Conformal Counterfactual Inference under Hidden Confounding

Zonghao Chen [1,*] Ruocheng Guo [2,*] Jean-François Ton[2] Yang Liu[2]

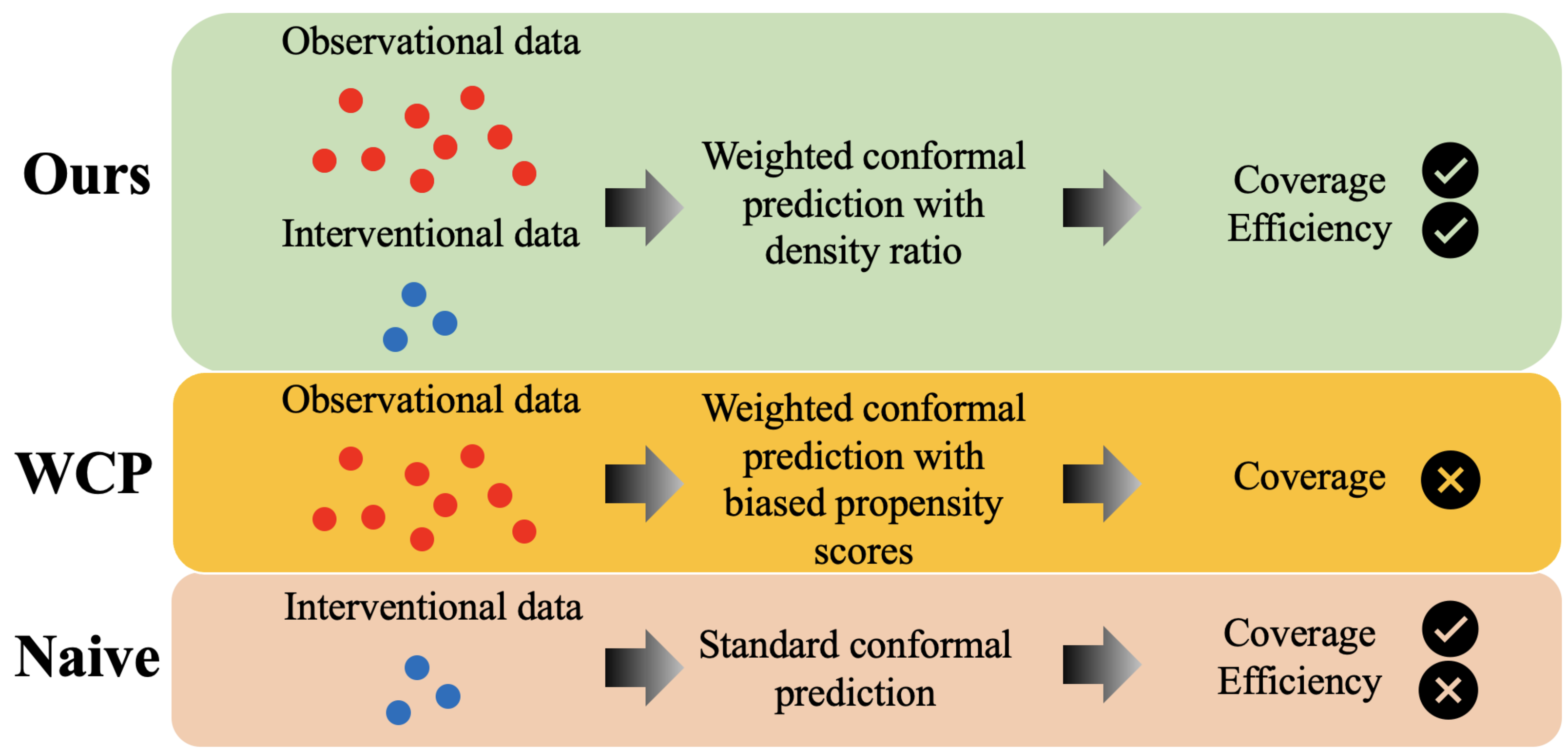
[1] University College London [2] ByteDance Research

TL; DR

- We propose a new algorithm, based on conformal prediction, to construct confidence intervals for heterogenous causal effects with marginal coverage guarantees, even under hidden confounding.
- Coverage: For a pre-specified target coverage α, we have

$$1-\alpha \leq \mathbb{P}(y \in C(x)) \leq 1-\alpha+\frac{1}{n}$$

- Efficiency: Width of $C(x)$.

