SEARCHING OPTIMAL ADJUSTMENT FEATURES FOR TREATMENT EFFECT ESTIMATION

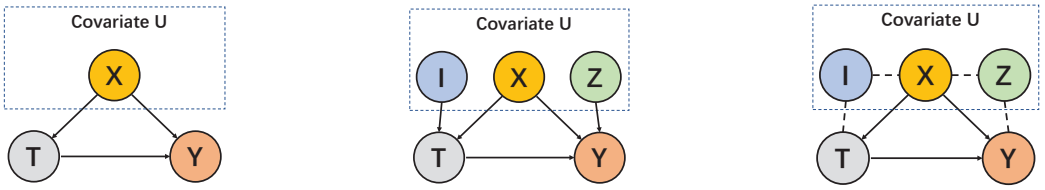
Anonymous authors
Paper under double-blind review

ABSTRACT

In causal inference, it is common to adjust for confounding variables in the treatment effect estimation. Due to the absence of prior knowledge, a common but brute-force approach for practitioners is to include every observed covariate for adjustment. Nevertheless, aside from the confounders, the collected covariates in practical applications often contain extra variables partially correlating to the treatment (treatment-only variables, e.g., instrumental variables) or the outcome (outcome-only variables, e.g., precision variables). Meanwhile, previous literature shows that adjusting treatment-only covariates hurts the treatment effect estimation, while adjusting outcome-only covariates partially correlating to the outcome brings benefits. Consequently, it is meaningful to find an optimal adjustment set rather than the brute-force approach for more efficient treatment effect estimation. To this end, we establish a metric named OAF, which is computationally tractable, to measure the optimality of the adjustment set. From the non-parametric viewpoint, we theoretically show that our metric can be seen as a functional of the adjustment set, which is minimized if and only if the adjustment features contain the confounders and the outcome-only covariates. As optimizing the OAF metric is a combinational optimization problem, we incorporate the Reinforcement Learning (RL) to search for the corresponding optimal adjustment set. More specifically, we adopt the encoder-decoder model as the actor to generate the binary feature mask on the original covariates, which serves as the differentiable policy. Meanwhile, the proposed OAF metric serves as the reward to guide the policy update. Empirical results on both synthetic and real-world datasets demonstrate that (a) our method successfully searches the optimal adjustment features and (b) the searched adjustment features achieve more precise estimation of the treatment effect.

1 INTRODUCTION

Causal inference Imbens & Rubin (2015); Pearl et al. (2000), which refers to infer the variation of potential outcomes by intervening treatments, is a fundamental research area in decision-making Zhang et al. (2021); Zou et al. (2022); Fernández-Loría & Provost (2022) and interpretable artificial intelligence Zhuang et al. (2020); Karimi et al. (2020). In this paper, we perform analysis under the potential outcome framework Imbens & Rubin (2015), and aim to estimate the average effect of intervening the (binary) treatment $\mathbf{T}$ on the outcome $\mathbf{Y}$ given a set of covariates, as shown in Figure 1(c). For example, a researcher attempt to assess the average treatment effect (ATE) of a drug ($\mathbf{T}$) on patients' recovery ($\mathbf{Y}$) from population data given some patients' characteristics. One fundamental problem of causal inference is the non-random treatment assignment between the control and treated groups, where the treatment is assigned with some explicit/implicit assignment policy manifested as correlations with some predictive covariates called confounders ($\mathbf{X}$ in Figure 1(c)) Wu & Fukumizu (2021); Zou et al. (2022). As a consequence, vanilla learning methods will introduce systematic bias without considering diverse treatment assignments across different groups Imbens & Rubin (2015). To overcome this issue, the randomized control trial (RCT) provides the golden standard Booth & Tannock (2014), while the ethical problems or the expensive practical cost become the obstacle to performing RCT in realistic cases. Fortunately, observational studies provide the possible alternative to infer the treatment effect from the Imbens & Rubin (2015); Zou et al. (2022).



(a) Brute-force. (b) Separation. (c) Ours.

Figure 1: The distinction among the settings, where arrows and dashed lines refer to the the causal relationship and the correlation, respectively. (a) Setting of brute-force adjustment, where each covariate is considered as the confounder for adjustment. (b) Setting of the previous separation approach, which only allows $\mathbf{I}$ and $\mathbf{Z}$ to be pre-treatment and pre-outcome variables. (c) Setting of our approach, which $\mathbf{I}$ and $\mathbf{Z}$ to be pre-treatment/post-treatment and pre-outcome/post-outcome or both. Meanwhile, the correlations between $\mathbf{I}, \mathbf{Z}$ and $\mathbf{X}$ are allowed as well.

Although remarkable progress has been achieved for average treatment effect estimation, an important but easily overlooked problem often confuses the practitioner in realistic applications: the collected covariates usually contain extra variables aside from the confounders $\mathbf{X}$, which profoundly affects the treatment effect estimationKuang et al. (2020); Hassanpour & Greiner (2019) (as shown Figure 1(b) and 1(c)). Recalling the drug-recovery example, the drug analyzer often collects the covariates $\mathbf{U}$ to be as abundant as enough such that all the confounders $\mathbf{X}$ (e.g., gender or age) are observed ($\mathbf{X} \subseteq \mathbf{U}$). At the same time, extra variables are introduced aside from $\mathbf{X}$, where can be divided ($\mathbf{U} \backslash \mathbf{X}$) into two types: (a) the treatment-only variables $\mathbf{I}$, denoting the extra variables partially correlating to the treatment $\mathbf{T}$ (e.g., income); (b) the outcome-only variables $\mathbf{Z}$, denoting the extra variables partially correlating to the outcome $\mathbf{Y}$ (e.g., living environment). Notably, we consider a more general setting as shown in Figure 1(c), where we allow (a) $\mathbf{I}$ and $\mathbf{Z}$ to be pre-treatment/post-treatment and pre-outcome/post-outcome variables or both, (b) the existence of correlations between $\{\mathbf{Z},\mathbf{I}\}$ and $\mathbf{X}$. According to previous literature in parametric/non-parametric settings Rotnitzky & Smucler (2020); Hahn (1998); Cochran (1968), adjusting $\mathbf{I}$ will decrease the precision, while adjusting $\mathbf{Z}$ will benefit the estimation. Therefore, even though the estimation is unbiased ($\mathbf{X}$ belongs to the adjustment set), the choice of different adjustment features selected from the covariates still plays a vital role in determining the performance of ATE estimation.

However, due to the lack of prior guidance, the most common approach for the practitioner is to include each covariate into the adjustment feature set Shalit et al. (2017); Shi et al. (2019), which we call the brute-force approach. Due to the (potential) large asymptotic variance, such a brute-force approach is inefficient with poor performance in some real-world cases Hassanpour & Greiner (2019). To overcome this issue, previous approaches Kuang et al. (2020); Hassanpour & Greiner (2019) have attempted to separate the confounders from the precision variables (pre-outcome variables, a special case of $\mathbf{Z}$) or instrumental variables (pre-treatment variables, a special case of $\mathbf{I}$). However, two drawbacks prevent these strategies to be applied in realistic scenes. To be first, their problem settings only consider the instrumental variables and the precision variables (as shown in Figure 1(b)), which is a narrow branch of our setting. By contrast, we allow a more general setting in this paper (as shown in Figure 1(c)). Second, such heuristic approaches cannot clarify what adjustment features are expected by their methods and how the selected adjustment features affect the estimation, while our approach is well supported by semi-parametric inference theory.

In this paper, we focus on separating the treatment-only variables $\mathbf{I}$ from the confounders $\mathbf{X}$ and outcome-only variables $\mathbf{Z}$ for more efficient ATE estimation. Motivated by the related advances of semi-parametric inference Van der Laan et al. (2011); Karimi et al. (2020), we establish a computational tractable metric, named Optimal Adjustment Features (OAF), to empirically describe the asymptotic variance of the ATE estimation. Meanwhile, in the non-parametric regime, we theoretically show that the asymptotic variance decreases within the supplementation of $\mathbf{Z}$ or the deletion of $\mathbf{I}$ into the adjustment set. Therefore, the proposed OAF metric can be considered as a function of the adjustment features, and the minimization of the variance metric implies that the optimal adjustment features ($\{\mathbf{Z}, \mathbf{X}\}$) is selected. As our OAF varies discretely within the change of adjustment features, we treat the minimization of OAF as a combinational optimization problem.

Regarding the optimization efficiency, we introduce the reinforcement learning (RL) and propose a policy gradient based optimization framework named OAFP. More specifically, we construct the actor with an encoder-decoder model Bello et al. (2016) to generate the binary feature mask on the original covariates, where the feature mask serves as the differentiable policy. On the other hand, the OAF metric plays the role of the reward function to guide the policy gradient (e.g., the update of the feature mask). In summary, our contributions are highlighted as follows:

i We propose a computational tractable metric, named OAF, to measure the optimality of the adjustment features for treatment effect estimation with non-parametric theoretical guarantee;

ii We design a reinforcement learning framework,named OAFP, to optimize the proposed metric and generate the corresponding feature mask for selecting adjustment features;

iii Extensive results on both synthetic and real-world datasets verify that: (a) our method can efficiently search the optimal adjustment features, (b) the searched adjustment features significantly improves the precision of treatment effect estimation.

2 Related Work

2.1 Confounder Balancing

To estimate ATE/CATE, statistical methods focus on balancing the confounder across different groups via diverse strategies, including reweighting Kuang et al. (2020), matching Stuart (2010) or covariate alignment Athey et al. (2018). To overcome the model misspecification for the high-dimensional data, a bunch of machine learning methods is further combined to capture the non-linear relationships among variables Van der Laan et al. (2011); Zou et al. (2022); Lim (2018); Qian et al. (2021); Yao et al. (2018); Shalit et al. (2017); Wager & Athey (2018). In detail, the representative non-parametric approach is to discretely fit the potential outcome using a regression tree or random forest (e.g., CF tree or CF forest) Wager & Athey (2018). The typical semi-parametric approaches includes TMLE Van der Laan et al. (2011), doubly-robust methods Karimi et al. (2020) and DragonNet Shi et al. (2019), which is asymptotically unbiased and efficient. The mainstream of deep methods models the confounder balancing as the domain adaptation problem, which learns the group invariant representation by minimizing the distribution divergence across different treatment arms Shalit et al. (2017); Yao et al. (2018). Besides, some methods also use sample-wise reweighting to make treatment and confounder independent in the representation space Qian et al. (2021).

2.2 Covariate Separation

Recent methods have already noticed the problem of separating confounder from the instrumental/precision variables Kuang et al. (2020); Hassanpour & Greiner (2019). For instance, Kuang et al. (2020) proposed a data-driven variance reduction approach Kuang et al. (2020) named DVD to separate the confounders from the precision variables, while DVD does not consider the treatment-only variables. To overcome this gap, Hassanpour & Greiner (2019) introduces the instrumental variables with non-linear deep networks to achieve disentanglement in the representation space. However, our paper contrasts the above-mentioned methods from three aspects: (a) they only consider the instrumental variables and precision variables, while we allow a much broader setting for $\mathbf{I}$ and $\mathbf{Z}$ in Figure 1(c), (b) they are lack of theoretical understanding on how their methods achieve variable separation for better ATE estimation, while our method is well supported by the semi-parametric inference theory; (c) Hassanpour & Greiner (2019) achieves disentanglement in the representation space; while our methods directly separate $\mathbf{I}$ from $\{\mathbf{Z},\mathbf{X}\}$ among the original covariates.

3 Establishing the variance metric

3.1 Semi-parametric Inference for ATE estimation

Setup. For concreteness, we consider the estimation of the average effect of a binary treatment. Suppose the data we own is generated independently and identically: $\{Y_i, U_i, T_i\}_{i=1}^n \sim P$, where P,n and $\mathbf{U}$ refer to the underlying joint distribution density, the sample size and the collected covariates,

respectively. Following notations in Imbens & Rubin (2015), we define the potential outcome under the treatment arm $\mathbf{T} = \mathbf{t}$ as $\mathbf{Y}(\mathbf{t})$ (We use upper-case (e.g. $\mathbf{T}$) to denote random variables, and lower-case (e.g. $\mathbf{t}$) for realizations.). Then the average treatment effect (ATE) equals to the expected difference between the treated ($\mathbf{T} = 1$) and the control ($\mathbf{T} = 0$) groups: $\gamma(P) = \mathbb{E}[\mathbf{Y}(\mathbf{T} = 1) - \mathbf{Y}(\mathbf{T} = 0)]$, where we refer ATE as $\gamma(P)$ for the convenience of later analysis. Given the collected covariates $\mathbf{U} = \{\mathbf{Z}, \mathbf{X}, \mathbf{I}\}$, one has to select $\mathbf{V} \subseteq \mathbf{U}$ as the adjustment feature set for ATE estimation. To guarantee the validity of $\mathbf{V}$, three prior assumptions should be satisfied: [a] **Stable Unit Treatment Value**: $Y_i(\mathbf{t})$ for sample i is independent of the treatment assignments on sample $j \neq i$; [b] **Unconfoundedness**: $Y(\mathbf{t}) \perp\!\!\!\perp \mathbf{T} \mid \mathbf{V}$; [c] **Overlap**: For arbitrary $\mathbf{V} \in \mathcal{V}$, $p(\mathbf{t} \mid \mathbf{V})$ for $\mathbf{t} \in \{0, 1\}$, where $\mathcal{V}$ is the domain of $\mathbf{V}$. When the above-mentioned assumptions are mentioned, the selected $\mathbf{V}$ supports the unbiased estimation of ATE via diverse methods. For instance, the outcome regression (stratification) estimate $\gamma(P) = m_{\mathbf{V}}^{\mathbf{T}=\mathbf{1}}(\mathbf{Y}) - m_{\mathbf{V}}^{\mathbf{T}=\mathbf{0}}(\mathbf{Y})$, where $m_{\mathbf{V}}^{\mathbf{T}=\mathbf{t}}(\mathbf{Y}) = \mathbb{E}[\mathbf{Y} \mid \mathbf{T} = \mathbf{t}, \mathbf{V}]$ refers to the conditional outcome. Alternatives include using the propensity score $\pi^{\mathbf{T}}(\mathbf{V}) = P(\mathbf{T} = \mathbf{t} \mid \mathbf{V})$ for inverse-reweighting. The adjustment set $\mathbf{V}$ satisfying the above three principles is valid, and invalid otherwise.

Semi-parametric Inference for ATE Estimation. Beyond estimating the whole underlying distribution P, previous literature in semi-parametric inference Van der Laan et al. (2011) concerns estimating the ATE parameter γ as a functional of the underlying density P. Moreover, we denote the estimated density from $\{Y_i, U_i, T_i\}_{i=1}^n$ as $\hat{P}$ (via diverse machine learning methods) and the empirical distribution of P as P_n. In the case that γ is pathwise differentiable to P (this holds for ATE) and the underlying statistical model is convex, the following convergence result is obtained through Central Limit Theorem (CLT) once one of $\pi^{\mathbf{T}}(\mathbf{V})$ and $m_{\mathbf{V}}^{\mathbf{T}}$ is consistent:

$$\sqrt{n}\left(\gamma\left(\widehat{P}\right) - \gamma(P)\right) \xrightarrow{d} N(0, \text{Var}[D^{\text{eff}}(\mathbf{V})]), \tag{1}$$

where $\xrightarrow{d}$ refers to the convergence in distribution. The function $D^{\text{eff}}(\mathbf{V})$ of $\mathbf{V}$ denotes the efficient influence curve Van der Laan et al. (2011), which has an unique expression Hines et al. (2022):

$$D^{\text{eff}}(\mathbf{V}) = \frac{\mathcal{I}(\mathbf{T} = 1) - \mathcal{I}(\mathbf{T} = 0)}{\pi^{\mathbf{T}}(\mathbf{V})}(\mathbf{Y}(\mathbf{t}) - m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y})) + m_{\mathbf{V}}^{\mathbf{T}=1}(\mathbf{Y}) - m_{\mathbf{V}}^{\mathbf{T}=0}(\mathbf{Y}) - \gamma(P). \tag{2}$$

Optimality of Adjustment Features. The above conclusion reflects two critical intuitions: (a) the Cauchy–Schwarz inequality and Cramer-Rao bound Van der Laan et al. (2011) guarantees that $D^{\text{eff}}(\mathbf{V})$ achieves the efficient estimation (with optimal asymptotic variance as $\text{Var}[D^{\text{eff}}(\mathbf{V})]$) of γ with respect to each $\mathbf{V}$; (b) different $\mathbf{V}$ determines different $\text{Var}[D^{\text{eff}}(\mathbf{V})]$, which further determines the ATE estimation. Notably, previous theoretical researches have already established the connection between the optimality of the adjustment features $\mathbf{V}$ and the minimization of asymptotic variance: $\text{Var}[D^{\text{eff}}(\mathbf{V})]$ is minimized if and only if $\mathbf{V} = \{\mathbf{X}, \mathbf{Z}\}$ Rotnitzky & Smucler (2020); Hahn (1998). However, two drawbacks restrict the practicality of their methods: (a) they require prior causal graphs to guide the choice of adjustment sets; (b) they utilize the original expression of efficient influence curve, namely D^{eff}, to derive the asymptotic variance, which is not computationally tractable.

3.2 Theoretical properties of Our proposed metric

Different from Rotnitzky & Smucler (2020), we adopt the decomposed version (in Chapter 6.2 in Van Der Laan & Rubin (2006)) of the efficient influence curve (2) as $D_d^{\text{eff}}(\mathbf{V})$, which is computational tractable and also satisfies the linear asymptotic results in (1):

$$D_d^{\text{eff}}(\mathbf{V}) = \frac{\mathcal{I}(\mathbf{T} = 1) - \mathcal{I}(\mathbf{T} = 0)}{\pi^{\mathbf{T}}(\mathbf{V})}(\mathbf{Y}(\mathbf{T}) - m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y})), \tag{3}$$

where the validity of such decomposition is supported in the following lemma:

Lemma 3.1 (Validity of $D_d^{\text{eff}}(\mathbf{V})$)**.** *Similar to D^{eff}, $\widehat{\gamma}(P)$ is asymptotically linear with D_d^{eff}, and $\sqrt{n}\left(\gamma\left(\widehat{P}\right) - \gamma(P)\right) \xrightarrow{d} N(0, \text{Var}[D_d^{\text{eff}}(\mathbf{V})])$.*

Moreover, we strengthen the viewpoint that the asymptotic variance is critical for the precision of estimating ATE in the case of finite samples using the following proposition:

Lemma 3.2. *Suppose that the cumulative distribution function F_n of $\gamma\left(\widehat{P}\right) - \gamma\left(P\right)$ is continuous within the sample size n increasing, then for any $\alpha \geq 0$ and n,*

$$P(|\gamma\left(\widehat{P}\right) - \gamma\left(P\right)| \geq \alpha) \leq \delta_n + 1 - F\left(\frac{\sqrt{n}\alpha}{\sqrt{Var[D_d^{eff}(\mathbf{V})]}}\right), \tag{4}$$

where F refers to the cumulative distribution function of the normal distribution $N(0,1)$ and $\delta_n = \sup|F_n - F|$ describes the point-wise convergence of $\{F_n\}$ to F with increasing n. According to the above lemma, we conclude that smaller $\text{Var}[D_d^{\text{eff}}(\mathbf{V})]$ implies the smaller right-side in equation 4, which further results in more precise $\gamma\left(\widehat{P}\right)$. Therefore, choosing different adjustment features $\mathbf{V}$ from the covariate set $\mathbf{U}$ determines different asymptotic variance $\text{Var}[D_d^{\text{eff}}(\mathbf{V})]$, which further affects the precision of ATE estimation. Naturally, we propose our metric named Optimal Adjustment Features (OAF), as a functional of the adjustment features $\mathbf{V} \mapsto R_+$: $\mathcal{R}^{\text{OAF}}(\mathbf{V}) = \text{Var}[D_d^{\text{eff}}(\mathbf{V})]$. Nevertheless, one might be still confused about how $\mathcal{R}^{\text{OAF}}$ varies within $\mathbf{V}$ changing. We provide theoretical insights to answer this problem using the following theorem:

Theorem 3.3 (Connections between $\mathcal{R}^{\text{OAF}}$ and $\mathbf{V}$)**.** *We denote the selected features for adjustment as $\mathbf{V} \subseteq \{\mathbf{X} \cup \mathbf{I} \cup \mathbf{Z}\}$. Meanwhile, we denote the optimal adjustment set as $\mathbf{V}_0 = \{\mathbf{X} \cup \mathbf{Z}\}$. Then the optimality of our reward is stated from the following three sub-theorems:*

(a) If $\mathbf{V}$ is a valid adjustment set, then $\mathcal{R}^{OAF}(\mathbf{V}') \leq \mathcal{R}^{OAF}(\mathbf{V})$ holds for $\mathbf{V}' = \mathbf{V} \cup \mathbf{Z}'$, where $\mathbf{Z}' \subseteq \mathbf{Z}$.

(b) If $\mathbf{V}$ is a valid adjustment set, then $\mathcal{R}^{OAF}(\mathbf{V}) \leq \mathcal{R}^{OAF}(\mathbf{V}')$ holds for any $\mathbf{V}' = \mathbf{V} \cup \mathbf{I}'$, where $\mathbf{I}' \subseteq \mathbf{I}$.

(c) We assume that the $\{\mathbf{X} \cup \mathbf{I} \cup \mathbf{Z}\}$ contains all the parents of $\mathbf{Y}$, which implies that $\mathbf{Z}$ contains all the outcome-precision variables of $\mathbf{Y}$. Then $\mathcal{R}^{OAF}(\mathbf{V_0}) \leq \mathcal{R}^{OAF}(\mathbf{V}')$ holds for any $\mathbf{V}'$ which is not a valid adjustment set.

Remark. Overall, our theorem reflects that $\mathcal{R}^{\text{OAF}}(\mathbf{V}) = \text{Var}[D_d^{\text{eff}}(\mathbf{V})]$ achieves the minimum if $\mathbf{V} = \{\mathbf{X} \cup \mathbf{Z}\}$. Meanwhile, we argue that if $\mathcal{R}^{\text{OAF}}(\mathbf{V}) = \text{Var}[D_d^{\text{eff}}(\mathbf{V})]$ then $\mathbf{V} = \{\mathbf{X} \cup \mathbf{Z}\}$ is the optimal adjustment features. To be specific, $\mathbf{V}$ must equal to $\{\mathbf{X}, \mathbf{Z}\}$ when $\mathcal{R}^{\text{OAF}}(\mathbf{V})$ achieves the minimum in the case that all the inequalities in Theorem 3.3 strictly hold (otherwise $\mathcal{R}^{\text{OAF}}(\{\mathbf{X}, \mathbf{Z}\}) < \mathcal{R}^{\text{OAF}}(\mathbf{V})$ contradicts the assumption). The case that some equalities hold is meaningless, since Lemma 4 implies that any valid adjustment features achieve the minimal asymptotic variance is optimal for ATE estimation. Finally, we claim that the proposed $\mathcal{R}^{\text{OAF}}(\mathbf{V})$ achieves the minimum if and only if $\mathbf{V} = \{\mathbf{X}, \mathbf{Z}\}$ are the optimal adjustment features.

3.3 Empirical estimation for computation

Recalling the empirical data $\{Y_i, U_i, T_i\}_{i=1}^n$, it is necessary to find an unbiased estimation of $\mathcal{R}^{\text{OAF}}(\mathbf{V}) = \text{Var}[D_d^{\text{eff}}(\mathbf{V})]$ in the case of finite samples. Fortunately, the M-estimation theory Stefanski & Boos (2002) provides the empirical sandwich estimator as an unbiased solution. Although the influence curve approach is more general than the M-estimator approach, they are equivalent in the case of ATE estimation Stefanski & Boos (2002). More specifically, supposing that $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ and $\widehat{m_{\mathbf{V}}^{\mathbf{T}=\mathbf{1}}(\mathbf{Y})}$ represents the estimated propensity score and the conditional outcome, respectively, the corresponding empirical M-estimator can be written as $\widehat{\phi}(\gamma) = \frac{\mathcal{I}(\mathbf{T}=1)-\mathcal{I}(\mathbf{T}=0)}{\widehat{\pi^{\mathbf{T}}(\mathbf{V})}}(\mathbf{Y}(\mathbf{t}) - \widehat{\pi^{\mathbf{T}}(\mathbf{V})})$, where the "sandwitch" terms can be further calculated as $\widehat{A}(\gamma) = I$ (I is identity matrix) and $\widehat{B}(\gamma) = \frac{1}{n}\sum_{i=1}^n \widehat{\phi}_{i=1}(\gamma)^2$. Finally, the empirical estimation of our metric, namely $\widehat{\mathcal{R}}^{\text{OAF}}(\mathbf{V})$, is derived as follows [1]:

$$\widehat{\mathcal{R}}^{\text{OAF}}(\mathbf{V}) = \widehat{A}(\gamma)\widehat{B}(\gamma)\widehat{A}(\gamma)^T = \frac{1}{n}\sum_{i=1}^n \left(\frac{\mathcal{I}(\mathbf{T}=1) - \mathcal{I}(\mathbf{T}=0)}{\widehat{\pi^{\mathbf{T}}(\mathbf{V})}}(\mathbf{Y}(\mathbf{T}) - \widehat{\pi^{\mathbf{T}}(\mathbf{V})})\right)^2. \tag{5}$$

[1] In fact, the term is similar to the additional term of doubly-robust methods (e.g., AIPW) or the iteration term in TMLE, as both TMLE and AIPW tunes the estimator or estimated distributions to compensate for the term $P_n D_d^{\text{eff}}(\mathbf{V})$ Hines et al. (2022) such that the error term convergences to a zero-mean Gaussian distribution.

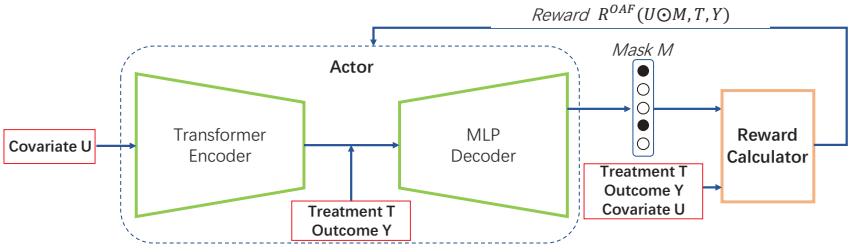


Figure 2: The framework of our reinforcement learning method.

4 Reinforcement Learning for Optimization

As mentioned above, the empirical variance metric $\widehat{\mathcal{R}}^{\text{OAF}}(\mathbf{V},\mathbf{T},\mathbf{Y})$ in equation 5 varies discretely with different adjustment features $\mathbf{V}$, where we rewrite $\widehat{\mathcal{R}}^{\text{OAF}}(\mathbf{V},\mathbf{T},\mathbf{Y})$ here to strengthen the point that the calculation of $\widehat{\mathcal{R}}$ depends on $\mathbf{T},\mathbf{Y}$ as well. Hence, it is difficult to optimize $\widehat{\mathcal{R}}^{\text{OAF}}(\mathbf{V},\mathbf{T},\mathbf{Y})$ in a differentiable approach. As an alternative, we consider the minimization of $\widehat{\mathcal{R}}^{\text{OAF}}$ as a combinational optimization problem. Motivated by the recent advances in neural combinational search area Zhu et al. (2019); Bello et al. (2016), we use the reinforcement learning (RL) to efficiently search $\mathbf{V}$. To this end, we define the binary feature mask $\mathbf{M}$ on the original covariates $\mathbf{U} = \{\mathbf{Z},\mathbf{X},\mathbf{I}\}$ such that the ultimate goal is to find $\mathbf{M}$ corresponding to the optimal adjustment features $\{\mathbf{Z},\mathbf{X}\}$. We suppose the policy for mask generation is $q_\Phi(\cdot \mid \{\mathbf{T},\mathbf{Y},\mathbf{U}\})$, where Φ is the network parameter. Then the expected reward is defined to be our training objective as follows:

$$J(\psi \mid \mathbf{s}) = \mathbb{E}_{\mathbf{M}\sim q_\Phi(\cdot\mid\{\mathbf{T},\mathbf{Y},\mathbf{U}\})} - \mathcal{R}^{\text{OAF}}(\mathbf{U}\odot\mathbf{M},\mathbf{T},\mathbf{Y}), \tag{6}$$

where we use the notation $\odot$ to denote the selection of $\mathbf{V}$ from $\mathbf{U}$ by $\mathbf{M}$. In detail, we adopt the policy gradient method with variance reduction (reinforcement) to optimize the objective in equation 6. Previous work for combinational optimization adopts the parametric approach by building a critic network to estimate the reward and reduce the variance Zhu et al. (2019); Bello et al. (2016). However, the critic can estimate the reward accurately only when the reward design is relatively simple (e.g., the traveling salesman problem Bello et al. (2016)). By contrast, the $\widehat{\mathcal{R}}^{\text{OAF}}(\mathbf{V},\mathbf{T},\mathbf{Y})$ in our problem is more complex, which is calculated upon two estimators $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ and $\widehat{m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y})}$. Therefore, we alternatively use the non-parametric approach as reinforcement Williams (1992) to calculate the gradient of $J(\psi \mid \mathbf{s})$ with respect to ψ. Moreover. we also add an entropy regularization term to encourage the exploration of the actor during the search process Zhu et al. (2019).

Regarding the implementations, we follow previous paradigms Bello et al. (2016) and build the actor network in the encoder-decoder architecture, as shown in Figure 2. The encoder is a multi-block transformer and the decoder is a Multi-layer-perception (MLP) perception. We leave the detailed settings of the actor network in the appendix. To improve the efficiency during the optimization, we sample K arrays $\{B_1,B_2,\ldots,B_K\}$ as a batch, where $B_i = \{\mathbf{t}_i,\mathbf{u}_i,\mathbf{y}_i\}_{i=1}^{n_b}$ with n_b as the sample size for each array. As such operation implies the computation of $\widehat{\mathcal{R}}^{\text{OAF}}(\mathbf{V},\mathbf{T},\mathbf{Y})_i$ for each B_i, calculating the reward becomes more time-consuming than updating the actor network, especially in the case that $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ and $\widehat{m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y})}$ are non-linear estimators. To alleviate this problem, we training $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ and $\widehat{m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y})}$ in parallel with multiples processes.

5 Experiments

5.1 Benchmarks and Baselines

To evaluate the effectiveness of the proposed method, we conduct experiments on three datasets including the synthetic data, the semi-synthetic IHDP dataset Hill (2011) and the real-world Twins dataset Almond et al. (2005), respectively. Details are present in the appendix for saving space.

Synthetic. Our synthetic datasets are generated according to the following process, which takes as input the total sample size N, the feature dimension d of the covariate U. In general, we

first generate the pre-treatment part of $\mathbf{Z}$, the confounders $\mathbf{X}$ with the post-treatment part of $\mathbf{I}$. Then $\mathbf{Y}$ and $\mathbf{T}$ are generated, where the post-treatment part of $\mathbf{I}$ and the post-outcome part of $\mathbf{Z}$ are further generated. To be specific, we first generate $\mathbf{X}$ with feature size d_x, the pre-treatment $\mathbf{I}^e$ with size d_{Ie} and the pre-outcome $\mathbf{Z}^e$ with size d_{Ze}: $\mathbf{X}_1, \cdots, \mathbf{X}_{d_x}, \mathbf{Z}^e_1, \cdots, \mathbf{Z}^e_{d_{ze}}, \mathbf{I}^e_1, \cdots, \mathbf{I}^e_{d_{Ie}} \overset{iid}{\sim} \mathcal{N}(0,1)$. The treatment $\mathbf{T}$ is then sampled from the logistic transformation of $\mathbf{I}^e$ and $\mathbf{X}$ as $\mathbf{T} \sim \text{Bernoulli}\left(1/\left(1+\exp\left(-\left(\mathbf{I}^T\mathbf{X}+\mathbf{I}^T\mathbf{I}^e\right)\cdot r\right)\right)\right)$, where $r = \frac{d_x+d_{Ie}}{20}$ is the scaling factor. Meanwhile, following previous protocols Kuang et al. (2020), the outcome $\mathbf{Y}$ is generated under both the linear and non-linear setting. More specifically, the linear generation of $\mathbf{Y}$ is $Y = \mathbf{X}\beta_{xy} + \mathbf{z}^e\beta^{zy} + \mathbf{T} + \sigma^Y$, where the non-linear generation is $Y = \mathbf{X}\beta_{xy} + \sum_{i=1}^{d_{ze}} \mathbf{z}^e_i \mathbf{z}^e_{i+1} \cdot \beta^{zy}_i + \mathbf{T} + \sigma^Y$ (the term $i+1$ is modulated by d_{ze}). Furthermore, the post-treatment variables $\mathbf{I}^o$ and the post-outcome variables $\mathbf{Z}^o$ are generated as $\mathbf{I}^o = \beta^{Io}\mathbf{T} + \sigma^{Io}$ and $\mathbf{Z}^o = \beta^{Zo}\mathbf{Y} + \sigma^{Zo}$. Overall, the ATE for the synthetic dataset is 1 and the covariate is $\mathbf{U} = \{\mathbf{X}, \mathbf{Z}^e, \mathbf{Z}^o, \mathbf{I}^e, \mathbf{I}^o\}$. To increase the challenging of separating $\mathbf{I}$ from $\{\mathbf{Z}, \mathbf{X}\}$, we set $d_{Io} = 0.3d$, $d_{Ie} = 0.2d$, $d_X = 0.3d$, $d_{Zo} = 0.1d$ and $d_{Ze} = 0.1d$ by enlarging the ratio of $\mathbf{I}$. Besides, the sample size N is set to 2000.

IHDP. Based on the original RCT data, the selection bias is introduced by Hill (2011) via removing a non-random subset of the treated population. The resulting dataset contains 747 instances (608 control, 139 treated) with 25 covariates collected from the real-world Shi et al. (2019). We first choose the 5 continuous covariates as the confounders $\mathbf{X}$. Then we randomly choose half of the discrete covariates as $\mathbf{I}$, where the rest discrete covariates are set as $\mathbf{Z}$. The generation of $\mathbf{Y}$ further follows the surface B setting in Hill (2011).

Twins. The original Twins dataset is derived from the all twins born in the USA between the year of 1989 and 1991 Almond et al. (2005). Following previous protocol Shi et al. (2019), we consider 28 variables related to parents, pregnancy, and birth, where the outcome is the children's mortality after one year. To develop $\mathbf{I}$, we construct 5 pre-treatment variables and 5 post-treatment variables to generate the 38-dimension $\mathbf{U}$. The treatment $\mathbf{T}$ is then generated via the logistic function.

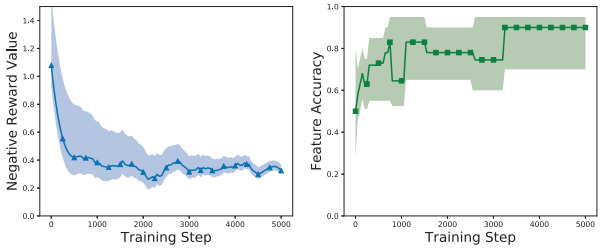
Baselines. The baselines we compared in this paper can be summarized into three classes: (a) Statistical methods, which include the direct difference method Kuang et al. (2020), the inverse propensity score reweighting (IPW) Austin & Stuart (2015), Augmented IPW (AIPW) Van Der Laan & Rubin (2006) and the TMLE method Van Der Laan & Rubin (2006); (b) Machine Learning methods including the DragonNet Shi et al. (2019), Generative adversarial Network (GANITE) Yoon et al. (2018), the Bayesian regression Tree (BART) Hill (2011) and the orthogonal regularized network (DNOUT) Hatt & Feuerriegel (2021); (c) previous covariate disentanglement/separation methods including the LASSO regularized AIPW (AIPW-L), the DVD method in Kuang et al. (2020), the DR-CFR method in Hassanpour & Greiner (2019) and the multi-environment invariant method NICE Shi et al. (2021).

Implementations of our method. Roughly speaking, we implement both the linear and the non-linear versions of our method, respectively. For the linear implementation, we implement the $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ as the logistic regression and $\widehat{m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y})}$ as the linear regression to search the adjustment features. The downstream estimator for ATE estimation is the doubly-robust AIPW. For the non-linear implementation, we build a two-layer MLP as $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ with a four-layer MLP as $\widehat{m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y})}$for searching features, with the DragonNet as the downstream estimator for estimating ATE. To ease the notation, we name our method OAFP_L implemented for the linear case, and OAFP_N implemented for the non-linear case. Our implementation in Python will be released in public once accepted.

Metric. We mainly focus on two metrics: the bias of ATE and the accuracy of feature selection. The former metric is quantified by $\epsilon_{\text{ATE}} = |\text{ATE} - \widehat{\text{ATE}}|$, where $\text{ATE} = \frac{1}{N}\sum_{i=1}^{N} Y_i^1 - \frac{1}{N}\sum_{j=1}^{N} Y_j^0$ is the underlying truth. Notably, as the underlying ATE for Twins is close to zero (0.025), we report the relative error as $\epsilon_{\text{ATE}} = \frac{|\text{ATE}-\widehat{\text{ATE}}|}{\text{ATE}}$ for Twins dataset. For the latter metric, we use $\text{Acc} = \frac{|\widehat{\mathbf{M}}-\mathbf{M}_0|_1}{d}$ to measure the feature accuracy, where $\widehat{\mathbf{M}}$ refers to the optimized feature mask and $\mathbf{M}^0$ refers to the ground truth feature mask with $\mathbf{M}^0_i = 0$ when $\mathbf{U}_i \in \mathbf{I}$ and $\mathbf{M}^0_i = 1$ otherwise.

5.2 Results and Analysis

In this section, we first propose three questions on the evaluation of the proposed OAFP method: (a) Whether OAFP searches the adjustment features accurately; (b) Whether the adjustment features



(a) Reward (b) Accuracy

Figure 3: Reward and feature accuracy curves in the setting of non-linear synthetic data with d = 20.

Dataset	Fs_Acc	R_err
S-20-l	95.0%	0.02
S-40-l	92.5%	0.04
S-80-l	90.0%	0.11
S-20-n	95.0%	0.06
S-40-n	95.0%	0.10
S-80-n	90.0%	0.13
IHDP	92.0%	0.11
Twins	94.7%	0.13

Table 1: Results on feature search, where S-20-l refers to the synthetic data with 20 features in the linear setting (n refers to non-linear).

Table 2: Non-Linear simulation results. The metrics are Mean±STD over 10 repeated experiments. The best performance is marked bold.

Settings		In_sample Prediction			Out_of_sample Prediction		
Feature Dimension		20	40	80	20	40	80
Statistical	Direct	4.69±0.62	7.09±0.68	8.92±0.76	5.23±0.41	6.28±1.41	9.28±1.32
	IPW	0.99±4.50	1.27±3.13	4.36±2.37	1.33±1.92	2.22±5.39	4.51±3.33
	AIPW	1.32±1.95	0.99±0.27	2.35±0.83	0.21±1.19	0.55±0.47	3.88±1.23
	TMLE	0.42±0.11	0.59±0.07	0.62±0.02	0.50±0.12	0.66±0.18	0.81±0.20
Machine	DragonNet	0.19±0.19	0.20±0.14	0.57±0.38	0.99±0.16	0.84±0.70	0.87±1.02
	GANITE	0.80±0.01	0.87±0.01	0.99±0.01	0.99±0.01	1.08±0.01	1.10±0.01
	DNOUT	0.47±0.01	0.62±0.04	0.92±0.09	0.50±0.02	0.61±0.05	0.95±0.09
	BART	0.92±0.20	2.03±0.27	2.89±0.98	0.92±0.20	2.25±0.16	2.98±1.10
Decomposed	AIPW_L	0.59±0.10	0.66±0.05	0.89±0.10	0.54±0.29	0.74±0.13	0.96±0.22
	DVD	0.95±0.03	0.83±0.01	0.76±0.01	1.06±0.08	0.64±0.01	1.05±0.73
	DR-CFR	0.88±0.08	1.18±0.16	2.08±0.69	1.28±0.08	1.69±0.73	1.52±0.51
	NICE	1.08±0.32	1.24±0.60	1.81±0.22	1.10±0.37	1.23±0.41	1.93±0.35
Ours	**OAFP_L**	**0.03±0.13**	**0.12±0.10**	**0.23±0.13**	**0.24±0.22**	**0.20±0.13**	**0.32±0.34**
	OAFP_N	**0.01±0.10**	**0.09±0.07**	**0.13±0.11**	**0.15±0.09**	**0.16±0.07**	**0.14±0.08**

searched by our OAFP achieve better ATE estimation compared to baselines; (c) Whether the search process of OAFP is efficient on the time cost.

Results on searching features. To answer the first question, we report results on Results on feature search (Fs_Acc) with the relative error (R_err) in Table 1, where R_err = $\frac{|R(\widehat{\mathbf{V}})-R(\mathbf{V})|}{R(\mathbf{V})}$ measure the relative distance between the optimal variance metric as $R(\mathbf{V})$ and the metric for our searched features $\widehat{\mathbf{V}}$ as $R(\widehat{\mathbf{V}})$. Notably, we search non-linear synthetic data, IHDP, and twins using the non-linear OAFP-N, while the linear synthetic dataset is searched using OAFP-L. The feature accuracy in Table 1 reflects that our method OAFP_L and OAFP_N successfully search the optimal adjustment features $\{\mathbf{Z}, \mathbf{X}\}$ in linear and non-linear cases, respectively. Meanwhile, the relative error of the variance metric R_err in Table 1 also reflects that our searched adjustment features achieve the empirical asymptotic variance close to the optimal one achieved by $\{\mathbf{Z}, \mathbf{X}\}$. Moreover, we provide an intuitive illustration of how the reward $\widehat{\mathcal{R}}$ and the feature accuracy vary in the training process during the search process in Figure 3(a) and Figure 3(b), respectively. As shown in 3(a), the average reward converges stably under the threshold at around 4000 steps, where the corresponding feature accuracy also achieves 95% after 3000 steps. Besides, the reasons behind that the feature accuracy cannot achieve 100% can be attributed to (a) error between the empirical $\widehat{\mathcal{R}}$in equation 5 and $\mathcal{R}$; (b) the effect of some covariates are too small in the underlying structural equation such that their existence or not is less important.

Results on ATE estimation. We then report the downstream results on ATE estimation in Table 2, Table 3 (results on linear simulation is present in appendix), respectively. For results on non-linear simulation, our method, OAFP_L and OAFP_N, achieve significant improvement in the estimation performance compared to other baselines. Results on the semi-synthetic IHDP and the real-world

Table 3: Results on IHDP and Twins datasets. The metrics are Mean±STD over 10 repeated experiments. The best performance is marked bold.

Benchmark		IHDP		**Twins**	
Settings		In_sample	Out_of_sample	In_sample	Out_of_sample
Statistical	Direct	3.36±3.70	3.70±3.36	1.50±0.03	4.34±0.14
	IPW	3.48±5.92	3.48±5.91	1.79±0.05	9.29±0.39
	AIPW	1.82±2.99	1.82±2.99	1.79±0.05	9.29±0.39
	TMLE	2.71±1.80	2.52±1.07	1.76±0.02	4.01±0.02
Machine	DragonNet	1.19±1.04	1.37±0.95	1.05±0.01	1.03±0.01
	GANITE	5.40±0.04	5.60±0.01	15.60±0.08	19.6±0.19
	DOUT	3.16±1.41	3.08±1.26	2.04±0.02	2.20±0.02
	BART	3.12±2.42	3.28±2.60	0.95±0.03	0.97±0.03
Decomposed	AIPW_L	1.85±2.64	1.85±2.64	1.03±0.03	3.35±0.12
	DVD	2.79±0.82	0.73±0.03	1.42±0.01	7.78±0.05
	DR-CFR	2.45±1.05	1.74±1.00	3.64±0.03	6.00±0.01
	NICE	2.75±3.91	2.68±2.25	42.92±0.02	53.12±1.84
Ours	**OAFP_L**	**1.14±0.47**	**1.24±0.33**	**0.65±0.02**	**1.98±0.06**
	OAFP_N	**0.28±0.07**	**0.29±0.09**	**0.30±0.01**	**0.43±0.01**

Twins further verify the superiority of our method. Meanwhile, the poor performance for methods without covariate separation also strengthens our view that the existence of treatment-only variables $\mathbf{I}$ will hurt the ATE performance in finite-sample cases. Notably, our linear implementation OAFP_L performs less accurately than some deep methods (e.g., DragonNet) due to the model misspecification problem for the IHDP dataset.

