily all) customers. This practice helped achieve a quick consensus on the critical requirements that will be initially implemented in the platform.

5.3.2 Balanced Platform practices

Both organizations worked very diligently to identify variation points (which are locations in the design where product variants will differ from each other). For example, at SCCo, different customers required different types of mechanisms to transfer data between their ERP and the warehouse management systems. The platform was designed in such a way that these options could be exercised during product derivation. During the development of a product family at SCCo, it was realized that the comprehensive development of the platform will delay product launch beyond what was acceptable to the marketing team. In order to take advantage of a market opportunity, SCCo management allowed customer-specific customizations of the platform itself.

Both organizations recognized the need to evolve the platform in multiple directions to support different groups of customers. Therefore, instead of just creating one platform from which all products will be derived, a family of platforms was derived from the base platform. Both organizations, however, recognized the need to carefully manage the evolution of their platforms.

5.3.3 Adaptable Infrastructure

When significant investments are required to develop or acquire the infrastructure in the development of product variants that are specific to a customer, both organizations persuaded the customers to invest in and maintain that infrastructure. This helped minimize not only the investment involved, but also the complexity and expertise required in the acquisition and maintenance of the infrastructure. Thus, both organizations were able to focus on the common infrastructure that is required across all (or most) members of the family, rather than on customer-specific requirements.

5.3.4 Balanced Reuse practices

While the benefits of incorporating functionality (and achieving planned reuse) through product platform were well recognized, both organizations were also careful not to inundate the platform with features that are not applicable to a significant majority of their clients. In fact, the platform development team was shielded from individual product development teams so that

itcould independently evaluate the merits of including any functionality into the platform without getting influenced by political pressures.

SCCo also attempted to reuse variant artifacts across multiple clients. While typical product platform development views variant artifacts as unique to specific product variants, SCCo tried to opportunistically reuse variant components that were incorporated in other product variants.

5.3.5 Controlled Integration

While communication with customers was considered critical, clear boundaries were established for restricting the overlaps between ECCo-developed artifacts and customer-developed artifacts. Since its customers often integrated their own modules into the product, ECCo had to ensure that these artifacts are kept separate so that they do not 'contaminate' the platform or the variants developed at ECCo. In contrast, SCCo facilitated considerably stronger integration, while still maintaining boundaries across the products.

Application engineering (product derivation) and domain engineering (platform development) teams were managed through a tightly controlled team structure. While separation of these teams was considered critical to maintain the stability of the product platform, mechanisms were put in place to enable communication across these teams at SCCo.

5.4 Ambidexterity

1. *Ambidexterity through platform development:*Evidence from our study points to the notion thatambidexterity is achieved by developing a small platform initially that addresses the common needs of a few customers and then incrementally evolving the platform to increase its coverage and by retaining the flexibility to derive numerous product variants.
2. *Ambidexterity through process management:* Managing the development process such that it can be easily adapted to accommodate different levels of product variety and dynamic customer needs can be conceptualized as another dimension of ambidextrous behavior.
3. *Ambidexterity through adaptable structure:* Ambidexterity may be achieved in software product development by creating an organization that is structured into domain and application engineering teams, and by structuring the communication across these teams.

6. Cross-case Comparison

While both SCCo and ECCo shared the same objectives in that they focused on developing a product family with a common platform that is shared across product variants, they demonstrated considerably different approaches to developing the family and to addressing the conflicting demands that arose in developing an adaptable platform as well as catering to unique customer needs. SCCo followed a bottom-up approach in that they focused initially on a limited set of features in the product platform selectively by focusing on one industry. They then expanded the platform incrementally to accommodate the needs of various industries leading to a versatile platform that was adaptable in such a way that product variants could be developed efficiently and could handle customer-specific requirements. ECCo, on the other hand, took a top-down approach in that it established the coverage of the platform so that it can accommodate the needs of a range of customer segments. While this required additional upfront investment and resources, they focused more on the stability of the platform and shared the notion that upfront

investment in developing a well-architected platform could lead to advantages during product derivation. Also, their product family was considerably more mission critical because it could impact on road performance of automobiles. This prompted ECCo to carefully develop a high quality platform that will enable them to derive product variants more efficiently than SCCo. Also, while SCCo could afford to use considerably complex design techniques to incorporate flexibility in the product platform, ECCo was limited to certain techniques due to various system constraints.

7. Contributions to Theory and Practice

Findings from our study contribute to the conceptualization of ambidexterity in the context of software PFD. While past research has extensively discussed ambidexterity in a general organizational context, our study is one of the first attempts at understanding ambidexterity in the context of managing the software product development process. Our research is also unique in its attempt at empirically linking balanced product family development practices to the theory of organizational ambidexterity. The balanced practices that are identified in our framework are novel to the literature on ambidexterity that identifies important antecedents for ambidexterity. Our findings add to the conversation in the literature on organizational ambidexterity that focuses on the various approaches to achieve ambidexterity (structural vs. contextual ambidexterity) and the systems and processes that need to be put in place to achieve ambidexterity. Our findings suggest that while a partially structural approach works well in PFD projects, the collaboration between the structures (domain and application engineering) is more pronounced and carefully managed.

Findings from our study contribute to software development organizations that face significant challenges in managing variety under quality, cost, and schedule constraints. Stakeholders involved in software product family development can use the specific practices discussed in this research to implement a balanced product family approach to software development.

8. References

1. Clements, P., and Northrop, L. *Software product lines: Practices and patterns*. Sei series in software engineering, Upper Saddle River, NJ: Addison-Wesley, 2002.
2. Cusumano, M.A., and Yoffie, D.B. Software development on internet time. *IEEE Comput.*, 32, 10 (1999), 60-69.
3. Eisenhardt, K.M. Building theories from case study research. *Acad. Management Rev.*, 14, 4 (1989), 532-550.
4. Gibson, C.B., and Birkinshaw, J. The antecedents, consequences, and mediating role of organizational ambidexterity. *Acad. Management J.*, 47, 2 (2004), 209.
5. Gupta, A.K.; Smith, K.G.; and Shalley, C.E. The interplay between exploration and exploitation. *Acad. Management J.*, 49, 4 (2006), 693-706.
6. Schmid, K., and Verlage, M. The economic impact of product line adoption and evolution. *IEEE Software*, 19, 4 (2002),
7. Tushman, M.L., and O'Reilly Iii, C.A. Ambidextrous organizations: Managing evolutionary and revolutionary change. *California Management Rev.*, 38, 4 (1996), 8.
8. Weiss, D.M., and Lai, C.T.R. *Software product-line engineering: A family-based software development process*. Reading, MA: Addison-Wesley, 1999.

9. Yin, R.K. *Case study research: Design and methods*. Third ed., Applied social research methods series, ed. Leonard Bickman and Bebra J. Rog, Vol. 5, London: Sage Publications, 2003.

An Integration Process Model for Supervisor Recommendation Services

Mingyu Zhang[1,2], Jian Ma[2], Chen Yang[1,2], Hongbing Jiang[1,2], Zhiying Liu[1], Jianshan Sun[1,2]
[1]School of Management, University of Science and Technology of China, Hefei, China
[2]Department of Information Systems,City University of Hong Kong, Hong Kong SAR, China
simileyu@mail.ustc.edu.cn, isjian@cityu.edu.hk, yangc0201@gmail.com,
jhbmx@foxmail.com, liuzhiy@ustc.edu.cn, sunjs9413@gmail.com

Abstract

Effectively identifying suitable research supervisors through mining massive amount of related information is a vital and challenging taskwhich requires lots of efforts and knowledge. The underlying reason of the phenomenon is that the information proliferation and disorientation have brought significant challenges on finding relevant information.Current approaches to identify research supervisors mainly focus on assessing quality of researchers and neglectsocial connections. Furthermore, they paid more attention to objective factors while scant attention to subjective information. To overcome these deficiencies, athree-dimensional frameworknamed RAF-E (research analytics framework for education) is raised to integrate objective and subjective information for personal profiling process and personalized supervisor-recommendation services. A prototype system is implemented to provide the student-supervisor recommendation services via our proposed approach.

1. Introduction

Finding research supervisors for new research students is very important in the education field. First, it influences many people. For example, there are almost 560 thousands new Ph.D. and master students in China every year. As new research students, most of them are confused about how to find a suitable research guider for themselves. Second, solid evidences have shown that match/mismatch with supervisors has a bearing on students' achievements[1-3]. For most of research students, when they begin to pursue a research degree, they oftenhave little understanding of scientific research. They need proper guidance. Design a method to provide services of selecting an appropriate research guider benefits new research students.

Traditionally, the decision of selecting a supervisor is made by students themselves based on their limited experience, where asymmetric information exists and leads to mismatching problems.Actually, there is much mismatch between students and supervisors, which can be mostly attributed to the thoughtless selections done by students and the random allocations of supervisors made by universities' officers after students'entrance. Consequently, it is a potential demand to identify a novel method to offer supervisor-recommendation services.

In order to bridge the gap, we propose a novel approach named RAF-E, combining the advantages of content-based method and collaborative filtering method[4]. It is based on a

systematic research analytics framework (RAF)[5] consisting of three dimensions, *Relevance*, *Connectivity* and *Quality*.We develop a process model to build the supervisors' and students' profiles, integrating subjective and objective information.

We make two contributions in our work. Firstly, we integrate subjective and objective information to construct a three-dimensional framework applying to recommend supervisors. And it has been exploited on Scholarmate[1] to do student-supervisor recommendation services intelligently. Secondly, we develop a process model for our proposed recommendation strategy. It takes comprehensive factors into consideration.

Subsequent sections develop the detailed demonstration further. In section 2, we describe the literature review. In section 3, we show the details of our proposed method. Subsequently, we introduce the system implementation in section 4. Finally, we make a conclusion and point out the limitations. And we alsoput forward the future research work in section 5.

2. Literature Review

Numerous articles have indicated that the students will be more likely to succeed when their research interests are matched with selected supervisors' expertise[3]. And the improper selection may cause students' demotivation, negatively influencing their academic achievements[2]. And "style war" [6] will occur when the compatibility between the supervisor and the student is poor.

Furthermore, there are various literature concentrating on methods of supervisor-selection, such as multi-criteria-decision making[7], analytical hierarchy process[8], and analytic network process[9], which mainly deal with limited candidates. Unfortunately, they merely emphasize objective factors and overlook subjective factors. Furthermore, they stress supervisors' quality evaluationwithout considering comprehensive factors from both sides. Moreover, they cannot provide personalized recommendations for individuals. In addition, although recommendation techniques have been widely applied to do people-to-people recommendations, there isa specific group whose demands have been neglected. As new research students are not familiar with scientific research, the generic methods are not appropriateforsatisfying their specific demands [10].

Since the inception of web2.0, people are facing too much information to make an optimized decision.As a wide range of technological applications are growing at an unprecedented rate, it is a chance to make good use of information on social network to provide supervisor-recommendation services for new research students.

3. Proposed Method

To illustrate clearly, an overview of our proposed method is shown in Figure 1.

[1]www.scholarmate.com

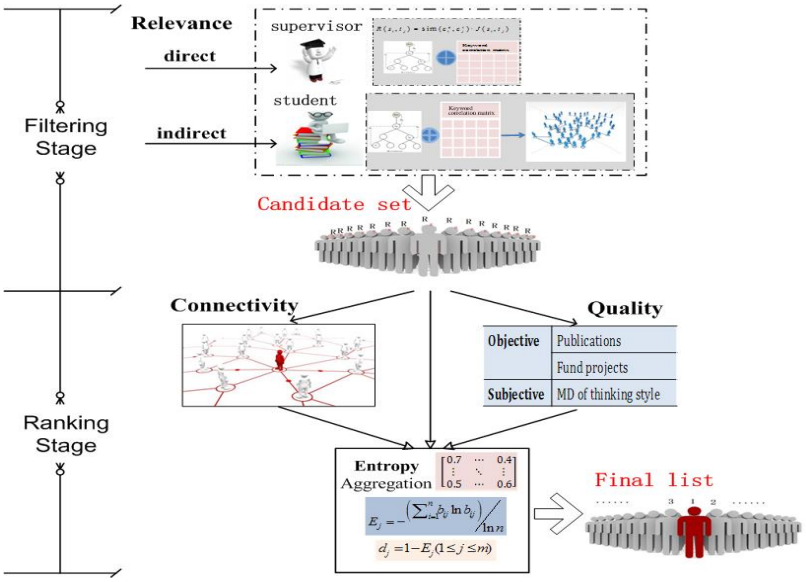


Figure 1.Framework ofproposed method

There are three dimensions: *Relevance*, *Connectivity*, *Quality* and two stages: *Filtering* and *Ranking*. In *Filtering* stage, *Relevance* is utilized to obtain the candidate set, which plays the role of filtering irrelevant subjects by directly matching the given student with supervisors and indirectly matching with supervisors' pre-existing students. In*Ranking*stage, the candidates will be also measured from *Connectivity* and *Quality* dimensions. Finally,scores obtained from three respective dimensions will be aggregated by the method of entropy [11] and the final recommended list will be obtained.

Figure 2 denotes the process model of our proposed recommendation strategy. At the beginning of the recommendation process, we will collect the personal subjective and objective information. Subsequently, there are two ways to measure the *Relevance* score and the initial candidate set will be obtained. Oneway is matching the target student with supervisors, which is called direct matching. And the other is matching the target student with supervisors' pre-existing students, which is called indirect matching. Undoubtedly, the candidates will also be measured by the *Connectivity* and *Quality* dimensionsrespectively. There are two sub processes: social connectivity filtering and quality ranking. After above processes, we get a *Relevance* score, *Connectivity* score and *Quality*score which have to be aggregated in the next step. The method of entropy [11] will be employed to aggregate the scores from three dimensions. And the final users' comments can be treated as verification of the recommendation method.

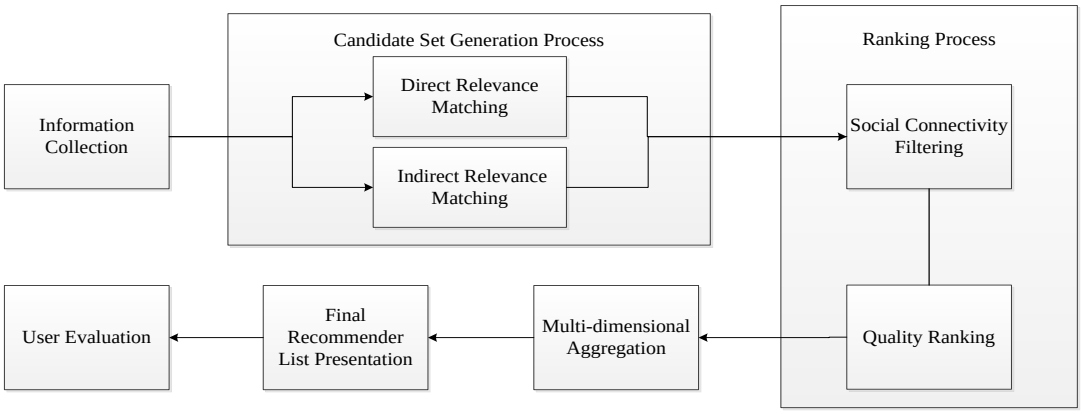


Figure 2.The profiling process model for supervisor recommendation

As shown in Figure 3, we combine objective and subjective information during both the profiling and similarity calculation processes. A student's objective information includes his major, taken courses and read references which can be collected in the folder of "References" on Scholarmate. And a supervisor's objective information contains his department in university, taken courses, publications and projects. The information mining from virtual social network is also objective. In the meantime, the student's subjective information embraces self-report research interests represented by suggested keywords. And the supervisor's subjective information contains the self-report research fields represented by standard keywords. We do also measure their thinking styles[12] and calculate the match degree as well, which is one of the subjective information.

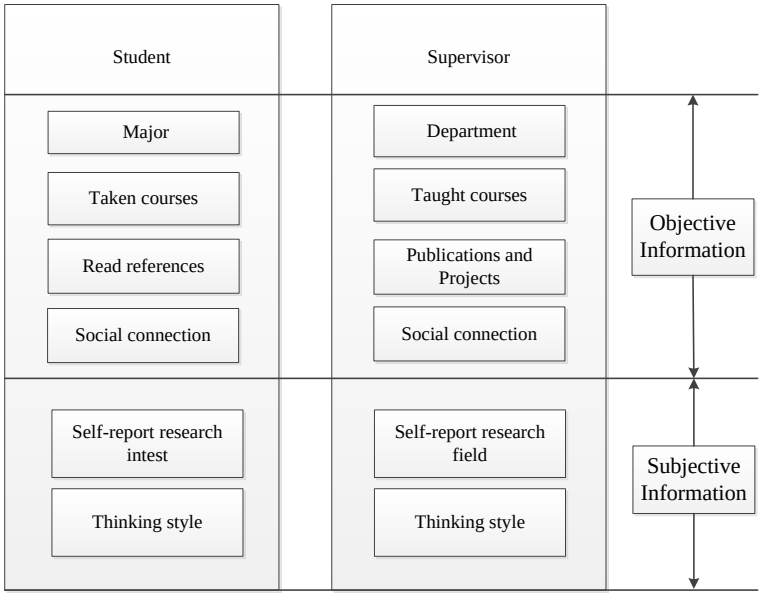


Figure 3. Objective and subjective information integration

4. System Implementation

We apply our method to furnish student-supervisor recommendation services on Scholarmate, which is an online social network platform for academic researchers. The student-supervisor recommendation services rely on the information from users' smart research CVs and their homepages. The smart CV structures the information, embracing the educational experience, academic experience, publications and so on. Furthermore, the authenticity of research publications can be tested, such as papers and funded projects which are the most representative parts of a researcher.

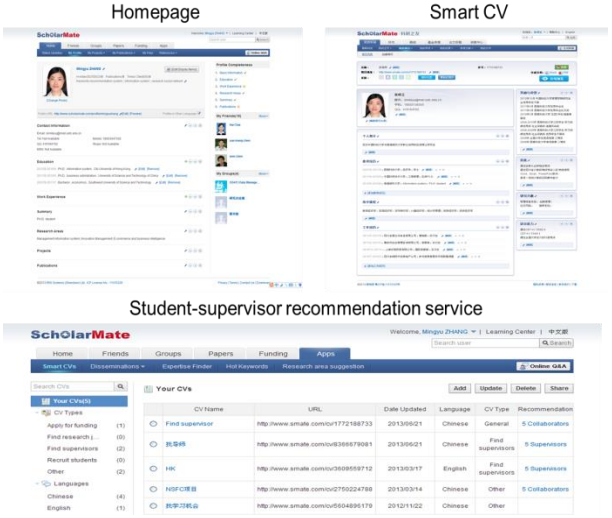


Figure 4.Student-supervisor recommendation interface in Scholarmate

5. Conclusion and Feature Work

In conclusion, we identify anintegrationprocess model to provide supervisor-recommendation services for students, and it is especially useful for new students to select a research guider for themselves. Subsequently, we put forward a model to combining objective and subjective information from bilateral sides. Furthermore, the personalized student-supervisor recommendation services are carried out on Scholarmate.

In the future, the detailed algorithms will be demonstrated. And we willalso test the proposed methods with a user study, and the performance will be measured by students' perceived satisfaction. We plan to make use of the metrics, such as average rate (AR) and normalized discounted cumulative gain (nDCG) to compare the performance of our proposed method and that of baseline methods. Meanwhile, the generality of our proposed RAF has not been tested. In the forthcoming study, our method will be verified in abroader scope of situations.

Reference

1. Armstrong, S.J., *The impact of supervisors' cognitive styles on the quality of research supervision in management education.* British Journal of Educational Psychology, 2004. **74**(4): p. 599-616.
2. Fang, S., *A Survey of Teacher-Student Style Mismatches.* Higher Education of Social Science, 2012. **3**(1): p. 5-12.
3. McAlpine, L. and J. Norton, *Reframing our approach to doctoral programs: an integrative framework for action and research.* Higher Education Research & Development, 2006. **25**(1): p. 3-17.
4. Adomavicius, G. and A. Tuzhilin, *Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions.* Knowledge and Data Engineering, IEEE Transactions on, 2005. **17**(6): p. 734-749.
5. Silva, T., et al., *A social network-empowered research analytics framework for project selection.* Decision Support Systems, 2013.
6. Wallace, B. and R. Oxford, *Disparity in Learning Styles and Teaching Styles in the ESL Classroom: Does This Mean War?* AMTESOL Journal, 1992. **1**(1): p. 45-68.
7. Datta, S., et al., *Use of compromise ranking method for supervisor selection: A multi-criteria decision making (MCDM) approach.* International Journal of Vocational and Technical Education, 2009. **1**(1): p. 7-13.
8. Ray, S. and G. Marakas, *Selecting a doctoral dissertation supervisor: Analytical hierarchy approach to the multiple criteria problem.* International Journal of Doctoral Studies, 2007. **2**(1): p. 23-32.
9. Momeni, M., et al., *Selection Process of Supervisor for Doctoral Dissertation Using Analytical Network Process (ANP): An Iranian Study.* Journal of Management and Strategy, 2011. **2**(2): p. p63.
10. Yang, Z., et al. *Expert2bole: From expert finding to bole search*. in *Proceeding of the 15th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'09)*. 2009.
11. Xu, X., *A note on the subjective and objective integrated approach to determine attribute weights.* European Journal of Operational Research, 2004. **156**(2): p. 530-532.
12. Clarke, T.A., et al., *Thinking styles: teaching and learning styles in graduate education students.* Educational Psychology, 2010. **30**(7): p. 837-848.

An Architecture for Integrating Cloud Computing and Process Management

Daniel E. O'Leary
University of Southern California
Oleary@usc.edu

Abstract

This paper presents an architecture for integrating process management in a cloud computing setting. The public / private approach used in RosettaNet is used as the conceptual basis to capture information about data, documents and processes. The approach is then examined using information in the context of cloud-based software to illustrate different characteristics about the approach.

