Ant Miner: A Hybrid Pittsburgh Style Classification Rule Mining Algorithm

Bijaya Kumar Nanda, BIET, Odisha, India
Satchidananda Dehuri, Fakir Mohan University, Balasore, India

https://orcid.org/0000-0003-1435-4531

ABSTRACT

In data mining the task of extracting classification rules from large data is an important task and is gaining considerable attention. This article presents a novel ant miner for classification rule mining. The ant miner is inspired by researches on the behaviour of real ant colonies, simulated annealing, and some data mining concepts as well as principles. This paper presents a Pittsburgh style approach for single objective classification rule mining. The algorithm is tested on a few benchmark datasets drawn from UCI repository. The experimental outcomes confirm that ant miner-HPB (Hybrid Pittsburgh Style Classification) is significantly better than ant-miner-PB (Pittsburgh Style Classification).

KEYWORDS

Ant Colony Optimization, Ant Miner, Ant Miner-HPB, Ant Miner-PB, Classification Rule Mining, Data Mining, Pheromone, Simulated Annealing

1. INTRODUCTION

Learning classification rules from instances have been a subject of intense study in data mining (Han & Kamber, 2006; Dehuri, Ghosh & Ghosh, 2008; Panda, Dehuri, & Patra, 2015). Based on various theories and techniques, many different algorithms have been proposed to generate classification rules (Dehuri & Mall, 2006; Dehuri, et al., 2008, & Kalia, et al., 2018). However, there are three common and important factors for classification rule learning: higher predictive accuracy, smaller rule sets, and shorter running time (Freitas, 2002;). More precisely, higher predictive accuracy means more efficient application; smaller rule sets enable better understanding for the user, and shorter running time means that the algorithm can be applied to online systems and address the scalability issue (Kalia, et al., 2018).

Ant miner is an application for extracting classification rules from data by simulating the behaviours of real ant colonies (Parpinelli, Lopes, & Freitas, 2001). Aiming at the insufficiencies of ant-miner, researchers have proposed some improvement strategies resulting new versions.
of ant-miner including new heuristic function formula, a new pheromone updating method, and a state transition rule, which can get higher accuracy rate (Otero, Freitas & Johnson, 2008; Liu, Abbass, & McKay, 2004; Parpinelli, Lopes, & Freitas, 2002). A comparative analysis and survey of ant colony optimization based rule miner can be obtained in (Ali & Shahzad, 2017). On the other hand, in rule-based classifier, many objectives are involved and regain simultaneous optimization, for example, minimization of the number of rules or length of rule, maximization of classification accuracy of the rule, maximization of interestingness, etc (Dehuri, Ghosh, & Ghosh, 2008). In lieu of these criterions, classification rule mining problem is not restricted as single objective problem. It is also an attractive field of research. For many objective optimization researches, although we realize this classification rule mining is a multi-objective optimization problem but by giving higher priority to the objective of classification accuracy, a novel algorithm has been developed by us. We explore the search space by giving proper attention to both good and bad solution with certain probabilities (this concept is derived from simulated annealing (Aarts & Korst, 1989; Laarhoven & Aarts, 1987). However, overall rule discovery process is motivated by ant colony optimization method (Mahapatra & Patnaik, 2018; Angus & Woodward, 2009). The combined approach of both stochastic approaches drives us to uncover hidden pattern very effectively.

2. PRELIMINARIES

In this section, we discuss two basic algorithmic paradigms like ant colony optimization and simulated annealing in following Subsections.

3. ANT COLONY OPTIMIZATION

Since the early 1990’s, several collective behavior (like social insects and bird flocking) inspired algorithms have been proposed and applied in many optimization problems with inherent intractability (Abraham, Grosan, & Ramos, 2006; Bonabeau, Dorigo, & Theraulaz, 1999; Tao, 2018). Ant Colony Optimization (ACO) (Blum, 2005; Dorigo & Stutzle, 2006; Dorigo, Maniezzo, & Colomi, 1996; Bonabeau, Dorigo & Theraulaz, 1999) is one of the most popular algorithms of them and was introduced around 1990. Ants are social insects, living in colonies and exhibit an effective collective behavior. Although each ant is relatively a simple insect with limited individual abilities, a swarm of ants has the ability to find the shortest path from their nest to food. Further, it was discovered that most of the communication amongst individual ants is based on the use of a chemical, called pheromone that is dropped on the ground. As ants walk from a food source to the nest, pheromone is deposited on the ground, creating in this way a pheromone trail on the path used. Shorter paths will be traversed faster and by consequence, will have stronger pheromone concentration than longer paths over a given period of time. The more pheromone path contains, the more attractive it becomes to be followed by other ants. Hence, as time goes by, more and more ants will prefer the shorter path, which will have more and more pheromone. At the end, almost all ants will be following a single path which usually will represent the shortest path between the food source and the nest.

An ant probabilistically chooses a path to follow based on heuristic information and the amount of pheromone deposited by previous ants. The inter-active process of building candidate solutions and updating pheromone values allows an ACO algorithm to converge to optimal or near optimal solutions (Liu, Zhang, & Yu, 2019; Al-Behadili, 2018). Algorithm 1 presents a high-level pseudo code of a basic ACO procedure, comprising four main steps: Initialization (), Construct-Ant-Solution (), Local-Search (), and Update-Pheromone ().
Asynchronous P Systems
Tudor Blănescu, Radu Nicolescu and Huiling Wu (2014). Natural Computing for Simulation and Knowledge Discovery (pp. 90-108).

www.igi-global.com/chapter/asynchronous-p-systems/80058?camid=4v1a