Abstract

Association Rule Mining (ARM) is an important data mining technique that discovers associations amongst itemsets in transactional datasets. Often the generated set of rules is too large to be understood by a human. Discovery of a smaller set of closed itemsets leading to a compact set of non-redundant rules is an important area of research. Concept lattice has been used as a data structure for storing closed itemsets and their corresponding transactions. It has also been proved that non-redundant association rules can be generated efficiently using concept lattice. However, high memory and computational requirements of concept lattice based algorithms prohibit their use in large and incremental datasets. Although the set of non-redundant association rules is relatively small as compared to the set of rules generated using frequent itemsets, still it is too large for useful interpretation. Various interestingness measures are employed to prune the rule set. Indeed, the choice of appropriate interestingness measures for pruning the set of non-redundant rules is an important issue in association rule mining.

In this thesis, we address three problems. Firstly we consider the problem of designing an efficient algorithm for mining of closed itemsets in static and incremental datasets. Next, we target efficient mining of non-redundant association rules. Finally, we design accurate and easy to comprehend associative classifiers.

In order to solve the above mentioned problems, we propose a novel and compact data structure based on lattice called CILattice. The proposed data structure enables efficient discovery and storage of closed itemsets, and also conducive to incremental updates. Thus we use CILattice as synopsis in data streams for discovery of all closed itemsets in sliding and damped window models. An algorithm has been proposed for discovery of non-redundant association rules using CILattice that traverses the lattice only once. We also make use of CILattice for computing multiple interestingness measures for pruning the association rules.

For building an associative classifier, we introduce the notion of closedCARs - Class association rules based on closed itemsets. An algorithm is developed to
build accurate associative classifiers using closedCARs. As comprehensibility of an
associative classifier depends on the number of rules and the number of conditions
in each rule, these criteria are used in a multiobjective framework to improve the
comprehensibility of the classifiers.

Extensive experiments on real-life and synthetic datasets demonstrate the ef-
fectiveness of the algorithms proposed in the thesis. The proposed algorithms have
also been compared to the competitive algorithms.