Chapter 12

Association Rules Evaluation by a Hybrid Multiple Criteria Decision Method

Zhen Zhang
Dalian University of Technology, China

Chonghui Guo
Dalian University of Technology, China

ABSTRACT

Various rules can be generated from databases by using association rule algorithms, but only a small number of these rules may be selected for implementation due to the limitations of resources. Accordingly, evaluating the quality of these rules becomes a hot topic in the data mining field. Based on multiple criteria decision theory, a framework for evaluating the mined association rules using TOPSIS method with combination weights is proposed, which takes into account both objective interestingness measures and the users’ domain information. An example of market basket analysis is applied to illustrate the applicability of this method.

INTRODUCTION

Rapid advances in data collection and storage technology have enabled organizations to accumulate vast amounts of data. To extract useful information from large databases, data mining has become an essential tool for data analysis. In recent years, data mining has been widely applied in many fields, such as business, science and engineering (Kriegel et al., 2007; Tan, Steinbach, & Kumar, 2006).

DOI: 10.4018/978-1-4666-3998-0.ch012

Association rule mining, known as one of the most important techniques in data mining field, is useful for discovering interesting relationships hidden in large data sets. Various rules can be mined from databases by using association rule algorithms (Agrawal, Imielinski, & Swami, 1993). But only a small number of rules may be selected for implementation due to the limitations of business resources (Choi, Ahn, & Kim, 2005). Thus, the interestingness issue has been identified as an important problem in data mining, which refers to finding association rules that are interesting or useful to the users, not just any possible rule (Liu, Hsu, Chen, & Ma, 2000).
To deal with this problem, many works have been done in recent years. The previous studies focus on two aspects. One aspect is to discover new objective interestingness measures which are based on probability, statistics, or information theory (Geng & Hamilton, 2006), such as support, confidence used in association rule mining. In order to obtain interesting rules for the users, the problem of how to select a right interestingness measure is also discussed by Lenca, Meyer, Vaillant, and Lallich (2008) and Tan, Kumar, and Srivastava (2004). Another aspect deals with the discovery of subjectively interesting association rules, which takes into account not only the raw data, but also the users’ behaviors and background knowledge. In order to find interesting association rules, these methods usually require the users to input domain knowledge manually as constraints, or distinguish rules as interesting or uninteresting by interacting with the data mining system (Fodeh & Tan, 2007; Liu et al., 2000; Malhas & Al Aghbari, 2009; Padmanabhan & Tuzhilin, 1998; Silberschatz & Tuzhilin, 1996; Taniar, Rahayu, Lee, & Daly, 2008).

Although many studies have been done on discovering interesting association patterns, it may be difficult to select an appropriate interestingness measure before data mining is performed. It is more appropriate to carry out post-processing for the mined association rules. Furthermore, the result the users concern may involve many aspects of the mined patterns, so the evaluation of association rules sometimes should be a multiple criteria decision problem. Choi et al. (2005) used ELECTRE-II approach combined with group AHP for association rules prioritization which considered both objective criteria and subjective preferences of users. They believed the proposed method made synergy with decision analysis techniques for solving problems in data mining field. Chen (2007) developed their work and proposed a DEA based method to rank association rules. This method firstly uses a DEA model to obtain efficient association rules, and then applies another discriminated model to rank the efficient association rules. Obviously, this method needs to solve considerable numbers of linear programming models, and includes redundant computations and considerations. What’s more, this method can’t rank all the association rules (Toloo, Sohrabi, & Nalchigar, 2009). Based on Chen’s work, Toloo et al. (2009) used an integrated DEA model which was able to identify the most efficient association rule by solving just one mixed integer linear programming and proposed a new method for ranking association rules with multiple criteria. Although this method can rank all the association rules, the computation may become complex when the number of association rules is very large.

As a classical multiple criteria decision method, TOPSIS method can help decision makers organize the problems to be solved, and carry out analysis, comparisons and rankings of the alternatives easily. Besides, it is computationally efficient, and allows criteria weights to be incorporated into the decision process (Wang & Chang, 2007). Thus based on the previous work, this paper attempts to apply TOPSIS method with combination weights to evaluate the quality of association rules, and propose a new framework for association rules evaluation.

The rest of this paper is organized as follows. First of all, association rule mining is described briefly, and then TOPSIS method with combination weights is introduced in detail. After that, we propose a framework for association rules evaluation, and an example of market basket analysis is applied to illustrate the applicability of our method. Finally, we conclude this paper.

ASSOCIATION RULE MINING

Association rule mining proposed by Agrawal et al. (1993) is a useful technique for discovering relationships between items in databases. It has been widely used in business applications such as cross-marketing, attached mailing and catalog design (Choi et al., 2005).
Related Content

Do We Mean Information Systems or Systems of Information?

The Influence of Emotional Intelligence on Tacit Knowledge Sharing in Service Organizations

Tacit Knowledge: Conceptualizations and Operationalizations

Fuzzy Multiple Criteria Workflow Robustness and Resiliency Modeling with Petri Nets

On Knowledge Management: A Field Study
Peter H. Carstensen and Ulrika Snis (2000). *Internet-Based Organizational Memory and Knowledge Management* (pp. 170-199). [www.irma-international.org/chapter/knowledge-management-field-study/24679/](http://www.irma-international.org/chapter/knowledge-management-field-study/24679/)