Using Bayesian Networks for Selecting Classifiers in GP Ensembles

Claudio De Stefano,
Francesco Fontanella,
Alessandra Scotto di Freca
Universitá di Cassino – ITALY
{destefano,fontanella,a.scotto}@unicas.it

Gianluigi Folino
ICAR, CNR – ITALY
folino@icar.cnr.it

ABSTRACT

In this paper we present a novel approach for combining GP-based ensembles by means of a Bayesian Network. The proposed system is able to effectively learn decision tree ensembles using two different strategies: decision trees ensembles are learned by means of boosted GP algorithm; the responses of the learned ensembles are combined using a Bayesian network, which also implements a selection strategy that reduces the size of the built ensembles.

Categories and Subject Descriptors: I.2.8 ARTIFICIAL INTELLIGENCE: Problem Solving, Control Methods, and Search

General Terms: Algorithms.

1. INTRODUCTION

In the last years, in order to further improve the performance of GP–based classification systems, ensemble techniques have been taken into account. They try to effectively combine the responses provided by several classifiers operating on the same feature space in order to improve the overall classification accuracy. A key issue is to ensure that classifiers in the ensemble be appropriately diverse, so as to avoid correlated errors. In fact, as the number of classifiers increases, it may happen that a correct classification provided by some classifiers is overturned by the convergence of other classifiers on the same wrong decision. This event is much more likely in case of highly correlated classifiers and may reduce the performance obtainable with any combination strategy.

In this paper, we present a new high performance classification system, based on a GP ensemble of classifiers, able to deal with large data sets and to maintain diversity among the classifiers. For this purpose, we built a two–module system that combines the BoostCGPC algorithm [2], which produces a high performing ensemble of decision tree classifiers, with the BN (Bayesan Network) based approach to perform classifier combination [1]. The proposed system exploits the advantages provided by both techniques and allows to strongly reduce the number of classifiers in the ensemble. More specifically, the diversity among the ensemble classifiers is achieved by following two different approaches: the boostCGPC evolves diverse classifiers (decision trees) by means of a boosting technique; the BN module evaluates classifiers diversity by estimating the statistical dependencies of the responses they provide. Such ability is used to select, among the classifiers provided by the BoostCGPC module, the minimum number of them required to effectively classify the data at hand. Moreover, the responses provided by the selected classifiers are effectively combined by means of a rule inferred by the BN module.

In order to assess the effectiveness of the proposed system, several experiments have been performed. The results have been compared with those obtained by the BoostCGPC approach using a weighted majority vote combining rule. Moreover, a diversity analysis of the selected classifiers has been carried out taking into account two diversity measures.

2. SYSTEM ARCHITECTURE

The proposed system consists of two main modules: the first one builds an ensemble of decision tree classifiers (experts) by means of the BoostCGPC algorithm (Fig. 1). The second one uses a BN combiner to implement the combining rule that produces the final output of the whole system (Fig. 2). More specifically, unknown samples are recognized using a two–step procedure: (i) the feature values describing the unknown sample are provided to each of the ensemble classifiers built by the BoostCGPC module; (ii) the set of responses produced is given in input to the BN module. Such module labels the sample with the most likely class, among those of the problem at hand, given the responses collected by the first module. Also the learning phase requires two steps. In the first step, the BoostCGPC module is trained using a data set containing labeled samples described by their feature values. This learning is carried out, by means of a boosting–based technique. In the second step, the re-

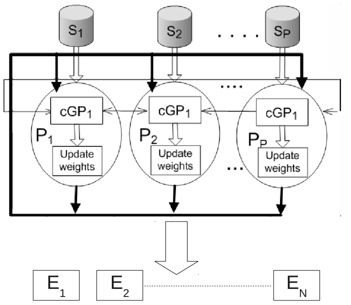


Figure 1: The boost module.

GECCO'11, July 12–16, 2011, Dublin, Ireland.
ACM 978-1-4503-0690-4/11/07.

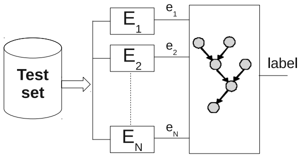


Figure 2: The BN module.

sponses provided by the set of decision trees built in the first step are used to learn the BN of the second module. The learned BN, given the set of responses concerning a sample, is able to estimate, for each class of the problem, the corresponding probability. Note that the BN is learned by means of a supervised procedure that requires to observe both the "true" class label c, and the set of responses provided by the classifiers for each training sample.

