PHSX 499: Selected Paper Outline v.2

William Jardee

March 3, 2022

This is an analysis of the work published in 2020 by Thibaut Vidal and Maximilian Schiffer, **Born-Again Tree Ensembles**. I go through section by section and highlight what I believe is the "topic sentence" of the paragraph. In some cases there are a couple topic sentences for a section, it should be clear why I chose multiple. The paper can be found at citation [1].

1 Introduction

"Tree ensembles constitute a core technique for prediction and classification tasks."
"Currently, there exists a trade-off between the interpretability and the performance of tree (ensemble) classifiers."

1.1 State of the Art

Thinning tree ensembles

"Thinning tree ensembles has been studied from different perspectives and divides in two different streams, i) classically thinning a tree ensemble by removing some weak learners from the original ensemble and ii) replacing a tree ensemble by a simpler classifier, e.g., a single decision tree."
"In the field of neural networks, related studies were done on *model compression.*"

Decision trees

"since the 1990's, some works focused on constructing decision trees based on mathematical programming techniques."

1.2 Contributions

"With this work, we revive the concept of born-again tree ensembles and aim to construct a single -minimum-size-tree that faithfully reproduces the decision function of the original tree ensemble."

2 Fundamentals

"In this section, we introduce some fundamental definitions."

3 Methodology

"In thiss section, we introduce a dynamic programming (DP) algorithm which optimally solves Problem 1 for many data sets of practical interest."

4 Computational Experiments

"The goal of our computational experiments is fourfold: ..." *The authors then go on to outline the four experiments that are conducted.*

4.1 Data Preparation

"We focus on a set of six datasets from the UCI machine learning repository [UCI] and from previous work by ..."

4.2 Computational Effort

"In a firsts analysis, we evaluate the computation time of Algorithm 1 for different data sets and size metrics."
"In our second analysis, we focus on the FICO case and randomly extract subsets of samples and features to produce smaller data sets."
"We observe that the computational time fo the DP algorithm is strongly driven by the number of features, with an exponential growth relative to their parameter."

4.3 Complexity of the Born-Again Trees

"We now analyze the depth and number of leaves of the born-again trees for different objective functions and datasets in Table 2."

4.4 Post-Pruned Born-Again Trees

"To circumvent this issue, we suggest to apply a simple post-pruning step to eliminate inexpressive tree sub-regions."

4.5 Heuristic bone-Again Trees

"Accordingly, we take a first step towards scalable heuristic algorithms in the following."

5 Conclusions

"In this paper, we introduced an efficient algorithm to transform a random forest into a single, smallest possible, decision tree."
"As a perspective for future work, we recommend to progress further on solution techniques for the born-again tree ensembles problem, proposing new optimal algorithms to effectively handle larger datasets as well as fast and accurate heuristics."

References

[1] Thibaut Vidal and Maximilian Schiffer. Born-again tree ensembles. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9743–9753. PMLR, 13–18 Jul 2020.