Applied Soft Computing 12 (2012) 1068–1086

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Mining the change of customer behavior in fuzzy time-interval sequential patterns

Tony Cheng-Kui Huang*

Department of Business Administration, National Chung Cheng University, 168, University Rd., Min-Hsiung, Chia-Yi, Taiwan, Republic of China

ARTICLE INFO

Article history:
Received 12 May 2011
Received in revised form
19 September 2011
Accepted 7 November 2011
Available online 23 November 2011

Keywords:
Data mining
Change mining
Fuzzy sets
Sequential patterns
Time-interval

ABSTRACT

Comprehending changes of customer behavior is an essential problem that must be faced for survival in a fast-changing business environment. Particularly in the management of electronic commerce (EC), many companies have developed on-line shopping stores to serve customers and immediately collect buying logs in databases. This trend has led to the development of data-mining applications. Fuzzy time-interval sequential pattern mining is one type of serviceable data-mining technique that discovers customer behavioral patterns over time. To take a shopping example, (Bread, *Short*, Milk, *Long*, Jam), means that Bread is bought before Milk in a *Short* period, and Jam is bought after Milk in a *Long* period, where *Short* and *Long* are predetermined linguistic terms given by managers. This information shown in this example reveals more general and concise knowledge for managers, allowing them to make quick-response decisions, especially in business. However, no studies, to our knowledge, have yet to address the issue of changes in fuzzy time-interval sequential patterns. The fuzzy time-interval sequential pattern, (Bread, *Short*, Milk, *Long*, Jam), became available in last year; however, is not a trend this year, and has been substituted by (Bread, *Short*, Yogurt, *Short*, Jam). Without updating this knowledge, managers might map out inappropriate marketing plans for products or services and dated inventory strategies with respect to time-intervals. To deal with this problem, we propose a novel change mining model, *MineFuzzChange*, to detect the change in fuzzy time-interval sequential patterns. Using a brick-and-mortar transactional dataset collected from a retail chain in Taiwan and a B2C EC dataset, experiments are carried out to evaluate the proposed model. We empirically demonstrate how the model helps managers to understand the changing behaviors of their customers and to formulate timely marketing and inventory strategies.

1. Introduction

In today's world, where the market is competitive and products are multitudinous, various marketing strategies have been developed to attract customers to endorse products or services [26]. In such a situation, customer switching behavior [23,24,40] takes place all the time. Thus, managers often have to make changes in their marketing plans. In many cases, they might have no idea about how and where to start comprehending these changes and how they come about. To address this problem, they often adopt traditional approaches, such as market surveys; yet, such approaches time-consuming and costly and they are unable react to changes immediately. As a result, comprehending changes of customer behavior accurately and responding to customer need in a timely manner has become an essential problem. There are two main explanations for the significance of this problem [30]: (1) to follow trends based on changes, managers like to know in which direction trends are proceeding so as not to be left behind. They analyze changes of customer behavior to provide products and services corresponding to the changing needs of the customers. (2) To stop or to delay undesirable changes, managers also like to about undesirable changes as soon as possible so that they can develop remedial measures to stop or delay the pace of such changes.

As the problem states, we investigate two research questions. First, how managers can be responsive to changes of customer behavior in a dynamic market? Second, how managers can detect and utilize the changes they indentify in customer behavioral patterns in order to respond in an accurate and timely manner? In the management of electronic commerce (EC), many companies have developed on-line shopping stores to serve customers and immediately collect buying logs in databases. Therefore, managers can segment databases in accordance with different time-periods to analyze what is changing and how it has been changed [5]. According to changes in customer behavior, managers can provide the right products or services over different time-periods.

When it comes to the topic of changing behaviors, the primary task is to adopt practical tools to discover customer behavioral patterns. As a result, many data-mining techniques have been

* Tel.: +886 5 2720411x34319; fax: +886 5 2720564.
E-mail address: bmahck@ccu.edu.tw

1568-4946/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2011.11.017

proposed to discover useful information, such as product bundling [50], RFM (Recency, Frequency, and Monetary) sequential patterns [13], product recommendation [22,25], cross-selling [31], customer profiling [36,49], personalized marketing [8], personal moving profiles in wireless networks [21], intrusion detection [47], mental health [16], and fuzzy time-interval sequential patterns [10]. Among the numerous data-mining techniques applied in EC, mining sequential patterns plays an important role in helping managers discover customer behavior over time. The problem of mining sequential patterns was first introduced in the mid-1990s. It consists of discovering a set of subsequences that occur frequently in a sequence database [1]. An *Amazon.com* shopping example of a sequential pattern is described as follows: having bought the book Alice's Adventures in Wonderland, a customer returns to buy the book Introduction to Logic and then the book Queen Victoria: A Personal History. A mapping of letters and items' names is given as: (*a*, *Alice's Adventures in Wonderland*), (*b*, *Introduction to Logic*), (*c*, *Queen Victoria: A Personal History*), and (*d*, *The Lord of the Rings: the Fellowship of the Ring*). Moreover, Chen et al. [9] proposed a generalization of sequential patterns, called *time-interval sequential patterns*, which reveals not only the order of items, but also the time-intervals between successive items. A time-interval sequential pattern in a database can be described as follows: having bought *a*, a customer returns to buy *b* in two months and then *c* in a half month. The *Amazon.com* manager, therefore, can refer to this type of pattern and develop his/her inventory or marketing plans.

