Causal Forest vs. Naïve Causal Forest in Detecting Personalization: An Empirical Study in ASSISTments

Biao Yin
Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA 01609
byin@wpi.edu

Anthony F. Botelho
Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA 01609
abotelho@wpi.edu

Thanaporn Patikorn
Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA 01609
tpatikorn@wpi.edu

Neil T. Heffernan
Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA 01609
nth@wpi.edu

Jian Zou
Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA 01609
jzou@wpi.edu

ABSTRACT

It is widely understood that students learn in a variety of different ways and what is beneficial for one student may not necessarily help another. This work observes the effectiveness of Causal Forests as they compare to a new method we present called Naïve Causal Forests. This new method, aimed to be a simpler, more intuitive approach to identifying heterogeneous effects, is developed to better understand the strengths and limitations of the Causal Forest method. We apply these techniques to real student data on three RCTs run within the ASSISTments online learning platform.

Keywords

Personalization, Heterogeneous Treatment Effects, Randomized Controlled Trials, Causal Forest, Random Forest

1. INTRODUCTION

The idea that students approach learning in differing ways is not a new concept to researchers in the field of education, but how to leverage these computer-based systems for individualized learning is not always clear. Individualization, also referred to as personalization, also exists outside the field of education as well. In other fields, this idea is described through heterogeneous treatment effects, as the effect of a particular treatment or intervention is not often homologous across all individuals. The introduction of computer-based systems in the classroom makes it feasible to supply aid to individuals allowing the teacher to focus on helping those students struggling most.

Recently, a technique known as a Causal Forest (CF) [8] has been developed, applying random forests to the task of identifying heterogeneous effects. This work explores a new, more intuitive method for identifying heterogeneity as it compares to the more complex CF method. This new method, called Naïve Causal Forest (NCF), attempts to employ a simpler approach based on the structure of CF to answer: 1. To what extent, if any, does the Causal Forest method outperform our simpler, more intuitive approach to identifying heterogeneous treatment effects in real student data? and 2. Do these models converge to large differences when compared using increasing sample sizes?

2. DATASET

The dataset used to build and evaluate our method is comprised of student information on 3 randomized control trials (RCTs) run within the ASSISTments online learning platform [2] from a previously published dataset [5]. ASSISTments is a free web-based platform where a recent efficacy trial found the system to be effective in improving student learning [4], motivating further study to better understand student behavior and measure effects within the platform.

After filtering the data to remove students with missing values, the Experiment 1 contains 519 students, the Experiment 2 contains 833 students, and Experiment 3 contains 1118 students.

3. METHODOLOGY

The Causal Forest (CF) method [8] has established itself as a viable model for identifying heterogeneous effects, for which we do not refute, but rather we wish to explore the benefits of this more complex method to a simpler, more intuitive approach. CF uses estimates of treatment effects within the splitting rule of a random forest algorithm; an "honest" variant uses a holdout set to estimate the effect for each split. Heterogeneous effects can be determined by observing students who then are grouped into different leaves of the generated trees. Our new method, which we have called Naïve Causal Forest, aims to implement a simpler approach that excludes the use of condition from the random forest until students are grouped into each leaf, where then an average treatment effect is calculated across each subgroup. In both methods, each tree has a "vote" as to what condition will benefit the students most.

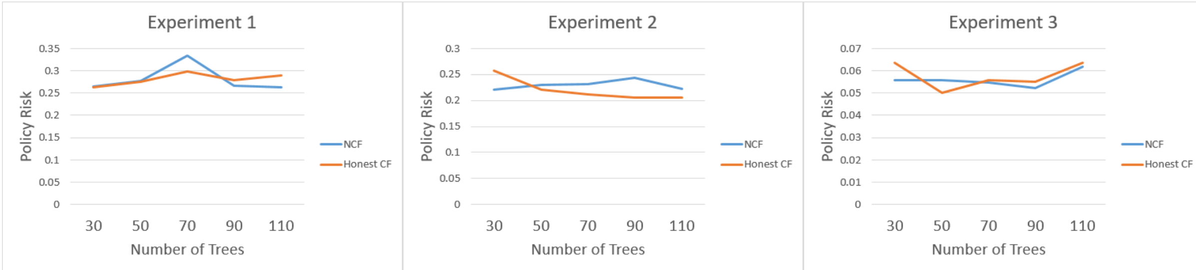

Figure 1: The 10-fold cross validation results for experiments 1 and 2 comparing NCF to an honest CF model. No reliable differences are found between the two methods, and both appear consistent with increases to the number of generated trees.