Introduction

The purpose of this paper is to provide an architecture for integrating cloud computing and process management. In particular, the purpose of that architecture is to provide cloud architectures that meet the minimum process and informational requirements for smaller or less complex user organizations while still allowing larger more complex organizations to add additional processes and data requirements over and above the base processes provided. The approach is based on separating processes into the equivalent of RosettaNet's "public" and "private" processes, where public processes provide some minimal required level of information, documents and work flow, but private processes, generated by the cloud provider (or customers), can be included over and above public processes to provide additional capabilities.

Background - RosettaNet Workflow: Public and Private Processes

RosettaNet was developed as part of a set of efforts to facilitate business to business (B2B) e-business, also referred to as "digitization." The "RosettaNet Consortium," is an independent and nonprofit consortium of some of the major companies in information technology, electronic components, and semiconductor manufacturing. The RosettaNet efforts are designed to create and implement industry-wide, open e-business process standards. Those standardized processes are designed to facilitate the electronic business interfaces used between participating supply chain partners.

In RosettaNet[2], "public" processes define information, documents and data flows that virtually all users must conform to as part of a particular process, for example, one firm ordering goods from another firm, while "private" processes provide specific information, documents and work flow to the user organization processes beyond the public process. This approach promulgates minimizing data and processes specified as part of the "public" process, while still allowing additional data and processes for private use.

As an example, of public and private processes, as seen in figure 1, RosettaNet-based workflows

[2] A similar set of public and private processes is discussed in McAfee (2001).

process information as follows[3]:

1. A customer private workflow initiates a RosettaNet message. Data is retrieved and formatted into a RosettaNet message structure, and is forwarded to the public workflow that implements the customer role.

The private workflow can generate more information than is required by the public process. The public process does not need to employ all of the information generated in the private process, but could include additional discretionary capabilities. The RosettaNet message requires only certain minimal data that meets the needs of the public workflow.

2. The public workflow process creates the appropriate RosettaNet message. The message is sent to the public workflow implementing the product supplier.
3. The product supplier public workflow receives the message, processes the information, and then passes customer information and message content to the appropriate private workflow process.
4. The product supplier private process resolves the message content and generates a reply that is passed back to the product supplier public process.

The product supplier private workflow process can employ or generate additional information over and above the basic information provided as part of the public workflow. The public process does not need to use all of the information generated by the private process, but could include additional discretionary information that may not be used by the private process.

5. The product supplier public process creates the appropriate RosettaNet reply message and sends it to the customer.
6. The customer public process receives the reply message, the information, and then passes product supplier information and message content to the appropriate private process.
7. The private process resolves content of the reply message.

Cloud Computing Structure

Although RosettaNet provides the historical backdrop base of public and private processes, there is an alternative interpretation of that structure that can be used to interface cloud computing and process management. We will describe a cloud computing structure that provides both the minimal required information and workflow for some process, but also provides organizations with the opportunity for additional capabilities in the same cloud computing environment. That architecture will mirror the public – private RosettaNet structure.

Public Processes

From a cloud computing, architectural design perspective, the public process can represent that portion of the particular process that each firm that uses the cloud software is required to use.

[3] BEA Systems, BEA Web Logic Integration: Implementing RosettaNet for B2B Integration, Release 7, June 2002, pp. 2-2 to 2-4, http://docs.oracle.com/cd/E13214_01/wli/docs70/pdf/rosnet.pdf.

That public process often is minimal in that it is designed so that virtually every organization needs the data and the information embedded in that portion of the process. Accordingly, this portion of the data and information flow is universally both minimal and "required." As an example, required accounting information or critical decision making information is likely to drive the public processes. Further, the required public information is likely to include necessary legal promulgated information or other transaction information.

However, cloud computing needs to allow more than just the minimal required workflow and process information, otherwise they unnecessarily limit their base of users or the capabilities of those users. Although potentially many users will need just the required information, other potential users may need to include more information and workflow in their processes, in order to allow for additional capabilities. As a result, public processes could include information that is "optional." For example, the public process could allow the user multiple additional data capabilities.

Private Processes

Private processes, that supplement public processes, can be used to provide additional depth. Since the cloud provider has deep knowledge of the public process they are in a position to build addition capabilities into private processes that can be made available to users as software as a service, over and above the base public process. The cloud provider can use their knowledge of the base public process and requirements of other users to generate additional private capabilities that interface with the public processes

Such private processes can take at least two forms. First, the cloud computing provider can provide alternatives and add-ons to the public process. In particular, "private" processes can be developed that multiple users can employ to supplement existing public processes, but that are not required. Second, individual firms can build their own private processes to interface with either public processes or cloud provider private processes.

Relationship of Private Processes to Public Processes

As illustrated in figure 2, there are a number of different types of relationships that can be captured by integrating public and private processes. For example, in part B of figure 2, the private process provides additional follow-on capabilities, while in part C, the private process encompasses the public process and provides additional capabilities.

Cloud Computing Providers

This architectural approach can be used in either public or private cloud settings. If the concern is a public cloud then the provider potentially is offering services to multiple external clients. If the concern is a private cloud then the provider generates solutions potentially for different branches, divisions or departments of an organization. In either case, the user base is likely to include disparate information consumers and process users.

Case Study: Workday

In order to illustrate these issues, this paper will briefly investigate requirements established by the cloud-based software firm, Workday. Workday has argued that they try to "…keep things simple for our customers."[4] However, they also advocate that "… continuous change is the 'new normal' for enterprise software."[5]

One approach to keeping things "simple" is to develop some key processes that "require" minimal data, documents and process support, but that can allow discretionary capabilities. Such an approach can be done using public processes that have minimal requirements but supplemental capabilities. Since there is a need for almost continuous change, there needs to be a means to "deliver" the software changes. Replacing a public process with another public process or another version of that process provides a means of rapidly adapting to change.

Further, cloud-based software firms need to be able to make wide-scale changes to the software without disrupting use. For example, as noted by Workday

> When we started the company we thought "easier to change" meant two things. First, we envisioned cloud applications that our customers could configure without breaking our ability to upgrade their changes. Second, we wanted a platform that allowed us to continuously deliver new innovations to our customers.

Having all customers on public-based processes and private-processes that Workday generated, would allow Workday to make such changes easily and be able to make changes in those processes without disrupting their users.

Further, public processes can be developed to provide users with the ability to employ supplemental discretionary capabilities. As an example,[6] "Workday captures transactions as business events in order to create a comprehensive financial and operational view of a customer's business. Customer-defined Worktags are tagged to transactional data and used to identify the key dimensions of the business that management would like to track and analyze, such as customer, product, region, and project." Customers do not need to provide the data but it is built into processes that all users have access to (e.g., public processes) if they determine to use it.

In addition, cloud-based firms can provide the equivalent of private processes that interface with their public processes (so not all users need to employ such processes). For example,[7] in what would be equivalent to a public process, Workday users can "… add custom fields to Workday objects, and create validations associated with custom fields. Customers also can now create their own labels. A simple example of this is a label that might better fit their brand or culture than what we offer as standard in Workday—such as 'team member' instead of 'worker.' "

Summary

[4]Swete, S., Applications first at workday, August 29, 2011, http://blogs.workday.com/Blog/applications_first_at_workday.html
[5]Swete, S., Embracing Continuous Change at Workday, September 11, 2013, http://blogs.workday.com/Blog/embracing_continuous_change_at_workday.html
[6]Nittler, M., The surprisingly sexy chart of accounts, March 8, 2012, http://blogs.workday.com/Blog/surprisingly_sexy_chart_of_accounts.html
[7]Swete, S., How we've reimagined software customization at Workday, April 16, 2013, http://blogs.workday.com/Blog/reimagined_software_customization.html

A public process can be used to delineate those processes and data that must be used or provided by each customer of some cloud application. Since all organizations that subscribe to the cloud must use these information flows and data, the public processes must be virtually universally used by large and small organizations and by some set of industries. Accordingly, the public process is effectively required by all organizations. The public process provides "minimal" data, workflows and capabilities. Because they are of universal interest, virtually all organizations that consider the cloud software find a portion of their required needs met in this public approach.

However, because this public portion is not likely to meet all of the requirements for all users, public processes may need to have some discretionary capabilities. Since the cloud provider has deep knowledge of the public process they are in a position to provide additional private applications to interface with and provide greater depth. Those private applications can be generated to provide a range of capabilities to meet the requirements of larger and more complex organizations.

References

McAfee, A., Extricity, Harvard Business School, March 13, 2003, 9-601-113

Oracle Solutions, "Introducing RosettaNet Solutions," http://docs.oracle.com/cd/E13214_01/wli/docs102/tpintro/rosettanet.html#wp1056112

Oracle Solutions, "Using Workflows with RosettaNet," http://docs.oracle.com/cd/E13214_01/wli/docs70/rosnet/workflow.htm#1324261

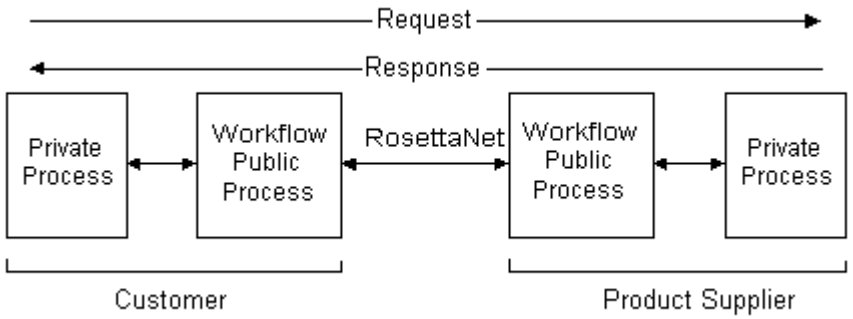


Figure 1: RosettaNet Flow of Information

(http://docs.oracle.com/cd/E13214_01/wli/docs70/rosnet/workflow.htm#1324261)

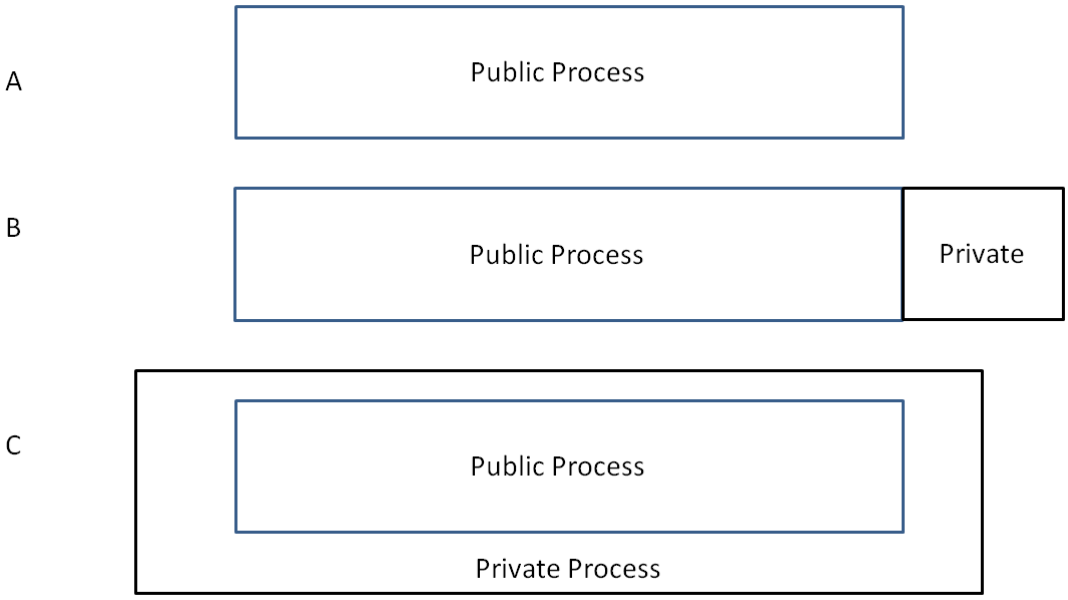


Figure 2: Sample Public/Private Process Configurations

Third-Party Recommendation From Online Recommendation Agents: The Think Aloud Method and Verbal Protocol Analysis

Yani Shi
City University of Hong Kong
yanishi2@cityu.edu.hk

Chuan-Hoo Tan
City University of Hong Kong
ch.tan@cityu.edu.hk

Choon Ling Sia
City University of Hong Kong
iscl@cityu.edu.hk

Abstract： Online product recommendation is viewed as an important way to facilitate online shoppers' decision-making. A recommendation is deemed to be successful if it persuades a consumer to accept the suggested product. To this end, it is of great importance to understand a consumer's cognitive process during online shopping in the presence of product recommendation agent. This study employs think-aloud method and performs verbal protocol analysis to examine the impact of third-party (other consumers and/or domain experts) recommendations on consumer decision-making process. The findings of the impact of third-party recommendation on this cognitive process contribute to both recommendation agent literature and practice.
Keywords: Recommendation Agent, Think Aloud, Verbal Protocol, Consumer Review, Expert Review

Introduction

Online retailers often employ recommendation agents (RAs) to provide online product recommendations, with the objective of not only to support a consumer's decision-making but also to positively influence product choice(Xiao & Benbasat, 2007). However, product recommendations are not always well accepted and sometime consumers may even react negatively(Fitzsimons & Lehmann, 2004). A likely cause of it is the lack of sufficient reasons provided to justify for the recommendations; thus they are not perceived to be persuasively enough. A way to address this is to anchor on the third-party product reviews, such as the expert reviews and the consumer reviews, because such information is often utilized by consumers to make their purchase decisions (Y. Chen & Xie, 2005; Y. B. Chen & Xie, 2008). Expert reviews refer to the product evaluations made by product domain experts while the consumer reviews are written by product customers reporting their post-consumption assessment. Since third-party generated information serves as an important source of reference for consumer decision-making, how the additions of such information in RA's recommendation influence consumer decision-making process?

To answer this question, we need to understand the cognitive processes during which a consumerassesses the product recommendations. Most previous studies on recommendation agents mainly leverage on variables such as the decision effort (e.g., decision time, extent of product search), the product evaluation (e.g., the consumer's ratings of the product alternatives

recommended) (Todd & Benbasat, 1987; Xiao & Benbasat, 2007), among others. The understanding of the entire cognitive thinking process remains lacking. Elaborately, the intervening process of online product recommendation is still considered as an unexplored "black box". For this reason, this research employs the "process tracing" method so as to gain a more wholesome understanding of how consumers go about making decisions when recommendations, which are based on third-party reviews.

Literature Review

Recommendations that are made based on expert reviews are termed as expert recommendation while those that are provided based on consumer reviews are, likewise, labeled as consumer recommendation. An expert recommendation,which is based on product evaluation or assessment by domain experts (Gershoff, Mukherjee, & Mukhopadhyay, 2003), is typically made with product attributes as the focus of justification (Lee, Kim, & Chan-Olmsted, 2010). Comparatively, a consumer recommendation, which is based on post-product consumption evaluation, is typically offered with the experience as the focus of justification (Senecal & Nantel, 2003). Indeed, a consumer review reflects a consumer's usage situations and evaluation of the product performance from a user's perspective (Bickart & Schindler, 2001). It is highlighted that when consumers have difficulties in evaluating product attribute information and expert review information, online consumer review could be a good information source (Alba & Hutchinson, 1987).

There is a lack of studies on embedding product reviews into RA, thus we are unable to reliably predict their consequential impact on consumer decision-making. Indeed, our understanding of the works on product reviews reveals two gaps that also serve as opportunities for research. First, previous studies mainly focus on the impacts of expert/consumer review on product sales or consumers' product judgments(Bickart & Schindler, 2001; Park, Lee, & Han, 2007), and there are few studies that investigate leveraging expert reviews and consumer reviews as recommendation source in online product recommendation. Second, some studies suggested that consumer review is likely to be more credible while other studies believe that expert review is of higher expertise(Brown & Reingen, 1987; Reddy, Swaminathan, & Motley, 1998).

Research Methodology

In order to understand consumer decision-making process, we adopted the "think aloud" method to capture their verbal protocols in a lab experiment setting. "Think aloud" method was developed based on protocol analysis techniques coined by Ericsson and Simon (Ericsson & Simon, 1993; Van Someren, Barnard, & Sandberg, 1994). Think aloud protocols involve subjects thinking aloud as they are performing tasks. Subjects are asked to speak out whatever they are thinking, doing and feeling during the process of task performing (Cooper-Martin, 1993; Todd & Benbasat, 1992).

In our experiment, 5-6 subjects were randomly assigned to every treatment group. Four types of recommendations were presented for four treatment groups: (1) arecommendation with only product information, (2) a recommendation with product information and consumer reviews, (3)

a recommendation with product information and expert reviews, and (4) a recommendation with product information and both consumer reviews and expert reviews. Before conducting the experiment tasks, subjects were primed with a shopping scenario in which they were going to purchase digital products for themselves. They were required to choose one product from each online store. There were four online stores selling cell phones, digital cameras, laptops and mp3 players, respectively. The sequence of purchasing tasks was randomly assigned. These four product categories were selected because they frequently appear in online shopping websites. A product recommendation was presented when a subject has short-listed several options in the shopping car before making a final decision. Real product data was used in the experiment. Product information, consumer reviews and expert reviews were gathered from an IT portal website using a self-developed web crawler.

Verbal protocols were recorded concurrently with the experimental session via Morae Recorder (software by TechSmith). Following the instruction of administering the "think aloud" method, if there is a period of silence (more than 10 seconds typically), the experimenter would prompt the subject to verbalize, which is the only intervention during think aloud session (Todd & Benbasat, 1987; Van Someren, et al., 1994). The prompt should be neutral and unobtrusive by simply asking the subject to speak out what he/she is doing and thinking during the task session.

Data Analysis and Research Findings

Subjects' individual characteristics, such as age, gender, computer experience and online shopping experience, were controlled through randomization. Further checks indicate that there is no significant differences among subjects in all four treatments in terms of age ($F=1.08$, $p>0.1$), computer experience ($F=3.018$, $p>0.05$), and online shopping experience ($F=0.081$, $p>0.1$). There was no significant difference across treatment groups in terms of gender ratio, based on the Kruskal-Wallis test ($\chi 2=4.313$, $p>0.1$).

Manipulation check was conducted to ensure that our manipulation of recommendation source was successful in the experiment.Recommendation source manipulation was checked by asking the subjects whether they thought the recommendation source was from consumer reviews, expert reviews or both. All subjects correctly answered the recommendation sources. As a result, our manipulation of the two independent variables was successful.

An initial data analysis, scanning, was conducted based on the verbal protocols collected in the experiment. As the most straightforward method, scanning "examines the verbal protocols for (frequently anecdotal) information that assists in interpreting quantitative observations" (Todd & Benbasat, 1992).

When a recommendation was presented without third-party product reviews, we found that the subjects did not pay much attention to the recommendation. For instance, a subject doubted by articulating "why the website recommended this product to me?" and closed the recommendation window immediately. This confirms that recommendation is not persuasive if no supportive information or explanation is provided (Gregor & Benbasat, 1999).

Verbal protocols provide rich information on the effect of expert recommendation. For example, a subject switched his choice to an expert recommended laptop. By comprehensively describing

the product features, the expert recommendation provided detailed information on the recommended laptop and comparisons with other products. After reading the recommendation, a subject could speak out several advantages of the recommended product and found that the recommended product would perform better than other alternatives in his consideration set. As a result, he chose the recommended product finally. However, some subjects may not be persuaded by the expert recommendation because they were not interested in the advantages mentioned in the expert review, which suggests the importance of content matching in reviews with consumer's preference

Compared to expert recommendation, consumer recommendation performed worse. Consumer reviews were perceived as subjective rather than objective. Some subjects said that, although other consumers' reviews were honest, they were mainly based on personal preference and there was a lack of supportive evidences provided with review comments.

The subjects, who received recommendations based on both consumer review and expert review, mentioned that "there were so many reviews but I did not want to read all of them". The recommendation was a little bit late, which could be one of the reasons that the subject did not want to spend much time in reading the reviews carefully.It is suggested that too much information may not always benefit consumer decision-making. We also found that subjects had already formed specific preferences and the recommended products often did not match their preferences. It was more difficult for them to change preferences so that recommended products were less likely to be added into consideration set, no matter it was consumer recommendation or expert recommendation.

Discussion and Conclusion

There are limitations in this study which serve as suggestions for future research. First, in this experiment, the recommendation was presented toward the later period of consumer decision-making process. The situations may be different if the recommendation is provided at other stages of consumer decision making process, which deserves further investigation. Second, because of the nature of think aloud method, this study may have limited generalizability as the sample size is small. However, the findings could be leveraged in a large-scaled study to examine its generalizability.

This study contributes to the existing literature by examining the effectiveness of third-party sources in online product recommendation. It is found that expert recommendations play a more important role in influencing consumer decision-making than consumer recommendation. It is also suggested to practitioners that they should attach importance to the selection of reviews with product recommendation. The study also implies the value of understanding the cognitive process via think aloud method and verbal protocol analysis. The use of protocol analysis to understand the principles of consumer information processing could contribute to the future development of online RAs.

References

Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of Consumer Expertise. *Journal of Consumer Research,*

13(4), 411-454.

Bickart, B., & Schindler, R. M. (2001). Internet Forums as Influential Sources of Consumer Information. *Journal of Interactive Marketing, 15*(3), 31-40. doi: 10.1002/dir.1014

Bowman, D., & Narayandas, D. (2001). Managing Customer-Initiated Contacts With Manufacturers: The Impact on Share of Category Requirements and Word-of-Mouth Behavior. *Journal of Marketing Research, 38*(3), 281-297.

