Chapter 7

Concluding Remarks

Association rule mining discovers associations amongst itemsets of a given transactional dataset. The set of association rules generated using traditional methods is often too large to be understood and interpreted by a human. Closed itemsets have been studied in literature for discovery of non-redundant rules. The use of concept lattice as a data structure facilitates generation of non-redundant rules. However, concept lattice is a memory intensive data structure and the algorithms based on it are computationally expensive. Hence it cannot be used efficiently in large and incremental datasets.

Support and Confidence are the two commonly used interestingness measures for pruning the set of non-redundant association rules. However, these measures may not be appropriate for the application in hand. Other rule evaluation measures like recall, certainty factor, variance, and information gain have been proposed in literature which may be useful in different applications of association rule mining. In general association rule mining may be viewed as a multiobjective problem rather than a biobjective problem. The choice of appropriate interestingness measures for pruning the set of non-redundant rules remains a major issue.

In this thesis, we have addressed the problems of efficient mining of closed itemsets in both static and incremental datasets, efficient mining of non-redundant association rules, computation of multiple interestingness measures and discovery of accurate associative classifiers. We summarize the research contributions of this
thesis and suggest directions for future work.

7.1 Summary of the Thesis

In summary,

1. we have proposed an efficient algorithm, *mineCIL*, for generating closed itemsets in static datasets. As the algorithm reduces the repeated computation of closed itemsets, it is faster than the competing algorithms.

2. we have proposed a novel and compact data structure based on lattice called *CILattice*. Procedures for incremental update of *CILattice* have also been proposed. Using *CILattice* as a synopsis, algorithms for discovery of all closed itemsets in sliding window and all recent closed itemsets in damped window of data streams have been proposed.

3. an efficient algorithm, *MARM_DS*, for discovery of non-redundant association rules from *CILattice* has been proposed. We have shown that *CILattice* can be used for computation of multiple interestingness measures like recall, correlation coefficient, accuracy, specificity, certainty factor.

4. we have proposed an algorithm, *ACCI*, for building associative classifiers using closed itemsets. The proposed algorithm generates a set of class association rules based on closed itemsets (ClosedCARs) using *CILattice* and subsequently removes the rule conflicts. A subset of non-conflicting ClosedCARs which covers the entire training set is chosen as a classifier. Further, multiobjective genetic algorithm NSGA-II has been used to improve the comprehensibility of the generated associative classifier.

5. all algorithms are backed by extensive experimentation on real life and synthetic datasets.

**Future Work** In case of high speed data streams, an attempt to give exact results i.e. discovering all closed itemsets, may lead to loss of data because per-
transaction-processing time increases due to enormous size of the synopsis. Although capturing approximate results reduces the problem of enormous size of the synopsis, it introduces error in the reported results. One can explore the possibility of discovering approximate closed itemsets in data streams with bounded error. One can also examine the usability of the proposed algorithms in real-time applications where multiple interestingness measures can be used for discovery of multiobjective association rules.