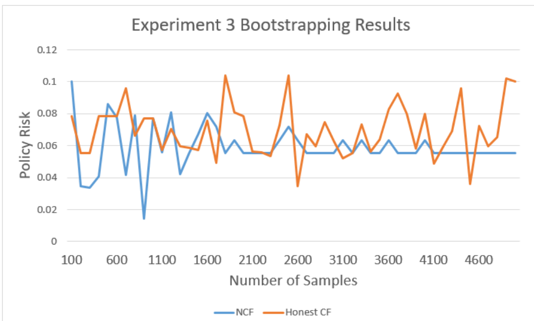

Figure 2: Experiment 3 bootstrapping results comparing NCF to two Causal Forest models.

We compare CF, implemented in R [3] using a Causal Tree package [1], and NCF in their ability to identify heterogeneous effects for the purpose of maximizing completion of the assignment. We calculate the Odds Ratio [7] within each leaf to identify which condition corresponds with the higher student completion rate within each leaf. We evaluate our models using a measure known as policy risk [6], where a lower value indicates better performance. This metric is used to compare the two methods for each experiment as the metric is not directly comparable across experiments.

4. DISCUSSION AND FUTURE WORK

The result of our 10-fold cross validation analysis can be seen in Figure 1. Both models use a minimum leaf size of 30, and are evaluated over several model complexities. In all three experiments, it is found that the CF and NCF model exhibit no reliable differences. It is also the case, however, that no significant heterogeneous effects are found by either method. Figure 2 illustrates how the methods converge with increasing sample sizes using a bootstrapping method of sampling with replacement on the largest experiment.

We compare in this work the Causal Forest method for identifying heterogeneous treatment effects to our Naïve Causal Forest method and find no reliable differences between the simpler and more complex methods. It is expected, and planned for future work, that applying these methods to experiments with larger sample sizes may show statistic reliability.

We also found that the CF model exhibited stable policy risk over increases to model complexity. This is a desirable quality of a prediction model, as it is data driven and less sensitive to changes in model structure. We found that the CF model exhibited non-converging behavior when bootstrapping, but may additionally be caused by insufficient variation or lack of heterogeneity in the dataset.

5. ACKNOWLEDGMENTS

We thank multiple current NSF grants (IIS-1636782, ACI1440753, DRL-1252297, DRL-1109483, DRL-1316736, DGE-1535428 & DRL-1031398), the US Dept. of Ed (IES R305A120125 & R305C100024 and GAANN), and the ONR.

6. REFERENCES

[1] S. Athey, G. Imbens, and Y. Kong. *causalTree: Recursive Partitioning Causal Trees*, 2016. R package version 0.0.

[2] N. T. Heffernan and C. L. Heffernan. The assistments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4):470–497, 2014.

[3] R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[4] J. Roschelle, M. Feng, R. F. Murphy, and C. A. Mason. Online mathematics homework increases student achievement. *AERA Open*, 2(4), 2016.

[5] D. Selent, T. Patikorn, and N. Heffernan. Assistments dataset from multiple randomized controlled experiments. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*, pages 181–184. ACM, 2016.

[6] U. Shalit, F. Johansson, and D. Sontag. Estimating individual treatment effect: generalization bounds and algorithms. *arXiv preprint arXiv:1606.03976*, 2016.

[7] M. Szumilas. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19:227, 2010.

[8] S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *arXiv preprint arXiv:1510.04342*, 2015.