Applied Soft Computing 58 (2017) 75–91

Contents lists available at ScienceDirect

Applied Soft Computing

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

# Considering diversity and accuracy simultaneously for ensemble pruning

Qun Dai*, Rui Ye, Zhuan Liu

*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*

ARTICLE INFO

*Article history:*
Received 19 December 2015
Received in revised form 14 February 2017
Accepted 20 April 2017
Available online 30 April 2017

*Keywords:*
Ensemble selection
Greedy ensemble pruning (GEP) algorithm
Simultaneous diversity & accuracy (SDAcc)
Diversity-focused-two (DFTwo)
Accuracy-reinforcement (AccRein)

ABSTRACT

Diversity among individual classifiers is widely recognized to be a key factor to successful ensemble selection, while the ultimate goal of ensemble pruning is to improve its predictive accuracy. Diversity and accuracy are two important properties of an ensemble. Existing ensemble pruning methods always consider diversity and accuracy separately. However, in contrast, the two closely interrelate with each other, and should be considered simultaneously. Accordingly, three new measures, i.e., Simultaneous Diversity & Accuracy, Diversity-Focused-Two and Accuracy-Reinforcement, are developed for pruning the ensemble by greedy algorithm. The motivation for Simultaneous Diversity & Accuracy is to consider the difference between the subensemble and the candidate classifier, and simultaneously, to consider the accuracy of both of them. With Simultaneous Diversity & Accuracy, those difficult samples are not given up so as to further improve the generalization performance of the ensemble. The inspiration of devising Diversity-Focused-Two stems from the cognition that ensemble diversity attaches more importance to the difference among the classifiers in an ensemble. Finally, the proposal of Accuracy-Reinforcement reinforces the concern about ensemble accuracy. Extensive experiments verified the effectiveness and efficiency of the proposed three pruning measures. Through the investigation of this work, it is found that by considering diversity and accuracy simultaneously for ensemble pruning, well-performed selective ensemble with superior generalization capability can be acquired, which is the scientific value of this paper.



## 1. Introduction

Ensemble learning has been successfully used as a desirable learning scheme for many regression and classification problems because of its potential to greatly increase predictive accuracy [1–9]. An ensemble refers to a group of base learners whose decisions are aggregated with the goal of achieving better performance than its constituent members [9,10]. Typically, ensemble learning algorithm consists of two main stages: first, the production of multiple base classifiers for one specific task, and then the combination of these classifiers to get a final predictive decision [1,9,11–13].

However, it is obvious that combining all of the classifiers in an ensemble adds a lot of computational overheads [9,12,13]. Both theoretical and empirical studies have shown that instead of using the whole ensemble, a subset of the ensemble can achieve equivalent or even better generalization performance [7,9,14]. Therefore, an additional intermediate stage that deals with the selection of an appropriate subensemble prior to its combination has to be considered [1,7,9,13,15–23]. This stage is generally termed as ensemble pruning [3,5], selective ensemble [7,16], ensemble selection [24] or ensemble thinning [1].

The problem of pruning an ensemble of classifiers has been proven to be NP-complete [25]. Finding the best subensemble through exhaustive searching is not feasible for the original ensemble with large or even moderate size [26]. The greedy algorithms [27–29], also known as hill climbing algorithms, which reduce the search space appropriately, seem to be a good choice. Various greedy ensemble pruning (GEP) algorithms have been proposed, just as in Refs. [1,2,24,30].

There are two key elements for GEP algorithms: the search direction and evaluation measure [28,30]. Typically, there are two search directions, i.e., forward expansion and backward shrinkage. Many researchers have compared the effect of the two different search directions in their research works [1–3,5,24,28,30]. As to the factor of evaluation measure, various evaluation measures have been proposed to guide the pruning algorithm and have made a success. The measures of uncertainty weighted accuracy (UWA) [30], complementariness (COM) [2] and concurrency (CON) [1], which will be discussed in detail in Section 2.2, are the three diversity measures particularly suitable for GEP algorithm. It is worth to mention

* Corresponding author.
*E-mail address:* daiqun@nuaa.edu.cn (Q. Dai).

http://dx.doi.org/10.1016/j.asoc.2017.04.058
1568-4946/

that UWA can be seen as an improvement to COM and CON, and it leads to better performance compared to the other two.

It is found through our investigation that, the evaluation measures adopted in GEP algorithms are usually focused on either the ensemble diversity or its accuracy solely, while ignore the counterpart. However, actually, both diversity and accuracy are crucial to the ensemble, and they interrelate with each other closely. A successful ensemble should possess a sufficiently high level of diversity, while an ensemble shows the success in its good predictive accuracy. If an ensemble possesses better predictive accuracy than its constituent members, its diversity is supposed to reach a high enough level. Equivalently, if an ensemble does not possess sufficiently high error-correction ability among its components, combination will do little to improve its classification performance. Accordingly, diversity and accuracy should not be considered separately, but rather, they should be taken into account simultaneously. This is the major argument of this study. We found in this work that through considering diversity and accuracy simultaneously for ensemble selection, pruned ensemble with excellent classification performance and superior generalization capability can be achieved, which is just the scientific value added of our paper.