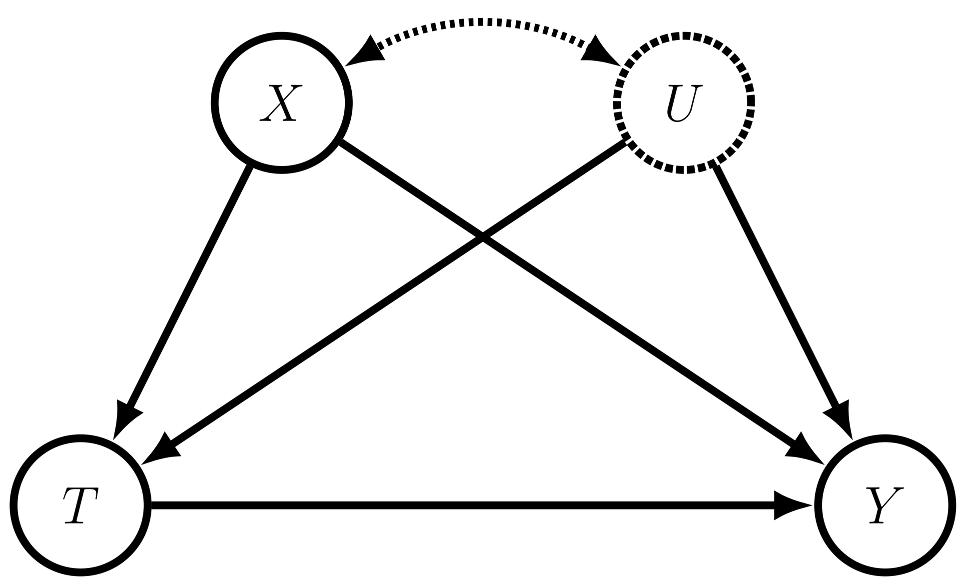


Our Contributions

- Our method constructs confidence intervals for heterogenous causal effects through weighted conformal prediction (WCP).
- We overcome hidden confounding through density-ratio adjustments given access to a small fraction of interventional data.
- We verify our method across synthetic and real-world data, including recommendation systems, in terms of both coverage and efficiency.

A Motivating Example

- What is the effect of the treatment T (pills or surgery) on the outcome Y (recovery rate), given both observed confounding X (severity of disease) and unobserved confounding U (types of disease)?
- For an individual i, what is the estimated heterogenous treatment effect? What is the confidence interval of the estimate?



Weighted Transductive Conformal Prediction (wTCP)

- A merged dataset $\mathcal{D} = \{(x_i, y_i)_{i=1}^n \sim P_{X,Y}\} \cup \{(x_i, y_i)_{i=n+1}^{n+m} \sim P'_{X,Y}\}$.
- Define weight functions $w(x,y) = 1$ if $(x,y) \sim P_{X,Y}$ and $w(x,y) = \frac{dP_{X,Y}}{dP'_{X,Y}}(x,y)$ if $(x,y) \sim P'_{X,Y}$.
- $(p_i)_{i=1}^{|\mathcal{D}|}$ is the "normalized" weight functions.
- For a test data $x \in P'_X$ and $\bar{y} \in \mathcal{Y}$, augment $\bar{\mathcal{D}} = \mathcal{D} \bigcup \{x, \bar{y}\}$. Fit a regressor $\bar{\mu}$ on $\bar{\mathcal{D}}$, compute the conformity scores $s(x_i, y_i) = |\bar{\mu}(x_i) - y_i|$.
- The weighted empirical distribution of conformity scores.

$$\widehat{F} = \sum_{i=1}^{|\mathcal{D}_c|} p_i \delta_{s(x_i, y_i)}$$

- The conformal interval is $C(x) = \{\bar{y} \in \mathcal{Y} : s_x^{\bar{y}} \leq q_{\widehat{F}}\}$, where $q_{\widehat{F}} = \mathrm{Quantile}(1-\alpha; \widehat{F})$.

