# Regret Minimization for Causal Inference on Large Treatment Space

**Akira Tanimoto**
NEC Corporation, Kyoto University, RIKEN

**Tomoya Sakai**
NEC Corporation, RIKEN

**Takashi Takenouchi**
Future University Hakodate, RIKEN

**Hisashi Kashima**
Kyoto University

## Abstract

Predicting which action (treatment) will lead to a better outcome is a central task in decision support systems. To build a prediction model in real situations, learning from observational data with a sampling bias is a critical issue due to the lack of randomized controlled trial (RCT) data. To handle such biased observational data, recent efforts in causal inference and counterfactual machine learning have focused on debiased estimation of the potential outcomes on a binary action space and the difference between them, namely, the conditional average treatment effect. When it comes to a large action space (e.g., selecting an appropriate combination of medicines for a patient), however, the regression accuracy of the potential outcomes is no longer sufficient in practical terms to achieve a good decision-making performance. This is because a high mean accuracy on the large action space does not guarantee the nonexistence of a single potential outcome misestimation that misleads the whole decision. Our proposed loss minimizes the classification error of whether or not the action is relatively good for the individual target among all feasible actions, which further improves the decision-making performance, as we demonstrate. We also propose a network architecture and a regularizer that extracts a debiased representation not only from the feature but also from the biased action for better generalization on large action spaces. Extensive experiments on synthetic and semi-synthetic datasets demonstrate the superiority of our method for large combinatorial action spaces.

Proceedings of the 24[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130. Copyright 2021 by the author(s).

## 1 INTRODUCTION

Predicting individualized causal effects is an important issue in many domains for decision-making. For example, a doctor considers which medication would be the most effective for a patient, a teacher considers which problems are most effective for helping student learn, and a retail store manager considers which assortment of items would improve the overall store sales. To support such decision-making, we advocate providing a prediction of which actions will lead to better outcomes.

Recent efforts in causal inference and counterfactual machine learning have focused on making predictions of the potential outcomes that correspond to each action for each individual target on the basis of observational data. Observational data consists of features of targets, past actions actually taken, and their outcomes. We have no direct access to the past decision-makers' policies, i.e., the mechanism of how to choose an action under a given target feature. Unlike in normal prediction problems, pursuing high-accuracy predictions only with respect to the historical data carries the risk of incorrect estimates due to the sampling bias in the past policies. This bias may cause *spurious correlation* (Simon, 1954; Pearl, 2009), which might mislead the decision-making. For those cases where real-world experiments such as randomized controlled trials (RCTs) or multi-armed bandit are infeasible or too expensive, causal inference methods provide debiased estimation of potential outcomes from observational data.

While most of the existing approaches assume limited action spaces such as a binary one, as in conditional average treatment effect (CATE) estimation, there are

many real-world situations where the number of options is large. For example, doctors need to consider which combination of medicines will best suit a patient.

For such cases, it is difficult to apply existing methods (as in (Shalit et al., 2017; Yoon et al., 2018; Schwab et al., 2018; Lopez et al., 2020)) for two reasons. One is the issue of sample-efficiency for large action spaces. Since the sample sizes for each action would be limited, building models for each action (or using a multi-head neural network), which existing methods adopt, is not sample-efficient. The other reason is the gap between the decision-making performance and the regression accuracy of the potential outcome. Even if we manage to achieve the same level of regression accuracy as when the action space is limited, the same decision-making performance is no longer guaranteed in a large action space, as we demonstrate in Section 4. This is because, in a nutshell, the overestimated potential outcome of only a single action may mislead the decision, even though it has only a small impact on the mean regression accuracy over all actions.

To achieve informative causal inference for decision-making in a large action space, we propose solutions for the above two issues. For the sample-efficiency, we propose extracting representations not only from features but also from actions. We extend two existing representation-based causal effect inference methods, respectively, to balance the representation distribution to be similar to that in the randomized trials.

For the gap between the decision performance and the regression accuracy, we prove that we can further improve the decision performance by minimizing the classification error of whether or not each action is relatively good for each target, in addition to the regression error (MSE). Unlike the recommendation problems in which ranking losses can be used, we cannot directly observe whether the action is relatively good or not since only one action and its outcome is observed for each target. We therefore propose a proxy loss that compares the observed outcome to the estimated conditional average performance of the past decision-makers, which is estimated by regular supervised learning.

In summary, our proposed method minimizes both the classification error and the MSE by using debiased representations of both the features and the actions. We demonstrate the effectiveness of our method through extensive experiments with synthetic and semi-synthetic datasets.

| $x$ | $a$ | | $Y_a$ | | | | | | | | $y$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $a_0$ | 0 | | | | 1 | | | | |
| | | $a_1$ | 0 | | 1 | | 0 | | 1 | | |
| | | $a_2$ | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | |
| $x_1$ | $(0,0,1)$ | | – | 1 | – | – | – | – | – | – | 1 |
| $x_2$ | $(0,1,0)$ | | – | – | 3 | – | – | – | – | – | 3 |
| $x_3$ | $(0,0,0)$ | | 4 | – | – | – | – | – | – | – | 4 |
| $x_4$ | $(1,0,1)$ | | – | – | – | – | – | 6 | – | – | 6 |

Figure 1: An example data table for our causal inference on a combinatorial action space. Dashes indicate missing entries. Only factual outcomes are observed (when $a = a'$, $y_{a'}$ is observed) and the counterfactual records $\{y_a\}_{a \neq a'}$ are missing.

## 2 PROBLEM SETTING

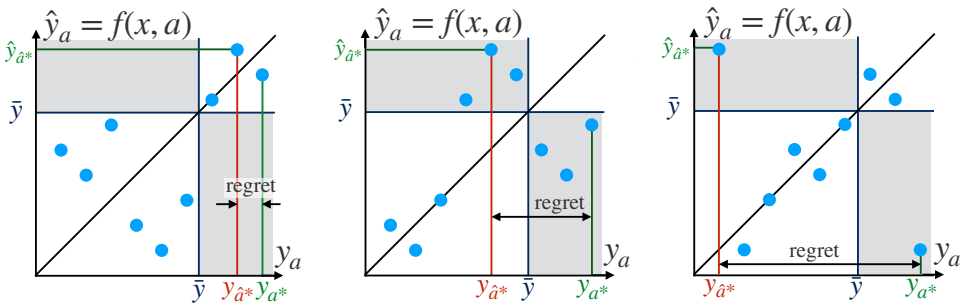
In this section, we formulate our problem and define a decision-focused performance metric. Our aim is to build a predictive model to inform decision-making. Given a feature vector $x \in \mathcal{X} \subset \mathbb{R}^d$, the learned predictive model $f$ is expected to correctly predict which action $a \in \mathcal{A}(x)$ leads to a better outcome $y \in \mathcal{Y} \subset \mathbb{R}$, where $\mathcal{A}(x)$ is a feasible subset of a finite action space $\mathcal{A}$ given $x$. We hereafter assume that the feasible action space does not depend on the feature, i.e., $\mathcal{A}(x) = \mathcal{A}$, for simplicity. A typical case of large action spaces is when an action consists of multiple causes, i.e., $\mathcal{A} = \{0,1\}^m$ (combinatorial action space).

We assume there exists a joint distribution $p(x, a, y_1, \ldots, y_{|\mathcal{A}|}) = p(x)\mu(a|x)p(y_1, \ldots, y_{|\mathcal{A}|}|x)$, where $\mu(a|x)$ is the unknown decision-making policy of past decision-makers, called propensity, and $y_1, \ldots, y_{|\mathcal{A}|}$ are the potential outcomes corresponding to each action. The observed (factual) outcome $y$ is the one corresponding to the observed action $a$, i.e., a training instance is the triplet $(x_n, a_n, y_{a_n})$, where $n$ denotes the instance index, and the other (counterfactual) potential outcomes are regarded as missing, as shown in Fig. 1.

We make the following assumptions on the distributions of the observational data.

- $(y_1, \ldots, y_{|\mathcal{A}|}) \perp a|x$ (unconfoundedness)
- $\forall a \in \mathcal{A}$ and $\forall x$, $0 < \mu(a|x) < 1$ (overlap)

These are commonly required to identify causal effects (Imbens and Wooldridge, 2009; Pearl, 2009).




(a) Zero $\mathrm{ER}^u_\mu$ (b) High $\mathrm{ER}^u_\mu$ (c) Low $\mathrm{ER}^u_\mu$, high $\mathrm{MSE}^u$

Figure 2: Example scatter plots of true vs. predicted potential outcomes for a target (fixed $x$) for different models. Each point corresponds to an action. $\mathrm{ER}^u_\mu$ corresponds to the rate of instances in the shaded areas. Assuming that the predicted best action $\hat{a}^* := \arg\max_a f(x,a)$ is adopted, minimizing the difference between its outcome $y_{\hat{a}^*}$ and the true optimal outcome $y_{a^*}$ (regret) is our aim (see the definition in Section 4).

# 3 REGRET MINIMIZATION NETWORK: DEBIASED POTENTIAL OUTCOME REGRESSION AND CLASSIFICATION

For this problem of estimating the action evaluation model, we propose our regret minimization network (RMNet), which consists of two parts: 1) a decision-focused risk to reduce the gap between the decision-making performance and the regression accuracy, and 2) representation balancing methods for debiased and sample-efficient learning.

## 3.1 Decision-Focused Risk

Most of the existing causal effect inference methods aim at minimizing the MSE of the treatment effect (a.k.a. the precision in estimation of heterogeneous effect (PEHE) (Hill, 2011)) in the binary treatment setting. In multiple treatment settings, a typical performance measure is the MSE averaged uniformly over all the actions (Schwab et al., 2018; Yoon et al., 2018):

$$\mathrm{MSE}^u(f) = \mathop{\mathbb{E}}_x \left[ \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} \mathop{\mathbb{E}}_{y_a|x} [(y_a - f(x,a))^2] \right]. \quad (1)$$

We refer to $\mathrm{MSE}^u$ as MSE, or specifically the uniform MSE, in this paper.

