Relaxing the Accurate Imputation Assumption in Doubly Robust Learning for Debiased Collaborative Filtering

Haoxuan Li [1] Chunyuan Zheng [1] Shuyi Wang [2] Kunhan Wu [3] Hao Wang [4]
Peng Wu [5] Zhi Geng [5] Xu Chen [6] Xiao-Hua Zhou [1]

Abstract

Recommender system aims to recommend items or information that may interest users based on their behaviors and preferences. However, there may be sampling selection bias in the data collection process, i.e., the collected data is not a representative of the target population. Many debiasing methods are developed based on pseudo-labelings. Nevertheless, the validity of these methods relies heavily on accurate pseudo-labelings (i.e., the imputed labels), which is difficult to satisfy in practice. In this paper, we theoretically propose several novel doubly robust estimators that are unbiased when either (a) the pseudo-labelings *deviate from* the true labels with an arbitrary user-specific inductive bias, item-specific inductive bias, or a combination of both, or (b) the learned propensities are accurate. We further propose a propensity reconstruction learning approach that adaptively updates the constraint weights using an attention mechanism and effectively controls the variance. Extensive experiments show that our approach outperforms the state-of-the-art on one semi-synthetic and three real-world datasets.

1. Introduction

By analyzing users' historical behaviors and preferences, recommender system (RS) predicts and recommends items or information that users may like (Rui et al., 2022; Li et al., 2024). However, as users are free to choose which item to rate, the collected data is always not a representative of the target population (or inference space) (Schnabel et al., 2016; Yang et al., 2021; Saito and Nomura, 2022; Yang et al., 2023; Wang et al., 2023c), and similar findings occur with tasks such as post-view click-through & conversion rate (CTCVR) prediction (Ma et al., 2018; Wang et al., 2022a), and uplift modeling (Saito et al., 2019; Sato et al., 2019; 2020). Since the labels are observable only in the collected data and missing in the target population, this poses a great challenge to achieve unbiased learning (Wang et al., 2020a).

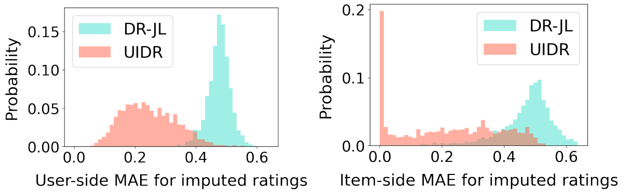


Figure 1. Inaccuracy of imputed ratings on the large scale industrial dataset KuaiRec. The user (item)-side mean absolute error (MAE) is the average difference between the imputed rating and the true rating on all the observed interactions for user u (item i).

To address this problem, the error-imputation-based (EIB) methods (Hernández-Lobato et al., 2014) first impute the missing labels and then train the prediction model using both the observed labels and pseudo-labelings. However, as the pseudo-labeling model is trained with observed data while deployed in the missing data, it is difficult to obtain accurate pseudo-labelings, leading to sub-optimal performance (Dai et al., 2022). The inverse-propensity-scoring (IPS) methods inversely weight the prediction error for each sample with observed rating using the propensity of collecting user rating (Schnabel et al., 2016), but it is empirically difficult to set proper propensity scores and theoretically has greater variance (Saito, 2020). By utilizing both pseudo-labeling and propensity models, the doubly robust (DR) methods are proposed to weaken the unbiasedness condition of the EIB and IPS estimators (Wang et al., 2019), with many enhanced DR approaches developed (Guo et al., 2021; Dai et al., 2022; Wang et al., 2022a; Song et al., 2023; Li et al., 2023a; Zhang et al., 2024). The advantage of DR estimators is attributed to the property of double robustness, i.e., it is unbiased if either the learned propensities or the pseudo-labelings are accurate. We summarize the unbiasedness condition of the previous debiasing estimators in Table 1 (see Appendix A

[1] Peking University [2] University of Pennsylvania [3] Carnegie Mellon University [4] Zhejiang University [5] Beijing Technology and Business University [6] Renmin University of China. Correspondence to: Xu Chen, Xiao-Hua Zhou <xu.chen@ruc.edu.cn, azhou@math.pku.edu.cn>.

