Pattern Recognition 60 (2016) 692–705

Contents lists available at ScienceDirect

Pattern Recognition

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

Progressive subspace ensemble learning

Zhiwen Yu [a,*], Daxing Wang [a,*], Jane You [b], Hau-San Wong [c], Si Wu [a], Jun Zhang [a], Guoqiang Han [a]

[a] School of Computer Science and Engineering, South China University of Technology, China
[b] Department of Computing, Hong Kong Polytechnic University, Hong Kong
[c] Department of Computer Science, City University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 16 July 2015
Received in revised form
15 May 2016
Accepted 19 June 2016
Available online 21 June 2016

Keywords:
Ensemble learning
Classifier ensemble
Random subspace
AdaBoost
Decision tree

ABSTRACT

There are not many classifier ensemble approaches which investigate the data sample space and the feature space at the same time, and this multi-pronged approach will be helpful for constructing more powerful learning models. For example, the AdaBoost approach only investigates the data sample space, while the random subspace technique only focuses on the feature space. To address this limitation, we propose the progressive subspace ensemble learning approach (PSEL) which takes into account the data sample space and the feature space at the same time. Specifically, PSEL first adopts the random subspace technique to generate a set of subspaces. Then, a progressive selection process based on new cost functions that incorporate current and long-term information to select the classifiers sequentially will be introduced. Finally, a weighted voting scheme is used to summarize the predicted labels and obtain the final result. We also adopt a number of non-parametric tests to compare PSEL and its competitors over multiple datasets. The results of the experiments show that PSEL works well on most of the real datasets, and outperforms a number of state-of-the-art classifier ensemble approaches.

1. Introduction

Nowadays, more and more researchers pay attention to ensemble learning [1–10,62–67], due to its high performance and robustness in various machine learning tasks. A number of ensemble learning algorithms have been proposed in recent years, such as hybrid adaptive ensemble learning [1], ensemble learning based on random linear oracle [2], bagging [3], random subspace [4], classifier ensemble based on neural networks [5,6], boosting [7], fuzzy classifier ensemble [8], rotation forest [9], random forest [10], and heterogeneous ensemble of classifiers [68,69]. Ensemble learning approaches have been successfully applied in the field of data mining [11,12], text categorization [13,14], bioinformatics [15,16] and face image classification [17,18], along with many other tasks. In the field of data mining, Polikar et al. [11] introduced a new classifier ensemble framework named Learn++.MF to address the missing value problem in data mining. It classifies missing values with multiple classifiers. Galar et al. [12] designed a new classifier ensemble algorithm named EUSBoost which incorporates random sub-sampling with boosting in the ensemble to handle the data imbalance problem. In the field of text categorization, Liu et al. [13] designed a novel classifier ensemble framework named BAM-Vote Box, and introduced a new feature selection algorithm for text categorization task. Saeedian et al. [14] designed a new spam detection algorithm using the ensemble learning framework based on clustering and weighted voting. In the field of bioinformatics, Liu et al. [15] introduced a new ensemble learning algorithm which adopts mutual information to search for important gene subsets, and then use the ensemble method to classify cancer gene expression data. Plumpton et al. [16] designed a new ensemble framework to classify streaming functional magnetic resonance images and measure brain activities. In the field of face image classification, Connolly et al. [17] introduced a face recognition framework which combines multiple neural network classifiers and produces a progressive ensemble learning algorithm for face recognition from video data. Zhang et al. [18] introduced RDA, an extension of LDA, to analyze face image data in a subspace. Ahmadvand et al. [57] adopted the ensemble classifier for MRI brain image segmentation. In summary, both theoretical and empirical analysis show that under certain conditions, ensemble learning achieves better performance and robustness than single classifiers, due to its capability to integrate more information encapsulated in multiple learners.

While there are different kinds of ensemble learning approaches, most of them only consider the distribution of the data in the sample space, or the distributions of the attributes in the

* Corresponding authors.
E-mail address: zhwyu@scut.edu.cn (Z. Yu).

http://dx.doi.org/10.1016/j.patcog.2016.06.017
0031-3203/

feature space. However, the two sets of distributions should be combined to achieve a better result in most of the cases. To address this limitation, we propose a progressive subspace ensemble learning approach (PSEL) which explores both the data sample space and the feature space at the same time. Specifically, PSEL adopts the random subspace technique to generate a set of subspaces. Each subspace is used to train a classifier in the original ensemble. Then, a progressive selection process is adopted to select the classifiers based on two cost functions which incorporate current and long-term information. Finally, a weighted voting scheme is used to combine the predicted results from individual classifiers of the ensemble, and generate the final result. The properties of PSEL are analyzed theoretically. We also adopt a number of non-parametric tests to compare PSEL and its competitors on multiple datasets. Our experiments show that PSEL works well on the real datasets, and outperforms most of the state-of-the-art classifier ensemble approaches.

The contribution of the paper is twofold. First, the progressive subspace ensemble learning approach is proposed, which not only explores the data sample space, but also takes the feature space into consideration. Second, two cost functions which incorporate current and long-term information are designed to perform the progressive selection process.

