Abstract

There are lots of data mining tasks such as association rule, clustering, classification, regression and others. Among these tasks association rule mining is most prominent. One of the most popular approaches to find frequent item set in a given transactional dataset is Association rule mining. Frequent pattern mining is one of the most important tasks for discovering useful meaningful patterns from large collection of data. The FP Growth algorithm is currently one of the fastest approaches to frequent item set mining. This paper proposed an efficient and improved FP Tree algorithm which used a projection method to reduce the database scan and save the execution time. The advantage of PFP Tree is that it takes less memory and time in association mining. Experimental result showed that the improved PFP Tree algorithm performs faster than FP growth Tree algorithm and partition projection algorithm. It is more efficient and scalable in the case of large volume of data. The effectiveness of the method has been justified over a sample our one super market database.
- Li Haoyuan, Yi Wang, Zhang Dong, Zhang Ming, Chang Edward, "PFP: Parallel FP Growth for query Recommendation";
- Jiawei Han, M. Kamber, "Data Mining-Concepts and Techniques"; Sam Francisco 2009, Morgan Kanufmann Publishers.
- Jiawei Han, Jian Pei, Runying Mao, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach"; Data Mining and Knowledge Discovery, April 2001, Kluwer Academic Publishers, Manufactured in The Netherlands.
- J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and Hsu, M. -C. FreeSpan, Frequent pattern-projected sequential pattern mining; ACM SIGKDD, 2010.
- J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation"; SIGMOD 2000, pp 1-12.
- R. Agrawal, C. C. Aggarwal, and V. V. V. Prasad, "A Tree Projection Algorithm For Generation of Frequent Itemsets"; Journal on Parallel and Distributed Computing[(Special Issue on High Performance Data mining)], 2010.

Index Terms

Computer Science Database Management And System
Partition projection.