Statistical Science
2006, Vol. 21, No. 1, 15–18
DOI 10.1214/088342306000000024

Comment: Classifier Technology and the Illusion of Progress

Jerome H. Friedman

This paper provides a valuable service by asking us to reflect on recent developments in classification methodology to ascertain how far we have progressed and what remains to be done. The suggestion in the paper is that the field has advanced very little over the past ten or so years in spite of all of the excitement to the contrary.

It is of course natural to become overenthusiastic about new methods. Academic disciplines are as susceptible to fads as any other endeavor. Statistics and machine learning are not exempt from this phenomenon. Often a new method is heavily championed by its developer(s) as the "magic bullet" that renders past methodology obsolete. Sometimes these arguments are accompanied by nontechnical metaphors such as brain biology, natural selection and human reasoning. The developers become gurus of a movement that eventually attracts disciples who in turn spread the word that a new dawn has emerged. All of this enthusiasm is infectious and the new method is adopted by practitioners who often uncritically assume that they are realizing benefits not afforded by previous methodology. Eventually realism sets in as the limitations of the newer methods emerge and they are placed in proper perspective.

Such realism is often not immediately welcomed. Suggesting that an exciting new method may not bring as great an improvement as initially envisioned or that it may simply be a variation of existing methodology expressed in new vocabulary often elicits a strong reaction. Thus, the messengers who bring this news tend to be, at least initially, unpopular among their colleagues in the field. It therefore takes courage to provide this type of service, and Professor Hand is to be congratulated for this thoughtful article.

Of course, simply because new methodologies are often overhyped does not necessarily imply that they do not, at least sometimes, represent important progress. In the case of classification, I believe that there have been major developments over the past ten years that have substantially advanced the field, both in terms of theory and practice. Although I find myself in agreement with most of the premises of this article, I do not see how they lead to the implication that such advances are "largely illusionary."

There appear to be three main premises presented in the article. First, the improvements realized by the newer methods over the previous ones are less than those achieved by the previous ones over their predecessors, presumably no methodology at all. Second, the evidence often presented (at least initially) in favor of the superiority of the newer methods is often suspect. Finally, the newer methods do not solve all of the outstanding important problems that remain in the field of classification. In my view these observations are correct and underappreciated in the field. The article does an important service by illustrating them so forcefully. However, the truth of these assertions does not imply lack of important progress; only that low-lying fruit is often easier to gather, we should be more thorough concerning validation when initially presenting new procedures and there is still important work to be done.

One of the main assertions in the paper is that, in many applications, older methods often yield error rates comparable to the more modern ones. This is of course true and is intrinsic to the classification problem, especially when the metric used to measure performance is based on error rate. First, there is the irreducible error caused by the fact that the predictor variables $\mathbf{x}$ often do not contain enough information to specify a unique value for the outcome variable y. At best, they specify a probability distribution of possible values $\Pr(y|\mathbf{x})$ which is hopefully different for differing values of $\mathbf{x}$, indicating some predictive power. This phenomenon afflicts all prediction problems. A second phenomenon is peculiar to classification; it is not necessary to accurately estimate $\Pr(y|\mathbf{x})$ to achieve minimal error rate. All that is required of the estimates $\widehat{\Pr}(y|\mathbf{x})$ is

$$\arg\max_y \widehat{\Pr}(y|\mathbf{x}) = \arg\max_y \Pr(y|\mathbf{x}). \tag{1}$$

Jerome H. Friedman is Professor, Department of Statistics and Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305, USA (e-mail: jhf@stanford.edu).