Chapter 4

Discrete Artificial Bee Colony Optimization Algorithm for Financial Classification Problems

Yannis Marinakis  
Technical University of Crete, Greece

Magdalene Marinaki  
Technical University of Crete, Greece

Nikolaos Matsatsinis  
Technical University of Crete, Greece

Constantin Zopounidis  
Technical University of Crete, Greece

ABSTRACT

Nature-inspired methods are used in various fields for solving a number of problems. This study uses a nature-inspired method, artificial bee colony optimization that is based on the foraging behaviour of bees, for a financial classification problem. Financial decisions are often based on classification models, which are used to assign a set of observations into predefined groups. One important step toward the development of accurate financial classification models involves the selection of the appropriate independent variables (features) that are relevant to the problem. The proposed method uses a discrete version of the artificial bee colony algorithm for the feature selection step while nearest neighbour based classifiers are used for the classification step. The performance of the method is tested using various benchmark datasets from UCI Machine Learning Repository and in a financial classification task involving credit risk assessment. Its results are compared with the results of other nature-inspired methods.

INTRODUCTION

Artificial Bee Colony (ABC) optimization algorithm is a population-based swarm intelligence algorithm that was originally proposed by Karaboga and Basturk (2007, 2008) and it simulates the foraging behaviour that a swarm of bees perform. In this algorithm, there are three groups of bees, the employed bees (bees that determine the food source (possible solutions) from a prespecified set of food sources and share this information (waggle dance) with the other bees in the hive),

DOI: 10.4018/978-1-4666-2145-9.ch004
the onlookers bees (bees that based on the information that they take from the employed bees they search for a better food source in the neighborhood of the memorized food sources) and the scout bees (employed bees that their food source has been abandoned and they search for a new food source randomly). The initially proposed Artificial Bee Colony optimization algorithm is applied in continuous optimization problems. In our study, as the feature selection problem is a discrete problem, we made some modifications in the initially proposed algorithm in order the algorithm to be suitable for solving this kind of problems.

The development of financial classification models is a complicated process, involving careful data collection and pre-processing, model development, validation and implementation. Focusing on model development, several methods have been used, including statistical methods, artificial intelligence techniques and operations research methodologies. In all cases, the quality of the data is a fundamental point. This is mainly related to the adequacy of the sample data in terms of the number of observation and the relevance of the decision attributes (i.e., independent variables) used in the analysis.

The latter is related to the feature selection problem. Feature selection refers to the identification of the appropriate attributes (features) that should be introduced in the analysis in order to maximize the expected performance of the resulting model. This has significant implications for issues such as (Kira & Rendell, 1992) the noise reduction through the elimination of noisy features, the reduction of the time and cost required to implement an appropriate model, the simplification of the resulting models and the facilitation of the easy use and updating of the models.

The basic feature selection problem is an optimization problem, with a performance measure for each subset of features, which represents expected classification performance of the resulting model. The problem is to search through the space of feature subsets in order to identify the optimal or near-optimal one with respect to the performance measure. Unfortunately, finding the optimal feature subset has been proved to be NP-hard (Kira & Rendell, 1992). Many algorithms are, thus, proposed to find the suboptimal solutions in comparably smaller amount of time (Jain & Zongker, 1997). Branch and bound approaches (Narendra & Fukunaga, 1977), sequential forward/backward search (Aha & Bankert, 1996; Cantu-Paz, Newsam & Kamath, 2004) and filters approaches (Cantu-Paz, 2004) deterministically search for the suboptimal solutions. One of the most important of the filter approaches is the Kira and Rendell’s Relief algorithm (Kira & Rendell, 1992). Stochastic algorithms, including simulated annealing (Siedlecki & Sklansky, 1988; Lin, Lee, Chen & Tseng, 2008), scatter search (Chen, Lin & Chou, 2010; Lopez, Torres, Batista, Perez & Moreno-Vega, 2006), ant colony optimization (Al-Ani, 2005a, 2005b; Parpinelli, Lopes & Freitas, 2002; Shelokar, Jayaraman & Kulkarni, 2004), GRASP (Yusta, 2009), tabu search (Yusta, 2009), particle swarm optimization (Lin & Chen, 2009; Lin, Ying, Chen & Lee, 2008; Pedrycz, Park & Pizzi, 2009) and genetic algorithms (Cantu-Paz, Newsam & Kamath, 2004; Rokach, 2008; Yusta, 2009) are of great interest recently because they often yield high accuracy and are much faster.

In this paper, a nature inspired intelligent technique, that uses the Discrete Artificial Bee Colony optimization algorithm (DABC) is presented and analyzed in detail for the solution of the feature selection problem. A hybridized version of the Discrete Artificial Bee Colony algorithm has been presented in (Marinakis, Marinaki & Matsatsinis, 2009a). The algorithm is combined with three nearest neighbour based classifiers, the 1-nearest neighbour, the k-nearest neighbour and the weighted k-nearest neighbour classifier. The algorithm is applied to a data set involving financial decision-making problems which involves credit
Related Content

A Generalized Graph Strict Strong Coloring Algorithm
[www.igi-global.com/article/generalized-graph-strict-strong-coloring/64642?camid=4v1a](www.igi-global.com/article/generalized-graph-strict-strong-coloring/64642?camid=4v1a)

Melanocytic Lesions Screening through Particle Swarm Optimization
[www.igi-global.com/chapter/melanocytic-lesions-screening-through-particle-swarm-optimization/82698?camid=4v1a](www.igi-global.com/chapter/melanocytic-lesions-screening-through-particle-swarm-optimization/82698?camid=4v1a)

An Adaptive Second Order Neural Network with Genetic-Algorithm-based Training (ASONN-GA) to Forecast the Closing Prices of the Stock Market

Systematic Redaction for Neuroimage Data
[www.igi-global.com/article/systematic-redaction-neuroimage-data/72876?camid=4v1a](www.igi-global.com/article/systematic-redaction-neuroimage-data/72876?camid=4v1a)