December 2006

Organizational Adoption of Data Mining

Abhijit Dutt
University of Wisconsin-Milwaukee

Organizational Adoption of Data Mining

Abhijit Dutt
University of Wisconsin-Milwaukee
adutt@uwm.edu

ABSTRACT

Unlike other IT innovations, data mining is not only an IT innovation but also a research methodology. Adoption and use of data mining in an organization could potentially raise unique privacy and ethical issues. Drawing on prior literature in the area of organizational adoption of IT innovation, specifically, the "diffusion of innovation" perspective, and research in the area of privacy and ethics, we propose a new model for understanding adoption of data mining in organizations. Plans for our data collection (both qualitative and quantitative) efforts are also discussed. We believe that our model makes significant contribution to the current literature by identifying the unique role of some of the ethical and privacy issues on data mining adoption in organizations.

Keywords

Data mining, IT innovation adoption, Diffusion of innovation, knowledge creation, Privacy, Ethics.

INTRODUCTION

During the last decade rapid improvements in computing technology in general and in database technology in particular have made it possible for businesses to store a variety of information. Widespread organizational use of database systems has led to the availability of a vast amount of historical data in firms. In recent years, many organizations have started using "data mining" to extract meaningful patterns and rules from such data, and in some cases to predict a variable of interest. Data mining has been used in a number of different areas ranging from consumer behavior to predicting bankruptcy (Berry and Linoff 2004). In this paper, we propose a model that identifies different factors such as organizational characteristics and ethics/privacy related issues which should be considered prior to adopting data mining in organizations. In other words, drawing on Fichman's (2004) approach to model building in IT innovation research, we identify a direct set of antecedents of the adoption of data mining. We assume that each antecedent has an independent effect contributing to the explanation of the variance in the dependent variable.

INNOVATION AND ADOPTION OF INNOVATION

Innovation could be studied at different levels focusing on individuals, firms or even on entire industries. Our focus is at the organizational level, and we define innovation as "an idea or behavior new to the adopting organization" (Damanpour, 1996: p. 694). According to Swanson, IS innovation may be defined as "innovation in the organizational application of digital computer and communication technologies (now commonly known as information technology or IT)" (Swanson, 1994: 1072).

A review of the IS innovation literature suggests that a majority of the IS innovation adoption studies have focused on the innovation itself and has paid limited attention to how those innovations could bring changes to an organization (Lyytinen and Rose 2003). Studies have shown that there are fundamental differences between IS innovations (Swanson 1994; Grover et al. 1997), and each innovation and its effect on organizations should be studied separately. We argue that data mining is an unique IT innovation.

First, data mining allows users to uncover new knowledge from available data (Berry and Linoff 2005) and could thus be considered a research methodology. Most information systems help automate business processes or reduce human intervention. On the other hand, some information systems such as decision support systems, expert systems can also serve as a repository of expert knowledge. Unlike other IS, data mining can not only help in the analysis of data but also can be instrumental in the creation of new knowledge. This could potentially raise unique challenges such as sharing of knowledge.

Second, unlike other IS innovations, data mining can be very intrusive and thus raise privacy and ethical issues. In view of the existing privacy laws and the generally accepted ethical standards, businesses need to consider these issues carefully before adopting data mining.

Third, for most IS innovations, adoption by individual users is very important even if one is interested in adoption by an organization. However, individual adoption will not play such an important part in data mining adoption although individual adoption behavior could play a significant part.

THE DATA MINING ADOPTION MODEL

The diffusion of innovation (DOI) theory, that has been used extensively in prior research, focuses on the perceived characteristics of the innovation (Chwelos et al., 2004). In the MIS literature, there has been numerous theoretical perspectives explaining individual acceptance/adoption of information systems (e.g., TAM, DTPB). Venkatesh et al. (2003) integrated constructs from different models and introduced the unified theory of acceptance and use of Technology (UTAUT). Our model of data mining adoption draws on the DOI theory instead of TAM or UTAUT, because of the following reasons. First, we are interested in organizational adoption instead of individual adoption, and most organization adoption studies have taken DOI perspective. Second, there are many similarities between TAM and Diffusion of Innovation theory and according to Fichman (1992) "TAM's perceived usefulness and perceived ease of use are essentially the same as diffusion theory's relative advantage and complexity" (page 202). Third, the main criticism about the DOI perspective is that it does not include organizational and environmental perspectives and those perspectives are important in the case of interorganizational systems. However, data mining is not an interorganizational system, and thus, the above criticism does not hold in this paper. Our preliminary model of data mining adoption is shown in Figure 1. Below, we discuss each construct in the model.