Results on the efficiency of our method. To verify how our RL framework improves the searching efficiency, we compare the search process of our OAFP to that of the brute-force approach (traversing the powerset of $\mathbf{U}$ and find the minimal $\mathcal{R}$) in Table 4. Obviously, it is meaningful for us to design the RL framework as it significantly reduces the time cost for searching the optimal adjustment features. (The brute-force approach is even impossible when the feature dimension is larger than 20.) Besides, we also provide ablation studies on the effect of varying sample size and $\mathbf{X}$-$\mathbf{I}$-$\mathbf{Z}$ ratio on searching the features and the ATE estimation in our appendix.

Table 4: Comparison on the time cost (hours) of searching for features, where Syn_20_l refers to synthetic data with 20 covariates generated under the linear setting. For the brute-force approach, we calculate its time cost as the multiplication of the average training time by 2^n.

Method	Syn_20_l	Syn_40_l	Syn_20_n	Syn_40_n
Ours	0.22	0.27	28.055	41.94
Brute-force	11.65	$1.83{\cdot}10^7$	$2.94{\cdot}10^3$	$4.61{\cdot}10^9$

6 FUTURE WORKS AND CONCLUSION

In this paper, we study the problem of estimating average treatment effect (ATE) from observational studies when the collected covariates ($\mathbf{U}$) contain the treatment-only variables ($\mathbf{I}$) and the outcome-only variables ($\mathbf{Z}$) aside from the confounder X. Based on the semi-parametric inference, we show that separating $\mathbf{I}$ from $\{\mathbf{Z}, \mathbf{X}\}$ brings benefits for ATE estimation. To this end, we establish a variance metric to measure the optimality of any adjustment features, and design an RL-based framework to efficiently optimize the proposed metric. Extensive experimental results also verified that the proposed method successfully identifies the optimal features with precise ATE estimation.

However, two problems still remain for further efforts. (a) The existence of selection bias. Aside from the confounding bias, the selection bias Bareinboim & Pearl (2012) hurts the ATE estimation when we adjust the common effects of both the treatment and the outcome. One possible solution is to treat the inverse propensity score of the treatment (IPT) as the target parameter and derive the corresponding efficient influence curve. (b) Estimating the individualized treatment effect (ITE). As a closed form of efficient influence curve on ITE estimation is difficult to derive, it remains an open but interesting problem.

REFERENCES

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.

Susan Athey, Guido W Imbens, and Stefan Wager. Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):597–623, 2018.

Peter C Austin and Elizabeth A Stuart. Moving towards best practice when using inverse probability of treatment weighting (iptw) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in medicine*, 34(28):3661–3679, 2015.

Elias Bareinboim and Judea Pearl. Controlling selection bias in causal inference. In *Artificial Intelligence and Statistics*, pp. 100–108. PMLR, 2012.

Irwan Bello, Hieu Pham, Quoc V Le, Mohammad Norouzi, and Samy Bengio. Neural combinatorial optimization with reinforcement learning. *arXiv preprint arXiv:1611.09940*, 2016.

CM Booth and IF Tannock. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *British journal of cancer*, 110(3):551–555, 2014.

William G Cochran. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, pp. 295–313, 1968.

Carlos Fernández-Loría and Foster Provost. Causal decision making and causal effect estimation are not the same… and why it matters. *INFORMS Journal on Data Science*, 2022.

Jinyong Hahn. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, pp. 315–331, 1998.

Negar Hassanpour and Russell Greiner. Learning disentangled representations for counterfactual regression. In *International Conference on Learning Representations*, 2019.

Tobias Hatt and Stefan Feuerriegel. Estimating average treatment effects via orthogonal regularization. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 680–689, 2021.

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

Oliver Hines, Oliver Dukes, Karla Diaz-Ordaz, and Stijn Vansteelandt. Demystifying statistical learning based on efficient influence functions. *The American Statistician*, pp. 1–13, 2022.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Amir-Hossein Karimi, Julius Von Kügelgen, Bernhard Schölkopf, and Isabel Valera. Algorithmic recourse under imperfect causal knowledge: a probabilistic approach. *Advances in neural information processing systems*, 33:265–277, 2020.

Kun Kuang, Peng Cui, Hao Zou, Bo Li, Jianrong Tao, Fei Wu, and Shiqiang Yang. Data-driven variable decomposition for treatment effect estimation. *IEEE Transactions on Knowledge and Data Engineering*, 2020.

Bryan Lim. Forecasting treatment responses over time using recurrent marginal structural networks. *advances in neural information processing systems*, 31, 2018.

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.

Safoora Masoumi and Saeid Shahraz. Meta-analysis using python: a hands-on tutorial. *BMC medical research methodology*, 22(1):1–8, 2022.

Judea Pearl et al. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19 (2), 2000.

Zhaozhi Qian, Alicia Curth, and Mihaela van der Schaar. Estimating multi-cause treatment effects via single-cause perturbation. *Advances in Neural Information Processing Systems*, 34:23754–23767, 2021.

Andrea Rotnitzky and Ezequiel Smucler. Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *J. Mach. Learn. Res.*, 21(188):1–86, 2020.

Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International Conference on Machine Learning*, pp. 3076–3085. PMLR, 2017.

Claudia Shi, David Blei, and Victor Veitch. Adapting neural networks for the estimation of treatment effects. *Advances in neural information processing systems*, 32, 2019.

Claudia Shi, Victor Veitch, and David M Blei. Invariant representation learning for treatment effect estimation. In *Uncertainty in Artificial Intelligence*, pp. 1546–1555. PMLR, 2021.

Leonard A Stefanski and Dennis D Boos. The calculus of m-estimation. *The American Statistician*, 56(1):29–38, 2002.

Elizabeth A Stuart. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1):1, 2010.

Mark J Van Der Laan and Daniel Rubin. Targeted maximum likelihood learning. *The international journal of biostatistics*, 2(1), 2006.

Mark J Van der Laan, Sherri Rose, et al. *Targeted learning: causal inference for observational and experimental data*, volume 10. Springer, 2011.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256, 1992.

Pengzhou Wu and Kenji Fukumizu. \beta-intact-vae: Identifying and estimating causal effects under limited overlap. *arXiv preprint arXiv:2110.05225*, 2021.

Liuyi Yao, Sheng Li, Yaliang Li, Mengdi Huai, Jing Gao, and Aidong Zhang. Representation learning for treatment effect estimation from observational data. *Advances in Neural Information Processing Systems*, 31, 2018.

Jinsung Yoon, James Jordon, and Mihaela Van Der Schaar. Ganite: Estimation of individualized treatment effects using generative adversarial nets. In *International Conference on Learning Representations*, 2018.

Shengyu Zhang, Dong Yao, Zhou Zhao, Tat-Seng Chua, and Fei Wu. Causerec: Counterfactual user sequence synthesis for sequential recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 367–377, 2021.

Shengyu Zhu, Ignavier Ng, and Zhitang Chen. Causal discovery with reinforcement learning. *arXiv preprint arXiv:1906.04477*, 2019.

Yueting Zhuang, Ming Cai, Xuelong Li, Xiangang Luo, Qiang Yang, and Fei Wu. The next breakthroughs of artificial intelligence: The interdisciplinary nature of ai. *Engineering*, 6(3):245, 2020.

Hao Zou, Bo Li, Jiangang Han, Shuiping Chen, Xuetao Ding, and Peng Cui. Counterfactual prediction for outcome-oriented treatments. In *International Conference on Machine Learning*, pp. 27693–27706. PMLR, 2022.

A APPENDIX

A.1 THEORETICAL PROOF

Lemma A.1 (Properties of decomposed efficient influence curve)**.** *To facilitate following analysis, we present several properties of the decomposed efficient influence curve* $D_d^{eff} = \frac{\mathcal{I}(\mathbf{T}=1)-\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}(\mathbf{Y} - m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y}))$ *in the condition that* $\mathbf{V}$ *is a valid adjustment set:*

a. $\mathbb{E}[D_d^{eff}(\mathbf{V})] = 0$

b. $Var[D_d^{eff}(\mathbf{V})] = \mathbb{E}[(D_d^{eff}(\mathbf{V}))^2]$

Proof. a.

$$\begin{aligned}\mathbb{E}[D_d^{\text{eff}}(\mathbf{V})] &= \mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)-\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}(\mathbf{Y} - m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y}))\right] \\ &= \mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}}(\mathbf{V})}(\mathbf{Y} - m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y}))\right] - \mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}(\mathbf{Y} - m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y}))\right],\end{aligned}$$

where we calculate the expectation here with respect to the joint density P. We then expand the first term as follows:

$$\begin{aligned}\mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}}(\mathbf{V})}(\mathbf{Y} - m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y}))\right] &= \mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}=1}(\mathbf{V})}(\mathbf{Y}(1) - m_{\mathbf{V}}^{\mathbf{T}=1}(\mathbf{Y}))\right] \\ &= \mathbb{E}_{\mathbf{V}}\mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}=1}(\mathbf{V})}(\mathbf{Y}(1) - m_{\mathbf{V}}^{\mathbf{T}=1}(\mathbf{Y})) \mid \mathbf{V}\right] \\ &= \mathbb{E}_{\mathbf{V}}\left\{\mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}=1}(\mathbf{V})} \mid \mathbf{V}\right]\mathbb{E}\left[(\mathbf{Y}(1) - m_{\mathbf{V}}^{\mathbf{T}=1}(\mathbf{Y})) \mid \mathbf{V}\right]\right\} \\ &= 0,\end{aligned}$$

where the first equality is due to the consistency of $\mathbf{Y}$, the second equality is due to the tower property of expectation. Meanwhile, the third equality is due to the fact that $\mathbf{Y}(\mathbf{t}) \perp\!\!\!\perp \pi^{\mathbf{T}}(\mathbf{V}) \mid \mathbf{V}$. Finally, $\mathbb{E}\left[(\mathbf{Y}(1) - m_{\mathbf{V}}^{\mathbf{T}=\mathbf{1}}(\mathbf{Y})) \mid \mathbf{V}\right] = 0$ derives the last equality.

b.

$$\text{Var}[D_d^{\text{eff}}(\mathbf{V})] = \mathbb{E}[(D_d^{\text{eff}}(\mathbf{V}))^2] - (\mathbb{E}[D_d^{\text{eff}}(\mathbf{V})])^2 = \mathbb{E}[(D_d^{\text{eff}}(\mathbf{V}))^2].$$

□

Lemma A.2 (Validity of $D_d^{\text{eff}}(\mathbf{V})$)**.** *Similar to* D^{eff}*,* $\widehat{\gamma}(P)$ *is asymptotically linear with* D_d^{eff}*, and* $\sqrt{n}\left(\gamma\left(\widehat{P}\right) - \gamma(P)\right) \xrightarrow{d} N(0, Var[D_d^{eff}(\mathbf{V})])$.

Proof. First, we present the original derivation of D^{eff} as follows for convenience:

$$D^{\text{eff}}(\mathbf{V}) = \frac{\mathcal{I}(\mathbf{T}=1)-\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}(\mathbf{Y}(\mathbf{t}) - m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y})) + m_{\mathbf{V}}^{\mathbf{T}=1}(\mathbf{Y}) - m_{\mathbf{V}}^{\mathbf{T}=0}(\mathbf{Y}) - \gamma(P). \tag{7}$$

Then a previous proved conclusion is provided such that the estimator $\widehat{\gamma}(P)$ is asymptotically linear with influence curve as D^{eff} Van Der Laan & Rubin (2006):

$$\gamma\left(\widehat{P}\right) - \gamma(P) = \frac{1}{n}\sum_{i=1}^{n} D_i^{\text{eff}}(\mathbf{V}) + \mathcal{O}(\frac{1}{\sqrt{n}}). \tag{8}$$

Moreover, the decomposition of D^{eff} is also proposed in Van Der Laan & Rubin (2006):

$$\begin{cases} D_d^{\text{eff}} = D^{\text{eff1}} = \dfrac{\mathcal{I}(\mathbf{T}=1)-\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}(\mathbf{Y}(\mathbf{t}) - m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y})) \\ D^{\text{eff2}} = m_{\mathbf{V}}^{\mathbf{T}=1}(\mathbf{Y}) - m_{\mathbf{V}}^{\mathbf{T}=0}(\mathbf{Y}) - \gamma(P), \end{cases} \tag{9}$$

where the second term equals to zero under the integral of the empirical distribution P_n: $P_n D^{\text{eff2}} = 0$ Van der Laan et al. (2011). Thus we conclude that $\widehat{\gamma}(P)$ is asymptotically linear with influence curve as D^{eff2}: $\frac{1}{n}\sum_{i=1}^{n} D_i^{\text{eff}}(\mathbf{V}) = \frac{1}{n}\sum_{i=1}^{n} D_i^{\text{eff2}}(\mathbf{V}) + \mathcal{O}(1)$. Meanwhile, combined with previous conclusion in Lemma equation A.1 that $\mathbb{E}[D_d^{\text{eff}}(\mathbf{V})] = 0$, we have the following derivation:

$$\begin{aligned}\lim_{n\to+\inf} \sqrt{n}\left(\gamma\left(\widehat{P}\right) - \gamma(P)\right) &= \lim_{n\to+\inf}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(D_i^{\text{eff1}}(\mathbf{V}) + D_i^{\text{eff2}}(\mathbf{V})\right) + \sqrt{n}\mathcal{O}(\frac{1}{\sqrt{n}})\right\} \\ &= \lim_{n\to+\inf}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_i^{\text{eff1}}(\mathbf{V})\right\},\end{aligned} \tag{10}$$

where the CLT further implies that $\sqrt{n}\left(\gamma\left(\widehat{P}\right) - \gamma(P)\right) \xrightarrow{d} N(0, \text{Var}[D_d^{\text{eff}}(\mathbf{V})])$. □

Lemma A.3. *Suppose that the cumulative distribution function (CDF) F_n of $\gamma\left(\widehat{P}\right) - \gamma(P)$ is continuous within the sample size n increasing, then for any $\alpha \geq 0$ and n,*

$$P(|\gamma\left(\widehat{P}\right) - \gamma(P)| \geq \alpha) \leq \delta_n + 1 - F\left(\frac{\sqrt{n}\alpha}{\sqrt{Var[D_d^{eff}(\mathbf{V})]}}\right), \tag{11}$$

Proof. We first claim that although we suppose the continuity of the CDF, the similar conclusion can be extended to CDFs with non-left-continuous points as well. As the term δ_n controls the convergence of the series $\{F_i\}_{i=1}^{n}$ to F, results in Lemma A.2 imply that for any α in the support of P, the following inequality holds:

$$|F(\alpha) - F_n(\alpha)| \leq \delta_n \Longrightarrow 1 - F_n(\alpha) \leq 1 - F(\alpha) + \delta_n. \tag{12}$$

where F is the CDF of the normal distribution $N(0, \text{Var}[D_d^{\text{eff}}(\mathbf{V})])$. Meanwhile, we observe that $N(0, \text{Var}[D_d^{\text{eff}}(\mathbf{V})]) \overset{d}{=} Z * \sqrt{\text{Var}[D_d^{\text{eff}}(\mathbf{V})]}$, where $\overset{d}{=}$ refers to the in-distribution equality and $Z \sim N(0,1)$. Therefore, the above inequality can be derived as follows:

$$\begin{aligned}P(X \geq \alpha) &\leq \delta_n + 1 - P(X \geq \alpha) \\ &\leq \delta_n + 1 - P(Z * \sqrt{\text{Var}[D_d^{\text{eff}}(\mathbf{V})]} \geq \alpha),\end{aligned} \tag{13}$$

where the final conclusion is obtained when we further let $X = \sqrt{n}|\gamma\left(\widehat{P}\right) - \gamma(P)|$ and $\alpha_0 = \sqrt{n}\alpha$. □

Theorem A.4 (Connections between $\mathcal{R}^{\text{OAF}}$ and $\mathbf{V}$)**.** *We denote the selected features for adjustment as $\mathbf{V} \subseteq \{\mathbf{X} \cup \mathbf{I} \cup \mathbf{Z}\}$. Meanwhile, we denote the optimal adjustment set as $\mathbf{V}_0 = \{\mathbf{X} \cup \mathbf{Z}\}$. Then the optimality of our reward is stated from the following three sub-theorems:*

- *(a) If $\mathbf{V}$ is a valid adjustment set, then $\mathcal{R}^{OAF}(\mathbf{V}') \leq \mathcal{R}^{OAF}(\mathbf{V})$ holds for any $\mathbf{V}' = \mathbf{V} \cup \mathbf{Z}'$, where $\mathbf{Z}' \subseteq \mathbf{Z}$.*
- *(b) If $\mathbf{V}$ is a valid adjustment set, then $\mathcal{R}^{OAF}(\mathbf{V}) \leq \mathcal{R}^{OAF}(\mathbf{V}')$ holds for any $\mathbf{V}' = \mathbf{V} \cup \mathbf{I}'$, where $\mathbf{I}' \subseteq \mathbf{I}$.*
- *(c) We assume that the $\{\mathbf{X} \cup \mathbf{I} \cup \mathbf{Z}\}$ contains all the parents of $\mathbf{Y}$, which implies that $\mathbf{Z}$ contains all the outcome-precision variables of $\mathbf{Y}$. Then $\mathcal{R}^{OAF}(\mathbf{V_0}) \leq \mathcal{R}^{OAF}(\mathbf{V}')$ holds for any $\mathbf{V}'$ which is not a valid adjustment set.*

Overall, the above reflects two things (1) the deletion of $\mathbf{I}$ from $\mathbf{V}$ or the supplementation of $\mathbf{Z}$ into $\mathbf{V}$ result in the decrease of $\mathcal{R}^{\text{OAF}}(\mathbf{V})$ for valid $\mathbf{V}$; (2) when $\mathbf{V}'$ is invalid, $\mathcal{R}^{\text{OAF}}(\mathbf{V}')$ is larger than that of any valid adjustment set containing $\mathbf{Z}$. There is another intuitive explanation for conclusion (2) such that the residual term in equation 3 will become extremely large since $\mathbf{V}'$ loses some of the predictors in $\mathbf{X}$.

Proof. Some of the techniques in our proof here are inspired by Karimi et al. (2020).

(a) First, $\mathbf{V}' = \mathbf{V} \cup \mathbf{Z}'$ implies that $\mathbf{Z}' \perp\!\!\!\perp \mathbf{T} \mid \mathbf{V}$. Then $\pi^{\mathbf{T}}(\mathbf{V}) = \pi^{\mathbf{T}}(\mathbf{V}')$ holds, which further derives the following equation:

$$D_d^{\text{eff}}(\mathbf{V}) = D_d^{\text{eff}}(\mathbf{V}') + \underbrace{\frac{\mathcal{I}(\mathbf{T}=1) - \mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}(m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y}))}_{D_M}.$$

Then we obtain the fact that $\mathbb{E}[D_M D_d^{\text{eff}}(\mathbf{V}')] = 0$:

$$\begin{aligned}
\mathbb{E}[D_M D_d^{\text{eff}}(\mathbf{V}')] &= \mathbb{E}_{\mathbf{V}'}\mathbb{E}[D_M D_d^{\text{eff}}(\mathbf{V}') \mid \mathbf{V}'] \\
&= \mathbb{E}_{\mathbf{V}'}\left[\sum_{\mathbf{t}\in\{0,1\}} (2\mathbf{t}-1)(m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y}))\mathbb{E}[\mathbf{Y}^{\mathbf{t}} - m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y}) \mid \mathbf{V}']\mathbb{E}[\frac{(\mathcal{I}(\mathbf{T}=\mathbf{t}))^2}{(\pi^{\mathbf{T}}(\mathbf{V}'))^2} \mid \mathbf{V}']\right] \\
&= 0,
\end{aligned}$$

where the first equality is due to the tower property, the second equality is due to the fact that both $\mathbf{V}$ and $\mathbf{V}'$ are valid adjustment sets, and the third equality is due to the fact that $\mathbb{E}[\mathbf{Y}(\mathbf{t}) - m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y}) \mid \mathbf{V}'] = 0$. Meanwhile, we derive the expectation of the term D_M as follows:

$$\begin{aligned}
\mathbb{E}[D_M] &= \mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1) - \mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}(m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y}))\right] \\
&= \underbrace{\mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}}(\mathbf{V})}(m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y}))\right]}_{D_M^1} - \underbrace{\mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}(m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y}))\right]}_{D_M^2},
\end{aligned}$$

where the term D_M^1 is then simplified as follows:

$$\begin{aligned}
D_M^1 &= \mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}=1}(\mathbf{V})}(m^{\mathbf{T}=1}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}=1}_{\mathbf{V}}(\mathbf{Y}))\right] \\
&= \mathbb{E}_{\mathbf{V}}\mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}=1}(\mathbf{V})}(m^{\mathbf{T}=1}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}=1}_{\mathbf{V}}(\mathbf{Y})) \mid \mathbf{V}\right] \\
&= \mathbb{E}_{\mathbf{V}}\left[\mathbb{E}\left[(m^{\mathbf{T}=1}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}=1}_{\mathbf{V}}(\mathbf{Y})) \mid \mathbf{V}\right]\mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}=1}(\mathbf{V}')} \mid \mathbf{V}\right]\right],
\end{aligned}$$

where the term $\mathbb{E}\left[(m^{\mathbf{T}=1}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}=1}_{\mathbf{V}}(\mathbf{Y})) \mid \mathbf{V}\right] = 0$ due to the fact that $\mathbb{E}[m^{\mathbf{T}=1}_{\mathbf{V}'}(\mathbf{Y}) \mid \mathbf{V}] = m^{\mathbf{T}=1}_{\mathbf{V}}(\mathbf{Y})$. The simplification of D_M^2 is similar to that of D_M^1. Thus we derive that $\mathbb{E}[D_M] = 0$. Finally, we derive the formulation of $\text{Var}(D_d^{\text{eff}}(\mathbf{V}))$ as follows:

$$\begin{aligned}
\text{Var}(D_d^{\text{eff}}(\mathbf{V})) &= \text{Var}(D_d^{\text{eff}}(\mathbf{V}')) + \text{Var}(D_M) \\
&= \text{Var}(D_d^{\text{eff}}(\mathbf{V}')) + \mathbb{E}\left[D_M^2\right] \\
&= \text{Var}(D_d^{\text{eff}}(\mathbf{V}')) + \mathbb{E}\left[\left(\frac{\mathcal{I}(\mathbf{T}=1) - \mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}(m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y}))\right)^2\right] \\
&= \text{Var}(D_d^{\text{eff}}(\mathbf{V}')) + \mathbb{E}_{\mathbf{V}}\left[\mathbb{E}\left[\left(\frac{\mathcal{I}(\mathbf{T}=1) - \mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}\right)^2 \mid \mathbf{V}\right]\mathbb{E}\left[(m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y}) - m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y}))^2 \mid \mathbf{V}\right]\right] \\
&= \text{Var}(D_d^{\text{eff}}(\mathbf{V}')) + \mathbb{E}_{\mathbf{V}}\left[\text{Var}(m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y}) \mid \mathbf{V})\left(\frac{1}{p(\mathbf{T}=1 \mid \mathbf{V})} + \frac{1}{p(\mathbf{T}=0 \mid \mathbf{V})}\right)\right],
\end{aligned}$$

where the last equality is due to the fact that $\mathbb{E}[m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y}) \mid \mathbf{V}] = m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y})$ with some algebra on the term $\mathbb{E}\left[\left(\frac{\mathcal{I}(\mathbf{T}=1)-\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V})}\right)^2 \mid \mathbf{V}\right]$.

