Information Fusion 31 (2016) 26–42

Contents lists available at ScienceDirect

Information Fusion

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

# Improved classification with allocation method and multiple classifiers

Sašo Karakatič*, Vili Podgorelec

*Faculty of Electrical Engineering and Computer Science, Institute of Informatics, University of Maribor, Slovenia*

ARTICLE INFO

*Article history:*
Received 14 May 2015
Revised 15 December 2015
Accepted 19 December 2015
Available online 29 December 2015

*Keywords:*
Classification
Ensemble methods
Allocation
Anomaly detection

ABSTRACT

Classification is the most used supervized machine learning method. As each of the many existing classification algorithms can perform poorly on some data, different attempts have arisen to improve the original algorithms by combining them. Some of the best know results are produced by ensemble methods, like bagging or boosting. We developed a new ensemble method called allocation. Allocation method uses the allocator, an algorithm that separates the data instances based on anomaly detection and allocates them to one of the micro classifiers, built with the existing classification algorithms on a subset of training data. The outputs of micro classifiers are then fused together into one final classification. Our goal was to improve the results of original classifiers with this new allocation method and to compare the classification results with existing ensemble methods. The allocation method was tested on 30 benchmark datasets and was used with six well known basic classification algorithms (J48, NaiveBayes, IBk, SMO, OneR and NBTree). The obtained results were compared to those of the basic classifiers as well as other ensemble methods (bagging, MultiBoost and AdaBoost). Results show that our allocation method is superior to basic classifiers and also to tested ensembles in classification accuracy and *f*-score. The conducted statistical analysis, when all of the used classification algorithms are considered, confirmed that our allocation method performs significantly better both in classification accuracy and *f*-score. Although the differences are not significant for each of the used basic classifier alone, the allocation method achieved the biggest improvements on all six basic classification algorithms. In this manner, allocation method proved to be a competitive ensemble method for classification that can be used with various classification algorithms and can possibly outperform other ensembles on different types of data.



## 1. Introduction

Classification is a technique where an algorithm learns a classification model on a given dataset of labeled instances of how to predict the labels (or classes) of new, unseen (not yet labeled) instances. Many different classification approaches have been proposed and used for solving real life problems, ranging from statistical methods to machine learning techniques as linear classifiers (Naive Bayes classifier and logistic regression), distance estimations (*k*-nearest neighbours), support vector machines, rule and decision tree based methods, and the neural networks, to name a few. The prediction performance of these basic methods can be improved significantly with the use of ensembles – a committee of models, where several models collaborate in classification process with the goal of more stable and more accurate classification results. This process is called *ensemble learning* and has been studied extensively [5,8,30] and [36] with both empirical [18,30,33] and theoretical [16,28] evidence that combination of classifiers can produce superior results in comparison to basic classifiers alone. Most used and more popular ensemble methods are following: boosting methods AdaBoost [11] and its extension MultiBoost [48], bagging [3], arbitrating [34], stacked generalization [49], delegation [9], averaged decision trees [23] and weighting of individual classifiers [46]. The nature of such methods is that the final classifier is combined from multiple basic classifiers, where each one of them is constructed on specific settings or subset of instances and none of them has a full set of information about the problem.

We propose a novel ensemble method of *allocation* where multiple micro classifiers are used and together they form a large macro classifier. As with previously mentioned ensemble methods, where numerous classifiers get each of their own set of instances to be classified, here also each micro classifier gets its set of instances, but in contrast to other methods – sets of instances are not chosen at random and sets do not overlap over each other. In the beginning of proposed allocation technique, the allocation method splits an original dataset of instances and allocates each of these sets to one specialized classifier, which is trained to deal with such instances. As opposed to delegation method where the classifier alone determines when it is no longer capable of classification of particular instance and thus delegates them to another

* Corresponding author. Tel.: +386 31869109.
*E-mail address:* karakatic@gmail.com, saso.karakatic@um.si (S. Karakatič).

http://dx.doi.org/10.1016/j.inffus.2015.12.006
1566-2535/

classifier, our method eliminates this need where classifiers must be modified so that they can learn when to delegate. Instead, it introduces the allocator that reviews instances and allocates them to one of the specialized micro classifiers. Each micro classifier works independently from the allocator and other micro classifiers and do not need any modification – any existing classification algorithm can be used. Allocation method is thus a parallel method, where no sampling is done and can efficiently deal with bi-class and multi-class instances. The type of classifier plays no role in this process as will be shown later when we analyze allocation method on several classification algorithms. The key point in the process is the splitting of original data in adequate sets. In this paper we propose the allocation method with two micro classifiers and the allocator based on one-class support vector machine method (SVM) used as anomaly detection method. In general, other methods and arrangements (more than two micro classifiers) could be used for allocation, but were not tested and are out of scope of this paper.

