Abstract

Association Rule mining is a sub-discipline of data mining. Apriori algorithm is one of the most popular association rule mining technique. Apriori technique has a disadvantage that before generating a maximal frequent set it generates all possible proper subsets of maximal set. Therefore it is very slow as it requires many database scans before generating a maximal frequent itemset. In the method proposed in this paper entire database is scanned only once. Frequency count of all distinct transactions is stored in a hash map. Algorithm maintains an array of tables such that each table in the array contain frequency count of all potential k-itemsets. Binary search and the concept of longest common subsequence are used to efficiently extract maximal frequent itemset. Experimental results show that proposed algorithm performs better than apriori algorithm.

References

1. R. Agrawal, T. Imielinski, and A. Swami, “Mining Association Rules between Sets of Items
A Novel Association Rule Algorithm to Discover Maximal Frequent Item Set

in Large Database,” Proceedings of the 1993 ACM SIGMOD
of the 20th International Conference on Very Large Data Bases, 1994, pp. 487-499.
4. J. Han, J. Pei, and Y. Yin, “Mining Frequent Patterns without Candidate Generation,”
Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data,
2000, pp. 1-12.
Proceedings of the tenth international conference on Information and knowledge management
(pp. 474-481). ACM.
7. J. S. Park, M. S. Chen, and P. S. Yu, "Using a Hash-Based Method with Transaction
Trimming for Mining Association Rules," IEEE Trans. on Knowledge and Data Engineering, Vol.
www.ncsa.uiuc.edu/UserInfo/

Index Terms

Computer Science Algorithms

Keywords

Association rule mining ,Apriori algorithm, Frequent itemset, Hashing, Longest common
subsequence