(b) First, $\mathbf{V}' = \mathbf{V} \cup \mathbf{I}'$ and $\mathbf{I}' \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{V}, \mathbf{T}$ imply that $m^{\mathbf{T}}_{\mathbf{V}}(\mathbf{Y}) = m^{\mathbf{T}}_{\mathbf{V}'}(\mathbf{Y})$ holds. Then we derive the following decomposition of $\text{Var}(D_d^{\text{eff}}(\mathbf{V}'))$:

$$\text{Var}(D_d^{\text{eff}}(\mathbf{V}')) = \text{Var}(\mathbb{E}\left[D_d^{\text{eff}}(\mathbf{V}') \mid \mathbf{T}, \mathbf{V}, \mathbf{Y}\right]) + \mathbb{E}\left[\text{Var}\left(D_d^{\text{eff}}(\mathbf{V}') \mid \mathbf{T}, \mathbf{V}, \mathbf{Y}\right)\right], \quad (14)$$

where the term $\mathbb{E}\left[D_d^{\text{eff}}(\mathbf{V}') \mid \mathbf{T}, \mathbf{V}, \mathbf{Y}\right]$ is simplified as follows:

$$\begin{aligned}
\mathbb{E}\left[D_d^{\text{eff}}(\mathbf{V}') \mid \mathbf{T}, \mathbf{V}, \mathbf{Y}\right] &= \mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)-\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V}')}(\mathbf{Y}-m_{\mathbf{V}'}^{\mathbf{T}}(\mathbf{Y})) \mid \mathbf{T}, \mathbf{V}, \mathbf{Y}\right] \\
&= \mathbb{E}\left[\frac{\mathcal{I}(\mathbf{T}=1)-\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V}')}(\mathbf{Y}-m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y})) \mid \mathbf{T}, \mathbf{V}, \mathbf{Y}\right] \\
&= \sum_{\mathbf{t}\in\{0,1\}} (2\mathbf{t}-1)\mathcal{I}(\mathbf{T}=\mathbf{t})(\mathbf{Y}-m_{\mathbf{V}}^{\mathbf{t}}(\mathbf{Y}))\mathbb{E}\left[\frac{1}{\pi^{\mathbf{T}}(\mathbf{V}')} \mid \mathbf{T}, \mathbf{V}\right] \\
&= \sum_{\mathbf{t}\in\{0,1\}} (2\mathbf{t}-1)\mathcal{I}(\mathbf{T}=\mathbf{t})(\mathbf{Y}-m_{\mathbf{V}}^{\mathbf{t}}(\mathbf{Y}))\frac{1}{\pi^{\mathbf{T}}(\mathbf{V})} \\
&= D_d^{\text{eff}}(\mathbf{V}).
\end{aligned}$$

Then, we apply the results in equation 14 and simplify the expression of $\text{Var}(D_d^{\text{eff}}(\mathbf{V}'))$ as follows:

$$\begin{aligned}
\text{Var}(D_d^{\text{eff}}(\mathbf{V}')) &= \text{Var}(D_d^{\text{eff}}(\mathbf{V})) + \mathbb{E}\left[\text{Var}\left(D_d^{\text{eff}}(\mathbf{V}') \mid \mathbf{T}, \mathbf{V}, \mathbf{Y}\right)\right] \\
&= \text{Var}(D_d^{\text{eff}}(\mathbf{V})) + \mathbb{E}\left[\text{Var}\left(\frac{\mathcal{I}(\mathbf{T}=1)-\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}}(\mathbf{V}')}(\mathbf{Y}-m_{\mathbf{V}'}^{\mathbf{T}}(\mathbf{Y})) \mid \mathbf{T}, \mathbf{V}, \mathbf{Y}\right)\right] \\
&= \text{Var}(D_d^{\text{eff}}(\mathbf{V})) + \mathbb{E}\left[\sum_{\mathbf{t}\in\{0,1\}} \mathcal{I}(\mathbf{T}=\mathbf{t})\left(\mathbf{Y}-m_{\mathbf{V}'}^{\mathbf{T}}(\mathbf{Y})\right)^2 \text{Var}\left(\frac{1}{\pi^{\mathbf{T}}(\mathbf{V}')} \mid \mathbf{T}, \mathbf{V}, \mathbf{Y}\right)\right].
\end{aligned}$$

(c) Once the variable set $\mathbf{Z}$ contains all the parents of $\mathbf{Y}$, we can write the structural equation of $\mathbf{Y}$ as $\mathbf{Y} = f_{\mathbf{Y}}(\mathbf{T}, \mathbf{X}, \mathbf{Z})$. Then the proposed OAF metric of $\mathbf{V}_0$ can be derived as follows:

$$\text{Var}(D_d^{\text{eff}}(\mathbf{V}_0)) = \underbrace{\mathbb{E}\left[\left(\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}=1}(\mathbf{V}_0)}(\mathbf{Y}-m_{\mathbf{V}_0}^{\mathbf{T}=1}(\mathbf{Y}))\right)^2\right]}_{K_1} + \underbrace{\mathbb{E}\left[\left(\frac{\mathcal{I}(\mathbf{T}=0)}{\pi^{\mathbf{T}=0}(\mathbf{V}_0)}(\mathbf{Y}-m_{\mathbf{V}_0}^{\mathbf{T}=0}(\mathbf{Y}))\right)^2\right]}_{K_2},$$

where we further expand K_1 as follows:

$$\begin{aligned}
K_1 &= \mathbb{E}_{\mathbf{V_0}}\left[\mathbb{E}\left[\left(\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}=1}(\mathbf{V}_0)}\right)^2 \mid \mathbf{V_0}\right]\mathbb{E}\left[\left(\mathbf{Y}(\mathbf{t}=1)-m_{\mathbf{V}_0}^{\mathbf{T}}(\mathbf{Y})\right)^2 \mid \mathbf{V_0}\right]\right] \\
&= \mathbb{E}_{\mathbf{V_0}}\left[\mathbb{E}\left[\left(\frac{\mathcal{I}(\mathbf{T}=1)}{\pi^{\mathbf{T}=1}(\mathbf{V}_0)}\right)^2 \mid \mathbf{V_0}\right]\mathbb{E}\left[\left(\mathbf{Y}(\mathbf{t}=1)-f_{\mathbf{Y}}(\mathbf{T}, \mathbf{X}, \mathbf{Z})\right)^2 \mid \mathbf{V_0}\right]\right] \\
&= 0,
\end{aligned}$$

where the second equality holds due to the fact that $\mathbf{Z}$ contains all the parents of $\mathbf{Y}$. Similar to above derivation, we obtain that $K_2 = 0$. Furthermore, we conclude that $\text{Var}(D_d^{\text{eff}}(\mathbf{V}_0)) = 0$, which indicates that $\text{Var}(D_d^{\text{eff}}(\mathbf{V}_0)) \leq \text{Var}(D_d^{\text{eff}}(\mathbf{V}'))$.

□

B Experimental Details

B.1 Details on Datasets

Synthetic In detail, we set the coefficient $\beta^{xy} = 4$ and $\beta^{zy} = -2$ for more significant difference between the effects of $\mathbf{X}$ and $\mathbf{Z}^e$ on $\mathbf{Y}$. Meanwhile, we sample β^{Io}, β^{Zo} from $U(0,1)$, together with σ^Y, σ^{Io} and σ^{Zo} sampled from $N(0,2)$.

IHDP We follow the classical surface-B setting in Hill (2011) to generate the IHDP dataset with the real-world 25 covariates. In detail, we set 5 continuous covariates (as all) as the confounders $\mathbf{X}$.

Table 5: Linear simulation results. The metrics are Mean±STD over 10 repeated experiments. The best performance is marked bold.

Linear Simulation							
Settings		In_sample Prediction			Out_of_sample Prediction		
Feature Dimension		20	40	80	20	40	80
Statistical	Direct	5.33±0.53	6.88±0.69	8.66±0.85	5.54±1.41	7.36±1.26	6.25±1.84
	IPW	0.68±2.20	0.93±2.90	1.11±2.59	0.87±2.33	1.51±3.11	2.28±3.87
	AIPW	0.30±0.64	0.93±2.90	1.11±2.59	0.27±2.00	0.76±1.96	1.39±1.41
	TMLE	0.25±0.07	0.58±0.11	0.61±0.05	0.48±0.10	0.60±0.13	0.65±0.11
Machine	DragonNet	0.05±0.48	0.29±0.15	0.49±0.37	0.93±0.42	0.90±0.37	1.05±0.86
	GANITE	0.86±0.00	0.97±0.00	1.01±0.00	0.99±0.00	1.00±0.00	1.01±0.00
	DNOUT	0.46±0.03	0.51±0.03	0.64±0.14	0.40±0.02	0.49±0.04	0.62±0.11
	BART	1.02±0.12	2.13±0.15	2.56±0.61	1.51±1.00	1.99±0.28	2.87±0.57
Decomposed	AIPW_L	0.50±0.41	0.57±0.49	0.64±0.35	0.60±0.34	0.67±0.38	0.91±0.52
	DVD	1.02±0.15	0.74±0.46	0.84±0.14	1.06±0.04	0.70±0.00	0.91±0.00
	DR-CFR	0.66±0.26	0.44±0.18	0.30±0.13	0.69±0.12	0.68±0.08	0.46±0.06
	NICE	1.02±0.09	1.05±0.11	1.34±1.08	0.90±0.18	0.96±0.13	1.18±0.84
Ours	**OAFP_L**	**0.03±0.01**	**0.01±0.01**	**0.08±0.02**	**0.14±0.03**	**0.15±0.07**	**0.12±0.08**
	OAFP_N	**0.01±0.02**	**0.02±0.01**	**0.01±0.01**	**0.06±0.08**	**0.03±0.05**	**0.02±0.03**

Meanwhile, we randomly select half of the rest 20 discrete variables as $\mathbf{I}$, with the rest as $\mathbf{Z}$. To this end, we select $\mathbf{Z}$ or $\mathbf{I}$ from the Bernouli distribution $B(0.5)$ for each discrete variable. The effect coefficients of $\mathbf{X}$ and $\mathbf{Z}$ on $\mathbf{Y}$, namely β_{xy} and β_{zy}, is generated in the same protocol in Hill (2011). The effect coefficients of $\mathbf{I}$ and $\mathbf{X}$ on $\mathbf{T}$, namely β_{it} and β_{xt}, are generated from $U(-2,2)$ as the uniform distributions. Furthermore, the $\mathbf{Y}_1$, $\mathbf{Y}_0$ and $\mathbf{T}$ are generated as follows:

$$\begin{cases} \mathbf{Y}_1 = \beta_{xy}^T\mathbf{X} + \beta_{zy}^T\mathbf{Z} - \omega + N(0,1), \\ \mathbf{Y}_0 = \exp(\beta_{xy}^T\mathbf{X} + \beta_{zy}^T\mathbf{Z}) + N(0,1), \\ \mathbf{T} \sim \text{ Bernoulli } \left(1/\left(1+\exp\left(-\left(\beta_{xt}^T\mathbf{X} + \beta_{it}^T\mathbf{I}\right)\right)\right)\right) \end{cases}, \tag{15}$$

where ω refers to the term to keep the Average Treatment Effect on the Treated (ATT) close to 4 Hill (2011). Here due to the reason that the covariates $\mathbf{X}$ are fixed, we do not distinguish pre-outcome and post-outcome variables in IHDP.

