6.1 Conclusions

In this dissertation a study of different issues of association rule mining is presented. After a careful study of the different association rule mining algorithms it was found that all of them treat the problem as a single objective one, where finally the confidence of the rules are maximized. All of those algorithms are based on the approach pioneered by Agrawal et. al. [AIS93]. Generation of the rules are done in two phases namely frequent itemset generation and rule generation. First phase being the most crucial one, the existing algorithms have given more importance on it. A very few algorithms were found that attended the second phase. In the chapter 2, an efficient algorithm to attend the rule generation phase was presented. From the various experiments it has been found that the new algorithm works faster than all the other existing algorithms.

Though the association rule mining was handled as single objective prob-
lem by the existing algorithms, from the study of the association rules it was found that it should be treated as a multi-objective problem. Considering comprehensibility and interestingness as two other objectives along with confidence, association rule mining problem can be handled as multi-objective problem. After testing several expressions finally, two expressions to quantify comprehensibility and interestingness were formulated. In Chapter 3, two algorithms based on Pareto genetic algorithm are presented to extract the association rules from the database. In literature, Pareto genetic algorithm was found to be a better technique to handle the multi-objective problem. Out of the two algorithms presented in the Chapter 3, second one has been found to be more efficient.

During the study of the association rule mining algorithms, it was observed that the performance of the algorithms are affected by the dimensionality of the database. But almost all databases contain some irrelevant attributes (dimensions) in it. Several works have been found in the literature to attend this issue of data mining, commonly known as dimensionality reduction. Need of a dimensionality reduction technique suitable for the association rule mining technique has lead to the work presented in Chapter 4. Here two efficient algorithms for dimensionality reduction, having their own strengths are presented. Depending on the decision maker’s need any of the algorithms can be used to reduce the size of the database in terms of attributes.

Mining association rules over an incremental database is yet another challenging problem of data mining. Several works based on the approach due to Agrawal et.al., can be found in the literature. Incremental rule mining being an allied area of association rule mining, a need of an algorithm to extract
the rules considering the multiple objectives was felt. This need has lead us to the development of the multi-objective incremental rule mining algorithm presented in the Chapter 5. From the various experiments it has been found that the proposed algorithm is efficient to extract the association rules using multiple objectives.

From the study and various experiments carried out during this dissertation work it is found that the association rule mining problem should be handled as a multi-objective problem rather than single objective one. To attend different issues related to the multi-objective association rule mining some efficient techniques are developed and presented here.

6.2 Future works

In this dissertation the association rule mining is treated as a multi-objective problem. And several techniques to attend different issues of association rule mining are presented here. Still there are some works left unattended and some works may need more tuning.

- Comprehensibility of the association rule is used as a measure of the rules. But it is difficult to quantify a subjective measure like comprehensibility. After considering a number of expressions, it was found that the expression used in this dissertation gives a better representation of the comprehensibility of association rules. It is not ensured that this is the best expression to quantify it. So this expression needs some more attention.
• Similarly expression used for the other subjective measure, i.e. interestingness also needs some attention.

• Most of all, the algorithms presented here are tested over some continuous(numeric) valued databases only. But, by nature, association rule mining is not restricted to continuous(numeric) valued databases. So the testing of the algorithms over categorical(nominal) valued database as well as mixed valued database is left to be done.