Abstract

Mining for association rules between items in large transactional databases is a central problem in the field of knowledge discovery. It has crucial applications in decision support and marketing strategy. Centralized and Distributed Association Rules Mining (DARM) include two phases of frequent itemset extraction and strong rule generation. The most important part of ARM is Frequent Itemsets Mining (FIM) and because of its importance in recent years, there have been many algorithms implemented for it. In this paper, we have focused on distributed Apriori-Like frequent itemsets mining and proposed a distributed algorithm, called New Dynamic Distributed Frequent Itemsets Mining (NDD-FIM), for geographically distributed data sets. NDD-FIM has a merger site to reduce communication overhead and eliminates size of dataset partitions dynamically. The experimental results show that our algorithm generates support counts of candidate itemsets quicker than other DARM algorithms and reduces the size of average transactions, datasets, and message exchanges.
A New Dynamic Distributed Algorithm for Frequent Itemsets Mining

- Agrawal, R., Shafer, J. C. 1996. Parallel mining of association rules, IEEE Transactions on Knowledge and Data Engineering, 8(6) 962-969.
- Cheung, D., Xiao, Y. 1998. Effect of data skewness in parallel mining of association rules, In 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Melbourne, Australia, April, 48-60.
- Cheung, D. W., Ng, V. T., Fu, A. W., Fu, Y. 1996. Efficient mining of association rules in distributed databases, IEEE Transactions on Knowledge and Data Engineering, 8(6) 911-922.

Index Terms

Computer Science
Algorithms
Keywords

Distributed Data Mining  Frequent Itemsets  Association Rule  Apriori Algorithm