A Study on the Importance of Selection Pressure and Low Dimensional Weak Learners to Produce Robust Ensembles

Danilo V. Vargas Kyushu University, Japan vargas@cig.ees.kyushuu.ac.jp Hirotaka Takano Kyushu University, Japan takano@cig.ees.kyushuu.ac.jp Junichi Murata Kyushu University, Japan murata@cig.ees.kyushuu.ac.jp

ABSTRACT

Ensembles of classifiers have been studied for some time. It is widely known that weak learners should be accurate and diverse. However, in the real world there are many constraints and few have been said about the robustness of ensembles and how to develop it. In the context of random subspace methods, this paper addresses the question of developing ensembles to face problems under time constraints. Experiments show that selecting weak learners based on their accuracy can be used to create robust ensembles. Thus, the selection pressure in ensembles is a key technique to create not just effective ensembles but also robust ones. Moreover, the experiments motivate further research on ensembles made of low dimensional classifiers which achieve general accurate results.

Categories and Subject Descriptors

I.2.6 [Artificial Intelligence]: Learning

General Terms

Algorithms, Performance

Keywords

Evolution of Ensembles; Ensembles of Classifiers; Neural Networks; Robustness of Ensembles

1. INTRODUCTION

Ensembles of classifiers have been studied for some time motivated by the idea that a set of classifiers can be combined to reduce the variance and bias of the prediction [1].

In the real world, robustness is a very important characteristic because problems must be solved under various constraints and often unplanned problem characteristics. However, there are few results on the robustness of ensembles in the literature. Mainly the robustness against noise in the data [5], [3] and robustness against missing data were investigated [5].

This paper concerns with the random subspace method of constructing ensembles. Random subspace method (also called attribute bagging) is a technique of producing ensembles by subsampling without replacement from the set of features [4],[2]. In other words, in an ensemble created by

Copyright is held by the author/owner(s).

the random subspace method the number of features used by the classifiers are randomly chosen but kept always smaller than the total number of available features of the problem.

Experiments will be described in Section 2 while Section 3 will show the results and verify the effects of time constraints on two random subspace methods: one method with selection pressure and the other method without it. The absence of genetic operators makes the results general enough to be extended to most evolved ensembles with similar selection pressures.

2. EXPERIMENTS

This Section will describe the settings of the experiments. Both ensembles are composed of multilayer perceptrons using a majority voting mechanism. They differ in the fact that one ensemble will use a simple selection pressure (i.e. a few but accurate classifiers) while the other will use all the classifiers (more diverse and larger ensemble). Time constraints means the necessity of a quick real-time response with little learning time. It will be simulated with few iterations as well as few hidden nodes available.

The selection pressure consists of the simple selection of a number of classifiers, hereby called "number of predated", to be excluded for each evolutionary cycle (iteration). In the selection process the accuracy of the neural networks are compared on an unseen 20% of the training samples, using the remaining 80% of the training samples for the actual training. Excluded classifiers ("predated classifiers") are the ones which achieved the worst accuracy on the unseen 20% of the training dataset. The same samples were used to train all neural networks, therefore no increase in diversity is caused by the evolution. At the same time, the accuracy measured on the 20% of the training samples is not biased by the samples (they do not vary for all classifiers) and therefore the selection is also not dependent on the samples.

Random subspace methods normally use a uniform distribution to define the number of variables used by the classifiers. But much details of the method are lost this way (e.g., the information of weather low or high dimensional classifiers are more relevant). Experiments will plot an extensive curve of the behavior of ensembles with the same number of variables (the terms dimensions and number of variables will be used indistinctly). Notice that each classifier has its unique set of variables chosen by sampling without replacement from the set of variables. Every test will be repeated over 30 trials with different 80/20 train/test splits. Moreover, throughout the tests the parameters are kept the same. They are described in Table 1. The number of neural net-

GECCO'13 Companion, July 6–10, 2013, Amsterdam, The Netherlands. ACM 978-1-4503-1964-5/13/07.



Figure 1: Wine (left), Vowel (middle) and Segmentation (right) problems solved under time constraints.

Table 1: Parameters	
	Multilayer Perceptron
Initial weights	[-0.1, 0.1]
Weight decay	5e - 4
Hidden nodes	3
Output units	logistic function
	Evolutionary Ensemble
Number of models	10
Evolutionary cycles	20
Number of predated	1
	Ensemble
Number of models	30

work learning iterations are 20 for Wine and 100 for both Vowel and Segmentation problems.

3. DISCUSSION AND CONCLUSIONS

Figure 1 shows the experiments for three datasets of the UCI machine learning repository: Wine, Vowel and Segmentation. There is a better accuracy in all datasets for the evolved ensemble. Therefore, a few selected (more accurate) classifiers are better than a large number of diverse classifiers. Note that both methods evaluate the same number of weak learners (30 neural networks), however, the evolved one decides to exclude the least accurate while the other ensemble decides to keep all of the neural networks.

Moreover, ensembles made of low dimensional networks achieved in general higher accuracy. This fact demonstrates the importance of the low dimensional classifiers. In fact, although obvious to some, it is important to mention that low dimensional classifiers do not suffer from overfitting, underfitting, higher classifier complexity and steep learning time.

Ensembles made of low dimensional classifiers should not be confused with low dimensional ensembles. Probably, ensembles made of low dimensional classifiers use all dimensions though some are parametric (used explicitly inside the classifiers' model) while the majority of the dimensions are non-parametric (used as a consequence of the voting mechanism). To illustrate the difference between the use or not of the non-parametric dimensions, Figure 2 shows the evolved ensemble and the average accuracy of its weak learners.



Number of Variables per Model

Figure 2: Comparison between the accuracy of the evolved ensemble and the individuals alone at the Segmentation dataset under time constraints.

4. **REFERENCES**

- L. Breiman. Bagging predictors. Machine learning, 24(2):123–140, 1996.
- [2] R. Bryll, R. Gutierrez-Osuna, and F. Quek. Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern recognition*, 36(6):1291–1302, 2003.
- [3] B. He and K. C.-C. Chang. Making holistic schema matching robust: an ensemble approach. In *Proceedings* of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, pages 429–438. ACM, 2005.
- [4] T. K. Ho. The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(8):832–844, 1998.
- [5] P. Melville, N. Shah, L. Mihalkova, and R. Mooney. Experiments on ensembles with missing and noisy data. *Multiple Classifier Systems*, pages 293–302, 2004.