Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 35 (2008) 245–253
Expert Systems with Applications
www.elsevier.com/locate/eswa

Using back-propagation to learn association rules for service personalization

Yo-Ping Huang [a,*], Wei-Po Chuang [a], Ya-Hui Ke [a], Frode-Eika Sandnes [b]

[a] Department of Computer Science and Engineering, Tatung University, Taipei 10451, Taiwan
[b] Faculty of Engineering, Oslo University College, Oslo, Norway

Abstract

A novel, practical and efficient strategy for providing personalization for online shoppers is proposed. Based on shoppers' previous purchasing behavior and the customer's previous choices the strategy is capable of suggesting relevant and desirable products to each customer accurately. The strategy is based on training a back-propagation neural network with association rules that are mined from a transactional database. Unlike most strategies that only consider the relationship between the purchased items, the proposed strategy also incorporates additional influential attributes such as the price of items and their merchandise category. A powerful confidence estimate is used to rank the suggestions. Experimental evidence is provided to demonstrate the effectiveness of the proposed strategy.

Keywords: Data mining model; Back-propagation neural network; Fuzzy membership degree; Association rules

1. Introduction

Recently, the need and importance of collecting and analyzing commerce data has increased rapidly. The widespread use of bar codes for most commercial products, the computerization of many business and government transactions, and the advances in data collection tools have provided us with huge amounts of transaction data. Millions of databases are used in business management, scientific and engineering data management, and other applications. The number of such databases grows rapidly because of the availability of powerful and affordable database systems.

Data mining has attracted much attention in the business intelligence research communities and the commercial sector in recent years. Data mining is the process of extracting useful information from large databases (Agrawal & Shafer, 1996; Chen, Han, & Yu, 1996; Chi, Xia, Yang, & Muntz, 2005; Han & Kamber, 2001; Lin & Kedem, 2002; Rajagopalan & Isken, 2001; Yen & Lee, 2004). Discovery of association rules from large databases is one important facet of data mining. Given a database of sales transactions, techniques can be exploited to discover the implicit relationship between the transactions. In today's competitive business world, data mining provides an efficient and cost effective way for companies to realize personalized service to its customers – providing the company with a competitive edge and ultimately increasing revenue and market share. Association rule mining and collaborative filtering are the two main approaches applied to recommender systems (Adomavicius & Tuzhilin, 2005; Cho, Kim, & Kim, 2002; Hsu, Chung, & Huang, 2004; Kim & Yum, 2005), and the work described herein falls into the family of association rule approaches. To learn the association rules derived using the data-mining model, several artificial intelligence methods are employed. The back-propagation model (Werbos, 1974) is used for predicting customer's purchasing behavior, because of its powerful feedback mechanism. Before applying the back-propagation mechanism,

* Corresponding author. Tel.: +886 2 21822928x6566; fax: 886 2 25943699.

E-mail addresses: yphuang@ttu.edu.tw (Y.-P. Huang), webber@fuzzy.cse.ttu.edu.tw (W.-P. Chuang), yhke@fuzzy.cse.ttu.edu.tw (Y.-H. Ke), Frode-Eika.Sandnes@iu.hio.no (F.-E. Sandnes).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2007.06.035

the representation of the training patterns should be considered carefully. ART-2 (adaptive resonance theory) is used to preprocess the input patterns in this study. ART-2 can learn to categorize either analog or binary data (Carpenter & Grossberg, 1987a; Carpenter & Grossberg, 1987b). Also, AprioriTid (Agrawal & Srikant, 1994; Borgelt et al., 2002; Chan & Au, 1997; Huang, Ouyang, Ke, & Lin, 2001; Huang, Ke, Ouyang, & Lin, 2002; Tung, Lu, Han, & Feng, 2003; Wur & Leu, 1999; Zu, Lu, & Shi, 1997) is used to extract the association rules that are used as training patterns for the back-propagation model.

