Editorial

Boosting – An Unusual Yet Attractive Optimiser

T. Hothorn
University of Zurich, Zurich, Switzerland

Summary
This editorial is part of a For-Discussion-Section of Methods of Information in Medicine about the papers “The Evolution of Boosting Algorithms – From Machine Learning to Statistical Modelling” [1] and “Extending Statistical Boosting – An Overview of Recent Methodological Developments” [2], written by Andreas Mayr and co-authors. It preludes two discussed reviews on developments and applications of boosting in biomedical research. The two review papers, written by Andreas Mayr, Harald Binder, Olaf Gefeller, and Matthias Schmid, give an overview on recently published methods that utilise gradient or likelihood-based boosting for fitting models in the life sciences. The reviews are followed by invited comments [3] by experts in both boosting theory and applications.

doi: 10.3414/ME13-10-0123

Correspondence to:
Torsten Hothorn
Epidemiology, Biostatistics and Prevention Institute
University of Zurich
Hirschengraben 84
8001 Zürich
Switzerland
E-Mail: Torsten.Hothorn@uzh.ch

This issue of Methods of Information in Medicine features a pair of discussed review papers [1, 2] on boosting methods in biomedical research. The authors, Drs. Mayr, Binder, Gefeller, and Schmid, present the main developments of boosting in both the machine learning and statistics communities with a focus on applications in biomedical research. After a short introduction to the history of boosting, especially from a statistical point of view, the authors summarise contemporary research results that were published after the first statistical review paper on boosting [4]. Especially noteworthy is that fact that two, at first sight rather distinct, boosting algorithms, gradient and likelihood boosting, are presented in the same language.

The review papers are discussed by Peter Bühlmann, Jan Gertheiss, Stefanie Hieke and Martin Schumacher, Thomas Kneib, Shuangge Ma, Gerhard Tutz, Zhu Wang and Ching-Yun Wang, and Andreas Ziegler [3]. The discussants were instrumental in establishing boosting in the statistics and biostatistics communities and published both theoretical and applied contributions. They made an excellent job summarising, highlighting, and extending the most important topics discussed by Drs. Mayr, Binder, Gefeller, and Schmid, so instead of adding another overview I will try to answer the question “Why and when should I consider using boosting?” in the light of the contributions published in this issue.

In my opinion, gradient boosting is like a Swiss knife. It is not the ideal tool for any task but if you know how to handle it, it will solve your problem eventually. This is also my reading of the extensive empirical comparison of different approaches to fit structured additive models performed by Scheipl et al. [5].

But there is more to it. Boosting is an unusual optimiser, because the target function is optimised with some implicit penalisation those impact varies with the number of boosting iterations. This lack of explicit penalty terms in the target function is the main difference between boosting and other penalization methods, such as Lasso and friends. What may seem as a weakness from a theoretical point of view is in fact a big advantage. The flexibility we obtain from refraining to specify an explicit penalty allows us to fit many known and (yet) unknown statistical models in a uni-
fied computational framework. This, in turn, has two consequences. Since there is less need to write additional code, the overall time and thus costs associated with fitting a model can be dramatically reduced. Second, the likelihood of implementation errors decreases when welltested infrastructure can be reused.

Apart from these rather convenient computational issues, we also benefit from such unified computational tools theoretically. For more than 100 years, what statisticians could do was limited by asymptotic theory. Later, the availability of optimisation or integration procedures (at the Bayesian side of the street), or both (in mixed models) defined what could be done. With boosting, lack of a highly flexible optimiser is no excuse anymore for using overly restrictive statistical models. The ability to quickly change the loss function (e.g. to one not involving the proportional hazards assumption, Johnson and Long [6]) or the model-structure (e.g. using wavelets, Dubossarsky et al., [7]) allows studying the properties of novel statistical models without much technical difficulties. In the same spirit, the whole business of model diagnostics is greatly simplified, because what keeps you from quickly boosting an additive non-proportional hazards model in order to check if it fits your data better than a linear Cox model? If the two models perform roughly equivalent, you know that assuming proportional hazards and linearity won’t hurt.

In this line, the most interesting part for me is section 4.2. in Mayr et al. [1], where models that go way beyond the (G)L/AM framework are discussed. Implementing tailored optimisers for GAMLSS or CTMs takes a lot of time and experience, but coming up with a ”quick-and-dirty” boosting implementation is (to those in the know) straightforward. And we can even go further. Without much limitation imposed by algorithms for model fitting, we can develop new and problem-specific models really fast, greatly improving our understanding of the processes we aim to model. Once we know exactly which model is most appropriate, we may want to implement and employ a more problem-specific, more stable, or simply faster optimiser. With boosting, this tedious and often time-consuming task can be put last (not first) on your research agenda.

There is, as Jan Gertheiss nicely pointed out, a downside. There are not many followers of these ideas yet. If you browse through mainstream statistics journals you will get the impression that papers solving (seemingly) highly specialised problems using hand-crafted methods are preferred by editors and reviewers over unifying approaches. Since I’m sure you think I exaggerate a little bit, I’ll end with a quote from a review I recently received as a co-author of a paper describing a unified theoretical and computational framework for functional data analysis (using boosting, of course):

“The extreme generality of the proposed framework is my biggest concern. The argument for a unified approach to all regressions involving functional data is that each particular model is just a special case, but the expense is that each of these specific cases then seems to require quite a bit of unique attention. […] The usefulness of this isn’t made clear, other than that generality for its own sake is implicitly preferred.”

This review, solicited and communicated by a prestigious journal in computational statistics, underlines the importance of the contribution by Dr. Mayr and his colleagues – it’s just time for the next step.

References