On the other hand, there is a gap between the decision performance and the regression accuracy ($\mathrm{MSE}^u$). Specifically, we do not necessarily have to accurately estimate the outcomes of candidate actions, but it is enough to identify better actions among others to achieve a higher decision-making performance. This is analogous to the personalized ranking approach in recommender systems (Rendle et al., 2009), in which pairwise comparison of the item preference for each target user is considered.

The pairwise ranking approach (Joachims, 2002; Burges et al., 2005) measures the consistency between the actual and predicted orders by means of the error rate of pairwise comparison as

$$\mathrm{ER}_{\mathrm{rank}}(f) = \mathop{\mathbb{E}}_{i,j} \left[ I(y_i \geq y_j \oplus f(x_i) \geq f(x_j)) \right],$$

where $\oplus$ denotes the logical XOR. However, we cannot apply a regular pairwise loss, since we typically only have the outcome for one action observed among the feasible actions. Instead, we propose minimizing the following comparison loss to the average performance of the past decision-makers as the personalized baseline for the target ($x$):

$$\mathrm{ER}^u_\mu(f) = \mathop{\mathbb{E}}_x \left[ \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} I(y_a \geq \bar{y} \oplus f(x,a) \geq \bar{y}) \right], \quad (2)$$

where $\bar{y} = \mathbb{E}_{a \sim \mu(a|x)} [Y_a | x]$ is the average performance of the past decision-makers under $x$. As shown in Fig. 2, minimizing $\mathrm{ER}^u_\mu$ leads to better models in terms of decision performance. The MSE is the same in Fig. 2(a) and Fig. 2(b), and thus MSE cannot be used to determine which of these prediction models is better. Minimizing $\mathrm{ER}^u_\mu$ enables us to correctly choose the model in Fig. 2(a) with a high decision performance (small regret).

Replacing the expected value $\bar{y}$ with its estimation $g(x)$ and the 0-1 loss with cross entropy, we get the following risk:

$$\widetilde{\mathrm{ER}}^u_g(f) = \mathop{\mathbb{E}}_x \left[ -\frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} \{ s \log v + (1-s) \log(1-v) \} \right], \quad (3)$$

where $s := I(y - g(x) \geq 0)$ and $v := \sigma(f(x,a) - g(x))$, and $g(x) \simeq \mathbb{E}_{a\sim\mu(a|x)}[Y_a|x]$ is the estimated average performance of the past decision-makers. We first fit $g$ with the standard supervised learning procedure from $\{(x_n, y_n)\}$ and then plug it into (3).

Not only the classification error but also the regression error (MSE) matters to the decision-making performance. This is because even with high classification accuracy, decisions might be misleading if only one misclassified action $a$ is predicted as the best ($\hat{y}_a$ is the highest among others $\{\hat{y}_{a'}\}_{a'}$) but is actually quite bad ($y_a$ is quite low), as in Fig. 2(c).

Therefore, we propose minimizing a combination of both of the regression and classification risks, i.e., the geometric mean of them.

$$L^u(f; g) = \sqrt{\widetilde{\mathrm{ER}}^u_g(f) \cdot \mathrm{MSE}^u(f)}. \tag{4}$$

The reason we chose the geometric mean will be explained theoretically in Section 4. Intuitively, it is sufficient to make one of these losses small, e.g., if the classification loss is zero, good decisions can be made even if the MSE is large. As shown in Fig. 2(a), a model that achieves $\mathrm{ER}^u_\mu = 0$ (thus the geometric mean is also zero) can at least outperform the past decision-makers on average ($y_{\hat{a}^*} \geq \bar{y}$) no matter how large the $\mathrm{MSE}^u$ is.

### 3.2 Debiased and Sample-Efficient Learning

While accessible observational data taken from $p(x, a)$ is biased by the propensity $\mu(a|x)$, our target expected risk $L^u(f; g)$ is averaged over all actions uniformly, i.e., $p^u(x, a) = p(x)p^u(a)$, where $p^u(a) = \mathrm{Unif}(\mathcal{A})$ is the discrete uniform distribution. In this section, therefore, we construct two debiasing methods for the sampling bias that performs domain adaptation from $p(x, a)$ to $p^u(x, a)$ as extensions of two existing approaches. Also, we propose network architectures that extract representations from both the feature and the action for better generalization in a large action space.

There are two major approaches for debiased learning in individual-level causal inference. One is a density estimation-based method called inverse probability weighting using propensity score (IPW) (Austin, 2011), in which each instance is weighted by the inverse propensity $1/\mu(a_n|x_n)$. Since the expected risk matches that of the RCT, a good performance can be expected asymptotically under accurate estimation of $\mu$ or when it is recorded as in logged bandit problems. However, in observational studies where the propensity has to be estimated and plugged in, its efficacy would easily decrease (Kang et al., 2007). It becomes further difficult when it comes to a large treatment space. Zou et al. (2020) proposed assuming an intrinsic low-dimensional structure for combinatorial treatment assignments (bundle treatments) $a \in \{0, 1\}^p$ and estimating weights on the latent space. While in this study we examine a general case of large treatment spaces without additional assumptions, it may be necessary to introduce such assumptions to consider such a huge treatment space of combinatorial interventions.

The other approach is representation balancing (Shalit et al., 2017; Johansson et al., 2016; Lopez et al., 2020), in which a representation extractor of the feature $\phi_x$ is trained to eliminate the effect of confounding as well as to preserve the relation to the outcome. Shalit et al. (2017); Johansson et al. (2016) proposed regularizing the conditional probabilities of representations $\{p(\phi_x|a)\}_a$ to be similar to each other by means of the integral probability metric (IPM) regularizer (Müller, 1997; Sriperumbudur et al., 2012) (as in Fig. 3(a)) for limited action spaces such as the binary space $\mathcal{A} = \{0, 1\}$. Lopez et al. (2020) proposed regularizing the representation $\phi_x$ to be independent from the action $a$ by means of the Hilbert-Schmidt Independence Criterion (HSIC) (Gretton et al., 2005, 2008) for real-valued action space $\mathcal{A} \subset \mathbb{R}$. We extend this approach to large treatment spaces.

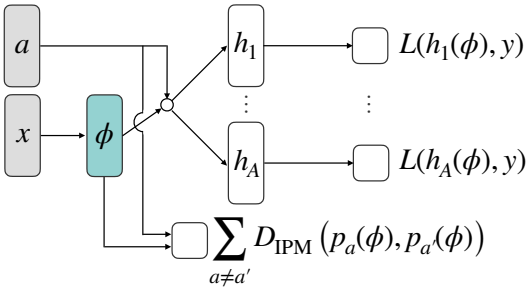
To deal with a large treatment space, we propose performing representation extraction from the treatment $a$ as well as the feature $x$. RMNet-IPM (Fig. 3(b)) extracts the joint representation $\phi_{x,a}$ from $x$ and $a$, which is regularized to be distributionally similar to that of the RCTs $p^u(\phi_{x,a})$. That is, IPM measures the discrepancy between the distributions

$$p(\phi_{x,a}) := \int \sum_{a'} p(\phi_{x,a}|x', a')\mu(a'|x')p(x')\mathrm{d}x',$$
$$p^u(\phi_{x,a}) := \int \sum_{a'} p(\phi_{x,a}|x', a')p^u(a')p(x')\mathrm{d}x',$$

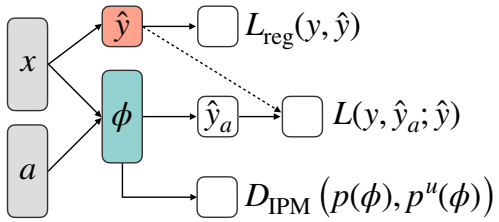
where $p(\phi_{x,a}|x', a') = \delta(\phi_{x,a} - \phi(x', a'))$. IPM is defined for a pair of distributions $(p_1, p_2)$ over $\mathcal{S}$ and a function family $G$ as follows.

$$\mathrm{IPM}_\mathrm{G}(p_1, p_2) = \sup_{g\in\mathrm{G}} \left| \int_\mathcal{S} g(s)\,(p_1(s) - p_2(s))\,\mathrm{d}s \right|.$$

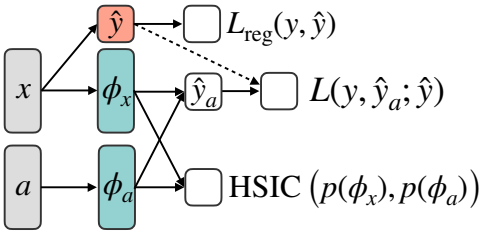
We adopt the set of 1-Lipschitz functions as $G$ (as in (Shalit et al., 2017)), after which IPM is equivalent to the Wasserstein distance. Specifically, we use an entropy relaxation of the exact Wasserstein distance, called Sinkhorn distance (Cuturi, 2013), to ensure the compatibility with the gradient-based optimization. This discrepancy upper-bounds the gap between our target risk (4), which is averaged over the uniform distribution with respect to action $p^u(x, a) = p(x)p^u(a)$, and the one of observational distribution $p(x, a)$. The-




(a) Counterfactual regression (CFR)




(b) RMNet-IPM (proposed)




(c) RMNet-HSIC (proposed)

Figure 3: Network structures of counterfactual regression for CATE (Shalit et al., 2017; Schwab et al., 2018) and our proposed methods. A broken line indicates no backpropagation.

oretical analysis for this point can be found in Appendix B.

Note that minimizing the discrepancy between $p(\phi_{x,a})$ and $p^u(\phi_{x,a})$ and preserving the causal relation are not necessarily incompatible. In this sense, our approach, which directly regularizes the representation distribution $p(\phi_{x,a})$ to be similar to that taken from RCTs $p^u(\phi_{x,a})$, provides a weaker and sufficient condition for this domain adaptation problem. We discuss this point in Appendix C.

