Customer Relationship Management and Knowledge Discovery in Database

Jounghae Bang
Kookmin University, Korea

Nikhilesh Dholakia
University of Rhode Island, USA

Lutz Hamel
University of Rhode Island, USA

Seung-Kyoon Shin
University of Rhode Island, USA

INTRODUCTION

Customer relationships are increasingly central to business success (Kotler, 1997; Reichheld & Sasser, 1990). Acquiring new customers is five to seven times costlier than retaining existing customers (Kotler, 1997). Simply by reducing customer defections by 5%, a company can improve profits by 25% to 85% (Reichheld & Sasser, 1990). Relationship marketing—getting to know customers intimately by understanding their preferences—has emerged as a key business strategy for customer retention (Dyche, 2002).

Internet and related technologies offer amazing possibilities for creating and sustaining ideal customer relationships (Goodhue, Wixom, & Watson, 2002; Ives, 1990; Moorman, Zaltman, & Deshpande, 1992). Internet is not only an important and convenient new channel for promotion, transactions, and business process coordination; it is also a source of customer data (Shaw, Subramaniam, Tan, & Welge, 2001). Huge customer data warehouses are being created using advanced database technologies (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

Customer data warehouses by themselves offer no competitive advantages: insightful customer knowledge must be extracted from such data (Kim, Kim, & Lee, 2002). Valuable marketing insights about customer characteristics and their purchase patterns, however, are often hidden and untapped (Shaw et al., 2001). Data mining and knowledge discovery in databases (KDD) facilitate extraction of valuable knowledge from rapidly growing volumes of data (Mackinnon, 1999; Fayyad et al., 1996).

This article provides a brief review of customer relationship issues. The article focuses on: (1) customer relationship management (CRM) technologies, (2) KDD techniques, and (3) Key CRM-KDD linkages in terms of relationship marketing. The article concludes with the observations about the state-of-the-art and future directions.

BACKGROUND

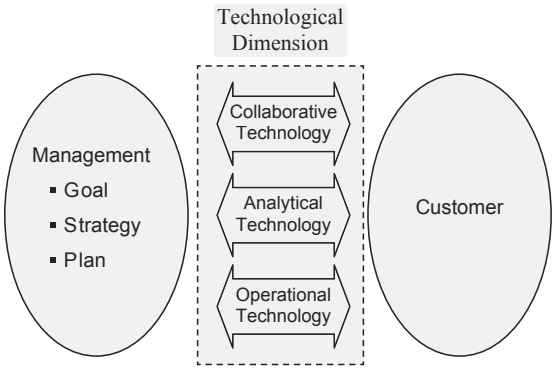
CRM Technologies

CRM is interpreted in a variety of ways (Goodhue et al., 2002; Winer, 2001; Wright, 2002). In some cases, CRM simply entails direct e-mails or database marketing. In other cases, CRM refers to CICs (customer interaction centers) and OLAP (online analytical processing), which is referred to as various types of online query-driven analyses for examining stored data. Overall, CRM can be seen as a core business strategy to interact with, create, and deliver value to targeted customers to improve customer satisfaction and customer retention at a profit. It is grounded in high quality customer data and enabled by information technology (Ang & Buttle, 2002).

Three core dimensions characterize buyer-focused CRM systems: customers, management, and technologies. *Customer* service and related issues must be included in the design, implementation, and operation of any CRM system. Organizations benefit from CRM only when such systems benefit their customers—using CRM merely as a sales or customer service solution is a recipe for failure (Davids, 1999). *Management's* articulation and tracking of customer relationship goals, plans, and metrics is an essential CRM component (Ang & Buttle, 2002; Greenberg, 2002). Successful CRM implementations rely on management goals, strategies, and plans that reflect customer commitment and promote a customer-responsive corporate culture at all levels of the organization (Ang & Buttle, 2002; Smith, 2001). *Technologies* for facilitating collaborative, operational, and analytical CRM activities are the manifest aspects of CRM (Goodhue et al., 2002).

