CHAPTER 7

CONCLUSIONS AND FUTURE ENHANCEMENTS

7.1 CONCLUSIONS

In this research work, new temporal mining algorithms have been proposed for providing efficient time sensitive techniques for mining frequent items from data sets. For this purpose, a Temporal FP-Tree algorithm has been proposed in this work to discover frequent patterns during the time intervals specified by calendar schemas. The main advantage of this proposed algorithm is that it discovers all possible temporal association rules with necessary support and confidence from a set of time stamped transactions. These rules can be used to predict the future using past and current data.

A Hash based linear probing algorithm has been proposed in this work to generate frequent itemsets with the primary aim of optimising the memory utilisation. It provides a feature to handle collisions in hashing and thereby improve the performance of the algorithm with respect to time. An improved Apriori algorithm is proposed in this work that uses this hashing technique for effective storage of mined information. Moreover, through pruning candidate itemsets by the infrequent itemsets, the present algorithm reduces the number of database scans and also the redundancy in storage while generating subsets.

This algorithm uses Quadratic Probing for mining frequent itemsets, where the proposed Hash Based Frequent Itemset - Quadratic
Probing algorithm incorporates the vertical data format representation of the database and hence it leads to easy manipulations of the hash data structure. The major advantages are 40% reduction in memory usage and corresponding reduction in the number of scans.

The Three dimensions Item Matrix Vector algorithm proposed in this work incorporates a Three-dimensional Matrix and a Bit Vector matrix. This algorithm scans the preprocessed database once, and does not create candidate itemsets and hence reduces the number of scans and in turn reduces the I/O cost. From the experiments conducted in this research work it has been observed that the proposed Three dimensions Item Matrix Vector algorithm is efficient in terms of processing cost and memory utilization.

The existing Clustering and Graph based Association Rule algorithm uses more space and consumes time. However in many applications, it is necessary to handle volume of data. In such a situation this proposed algorithm is more effective in terms of space as well as time for mining frequent itemsets. Enhanced Cluster based Bit Vector Association Rule is the new algorithm proposed in this research work to enhance the classical Apriori algorithm. The distinguished feature of this algorithm is that it uses a single table for mining k-frequent itemsets unlike the other existing algorithm such as CGAR and CBVAR where multiple tables are used for this purpose and hence the space and time will be optimal.

The existing Hybrid Temporal Pattern Mining uses more space and consumes time. However, in some applications it is necessary to handle large volume of legacy data. In such a situation the proposed Temporal Pattern Mining, provides better performance in terms of time, space and discovers patterns from hybrid event sequences. In this research work, experiments have been carried out on real time data to verify the efficiency of the proposed algorithm. The experimental results obtained from this work show that this
algorithm is more efficient than that of the existing algorithm in terms of effective decision making. This is due to the fact that the proposed Temporal Pattern Mining reduces the number of database scans.

Finally, a user preference database has been proposed in this work for automating the threshold setting with respect to support. This enables the system to provide optimal results without user interaction. Moreover, most users are not familiar in providing support threshold and hence a number of attempts are necessary to get better results with the existing algorithm. On the other hand, the proposed user preference database not only helps to find an optimal value for threshold but also provides a uniform value unlike user interactions. The major applications of this work include effective analysis and prediction on diabetic’s diseases for early treatment, advice on tea and coffee exports and recommendation of bakery items based on association rules.

7.2 FUTURE ENHANCEMENTS

Future work in this direction can be the provision of a facility for distributed data mining with intelligent agent based communication. Moreover, the generated association rules may be stored permanently in a knowledge base and can be used to perform the temporal reasoning task namely analysis of the past, prediction of the future, planning and learning effectively. This work has focused mainly on single level association rule mining technique for temporal databases. However, this work can be further extended to provide multi-level association rule mining techniques.