Prioritization of association rules in data mining: Multiple criteria decision approach

Duke Hyun Choi[a,*], Byeong Seok Ahn[b], Soung Hie Kim[a]

[a]Graduate School of Management, Korea Advanced Institute of Science and Technology (KAIST), 207-43 Cheongryangri-Dong, Dongdaemun, Seoul 130-012, South Korea
[b]Department of Business Administration, Hansung University, 389 Samsun-Dong 2, Sungbuk, Seoul 136-792, South Korea

Abstract

Data mining techniques, extracting patterns from large databases are the processes that focus on the automatic exploration and analysis of large quantities of raw data in order to discover meaningful patterns and rules. In the process of applying the methods, most of the managers who are engaging the business encounter a multitude of rules resulted from the data mining technique. In view of multi-faceted characteristics of such rules, in general, the rules are featured by multiple conflicting criteria that are directly related with the business values, such as, e.g. expected monetary value or incremental monetary value.

In the paper, we present a method for rule prioritization, taking into account the business values which are comprised of objective metric or managers' subjective judgments. The proposed methodology is an attempt to make synergy with decision analysis techniques for solving problems in the domain of data mining. We believe that this approach would be particularly useful for the business managers who are suffering from rule quality or quantity problems, conflicts between extracted rules, and difficulties of building a consensus in case several managers are involved for the rule selection.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Rule prioritization; Rule conflict; Association rule mining; ELECTRE

1. Introduction

Under intense competition forcing companies to develop and maintain competitive marketing activity, techniques from the disciplines of both data mining and decision analysis have been extensively used in the development of computerized decision aids. Data mining techniques, extracting patterns from large databases, are the processes that focus on the automatic exploration and analysis of large quantities of raw data in order to discover meaningful patterns and rules (Agrawal, Imielinski, & Swami, 1993; Agrawal & Srikant, 1994; Song, Kim, & Kim, 2001). On the other hand, decision analysis is concerned with applying decision theory to real-world problems to help companies through the decision making process in which decision makers' preferences that are important for judging the

desirability of alternative outcomes, are associated with criteria in a problem solving or decision making situation (White, 1990). Whilst separately developed, both data mining and decision analysis have conceptual basic denominators, providing systematic methods for problem solving and decision making in view of objectives (decision aiding) and delivery vehicle (the computer).

From the perspective of many types of practical decision aiding applications, however, both data mining and decision analysis techniques have some limitations. Particularly, in decision support system development, there is little effort for generating synergies with complementing each other's limitations. More specifically, user preferences, which play a key role in decision aids with decision analysis, are not explicitly considered in the current generation of data mining systems. Even if they are (indirectly) addressed, they are constrained to the preferences of the data mining engineers by the use of threshold values rather than the decision makers' preferences that should be continuously adjusted to the current dynamic business environment.

Normative decision analysis, on the other hand, is usually built around a prescriptive and rigid problem

* Corresponding author. Tel.: +82 2 958 3684; fax: +82 2 958 3604.
E-mail addresses: dhchoi@kgsm.kaist.ac.kr (D.H. Choi), bsahn@hansung.ac.kr (B.S. Ahn), seekim@kgsm.kaist.ac.kr (S.H. Kim).

structure called a decision analysis model. Decision analysis is not compatible with extracting knowledge from large corporate databases of nowadays, since it does not focus on the automotive generation of meaningful knowledge from raw data. It can be, thus, meaningful to complement the techniques from data mining with those from decision analysis: data mining for alternatives (i.e. rules) generation from large database and decision analysis for prioritizing those alternatives by reflecting decision makers' preferences.

In existing data mining techniques, there exist some situations that make necessary the prioritization of rules for selecting and concentrating on more valuable rules due to the number of qualified rules (Tan & Kumar, 2000) and limited business resources. Even though the purpose of data mining is rule (pattern) extraction that is valuable for decision making, patterns are deemed 'interesting' just on the basis of passing certain statistical tests such as support/ confidence in data mining. To the enterprise, however, it remains unclear how such patterns can be used to maximize business objectives. The major obstacle lies in the gap between *statistic-based summaries* (the statistic-based pattern extraction) extracted by traditional rule mining and a *profit-driven action* (the value-based decision making) required by business decision making (Wang, Zhou, & Han, 2002) which is characterized by explicit consideration of conflicts of business objectives (Bar & Feigenbaum, 1981; Hayes-Roth, Waterman, & Lenat, 1983) and by multiple decision makers' involvement for corporate decision making.

In summary, the research objective is to prioritize association rules resulted from the data mining, taking into account their business values by explicitly incorporating the conflicting criteria of business values and by the managers' preference statements toward their trade-off conditions. For this purpose, a decision analysis method, e.g. the Analytic Hierarchy Process (AHP) is applied to aggregate the opinions of the group decision makers on what are the relevant criteria for evaluating business values of rules and relative importance of those criteria. Association rule mining, one of the data mining techniques, is then performed to capture a set of competing rules with their various business values and those are, in turn, used as input for the rule prioritization. Thus, the final rule selected from the appropriate decision method, e.g. ELECTRE reveals meaningful result obtained from the use of machine learning and human intelligence.

2. Research background

2.1. Business values of rules

In supporting decision making for the real business activities, decision makers are embarrassed when they encounter too many rules extracted and (or) those rules are of little interest for the decision making. The quality and quantity problems of these rules in traditional data mining applications stem from the fact that rules are extracted and selected only with statistical criteria, not concerning with the business values that are closely related with business strategies. Since one of the most important considerations when firms extract rules from large database is the business values that the firms could expect to obtain by decision making relied on those rules, some ways for reflecting business requirements in the process of rule selection, thus, are necessary for effective decision aids.

The importance of each of business values, on the other hand, varies according to the environments with which firms face and their business strategies. For example, after an Internet shopping mall engages in business, the frequency of rules might be considered relatively more important than others since more frequent transactions with the customers may imply wide acknowledgement in the market. Later, the monetary value of rules might become increasingly important with the purpose of maximizing profit from the loyal customers. For the firms dealing with the fashionable items such as clothes, music CD, or computer games, it might be considered to be important to capture trends in customers purchasing patterns since such items fall into the category having relatively short life cycle.

Recognizing those various requirements, researchers in data mining domain began to give attention to the business values in pattern extraction. Some researchers focused on discovering time trends and differences between datasets (Dong & Li, 1999; Song et al., 2001) or capturing trends of patterns by highlighting recent dataset (Zhang et al., 2003). Another group of studies deal with pruning or ranking association rules according to their frequency, statistical interdependence, and interestingness measures (Hilderman & Hamilton, 2001; Tan & Kumar, 2000). Recommending products to customers by considering expected profit (Kitts et al., 2000), selecting products based on association rules and product-specific profitability (Brijs, Goethals, Swinnen, Vanhoof, & Wets, 2000), or recommending target items and promotion strategy with the goal of maximizing the net profit encompasses the third kind of researches (Wang et al., 2002). The business values, reflecting each group of studies, can be largely categorized by three types including time trends (recency), statistical significance (frequency), and profit (monetary value), respectively. The related research efforts with the types of business values considered are shown in Table 1.

2.2. Association rule mining

Association rule mining has been widely used from traditional business applications such as cross-marketing, attached mailing, catalog design, loss-leader analysis, store layout, and customer segmentation (Agrawal et al., 1993; Srikant & Agrawal, 1995) to e-business applications such as the renewal of web pages (Cooley, Mobasher, & Srivastava,