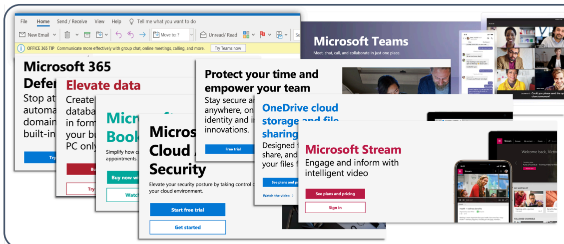
Split-Treatment Analysis to Rank Heterogenous Causal Effects for Prospective Interventions

Yanbo Xu (Georgia Tech), Divyat Mahajan (Microsoft Research India)
Liz Manrao (Microsoft), Amit Sharma (Microsoft Research India), Emre Kıcıman (Microsoft Research AI)

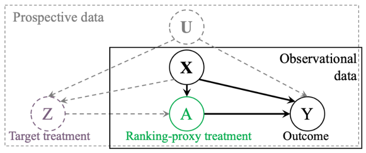
Motivation



- ✗ So many features to recommend!
- ✗ Not all such messages are useful for every individual!
- ✗ Unaffordable or detrimental to run active experiments on all of them!

✓ **Split-Treatment!**
Use logged behavioral data to identify who are likely to benefit from a novel intervention.

Identification of **Split-Treatment**



- Z: Prospective treatment
- A: Proxy treatment
- Y: Outcome
- X: Observed Confounder
- U: No-unmeasured unobserved Confounder

➢ **Assumption1 (Ignorability):** $P(Y|do(a), x) = P(Y|a, x)$ ➢ **Assumption2 (Compliance):** $\mathbb{E}_{x\in\mathcal{G}}[\text{Compliance}(x)] > 0$

$$\text{ITE}^{(z)}(x) = \mathbb{E}[Y|do(z=1), x] - \mathbb{E}[Y|do(z=0), x] = \underbrace{\left(P(a=1|z=1,x) - P(a=1|z=0,x)\right)}_{\text{Compliance}(x)} \cdot \underbrace{\left(\mathbb{E}[Y|a=1,x] - \mathbb{E}[Y|a=0,x]\right)}_{\text{ITE}^{(a)}(x)}.$$

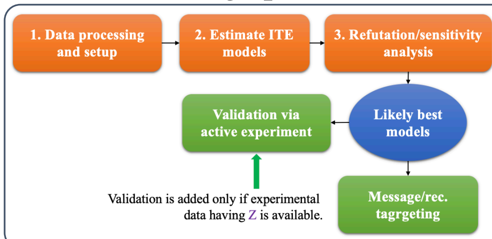
$$\text{CATE}_{\mathcal{G}}^{(z)} = \mathbb{E}_{x\in\mathcal{G}}[\text{ITE}^{(z)}(x)] \propto \mathbb{E}_{x\in\mathcal{G}}[\text{ITE}^{(a)}(x)]$$

We pick a proxy treatment A such that:

➡ A exists, with some natural variation, in our observational logs.

➡ The effect of Z on Y should be mediated through A.

Estimation using **Split-Treatment**



2. Inverse probability of treatment weighting (IPTW) to adjust ITE estimation bias.

$$\sum_{i=1}^{n} w_i \mathcal{L}(y_i, f(x_i, a_i)),$$

$$w_i = a_i \cdot \frac{\hat{P}(a_i=1)}{\hat{P}(a_i=1|x_i)} + (1-a_i) \cdot \frac{1-\hat{P}(a_i=1)}{1-\hat{P}(a_i=1|x_i)}.$$

⬇

$$\hat{\text{ITE}}^{(a)}(x) \equiv f(x, a=1) - f(x, a=0).$$

3. Sensitivity analyses to eliminate unreliable models in the absence of experimental validation.

- *Placebo* test

  Place a random variables as the treatment A

  ➡ Test if an estimator returns zero causal effect.

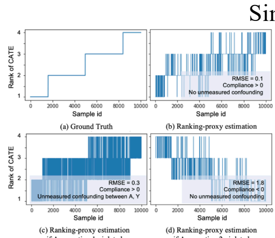
- *Unobserved-confounding* test

  Add a new confounder to the feature set with varying degrees of its effect on A and Y.

  ➡ Test if an estimator is less sensitive to the varying degrees of effect of the new confounder.

Experiments and Results

Simulation



Violation of the two Assumptions: Comparison between the ground-truth rank and the proxy-estimated rank in simulations with or without violation of the assumptions.

IPTW-LR is less sensitive to unobserved confounding. Box plots are for 5 runs with different degrees of confounding.

Unobserved-confounding analysis: Comparison between estimated causal effect with and without unobserved confounding, for two causal models.

Conclusion

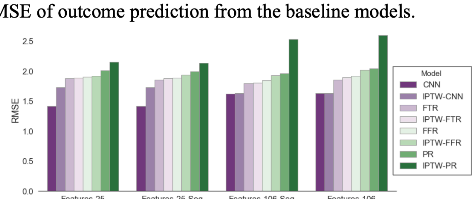
➢ We presented a practical, observational analysis pipeline for
  - Identifying individuals likely to benefit from a novel treatment Z
  - Using proper causal analysis of existing logs that contain proxy treatment A

➢ A key contribution:
  - Refutation tests and sensitivity analyses enable a principled a priori identification of the feature selection and elimination of unreliable algorithmic design.

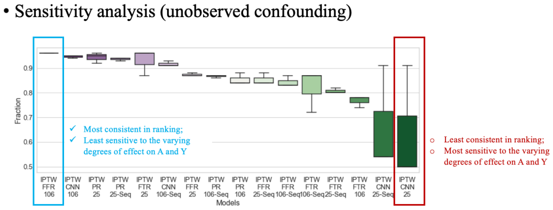
➢ We validated our analysis with an A/B experiment in a large real-world setting.

Real-world data

- RMSE of outcome prediction from the baseline models.



- Sensitivity analysis (unobserved confounding)



Fraction of the top 50-percentile individuals that remain in the top 50-percentile after adding an observed confounder. Box plots are for 3 runs with different degrees of confounding.

- Validation on experimental data

Best model (IPTW-FFR) picked by the sensitivity analysis

Worst model (IPTW-CNN) picked by the sensitivity analysis