TTML Project Proposal

Yijie Guo , Wenbo Jing, Mu Li, Zichang Ye

yg2418@nyu.edu
ml4844@nyu.edu
wj2093@nyu.edu
zy1545@nyu.edu

March 2021

1 Introduction

In the class, we have discussed the S-Learner, the T-Learner and the X-Learner as the Meta-algorithms for the CATE estimation. In this project, we would dive deeper into these CATE estimators. First, we would reproduce some results from the authors' simulations and the Get-Out-The-Vote (GOTV) experiment. Then, we aim to improve the X-Learner's performance using complex base learners and reduce the biases of the CATE estimators via transforming the input space. Lastly, if we obtain good results from the experiments, we could further test our improvements on new datasets and attempt to provide theoretical proofs.

2 Experiment Results Reproduction

The authors claimed that their X-Learner with the honest RF and BART would outperform S-Learner and T-Learner to provide the lowest Mean Squared Error (MSE) under certain synthetic cases and proved it again in their GOTV experiment. Thus, we would reproduce some of their simulation studies and the GOTV experiment as well, using those results as the baseline for our methods with different base learners to offer better performance.

Next, we decide to reproduce the CI-Simulation 1 and CI-Simulation 2. These two simulations were the experiments that the authors tried to compute the 95% confidence intervals (CI) using the bootstrap given the GOTV data set. However, they claimed that the resulting CI could not achieved nominal coverage due to the biases from the CATE estimators [2]. Therefore, we want to run these two simulations as well to establish a baseline for our methods in reducing the biases.

3 Performance Improvements

3.1 X-Learner Using Other Base Learners

While the original X-Learner using the honest RF and BART has performed well, we decide to further improve it by changing the base learner to other algorithms, like the regular RF, stochastic gradient boosting (boosting), support vector machines (SVMs) and the neural network (NN), such that we want to acquire lower MSE scores, or obtain better convergence rate in the same GOTV experiment.

3.2 Reduce CATE Estimator Biases

In addition, we would like to transform the input space from continuous data to discrete data. Generally, the discretization or the binarization would reduce biases as they work as regularizers. In this case, we would also like to see if these transformations would adjust the biases from the CATE estimators, such that the resulting bootstrap confidence intervals can achieve the nominal coverage.

4 Potential Next Steps

Depending on the results from Section 3 experiments, we would like to conclude with theoretical explanations as we understand the reasons behind these experiment results. Also, [1] mentions that the original paper fails to consider the effect of selection bias. It would be valuable for us to examine whether selection bias could significantly harm our model's performance.

5 Preliminary Results

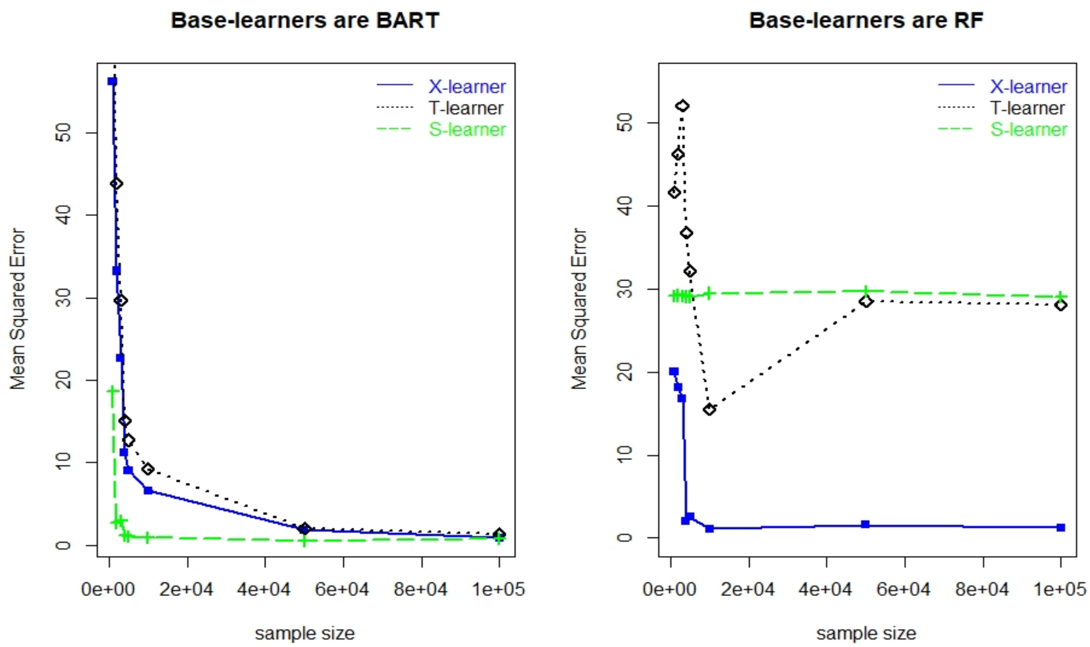
We started by reproducing the first and second simulation studies (referred to as SI1 and SI2) in [2], which compared the performance of the meta-learners under unbalanced piece-wise linear setting and balanced complex linear setting, respectively. Our reproduced results are shown in Figure 1, which mainly reproduced the patterns of the original results.

