ABSTRACT

The analysis of association rules composes a very important task in the process of data mining. Association rules are an important method of regularities within data which have been exclusively studied by the data mining society. The extensive objective here is to find intermittent concurrence of items within a set of transactions. The found concurrence items are called associations.

Mining frequent patterns from a given transaction dataset is not a negligible task. Based on the user specified minimum support value the set of items that occur frequently has to be identified. An important issue involved here is that the time taken to compute the frequent item sets, because when it involves large databases there might be lot of possible item sets, which need to be evaluated. A different approaches in the algorithm tends to allow efficient discovery of frequent patterns.

But generation of candidate item set was large and more time was required to generate associative rules. To overcome the slowness in associative rule pruning various strategies and approaches were discussed in literature to increase and improve the speed of rule formation. In addition the job of mining all frequent associations in very large datasets is more challenging. The search space is exponential in terms of number of attributes and with thousands of records of dataset.
This research work presented an improved association rule mining algorithm that minimizes the number of candidate item sets during the process of generating association rules resulting in efficient pruning time and optimizing the search space. One distinct advantage of this model is that the work, first analyze the scalability issues of association rule mining in large data sets. Parallel pruning technique is deployed to mine the large transactional items simultaneously at different level of items sets to improve the execution speed for generating frequent items and association rules.

Secondly the proposal presented the Fuzzy based Optimal Search Space Pruning (FOSSP) technique to generate more sensitive item-value pairs. The rule obtained from FOSSP is then used to extract maximum information from the large transactional data sets which is again subjected to parallel item set pruning.

In addition, the improvement is made in the direction of parallel pruning accomplished through thread mechanism for support values of 20% – 40% on car data set and 10% – 40% on bank data set, keeping in mind that the main memory is utilized in the best manner possible, without any thrashing. Simulations are performed to evaluate the efficiency in terms of scalability and providing optimal search space. The effectiveness of optimal support value results in an improvement in a range of 15% - 20% for car dataset. This work finally describes that FOSSP provided optimal search size for larger data sets when compared to traditional fuzzy Apriori.