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