Expert Systems With Applications 62 (2016) 91–103

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A hybrid approach of differential evolution and artificial bee colony for feature selection

Ezgi Zorarpacı, Selma Ayşe Özel*

Çukurova University, Faculty of Engineering and Architecture, Department of Computer Engineering, 01330, Balcalı, Sarıçam, Adana, Turkey

ARTICLE INFO

Article history:
Received 16 March 2015
Revised 13 May 2016
Accepted 2 June 2016
Available online 8 June 2016

Keywords:
Feature selection
Hybrid optimization
Artificial bee colony
Differential evolution
Classification

ABSTRACT

"Dimensionality" is one of the major problems which affect the quality of learning process in most of the machine learning and data mining tasks. Having high dimensional datasets for training a classification model may lead to have "overfitting" of the learned model to the training data. Overfitting reduces generalization of the model, therefore causes poor classification accuracy for the new test instances. Another disadvantage of dimensionality of dataset is to have high CPU time requirement for learning and testing the model. Applying feature selection to the dataset before the learning process is essential to improve the performance of the classification task. In this study, a new hybrid method which combines artificial bee colony optimization technique with differential evolution algorithm is proposed for feature selection of classification tasks. The developed hybrid method is evaluated by using fifteen datasets from the UCI Repository which are commonly used in classification problems. To make a complete evaluation, the proposed hybrid feature selection method is compared with the artificial bee colony optimization, and differential evolution based feature selection methods, as well as with the three most popular feature selection techniques that are information gain, chi-square, and correlation feature selection. In addition to these, the performance of the proposed method is also compared with the studies in the literature which uses the same datasets. The experimental results of this study show that our developed hybrid method is able to select good features for classification tasks to improve run-time performance and accuracy of the classifier. The proposed hybrid method may also be applied to other search and optimization problems as its performance for feature selection is better than pure artificial bee colony optimization, and differential evolution.

1. Introduction

Feature selection, also known as dimensionality reduction, is the method of selecting an optimum subset of relevant features which represents original feature set with the least error for learning a classification model. Thanks to feature selection techniques, we have some benefits such as improved model interpretability, shorter training times, and enhanced generalization by reducing overfitting when constructing classification models (He, Zhang, Sun, & Dong, 2009). As a result of these benefits, many feature selection methods have been proposed in the literature. These methods are traditionally categorized as *wrapper* and *filter* techniques according to how the method is used. When a classifier is used to evaluate the generated feature subsets, the feature selection method is called as a *wrapper* approach. When feature subsets are evaluated according to their information content or statistical measures, the feature selector is called as a *filter* method (Palasinamy & Kanmani, 2012). Filter methods are usually faster than wrapper methods since they have lower computational cost. However, wrapper methods have usually better performance than the filter methods since they select more representative features from the original feature set (Grande, Del Rosario Suárez, & Villar, 2007).

Theoretically, a feature selection method must search through the subsets of features, and find the best one among all possible subsets according to a certain evaluation criterion. If n features exist in the feature set, optimal feature selection process needs to evaluate 2^n feature subsets to specify the best subset. However, this operation cannot be possible since it is too costly and restrictive practically. Instead of the best one, a (sub)optimum feature subset not reducing or least reducing classification outcome may be accepted. Heuristic and random search methods can be applied to find these (sub)optimum subsets. Various metaheuristic search methods including tabu search (TS), simulated annealing (SA), genetic algorithm (GA), particle swarm optimization (PSO), ant colony optimization (ACO), differential evolution (DE), and artificial bee

* Corresponding author.
E-mail addresses: ezorarpaci@cu.edu.tr (E. Zorarpacı), saozel@cu.edu.tr, saozel@gmail.com (S.A. Özel).

http://dx.doi.org/10.1016/j.eswa.2016.06.004
0957-4174/

colony (ABC) are used to search feature subset space for selecting (sub)optimal feature set (Frohlich, Chapelle, & Scholkopf, 2003). These heuristic models employ different strategies to keep the balance between exploration and exploitation for searching. While exploration provides to discover distinct areas in the search space, exploitation allows scanning the local search space for better solutions. In some of the above mentioned metaheuristic search techniques the exploration process; while in some others the exploitation process perform better. For this reason, hybrid methods can be used to increase the performance of the search algorithm. In hybridization, good properties of at least two techniques are combined to enhance the performance of each technique. In this study, our aim is to use ABC and DE, which are recently proposed and successful metaheuristic techniques, to build a new hybrid wrapper method to improve the performance of general classification tasks. In literature, different forms of hybridization of heuristic algorithms were developed for feature selection problem, but only a few of them includes hybrid of ABC and DE algorithms.

DE (Storn & Price, 1997) is a metaheuristic algorithm, introduced by Rainer Storn and Kenneth Price, finds an (sub)optimal solution for a problem by iteratively trying to improve a candidate solution according to a given measure of quality. DE is used for multi-dimensional real-valued optimization problems. Also, DE has several advantages compared to other optimization techniques such as it requires settings for less number of parameters; it is fast, robust, and applicable to high-dimensional complex optimization problems. Even though DE is efficient, it has some drawbacks such as unstable convergence, and easy to hang out with local optimum (Wu, Lee, & Chien, 2011).

