ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
HAROKOPIO UNIVERSITY

.dit/*
Informatics & Telematics

C-XGBoost: A tree boosting model for causal effect estimation

Niki Kiriakidou, Ioannis E. Livieris and Christos Diou

Harokopio University of Athens, Department of Informatics and Telematics

University of Pireaus, Department of Statistics & Insurance Science

27 – 30 June, 2024

AIAI ARTIFICIAL INTELLIGENCE APPLICATIONS & INNOVATIONS

Table of contents

Causal inference

Causal inference

- Causal inference has been a crucial research topic in many scientific fields such as medicine, education, economics, engineering

Primary goal

- Identification of the effect of an intervention, which is often called “treatment” to an outcome

Challenge

- This process requires the study of the difference between alternative choices of intervention (not possible to study directly).
- The only available observed outcome is the one of the action actually taken, while the rest remain unknown (counterfactual outcomes).

Problem

Problem

- Estimation of treatment effects from observational data.

Average Treatment Effect (ATE)

$$\psi = E[Y|X,T = 1] - E[Y|X,T = 0],$$

where

· X is a d −dimensional space of covariates

· T is the binary treatment

· Y is the outcome for a single sample

Notations

Y_0 denotes the outcome for a sample when $T = 0$.

Y_1 denotes the outcome for a sample when $T = 1$.

- If $T = 0$ then:
  - Y_0 **factual** outcome
  - Y_1 **counterfactual** outcome

- If $T = 1$ then:
  - Y_1 **factual** outcome
  - Y_0 **counterfactual** outcome

- In each of the above cases we need to **estimate** the <u>counterfactual outcome</u>.

Related Work

NN-based models

- **TARNet (Treatment Agnostic Representation Network)**

Shalit, U., Johansson, F. D., & Sontag, D. (2017). Estimating individual treatment effect: generalization bounds and algorithms. In *International Conference on Machine Learning* (pp. 3076-3085). PMLR.

- **Dragonnet**

Shi, C., Blei, D., & Veitch, V. (2019). Adapting neural networks for the estimation of treatment effects. *Advances in neural information processing systems*, *32*.

- **kNN-Dragonnet**

Kiriakidou, N., & Diou, C. (2023). Integrating Nearest Neighbors with Neural Network Models for Treatment Effect Estimation. *International journal of neural systems*, *33*(7).

Tree-based models

- **R-Forest**

Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, *116*(10).

- **C-Forest**

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523).

Motivation

Current status

- NN-based models constitute the most effective causal inference models.
- *Advantage*: Lack of requirement of any information about the treatment value for the estimation of the potential outcomes due to their sophisticated architecture design.

NN-based models ***superiority*** on tabular data is under question

- NN-based models are usually biased to overly smooth solutions and toward low-frequency functions, while decision tree-based models are better at handling irregular patterns in the data (non-smooth target functions).
- NN-based models are considerably affected by uninformative features in contrast to tree-based models.

Grinsztajn, L., Oyallon, E., Varoquaux, G.: Why do tree-based models still outperform deep learning on typical tabular data? Advances in Neural Information Processing Systems 35, 507–520 (2022)

