Chapter 8

Conclusion and Future work

The basic aim of incorporating this chapter into present thesis is to highlight the outcome of the present research work. These outcomes are represented in section 8.1. Further during this research journey; author feels that there are certain areas which remains untouched which can be pursued as the future scope of the work as discussed in section 8.2.

8.1 Conclusion

The knowledge discovery from rapidly growing data can be handled by data mining techniques. There are various data mining techniques like association rules, classification, clustering, correlations, sequential patterns and many more. In this research work the author concentrates on association rules mining algorithm for discovery of frequent patterns.

To carry on present work theoretical concepts of data mining are discussed in detail which includes data mining life cycle, models and functionalities. Potential application areas are also verified and conclude that data mining algorithms are highly applicable in immeasurable applications like market basket analysis, item placement, attached mailing, fraud detection, medical research and many more.

Available literature belonging to data mining demonstrates that mining single level and multiple-level knowledge both are practical and desirable. This research work is expanding the scope of the study of mining association rules from single level to multiple concept levels. Since single level association rules are providing the general information, which is not helpful to get concise and specific knowledge. Existing algorithms that accomplish the task of mining association rules at multiple levels of abstraction with Apriori based methods are not efficient to achieve time efficiency. The repetitive numbers of database scan are required to generate the candidate sets. Accomplished methods that employ FP-growth approach are inefficient in the way that they use a large amount of computer memory to store conditional trees for frequent itemsets mining. However most kind of methods produce redundant multiple levels association rules in association rule discovery process.

The research work has been implemented using widespread high level and object oriented language-Java. Seeing the vast applicability and benefits of association rules, the proposed work is tested on well-known datasets. The dataset used in this study has been taken from UCI Repository of Machine Learning databases available on line. In this study basically four datasets of various sizes and with different number of attributes are used. These real world
datasets are Breast-cancer, Credit-g, Mushroom and Soybean. The two types of files have been provided to the algorithm as input first is XML file, which is required to provide the concept hierarchy of data. Second is the actual ARFF (Attribute-Relation File Format) data file. Detail study has been carried out on each and every concept of multiple level association rule mining algorithm covered in this research work.

Comparative analysis of several algorithms of single level and multiple levels has been conducted for mining frequent patterns. The heated discussion about shortcoming of evolutionary algorithms leads to some improvements. To mine multiple level association rules, a hierarchy information encoded transaction table is required instead of the original transaction table. The concept hierarchy encoding reduces the size of the encoded transaction table by removing or merging the identical items. Thus, it is often essential to use an encoded table.

The traditional algorithms for mining association rules may apply the Apriori algorithm to examine data items at multiple levels of abstraction under the various minimum support and minimum confidence thresholds. This direction is simple, but it may lead to some undesirable results. Apriori uses repeated scans of the database to match and tally the patterns of candidates. It can be potentially time consuming for large databases. MLT2_L1 uses Apriori to generate frequent itemsets at each concept levels. That makes MLT2_L1 costly in terms of execution time. On the other hand LWFT algorithm makes use of counting inference approach to generate frequent itemsets at multiple levels of concept hierarchy. But in worst case, this algorithm behaves like Apriori.

The author has proposed an efficient algorithm for mining frequent patterns at single and multiple levels. The single level mining algorithms has been named as TransTrie and the multiple level mining algorithm has been named as MLTranTrie. These algorithms are used to discover regularly occurring patterns by utilizing transposed dataset. The transposed dataset has been stored in bit format. This bit format database has been used to handle the large databases effectively. Proceeding further, the algorithm utilizes the advantage of an advance data structure i.e. Trie, by the help of which different items of the dataset has been stored along with their support value.

The results of the proposed algorithm MLTransTrie have been discussed. On the basis of given support and confidence, the MLTransTrie algorithm discovers all the possible multiple level association rules from a set of transactions, by reducing the number of scans and execution time. On the basis of results, it has been found that MLTransTrie algorithm works well to generate the frequent patterns and association rules. It has been noticed that
the execution time of the algorithm depends on the size and complexity of concept hierarchy. Due to this it varies for different datasets. Number of association rules discovered depends on value of parameters at each level like support, delta and confidence. The experimental results have shown that proposed algorithm perform tremendous as compare to existing algorithms for their support threshold.

This work is a contribution in the field of discovery of knowledge at multiple-levels in the form of association rules that enhances the ease and comprehensibility of the users.

### 8.2 Future Work

During present research work; the author feels that there are few areas where upcoming researchers can conduct their research work to contribute to society, academics and research environment. Such areas are given as below:

- The upcoming researchers can explore such methods/algorithms which can produced single level and multiple level association rules without candidate set generation approach so that it may consumes less time and memory.
- They can carry out their research work on big data environment; where association rules mining may be useful for frequent patterns discovery.
- In the modern mobile era the upcoming researchers can investigate such rules set theory which can answer mobile users query promptly.
- Future researchers can design such algorithms which may be implemented through a mobile device to ascertain knowledge discovery process.
- Researchers may design such algorithm which can work efficiently on existing data structure for efficient utilization of memory.
- More commonly, researchers can use the various data mining technique on all most all the real life applications to produce appropriate knowledge by using various parameters like support, confidence, lift, leverage etc efficiently.