Abstract

One of the most popular data mining approach to find frequent itemset in a given transactional dataset is Association rule mining. The important task of Association rule mining is to mine association rules using minimum support value which is specified by the user or can be generated by system itself. In order to calculate minimum support value, every time the complete database has to be scanned for each item in the transaction. This decreases the time complexity of the algorithm. Here we proposed a new algorithm which scan the database once and create a cache database for each transaction using hash map. This cache copy is then used to search for frequent item sets. Due to which the overhead of scanning complete database for each item is reduced, and efficiency is increased.

References

- K. Vanitha and R. Santhi, "Using Hash Based Apriori Algorithm to Reduce the

- B. Liu, W. Hsu, Y. Ma, Mining association rules with multiple minimum supports, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99), San Diego, CA, USA, 1999.
- M. C. Tseng, W. Y. Lin, &quot;Mining generalized association rules with multiple minimum supports&quot;, International Conference on Data Warehousing and Knowledge Discovery (DaWaK'00), Munich, Germany, 2001, pp. 11 – 20.
- H. Mannila, &quot;Database methods for data mining&quot;, Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD &apos;98) tutorial, New York, NY, USA, 1998.
- W. Lee, S. J. Stolfo, K. W. Mok, &quot;Mining audit data to build intrusion detection models&quot;, Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD &apos;98), New York, NY, USA, 1998.
- J. Han, Y. Fu, &quot;Discovery of multiple-level association rules from large databases&quot;, Proceedings of the 21st Very Large DataBases Conference (VLDB'95), Zurich, Switzerland, 1995, pp. 420–431

Index Terms

Computer Science

Artificial Intelligence
Keywords
Apriori  cache database  hash map  scanning  time  time complexity