ARTICLE IN PRESS

Pattern Recognition ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Pattern Recognition

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

Dynamic selection of classifiers—A comprehensive review

Alceu S. Britto Jr.[a,b,*], Robert Sabourin[c], Luiz E.S. Oliveira[d]

[a] Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, PR, Brazil
[b] Universidade Estadual de Ponta Grossa (UEPG), Ponta Grossa, PR, Brazil
[c] École de technologie supérieure (ÉTS), Université du Québec, Montreal, QC, Canada
[d] Universidade Federal do Paraná (UFPR), Curitiba, PR, Brazil

ARTICLE INFO

Article history:
Received 28 August 2013
Received in revised form
3 May 2014
Accepted 7 May 2014

Keywords:
Ensemble of classifiers
Dynamic selection of classifiers
Data complexity

ABSTRACT

This work presents a literature review of multiple classifier systems based on the dynamic selection of classifiers. First, it briefly reviews some basic concepts and definitions related to such a classification approach and then it presents the state of the art organized according to a proposed taxonomy. In addition, a two-step analysis is applied to the results of the main methods reported in the literature, considering different classification problems. The first step is based on statistical analyses of the significance of these results. The idea is to figure out the problems for which a significant contribution can be observed in terms of classification performance by using a dynamic selection approach. The second step, based on data complexity measures, is used to investigate whether or not a relation exists between the possible performance contribution and the complexity of the classification problem. From this comprehensive study, we observed that, for some classification problems, the performance contribution of the dynamic selection approach is statistically significant when compared to that of a single-based classifier. In addition, we found evidence of a relation between the observed performance contribution and the complexity of the classification problem. These observations allow us to suggest, from the classification problem complexity, that further work should be done to predict whether or not to use a dynamic selection approach.

1. Introduction

Classification is a fundamental task in Pattern Recognition, which is the main reason why the past few decades have seen a vast number of research projects devoted to classification methods applied to different fields of the human activity. Although the methods available in the literature may differ in many respects, the latest research results lead to a common conclusion; creating a monolithic classifier to cover all the variability inherent to most pattern recognition problems is somewhat unfeasible.

With this in mind, many researchers have focused on Multiple Classifier Systems (MCSs), and consequently, many new solutions have been dedicated to each of the three possible MCS phases: (a) generation, (b) selection, and (c) integration, which are represented in Fig. 1. In the first phase, a pool of classifiers is generated; in the second phase, one or a subset of these classifiers is selected, while in the last phase, a final decision is made based on the prediction(s) of the selected classifier(s). It is worth noting that such a representation is not unique, since the selection and integration phases may be facultative. For instance, one may find MCS where the whole pool of classifiers is used without any selection or systems where just one classifier is selected from the pool, making the integration phase unnecessary.

In a nutshell, recent contributions with respect to the first phase indicate that the most promising direction is to generate a pool of accurate and diverse classifiers. The authors in [1] state that a necessary and sufficient condition for an ensemble of classifiers to be more accurate than any of its individual members is for the classifiers to be accurate and diverse. Dietterich [2] explains that an accurate classifier has an error rate lower than the random guessing on new samples, while two classifiers are diverse if they make different errors on new samples. The rationale behind this is that the individual accurate classifiers in the pool may compete each other by making different and perhaps complementary errors. As for the selection phase, interesting results have been obtained by selecting specific classifiers for each test pattern, which characterizes a dynamic selection of classifiers, instead of using the same classifier for all of them (static selection). Moreover, additional contributions have been observed when ensembles are selected instead of just one single classifier. In such a case,

* Corresponding author at: Post-graduate Program in Informatics (PPGIa), Pontifical Catholic University of Parana Rua Imaculada Conceição, 1155, Curitiba (PR), 80215-901, Brazil. Tel.: +55 41 3271 1669; fax: +55 41 3271 2121.
E-mail addresses: alceu@ppgia.pucpr.br (A.S. Britto Jr.), robert.sabourin@etsmtl.ca (R. Sabourin), lesoliveira@inf.ufpr.br (L.E.S. Oliveira).

http://dx.doi.org/10.1016/j.patcog.2014.05.003
0031-3203/

Fig. 1. The possible phases of a Multiple Classifier System.

the outputs of the selected classifiers must be combined and the third phase of the MCS is necessary. The main contributions for this phase have been comprised of different strategies combining the classifiers and the assumption that the best integration choice is usually problem depended.

The focus of this paper is on the second phase of an MCS, particularly, the approaches based on dynamic selection (DS) of classifiers or ensembles of such classifiers. Despite the large number of DS methods available in the literature, there is no comprehensive study available to those wishing to explore the advantages of using such an approach. In addition, due to the high computational cost usually observed in the DS solutions, its application is often criticized. In fact, the decision as to whether or not to use DS is still an open question.