Brown, J. J., & Reingen, P. H. (1987). Social Ties and Word-of-Mouth Referral Behavior. *The Journal of Consumer Research, 14*(3), 350-362.

Chen, Y., & Xie, J. (2005). Third-Party Product Review and Firm Marketing Strategy. *Marketing Science, 24*(2), 218-240.

Chen, Y. B., & Xie, J. H. (2008). Online Consumer Review: Word-of-Mouth as a News Element of Marketing Communication Mix. *Management Science, 54*(3), 477-491. doi: Doi 10.1287/Mnsc.1070.0810

Cooper-Martin, E. (1993). An Extension of the Congruence Hypothesis: The Effects of Real Products, Branching Format, Similarity, and Involvement. *Psychology and Marketing, 10*(5), 433-447.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal Reports as Data*: the MIT Press.

Fitzsimons, G. J., & Lehmann, D. R. (2004). Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses. *Marketing Science, 23*(1), 82-94. doi: Doi 10.1287/Mksc.1030.0033

Gershoff, A. D., Mukherjee, A., & Mukhopadhyay, A. (2003). Consumer Acceptance of Online Agent Advice: Extremity and Positivity Effects. *Journal of Consumer Psychology, 13*(1-2), 161-170. doi: Doi: 10.1207/s15327663jcp13-1&2_14

Gregor, S., & Benbasat, I. (1999). Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. *Mis Quarterly, 23*(4), 497-530.

Lee, C., Kim, J., & Chan-Olmsted, S. M. (2010). Branded Product Information Search on the Web: The Role of Brand Trust and Credibility of Online Information Sources. *Journal of Marketing Communications*, First published on: 05 October 2010 (iFirst).

Park, D. H., Lee, J., & Han, I. (2007). The Effect of Online Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. *International Journal of Electronic Commerce, 11*(4), 125-148. doi: Doi 10.2753/Jec1086-4415110405

Reddy, S. K., Swaminathan, V., & Motley, C. M. (1998). Exploring the determinants of Broadway show success. *Journal of Marketing Research, 35*(3), 370-383.

Senecal, S., & Nantel, J. (2003). *Online Influence of Relevant Others: A Framework*. Paper presented at the Sixth International Conference on Electronic Commerce Research, Dallas, TX.

Todd, P., & Benbasat, I. (1987). Process Tracing Methods in Decision Support Systems Research: Exploring The Black Box. *MIS Quarterly*, 493-512.

Todd, P., & Benbasat, I. (1992). The Use of Information in Decision Making: an Experimental Investigation of the Impact of Computer-Based Decision Aids. *MIS Quarterly*, 373-393.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*: Citeseer.

Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *Mis Quarterly, 31*(1), 137-209.

How Patient Adopt Online Healthcare Information? An Empirical Study of Online Q&A Community

Jiahua Jin, Harbin Institute of Technology, **jiahuaking@gmail.com**
Xiangbin Yan, Harbin Institute of Technology, **xbyan@hit.edu.cn**
Yumei Li, Harbin Institute of Technology, **yumeili.hit@gmail.com**
YiJun Li, Harbin Institute of Technology, **liyijun@hit.edu.cn**

Abstract: Based on dual-process theories and knowledge adoption model, we proposed a healthcare information adoption model in online communities. This model highlights that emotional support is an antecedent variable of adoption likelihood, and competition among repliers moderate the relationship between the antecedent variables and adoption likelihood. Empirical data was crawled from the healthcare module of Chinese biggest Q&A community –*BaiduKnow.*Binary logistics regression model and hierarchical regression approach were employed to test the proposed conceptual model. Results indicate that emotional support have significant and positive impact on healthcare information adoption likelihood, and the importance of source credibility on knowledge adoption will be stronger under conditions of higher competition among the repliers.
Keywords:online healthcare community; healthcare information adoption; user-generated content; hierarchical regression

1 Introduction

Online healthcare communities where members share common problems, help each other towards mutual goals, and support each other through good and bad times have become veritable hubs of user-generated content. However, as with most venues of user-generated content, there is need to constantly make adoption evaluations as one sifts through enormous amounts of healthcare content. Misinformation is a big worry when patient search for healthcare assistance through online communities. In addition, too much information may confuse healthcare information seekers. Thus, which determinants may affect users' healthcare information adoption behavior and how do they locate information that is most valuable in their medical advice or decision making are big concerns for not only information seeker, but also website designer and information provider. Previous researches related to online healthcare communities mainly focus on the motivation of patient participation and mechanism of community operation. For example, Kane and Ransbotham(Kane et al. 2010) analyzed how people work together to create peer-produced medical information in social media platforms. Yan and Tan (2010)propose an inhomogeneous Partially Observed Markov Decision Process model to study the helpfulness of an online healthcare community to patients' health condition dynamics. However, few researchers concern about how information seeker adopt online healthcare information and which factors affect their adoption behavior. Based on classical knowledge adoption model and characters of healthcare information, this paper tries to explore how patients adopt healthcare information on online healthcare communities.

2 Literature Review and Theoretical Development

Messages providing by different members are always different in words, tone, perspective, and even logic. Previous researches employed "argument quality" to measure the validity of message (Angst et al. 2009; Sussman et al. 2003). Their research indicated that high argument quality is generally relevant to high possibility of adoption. The more truthful, relevant, and helpful the information contained in the message is, the higher quality the member will perceive the content-based arguments to be, and the more likely it will be that the member will adopt this message(Sussman et al. 2003). Berkman(2000)categorized social support for patient into four forms, namely emotional support, informational support, companionship and instrumental assistance. In online communities, patients usually contact through text-based messages, only information support and emotional support can be expressed through the message. We therefore test:

*Hypothesis 1: The higher the **information quality**, the higher likelihood the message will be adopted.*

*Hypothesis 2: The more the **emotional support**, the higher likelihood the message will be adopted.*

According to elaboration likelihood model, people often use cues pertaining to the message's source when they are unable or unwilling to expend the effort to elaborate on the message content (Petty et al. 1986). In online communities, the message is displayed in a threaded format. Cues such as the popularity of the thread, the relationships among participants, participants' profile (for example, reputation, interest, and skill), the interaction patterns and the evolution of the thread over time, and so on are visible on the interface. Those cues may effect as source credibility to measure the validity of message. We therefore test:

*Hypothesis 3: The higher the **sourcecredibility**, the higher likelihood the message will be adopted.*

Due to the openness of online community, any user can participate in the process of topic discussion. More than one solution is provided, and the recipient would choose the best answer for his/her question. In this paper, we regard members who participant in the topic discussion compete with each other to provide problem solutions. We assume that competition among the participants moderate the effect between adoption likelihood and its antecedents. Because of increased competition, customers would increase their demands with respect to information quality, emotional support and source credibility. We therefore test:

*Hypothesis 4A: The importance of **information quality** on knowledge adoption will be stronger under conditions of higher **competition** among the participants.*

*Hypothesis 4B: The importance of **emotional support** on knowledge adoption will be stronger under conditions of higher **competition** among the participants.*

*Hypothesis 4C: The importance of **source credibility** on knowledge adoption will be stronger under conditions of higher **competition** among the participants.*

Recipient expertise and recipient involvement are two of the most researched determinants of elaboration likelihood (Sussman et al. 2003). In online communities, it's hard to observe

recipient expertise, but we can follow recipient's past experience recorded by the system to measure the degree of involvement, which include the involvement of the online community and the involvement of the discussing thread. Receivers that are highly involved with message issue are likely to engage in high elaboration, while those that are not involved will be less likely to engage in elaboration and more likely to be influenced by peripheral cues (Stamm et al. 1994). We therefore test:

*Hypothesis 5A: The importance of **information quality** on knowledge adoption will be stronger under conditions of higher **involvement** of recipient.*

*Hypothesis 5B: The importance of **emotional support** on knowledge adoption will be weaker under conditions of higher **involvement** of recipient.*

*Hypothesis 5C: The importance of **source credibility** on knowledge adoption will be weaker under conditions of higher **involvement** of recipient.*

3 Empirical results and analysis

3.1 Research context and data corpus

This study took place at *Baidu Know*, which is the most popular online Q&A communility in China and where user puts a question and motivates other members to supply answers. Healthcare module is one of the most popular module of *Baidu Know*, where questions about healthcare problems are proposed and members with related knowledges or similar experiences could provide answers for questions. The website records every detail of questions , answers, and members.The data corpus, which includes texts posted by members from March 2013, consists of 1722 threads.A thread is a collection of one question and several answers. There are two types of member in each thread – recipient and replier. The recipient is the member posting question and make adoption decision, and the replier is the member providing solutions.

3.2 Variable description

As stated in the second section, there are 5 latent variables, namely information quality, emotional support, source credibility, replier competition and recipient involvement. Based on data collected from *Baidu Know*, each variable will be measured as follow:

Information quality: The solution provided by the replier is the key content that recipient cared. We will evaluate the information quality in the next aspects (Otterbacher 2009; Radev et al. 2004). With considering the relevance between the question and the replies, we apply (1) the cosine similarity between the reply and its question, (2) the centroid (textual centrality) score of the reply(Radev et al. 2004), (3) theoverlap between the reply and the question. Taking into account the linguistic characteristics of the replies, we apply (4) the ratio of object sentence to all sentences of a reply, (5) the average number of words in every sentence in the reply.Considering the amount of information in a reply, we apply, (6) the unique words in the reply.Considering the timeliness of reply, we apply (7) the time span between the reply and its question.

Emotional support: Semantic analysis was employed to abstract emotion element contained in the reply. First, we identify the subjective sentences of the messages (Ye et al. 2007). Two

variables are employed to represent emotional support of reply, namely (8) the sentiment score of the reply, (9) the ratio of subject sentence to all sentences of a reply.

Source credibility:*Baidu Know* has an authoritative reputation system. Every member is assigned to certain level according to his contribution to the communities and help to other members. According to ELM, member's reputation can be used to measure source credibility, and impact recipient' adoption likelihood. In this paper, we apply (10) the reputation level of the replier, (11) the number of the adoption of the replier, and (12) the adoption ratio of the replier.

Replier competition: It is certain that recipient will be benefited from the fierce competition among repliers. In this paper, we apply three variables to evaluate the competition, including (13) the number of repliers, (14) the number of replies, the (15) average time span between consecutive replies and (16) the reputation level of the recipient.

Recipient involvement: There are two levels of involvement, first level of involvement is the extent to which recipient takes part in community activities, and the second is the extent to which recipient involves in the discussing thread. According to information provided by *Baidu Know*, we apply 4 variables to measure recipient involvement. The first 3 variables, which represent the first level of involvement, are (17) the reputation level of the recipient, (18) the adoption number of the recipient, (19) the adoption ratio of the recipient and number of questions proposed by the replier.The rest variable representing second level is (20) the number of sub-questions during the thread discussion process.

3.3 Hypothesis test

We test the hypothesized relationships among the constructs using binary regression model with the software program SPSS18.0. Hierarchical regression approach is employed to test the moderating effects of moderator variables (Angst et al. 2009; Baron et al. 1986). Table 1 presents the results of the binary logistics regression. As discussed, the dependent variable is a recipient's decision to adopt a solution in a thread. The third column (Model 1) shows regression on all antecedent variables, namely *information quality*, *emotional support* and *source credibility*. Model 2A and 2B illustrate the results with moderator variable *replier competition.* Model 3A and 3B presents the results with moderator variable *recipient involvement.*

Regression results are shown in third column (model1) in table 1, all of the three antecedent variables are significant correlated to the likelihood of adoption: the beta coefficient for information quality was 0.99 ($p<0.01$), while that of emotional support and source credibility is 0.44 ($p<0.01$) and 0.42 ($p<0.01$) respectively. Thus, hypothesis1, 2 and 3 are supported. Also, the coefficient of *information quality* is as twice big as that of *emotional support* and *source credibility*, indicating *information quality* is the most import factor for recipient's adoption decision. *Emotional support* and *source credibility* have similar impact on adoption decision.

According to model 2A and 2B, only hypothesis 4C was supported ($p<0.01$), namely the importance of source credibility on knowledge adoption will be stronger under conditions of higher competition among the repliers. The results of model 3A and 3B indicate that moderating effects of recipient involvement between information quality and adoption likelihood, and source credibility and adoption likelihood are significant at the 0.1 level of significance. In contrast to

Hypothesis 5C, the importance of source credibility on knowledge adoption will be stronger under conditions of higher involvement of recipient. So only hypothesis 5A was supported, namely the importance of information quality on knowledge adoption will be stronger under conditions of higher involvement of recipient.

Table 1Binary logistics regression results for adoption decision

	Variable	Model 1	Model 2A	Model 2B	Model 3A	Model 3B
		Just *IQ*, *ES*, *SC*	Model 1 plus *RC*	Model 2A plus *RC*× *IQ*, *RC*× *ES*, *RC*× *SC*	Model 1 plus *RI*	Model 3A plus *RI*× *IQ*, *RI*× *ES*, *RI*× *SC*
Main effect	*IQ*	0.99***	0.92***	0.91***	0.99***	0.98**
	ES	0.44***	0.44***	0.45***	0.44***	0.45**
	SC	0.42***	0.40***	0.45***	0.43***	0.44**
	RC	--	-0.54***	-0.56***	--	--
	RI	--	--	--	-0.12**	-0.13**
Interaction effect	*PC*× *IQ*	--	--	-0.04	--	--
	PC× *ES*	--	--	-0.05	--	--
	PC× *SC*	--	--	0.26***	--	--
	RI× *IQ*	--	--	--	--	0.05*
	RI× *ES*	--	--	--	--	-0.08
	RI× *SC*	--	--	--	--	0.02*
Constant	b_0	-0.01	-0.04	-0.04	-0.01	-0.02
Model evaluation	R^2	0.32	0.35	0.36	0.33	0.34
	Percentage correct	71.7	73.1	73.3	72.1	72.3

IQ – information quality; ES – emotional support; SC – source credibility; RC – replier competition; RI – recipient involvement.***p<0.01, ** p<0.05, *p<0.10.

4 Conclusion and Discussion

Although knowledge adoption has been widely researched, little attention was focued on healthcare knowledge adoption. Furthemore, most of the researches related to knowledge adoption employ questionare to collect emperical data. In this paper, we collected data from helathcare module of Chinese biggest online Q&A community — *Baidu Know*. Emperical result indicated that information quality, emotional support and souce credibility have positive and direct effect on adoption decision. Competition among repliers would moderate the relation between source credibility and adoption likelihood, namely the importance of source credibility on knowledge adoption will be stronger under conditions of higher competition among repliers. Similarly, the extent of recipient involvement also has positive moderating effect between information quality and adoption likelihood, source credibility and adoption likelihood.

References

Angst, C. M., and Agarwal, R. 2009. "Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion," *MIS quarterly* (33:2), pp 339-370.

Baron, R. M., and Kenny, D. A. 1986. "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations," *Journal of personality and social psychology* (51:6), p 1173.

Berkman, L. F., Glass, T., Brissette, I., and Seeman, T. E. 2000. "From social integration to health: Durkheim in the new millennium," *Social science & medicine* (51:6), pp 843-857.

Hesse, B. W., Nelson, D. E., Kreps, G. L., Croyle, R. T., Arora, N. K., Rimer, B. K., and Viswanath, K. 2005. "Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey," *Archives of internal medicine* (165:22), p 2618.

Kane, G. C., and Ransbotham, S. 2010. "It's a Network, Not an Encyclopedia: The Quality of Peer Produced Medical Information on Wikipedia," in *Workshop on Health IT and Economics*: Maryland.

Petty, R. E., and Cacioppo, J. T. 1986. "The elaboration likelihood model of persuasion," in *Communication and Persuasion*, Springer, pp. 1-24.

Otterbacher, J. Year. "'Helpfulness' in online communities: a measure of message quality," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM2009, pp. 955-964.

Radev, D. R., Jing, H., Styś, M., and Tam, D. 2004. "Centroid-based summarization of multiple documents," Information Processing & Management (40:6), pp 919-938.

Stamm, K., and Dube, R. 1994. "The relationship of attitudinal components to trust in media," *Communication Research* (21:1), pp 105-123.

Sussman, S. W., and Siegal, W. S. 2003. "Informational influence in organizations: an integrated approach to knowledge adoption," *Information Systems Research* (14:1), pp 47-65.

Yan, L., and Tan, Y. 2010. "An Empirical Study of Online Supports among Patients," Available at SSRN: http://ssrn.com/abstract=1697849.

Ye, Q., Zhang, Z., and Luo, Z. 2007. "Automatically Measuring Subjectivity of Chinese Sentences for Sentiment Analysis for Reviews on the Internet," China Journal of Information Systems:01), pp 79-91.

Exploring Impact of Research Social Networking Services on Research Dissemination

Haidan Liang[1,2], Jian Ma[1], Thushari Silva[1], Mingyu Zhang[1,2], Xu Yunhong[3],Huaping Chen[2] ,
[1]Department of Information Systems, City University of Hong Kong
[2]School of Management, University of Science and Technology of China
[3]Faculty of Management and Economics, Kunming University of Science and Technology
haidan@mail.ustc.edu.cn, isjian@cityu.edu.hk, tpsilva2@cityu.edu.hk,
simileyu@mail.ustc.edu.cn, xuyunhong@gmail.com, hpchen@ustc.edu.cn

Abstract
Research social networking services in research 2.0 have greatly changed the way that researchers share and disseminate scientific knowledge. Effective research dissemination among distributed scholars is crucial to achieve high scientific impact. Researchers used to overlook the effect of social matrices produced via research networking services in scientific knowledge dissemination hence research impact. This study explores the effect of social networking services on scientific impact by assessing the relative explanatorypower of two perspectives: social constructivist approach and word-of-mouth mechanism. Using panel data on 8,411 research articles collected from 45 journals, this study adopted random effect modeling to disentangle impact between article and social level factors on citations. This research strategy resulted in number of both expected and surprising findings. The primary determinants for citations are found to be social network level factors accounting to 58% of the variance in citations. The impact of open accessibility on citation is moderated by social matrices. Thus in conclusion, social networking matrices such as number of likes, shares, downloads and recommendations influence citation pattern of open access articles.

1. Introduction

In the free market of ideas (e.g. scholarly publications), the shoppers (i.e. scholars) are inclined to be more informed than consumers in general market in determining the most desirable (i.e. selecting suitable articles for citation) to acquire. Scholars tend to select previous work based on both intrinsic factors such as quality of the work and extrinsic factors such as quality of the journals in which the work have been published in. With the emergence of research 2.0 technologies, tremendous opportunities for convenient access to an increasing amount of literature have come forth while changing the way of selecting articles for citations. Research social networking services are such tools that have reshaped the ways how researchers could exchangerelated information. For an instance, recommendation services(e.g. Recommendation made by friends) in research social network supports for sharing research article among users and/or friends. This recommendation improves the awareness of the paper hence improves the usability. Ultimately, it improves citations (Alishah et al., 2013).

Previous research has investigated various factors that may influence citations. These factors can

be classified into two main categories: article-level factors and journal-level factors. Article-level factors focus on the content of the article and the influence of authors, which can be considered as intrinsic factors. They cover length of the article, popularity of the topic, and the reputation of the authors in related areas(Leimu & Koricheva, 2005).Some researchers found that multi-author papers are cited more frequently than single-author articles (Figg et al., 2006; Hsu & Huang, 2011), while other researchers support the contrary (Bornmann, Schier, Marx, & Daniel, 2012). (Di Vaio, Waldenström, & Weisdorf, 2012)observed that authors who are full professors at economics and history departmentsare more likely to receive citations than othersin the same field. Journal-level factors investigate factors like journals' prestige (e.g., journal's impact factor or ranking of the journal), journals' interdisciplinary degree and journals' self-citation rate. Previous research has demonstrated that prestige or average citation rate of the journal is one of the most important factors which drive citations (Larivière & Gingras, 2010; Mingers & Xu, 2010;Peng & Zhu, 2012). Most of these researches overlooked the effect of social networking services towards the citations as well as how social networking matrices moderate the effect of open accessibility towards citations.Thus, by drawing social constructivist approach and word-of-mouth mechanism, this study attempts to step further on the basis of existing research on citation, and explore the role of research social networking services in scientific knowledge dissemination hence citation improvement.

The rest of the paper is structured as follows. Section 2 presents theoretical framework and research hypothesis. Section 3 reports results & analysis of the collected panel data. We conclude the paper with conclusion and future work.

2. Research Hypothesis

This study investigates factors influencing citations by integrating two perspectives: Article—level and Social network-level. It draws on two approaches: social constructivist approach and Electronic-Word-of-Mouth (EWOM) mechanism on citing behavior to discover causal forces of citations. Based on the constructed hypotheses, the conceptual research model can be presented as in figure 1. It is developed by extending the model proposed by (Peng & Zhu, 2012).

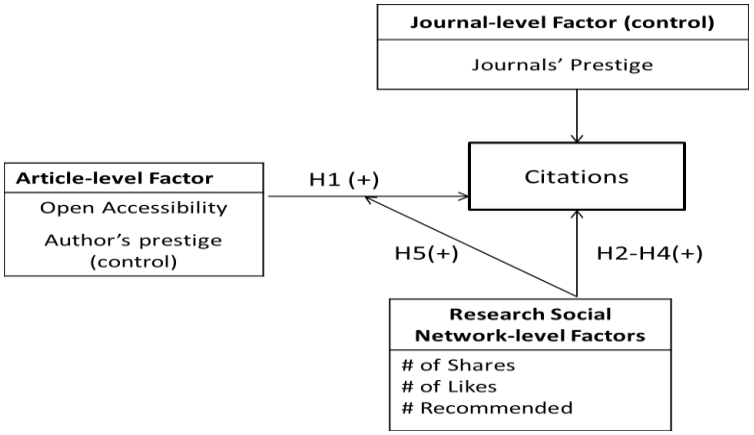


Figure 1: Conceptual research model

2.1 Article-level Factors: Open Accessibility

It has been found established in previous literature that, accessibility to full text of an article i.e. open accessibility is one factor which could influence citation impact(Gargouri et al., 2010). In this research we extend the model proposed by Peng et al., (2012) and formalize open accessibility as one of extrinsic feature which characterize a scientific article.