Proposed approach

Contribution

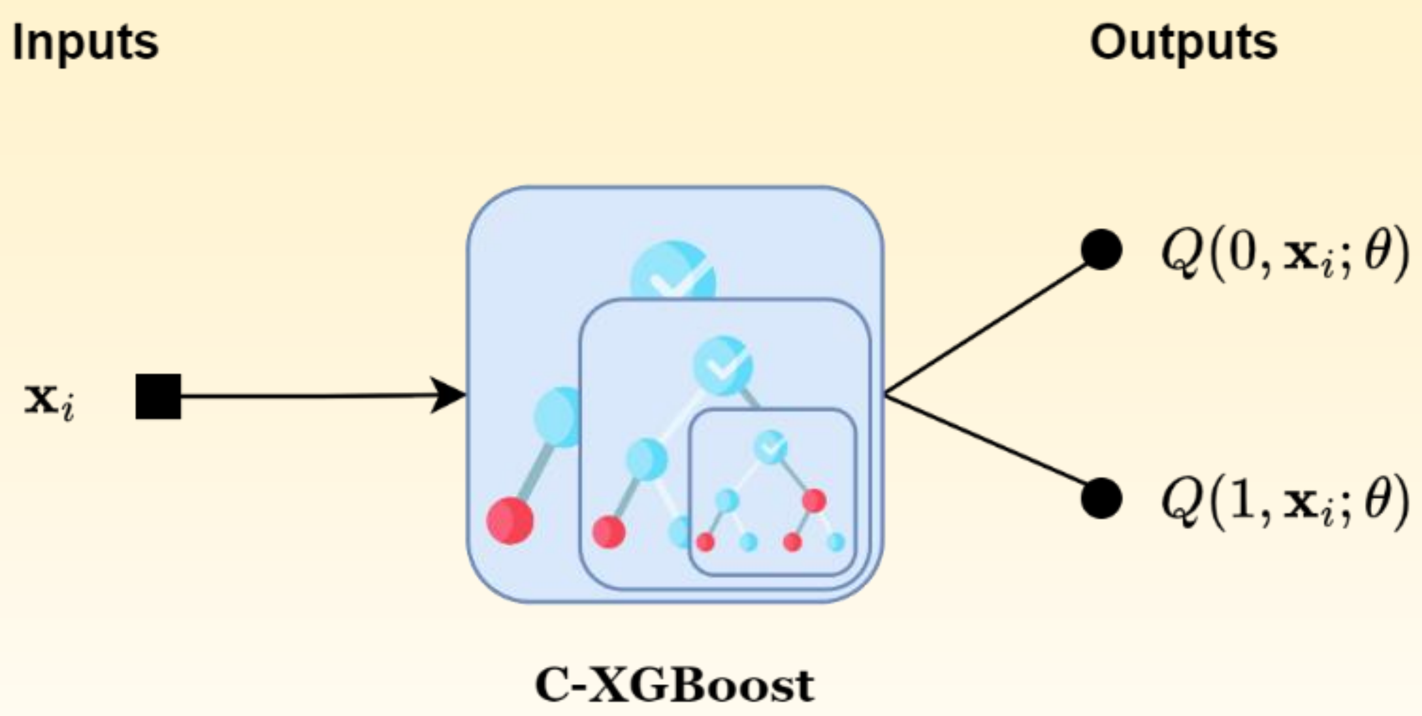
- **C-XGBoost:** A new causal inference model for the prediction of potential outcomes
- New loss function for training the proposed C-XGBoost model

Motivation

- Exploit the strong prediction abilities of eXtreme Gradient Boosting (XGBoost) algorithm
- The notable property of NN-based causal inference models to learn representations that are useful for estimating the outcome for both the treatment and non-treatment cases.

Scope

- *Reduce bias* and *increase the prediction accuracy*



- Inputs
  - X: vector of covariates

- Outputs
  - $Q(1, x_i; \theta)$: conditional outcome for treatment group
  - $Q(0, x_i; \theta)$: conditional outcome for control group

New Loss Function

- For training C-XGBoost model, we define the following loss function, which constitutes a modification of mean square loss:

$$\mathcal{L}(\mathbf{X};\theta) = \frac{1}{n}\sum_i \mathcal{L}_1(\mathbf{x}_i;\theta) + \mathcal{L}_2(\mathbf{x}_i;\theta),$$

where

$$\mathcal{L}_1(\mathbf{x}_i;\theta) = (1-t_i)(Q(0,\mathbf{x}_i;\theta) - y_i)^2,$$
$$\mathcal{L}_2(\mathbf{x}_i;\theta) = t_i(Q(1,\mathbf{x}_i;\theta) - y_i)^2.$$

Notice that since C-XGBoost does not provide automatic differentiation the gradient of $\mathcal{L}$ is calculated by

$$\frac{\partial \mathcal{L}}{\partial Q(t_i,\mathbf{x}_i;\theta)} = \begin{cases} 2(1-t_i)(Q(0,\mathbf{x}_i;\theta) - y_i), & t_i = 0; \\ 2t_i(Q(1,\mathbf{x}_i;\theta) - y_i)), & t_i = 1. \end{cases}$$

while all the elements of the Hessian matrix are equal to 2.