Proceedings of the 41^{st} International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

Table 1. Comparison of the various debiasing methods, where $\hat{p}$ and $\tilde{r}$ denotes the learned propensities and pseudo-labelings, respectively. The red and blue color highlight the unbiasedness of the proposed DR estimator under arbitrary user-specific inductive bias $f(b_u)$ and item-specific inductive bias $g(b_i)$, where f and g are two arbitrary real value functions.

Method	Unbiasedness Condition
EIB	$\tilde{r}_{u,i} = r_{u,i}$
IPS, Multi-IPS, ESCM^2-IPS	$\hat{p}_{u,i} = p_{u,i}$
DR, Multi-DR, DR-JL, MRDR, ESCM^2-DR	$\hat{p}_{u,i} = p_{u,i}$ or $\tilde{r}_{u,i} = r_{u,i}$
User-DR (ours) Item-DR (ours) User-Item-DR (ours)	$\tilde{p}_{u,i} = p_{u,i}$ or $\tilde{r}_{u,i} = r_{u,i} + f(b_u)$, for all f and b_u $\tilde{p}_{u,i} = p_{u,i}$ or $\tilde{r}_{u,i} = r_{u,i} + g(b_i)$, for all g and b_i $\tilde{p}_{u,i} = p_{u,i}$ or $\tilde{r}_{u,i} = r_{u,i} + f(b_u) + g(b_i)$, for all f, g, b_u and b_i

Note: b_u and b_i are arbitrary user-specific and item-specific inductive biases (see Section 3 for more details). Since the proposed methods require reconstructing the learned propensities, so we use $\tilde{p}_{u,i}$ instead of $\hat{p}_{u,i}$ to distinguish.

for more detailed discussions on related work).

Despite the double robustness providing additional protection against inaccurate pseudo-labelings, recent studies have shown that DR methods are highly sensitive to inaccurate pseudo-labelings. Specifically, when the learned propensities are slightly inaccurate, the DR estimator can be severely biased with inaccurate pseudo-labelings (Kang and Schafer, 2007; Molenberghs et al., 2015; Vermeulen and Vansteelandt, 2015; Seaman and Vansteelandt, 2018). These inaccurate pseudo-labels would as a result lead to biased prediction models during the training phase (Mansoury et al., 2020; Krauth et al., 2022; Wen et al., 2022). Therefore, it is essential to develop novel DR estimators with relaxed unbiasedness conditions on the accurate pseudo-labelings.

To this end, in this paper we theoretically propose several novel DR estimators that are *unbiased under inaccurate pseudo-labelings*, named User-DR, Item-DR, and User-Item-DR. As shown in Table 1, our theoretical analysis proves that the User-Item-DR estimator is unbiased as long as the pseudo-labelings *deviate* the true labels with an *arbitrary user-specific* and *item-specific inductive bias*. Whereas the corresponding unbiasedness condition of previous DR estimators requires the pseudo-labelings to be *strictly equal* to the true labels, which is much stronger than that of the proposed User-Item DR estimator. In addition, similar to the DR estimators, the proposed DR estimators are unbiased if the learned propensities are accurate. Figure 1 shows the inaccuracy of the imputed rating on a large scale industrial dataset. It can be found that the proposed User-Item-DR method has a much lower bias compared to the doubly robust joint learning (DR-JL) method, which provides the empirical evidence of the effectiveness of the proposed method.

We further propose a propensity reconstruction learning approach that alternatively updates the propensity model, the imputation model, and the prediction model for debiased learning. To enable the proposed doubly robust estimators to be unbiased with user- and item-specific inductive biases, we introduce additional constraints in the learning phase of the propensity model, where the constraint weights are adaptively updated using an attention mechanism. We also theoretically show that such propensity reconstruction approach can effectively achieve a better bias-variance trade-off.

The main contributions of this paper are:

- We theoretically propose novel DR estimators that are unbiased when the pseudo-labelings deviate from the true labels with an arbitrary and unknown user-specific inductive bias, item-specific inductive bias, or a combination of both.
- We further propose a propensity reconstruction learning approach that adaptively updates the constraint weights using an attention mechanism, and theoretically show that such approach can achieve a better bias-variance trade-off.
- We perform semi-synthetic experiments to verify the effectiveness of the proposed methods for arbitrary user-specific and item-specific inductive bias, while previous methods fail to unbiasedly estimate the ideal loss. We also conducted extensive experiments on three real-world datasets to demonstrate the advantages of the proposed methods.