We also attempted to replicate the performances of these learners on the GOTV dataset, however using a subset of 10,000 data points to prove the concept. We presented the variance, bias, and RMSE of the learners in Figure 2. The plots of variance and RMSE seem to resemble the original figure from the paper, yet the plot of bias doesn't. We intend to explore this issue in the following weeks.

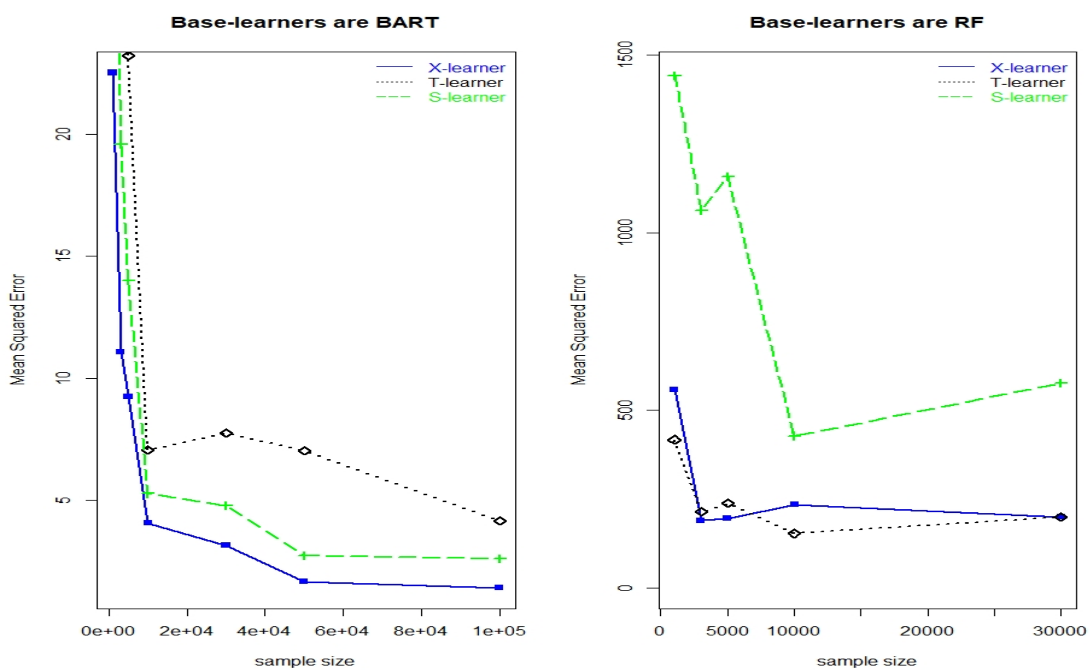
References

[1] Ahmed Alaa and Mihaela Schaar. Limits of estimating heterogeneous treatment effects: Guidelines for practical algorithm design. pages 129–138, 2018.

[2] Sören R. Künzel, Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, 2019.

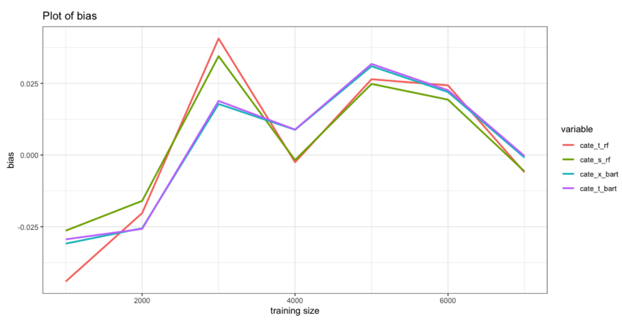


(a)

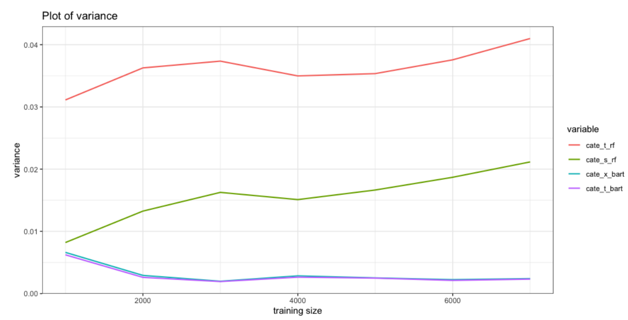


(b)

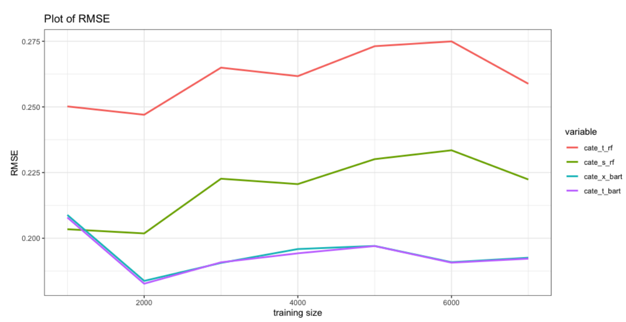
Figure 1: Mean Squared Error of Meta-learners in (a) SI1 (b) SI2 against sample size



(a)



(b)



(c)

Figure 2: Replication of Performance on Subset of GOTV