In order to achieve greater research impact, authors and publishers should make their contributions freely available for the scientific community(Lawrence, 2001). It has been found that "about two-thirds of all articles submitted to arXivwhich is one of the major Open Access digital libraries, since 1991 were ultimately cited somewhere between 1998 and 1999" (Brown, 2001). Davis and Fromerth(Davis & Fromerth, 2007) has conducted analysis on 2765 articles which are published on four mathematics journal and found that articles in arXiv received 35% more citation on average than those for non-deposited articles. According to social constructivist approach, the extent to which articles are cited is the outcome of the ascriptive process in which extrinsic characteristics including open availability matters, and not only the intrinsic characteristics (Van Dalen & Henkens, 2001). Thus treating open accessibility as one of the extrinsic characteristic, and following empirical observations we hypothesize 'Hypothesis 01' as follows.

H1: Articles which are openly accessible would receive more citations than those which are not openly accessible

2.2 Research Social Network -Level Factors

Academic information propagation in a research social network is similar to what is happening in an e-market. In the e-market context consumers tend to buy products which have gained high number of reviews. Similarly, in the free market of ideas (i.e. scientific knowledge), researchers tend to select articles which have been downloaded or used by many users (Chu & Krichel, 2007). This can be interpreted as researchers tend follow what others have usedandthis behavior could affect the citation patterns of the scientific articles. This behavior is supported with social networking services (e.g. shares, likes and recommends) in many research social networking web sites. These services influence the citation impact of scientific articles (Alishah et al., 2013). In parallel, social-media based metrics -sometimes called altmetrics(Priem, Groth, & Taraborelli, 2012) have been developed recently to measure the impact of scientific information propagation in social networks.

There are multiple ways in research social network sites to support scientific information propagation. For an instance, SchoalrMate(www.scholarmate.com) has provided services/tools including 'Like', 'Share' and 'Recommend' to disseminate research articles among its users. Similarly, Tweeter mentions about articles have been used to share scientists' view of the articles among the readers. Furthermore it has been proved empirically that Tweets can be used to predict the citation impact of scientific articles (Eysenbach, 2011). These social networking services/tools can also be treated as EWOM processes of scientific information dissemination which promote information usability. Based on these observations we hypothesize as follows. ;

H2: The higher the number of likes an article has, the more citations that article receives

H3: The higher the number of shares an article has, the more citations that article receives

According to previous research of (Brody, Harnad, & Carr, 2006), authors citing a paper is a multiple-stage decision. First, the authors should access and be familiar with the paper. Second, the authors need to believe that the paper is relevant to their own article. Third, the authors should believe the paper is important enough to be cited explicitly. From the citation process, we can see that if a research paper can be propagated to many researchers, i.e. if the paper has more shares, likes or downloads then the probability that this paper is cited by others will be high. This could be explained in two aspects. First, research social network may expand researchers' channels to receive information. More researchers may be aware of the existence of a particular research paper. Letting more researchers know about the paper may increase the probability of citations. For example, (Di Vaio et al., 2012) demonstrated that diffusion of research-publication of working papers, as well as conference and workshop presentations has positive impact on the citation rate. Second, similar to EWOM in e-commerce, researchers' feelings toward a research paper may also be influenced by other researchers. In other words, if the paper is recommended by other researchers who are users/friends or friend's friend then a particular paper will be getting more attention and high chance of future citation. Following these facts and observations we hypothesize as follows.

H4: If the article is recommended by others then that article receives high citations

Recent research has investigated the effect of social media such as tweeter mentions on openly accessible articles. The recent investigation carried out by Shuai(Shuai, Pepe, & Bollen, 2012) found that tweets could positively influence citation impact of the openly accessible articles.Similarly, there is an interaction effect between open accessibility of the article and social networking services such as 'like', 'Share' and 'Recommend". Based on this conclusion we hypothesize that:

H5: The relationship between open accessibility and citation will be moderated by social network level factors including (a) number of likes (b) number of shares and (c) number of recommends

3 Results & Analysis

We have collected panel data of 17,456 articles which were published in 189 journals. To ensure adequate explaining power of analysis, only those journals which have more than 30 articles from the Information Systems and Management Science discipline in 2000-2007 were included in the study. Finally, 8,411 scientific articles published in 45 journals were retained for further analysis. We estimated a random effects model of the determinants of citation. Figure 2 depicts the supported hypothesis and their path coefficients. We found that there is significant effect of social network level factors on citation impact. According to results all tested hypothesis are supported at $p < 0.001$.

	Model 1 (Control Variable)	**Model 2 (random coefficient model)**		**Model 3 (slopes-as-outcomes effect)**
		Block 1	**Block 2**	
Journal Prestige (IF)	0.245***	0.267***	0.27***	0.29***

	(0.064)	(0.0647)	(0.071)	(0.069)
Author's Prestige (h-index)	0.389*** (0.074)	0.415*** (0.054)	0.428*** (0.064)	0.437*** (0.068)
Article level				
Open accessibility		**0.487*** ** (0.054)	0.586*** (0.061)	0.611*** (0.059)
Network level				
# of like			**0.798*** ** (0.045)	0.841*** (0.045)
# of share			**0.427**** (0.057)	0.478*** (0.057)
# of recommend			**0.598**** (0.087)	0.611*** (0.0759)
Cross-level interactions				
Open accessibility * # of likes				**0.859*** ** (0.036)
Open accessibility * # of shares				**0..478*** ** (0.079)
Open accessibility * # of recommend				**0.651*** ** (0.090)
Model fit				
Explained variance(R^2)[1]	9%	28%	58%	67%

Figure 2: Summary of Results

In summary,themore the article is openly accessible, the more the citations it would get (path coefficient is 0.487, p<0.001).The more the number of likes, the more the citation an article would have (0.798, p<0.001). If an article has more number of shares then it would receive more number of citations (0.427, p<0.001). Similarly more number of recommend, would bring more citation to an article (0.598, p<0.001).With regard to the hypotheses on the moderation effect we found that, the positive relationship between article-level factors (open accessibility) and citations is positively moderated by all three social network-level factors.

4 Conclusion

The objective of this research is to understand the role of social networking services in increasing the impact of research paper as well as how social networking matrices (likes, shares

and recommends) affect positive relationship of open accessibility towards citation. In this research open accessibility is treated as an extrinsic character of an article extending the idea proposed in current literature. This research resulted in a number of both expected and surprising findings. The primary determinants for citations are found to be social network level factors especially likes, shares and recommend accounting to 58% of the variance in citations. Furthermore it has been found that social networking services moderate the effect of open accessibility towards citation significantly. In summary, this study offers several important theoretical implications. First, this study extends the current literature by explaining factors which influence citations from social networking perspective. Second, our study contributes in understanding the role of social networking services in citations. We not only investigate the direct effect of social recommendation factors on citation, we also systematically examine interaction effects of social recommendation factors on the impacts of article-level factors. A managerial implication of this research is that authors or publishers will consider connecting journal websites with research social networking services in order to gain effective knowledge dissemination hence citations and to improve journal prestige. Future research involves the investigation of the effect of other characteristics of research social network such as social capital on citation impact.

References

Alishah, K., Hadi, M., Hosseinian, S., Hosseini-Nami, S. M. A., Hosseini, Z., Karimi, A., et al. (2013). Impact of Wikipedia on citation trends.

Bornmann, L., Schier, H., Marx, W., & Daniel, H.-D. (2012). What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics, 6*(1), 11-18.

Brody, T., Harnad, S., & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Society for Information Science and Technology, 57*(8), 1060-1072.

Brown, C. (2001). The E-volution of preprints in the scholarly communication of physicists and astronomers. *Journal of the American Society for Information Science and Technology, 52*(3), 187-200.

Chu, H., & Krichel, T. (2007). Downloads vs. citations: relationships, contributing factors and beyond.

Davis, P. M., & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics, 71*(2), 203-215.

Di Vaio, G., Waldenström, D., & Weisdorf, J. (2012). Citation success: evidence from economic history journal publications. *Explorations in Economic History, 49*(1), 92-104.

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of medical Internet research, 13*(4).

Figg, W. D., Dunn, L., Liewehr, D. J., Steinberg, S. M., Thurman, P. W., Barrett, J. C., et al. (2006). Scientific collaboration results in higher citation rates of published articles. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 26*(6), 759-767.

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., et al. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE, 5*(10), e13636.

Hsu, J.-w., & Huang, D.-W. (2011). Correlation between impact and collaboration. *Scientometrics, 86*(2), 317-324.

Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American society for information science and technology, 61*(2), 424-427.

Lawrence, S. (2001). Online or invisible. *Nature, 411*(6837), 521.

Leimu, R., & Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution, 20*(1), 28-32.

Mingers, J., & Xu, F. (2010). The drivers of citations in management science journals. *European Journal of Operational Research, 205*(2), 422-430.

Peng, T. Q., & Zhu, J. J. (2012). Where you publish matters most: A multilevel analysis of factors affecting citations of internet studies. *Journal of the American Society for Information Science and Technology, 63*(9), 1789-1803.

Priem, J., Groth, P., & Taraborelli, D. (2012). The altmetrics collection. *PLoS ONE, 7*(11), e48753.

Shuai, X., Pepe, A., & Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations. *PLoS ONE, 7*(11), e47523.

Van Dalen, H. P., & Henkens, K. (2001). What makes a scientific article influential? The case of demographers. *Scientometrics, 50*(3), 455-482.

Business Rules, Decisions and Processes: Five Reflections upon Living Apart Together

Jan Vanthienen, Filip Caron and Johannes De Smedt
Department of Decision Sciences & Information Management, Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven Belgium

{Jan.Vanthienen, Filip.Caron, Johannes.DeSmedt} @kuleuven.be

Abstract

Information Systems must provide flexible support for business processes, which should be compliant to the (intra-organizational) business policies and procedures, to regulations and to imposed protocols. Some separation of concerns, i.e. making the imposed constraints explicit, has been observed in order to comply with this dual objective of flexibility and compliance. In this paper, a list of five reflections is used to examine the coexistence and the relationship between business rules, decisions and processes. These considerations will be useful when evaluating how to become less dependent on rigid process models containing large parts of business and decision logic.

Keywords: Business Processes, Business Rules, Business Decisions

Introduction

Contemporary socio-economic factors – i.e. globalization, mergers and acquisitions – have resulted in a need for standardizing and streamlining business operations. On the other hand, businesses are increasingly facing demands for custom services and flexibility.

Designing information systems that provide support for operational business processes with the right level of process flexibility, compliance, efficiency and effectiveness can be a challenging task [1,2]. This position paper describes five reflections on the coexistence between business rules and processes in order to obtain the aforementioned qualities. Each of the reflections shows a different focus in the business process management research, in increasing order of involvement, which results in differences in the extent to which the desirable characteristics are present in the resulting business process model.

The five reflections can be used as discussion points when deciding to what extent the business logic (rules, decisions) can or should be separated from the business process. Furthermore, different types can be combined and therefore deliver the desired fit in different situations throughout the lifecycle of the business environment.

This paper is structured as follows: in the next section the five reflections, which deal with various forms of business logic modeling, are elaborated. A conclusion summarizes the body of the document and provides some afterthoughts.

An overview of reflections on rules, decision and processes

Reflection 1: Too detailed decision paths clutter the process: Keep it simple

A first point of reflection relates to the intertwining nature of processes and decisions. A process model should not be the direct mapping of a decision tree (as in Figure 1), as multiple cascaded decision diamonds could be merged into one overall decision point (when age > 18 and medical record is clean, accept the client, otherwise decline). Otherwise the business logic takes away the focus from the real process tasks. Of course the example is trivial, but it illustrates how some processes overemphasize decision paths.

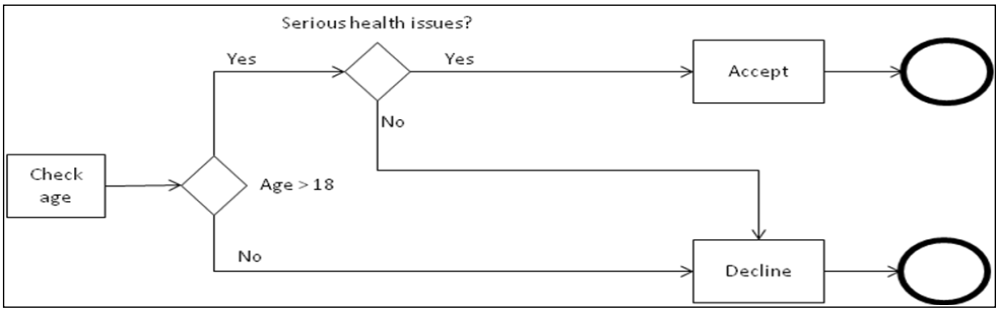


Figure 1: Process model with cascaded decision diamonds

Reflection 2: Hardcoding the rules may become inflexible. Keep them apart

This second reflection entails a fully-fledged co-existence between business processes and decision rules. Universal stable sequential aspects of the operations are being specified in imperative flow models, whereas decisions rules are deliberately withdrawn from the flow specification. Separating these business rules, such as calculations or preconditions, enables them to evolve independently and consequently results in higher levels of operational flexibility [2]. Research has indicated the potentially high volatility of business strategies requiring adequate changes in the business rules leading to a decision, such as calculation rules or preconditions. As a result this co-existence may better cope with the flexibility required by the idiosyncrasies that contemporary organizations face. Of course this seems to hide some of the detailed execution paths or introduces variation, and therefore seems to take away the comfort of nicely drawn execution paths.

Figure 2 revisits part of the example from the previous reflection. A change in the rules will no longer impose immediate changes to the flow logic and therefore creates an agile and maintainable environment. Separating rules and decisions from the process simplifies the process model (i.e. separation of concerns).

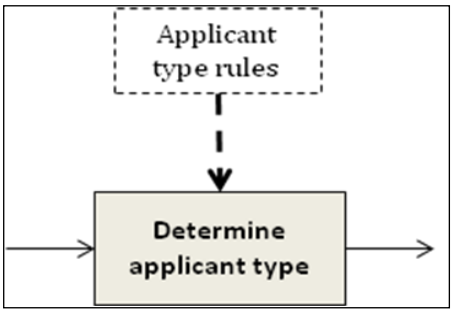


Figure 2: Offloading the process

Reflection 3: Modeling Decisions/Rules is a separate modeling task

Over the years, numerous notations and frameworks have been proposed to formalize rules and decisions. Techniques such as decisions requirements analysis [6] have proven to be able to more clearly separate decisions, business knowledge such as rules and decision tables, and knowledge sources. The approach takes the business owner(s) through various stages to efficiently identify the decision points and their inputs (business knowledge and data) and offers a comprehensive view of the process(es) that will support the obtained decision model.

Decisions are typically based upon a number of business (decision) rules that describe the premises and possible outcomes of a specific situation. Since these decisions guide the activities and workflows of all process stakeholders (participants, owners), they should be regarded as first-class citizens in business process management. Sometimes, the entire decision can be included as a decision activity or as a service (a decision service). Typical decisions of this kind are: creditworthiness of the customer in a financial process, claim acceptance in an insurance process, eligibility decision in social security, etc.

Recently, extensions within the range of business knowledge representation frameworks have been provided. These include the forthcoming OMG Decision Modeling & Notation (DMN) standard, which will represent decision requirements and different decision logic representations such as decision tables. These tables are an excellent tool for business users to model their decisions and provide an intuitive interface for representing business knowledge.

Note that the setup of this reflection is the further extraction of business logic from the overall process. Where the previous approach only modeled the separate decision points, this reflection considers a level of separation which aims to make the business knowledge separate and reusable across the process(es). The model of the decision (e.g. in complex legal documents) can be modeled separately and this knowledge can be used at the appropriate places in the process.

Reflection 4: Sometimes the entire process is basically the execution of a complex decision: there may be multiple ways to 'process' a decision

Before we install or reengineer a process for a complex decision, it might be good practice to study the decision and data requirements. Depending on the decision logic, we could (automatically?) design an optimal process where optimality is defined in terms of process criteria.

An example of this approach is described in [3]. It introduces models illustrating the relationships between decisions, and provides ways to derive business processes to facilitate corresponding decision making. This approach will increase the flexibility, traceability and maintainability of the underlying decision making processes, while at the same time, minimizing the impact from changes caused by modification of specific decision logic. Business decision management should have clearly stated goals during the entire process of eliciting, analyzing, defining, tracking, evaluating, and documenting business decisions. [8] defines five criteria to ensure better decision yield – that is, the impact of decisions on business results: precision, cost, speed, agility and consistency. Moreover, the transformation patterns allow for added operational agility when the characteristics (e.g. time and cost) for obtaining lower level elements change.

Reflection 5: Rule-driven, declarative or Intelligent BPM

Whereas the first reflection explicitly deals with a far reaching sequential structuring of the business operations, the focus in this fifth reflection shifts to expressing the business rules. Attention is put on capturing regulatory and internal directives in rules of different forms (e.g. event conditions and logical expressions). With a minimum specification of the relevant business concerns, maximal allowable freedom is left for letting the exact activity sequence of a process instance grow organically [5]. Moreover, business operations that are modeled according to these principles have the advantage that compliance with internal and external directives can be easily demonstrated. Several business process research subdomains are compromised in this reflection, including declarative business process management [9], ad-hoc business processes [14] and adaptive case management [15].

Business processes that are characterized by a dynamic, human-centric and non-standardized setting, will benefit from the flexibility that could potentially be provided by declarative process modeling (e.g. healthcare processes while general medical principles are the same for all patient, each case will be different due to complications, patient conditions, etc.).

Conclusion

This paper provides an overview of five important reflections relating to the coexistence of business decisions, rules and processes. These considerations could be useful when evaluating how to become less dependent on rigid process models containing large parts of business logic. The construction of more comprehensible and agile outcomes is facilitated by using the appropriate distinct models.

Future research will focus on the development of mechanisms for the integration of business processes,

decisions and rules. We currently examine how business rules can be translated into a uniform event mechanism, such that the event handling could provide an integrated enforcement of business rules of many kinds (including data rules, process rules, timing rules and authorization rules). The creation of a tool that supports the transformation of business constraints specified in an extended process rule enabled version of SBVR (Semantics of Business Vocabulary and Business Rules) into Event-Condition-Action rules, enables the creation of a non-overly restrictive execution model that is compliant with the imposed directives. Moreover, the mechanism could significantly simplify a compliance assessment.

References

[1] De Roover W, Vanthienen J, Unified patterns to transform business rules into an event coordination mechanism (2010), pp. 61 - 73, *International Workshop on Event-Driven Business Process Management (edBPM'10)*, BPM, Hoboken, NJ, USA.

[2] Vanthienen J (2007). How Business Rules (Re)define Business Processes: A Service Oriented View, *10th International Business Rules Forum*, Orlando, FL (USA), Oct. 21-25.

[3] Wu F, Priscilla L, Gao M, Caron F, De Roover W, Vanthienen J (2012), Modeling decision structures and dependencies, *Proceedings of the Third International Workshop on Semantics and Decision Making (SeDes 2012)),* Lecture Notes in Computer Science, vol. 7567, pp. 525 - 533, International Workshop on Semantics and Decision Making (SeDes 2012) (Rome (Italy)).

[4] Caron F, Vanthienen J (2012), Moving across paradigms between the process design and enactment phase in enterprise information systems. *Proceedings of the 14th International Conference on Enterprise Information Systems*, vol. 3, pp. 218 - 223, International Conference on Enterprise Information Systems (Wroclaw (Poland)).

[5] Goedertier S, Vanthienen J. and Caron F (2013), Declarative Business Process Modeling: Principles and Modeling Languages. *Enterprise Information Systems,* (ahead-of-print), 1-25.

[6] Fish A (2012), Knowledge Automation: How to Implement Decision Management in Business Processes. Hoboken: New Jersey, John Wiley & Sons.

[7] Ross, RG (2005), Business Rule Concepts - Getting to the Point of Knowledge, *Business Rules Solutions*, LLC; 2nd edition (2005).

[8] Rohde F, (2005), Little decisions add up. *Harvard Business Review 83(6)*, 24-+.

[9] Braubach L, Pokahr A, Jander K, Lamersdorf W, Burmeister B (2010), Goal-Oriented Process Modelling, *Intelligent Distributed Computing IV*, vol 215, pp. 77-87. Springer, Heidelberg.

[10] Madhusudan T, Zhao JL, Marhsall B (2004), A case-based reasoning framework for workflow model management, *Data & Knowledge Engineering*, 50(1), 87-115.

[11] Pesic M, van der Aalst WMP (2006), A Declarative Approach for Flexible Business Process Management, *Business Process Management Workshops*, edited by J Elder and S Dustar, pp. 169-180, Berlin, Springer.

[12] Sadiq S, Orlowska ME, Sadiq W (2005), Specification and Validation of Process Constraints for Flexible Workflows, *Information Systems*, 30(5), pp. 349-378.

[13] Lu R, Sadiq S, Governatori G (2009), On Managing Business processes Variants, *Data & Knowledge Engineering*, 68(7), pp. 642-664.

[14] Cantara M (2013), Leverage Dynamic and Ad Hoc Processes Now for Business Adaptability, *Research note for Gartner.*

[15] Swenson, K., Palmer, N., 2010. Mastering the Unpredictable: How Adaptive Case Management Will Revolutionize the Way That Knowledge Workers Get Things Done, Meghan-Kiffer Press.

