Using a projection-based approach to mine frequent inter-transaction patterns

Chun-Sheng Wang\(^a\), Kuo-Chung Chu\(^b\,*\)

\(^a\)Department of Information Management, Jinnwen University of Science and Technology, No. 99, An-Chung Road, Hsin-Tien, Dist. New Taipei City, Taiwan, ROC
\(^b\)Department of Information Management, National Taipei University of Nursing and Health Sciences, No. 365, Min-Te Road, Taipei, Taiwan, ROC

**Abstract**

In this paper, we propose an algorithm called PITP-Miner that utilizes a projection based approach to mine frequent inter-transaction patterns efficiently. The algorithm only searches for local frequent items in a projected database that stores potential local inter-transaction items and partitions the database into a set of smaller databases recursively. In addition, two pruning strategies are designed to further condense the partitioned databases and thus accelerate the algorithm. Our experiment results demonstrate that the proposed PITP-Miner algorithm outperforms the ITP-Miner and FITI algorithms in most cases.

© 2011 Published by Elsevier Ltd.

**1. Introduction**

Mining association rules from large databases have received considerable attention in recent years and numerous studies have been conducted (Agrawal, Imielinski, \& Swami, 1993; Han \& Kamber, 2006; Lee, Lin, \& Wang, 2006; Lee, Hong, Ko, Tsao, \& Lin, 2007). Traditional association rule mining methods are intra-transactional in nature, so they only consider associations that occur within the same transaction. In contrast, an inter-transaction association rule can represent the associations of items within a transaction, as well as the associations of items across different transactions along a certain dimension (Lu, Feng, \& Han, 2000). For example, an inter-transaction association rule might state: "When the prices of Intel and IBM go up, 85% of the time, the price of SUN will increase on the same day". However, an inter-transaction can extend such a rule to: "When the prices of Intel and IBM go up, 85% of the time, the price of SUN will increase two days later".

Lu, Han, and Feng (1998) first used inter-transaction association rules to predict stock market movements, and Feng, Dillon, and Liu (2001) applied such rules to meteorological data study. Subsequently, Li, Feng, and Wong (2005) extended inter-transaction association rules to a more general form of association rules, called generalized multidimensional inter-transaction association rules. Several algorithms have been proposed for mining frequent inter-transaction patterns from large databases. For example, Lu et al. (2000) introduced the E/EH-Apriori (Extended/Extended Hash-based Apriori) algorithm, in which E-Apriori uses the Apriori property to discover frequent inter-transaction patterns and EH-Apriori uses a hashing technique to prune the number of candidates of length-2. Feng, Yu, Lu, and Han (2002) described a constraint-based inter-transaction association rules mining approach that uses several optimization techniques to enhance the EH-Apriori algorithm under rule templates. The following year, Tung, Lu, Han, and Feng (2003) proposed the FITI (First-Intra-Then-Inter) algorithm, which is implemented in two phases. First, the Apriori algorithm (Agrawal \& Srikant, 1994) is executed to discover frequent intra-transaction patterns, which are then used to generate frequent inter-transaction patterns in the second phase. More recently, Lee and Wang (2007) presented an algorithm called ITP-Miner, which uses vertical data structure dat-lists and an ITP-tree to mine all frequent inter-transaction patterns in a depth-first search manner. It has been shown that ITP-Miner outperforms previous inter-transaction mining algorithms.

The task of mining all frequent inter-transaction patterns in very large databases is quite challenging because the transaction boundaries are broken, and the search space increases exponentially when some of the mining parameters change. In this paper, we propose a projection-based approach called the PITP-Miner algorithm for efficient mining of frequent inter-transaction patterns in a large transaction database. The approach is based on a divide-and-conquer, pattern-growth principle, which means that the algorithm searches along a structure called a PITP-tree in a depth-first search manner. In a PITP-tree, each node represents a frequent inter-transaction pattern and an associated projected database that stores local potential inter-transaction items. In the algorithm, a projected database is recursively projected into a set of smaller projected sub-databases, after which frequent inter-transaction patterns are grown in each sub-database by searching only local frequent fragments. In addition, we introduce two pruning strategies, called ancestor node pruning and hash table pruning, to further condense the projected sub-databases and thus accelerate the algorithm. A performance study shows that, in most cases, the proposed PITP-Miner algorithm outperforms the ITP-Miner and FITI algorithms.
The contribution of this paper is threefold. First, we propose an effective projection-based algorithm for mining frequent inter-transaction patterns efficiently. Second, we design two pruning strategies to reduce the search space and thus speed up the algorithm. Third, we demonstrate via experiments that the proposed algorithm outperforms the ITP-Miner and FITI algorithm in most cases.

The remainder of this paper is organized as follows. Section 2 defines the frequent inter-transaction pattern mining problem and introduces some notations. In Section 3, we present the PITP-Miner algorithm. Section 4 describes the experiments and their performance results. Finally, we present our conclusions and indicate some future research directions in Section 5.

2. Problem description

In traditional association mining models, each transaction in a database contains a set of items. Although transactions may occur in different contexts, such as time and location, traditional models ignore this contextual information because the patterns are intra-transactional in nature. However, if we are interested in inter-transaction patterns across multiple transactions, the context in which a transaction occurs is important.