2. Preliminaries

Let $\mathcal{U} = \{u_1, u_2, \dots, u_m\}$ be the set of m users, $\mathcal{I} = \{i_1, i_2, \dots, i_n\}$ be the set of n items, and $\mathcal{D} = \mathcal{U} \times \mathcal{I}$ be the set of all user-item pairs. Denote $\mathbf{R} \in \mathbb{R}^{m\times n}$ as the rating matrix of all user-item pairs, where $r_{u,i}$ indicates the rating of user u on item i. Let $x_{u,i}$ be the feature of user u and item i, and $\hat{\mathbf{R}} \in \mathbb{R}^{m\times n}$ be the rating prediction matrix for $\mathbf{R}$, where $\hat{r}_{u,i} = f(x_{u,i};\theta)$ is the predicted rating induced by a prediction model, θ is the parameter. Let $o_{u,i}$ be the indicator of whether user u rated item i, $\mathcal{O} = \{(u,i) \in \mathcal{D} | o_{u,i} = 1\}$ be the user-item index set with observed ratings, and $\mathbf{R}^o = \{r_{u,i} \in \mathbf{R} | o_{u,i} = 1\}$ be the observed ratings. If $\mathbf{R}$ is fully observed, then the prediction model $f(x_{u,i};\theta)$ can be trained by minimizing the ideal loss

$$\mathcal{L}_{\text{ideal}}(\theta) = \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}} \delta_{u,i},$$

where $\delta_{u,i} = r_{u,i}\delta^{(1)}(\hat{r}_{u,i}) + (1 - r_{u,i})\,\delta^{(0)}(\hat{r}_{u,i})$ is the prediction error, and $\delta^{(r)}(\hat{r}_{u,i})$ is a pre-defined loss function for $r = 0, 1$. For example, $\delta^{(r)}(\hat{r}_{u,i}) = -r\log\hat{r}_{u,i} - (1-r)\log(1-\hat{r}_{u,i})$ represents the cross-entropy loss. However, optimizing the ideal loss is infeasible, as $r_{u,i}$ is observable only when $o_{u,i} = 1$. A naive method is to optimize the prediction model directly using the user-item pairs with observed ratings, but this will incur sample selection bias because the user-item pairs with observed ratings are no longer representative of all user-item pairs.

To address this problem, many debiasing methods have been proposed by designing unbiased estimators of the ideal loss. For example, the EIB method directly imputes the label $r_{u,i}$ corresponding to missing events, with the estimator

$$\mathcal{L}_{\mathrm{EIB}}(\theta) = \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left[o_{u,i}\delta_{u,i} + (1-o_{u,i})\hat{\delta}_{u,i}\right],$$

where $\hat{\delta}_{u,i} = \tilde{r}_{u,i}\delta^{(1)}(\hat{r}_{u,i}) + (1-\tilde{r}_{u,i})\,\delta^{(0)}(\hat{r}_{u,i})$ is the imputed error, $\tilde{r}_{u,i}$ is the pseudo-labeling for estimating $r_{u,i}$ given by a labeling-imputation model. Clearly, $\mathcal{L}_{\mathrm{EIB}}(\theta)$ is an unbiased estimator of the ideal loss when all the pseudo-labelings are accurate, i.e., $\tilde{r}_{u,i} = r_{u,i}$ for $(u,i) \in \mathcal{D}\setminus\mathcal{O}$. Nevertheless, the EIB method usually has sub-optimal performance in practice due to the difficulty of obtaining accurate pseudo-labelings for the missing ratings (Guo et al., 2021). By additionally introducing the propensity $p_{u,i} = \mathbb{P}(o_{u,i} = 1|x_{u,i})$, the DR estimator is proposed as

$$\mathcal{L}_{\mathrm{DR}}(\theta) = \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\Big[\hat{\delta}_{u,i} + \frac{o_{u,i}(\delta_{u,i} - \hat{\delta}_{u,i})}{\hat{p}_{u,i}}\Big],$$

where $\hat{p}_{u,i}$ is the propensity model for estimating $p_{u,i}$. Despite theoretically being doubly robust, i.e., unbiasedness holds when either the learned propensities or the pseudo-labelings are accurate for all user-item pairs, however, it has been widely shown that the DR estimator would result in severe bias under inaccurate pseudo-labelings if the learned propensities are slightly inaccurate (Tan, 2007; Molenberghs et al., 2015; Seaman and Vansteelandt, 2018).