Unlike conventional models that only consider items that appear simultaneously in the transactions, we propose a new model that incorporates additional factors including purchase price and product category. Therefore, fuzzy membership degree, back-propagation and ART-2 are used to derive an effective prediction model. The prediction model is subsequently used to recommend appropriate items to customers and hence provide them with a better personalized service. Users are therefore presented with a larger array of relevant and attractive products.

In this paper, we present a model for predicting what the customers will buy next and from what categories. The purpose of this prediction model is to provide a better personalized service to prospective customers. A well-trained back-propagation model can help decision makers determine the requirements of their customers. As a result, a company can recommend what its customers want in the near future.

This paper is organized as follows. In Section 1, we gave an introduction to this study. In Section 2, related work on association rules and neural networks are surveyed. Section 3 investigates the coding scheme and methodology of prediction model. Section 4 shows the experimental results. The last section concludes our work.

2. Related work

In this study, an improved data-mining model for personalized service provisioning is proposed. Underlying techniques are introduced in this section, including association rules, ART-2, and back-propagation.

2.1. Association rules extracted from large databases

The process of mining associations can be divided into two steps: (1) find all frequent itemsets; and (2) generate strong association rules from the frequent itemsets (Agrawal & Srikant, 1994; Chen et al., 1996). Two criteria are used to measure the frequent itemsets— minimum support and confidence. The support value of grouping item A with item B is defined as follows:

$$\text{Support}(\text{A} \Rightarrow \text{B}) = \frac{\#\_\text{tuples\_containing\_both\_A\_and\_B}}{\#\ \_\text{transactions}}. \quad (1)$$

Similarly, the confidence value of relating item A to item B is defined as follows:

$$\text{Confidence}(\text{A} \Rightarrow \text{B}) = \frac{\#\_\text{tuples\_containing\_both\_A\_and\_B}}{\#\ \_\text{tuples\_containing\_A}}. \quad (2)$$

Based on the assigned support and confidence values, one can determine whether interesting items appear frequent enough in the transaction database. Strategies for efficient retrieval of interesting items in large databases have attracted much attention in the past decade. Different approaches for discovering association rules are proposed (Agrawal & Shafer, 1996; Chi et al., 2005; Lin & Kedem, 2002; Yen & Lee, 2004). A widely cited technique is the Apriori algorithm (Agrawal & Srikant, 1994), which finds frequent itemsets by using candidate generation. The Apriori algorithm employs an iterative approach known as level-wise search, where $(k-1)$-itemsets are used to explore k-itemsets, to derive large itemsets. A two-step process, which consists of joining and pruning actions that look at how L_{k-1} is used to find L_K, is required. The joining step finds L_k, a set of candidate k-items generated by joining L_{k-1} with themselves. In the pruning step all candidates (C_k) having items no less than the minimum support are frequent by definition. AprioriTid (Chen et al., 1996) is a refined version of the original Apriori algorithm. The merit of the AprioriTid is that the database is not used for counting the support values after the first pass. Thus, it is more efficient than the original Apriori algorithm. The Boolean method is another approach for obtaining association rules (Wur & Leu, 1999). In the Boolean algorithm two major steps are needed. First, both logical OR and AND operations are used to compute the frequent item sets. Second, logical AND and XOR operations are applied to derive all interesting association rules. All the mentioned algorithms are capable of deriving the association rules without difficulty. Due to its efficiency, AprioriTid was used in this study.

The support and confidence values are useful for determining the association rules in many applications. However, using a support-confidence framework as a measurement can also be misleading in that it may consider a rule $A \Rightarrow B$ as interesting when, in fact, the occurrence of item-A does not imply the occurrence of item-B (Han & Kamber, 2001). To avoid this situation, the interestingness of association rules is analyzed by the means of correlation. Denote $P(A)$ as the probability of occurrence of A in the transaction database. The correlation between the occurrence of A and B can be defined as follows:

$$\text{corr}_{A,B} = \frac{P(A \cap B)}{P(A)P(B)}. \quad (3)$$

If the calculated value from Eq. (3) is higher than 1, equal to 1, or less than 1, it means the occurrence of A and B is positively correlated, uncorrelated, or negatively corre-