Datasets

Synthetic collection of datasets

- It consists of a collection of toy causal-inference classification datasets
  Louizos, Christos, et al. "Causal effect inference with deep latent-variable models." *Advances in neural information processing systems* 30 (2017).
- Covariates: 1000
- Samples in each dataset: 5000

ACIC collection of datasets

- It consists of a collection of semi-synthetic datasets developed for the 2018 Atlantic Causal Inference Conference competition data
  Shimoni, Yishai, et al. "Benchmarking framework for performance-evaluation of causal inference analysis." arXiv preprint arXiv:1802.05046 (2018).
- For each setting in the data generation process, we randomly selected 5 and 11 datasets of size 5000 and 10000, respectively

Numerical experiments

Evaluated models

- “R-Forest” (Künzel et al., 2019)
- “C-Forest” (Wager and Athey, 2018)
- “Dragonnet” (Shi et al. 2019)
- “kNN - Dragonnet” (Kiriakidou & Diou, 2023)
- “C - XGBoost”

Evaluation methodology

- Performance profiles of Dolan & More
- Statistical analysis (FAR – Finner post-hoc test)

Performance metrics

Absolute error of the estimation of Average Treatment Effect - $|\varepsilon_{ATE}|$

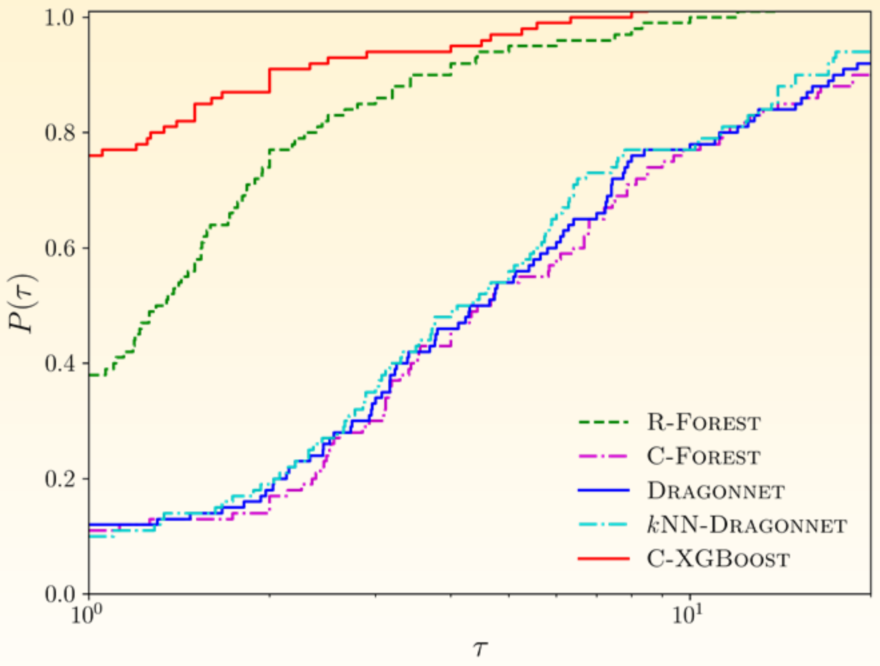
- Measures treatment effect for the average population
- Compare evaluated models as estimators

Precision in Estimation of Heterogeneous Effect - *PEHE*

- Measures treatment effect on individual level
- Compare evaluated models as predictors