RMNet-HSIC (Fig. 3(c)) extracts each representation $\phi_x$ and $\phi_a$ from $x$ and $a$ separately, and they are regularized to be independent from each other by minimizing $\mathrm{HSIC}(p(\phi_x), p(\phi_a))$. HSIC can be defined as a special case of the (squared) maximum mean discrepancy (MMD), which is an instance of the IPM with the class of norm-1 reproducing kernel Hilbert space (RKHS) functions, as follows:

$$\mathrm{HSIC}(p(\phi_x), p(\phi_a)) = \mathrm{MMD}^2(p(\phi_x, \phi_a), p(\phi_x)p(\phi_a)).$$

**Algorithm 1** Regret minimization network

**Input:** Observational data $D = \{(x_n, a_n, y_n)\}_n$, a hyperparameter $\alpha$
**Output:** Trained network parameter $W$

Train $g$ by an arbitrary supervised learning method with $D' = \{(x_n, y_n)\}_n$, e.g.: $g = \arg\min_{g'} \sum (y_n - g'(x_n))^2$.
**if** Method is RMNet-HSIC **then**
  Set weight $\beta_n = 1/\hat{p}(a_n)$ for each instance, where $\hat{p}(a_n)$ is the count $|\{n \in D \mid a = a_n\}|$.
**else**
  Set $\beta_n = 1$ for all $n$.
**end if**
**while** Convergence criteria is not met **do**
  Sample mini-batch $\{n_1, \ldots, n_b\} \subset \{1, \ldots, N\}$.
  Calculate the gradient of the supervised loss $L$ in (5): $g_1 = \nabla_W \frac{1}{b} \sum L(f(x_{n_i}, a_{n_i}; W), y_{n_i}; g(x_{n_i}), \beta_n)$.
  Calculate the gradient of the representation balancing regularizer: $g_2 = \nabla_W D_{\mathrm{bal}}(\{\phi(x_{n_i}, a_{n_i}; W)\})$.
  Obtain step size $\eta$ with an optimizer (e.g., Adam (Kingma and Ba, 2015)).
  $W \leftarrow [W - \eta(g_1 + \alpha g_2)]$.
  Check convergence criterion.
**end while**
**return** $W$

This means the joint distribution is being separated, i.e., $p(\phi_x, \phi_a) = p(\phi_x)p(\phi_a)$, but it does not mean the consistency with the RCTs $p^u(\phi_{x,a}) = p(\phi_x)p^u(\phi_a)$. To compensate $p(\phi_a)$, we weight the loss according to the estimated marginal probability of the actions $\beta = 1/\hat{p}(a)$.

The resulting objective function is

$$\min_f \frac{1}{N} \sum_n L(f(x_n, a_n), y_n; g(x_n), \beta_n) + \alpha \cdot D_{\mathrm{bal}}(\{\phi(x_n, a_n)\}_n) + \mathfrak{R}(f), \tag{5}$$

where $L$ is the empirical instance-wise version of (4), $D_{\mathrm{bal}}$ is the balancing regularizer (IPM or HSIC), and $\mathfrak{R}$ is a regularizer. The resulting learning flow is shown in Algorithm 1.

## 4 RELATION BETWEEN PREDICTION ACCURACY AND DECISION-MAKING PERFORMANCE

In this section, we analyze our decision-focused performance metric. This analysis demonstrates the difficulty of maximizing the decision performance only by

minimizing the regression error when the action space is large. At the same time, however, it is shown that we can further minimize the upper-bound of the regret by minimizing a classification error, which justifies our proposed loss (4) in Section 3.1.

Here we define the decision performance of a model $f$ as the simple average of the potential outcomes for the top-$k$ predicted actions by $f$. We call that performance metric the mean cumulative gain (mCG), and also define its difference from the oracle's performance (regret).

$$\mathrm{mCG}_k(f) := \frac{1}{k}\mathop{\mathbb{E}}_x\left[\sum_{a:\mathrm{rank}(f(x,a))\le k} y_a\right], \tag{6}$$

$$\mathrm{Regret}_k(f) := \frac{1}{k}\mathop{\mathbb{E}}_x\left[\sum_{a:\mathrm{rank}(y_a)\le k} y_a\right] - \mathrm{mCG}_k(f), \tag{7}$$

where $\mathrm{rank}(\cdot)$ is the rank among all the feasible actions, e.g., $\mathrm{rank}(f(x,a);\{f(x,a')\}_{a'}) := |\{a' \mid f(x,a') \ge f(x,a), a' \in \mathcal{A}\}|$. Here, $(1-\mathrm{mCG}_{k=1}(f))$ is known as the policy risk (Shalit et al., 2017). Since the first term in (7) is constant with respect to $f$, the mCG and the regret are two sides of the same coin as the performance metrics of a model.

The relation between the regret and the regression and classification accuracies is the following (full proof and analysis on the tightness can be found in Appendix A).

**Proposition 4.1.** *The regret in* (7) *will be bounded with uniform MSE in* (1) *as*

$$\mathrm{Regret}_k(f) \le \frac{|\mathcal{A}|}{k}\sqrt{\mathrm{ER}^u_k(f)\cdot\mathrm{MSE}^u(f)}, \tag{8}$$

*where* $\mathrm{ER}^u_k(f)$ *is the top-k classification error rate, i.e.,*

$$\mathrm{ER}^u_k(f) := \mathop{\mathbb{E}}_x\left[\frac{1}{|\mathcal{A}|}\sum_{a\in\mathcal{A}} I\left((\mathrm{rank}(y_a)\le k)\oplus(\mathrm{rank}(f(x,a))\le k)\right)\right].$$

*Proof Sketch.* Let $s(x,a) := I(\mathrm{rank}(y_a)\le k) - I(\mathrm{rank}(f(x,a))\le k)$ denote the classification error. Then, we have

$$\begin{aligned}
&k\cdot\mathrm{Regret}_k(f)\\
&= |\mathcal{A}|\mathop{\mathbb{E}}_{x,a\sim p^u(x,a)}[s(x,a)y_a]\\
&\le |\mathcal{A}|\mathop{\mathbb{E}}_{x,a\sim p^u(x,a)}[s(x,a)(y_a - f(x,a))] \qquad (9)\\
&\le |\mathcal{A}|\sqrt{\mathop{\mathbb{E}}_{x,a\sim p^u(x,a)}[s(x,a)^2]\mathop{\mathbb{E}}_{x,a\sim p^u(x,a)}[(y_a-f(x,a))^2]} \qquad (10)\\
&= |\mathcal{A}|\sqrt{\mathrm{ER}^u_k(f)\cdot\mathrm{MSE}^u(f)}.
\end{aligned}$$

Equation (9) is from the definition of $s(x,a)$ and (10) is from the Cauchy-Schwarz inequality. By dividing both sides by $k$, we get the proposition. □

Since $\mathrm{ER}^u_k(f) \le 1$ for any $f$, we see that only minimizing the uniform MSE as in existing causal inference methods leads to minimizing the regret. However, if $|\mathcal{A}|/k$ is large, the bound would be loose, and only unrealistically small $\mathrm{MSE}^u$ provide a meaningful guarantee for the regret.

At the same time, we see that the bound can be further improved by minimizing the uniform top-$k$ classification error rate $\mathrm{ER}^u_k(f)$ simultaneously, which leads to our proposed method. Let $k'$ be the past decision-makers' average performance, i.e., $y_{a^*_{k'+1}} \le \mathbb{E}_{a\sim\mu(a|x)}[y_a|x] \le y_{a^*_{k'}}$. Then, the proposed method can be interpreted as minimizing the upper-bound of $\mathrm{Regret}_{k'}$. While training a model for a particular $k$ is an interesting direction, the proposed method is not so sensitive to the difference between the decision-making performance of the data $k'$ and the actual $k$ to be evaluated, as we will see in Section 5. Another interesting direction is optimizing $k$ or the decision-making policy. The $\mathrm{mCG}_k$ can be interpreted as the expected performance (reward) of the following plug-in policy that takes an action uniformly at random from the predicted top-$k$ actions.

$$\pi^f_k(a|x) := \begin{cases} 1/k & \text{if } \mathrm{rank}(f(x,a);\{f(x,a')\}_{a'}) \le k \\ 0 & \text{otherwise,}\end{cases}$$

Therefore, choosing $k$ means choosing a policy. If we choose $k$ greater than 1, the oracle's performance (the first term in (7)) would be smaller, but the upper bound of the regret (8) would be larger. Thus there may exist an optimal $k > 1$ that maximizes the overall performance of the decision-making.

# 5 EXPERIMENTS

We investigated the effectiveness of our method through synthetic and semi-synthetic experiments. We newly designed both datasets for the problem setting with a large action space.

## 5.1 Experimental Setup

**Compared Methods** We compared our proposed method (RMNet) with ridge linear regression (OLS), random forests (Breiman, 2001) (RF), k-nearest neighbor (kNN), Bayesian additive regression trees (BART) (Hill, 2011), naive deep neural network (S-DNN), naive DNN with multi-head architecture for each action (M-DNN) (a.k.a. TARNET (Shalit et al., 2017)), RankNet (Burges et al., 2005), and a straightforward

extension of the existing action-wise representation-balancing method (counterfactual regression network (CFRNet)) (Shalit et al., 2017). We also made an ablation study to clarify the contributions of each component. The strength of representation-balancing regularizer $\alpha$ in CFRNet and the proposed method was selected from $[0.1, 0.3, 1.0, 3.0, 10.0]$. Other specifications of the DNN parameters can be found in Appendix D.

