Pattern Recognition 60 (2016) 121–129

Contents lists available at ScienceDirect

Pattern Recognition

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

Feature selection using Forest Optimization Algorithm

Manizheh Ghaemi [a,*], Mohammad-Reza Feizi-Derakhshi [b]

[a] Faculty of Electrical and Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran
[b] Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

ARTICLE INFO

Article history:
Received 17 August 2015
Received in revised form
26 March 2016
Accepted 11 May 2016
Available online 24 May 2016

Keywords:
Feature selection
Forest Optimization Algorithm (FOA)
KNN classifier
Dimension reduction
FSFOA

ABSTRACT

Feature selection as a combinatorial optimization problem is an important preprocessing step in data mining; which improves the performance of the learning algorithms with the help of removing the irrelevant and redundant features. As evolutionary algorithms are reported to be suitable for optimization tasks, so Forest Optimization Algorithm (FOA) – which is initially proposed for continuous search problems – is adapted to be used for feature selection as a discrete search space problem. As the result, Feature Selection using Forest Optimization Algorithm (FSFOA) is proposed in this article in order to select the more informative features from the datasets. The proposed FSFOA is validated on several real world datasets and it is compared with some other methods including HGAFS, PSO and SVM-FuzCoc. The results of the experiments show that, FSFOA can improve the classification accuracy of classifiers in some selected datasets. Also, we have compared the dimensionality reduction of the proposed FSFOA with other available methods.

1. Introduction

One of the inevitable steps in knowledge discovery is data mining and the knowledge obtained as the result of data mining is used in many trends; like business and medical use [6,15,20,37]. These days, there has been an increase in the number of collected and stored features in databases but not all the features are useful for data mining, so that some of the features are completely irrelevant or redundant [8,10,36,23]. These features not only have no use in the process of knowledge discovery, but also they increase the complexity and incomprehensibility of the results. So, feature selection helps to reduce the dimensionality of the datasets before data mining. In large databases with many features to handle, when there are n features, time complexity to evaluate all the subsets of features is exponential $(O(2^n))$ [31], which is practically impossible. So, feature selection methods are the bases for data mining to keep useful features for latter learning tasks alongside the ignoring of the most irrelevant and less important ones [11]. In fact feature selection techniques ignore the irrelevant features so, learning process can be done more efficiently. It is also proved that feature selection increases the classification accuracy of machine learning algorithms like KNN classifier [11].

Feature selection is the special case of feature weighting problem [34]. Many studies have shown the beneficial effect of feature weighting [1,9,27,30,32,33]. In feature weighting problem, features are assigned a value which shows their importance in the machine learning process but, in feature selection problem a feature is either retained or deleted and the weights are limited to just '0' and '1'. In fact, feature selection algorithms are a proper subset of feature weighting algorithms which use binary weights (i.e., 0 or 1).

It has been proved that feature selection has an impact on the accuracy and complexity of the classifiers [11]. The mostly used criterions for evaluating the selected feature subset is classification accuracy (CA) on new instances (test dataset). In fact, we expect that dimensionality reduction with the help of feature selection will increase classification accuracy or at least it remains the same.

The objective of this paper is to select the useful features of the datasets with the help of FOA as a new evolutionary algorithm. As FOA is reported to be suitable in continuous search spaces, in this article we have attempted to investigate the performance of FOA in feature selection (FS) as a discrete search problem and we have introduced a method named as Feature Selection using Forest Optimization Algorithm (FSFOA). In fact, FSFOA searches for the best feature subset with the objective of improving the classification accuracy of some classifiers as learning algorithms including KNN, C4.5 and SVM classifiers. The contribution of this paper is twofold: adapting FOA for solving discrete problems and also solving feature selection problem with the help of discrete FOA which leads to the proposed FSFOA method.

The rest of this paper is organized as follows. In Section 2, an overview of feature selection methods is presented. An overview of Forest Optimization Algorithm (FOA) is given in Section 3. In

* Corresponding author.
E-mail addresses: m_ghaemi@ee.kntu.ac.ir (M. Ghaemi), mfeizi@tabrizu.ac.ir (M.-R. Feizi-Derakhshi).

http://dx.doi.org/10.1016/j.patcog.2016.05.012
0031-3203/

Section 4, the application of FOA for feature selection (FSFOA) is presented and Section 5 is devoted to the experiments and results on the proposed FSFOA. Finally, Section 6 summarizes the main conclusions.

2. An overview of feature selection methods

Many researchers have addressed feature selection (FS) problem up to now and also more attempt is needed to further speed up the process of selecting informative and useful features in databases for data mining.

The earliest methods in FS literature based on the machine learning algorithms are filters [11,12]. In all the filters, heuristic techniques based on the general characteristics of data such as information gain and distance is used instead of learning algorithms. Another approach in feature selection is wrapper methods [11,19]. In contrary to filters, wrappers use learning algorithms to investigate the worthy of the selected features [41]. Generally, wrappers produce better results than filters; because while using wrapper approach, the relationship between the learning algorithm and the training data is considered. The well-known drawback of wrappers is that they are slower than filters; because the learning algorithm must be repeatedly executed for every selected feature subset. Sometimes a hybrid of filter-wrapper methods is used. Hybrid methods integrate feature selection within the learning algorithm in order to exploit the advantages of both wrappers and filters [11]. Ignoring the filter or wrapper approach for feature selection methods, they can match any of the following groups: complete search, heuristic search and meta-heuristic methods.