Numerical results for Synthetic collection of datasets

Numerical experiments - $|\varepsilon_{ATE}|$

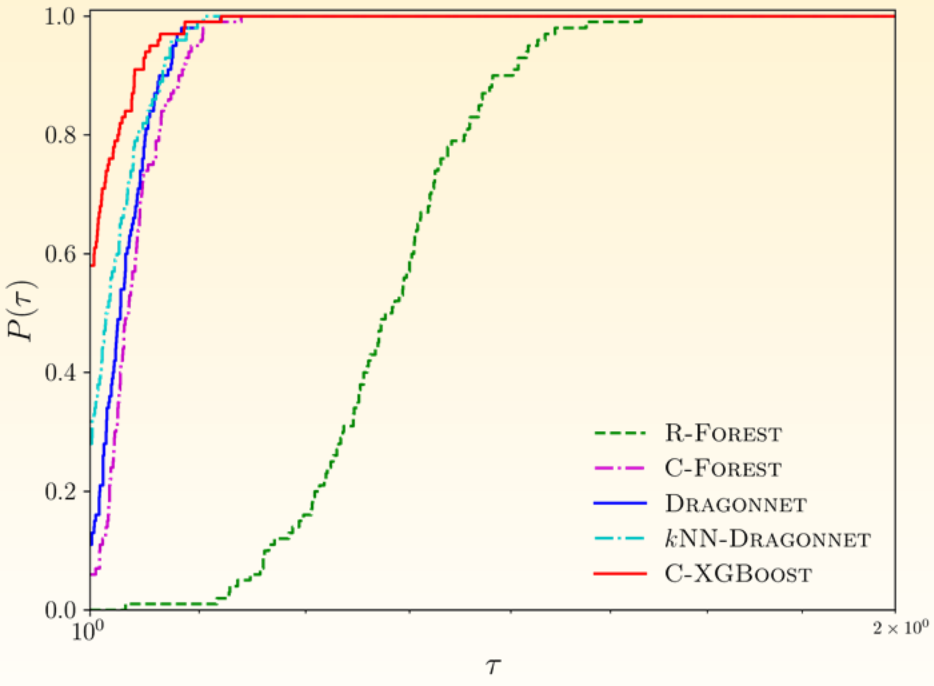


Model	FAR	Finner post-hoc test	
		p_F-value	H_0
C-XGBoost	86.41	-	-
R-Forest	125.85	0.053579	Failed to reject
kNN-Dragonnet	303.41	0.000000	Reject
Dragonnet	350.82	0.000000	Reject
C-Forest	386.02	0.000000	Reject

Finding/Conclusions

- C-XGBoost exhibit the best performance, outperforming all models
- Statistically considerable differences among the performance of C-XGBoost and kNN – Dragonnet, Dragonnet and C-Forest

Numerical experiments - *PEHE*



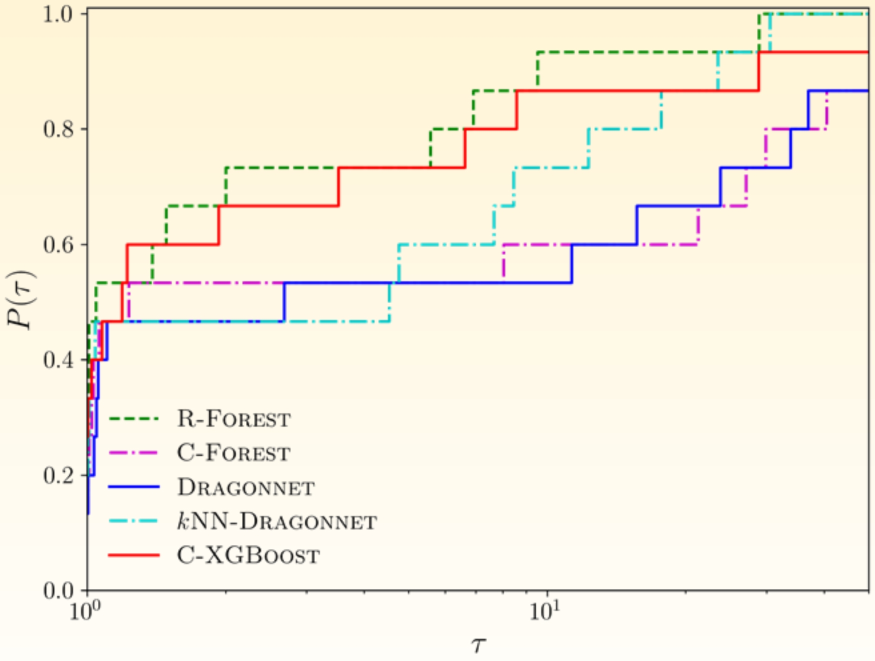
Model	FAR	Finner post-hoc test	
		p_F-value	H_0
C-XGBoost	139.96	-	-
kNN-Dragonnet	187.66	0.019582	Reject
Dragonnet	219.24	0.000139	Reject
C-Forest	255.28	0.000000	Reject
R-Forest	450.37	0.000000	Reject

Finding/Conclusions

- C-XGBoost reports the best performance in terms of both efficiency and robustness
- Significant statistical differences in the performance of proposed and SoA models

