Abstract

Frequent Patterns are very important in knowledge discovery and data mining process such as mining of association rules, correlations etc. Prefix-tree based approach is one of the contemporary approaches for mining frequent patterns. FP-tree is a compact representation of transaction database that contains frequency information of all relevant Frequent Patterns (FP) in a dataset. Since the introduction of FP-growth algorithm for FP-tree construction, three major algorithms have been proposed, namely AFPIM, CATS tree, and CanTree, that have adopted FP-tree for incremental mining of frequent patterns. All of the three methods perform incremental mining by processing one transaction of the incremental database at a time and updating it to the FP-tree of the initial (original) database. Here in this paper we propose a novel method to take advantage of FP-tree representation of incremental transaction database for incremental mining. We propose “Batch Incremental Tree (BIT)” algorithm to merge two small consecutive duration FP-trees to obtain a FP-tree that is equivalent of FP-tree obtained when the entire database is processed at once from the beginning of the first duration
to the end of the second duration. For large databases, our experimental results show significant reduction in runtime of the BIT algorithm compared to the runtime of sequential incremental algorithms.

Reference

- D W Cheung, J. Han, V.T. Ng, and C.Y. Wong,”Maintenance of discovered association rules in large databases: an incremental updating technique”. In Proc. of ICDE 1996, pp. 106–114.
- J. Han, J. Pei, Y. Yin and R. Mao, “Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach”. Data Mining and Knowledge Discovery, 8(1), 2004, pp.53-87.
- J. Han, J. Pei, and Y. Yin , “Mining Frequent Patterns without Candidate Generation”. In Proc. of SIGMOD 2000,pp.1-12

Index Terms

Computer Science
Databases
Key words

Batch Incremental Mining
Batch Incremental tree
Sequential Incremental Mining
minSup