We define a model for inter-transaction pattern mining as follows. An itemset $I = \{i_1, i_2, \ldots, i_s\}$ is a set of distinct items. When there is only one item in an itemset, the parentheses can be omitted; that is, $(i)$ can be written as $i$. Items in an itemset are listed in alphabetical order, i.e., we write $(a, b, d)$ instead of $(d, a, b)$. A transaction database $D = \{(t_1, t_2, \ldots, t_n)\}$, where $|D|$ is the number of transactions in $D$ and $t_i (1 \leq i \leq |D|)$ is a transaction of the form $(\text{Dat}, \text{Itemset})$. $\text{Dat}$ is the domain attribute of $t_i$ that describes the contextual information, such as the time stamp or space location associated with $t_i$. As $\text{Dat}$ is an ordinal, it can be divided into intervals of equal length. For example, time is a domain attribute that can be divided into days, weeks, etc. These intervals can be represented by non-negative integers $0, 1, 2,$ and so on. To demonstrate our proposed inter-transaction pattern mining algorithm, we use a transaction database containing four transactions, as shown in Table 1.

Table 1: A transaction database.

<table>
<thead>
<tr>
<th>Transaction</th>
<th>Dat</th>
<th>Itemset</th>
<th>Megatransaction (maxspan = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>0</td>
<td>(a, c)</td>
<td>$(a^2, c^2, a^1, c^1, d^1)$</td>
</tr>
<tr>
<td>$t_2$</td>
<td>1</td>
<td>(a, c, d)</td>
<td>$(a^2, c^2, d^2, c^1)$</td>
</tr>
<tr>
<td>$t_3$</td>
<td>2</td>
<td>(c)</td>
<td>$(c^2, a^1, b^1, c^1, d^1)$</td>
</tr>
<tr>
<td>$t_4$</td>
<td>3</td>
<td>(a, b, c, d)</td>
<td>$(a^2, b^2, c^2, d^2)$</td>
</tr>
</tbody>
</table>

In a transaction database, let $x$ be a pattern, and $T_x$ be a set of megatransactions in $D$, where each megatransaction in $T_x$ contains $x$. The support of $x$, $\text{sup}(x)$, is defined as $|T_x|$. If $\text{sup}(x)$ is not less than the user-specified minimum support threshold $\text{minsup}$, $x$ is called a frequent pattern. An inter-transaction association rule is written in the form of $x \rightarrow \beta$, where $x$ and $\beta$ are frequent patterns: $x \cap \beta = \varnothing$ and $\text{conf}(x \rightarrow \beta)$ is not less than the user-specified minimum confidence; and the confidence of the rule $\text{conf}(x \rightarrow \beta)$ is defined as $\text{sup}(x \cup \beta)/\text{sup}(x)$. The problem of inter-transaction pattern mining involves finding all frequent inter-transaction patterns in a transaction database with respect to the user-specified $\text{minsup}$ and maxspan thresholds.

Let us consider the example shown in Table 1. Assume that maxspan $= 1$, minsup $= 2$ and $X = \{(c^2, a^1)\}$. Since the megatransactions formed by the first and second transactions, $(a^2, c^2, a^1, c^1, d^1)$, contains $X$, sup($X$) is set to 1. In addition, the megatransaction formed by the third and fourth transactions, $(c^2, a^1, b^1, c^1, d^1)$, contains $X$. Thus, sup($X$) is incremented by 1, that is, sup($X$) $= 2$. Consequently, $X$ is a frequent pattern. Let another pattern $Y = \{(c^2, a^1, c^1)\}$, then, we can find sup($Y$) $= 2$, so $\text{conf}(\{(c^2, a^1)\} \rightarrow \{(c^2, a^1, c^1)\}) = 2/2 = 100\%$.

Table 2 shows the complete set of frequent inter-transaction patterns mined from the transaction database in Table 1 with maxspan $= 1$ and minsup $= 2$. The number after each pattern represents the pattern's support. The total number of frequent inter-transaction patterns is 16.

3. The proposed method

3.1. A projection-based approach

In this section, we introduce a new projection-based approach for mining frequent inter-transaction patterns in a transaction database. First, we define the concept of projected databases.

Table 2: The patterns mined from Table 1 with maxspan $= 1$ and minsup $= 2$.

<table>
<thead>
<tr>
<th>Length</th>
<th>Frequent inter-transaction patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$(a^3)^2, (a^3, c^2)^2, (c^2, a)^2$</td>
</tr>
<tr>
<td>2</td>
<td>$(a^2, c^2)^2, (a^2, a^3)^2, (c^2, a)^2, (c^2, d)^2, (c^2, c^1)^2, (c^2, d)^2$</td>
</tr>
<tr>
<td>3</td>
<td>$(a^2, c^2, d^2)^2, (a^2, c^2, c^1)^2, (c^2, a^1, d^2)^2, (c^2, c^1, d^2)^2$</td>
</tr>
<tr>
<td>4</td>
<td>$(c^2, a^1, c^1, d^1)^2$</td>
</tr>
</tbody>
</table>
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات