Applied Computing and Informatics (2016) 12, 134–141

Saudi Computer Society, King Saud University

Applied Computing and Informatics

(http://computer.org.sa)
www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Discovery of temporal association rules with hierarchical granular framework

Tzung-Pei Hong [a,b,*], Guo-Cheng Lan [c], Ja-Hwung Su [d], Pei-Shan Wu [b], Shyue-Liang Wang [e]

[a] *Department of Computer Science and Information Engineering, National University of Kaohsiung, 811, Taiwan, ROC*
[b] *Department of Computer Science and Engineering, National Sun Yat-Sen University, 804, Taiwan, ROC*
[c] *Department of Computer Science and Information Engineering, National Cheng Kung University, 701, Taiwan, ROC*
[d] *Department of Information Management, Cheng Shiu University, 833, Taiwan, ROC*
[e] *Department of Information Management, National University of Kaohsiung, 811, Taiwan, ROC*

Received 4 November 2014; revised 24 November 2015; accepted 28 January 2016
Available online 2 February 2016

KEYWORDS

Data mining;
Association-rule mining;
Temporal association rules;
Item lifespan;
Time granules

Abstract Most of the existing studies in temporal data mining consider only lifespan of items to find general temporal association rules. However, an infrequent item for the entire time may be frequent within part of the time. We thus organize time into granules and consider temporal data mining for different levels of granules. Besides, an item may not be ready at the beginning of a store. In this paper, we use the first transaction including an item as the start point for the item. Before the start point, the item may not be brought. A three-phase mining framework with consideration of the item lifespan definition is designed. At last, experiments were made to demonstrate the performance of the proposed framework.

1. Introduction

Data mining can help derive useful knowledge from databases. Among its technology, association-rule mining [1,3,28] considers frequency relationship among items and is commonly applied to many applications. A transaction usually includes the items bought and the time of its occurrence. Besides, the periods for items to be exhibited are also important. Some researches about temporal data mining were thus presented [27]. For example, the time period for an item may be the entire time interval of a database [5], the duration from the first occurring time of the item to the end of a database [20], or the on-shelf time periods of the item [8]. However, an infrequent item for the entire time interval may be frequent within part of the time.

* Corresponding author at: Department of Computer Science and Information Engineering, National University of Kaohsiung, 811, Taiwan, ROC.
E-mail addresses: tphong@nuk.edu.tw (T.-P. Hong), rrfoheiay@gmail.com (G.-C. Lan), bb0820@ms22.hinet.net (J.-H. Su), grani1987@gmail.com (P.-S. Wu), slwang@nuk.edu.tw (S.-L. Wang).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.aci.2016.01.003
2210-8327

In this paper, we thus organize time into granules and consider temporal data mining for different levels of granules. We use the first transaction including an item as the start point for the item. We propose a three-phase mining framework with consideration of the above item lifespan definition to mine temporal association rules with time granules from a temporal database. According to the definition of item lifespan, in the first phase, each elementary time interval is processed. The temporal frequent itemsets within the above intervals are first found, and then the itemsets are identified as candidate temporal frequent ones in all the time granules of the upper level of the hierarchy. These candidates are then judged for being temporal or not at each level of granules. Additional database scans may be needed to find the actual supports of the candidates. In the third phase, the possible candidate association rules are derived from the temporal frequent itemsets at each level. Their confidence values are then calculated and compared with the minimum confidence value to get the final temporal association rules.

The organization of the paper is stated below. Related works are given in Section 2. The problem to be solved is described in Section 3. The proposed algorithm with consideration of the first transaction appearance period is presented in Section 4. The performance of the proposed approach is shown in Section 5. Conclusions and future works are finally given in Section 6.

2. Review of related works

Temporal data mining is popular in recent years. It analyzes temporal data to get patterns or regularities. There are many techniques included in temporal data mining. Sequential association mining [2], cyclic association mining [22], stock trading rule mining [11], patent mining [12], clinical mining [25], image time series mining [15], software adoption and penetration mining [23], temporal utility mining [9,29], fuzzy temporal mining [6,16,17], and calendar association mining [21] all belong to it. There are also a variety of applications for temporal data mining. For example, Patnaik et al. used temporal data mining to efficiently manage the cooling system in data centers [24], and Rashid et al. adopted it for finding the correlation among sensor data [26].

Chang et al. considered the temporal mining problem of products exhibited in a store [5]. They proposed the concept of common exhibition to find patterns. In a common exhibition period, all the items in an itemset need to be on the shelf at the same time. Lee et al. then used it to discover general temporal association rules for publication databases [20]. Ale and Rossi then considered the transaction periods of products [4], instead of their exhibition periods, for finding temporal association rules. Besides, different products may have different on-shelf properties. For example, a popular product may be sold out quickly, and then be supplied and on shelf soon. It is thus intermittently on-shelf and off-shelf in the entire time [18].

As to hierarchical temporal mining, Li et al. proposed an approach to discover calendar-based temporal association rules [21]. That approach could mine rules according to different calendar constraints including years, months and days. Chen et al. proposed a hierarchical strategy for video event detection from video databases [7]. They divided the frequent actions into two types, namely pre-actions and post-actions by pre- and post-temporal windows. Fang and Wu used granules of features to speed up the mining process of association rules [10].

In this paper, we consider the phenomenon that an itemset may not be frequent in the entire time interval, but may be frequent in a partial time interval. We thus organize the time into different levels of granules and find the temporal association rules at each level. This paper is extended from our previous work [19] with different consideration of effective time intervals. Here we use the first occurring transaction of an item as the start point for the item. Before the start point, the item may not be brought since it is not ready. This definition is of the benefit that it is not necessary to require the exact on-shelf time of each item in advance.

3. Problem statement and definitions

To describe the problem of hierarchical temporal association rule mining clearly, assume a temporal database (abbreviated as *TDB*) in Table 1 is given. Four items are included in the transactions, denoted A to D.

In addition, there is a pre-defined hierarchy with time granules in three levels, in which there are four basic time periods, denoted as p_1 to p_4, and the time granules are in three levels in the hierarchy, as shown in Fig. 1. Based on Fig. 1 and Table 1, $\{C\} \rightarrow \{D\}$ is one of hierarchical temporal association rules occurring in the time granule p_{12}. The goal of this paper was to mine such temporal association rules, and the detailed definitions and examples will be described as follows.

The terms related to the hierarchical temporal mining under the first occurring transaction periods of items are explained below.

Definition 1. $P = \{p_1, p_2, \ldots, p_j, \ldots, p_n\}$ is a set of mutually disjoint time periods, where p_j denotes the j-th time period in the whole set of periods, P.

Definition 2. Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of items appearing in a database. If $X \subseteq I$, then X is called an itemset.

Definition 3. Let X be an itemset and t be a time stamp. A transaction T is a pair (X, t).

Table 1 An example of a temporal database.

Period	TID	Items
p_1	$Trans_1$	D
	$Trans_2$	C, D
	$Trans_3$	C
	$Trans_4$	D
p_2	$Trans_5$	A, C, D
	$Trans_6$	A, B, C, D
	$Trans_7$	B, C, D
	$Trans_8$	A, D
p_3	$Trans_9$	B
	$Trans_{10}$	A, C
	$Trans_{11}$	A, B, C
	$Trans_{12}$	B, C
p_4	$Trans_{13}$	B, D
	$Trans_{14}$	B, C, D
	$Trans_{15}$	B
	$Trans_{16}$	B, C, D