Use of Business Network Analysis for Identifying Financial Fraud Features

Chen Zhu, Jinbi Yang, Wenping Zhang, Choon Ling Sia, Stephen S.Y. Liao
Department of Information Systems, City University of Hong Kong, Hong Kong
Email addresses: chenzhu2-c@my.cityu.edu.hk, yjinbi2-c@my.cityu.edu.hk, wzhang23-c@my.cityu.edu.hk, iscl@cityu.edu.hk, issliao@cityu.edu.hk.

Abstract: Constructions of fraud features in existing studies focus on accounting measures and financial ratios extracted from the financial statements. Accordingly, there have been some considerable fraud sources under-explored. This highlights the theoretical and practical values toward social context fraud features mining and analysis. Following the design science research framework,this research-in-progress paper aims to contribute two kinds of artifacts. First, we utilizethe business network analysis to detect latent contextual fraud features. Second, a framework is designed to identify fraud features from both financial statement and various social media contents and conduct automated financial fraud detection. Our preliminary experimental results confirm that the business network analysis could find some potential fraud evidence for decision making.
Keywords: Financial fraud detection, fraud feature, business network

INTRODUCTION

The definition of the mentioned financial fraud is not limited to fraudulent financial statement, but the general corporate fraud, which involves the following activities (FBI 2007;Ngai et al. 2011), "1) falsification of financial information, 2) self-dealing by corporate insiders, and 3) obstruction of justice designed to conceal any of the above noted types of criminal conduct." With respect to the fraud features utilized, existing studies unvaryingly employed variables extracted from firm's financial statement, including annual statements and quarterly statements (Jans et al. 2010;Cecchini et al. 2010;Ngai et al. 2011;Abbasi et al. 2012) to detect possible activities of corporate fraud. However, the existing methods are limited in detecting the other two kinds of fraud features. Some market research institutions (e.g., Muddy Waters Research Group, Citron Research) have dedicated resources to investigate possible financial frauds involved with Chinese companies that went public inthe North American. Although some of these studies are testified by SEC investigation to be successfully revealing type 2) and type 3) fraud featured activities, the conclusion they reached is not alwaysaccurate. For the sake of fairness, not only investors but also the companies should be protected in the global capital market. The development of an advanced financial fraud detection tool that enables identify each type of the common fraud featureswould be widely applicable to improve deception detection.

With the development of social network analysis (SNA), new shoots of business network discovery and analysis sprout very fast recently. Business network indicates a company's relationships with other companies, and it could reflect a company's status and performance in a specific business sector by compared to competitors or revealed from customers and suppliers (Bernstein 2003). Fraudulent behavior is kind of learning process that is influenced by self-

motivation and environmental incentives. For decades, social relations through business network and social network have not been studied in a social cognitive perspective. Morrison (2002) suggests that characteristics of the social network structures are demonstrated to be related to three different indicators of learning: organizational knowledge, task mastery and role clarity.Identifying and detecting "whiskers of useful information" from those unstructured online content is a huge challenge but holds great opportunities if properly done.

Therefore, this research-in-progress paper aims to tackle the problem of how we canidentify latent fraud features from various social media contents other than financial statementby:(1) designing an analytic framework that can automate the financial fraud detection process, and (2) deploying the business network mining method to identify latent features of financial fraud.

RELATED WORKS

FINANCIAL FRAUD DETECTION (FFD)

Financial fraud detection research has drawn attention of both researchers and practitioners. Previous studies have attempted to build models that could predict the presence of management frauds in financial statements. Kirkos et al. (2007) compare three models and find Bayesian Belief Networkmodel can achieve the best performance to correctly classify 90.3% of the overall samples made up with Greek firms. By using a SVM classifier, Cecchini et al. (2010) attain a fairly good fraud detection accuracy of 80%. The variables used in those articles are mostly and solely made of financial ratios based upon prior research work linked to the topic of fraudulent financial statement (FFS).A theoretical explanation of deception behavior calls for research not only focusing on first hand evidence, but also analyzing in-depth relationship (Miller and Stiff 1993). However, existing FFD technology studies could be reckoned as awide range of applications but the associated theoretical development is limited. On one hand, they use only information from thefinancial statements. Existing methods primarily focus on analysis of structured data from finance statements, while largely ignoring textual information about those related companiesand market activity data, including but not limited to news reports and various social media contents. On the other hand, although some studies such as Abbasi et al. (2012)take industry context into consideration when developing alearningframework; very few studies have covered the social context in their financial fraud research.

BUSINESS NETWORK ANALYSIS

According to Porter's five forces model (Porter 1980), to analyzea firm's supplier and rivalry forces (e.g. number of suppliers and rivals the firm has) is one of the effective ways to learn its operation situations (i.e. competences). Subsequently, the complementary force (e.g. the number of collaborators a firm has) is added as a sixth force by Brandenburger and Nalebuff (1997). Researchers also found that a firm's behavior are affected by its business relationships (i.e. competitions and collaborations between firms(Gnyawali and He2006). The challenge lies on how to extract such business relationships, which may be latent, effectively. In this regards, Zhang et al. (2013) propose a generativeprobabilistic model to extract dynamic business relationships, including latent ones, from online financial articles. It brings us a chance to study

the financial fraud behavior in the whole business networks instead of being isolated. Based on previous studies,our study tackles thefollowing issues: (1) the adoption of latent fraud features detection fromvarious company related information sources, (2) the inclusion of social context as environment consideration when designing the model of financial fraud behavior.

HYPOTHESES

It is suggested by Miller and Stiffin (1993) a model of deceptive communication that the social context influences people to deceive. Social context is described by three aspects: ***familiarity of the parties, context of the relationship, and the status of the relationship***. Familiarity reflects the length and depth of the relationship; Context of relationship reflects familial, organizational, or friendship; the status indicates superior-subordinate or equal relationship. The motivation for the deceiver to deceive affects the actual deceptive behavior and the potential to deceive. Since prior FFD studies only focus on technology applications for financial statement detection,very few studies emphasize environmental influence. Experience of deception detection research pushes us to explore more the relationship among firm itself (ego) and alters (as actors tie with ego in ego network), and potentialfraudbehavior.

Developing relationshipinbusiness network is regarded as strategic skills and momentous demands for management.Snehota and Hakansson (1995) have noted that both personal and organizational relationships are shown to be critical to create effective linkages through daily execution of the partnership and for a long-term perspective. They found that the interaction with external cooperation partners is in form of well gathered strategic information. Therefore, operational levels trading information dissemination could effectively reveal the collaboration partner, as well as their reciprocal influence on organizational behavior. This calls for a clear specification of the ***status and context*** of a firm's position and its impact on organizational behavior within a specified business relationship network. Network ***centrality***measures the relative importance of nodes and edges in a graph in disciplines like sociology (Gabbay and Leenders 2001). We propose hypotheses 1a to 3b focusing on the relationship between the network properties and the fraudulent behavior based on social network analysis.

Hypotheses 1a, 2a, 3a: The network attributes (1a: centrality, 2a: context, 3a: status) of the manager's social network will be significantly related to the fraudulent behavior of his/her organization.

Hypotheses 1b, 2b, 3b: The network attributes (1b: centrality, 2b: context, 3b: status) of a firm's social network will be significantly related to the fraudulent behavior of the organization.

It is found by prior researchers that certain industries have significantly different values for accounting measures as compared with other industries (Abbasi et al. 2012). They found that industry contextual information significantly outperformed the baseline classifiers with no concern of this feature. While they consider the industry as a whole cluster to abstract their common features, the business network intends to look for individual node's connections among other factors, such as customers and partners. It could also reveal the latent relationships of a firm's different products with their competitors. The use of industry measures without inner relationship would result in parsimonious feature coverage. Accordingly, we propose that fraud

cases might have irregular value patterns that are unique and distinctly different from their industries.

RESEARCH METHODOLOGY

Wehave integrated various information sources, including public information, latentfeature information and private information, with different IT artifacts (i.e. statistics, text mining, social media, etc.). As discussed above, a firm's behavior is always influenced by its related firms (Gnyawali and He2006), it is not reasonable to consider the firm isolated when detecting the financial frauds. In our framework, we use text mining and statistical methods to extract abnormal events from public information such as financial statement, expertise online analysis, financial news, stock data, etc.These abnormal events are used as features for our final classifiers.The business relationships, which can be extracted from public information, for instance financialarticles, are considered as latent information.

Based on the assumption that firms co-occurring within a short distance (i.e. one sentence) may have some kind of relationship, and the relationship can be indicated by the left words within the distance, firm pairs, which are potential competitors or collaborators, are extracted from financial articles and represented by a series of terms. Then business relationship indicators, including competitive indicators and collaborative indicators respectively, are adopted to count the number of different indicators embedded in the firm pairs' representation terms. Relationship scores for firm pair i (RS_i) are calculated according to the statistics by the following formula, where COL_i and COM_i stands for the number of collaborative indicators and competitive indicators.

$$RS_i = \frac{COL_i - COM_i}{COL_i + COM_i}$$

Finally, thresholds empirically selected to distinguish collaborative and competitive firm pairs. By combing all the firms' pairs, which may be collaborative, competitive, or others (we defined as neural in our experiments), business networks for the concerned firms can be constructed.

PRELIMINARY RESULTS

Preliminary experiments were carried out based on a set of financial news and annotated articlesfrom Reuters finance (www.reuters.com/finance). DGW (Duoyuan Global Water, Inc.) is selected from the alleged fraudulent companies to be the target into study. We used 30 seeding competitive indicators and 34 seeding collaborative indicators to construct the relationship indicator snippets for latent feature mining. Crawler program was used and collected around 2,000 financial news articles and related reports for the period from 2008 to 2012, among which 431 business relationships were identified. Figure 1 shows the 424business networks of water treatment industry, except DGW. Four nodes are marked red in this industry because they have the most connections. The right upper figure is DGW's network. It is shown separately because its network is found isolated from the others of this industry. When a company operates effectively, it should be active in the particular business sector. Accordingly, it is shown well connected with many other companies in a business network diagram. As opposite, it reveals

potential abnormal operations of DGW, which calls for attention from auditors or investors.

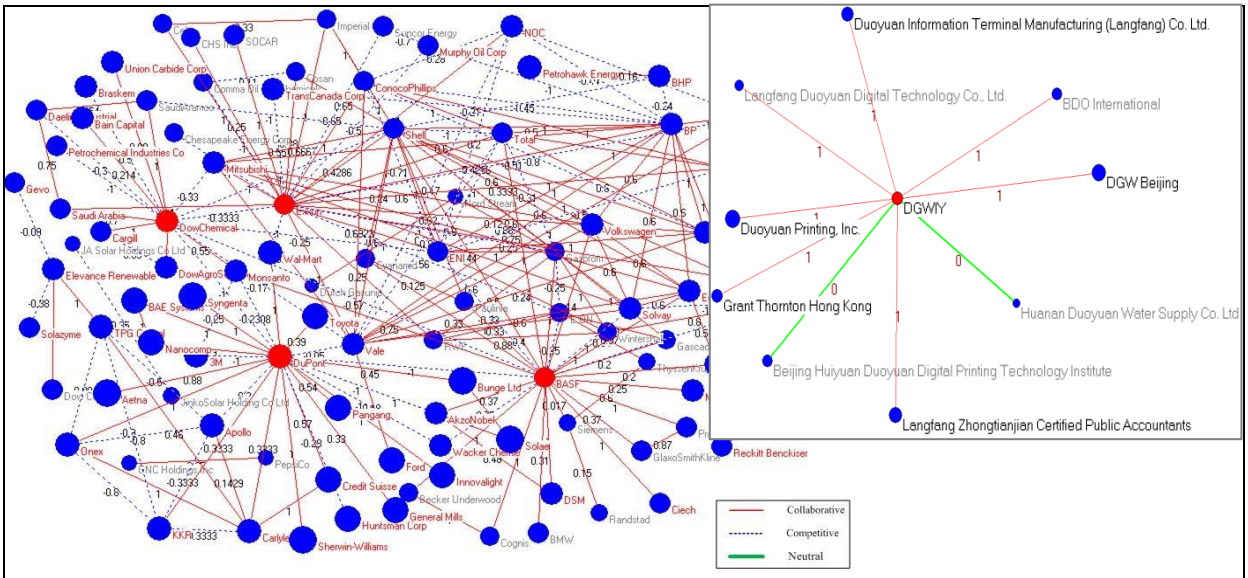


Figure 1. The Business Network of Water Treatment Industry and The Business Network of DGW (right)

CONCLUSIONS AND FUTURE WORK

This study conceptualizes the nature and the role of business network influence as a means for understanding the mechanisms of financial fraud behavior. Taking the extended two-level social context theoretical model as ground, we suggest four hypotheses for future research into the area of contextual analysis of firm's fraud behavior. Empirical investigation would provide more valuable findings of latent features of financial fraud from public and internal, structured and unstructured company information. We are currently testing the effects of detected business network facts as contextual factors on several firms' successful fraudulent reporting. We expect to present the preliminary results at the workshop. The research model and framework that we developed could be studied in other organizational decision making area. Further examination of the financial fraud mechanisms can improve the effectiveness of fraud detection and firm's creditability analysis.

REFERENCES

Abbasi, A., Albrecht, C., Vance, A.O., and Hansen, J.V. 2012."Metafraud: A Meta-Learning Framework for Detecting Financial Fraud," *Management Information Systems Quarterly* (36:4), pp. 1293-1327.

Bandura, A., Ross, D., and Ross, S.A. 1963."Imitation of Film-Mediated Aggressive Models," *Journal of abnormal and social psychology* (66:1), pp. 3-11.

Bernstein,A.,Clearwater,S.,andProvost,F.2003."The Relational Vector-Space Model and Industry Classification," *Proceedings of the IJCAIworkshop on learning statistical models from relational data*.

Brandenburger, A.M., and Nalebuff, B.J. 1997."Co-Opetition: A Revolution Mindset That

Combines Competition and Cooperation: The Game Theory Strategy That's Changing,").
Cecchini, M., Aytug, H., Koehler, G.J., and Pathak, P. 2010. "Detecting Management Fraud in Public Companies," *Management Science* (56:7), pp. 1146-1160.
FBI, Federal Bureau of Investigation, Financial Crimes Report to the Public Fiscal Year, Department of Justice, United States,2007.
Gabbay, S.M., and Leenders, R.T.A. 2001.*Social Capital of Organizations: From Social Structure to the Management of Corporate Social Capital*. Emerald Group Publishing Limited.
Gnyawali, D.R., and He, J. 2006. "Impact of Co-Opetition on Firm Competitive Behavior: An Empirical Examination," *Journal of Management* (32:4), pp. 507-530.
Jans, M., Lybaert, N., and Vanhoof, K. 2010. "Internal Fraud Risk Reduction: Results of a Data Mining Case Study," *International Journal of Accounting Information Systems* (11:1), pp. 17-41.
Kirkos, E., Spathis, C., and Manolopoulos, Y. 2007. "Data Mining Techniques for the Detection of Fraudulent Financial Statements," *Expert Systems with Applications* (32:4), pp. 995-1003.
Miller, G.R., and Stiff, J.B. 1993. *Deceptive Communication*.Sage Publications Newbury Park.
Morrison, E.W. 2002. "Newcomers'relationships: The Role of Social Network Ties During Socialization," *Academy of management Journal* (45:6), pp. 1149-1160.
Ngai, E., Hu, Y., Wong, Y., Chen, Y., and Sun, X. 2011. "The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature," *Decision Support Systems* (50:3), pp. 559-569.
Porter, M.E. 1980. "Competitive Strategy: Techniques for Analyzing Industries and Competitors. 1980," *External links*.
Snehota, I., and Hakansson, H. 1995. *Developing Relationships in Business Networks*.RoutledgeLondres.
Zhang, W., Lau, R.Y., Liao, S.S., and Kwok, R.C.-W. 2013. "A Probabilistic Generative Model for Latent Business Networks Mining."

Ever-Changing Workarounds: A Model for Workaround Management Lifecycle in Healthcare Workflow

Shaokun Fan, Yu Tong, J. Leon Zhao
Department of Information Systems, City University of Hong Kong
{shaokfan,yutong, jlzhao}@cityu.edu.hk

Abstract: When users do not use an information system in the intended ways, workarounds occur. Research on healthcare information technologies finds that workarounds have important impact on healthcare risks and service quality. However, there has been limited study of the mechanisms for workarounds management. This work adds a fundamental step for workaround management by proposing a conceptual model of workaround management lifecycle in healthcare routines. Drawing upon literature of organizational routines and business process management, our model suggests that the key to successful workaround management is to use IT as the controller for different types of workarounds and continuously evolve the system when new workarounds emerge.

1. Introduction

With the touted potential to improve healthcare service efficiency and quality, healthcare information technologies (HITs) are gaining immense attention by hospitals and governments worldwide. However, the realization of potential benefits is threatened in reality because healthcare users are often found to adapt the intended way of using a HIT, i.e., workarounds occur(Koppel et al. 2008). An important source of HIT workarounds stems from the paradoxical tension between system standardization and healthcare complexity. On the one hand, the design and implementation of HITs aim to standardize the healthcare work processes. One the other hand, the nature of healthcare work within hospitals inherently involves a multitude of contingencies and interruptions. Hence, workflows in hospitals are advised to retain certain extent of flexibility and redundancies.

Workarounds may possibly trigger further improvement on the system and/or healthcare processes. However, some workarounds could also serve as surprises to the hospital executives and even bring adverse effects on the healthcare work such as new medication errors andlow service quality(Ash et al. 2007; Koppel et al. 2008). Given the paradoxical tension and prominent role of workarounds, a deep understanding of HIT-related workarounds is needed for managing workarounds in healthcare routines. While workarounds are widely recognized in hospitals, how to appropriately manage them is still a challenging task. Extant literature mainly views workarounds as a post-implementation phenomenon and focuses on causes and consequences (Goh et al. 2011; Nadhrah et al. 2013). While these studies provide valuable contribution, they usually view social and technological components in isolation. In this paper, we complement the existing literature by proposing a model of workarounds-based system development lifecycle in healthcare management. We contend that HIT development as a continuous interaction process between users, managers and designers. Workarounds should be

both the initiator and result of HIT implementation.

2. Literature Review

2.1. Organizational Routine

Defined as repetitive patterns of independent actions among organizational actors (Feldman and Pentland 2003), routines play a central role in organizational studies. Although scholars agree about the pervasiveness of routines in organizations, they differ considerably in theorizing how routines function, which result in the "(n)ever-changing world" paradox. The "never-changing world" perspective treats routines as fixed entities designed to generate consistent actions and efficient results (March and Simon 1958, Nelson and Winter 1982). This work emphasizes that patterns of actions stabilize in organizations and become increasingly repetitive. By contrast, the "ever-changing world" perspective in routines emphasizes the participation of actors who change their performances of a routine to respond to changes in the context (Feldman 2000, Howard-Grenville 2005, Salvato 2009). Work from this perspective therefore emphasizes the dynamic characteristics of routines and helps us understand the lifecycle of workarounds management.

2.2. Workarounds in Healthcare Routines

Workaroundscan be understood from the notion of deviance, which represents behaviors departing from certain norms (Dodge 1985). Deviance was conventionally conceptualized as a negative set of behaviors based on the assumption that an individual ought to comply with the norms from certain reference groups. Recent researchers advocate that deviance could also be beneficial to the organization. Spreitzer and Sonenshein (2004) define positive deviance as "intentional behaviors that depart from the norms of a referent group in honorable ways" (p. 828).

Workarounds can be classified into three categories based on the causes: exception, functional misfit, and user motivation. First, some researchers consider workarounds as informal temporary practices developed by users for conducting exceptions to workflow (Kobayashi 2005). Exceptions are usually caused by unexpected event (e.g. special requirements, system failures or emergent cases) that are not anticipated by normal workflow. Given the ad hoc nature of exceptional events, this type of workarounds does not happen frequently in routines. Second, workarounds may also happen frequently when the intended process cannot satisfy the functional requirements. That definition associates a workaround with work processes that are poorly designed or fail to keep up with changes in the context (Morath and Turnbull 2005). Workarounds can be treated as a phenomenon to create and execute an alternate path to a goal when a path to that goal is blocked (Koopman and Hoffman 2003). The third type of workarounds is caused by user motivations. Poelmans (1999) observed that in many cases "end-users work around the system to save time and/or efforts". Besides time or effort savings, goals, workarounds may be motivated by payoff expectations (Nadhrah and Michell 2013).

Prior literature often takes a black-box approach to study workarounds, resulting in a lack of theoretical bases (Ciborra, 2002). However, HIT-induced workarounds have very important impacts on healthcare service quality and safety (Murray, 2001). As a result, this leads to a call for a deeper understanding of workarounds and theimplicationsto IS implementations (Orlikowski and Yates, 2006). In this study, we propose a conceptual model to develop a theoretical understanding of the social-technical aspects of HIT workarounds management.

3. A Model for Workarounds Management Lifecycle

Figure 1 depicts the proposed workaround management lifecycle model, which attempts to articulate workarounds management as an approach to continuous improvement of healthcare information systems. The logic of the workaround management lifecycle model draws upon the lens of structuration theory and adaptive structuration. Adaptive structuration theory (DeSanctis and Poole 1994), originated from structuration theory(Giddens 1984), recognizes the mutual influence of user actions and system structure in organizations. System structure refers to the structural features and spirit of a system that serve as templates for accomplishing system-related tasks (DeSanctis and Poole 1994). During the system design and implementation phases, designers and/or organizational management often create the expected system structure, which influence the actions performed by system users. However, system users' actions are not determined by the created system structure. They can appropriate the expected structure (e.g., workarounds). As a result of appropriation, updates on the system and/or related rules could emerge (Jones and Karsten 2008).