3. Proposed Method

In this section, we first propose User-DR, Item-DR, and User-Item-DR estimators in Section 3.1, and theoretically show the unbiasedness of the proposed estimators for arbitrary user-specific and item-specific inductive biases, which greatly weakens the unbiasedness condition of the previous DR estimators on pseudo-labelings. In Section 3.2, we show that the variances of the proposed estimators are highly controllable, provided that the reconstructed propensities do not differ much from the original propensities. In Section 3.3, we further propose a propensity reconstruction learning approach to effectively achieve unbiased learning.

3.1. User-DR, Item-DR, and User-Item-DR Estimators

In contrast to previous DR estimators that directly use $\hat{p}_{u,i}$ as propensities, where $\hat{p}_{u,i}$ are obtained by performing a binary classification on $o_{u,i}$ using $x_{u,i}$ (Wang et al., 2019; Saito, 2020; Guo et al., 2021), given a prediction model $\hat{r}_{u,i}$, the proposed User-DR estimator first learns a constrained propensity model $\tilde{p}_{u,i}$ that satisfies for all $u \in \mathcal{U}$, we have

$$\sum_{i\in\mathcal{I}}\left(\frac{o_{u,i}}{\tilde{p}_{u,i}} - 1\right)\left(\delta^{(1)}(\hat{r}_{u,i}) - \delta^{(0)}(\hat{r}_{u,i})\right) = 0, \qquad (1)$$

where $\delta^{(r)}(\hat{r}_{u,i})$ is a pre-defined loss function for $r = 0, 1$, such as the cross-entropy loss $\delta^{(r)}(\hat{r}_{u,i}) = -r\log\hat{r}_{u,i} - (1-r)\log(1-\hat{r}_{u,i})$. Then the User-DR estimator is

$$\mathcal{L}_{\mathrm{UDR}}(\theta) = \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\Big[\hat{\delta}_{u,i} + \frac{o_{u,i}(\delta_{u,i} - \hat{\delta}_{u,i})}{\tilde{p}_{u,i}}\Big],$$

which adopts a similar form to the DR estimator, but requires the learned propensities $\tilde{p}_{u,i}$ satisfying the above constraints in Eq. (1).

Now, we prove that the constraints in Eq. (1) can effectively alleviate the inaccurate pseudo-labelings problem in the previous DR estimators. Formally, the bias of the User-DR estimator is

$$\begin{aligned}
&\mathrm{Bias}(\mathcal{L}_{\mathrm{UDR}}(\theta)) = \mathcal{L}_{\mathrm{ideal}}(\theta) - \mathbb{E}(\mathcal{L}_{\mathrm{UDR}}(\theta))\\
&= \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\delta_{u,i} - \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\mathbb{E}\Big[\hat{\delta}_{u,i} + \frac{o_{u,i}(\delta_{u,i} - \hat{\delta}_{u,i})}{\tilde{p}_{u,i}}\Big]\\
&= \mathbb{E}\Big[\frac{1}{|\mathcal{D}|}\sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{I}}\left(\frac{o_{u,i}}{\tilde{p}_{u,i}} - 1\right)(\hat{\delta}_{u,i} - \delta_{u,i})\Big]\\
&= \mathbb{E}\Big[\frac{1}{|\mathcal{D}|}\sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{I}}\left(\frac{o_{u,i}}{\tilde{p}_{u,i}} - 1\right)\\
&\qquad\Big\{\left(\delta^{(1)}(\hat{r}_{u,i}) - \delta^{(0)}(\hat{r}_{u,i})\right)(\tilde{r}_{u,i} - r_{u,i})\Big\}\Big].
\end{aligned}$$