Our Methodology: wTCP-DR

- Our proposed method *weighted Transductive Conformal Prediction with Density Ratio estimation* (wTCP-DR).
- n observational and m interventional samples and the test sample is x_{n+m+1}.

$$\begin{aligned} (x_i^O, y_i^O)_{i=1}^n &\sim p^O(x,y) = p^O(y \mid x,t) p(x \mid t) \\ (x_i^I, y_i^I)_{i=n+1}^{n+m} &\sim p^I(x,y) = p^I(y \mid x,t) p(x) \end{aligned} \tag{1}$$

Naive Method

- Only uses m interventional data $(x_i^I, y_i^I)_{i=n+1}^{n+m} \sim p^I(x,y)$.
- Has wide confidence interval. (Efficiency NO)
- Has marginal coverage guarantee. (Coverage YES)

WCP (Lei et al, 2021)

- Combines m interventional data $(x_i^I, y_i^I)_{i=n+1}^{n+m} \sim p^I(x,y)$ and n observational data.
- Learns covariate shift adjustment (weights w) through propensity score estimation.
- Ignores hidden confounding.
- Has narrow confidence interval. (Efficiency YES)
- Has marginal coverage guarantee. (Coverage NO)

wTCP-DR (Ours)

- Combines m interventional data $(x_i^I, y_i^I)_{i=n+1}^{n+m} \sim p^I(x,y)$ and n observational data.
- Learns covariate shift adjustment (weights w) through density ratio estimation.
- A data-driven approach to adjust for hidden confounding.
- Has narrow confidence interval. (Efficiency YES)
- Has marginal coverage guarantee. (Coverage YES)

- Transductive conformal prediction (TCP) is computationally more expensive than split conformal prediction (SCP).
- We propose two cheaper variants of wTCP-DR: wSCP-DR(inexact) and wSCP-DR(exact).

Experimental Results

Synthetic Dataset

Method	Coverage $Y(0)$ ↑	Interval $Y(0)$ ↓	Coverage $Y(1)$ ↑	Interval $Y(1)$ ↓	Coverage ITE ↑	Interval ITE ↓
wSCP-DR(Inexact)	0.891 ± 0.026	$\underline{0.414} \pm 0.008$	0.889 ± 0.019	**0.421** ± 0.013	0.942 ± 0.017	**0.835** ± 0.016
wSCP-DR(Exact)	0.934 ± 0.026	0.496 ± 0.010	0.935 ± 0.023	$\underline{0.503} \pm 0.010$	0.957 ± 0.018	0.998 ± 0.015
wTCP-DR	0.899 ± 0.028	**0.386** ± 0.013	0.923 ± 0.015	0.576 ± 0.066	0.953 ± 0.015	$\underline{0.962} \pm 0.074$
WCP	0.572 ± 0.039	0.222 ± 0.007	0.608 ± 0.042	0.227 ± 0.009	0.710 ± 0.027	0.449 ± 0.012
Naive	0.932 ± 0.018	0.508 ± 0.042	0.930 ± 0.023	0.560 ± 0.049	0.952 ± 0.018	1.068 ± 0.098

Recommendation Datasets: Yahoo and Coat

	Yahoo		Coat	
Method	Coverage ↑	Interval ↓	Coverage ↑	Interval ↓
wSCP-DR(Inexact)	0.892 ± 0.019	$\mathbf{4.353} \pm 0.019$	0.919 ± 0.008	**3.787** ± 0.045
wSCP-DR(Exact)	0.952 ± 0.001	5.140 ± 0.001	0.959 ± 0.001	4.565 ± 0.228
wSCP-DR*(Inexact)	0.892 ± 0.020	$\mathbf{4.353} \pm 0.020$	0.919 ± 0.008	$\underline{3.789} \pm 0.046$
wSCP-DR*(Exact)	0.952 ± 0.001	5.140 ± 0.001	0.960 ± 0.001	4.571 ± 0.233
WCP-NB	0.825 ± 0.002	4.036 ± 0.002	0.912 ± 0.005	3.635 ± 0.040
Naive	0.899 ± 0.001	6.047 ± 0.001	0.896 ± 0.003	7.725 ± 0.018

Full paper: `https://arxiv.org/abs/2405.12387`
Code: `https://github.com/rguo12/KDD24-Conformal`

@Hudson19990518