Collaborative CRM systems refer to any CRM function that provides a point of interaction between the customer and the marketing channel (Greenberg, 2002). Web-based

Figure 1. Alignment of three dimensions of CRM



Internet, and in some cases mobile commerce systems, offer multiple "touch points" for reaching the customers. In employing the Web and mobile technologies, it is important to ensure that such technologies enhance older, preexisting channels (Johnson, 2002). *Operational CRM* systems refer to technologies that span the ordering-delivery cycle (Goodhue et al., 2002). Operational CRM is concerned with automating the customer-facing parts of the enterprise (Ang & Buttle, 2002). Since the sales process depends on the cooperation of multiple departments performing different functions, integration of all such functions is critical for operational CRM systems (Earl, 2003; Greenberg, 2002). *Analytical CRM* systems analyze customer data warehouses so that the firm can detect valuable patterns of customers' purchasing behavior. Off-line data mining of customer data warehouses as well as online analytical processing (OLAP) can aid in applications such as campaign management, churn analysis, propensity scoring, and customer profitability analysis (Goodhue et al., 2002). It is this component of CRM that has a clear linkage to KDD methods.

KDD Techniques

Since multiple data formats and distributed nature of knowledge on the Web make it a challenge to collect, discover, organize, and manage CRM-related customer data (Shaw et al., 2001), KDD methods are receiving attention in relationship marketing contexts (Fayyad et al., 1996; Mackinnon, 1999). Massive databases are commonplace, and they are ever growing, dynamic, and heterogeneous (Mackinnon & Glick, 1999). Systematic combining of data mining and knowledge management techniques can be the basis for advantageous customer relationships (Shaw et al., 2001).

KDD is defined as the process of data selection, sampling, pre-processing, cleaning, transformation, dimension reduction, analysis, visualization, and evaluation (Mackinnon, 1999). As a component of KDD (Fayyad et al., 1996), data mining can be defined as the process of searching and analyzing data in order to find latent but potentially valuable information (Shaw et al., 2001).

KDD constitutes the overall process of extracting useful knowledge from databases. It is a multidisciplinary activity with the following stages (Brachman, Khabaza, Kloesgen, Piatetsky-Shapiro, & Simoudis, 1996; Bruha, Kralik, & Berka, 2000; Fayyad et al., 1996):

- Selecting the problem area and choosing a tool for representing the goal to be achieved
- Collecting the data and choosing tools for representing objects (observations) of the dataset
- Preprocessing of the data: integrating and cleaning data
- Data mining: extracting pieces of knowledge
- Post-processing of the knowledge derived: testing and verifying, interpreting, and applying the knowledge to the problem area at hand

In Web-based relationship marketing, three distinct categories of data mining have emerged: Web content mining, Web structure mining, and Web usage mining (Jackson, 2002). Web usage mining is also referred to as clickstream analysis (Edelstein, 2001). Valuable information hidden in the clickstream data of many e-commerce sites can provide sharp diagnostics and accurate forecasts, allowing e-commerce sites to profitably target and reach key customers (Moe & Fader, 2001). Therefore, many detailed studies have been conducted on Web usage mining. For example, Web access pattern tree (WAP-tree) mining is one of the sequential pattern mining techniques for Web log access sequences (Ezeife & Lu, 2005).

Such Web-based CRM systems require large, integrated data repositories and advanced analytical capability. Even though there are many success stories, Web-based CRM projects continue to be expensive and risky undertakings. OLAP refers to the various types of query-driven analysis for analyzing stored data (Berry & Linoff, 1997). Data mining and OLAP can be seen as complementary tools (Jackson, 2002). Both Web-based CRM systems and OLAP, in general, involve vast volumes of both structured and unstructured data. One common challenge with managing this data is to incorporate unstructured data into a data warehouse. Traditional database systems are not designed for unstructured data.

Research in KDD in general is intended to develop methods and techniques to process a large volume of unstructured data in order to retrieve valuable knowledge (which is "hidden" in these databases) that would be compact and abstract, yet understandable and useful for managerial applications (Bruha et al., 2000).