Statistical Science
2006, Vol. 21, No. 1, 24–26
DOI 10.1214/088342306000000033

Elaboration on Two Points Raised in "Classifier Technology and the Illusion of Progress"

Robert C. Holte

1. INTRODUCTION

This short note elaborates two points raised in David Hand's target article. First, I provide additional evidence that simple classification rules should be given serious consideration in any application and that there are often diminishing returns in considering increasingly complex classifiers. Second, I refine Hand's basic argument that small improvements in performance are irrelevant because of the uncertainty about many aspects of the situation in which the classifier will be deployed. In particular, I briefly describe a recently developed method for analyzing and comparing classifier performance when the class ratios and misclassification costs are unknown. This does not refute his general argument, but it does provide an important exception to it.

2. SIMPLICITY-FIRST METHODOLOGY AND DIMINISHING RETURNS

Hand (Section 2.3) cites my 1993 study [4] in which the accuracy of one-level decision trees, which classify examples based on the value of a single feature, was compared to the accuracy of the decision trees learned by C4.5 [8], a state-of-the-art decision tree learning algorithm. The article caused quite a stir, because nobody at the time suspected that most of C4.5's classification accuracy could be achieved, on many of the standard test data sets, by building just the first level of the decision tree. The overall conclusion of my 1993 article is the same as Hand's—not that the more complex decision rules should be cast aside, but that the simple decision rules should not be dismissed out of hand. One can never tell, a priori, how much of the structure in a domain can be captured by a very simple decision rule, and since simplicity is advantageous for both theoretical and practical reasons, it is incumbent on a responsible experimentalist or practitioner to begin with the simplest decision rules. Only if they prove unacceptable should more complex decision rules be considered. I coined the term "simplicity-first methodology" to describe this systematic approach of proceeding from simple to more complex decision rules.

In a follow-up paper [1], Maass and Auer developed an efficient algorithm for constructing a decision tree of fixed depth d, with the minimal error rate on the training data, and we proved theoretical bounds on the generalization error rate of this decision tree. This empirical study showed that the performance advantage of C4.5 over one-level trees in my original study [4] greatly diminishes when depth is increased to two, with the two-level trees actually being superior to C4.5's trees on 4 of the 15 data sets in the study.

Table 1 herein compares the accuracies achieved when $d = 0$, $d = 1$ and $d = 2$. These accuracies are averages of nine repetitions of 25-fold cross-validation on each data set. The $\Delta(1–0)$ column gives the accuracy improvement achieved by moving from a zero-level tree, which classifies all examples according to the majority class, to a one-level tree, and the $\Delta(2–1)$ column gives the accuracy improvement achieved by moving from a one-level tree to a two-level tree. Comparing these two columns, we see clear confirmation of Hand's observation that increasing complexity produces diminishing returns on accuracy improvement in many domains.

There have been other studies that showed that simple classifiers perform well on standard test data sets. Domingos and Pazzani [2] showed that a naive Bayesian classification algorithm significantly outperformed state-of-the-art systems for decision tree learning, decision rule learning and instance-based learning in a substantial number of the 28 data sets in their study. Kohavi [5] showed that wrapper-based feature selection, combined with a majority classifier, can produce simple classifiers that are as accurate as C4.5's trees in many cases. Linear discriminants (perceptrons)

Robert C. Holte is Professor, Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, T6G 2E8 (e-mail: holte@cs.ualberta.ca).