Knowledge management and data mining for marketing

Michael J. Shaw [a,b,c,*], Chandrasekar Subramaniam [a], Gek Woo Tan [a],
Michael E. Welge [b]

[a] *Department of Business Administration, University of Illinois at Urbana-Champaign, Urbana, IL, USA*
[b] *National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign, Urbana, IL, USA*
[c] *Beckman Institute, University of Illinois at Urbana-Champaign, Room 2051, 405 N. Mathews Avenue, Urbana, IL 61801, USA*

Abstract

Due to the proliferation of information systems and technology, businesses increasingly have the capability to accumulate huge amounts of customer data in large databases. However, much of the useful marketing insights into customer characteristics and their purchase patterns are largely hidden and untapped. Current emphasis on customer relationship management makes the marketing function an ideal application area to greatly benefit from the use of data mining tools for decision support. A systematic methodology that uses data mining and knowledge management techniques is proposed to manage the marketing knowledge and support marketing decisions. This methodology can be the basis for enhancing customer relationship management. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Data mining; Knowledge management; Marketing decision support; Customer relationship management

1. Introduction

In recent years, the advent of information technology has transformed the way marketing is done and how companies manage information about their customers. The availability of large volume of data on customers, made possible by new information technology tools, has created opportunities as well as challenges for businesses to leverage the data and gain competitive advantage. Wal-Mart, the largest retailer in the U.S., for example, has a customer database that contains around 43 tera-bytes of data, which is larger than the database used by the Internal Revenue Services for collecting income taxes [10]. The Internet and the World Wide Web have made the process of collecting data easier, adding to the volume of data available to businesses. On the one hand, many organizations have realized that the knowledge in these huge databases are key to supporting the various organizational decisions. Particularly, the knowledge about customers from these databases is critical for the marketing function. But, much of this useful knowledge is hidden and untapped. On the other hand, the intense competition and increased choices available for customers have created new pressures on marketing decision-makers and there has emerged a need to manage customers in a long-term relationship. This new phenomenon, called customer relationship management, requires that the organizations tailor their products and ser-

---
* Corresponding author. Beckman Institute, University of Illinois at Urbana-Champaign, Room 2051, 405 N. Mathews Avenue, Urbana, IL 61801, USA. Tel.: +1-217-244-1266; fax: +1-217-244-8371.

E-mail address: m-shaw2@uiuc.edu (M.J. Shaw).

vices and interact with their customers based on actual customer preferences, rather than some assumed general characteristics [21,22]. As organizations move towards customer relationship management, the marketing function, as the front-line to interact with customers, is the most impacted due to these changes. There is an increasing realization that effective customer relationship management can be done only based on a true understanding of the needs and preferences of the customers. Under these conditions, data mining tools can help uncover the hidden knowledge and understand customer better, while a systematic knowledge management effort can channel the knowledge into effective marketing strategies. This makes the study of the knowledge extraction and management particularly valuable for marketing.

Developments in database processing [6,13,15,28], data warehousing [16,18], machine learning [4,12,25] and knowledge management [2,14,24] have contributed greatly to our understanding of the data mining process. More recent research on data mining and knowledge discovery [20,26,27] has further enhanced our understanding of the application of data mining and the knowledge discovery process. But, most research has focused on the theoretical and computational process of pattern discovery and a narrow set of applications such as fraud detection or risk prediction. Given the important role played by marketing decisions in the current customer-centric environment, there is a need for a simple and integrated framework for a systematic management of customer knowledge. But, there is a surprising lack of a simple and overall framework to link the extraction of customer knowledge with the management and application of the knowledge, particularly in the context of marketing decisions. While data mining studies have focused on the techniques, customer relationship studies have focused on the interface to the customer and the strategies to manage customer interactions. True customer relationship management is possible only by integrating the knowledge discovery process with the management and use of the knowledge for marketing strategies. This will help marketers address customer needs based on what the marketers know about their customers, rather than a mass generalization of the characteristics of customers.

We address this issue in this paper by presenting an integrated framework for knowledge discovery and management, in the context of marketing decisions. Our paper is further organized as follows. First, we present a taxonomy of data mining tasks and discuss knowledge management as an iterative process (Section 2). We then survey different types of potentially useful marketing and customer knowledge discovered by data mining (Section 3). Marketing decisions based on discovered customer knowledge leads to knowledge-based marketing (Section 4). We close our discussion by identifying the emerging issues to be addressed in the process of managing the discovered marketing knowledge (Section 5).

