

Applications for Data Mining Techniques in Customer Relationship Management

Natalie Clewley
Brunel University, UK

Sherry Y. Chen
Brunel University, UK

Xiaohui Liu
Brunel University, UK

INTRODUCTION

With the explosion in the amount of data produced in commercial environments, organizations are faced with the challenge of how to collect, analyze, and manage such large volumes of data. As a consequence, they have to rely upon new technologies to efficiently and automatically manage this process. Data mining is an example of one such technology, which can help to discover hidden knowledge from an organization's databases with a view to making better business decisions (Changchien & Lu, 2001).

Data mining, or knowledge discovery from databases (KDD), is the search for valuable information within large volumes of data (Hand, Mannila & Smyth, 2001), which can then be used to predict, model or identify interrelationships within the data (Urtubia, Perez-Correa, Soto & Pszczolkowski, 2007). By utilizing data mining techniques, organizations can gain the ability to predict future trends in both the markets and customer behaviors. By providing detailed analyses of current markets and customers, data mining gives organizations the opportunity to better meet the needs of its customers.

With such significance in mind, this chapter aims to investigate how data mining techniques can be applied in customer relationship management (CRM). This chapter is organized as follows. Firstly, an overview of the main functionalities data mining technologies can provide is given. The following section presents application examples where data mining is commonly applied within the domain, with supporting evidence as to how each enhances CRM processes. Finally, current issues and future research trends are discussed before the main conclusions are presented.

BACKGROUND

Data mining methods can generally be grouped into four categories: classification, clustering, association rules and

information visualization. The following subsections will describe these in further detail.

Classification

Databases are full of hidden information that can help to make important business decisions. Classification involves using an algorithm to find a model that describes a data class or concept (Han & Kamber, 2006). By identifying a series of predefined labels, items can be categorized into classes according to its attributes (e.g., age or income). Thus, it is a useful technique for identifying the characteristics of a new item. For example, in the case of a bank loan clerk, classification is useful for predicting whether loan applicants are a "safe" or "risky" investment for the bank based on the class that they belong to. Popular classification techniques include Decision Trees and Bayesian Networks.

Clustering

Where classification is thought of as a supervised learning technique because it uses a set of predefined class labels, clustering is an unsupervised learning technique. Because no assumptions are made about the structure of the data, clustering can uncover previously hidden and unexpected trends or patterns. Clustering involves grouping items into "natural" clusters based on their similarities (Hand et al., 2001). Each item in a cluster is similar to those within its cluster, but dissimilar to those items in other clusters. In this way, clustering is commonly used to identify customer affinity groups with the aim of targeting with specialized marketing promotions (section 3.2.2). Common clustering techniques include K-means and Kohonen Networks.

Association Rules

Association rules are mainly used to find relationships between two or more items in a database. Association rules are

expressed in the form ($X \rightarrow Y$), where X and Y are both items. In a set of transactions, this means that those containing the items X , tend to contain the items Y . Such an association rule is usually measured by *support* and *confidence*, where the *support* is the percentage of both X and Y contained in all transactions and the *confidence* is calculated by dividing the number of transactions supporting the rule by the number of transactions supporting the rule body (Zhang, Gong & Kawamura, 2004). For example, this technique is commonly used to identify which items are regularly purchased together or to identify the navigational paths of users through an online store. The discovery of such relationships can help in many business decisions, such as customer shopping behavior analysis, recommendations, and catalog design (Han & Kamber, 2006).

Information Visualization

Information visualization is based on an assumption that human beings are very good at perceiving structure in visual forms. The basic idea is to present the data with some graphics, for example, 2D graphics and 3D graphics, allowing the human to gain insight from the data, draw conclusions, and directly interact with the data (Ankerst, 2001). Since the user is directly involved in the exploration process, shifting and adjusting the exploration goals is automatically done if necessary (Lopez, Kreuseler & Schumann, 2002). This approach is especially useful when little is known about the data and the exploration goals are vague, for example, analyzing the path of customers through an online store.

Data Mining and Its Applications in CRM

Customer relationship management (CRM) is thought to be an increasingly important success factor for e-business. CRM is the process of managing interactions between a company and its customers. Initially, this involves segmenting the market to identify customers with high-profit potential, from which marketing strategies are designed to favorably impact the behavior of customers in these segments (Berson, Smith & Thearling, 2000).

