URN etd-0407109-183722-1858
Author Chung-chieh Su
Author's Email Address No Public.
Statistics This thesis had been viewed 36 times. Download 0 times.
Department
Year 2005
Semester 2
Degree Master
Type of Document Master's Thesis
Language zh-TW.Big5 Chinese
Title Mining Semi-Frequent and Proportionally Fault-Tolerant Association Rule Algorithms and Applications
Date of Defense 0000-00-00
Page Count 198

Keyword
semi-frequent association rule
proportionally fault-tolerant association rule
cluster
mask
index

Abstract
Data mining technologies are becoming more and more mature so that applied on various fields wide. On classical association rules, we can get the relations between items, but decision makers need more information from the rules to make correct decisions. Hence, in this paper, we propose several mining algorithms for many purposes, these purposes include semi-frequent association rule, proportionally fault-tolerant association rule, specified items association rule, multiple-level association rule and semi-frequent multi-time interval sequential pattern.
First, we propose two algorithms MCSF(Using Mask and Cluster to Mine Semi-Frequent Association Rules) and IDSF(Using Index Decomposition to Mine Semi-Frequent Association Rules) for semi-frequent association rule. The main process of MCSF is that each item in transactions is divided two clusters and then the clusters decompose to generate itemsets by mask and joining. The main processes of IDSF are as following: it stores indexes of semi-frequent items in transaction arrays into SFItable(Semi-Frequent Items Index Table). Then, transactions directly decompose to generate itemsets by SFItable and an index method. In order to reduce the number of unnecessary itemsets, both algorithms add the filtration mechanism. Among the algorithms, although there are some differences between MCSF and IDSF, the filtration mechanisms of them are the same. In addition, we divide the large database into many sub-databases which can be loaded in memory for mining the large database. We slightly modify some processes of MCSF and IDSF so that the algorithms can be processed in the large database.
In the proportional fault-tolerant association rule field, we propose an algorithm FT-GDA(Mining Proportional Fault-Tolerant Association Rules by GDA Algorithm). FT-GDA is similar to MCSF. Both use a mask technique to generate itemsets. In mining process, in order to avoid the question which is scanning database repeatedly to count fault-tolerant supports of itemsets, FT-GDA records TID(Transaction ID) and uses the operations of union and intersection to count the fault-tolerant supports. Besides, TID can also uses to filter and check the unnecessary itemsets. FT-GDA uses the same method to mine the large database.
In the specified items association rule field, we propose two algorithms SI_MCSF(Mining Specified Items Association Rules by MCSF Algorithm) and SI_IDSF(Mining Specified Items Association Rules by IDSF Algorithm). SI_MCSF and SI_IDSF are based on MCSF and IDSF algorithms

respectively. The only difference between specified items association rules and semi-frequent association rules is that specified items are users specified, but semi-frequent items are filtered out. The other processes are the same. Both algorithms are also applied to mine the large database.
Further, we propose an algorithm which is based on MCSF to mine multiple-level association rules. The algorithm is named ML_MCSF(Mining Multiple-Level Association Rules by MCSF Algorithm). Because the multiple-level association rule has the relation of level, ML_MCSF sets the number of clusters by the number of levels. Then, it finds out the items of each level of merchandises in transactions and the item is stored into the cluster of its own level. In order to promise that the rules are complete, before generating itemsets, we count all possible combinations of the different length of decomposing. And then each cluster can decompose and join to generate itemsets by the combination.
IDSF algorithm has an advantage which is to decompose transactions to generate sequential patterns and avoid breaking the order of sequences of the rules. In semi-frequent multi-time interval sequential patterns we propose a new algorithm MI_IDSF (Mining Semi-Frequent Multi-Time Interval Sequential Patterns by IDSF Algorithm) algorithm which is based on IDSF to mine the Semi-Frequent Multi-Time Interval Sequential Patterns. The main processes of MI_IDSF are as following: in generating 2-sequence process, MI_IDSF find out all time intervals between items. And then it can directly generate k-sequence by IDSF, where k is greater than 3. At the end of processes, we just join every sequence with its time interval together to generate a complete pattern.
In the experiments, we compare each algorithm in different parameters, fact databases and large databases. The experimental results show that our algorithms have great performance and low memory usage.
Advisory Committee Jen-peng Huang - advisor

Files No Any Full Text File.
Date of Submission 2006-07-24