Neurocomputing 196 (2016) 14–22

Contents lists available at ScienceDirect

Neurocomputing

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

Untrained weighted classifier combination with embedded ensemble pruning

Bartosz Krawczyk*, Michał Woźniak

Department of Systems and Computer Networks, Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland

ARTICLE INFO

Article history:
Received 22 April 2015
Received in revised form
3 February 2016
Accepted 16 February 2016
Available online 4 March 2016

Keywords:
Machine learning
Classifier ensemble
Combination rule
Ensemble pruning
Weighted aggregation

ABSTRACT

One of the crucial problems of the classifier ensemble is the so-called combination rule which is responsible for establishing a single decision from the pool of predictors. The final decision is made on the basis of the outputs of individual classifiers. At the same time, some of the individuals do not contribute much to the collective decision and may be discarded. This paper discusses how to design an effective combination rule, based on support functions returned by individual classifiers. We express our interest in aggregation methods which do not require training, because in many real-life problems we do not have an abundance of training objects or we are working under time constraints. Additionally, we show how to use proposed operators for simultaneous classifier combination and ensemble pruning. Our proposed schemes have embedded classifier selection step, which is based on weight thresholding. The experimental analysis carried out on the set of benchmark datasets and backed up with a statistical analysis, proved the usefulness of the proposed method, especially when the number of class labels is high.

1. Introduction

Machine learning becomes an irreplaceable tool for managing the data flood in the big data era. The amount, complexity and velocity of generated data greatly exceeds perceptive abilities of any human being. Therefore, designing novel and efficient methods for automated learning from data [9,38] is still the focus of intense research.

For a considered recognition task, we may often have more than a single classifier available. What is interesting, the number of misclassified objects by all individual classifiers is typically small. From this we can conclude that even if individual classifiers do not have high quality, their union could form a reasonably good compound classifier. The considered approach is called a multiple classifier system (MCS), combined classifier or classifier ensemble and is considered as one of the most vital fields in the contemporary machine learning [43].

Forming an ensemble requires an input pool of classifiers and a method for combining their individual outputs into a single committee decision. Optionally, one may assume that not all of classifiers from the pool are significantly important, and perform a pruning step in order to discard some learners. In recent years, many different schemes for this task were presented, because each of them is subject to some restrictions. Many approaches are computationally expensive and cannot be used for real-time classification or stream mining. Therefore, despite almost two decades of progress there are still novel ensemble approaches being frequently proposed in the literature.

This work focuses on crucial ensemble forming steps: pruning and weighted classifier combination.

For most real-life classification tasks we can create/collect a large number of classifiers. However, for ensemble to work properly it should be formed by mutually complementary models of high individual quality. Adding new classifiers that do not exploit a new area of competence do not improve the ensemble, only increases the computational cost and reduces its robustness. The problem lies on how to select a useful subgroup from a large pool of classifiers at hand. Due to the high computational complexity of full-search over all of the possible classifier subgroups, several heuristic groups of methods were proposed [39]. However, one should note that these methods are highly dependent on properly established evaluation criteria and require significant processing time. These properties can become prohibitive in certain application areas.

When having selected a number of competent classifiers, one need to design a combination rule in order to establish a collective decision of the ensemble [35]. Such a mechanism should be able to exploit the individual strengths of classifiers in the pool, while at the same time minimizing their drawbacks [34]. For many years

* Corresponding author.
E-mail addresses: bartosz.krawczyk@pwr.edu.pl (B. Krawczyk), michal.wozniak@pwr.edu.pl (M. Woźniak).

http://dx.doi.org/10.1016/j.neucom.2016.02.040
0925-2312/

voting algorithms were the ones most popular. However, for more complex data one requires flexible approaches that can adjust their combination methods to the properties of analyzed datasets. Trained combiners have gained a significant attention of the machine learning community [45]. However, such aggregation methods require additional training time and access to separate subset of examples—and once again this requirements can become prohibitive in certain application areas.

In this work, we introduce a novel method for weighted combination with embedded ensemble pruning step based on our previous works on efficient, fast and simple combination rules [26]. We propose novel weighted aggregation operators which do not require learning and have embedded pruning procedure that do not require any criterion to work. We work on modification of two popular operators: average of supports and maximum of supports. Their main drawback lies in lack of robustness to weak and irrelevant classifiers, and in minimizing the influence of other ensemble members. By using a Gaussian function to estimate the weights for the entire ensemble, we achieve a smooth method for reducing, but not eliminating the influence of weaker classifiers. At the same time by adjusting a threshold on the value of weights, we are able to prune the ensemble by discarding incompetent learners.

