Applied Soft Computing 41 (2016) 265–274

Contents lists available at ScienceDirect

Applied Soft Computing

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

Mining fuzzy temporal association rules by item lifespans☆

Chun-Hao Chen [a], Guo-Cheng Lan [b], Tzung-Pei Hong [c,d,*], Shih-Bin Lin [d]

[a] Department of Computer Science and Information Engineering, Tamkang University, Taipei 251, Taiwan
[b] Department of Computer Science and Information Engineering, National Cheng Kung University, 701, Taiwan
[c] Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan
[d] Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan

ARTICLE INFO

Article history:
Received 2 July 2012
Received in revised form
30 November 2015
Accepted 8 January 2016
Available online 13 January 2016

Keywords:
Fuzzy set
Fuzzy data mining
Fuzzy temporal association rule
Item lifespan

ABSTRACT

Data mining is the process of extracting desirable knowledge or interesting patterns from existing databases for specific purposes. In real-world applications, transactions may contain quantitative values and each item may have a lifespan from a temporal database. In this paper, we thus propose a data mining algorithm for deriving fuzzy temporal association rules. It first transforms each quantitative value into a fuzzy set using the given membership functions. Meanwhile, item lifespans are collected and recorded in a temporal information table through a transformation process. The algorithm then calculates the scalar cardinality of each linguistic term of each item. A mining process based on fuzzy counts and item lifespans is then performed to find fuzzy temporal association rules. Experiments are finally performed on two simulation datasets and the foodmart dataset to show the effectiveness and the efficiency of the proposed approach.

1. Introduction

Data mining is commonly used for extracting association rules from transaction data. An association rule is an expression $X \rightarrow Y$, where X and Y are a set of items. It means that in the set of transactions, if all the items in X exist in a transaction, then there is a high probability that Y is also in the transaction [1]. Most previous studies have focused on binary-value transaction data. Transactions in real-world applications, however, usually consist of quantitative values. Also, each item may have a publication time interval. This means that each item has a lifespan. Thus, designing a sophisticated data-mining algorithm able to deal with this type of data is a challenge for researchers in this field.

The fuzzy set theory has been used more and more frequently in intelligent systems because of its simplicity and similarity to human reasoning [7,13,19,28,29]. Several fuzzy data mining algorithms for inducing rules from a given set of data have been designed and used with good results for specific domains [6,11,15,27]. For example, Chan et al. proposed an *F-APACS* algorithm to mine fuzzy association rules [6]. Kuok et al. proposed a fuzzy mining approach to handle numerical data in databases and derived fuzzy association rules [12]. At nearly the same time, Hong et al. proposed a fuzzy mining algorithm to mine fuzzy rules from quantitative transaction data [11]. Yue et al. then extended the above concept to find fuzzy association rules with weighted items [27].

However, those fuzzy data mining approaches did not consider item lifespans. Although Lee proposed two algorithms for discovering fuzzy temporal association rules and fuzzy periodic association rules using fuzzy calendar algebra [21], the lifespan of each item was not considered. Generally speaking, each product has its own lifespan in a supermarket. If the traditional data mining approach is applied to mine rules, then some new products will not become frequent itemsets because the supports of those items cannot reach the minimum support threshold such that some associations among them will be missed. In this paper, we thus propose a fuzzy mining algorithm for deriving linguistic temporal association rules from a given temporal transaction database by considering item lifespans. It first transforms each quantitative value into fuzzy sets using given membership functions. During the transformation process, the lifespans of the items are collected and recorded in the temporal information table. The algorithm then calculates the scalar cardinality of each linguistic term of the items. The support values of items are calculated according to their corresponding periods such that the real importance of each item can be

☆ This is a modified and expanded version of the paper “Mining fuzzy temporal knowledge from quantitative transactions,” presented at *International Conference on System Science and Engineering*, 2011, Macao.

* Corresponding author at: Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan.
Tel.: +886 75919191.
E-mail addresses: chchen@mail.tku.edu.tw (C.-H. Chen), rrfoheiay@gmail.com (G.-C. Lan), tphong@nuk.edu.tw (T.-P. Hong), shinbinxp@gmail.com (S.-B. Lin).

http://dx.doi.org/10.1016/j.asoc.2016.01.008
1568-4946/

represented. The mining process, which is based on fuzzy counts, is then performed to find fuzzy temporal association rules, which can not be mined by the existing (traditional) fuzzy rule mining approaches since all the transactions are considered to calculate support values of items. Experiments are also performed using two simulation datasets and the foodmart dataset [12,22] to show the effectiveness and efficiency of the proposed approach. There are two contributions of the proposed approach. The first one is that we propose in fuzzy data mining the first work which takes the lifespans of products into consideration such that the associations of items can be exhibited more correctly. The second one is that by the proposed approach more useful fuzzy rules can be derived in terms of average supports and confidences through the experimental evaluation on some simulation and real datasets.

The rest of this paper is organized as follows. The problem statement and some definitions are described in Section 2. Some related mining approaches are reviewed in Section 3. The details of the proposed fuzzy temporal association rule mining algorithm are given in Section 4. An example that illustrates the use of the proposed algorithm is described in Section 5. Experiments that demonstrate the performance of the proposed approach are stated in Section 6. Conclusions and suggestions for future work are discussed in Section 7.

2. Problem statement and definitions

Since new products are continuously added to the market, each item in a transaction has its own lifespan. Some related terms such as temporal item, temporal transaction database, and temporal support are defined as follows [8].