Customer Analysis

Customer Segmentation

Customer segmentation is the process of splitting customers into homogeneous groups based on their common attributes (Böttcher, Spott, Nauck & Kruse, 2007). More specifically, clustering algorithms are used to discover groups of customers based on such attributes. By doing so, each cluster contains customers who share similar attribute values. Benefits of

clustering in this way include a deeper understanding of the needs and behaviors of individual customer groups, allowing the implementation of systems for the purpose of personalized recommendations (section 3.2.1) and targeted marketing (section 3.2.2). For example, Vellido, Lisboa, and Meehan (1999) segmented customers from the data gathered in an online survey. Initially, they conducted a factor analysis of the observable data and then the factor scores were clustered by using the Self Organizing Maps. Five clusters were discovered: cost conscious, complexity avoiders, unconvinced, convinced and security conscious. These clusters allowed researchers to identify which groups of customers frequently used online shopping and those that would be more open to specific marketing campaigns.

However, even though customers can be easily segmented into groups in such a manner, today's fluctuating markets mean that segmentation often becomes obsolete very quickly. Ha (2007) overcame this by providing a method that monitors constantly changing customer needs. After performing Kohonen Network clustering to divide the customers into four dominant clusters, the author focused on keeping track of customer shifts among the segments to monitor the changes over time. Behavior patterns of customers in each of the segments were then predicted through the use of transition paths.

In addition, Böttcher et al. (2007) present a system for customer segmentation which accounts for the dynamic nature of a market based on "interestingness". Their system focuses on the discovery of frequent item-sets and the analysis of their change over time, which provides detailed knowledge about how customer behavior evolves over time. They successfully applied their system to two problem domains in a telecommunications company: customer analytics and network usage. The former aimed to identify the factors likely to drive customer satisfaction in the future, whereas the latter aimed to understand the drivers of change in customer behavior whilst using the services.

Click Stream Analysis

In addition to grouping the types of users through transaction or personal data, click streams, the paths visitors take through a website, also provide valuable information. Analyzing click stream data can show retailers how visitors navigate their way around an online shop, which is useful towards understanding the effectiveness of marketing efforts, that is, how customers find the online store, what products they view and what they purchase (Lee, Podlaseck, Schonberg & Hoch, 2001).

Lee et al. (2001) conducted a study to analyze click streams with a view to evaluating the effectiveness of online merchandising tactics. By using an interactive visualization system, they were able to break down the click streams into individual customer shopping steps to highlight potential

3 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/applications-data-mining-techniques-customer/13571

Related Content

Learning Styles and Adaptive ICT-Based Learning Environment

Zlatko J. Kovacic (2008). *Information Communication Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 413-429).

www.irma-international.org/chapter/learning-styles-adaptive-ict-based/22676

Institutional Repository as a Knowledge Management Tool for the Enhancement of Library Visibility in the 21st Century: A Case of Midlands State University

Austin Tonderai Nyakurerwa (2021). *Handbook of Research on Information and Records Management in the Fourth Industrial Revolution* (pp. 81-93).

www.irma-international.org/chapter/institutional-repository-as-a-knowledge-management-tool-for-the-enhancement-of-library-visibility-in-the-21st-century/284719

Techno-Eustress and Techno-Distress: A Metaverse Investigation

Helmi Issa, Roy Dakroub, Hussein Lakkisand Jad Jaber (2022). *Information Resources Management Journal* (pp. 1-21).

www.irma-international.org/article/techno-eustress-and-techno-distress/314575

Direct-Indirect Link Matrix: A Black Box Testing Technique for Component-Based Software

Saurabh Rawat, Anushree Sahand Ankur Dumka (2020). *International Journal of Information Technology Project Management* (pp. 56-69).

www.irma-international.org/article/direct-indirect-link-matrix/265139

A Novel Node Management in Hadoop Cluster by Using DNA

Balaraju J. and P.V.R.D. Prasada Rao (2021). *International Journal of Information Technology Project Management* (pp. 38-46).

www.irma-international.org/article/a-novel-node-management-in-hadoop-cluster-by-using-dna/288711