Neurocomputing 138 (2014) 347–357

Contents lists available at ScienceDirect

Neurocomputing

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

Integrating complementary techniques for promoting diversity in classifier ensembles: A systematic study

Diego S.C. Nascimento [a,c], André L.V. Coelho [b,*], Anne M.P. Canuto [c]

[a] Federal Institute of Rio Grande do Norte, Brazil
[b] Graduate Program in Applied Informatics, Center of Technological Sciences, University of Fortaleza, Brazil
[c] Department of Informatics and Applied Mathematics, Federal University of Rio Grande do Norte, Brazil

ARTICLE INFO

Article history:
Received 22 January 2013
Received in revised form
21 July 2013
Accepted 23 January 2014
Communicated by A. Abraham
Available online 20 February 2014

Keywords:
Classifier ensembles
Diversity
Feature selection
Bagging
Heterogeneous models

ABSTRACT

Various studies have provided theoretical and empirical evidence that diversity is a key factor for yielding satisfactory accuracy-generalization performance with classifier ensembles. As a consequence, in the last years, several approaches for boosting reasonable levels of diversity have been investigated, ranging from the use of data resampling techniques to the use of different types of classifiers as ensemble components. However, little work has been pursued on the combination of diversity-promoting techniques into a single conceptual framework. The aim of this paper is thus to empirically assess the impact of using, in a sequential manner, three complementary approaches for enhancing diversity in classifier ensembles. For this purpose, simulations were conducted on 15 well-known classification problems with ensemble models composed of up to 10 different types of classifiers. Overall, the results evidence the usefulness of the proposed integrative strategy in incrementing the levels of diversity progressively.

1. Introduction

In a typical ensemble setting, each input pattern is handled redundantly by different classification modules (ensemble components or, simply, components), which separately produce their estimates of the class label [1,2]. These individual results are then fused by a combination module, which is responsible for producing the final ensemble decision. Usually, for classification tasks, simple methods, such as the majority voting, can be employed as combination module [1,3].

As already pointed out by Kuncheva [1,4], the diversity among the decisions produced by the components seems to be a key factor in order to achieve high levels of accuracy in ensembles. In other words, there is no gain in combining identical components and, therefore, diversity is an important issue to take into consideration when designing ensemble models.

In this work, we report on an empirical study investigating an integrative approach for enhancing diversity in classifier ensembles. The main idea is to combine, in an incremental way, complementary techniques (namely feature selection, data resampling, and heterogeneous architectures) to increase the diversity level of these systems and to reduce their generalization error. The main aim of this analysis is thus to observe the performance (in terms of accuracy and diversity) of the resulting ensemble models when we incrementally introduce techniques for promoting diversity.

The rest of the paper is structured as follows. In Section 2, we briefly survey some studies related to classifier ensembles, while a general description of the role of diversity in this context is presented in Section 3. We describe the three stages of the proposed integrative approach for incrementing the ensemble diversity gradually in Section 4. Then, in Section 5, we outline the way the computational experiments were set up, whereas in Section 6, the results obtained by the two first stages and by the whole approach are presented and discussed, taking as baseline the levels of diversity and generalization error delivered by standard Bagging models. Finally, Section 7 concludes the paper and provides remarks on future work.

2. Related work

Theoretical and empirical studies have been widely conducted on ensemble systems [1,5–11]. In particular, different strategies are now available aiming at enhancing diversity in ensembles [12,3,10,11,7,13], ranging from those manipulating the training data to those that operate on the architectural setting [14,15].

* Principal corresponding author. Tel.: +558534773268; fax:+558534773061.
E-mail addresses: diego.nascimento@ifrn.edu.br (D.S.C. Nascimento), acoelho@unifor.br (A.L.V. Coelho), anne@dimap.ufrn.br (A.M.P. Canuto).

http://dx.doi.org/10.1016/j.neucom.2014.01.027
0925-2312

For the first case, in [10], the authors presented an approach, called Bagging++, which selects heterogeneous structures of classifier ensembles. In general, the authors pointed out that the combination of different techniques for enhancing diversity (resampling and heterogeneous components) in Bagging obtained better accuracy levels, when compared with the standard Bagging. They also proposed an extension in [11] in which a fifth classification algorithm and a pruning process were included to eliminate the redundant components. Nonetheless, these studies did not use any feature selection procedure. On the other hand, Ho [7] investigated the idea of randomly selecting features for ensembles composed of decision trees (C4.5), making use of the Boosting method.

In summary, there are already several studies investigating different approaches for enhancing the levels of diversity among components in ensemble systems. However, very little has been done on the combination of complementary techniques for enhancing diversity in order to increase the performance of the resulting models even further. Unlike the studies reviewed, this paper proposes and deeply assesses an integrative approach using three complementary techniques (namely feature selection, resampling, and heterogeneous structure) to enhance diversity in ensembles progressively. This is the main contribution of this paper. In the next section, we will describe the role of diversity in ensembles, focusing on different mechanisms that can promote diversity.

