A Numerical Study of Three Ensemble Methods

A Mathematical and Data-Driven Approach to Intrusion Detection for High-Performance Computing

Orianna DeMasi, Juan Meza, and David Bailey
Lawrence Berkeley National Laboratory
1 Cyclotron Road, Berkeley, CA 94720

Ensemble methods for supervised machine learning have become popular due to their ability to accurately predict class labels with groups of simple, lightweight "base learners." While ensembles offer computationally efficient models that have good predictive capability they tend to be large and offer little insight into the patterns or structure in a dataset. We consider an ensemble technique that returns a model of ranked rules. The model accurately predicts class labels and has the advantage of indicating which parameter constraints are most useful for predicting those labels. The selection of important rules is shown on a dataset of supernova where the misclassification error rate was less than 5% before attribute selection and even lower after. We also compare the rule ensemble method with boosting and bagging, which are two well known ensemble techniques that use decision trees as base learners.

We modified an existing implementation that had not yet been subjected to experimentation, and tested and compared it against other rule ensembles and multi-class problems. To the best of our knowledge, no other implementations have yielded results comparing the rule ensemble with other methods on multi-class problems. We extended the method to address multiple class classification problems with one-versus-all classification, and tested against classical machine learning datasets from the UC Irvine machine learning repository and a dataset of potential supernovas.

The rule ensemble method builds a set of decision trees, uses each node as a rule, and finds coefficients to combine the rules in a linear model. The advantage to using rules rather than variables, as in multiple linear regression, is that rules can find relationships and hierarchies between attributes. The rules in the decision trees get more complex the deeper the tree is grown and also are able to have limited support in the parameter space, so they only affect observations that fall in that space. Complex rules can be seen as discrete correlations. The post-processing of the rules allows for overly simplified correlations to be removed from the model. Thus, in contrast to a standard linear regression, some variable interactions can be captured by the rule ensemble method without any *a priori* assumption that they exist or taking the expense of including terms for all potential correlations.

Rules with large weights have a larger effect on the model and can be thought of as more important than other rules. The importance of rules indicates the importance of the features that the rule is defined on and thus alerts the user to the importance of certain features in a given dataset. Traditional algorithms that use ensembles of decision trees, such as boosting and bagging, are not able to provide much insight into the importance of certain variables of a dataset because there is no ranking or weighting of rules. The rule ensemble method is shown to perform well compared to boosting and bagging on the UC Irvine datasets and the success of the rule ranking is shown on a dataset of images of potential supernova. From the supernova dataset, the rule ensemble method selected 21 of the 39 attributes and a new model built only with the selected attributes had a lower overall misclassification rate.

This research was supported in part by the Director, Office of Computational and Technology Research, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy, under contract number DE-AC02-05CH11231