PERGAMON

Expert Systems with Applications 21 (2001) 157–168

Expert Systems with Applications

www.elsevier.com/locate/eswa

Mining the change of customer behavior in an internet shopping mall

Hee Seok Song[a,*], Jae kyeong Kim[b], Soung Hie Kim[a]

[a]*Graduate School of Management, Korea Advanced Institute of Science and Technology, 207-43, Cheongryangri, Dongdaemun, Seoul, 130-012, South Korea*

[b]*School of Business Administration, KyungHee University, no. 1, Hoeki-Dong, Dongdaemoon, Seoul, 130-701, South Korea*

Abstract

Understanding and adapting to changes of customer behavior is an important aspect for a internet-based company to survive in a continuously changing environment. The aim of this paper is to develop a methodology which detects changes of customer behavior automatically from customer profiles and sales data at different time snapshots. For this purpose, we first define the three types of changes as emerging pattern, unexpected change and the added/perished rule, then, we develop similarity and difference measures for rule matching to detect all types of change. Finally, the degree of change is evaluated to detect significantly changed rules. Our proposed methodology can evaluate the degree of changes as well as detect all kinds of change automatically from different time snapshot data. A case study on an internet shopping mall for evaluation of this methodology is also provided.

Keywords: Data mining; Association rule mining; Change mining

1. Introduction

Understanding and adapting to changes of customer behavior is an important aspect of surviving in a continuously changing environment. Especially for internet-based companies, knowing what is changing and how it has been changed is of crucial importance because it allows businesses to provide the right products and services to suit the changing market needs (Liu, Hsu, Han & Xia, 2000). More specifically, most decision makers in internet-based companies have a strong need to know and adapt to the answers to following questions about their customers.

- Which customer group's sales are gradually increasing?
- Which customer groups moved from product A to product B over the years?
- Whether certain groups of customers had gradually emerged to be the major buyers over the years?

Data mining is the process of exploration and analysis of large quantities of data in order to discover meaningful patterns and rules, but much of existing data mining research has focused on devising techniques to build accurate models and to discover rules. Relatively little attention has been paid to mining changes in databases collected over time (Liu et al., 2000). Most data mining techniques such as association rules, decision trees and neural networks cannot be applied alone to answer the above research questions, because they cannot handle dynamic situations well. Also, most data mining techniques usually ignore rare items which have a small frequency of occurrence, but rare items which have a large growth rate or decreasing rate may give very significant implications to managers in changing environment. These are the reasons why we are motivated to develop other methodology to detect change. Association rule mining finds interesting association relationships among a large set of data items (Agrawal, Imielinski & Swami, 1993a,b; Agrawal & Srikant, 1994). With massive amounts of data continuously being collected and stored, many industries are becoming interested in mining association rules from their databases. Association rule mining is used as a basic mining methodology in our research.

Detected changes can be usefully applied to plan various niche marketing campaigns. For example, in a shop, people used to buy beer and snacks together—now they still buy beer, but seldom buy snacks. The shop manager needs to know this information so that he/she can find out the reason for this and design some catalysts to attract customers to buy snacks again. As an another example, if a manager can find out that a certain customer's preference has moved from a medium-size car to a large-size car, then that manager can establish a trade-in plan for customers who have a medium-size car and have the intention of buying a large-size car for replacement. In this paper, we develop a methodology

* Corresponding author. Tel.: +82-2-958-3670; fax: +82-2-958-3604.
E-mail address: hssong@kgsm.kaist.ac.kr (H.S. Song).

0957-4174/01/$ - see front matter
PII: S0957-4174(01)00037-9

which detects changes automatically from customer profiles and sales data at different periods of time. The most common approach to discover changes between two datasets is to generate rules from each dataset and directly compare the rules by rule matching, but this is not a simple process because of the following reasons. First, some rules cannot be easily compared due to different rule structures. Second, even with matched rules, it is difficult to know what kind of change and how much change has occurred. To simplify these difficulties, we first define three types of changes as emerging pattern, unexpected change and the added/perished rule, then we develop similarity and difference measures for rule matching to detect all types of change. Finally, the degree of change is evaluated to detect significantly changed rules. The proposed methodology can evaluate degree of changes as well as detect all kinds of changes automatically from different time snapshot data. Furthermore, the methodology can also be applicable for discovery of different characteristics from different categorical data.

