ABSTRACT

Majority of the research activities in general concentrates on effective approaches for itemset extraction, which describes the most computationally thorough knowledge extraction process in association rule mining. During the recent development, disk-based extraction techniques have been implemented to carry out the extraction from huge data sets, but still there is no approach developed to carry out the extraction from data stored in flat files.

Itemset-Mine index (IMine index) provides well compressed and offers complete representation of all the available transactional data and helps in efficient itemset extraction from a relational DBMS. It has been widely accepted that frequent itemset mining plays a considerable role in many significant data mining operations. On the other hand, frequent item set mining often produces an extremely large number of frequent itemsets and rules, which condenses not only the effectiveness but also the value of mining because users have to examine through a huge number of mined rules to discover useful ones.

The proposed methodology mainly focuses on enhancing IMine by improving the frequent itemset extraction techniques by providing better results with reduced extraction time and low reading cost. The methodologies used in the proposed technique are,

1. A Novel Prefix Hash Tree based IMine Indexing
2. An Effective Technique for IMine Indexing using FP-Bonsai Tree
3. Fuzzy Decision Tree based Effective IMine Indexing
Three standard datasets were selected from the UCI machine learning repository datasets for the purpose of evaluation. The datasets used for the evaluation purpose are Page Blocks Classification Dataset, Letter Recognition Dataset and US Census Data (1990). The performance of the proposed itemset mining techniques are evaluated on the basis of reading cost, frequent item set extraction time and memory consumption.