2. Data mining tasks

Data mining is the process of searching and analyzing data in order to find implicit, but potentially useful, information [3,8,9]. It involves selecting, exploring and modeling large amounts of data to uncover previously unknown patterns, and ultimately comprehensible information, from large databases. Data mining uses a broad family of computational methods that include statistical analysis, decision trees, neural networks, rule induction and refinement, and graphic visualization. Although, data mining tools have been available for a long time, the advances in computer hardware and software, particularly exploratory tools like data visualization and neural networks, have made data mining more attractive and practical.

Pattern extraction is an important component of any data mining activity and it deals with relationships between subsets of data. Formally, a *pattern* is defined as [8]:

> *A statement S in L that describes relationships among a subset of facts F_s of a given set of facts F, with some certainty C, such that S is simpler than the enumeration of all facts in F_s.*

Data mining tasks are used to extract patterns from large data sets. The various data mining tasks can be broadly divided into five categories as summarized in Fig. 1. The taxonomy reflects the emerging role of data visualization as a separate data
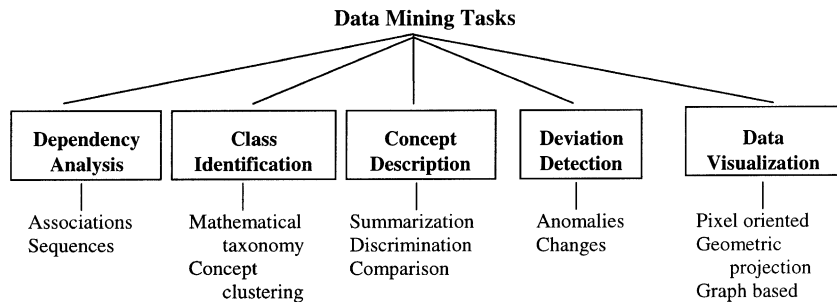
Data Mining Tasks

Dependency Analysis	Class Identification	Concept Description	Deviation Detection	Data Visualization
Associations Sequences	Mathematical taxonomy Concept clustering	Summarization Discrimination Comparison	Anomalies Changes	Pixel oriented Geometric projection Graph based

Fig. 1. A taxonomy of data mining tasks.

mining task, even as it is used to support other data mining tasks. Different data mining tasks are grouped into categories depending on the type of knowledge extracted by the tasks. The identification of patterns in a large data set is the first step to gaining useful marketing insights and making critical marketing decisions. The data mining tasks generate an assortment of customer and market knowledge which form the core of knowledge management process. The specific task to be used is determined by the marketing problem on hand and the following discussion will provide concrete marketing examples to illustrate how the data mining tasks are used.

2.1. Dependency analysis

The primary type of dependency knowledge is the association between sets of items stated with some minimum specified confidence [1]. This is also called "market basket analysis" [3] and gives us the relationship between different products purchased by a customer. This type of knowledge can be useful in developing marketing strategies for promoting products that have dependency relationships in the minds of the customers. For example, rules that have P (e.g., "sausage") in the antecedent and Q (e.g., "mustard") in the consequent may help determine the additional items that have to be sold together with P (i.e., sausage), in order to make it highly likely that Q (i.e. mustard) will also be sold.

2.2. Class identification

Class identification groups customers into classes, which are defined in advance. There are two types of

class identification tasks — mathematical taxonomy and concept clustering. Mathematical taxonomy algorithms produce classes that maximize similarity within classes but minimize similarity between classes. For example, a food store can classify its customers based on their income or past purchase amounts and then target its marketing efforts accordingly. A drawback of this task is its inability to use background information, such as domain knowledge, to facilitate clustering [9]. Concept clustering overcomes this limitation and determines clusters according to attribute similarity as well as conceptual cohesiveness as defined by domain knowledge. Users provide the domain knowledge by identifying useful clustering characteristics. For example, based on the session log data of Internet users, an Internet based company can classify the web users into "email only" users, "serious surfers," and "fun and entertainment surfers."

2.3. Concept description

Concept description is a technique to group customers based on domain knowledge and the database, without forced definitions of the groups. Concept description can be used for summarization, discrimination, or comparison of marketing and customer knowledge. Data summarization is the process of deriving a characteristic summary of a data subset that is interesting with respect to domain knowledge and the full data file. Technically, summarization of a concept *A* is performed by scanning all tuples that satisfy *A* and computing for all fields, in parallel, statistics on their values [23]. Using summarization, a marketer can learn about customer characteristics

by grouping them according to their occupation, income, spending patterns and types of purchases, and build customer profiles. Discrimination describes qualities sufficient to differentiate records of one class from another [9]. For example, the color of the car might be used to distinguish whether or not a sales person is from the Midwest. It can be done by a discrimination algorithm [11]. Comparison describes the class in a way that facilitates comparison and analysis with other records. For example, a prototypical Midwest sales person might own a blue car, have increased sales, and average 100 phone calls a week. This description might serve as the basis against which individual sales people are judged. Comparison analysis can be done by statistical or visualization techniques.