ABC (Karaboga, 2005), proposed by Dervis Karaboga, is an optimization algorithm that mimics the intelligent foraging behavior of honey bee swarm. ABC has good properties such as it is easy to implement; and it has strong robustness, high flexibility, and fewer control parameters (Bolaji, Khader, Al-Betar, & Awadallah, 2013). Also, ABC shows good performance at exploitation with onlooker bee processing phase. However, this process may increase time required for convergence and the algorithm cannot exhibit its own real performance (Gao & Liu, 2011). To obtain better results than standard ABC and DE algorithms, powerful properties of these two algorithms can be combined.

In this paper our aim is to develop a hybrid method from ABC and DE algorithms to provide a solution to the feature selection problem of general classification tasks in data mining. In the literature, different hybrid methods which involve both ABC and DE have been proposed to solve optimization problems other than feature selection problem. In this paper, our hybrid method that is based on ABC and DE for feature selection problem contains a new binary neighborhood search mechanism and a new modified onlooker bee process for the ABC algorithm; as well as it has a new binary mutation phase for the DE algorithm. We use DE since it is employed to solve various search problems without performing too much local search to overcome high computational time. To improve the generated solution for feature selection problem without increasing computational time, we use the neighborhood search phase of the ABC algorithm with a new binary neighborhood operator. Therefore we merge the fast convergence nature of the DE algorithm with the new binary mutation phase. Besides, a new modified onlooker bee process is proposed to prevent the drawbacks of trapping to the local optimum of the new mutation phase of the DE algorithm. The proposed methods are not only used for feature selection problem, but also applicable to other binary optimization problems. The performance of the proposed hybrid algorithm on the feature selection problem is a good indicator which shows that it can be used to handle other binary optimization problems.

To measure the performance of our new method, J48, Naive-Bayes, and RBF Networks classifiers are applied and fifteen datasets from the UCI machine learning repository are used. The proposed method is compared with well-known feature selection algorithms such as chi-square (CHI), information gain (IG) and correlation feature selection (CFS), and some recent feature selection methods that involve metaheuristics including harmony search and stochastic search algorithms for feature selection (Nekkaa & Boughaci, 2015), a binary ABC with DE operators (Hancer, Xue, Karaboga, & Zhang, 2015), a novel system for feature selection based on gravitational search (Xiang et al., 2015), an unsupervised feature selection method based on improved version of DE (Bhadra & Bandyopadhyay, 2015), the relevance-redundancy filter method based on ACO for feature selection (Tabakhi & Moradi, 2015), as well as standard ABC and DE algorithms. It is important to note that ABC is often used for feature selection problem; however there are only a few studies which involve hybridization of it with DE algorithm as emphasized in the previous works section.

This paper is organized as follows: Initially, previous works related to feature selection are described in Section 2. The proposed hybrid method of ABC and DE for feature selection is then presented in detail in Section 3. Experimental results obtained through the proposed hybrid method, standard ABC and DE algorithms, and the comparison of the proposed hybrid method with CHI, IG, CFS, and other metaheuristic based feature selection systems are given and discussed in Section 4. At the end of the paper, final remarks and directions for future works are presented.

2. Related works

A prominent problem of the classification tasks is the high dimensionality of the feature space (Shang et al., 2007). In this case we should incorporate only relevant features and discard less relevant or irrelevant of them. To cope with this selection process, several search strategies were developed. Referring to literature, search methods based on rankings with feature scoring measures such as IG, CHI etc. are used for selecting attributes. The major disadvantage of these filter methods is that they neglect the mutual effect with the classifier model, and this may lead to decreased classification performance when compared to other types of feature selection methods. While filter methods handle the issue of finding a good feature subset without depending on the classifier model, wrapper methods accommodate the classifier model into their search strategy. In this manner, it is possible that a search procedure evaluates a specific subset of features by using classification model and can obtain better classification performance. Since the size of the feature subset space is proportional to 2^n, where n is the number of features in the feature set, the computational time for search strategy which finds the best feature subset among all 2^n feature subsets is too much in wrapper methods. Therefore meta-heuristic search methods, which discover a (sub)optimum feature subset, can be used with a classifier model to construct a wrapper method (Saeys, Inza, & Larrañaga, 2007). In the literature several metaheuristic search methods including ABC and DE were proposed for feature selection problem.

2.1. Artificial bee colony and differential evolution based feature selection methods

First feature selection algorithm based on Differential Evolution (DE) is proposed by Khushaba, Al-Ani, and Al-Jumaily (2008) who use a real number optimizer and applies DE operators to the indices of the features which cause the same features to be encountered multiple times in the solution vector. To prevent having the same feature multiple times, feature distribution factors are used by applying a roulette wheel weighting scheme. The proposed algorithm is compared with GA and PSO by using the number of