Data mining adoption

By data mining adoption we mean decision to use data mining methodology in a firm's day to day operation, and to obtain benefit from it. We believe that several factors such as characteristics of the organization, and concerns regarding privacy and ethics will affect adoption of data mining.

Strategic importance of IS

Prior research has suggested that organizations where IS is considered to be of strategic importance tend to adopt IS innovations (Swanson 1994; Grover et al. 1997). Thus, we posit that organizations where IS is considered to be of strategic importance will be more likely to adopt data mining. This leads to our first hypothesis:

H1: Strategic importance of IS in an organization is positively related to adoption of data mining.

Organizational absorptive capacity

"Absorptive capacity" is defined as an organization's capability to identify, assimilate, share and exploit knowledge from the environment (Cohen and Levinthal, 1990). In a recent work, Ko et al. (2005) studied knowledge transfer from consultants to clients during development and implementation of an Enterprise Resource Planning (ERP) system. They found a significant relationship between individual absorptive capacity and knowledge transfer. Ravichandran (2005) also emphasized the importance of "absorptive capacity," and drawing on prior research suggested that organizational absorptive capacity can explain variance in the adoption of IT innovations. Data mining can be instrumental in knowledge creation in organizations, and successful use of data mining will involve not only IT experts but also domain experts. This leads to our second hypothesis:

H2: Organizational absorptive capacity is positively related to adoption of data mining.

Anticipated benefits of data mining

In prior IS innovation studies (e.g., those involving EDI), anticipated benefits (perceived benefits) of the particular innovation was found to be of significant importance (Premkumar and Ramamurthy, 1995; Chwelos et al. 2001). This leads to our third hypothesis:

H3: Anticipated benefits of data mining in an organization are positively related to adoption of data mining.

Data Privacy concerns

While discussing data mining adoption it is necessary to discuss the underlying data on which data mining will be performed. Depending on the type of data which will be used, data mining could uncover some sensitive information. A business has to make sure that "privacy issues" are handled appropriately. We posit that a higher data privacy concerns will be detrimental to data mining adoption, leading to our fourth hypothesis:

H4: Data Privacy concerns of an organization are negatively related to adoption of data mining.

A construct related to this issue (i.e., "concern for information") was empirically tested by Stewart and Segars (2002). However, Greenway and Chan (2005) noted that there is limited research involving organizational privacy behaviors. They distinguished between organizational privacy and individual privacy, and identified three themes in the organizational privacy research: "information privacy as organizational liability, information privacy as an organizational decision outcome, and information privacy as an organizational ethical imperative" (page 175).