2.4. Deviation detection

Deviations are useful for the discovery of anomaly and changes. Anomalies are things that are different from the normal. For example, compare a group of similar sales people and identify those who stand apart from the average, either in a positive or a negative way. Note that we need to adjust the various factors of the group before comparison. Anomalies can be detected by analysis of the means, standard deviations, and volatility measures from the data. In addition to anomalies, variables or attributes may have significantly different values from the previous transactions for the same customer or group of customers. A credit card company may find a sudden increase in the credit purchases of an individual customer. This change in behavior can be a result of a change in the status of the customer, and not necessarily a fraud. Thus, confirmation of the "change" is made after investigation and the knowledge is updated.

2.5. Data visualization

Data visualization software allow marketers to view complex patterns in their customer data as visual objects complete in three dimensions and colors. They also provide advanced manipulation capabilities to slice, rotate or zoom the objects to provide varying levels of details of the patterns observed. To explore the knowledge in database, data visualization can be used alone or in association with other tasks such as dependency analysis, class identification, concept description and deviation detection. Keim [17] provides an elaborate analysis of visualization techniques for mining large databases and classifies visualization techniques into pixel-oriented, geometric projection and graph-based. The pixel oriented technique maps each data value to a colored pixel and presents the data values belonging to each attribute in separate windows. Geometric projection techniques aim at finding "interesting" projections of multidimensional data set. The basic idea of the graph-based technique is to effectively present a large graph using specific layout algorithms, query languages, and abstraction techniques. Examples of graph based representations are 2-dimensional graphs, 3-dimensional graphs, Hygraphs and SeeNet [7].

3. Knowledge management process

Knowledge discovery and learning is an iterative process that extends the collection of data mining techniques into a knowledge management framework (Fig. 2). Though data mining techniques are usually applied to the complete database, it is possible to mine a statistically representative sample of the data. Hence, the first step in this process is the decision to
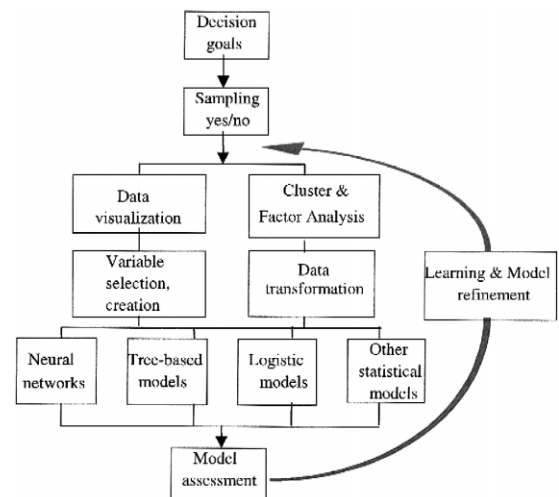

Fig. 2. Knowledge management process.

sample or use the complete database for mining. Once this decision is taken, the next step is to explore the data using tasks such as data visualization. The purpose of this step is to get a first feel of the data in order to select the appropriate variables and data mining tasks. To mine the data set, the marketer may use one or more of several data mining techniques such as neural networks, tree-based methods, rule induction methods, or other statistical models. The outcome of the data mining efforts is evaluated to identify the usefulness of the resulting patterns to the solution of the marketing problem and the accuracy of prediction of future customer behavior from a known set of data. This assessment gives further insights into the data set and helps the marketer to refine the data mining model. The iterative learning process continues until the model is acceptable. While existing knowledge discovery frameworks [8] focus on the discovery as the goal, the data mining model that help extract the patterns is also equally important. A systematic way to retain, refine and use the data model, as shown in our framework, is crucial for effective decision making in the future.

The process of choosing the target goals of knowledge discovery and techniques for data mining on a specific set of data is still unstructured and based on judgment. For example, from point-of-sale data, the marketer may start with identification of purchase patterns of different segments. The segment characteristics are then specified to explore the purchase patterns for each segment. Significant differences in purchase patterns in one segment and similarities in purchase patterns among more segments may lead to refinement of the segment, usually by

adding another dimension. In doing so, the marketer may discover interesting segments, with specific demographic characteristics. In an age of fast changing customer preferences, the right data mining tasks help the marketer rapidly zero-in on and leverage these interesting segments.

