Classifiers ensemble selection

Mariano Ferrero
Universidad Politécnica de Madrid, Spain

Abstract

When a ensemble of classifiers is generated it has been shown that pruning it into a small number has better results. This article presents a review of some of the existing classifiers selection methods, both static and dynamic.

KEY WORDS: Meta-learning; Ensemble of classifiers; Dynamic ensemble selection; Static ensemble selection

1 Introduction

Within the area of machine learning, one of the most covered topics in recent decades is classification. While the technical proposals and approaches differ from each other, the idea of using a single classifier to cover all the diversity that a specific problem may contain cannot be deemed as something honest in most cases.

Because of this, various Multiple Classifier Systems (MCS) have been proposed in recent years. These systems consist of three potential phases: generation, selection and integration. The first one is based on generating a set of classifiers, being some of the most well known techniques Bagging (Breiman (1996)) and Boosting (Freund and Schapire (1996)). In the selection phase a filter on the classifiers generated set is applied, and finally, in the integration phase, predictions made by selected classifiers are combined in some way to produce a single output.

Selection phase is based on what is mentioned in the article by Zhou ***et al.*** (2002), which arises that after generating a set of classifiers, it is convenient to use a subset of them above the use of the set as a whole. For this reason, this work intends to conduct a review of some of the techniques and methods proposed so far for this phase.

2 State of the art

The selection of a group of base classifiers may be out in a static or dynamic way, depending on whether a same set of classifiers are used for all instances unclassified (static) or a subset is obtained specifically for each of these instances (dynamic). The following are some of the existing proposals of each of these two categories.

2.1 Static Ensemble Selection

Since the selection of classifiers may pose as an optimization problem, where the goal is to find the combination of classifiers that together produce the best result, Kim and Oh (2008) propose using a hybrid genetic algorithm with binary encoding. Each individual's chromosomes represent a base-classifier, indicating that it should be included or not with values of 1 and 0 respectively. It is called hybrid because two proposed local search methods that seek to improve the produced offspring are used after mutation is applied.

Another distinct proposal is to use Q statistic to focus the selection (Yang (2011)), taking into account not only precision but also diversity of results. For this, the classifier which obtains the best performance is selected and then those classifiers that differ to a greater extent, using the statistical, are also chosen.

A prior but more general work to aforementioned articles was developed by Ruta and Gabrys (2005). This article presents several experiments using various selection criterions (including Q statistics) and search algorithms (genetic algorithms included), concluding that in this case best results were obtained by using Majority Vote Error (MVE) as selection criterion and Forward Search (FS) or Backward Search (BS) as search algorithms.

On the other hand, Lin ***et al.*** (2014) present a selection method based on partitional clustering, where individuals are represented by classifiers and distances among them is obtained from their diversity. Then, they use a strategy which allows to modify iteratively the number of considered classifiers according to a measure that take into account diversity and accuracy.

Although the techniques mentioned above represent different approaches, it is worth noting that the goal pursued is similar: find the set of classifiers which obtain better performance and, in turn, present greatest diversity.

2.2 Dynamic Ensemble Selection

The main goal of these techniques is trying to select the subset of classifiers that obtain a better result for each instance in the generalization phase.

One of the ideas used to address this issue is based on getting k-neighbors of the incoming instance and then select a subset of classifiers according to performance on these instances. Mousavi and Eftekhari (2015) presented a hybrid method that combines the previous idea with a pre-filter of classifiers using a genetic algorithm. Meanwhile, Ko ***et al.*** (2008) propose different methods, which are called KNORA (K-Nearest-ORAcles), using some or all of the classifiers depends on the number of hits on the k-neighbors.

On the other hand, some methods proposed a slightly extremist selection, delivering only one classifier per instance. Continuing with the idea previously mentioned, Giacinto and Roli (2001) proposed to use the classifier that get the better accuracy on the k-neighbors instances. Todorovski and Džeroski (2003) have defined Meta Decision Trees, which unlike the ordinary trees instead of represent labels in the leaves, they indicate the classifier that should be used.

Cruz ***et al.*** (2015) propose a method called META-DES. It consists in generate a number of base-classifiers using bagging. Then, a meta-classifier is trained, where each instance represents meta-features obtained from the training set for each base-classifier. In the generalization phase, the meta-features are obtained for each base-classifier, and the meta-classifier determines if it is taken into account in the final set or not. It is worth noting that some of these meta-features are generated from the same idea of obtaining k-neighbors.

Finally, one of the most interesting proposal is to tackle the problem as if it were a multilabel one, where each labels represent one base-classifier (Markatopoulou *et al.* (2015)). The key aspect about this proposal is that it allows to use many of the techniques presented to resolve this other type of problems.

3 Conclusions and future research

In this article different techniques were presented for static and dynamic classifiers selection, being static ones which has attracted most attention in recent years. It is worth noting that the proposals are diverse, even some of them using hybrid approaches that combine several existing techniques to address the problem. This realizes that new methods could emerge in the coming years using combinations that have not been taken into account yet.

Classifiers ensemble selection techniques provides, in most cases, best results when compared to the use of all of initially generetad classifiers, verifying Zhou idea.

While many of the proposed methods perform a comparison against similar techniques, experiments that will do the same with some of the largely used today techniques, such as bagging and boosting, were not found. Carry out this comparison would be beneficial as a comparison if the advantages of using a static or dynamic selection is appropriate, given that also presents a series of inherent disadvantages, as for example the computational cost associated with this task.

References

Breiman, B. (1996) Bagging Predictors. *Machine Learning*, 24, 123-140.

Cruz, R.M.O., Sabourin, R., Cavalcanti, G.D.C. and Ren, T.I. (2015) META-DES:A dynamic ensemble selection framework using meta-learning. *Pattern Recognition*, 48, 1925-1935.

Freund, Y. and Schapire, R.E. (1996) Experiments with a New Boosting Algorithm. In *Proceedings of the 13th International Conference on Machine Learning (ICML)*, Morgan Kaufmann, pp. 148-156.

Giacinto, G. and Roli, F. (2001) Dynamic classifier selection based on multiple classifier behaviour. *Pattern Recognition*, 34, 1879-1881.

Kim, Y. and Oh, I. (2008) Classifier ensemble selection using hybrid genetic algorithms. *Pattern Recognition Letters*, 29, 796-802.

Ko, A.H.R., Sabourin, R. and Souza Britto, A. (2008) From dynamic classifier selection to dynamic ensemble selection *Pattern Recognition*, 41, 1718-1731.

Lin, C., Chen, W., Qiu, C., Wu, Y., Krishnan, S. and Zou, Q. (2014) LibD3C: Ensemble classifiers with a clustering and dynamic selection strategy. *Neurocomputing*, 123, 424–435.

Markatopoulou, F., Tsoumakas, G. and Vlahavas, I. (2015) Dynamic ensemble pruning based on multi-label classification. *Neurocomputing*, 150, 501-512.

Mousavi, R. and Eftekhari, M. (2015) A new ensemble learning methodology based on hybridization of classifier ensemble selection approaches. *Applied Soft Computing*, 37, 652-666.

Ruta, D. and Gabrys, B. (2005) Classifier selection for majority voting. *Information Fusion*, 6, 63-81.

Todorovski, L. and Džeroski, S. (2003) Combining Classifiers with Meta Decision Trees. *Machine Learning*, 50, 223-249.

Yang, L. (2011) Classifiers selection for ensemble learning based on accuracy and diversity. *Procedia Engineering*, 15, 4266-4270.

Zhou, Z., Wu. J. and Tang, W. (2002) Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, 137, 239-263.