Decision Support Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Decision Support Systems

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

Multiobjective sparse ensemble learning by means of evolutionary algorithms

Jiaqi Zhao[a], Licheng Jiao[b], Shixiong Xia[a,*], Vitor Basto Fernandes[c,d], Iryna Yevseyeva[e], Yong Zhou[a], Michael T.M. Emmerich[f]

[a] School of Computer Science and Technology, China University of Mining and Technology, No. 1, Daxue Road, Xuzhou, Jiangsu 221116, China
[b] Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University, Xi'an, Shaanxi 710071, China
[c] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, University Institute of Lisbon, Av. das Forças Armadas, Lisboa 1649-026, Portugal
[d] School of Technology and Management, Computer Science and Communications Research Centre, Polytechnic Institute of Leiria, Leiria 2411-901, Portugal
[e] Faculty of Technology, De Montfort University, Gateway House 5.33, The Gateway, Leicester LE1 9BH, UK
[f] Multicriteria Optimization, Design, and Analytics Group, LIACS, Leiden University, Niels Bohrweg 1, Leiden 2333-CA, The Netherlands

ARTICLE INFO

Keywords:
Ensemble learning
Sparse representation
Classification
Multiobjective optimization
Change detection

ABSTRACT

Ensemble learning can improve the performance of individual classifiers by combining their decisions. The sparseness of ensemble learning has attracted much attention in recent years. In this paper, a novel multi-objective sparse ensemble learning (MOSEL) model is proposed. Firstly, to describe the ensemble classifiers more precisely the detection error trade-off (DET) curve is taken into consideration. The sparsity ratio (*sr*) is treated as the third objective to be minimized, in addition to false positive rate (*fpr*) and false negative rate (*fnr*) minimization. The MOSEL turns out to be augmented DET (ADET) convex hull maximization problem. Secondly, several evolutionary multiobjective algorithms are exploited to find sparse ensemble classifiers with strong performance. The relationship between the sparsity and the performance of ensemble classifiers on the ADET space is explained. Thirdly, an adaptive MOSEL classifiers selection method is designed to select the most suitable ensemble classifiers for a given dataset. The proposed MOSEL method is applied to well-known MNIST datasets and a real-world remote sensing image change detection problem, and several datasets are used to test the performance of the method on this problem. Experimental results based on both MNIST datasets and remote sensing image change detection show that MOSEL performs significantly better than conventional ensemble learning methods.

1. Introduction

The idea of ensemble learning methods [1] is to construct a set of classifiers with base learning algorithms and then classify new data points by taking a (weighted) vote of their predictions. Generally, ensemble methods combine the prediction of individual methods and can obtain better predictive performance than any individual method alone. Ensemble learning methods have attracted much attention in recent years. Not only have many ensemble algorithms been proposed [2,3], but also ensemble learning methods have been applied to many areas [4,5], such as medical information processing [1] and satellite image classification [6].

In general, an ensemble learning algorithm is constructed in two steps, i.e., training a number of component classifiers and then combining the predictions of the components. The most prevailing approaches for training component classifiers are bagging [7], boosting [8], random subspace [9], and rotation forest [10]. Recently, research has drawn attention to multiobjective optimization of ensemble learning [11,12] and several evolutionary multiobjective algorithms (EMOAs) have been used to deal with it. Generally, most of this work is trying to obtain a set of classifiers with good performance on both diversity and accuracy by using multiobjective optimization algorithms with different objectives. The multiobjective deep belief networks (DBNs) ensemble method was proposed in [13], in which an MOEA was applied to evolve multiple DBNs by considering accuracy and diversity as two conflicting objectives. A divide-and-conquer based optimization framework for ensemble classifiers generation was proposed in [12], in which the accuracy of each class was treated as the objectives to describe the performance of classifiers. Besides, maximizing the ensemble size is also taken as an additional objective. The

* Corresponding author.
E-mail address: shixiongxia.cumt@outlook.com (S. Xia).

https://doi.org/10.1016/j.dss.2018.05.003
Received 21 November 2017; Received in revised form 26 May 2018; Accepted 26 May 2018

Pareto image features were applied for candidate classifiers generation in [14] by using a multiobjective evolutionary trace transform algorithm. These methods do not consider the redundancy between classifiers and the efficiency of ensemble learning, as it requires a significant amount of memory to store the candidates of classifiers and lots of computation time is also needed to predict the label of each new input instance.

In this paper, we focus on combining the predictions of component classifiers by finding several appropriate sparse weight vectors for them. Many works have addressed the complexity of ensemble classifiers by reducing the number of classifiers in the component candidate set. The relationship between the ensemble learning and its component classifiers is analyzed in [15], which reveals that a better performance can be obtained by ensembling many instead of all the available classifiers. A genetic algorithm is adopted to evolve the weights of the component classifiers, showing that it can generate ensemble classifiers with small sizes but good generalization ability. The theoretical and empirical evidence in [16] suggests that a smaller ensemble size can often obtain better performance than a larger ensemble. It is, therefore, possible to obtain an ensemble which minimizes the number of individual classifiers and preserves or improves the performance of attributes, such as accuracy and cost of misclassification. However, only the accuracy is considered in this method, the result contains redundant classifiers, as the sparsity of ensemble classifiers is not considered. Several pruning strategies are analyzed in [17], including reduction error (RE), Kappa pruning (KP), complementarity measure (CM) and margin distance (MD). Matching pursuit (MP) is used to prune the ensemble classifiers in [18] by balancing the diversity and the individual accuracy. In these methods, the greedy strategy is used to search for the optimal classifiers set and it is easy to fall into the local extremum.