Numerical results for
ACIC collection of datasets

Numerical experiments - $|\varepsilon_{ATE}|$

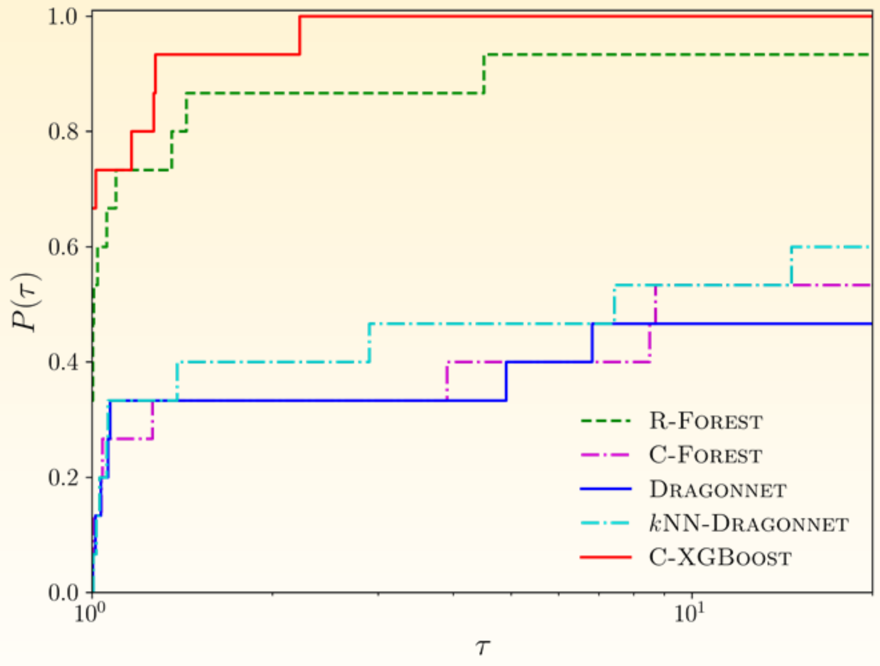


Model	FAR	Finner post-hoc test	
		p_F-value	H_0
R-Forest	24.26	-	-
C-XGBoost	26.86	0.743890	Failed to reject
kNN-Dragonnet	39.13	0.081474	Failed to reject
Dragonnet	47.93	0.005873	Reject
C-Forest	51.80	0.002161	Reject

Finding/Conclusions

- C-XGBoost and R-Forest report the best performance
- R-Forest, C-XGBoost and kNN-Dragonnet report the top ranking and there are no statistical significant differences

Numerical experiments - *PEHE*



Model	FAR	Finner post-hoc test	
		p_F-value	H_0
C-XGBoost	23.06	-	-
R-Forest	23.13	0.993316	Failed to reject
kNN-Dragonnet	41.53	0.026997	Reject
C-Forest	49.46	0.001817	Reject
Dragonnet	52.80	0.000747	Reject

Finding/Conclusions

- C-XGBoost presents the top performance in terms of robustness and efficiency
- C-XGBoost and R-Forest present the highest ranking score - There are statistical significant differences between them and the rest SoA causal inference models.

Conclusions & Future work

Contribution: C-XGBoost

- New tree-based model for the prediction of potential outcomes for causal effect estimation.
- A new loss function was proposed for efficiently training C-XGBoost model

Motivation

- Exploit the strong prediction abilities of XGBoost algorithm
- The remarkable property of NN-based causal inference models to learn representations that are useful for estimating the outcome for both the treatment and non-treatment cases.

Conclusions & Future work

Findings

- **Experimental analysis:** C-XGBoost model outperformed state-of-the-art tree-based and NN-based models on two semi-synthetic collections of datasets
- **Limitation:** considered the potential influence of hyperparameters settings to the efficiency of proposed model since its sensitivity to different configuration settings is unclear.

Future work

- Further enhancing the robustness and predictive accuracy of the proposed C-XGBoost model by incorporating a regularization procedure for inducing bias.
- Provide explainability functionalities and tools and provide insights into the decision-making process.

Thank you for your attention

Contact info:

Niki Kiriakidou (kiriakidou@hua.gr)
Dr Ioannis E. Livieris (livieris@unipi.gr)
Dr Christos Diou (cdiou@hua.gr)