3. EXPERIMENTAL RESULTS

The proposed approach has been tested on five well-known data sets: *Census*, *Segment*, *Adult*, *Phoneme* and *Covtype*. The BoostGCPC module used standard GP parameters and a population of 100 individuals for node. The original training set has been partitioned among 5 nodes and respectively 5 and 10 rounds of boosting, with 100 generations for round, have been used to produce respectively 25 and 50 classifiers on 5 nodes. All results were obtained by averaging over 30 runs. The results achieved by our approach (hereafter BN-BoostCGPC) have been compared with those obtained by the BoostCGPC approach, which uses the wighted majority rule for combining the ensemble responses. The comparison results are shown in Tab. 1.
In order to statistically validate the comparison results, we performed the two–tailed t–test($\alpha = 0.05$) over the 30 carried out runs. The values in bold in the test set error columns highlight, for each ensemble, the results which are significantly better according to the two–tailed t–test. The proposed approach achieves better performance on the majority of the considered ensembles while, in the remaining cases, the performance are comparable. It is also worth noticing that the most significant improvements have been obtained on Adult, Census and Covtype data sets, which are the largest ones among those considered. Finally, it is worth to remark that the results of our system are always achieved by using only a small number of the available classifiers.

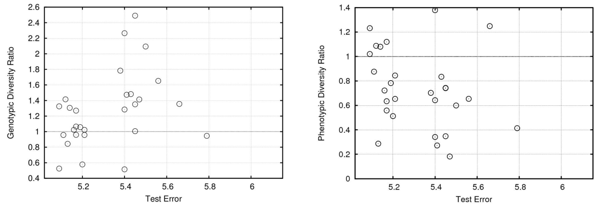
After assessing the effectiveness of proposed approach, we investigated the ability of the BN to achieve such performances by using only a very limited number of classifiers. To this aim, we studied the relationship between the diversity of the selected trees and the classification error. We adopted two metrics, the first considering the genotypic diversity in the ensembles and the other keeping into account the phenotypic diversity.

Table 1: Comparison results.

Datasets	**ens.**	**BoostCGPC**		**BN-BoostCGPC**	
		test err.	**#sel.**	**test err.**	**#sel.**
Adult	**25**	16.94	25	**16.28**	3.4
	50	18.23	50	**16.99**	3.8
Segment	**25**	12.69	25	**11.68**	2.9
	50	12.06	50	11.99	2.9
Phoneme	**25**	18.87	25	19.23	3.2
	50	20.04	50	19.51	7.8
Census	**25**	8.89	25	**5.27**	4.3
	50	9.07	50	**5.37**	3.9
Covtype	**25**	35.32	25	**33.44**	3.3
	50	33.76	50	33.65	6.2

The genotypic diversity we adopted evaluates the structural diversity between two trees. Given two trees, their genotypic diversity is computed by overlapping them at the root node and, recursively, for each pair of nodes at matching positions, the difference between the corresponding symbols is computed and combined in a weighted sum.

As phenotypic diversity measure, we used a disagreement measure, named *kappa statistics* (k). Such measure considers the class label outputs provided by the two classifiers to be compared and estimates the probability that they give the same responses. A value of $\kappa = 0$ indicates that the two classifiers are different, while a value of $\kappa = 1$ means that the two classifiers agree on each example.



(a) Genotypic (b) Phenotypic

Figure 3: Genotypic (top) and Phenotypic (bottom) diversity ratio vs. Test error for the Census dataset (25 trees).

In order to assess the diversity of the selected trees, with respect to the average diversity of all the trees of a given ensemble, we proceeded as follows. For each ensemble, we computed the two diversities for every couple of trees, then we computed the average diversity $\overline{d_a}$ over all the couples and the average diversity $\overline{d_s}$ of the selected trees. Finally, the *Diversity ratio* $d_r = \overline{d_s}/\overline{d_a}$ has been calculated. As a consequence, values of $d_r > 1$ (< 1) indicate that the trees selected by the BN present a greater genotypic (phenotypic) diversity than the average one of all the trees of the considered ensemble.

Figure 3 shows the diversity ratio versus the test error for the Census dataset, where each circle represents one of the 30 runs performed and a straight line ($d_r = 1$) separates the plotting area. From both the plots, it can be noted that in most of the cases the selected trees are, on average, more different each other, than all the couples of trees of the considered ensemble.

4. REFERENCES

[1] C. De Stefano, F. Fontanella, C. Marrocco, and A. S. di Freca. A hybrid evolutionary algorithm for bayesian networks learning: An application to classifier combination. In *EvoApplications (1)*, pages 221–230, 2010.

[2] G. Folino, C. Pizzuti, and G. Spezzano. Training distributed gp ensemble with a selective algorithm based on clustering and pruning for pattern classification. *IEEE Trans. Evolutionary Computation*, 12(4):458–468, 2008.