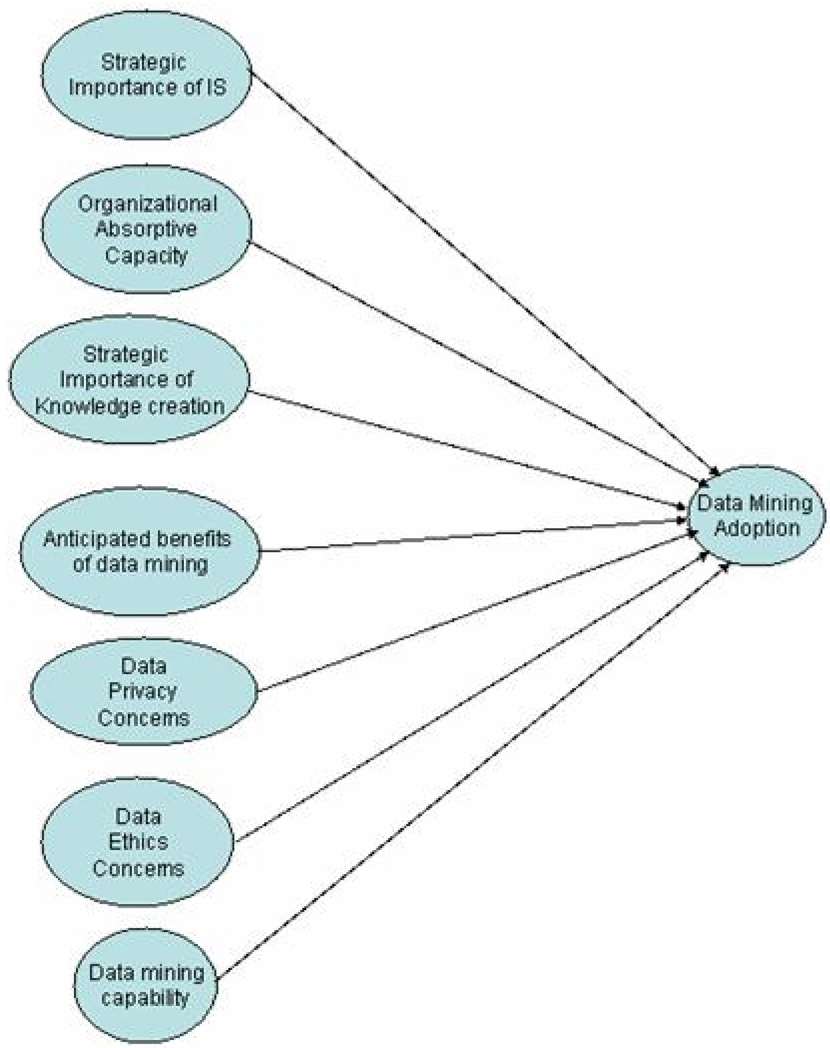

Figure 1: A Model for Data Mining Adoption

Drawing on prior research, it is possible to differentiate between different data sets where privacy will be important. As an example, "health data" of employees are more private than "IS use" data. We shall develop an instrument which will help us quantify the "Data Privacy concerns". There are many factors which will be important in determining "Data Privacy concerns". As an example, in the current global economy it is very important that we consider different environments under which a firm needs to operate depending on its geographical location. Most western countries have strict privacy regulations. Hence, depending on the country, same data could have different privacy concerns.

Data Ethics concerns

Data ethics concerns are also related to data privacy concerns. For example, if private and or sensitive information are uncovered about employees in an organization, it is very important that employees in that organization do not misuse such information and act ethically. Private information could be used by an organization (or employees in an organization) to discriminate against a group of employees. This is an example of individual behavior which could potentially be important in our study. Banerjee et al. (1998) developed a model for ethical behavior of IS personnel emphasizing the role of attitude, ethical behavior, and moral development. Ethical concerns in an organization will negatively affect adoption of data mining, leading to our fifth hypothesis

H5: Data ethics concerns in an organization are negatively related to adoption of data mining.

Strategic importance of knowledge creation

According to the knowledge-based perspective of organization, knowledge could be used as a strategic advantage (Davenport and Prusak 1998). Organizations who value knowledge and consider knowledge creation to be of strategic importance, will find data mining attractive. This leads to our sixth hypothesis:

H6: Strategic importance of knowledge creation in an organization is positively related to adoption of data mining.

Data mining capability

If an organization already possesses data mining capability then the cost for adopting data mining will be less. It could be argued that unless an organization adopts data mining an organization cannot develop data mining capability. However, it should be recognized that data mining is a multi-faceted innovation and data mining capability could be developed in many different ways. It is possible that an organization could have data mining capability before adoption of data mining. Expertise in the use of statistical software could translate into higher data mining capability. This leads to our seventh hypothesis

H7: Data mining capability is positively related to adoption of data mining.

Bharadwaj et al. (1999) defined and empirically verified a construct termed as “IT capability”. They found thirty distinct capabilities and those capabilities were classified into six different categories. Similarly, we shall develop a construct “data mining capability” which will help us measure how capable an organization is for undertaking data mining.

RESEARCH METHODOLOGY

First, instruments will be developed for measuring the new constructs introduced in our study. Second, for the constructs which are not new we shall adapt the available instruments. Third, we shall interview a few data mining consultants who will help us refine and calibrate our research model. Fourth, for testing the research model we shall use survey methodology with questionnaires. We shall identify companies who considered adoption of data mining.

