Abstract

The problem of mining association rules has attracted a lot of attention in the research community. Several techniques for discovery of association rules have been discussed in literature. These algorithms, motivated by Agrawal's approach, handles the rule mining problem as a single objective problem. However, the rule mining based on single objective have some common drawbacks, i.e. rules may be difficult to understand due to the involvement of a lot of conditions, some discovered rules may not be interesting to the user as they were not actually hidden. To overcome these limitations, this dissertation presents an approach for rule mining using multiple objectives so that some interesting and useful rules can be extracted. For this - three different objectives namely predictive accuracy, comprehensibility and interestingness are considered. As a result, the proposed approach is found to be better than the classical approach as some previously unknown, potentially useful and ultimately understandable rules can be discovered.

Moreover, the single objective based classical rule mining approach can be found to be impractical as sometimes they may produce a huge number of rules, that in turn makes the job of decision makers tougher one in deciding which rule to use which to discard. However the proposed approach will give only a few valid rules, which will help the decision maker as the choice is limited.

It was found that Pareto Based Genetic Algorithm acts as an efficient tool in handling the multi-objective problems. Since association rule mining is treated as a multi objective problem, the proposed approach uses Pareto based multi objective genetic algorithm to extract the rules.
The dimensionality of the databases plays an important role in the data mining tasks including association rule mining. Appropriate feature selection or dimensionality reduction techniques can save the cost of computation during the association rule mining over high dimensional space to a grate extent. Hence, this dissertation presents two effective techniques as a preprocessing task to reduce the dimensionality of the databases before applying any data mining techniques.

Rule mining being a time consuming job, it is not appreciated to extract the rules again and again from the whole database, every time the database is updated. So the task of extracting the rules from an incremental database received a lot of research attention. However, my survey reveals that most of the existing works are basically based on the classical approach. To meet this challenge, this dissertation also presents a technique of incremental mining based on multi objective approach.

*Keywords — Data mining, Association rule mining, Multi objective rule mining, Genetic algorithms for rule mining, Pareto based rule mining, GA based incremental mining*