Proceedings of Machine Learning Research 189, 2022 ACML 2022

Robust Direct Learning for Causal Data Fusion (Supplementary Material)

Xinyu Li[1,2] XINYU.LI@PKU.EDU.CN
Yilin Li[1] YILINLI@PKU.EDU.CN
Qing Cui[2] CUIQING.CQ@ANTGROUP.COM
Longfei Li[2] LONGYAO.LLF@ANTGROUP.COM
Jun Zhou[2] JUN.ZHOUJUN@ANTFIN.COM
[1] *School of Mathematical Sciences, Peking University, Beijing, China*
[2] *Ant Group, Hangzhou, China*

Editors: Emtiyaz Khan and Mehmet Gönen

In this Supplementary Material, we present technical proofs of Theorem 4 in Section S.1, and further details of simulation studies in Section S.2.

S.1. Proof of Theorem 4

Proof Under Assumptions 1, 2 and 3, we have

$$
\begin{aligned}
&E\left[\frac{w_S(X)}{\widetilde{p}_{A|S}(X,Z_S)}\left\{\frac{Y-\widetilde{m}_S(X,Z_S)}{A}-l(X)\right\}^2\right]\\
=&E\left[\frac{w_S(X)}{\widetilde{p}_{A|S}(X,Z_S)}\left\{Y-\widetilde{m}_S(X,Z_S)-Al(X)\right\}^2\right]\\
=&\sum_{s\in\mathcal{S}}E\left[\frac{w_s(X)}{\widetilde{p}_{A|s}(X,Z_s)}\left\{Y-\widetilde{m}_s(X,Z_s)-Al(X)\right\}^2\mid S=s\right]P(S=s)\\
=&\sum_{s\in\mathcal{S}}E\left[w_s(X)\sum_{a\in\{-1,1\}}\frac{\mathbb{1}(A=a)}{\widetilde{p}_{a|s}(X,Z_s)}\left\{Y(a)-\widetilde{m}_s(X,Z_s)-al(X)\right\}^2\mid S=s\right]P(S=s)\\
=&\sum_{s\in\mathcal{S}}E\left[w_s(X)\sum_{a\in\{-1,1\}}\frac{p_{a|s}(X,Z_s)}{\widetilde{p}_{a|s}(X,Z_s)}\left\{m_s+a\delta(X,Z_s)+\varepsilon_s-\widetilde{m}_s-al(X)\right\}^2\mid S=s\right]P(S=s),
\end{aligned}
$$

where the first equality follows from $A\in\{-1,1\}$, the second from the law of total probability, the third from $Y=\sum_a \mathbb{1}(A=a)Y(a)$ and the fourth from $Y=m_S(X,Z_S)+A\delta_S(X,Z_S)+\varepsilon_S$.

Thus it suffices to show that for each data source $S=s$, if either of the two conditions in Theorem 4 holds, then

$$\delta(X) \in \arg\min_l E\left[w_s(X)\sum_{a\in\{-1,1\}}\frac{p_{a|s}(X,Z_s)}{\widetilde{p}_{a|s}(X,Z_s)}\left\{m_s + a\delta(X,Z_s) + \varepsilon_s - \widetilde{m}_s - al(X)\right\}^2 \mid S=s\right].$$

In the following, we consider the cases in which each of the two conditions holds.

(i) If $\widetilde{p}_{a|s}(X,Z_s) = p_{a|s}(X,Z_s)$, we have

$$\begin{aligned}
&\arg\min_l E\left[w_s(X)\sum_{a\in\{-1,1\}}\frac{p_{a|s}(X,Z_s)}{\widetilde{p}_{a|s}(X,Z_s)}\left\{m_s + a\delta(X,Z_s) + \varepsilon_s - \widetilde{m}_s - al(X)\right\}^2 \mid S=s\right]\\
=&\arg\min_l E\left[w_s(X)\sum_{a\in\{-1,1\}}\left\{m_s - \widetilde{m}_s + \varepsilon_s + a\delta(X,Z_s) - al(X)\right\}^2 \mid S=s\right]\\
=&\arg\min_l E\left[w_s(X)\left\{2l^2(X) - 4l(X)\delta(X,Z_s)\right\} \mid S=s\right].
\end{aligned}$$

Then Equation (3) in Theorem 4 follows because

$$\delta(X) = E\{\delta(X,Z_s)\mid X, S=s\} \in \arg\min_l E\left[l^2(X) - 2l(X)\delta(X,Z_s) \mid X, S=s\right].$$

(ii) If $\widetilde{m}_s(X,Z_s) = m_s(X,Z_s)$, we have

$$\begin{aligned}
&\arg\min_l E\left[w_s(X)\sum_{a\in\{-1,1\}}\frac{p_{a|s}(X,Z_s)}{\widetilde{p}_{a|s}(X,Z_s)}\left\{m_s + a\delta(X,Z_s) + \varepsilon_s - \widetilde{m}_s - al(X)\right\}^2 \mid S=s\right]\\
=&\arg\min_l E\left[w_s(X)\sum_{a\in\{-1,1\}}\frac{p_{a|s}(X,Z_s)}{\widetilde{p}_{a|s}(X,Z_s)}\left\{a\delta(X,Z_s) + \varepsilon_s - al(X)\right\}^2 \mid S=s\right],\\
=&\arg\min_l E\left[w_s(X)\left\{2\left(\frac{p_{1|s}}{\widetilde{p}_{1|s}} + \frac{p_{-1|s}}{\widetilde{p}_{-1|s}}\right)l^2(X) - 4\left(\frac{p_{1|s}}{\widetilde{p}_{1|s}} + \frac{p_{-1|s}}{\widetilde{p}_{-1|s}}\right)l(X)\delta(X,Z_s)\right\} \mid S=s\right].
\end{aligned}$$