MANAGERIAL IMPLICATIONS AND CONCLUSIONS

In this paper, we introduced a new model identifying the factors affecting the adoption of data mining in organizations. Our model also highlights the two dimensions of data mining, one as an IT system, and the other as an instrument for knowledge creation. Drawing on prior research, we also identified a few key issues such as privacy and ethical concerns that can pose as challenges during the adoption of data mining, because of its uniqueness. The results from this research could be used by organizations to quickly identify the issues they need consider while considering the adoption of data mining.

ACKNOWLEDGMENTS

The author expresses his gratitude to Dr. Saonee Sarker of Washington State University and Dr. K. Ramamurthy of University of Wisconsin-Milwaukee for helpful suggestions.

REFERENCES

1. Banerjee, D., Cronan, T. P., and Jones, T. W. (1998) Modeling IT Ethics: A Study in Situational Ethics, *MIS Quarterly*, 22(1), 31-60.
2. Berry, M. J. A. and Linoff, G. S. (2004) *Data Mining Techniques*, Second Edition, Wiley Publishing Inc.
3. Bharadwaj, A., Sambamurthy, V., and Zmud, R.W. (1999) IT Capabilities: Theoretical Perspectives and Empirical Operationalization, *Proceedings of the 20th International Conference on Information Systems*, 378-385.
4. Chwelos, P., Benbasat, I. and Dexter, A. S. (2001) Research Report: Empirical Test of an EDI Adoption Model, *Information Systems Research*, 12(3), 304-321.
5. Cohen, W.M. and Levinthal, D.A. (1990) Absorptive Capacity: A New Perspective on Learning an Innovation, *Administrative Science Quarterly*, 35 (1), 128-152.
6. Damanpour, F. (1996), Organizational Complexity and Innovation: Developing and Testing Multiple Contingency Models, *Management Science*, 42(5), 693-716
7. Davenport, T. H. and Prusak, L. (1998). *Working knowledge: how organizations manage what they know*, Harvard Business School Press, Boston, Mass.
8. Fichman, R.G. (1992) Information Technology Diffusion: A Review of Empirical Research, *Proc. of the 13th International Conference on Information Systems (ICIS*), Dallas, TX, 195-206.
9. Fichman, R.G. (2004) Going Beyond the Dominant paradigm for Information Technology Research: Emerging Concepts and Methods, *Journal of the Association for Information Systems*, 5(8), 314-355.
10. Greenaway, K. J. and Chan, Y. E. (2005) Theoretical Explanations for Firms' Information Privacy Behaviors, *Journal of the Association for Information Systems*, 6(6), 171-198.
11. Grover, V., Fiedler, K., and Teng, J. (1997) Empirical Evidence on Swanson's Tri-Core Model of Information Systems Innovation, *Information Systems Research*, 8 (3), 273-288.
12. Ko, D., Kirsch, L. and King, W. R. (2005) Antecedents of Knowledge Transfer from Consultants to Clients in Enterprise System Implementations, *MIS Quarterly*, 29(1), 59-85.
13. Lyytinen, Kalle, and Rose, Gregory. M. (2003) The Disruptive Nature of Information Technology Innovations: The Case of Internet Computing in Systems Development Organizations, *MIS Quarterly*, 27(4), 557-595.
14. Premkumar, G. and Ramamurthy, K. (1995) The Role of Interorganizational and Organizational Factors on the Decision Mode for Adoption of Interorganizational Systems, *Decision Sciences*, 26 (3), 303-336.
15. Ravichandran, T. (2005) Organizational Assimilation of Complex Technologies: An Empirical Study of Component-Based Software Development, *IEEE Transactions on Engineering Management*, 52(2), 249-268.
16. Stewart, K. A. and Segars, A. H. (2002) An Empirical Examination of the Concern for Information Privacy Instrument, *Information Systems Research*, 13(1), 36-49.
17. Swanson, E. B. (1994) Information Systems Innovation Among Organizations, *Management Science,* 40 (9), 1069-1092.
18. Venkatesh, V., Morris, M.G., Davis, G.B., and Davis, F.D. (2003) User Acceptance of Information Technology: Toward a Unified View, MIS Quarterly, 27 (3), 425-478.