A New Approach to Associative Classification Based on Binary Multi-Objective Particle Swarm Optimization

Madhabananda Das  
KIIT University, India  

Rahul Roy  
KIIT University, India  

Satchidananda Dehuri  
Fakir Mohan University, India  

Sung-Bae Cho  
Yonsei University, Korea

ABSTRACT

Associative classification rule mining (ACRM) methods operate by association rule mining (ARM) to obtain classification rules from a previously classified data. In ACRM, classifiers are designed through two phases: rule extraction and rule selection. In this paper, the ACRM problem is treated as a multi-objective problem rather than a single objective one. As the problem is a discrete combinatorial optimization problem, it was necessary to develop a binary multi-objective particle swarm optimization (BMOPSO) to optimize the measure like coverage and confidence of association rule mining (ARM) to extract classification rules in rule extraction phase. In rule selection phase, a small number of rules are targeted from the extracted rules by BMOPSO to design an accurate and compact classifier which can maximize the accuracy of the rule sets and minimize their complexity simultaneously. Experiments are conducted on some of the University of California, Irvine (UCI) repository datasets. The comparative result of the proposed method with other standard classifiers confirms that the new proposed approach can be a suitable method for classification.

DOI: 10.4018/978-1-4666-2145-9.ch014
A New Approach to Associative Classification Based on Binary Multi-Objective Particle Swarm Optimization

INTRODUCTION

Association and classification rule mining are two most promising data mining techniques for designing a classifier to solve the classification task in various domains of medical science (Giannopoulou, 2008, Huap et al., 2009), marketing management (Groth, 1997), telecommunication (Sasisekharan, Seshadri, & Weiss, 1996), disaster management (Shao & Fu, 2008), intrusion detection (Pei et al., 2004), life science (Wong & Li, 2006), insurance, biometric (Gutierrez et al., 2002), etc. The objective of association rule mining, in its basic form, extracts a set of the individual descriptive rules from the database with minimum support and confidence and the objective of classification rule mining is to design a compact and efficient predictive rule sets from the extracted rules. The integration of association rule mining with the classification rule mining has helped in overcoming some of the basic problems of classification rule mining which are illustrated below:

1. The integrated framework helps in better understandability of the classification problem. The standard classification system generates compact rules by using dominant biases and heuristics which sometimes result in generation of rules that may not be in agreement with the user, leaving behind many understandable rules undiscovered. With this framework, all the understandable rules are discovered using association rule mining and the compact classifier is designed from these rules thereby increasing the better understandability of the rules used by the classifier.

2. Another issue of rule discovery problem is the generation of interesting and useful rules. In quest of generating a compact classifier, the classification rule mining algorithm may leave behind many interesting and useful rules.

3. Classification rule requires the entire database to be loaded to the main memory for rule discovery which is not the case with associative classification rule mining.

This integrated framework for designing a classifier is referred to as associative classification rule mining or simply associative classification (Liu et al., 1998) (de la Iglesia, Reynolds, & Rayward-Smith, 2005; de la Iglesia, Richards, Philpott, & Rayward-Smith, 2006). Associative classification rule mining usually consists of two phases: rule extraction and rule selection. In the rule extraction phase, a large number of classification rules are extracted from a data set using an association rule mining technique that satisfies the pre-specified threshold values of the minimum support and confidence. In rule selection phase, a part of extracted rules are selected or targeted to design an accurate and compact classifier. The accuracy of the designed classifier usually depends on the specification of the minimum support and confidence. In the context of classification rule mining, their tuning has been discussed in the literatures (Coenen, Leng, & Zhang, 2005; Bayardo & Agrawal, 1999).

The initial framework for associative classification rule mining was proposed by Liu et al. (1998). Thereafter, many researchers have used the two phase associative rule mining for designing classifiers. Most of the two phase associative algorithms differ in the rule selection phase. All the statistical based associative classification rule mining concentrate on pruning of the rules using various heuristic algorithms in rule selection phase. These heuristic approaches prune the rules in order to make classifier compact without taking into account the interestingness of the rule, thereby compromising with the classification accuracy. However, the actual problem lied with the rule generation phase, which uses the exhaustive techniques like Apriori algorithms for generation of rules. For this reason, huge set of rules are
Related Content

Online Clustering and Outlier Detection
[www.igi-global.com/chapter/online-clustering-outlier-detection/69896?camid=4v1a](www.igi-global.com/chapter/online-clustering-outlier-detection/69896?camid=4v1a)

A Two-Phase Scheduling Method Combined to the Tabu Search for the DARP
[www.igi-global.com/article/a-two-phase-scheduling-method-combined-to-the-tabu-search-for-the-darp/114203?camid=4v1a](www.igi-global.com/article/a-two-phase-scheduling-method-combined-to-the-tabu-search-for-the-darp/114203?camid=4v1a)

DIMMA: A Design and Implementation Methodology for Metaheuristic Algorithms - A Perspective from Software Development
[www.igi-global.com/article/dimma-design-implementation-methodology-metaheuristic/51678?camid=4v1a](www.igi-global.com/article/dimma-design-implementation-methodology-metaheuristic/51678?camid=4v1a)

New Genetic Operator (Jump Crossover) for the Traveling Salesman Problem