Almuallim and Dietterich presented FOCUS method which completely searches the search space up to reaching to the smallest set of features that divides the training data into pure classes [2,3]. But with n features to handle, there are $(2^n)-1$ possible subsets of features so, evaluating all of the subsets is practically impossible in datasets with many features. As the result, complete search methods are seldom used for feature selection in large datasets with many features.

Heuristic methods of feature selection problem include greedy hill climbing algorithm [25,26], branch and bound method, beam search and best first algorithm. Greedy hill climbing algorithm evaluates all local changes in order to select the relevant features [11,25]. SFS (Sequential Forward Selection) and SBS (Sequential Backward Selection) are two kinds of hill climbing methods. SFS starts with an empty set of selected features and each step of the algorithm adds one of the informative features to the selected set; but, SBS starts will the full set of features and in each step, one of the redundant or irrelevant features is omitted. Bi-directional search is another method which considers both adding and deleting the features simultaneously [11]. The main drawback of both SFS and SBS algorithms is the "nesting effect" problem; which means that while a change is considered positive (either addition or deletion of a feature), there is no chance of re-evaluating that feature. Later in order to overcome the "nesting effect" of SFS and SBS algorithms, SFFS (Sequential Forward Floating Selection) and SBFS (Sequential Forward Floating Selection) were introduced [24]. Best first search is another method which like hill climbing considers local changes in the search space but, it allows backtracking in the search space unlike hill climbing methods [11].

Heuristic algorithms perform better than complete search methods while comparing time complexities, but recently meta-heuristic algorithms like Genetic Algorithm (GA), Particle Swarm Intelligence Optimization (PSO) and Ant Colony Optimization (ACO) show more desirable results. The main advantage of the meta-heuristic methods is their acceptable time complexity. Due to the random nature of meta-heuristic search methods, the application of genetic algorithms, particle swarm optimization algorithm and ant colony optimization in feature selection domain have shown promising results [18]; some of which are summarized in the following.

Hamdani et al. proposed a new algorithm based on hierarchical genetic algorithms with bi-coded chromosome representation and new evaluation function [13]. In order to minimize the computational cost and also speed up the convergence speed, they used a hierarchical algorithm with homogeneous and heterogeneous population. In another attempt, Zhu et al. proposed a new algorithm which is a combination of genetic algorithm and local search method [40]. At first, GA population is generated randomly, then local search is applied to all of the individuals of the population in order to improve the classification accuracy and speed up the searching process. Tan et al. used SVM (Support Vector Machine) based on wrapper approach [31] in GA. In their proposed algorithm, GA searches for the best feature subset and the classification accuracy of SVM guides the search process. Gheyas et al. combined both simulated annealing (SA) and GA to use the advantages of both SA and GA [10]. In their proposed SAGA, GA helps to escape from local optimum of SA with the crossover operator. Nemati et al. proposed a new hybrid algorithm of GA and ACO in order to use the advantages of both algorithms [22]. In their algorithm, ACO performs a local search, while GA is used to perform a global search. Sivagaminathan et al. used ACO which searches for near-optimum solution and ANN is used as a classifying function [28]. ElAlami et al. proposed an algorithm based on GA, which optimizes the output nodes of ANN [7]. In their method, ANN is used to give a weight to each of the features and GA finds the optimal relevant features. Kabir et al. proposed a new hybrid algorithm that combines GA with local search method (HGAFS) [17]. Their proposed method selects the feature subset with a limited size; which is the important aspect of their method. Their method is a wrapper based method that uses both GA and ANN. In another attempt, Tabakhi et al. presented an unsupervised feature selection method based on ant colony optimization, called UFSACO [29]. Their proposed UFSACO is a filter-based method and the search space is represented as a fully connected undirected weighted graph. Xue et al proposed a series of methods based on PSO with novel initialization and updating mechanisms [35]. In their proposed algorithm, three new initialization strategies and three new personal best and global best updating mechanisms in PSO are presented to develop novel feature selection approaches; in which, maximizing the classification performance, minimizing the number of features and reducing the computational time are the main goals.

Despite good progress in solving feature selection problem, more study is also welcomed to further optimize the solutions. In all the proposed methods, one should choose either computationally feasible or optimality of the selected features. Further research is needed to develop more promising methods for feature selection with the aim of providing very good results. In the present work, FSFOA algorithm is proposed to further optimize the results of feature selection methods in the case of improving classification accuracy.

3. An overview of the Forest Optimization Algorithm (FOA)

Forest Optimization algorithm is an evolutionary algorithm, which is inspired by the procedure of a few trees in the forests [9]. FOA is proposed to solve continuous search space problems, but in this article we have attempted to adjust it to use in discrete search space problems like feature selection. FOA involves three main stages: 1 – Local seeding of the trees, 2 – Population limiting, and 3