CHAPTER 7

CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION

The traditional association rules are Intra-transactional since they only capture associations among items within the same transactions, where the notion of the transaction could be the items bought by the same customer, the atmospheric events that take place at the same time, and so on. However, an Inter-transactional association rule can represent not only the associations of items within transactions, but also the associations of items among different transactions along certain dimensions.

Existing algorithms for mining association rules developed so far are based on Breadth-First Search (BFS) approaches. In the BFS, the frequent itemsets are generated level by level. The database is scanned at each level to determine the support value for each generated candidate itemsets. This will increase execution time and memory consumption. But the existing Apriori-like approaches perform better on smaller datasets compared to larger datasets. The performance of such algorithms is also good at larger support values. The meanings of the generated association rules are not well understandable in most cases.
The main focus of this research work is to improve the prediction rate of the Inter-transaction association rules, reduce the execution time for finding the association rules, reduce the search space and generate stronger, useful, abstract and meaningful association rules.

To obtain the first objective of improving the prediction rate of association rules, a new approach called Modified Apriori Algorithm (MAA) is proposed in this work. The MAA approach enhances the traditional association rule framework with Inter-transactional associations from Intra-transactional associations to ensure its predictive capability in the tasks such as weather forecasting and stock market analysis. The MAA is used to discover frequent Inter-transaction itemsets. To enhance the algorithm’s efficiency, a hashing technique is used to reduce the number of candidate 2-itemsets generation. The MAA is applicable to both single-station and multi-station meteorological data sets.

The test results of meteorological data show that the numbers of predicted Inter-transactional association rules are more than that of the Intra-transaction association rules with certain support and confidence thresholds. The traditional Apriori algorithm predicts only 2 Intra-transactional association rules while 842 Inter-transaction association rules are predicted by the MAA approach with support value of 50% and confidence value of 95% along with max_span value of 11. This MAA approach also finds out some interesting association relationships which are not to be found in existing methodologies. This approach also concentrates on the execution time for generating such a large number of Inter-transaction association rules. Compared to existing Apriori algorithms, the MAA approach requires lesser time.
The MAA approach generates a large number of Inter-transaction association rules which require a large amount of memory space for storing and processing the intermediate candidate itemsets and Inter-transaction association rules. So, the next module CBARM (Crumb Based Association Rule Mining) is proposed to address this problem.

Several algorithms have been proposed to solve the problem of generating large number of association rules. At present, the FITI algorithm is the state-of-art in the Inter-transaction association rule mining, but it generates many unneeded combinations of items because the set of generated extended items is much larger than the set of items. To solve this problem, an alternative approach CBARM is proposed, where a crumb is a group of identical transactions that meet a certain condition.

The CBARM approach is implemented to reduce the generation of frequent patterns. This reduces number of frequent patterns dramatically and improves the efficiency of the system in terms of execution time. There are two sets of industries ie Bank vs Insurance and Food beverage vs Tobacco retailing are taken into account for comparing the results of the proposed and existing FITI approaches. The proposed CBARM approach only generates one third of association rules compared to existing FITI. This approach yields less execution time, but it will be possible to reduce the execution time further by the mining only on closed itemsets.

To obtain further reduction in execution time and memory space consumption, the concept of closed itemsets is incorporated into Inter-transaction association rule mining in the proposed approach called IAR (Inter-transaction Association Rule) miner. The IAR miner addresses the problem of generating large number of itemset results in Inter-transaction association rule mining. There are two steps in this approach, in the first step, it scans the database once and finds the frequent items. In the second step, it
enumerates Inter-transaction itemsets using an Itemset-Dataset tree, called ID-tree.

The IAR Miner applies effective pruning strategies and avoids rich candidate set generation and repeated counting of support values. The experiments are conducted with two different datasets and the performance of the proposed IAR miner algorithm is compared to existing algorithms. The basic experiments are conducted in terms of execution time and memory consumption of the algorithm with different minimum support values and standard max_span values. The execution time of the proposed IAR miner is nearly half of that of the ITP miner and far less than those of other existing algorithms. The performance of the proposed algorithm is also better in occupying less memory space compared to the existing works.

The performance of the proposed IAR algorithm is investigated with number of transactions, average length of transactions and number of items as a scalability measure. The IAR miner needs only half of the time required for the existing ITP miner for mining the association rules and require much lesser execution time though number of transactions gets increased when compared to other existing algorithms. The same kinds of results are obtained when the IAR miner is implemented with average length of transactions and number of items as a measurement. The performance study on the synthetic datasets shows that the IAR Miner algorithm is more efficient than the EH-Apriori, FITI, ClosedPROWL and ITP-Miner algorithms in terms of memory consumption and execution time. The limitations of this algorithm are that the generated rules are not abstract, meaningless in few cases and more detailed.

To address the aforementioned problems, Categorized and Bounded Inter-Transaction (CBIT) algorithm is finally proposed. The CBIT algorithm well understands the user situation and observes the comprehension of every
domain. The abstraction and meaningfulness of the generated association rules are obtained by the introduction of taxonomy with generalization. There are four significant processes in the CBIT algorithm namely Determination of top-level taxonomy, Bounding of Inter-transactional domains, Defining the length of interval and Depictions of an itemset.

Transaction databases will be clearly placed in their position, the entries are studied, and they are named according to transformations and then stored. Then the algorithm derives the association rules from the datasets by examining the number of observations of same transaction in association with the pair components. The experiments are conducted to study the performance of the proposed CBIT algorithm in terms of execution time and memory consumption. The proposed CBIT requires lesser execution time because it can derive generalized structure for the given database by using taxonomy with generalization. The memory space requirement for generation of Inter-transaction association rules is less as a boundary is set for number of scans and thus restricts enormous wastage of memory and other computational resources.

In this research work, An Adept Strategy for Mining of Association Rules from Inter-Transactional Domains, has been designed with the help of the four modules namely Modified Apriori Algorithm (MAA), Crumb Based Association Rule Mining (CBARM), Inter-transaction Association Rule miner (IAR) and Categorized and Bounded Inter-Transaction (CBIT).

The problem of lower prediction rate in the association rules mining from multidimensional domains, is addressed by the MAA module. The scope of mining from Intra-transaction rules from Inter-transaction rules can be achieved through this module. A higher prediction rate of 99.7% and only a moderate rate of improvement in the execution time can be obtained through this module.
In order to improve the execution time, the CBARM module is proposed. The crumbs are included to mine the Inter-transaction association rules and only a minimum number of 42 association rules are generated as compared to the existing FITI’s 240 rules and this will improve the efficiency of the system in terms of execution time.

In order to get further reduction in execution time with less memory consumption, the concept of closed itemset mining is incorporated in the third proposed module IAR miner. The problem of generating large number of itemsets result in Inter-transaction rule mining is amicably solved in this module with the 14% reduction in execution time as compared to the comparatively better existing scheme ITP-Miner due to the reduction in number of generated association rules.

Finally, to generate meaningful and abstract association rules, the CBIT module is proposed by combining the taxonomy concept with closed itemset which is accomplished by the efficient declarations of ancestors which would enhance the futuristic computations of support and confidence values of every transaction at later stages of mining process along with generalization. The execution time is improved by 30% over the existing schemes with 50% reduction in memory consumption.

7.2 FUTURE SCOPE

Though this research work has been successful in addressing the problem of achieving higher prediction rate, less execution time and lower memory consumption with the help of proposed modules, this research work has not taken the following aspects into account:

- Setting up multiple minimum support values for the items.
- Generation of only user interested association rules.
• Setting up of different minimum support values of various items present in the database.

In future, further research work can be proceeded in the aforementioned aspects.