Generalized gradient boosting for causal inference

Pingfan Su[1], Joakim Andersen[2], Qingyuan Zhao[2], Chengchun Shi[1]

[1]London School of Economics [2]Cambridge University

Objectives

- goal: learn the causal effects of treatment variables on the outcome from confounded data.
- algorithm: we propose ggboosting, a computationally cheap approach which generalizes to all problems with confounding factors.
- performance: can reach the state-of-the-art methods and can further boost the performances for a wide range of initialization models.

Motivation

Traditional methods such as regression analysis primarily indicates correlation while unmeasured confounders can lead to spurious correlation. A perfect way to eliminate impacts from confounding is Randomized Controlled Trials (RCTs). However, RCTs are not always feasible in practice due to various restrictions. IV have become one of the most popular method for scientists who work on discovering causal effects. The idea behind IV is intuitive: find an instrument Z that only influences Y via X and uncorrelated with ϵ.

$$Y = g(X) + \epsilon, \quad \mathbb{E}[\epsilon|Z] = 0$$

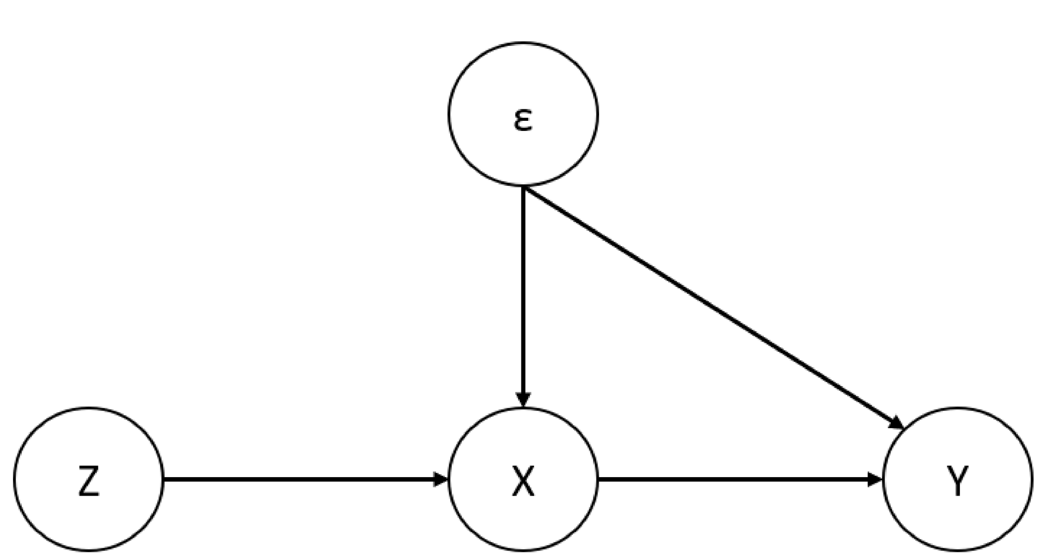

Figure 1:IV DAG

Gradient boosting

- Accuracy: Can achieve high accuracy in prediction tasks.
- Scalability: Can be parallelized and scaled to handle large datasets.
- Flexibility: Can be applied to a wide range of problems.

$$\hat{g} = \arg\min_g \sum_{i=1}^{n} L(y_i, g(x_i)) = L(g)$$

$$g_t = g_{t-1} - \eta \nabla L(g)|_{g_{t-1}}$$

$$\hat{g} = \sum_{m=0}^{M} g_m, \quad g_m \in \mathbb{R}^n$$

To avoid overfit, fit a base learner (e.g. decision trees) to the gradient and update on the predictions from the learner.

Algorithms

$$\{\hat{g}, \hat{h}\} = \arg\min_g \max_h |\mathbb{E}[Y - g(X)h(Z)]|$$

$$\{\hat{g}, \hat{h}\} = \arg\min_g \max_h \frac{1}{n}\sum_{i=1}^{n} |(y_i - g(x_i))h(z_i)|$$

$$= \arg\min_g \max_h \frac{1}{n} L(g, h)$$

Learn target function $g(\cdot)$ by gradient boosting:

$$g_t = g_{t-1} - \eta \nabla L(g, h)|_{g_{t-1}}$$

Flexible function classes for $h(\cdot)$:

Boosting trees:

$$h_t = h_{t-1} + \eta \nabla L(g, h)|_{h_{t-1}}$$

RKHS:

$$\hat{g} = \arg\min_g \mathbb{E}[(Y - g(X))(Y' - g(X'))k(Z, Z')]$$

NNs:

$$\theta_t = \theta_{t-1} + \eta \nabla_\theta L(g, h)$$

Results

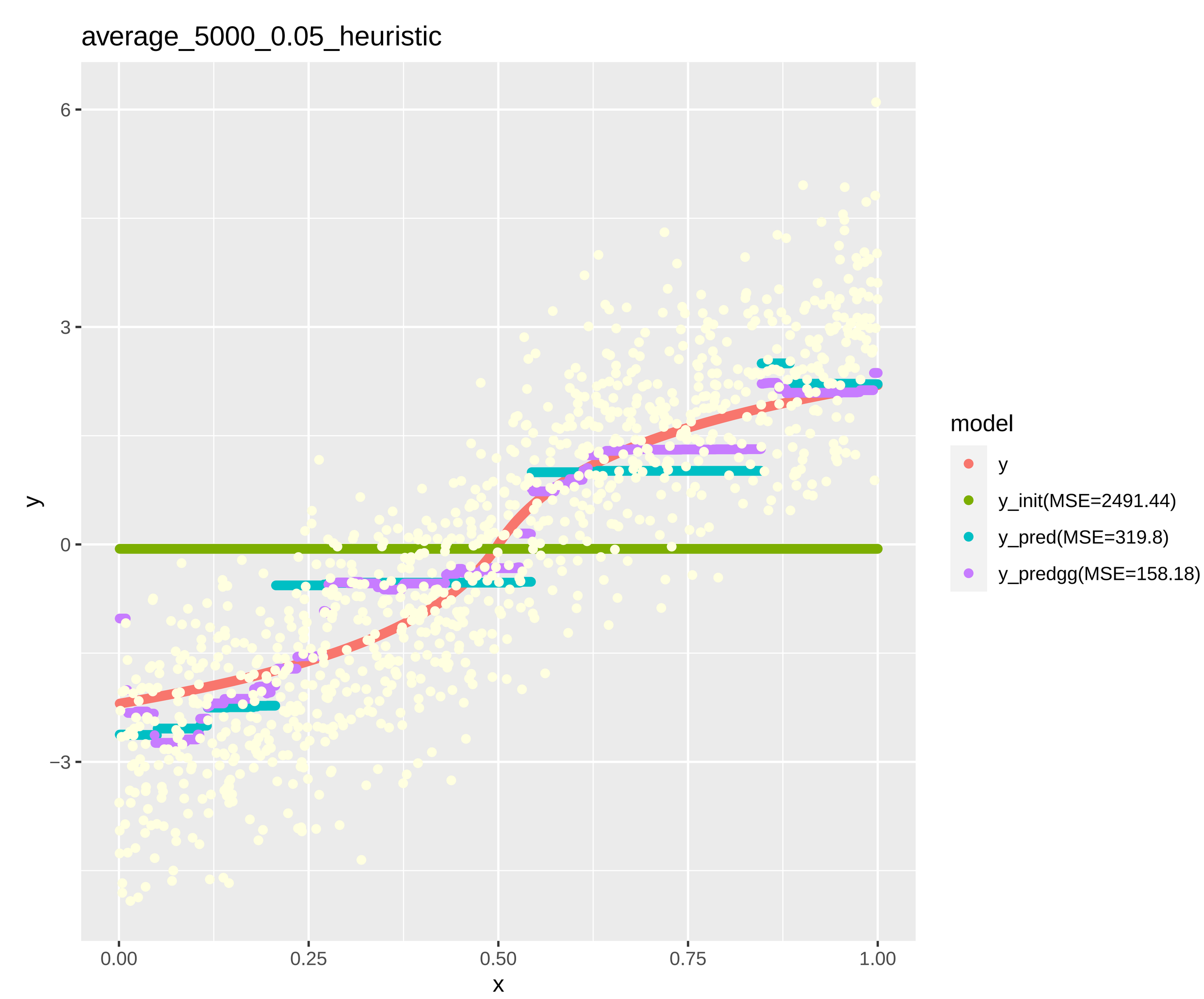

Figure 2:Average initialization

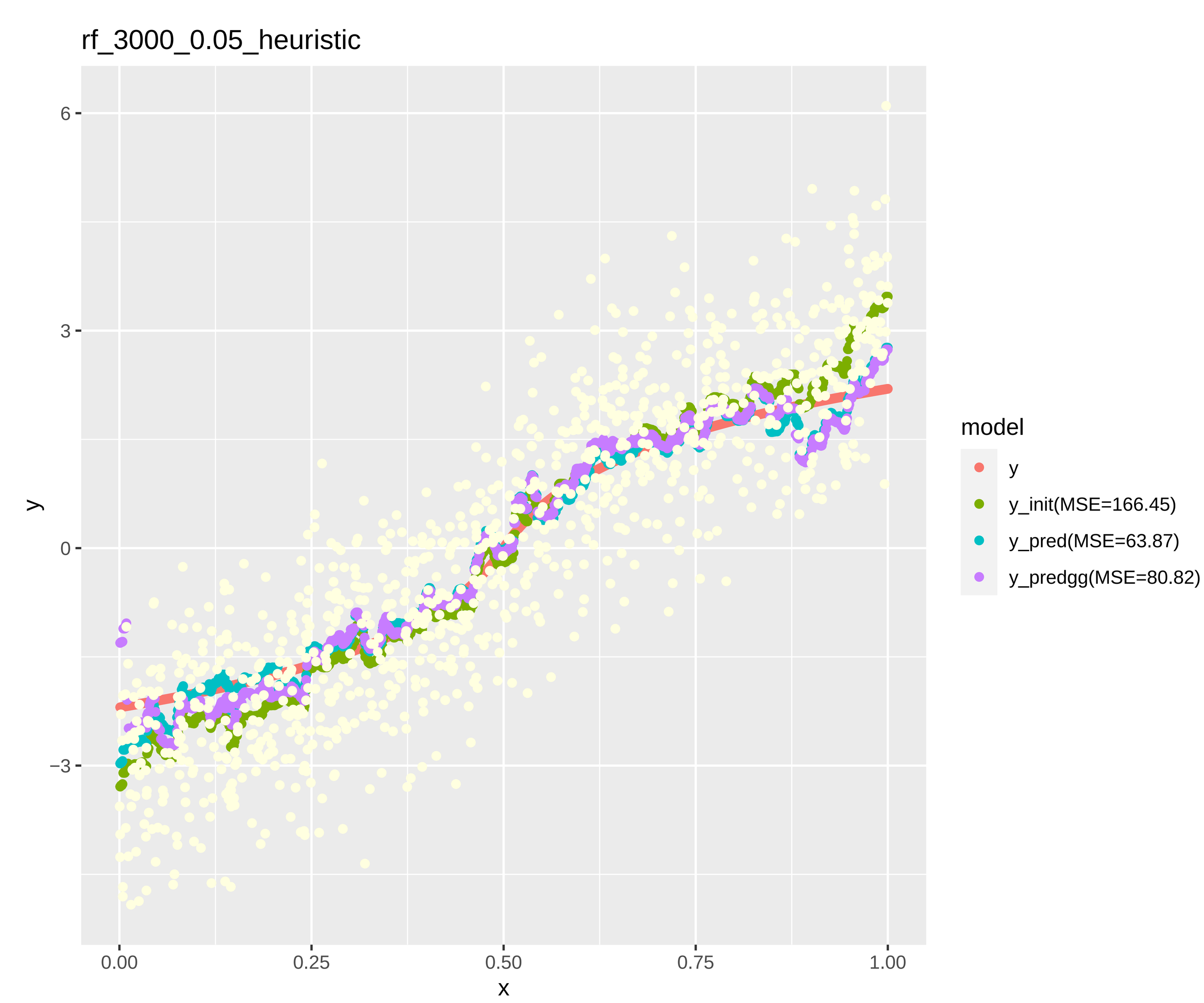

Figure 3:Random forest initialization

Simulation design

The underlying structural function is

$$g(x) = \ln(|16x - 8| + 1) * sgn(x - 0.5)$$

and data are drawn from

$$\begin{pmatrix}\epsilon \\ V \\ W\end{pmatrix} \overset{\text{i.i.d}}{\sim} N\left(\begin{pmatrix}0 \\ 0 \\ 0\end{pmatrix}, \begin{pmatrix}1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1\end{pmatrix}\right)$$

$$X = \Phi\left(\frac{W + V}{\sqrt{2}}\right)$$

$$Z = \Phi(W)$$

Consistency

$\mathbb{E}[|Y|^2] < \infty$, $\mathbb{E}[\sup_{f\in\mathcal{F}} |f(X)|^2] < \infty$, $\mathcal{F}$ is compact and convex, then $\hat{f} \xrightarrow{P} f^*$

Conclusion

We propose a generalized gradient boosting framework for estimating causal effects where unmeasured confounding exists. Different from other related works, our method gives the first discrete estimators for learning causal relationship and can generalize to a wide range of function classes. Our method performs well on the simulated causal problem and show the potential for improving other estimators.

References

[1] Imbens Guido Angrist Joshua and Rubin Donald. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

[2] Lewis Greg et al Dikkala Nishanth. Minimax estimation of conditional moment models. *Advances in Neural Information Processing Systems*, 33:12248–12262, 2020.

[3] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.