**Evaluation** We used the normalized mean gain (NMG) as the main metric, defined as follows.

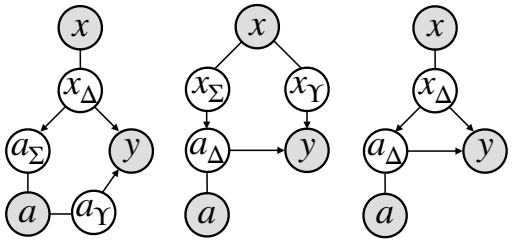
$$\mathrm{NMG} := \sum_x y_{\hat{a}^*}(x) \Big/ \sum_x y_{a^*}(x),$$

where $\hat{a}^*$ and $a^*$ are the predicted and true best actions for each $x$, respectively. The NMG is proportional to the mean CG ($k = 1$) (6). We can see NMG $\leq 1$. Since we have standardized the outcome, the chance rate is NMG $= 0$. In addition to NMG, we have also evaluated with respect to $\mathrm{MSE}^u$ and $\mathrm{ER}^u_{k=1}$. The validation and the model selection were based on the NMG. For those cases where the complete validation dataset to compute NMG is not accessible, an alternative validation strategy needs to be considered, e.g., imputing missing values by 1-NN or BART (as in (Hassanpour and Greiner, 2019)) or constructing a special method (such as the counterfactual cross-validation in (Saito and Yasui, 2020)).

**Infrastructure** All the experiments were run on a machine with 28 CPUs (Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz), 250GB memory, and 8 GPUs.

### 5.2 Synthetic Experiment

**Dataset** We prepared four biased datasets with sampling bias in total to examine the robustness of the proposed and baseline methods. For a detailed description of the generation process, see Appendix D. The feature space and the action space are fixed to $\mathbb{R}^5$ and $\{0, 1\}^5$, respectively. The true causal models are set as follows. Three settings (called Quadratic) have a relation $y_a(x) = a_\Upsilon^2 - 2x_\Upsilon + \varepsilon$, where $a_\Upsilon = w_a^\top a$ and $x_\Upsilon = w_x^\top x$ are the one-dimensional representations of $a$ and $x$, respectively, and where $w_a, w_x \sim N(0, 1/5)^5$. The last setting (called Bilinear) has a bilinear relation $y = x^\top W a + \varepsilon$, where $W \sim N(0, 1/25)^{5\times 5}$. For training, only one action and the corresponding outcome for each $x$ are sampled as $p(a|x) \propto \exp(10\,|x_\Sigma - a_\Sigma|)$, where $x_\Sigma$ and $a_\Sigma$ are additional representations of $x$ and $a$. The three settings for the quadratic patterns correspond to the relation between $\cdot_\Sigma$ and $\cdot_\Upsilon$ as illustrated in Fig. 4(a)–(c), i.e., $x_\Sigma = x_\Upsilon$ (=: $x_\Delta$) in Setups A and C, and $a_\Sigma = a_\Upsilon$ (=: $a_\Delta$) in Setups B and C. These relations of variables were designed to reproduce spurious correlations, which mislead the decision-making as follows. In Setup A, $a_\Sigma$ would have dependence on $y$ through its dependence on $x_\Delta$ despite $a_\Sigma$ itself having no causal relation to $y$. In the same manner, in Setup B, $x_\Sigma$ would have dependence on $y$ through $a_\Delta$, and the causal effect of $a_\Delta$ may appear discounted. Setup C has both effects. The sample sizes for $x$ were 1,000 for training, 100 for validation, and 200 for testing.




(a) Setup A (b) Setup B (c) Setup C

Figure 4: Data generation models for synthetic experiment. Shaded variables denote the accessible variables in training. Non-shaded variables are latent one-dimensional representations of $x$ and $a$.

Table 1: Synthetic results on NMG (larger is better and the maximum is one) and its standard error in ten data generations. Best and second-best methods are in bold.

| Method | Quadratic-A | Quadratic-B | Quadratic-C | Bilinear |
|---|---|---|---|---|
| OLS | 0.35 ± 0.13 | 0.74 ± 0.10 | 0.73 ± 0.12 | 0.02 ± 0.02 |
| RF | 0.71 ± 0.08 | 0.24 ± 0.02 | 0.91 ± 0.04 | 0.67 ± 0.03 |
| kNN | 0.58 ± 0.05 | 0.33 ± 0.04 | 0.53 ± 0.07 | 0.59 ± 0.03 |
| BART | 0.53 ± 0.12 | **0.91 ± 0.05** | **0.99 ± 0.00** | 0.14 ± 0.07 |
| M-DNN | 0.46 ± 0.09 | 0.42 ± 0.12 | 0.57 ± 0.12 | −0.01 ± 0.04 |
| S-DNN | 0.63 ± 0.08 | 0.43 ± 0.07 | 0.60 ± 0.08 | 0.58 ± 0.09 |
| CFRNet | 0.46 ± 0.08 | 0.43 ± 0.12 | 0.63 ± 0.13 | −0.01 ± 0.04 |
| RankNet | 0.62 ± 0.09 | 0.70 ± 0.05 | 0.68 ± 0.08 | **0.74 ± 0.04** |
| RMNet-IPM | **0.86 ± 0.04** | 0.84 ± 0.03 | 0.82 ± 0.05 | **0.77 ± 0.04** |
| RMNet-HSIC | **0.90 ± 0.02** | **0.88 ± 0.05** | **0.86 ± 0.07** | 0.14 ± 0.03 |

**Results** The results listed in Table 1 show that our proposed method achieved the best or comparable performance under all settings, while the other methods varied in performance across settings. We analyze the reason of the poor performance of RMNet-HSIC in Bilinear in the ablation study in Section 5.4.

### 5.3 Semi-Synthetic Experiment

**Dataset (GPU Kernel Performance)** For the semi-synthetic experiment, we used the SGEMM GPU kernel performance dataset (Nugteren and Codreanu, 2015; Ballester-Ripoll et al., 2019), which has 14 feature attributes of GPU kernel parameters and four target attributes of elapsed times in milliseconds for four independent runs of each combination of parameters.

Table 2: Semi-synthetic results on NMG with the standard error in ten different samplings of the training data. The $\mathrm{MSE}^u$ and $\mathrm{ER}^u_{k=1}$ are also shown. Best and second-best methods are in bold.

| | Normalized mean gain | | | | $\mathrm{MSE}^u$ | | | | $\mathrm{ER}^u_{k=1}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lvert\mathcal{A}\rvert$ / Method | 8 | 16 | 32 | 64 | 8 | 16 | 32 | 64 | 8 | 16 | 32 | 64 |
| OLS | −0.04 ± 0.15 | −0.08 ± 0.20 | −0.10 ± 0.13 | −0.01 ± 0.10 | 1.12 | 1.89 | 1.70 | 5.86 | 0.221 | 0.116 | 0.061 | 0.031 |
| RF | 0.24 ± 0.08 | 0.33 ± 0.07 | 0.33 ± 0.05 | 0.38 ± 0.05 | 1.03 | 0.87 | 0.93 | 1.07 | 0.214 | 0.114 | 0.059 | 0.030 |
| kNN | 0.35 ± 0.04 | 0.39 ± 0.04 | 0.33 ± 0.04 | 0.39 ± 0.02 | **0.59** | **0.64** | **0.64** | 0.63 | 0.211 | 0.113 | 0.059 | 0.030 |
| BART | −0.05 ± 0.13 | 0.13 ± 0.13 | 0.13 ± 0.10 | 0.04 ± 0.09 | 1.06 | 1.05 | 1.15 | 1.63 | 0.222 | 0.116 | 0.060 | 0.031 |
| M-DNN | 0.40 ± 0.05 | 0.48 ± 0.06 | 0.30 ± 0.07 | 0.37 ± 0.05 | 0.78 | 0.83 | 0.82 | 0.84 | 0.211 | 0.113 | 0.059 | 0.030 |
| S-DNN | 0.28 ± 0.09 | 0.25 ± 0.10 | 0.32 ± 0.07 | 0.45 ± 0.05 | 0.75 | **0.64** | 0.74 | **0.74** | 0.212 | 0.114 | 0.059 | 0.029 |
| CFRNet | 0.50 ± 0.06 | 0.39 ± 0.14 | 0.39 ± 0.10 | 0.35 ± 0.05 | 0.78 | 0.80 | 0.87 | 0.86 | 0.210 | 0.113 | 0.058 | 0.030 |
| RankNet | 0.35 ± 0.07 | 0.29 ± 0.09 | 0.38 ± 0.06 | 0.45 ± 0.05 | 6.08 | 10.13 | 8.47 | 2.42 | 0.210 | 0.113 | 0.058 | 0.029 |
| RMNet-IPM | **0.68** ± 0.01 | **0.61** ± 0.05 | **0.61** ± 0.04 | **0.51** ± 0.06 | 0.76 | 0.81 | 0.85 | 0.75 | **0.204** | **0.109** | **0.055** | **0.029** |
| RMNet-HSIC | **0.59** ± 0.04 | **0.57** ± 0.06 | **0.55** ± 0.06 | **0.69** ± 0.06 | **0.48** | 0.66 | **0.61** | **0.39** | **0.207** | **0.109** | **0.056** | **0.028** |

We used the inverse of the mean elapsed times as the outcome, resulting in 241.6k instances in total. By treating some of the feature attributes as action dimensions, we obtained a *complete* dataset, which has all the entries (potential outcomes) in Fig. 1 observed. Then we composed our semi-synthetic dataset by biased subsampling of only one action $a$ and the corresponding potential outcome $y_a$ for each $x$. The details of this preprocess can be found in Appendix D.

The sampling policy in the training data was $p(a|x,y) \propto \exp(-10|y-[x^\top, a^\top]^\top w|)$, where $w$ is sampled from $\mathcal{N}(0,1)^{d+m}$. This policy reproduces a spurious correlation; that is, a random projection of the feature and the action $[x^\top, a^\top]^\top w$ is likely to have little causal relation with $y$ but does have a strong correlation due to the sampling policy. This policy also depends on $y$, which violates the unconfoundedness assumption. However, the dataset we used has a low noise level, i.e., $y \simeq g(x,a)$ for some function $g$, and thus the violation is limited, i.e., $p(a|x,y) \simeq p(a|x,g(x))$.