Sparse ensembles were proposed in [19]. The outputs of multiple classifiers were combined by using a sparse weight vector. The *hinge loss* and the *1-norm* regularization were exploited to calculate the sparse weight vector, formulated as a linear programming problem. However, the *1-norm* metric cannot describe the sparseness of ensemble classifiers precisely. This is because a weight vector with a group of small values can improve the performance of *1-norm* measurement but cannot improve the performance of sparseness. The *0-norm* metric can describe the sparseness more precisely [20]. The sparse ensemble learning is applied for synthetic aperture radar (SAR) image classification in [6] and for Youtube videos classification in [21]. The *0-norm* learning can be regarded as an NP-hard problem, it is still an open problem to search the global optimum.

Compressed sensing (CS) [22] was brought to ensemble learning in [23]. It explores the globally optimal subset of classifiers for a given ensemble. To solve the compressed sensing problem, a sparse weighting vector which contains many zeros should be generated first, and then appropriate weights should be provided for the remaining classifiers according to their relative importance. Several popular methods such as SpaRAS [24], OMP [25], FISTA [26], and PFP [27] are used to tune the weight vector of ensemble classifiers. In [23] it is shown that compressed sensing ensembles are often as accurate as, or more accurate than, conventional ensembles, although they use only small subsets of the total set of classifiers. However, the sparseness should be set in advance when using the compressed sensing methods. Meanwhile, the characteristics of the unbalanced data classification were not taken into consideration.

The contributions and drawbacks of the most related works of literature are listed in Table 1. Above all, the drawbacks of these methods are listed in the following: 1) The optimization algorithms used were easily trapped into local extremum; 2) only accuracy metric cannot describe ensemble performance precisely; and 3) the relationship between sparsity ensemble weights and ensemble performance was not analyzed in depth.

In this paper, we propose the novel concept of a multiobjective sparse ensemble learning (MOSEL) method, in which the relationship between the sparsity and the classification performance is explained. To accurately describe the performance of ensemble classifiers, the detection error trade-off (DET) [28] performance is taken into consideration by adopting the false positive rate (*fpr*) and the false negative rate (*fnr*) simultaneously. Besides, the sparsity ratio (*sr*) of ensemble classifiers is treated as the third objective to be minimized. The DET can describe the classifiers more precisely than the accuracy metric especially for unbalance data classification problems [28]. Besides, the evolutionary multiobjective algorithm (EMOA) [29] technique is first applied to evolve the combining weights of ensemble component classifiers. With the technique of tri-objective ensemble learning, we can obtain a set of ensemble classifiers with different sparseness, rather than an ensemble classifier with a certain sparseness that is previously set. The sparsity and the error rates of ensemble classifiers are explainable, and their trade-offs are quantifiable in the augmented DET (ADET) space.

We analyze the properties of the ADET for sparse ensemble learning and several state-of-the-art many-objective optimization algorithms are applied to solve multiobjective ADCH maximization problems, including the two-archive algorithm (Two_Arch2) [30], which focuses on convergence and diversity separately, the decomposition based algorithms, such as NSGA-III [31], the evolutionary algorithms based on both dominance and decomposition (MOEA/DD) [32], the reference vector guided evolutionary algorithm (RVEA) [33], an indicator based evolutionary algorithm with a reference point adaptation (AR-MOEA) [34], and 3D convex-hull-based evolutionary multiobjective optimization algorithm (3DFCH-EMOA) [35,36]. By using EMOAs, we can obtain a set of potentially optimal ensemble classifiers with different *sr-fpr-fnr* trade-offs.

The remaining paper is organized as follows. Section 2 gives a brief introduction to multiobjective optimization of a sparse ensemble method. Section 3 presents the results of several classification problems with MNIST [37] and remote sensing change detection datasets, and Section 4 provides concluding remarks.

2. Multiobjective sparse ensemble learning

2.1. Ensemble learning

The idea of a *sparse ensemble* of classifiers is to combine the predictions of all classifiers in the candidate set using a sparse weight vector. The sparse vector has many elements with the value of zero and only classifiers corresponding to nonzero weights are selected for the ensemble. To improve the performance of the ensemble classifier and to reduce the memory demand for the components, it is required to select an optimal subset of classifiers and the corresponding weights vector for this subset. The problem of seeking sparse weights vectors can be modeled as a combinatorial optimization problem, which can be solved by evolutionary algorithms [30].

In this paper, we only consider binary supervised ensemble classification problems. With a set of training samples $X_{tr} = \{(x_j,y_j)|x_j \in R^d, y_j \in \{-1,+1\}, j = 1,2,\ldots,M_{tr}\}$, where y_j is the class label corresponding to a given input x_j, d is the dimensionality of sample of features, and M_{tr} is the number of instances. Note that in this work we only consider binary classification problems and we set the labels as $\{-1,1\}$, where 1 represents positive category and -1 represents negative category, given a set of classifiers $\{C_1(x),C_2(x),\ldots,C_N(x)\}$, where $C_i(x)$ is the i-th classifier in the candidate ensemble set. Usually, the classifier $C_i(x)$ is obtained by using the training dataset X_{tr} with the strategy of random selection of the features or the instances.

A classifier can be obtained by using a training dataset with a machine learning algorithm, which can be described as an estimate of the unknown function $y = f(x)$. The classifier $C_i(x)$ is a hypothesis $f_i(x)$ about the true function $f(x)$, which can predict the class label y for a new input vector x from a testing dataset X_{ts} or a validation dataset X_{val}.