3.1. Issues in knowledge management

One of the important issues in knowledge management is the organization, distribution and refinement of knowledge. Knowledge can be generated by data mining tools, can be acquired from third parties, or can be refined or refreshed knowledge. The collected knowledge can then be organized by indexing the knowledge elements, filtering based on content and establishing linkages and relationships among the elements. This knowledge is then integrated into a knowledge base and distributed to the decision support applications. The insights gained by the decision support applications are used to refine the existing knowledge and feedback into knowledge organization. We show this organization, distribution and refinement as a process in Fig. 3. Maximizing the effectiveness of this knowledge management process requires proper definition of the knowledge elements and measures so that marketing knowledge can be shared across distributed applications and can be delivered to decision-makers over networks such as the Internet and the Intranet.

A second most important issue in knowledge management is knowledge integration from disparate sources. Knowledge for marketing decision support can come from three major sources — customer
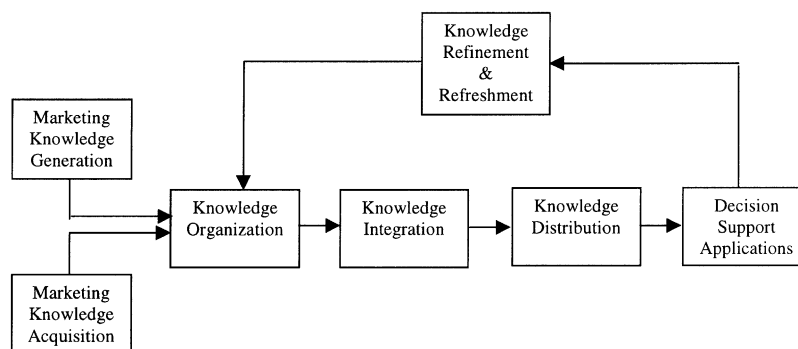

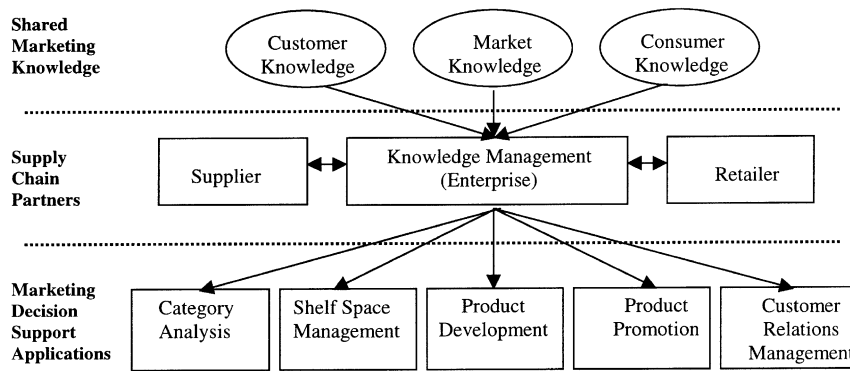Fig. 3. Knowledge organization, distribution and refinement.

Fig. 4. Integrated knowledge management system for marketing.

knowledge from the retailer, consumer knowledge from market research and market knowledge from third-party data providers (Fig. 4). Increasingly, this knowledge is shared by the enterprise with its supply chain partners such as suppliers and retailers. Information technology and the Internet have enabled and increased this sharing of knowledge. One of the classic examples of this shared operation is the partnership between Procter & Gamble (P&G) and Wal-Mart. The Wal-Mart/P&G business team uses concepts such as common data highway, joint scorecards, and customer table checking to share knowledge for mutual benefit of both partners [10]. Going further, Wal-Mart has developed a tool that allows Wal-Mart to share data with its key vendor partners and carriers. P&G has extended some of its continuous replenishment systems to other customers. As an enterprise develops global supply chain partnerships, the critical marketing knowledge crosses traditional organizational boundaries. In such a scenario, ownership and access to the marketing knowledge, standards of knowledge interchange, and sharing of applications become critical success factors.

