ABSTRACT

In today’s internet era, a vast amount of knowledge is stored in the web and database. Due to the availability of huge knowledge repositories, getting the relevant information is a challenging task and hence it must be mined and extracted. Association Rule Mining is one approach for extracting useful knowledge from datasets or database which includes frequent patterns and association rules between the items or attributes of a dataset with varying levels of strength.

In this thesis, an intelligent architecture for effective Frequent Itemset Mining and Association Rule Mining has been proposed and implemented in order to provide effective mining for both temporal database and conventional database. The proposed architecture consists of six different components namely User interface, Data Set, Rule Base, Decision Manager, Frequent Itemset Generator component and User Profile Manager model which uses the User Preference Database (UPDB) for fixing the support threshold. The Frequent Itemset Generator provides frequent itemsets using the standard data structure component - Temporal Frequent Pattern Tree algorithm and Hashing with Quadratic Probing for effective data mining. These mining methods are used to find frequent itemsets from temporal database as well as conventional transaction database respectively.
Moreover, the Frequent Itemset Generator provides additional techniques for effective frequent itemset generation namely the basic Bit Vector Matrix representation, Enhanced Cluster based Bit Vector Association Rule Mining and Temporal Pattern Mining algorithm. These algorithms incorporate clustering, bit vector and variable threshold during mining. Moreover, it reduces the number of computations in the post-processing phase and also provides more flexibility in the process of effective decision making in many applications.

The User Preference Database method which is proposed and implemented in this work efficiently tracks the database and gives the support range to the inexperienced user with less comparison. The main advantage of this database is that it does not require statistics about the past transactions. Standard UCI data set is used as the input data for testing the temporal frequent itemset mining algorithms. Also, the bakery transaction data set is used as the input data for validating the frequent itemset generation. The rule base is capable of providing rule matching and rule firing techniques, so that accuracy level is increased and are based on storage and retrieval.

The architecture of the proposed system has a decision manager component and this module has the overall control. This decision making module uses a decision agent which performs many functions including the selection of a suitable frequent itemset generation algorithm from the five algorithms present in the Frequent Itemset Generator using production rules. This module uses a set of intelligent agents namely the data set selection agent, decision making agent and action agent to make a final decision.
The main contribution of this research work is that new temporal Association Rule Mining algorithms have been proposed and implemented to enhance the capabilities of the existing Association Rule Mining algorithms in terms of the number of scans and memory utilisation.