Induction of Decision Trees from Inconclusive Data
(extended abstract)

Scott Spangler
Knowledge Engineering Group
Advanced Engineering Staff
GM Technical Center
Warren, MI 48090

Usama M. Fayyad
Artificial Intelligence Group
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

Ramasamy Uthurusamy
Computer Science Department
GM Research Laboratories
GM Technical Center
Warren, MI 48090

1. Introduction

The automated induction of diagnostic rules from large databases promises to be one of the most successful areas of application of machine learning. However, real-world databases pose additional problems that make the induction process significantly more difficult. In a databases where cases are stored as vectors of attribute values, some values may be missing or unknown for some cases (examples). Furthermore, some attributes may be "noisy" or their values may be recoded erroneously. Some research has been conducted to address these issues [Quinlan 86]. In our experience with a large database of automobile repair cases, we have come across an instance of what we believe to be a new class of data that requires a special kind of handling. We refer to such training data sets as *inconclusive data*.

Most learning algorithms make the implicit assumption that the attributes used to describe examples contain enough information to classify each example in a single class. The learning algorithm would then attempt to obtain a perfect classification of the training set. If this is not possible, some post processing may take place to attempt to resolve the outcome of a rule to a single class. In some domains, the levels of noise may be known to be low. However, the attributes may not contain enough knowledge to resolve every example to a single class. In such a case, the best that could be expected from a classification rule is to narrow down the possible classification to a few classes. More formally, a rule that utilizes only the given attributes to classify examples in a single class will have a true probability of classification error that is bounded from below by some number greater than zero. Though this characterization cannot be applied to arbitrary data sets to determine whether they should be considered inconclusive, it does identify the notion of inconclusiveness as a problem to be handled via special methods that do not make assumptions of noise as in [Breiman et al. 84; Quinlan 86]. We outline such a method below under our description of the INFERULE algorithm.

Given a training set of data which is known to have low levels of noise, a clear symptom of inconclusiveness is that induction algorithms produce classification rules that classify the data set perfectly, but they achieve very low accuracy when used to classify a separate set of test examples. The goal in such a situation is to obtain rules that predict multiple classifications since a single classification can only be obtained at the unacceptable expense of overspecialization.

2. The INFERULE Algorithm

A popular and efficient approach to the induction of a set of rules for classifying a training set of examples is to generate a decision tree[Breiman et al. 84; Quinlan 86]. The INFERULE algorithm is a modification of Quinlan's ID3 algorithm. The two algorithms differ in several respects. The details are discussed at length in [Spangler et al. 89]. For the purposes of this abstract it suffices to outline the difference in the attribute selection criterion. INFERULE utilizes an attribute evaluation criterion that evaluates an attribute-value pair's worth by measuring the effect the split due to the pair has on changing the class distribution in a subset of examples at a node in the tree. The split which results in the distribution having the farthest "distance" from the original class distribution is considered the best split. Furthermore, if none of the splits exceed a certain *minimum distance value* then that is taken as evidence that no further splitting of the node is warranted and the node is deemed a leaf. This provides INFERULE with a mechanism for halting specialization rather than resorting to irrelevant splits that result in overspecialization.

An alternative approach would be to generate a full decision tree and then prune it back. There are many reasons why this is not desirable or applicable in our case. The type of pruning done in [Breiman et al. 84] assumes a single (usually the most probable) class for a rule and prunes with respect to that class. This assumption does not hold for inconclusive data where multiple classes are desirable.

The Chi-square pruning employed in [Quinlan 86] requires that at least 4 examples of each class be present at a node. In our domain, there is a very large number of classes (over 200) and this condition was violated 95% of the time. We found pessimistic pruning [Quinlan 86b] to be not very effective (see results in next section). The other disadvantage of the pruning approach lies in the extra computational cost associated with it in terms of time and space.

3. The Results

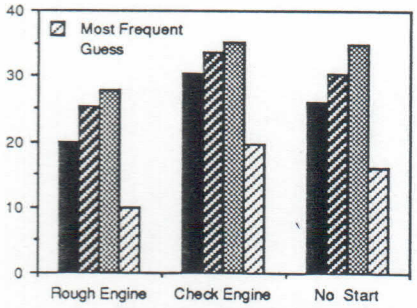
We have applied the INFERULE algorithm to a large database of automobile repair cases. The diagnostic rules obtained were used to classify a separate set of test data. The figures below show the results obtained by averaging over several runs. INFERULE is compared against ID3 and ID3 with pessimistic pruning [Quinlan 86b]. Two remarks need to be made regarding the result charts. In the percentage correctness chart, the INFERULE column does not reflect one of the primary advantages of the INFERULE algorithm, namely suggesting multiple classifications per example. The graph shows only the percentage of the time for which the class with the highest confidence factor matched the correct classification. The percentage of test cases for which the actual outcome was one of those suggested by INFERULE averaged between 80% and 90%. Thus including multiple guessing would enhance performance significantly. The second thing to note is that generating a full decision tree and then pruning it back is a time consuming process. INFERULE generates an already pruned tree without incurring the extra cost of pruning.

We translated the rules obtained into M1 expert system shell language and presented them in the form of a simplistic expert system to the technical assistance staff who use the database as part of their daily operations. We received very positive feedback regarding the usefulness of the rules as an aid to them in performing their diagnostic tasks.

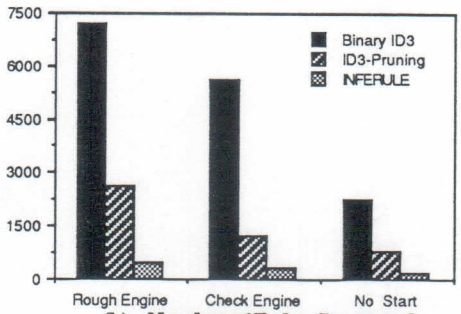
4. Conclusions

The primary purpose of this paper is not to introduce yet another variant of ID3. Instead, we hope to influence future rule induction research to focus more on inconclusive data sets. After studying several diagnostic databases at General Motors, we are convinced that inconclusiveness is a characteristic that rule induction algorithms must be able to deal with if they are to have practical application in industry.

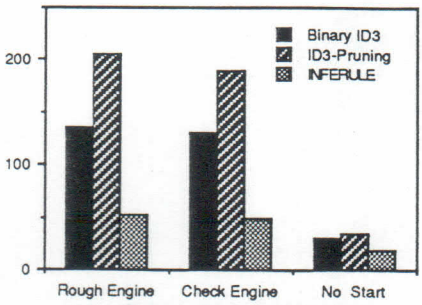
The proper way of handling inconclusive data sets is to produce probabilistic, rather than categorical, classification rules. The assumption that rules with a single outcome are desirable should be abandoned when it is not statistically supported. This introduces the need for a criterion to determine when to stop the specialization process in top-down tree generation. INFERULE addresses this issue and illustrates that the predicted improvement in performance is indeed achieved.



(a) Percentage of Corretness (%)



(b) Number of Rules Generated



(c) Time to Run (in Minutes)

References

Breiman et al. 84 Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. *Classification and Regression Trees*. Wadsworth: Belmont, CA 1984.

Quinlan 86 Quinlan, J.R. "Induction of Decision Trees." *Machine Learning 1 (1)*. 1986.

Quinlan 86b Quinlan, J.R. "Simplifying decision trees." MIT AI Memo No. 930. 1986.

Spangler et al. 89 Spangler, S., Fayyad, U.M., and Uthurusamy, R. "Induction of decision trees from inconclusive data." Proceedings of the Sixth International Workshop on Machine Learning. Ithaca, NY 1989.