Such an approach, however, could cause a sharp boundary problem. That is, when a time-interval is near the boundary of two adjacent ranges, we either ignore or overemphasize it. For instance, let a time-interval of ti_1 be $1 \leq tg < 4$ and that of ti_2 be $4 \leq tg < 7$, where tg is the time gap between two successive items. Then, if the time gap between items *a* and *b* is near 4, either a little larger or smaller, it is difficult to say whether the time gap is in ti_1 or in ti_2. Hence, the case can only be one hundred percent in ti_1 or in ti_2. This difficulty can be adequately tackled by using fuzzy techniques, since fuzzy set theory allows the time gap to be 50% in ti_1 and 50% in ti_2 at the same time. Additionally, other patterns might be discovered from the same database, such as: having bought *a*, a customer returns to buy *b* in three months and then *c* in one month and having bought *a*, a customer returns to buy *b* in two and a half months and then *c* in one month. These detailed but similar patterns make it cumbersome for managers to make decisions. Therefore, a fuzzy extension, called *fuzzy time-interval sequential patterns* (FTSP), was proposed by Chen and Huang [10] to find the fuzzy time-interval between items in sequential patterns. A fuzzy time-interval sequential pattern takes the following form.

Having bought *a*, a customer returns to buy *b* in a *Long* period and then *c* in a *Short* period.

The pattern can be represented by (*a*, *Long*, *b*, *Short*, *c*) as well. This simple example indicates that the fuzzy concept is better than the partition method because fuzzy sets provide smooth transitions between members and non-members of a set.

As mentioned in the above introduction, we know that there are several advantages to applying fuzzy time-interval sequential patterns in EC. First, the knowledge of managers in decision making can be represented more naturally and appropriately by fuzzy logic, and partitioning and representing time information are types of manager knowledge. Second, it is widely acknowledged that many real-life situations are intrinsically fuzzy, and partition of time information is one such situation. Third, using linguistic terms is simple and easy for managers. As part of executive work activities, managers often take care of strategic issues and long-term trends [48]. Therefore, verbal communication with linguistic terms is preferred for the exchange of soft knowledge. Moreover, since customer requirements in EC markets are often changing, managers need to gain more general and concise knowledge, not precise knowledge in the form of figures, for making quick-response decisions. Accordingly, any manager can first define the linguistic terms that are meaningful and understandable to them and then finally discover specific knowledge.

Although applying this technique to mining FTSP is workable, it still fails to take into account customer behavioral changes in fast-changing EC environments. For example, a fuzzy time-interval sequential pattern (*a*, *Long*, *b*, *Short*, *c*) may be available for the previous year. The pattern, however, is not necessary a trend this year, and it may be substituted by (*a*, *Short*, *b*, *Long*, *d*). If managers cannot capture the customer behavioral change in time, two failed beliefs between the previous year and the present year still exist in their minds, including:

(1) They still believe that a customer will buy *b* after buying *a* in a *Long* period. In fact, the period has been changed from *Long* to *Short*.
(2) After buying *a* and *b*, a customer will buy *c* in a *Short* period; however, the item has been changed from *c* to *d* and the period is not *Short* but *Long*.

Without updating this knowledge, managers fail to provide appropriate products or services to customers and may adopt inappropriate inventory strategies with respect to time-intervals. No studies, to our knowledge, have yet to address the issue of changes in fuzzy time-interval sequential patterns so that managers can detect and utilize these types of changes. Therefore, our research goal is to propose a novel change mining model, *MineFuzzChange*, to detect FTSP change.

The remainder of this paper is organized as follows. Section 2 reviews related works. Section 3 defines a similarity measurement for fuzzy time-interval sequential patterns. Section 4 presents the *MineFuzzChange* model for mining changes in fuzzy time-interval sequential patterns. Section 5 shows the experimental results of the proposed model. Conclusions are drawn in Section 6.

2. Related works

Previous works have addressed the change problem in two ways: (1) by adopting a maintenance approach when new transactions are constantly added to the original database [7,17,19,32,33,35,44] and (2) by comparing different databases to discover change patterns [11,14,30,43,46].

The first way is to develop approaches to improve the accuracy of rules or patterns in a changing environment. The way rules or patterns are discovered is adjusted when new data are recorded to the database. These approaches are used to maintain existing knowledge but do not provide any extra information; this is, managers or users obtain the final outcomes but cannot understand the evolutionary processes as customer behaviors are changing. By reviewing previous studies, we have found that the above approach has contributed to two topics, mining sequential patterns and association rules. On the topic of mining sequential patterns, Masseglia et al. [35], El-Sayed et al. [17], and Lin and Lee [32] developed efficient strategies to improve performance in the incremental mining of patterns. In these studies, experimental evaluations were performed to show the computational efficiency and scalability of the proposed algorithms. In addition, Subramanyam and Goswami [44] proposed a fuzzy data mining algorithm for incremental mining of sequential patterns from quantitative databases. On the topic of mining association rules, Cheung et al. [7], Lee et al. [33], and Hong et al. [20] also focused performance improvement to develop rule maintenance methods. The experimental results of the above studies show that their methods were proposed to address the issues of