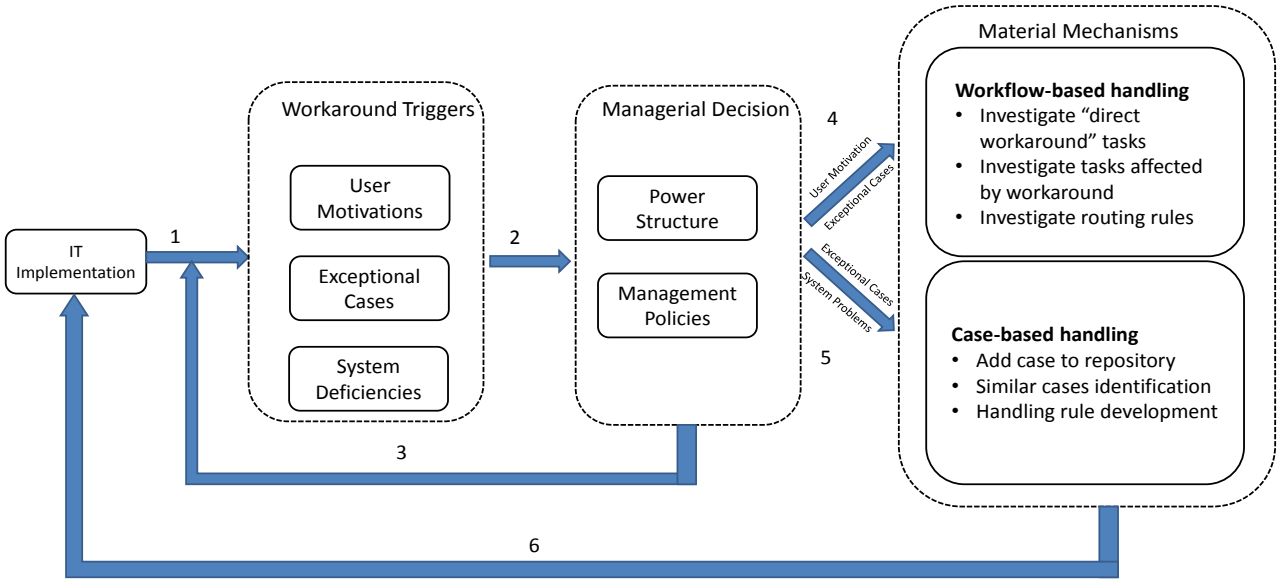


Figure1. Workaround Management Lifecycle Model

The workaround management lifecycle model is developed based on research and observation of HITs. We only consider workarounds that are induced by HITs. Other types of workaround (such as change of personnel, policy, etc) will not be considered in this paper. *Workarounds* are a post-implementation phenomenon widespread in healthcare organizations. Alongside the increasing computerization of healthcare services, there is a parallel increase in HIT-related workarounds (Murray, 2001). Workaround management is a particularly important issue in healthcare context as healthcare professionals are extremely sensitive to potential negative consequences (such as risks) that are caused by workarounds. The workaround

management lifecycle model asserts that HIT development is an iterative process in which the system should be continuously improved based on workarounds that happens in system usage. IT implementation leads to three types of workaround triggers. Based on social decisions, workarounds can be handled by either managerial control or material handling mechanisms (workflow-based or case-based). Material handling mechanisms lead to technical enhancement to the system. Next, we will explain the major stages (numbered arrows in Figure 1) in the workaround management lifecycle.

While HITs aim to standardizing and automating healthcare routines via formal workflows, they may cause many problems which also enable workarounds. After a new system is implemented, users may have difficulties in using the system for reasons such as inefficient process design, poor system interface, and inadequate user training (Halbesleben et al., 2008; Vogelsmeier et al., 2008). For example, if a system requires verification steps that are considered unnecessary by users, people want to save time and skip the verification tasks. Workarounds may be also caused by exceptions. Exceptions are usually caused by unexpected events (e.g. special requirements or emergent cases) that are not anticipated by the design of the system workflow. Given the ad hoc nature of exceptional events, this type of workaround does not happen frequently in routines. System deficiencies may also lead to workarounds. For example, people have to go with manual process when the system is down or has runtime errors. For the path from IT usage to work around triggers, we argue that

Arrow 1: The use of newly implemented HIT functions can lead to three types of workaround triggers: user motivation, exceptional cases and system deficiencies.

After workarounds occur, managerial decisions should be made based on the power structure of participants and management policies. Power structure means the power distribution between a workaround actor and the affected actors. For example, physicians inherently possess higher power over nurses in hospitals. Hence, hospital management may tolerant physicians' workarounds even when these workarounds burden nurses' work. The other way may not work. Management policies mean organization's regulations and rules. For instance, in hospitals, the management can rely on certain rules to decide the workarounds handling methods: 1) whether the reported workarounds are executed by large amount of users and can benefit the healthcare process; 2) whether certain exceptional cases have the indications that could potentially trigger mass breakouts, e.g., SARS. Thus, we argue that

Arrow 2: The workaround handling mechanism will be decided by managerial decisions that are made based on power structure and management policies.

Workaround can be managed with either technical (material perspective) or non-technical (social perspective) approaches. The social control mainly uses managerial intervention that does not require any changes to the system. Sometimes, managers may decide to do nothing (e.g., be tolerant about physicians' system delegation). Most of the time, managerial decision for social control will lead to creation of new management policies (e.g., provide monetary incentives to users who do not perform workarounds). Material handling mechanism tries to improve the system that can better manage the workarounds. If a workaround is not appropriately managed, the potential risks may rise. The objective of workaround management is to avoid

workaroundswith negative impact and enable workarounds with positive impact.

Arrow 3:Social mechanisms that does not involve technical enhancement can be used for workaround handling.

Depending on the triggers of workarounds, two major material mechanisms can be used to handle workarounds, cased-based handling and workflow-based handling.*Workflow-based handling* relies on contemporary workflow management system (WfMS) that enables the modeling, execution, and monitoring of business processes(Georgakopoulos et al. 1995). When an actor (nurses or physicians) is working on a particular process step (i.e., activity), typically, in WfMS-based applications, only data needed for executing this activity is visible to the actor, but no other workflow data. This handling paradigm is typically strong in control but weak in flexibility. *Case-based handling* is different from workflow-based handling(Madhusudan et al. 2004; Mutschler et al. 2008;Van der Aalst et al. 2005). The central concept behind case-based handling is that every individual case and its process are stored in the system. Usually, case-based handling presents all data about a case at any time to all the users. Case-based handling orchestrates the execution of activities based on processes of similar cases. This handling paradigm is strong in flexibility and weak in control. Three major steps of case-based handling and workflow-based handling are listed in Figure 1. ("Direct workaround tasks" are defined as the tasks that have workaround; "tasks affected by workaround" are tasks that don't have workaround but the execution of these tasks are changed by workarounds.) The steps for both handling mechanisms are conducted sequentially and allow retrospective backtracking. Due to page limit, the details of the workaround handling steps are omitted.

Arrow 4&5. Two material mechanisms can be used to handle workarounds, depending on the triggers of workarounds: cased-based handling and workflow-based handling.

- *User motivation triggered workarounds require workflow-based handling*
- *System error triggered workarounds require case-based handling*
- *Exceptional case triggered workarounds can be handled in either way, depending on the frequency of exceptional cases and expected trend of frequency in the future.*

User motivation caused workarounds requires workflow-based handling. User motivation usually does NOT change for the same workflow. If the healthcare system stays the same, user motivation triggered workarounds are very likely to happen again. The changes to workflow includes the ways of task execution (e.g., new system user interface), task assignment (inclusion of duplicate users), and/or process logic(e.g., sequence of taking a medication order). System error caused workaround requires case-based handling. System errors happen rarely and each occurrence can be treated as a case. Exceptional case caused workaround may be handled in either way, depending on the frequency of exceptional cases and expected trend of frequency in the future. For exceptional cases that happen more and more, workflow-based handling should be used. Workflow based handling will need to change the configuration of system and redesign the workflow. Case-based handling will need to add the case and rules to the repository. Both of these handling methods require the development and implementation of new IT artifacts. With several rounds of workaround management, the system can be continuously improved.

Arrow 6. Material workaround handling mechanisms lead to technical enhancement and implementation of new IT artifacts.

4. Conclusions

In this paper, we propose a model of workarounds management lifecycle based on the theories of organizational routines and business process management. We model workarounds management as an approach to continuous improvement of healthcare information systems. It sets forth a theoretical framework for further empirical studies to investigate workaround management. In this study, we incorporate managerial and technical principals into the workarounds management lifecycle. The workaround lifecycle model advances the IS design and post-adoption literature by focusing on the deviated IS use behaviors and their contribution to system improvement. Future work includes developing falsifiable hypotheses based on the workaround management lifecycle model and collecting empirical data to validate framework.

References are available upon request.

Exploring the Potential of User Segmentation in CQA

Research in progress

Jiawei Wang
Department of Information Systems
USTC-CityU Joint Advanced Research Centre
166 Ren'ai Road, Suzhou, China
wangjw2@mail.ustc.edu.cn

Yiyang Bian
Department of Information Systems
USTC-CityU Joint Advanced Research Centre
166 Ren'ai Road, Suzhou, China
bianyiyang321@gmail.com

Yujing Xu
Department of Information Systems
USTC-CityU Joint Advanced Research Centre
166 Ren'ai Road, Suzhou, China
xyjsgdhr@gmail.com

Stephen SY Liao
Department of Information Systems
City University of Hong Kong
83 Tat Chee Avenue, Hong Kong
issliao@cityu.edu.hk

Abstract

Community Question and Answering (CQA) plays a vital role in online problem solving and knowledge sharing. CQA service facilitates users to post questions and get right answers. In this sense, its vitality heavily relies on active users possessing valuable expertise in domains from which questions are proposed. In practice, system assigns unsolved questions to appropriate users – namely, question recommendation which is a crucial process in providing CQA service. Previous research mainly focused on developing techniques for associating each question with potential active user based on overall users, leading to the prohibitive system cost. With the expansion of CQA, user management becomes difficult. In this work, we complement previous research by proposing an interesting process - user segmentation- for efficient question recommendation and user management before the question recommendation process. Use segmentation makes the reduction of system cost in question recommendation possible. Besides, it helps community organizers develop and maintain relationships between the community and users in groups of different characteristics. By applying fuzzy c-means clustering, active users may be classified into different groups representing different preferences in providing answers. A dataset collected from Stack Overflow might help conduct nice user segmentation and inform the design of future practical question recommendation mechanism and user management.

Keywords*: User segmentation, question recommendation, fuzzy c-means clustering.*

1. INTRODUCTION

Community question answering (CQA), such as Yahoo! Answer, Stack Overflow, Quaora, and TurboTax, is a type of online social networks (OSN) and contributes greatly to problem solving and knowledge sharing. In these CQAs, people are able to post questions that cannot be solved by searching the web content and obtain answers from other members. Besides, CQA provides a platform for all Internet users to exchange and share their knowledge. Therefore, a CQA of high vitality and activity relies heavily on values of questions and answers it publishes. Although the CQA's prevalent, there still remains a question: how to make this service to solve users' question efficiently?

In practice, there are active users who are the major drivers for community's development and prior studies usually define active users as those who are active in responding to questions and giving answers of high

credibility(Pal et al. 2011). Questions are volatile and new questions come every second in popular CQAs. Thus, it's time-consuming for users to find the questions that they relate to. As a result, although there are lots of active users, a number of questions remain unsolved or without accepted answers. To improve the users' management of CQA portals and expedite the answering of new questions, question recommendation emerges. Question recommendation is a mechanism that exposes the right question to the right users (Dror et al. 2011).

There is another fact that active users receiving the recommended question might not answer these questions in a timely manner. People provide answers in different patterns. For example, some prefer to answer questions at any moment of a day while some others only want to response in the evening. Some active users even have a question selection bias and prefer to answer questions from which they have a higher chance of making a valuable contribution (Pal et al. 2010). A better understanding of users' behaviour can help the CQA portal run user management efficient. Therefore, modelling users' answering behaviours is of great importance. To our best knowledge, research in question recommendation run the model based on overall active users for each question, which leads to a prohibitive system cost. Besides, in order to achieve the integration of social behaviour of users (e.g. voting, making comments.) with their questions and answers, researchers tend to combine topic models and link analysis However, this task in non-trivial when dealing with volatile questions and answers.

Moreover, rapid expansion of CQA inhibits an efficient management of the community. Organizers need to nurture and promote relationships between users and the community for the sake of the community's sustained development. With such a large scale, community possibly lost focus. These considerations turn the need for CQA users' classification in an appealing idea from both an academic and practitioner approach(Alarcon-del-Amo et al. 2011).

Introducing the user segmentation into CQA area can bridge the aforementioned gaps. User segmentation or customer segmentation is a strategy for enhancing sales in direct marketing(Seret et al. 2012). By dividing users into different groups, company can devise and tweak its policies to attract users with different preferences (Ozer 2001). We claim that for the sake of practical recommendation and efficient user management, the benefit of introducing user segmentation in CQA is two-fold. Firstly, question recommendation would be much efficient if we only need to match the question with limited number of user cliques represented by group centres. Hence, the question would be routed to users in the most appropriate group. Secondly, maintaining relationships with actives users would be easily and appropriated fulfilled if we know user's preference in terms of interest, answering preferences, cultural background, and geographical location, etc.

2. METHODOLOGY

In order to perform user segmentation, we design a methodology as the process model shown in figure 1.

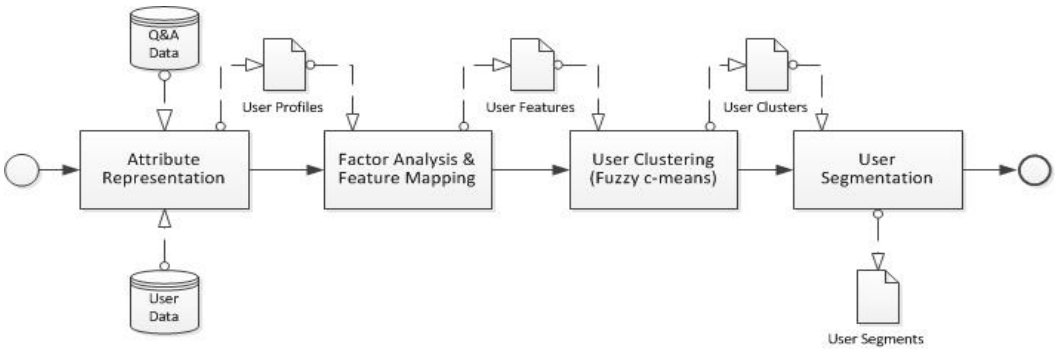


Figure 1. A Process Model of User Segmentation

Attributes Representation

For each active user, we describe him/her from four perspectives: demogrphic informaiont, personal interests, answering patterns, and authority. For detailed attributes for each perspetive, please refer to table 1.

Demogrphic Information	Gender, age, location.
Interests	Topics extracted from collection of questions and answers with topic models.
Answering Patterns	Prefered time peirod, median length of words in answers, percentage of answers with codes, percentage of answers with linkage, toatal # of answers, median of # of prior answers, median length of words in questions answered.
Authority	Reputation score, percentage of answers been accepted, # of upvotes, # of downvotes, # of favorites, # of pageviews.

Table 1. Attributes Representation

Interests of users are hidden patterns and can be revealed from users' questions and answers. We regard collection of questions and answers from all users as a corpus of documents, from which topics can be explored with Latent Dirichlet Allocation (LDA) model . LDA is a latent topic model and it assumes each document is a mixture of serveral topics with different weights (Blei et al. 2003). Therefore, each active user's passion for different topics is quantified with weights. Please note that we understand there are some attributes that we have not leveraged yet and plan to consider in the future.

Factor Analysis & Feature Generation

High correlation might exist in the above attributes and factor analysis is an efficient approach for eleminating redundant variables. In the context of CQA, we refine our attributes generated (see table 1) with factor analysis and find the higher level representative features for furthur user segmetation.

UserClustering

Fuzzy c-means (FCM) is a widely used method of clustering which allows one piece of data to belong to two or more clusters and it remains one of the general purpose fuzzy clustering techniques (Bezdek, 1984). In the context of CQA, this can represent similarity of one user shares with each cluster with the membership function. This method (Dunn, 1973) is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \qquad , \quad 1 \le m < \infty$$

where m is the weight, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the *i*thobservation, c_j is the center of the cluster j, and ||*|| is any norm expressing the similarity between any measured data and the center. Fuzzy clustering is solved by an iterative optimization of the objective function shown above, with the update of membership u_{ji} and the cluster centers c_j by:

$$u_{ji} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_j - c_k\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}}, \; c_j = \frac{\sum_{i=1}^{N} u_{ji}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ji}^{m}}$$

This iteration will stop when$max_{ij}\left\{\left|u_{ji}^{(s+1)} - u_{ji}^{(s)}\right|\right\} < \varepsilon$, where ε is a termination criterion between 0 and 1, and *s*is the iteration step. This procedure converges to a local minimum or a saddle point of J_m. The algorithm is composed of the following steps:

1. *Initialize U=[u_{ji}] matrix, $U^{(0)}$*
2. *At k-step: calculate the centers vectors $C^{(k)}$=[c_j] with $U^{(k)}$,* $c_j = \frac{\sum_{i=1}^{N} u_{ji}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ji}^{m}}$
3. *Update $U^{(s)}$, $U^{(s+1)}$*
4. $u_{ji} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_j - c_k\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}}$
5. *If || $U^{(s+1)}$ - $U^{(s)}$||< ε then STOP; otherwise return to step 2.*

According to these steps, the center of each cluster obtained represents each user group, which creates the ability to model users' different preferences.

3. DATA SET

In order to illustrate our design, we select one popular CQA: Stack Overflow to explore the attributes reflecting active users and examine the utility of these attributes in identifying interesting user segments. We downloaded the data dump from its official website[8] and parsed the data set with regards to user personal information, questions, and answers. Since we conduct the use segmentation with active users who are more involvedin the community, we select users with reputation scores above the sample median for our future experiment. Experiments on this data set will be conducted in the next phase of our research.

4. CONCLUSION

In this work, we tap on the potential of user segmentation in CQA for efficient question recommendation and user management. In order to build the profile for users, we sketch user from four perspectives: demographic, interests, answering patterns, and authority. By applying FCM technique, we tend to fuzzily subdivide users

[8]http://blog.stackoverflow.com/2009/06/stack-overflow-creative-commons-data-dump/

into different segments with different preferences. Being similar to FCM's application in direct marketing, CQA organizers match the question with each segment and can reduce system cost in question recommendation. Most importantly, organizers can adapt community's policies for users in each segment. We select dataset from a popular CQA – Stack Overflow– for future analysis.

REFERENCES

Alarcon-del-Amo, M. C. Lorenzo-Romero and M. A. Gomez-Borja (2011). "Classifying and Profiling Social Networking Site Users: A Latent Segmentation Approach." *Cyberpsychology Behavior And Social Networking* 14(9): 547-553.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent dirichlet allocation," *the Journal of machine Learning research* (3), pp 993-1022.

Dror, G., Koren, Y., Maarek, Y., and Szpektor, I. Year. "I want to answer; who has a question?: Yahoo! answers recommender system," Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM2011, pp. 1109-1117.

Ozer, M. 2001. "User segmentation of online music services using fuzzy clustering," *Omega* (29:2) 4//, pp 193-206.

Pal, A., Farzan, R., Konstan, J. A., and Kraut, R. E. 2011. "Early detection of potential experts in question answering communities," in *User Modeling, Adaption and Personalization*, Springer, pp. 231-242.

Pal, A., and Konstan, J. A. Year. "Expert identification in community question answering: exploring question selection bias," Proceedings of the 19th ACM international conference on Information and knowledge management, ACM2010, pp. 1505-1508.

Seret, A., Verbraken, T., Versailles, S., and Baesens, B. 2012. "A new SOM-based method for profile generation: Theory and an application in direct marketing," *European Journal of Operational Research* (220:1) 7/1/, pp 199-209.

Investigation on the relationship between China's social media and the purchases made on the e-Commerce platforms for better promotion strategies

Terence Chun-Ho Cheung
Department of Information Systems, City University of Hong Kong, Hong Kong SAR
is.tc@cityu.edu.hk

ABSTRACT

This paper summarizes some observations from the temporal trends of the discussion volume in China's social media against the actual purchase transactions made in China's e-Commerce platforms such as Taobao and Tmall. Such temporal trends could be used as a reference for better promotion strategies in any industry, for example, beauty industry chosen in this study, and for any campaigns in specified festivals. The temporal trends in the Q1 of both 2012 and 2013 could help demonstrate the informal communication among peer customers, i.e. electronic word-of-mouth, as kind of their social media footprints for obviously supporting actual purchases made in the periods of Chinese New Year and the 3.8 Women's Day in 2012 and 2013. Lastly, a hypothesis is set and tested that there is significant positive relationship between number of posts relevant to skin care topics in social media and actual purchases made in an e-Commerce platform in China.

