Title
Reply to Nock and Nielsen: On the work of Nock and Nielsen and its relationship to the additive tree

Journal
Proceedings of the National Academy of Sciences of the United States of America, 117(16)

ISSN
0027-8424

Authors
Valdes, Gilmer
Luna, José Marcio
Gennatas, Efstathios D
et al.

Publication Date
2020-04-21

DOI
10.1073/pnas.2002399117

Peer reviewed

LETTER

LETTER

REPLY TO NOCK AND NIELSEN: On the work of Nock and Nielsen and its relationship to the additive tree

Gilmer Valdes[a,b,1], José Marcio Luna[c], Efstathios D. Gennatas[a], Lyle H. Ungar[d], Eric Eaton[d], Eric S. Diffenderfer[c], Shane T. Jensen[e], Charles B. Simone II[f], Jerome H. Friedman[g], and Timothy D. Solberg[a]

The observation that decision trees are boosting algorithms, as cited in our work (1) and acknowledged by Nock and Nielsen (2), was first established by refs. 3 and 4. This was later used by refs. 5 and 6 to develop, to the best of our knowledge, the first decision tree algorithms based purely on boosting. This work, cited in our article, precedes refs. 7 and 8 cited by Nock and Nielsen (2). The original and important contributions of refs. 7 and 8 as they pertain to this discussion was to theoretically prove convergence rates for decision tree algorithms built with boosting, along with the generalization that all decision tree algorithms have an equivalent boosting algorithm. This important theoretical result applies to the AddTree (1), and we thank the authors for bringing it to our attention, particularly as it provides readers with a deeper understanding of our contribution.

Nock and Nielsen (2) indicate they were the first to establish a theoretical connection between additive models (represented by boosting) and full interaction models (represented by decision trees) in a master algorithm. However, this neglects the established connection between single decision trees and boosting (3–6). In contrast, our claim that we discovered a connection between additive and full interaction models does not rely on the fact that decision trees are boosting algorithms. Our central result is that these two models can be joined in a master algorithm by a single parameter, lambda, that controls the weight decay of the observations during recursive partitioning.

We 1) show how Classification and Regression Trees (CART) can be considered the greediest version of GBS (lambda equal 0) with the highest variance (figure 4 in ref. 1), and consequently the lowest accuracy on average, and 2) design an algorithm, the additive tree (AddTree), that exploits this parameter to obtain the same topology as CART (which makes it interpretable) but improves its accuracy on expectation (by effectively controlling the bias–variance tradeoff). Although other decision tree algorithms have been designed that improved CART accuracy (9), including the oblique decision tree mentioned in the letter, they had not done so by maintaining the topology and as such the interpretability.

There is another point worth highlighting from the letter. The AddTree allows any type of partition in the leaves (including oblique) and model at the terminal nodes, and thus the previous decision tree algorithm proposed by Nock and Nielsen is not more general than the AddTree. It is more general than the version of the AddTree investigated by us in ref. 1, which was chosen to only include stumps in the partitions to result in the same topology as CART and keep its interpretability. In closing, we thank all of the researchers who contributed to efforts on which our work is based, whether explicitly cited in our article or not.

1 J. M. Luna *et al.*, Building more accurate decision trees with the additive tree. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19887–19893 (2019).
2 R. Nock, F. Nielsen, The phylogenetic tree of boosting has a bushy carriage but a single trunk. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8692–8693 (2020).
3 M. Kearns, Y. Mansour, "On the boosting ability of top-down decision tree learning algorithms" in *Proceedings of the Twenty-Eight Annual ACM Symposium on Theory of Computing* (Assoc. for Computing Machinery, New York, 1996), pp. 459–468.

[a]Department of Radiation Oncology, University of California, San Francisco, CA 94115; [b]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94115; [c]Department of Radiation Oncology, University of Pennsylvania, Philadelphia, PA 19104; [d]Department of Computing and Information Science, University of Pennsylvania, Philadelphia, PA 19104; [e]Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104; [f]Department of Radiation Oncology, New York Proton Center, New York, NY 10035; and [g]Department of Statistics, Stanford University, Stanford, CA 94305

Author contributions: G.V., J.M.L., E.D.G., L.H.U., E.E., E.S.D., S.T.J., C.B.S., J.H.F., and T.D.S. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: gilmer.valdes@ucsf.edu.

First published April 7, 2020.

4 Y. Freund, R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to Boosting" in Computational Learning Theory. EuroCOLT 1995. P. Vitányi, Ed. (Lecture Notes in Computer Science, Springer, Berlin, 1995), vol. 904, pp. 23–37.
5 E. Grossmann, "AdaTree: Boosting a weak classifier into a decision tree" in *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, Z. Zhu, R. Kumar, Y.-P. Hung, R. Haralick, A. Hanson, Eds. (IEEE, 2004), p. 105.
6 Z. Tu, "Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering" in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, S. Ma, H.-Y. Shum, W. T. Freeman, L. Van Gool, S. Chaudhuri, Eds. (IEEE, 2005), vol. 2, pp. 1589–1596.
7 C. Henry, R. Nock, F. Nielsen, "Real boosting a la carte with an application to boosting oblique decision trees" in *IJCAI International Joint Conference on Artificial Intelligence*. (IJCAI, 2007), pp. 842–847.
8 R. Nock, F. Nielsen, Bregman divergences and surrogates for learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 2048–2059 (2009).
9 W.-Y. Loh, Improving the precision of classification trees. *Ann. Appl. Stat.* **3**, 1710–1737 (2009).