Abstract

Association Rule Mining Using Ant Colony Optimization

Data mining is one of the extensively researched areas for the last 25 years. Mining of useful knowledge leverages a number of benefits to decision makers for making policy decisions and for designing marketing strategy. The task of data mining is composed of several techniques such as association rule mining, classification rule mining, pattern mining, sequence discovery, clustering and generalization. Data mining has shown its influence in several different fields. Applications of data mining include analysis of transaction data, analysis of medical data, web usage mining and analysis of crime data for detecting frauds. Association rule mining is the technique of data mining that deals with finding associations among the different pieces of data. It aims at finding comprehensive and useful correlations between set of items in large database repositories. Literature reports several algorithms for finding association rules from databases ranging from the brute-force approaches that perform several passes over the database; to evolutionary approaches that performs one or two scans of database to discover the association rules. These approaches suffer from a number of limitations such as multiple scans of database, producing unmanageable number of rules, requiring large amount of storage space for intermediate data in producing desired results. An exhaustive survey of the literature for the past couple of decades has proved that still, the problem of finding association rules from large databases has not solved satisfactorily.

To resolve the existing problems, the research work in this thesis is dedicated to designing novel approaches for finding association rules that are based on the ACO meta-
heuristic. Ant Colony Optimization is a sub-field of swarm intelligence that has been shaped by a number of people who have made valuable contributions to the development and success of the field. ACO is inspired by the foraging behavior of ant colonies. The model is based on the feedback mechanism through which the information about the beneficial path is communicated between the agents. The ACO framework has proved its efficiency in several areas where some form of optimization needs to be performed. The research work in this thesis presents three novel approaches, two of which are based on the ACO model for mining association rules. All the three approaches are graph based. Representing the problem in the form of a graph is one of the preliminary requirements for the application of ACO meta-heuristic. To prove the practical significance of the approaches, the approaches have been implemented in Microsoft Visual Studio 2010 using C# and tested on real datasets taken from UCI repository. The efficiency in terms of reduced database scans and comprehensiveness of generated rules is evidenced from the results obtained from the implementation. All the approaches have been evaluated for two major criteria namely time consumed and the number of itemsets/rules produced by the approach. The results of the proposed approaches were compared with existing approaches.