Information Sciences 422 (2018) 462–479

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Pareto front feature selection based on artificial bee colony optimization

Emrah Hancer[a,b], Bing Xue[c,*], Mengjie Zhang[c], Dervis Karaboga[a], Bahriye Akay[a]

[a] Department of Computer Technology and Information Systems, Mehmet Akif Ersoy University, Burdur 15030, Turkey
[b] Department of Computer Engineering, Erciyes University, Kayseri 38039, Turkey
[c] School of Engineering and Computer Science, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand

ARTICLE INFO

Article history:
Received 11 October 2016
Revised 6 September 2017
Accepted 9 September 2017
Available online 12 September 2017

Keywords:
Feature selection
Classification
Multi-objective optimization
Artificial bee colony

ABSTRACT

Feature selection has two major conflicting aims, i.e., to maximize the classification performance and to minimize the number of selected features to overcome the curse of dimensionality. To balance their trade-off, feature selection can be handled as a multi-objective problem. In this paper, a feature selection approach is proposed based on a new multi-objective artificial bee colony algorithm integrated with non-dominated sorting procedure and genetic operators. Two different implementations of the proposed approach are developed: ABC with binary representation and ABC with continuous representation. Their performance are examined on 12 benchmark datasets and the results are compared with those of linear forward selection, greedy stepwise backward selection, two single objective ABC algorithms and three well-known multi-objective evolutionary computation algorithms. The results show that the proposed approach with the binary representation outperformed the other methods in terms of both the dimensionality reduction and the classification accuracy.

1. Introduction

Data mining is in the intersection of artificial intelligence, machine learning, statistics and database systems. It is basically the process of extracting valuable knowledge embedded in data and then transforming the knowledge into an understandable format for users through the steps, such as data pre-processing, management, post-processing and visualization [14]. Data mining and machine learning techniques can be mainly divided into unsupervised (e.g., clustering), supervised (e.g., classification) and reinforcement learning [14]. This paper focuses mainly on classification, which aims to learn a model based on a training set of instances and predict the class labels of unseen instances in the test set. Classification has been used in various real-world applications such as medical healthcare, image analysis, marketing and statistical problems [26,43]. However, the datasets, especially large dimensional ones, may comprise redundant, irrelevant and relevant features. This brings the problems of high complexity and poor learning performance in real-world applications [43].

One of the most common ways to overcome these problems is to apply feature selection [37]. Feature selection aims to select the most relevant/useful features which contribute to the constructed model more efficiently and effectively. Not only for the classification performance, it is also beneficial for simplifying the learned models and shortening the training

* Corresponding author at: School of Engineering and Computer Science, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand.
E-mail address: bing.xue@ecs.vuw.ac.nz (B. Xue).

http://dx.doi.org/10.1016/j.ins.2017.09.028
0020-0255/

time. However, finding relevant/useful features is not an easy task due to the huge search space and the complex interactions among features. Feature interaction may occur in two ways, three ways or more than three ways. An individually irrelevant feature may be beneficial for the classification/learning performance while being interacted with other features. An individually relevant feature may become redundant when it is interconnected with other features. Furthermore, there exist 2^n possible feature subsets for a n-dimensional dataset. It is impractical to intimately search all possible solutions for a large n. Accordingly, feature selection is an NP-hard combinatorial problem [37]. Even though a number of search techniques such as sequential forward and backward feature selection (SFS, SBS) [26] have been proposed, they may have premature convergence problems or intensive computational complexity. To alleviate these problems, evolutionary computation (EC) techniques which are population based solvers in the subclass of global optimization and artificial intelligence have been applied due to their global search potential. The mostly commonly applied techniques for feature selection are genetic programming (GP) [36], genetic algorithms (GAs) [32] and particle swarm optimization (PSO) [27,35,37]. EC techniques are particularly good at multi-objective optimization because their population based search mechanism can produce multiple trade-off solutions in a single run.

It can be inferred from the two main conflicting objectives of feature selection, i.e., the maximization of the classification accuracy and the minimization of the feature subset size, that feature selection can be treated as a multi-objective problem. Unfortunately, there exist just a few studies concerning multi-objective feature selection in the literature [43], i.e., most of the existing approaches are based on a single objective of maximizing the classification accuracy. One of the recent metaheuristics, artificial bee colony (ABC) [19] is an EC technique with many successful applications to solve different problems, which is a motivation to design ABC for multi-objective feature selection. Furthermore, ABC is easy implement, robust against initialization, and has the ability to explore local solutions with the low risk of local convergence. Our recent study [16] has shown that ABC can be used for multi-objective feature selection, but the method in [16] is for *filter* feature selection and the number of features in the datasets is small. The potential of ABC for multi-objective *wrapper* feature selection, which requires often a different approach from *filters* [21], and with a large number of features, has not been investigated yet.

1.1. Goals

The main goal of this paper is to improve an ABC-based feature selection approach to searching for a set of Pareto optimal solutions yielding a smaller feature subset size and a lower classification error percentage than the case that all features are used. To fulfill this goal, a new multi-objective ABC approach based on non-dominated sorting and genetically inspired search is proposed, and two different implementations of the proposed approach are developed: Bin-MOABC (binary version) and Num-MOABC (continuous version). Bin-MOABC and Num-MOABC are compared with two traditional approaches, two single objective ABC variants and three well-known multi-objective feature selection approaches on 12 benchmark datasets including a variety of features, classes and instances.

Specifically, the following objectives are investigated:

1. the performance of single objective ABC approaches on reducing the feature subset size and increasing the classification performance,
2. the performance of the proposed multi-objective ABC implementations on obtaining Pareto optimal solutions and comparisons with two traditional and two single objective ABC approaches,
3. the performance analysis of the proposed multi-objective ABC implementations versus existing multi-objective approaches, and
4. the effect of considering feature selection in binary domain (Bin-MOABC) and continuous domain (Num-MOABC) on the classification performance.

1.2. Organization

The organization of the rest of the paper is as follows. A general knowledge concerning the standard ABC algorithm and the recent studies on feature selection is provided in Section 2. The proposed feature selection approaches are explained in Section 3. The experimental design is described in Section 4 and the experimental results are presented with discussions in Section 5. Finally, the conclusions and the future trends are introduced in Section 6.

2. Background

In this section, ABC is described, the definition of multi-objective optimization is given, and then the recent research of the feature selection is briefly reviewed.

2.1. Artificial bee colony

ABC is a swarm intelligence algorithm that simulates the foraging behavior of a honey bee colony [19]. In the hive, three types of bees are assigned to the foraging task: employed bees, onlooker bees and scout bees. Employed bees are responsible