The above understanding about the ensemble diversity and accuracy motivates the proposal of three new evaluation measures, i.e., **S**imultaneous **D**iversity & **Acc**uracy (SDAcc), **D**iversity-**F**ocused-**Two** (DFTwo) and **Acc**uracy-**Rein**forcement (AccRein), which are the main contributions of this paper. The three new evaluation measures are also designed specifically for GEP algorithms, since related researches have shown that GEP algorithms are usually able to achieve good performance.

The motivation for devising SDAcc is to take into consideration the difference between the current subensemble and the current candidate classifier, and simultaneously, to consider the accuracy of both of them. Moreover, different from UWA, SDAcc attempts to correctly classify all the samples, including those samples which are very hard to classify, in order that, the generalization performance of the pruned ensemble could be further boosted. As is explained in detail in Section 3.1, both the design of SDAcc and the reasons for its design are different from that of UWA presented in Ref. [30].

The inspiration of the measure DFTwo comes from the concept that the ensemble diversity attaches more importance to the difference among the classifiers in an ensemble. Consequently, the design of DFTwo focuses on the difference between the subensembe and the candidate classifier.

The proposal of the measure AccRein takes into consideration the property of ensemble diversity and accuracy simultaneously, so that the generalization performance of the ensemble could be further improved. It reinforces the concern about the property of ensemble accuracy.

The rest of the paper is organized as follows. Section 2 introduces some background knowledge of the ensembles selection paradigm, and reviews the ensemble pruning algorithm based on the greedy selection strategy. The details of the proposed three evaluation measures, i.e., SDAcc, DFTwo and AccRein, are presented in Section 3. Section 4 presents the results of the experimental study. Finally, conclusions and discussions about future works are given in Section 5.

## 2. Literature review

### 2.1. Background knowledge of the ensemble selection paradigm

It is well-known that aggregating the predictive decisions of classifiers in an ensemble usually achieves better generalization performance in comparison with individual classifiers [2,31,32]. However, crucial defects exist in ensemble methods that, the system complexity is increased dramatically. Therefore, ensemble selection paradigm emerged, which can reduce the required computing time complexity and space complexity. And simultaneously, the selected subensembles often possess better predictive accuracy compared to the original ensembles. The working principle of ensemble selection paradigm is the so-called “over produce and choose” strategy. After a collection of member classifiers have been produced, a subset of them that are simultaneously precise and varied is chosen as the ensemble selection solution [17].

A lot of research work has been conducted on ensemble selection paradigm. Several heuristics were proposed by Margineantu and Dietterich for ensemble selection with AdaBoost algorithm on the basis of diversity and accuracy measures [3]. Banfield et al. pruned an ensemble using sequential backward contraction method [1]. Ensemble diversity and accuracy measures are also employed to discern the incapable classifiers, which will be deleted from the initial ensemble in sequence. Kiyoung Choi et al. presented an ensemble of multiple feature extractors and classifiers in hierarchy structure [33]. Reinforcement machine learning and Bayesian networks are aggregated into their system to improve the classification performance. In [34], Rafael M. O. Cruz et al. utilized meta-learning method to develop a dynamic ensemble selection algorithm. Five sets of meta-features were designed. Every set of meta-features corresponds to a distinct measurement which can be used to evaluate the competence of a single classifier.

Yiannis Kokkinos et al. formulated a distributed privacy-preserving data mining technique in large scale decentralized data locations, where a few neural network models were generated to construct an ensemble [35]. The optimal solution of classifiers was determined using their proposed confidence ratio affinity propagation in an asynchronous distributed and privacy-preserving calculating circle. Rafal Lysiak et al. proposed a probabilistic model exploiting measures of classifier competence and diversity. A subensemble was formed on the basis of the dynamic ensemble selection paradigm utilizing the two proposed measures [36]. Chee Peng Lim et al. presented an ensemble optimizer based on a multi-objective evolutionary algorithm to address feature selection and classification issues. The ensemble optimizer was formulated utilizing the modified micro genetic algorithm [37].

Quan Zou et al. designed a method D3C by combining k-means clustering, dynamic selection, circulating and a sequential search method [38]. Shenmin Song et al. developed a self-organizing reproduction mechanism based on the regularity property of multiobjective optimization problems, and designed a self-organizing multiobjective evolutionary algorithm based on decomposition with neighborhood ensemble [39]. An ensemble of multiple neuron neighborhood sizes is built to form the mating pools. In [40], Bin Yang et al. presented a flexible neural tree model in order to detect somatic mutations in tumour-normal paired sequencing data. An ensemble method for classification using radial basis function network as the nonlinear integration function is proposed so as to further improve the classification performance.

Diversity is an important property for an ensemble of base classifiers, which is usually measured on the basis of a specific pruning dataset [1]. The more uniformly distributed the errors are, the greater the ensemble diversity could be, and vice versa [1]. The property of diversity cares about the characteristics of difference and similarity among the base classifiers in an ensemble.

In [11], Zhou et al. theoretically analyzed the effect of diversity on the generalization performance of ensemble combined with voting method in the PAC-learning framework, and concluded that enhancing diversity could be a good way to realize selective ensemble learning. Empirical results have demonstrated that there exists correlation between the accuracy of an ensemble and the diversity