Information Fusion 34 (2017) 28–42

Contents lists available at ScienceDirect

Information Fusion

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

An efficient ensemble pruning approach based on simple coalitional games

Hadjer Ykhlef[a,*], Djamel Bouchaffra[b]

[a] Department of Computer Science, University of Blida, Algeria
[b] Design of Intelligent Machines Group, Centre de recherche des technologies avancées, Algeria

ARTICLE INFO

Article history:
Received 19 November 2015
Revised 29 May 2016
Accepted 7 June 2016
Available online 8 June 2016

Keywords:
Ensemble pruning
Simple coalitional game
Banzhaf power index
Weighted voting game
Diversity

ABSTRACT

We propose a novel ensemble pruning methodology using non-monotone Simple Coalitional Games, termed SCG-Pruning. Our main contribution is two-fold: (1) Evaluate the diversity contribution of a classifier based on Banzhaf power index. (2) Define the pruned ensemble as the minimal winning coalition made of the members that together exhibit moderate diversity. We also provide a new formulation of Banzhaf power index for the proposed game using weighted voting games. To demonstrate the validity and the effectiveness of the proposed methodology, we performed extensive statistical comparisons with several ensemble pruning techniques based on 58 UCI benchmark datasets. The results indicate that SCG-Pruning outperforms both the original ensemble and some major state-of-the-art selection approaches.

1. Introduction

Ensemble learning remains a challenging task within the pattern recognition and machine learning community [1–4]. A large body of literature has shown that a combination of multiple classifiers is a powerful decision making tool, and usually generalizes better than a single classifier [5–7]. Ensemble learning builds a classification model in two steps. The first step concerns the generation of the ensemble members (also called team, committee, and pool). To this end, several methods such as: boosting [5], bagging [6], random subspace [8], and random forest [9] have been introduced in the literature. In the second step, the predictions of the individual members are merged together to give the final decision of the ensemble using a combiner function. Major combining strategies include: majority voting [6], performance weighting [5], stacking [6], and local within-class accuracies [10]. Ensemble learning has demonstrated a great potential for improvement in many real-world applications such as: remote sensing [1], face recognition [2], intrusion detection [3], and information retrieval [4].

It is well-accepted that no significant gain can be obtained by combining multiple identical learning models. On the other hand, an ensemble whose members make errors on different samples reaches higher prediction performance [5,6]. This concept refers to the notion of *diversity* among the individual classifiers. Unfortunately, the relationship between diversity and the ensemble generalization power remains an open problem. As suggested by many authors [5,11,12], an ensemble composed of highly diversified members may result in a better or worse performance. In other words, diversity can be either harmful or beneficial and therefore requires an adequate quantification. As a matter of fact, it has been demonstrated that maximizing diversity measures does not necessarily have a positive impact on the prediction performance of the committee [13].

Despite their remarkable success, ensemble methods can negatively affect both the *predictive performance* and the *efficiency* of the committee. Specifically, most techniques for growing ensembles tend to generate an unnecessarily large number of classifiers in order to guarantee that the training error rate reaches its minimal value. This necessity may result in overfitting the training set, which in turn causes a reduction in the generalization performance of the ensemble. Furthermore, an ensemble made of many members incurs an increase in memory requirement and computational cost. For instance, an ensemble made of C4.5 classifiers can require large memory storage [14]; a set of lazy learning methods, such as k-nearest neighbors and K∗, may increase the prediction time. The memory and computational costs appear to be negligible for toy datasets, nevertheless they can become a serious problem when applied to real-world applications such as learning from data stream.

All the above reasons motivate the appearance of ensemble pruning approaches (also called ensemble shrinking, ensemble thinning, and ensemble selection). Ensemble pruning aims at ex-

[*] Corresponding author.
E-mail addresses: ykhlef.hadjer@gmail.com (H. Ykhlef), djamel.bouchaffra@gmail.com (D. Bouchaffra).

http://dx.doi.org/10.1016/j.inffus.2016.06.003
1566-2535/© 2016 Elsevier B.V. All rights reserved.

tracting a subset of classifiers that optimizes a criterion indicative of a committee generalization performance. Given an ensemble composed of n classifiers, finding a subset that yields the best prediction performance requires searching the space of $2^n - 2$ non-empty subsets, which is intractable for large ensembles. This problem has been proven to be NP-complete [7]. To alleviate this computational burden, many ensemble pruning approaches have been introduced in the literature. Most of these techniques fall into three main categories: ranking-based, optimization-based, and clustering-based approaches. Please, refer to the related work subsection for additional details.

