Formatting Instructions For NeurIPS 2022

Anonymous Author(s)
Affiliation
Address
email

Abstract

No abstract.

1 Datasets and Tool Packages for Causal Inference

In this section, we present the commonly used datasets and tool packages for causal inference experiments.

1.1 Datasets

The available dataset for causal inference is summarized in this section. It's hard to find a dataset with perfect ground truth for counterfactual experiment due to the fact that it is impossible to observe counterfactual outcome. Semi-synthetic datasets are often used in literature. IHDP, for example, is obtained by generating its own observed outcome following some certain generating process. The detailed description of these commonly used datasets are in the following.

1.1.1 IHDP

. This dataset is an often used dataset generated from the experiment conducted by Infant Health and Development Program[3] targeting on the premature infants with low birth-weight. In the experiment, infants were subject to the control group and experiment group. Infants in the control groups were care by their families while those in the experiment group received both high quality childcare by specialists[6]. The pre-treatment co-variates are composed by 25 variables measuring the infants' and their mothers' conditions, including birth weight, head size, prenatal care, mothers' age, education, medical situation etc. The outcome is these infants' cognitive test score. To simulate the selection bias, a biased subset of the experiment group should be removed while using.

1.1.2 Twins

. The Twins dataset is generated from twin birth rates in the United States between 1989 and 1991[1]. This dataset was proposed by C. Louizos et al. in [8]. The treatment of the Twins dataset is being the heavier one of the two, and the outcome is the death rate of it in its first year of life. In this dataset, only 11,984 pairs of twins weighting less than 2kg and with the same gender were selected, since the outcome was very rare (3.5% in total). The mortality rate of the selected infants is 18.9% for the lighter and 16.4% for the heavier. Therefore, the effect of the treatment is -2.5%. In this dataset, there are 40 pre-treatment co-variants evaluating the pregnancy of the twins, such as the pregnancy weeks, quality of care during gestation, alcohol consumption and tobacco usage during pregnancy nutrition status of their mother etc. In the Twins dataset, both the lighter(control outcome) and the heavier(treatment outcome) infant are being observed. To simulate the selection biases, users have to deploy their own treatment assignment. For example, in the original paper[8], a bias was created following equation (1)

$$
\begin{aligned}
W_i\mathbf{X}_i &\sim Bern(Sigmod(\mathbf{w}'\mathbf{X}_i) + n) \\
\mathbf{w} &\sim U(-0.1, 0.1)^{40\times 1} \\
n &\sim \mathbf{N}(0, 0.1))
\end{aligned}
\tag{1}
$$

1.1.3 Jobs

. The Jobs dataset is a combination of Lalonde experiment[7] and the PSID comparison group. Lalonde's study inspected the effect of a job training program (the treatment) on the participant's earnings, after years of the completion of the training. The pre-treatment data consists of a number of variables, including the participants' age, race, academic achieves and their earnings before the training. The outcome is their job status. This dataset can be downloaded from the website http://users.nber.org/ rdehejia/data/nswdata2.html

1.1.4 ACIC 2016-2019

. This is a series of datasets published by the Atlantic Causal Inference Conference committee while they're holding the causal inference data analysis challenge. These datasets focused on various causal inference problems and the details are as follows.

ACIC 2016. ACIC 2016 Challenge focused on which approaches to causal inference perform well in particular observational study settings. The committee provided 77 datasets with different degrees of non-linearity, sparsity and overlapping between treatment and outcome. The pre-treatment data is extracted from the Infant Health and Development Program containing 58 variables, and the outcomes, both factual and counter-factual, are generated by simulation. The selection bias is created by only keeping children with white moms. The result of the 2016 competition is summarized in [4], and the dataset can be found in https://drive.google.com/"le/d/0B7pG5PPgj6A3N09ibmFwNWE1djA/view

ACIC 2017. The theme of ACIC 2017 Challenge is exploring the estimation and inference for conditional average treatment effects (CATE) with targeted selection. The meaning of targeted selection is the likelihood of treatment that the individual received is a function of the individual's expected response while untreated. This selection will lead to strong confounding[5]. The pre-treatment data is extracted from the Infant Health and Development Program as well, but with only 8 variables. The outcomes and treatment are synthesized following 32 fixed data generating process with four types of errors. 250 independent replica were produced during each of the generating process, so there are 8,000 dataset in total.

ACIC 2018. The ACIC 2018 Challenge consists of two sub-challenge focusing on censoring and scaling. Censoring means not all the sample in the dataset has observed outcome, which means the dataset used for the censoring sub-challenge contains samples without outcomes, or only part of the outcomes. The goal of the scaling sub-challenge is to understand how run-time and memory requirements of the applied causal methods scale, as datasets get larger, and to understand which methods benefit from additional data and which do not. The datasets can be downloaded from https://www.synapse.org/#!Synapse:syn11294478/wiki/486304.

ACIC 2019. The goal of this chanllenge is to estimate the low and high dimension quasi-real-world dataset. The samples in these datasets have different pre-treatment and outcome dimensions, and the data generation code is available on Google Drive[1].

1.2 Codes and Tool Packages

In this section, we introduce some tool packages for causal inference, including Dowhy[9], CausalML, EconML[2] and CausalToolBox, and they're detailed below.

Dowhy[2]: Dowhy is an end-to-end python library for causal inference published by Microsoft. It provides a principled four-step interface for causal inference. It supports the estimation of causal effect with various identification methods, including front door, back door, instrumental variable and so on. Methods it implemented include Propensity-based Stratification, PSM, IPW and Regression.

[1]https://drive.google.com/file/d/1Qqgmb3R9Vt9KTx6t8i_5IbFenylsPfrK/view
[2]https://github.com/py-why/dowhy

References

[1] Douglas Almond, Kenneth Y Chay, and David S Lee. The costs of low birth weight. *The Quarterly Journal of Economics*, 120(3):1031–1083, 2005.

[2] Philipp Bach, Victor Chernozhukov, Malte S Kurz, and Martin Spindler. Doubleml-an object-oriented implementation of double machine learning in python. *J. Mach. Learn. Res.*, 23:53–1, 2022.

[3] Jeanne Brooks-Gunn, Fong-ruey Liaw, and Pamela Kato Klebanov. Effects of early intervention on cognitive function of low birth weight preterm infants. *The Journal of pediatrics*, 120(3):350–359, 1992.

[4] Vincent Dorie, Jennifer Hill, Uri Shalit, Marc Scott, and Dan Cervone. Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statistical Science*, 34(1):43–68, 2019.

[5] P Richard Hahn, Vincent Dorie, and Jared S Murray. Atlantic causal inference conference (acic) data analysis challenge 2017. *arXiv preprint arXiv:1905.09515*, 2019.

[6] Jennifer L Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011.

[7] Robert J LaLonde. Evaluating the econometric evaluations of training programs with experimental data. *The American economic review*, pages 604–620, 1986.

[8] Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. *Advances in neural information processing systems*, 30, 2017.

[9] Amit Sharma and Emre Kiciman. Dowhy: An end-to-end library for causal inference. *arXiv preprint arXiv:2011.04216*, 2020.

A Appendix

Optionally include extra information (complete proofs, additional experiments and plots) in the appendix. This section will often be part of the supplemental material.