In this scenario, we have three research questions, namely:

1. Are the performance results of the DS methods reported in the literature significantly better than those obtained by a single-based classifier approach?
2. Is there any relation between the classification complexity and the observed DS performance for a given problem?
3. Can we predict whether or not DS should be used for a given classification problem?

To answer these questions, we have reviewed several works on dynamic selection and performed a thorough statistical analysis of the results reported in the literature for different classification problems.

The motivation for investigating the possible existence of a relation between the DS contribution and the complexity of a classification problem is inspired by previous works in which the data complexity is used to better define the classifier models. An interesting work in this vein is presented in [3], in which the authors use geometrical characteristics of data to determine the classifier models. Two other interesting studies are presented in [4,5], where the authors characterize the behavior of a specific classifier approach considering problems with different complexities.

With this in mind, our contribution is two-fold that (a) presents a comprehensive review of the main DS methods available in the literature, providing a taxonomy for them and (b) performs a further analysis of the DS results reported in the literature to determine when to apply DS.

This paper is organized as follows. After this brief introduction, Section 2 presents the main basic concepts and definitions related to the dynamic selection of classifiers. Section 3 presents the state of the art of DS methods and describes the suggested taxonomy. The algorithms of some key examples of each category are presented based on the same notation to facilitate comprehension. Section 4 presents further analysis of the DS results available in the literature, in a bid to answer our research questions. Finally, Section 5 presents the conclusions and further works.

2. Basic concepts and definitions

This section presents the main concepts related to MCS and DS approaches, which represent the necessary background for the comprehension of the different works available in the literature.

The first concepts are related to the generation phase of the MCS. As described earlier, this first phase is responsible for the generation of a pool of base classifiers, considering a given strategy, to create diverse and accurate experts. A pool may be composed of homogeneous classifiers (same base classifiers) or heterogeneous classifiers (different base classifiers). In both cases, some diversity is expected. The idea is to generate classifiers that make different mistakes, and consequently, show some degree of complementarity. A comprehensive study of different diversity measures may be found in the work of Kuncheva and Whitaker [6]. The schemes to provide diversity are categorized in [7] as implicit, when there is no use of diversity measures during the generation process, or as explicit, in opposite cases.

In homogeneous pools, diversity is achieved by varying the information used to construct their elements, such as changing the initial parameters, using different subsets of training data (Bagging [8], Boosting [9]), or using different feature subspaces (Random Subspace Selection [10]). On the other hand, the basic idea behind heterogeneous pools is to obtain experts that differ in terms of the properties and concepts on which they are based.

Regarding the selection phase of an MCS, the main concepts are related to the type of selection and the notion of classifier competence. The type of selection may be static or dynamic, as explained earlier. The rationale behind the preference for dynamic over static selection is to select the most locally accurate classifiers for each unknown pattern. Both static and dynamic schemes may be devoted to classifier selection, providing a single classifier, or to ensemble selection, selecting a subset of classifiers from the pool.

Usually, the selection is done by estimating the competence of the classifiers available in the pool on local regions of the feature space. To that end, a partitioning process is commonly used during the training or testing phases of the MCS. In this process, the feature space is divided into different partitions, and the most capable classifiers for each of them are determined. In static selection methods, the partitioning is usually based on clustering or evolutionary algorithms, and it is executed during the training phase. This means that the classifier competence is always determined during the training phase of the system. Although it is possible to apply similar strategies for dynamic selection methods, what is mostly commonly seen with this approach is the use of a partitioning scheme based on the NN-rule to define the neighborhood of the unknown pattern in the feature space during the testing phase. In this case, the competence of each classifier is defined on a local region on the entire feature space represented by the training or validation dataset.

Regarding the competence measures, the literature reports several of them, which consider the classifiers either individually or in groups. This is the basis of the DS taxonomy proposed in the next section. It is worth noting that, basically, the individual-based measures most often take into account the classifier accuracy. However, the measures are conducted in different ways. For instance, one may find measures based on pure accuracy (overall local accuracy or local class accuracy) [11], ranking of classifiers [12], probabilistic information [13,14], classifier behavior calculated on output profiles [15–17], and oracle information [18,19]. Moreover, we may find measures that consider interactions among classifiers, such as diversity [20–22], ambiguity [23,24,17] or other grouping approaches [25].

The third phase of an MCS consists in applying the selected classifiers to recognize a given testing pattern. In cases where all classifiers are used (without selection) or when an ensemble is selected, a fusion strategy is necessary. For the integration of the classifier outputs, there are different schemes available in the literature. Complete details regarding the combination methods and their taxonomy are available in Jain et al. [26] and in Kittler et al. [27].

Please cite this article as: A.S. Britto Jr. et al., Dynamic selection of classifiers—A comprehensive review, Pattern Recognition (2014), http://dx.doi.org/10.1016/j.patcog.2014.05.003