Variable Selection and Model Choice in Survival Models with Time-Varying Effects

Benjamin Hofner, Thomas Kneib & Torsten Hothorn

Institut für Statistik, Ludwig-Maximilians-Universität München, Munich, Germany;
`benjamin.hofner@stat.uni-muenchen.de`

Abstract: Flexible hazard regression models based on penalised splines allow to extend the classical Cox-model via the inclusion of time-varying and nonparametric effects (Kneib & Fahrmeir 2007). Despite their immediate appeal in terms of flexibility, these models introduce additional difficulties when performing model choice and variable selection.

Boosting (cf. Bühlmann & Hothorn, 2008, and Tutz & Binder, 2006) supports model fitting for high-dimensional data. By using component-wise base-learners, variable selection and model choice can be performed in the boosting framework.

We introduce a boosting algorithm for survival data that permits the inclusion of time-varying effects in a parametric form or in a flexible way, using P-splines. Thus we can fit flexible, additive hazard regression models and have a fully automated procedure for variable selection and model choice at hand.

The properties and performance of the algorithm are investigated in simulation studies. In an application, we present the analysis of retrospective data of surgical patients with severe sepsis were the aim was to build a flexible prognostic model.

References

KNEIB, T. & FAHRMEIR, L. (2007) A mixed model approach for geoadditive hazard regression. *Scand. J. Statist.*, **34**, 207–228.

BÜHLMANN, P. & HOTHORN, T. (2008) Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, Accepted.

TUTZ, G. & BINDER, H. (2006) Generalized additive modelling with implicit variable selection by likelihood-based boosting. *Biometrics*, **62**, 961–971.

Keywords