Since

$$h(X) \in \arg\min_l E\left[\left\{\left(\frac{p_{1|s}}{\widetilde{p}_{1|s}} + \frac{p_{-1|s}}{\widetilde{p}_{-1|s}}\right)l^2(X) - 2\left(\frac{p_{1|s}}{\widetilde{p}_{1|s}} + \frac{p_{-1|s}}{\widetilde{p}_{-1|s}}\right)l(X)\delta(X,Z_s)\right\} \mid X, S=s\right],$$

where

$$\begin{aligned}
h(X) &= E\left\{\left(\frac{p_{1|s}}{\widetilde{p}_{1|s}} + \frac{p_{-1|s}}{\widetilde{p}_{-1|s}}\right)\delta(X,Z_s) \mid X, S=s\right\} \Big/ E\left\{\left(\frac{p_{1|s}}{\widetilde{p}_{1|s}} + \frac{p_{-1|s}}{\widetilde{p}_{-1|s}}\right) \mid X, S=s\right\}\\
&= E\{e_A(X,Z_s)\delta(X,Z_s)\mid X, S=s\},
\end{aligned}$$

and the fact that $\widetilde{p}_{a|s}(X,Z_s)$ satisfies the partial balance property with respect to $\delta(X,Z_s)$, Equation (3) in Theorem 4 follows and we complete the proof.

■

S.2. Further Details of Experiments

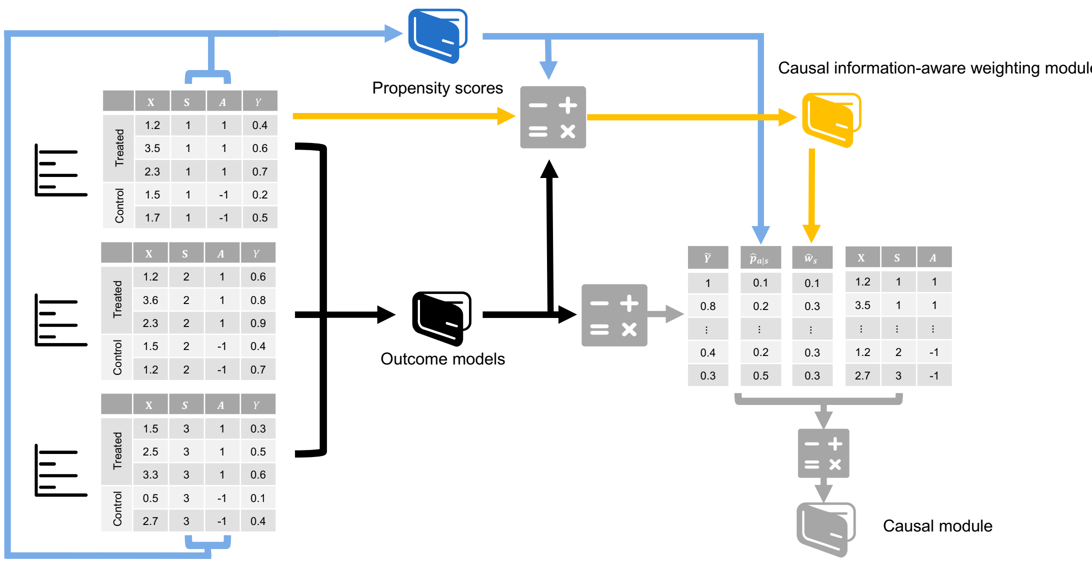

Figure 1: Illustration diagram of our proposed two-step algorithm WMDL, where the nuisance functions include the main effect functions, treatment propensity scores and information-aware weighting functions. The estimate of CATE is obtained by regressing the pseudo-outcomes on the covariates using weighted least squares. See Algorithm 1 in the main text.

The main effect functions m_s are specified as follows.

- Scenario I:

$$\begin{aligned}
m_1(X) &= 1 + X_1 \mathbb{1}(X_2 > 0.5) + 5X_3X_4, \\
m_2(X) &= -X_4 \mathbb{1}(X_4 > 0.5) + 5X_1X_2 - X_3, \\
m_3(X) &= \mathbb{1}(X_3 > 0.5)\mathbb{1}(X_2 > 0.5) - 2\mathbb{1}(X_1 < 0.5), \\
m_4(X) &= 2X_1^2 + 2X_2^2 + 2X_3^2 + 2X_4^2, \\
m_5(X) &= 3X_1X_2 + 3X_3X_4 + 3X_2X_3,
\end{aligned}$$

  and $m_{j+5}(X) = -m_j(X)$ for $j = 1, 2, \cdots, 5$.

- Scenario II: For $s = 1, \ldots, K$,

$$m_s(X, Z_s) = m_s(X) + 3Z_s \mathbb{1}(\mathcal{S}_1) + 2Z_s \mathbb{1}(s \le 5)\mathbb{1}(X_4 < 0.5) + Z_s \mathbb{1}(s > 5)\mathbb{1}(X_4 > 0.5).$$

To further support our method, we conduct more simulated experiments and present numerical results online, see the website https://github.com/xinyuli-stat/CausalDataFusion.