## ACO BASED DISCOVERY OF COMPREHENSIBLE AND ACCURATE RULES FROM MEDICAL DATASETS

ABDUL RAUF BAIG, WASEEM SHAHZAD, SALABAT KHAN AND FARIHA ALTAF

Department of Computer Science National University of Computer and Emerging Sciences 15-Mauve Area, Rohtas Road, Sector H-11/4, Islamabad, Pakistan rauf.baig@nu.edu.pk

Received June 2010; revised October 2010

ABSTRACT. In many real world applications, comprehensibility of the classifier is as important as its accuracy. The medical field is one of those where this requirement is more pronounced. It is not enough for users in this field to have an accurate classifier, and they also need to verify and analyze the logic of the classification process. It is difficult to have confidence in a black box type of classifier when the classification decision is a matter of life and death of a patient. In recent years, algorithms for classification rule discovery based on the ant colony optimization meta-heuristic (ACO) have been proposed, which fulfill both the requirements of high accuracy and comprehensibility. This paper reports some improvements in a recently proposed ACO based classification algorithm, called CAntMiner, whose main feature is a heuristic function based on the compatibility of pairs of attribute-values and class labels, and its application on medical datasets. We study the performance of the algorithm for twelve commonly used datasets and compare it with ten well known classification algorithms, three of which are ACO based. Experimental results show that the accuracy rate obtained by CAntMiner is better than that of the compared algorithms. We also discuss some other issues related to comprehensibility of the classifier building process.

**Keywords:** Ant colony optimization (ACO), Classification rules, Medical data mining, Swarm intelligence

1. Introduction. Vast quantities of medical data are added to clinical databases on a daily basis. One of the major challenges in this context is to automatically extract accurate and comprehensible knowledge from these large collections of raw data. Discovery of relationships and patterns within these data has the potential to provide new medical knowledge. The discovered knowledge can then be used to improve the quality of service. For example, it can be used by medical practitioners to avoid drugs with adverse effects, suggest less expensive but therapeutically equivalent alternatives, avoid redundant tests and replace expensive tests by cheaper tests. If the extracted knowledge is comprehensible, domain experts can further refine and enhance the discovered knowledge.

Some methods and algorithms already exist for the analysis of medical data. For example, in recent years, neural networks have been extensively applied to many problems [1] including medical related tasks [2-4]. Most applications of neural networks in medicine are related to classification tasks [2]. However, neural networks are incomprehensible and opaque to humans. In most of the medical applications, both comprehensibility and accuracy are required.

A well known example of a comprehensible classifier is the decision tree which builds a rule set [5,6]. Comprehensible classifiers are also learnt with the help of evolutionary algorithms (e.g., [7]). In recent years, a few Ant Colony Optimization (ACO) [8-10] based