The Annals of Applied Statistics
2010, Vol. 4, No. 4, 1638–1641
DOI: 10.1214/10-AOAS381
© Institute of Mathematical Statistics, 2010

REMEMBRANCE OF LEO BREIMAN

BY PETER BÜHLMANN

ETH Zürich

1. How I met Leo Breiman. In 1994, I came to Berkeley and was fortunate to stay there three years, first as a postdoctoral researcher and then as Neyman Visiting Assistant Professor. For me, this period was a unique opportunity to see other aspects and learn many more things about statistics: the Department of Statistics at Berkeley was much bigger and hence broader than my home at ETH Zürich and I enjoyed very much that the science was perhaps a bit more speculative.

As soon as I settled in the department, I tried to get in touch with the local faculty. Leo Breiman started a reading group on topics in machine learning and I didn't hesitate to participate together with other Ph.D. students. Leo spread a tremendous amount of enthusiasm, telling us about the vast opportunity we now had by taking advantage of computational power. Hearing his views and opinions and listening to his thoughts and ideas has been very exciting, stimulating and entertaining as well. This was my first occasion to get to know Leo. And there was, at least a bit, a vice-versa implication: now, Leo knew my name and who I am. Whenever we saw each other on the 4th floor in Evans Hall, I got a very gentle smile and "hello" from Leo. And in fact, this happened quite often: I often walked around while thinking about a problem, and it seemed to me, that Leo had a similar habit.

2. Witnessing three of Leo's fundamental contributions. I only got to know Leo Breiman in his late career. Nevertheless, between 1994 and 1997 when I was in Berkeley, I could witness Leo's exceptional creativity when he invented Bagging [Breiman (1996a)], gave fundamental explanations about Boosting [Breiman (1999)] and started to develop Random Forests [Breiman (2001)].

2.1. *Bagging.* I had the unique opportunity to listen to Leo Breiman when he presented Bagging during a seminar talk at UC Berkeley. I was puzzled and intrigued. At that time, I was working on the bootstrap and what Leo said didn't sound right to me: using the bootstrap language, he proposed to use $\hat{\theta}_{\text{Bag}} = \mathbb{E}^*[\hat{\theta}^*]$, where $\hat{\theta}$ is the output of a "complex algorithm" based on the original observations and $\hat{\theta}^*$ denoting the analogue based on the bootstrap sample. Trivially,

$$\hat{\theta}_{\text{Bag}} = \hat{\theta} + (\mathbb{E}^*[\hat{\theta}^*] - \hat{\theta}),$$

Received July 2010.
AMS 2000 subject classifications. Primary 62G08, 62G09; secondary 68T10.
Key words and phrases. Bagging, boosting, classification and regression trees, random forests.

and hence from this point of view, Leo has proposed to use the original estimator and *adding* the classical bootstrap bias correction estimate (instead of subtracting it). But this is not an appropriate view for the problem Leo was looking at, and—as usual—it turned out that he was right. Of course, Leo didn't present Bagging in this way: he argued via stability [Breiman (1996a, 1996b)] and that unstable estimators can be stabilized using the bootstrap. I still remember how Leo presented during the seminar talk many empirical examples, one batch of datasets after the other, demonstrating that Bagging improves the prediction performance by about 30%. It was great news! And also a kind of shock that nobody among the people in the audience or in the community had thought about it before.

After the seminar, I tried it out myself: it's so simple and easy to do! And indeed, Bagging worked when using CART trees or other "unstable" procedures. And in terms of prediction, it didn't do any harm for "stable" procedures. I have been fascinated by the idea, I started working on it and eventually, Bin Yu and I had some additional explanations why Bagging works [Bühlmann and Yu (2002)]—a tiny contribution in comparison to Leo's breakthrough.

2.2. *Arcing and Boosting*. In 1996, Freund and Schapire (1996) published their AdaBoost algorithm and they showed many empirical examples where their method performed exceptionally well. This caught a lot of attention, and maybe even more so than with Bagging, one wondered why such an ensemble method based on mysterious re-weighting works so well. Leo Breiman also got involved: he proposed a variant of Boosting called "Arcing" [Breiman (1998)] and then once more, he made a breakthrough: he formalized AdaBoost as a gradient descent optimization in function space where the gradient is estimated by a nonparametric procedure such as a CART tree [Breiman (1999)]. Many people, particularly from statistics, followed up on Leo's formalized framework [Friedman, Hastie and Tibshirani (2000); Bühlmann and Yu (2000); Friedman (2001); Bühlmann and Yu (2003)]. Part of my own research has built up on this result and Leo's result had a big and crucial influence on my research.

Leo's important and deep contribution in Boosting was about understanding the algorithm and not in terms of developing a new method. Maybe this was an interesting "outlier" in Leo's late career where he primarily was the designer of new methods and algorithms. But it fits perfectly into the picture: my remembrance of Leo is not only about his outstanding creativity but also about his analytical thinking regarding algorithms and machine learning—which is not a complete surprise given his mathematical background and training.