We begin by reviewing the concept of association rules which are a prerequisite of our research and the discussion of related works in Section 2. We define types of change and change detection problems to clarify our objectives in Section 3. In Section 4, we provide our methodology. A case study for evaluation and its business implications are presented in Section 5 and Section 6. Finally we summarize our contributions and outline areas for further research in the Conclusion section.

2. Background

2.1. Association rule mining

A typical association rule has an implication of the form $A \Rightarrow B$, where A is an itemset and B is an itemset that contains only a single atomic condition. The *support* of an association rule is the percentage of records containing itemsets A and B together. The *confidence* of a rule is the percentage of records containing itemset A that also contain itemset B. The support represents the usefulness of the discovered rule and the confidence represents certainty of the detected association rule. Fig. 1 shows two association rules of which the support is the same but the confidence of Rule 2 is larger than that of Rule 1.

Association rule mining finds all collections of items in a database whose confidence and support meet or exceed a prespecified threshold value. Apriori algorithm is one of the prevalent techniques used to find association rules (Agrawal et al., 1993a,b; Agrawal & Srikant, 1994; Agrawal, Srikant & Vu, 1997). Apriori operates in two phases. In the first phase, all large itemsets are generated. This phase utilizes the downward closure property of support. In other words, if k size (or length k) itemset is a large itemset, then all the itemsets below $(k-1)$ size must also be large itemsets.

Record ID	Items Bought
2000	A, B, C
1000	A, C
4000	A, D
5000	B, E, F
Discovered Association Rules	
Rule 1 : A ⇒ C (support: 50 %, confidence: 66.6 %)	
Rule 2 : C ⇒ A (support: 50 %, confidence: 100 %)	

Fig. 1. Dataset and discovered association rules.

Using this property, candidate itemsets of length k are generated from the set of large itemsets of length $(k-1)$ by imposing the constraint that all subsets of length $(k-1)$ of any candidate itemset must be present in the set of large itemsets of length $(k-1)$. The second phase of the algorithm generates rules from the set of all large itemsets. Please refer to the study of Agrawal et al. (1993a,b) for a more detail.

In this paper we use profile association rules to explain examples. A profile association rule is one type of association rule in which the left hand side of the rule consists of customer profile information, such as age, salary, years of education, and social status. The right hand side of the rules consists of customer behavior information, such as buying milk, diaper, beer, etc (Aggarwal, Sun & Yu, 1998). Since profile association rules give various chances to establish a target marketing strategy, we use the profile association rule in each example and case study, but the proposed methodology is not limited to special type of association rules.

2.2. Data mining in a changing environment

There are existing works that have been carried out on learning (Freund & Mansour, 1997; Helmbold & Long, 1994; Widmer, 1996) and mining (Bay & Pazzani, 1999; Ganti, Gehrke & Ramakrishnan, 1999; Han & Kamber, 2001; Liu et al., 2000; Nakhaeizadeh, Taylor & Lanquillon, 1998) in a changing environment. All the following related works focus on dynamic aspects or comparison between two different datasets or rules. They are clustered as six categories in this paper.

The first field of study that examines mining in a changing environment is rule maintenance (Cheung, Han, Ng & Wong, 1996a; Cheung, Ng & Tam, 1996b; Feldman, Aumann, Amir & Manila, 1997; Thomas, Bodagala, Alsabti & Ranka, 1997). The purpose of these studies is improving accuracy in a changing environment. For example, in the study of Cheung et al. (1996a,b), incremental updating techniques are proposed for the efficient maintenance of discovered association rules when new transaction data are added