The move from mass marketing to customer relationship marketing requires decision-makers to come up with specific strategies for each individual customer based on her profile. With traditional tools, it has been a complicated, laborious and painstaking job of identifying and pursuing such segmented markets. In today's environment of complex and ever changing customer preferences, marketing decisions that are informed by knowledge about individual customers become critical [22]. For example, catalog

retailers like Land's End and LL Bean use customer purchase patterns to compute the probability of purchase for each of the merchandise lines. Armed with this information, these firms send the customers only those catalogs for which the calculated purchase probability exceeds a threshold value. Data mining tools provide today's marketer with just the right kind of knowledge to take the appropriate marketing decisions. This true customer knowledge combined with today's interactive technology, such as the Web, can lead to successful relationship marketing and management of each segment in terms of its stage of development [5]. But, for effective customer-centric marketing strategies, the discovered knowledge has to be managed in a systematic manner. We call this process of tightly integrating marketing decisions with the customer knowledge gained from knowledge discovery as knowledge-based marketing. In the following section, we discuss how marketers can benefit from this knowledge-based marketing.

4. Knowledge-based marketing

Marketing decisions, such as promotions, distribution channels and advertising media, based on traditional segmentation approaches result in poor response rate and increased cost. Today's customers have such varied tastes and preferences that it is not possible to group them into large homogenous populations to develop marketing strategies. In fact, each customer wants to be served according to her individual and unique needs. Database marketing, char-

acterized by marketing strategies based on the great deal of information available from the transaction databases and customer databases became popular [14] and most organizations have built up massive databases about their customers and their purchase transactions. But, due to lack of appropriate tools and techniques to analyze these huge databases, a wealth of customer information and buying patterns is permanently hidden and unutilized in such databases. Knowledge-based marketing, which uses appropriate data mining tools and knowledge management framework, addresses this need and helps leverage knowledge hidden in databases. There are three major areas of application of data mining for knowledge-based marketing — (1) customer profiling, (2) deviation analysis, and (3) trend analysis.

4.1. Customer profiling

One of the useful knowledge about a customer is her profile, which is used to make several important marketing decisions. A customer profile is a model of the customer, based on which the marketer decides on the right strategies and tactics to meet the needs of that customer. Fig. 5 presents a customer profiling system that uses data mining tasks. While learning customer profiles, a marketer is interested in the customer demographic details as well as the characteristics of the purchase transactions of the customer. The data mining tasks used in customer profiling can be dependency analysis, class identification and concept description, and we present a list of transaction characteristics that can help the marketer construct useful customer profiles.
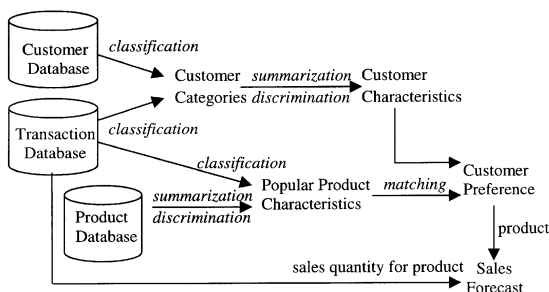

Fig. 5. Customer profiling system.

4.1.1. Frequency of purchases
How often does the customer buy your product or visit your shop? By knowing this, the marketer can build targeted promotions such as "frequent buyer programs."

4.1.2. Size of purchases
How much does the customer spend on a typical transaction? This information helps the marketer devote appropriate resources to the customer who spends more.

4.1.3. Recency of purchases
How long has it been since this customer last placed an order? The marketer may investigate the reasons a customer or a group has not purchased over a long period of time and take appropriate steps. Many times, this could be due to the customer having moved from that location or having shifted loyalty.

4.1.4. Identifying typical customer groups
The characteristics of each group can be obtained by class identification or concept description. For example, a profile indicating that the customer has purchased a new house may lead to the marketer offering a special deal for home furnishings. Knowing the customer and targeting the right deal gets a far better response rate than a general message.

4.1.5. Computing customer lifetime values
With customer profiling supported by data mining and knowledge discovery systems, a number of marketing activities can be enhanced, such as computing customer lifetime values, prospecting and success/ failure of marketing programs.

Customer lifetime values, a measure to understand what is happening to the size and value of a customer base, can be computed by using the customer profile information combined with the product and promotional statistics. Customer lifetime values are asset measures that can help marketers judge their expenditures by measuring a plan's efficiency in producing assets.

4.1.6. Prospecting
Customer profiles, especially their buying patterns, give clues to the marketer on prospective

customers. For example, consider the pattern "Purchase of toys for age group 3–5 years, is followed by purchase of kid's bicycle within 6 months about 90% of the time by high income customers" discovered by data mining. A marketer who has knowledge about the above pattern can identify the prospective customers for kid's bicycle based on toy purchase details and tailor the mail catalog accordingly, thus, increasing the prospect of sales.

4.1.7. Success / failure of marketing programs

Customer databases provide accurate information on the results of marketing programs. The marketer can use the patterns of purchase discovered from the database and the related marketing programs to measure the short-term and long-term effects of the programs.

