Chapter 6

Conclusion and Future Direction

6.1 Inferences

A compact bitmap representation of a set supported by the corresponding implementation of the different set-theoretic operations was used in the different stages of the rule mining. The Apriori algorithm was supported with the new representation. The experimental results were provided relevantly. The proposed representation outperforms the array representation of itemsets significantly when used in the Apriori algorithm.

The graph-based itemset mining algorithm is presented which gets the explicit benefit of the proposed itemset representation. For this algorithm also, the conventional representation of itemset remained on the loser side both in terms of time and memory.

In the mining of rules from the already mined frequent itemsets, the proposed representation of itemsets has brought little but observable improvements in time and space efficiency for the existing rule mining algorithms.

Hence from the different experimental results, it may be observed that the proposed representation may be found as an attribute to the association rule mining process.

6.2 Future Work

The data are increasing every day continuously so even the data in the database are also intended to change. As the dynamics of the dataset keep on changing, the

mining process also becomes more complex as compared to mining the traditional static dataset. Such complex issues can arise due to the generation of a candidate itemset, and the scanning of the original dataset for the process of validating the frequency of an item. The incremental method aims to update the original dataset to reflect the updated dataset without the need to re-model or scan the dataset from scratch.

According to the proposed itemset representation, the length of the itemset size is fixed according to the maximum size of the itemset present in the dataset. For the incremental dataset, the proposed method will not work. The length of the itemset cannot be changed if new itemsets are added to the dataset whose length is greater than the length of the maximum size of the itemset. For the future scope, we can work on the incremental dataset so that the proposed itemset can still be used for the newly added itemsets in the dataset.

In the incremental dataset, the new itemset is added only when the itemset is frequent in the original dataset. The support count of the itemset is updated in the updated dataset. When the itemset is infrequent in the incremental dataset but frequent in the original dataset, then the itemset may or may not be added only. The decision is taken after computing the support count of the itemset from the original and the incremental dataset. When the itemset is frequent in the incremental dataset, then the itemset and support count are added to the updated dataset. If the itemset is infrequent in the original dataset but frequent in the incremental database. The support count of the item is updated in the updated dataset. When the itemset is infrequent in both the original and incremental dataset, then the itemset will not be added to the updated dataset.