PROJECT MUSE®

A Melting Pot

Rob Tibshirani, Trevor Hastie

Observational Studies, Volume 7, Issue 1, 2021, pp. 213-215 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2021.0012>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/799737>

A Melting Pot

Rob Tibshirani

*Department of Statistics
Stanford University
Stanford, CA 94305, USA*

tibs@stanford.edu

Trevor Hastie

*Department of Statistics
Stanford University
Stanford, CA 94305, USA*

hastie@stanford.edu

Abstract

Leo Breiman’s article “Statistical Modeling: The two cultures” was timely and provocative. He advocated for Statisticians to learn about and appreciate a different “culture”: an algorithmic approach, as distinct from the familiar, stochastic, data modeling approach to Statistics. While we have appreciated and contributed to the algorithmic approach, we have always had a foot in both camps. Here we advocate for a “melting pot”, arguing that both approaches have their virtues, sometimes on the same problem.

Keywords: Algorithm, prediction, supervised learning, causal inference

We knew Leo well, and admired him for his seminal contributions to our field and for his zest for life. His last major contribution was the Random Forest, which was gaining popularity by the time of his death in July 2005. We wish he could see how popular it is now. Leo, Random Forests are “going like gangbusters”.

Our own work in statistical learning has been inspired by Leo and some of his views as expressed in this paper, and we have largely embraced the algorithmic modeling culture. However, it is fair to say that we have always had a foot in both the data modeling and the algorithmic modeling camps. We feel there is a clear role for both of these approaches, sometimes even working together on the same problem. We mention a few examples.

- *Causal inference* is a hot topic — comparing two treatments using observational (non-randomized) measurements on subjects. This cannot be cast as a simple supervised learning problem because we never observe the outcome for a given patient under both treatments. In addition, any reasonable analysis needs to account for selection bias.

The current state-of-the-art is to use probabilistic models, with assumptions such as potential outcomes and unconfoundedness, and the use of propensity scores. Given this model, one can use algorithmic supervised learning tools to estimate the key ingredients. Then the estimated model can be used for inferences about the treat-

ments. This program is nicely illustrated in the R-learner for estimating heterogeneous treatment effects Nie and Wager (2020).

- *Logistic regression* initially gained popularity in biostatistics and epidemiology. Its focus is the odds ratio, a reasonable surrogate for relative risk in the case of rare diseases. It comes with a big plus: it can be estimated from retrospective, case-control samples. This is a probabilistic result, that relies on Bayes formula and conditioning arguments. It allows us to treat the non-random case-control label as the target for estimation via logistic regression, and still get valid estimates of the parameters of interest. It also shows us that we can subsample the massive number of negative examples in web-based modeling of click-through rate, and still get valid predictions from logistic regression. Of course with such data we are free to use more complex algorithmic approaches to fit the response probability function, but the knowledge gained from logistic regression shows us what to do with it.
- *The Cox proportional hazards model* is ubiquitous in biostatistics and epidemiology, as well as in financial and industrial applications. It uses a probabilistic semi-parametric framework that cleverly allows us to model the relative-risk functions of interest, while allowing for an arbitrary baseline hazard. The probability model thus determines the loss function, and thereafter allows for any classical or machine-learning model to estimate the functions of interest.
- There are many examples of using probabilistic models to understand and explain the behavior and performance of algorithmic models. We list a few:
 - The hinge-loss plus penalty for support-vector machines, which makes a direct comparison possible with the log-likelihood for logistic regression Wahba et al. (2000); Hastie et al. (2009), and highlights their similarities and where they differ.
 - The interpretation of Adaboost as an additive logistic regression model Friedman et al. (2000), leading to the development of gradient boosting Friedman (2001).
 - Clear explanations for the “double descent” phenomena in machine learning, using simple parametric models (James et al., 2021; Hastie et al., 2020, for example).

To summarize, we need both data and algorithmic models — sometimes working together in the same problem — and we need subject matter expertise to know when to employ either or both. Even more, data modeling can help us to understand the properties of algorithmic modeling.

References

- J. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232, 2001.
- J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting (with discussion). *Annals of Statistics*, 28:337–307, 2000.
- T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Prediction, Inference and Data Mining*. Springer Verlag, New York, second edition, 2009.
- Trevor Hastie, Andrea Montanari, Saharon Rosset, and Ryan J. Tibshirani. Surprises in high-dimensional ridgeless least squares interpolation, 2020.
- G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning, 2nd Edition (p439)*. Springer, New York, 2021.
- Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects, 2020. To appear, *Biometrika*.
- G. Wahba, Y. Lin, and H. Zhang. GACV for support vector machines. In A.J. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 297–311, Cambridge, MA., 2000. MIT Press.