We split the feature set $\{x_n\}_n$ into 80% for training, 5% for validation, and 15% for testing. Then, for the training set, only one action $a$ and the corresponding outcome $y$ was taken for each $x$. The resulting training sample size for each setting of $m$ is listed in Table 5 in Appendix D. We repeated the training and evaluation process ten times for different splits and samplings of $a$.

**Results** The results listed in Table 2 show that our proposed methods outperformed the others in NMG in all cases. The decision performance (NMG) was more consistent with ER than MSE, indicating that ER as well as MSE needs to be considered. The performance of multi-head DNNs (M-DNN and CFRNet) decreased in larger action spaces, while single-head DNNs (S-DNN and the proposed methods) maintained their performance. This demonstrates the importance of sample efficiency by extracting the representation of both the feature and the action.

### 5.4 Ablation Study

We examined the effect of each component of the proposed method, i.e., the balancing regularizer ($D_{\mathrm{bal}}$), each component of the risk (MSE and ER), and the representation extraction from the action ($\phi_a$) and the reweighting with respect to the marginal distribution of the action ($\beta$) for RMNet-HSIC. Table 3 shows the results.

The effectiveness of $D_{\mathrm{bal}}$ was verified in the setting of $|\mathcal{A}| = 32$. Also, the effectiveness of ER was significant in the Bilinear setting. Extracting representation from the action ($\phi_a$) was quite effective in Semi-synthetic settings. The reweighting ($\beta$) was also effective in the Semi-synthetic settings, while it decreased the performance in the Bilinear setting. A possible reason is the estimation variance induced by plugging the estimated marginal distribution of the action $\hat{p}(a)$ into weights as its inverse, which is the same issue as the inverse propensity score weighting approach.

## 6 SUMMARY

In this paper, we have investigated causal inference on a large action space with a focus on the decision-making performance. We analyzed the decision-making performance brought about by a model through a simple prediction-based decision-making policy. We showed that the bound with only the regression accuracy (MSE) gets looser as the action space gets large, which demonstrates the difficulty of utilizing causal inference in decision-making in a large action space. At the same time, however, our bound indicates that minimizing not only the regression loss but also the classification loss leads to a better performance. From this viewpoint, our proposed methods

Table 3: Ablation study of the proposed methods (indicated by (†)) on semi-synthetic dataset. $D_{\text{bal}}$ indicates the type of balancing regularizer. MSE and ER are the used loss. $\phi_a$ indicates whether or not the representation is also extracted from the action, i.e., if $\phi_a$ is not checked, identity function is used for $\phi_a$ (i.e., $\phi_a = a$). $\beta$ indicates the reweighting with $1/\hat{p}(a)$, which is needed only in the HSIC-based methods (as explained in Section 3.2). Best and second-best methods are in bold.

| | | | | | Normalized mean gain | | |
|---|---|---|---|---|---|---|---|
| | | | | | Synthetic | Semi-synthetic | |
| $D_{\text{bal}}$ | MSE | ER | $\phi_a$ | $\beta$ | Bilinear | $\lvert\mathcal{A}\rvert = 32$ | $\lvert\mathcal{A}\rvert = 64$ |
| † IPM | ✓ | ✓ | ✓ | — | **0.77 ± 0.04** | **0.61 ± 0.04** | 0.51 ± 0.06 |
| IPM | | ✓ | ✓ | — | **0.73 ± 0.03** | **0.61 ± 0.05** | 0.58 ± 0.05 |
| IPM | ✓ | | ✓ | — | 0.55 ± 0.10 | 0.55 ± 0.05 | 0.49 ± 0.05 |
| None | ✓ | ✓ | ✓ | | 0.72 ± 0.03 | 0.39 ± 0.07 | 0.49 ± 0.06 |
| †HSIC | ✓ | ✓ | ✓ | ✓ | 0.14 ± 0.03 | 0.55 ± 0.06 | **0.69 ± 0.06** |
| HSIC | | ✓ | ✓ | ✓ | 0.11 ± 0.02 | 0.56 ± 0.07 | **0.72 ± 0.02** |
| HSIC | ✓ | | ✓ | ✓ | 0.16 ± 0.05 | 0.59 ± 0.05 | 0.68 ± 0.06 |
| HSIC | ✓ | ✓ | | ✓ | 0.04 ± 0.03 | 0.31 ± 0.08 | 0.23 ± 0.09 |
| HSIC | ✓ | ✓ | ✓ | | 0.51 ± 0.07 | 0.38 ± 0.07 | 0.49 ± 0.06 |
| HSIC | ✓ | ✓ | | | 0.63 ± 0.05 | 0.29 ± 0.07 | 0.22 ± 0.09 |

minimize both the MSE and the classification loss of whether or not the outcome is better than the average performance of the past decision-makers. Specifically, we adopt the cross-entropy with a teacher label indicating whether an observed outcome is better than the estimated average decision performance of the past decision-makers under a given feature. For the sample efficiency in a large treatment space, we proposed extracting representations from both the feature and the action. To generalize in the distribution of RCTs, we proposed two balancing regularizers that encourage the representation distribution to be similar to that of RCTs as extensions of existing approaches. Experiments on synthetic and semi-synthetic datasets, which were designed to have misleading spurious correlations, demonstrated the superior performance of the proposed methods with respect to the decision performance.

## Acknowledgements

TT was partially supported by JSPS KAKENHI Grant Numbers 20K03753 and 19H04071. HK was supported by the JSPS KAKENHI Grant Number 20H04244.

## References

P. C. Austin. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3):399–424, 2011.

R. Ballester-Ripoll, E. G. Paredes, and R. Pajarola. Sobol tensor trains for global sensitivity analysis. *Reliability Engineering & System Safety*, 183:311–322, 2019.

L. Breiman. Random forests. *Machine learning*, 45(1): 5–32, 2001.

C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*, pages 89–96, 2005.

M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.

A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer, 2005.

A. Gretton, K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, and A. J. Smola. A kernel statistical test of independence. In *Advances in neural information processing systems*, pages 585–592, 2008.

N. Hassanpour and R. Greiner. Counterfactual regression with importance sampling weights. In *IJCAI*, pages 5880–5887, 2019.

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

G. W. Imbens and J. M. Wooldridge. Recent developments in the econometrics of program evaluation. *Journal of economic literature*, 47(1):5–86, 2009.

T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142, 2002.

F. Johansson, U. Shalit, and D. Sontag. Learning representations for counterfactual inference. In *International conference on machine learning*, pages 3020–3029, 2016.

F. D. Johansson, D. Sontag, and R. Ranganath. Support and invertibility in domain-invariant representations. In K. Chaudhuri and M. Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 527–536. PMLR, 16–18 Apr 2019.

J. D. Kang, J. L. Schafer, et al. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science*, 22(4):523–539, 2007.

D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. In *ICLR 2015 : International Conference on Learning Representations 2015*, 2015.

R. Lopez, C. Li, X. Yan, J. Xiong, M. Jordan, Y. Qi, and L. Song. Cost-effective incentive allocation via structured counterfactual inference. In *AAAI 2020 : The Thirty-Fourth AAAI Conference on Artificial Intelligence*, volume 34, pages 4997–5004, 2020.

A. Müller. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, pages 429–443, 1997.

C. Nugteren and V. Codreanu. Cltune: A generic auto-tuner for opencl kernels. In *Embedded Multicore/Many-core Systems-on-Chip (MCSoC), 2015 IEEE 9th International Symposium on*, pages 195–202. IEEE, 2015.

J. Pearl. *Causality*. Cambridge university press, 2009.

S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 452–461, 2009.

Y. Saito and S. Yasui. Counterfactual cross-validation: Stable model selection procedure for causal inference models. In *ICML 2020: 37th International Conference on Machine Learning*, 2020.

P. Schwab, L. Linhardt, and W. Karlen. Perfect match: A simple method for learning representations for counterfactual inference with neural networks. *arXiv preprint arXiv:1810.00656*, 2018.

U. Shalit, F. D. Johansson, and D. Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3076–3085. JMLR. org, 2017.

H. A. Simon. Spurious correlation: A causal interpretation. *Journal of the American statistical Association*, 49(267):467–479, 1954.

B. K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, G. R. Lanckriet, et al. On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics*, 6:1550–1599, 2012.

J. Yoon, J. Jordon, and M. van der Schaar. GANITE: Estimation of individualized treatment effects using generative adversarial nets. In *International Conference on Learning Representations*, 2018.

Y. Zhang, A. Bellot, and M. van der Schaar. Learning overlapping representations for the estimation of individualized treatment effects. In *AISTATS*, pages 1005–1014, 2020.

H. Zhao, R. T. D. Combes, K. Zhang, and G. Gordon. On learning invariant representations for domain adaptation. volume 97 of *Proceedings of Machine Learning Research*, pages 7523–7532, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

H. Zou, P. Cui, B. Li, Z. Shen, J. Ma, H. Yang, and Y. He. Counterfactual prediction for bundle treatment. *Advances in Neural Information Processing Systems*, 33, 2020.

