Neurocomputing 171 (2016) 586–597

Contents lists available at ScienceDirect

Neurocomputing

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

Feature selection using multimodal optimization techniques

Shima Kamyab *, Mahdi Eftekhari

Department of Computer Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

ARTICLE INFO

Article history:
Received 7 July 2014
Received in revised form
6 June 2015
Accepted 30 June 2015
Communicated by Feiping Nie
Available online 16 July 2015

Keywords:
Multimodal optimization
Feature selection
Evolutionary feature selection

ABSTRACT

This paper investigates the effect of using Multimodal Optimization (MO) techniques on solving the Feature Selection (FSel) problem. The FSel problem is a high-dimensional optimization problem in the nature and thus needs a solver with high exploration power. On the other hand, if alternative optimal solutions could be provided for a problem, the implementation phase may become more selective depending on the cost and limitations of domain of the problem. The high exploration power and solution conservation capability of MO methods make them able to find multiple suitable solutions in a single run. Therefore, MO methods can be considered as a powerful tool of finding suitable feature subsets for FSel problem. In this paper, we made a special study on the use of MO methods in the feature selection problem. The binary versions of some existing Evolutionary Algorithm (EA) based MO methods like Dynamic Fitness Sharing (DFS), local Best PSO variants and GA_SN_CM, are proposed and used for selection of suitable features from several benchmark datasets. The results obtained by the MO methods are compared to some well-known heuristic approaches for FSel problem from the literature. The obtained results and their statistical analyses indicate the effectiveness of MO methods in finding multiple accurate feature subsets compared to existing powerful methods.

1. Introduction

Reduction of pattern dimensionality, via feature selection and feature extraction, is one of the most fundamental steps in data preprocessing. The aim of Feature Selection (FSel) is to choose a necessary and sufficient subset of features which is capable of describing the target concept while retaining the accuracy of classification in a dataset. Additional features may induce some disadvantageous effects on the classification process. First, they significantly slow down the learning process. Second, they deteriorate the classification accuracy by causing the classifier to over-fit the training data as irrelevant or redundant features may confound the learning algorithm [1]. In design of pattern classifiers, careful feature selection may improve both the quality and computation time of inducing subsequent models. Using fewer features often leads to both simpler and easier models to interpret and important insights into the application such as noise reduction [2].

The proposed feature selection methods in the literature have been categorized into four classes depending on how they evaluate the feature subsets: 1—Wrapper, 2—Filter, 3—Hybrid and 4—Embedded based methods.

Wrapper based methods aim at selecting those subsets of features which improve the performance of a predetermined learning model [3–7]. A wrapper model consists of two phases: Phase 1—feature subset selection, which selects the best subset using the accuracy of the classifier (on the training data) as a criterion. Phase 2—learning and testing, where a classifier is learned from the training data with the best feature subset, and is tested on the test data. Therefore, the wrapper approaches use the prediction performance of a model to assess the relative usefulness of subsets of features [8]. In [9], the feature selection problem is reformulated as a least-square regression problem equivalent to $\ell_{1,2}$-norm minimization on both loss function and regularization. This method uses the label of training data in its optimization procedure and therefore has a learning phase.

Filter approaches work based on the intrinsic properties of the data, rather than being biased toward a particular classifier [10–18]. The essence of filter methods is to seek the relevant features and eliminate irrelevant ones. They also consist of two phases: Phase 1—feature selection using measures such as information, distance, dependence, or consistency; no classifier is engaged in this phase. Phase 2—this is the same as in the wrapper's model, where a classifier is learned from the training data with the selected features, and is tested on the test data. In [19] a graph-based feature selection framework under Trace Ration criterion is proposed in which the feature selection problem is viewed as a special subspace learning task where the projection matrix is constraint to be selection matrix. In order to encode the relationship

* Corresponding author. Tel./fax: +91 343 3235901.
E-mail addresses: shima.kamyab@gmail.com (S. Kamyab), m.eftekhari@uk.ac.ir (M. Eftekhari).

http://dx.doi.org/10.1016/j.neucom.2015.06.068
0925-2312/

among, a graph structure data based on the proximity of data samples is used.

Hybrid methods attempt to take advantages of both wrapper and filter methods to obtain better feature sets [20–26]. In particular, hybrid methods are based on a sequential (e.g. two-step) approach where the first step is usually based on filter methods to reduce the number of features. Using this reduced set, in second step, a wrapper method is then employed to select desired number of features [1]. Embedded methods embed the feature selection procedure into the learning algorithm i.e., feature selection is occurring naturally as a part of learners [3]. From the literature, it seems more reasonable to use a wrapper (or embedded method) with a linear predictor as a filter and then train a more complex nonlinear predictor on the resulting variables [5]. Therefore, in this paper we used a wrapper approach with 1-NN classifier as predictor in the first phase of a wrapper based approach and then the resulted subsets were evaluated using more complex classifiers.

