Strategic Utilization of Data Mining

Chandra S. Amaravadi
Western Illinois University, USA

INTRODUCTION

In the past decade, a new and exciting technology has unfolded on the shores of the information systems area. Based on a combination of statistical and artificial intelligence techniques, data mining has emerged from relational databases and Online Analytical Processing as a powerful tool for organizational decision support (Shim et al., 2002).

A number of techniques are available to analyze warehouse data, including descriptive techniques such as data summarization; data visualization; clustering and classification; and predictive techniques such as regression, association, and dependency analyses (Jackson, 2002; Mackinnon & Glick, 1999). The technology is being extended to mine semi-structured data as well (Hui & Jha, 2000).

Applications of data mining have ranged from predicting ingredient usage in fast-food restaurants (Liu, Bhattacharyya, Sclove, Chen & Lattyak, 2001) to predicting the length of stay for hospital patients (Hogl, Muller, Stoyan & Stuhlinger, 2001). See Table 1 for other representative examples. Some of the important findings are:

1) Bankruptcies can be predicted from variables such as the "ratio of cash flow to total assets" and "return on assets" (Sung, Chang & Lee, 1999).
2) Gas station transactions in the UK average £20, with a tendency for customers to round the purchase to the nearest £5 (Hand & Blunt, 2001).
3) Sixty-nine percent (69%) of dissatisfied airline customers did not contact the airline about their problem (Marple & Zimmerman, 1999).
4) Sales in fast-food restaurants are seasonal and tend to peak during holidays and special events (Liu et al., 2001).
5) Patients in the age group > 75 are 100% likely to exceed the standard upper limit for hospital stay (Hogl et al., 2001).

BACKGROUND

A majority of data mining (DM) applications serve a managerial purpose. They are useful in finding information such as identifying loyal customers or patients who are likely to stay longer at hospitals. This usage can be extended to strategic decision making as well. According to Sabherwal and King (1991), a *strategic application* is one that has a profound influence on a firm's success, by either influencing or shaping the organization's strategy or by playing a direct role in the implementation or support of it. If DM could be utilized in shaping the firm's strategy, it could have a strategic impact. Let us then consider the process of strategic decision making (SDM). An interpretive view of this process involves *scanning* the environment for important events or information, *interpreting* these events as threats or opportunities, and *formulating* a response (Daft & Weick, 1984).

The interpretation stage involves some form of consensual validation, with managers comparing notes with subordinates or with peers (Daft & Weick, 1984). It is of particular interest since it involves modifying the *belief systems*, the summary of perceptions, observations, and experiences concerning the organization's resources, markets, and customers. For instance, an organization

Table 1. Examples of data mining applications

• Predicting supplies in fast food restaurants (Liu, Bhattacharyya, Sclove, Chen and Lattyak 2001).
• Quality of health care (Hogl, Muller, Stoyan and Stuhlinger 2001).
• Analyzing Franchisee sales (Chen, Justis and Chong 2003).
• Predicting customer loyalty (Ng and Liu 2001).
• Job shop scheduling (Koonce, Fang and Tsai 1997).
• Mining credit card data (Hand and Blunt 2001).
• Bankruptcy prediction (Sung, Chang and Lee 1999).

might have a perception that its product lines are aging. Customers switching to competitors' products could confirm this observation. There is empirical evidence that belief systems influence strategic decision making (Lorsch, 1989). In a study of 12 firms, Lorsch (1989) found that major strategic decisions were influenced by core beliefs that included financial goals, acceptable types of risks, and distinctive competence, among other things. Thus decisions regarding product lines, customers, and suppliers are influenced by perceptions concerning adequacy of product lines, type of customers, reliability of suppliers, and so forth. We will alternatively use the term *micro-theories* (MTs) to refer to these beliefs and will regard each as a strategic assumption to be tested by data mining.

The mining process, often labeled as "KDD" (Knowledge Discovery in Databases) can be "data driven" or "hypothesis driven" ("question driven"). *Data-driven* methods attempt to identify all possible patterns from the data, while *hypothesis-driven* methods attempt to verify whether or not a particular pattern exists (Hogl et al., 2001). Usually, organizations have more data than they can analyze. Question-driven approaches are computationally more tractable, especially when large data sets are involved, since the solution space is bounded. In this mode, KDD commences with a set of MTs that management is keen on verifying. The remainder of the process is the same for both approaches (Mackinnon & Glick, 1999). The next step is to select suitable data (see Figure 1). This is greatly facilitated if the analyst already has hypotheses to verify. Otherwise, data selection will involve an iterative process of selection followed by testing. The required data needs to be carefully selected from the warehouse or organizational databases. It is then cleaned and transformed by filling in missing values, changing "look up codes" (i.e., standardizing codes from numeric values to text or vice-versa: "1"–married; "2"–single), and ignoring outliers if necessary. Calculations such as totals, cost/item, and discount are also performed during this stage. The next step is testing and analysis where each MT is examined using the "selected" and "cleaned" data. The last step is the sharing of results with management, usually through formal reports or presentations or via an intranet.

MAIN THRUST

What sort of beliefs should an analyst select for testing purposes? Porter's framework is widely used by academics and practitioners to understand organizational strategy. It summarizes key competitive forces acting on an organization as follows: a) bargaining power of customers, b) bargaining power of suppliers, c) firm rivalry, d) threat of substitute products, and e) threat of new entrants. To compete effectively, a firm has to counteract these forces.

The framework is useful in organizing micro-theories as well since it describes the entities pertaining to the organization's *task environment*, which govern its inputs and outputs and therefore affect its performance.

As shown in Figure 2, a firm's beliefs can be organized by each of the entities in the firm's task environment, including suppliers, customers, competitors, and substitute products. For instance, how do customers perceive the products of a company? How does the firm feel about a particular substitute? Ultimately, these perceptions

Figure 1. Knowledge discovery process with micro-theories (Adapted from Mackinnon & Glick, 1999)