1.INTRODUCTION

The increasing huge amount of social media discussions and posts reflects that the "big data" could affect various aspects of our daily activities and represents many opportunities to the online sellers and also scientists to address fundamental questions about the complex world we inhabit [1-2]. Collective human trading behavior in financial market could be predicted and illustrated by massive behavioral data from Google Trends [3]. Investigation on electronic word-of-mouth (eWOM) is significantly and positively related to the revenue of a movie (i.e. information exchange in Facebook) [4]. It is both the precursor and outcome of retailing [5], and the total box office revenue of a movie can be predicted by, for example, average frequency, peak frequency, total number of posts, number of positive and negative posts, and all their relationshipsare found to be positive to each other [6]. Branded-related comments and product-related comments from social networking sites on public display are found to help increase sales, improve brand image and encourage customer referrals [7]. Ultimately, they help generate large volume of sales in e-Commerce. In China's social media, up to 40% SinaWeibo (a Twitter alike social media platform) users use Alibaba'sTaobao, a famous B2C e-Commerce platform in China. About 2% share of Weibo's traffic to Taobao.com and of about 3.5% Weibo users will help spread and share their shopping experience from Taobao [8]. Thus, social media data, such as those obtained from forums, blogs, Weibo, is expected to be interactively and promptly beneficial to shopping experience on Taobao. However, how should the online sellers plan ahead for their promotion strategies?How many days in advance to be ready for seasonal sales?And how many days the sales last for, such as during Chinese New Year (CNY), 3.8 Women's day, 10.1 China National Day and 11.11 Singles Day? An investigation on the

relationship between the China's social media and the actual purchases made on the e-Commerce platformis carried out and several phenomena in Q1 2013is shared in this study. This can provide references for better promotion strategies during similar seasonal or Chinese festivals. The objective of this study is to investigate whether the number of posts regarding to a particular industry, for example, beauty industry, in China's social media can affect the actual transactions created on the China's e-Commerce platforms, such as Taobao.This paper first gives a literature review and describes our research methodology followed with the data collection, analysis and results. Lastly, a conclusion, limitation and future research direction will be given.

2. LITERATURE REVIEW

Recent studies on the relationship between online consumers' reviews and the sales in game industry indicate that the online consumers' reviews influence the sales in games, especially for less popular products [9]. Through text-mining analysis, positive relationships between blogging activities and box office revenue of movies are found [10]. Amount of blogging activities is suggested to forecast sales revenue, measure effectiveness of traditional marketing, conduct consumer evaluation and review the adoption of new product. However, empirical investigation on eBay shows that the online community participation had mixed effects on customers' likelihoods of participating in buying and selling behaviors [11]. This does not affect the participation on the number of bids placed or the revenue earned, but gives a negative impact of participation on the number of listings and the amount spent. Thus, research on product sales depends on the balance between positive word-of-mouth and negative word-of-mouth. For example, discovered pre-consumption word-of-mouth (tweets) owns a larger effect than post-consumption word-of-mouth because of the increase of awareness effect and the persuasive effect of recipients [12]. It is noted that negative relationship between advertising and online word-of-mouth among consumers is found [13]. Managers are suggested to set budget on these strategically because both of them will affect final sales. Therefore, it is important to estimate the trend of demand from the text-mined social media in order to capture the right audiences at the right time at the rapid expansion e-Commerce market. This paper is a preliminary investigation on the relationship between the discussion posts in China's social media and actual purchases in China's e-Commerce platforms, and followed by some temporal phenomena observed between the trend of discussion volume in social media and the trend of actual e-commerce transactions volume. This also reveals the trend of how the discussion in social media related to the actual searches and buying behavior of consumer, and thus be used to predict consumer decision process.

3. RESEARCH METHODOLOGY AND HYPOTHESES

3.1 Experiment design and data collection

A statistical analysis will be used to analyze whether there is a relationship between the number of posts and transactions in the first quarter of 2013. Discussion posts and transaction trends of 62 international brands (except Maybelline) were monitored, and retrieved from three main sources of China's social media channels, namely, Weibo, forums and blogs Q&A. These are the most important discussion-based media in China with very active participation from local users every day. The online transaction index are retrieved from thebiggest

B2C e-commerce platform in mainland China, namely, Taobao, that climbs up to 51.3% [14] online sales in 2013 Q1.

3.2 Analysis and results

Descriptive statistics

To first compare the trends of discussion volume and transactions in beauty industry in Q1 of 2012 and 2013, data is collected using hot term extraction technology. The total numbers of relevant topics, posts and authors in the three channels are shown in Table 1. It is noted that an obvious shift from postings in forums/blogs Q&A to Weibo and about 4.24 times increases in Taobao transactions. There were totally 66,397 posts from weibo, forums and blogs Q&A contributed by 35,235 unique authorsin 2013 Q1. In 2012 Q1, there were totally 100,722 posts from weibo, forums and blogs Q&A contributed by 34,378 unique authors. From Table 2, there are 675.62 posts in average per day with standard deviation 450.02 posts in 2013 Q1, while 1,097.01 posts in average per day with standard deviation 351.19 posts in 2012 Q1.

Q1	topics	posts	authors	# of people viewed	topics	posts	authors	Skin care [護膚]
Channels	Forums and Blogs Q&A (159 websites)				Weibo			Transaction index (Taobao) [min, max]
2013	12,329	36,664	11,440	25,062,767	11,800	29,733	23,795	[7, 390]
Channels	Forums and Blogs Q&A (197 websites)				Weibo			Transaction index (Taobao) [min, max]
2012	36,159	95,462	30,679	307,272,196	4,108	5,260	3,699	[5, 92]

Table 1. China's social media activities in Q1 of 2012 and 2013

Year Q1	Av. posts	Av. Authors*	Av. posts	Av. Authors*	Av. posts	Av. Authors*	Skin care [護膚]
2013 / Channels	Forums and Blogs Q&A (159 websites)		Weibo		ALL: Forums and Blogs Q&A (159 websites) and Weibo		Transaction index (Taobao)
Mean	407.38	171.41	268.24	304.26	675.62	475.67	142.88
Standard deviation	224.77	91.71	335.86	661.40	450.02	680.78	61.93
2012 / Channels	Forums and Blogs Q&A (197 websites)		Weibo		ALL: Forums and Blogs Q&A (197 websites) and Weibo		
Mean	1039.21	1049.03	57.80	48.81	1097.01	1097.85	47.49
Standard deviation	356.40	351.19	65.90	55.65	351.19	352.75	19.91

**Same author can post in different days in Q1 of 2012 or 2013.*

Table 2. Descriptive statistics of China's social media activities in Q1 of 2012 and 2013

China e-Commerce reached RMB$352.1 billion in 2013 Q1, representing a year-over-year increase of 36.6% [14]. The Taobao transaction index increases in an average of more than 200% in 2013 Q1 comparing to that of 2012 Q1. A significant drop before CNY holidays in both years. In 2012 Q1, the lowest transaction index and discussion volume were found on the last day of CNY (年廿九), i.e. Jan 22, whereas it resulted in a transaction index fall of 5 points and only 417 discussion posts in total. Similarly, in 2013, the quarter lowest transaction index and discussion volume were on Feb 9 Feb (年廿九), resulted a transaction index fall to 7 points and 133 discussion posts. The trend for both discussion and transaction volumes resumed after CNY in both years and reached a climax on the 3.8 Women's Day. Vivid seasonal patterns of CNY and 3.8 Women's Day in both trends of discussion and transaction volume in Q1 of 2012 and 2013 can be seen in Fig. 1a. and 1b. Both trends dropped below the normal level of an average of 4 days before the start of CNY holidays. The transactions and discussion volume resumed after the 4th day of CNY (年初四). In 2012, it took 7 and 9 days respectively for the transactions and discussion volume resumed to normal levels while a faster recovery of about 6 days was noted for both trends in 2013.

Types of discussion on different categories of cosmetic products

In 2013 Q1, the number of relevant posts regarding "advices seeking" and "reviews sharing" were 5,331 and 10,648 extracted from 107 websites and 129 websites respectively.

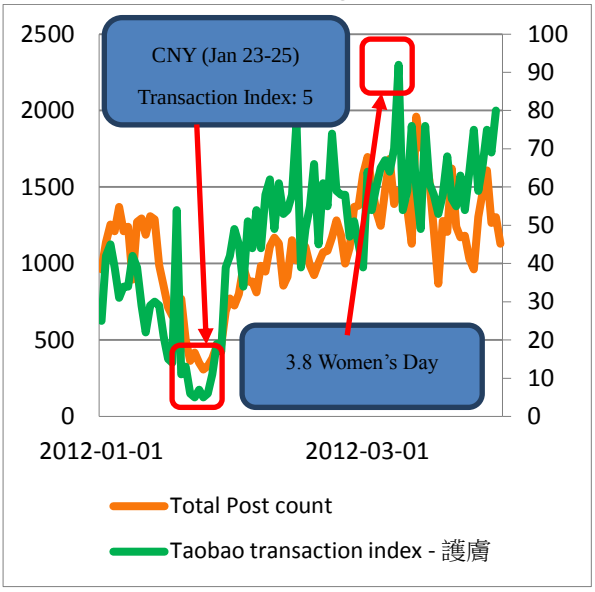
	topics	posts	authors	# of people viewed	topics	posts	authors	# of people viewed
Types	Advices seeking				Reviews sharing			
Channels	Forums and Blogs Q&A (107 websites)				Forums and Blogs Q&A (129 websites)			
2013 Q1	3,321	5,331	3,056	4,250,178	6,825	10,648	6,702	5,659,906

Table 3. Types of discussion extracted from China's social media activities in 2013 Q1

Discussion types	Advices seeking	Reviews sharing	Advices seeking	Reviews sharing
Rank/Channels	Weibo		Forums and blogs (107 websites)	Forums and blogs (129 websites)
Top 1st	*Cosmetic Tools (32%)*	*Cosmetic Tools (38%)*	Lotion & Cream	Facial Care Set
Top 2nd	*Facial Care Set (20%)*	*Facial Care Set (16%)*	Facial Care Set	Facial Mask
Top 3rd	*Facial Mask (12%)*	*Facial Mask (12%)*	Facial Mask	Lotion & Cream
Top 4th	*Lotion & Cream (6%)*	*Lotion & Cream (8%)*	Eye Care	Cosmetic Tools
Top 5th	BB Cream (6%)	Sun Block (7%)	Sun Block	Facial Cleansing

Table 4. Discussion types ranking in Weibo against that in Forum/blogs Q&A in 2013 Q1

In general, the ranking of relevant posts regarding both advices seeking and reviews sharing is consistent in Weibo while the 5 similar hot topics (e.g. Lotion & Cream, Facial Care Set) are ranked high in forums and blogs Q&A. Interestingly, "Cosmetic Tools" was not ranked within the top 5 in the type of "advices seeking" out of the 107 forums and blogs Q&A.



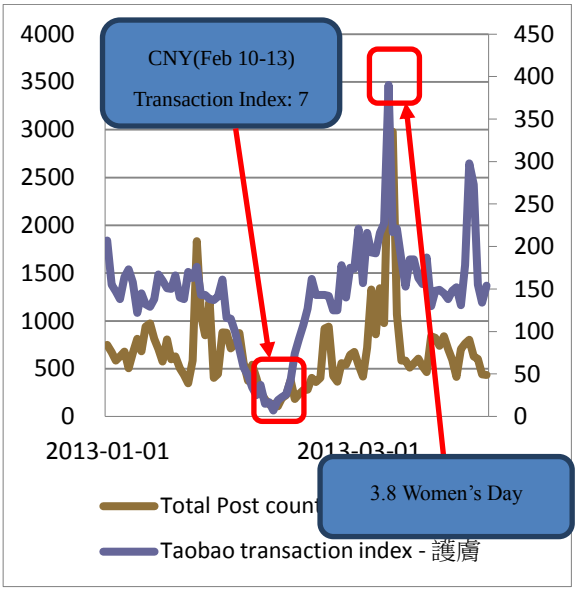


Fig. 1a and 1b.Taobao transaction index and trends of number of posts in China's social media such as Weibo, discussion forum and blogs Q&A in Q1 of 2012 and 2013.

Statistical Analysis on the Relationship between discussion volume and transaction index

The research question in this paper is whether the number of posts has relationship to the Taobao transaction index, particularly in 2013 Q1. Based on the scatter plot in Fig. 2, we can deduce a positive linear relationship between discussion volume and transaction index. Even though the association is not perfectly strong, the transaction index is apparently increasing with the online post count. In order to further investigate their relationship, statistical method would be applied in the following. The results could provide more accurate figures for the verification.

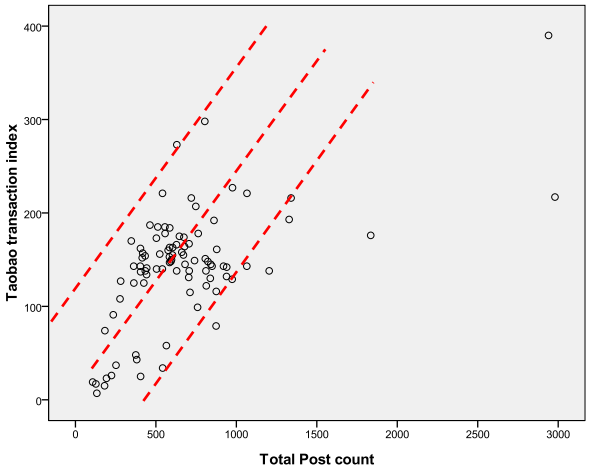


Fig. 2. Scatter Plot of daily no. of posts and daily transaction index in 2013 Q1

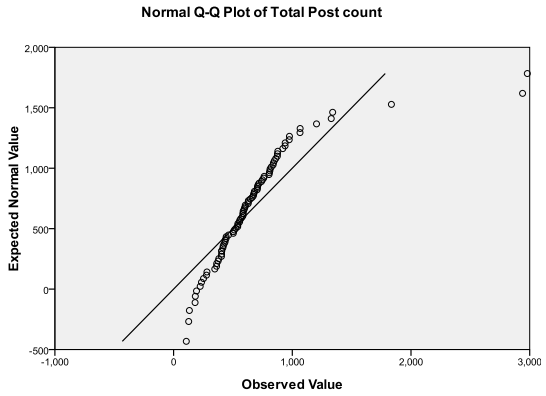


Fig. 3a Q-Q Plot of 2013 Q1 Daily No. of Posts

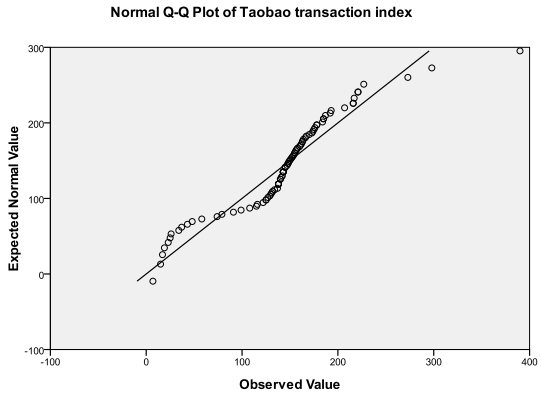


Fig. 3b Q-Q Plot of 2013 Q1 Daily Transaction Index

In Fig. 3a.and 3b, the two parameters (Total Post Count and Transaction Index) do not seem to be normally distributed. Both are deviated from the normal straight line to a large extent. As no assumption could be made about their population distributions, the non-parametric method, Spearman's rho, is selected for testing the strength of their correlation, i.e. hypothesis is made as below: H0: Correlation = 0 and H1: Correlation ≠ 0

			Total post count	Transaction index
Spearman's rho	Total post count	Correlation coeff	1.000	0.449**
		Sig. (2-tailed)	.	0.000
		N	90	90
	Transaction index	Correlation coeff	0.449**	1.000
		Sig. (2-tailed)	0.000	.
		N	90	90

Table 5. Test for significance using Spearman's Rank Correlation

As the correlation coefficient is **+0.449.**This reconfirms the positive relationship between discussion volume and transaction index. The p-value is very close to **zero**. At the 0.01 level of significance**, H_0 is rejected. We can therefore conclude that there is significantly positive association between discussion volume and transaction index. This is a clue that the current China's social media has a positive impact on the

consumer purchases in the beauty industry.

4.CONCLUSION, LIMITATION AND SUGGESTION FOR FUTURE RESEARCH

This paper describes and analyzes some temporal phenomena observed from the trends of discussion posts in China's social media and actual purchases in China's e-Commerce platform Taobao, and also shares a preliminary investigation on their relationship. It also reveals the trend of how the discussion in social media related to the actual searches and buying behavior of consumer, if possible, and thus be used to predict consumer decision process. A limitation in the research is whether the posts can explicitly manifest themselves in e-Commerce transactions. In addition, advices seeking and reviews sharing on various product categories can also be further investigated in details, for example, to find the relationship between Taobao transaction index and Weibo, and compared to the relationship between Taobao transaction index and forums, blogs Q&A, and the shifts. These can be the future research directions.

Acknowledgement: The hot-term extraction technology from discussion posts related to the beauty industry described in this paper is supported by K-Matrix Digital Intelligence Ltd.

References

1. N. A. Christakis and J. H Fowler, "*Connected: The surprising power of our social networks and how they shape our lives*," Hachette Digital, Inc., 2009.
2. G. King, "*Ensuring the data-rich future of the social sciences,*"Science (Washington) 331.6018 (2011): 719-721.
3. T. Preis, H.S. Moat and H.E. Stanley, "*Quantifying trading behavior in financial markets using Google Trends,*"Scientific reports 3 (2013).
4. J.P. Lo, "*The effectiveness of WOM by using Facebook as an implementation in movie industry*," 31 Aug. 2010.
5. W. Duan, B. Gu, A.B. Whinston, "*The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry*,"Journal of Retailing,84.2 (2008): 233-242.
6. V Kuznetsov, A Semyonov, "*Evaluation of effectiveness of an advertisement campaign by use of Twitter*,"American Academic & Scholarly Research Journal (AASRJ)5.5 (2013).
7. E.D. Spiegler, C. Hildebrand, F. Michahelles, "*Increasing Brand Attractiveness and Sales through Social Media Comments on Public Displays–Evidence from a Field Experiment in the Retail Industry*,"Pervasive Computing (2012): 443-460.
8. "*Alibaba and SinaWeibo Jointly Explore Social E-Commerce*," published in iResearch View (http://www.iresearchchina.com/views/5105.html) on 20 Aug 2013.
9. F. Zhu, X. Zhang, "*Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics*," Journal of Marketing 74.2 (2010): 133-148.
10. H. Onishi, P. Manchanda, "*Marketing activity, blogging and sales*,"International Journal of Research in Marketing29.3 (2012): 221-234.
11. R. Algesheimer, S. Borle, U.M. Dholakia, "*The impact of customer community participation on customer behaviors: An empirical investigation*," Marketing Science 29.4 (2010): 756-769.
12. H. Rui, Y. Liu, A Whinston, "*Whose and what chatter matters? The impact of tweets on movie sales*," The Impact of Tweets on Movie Sales (October 1, 2011) .NET Institute Working Paper 11-27 (2011).
13. J. Feng, P. Papatla, "*Advertising: Stimulant or Suppressant of Online Word of Mouth?"* Journal of Interactive Marketing 25.2 (2011): 75-84.

14. “国美 Q1 扭亏获净利 7633 万电商销售增 51.3%”, published in e-Commerce (http://www.iecnews.com/B2C/2013/0531/12724.html) on 31 May 2013.

Survival Analysis on Hacker Forums

Xiong Zhang Chenwei Li
Xiong.Zhang@my.cityu.edu.hkchenweli@cityu.edu.hk
Department of Information Systems, City University of Hong Kong, Hong Kong

Abstract: One representative hacker forum is selected to analyze users' reputation earning behaviors. Our analysis results suggest that both post-centered characteristics and network-centered characteristics are important for users to earn high reputation in the online forum. To earn high reputation, users are suggested to write more posts in detail to offer help to others and try to interact with others in more diverse topics.

Keyword: Learning Behavior, User Reputation, Survival Analysis, Social Network Analysis.

1. Introduction

Piracy loss due to hacking has increased significantly [1]. Hackers are always considered as mysterious in gray world[2]. Holt[3] refers to a hacker as "any individual with a profound interest in computers and technology that has used this knowledge to access computer systems with or without authorization from the system owners". Online forums are the important venues for hackers to learn, to communicate, even to collaborate for an attack [4]. As a social network, hackers interact with others to earn reputations in online cummunity. This motivate us to study hacking-related phenomenon in social network perspective. Both white-hat hackers and black-hat hackers may exist in the hacker forum. Some hackers are very knowledgeable and are very active in the forum. However, only a small number users can earn high reputation in the online forum compared with most other users of low reputation. In other words, in terms of high user reputation, only a small number of users can survive in the forum.Survival analysis is usually used in the analysis of time to event, such as the death of biological organisms and failure in mechanical systems[5]. Although we can't equate the high reputation with survival of hackers in the online forum, in this paper, we try to apply survival analysis to study the characteristics high reputation hackers should have.Although this online hacking forum doesn't kick users out, the forum has the policy that users can't ask for score from others without making contribution, therefore, it's not easy for users to gain high reputation. Users have to interact with others and contribute to the forum to earn respect and reputation from others.By studying hackers' reputation, we expect to study how users can gain higher reputation and what are the characteristics of high reputation hackers. The analysis result of this paper can be the foundation from which we can extend to study hacking's impact on firms and organizations.

In this paper, we try to study users' survival characteristics[6] in terms of high user reputation based on the raw data downloaded from a representative hacker forum. More specifically, we are interested in what would make a user a high-reputation hacker in online forum and how could they survive in this virtual community in terms of earning high reputation. We try to study this research question in both post content perspective and social network perspective. We do survival analysis using Cox proportional hazard regression model[7]. The

analysis results would provide insights on how users can survive and earn high reputation in the online forum.

2. Data Collection

The hack forum we choose has over six years of history. Up to now, this forum has more than 350 thousands registered members, in which, over 145000 members are pretty active. These members have made a total of more than 23 million posts in over 2 million threads. There are at most 2885 users online at the same time on July 22, 2012. In this forum, different topics, such as techniques, tutorials, movies, aggregate into dozens of sub-boards, of which 23 hack-related sub-boards focus on hacking related technologies, such as hacking tools, hacking tutorials, hacking issues in specific areas, proxies, decryption and so on. The raw data was downloaded from this forum in the time range of February 2007 to April 2012. From the raw data, we select the users who post more than 10 times, either initializing a thread or replying others. Those users who post less than 10 times in 6 years are out of our interest since not enough posting behaviors can be obtained for those users. We select users who comment or get comments for more than 3 times to remove those users who don't get recognized in this community. Finally, we have 1112 users and their 127628 posts.

