

# Balanced Tidset-based Parallel FP-tree Algorithm for the Frequent Pattern Mining on Grid System

Jiayi Zhou, 游坤明

Computer Science & Information Engineering

Computer Science and Informatics

yu@chu.edu.tw

## Abstract

Mining frequent patterns from transaction-oriented database is an important problem. Frequent patterns are essential for generate association rules, time series, etc. Most of frequent patterns mining algorithm can be classified into two categories: generate-and-test approach (Apriori-like) and pattern growth approach (FP-tree). In recent times, many methods have been proposed for solving this problem based on FP-tree, because this approach can reduce the number of database scan. However, even for pattern growth methods, the execution time grows rapidly when the database size is getting large and the given support is small. Therefore, parallel-distributed computing is a good strategy to solve this problem. Some parallel algorithms have been proposed, but the execution time is costly when the database size is large. In this paper, we proposed an efficient parallel and distributed mining algorithm—Balanced Tidset Parallel FP-tree (BTP-tree) algorithm on grid computing system. Grid system is a heterogeneous computing environment, our proposed method can balance the loading according to the tree depth and width. In order to exchange transactions efficiently, transaction identification set (Tidset) was used to directly select transactions instead of scanning database. BTP-tree, TPF-tree and PFP-tree were implemented and the datasets generated by IBM Quest Synthetic Data Generator are used to verify the performance of BTP-tree. The experimental results show that BTP-tree can reduce the execution time significantly and has better loading balance capability than TPF-tree and PFP-tree.

Keyword :