4.2. Deviation analysis

Knowledge of deviations from normal is extremely important to a marketer. A deviation can be an anomaly (fraud) or a change. In the past, such deviations were difficult to detect in time to take corrective action. Data mining tools provide powerful means such as neural networks for detecting and classifying such deviations. For example, a higher than normal credit purchase on a credit card can be a fraud (anomaly) or a genuine purchase by the customer (change).

Once a deviation has been discovered as a fraud, the marketer takes steps to prevent such frauds and initiates corrective action. If the deviation has been discovered as a change, further information collection is necessary. For example, a change can be that a customer got a new job and moved to a new house. In this case, the marketer has to update the knowledge about the customer. A marketer can use the deviation detection capability to query changes that occurred as a result of recent price changes or promotions.

4.3. Trend analysis

Trends are patterns that persist over a period of time. Trends could be short-term trends like the

immediate increase and subsequent slow decrease of sales following a sales campaign. Or, trends could be long-term, like the slow flattening of sales of a product over a few years. Data mining tools, such as visualization, help us detect trends, sometimes very subtle and hidden in the database, which would have been missed using traditional analysis tools like scatter plots. In marketing decisions, trends can be used for evaluating marketing programs or to forecast future sales.

4.3.1. Evaluate performance of products or marketing programs

The customer database provides an accurate record of the transactions. Marketers can use visualization tools to identify trends in sales, costs and profits by products, regions or markets in order to understand the impact of, say, a sales promotion. Data mining also provides statistical tools to precisely measure the performance of the various parameters of interest.

4.3.2. Forecast future sales

One of the popular uses of trends is forecasting future sales. Marketers are interested in knowing how various marketing programs affect future sales of their products. Data mining allows discovery of subtle relationships like a peak in sales of a product associated with a change in the profile of a particular group of customers.

The recent emphasis on customer relationship management has put the focus back on the customer. The four key steps for customer relationship management — (1) identifying the right customers, (2) differentiating among them, (3) interacting with and learning from existing customers, and (4) customizing the product or service to the needs of individual customers [22] — are based on knowing customers better. Current efforts on customer relationship management are focused on the customer interface and managing customer interactions. But inadequate knowledge about customers and the lack of a systematic knowledge management framework continue to hinder the efforts of organizations, particularly the marketing function, to manage their customer relationships. The knowledge management framework described in this paper can provide the basis for

organizations to effectively integrate the discovery of customer knowledge with their relationship management strategies.

5. Research challenges in marketing knowledge management

Knowledge management and data mining are still evolving fields and thus present interesting challenges for researchers and practitioners, with implications for the marketing function. Even as we present an integrated framework for knowledge management in the context of marketing, we realize that there are critical research challenges to be addressed. Some are related to the data mining techniques and the knowledge discovery process, while others are related to the management of knowledge. First, knowledge discovery through data mining is an *iterative learning* process similar to other knowledge generation processes, such as scientific discovery. The selection of data mining algorithms, hypotheses formation, model evaluation and refinement are key components of this discovery process. Because it takes cycles of trials and errors to progressively produce the most useful knowledge from the data mining, a learning by experimentation approach [19] can be useful to ensure that the process can eventually "discover" the useful knowledge. One of the research challenges is to make this process more structured and thus improve the productivity of the data mining efforts.

A second challenge is to manage knowledge that crosses organizational boundaries and is distributed across supply chain partners. Customer knowledge is typically distributed across supply chain partners, and marketing is an important beneficiary of this knowledge. But, managing the cross-organizational knowledge requires organizational and industry level efforts. The key research issues are the development of appropriate inter-organizational knowledge management models, protection of knowledge rights and distribution of knowledge benefits among partners.

A third challenge for knowledge management research is multiple classifications, a situation in which customers can belong to more than one category. Current data mining techniques have been shown to be limited in handling memberships to multiple classes [26]. The increasing complexity of the customer preferences makes this issue particularly relevant for marketers, as they may encounter customers with multiple membership and need reliable classification tools. A marketer may also want to use multiple memberships to gain important knowledge about customers, instead of simplifying the classes and losing valuable information.

Yet another important challenge is "Web mining." With the Internet emerging as the new channel for distribution of goods, promotion of products, handling of transactions, and coordination of business processes, the Web is emerging as an important and convenient source of customer data. But, the multiple data formats and distributed nature of the knowledge on the Web make it a challenge to collect, discover, organize and manage the knowledge in a manner that is useful for marketing decision support. As marketing depends more and more on the Web for customer data, Web mining need to be addressed as an important marketing knowledge management issue.