3. Model and Implementation

The panel of posting network includes the threads users initialize and posts they reply to others in different sub-boards. Post content is one of the key indicators of hacking knowledge of users. Post content is also the basis on which other privileged users give feedbacks and opinions. We propose that whether a post is written to offer help to others may have important influence on the comment authors can get. Therefore,we built a text mining system[8][9] to automatically classify post content in the database so that each post can be assigned to one of the two classes of help offering or not. In this paper, we consider the value assigned to each post by this system as the level of offering help of each post. The length of post content can serve as the level of user generated content.

Different posting networks exist in this forum on the basis of users' interests. Users who post consistently in one or several certain sub-boards more easily get recognized and furthermore get positive feedbacks from other senior users. We propose a measurement about user i's sub-board loyalty as follows, which is higher when user i concentratemore on certain sub-boards.

$$Subboard\_loyalty_i = \sum_{j \neq k} (p_j - p_k)^2$$

In which p_j in the percentage of user i's posts in sub-board j.

Users belong to five different clubs, among which, the users in the highest club can assign score within the largest range to others and the users in the lowest club can't assign any scores to others. According to this power hierarchy, we denote the highest club as level 4 and the lowest club as level 0.

Besides, we also analyze the online forum in a social network perspective. Between-ness

centrality[10][11]quantifies the times a user serves as bridge along the shortest path between two other users. Between-ness centrality is defined as

$$C\_betweenness_i = \sum_{s \neq t \neq i \in V} \frac{\delta_{st}(i)}{\delta_{st}}$$

In which δ_{st} is the total number of shortest paths from user s to user t and $\delta_{st}(i)$ is the number of those paths that pass through user i. We normalize between-ness centrality such that $C\_betweenness_i \in [0,1]$ by dividing through the number of pairs of users in the network not including user i.

In this paper, we consider all these factors as independent variables for survival analysis. The descriptive analysis of the variables used in this model is presented in Table 1.

Table 1: Descriptive Analysis of Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Comment count	1112	16.60072	39.24784	0	419
Header count	1112	11.60162	15.966	0	143
Reply count	1112	103.1718	162.0178	3	1944
Club membership	1112	0.803058	0.952457	0	3
Avg. offerhelp	1112	-0.900171	0.013936	-0.96455	-0.82321
Avg. post length	1112	199.6682	150.9032	41	1913
Sub-board loyalty	1112	6.432228	4.383272	1.0608	22
Between-ness centrality	1112	8.26E-05	0.000304	0	0.00426
Score per comment	1112	0.452338	0.497947	0	1

The correlation analysis between independent variables is presented in Table 2.

Table 2: Correlation Analysis of Variables

	1	2	3	4	5	6	7	8
Comment count	1							
Header count	0.2041	1						
Reply count	0.3381	0.4138	1					
Club membership	0.5681	0.0537	0.1695	1				
Avg. offerhelp	0.0499	0.0675	0.2216	-0.0003	1			
Avg. post length	-0.0691	0.0946	-0.0028	-0.0521	0.11	1		
Sub-board loyalty	-0.1332	0.0602	0.0947	-0.1283	0.0469	0.1165	1	
Between-ness centrality	0.8811	0.2607	0.3939	0.3472	0.0777	-0.0473	-0.0904	1

4. Analyses and Results

In this paper, we denote users whose score per comment is higher than the median of score per comment for all users as the users who survive in the online forum, in terms of high reputation. The survival time is defined as the time range between the date when the first post and the date when the last post wrote by users. The analysis results are shown in Table 3.

From the results of cox regression, the regression coefficients of comment count, header post count and sub board loyalty are positive. These factors serve as risk factors for users to earn high reputation in the online forum. On the other hand, the regression coefficients of reply post count, club membership, the average level of help offering, the average length of post content and the between-ness centrality are negative. These factors serve as protective factors for users to earn high reputation in the online forum.

Especially, our results indicate users who open more threads tend not to earn high reputation

in the online forum more easily, while users who reply more frequently tend to gain positive feedback from others in the forum. When we focus on post content, we find that users who reply to offer help to others can earn higher score compared with others in the forum. Besides, posting posts in much details can also help authors to gain positive feedback in the virtual community.

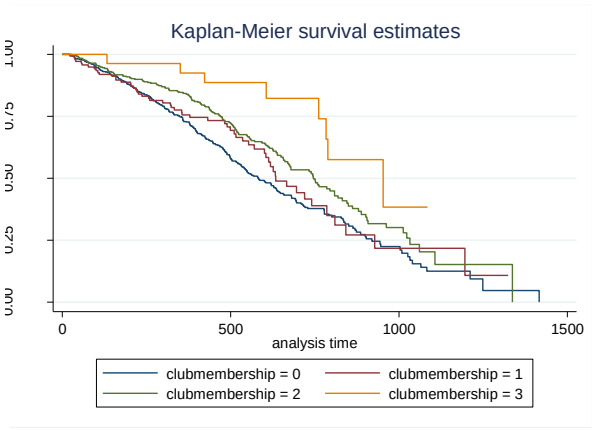
Table 3:Cox Regression Analysis

VARIABLES	50%	Z
Comment count	0.00263	0.00362
Header count	0.00877***	0.00308
Reply count	-0.00394***	0.000623
clubmembership	-0.0885	0.0647
Avgofferhelp	-5.169	3.278
Avg post length	-0.000234	0.000304
Sub-board loyalty	0.0303***	0.0101
Betweennesscentrality	-1,173*	658.1
Observations	1,112	1,112

*** p<0.01, ** p<0.05, * p<0.1

Besides, users are also advised not to post in only a few sub boards. Instead, users should extend their interests and write posts in other sub-boards such that they have the chance to interact with more users. In the social network perspective, users who play as the bridges among others can earn higher reputation in the forum compared with other users.

Figure 1: Kaplan-Meier Survival Estimates



To validate the robustness of the regression results, we take the factor club membership, which is not a significant factor, as an example to give the Kaplan-Meier survival estimates under different memberships. The Kaplan_Meiersurvivial curve is defined as the probability of surviving in a given length of time while considering time in many small intervals[12].The Cox regression result shows that focal users survive for longer time in terms of higher reputation when these users are in a higher club membership. This conclusion is also verified by the survival curve. Figure 1 shows the failure rates for users in different levels of club memberships. As we can see, users of higher level of club membership tend to live longer in the online forum in terms of higher reputation.

5. Conclusion and Future Work

In this paper, we try to study how users can earn high reputation in a hacker forum. We download the raw data from a representative hacker forum and do survival analysis using Cox proportional hazard regression model. Our regression results suggest that users should reply detailed posts to help others so that they can earn positive feedback in the forum more easily. In the social network perspective, users are also advised to broaden their interests and try to serve as the bridge among user interactions in the virtual community. The current work can be extended in following directions. First, our model will be verified in a larger database to check the robustness of our results. Second, we will further investigate how these risk factors and protective factors influence firms and organizations in business.

References

[1] G. Farrell and M. A. Riley, "Hackers Take $1 Billion a Year as Banks Blame Their Clients," Bloomberg, 5 August 2011. [Online]. Available: http://www.bloomberg.com/news/2011-08-04/hackers-take-1-billion-a-year-from-company-accounts-banks-won-t-indemnify.html. [Accessed 24 11 2013].

[2] Symantec, "2013 Norton Report: Cost per Cybercrime Victim Up 50 Percent," Symantec, Mountain View, Calif., 2013.

[3] T. J. Holt, "Hacks, cracks, and crime: An examination of subculture and social organization of computer hackers," *University of Missouri-St. Louis, Ph.D Dissertation,* 2005.

[4] V. Mookerjee, R. Mookerjee, A. Bensoussan and W. T. Yue, "When Hackers Talk: Managing Information Security Under Variable Attack Rates and Knowledge Dissemination," *Information Systems Research,* vol. 22, no. 3, pp. 606-623, 2011.

[5] "Wikipedia," [Online]. Available: http://en.wikipedia.org/wiki/Survival_analysis. [Accessed 24 11 2013].

[6] J. F. Lawless, Statistical Models and Methods for Lifetime Data, 2nd ed, Wiley-Interscience, 2002.

[7] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society, Series B (Methodological),* vol. 34, no. 2, pp. 187-220, 1972.

[8] T. Joachims, Learning to Classify Text Using Support Vector Machines-Methods, Theory and Algorithms, Kluwer Academic Publishers, 2002.

[9] M. F. Porter, "An Algorithm for Suffix Stripping," *Program,* pp. 130-137, 1980.

[10] K. Faust, "Centrality in Affiliation Networks," *Social Networks,* pp. 157-191, 1997.

[11] L. Freeman, "A set of measures of centrality based on betweenness," *Sociometry,* vol. 40, no. 1, pp. 35-41, 1977.

[12] A. DG, Analysis of Survival times.In:Practical statistics for Medical research, London: Chapman and Hall, 1992.

OWSDR: An Ontology-based Web Service Discovery and Selection System

Wenping Zhang and Raymond Y.K. Lau
Department of Information Systems
City University of Hong Kong
Hong Kong SAR
E-mail: {wzhang23, raylau}@cityu.edu.hk

Abstract: With the rapid growth of Web of Things (WoT) and corresponding Web services, there is a pressing need to develop an effective computational method for services discovery and recommendation. Despite personalized services discovery methods have been studied before, few attempts have been made to explore ontological user profiling and probabilistic language modeling approach for Web service contextualization and ranking. This paper makes a novel contribution in terms of developing an ontology-based user profiling method to improve the Web service discovery and recommendations. In particular, a novel probabilistic language model is developed to conduct Web service contextualization and ranking. Our preliminary experimental results reveal that the proposed service personalization approach outperforms a classical baseline method.
Keywords: Services Discovery, Ontological User Profiling, Language Modeling.

1. INTRODUCTION

The fundamental problems of Web services discovery are about the representations of the semantics pertaining to service queries and resources, and the prediction of the relevance of the target resources (services) with respect to a query [17]. With the rapid growth of WoT [5] and the Semantic Web [4], personalized services discovery and recommendation has become a hot research topic [8]. For the paradigm of the semantic Web, ontology has been playing a key role in formal knowledge representation to facilitate human and computer interactions [2,4]. Ontology refers to a formal specification of conceptualization; it may take the simple form of a taxonomy of concepts (i.e., light-weight ontology), or the more comprehensive representation of comprising a taxonomy as well as the axioms and constraints which characterize some prominent features of the real-world (i.e., heavy-weight ontology) [10]. Domain ontology is one kind of ontology which is used to represent the knowledge for a particular type of application domain, and it can be expressed by using formal semantic markup languages such as OWL [11].

Although ontology has been playing a key role in Semantic Web, leveraging ontology to enhance Web services personalization is a relative new topic. This paper illustrates a novel design and development of an ontology-based personalized service discovery and selection model. In particular, ontological user profiling is applied to capture users' possibly changing service requirements, and probabilistic language modeling is exploited to develop an effective mechanism for service query contextualization based on both current and past search service invocation history. As user queries are usually short e.g., around 2 words long on average [7],

query personalization and contextualization is essential for effective Web service discovery. Our preliminary experiments have shown that the proposed ontological user profiling and probabilistic language modeling methods are promising.

2. SYSTEM ARCHITECTURE

The system architecture of the Ontology-based Web Service Discovery and Recommendation System (OWSDR) is depicted in Figure 1. Users interact with their intelligent client stubs that will in-turn look up the most relevant Web services from the external service registries such as UDDI. The Query Processing and Logging module of OWSDR is responsible for accepting users' queries and managing the query and service invocation histories. After a query is accepted to OWSDR, the Query Contextualization module will refine the original query by referring to the specific user profile. Probabilistic language model is then applied to infer the most relevant service context for the query according to the user's ontological profile.

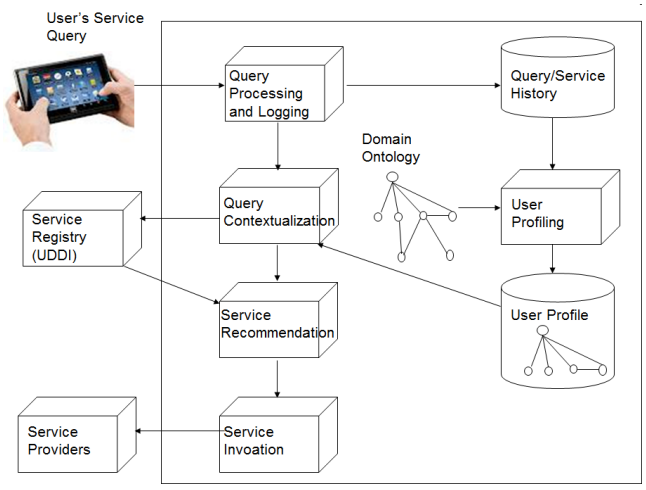


Figure 1. General System Architecture of OWSDR

After query contextualization, the refined query is sent to the external service registries for Web services selection. Potentially matching services are then returned to the Service Recommendation module of OWSDR for service ranking with respect to the contextualized query. The Service Invocation module selects the most relevant service in the ranked list and communicates with the external service provider. On the other hand, the User Profiling module is responsible for user profile creation and revision. An ontological user profile is first instantiated based on the ODP taxonomy. With reference to the query and service invocation history, the ontological user profile is updated to reflect the user's most recent interests. Our prototype system was developed using Java (J2SE v 1.4.2), Java Server Pages (JSP) 2.1, and Servlet 2.5 and operated under Apache Tomcat 6.0.

3. SYSTEM EVALUATION

To evaluate the effectiveness of the OWSDR system, we develop a collection of service descriptions with respect to 10 categories of the ODP ontology such as exchange, business, software, weather, health, stock, football, book, game, and education. By searching common Web services portals such as xmethods.net, webservicex.net, seekda.com, and so on, we collected 55 Web service descriptions for each category. Figure 3 is a snapshot view of the service descriptions extracted from xmethods.net. As a result, our dataset consists of 550 service descriptions (documents) for 10 ODP categories. For instance, for the weather category, the corresponding retrieved web service description is "Provide weather forecast for a given place". The performance evaluation metric used was top-*n* precision, where *n*=50 was applied in our experiment.

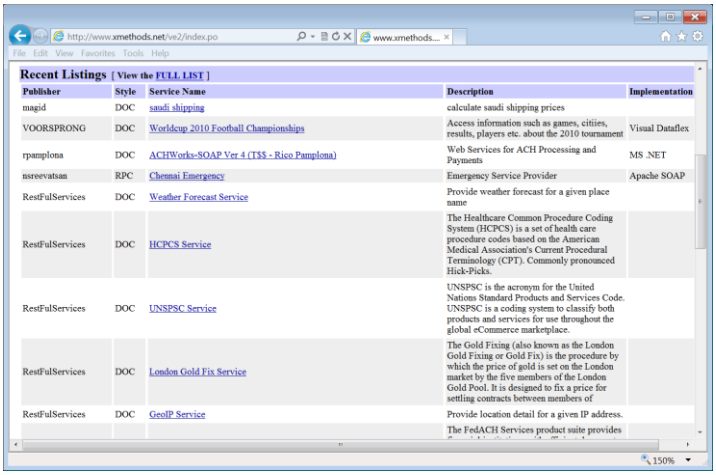


Figure 2. Web Service Descriptions at xmethods.net

For each service domain (ODP category), we applied five relevant service descriptions to train the user profiling module of the OWSDR system and establish the corresponding user profile. The keywords of the category such as "exchange" was taken as the raw query and passed to the contextualization module for expansion. The service recommendation module was then invoked to rank the remaining service descriptions. The top fifty service descriptions ranked by the ranking module was used to evaluate the performance (e.g., precision and recall) of the OWSDR system. A baseline system was also developed based on the vector space model and the cosine similarity measure [18]:

$$CosSim(q,d) = \frac{\sum_{i=1}^{|T|} w_{qi} \times w_{di}}{\sqrt{\sum_{i=1}^{|T|} (w_{qi})^2} \times \sqrt{\sum_{i=1}^{|T|} (w_{di})^2}} \quad (1)$$

where w_{qi} represents the TFIDF weight of the *i*th term in the query *q*, and w_{di} is the TFIDF weight

of the *i*th term in the service description d respectively. The set $T = \{t_1, t_2, ..., t_{|T|}\}$ represents the set of terms (i.e., the vocabulary) of our entire dataset.

For the baseline system, user profiling was conducted based on the vector space model and query contextualization was performed based on each user vector. In addition, profile revision was based on the classical Rocchio learning method [18] and the baseline system applied the cosine similarity function to rank the service descriptions with respect to each service category. Our preliminary experimental results are reported in Table 1. The average improvement in terms of top-*n* precision over the ten service categories is +21.77%.

Category	Baseline	OWSDR	Improvement
book	0.58	0.68	+17.24%
health	0.52	0.62	+19.23%
business	0.54	0.66	+22.22%
education	0.48	0.58	+20.83%
weather	0.56	0.64	+14.29%
exchange	0.44	0.56	+27.27%
football	0.46	0.58	+26.09%
game	0.48	0.54	+12.50%
stock	0.54	0.60	+11.11%
software	0.62	0.72	+16.13%
Average	**0.51**	**0.62**	**+21.77%**

Table 1. Comparative Performance of OWSDR

The main reason of such a significant performance improvement brought by OWSDR is due to the dynamic user profiling and service query contextualization processes. In particular, the semantics of service queries and service descriptions are taken into account while Web services are evaluated. Moreover, with a personalized and contextualized service query, more accurate service matching is performed by the system. The end result is that the overall effectiveness of Web service discovery and recommendation is improved.

4. CONCLUSIONS AND FUTURE WORK

With the rapid proliferation of WoT and semantic Web services, intelligent and personalized services discovery and recommendation methods are desirable for practical deployment of Web services in real-world application contexts. Even though personalized services discovery has been examined by researchers before, few attempts have been made to exploit ontological user profiling and probabilistic language modeling for service contextualization and service ranking. This paper makes a novel theoretical contribution in the sense that domain ontology and probabilistic language modeling have been applied to design an effective computational

methodology to enhance service discovery and recommendation in the real-world contexts. Our preliminary experimental results show that the proposed method is promising and it outperforms a classical baseline method by 21.77% in terms of top-*n* precision. Our future work will further examine and evaluate the dynamic user profiling method based on incremental users' relevance feedback. A large-scale usability study against the proposed system will also be conducted.

ACKNOWLEDGMENT

The work reported in this paper has been funded in part by the Strategic Research Grants of City University of Hong Kong (Project No. 7004120, 7003002, 7008138) and the Shenzhen Municipal Science and Technology R&D Funding - Basic Research Program (Project No. JCYJ20130401145617281).

References

[1] Abbar, S., et al. A Personalized Access Model: Concepts and Services for Content Delivery Platforms. Proceedings of the 10th International Conference on Information Integration and Web-based Applications & Services. 2008.

[2] Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web. Scientific American, 284(5):34-43, 2001.

[3] Bolchini, C., Quintarelli, E., and Rossato, R. Relational Data Tailoring Through View Composition. Proceedings of the 26th International Conference on Conceptual Modeling, 2007.

[4] Burstein, M. and McDermott, D. Ontology Translation for Interoperability among Semantic Web Services. AI Magazine, 26(1): 71-82, 2005.

[5] He, J., Zhang, Y., Huang, G., and Cao, J. A smart Web service based on the context of things. ACM Transactions on Internet Technology, 11(3):Article 13, 2012.

[6] Hsieh, P.-H. and Yuan, S.-T. Dynamic Semantic Location Modeling in Mobile Enterprise Applications.Proceedings of the 5th International Conference on Electronic Commerce. 2003.

[7] Jansen, B. and Spink, A. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. Information Processing and Management, 42(1):248-263, 2006.

[8] Jiang, X. and Tan, A-H. Learning and inferencing in user ontology for personalized semantic web services. Proceedings of the 15th international conference on World Wide Web, pp. 1067-1068, 2006.

[9] Keahey, K. and Freeman, T. Contextualization: Providing One-Click Virtual Clusters. Proceedings of the Fourth IEEE International Conference on eScience. 2008.

[10] Lau, R.Y.K., Song, D., Li, Y., Cheung, C.H., Hao, J.X. Towards A Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning. IEEE Transactions on Knowledge and Data Engineering, 21(6): 800-813, 2009.

[11] Lau, R.Y.K., Lai, C.L., and Li, Y. Fuzzy Ontology Mining and Semantic Information Granulation for Effective Information Retrieval Decision Making. International Journal of Computational Intelligence Systems, 4(1):54-65, 2011.

[12] Licker, P.S., Getting Advice From a Computer: Five Models and A Simulator SIGMIS Database, 22(3):1-13, 1991.

[13] Liu, X. and Croft, B. Cluster-based retrieval using language models. Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.186-193, 2004.

[14] Milic-Frayling, N., et al. On the Design and Evaluation of Web Augmented Mobile Applications.Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services. 2007.

[15] Mrissa, M., et al., A Context-Based Mediation Approach to Compose Semantic Web Services. ACM Transactions on Internet Technology. 8(1):1-23, 2007.

[16] Pon-Barry, H., et al. Contextualizing Learning in a Reflective Conversational Tutor. Proceedings of the IEEE International Conference on Advanced Learning Technologies. 2004.
[17] Rong, W., Liu, K., and Liang, L. Towards Personalized Ranking in Web Service Selection. Proceedings of the 2008 IEEE International Conference on e-Business Engineering, pp. 165-172, 2008.
[18] Salton, G. and McGill, M. Introduction to Modern Information Retrieval. McGraw-Hill, New York, 1983.