Based on these insights, this paper considers the problem of ensemble pruning as a Simple Coalitional Game (SCG). The proposed methodology aims at extracting sub-ensembles with moderate diversities while ignoring extreme scenarios: strongly correlated and highly diversified members. This mission is achieved in three steps: (1) We formulate ensemble pruning as a non-monotone SCG played among the ensemble members. (2) We evaluate the *power* or the *diversity contribution* of each ensemble member using Banzhaf power index. (3) We define the pruned ensemble as the *minimal winning coalition* constituted of the best ranked members. It is worth underscoring that the original definition of Banzhaf power index for non-monotone SCGs is intractable. Specifically, given a n-player game, the calculation of Banzhaf power index involves summing over 2^{n-1} coalitions, which is unfeasible for large values of n. To overcome this computational difficulty, we introduce a *new formulation of Banzhaf power index* for the proposed game, and show that its time complexity is pseudo-polynomial.

1.1. Related work

Tsoumakas et al. classified the ensemble pruning approaches into four categories [15]:

1.1.1. Ranking-based approaches

Methods of this category first assign a rank to every classifier according to an evaluation measure (or criterion); then, the selection is conducted by aggregating the ensemble members whose ranks are above a predefined threshold. The main challenge a ranking-based method faces, consists of adequately setting the criterion used for measuring every member's contribution to the ensemble performance. For instance, Margineantu and Dietterich introduced *Kappa pruning*, which selects a subset made of the most diverse members of the ensemble [14]. Specifically, it first measures the agreement between all pairs of classifiers using kappa statistic; it then selects the pairs of classifiers starting with the one which has the lowest kappa statistic (high diversity), and it considers them in ascending order of their agreement until the desired number of classifiers is reached.

Zheng Lu et al. proposed to estimate each classifier's contribution based on the diversity/accuracy tradeoff [16]. Then, they ordered the ensemble members according to their contributions in descending order. In the same regard, Ykhlef and Bouchaffra formulated ensemble pruning problem as an induced subgraph game [17]. Their approach first ranks every classifier by considering the ensemble diversity and the individual accuracies based on Shapley value; then, it constitutes the pruned ensemble by aggregating the top N members.

Galar et al. introduced several criterions for ordering ensemble members in the context of imbalanced classification [18]. They investigated and adapted five well-known approaches: Reduce error [14], Kappa pruning [14], Boosting-based [19], Margin distance minimization [20], and Complementarity measure [20].

1.1.2. Optimization-based approaches

This category formulates ensemble pruning as an optimization problem. A well-known method of this category is Genetic Algorithm based Selective ENsemble (Gasen) [21]. This technique assigns a weight to each classifier; a low value indicates that the associated individual member should be excluded. These weights are initialized randomly, and then evolved toward an optimal solution using *genetic algorithm*. The fitness function is computed based on the corresponding ensemble performance on a separate sample set. Finally, pruning is conducted by discarding members whose weights are below a predefined threshold.

Zhang et al. formulated ensemble pruning as a *quadratic integer programming* problem that considers the diversity/accuracy tradeoff [22]. Since this optimization problem is NP-hard, they used *semi definite programming* on a relaxation of the original problem to efficiently approximate the optimal solution.

Rokach introduced Collective Agreement-based ensemble Pruning (CAP), a criterion for measuring the goodness of a candidate ensemble [23]. CAP is defined based on two terms: member-class and member-member agreement. The first term indicates how much a classifier's predictions agree with the true class label, whereas the second term measures the agreement level between two ensemble members. This metric promotes sub-ensembles whose members highly agree with the class and have low inter-agreement among each other. Note that CAP provides only a criterion for measuring the goodness of a candidate ensemble in the solution space, and hence requires defining a search strategy like best-first or directed hill climbing [6,15].

1.1.3. Clustering-based approaches

The key idea behind this category consists of invoking a clustering technique, which allows identifying a set of representative *prototype* classifiers that compose the pruned ensemble. A clustering-based method involves two main steps. In the first step, the ensemble is partitioned into clusters, where individual members in the same cluster make similar predictions (strong correlation), while classifiers from different clusters have large diversity. For this purpose, several clustering techniques such as k-means [24], hierarchical agglomerative clustering [25], and deterministic annealing [26] have been proposed. In the second step, each cluster is separately pruned in order to increase the diversity of the ensemble. For example, Bakker and Heskes selected the individual members at the *centroid* of each cluster to compose the pruned ensemble [26].

1.1.4. Other approaches

This category comprises the pruning approaches that do not belong to any of the above categories. For example, Partlas et al. [27] considered the ensemble pruning problem from a *reinforcement learning* perspective; Martínez-Muñoz et al. used AdaBoost to prune an ensemble trained by Bagging [19].

1.2. Contributions and outline

The contribution of the proposed research is described by the following tasks:

(1) We propose a novel methodology for pruning an ensemble of learning models based on the minimal winning coalition and Banzhaf power index.
(2) We present a new representation for non-monotone SCGs and provide, under some restrictions, a pseudo-polynomial time algorithm for computing Banzhaf power index.
(3) We show the efficiency of the proposed methodology through extensive experiments and statistical tests using a large set of 58 UCI benchmark datasets.