From a computational perspective, the feature selection problem is generally difficult to solve. It is inherently a combinatorial optimization problem [27]. Hence, optimization techniques are suitable methods to resolve this problem. Recently, several optimization based methods are used and proposed in order to find relevant feature subsets among possible combinations of features [25,28–30]. Evolutionary Algorithms (EAs) are capable of exploring the search space comprehensively due to their population based structures, in a reasonable time. However, these methods suffer from premature convergence problem because of the tendency of their population to converge to its best found solution. This problem could be solved by controlling the diversity of the population.

Finding multiple optimum solutions is the objective of a class of optimization methods called Multimodal Optimization (MO) techniques. In such EA based approaches, some additional mechanisms are involved in the EAs in order to help the main search process to avoid converging toward a local optimum. These mechanisms, called niching techniques, empower the search process to find multiple global and local optima by controlling the diversity of the population and performing some local search behavior to fine-tune the obtained solutions. These attributes of MO methods encourage them to be considered as appropriate tools to find suitable solutions for the feature selection problem.

In this paper, a special study on the use of MO methods to the feature selection problem is made which was not performed before. To do this, the modification of some well-known EA based MO methods and also our proposed niching method called GA_SN_CM are used for feature selection task and are compared with some well-known EA based methods for feature selection to study the power of MO methods on improving the feature selection results.

The reminder of this paper is organized as follows: Section 3 provides a brief description about using EA based optimization methods for feature selection. In Section 3, MO methods which are used for feature selection in this paper are discussed. In Section 4, the described MO methods are applied to several test problems and the results are presented and discussed. Finally, Section 5 is dedicated to conclusions.

2. Feature selection using EA-based optimization methods

Searching for an appropriate feature subset, which can represent the target concept in a large feature set, is an NP-hard problem. If the feature subset is modeled as a binary string where, each feature is either selected or not for inclusion, the number of possible non-empty feature subsets is 2^n-1, where n is the number of all features. Generally, for such problems, the optimal solution cannot be guaranteed to be acquired except by performing an exhaustive search in the solution space. The use of meta-heuristic techniques allows us to obtain reasonably good solutions without being forced to explore the whole solution space. The quality of "heuristic" solutions, depends on the characteristics of applied method. Recent studies show that meta-heuristic techniques are among superior methodologies. In real-world applications, people are more interested in obtaining good solutions in a reasonable amount of time rather than being obsessed with optimal solutions. Therefore, we favor meta-heuristic methods that are efficient for dealing with real-world applications [8]. The most frequently used meta-heuristic strategies applied to the feature selection problem are EA based methods. Such methods in the process of feature selection do not suffer from the so called *Nesting effect* caused by traditional sequential feature selection methods like SBS[1] [31] and SFS[2] [32] methods. *Nesting effect* corresponds to the inability to reselect the discarded features (SBS) and discarding the selected features (SFS). Using such sequential schemes in feature selection may result in finding local optimum solutions for the FSel problem. EA-based methods overcome this drawback by having no restriction on selecting features in during their search process. Moreover, when the number of variables is large, due to the power of parallel selection of features, the computational time of EA based methods seems to be considerably less than the methods based on forward selection and backward elimination.

There are several studies which investigate the effectiveness of EA based methods in feature selection process. They adopted different kinds of EAs such as Genetic Algorithm (GA) [22,25,28,29], PSO Algorithm [2,33–35], Differential Evolution (DE) Algorithm [36], ACO algorithm [37] and so on. These approaches also would act as filter, wrapper, hybrid or embedded depending on the way they evaluate fitness of their population.

Proposing some improvements of the exploration and local search powers of EAs for feature selection in several available studies, shows that these powers play crucial roles in obtaining better results for the FSel problem. For instance, GRASP [29] is an iterative process, in which each of the iterations has two phases: construction and local search. In the construction phase, a feasible solution is built. Then its neighborhood is explored by the local search. The final result will be the best solution found over all iterations. Chuang et al. in [34] used a new modification of PSO algorithm called CatfishBPSO for the feature selection process. In CatfishBPSO, the so-called "catfish" effect introduces a competition function into a group of individuals. Catfish particles are introduced into the search space if the fitness of gbest cannot be improved over a number of consecutive iterations. These catfish particles are introduced at extreme positions of the search space and will initialize a new search for these extreme positions. The catfish particles open up new opportunities for finding better solutions, and guide the entire swarm to promising new regions of the search space [34]. Introducing catfish particles in CatfishBPSO algorithm helps it to avoid converging toward a local optimum solution by increasing the exploration power and diversity of its population. High classification accuracy of this heuristic optimization method for feature selection indicates the effectiveness of its mechanism. These methods inspired us to use Multimodal Optimization techniques which have high exploration power and good local search behavior for feature selection in order to find more desirable feature subsets. In [8] Memetic Algorithm (MA), which is a population based approach with a local search mechanism to improve the solutions,

[1] Sequential Backward Selection.
[2] Sequential Forward Selection.