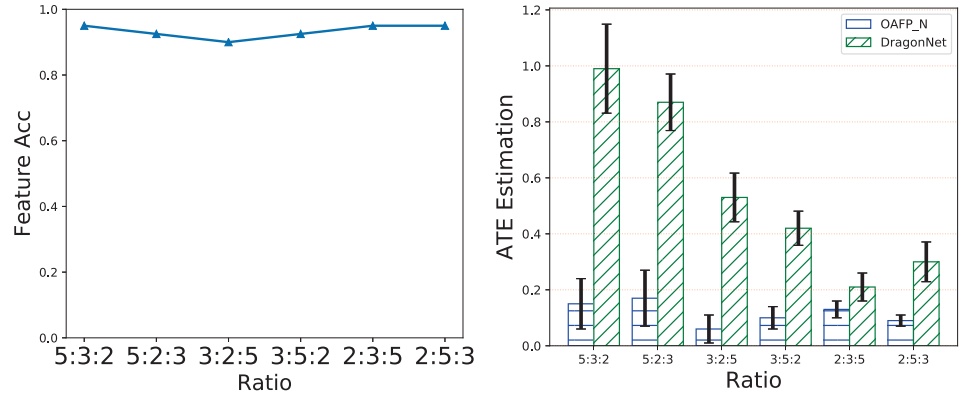
Twins The original Twins dataset is derived from the all twins born in the USA between the year of 1989 and 1991 Almond et al. (2005). The original treatment equaling to 1 indicates the heavier one in the twins, and vice versa. Following previous protocols Louizos et al. (2017), we select 28 variables related to parents, pregnancy, and birth, with the outcome recording the children's mortality after one year. Moreover, we pro-pressing the dataset by filtering out entries with the same-sex twins weighing less than 2000g or with missing features. Finally, we obtain 5271 samples for experiments. As the original treatment $\mathbf{T}$ is assigned randomly, the typical approach for observational studies is to re-simulating the treatment. To this end, we introduce 5 pre-treatment $\mathbf{I}^e$ and 5 post-treatment $\mathbf{I}^o$ by adding them to covariates, resulting the 38-dimension covariates. In detail, we sample $\mathbf{I}^e$ from $N(0,1)$. Then $\mathbf{T}$ is sampled from the Bernouli-logistic approach as follows:

$$\mathbf{T} \sim \text{ Bernoulli } \left(1/\left(1+\exp\left(-\left(\beta_{xt}^T\mathbf{X} + \beta_{it}^T\mathbf{I}^e\right) + N(0,0.5)\right)\right)\right), \tag{16}$$

with β_{it} and β_{xt} sampled from $U(-2,2)$. Moreover, $\mathbf{I}^o$ is generated as $\mathbf{I}^o = \beta^{Io}\mathbf{T} + \sigma^{Io}$, with $\beta^{Io} \sim U(-2,2)$ and $\sigma^{Io} \sim N(0,0.5)$.

B.2 Details on Implementation

Details on baselines For baselines we have compared in this paper, we exactly follow the optimal hyper-parameters with the original network architectures in their open-source implementations. Notably, the NICE Shi et al. (2021) method requires the multiple environments to support the identification of optimal adjustment features, where different environment is generated using distinct causal graph. For our problem, to adapt NICE method, we randomly split the training data into three environments to simulate the heterogeneous training domains.



(a) Feature Accuracy

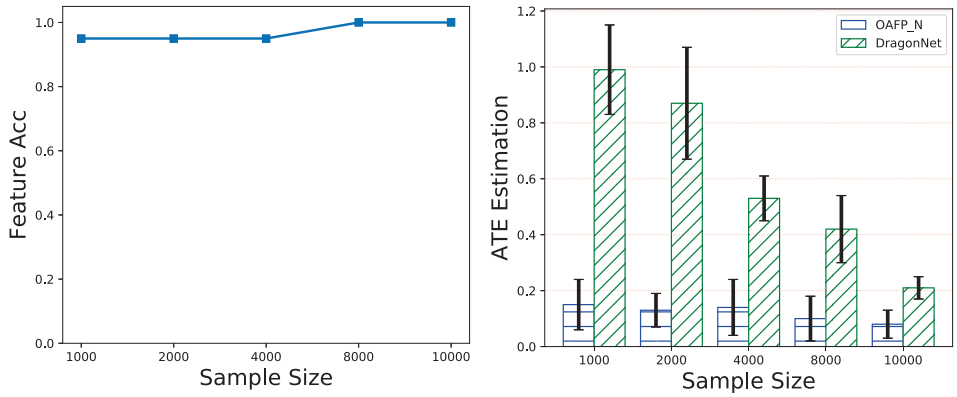
(b) ATE Estimation

Figure 4: Results with different variable ratio.

Details on estimating $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ and $\widehat{m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y})}$ For the linear implementation OAFP_L, we build $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ and $\widehat{m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y})}$ using the linear regression and logistic regression without any regularization tricks. Meanwhile, our implementation on the downstream estimator, namely the AIPW estimator, follows the Zepid package Masoumi & Shahraz (2022). For the non-linear implementation OAFP_N, we build $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ and $\widehat{m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y})}$ with two deep networks. The regression network for $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ consists of four MLP layers with the activation function as ELU, and the score network consists of three MLP layers with ELU as the activation function for the first two layers and $Sigmoid$ for the last layer. The optimizer we choose for $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ and $\widehat{m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y})}$ is the Adam optimizer , where the learning rate is 0.001 and 0.0005, respectively. Notably, we split an extra validation set from the training data such that $\widehat{\pi^{\mathbf{T}}(\mathbf{V})}$ and $\widehat{m_{\mathbf{V}}^{\mathbf{T}}(\mathbf{Y})}$ are evaluated on the validation part. Besides, we implement OAFP_L on a single Tesla V100 gpu. For OAFP_N, we compute $\widehat{\mathcal{R}}$ on a 8-gpu Tesla V100 cluster, where each batch array B_i is trained in a single process in parallel.

Details on our RL framework Our implementation follows the previous neural combinational search Bello et al. (2016); Zhu et al. (2019), where the encoder is a Transformer and the decoder is a multi-layer MLP. As shown in Figure 2, the Transformer takes the covariates U as input, with the total input size as $\mathcal{R}^{K\times d\times n_b}$ (K is the batch size, d is the dimension of $\mathbf{U}$ and n_b is the sample number for each array B_i in a batch). Under the alternative feed-forward by multi-layer MLPs and multi-head attention module in each block, the representation (the output of the encoder) owns the same shape as $\mathcal{R}^{K\times d\times n_b}$. Meanwhile, we concentrate $\mathbf{T}$ and $\mathbf{Y}$ with the representation of $\mathbf{U}$ learned from the Transformer, and then feed them into the MLP decoder. Finally, the MLP decoder sample the binary feature mask with the sigmoid functions. We set n_b = 512 and K = 64 throughout our experiments. To be specific, our Transformer encoder has two blocks. The MLP decoder has two linear layers with the $Relu$ activation function. Moreover, we adopt the reinforce approach Williams (1992) to non-parametrically reduce the variance of actor. To achieve this, we take the exponential average of the past reward as the baseline term and the scaling factor is set to 0.99. Meanwhile, the hyper-parameter for controlling the entropy term is set to 1.

Details on the downstream estimators AIPW and DragonNet For two downstream estimators, namely the AIPW and DragonNet, we implement AIPW using the zepid package, where DragonNet is reproduced with the original open-source implementation. In detail, AIPW uses the linear regression and logistic regression to estimate the outcome regression and the propensity score, respectively. The two estimators are then combined in the form of semi-parametric approaches Hines et al. (2022). As the deep version of AIPW, DragonNet ensembles the optimization of score prediction and outcome prediction into an end-to-end network with target regularization to satisfy the estimation equation Shi et al. (2019). The architecture of DragonNet consists of the 4-layer representation MLD layers activated by the ELU function, where the following three heads contains a single-layer MLP with $sigmoid$ as the score head, two three-layer with ELU as the outcome heads. The optimizer of DragonNet is Adam optimizer with the initial learning rate as 0.001.



(a) Feature Accuracy

(b) ATE Estimation

Figure 5: Results with different sample size.

B.3 Extra experimental results

First, we report the results of ATE estimation in the linear case in Table 5 here. The reported ATE estimation results also reflect that, the deletion of treatment-only variables plays a vital role to improve performance. Moreover, we perform some ablation studies on the data simulation to check whether the proposed method is affected by (a) the variation on the ratio of $\mathbf{I}$; (b) the variation in the sample size. We choose the non-linear synthetic data with the total dimension $d = 20$ for our ablation study. As the original ratio of $\mathbf{I}$:$\mathbf{X}$:$\mathbf{Z}$ is 5:3:2, we simulate the other 5 cases as $5:2:3$, $3:2:5$,$3:5:2$,$2:3:5$ and $2:5:3$ by tuning the ratio of $\mathbf{I}$. Related results are further presented in Figure 4, where our method substantially achieves the selection of the optimal adjustment set with accurate ATE estimation. Notably, an interesting phenomenon is that the performance gap between the original DragonNet and our OAFP_N decreases when the ratio of $\mathbf{I}$ decreases, which strengthens our viewpoint that including $\mathbf{I}$ is harmful for ATE estimation. To perform the ablation study on the variation of the sample size N, we choose the ratio of the variables as $5:3:2$ as in our paper and vary N from 1000 to 10000. The corresponding results are present in Figure 5. Obviously, our method is not sensitive to the variation of sample size, and the baseline does not achieve significant improvement within sample size increases. Besides, such results reflect that the harm brought by $\mathbf{I}$ cannot be alleviated by increasing the sample size.