2.3. *Random Forests*. A third fundamental contribution of Leo's late career is the development of Random Forests, and I have a special memory on this. I was at home in Switzerland and Don Geman gave a talk at ETH Zürich about using trees with randomization at the nodes [Amit and Geman (1997)]. I spoke with Don and told him about Leo's Bagging which randomizes the samples instead of the nodes in the tree but Don was convinced that node randomization is much better. Leo took this suggestion and carried it much further. In particular, he created the idea of incorporating "variable importance," knowing well in advance that people will use it in complex data problems with thousands of variables as in, for example, high-throughput molecular biology [Diaz-Uriarte and de Andres (2006); Menze et al. (2009)].

Random Forests is an astonishingly powerful "off-the-shelf" method. Whether we like such "off-the-shelf" procedures or not, Random Forests works extremely well in great generality, given that it is a pure machine learning algorithm which essentially does not even require the specification of a tuning parameter! There is virtually no competing method which can so easily deal with high-dimensional continuous, categorical or mixed data yielding powerful predictions and some "first-order" information about variable importance. There have been some attempts for better (mathematical) understanding of Leo's Random Forests [Lin and Jeon (2006); Biau, Devroye and Lugosi (2008)] and I tried myself some years ago. However, without having Leo's deep insights and intuition, it's maybe still a bit of a mystery why Random Forests works so well.

3. Being influenced by Leo. Leo's grand views, visions and his research had a profound influence on my own scientific life. My joint work with Adi Wyner on "Variable Length Markov Chains" [Bühlmann and Wyner (1999)], developed during my time in Berkeley, is a tree model and certainly inspired by CART [Breiman et al. (1984)]. Similarly, a tree-based GARCH model with Francesco Audrino [Audrino and Bühlmann (2001)] is an adaptation of CART. Much more obvious is the connection of my joint work with Bin Yu on Bagging and Boosting [Bühlmann and Yu (2000, 2002, 2003, 2006)]: it was Leo's excitement and his great ideas that stimulated my curiosity and my interest in these techniques and more generally in machine learning. My latest example is some joint work with Nicolai Meinshausen: what we call "Stability Selection" [Meinshausen and Bühlmann (2010)] is Leo Breiman's idea of Bagging, transferred from the problem of making predictions to the notion of variable and feature selection.

Leo Breiman, the pioneer of statistical machine learning: without him, my scientific life would have gone a different way, and I am tremendously thankful that I had the chance to know him personally.

Acknowledgments. I would like to thank Fred Hamprecht and Markus Kalisch for thoughtful comments.

REFERENCES

AMIT, Y. and GEMAN, D. (1997). Shape quantization and recognition with randomized trees. *Neural Comput.* **9** 1545–1588.

AUDRINO, F. and BÜHLMANN, P. (2001). Tree-structured generalized autoregressive conditional heteroscedastic models. *J. Roy. Statist. Soc. Ser. B* **63** 727–744. MR1872063

BIAU, G., DEVROYE, L. and LUGOSI, G. (2008). Consistency of random forests and other averaging classifiers. *J. Mach. Learn. Res.* **9** 2015–2033. MR2447310

BREIMAN, L. (1996a). Bagging predictors. *Mach. Learn.* **24** 123–140.

BREIMAN, L. (1996b). Heuristics of instability and stabilization in model selection. *Ann. Statist.* **24** 2350–2383. MR1425957

BREIMAN, L. (1998). Arcing classifiers (with discussion). *Ann. Statist.* **26** 801–849. MR1635406

BREIMAN, L. (1999). Prediction games and arcing algorithms. *Neural Comput.* **11** 1493–1517.

BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, CA. MR0726392

BÜHLMANN, P. and WYNER, A. J. (1999). Variable length Markov chains. *Ann. Statist.* **27** 480–513. MR1714720

BÜHLMANN, P. and YU, B. (2000). Discussion of "Additive logistic regression: A statistical view" by J. Friedman, T. Hastie and R. Tibshirani. *Ann. Statist.* **28** 377–386.

BÜHLMANN, P. and YU, B. (2002). Analyzing bagging. *Ann. Statist.* **30** 927–961. MR1926165

BÜHLMANN, P. and YU, B. (2003). Boosting with the L_2 loss: Regression and classification. *J. Amer. Statist. Assoc.* **98** 324–339. MR1995709

BÜHLMANN, P. and YU, B. (2006). Sparse boosting. *J. Mach. Learn. Res.* **7** 1001–1024. MR2274395

DIAZ-URIARTE, R. and ALVAREZ DE ANDRES, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* **7** 1–25.

FREUND, Y. and SCHAPIRE, R. E. (1996). Experiments with a new boosting algorithm. In *Proceedings of the Thirteenth International Conference on Machine Learning* 148–156. Morgan Kaufmann, San Francisco, CA.

FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. MR1873328

FRIEDMAN, J. H., HASTIE, T. and TIBSHIRANI, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion). *Ann. Statist.* **28** 337–407. MR1790002

LIN, Y. and JEON, Y. (2006). Random forests and adaptive nearest neighbors. *J. Amer. Statist. Assoc.* **101** 578–590. MR2256176

MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection (with discussion). *J. Roy. Statist. Soc. Ser. B* **72** 417–473.

MENZE, B. H., KELM, B. M., MASUCH, R., HIMMELREICH, U., BACHERT, P., PETRICH, W. and HAMPRECHT, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* **10** 1–16.

SEMINAR FÜR STATISTIK
ETH ZENTRUM, HG G17
CH-8092 ZURICH
SWITZERLAND
E-MAIL: buhlmann@stat.math.ethz.ch