The remainder of the paper is organized as follows. Section 2 discusses previous works related to ensemble learning. Section 3 describes the progressive subspace ensemble learning approach and the progressive selection process. Section 4 analyzes the proposed algorithm theoretically. Section 5 experimentally evaluates the performance of our proposed approach. Section 6 presents our conclusion and future work.

2. Related work

As an important branch of ensemble learning, classifier ensemble has drawn the attention of researchers from many fields. In recent years, classifier ensemble approaches based on different viewpoints have been developed. Some of them focus on generating new ensembles [19–24]. For example, Yu et al. [19] introduced a graph based semi-supervised ensemble algorithm which generates the ensemble by performing multiple rounds of dimensionality reduction. Ye et al. [20] designed an extension of the random forest algorithm which ensures that enough good features are selected in each subspace. Guo et al. [21] applied the rough set theory for dimensionality reduction in the ensemble framework. Zhang et al. [22] used the rotation space technique to increase the diversity of the classifiers in the random forest approach. Zhang et al. [23] incorporated label information into canonical correlation analysis and designed a new ensemble algorithm. Hllermeier et al. [24] developed a label ranking based voting framework, and performed theoretical analysis related to weighted voting. Zhang et al. designed a random vector functional link (RVFL) network based classifier [58] and a new oblique decision tree ensemble [59], which achieve good performance on the dataset from UCI machine learning repository. On the other hand, some researchers focus on the ensemble property. For example, Kuncheva [25] studied ensembles with Kappa-error diagrams. Wang et al. [26] analyzed the generalization and fuzziness properties of an ensemble. Other researchers consider the combination of different classifier ensemble approaches. For example, Nanni et al. [68] studied different combinations of ensemble techniques to improve the performance of AdaBoost. They also explored a heterogeneous ensemble which considers the combination of different classification approaches [69]. Zhang et al. [70] investigated the combination of rotation forest and AdaBoost. A survey of the different classifier ensemble techniques have been included in [34].

In addition, ensemble learning is applied in a number of research fields. For example, Rasheed et al. [27] used the classifier ensemble approach for electromyographic signal decomposition. Tian et al. [28] designed a classifier ensemble consisting of Haar-like and shapelet components for pedestrian detection. Guo et al. [29] proposed a two-stage pedestrian detection algorithm using AdaBoost and SVM. There are also applications in breast tumor discrimination [30], protein structural class prediction [31], network intrusion detection [32], and handwritten digit recognition [33] as well. Daliri [54] applied an extreme learning machine-based ensemble classifier for breast tissue classification. In addition, he also integrated multiple feature selection methods for cognitive state prediction in functional magnetic resonance imaging (fMRI) [55], and performed feature selection by applying swarm optimization and support vector machine in medical informatics [56].

While these ensemble learning approaches lead to performance improvement when compared to conventional techniques, most of them focus on the sample space or the feature space separately. Some of these works [64] consider exploring the sample space and the feature space at the same time.

Sometimes not all classifiers need to be selected for the ensemble. Some researchers suggest combining a suitable subset of the classifiers to pursue better performance [35–39]. For example, Hu et al. [35] applied the rough set theory to generate base classifiers, and then used an accuracy based search and prune method to select classifiers. Cruz et al. [36] introduced a dynamic framework which measures the competency of each classifier with a meta-learning approach, and select classifiers from the ensemble. Xiao et al. [37] developed a new dynamic classifier selection algorithm which considers both the accuracy and the diversity. Santos et al. [38] introduced an overproduce-and-choose strategy based ensemble member selection approach. Yang et al. [39] introduced a classifier ensemble selection method which also considers the accuracy and the diversity of the classifiers.

We focus on the random subspace technique [40], which has been successfully used for human detection [41], image retrieval [42], functional magnetic resonance imaging classification [43], traffic flow forecasting [44], and so on. The random subspace technique is one of the most popular ensemble learning approaches, which makes use of a combination of different subspaces to increase the diversity of the ensemble, and improve the performance of the ensemble learning approach. Fig. 1 provides an overview of the random subspace (RS) approach. Specifically, RS first generates B random subspaces using a predefined sampling rate $\kappa \in [0, 1]$. Then, the training samples are transformed from the full space to the corresponding random subspaces. The new training samples in each subspace are used to train a classifier. B trained classifiers corresponding to the B random subspaces will be obtained. Next, given a testing dataset, RS will obtain a set of predicted results for the samples from the individual classifiers in the ensemble. Finally, the majority voting scheme is adopted to

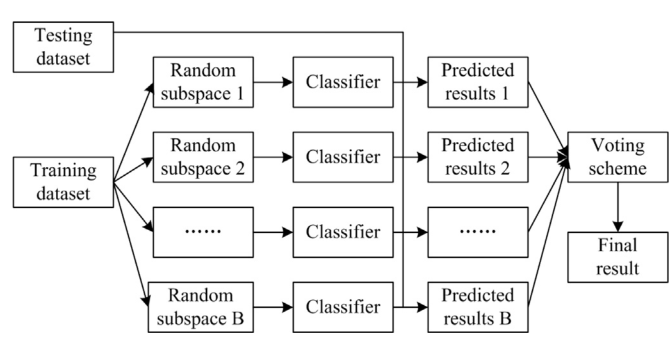

Fig. 1. An overview of the random subspace approach.