6. Conclusion

Though data mining techniques are used in several areas such as fraud detection, bankruptcy prediction, medical diagnosis, and scientific discoveries, their use for marketing decision support highlights unique and interesting issues such as customer relationship management, real-time interactive marketing, customer profiling and cross-organizational management of knowledge. In the current customer-centric business environment, it is our firm belief that there is a need for deeper understanding of use of data mining and knowledge management for marketing decision support. Towards that end, in this paper, we have shown how data mining can be integrated into a marketing knowledge management framework. With the availability of large volume of data, made possible by modern information technology, a major problem is to filter, sort, process, analyze and manage this data in order to extract the information relevant to the user. The growth in the size and number of existing databases far exceeds

human abilities to analyze such data using traditional tools and thus creates both a need and an opportunity for data mining tools. With the shift from mass marketing to one-to-one relationship marketing, one area that could greatly benefit from data mining is the marketing function itself. A systematic application of data mining techniques will enhance the knowledge management process and arm the marketers with better knowledge of their customers leading to better service to customers. To us, it is also clear that the Web technology will have a major impact on the practice of data mining and knowledge management, and that should present interesting challenges for future information systems research.

References

[1] R. Agrawal, T. Dimielinski, A. Swami, Database mining: a performance perspective, IEEE Transactions on Knowledge and Data Engineering 5 (6) (1993) 914–925.

[2] D.M. Amidon, Blueprint for 21st century innovation management, Journal of Knowledge Management 2 (1) (1998) 23–31.

[3] M.J.A. Berry, G. Linoff, Data Mining Techniques for Marketing, Sales, and Customer Support, Wiley, New York, 1997.

[4] J.P. Bigus, Data Mining with Neural Networks: Solving Business Problems — From Application Development to Decision Support, McGraw-Hill, New York, 1996.

[5] R.C. Blattberg, J. Deighton, Interactive marketing: exploiting the age of addressability, Sloan Management Review (1991) (Fall) 5–14.

[6] K.C.C.C. Chan, A.K.C. Wong, A statistical technique for extracting classificatory knowledge from databases, in: G. Piatetsky-Shapiro, W.J. Frawley (Eds.), Knowledge Discovery in Databases, MIT Press, Massachusetts, 1991, Chap. 6.

[7] S.G. Eick, G.J. Wills, Navigating large networks with hierarchies, Visualization '93 (San Jose, CA), 1993, pp. 204–210.

[8] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, From data mining to knowledge discovery: an overview, in: U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (Eds.), Advances in Knowledge Discovery and Data Mining, MIT Press, Massachusetts, 1996, Chap. 1.

[9] W.J. Frawley, G. Piatetsky-Shapiro, C.J. Matheus, Knowledge discovery in databases: an overview, AI Magazine 13 (3) (1992) 57–70.

[10] M. Graen, Technology in Manufacturer/Retailer Integration: Wal-Mart and Procter & Gamble, Private communication, 1999.

[11] J. Han, Y. Cai, N. Cercone, Knowledge discovery in databases: an attribute-oriented approach, Proceedings of the 18th VLDB Conference, 1992.

[12] C.W. Holsapple, R. Pakath, V.S. Jacob, J.S. Zaveri, Learning by problem processors: adaptive decision support systems, Decision Support Systems 10 (2) (1993) 85–108.

[13] M. Holsheimer, M.L. Kersten, A.P.J.M. Siebes, Data surveyor: searching the nuggets in parallel, in: U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (Eds.), Advances in Knowledge Discovery and Data Mining, MIT Press, Massachusetts, 1996, Chap. 18.

[14] H. Holtz, Databased Marketing — Every Manager's Guide to the Super Marketing Tool of the 21st Century, Wiley, New York, 1992.

[15] C. Hsu, C.A. Knoblock, Using inductive learning to generate rules for semantic query optimization, in: U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (Eds.), Advances in Knowledge Discovery and Data Mining, MIT Press, Massachusetts, 1996, Chap. 17.

[16] W. Inmon, Building the Data Warehouse, Wiley, New York, 1996.

[17] D.A. Keim, Visualization techniques for mining large databases: a comparison, IEEE Transactions on Knowledge and Data Engineering 8 (6) (1996) 923–937.

[18] S. Kelly, Data Warehousing: The Route to Mass Customization, Wiley, New York, 1996.

[19] F. Lin, M.J. Shaw, Active training of backpropagation neural networks using the learning by experimentation methodology, Annals of Operations Research 75 (1997) 105–122.

