An Algorithm for Mining Lower Closed Itemsets

Tian-rui Li, Ming Qing, Jun Ma, and Yang Xu

Department of mathematics, School of Science,
Southwest Jiaotong University, Chengdu, 610031, China
{trli, xuyang}@swjtu.edu.cn

Abstract. The generalized association rule base (GARB) presented by Li (2003) can efficiently solve the problem of quantity of rule in the process of acquiring rule by traditional association rule mining algorithms. Therefore, how to deduce all rules contained in the rule of GARB becomes an urgent issue in order to support more effective decision-making. In this paper, the notation of lower closed itemset of an itemset was proposed and some properties of it are proved. Then, it is concluded that the above problem can be solved if all the lower closed itemsets of frequent closed itemset (FCI) are obtained. Finally, an algorithm for mining all lower closed itemsets of an itemset was given and its validity is proved.

Introduction

Association rule mining is one of the most important research areas in data mining, to which many researchers have paid their attention[1-8]. Yet there still exist a big problem in process of acquiring rule by traditional mining algorithms, namely, the over-abundant acquired rules, which cause people hard to understand and eventually make use of this knowledge. Presently there are many methods focus on resolving this problem. Some researchers, for example, proposed the methods for reducing the amount of rules by deleting or combining derived rules such as reducing over-abundant rules by structured rule and summarizing related rule by clustering[2], the method of GSE-pattern to organize, generalize, and represent the mined rule[3]. Other researchers presented the methods of mining association rule based on constraints[4], which can reduce amount of rules by applying some special constraints provided by users to the mining process. Moreover, there are many other methods, such as mining multiple-level association rules[5], mining association rules with multiple minimum supports[6], mining $\phi$-association rule[7] and the like.

Although the aforementioned methods can reduce the amount of rules derived to some extent, the number of mined rules is still too big. Then, in [8], the generalized association rule base (GARB) was presented which can solve this problem efficiently because we can deduce all association rules of the database from it and keep

---

1 This work was supported by the National Natural Science Foundation of China (NSFC) and Basic Science Foundation of Southwest Jiaotong University.

---

© Springer-Verlag Berlin Heidelberg 2004
information in the course of deducing, i.e., the support and confidence of rule in the deductive course do not change. Also there is no redundant rule in the base, which means no rule can be obtained by adding or deleting item of antecedent (or consequent) to another rule in this base on the basis of keeping information. Therefore, how to deduce all rules contained in the rule of GARB becomes an urgent issue that should be solved as soon as possible in order to support more effective decision-making. In this paper, the notation of lower closed itemset of an itemset was proposed firstly, then an algorithm for mining all lower closed itemsets of an itemset was given, which can be used to solve this problem.

Preliminaries

Definition 1 \[1\]. Let \( I = \{i_1, i_2, \ldots, i_m\} \) be a set of \( m \) different items. An association rule is an implication of the form \( X \Rightarrow Y \), where \( \emptyset \neq X \subset I \), \( \emptyset \neq Y \subset I \), and \( X \cap Y = \emptyset \).

Definition 2 \[1\]. Let \( X \) be an itemset. If \( s \) percents of the transactions containing \( X \) in the transaction database, then the support of \( X \), denoted as \( \text{supp}(X) \), is \( s \).

Definition 3 \[8\]. An itemset \( X \) is closed if there exists no itemset \( Y \) such that \( Y \) is a proper superset of \( X \) and every transaction containing \( X \) also contains \( Y \).

Definition 4 \[8\]. A closed rule is an implication expression of the form \( X \Rightarrow Y \), where \( \emptyset \neq X \subset I \), \( Y = \emptyset \), and \( X \) is a closed itemset.

Definition 5 \[8\]. Association rule and closed rule are all called generalized association rule.

Definition 6. \[8\] An upper closed set of an itemset \( X \), denoted as \( X^- \) (called as upper closed itemset), is the smallest closed itemset containing it.

Definition 7. An lower closed set of an itemset \( X \), denoted as \( X_\) (called as lower closed itemset), is the largest closed itemset contained in it, which means that there exists no closed itemset \( Y \), a proper subset of \( X \), containing \( X_\).

Example 1. Suppose a transaction database is shown as Table 1. \{B\} is one of the lower closed itemsets of itemset \{B, C, E\}.

<table>
<thead>
<tr>
<th>Table 1. Transaction database</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>