# A Proof of Proposition 4.1

**Proposition A.1.** *The expected regret will be bounded with uniform MSE in (1) as*

$$\mathrm{Regret}_k(f) \leq \frac{|\mathcal{A}|}{k}\sqrt{\mathrm{ER}_k^u(f)\cdot \mathrm{MSE}^u(f)},$$

*where* $\mathrm{ER}_k^u(f)$ *is the top-k classification error rate, i.e.,*

$$\mathrm{ER}_k^u(f) := \mathop{\mathbb{E}}_x\left[\frac{1}{|\mathcal{A}|}\sum_{a\in\mathcal{A}} I\left((\mathrm{rank}(y_a)\leq k)\oplus(\mathrm{rank}(f(x,a))\leq k)\right)\right],$$

*where* $\oplus$ *denotes the logical XOR.*

*Proof.* We denote the true and the predicted $i$-th best action by $a_i^*$ and $\hat{a}_i^*$, respectively; i.e., $\mathrm{rank}(y_{a_i^*}) = \mathrm{rank}(f(x,\hat{a}_i^*)) = i$. For all $k \in [|\mathcal{A}|]$, the target-wise regret can be bounded as follows:

$$\begin{aligned}
&k\cdot\mathrm{Regret}_k(x)\\
&:= \sum_{i\leq k}\left(y_{a_i^*} - y_{\hat{a}_i^*}\right)\\
&\leq \sum_{i\leq k}\left(y_{a_i^*} - y_{\hat{a}_i^*}\right) + \sum_{i\leq k}\left(f_{\hat{a}_i^*} - f_{a_i^*}\right) \qquad (11)\\
&= \sum_{i\leq k}\left\{\left(y_{a_i^*} - f_{a_i^*}\right) - \left(y_{\hat{a}_i^*} - f_{\hat{a}_i^*}\right)\right\}\\
&= \sum_a \big\{(I(\mathrm{rank}(y_a)\leq k) - I(\mathrm{rank}(f_a)\leq k))\\
&\qquad\cdot(y_a - f_a)\big\},
\end{aligned}$$

where $f_a = f(x,a)$. Inequality (11) is from the definition of $\hat{a}_i^*$; i.e., $\sum_{i\leq k} f_{\hat{a}_i^*}$ is the summation of the top-$k$ $f_a$s out of $\{f_a\}_{a\in\mathcal{A}}$, which must be larger than or equal to the summation of $k$ $f_a$s that are not necessarily top-$k$, $\sum_{i\leq k} f_{a_i^*}$. Let us define a classification error $s$ and the regression error $e$ as

$$\begin{aligned}
s(x,a) &:= I(\mathrm{rank}(y_a)\leq k) - I(\mathrm{rank}(f_a)\leq k),\\
e(x,a) &:= y_a - f_a.
\end{aligned}$$

The r.h.s. is written as

$$\text{r.h.s.} = \sum_a s(x,a)e(x,a).$$

By taking the expectation with respect to $x$, we have

$$\begin{aligned}
&k\cdot\mathrm{Regret}_k(f)\\
&= \mathop{\mathbb{E}}_x\left[\sum_a s(x,a)e(x,a)\right]\\
&= |\mathcal{A}|\mathop{\mathbb{E}}_{(x,a)\sim p^u(x,a)}[s(x,a)e(x,a)]\\
&\leq |\mathcal{A}|\sqrt{\mathop{\mathbb{E}}_{(x,a)\sim p^u(x,a)}[|s(x,a)|^2]\cdot\mathop{\mathbb{E}}_{(x,a)\sim p^u(x,a)}[|e(x,a)|^2]} \qquad (12)\\
&= \left\{\mathop{\mathbb{E}}_x\left[\sum_a I(\mathrm{rank}(y_a)\leq k\oplus\mathrm{rank}(f(x,a))\leq k)\right]\right.\qquad (13)\\
&\qquad\left.\cdot\mathop{\mathbb{E}}_x\left[\sum_a (y_a - f(x,a))^2\right]\right\}^{1/2}\\
&= |\mathcal{A}|\sqrt{\mathrm{ER}_k^u(f)\cdot\mathrm{MSE}_k^u(f)},
\end{aligned}$$

where the inequality (12) comes from the Cauchy–Schwarz inequality and the equality (13) comes from the definitions of $s$ and $e$. By dividing both sides by $k$, we get the proposition. □

Note that our bound cannot be improved without additional assumptions on the true and assumed model classes of the causal mechanism $f(x,a)$ (and thus the true potential outcomes and its predictions). For any $|\mathcal{A}|$, $k \leq |\mathcal{A}|/2$, $\mathrm{ER}_k^u$, $\mathrm{MSE}^u$, and $\epsilon > 0$, there exist a joint distribution of potential outcomes and $x$, and a model $f$ that have the gap (the ratio) between both sides of the proposition is $(1+\epsilon)$.

Let us define a prototype of a potential outcome vector $\boldsymbol{y}^\kappa$ as

$$\begin{aligned}
\boldsymbol{y}^\kappa :=& (y_1,\ldots,y_{|\mathcal{A}|})\\
=& (\underbrace{1,\ldots,1}_{\kappa},\underbrace{-1,\ldots,-1}_{\kappa},0,\ldots,0),
\end{aligned}$$

that is, the first $\kappa$ dimensions are 1, the following $\kappa$ dimensions are $-1$, and the rest are 0. When the true outcome is $\boldsymbol{y} = t\boldsymbol{y}^\kappa$ for $t > 0$ and $\kappa \leq k$, and when the prediction of the model is bad (misleading) as $\hat{y} = -\epsilon\boldsymbol{y}$, the components of the r.h.s. would be

$$\begin{aligned}
\mathrm{MSE}^u &= 2\kappa t^2(1+\epsilon)^2/|\mathcal{A}|,\\
\mathrm{ER}_k^u &= 2\kappa/|\mathcal{A}|,
\end{aligned}$$

and thus the r.h.s. would be

$$\text{r.h.s.} = 2t\kappa(1+\epsilon)/k,$$

while the l.h.s. would be

$$\mathrm{Regret}_k = 2t\kappa/k.$$

The gap (the ratio) between them is $(1+\epsilon)$ for any $\epsilon$. Since we have two free parameters $\kappa$ and $t$, any $\mathrm{MSE}^u$ and $\mathrm{ER}^u_k$ can be (almost) achieved. At this point, the constraint $\kappa \in \mathbb{N}$ also causes a constraint on ER, but it can be removed ($\kappa := |\mathcal{A}|\mathrm{ER}^u_k/2$ for any $\mathrm{ER}^u_k$) as follows. We consider a domain partition $\mathcal{X}_1 \in \mathcal{X}$ and the potential outcomes as

$$p(\boldsymbol{y} = t\boldsymbol{y}^{\kappa_1}|x) = 1 \quad (x \in \mathcal{X}_1 \subset \mathcal{X}),$$
$$p(\boldsymbol{y} = t\boldsymbol{y}^{\kappa_2}|x) = 1 \quad (x \in \mathcal{X} \backslash \mathcal{X}_1),$$

where $\kappa_1 := \lfloor |\mathcal{A}|\mathrm{ER}^u_k/2 \rfloor$, $\kappa_2 := \lceil |\mathcal{A}|\mathrm{ER}^u_k/2 \rceil$, and the partition $\mathcal{X}_1$ can be determined to satisfy $\mathbb{E}_x\left[\mathrm{ER}^u_k(f,x)\right] = 2\kappa/|\mathcal{A}|$. Thus, for any $|\mathcal{A}|$, $k$, $\mathrm{MSE}^u$, and $\mathrm{ER}^u_k$, the bound cannot be improved without any assumption.

Our bound means that, when $\kappa = k \ll |\mathcal{A}|$ holds, despite this prediction $\hat{\boldsymbol{y}}$ being quite "accurate" in terms of $\mathrm{MSE}^u$, the decision is constantly misleading regardless of $|\mathcal{A}|/k$ (and thus $\mathrm{MSE}^u$). It could be improved when the spaces of $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ are limited and well-specified, but such specification of the model class is another big issue in real-world applications. We therefore conclude that minimizing only the regression accuracy $\mathrm{MSE}^u$ is insufficient in terms of decision performance when the treatment space is large, and minimizing the classification accuracy $\mathrm{ER}^u_k$ is also important.

## B Error analysis for representation-based domain adaptation from observational data to the uniform average on action space

By performing the representation-balancing regularization, our method enjoys better generalization through minimizing the upper bound of the error on the test distribution (under uniform random policy). We briefly show how minimizing the combination of empirical loss on training and the regularization of distribution (5) results in minimizing the test error. First, we define the point-wise loss function under a hypothesis $h$ and an invertible extractor $\phi(\cdot,\cdot)$, which defines the representation $\phi = \phi(x,a)$ with its inverse $(x,a) = \psi(\phi)$, as

$$\ell_h^{x,a}(\phi) := \int_{\mathcal{Y}} L(Y_a, h(\phi)) p(Y_a|x) \mathrm{d}Y_a.$$

Then, the expected losses for the training (source) and the test distribution (target) are

$$\epsilon^s(h) := \int_{\mathcal{X},\Phi} \sum_{a\in\mathcal{A}} \ell_h^{x,a}(\phi) p(\phi|x,a) p(x,a) \mathrm{d}\phi \mathrm{d}x,$$
$$\epsilon^t(h) := \int_{\mathcal{X},\Phi} \sum_{a\in\mathcal{A}} \ell_h^{x,a}(\phi) p(\phi|x,a) p^u(a|x) p(x) \mathrm{d}\phi \mathrm{d}x,$$

where $p(\phi|x,a) = \delta(\phi - \phi(x,a))$. We assume there exists $B > 0$ such that $\frac{1}{B}\ell_h^{x,a}(\phi) \in G$ for the given function space $G$. Then the integral probability metric $\mathrm{IPM}_G$ is defined for $\phi \in \Phi = \{\phi(x,a)|p(x,a) > 0\}$ as

$$\mathrm{IPM}_G(p_1, p_2) := \sup_{g\in G} \left| \int_\Phi g(\phi)(p_1(\phi) - p_2(\phi)) \mathrm{d}\phi \right|.$$

The difference between the expected losses under training and test distributions are then bounded as

$$\begin{aligned}
&\epsilon^t(h) - \epsilon^s(h) \\
&= \int_\Phi \ell_h^{\psi(\phi)}(\phi) \left(p^u(\phi) - p(\phi)\right) \mathrm{d}\phi \\
&= B \int_\Phi \frac{1}{B} \ell_h^{\psi(\phi)}(\phi) \left(p^u(\phi) - p(\phi)\right) \mathrm{d}\phi \\
&\leq B \sup_{g\in G} \left| \int_\Phi g(\phi) \left(p^u(\phi) - p(\phi)\right) \mathrm{d}\phi \right| \\
&= B \cdot \mathrm{IPM}_G\left(p(\phi), p^u(\phi)\right).
\end{aligned}$$

Although $B$ is unknown, the hyperparameter tuning of the regularization strength $\alpha$ in (5) can achieve the tuning of $B$.