Related methods have been reviewed in [15] and [50], ranging from the gating networks where mixed expert method is used based on the probabilities of density estimations [21] to mixtures of experts used with the expectation maximization algorithm [24]. In [14] researchers used the subspace projections to construct ensembles of classifiers. Similar research was done in [47] where researchers used the partitioning of the features to form different subspaces, which were used in training of ensembles. In contrast to our method, these methods use some kind of randomized splitting of the initial dataset, whereas our allocation method uses classification method (in the case of this paper the one-class SVM is used) for non-random splitting of the instances and therefore enabling micro classifiers to work with specific type of data. In [27] authors use one-class classification, which is also used as the allocator in our approach, but they combine multiple one-class models to classify the instances. Allocation method uses the one-class anomaly detection classification only to divide the initial dataset and continues to use other classification algorithms for inducing micro classifiers. One-class SVM method, similar to our implementation of allocation method, is used in anomaly detection system [45], where the authors used multiple SVM models to remove normal data and identify the outliers.

A number of improvements in boosting algorithms have been made, from previously mentioned AdaBoost and MultiBoost to other boosting methods where some kind of oversampling algorithm is used. RUSBoost [40] and its newer variant EUSBoost [13] use oversampling in order to optimize performance on skewed and highly imbalanced datasets. Allocation method does not use any over- or under-sampling and so does not reshape the data with artificial data instances to construct the classification model. Besides sampling, allocation method is not an iterative process, where classification methods are improved through steps. Micro classifiers in allocation method work independently of each other and do not vote for final classification decision; they rather specialize in one type of instances and full trust is given in their sole decision. The biggest difference of allocation method with regard to ensemble methods is in the fact, that each new instance is always classified only by one micro classifier, determined by allocation model. In this regard, the allocation method might not be considered as a classical ensemble (where each classifier contributes to decision on each instance), but rather as a specialization (where each classifier is solely responsible for its own subset of instances).

The contributions of this work are as follows:

- We propose a new allocation method for building a simple ensemble of two basic classifiers and one allocator. The allocator is based on one-class SVM for anomaly detection; it splits an original dataset into two disjoint subsets, of which each one is used to train one micro classifier using any of the existing classification algorithms.
- The allocation method proposes an easy-to-use approach, which does not require any further configuration or parameter setting from the user and works on any classification dataset. The allocation method thus ensures a simple and universal, yet competitive and scalable classification approach which provides very good classification performance.
- We perform extensive computational tests on a diverse set of benchmark datasets that demonstrate the strength of the proposed allocation method and show that it outperforms basic classifiers as well as standard classification ensemble methods.
- We discuss the possibility of extending our allocation method using different allocation approaches and algorithms, combinations of micro classifiers, and more than two micro classifiers in the ensemble.

The goal of our research was to test whether our technique of allocation is a valid method for solving classification problems and how it performs on different datasets and using different classification algorithms. Although partitioning of a dataset is quite a standard strategy for improving results on almost any specific domain, we wanted to show how the proposed allocation approach, which is a universal method independent of a domain, performs in general over multiple datasets from several domains and using several basic classifiers. The experiment was conducted where allocation method was tested on 30 standard benchmark datasets with the use of six well known classification algorithms used to build the micro classifiers. Results are compared to other popular ensemble techniques.

The remaining of this paper is organized as follows. Second chapter starts with review of similar methods where we compare the allocation method with the existing research body of ensemble methods. Next is the description of one-class SVM for anomaly detection, which is an important part of allocation method. After that the description and basic formulation of allocation method is presented. In third chapter we describe the layout of the experiment and present the results supplemented with the statistical analysis, and discussion. In the end we conclude with final remarks about allocation method and results of the experiment, and propose a development for future research.

## 2. Proposed improved classification with allocation method

Many researches indicate that removing outlier instances before the process of classification can significantly improve the classification model. The process of outlier removal was done in combination with traditional classification algorithms to achieve better classification accuracy [41]. Identifying and removing potentially problematic instances has also been done with PRISM filtering method and produced statistically significant improvements in classification metrics [43]. In another research authors used only particular data in classification process to improve classification of music genres [29]. In [44] authors used instance based learning algorithms and proved that datasets reduced of noise can lead to better classification models [44]. Recent survey additionally confirms that classification metrics suffer with the presence of noise in the dataset and so data instances should be chosen to produce the best performing models [10]. Another review overviews the instance selection methods and their contribution to the classification metrics improvements [32]. The improvements that these outliers removal methods made makes a case for their usage in the classification process.

Few researches take this approach to another level, with the theory that too large datasets do not perform well in construction of classification models, mainly for the same reasons as