3. Diversity in classifier ensembles

An ensemble system is composed of a set of N individual classifiers (ICs), organized in a parallel way, that receive the input patterns and send their output to a combination module, which in turn is responsible for providing the final output of the system. The use of ensembles can lead to an increase in the processing time required to perform the classification task, since ensemble models are self-evidently more complex than single classifiers [1]. Because of this, the use of ensembles generally has to be well justified in order to counterbalance their increased complexities.

According to Dietterich [15], it is possible to have a good approximation of the best classifier available in a pool of classifiers by simply averaging out the output of the different classifiers. The new classifier might not be better than the single best classifier, but this strategy can eliminate or at least decrease the risk of selecting an inadequate single classifier. Fig. 1, adapted from [15], gives a graphical illustration of this argument. This figure depicts two possible scenarios over a classification space H, which is composed of four hypotheses (classifiers), namely h_1, h_2, h_3, and h_4. These classifiers can be also regarded as functions operating on the input data in such a way that $h_i = f_i(\mathbf{x})$ should be interpreted as the application of the function f_i on the input data $\mathbf{x}$. The best classifier for this problem is denoted by $h = f$.

In Fig. 1, the outer region denotes the space of all possible classifiers, while the light gray region contains all classifiers with good performance on the training data. In this figure, the level of diversity among the classifiers is directly related to their positions on the space. In other words, the more diverse the two classifiers are, the more distant they will be in the classification space. The exact position of a classifier h_i depends on the associated hypothesis f_i (classification algorithm or parameter setting) as well as the training data, $\mathbf{x}$. In addition, the final outcome of an ensemble system can be obtained through a combination of the points (classifiers) of this figure.

In Fig. 1(a), the individual classifiers (h_1, h_2, h_3, h_4) are far from each other. In this case, the simple linear combination of their outputs can be associated with a large region within the classification space, possibly yielding a classifier that is closer to the ideal classifier. In contrast, when working with classifiers with low diversity, their hypotheses tend to be located very close in the inner dark gray region, as illustrated in Fig. 1(b). In this case, any type of aggregation will produce classifiers in a small part of the good classifiers region (dark region inside the inner region), leading to more difficulties in yielding classifiers close to the ideal one.

Diversity in classifier ensembles can be promoted through different conceptual mechanisms, such as the following ones [15]:

- *Different parameter settings of the classifiers*: According to this mechanism, diversity can be reached through the use of different initial parameter settings of the adopted classification method. In relation to Fig. 1, this means a change in the functions f_i associated to the hypotheses. These different parameter configurations may serve as different initial search biases while seeking the final hypothesis. Since the classifier training process is usually heuristic in nature, adopting this mechanism may lead to a diverse set of hypotheses associated with different positions of h_i in the classification space, in an analogy to Fig. 1. In other words, we can produce a set of classifiers that are diverse.
- *Different classification algorithms*: According to this mechanism, diversity can be promoted through the use of different types of classifiers. As in the previous case, this should change the function f_i associated to each hypothesis. That is, since different

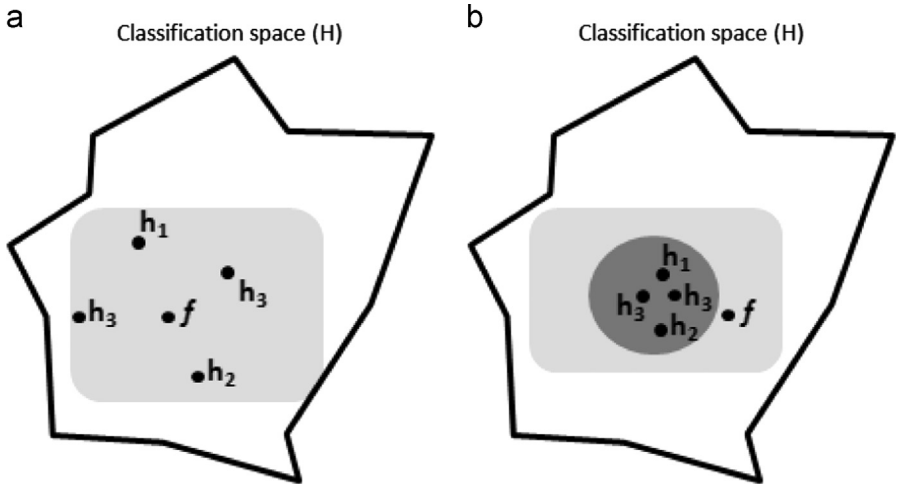

Fig. 1. Statistical reasons for combining classifiers – adapted from [15].