Exploiting Diversity for Natural Language Processing

John C. Henderson*
Department of Computer Science
Johns Hopkins University
Baltimore, MD 21218-2694
jhndrsn@cs.jhu.edu

The recent popularity of applying machine learning methods to computational linguistics problems has given rise to a large supply of trainable natural language processing systems. Most problems of interest have an array of off-the-shelf products or downloadable code implementing solutions of varying quality using varying techniques.

The task this thesis is concerned with is developing reasonable methods for combining the *outputs* of a diverse set of systems which all address the same task. The hope is that if the set has a high enough initial accuracy and independently assorted errors, we can produce a more accurate solution using an appropriate combining method. In addition, there are principles that initial system developers should keep in mind which will help them create a family of diverse systems. We are also interested in developing methods for increasing the diversity of a set of systems without sacrificing accuracy, for the sake of fruitful combination. Each task we approach will warrant a separate investigation into how to combine outputs, and hope lies in discovering the principles that are common among all tasks. We don't want to study just one learning method or task, instead we want to discover principles that can be applied universally, or to a broad class of problems.

Our proposed work can draw from a number of recent machine learning results involving combining outputs of systems. In the *bagging* technique, randomness is used as the basis of generating a diverse ensemble of classifiers, and results are combined using simple interpolation or majority voting depending on the type of predictor that is required (Breiman 1996). Bagging has been used in conjunction with a genetic algorithm that enforces diversity while it trains a collection of neural networks to boost accuracy (Opitz & Shavlik 1996). Given a set of classification systems, the *AdaBoost* system, in conjunction with certain types of classifier generators, can produce a new system that provably reduces the error of the ensemble (Freund & Schapire 1997). The principles these systems utilize can be applied to combining outputs of systems with more structure, such as those in the domain of natural language processing.

*The author was supported by NSF grant IRI-9502312.

The overall plan is to start with computational linguistics problems that mesh well with current machine learning methods and progress toward problems with outputs that are full of rich structural dependencies. The lessons we learn and the techniques we perfect in coping with the simpler systems will be applied to the more complex systems.

The first step in pursuit of this thesis is to take off-the-shelf machine learning combining methods, such as majority voting, bagging, or boosting, and apply them to off-the-shelf natural language processing systems. The initial domains we are exploring in this way are text filtering and prepositional phrase attachment. These problems are well suited to this type of exploration because they are easily reformulated as standard classification problems.

We will then move toward more structured problems, those that do not lend themselves to classification-style output, such as part-of-speech (POS) tagging. The POS tagging problem is not simple classification because the input is a tape of words and the output is a tape of parts of speech. The identity of a single tag is highly dependent on tags assigned to nearby words.

Finally, we will attack the parsing problem. This problem is quite different from standard classification problems, as the output has a tree structure and the evaluation metric is highly sensitive to mistakes in the shape of that tree. For parsing we will most likely need to construct entirely new combining methods.

See `http://www.cs.jhu.edu/~jhndrsn/thesis.html` for an updated summary of progress toward this thesis.

References

Breiman, L. 1996. Bagging predictors. In *Machine Learning*, volume 24, 123.

Freund, Y., and Schapire, R. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *JCSS* 55(1):119–139.

Opitz, D. W., and Shavlik, J. W. 1996. Generating accurate and diverse members of a neural-network ensemble. In *NIPS8*.