06

Accurate, interpretable statistical causal effect estimation

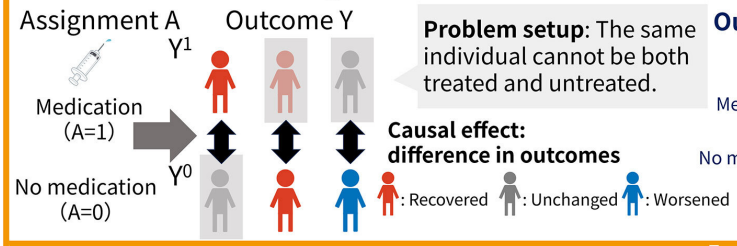
Who is affected by this policy, and why?

Abstract

To design important, individualized interventions such as medical treatments and targeted advertising, we need to understand who benefits from a policy, by how much, and why. We develop two methods for this goal. The first method focuses on accurate effect estimation. It aims to improve the prediction of outcomes with and without treatment by detecting and correcting hidden correlations between treatment assignment and outcomes—for example, when older patients are less likely to receive risky surgery and also tend to have worse outcomes. The second method focuses on explanation. It identifies which personal characteristics account for differences in treatment effects across individuals, while assessing the statistical significance of each characteristic. Compared with simple machine learning techniques that rely only on observed correlations, these methods aim to evaluate the treatment effects more accurately by capturing cause-effect relationships. By building accurate and interpretable causal effect estimation techniques that work even with limited data, our research aims to support data-driven decision-making in high-stakes settings. Ultimately, we hope this research will contribute to a future in which important decisions can be tailored more precisely, reliably, and effectively to each individual.

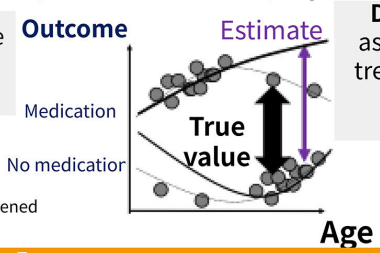
What is a causal effect?

Effect of treatment assignment on each individual



Two tasks for policy design

1) **Predict** the effects 2) **Explain** why they differ across individuals



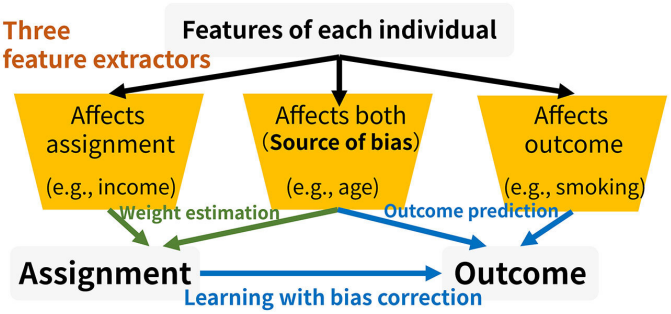
Difficulty 1: Biased treatment assignment—for example, when treatment choice differs by age—**may lead to biased effect estimation**

Difficulty 2: It is unclear **which characteristics cause the bias**

1. Debiased representation learning for effect prediction [1]

Feature representation learning for bias correction in high-dimensional data

1. Learn debiased feature representations from data and **minimize prediction errors weighted by the degree of bias**
2. **Use a differentiable weight estimation technique** to efficiently correct bias and improve effect estimation



Effect prediction results

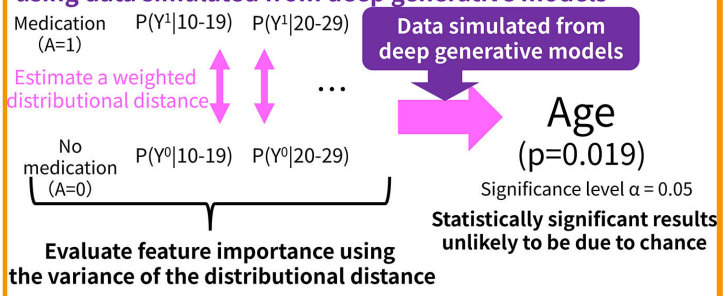
Effect of reading device assignment (mobile vs. PC) on reading time

	Proposed	Existing (Representation learning)	Existing (Two regression models)
Test RMSE	**2.10**	2.38	2.55

2. Feature discovery for treatment effect heterogeneity [2]

More accurate identification of features that influence treatment effects than existing techniques

1. Measure the feature importance by **how much the distance between the outcome distributions** with and without treatment **varies across feature values**
2. **Evaluate statistical significance** (p-values) of each feature **using data simulated from deep generative models**



Feature selection results

U.S. electronic health record data on the treatment of systemic inflammation

Existing	Proposed
Age, Gender	Age, Gender, **Smoking history**

Data-driven discovery of clinically important features reported in previous studies

References

[1] Y. Chikahara, K. Ushiyama, "Differentiable Pareto-Smoothed Weighting for High-Dimensional Heterogeneous Treatment Effect Estimation," *Proc. of The 40th International Conference on Uncertainty in Artificial Intelligence (UAI '24)*, 2024.
[2] Y. Chikahara, M. Yamada, H. Kashima, "Feature Selection for Discovering Distributional Treatment Effect Modifiers," *Proc. of The 38th International Conference on Uncertainty in Artificial Intelligence (UAI '22)*, 2022.
[3] S. Horii, Y. Chikahara, "Uncertainty Quantification in Heterogeneous Treatment Effect Estimation with Gaussian-Process-Based Partially Linear Model," *Proc. of The 38th AAAI Conference on Artificial Intelligence (AAAI '24)*, 2024.
[4] T. Iwata, Y. Chikahara, "Meta-learning for heterogeneous treatment effect estimation with closed-form solvers," *Machine Learning*, Vol. 113, pp. 6093-6114, 2024.

Contact

Yoichi Chikahara, Learning and Intelligent Systems Research Group, Innovative Communication Laboratory