## C Minimizing IPM while preserving the causal relation

We show that minimizing the discrepancy between $p(\phi_{x,a})$ and $p^u(\phi_{x,a})$ and preserving the causal relation are not necessarily in conflict with each other.

Let us consider an example of $p(x,a)$ shown in Table 4 and a representation $\phi(x,a) = x + a$. Then, for any $\epsilon \in (-1/9, 1/9)$, the representation distribution is calculated as, e.g., $p(\phi_{x,a} = 1) = p(x = 1, a = 0) + p(x = 0, a = 1) = 2/9$. In the same manner, we have

$$\begin{aligned}
&p(\phi_{x,a} = 0) = p(\phi_{x,a} = 4) = 1/9, \\
&p(\phi_{x,a} = 1) = p(\phi_{x,a} = 3) = 2/9, \\
&p(\phi_{x,a} = 2) = 1/3.
\end{aligned} \tag{14}$$

Also, the uniform target distribution is calculated as $p^u(x,a) := p(x)p^u(a) = 1/9$ for all $x, a \in \{0,1,2\}$, and the representation distribution is the same as (14), meaning $\mathrm{IPM}(p(\phi_{x,a}), p^u(\phi_{x,a})) = 0$. If the true causal relation can be written via $\phi_{x,a}$ as $y = h(\phi(x,a))$ for some $h$, then this representation extractor can achieve

Table 4: Example observational distribution $p(x, a)$ and its marginal distributions $p(x)$ and $p(a)$.

| $\mu(a\|x)p(x)$ | $a=0$ | $a=1$ | $a=2$ | $p(x)$ |
|---|---|---|---|---|
| $x=0$ | $1/9$ | $1/9+\epsilon$ | $1/9-\epsilon$ | $1/3$ |
| $x=1$ | $1/9-\epsilon$ | $1/9$ | $1/9+\epsilon$ | $1/3$ |
| $x=2$ | $1/9+\epsilon$ | $1/9-\epsilon$ | $1/9$ | $1/3$ |
| $p(a)$ | $1/3$ | $1/3$ | $1/3$ | |

IPM $= 0$ while preserving the causal relation, and thus it can still achieve $L^u = 0$.

On the other hand, the action-wise representation extraction approach (e.g., CFRNet (Shalit et al., 2017) in Fig. 3(a)) cannot achieve both the extraction of fully balanced representation $\sum D_{\mathrm{IPM}} = 0$ and the preservation of the relation in this case with $\epsilon \neq 0$. Only constant representation $\phi(x) = c$ for all $x \in \{0, 1, 2\}$ can achieve $\sum_{a,a' \in \mathcal{A}} D_{\mathrm{IPM}}(p(\phi_x|a), p(\phi_x|a')) = 0$, and then the true relation $y = h(x + a)$ is not expressible.

When the action-wise representation achieves IPM $= 0$, our representation $\phi(x, a)$ can also achieve IPM $= 0$ under an assumption that the marginal action distribution is uniform, i.e., $p(a) = p^u(a)$. By defining the representation of both the feature and action as a concatenation $\phi(x, a) = (\phi_x, a)$, we have

$$\begin{aligned} p\left(\phi_{x,a}\right) &= p\left(\phi_{x,} a\right) \\ &= p\left(\phi_x \mid a\right) p(a) \\ &= p\left(\phi_x\right) p(a) \\ &= p\left(\phi_x\right) p^u(a) \\ &= p^u\left(\phi_x, a\right) \end{aligned}$$

under $p(a) = p^u(a)$.

These facts demonstrate that our proposed regularizer encourages a weaker (but sufficient) condition than the action-wise representation-balancing approach.

Note that there is a potential issue with the use of a representation-balancing regularizer with such a non-invertible representation as $\phi(x, a) = x + a$. That is, an unobservable error term would be induced in the upper-bound (Johansson et al., 2019; Zhao et al., 2019). Thus, a minimization of only the observable error terms ($L + D_{\mathrm{bal}}$) may not lead to a minimization of the target error in such cases. Some countermeasures have been proposed to address this issue, such as adding a reconstruction loss of inputs to guarantee invertibility of the representation (Zhang et al., 2020), but in some cases, $D_{\mathrm{bal}} = 0$ is achieved only by using a non-invertible representation, as in the example in Table 4. Therefore, this point remains an area for improvement in future work.

# D Experimental details

**Synthetic data generation process** Our synthetic datasets are built as follows.

1 Sample $x \sim \mathcal{N}(0, 1)^d$, where $d = 5$.

2 Sample $a \in \{0, 1\}^m$, where $m = 5$, from $p(a|x) \propto \exp(10\,|x_\Sigma - a_\Sigma|)$, where $x_\Sigma$ and $a_\Sigma$ are the following.

- 2-1 In settings other than Setup B, $x_\Sigma = x_\Delta = w_x^\top x$, where $w_x \sim \mathcal{N}(0, 1/d)^d$.
- 2-2 In Setup B, $x_\Sigma = x_1$, i.e., only the first dimension in $x$ is used to bias $a$.
- 2-3 $a_\Sigma = w_a^\top a$, where $w_a \sim \mathcal{N}(0, 1/m)^m$.

3 Calculate the expected outcome $y_a = f(x, a)$, where we examine two types of functions $f$, namely, Quadratic and Bilinear. In the Quadratic setting, $f(x, a) = a_\Upsilon^2 - 2x_\Upsilon$, where $x_\Upsilon$ and $a_\Upsilon$ are one-dimensional representations of $x$ and $a$, respectively.

- 3-1 In Setup B, $x_\Upsilon = w_{x,2:d}^\top x_{2:d}$, where $x_{2:d}$ denotes all dimensions other than the first one ($x_\Sigma$).
- 3-2 In settings other than Setup B, $x_\Upsilon = x_\Sigma (=: x_\Delta)$.
- 3-3 In Setup A, $a_\Upsilon = w_a'^\top a$, where $w_a' \sim \mathcal{N}(0, 1/m)^m$.
- 3-4 In settings other than Setup A, $a_\Upsilon = a_\Sigma (=: a_\Delta)$.
- 3-5 In the Bilinear setting, $f(x, a) = x^\top W a$, where $W \sim \mathcal{N}(0, 1/(dm))^{(d,m)}$.

4 Sample the observed outcome $y \sim \mathcal{N}(y_a, 0.1)$.

**Details of semi-synthetic data** We transformed the target attributes of elapsed times into the average speed as the outcome, i.e., $y = \frac{4}{\sum z_i}$, where $\{z_i\}_{1:4}$ are the original elapsed times. Then we standardized $y$ and the feature attributes. Each feature attribute can take binary values or up to four different powers of two values. Out of 1,327k total parameter combinations, only 241.6k feasible combinations are recorded. We split these original feature dimensions into $a$ and $x$ as follows. The dimensions of the action space $m$ ranged from three to six, and the 8th, 11th, 12th, 13th, 14th, and 3rd dimensions are regarded as $a$ from the head in order (e.g., for $m = 3$, the 8th, 11th, and 12th dimensions in the original feature attributes are regarded as $a$). This split was for maximizing the overlap of $\mathcal{A}(x)$ among $\mathcal{X}$.

**Other DNN parameters** The detailed parameters we used for the DNN-based methods (S-DNN, M-DNN, CFRNet, RMNet-IPM, and RMNet-HSIC) were

Table 5: Training sample size for each setting.

| $m$ | $\lvert\mathcal{A}\rvert$ | $N_{\text{tr}}$ |
|---|---|---|
| 3 | 8 | 24,160 |
| 4 | 16 | 12,080 |
| 5 | 32 | 6,040 |
| 6 | 64 | 3,591 |

as follows. The backbone DNN structure had four layers for representation extraction and three layers for hypothesis with the width of 64 for the middle layers and the width of 10 for the representation $\phi_{x,a}$. In RMNet-HSIC, the representation $\phi_{x,a} = (\phi_x, \phi_a)$ was composed of representations of feature $(\phi_x)$ and action $(\phi_a)$, each of which had a width of 5. The batch size was 64 except for CFRNet, where it was 512 due to the need to approximate the distributions for each action. The strength of the L2 regularizer was $10^{-4}$. We used Adam (Kingma and Ba, 2015) as the optimizer with the learning rate of $10^{-4}$.

