Load Balancing Approach Parallel Algorithm for Frequent Pattern Mining

游坤明,Jiayi Zhou,Wei Chen Hsiao
Computer Science & Information Engineering
Computer Science and Informatics
yu@chu.edu.tw

Abstract

Association rules mining from transaction-oriented databases is an important issue in data mining. Frequent pattern is crucial for association rules generation, time series analysis, classification, etc. There are two categories of algorithms that had been proposed, candidate set generate-and-test approach (Apriori-like) and Pattern growth approach. Many methods had been proposed to solve the association rules mining problem based on FP-tree instead of Apriori-like, since apriori-like algorithm scans the database many times. However, the computation time is costly when the database size is large with FP-tree data structure. Parallel and distributed computing is a good strategy to solve this circumstance. Some parallel algorithms had been proposed, however, most of them did not consider the load balancing issue. In this paper, we proposed a parallel and distributed mining algorithm based on FP-tree structure, Load Balancing FP-Tree (LFP-tree). The algorithm divides the item set for mining by evaluating the tree's width and depth. Moreover, a simple and trusty calculate formulation for loading degree is proposed. The experimental results show that LFP-tree can reduce the computation time and has less idle time compared with Parallel FP-Tree (PFP-tree). In addition, it has better speed-up ratio than PFP-tree when number of processors grow. The communication time can be reduced by preserving the heavy loading items in their local computing node.

Keyword：FP-tree-data mining-association rules-parallel and distributed computing-load-balancing