Definition 1 *(Temporal Item).* Let $P = <p_1, p_2, \ldots, p_n>$ be a sequence of continuous time periods such that each period has a particular time granularity, e.g., day, week, month, year, etc. $\forall\, i, j, n, 1 \le i \le j \le n$, p_i occurs before p_j, which is denoted as $p_i \le p_j$. A temporal item x is an item with a starting period, denoted as $S(x)$. The starting period of a temporal itemset X is then defined as $S(X) = \max(\{S(x)\})$, where $x \in X$. The set of temporal items is denoted as I such that $\forall\, x \in I$, $S(x) \le p_n$.

Definition 2 *(Temporal Transaction Database).* A temporal transaction, T, is a subset of I, which means $T \in I$. The period in which T occurs is denoted $O(T)$. Then, $D = \{T \mid p_1 \le O(T) \le p_n\}$ is a temporal transaction database on I over P.

Definition 3 *(Temporal Support).* Let D be a temporal transaction database on I over P. Let X be a set of temporal items. The temporal support of X, with respect to the subset of D from the period p_i, denoted as $TSup(X, p_i)$, is defined as formula (1):

$$TSup(X, p_i) = \frac{|\{T | X \subseteq T, O(T) \ge p_i, T \in D\}|}{|\{T | O(T) \ge p_i, T \in D\}|} \quad (1)$$

Then the temporal support of X, denoted as $TSup(X)$, can be computed as $TSup(X, S(X))$. That is, the temporal support of itemset X is the ratio of the number of transactions that contain temporal itemset X to the total number of transactions that occur from $S(X)$.

From the above definitions, it can be seen that many previous studies have focused on handling binary-valued data. Transaction data, however, usually have quantitative values. We thus define the temporal quantitative item and temporal quantitative transaction database by extending Definitions 1 and 2. The definitions are as follows.

Definition 4 *(Temporal Quantitative Item).* Let $P = <p_1, p_2, \ldots, p_n>$ be a sequence of continuous time periods such that each period has a particular time granularity, e.g., day, week, month, year, etc. $\forall\, i, j, n, 1 \le i \le j \le n$, p_i occurs before p_j, which is denoted as $p_i \le p_j$.

Table 1
The five temporal quantitative transactions.

Period	TID	Items
P_1 (Aug-5)	$Trans_1$	(A, 5), (C,4);
	$Trans_2$	(A, 3), (B,2);
	$Trans_3$	(C,4);
P_2 (Aug-6)	$Trans_4$	(A, 3), (B, 2), (D,4);
	$Trans_5$	(A, 3), (B, 5), (D,4), (E,2);

Temporal quantitative item x' consists of temporal item x and quantitative value v, denoted as (x, v), where v is a positive real number. The temporal item is a special kind of temporal quantitative item. The starting period of a temporal quantitative item x', denoted as $S(x')$, is the same as temporal item x, which is $S(x)$. In the same way, the starting period of a temporal quantitative itemset X' is $S(X') = \max(\{S(x')\})$, where $x \in X$. The set of temporal quantitative items is denoted as I' s.t. $\forall\, x' \in I'$, $S(x') \le p_n$.

Definition 5 *(Temporal Quantitative Transaction Database).* Temporal quantitative transaction T' is a subset of I', which means $T' \in I'$. The period that T' occurs is denoted as $O(T')$. Then, $D' = \{T' \mid p_1 \le O(T') \le p_n\}$ is temporal quantitative transaction database I' over P.

For example, Table 1 is a temporal quantitative transaction database with five temporal quantitative transactions over two periods, $P = \{P_1, P_2\}$. The time granularity used in this example is day. There are five temporal items. That is $I = \{A, B, C, D, E\}$. Each temporal quantitative item has its own starting period, e.g., $S(A) = 1$ and $S(A, D) = 2$. Each temporal quantitative transaction also has its own starting period, e.g., the starting period $O(T')$ of the 4th transaction is 2.

In addition, since fuzzy theory is also commonly used to deal with quantitative values, in this paper, we thus utilize fuzzy theory to transform quantitative values into a more friendly representation that we can easily understand. We assume a given set of membership functions, and that quantitative value $v_j^{(i)}$, which is the quantitative value of temporal item x (=I_j) in transaction $T'^{(i)}$, can be transformed into a fuzzy set represented as $\left(\frac{f_{j_1}^{(i)}}{R_{j_1}} + \frac{f_{j_2}^{(i)}}{R_{j_2}} + \cdots + \frac{f_{j_l}^{(i)}}{R_{j_l}}\right)$, where R_{jk} is the kth fuzzy region (linguistic term) of temporal item I_j, $f_j^{(i)}$ is $v_j^{(i)}$'s membership value in region R_{jk}, and l (=$|I_j|$) is the number of fuzzy regions for I_j. Then, we define the fuzzy temporal support of a fuzzy region and fuzzy temporal confidence as follows:

Definition 6 *(Fuzzy Temporal Support).* Fuzzy temporal support of a fuzzy region R_{jk} of a temporal item x (=I_j) is defined as formula (2):

$$tFuzzySupport_{jk} = count_{jk} / |\{T' | O(T') \ge S(I_j), T' \in D'\}| \quad (2)$$

where $count_{jk}$ is the scalar cardinality of fuzzy region R_{jk}, which is the summation of membership values $f_{jk}^{(i)}$. In other words, the fuzzy temporal support of a fuzzy region of a temporal item is the ratio of its scalar cardinality to the number of transactions that occur from the starting period of temporal item x.

Let s be a temporal q-itemset with items $(s_1, s_2, \ldots, s_q)$, $q \ge 2$. Similarly, fuzzy temporal support of fuzzy regions R_s of temporal itemset X (=s) is defined as formula (3):

$$tFuzzySupport_s = count_s / |\{T' O(T') \ge S(s), T' \in D'\}| \quad (3)$$

where $count_s$ is the scalar cardinality of fuzzy region R_s, which is the summation of membership values $f_s^{(i)}$. It should be noted that if