Learning Multiple Models Without Sacrificing Comprehensibility

Pedro Domingos[1]

Department of Information and Computer Science
University of California, Irvine
Irvine, California 92697, U.S.A.
pedrod@ics.uci.edu
http://www.ics.uci.edu/~pedrod

Learning multiple models and combining their results can often lead to significant accuracy improvements over the single "best" model. This area of research has recently received much attention (e.g., Chan, Stolfo, and Wolpert 1996). However, as Breiman (1996) notes, when the models being combined are "human-readable" (as is the case with, for example, decision trees and rule sets), the cost of this procedure is the loss of the comprehensibility afforded by the single model. Not only is the complexity of m models m times greater than that of one, but it is difficult and tedious for a human to predict the output of the model ensemble, and thus to understand its behavior. This can be a significant disadvantage, since comprehensibility is often of paramount importance to make the learner's output acceptable to the users, to allow interactive refinement of the model produced, and to gain knowledge of the domain. This extended abstract describes and evaluates a method for retaining most of the accuracy improvements obtained by multiple model approaches, while still producing a single comprehensible model.

Suppose L is the learner being used, and that m models are learned, for example using Breiman's (1996) bagging procedure. The m models can be combined into one by reapplying L to learn the labeling of the instance space produced by the model ensemble. Specifically, an unlimited number of additional training examples can be produced by randomly generating attribute vectors, and assigning to each the class obtained by passing it through the model ensemble. L can then be applied to the expanded training set, and the resulting single model used as the global output. In order to avoid misleading the learner, some care should be taken to ensure that the distribution of "artificial" vectors is similar to that found in the original data.

Whether this approach produces a model that retains some (or most of) the model ensemble's accuracy gain over the original single model, without becoming overly complex, is a matter for empirical verification. In our study, C4.5RULES (Quinlan 1993) was used as the base learner, and bagging was applied in a form very similar to that of (Breiman 1996), with 25 bootstrap replicates and 20 runs. Twenty-six datasets from the Irvine repository were used. In each domain, 1000 examples were generated from the bagged ensemble as outlined above. The distribution in the original data was simulated by taking each rule in each model and uniformly generating a number of examples covered by it, in proportion to the number of times the rule was used in the original training set (and making sure, up to a maximum number of iterations, that the examples were not covered by rules preceding the current rule in the model). Thus the example-generating procedure has a bias similar to that of the learner itself, and a mismatch between the two is avoided.

These experiments showed the proposed approach to be remarkably successful. In the 22 domains where bagging improved accuracy (and where, therefore, it makes sense to apply our procedure), the combined model retained on average 60% of the accuracy gain produced by bagging (which was 3.5% on average),[3] while having an output complexity that was typically a small multiple of the single model's (2–6). Further, the accuracy and complexity of the combined model can be traded off by varying the pruning level of C4.5RULES during the final learning pass: for example, with a confidence factor of 10%, the accuracy decreases by only fractions of a percentage point, while the output complexity is typically reduced by 10-20%. With C4.5RULES, the combined model's complexity is also a function of the number of examples generated; with 1000 examples used for all domains, the larger complexity ratios tended to be observed in the smaller datasets. Thus the current results may still admit to substantial improvement.

References

Breiman, L. 1996. Bagging Predictors. *Machine Learning* 24:123–140.

Chan, P.; Stolfo, S.; and Wolpert, D. eds. 1996. *Proc. AAAI-96 Workshop on Integrating Multiple Learned Models.* Portland, OR: AAAI Press.

Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning.* San Mateo, CA: Morgan Kaufmann.

[1]Partly supported by a NATO scholarship.

[3]The combined model was more accurate than C4.5RULES with 99.9% confidence according to a Wilcoxon test.