Incremental Rule-Based Learning

Charles M. Higgins and Rodney M. Goodman

Department of Electrical Engineering, 116-81

California Institute of Technology

Pasadena, CA 91125

USA

In a system which learns to predict the value of an output variable given one or more input variables by looking at a set of examples, a rule-based knowledge representation provides not only a natural method of constructing a classifier, but also a human-readable explanation of what has been learned. Consider a rule of the form *if y then x* where y is a conjunction of values of input variables and x is a value of the output variable. The number of input variables in y is called the **order** of the rule. In previous work, a measure of the information content or "value" of such a rule has been developed (the J-measure [1],[2]). It has been shown in [3] that a classifier can be built from the rules obtained by a constrained search of all possible rules which performs comparably with other classifiers.

It is desirable that such a system learns *incrementally*. That is, the system should be able to modify what it has learned in the past to reflect more data as it is available. In the present work, we show an algorithm for finding rules which is incremental, less computationally intensive and more suitable for parallel implementation than [3] while using the same classifier architecture and retaining similar performance.

In order to discover the best rules describing a set of examples, we begin by converting each example into a very specific rule; that is, for each example a rule is made whose left-hand side (y) is composed of a conjunction of the values of the input variables in the example, and whose right-hand side (x) is the value of the output variable in the example. We will proceed to generalize this inital set of very specific rules. The J-measure is calculated for this rule and for the child-rules generated by removing each input variable in the rule successively. All child-rules have order one less than the original. The single rule with the greatest J-measure among the original rule and its child-rules is picked and the other rules are discarded. In the case of a tie, the rule with the lower order is chosen. This continues until the J-measure no longer increases. When this has been done for every example, a single rule exists for every original example. Duplicate rules are removed and the remaining rule set may be used in classification.

To learn incrementally, the algorithm in the previous paragraph is performed on the data immediately available to the classifier and each rule generated retains the example from which it was generated. When more data is available, the new examples are put into the algorithm and developed as in the previous paragraph. In addition, the previously generated rules are developed as above with one exception: child-rules for this set include those generated by *adding back* successively each previously removed attribute in case the rule has been generalized too far. The J-measure in all cases is calculated with respect to the concatenation of the two data sets.

Experimental results are shown comparing the algorithm with a back-propagation network, a first-order Bayes network, and the rule-based classifier of [3] on several well known data sets. Theoretical topography of the rule search space and variations on the algorithm are discussed. Parallel hardware implementation issues are also addressed.

References

[1] N. M. Blachman, "The amount of information that y gives about X," *IEEE Transactions on Information Theory*, vol. IT-14(1), 27-31, 1968.

[2] R. M. Goodman and P. Smyth, "The Induction of Probabilistic Rule Sets – The ITRULE Algorithm," *Proceedings of the 1989 International Workshop on Machine Learning*, San Mateo, CA: Morgan Kaufmann, pp. 129-132, 1989.

[3] R. M. Goodman, C. Higgins, and P. Smyth, "A Hybrid Rule-Based/Bayesian Classifier," *Proceedings of the 1990 European Conference on Artificial Intelligence*, Stockholm, Sweden, August 1990.