Statistical Science
2007, Vol. 22, No. 4, 477–505
DOI: 10.1214/07-STS242

Boosting Algorithms: Regularization, Prediction and Model Fitting

Peter Bühlmann and Torsten Hothorn

Abstract. We present a statistical perspective on boosting. Special emphasis is given to estimating potentially complex parametric or nonparametric models, including generalized linear and additive models as well as regression models for survival analysis. Concepts of degrees of freedom and corresponding Akaike or Bayesian information criteria, particularly useful for regularization and variable selection in high-dimensional covariate spaces, are discussed as well.

The practical aspects of boosting procedures for fitting statistical models are illustrated by means of the dedicated open-source software package **mboost**. This package implements functions which can be used for model fitting, prediction and variable selection. It is flexible, allowing for the implementation of new boosting algorithms optimizing user-specified loss functions.

Key words and phrases: Generalized linear models, generalized additive models, gradient boosting, survival analysis, variable selection, software.

1. INTRODUCTION

Freund and Schapire's AdaBoost algorithm for classification [29–31] has attracted much attention in the machine learning community (cf. [76], and the references therein) as well as in related areas in statistics [15, 16, 33]. Various versions of the AdaBoost algorithm have proven to be very competitive in terms of prediction accuracy in a variety of applications. Boosting methods have been originally proposed as ensemble methods (see Section 1.1), which rely on the principle of generating multiple predictions and majority voting (averaging) among the individual classifiers.

Later, Breiman [15, 16] made a path-breaking observation that the AdaBoost algorithm can be viewed as a gradient descent algorithm in function space, inspired by numerical optimization and statistical estimation. Moreover, Friedman, Hastie and Tibshirani [33] laid out further important foundations which linked AdaBoost and other boosting algorithms to the framework of statistical estimation and additive basis expansion. In their terminology, boosting is represented as "stagewise, additive modeling": the word "additive" does not imply a model fit which is additive in the covariates (see our Section 4), but refers to the fact that boosting is an additive (in fact, a linear) combination of "simple" (function) estimators. Also Mason et al. [62] and Rätsch, Onoda and Müller [70] developed related ideas which were mainly acknowledged in the machine learning community. In Hastie, Tibshirani and Friedman [42], additional views on boosting are given; in particular, the authors first pointed out the relation between boosting and ℓ^1-penalized estimation. The insights of Friedman, Hastie and Tibshirani [33] opened new perspectives, namely to use boosting methods in many other contexts than classification. We mention here boosting methods for regression (including generalized regression) [22, 32, 71], for density estimation [73], for survival analysis [45, 71] or for multivariate analysis [33, 59]. In quite a few of these proposals, boosting is not only a black-box prediction tool

Peter Bühlmann is Professor, Seminar für Statistik, ETH Zürich, CH-8092 Zürich, Switzerland (e-mail: buhlmann@stat.math.ethz.ch). Torsten Hothorn is Professor, Institut für Statistik, Ludwig-Maximilians-Universität München, Ludwigstraße 33, D-80539 München, Germany (e-mail: Torsten.Hothorn@R-project.org). Torsten Hothorn wrote this paper while he was a lecturer at the Universität Erlangen-Nürnberg.

Discussed in 10.1214/07-STS242A and 10.1214/07-STS242B; rejoinder at 10.1214/07-STS242REJ.