The main contributions of this paper are given below:

1. We introduce two novel untrained combination rules for forming efficient classifier ensembles. They use continuous outputs of base classifiers (support functions) and are based on popular *maximum* and *average* operators. Our methods provide a more robust combination, as they use a Gaussian function to assign weights to each base classifier. Therefore, we are able to control the level of influence of each classifier on the combination procedure.
2. We propose an unsupervised methodology for calculating weights for base classifiers in the ensemble. This way we are able to apply a weighted combination scheme, where weights assigned to each classifier are based on the individual model and class number without a need for an external validation set. This allows us to boost the classifier's influence over the classes where it is most competent, while reducing its role for classes that cannot be properly recognized by it. It is a highly suitable solution for scenarios where we do not have abundant data and cannot afford to use some of them to form a separate combiner training set. Additionally, this allows us to efficiently mine datasets with a large number of classes.
3. We propose to embed an unsupervised ensemble pruning step within the combination operators. It is based on thresholding the weights assigned to models and discarding classifiers with lowest weights assigned (i.e., non-competent ones). It does not require any external procedure or extensive computational effort to perform and is able to significantly reduce the size of the committee. This is highly suitable for scenarios with limited computational resources.

With the use of a wide selection of benchmark datasets with large number of classes we show the quality of proposed combiners. We present the results of pruning step that indicates the possibility of creating smaller, but efficient ensembles with our methods. Comparison with state-of-the-art untrained and trained combiners is backed-up with a rigorous statistical analysis that further proves the usefulness of the proposed ensemble fusion algorithms.

The remaining parts of this manuscript are organized as follows. Next section described the background and advantages of ensemble systems in machine learning. Section 3 presents a detailed description of our weighting method together with the embedded pruning mechanism. Section 4 depicts the set-up, datasets and experimental analysis. Section 5 summarizes the main findings, while the final section concludes this manuscript.

2. Classifier ensembles in machine learning

Classifier ensembles concentrate on the problem of efficient exploitation of different classifiers available for a considered recognition problem, believing that utilizing more than one single model can be beneficial for the formed system. This concept was first presented by Chow [6], who proved that the decision of independent classifiers with appropriately defined weights is optimal.

Let us now present some advantages of an MCS:

- The design of an MCS can be seen as following similar steps as the design of a canonical pattern recognition system [17]. In the standard approach, we concentrate on selecting the most informative features and choosing the best classification algorithm from the set of available ones. When forming a classifier ensemble, we aim to create a set of mutually complementary and individually accurate classifiers and assign an appropriate combination method, which can most efficiently combine their individual decisions [32].
- One may find numerous literature reports stating that MCSs are able to improve the overall performance when compared with the best individual classifier from the pool. This happens because they are able to exploit unique strengths of each of the individual classifiers. In some cases (e.g., when a majority voting is applied on a group of independent classifiers) the characteristics have been proven in an analytical way [40]. Additionally, an MCS protects against the selection of the worst classifier, when we have only a small training sample at our disposal [33].
- One should notice that some machine learning algorithms (e.g., C4.5 based on a top down induction decision tree concept) are de facto heuristic search algorithms. For them it is not guaranteed that the best possible model for a given dataset is found. One may alleviate this by the combined approach, which would start simultaneously searching from different points of the search space.
- Combined classifiers could be easily used in high-speed computing environments such as parallel and multithreaded computer architectures [8]. Another attractive area of application is distributed computing systems (P2P, GRID) [24], especially in the case of sensor networks [37] or multi-source datasets.

When designing an MCS one should take into consideration a number of important issues that can be grouped into the following problems:

1. How to select a pool of diverse and complementary individual classifiers for the ensemble?
2. How to design a combination rule that can exploit the strengths of the selected classifiers and combine their outputs optimally?
3. How to propose a suitable topology for a given system i.e., interconnections among classifiers in the ensemble.

We do not address the last issue because most of the combined classifiers are based on a parallel topology, which has a good methodological background [29] and is used in this work.

When selecting members to the committee one should assure that they work on different principles or utilize different components/data subsets. Apart from increasing the computational complexity, combining similar classifiers should not contribute much to the MCS under construction. An ideal ensemble consists