[20] C.J. Matheus, G. Piatetsky-Shapiro, D. McNeill, Selecting and reporting what is interesting: the KEFIR application to healthcare data, in: U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (Eds.), Advances in Knowledge Discovery and Data Mining, MIT Press, Massachusetts, 1996, Chap. 20.

[21] D. Peppers, M. Rogers, Enterprise One to One: Tools for Competing in the Interactive Age, Doubleday, New York, 1997.

[22] D. Peppers, M. Rogers, Is your company ready for one-to-one marketing? Harvard Business Review (1999) 151–160.

[23] G. Piatetsky-Shapiro, C.J. Matheus, Knowledge discovery workbench for exploring business databases, International Journal of Intelligent Systems 7 (1992) 675–686.

[24] M.C. Rumizen, Report on the second comparative study of knowledge creation conference, Journal of Knowledge Management 2 (1) (1998) 77–82.

[25] M.J. Shaw, Machine learning methods for intelligent decision support: an introduction, Decision Support Systems 10 (2) (1993) 79–83.

[26] W.E. Spangler, J.H. May, L.G. Vargas, Choosing data-mining methods for multiple classification: representational and performance measurement implications for decision support, Journal of Management Information System 16 (1) (1999) 37–62 (Summer).

[27] T.K. Sung, N. Chang, G. Lee, Dynamics of modeling in data mining: interpretive approach to bankruptcy prediction, Journal of Management Information System 16 (1) (1999) 63–85 (Summer).

[28] W. Ziarko, The discovery, analysis, and representation of data dependencies in databases, in: G. Piatetsky-Shapiro, W.J. Frawley (Eds.), Knowledge Discovery in Databases, MIT Press, Massachusetts, 1991, Chap. 11.

Michael J. Shaw is Hoeft Endowed Chair Professor in Information Technology Management and Director of the Center for Information System and Technology Management at the University of Illinois at Urbana-Champaign. He is also affiliated with the Beckman Institute for Advanced Science and Technology. He has published over 60 refereed scholarly papers in journals such as *Management Science*, *Information Systems Research*, *INFORMS Journal* on Computing, *Communications of the ACM*, *IEEE Inter Computing*, *IIE Transactions*, and *Decision Support Systems*. He is the chief editor of the recently published *Handbook on Electronic Commerce* and is also the editor of the forthcoming book, *Information-Based Manufacturing*, to be published by Kluwer Academic Publishers.

Chandrasekar Subramaniam is a doctoral student in Management Information Systems at the University of Illinois at Urbana-Champaign. Before joining the doctoral program, he was Assistant Professor in Information Systems for more than 4 years at Bharathidasan Institute of Management, India. While in India, he taught courses in information technology strategy, business process reengineering and information systems development. His current research interests are electronic commerce strategies for marketing, business-to-business electronic commerce and inter-organizational information systems.

Gek Woo Tan is Assistant Professor in the School of Computing at the National University of Singapore. She received her M.Sc. degree in Computer Science from Indiana University and Ph.D. in Business Administration from the University of Illinois at Urbana-Champaign. Dr. Tan's primary research focuses on information sharing in supply chain network. Her other research interests include electronic commerce, multi-agent simulations, and enterprise modeling.

Michael E. Welge is Director, Automated Learning Group at the National Center for Supercomputing Applications and a faculty member at the Center for Information Technology at the University of Illinois. He is also an adjunct researcher at the Laboratory for High-Performance Data Mining, Department of Computer and Information Sciences, Kansas State University. In these positions, Michael has been helping *Fortune 100* companies make sense of the data mining process. He has worked in the fields of data mining, mathematical modeling, applied AI, and stochastic simulation since 1982. He has developed models and systems that use applied technologies including decision trees, neural networks, genetic algorithms, self-organizing maps, association rules, and models that use more traditional statistical methods. This work spans a diverse set of data mining applications dealing with issues such as in manufacturing, medicine, fraud detection, customer relationship management, water treatment, business intelligence, telecommunications, risk management and finance. Current work includes algorithm design and development for the data mining environment D2K. Michael is frequently invited to lecture at Caterpillar, Allstate, Sears, Ford, Boeing, Motorola, and the University of Illinois School of Business, and has presented briefing and courses on data mining to the NSF Alliance, DOD and State of Illinois. He was awarded the 1995, 1997 and 1998 Industrial Grand Challenge awards for his work in the area of data mining with Motorola, Sears, and Caterpillar. Along with his team, Michael was the recipient of the Sears 1998 Partners in Progress and 1998 Caterpillar Annual Quality Improvement Awards.