## E Connection between $\text{ER}^u_k$ in Proposition 4.1 and $\text{ER}^u_\mu$ in (2)

We explain how we obtain $\text{ER}^u$ in (2) from $\text{ER}^u_k$ in Proposition 4.1. Recall $\text{ER}^u_k$ in Proposition 4.1:

$$\text{ER}^u_k(f) := \mathop{\mathbb{E}}_x \left[ \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} I\left( (\text{rank}(y_a) \le k) \oplus (\text{rank}(f(x,a)) \le k) \right) \right].$$

We can show the following:

$$\begin{aligned} & I\left( (\text{rank}(y_a) \le k) \oplus (\text{rank}(f(x,a)) \le k) \right) \\ &= I\left( (y_{a^*_k} \le y_a) \oplus (f(x, \hat{a}^*_k) \le f(x,a)) \right) \\ &= I\left( (y_{a^*_k} \le y_a) \oplus (y_{a^*_k} \le f(x,a) - f(x,\hat{a}^*_k) + y_{a^*_k}) \right) \\ &= I\left( (y_{a^*_k} \le y_a) \oplus (y_{a^*_k} \le f'(x,a)) \right), \end{aligned}$$

where $f'(x,a) := f(x,a) - f(x,\hat{a}^*_k) + y_{a^*_k}$. We then have

$$\text{ER}^u_k(f) = \mathop{\mathbb{E}}_x \left[ \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} I((y_a \ge y_{a^*_k}) \oplus (f'(x,a) \ge y_{a^*_k})) \right].$$

Here, the rank with $f'$ $(\text{rank}(f'(x,a)))$ is the same as that of $f$, but the difference is that $f'$ satisfies the condition $f'(x,\hat{a}^*_k) = y_{a^*_k}$. That is, the $k$-th largest value among $\{f'(x,a)\}_a$ equals to $y_{a^*_k}$. Although, since $y_{a^*_k}$ is unobservable, we relaxed the optimization of $f'$ in the function space that satisfies the condition into the optimization in the general function space. In addition, we used the average performance of the past decision-makers $(\mu)$ with respect to the target $(x)$ $\bar{y}$ instead of unobservable $y_{a^*_k}$, as

$$\text{ER}^u_\mu(f) = \mathop{\mathbb{E}}_x \left[ \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} I(y_a \ge \bar{y} \oplus f(x,a) \ge \bar{y}) \right].$$

In the end, for such $k'$ that satisfies $y_{a^*_{k'+1}} \le \bar{y} \le y_{a^*_{k'}}$ and for such $f'$ that satisfies $f'(x, \hat{a}^*_{k'}) = y_{a^*_{k'}}$, we have $\text{ER}^u_\mu(f') = \text{ER}^u_{k=k'}(f')$.

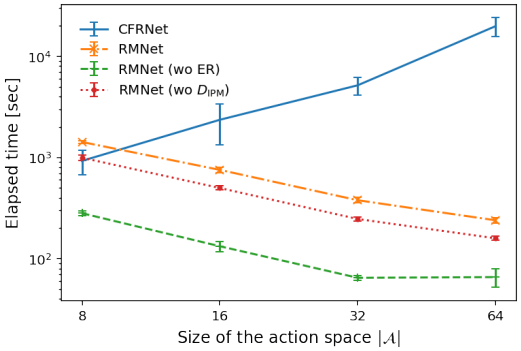



Figure 5: Elapsed time for training. Error bars indicate standard deviation.

## F Additional experimental results

**Elapsed times compared to CFR** Figure 5 shows the comparison in training time between the proposed method RMNet-IPM and CFRNet. For CFRNet, the elapsed time increased when the size of the action space $|\mathcal{A}|$ became large. The main reason for this is the calculation of distance between the representation distributions for each pair of actions $\sum_{a \ne a'} D_{\text{IPM}}\left(p_a(\phi), p_{a'}(\phi)\right)$ in Fig. 3(a). The decrease of the elapsed time for RMNet is mainly due to the sample sizes shown in Table 5.

**Semi-synthetic results for $k > 1$** Table 2 shows the results for $k = 1$. We also evaluated with respect to $k = 2$ and $k = 4$ as shown in Table 6 and Table 7, respectively. The metric for $k > 1$ is defined as the following normalized mean cumulative gain (NMCG) for $k$:

$$\text{NMCG}_k(f) := \mathop{\mathbb{E}}_x \left[ \sum_{a:\text{rank}(f(x,a)) \le k} y_a \right] \Big/ \mathop{\mathbb{E}}_x \left[ \sum_{a:\text{rank}(y_a) \le k} y_a \right].$$

The model selection is also performed with respect to $\text{mCG}_k$. The results were similar to that in $k = 1$, which demonstrates the robustness of the proposed methods with respect to the choice of $k$ (and thus the policy $\pi_k$).

Table 6: Semi-synthetic results on mean cumulative gain (mCG) and other metrics in $k = 2$. Best and second-best methods are in bold.

| | Normalized mean cumulative gain @ k=2 | | | | $\text{MSE}^u$ | | | | $\text{ER}^u_{k=2}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lvert\mathcal{A}\rvert$ | 8 | 16 | 32 | 64 | 8 | 16 | 32 | 64 | 8 | 16 | 32 | 64 |
| Method | | | | | | | | | | | | |
| OLS | 0.08 ± 0.15 | 0.01 ± 0.19 | −0.00 ± 0.12 | 0.03 ± 0.10 | 1.12 | 1.89 | 1.70 | 5.86 | 0.374 | 0.215 | 0.118 | 0.061 |
| RF | 0.27 ± 0.08 | 0.34 ± 0.07 | 0.33 ± 0.05 | 0.38 ± 0.05 | 1.03 | 0.87 | 0.93 | 1.07 | 0.358 | 0.205 | 0.111 | 0.057 |
| kNN | 0.27 ± 0.04 | 0.30 ± 0.06 | 0.30 ± 0.04 | 0.37 ± 0.02 | **0.59** | **0.64** | **0.64** | **0.63** | 0.356 | 0.206 | 0.112 | 0.057 |
| BART | 0.13 ± 0.13 | 0.18 ± 0.12 | 0.15 ± 0.10 | 0.09 ± 0.09 | 1.06 | 1.05 | 1.15 | 1.63 | 0.371 | 0.213 | 0.116 | 0.060 |
| M-DNN | 0.30 ± 0.08 | 0.43 ± 0.05 | 0.29 ± 0.06 | 0.34 ± 0.04 | 0.81 | 0.82 | 0.81 | 0.85 | 0.357 | 0.205 | 0.114 | 0.057 |
| S-DNN | 0.32 ± 0.08 | 0.27 ± 0.10 | 0.32 ± 0.07 | 0.45 ± 0.05 | 0.69 | 0.68 | 0.74 | 0.72 | 0.353 | 0.208 | 0.111 | 0.055 |
| CFRNet | 0.44 ± 0.09 | 0.41 ± 0.08 | 0.34 ± 0.07 | 0.35 ± 0.04 | 0.79 | 0.78 | 0.83 | 0.86 | **0.334** | 0.204 | 0.111 | 0.057 |
| RankNet | 0.37 ± 0.07 | 0.29 ± 0.09 | 0.37 ± 0.07 | 0.44 ± 0.05 | 6.98 | 11.34 | 8.26 | 2.60 | 0.349 | 0.205 | 0.109 | 0.055 |
| RMNet-IPM | **0.66 ± 0.01** | **0.43 ± 0.06** | **0.49 ± 0.04** | **0.50 ± 0.05** | 0.73 | 0.85 | 0.73 | 0.73 | **0.304** | **0.197** | **0.106** | **0.053** |
| RMNet-HSIC | **0.49 ± 0.07** | **0.49 ± 0.09** | **0.52 ± 0.05** | **0.65 ± 0.04** | **0.44** | **0.61** | **0.66** | **0.30** | 0.334 | **0.194** | **0.104** | **0.050** |

Table 7: Semi-synthetic results on mean cumulative gain (mCG) and other metrics in $k = 4$. Best and second-best methods are in bold.

| | Normalized mean cumulative gain @ k=4 | | | | $\text{MSE}^u$ | | | | $\text{ER}^u_{k=4}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lvert\mathcal{A}\rvert$ | 8 | 16 | 32 | 64 | 8 | 16 | 32 | 64 | 8 | 16 | 32 | 64 |
| Method | | | | | | | | | | | | |
| OLS | 0.18 ± 0.15 | −0.01 ± 0.15 | −0.00 ± 0.11 | 0.02 ± 0.07 | 1.12 | 1.89 | 1.70 | 5.86 | 0.471 | 0.373 | 0.221 | 0.117 |
| RF | 0.24 ± 0.08 | 0.34 ± 0.07 | 0.34 ± 0.05 | 0.36 ± 0.05 | 1.03 | 0.87 | 0.93 | 1.07 | 0.459 | 0.330 | 0.198 | 0.104 |
| kNN | 0.19 ± 0.05 | 0.26 ± 0.06 | 0.28 ± 0.04 | 0.36 ± 0.02 | **0.59** | 0.64 | **0.64** | **0.63** | 0.467 | 0.339 | 0.203 | 0.106 |
| BART | 0.11 ± 0.12 | 0.23 ± 0.11 | 0.16 ± 0.10 | 0.13 ± 0.09 | 1.06 | 1.05 | 1.15 | 1.63 | 0.485 | 0.350 | 0.212 | 0.114 |
| M-DNN | **0.42 ± 0.05** | 0.38 ± 0.06 | 0.28 ± 0.06 | 0.26 ± 0.04 | 0.79 | 0.82 | 0.82 | 0.85 | **0.418** | 0.334 | 0.207 | 0.110 |
| S-DNN | 0.28 ± 0.08 | 0.28 ± 0.10 | 0.31 ± 0.07 | 0.44 ± 0.05 | 0.68 | **0.59** | 0.79 | 0.69 | 0.451 | 0.339 | 0.198 | 0.098 |
| CFRNet | **0.46 ± 0.05** | 0.40 ± 0.06 | 0.30 ± 0.06 | 0.26 ± 0.03 | 0.79 | 0.79 | 0.86 | 0.86 | **0.408** | 0.327 | 0.204 | 0.111 |
| RankNet | 0.33 ± 0.07 | 0.28 ± 0.10 | 0.36 ± 0.06 | 0.44 ± 0.04 | 6.20 | 11.09 | 7.98 | 4.48 | 0.439 | 0.331 | 0.192 | 0.099 |
| RMNet-IPM | 0.39 ± 0.06 | **0.43 ± 0.07** | **0.43 ± 0.05** | **0.49 ± 0.05** | 0.69 | 0.77 | 0.67 | 0.72 | 0.422 | **0.318** | **0.188** | **0.095** |
| RMNet-HSIC | 0.35 ± 0.09 | **0.49 ± 0.07** | **0.47 ± 0.05** | **0.62 ± 0.02** | **0.65** | **0.53** | **0.55** | **0.35** | 0.438 | **0.305** | **0.180** | **0.087** |