The last equation holds directly from the definitions of $\hat{\delta}_{u,i}$ and $\delta_{u,i}$. On the one hand, similar to the previous DR estimators, the User-DR estimator is unbiased under either accurate pseudo-labelings $\tilde{r}_{u,i} = r_{u,i}$ or accurate learned propensities $\tilde{p}_{u,i} = p_{u,i} = \mathbb{P}(o_{u,i} = 1|x_{u,i})$ for all user-item pairs. On the other hand, when the pseudo-labelings model has a user-specific inductive bias, i.e., $\tilde{r}_{u,i} = r_{u,i} + f(b_u)$, multiplying both sides of the Eq. (1) by $f(b_u)$ and summing over all u yields

$$\begin{aligned}
\mathrm{Bias}(\mathcal{L}_{\mathrm{UDR}}(\theta)) = \mathbb{E}\Big[&\sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{I}}\left(\frac{o_{u,i}}{\tilde{p}_{u,i}} - 1\right)\\
&\left(\delta^{(1)}(\hat{r}_{u,i}) - \delta^{(0)}(\hat{r}_{u,i})\right)\cdot f(b_u)\Big] = 0.
\end{aligned}$$

We summarize and compare the unbiasedness conditions of previous DR estimators (Wang et al., 2019; Saito, 2020; Zhang et al., 2020; Guo et al., 2021; Wang et al., 2022a) and the proposed User-DR estimator below.

Lemma 1 (Wang et al. (2019))**.** *The DR estimator is unbiased when either* ***pseudo-labelings are accurate*** $\tilde{r}_{u,i} = r_{u,i}$ *or learned propensities are accurate* $\hat{p}_{u,i} = p_{u,i}$.

Theorem 1 (main result)**.** *The User-DR estimator is unbiased when either* ***pseudo-labelings deviate from the true labels with arbitrary user-specific inductive bias*** $\tilde{r}_{u,i} = r_{u,i} + f(b_u)$ *or learned propensities are accurate* $\tilde{p}_{u,i} = p_{u,i}$.

Since the improved theoretical guarantees for User-DR originate from constraints in Eq. (1), one may argue whether such constraints are too strong to be satisfied. In fact, a key observation is that when the learned propensities are accurate, i.e., $\tilde{p}_{u,i} = p_{u,i}$, then these constraints will be satisfied naturally, and $\mathcal{L}_{\mathrm{UDR}}(\theta)$ will degenerates to $\mathcal{L}_{\mathrm{DR}}(\theta)$, which does not impose additional constraints to reduce the accuracy of learned propensities. In contrast, if the learned propensities are inaccurate, i.e., $\tilde{p}_{u,i} \neq p_{u,i}$, the bias of the previous DR estimators will strictly depend on the accuracy of the pseudo-labelings, whereas the proposed User-DR reduces the influence of those inaccurate pseudo-labelings by learning an alternative propensity model that satisfies constraints in Eq. (1).

Similar to the construction of the User-DR, we propose the Item-DR estimator that is unbiased to item-specific inductive bias by replacing the constraints in Eq. (1) with

$$\sum_{u \in \mathcal{U}} \left(\frac{o_{u,i}}{\tilde{p}_{u,i}} - 1 \right) \left(\delta^{(1)}(\hat{r}_{u,i}) - \delta^{(0)}(\hat{r}_{u,i}) \right) = 0 \qquad (2)$$

for all $i \in \mathcal{I}$. Furthermore, the User-Item-DR estimator, which is robust to both user-specific and item-specific inductive biases, can be obtained by learning a propensity model satisfying constraints in both Eq. (1) and Eq. (2). Similar to Theorem 1, we have the following results.

Corollary 1. *(a) The Item-DR estimator is unbiased, if either (i)* $\tilde{r}_{u,i} = r_{u,i} + g(b_i)$*, or (ii)* $\tilde{p}_{u,i} = p_{u,i}$*;*

(b) The User-Item-DR estimator is unbiased, if either (i) $\tilde{r}_{u,i} = r_{u,i} + f(b_u) + g(b_i)$*, or (ii)* $\tilde{p}_{u,i} = p_{u,i}$.

It is also meaningful to consider inaccurate pseudo-labelings with item-specific inductive bias, e.g., item popularity bias and item exposure position bias. Notably, we would like to clarify that *even if the "user/item-specific inductive bias" conditions are not strictly satisfied, the biases of the proposed DR estimators are still strictly smaller than the previous DR,* as long as the existence of a user-specific constant $f(b_u)$ or an item-specific constant $g(b_i)$ such that the bias arises from the inaccurate pseudo-labelings $\{\tilde{r}_{u,i} : i \in \mathcal{I}\}$ or $\{\tilde{r}_{u,i} : u \in \mathcal{U}\}$ can be reduced. We illustrate this with a toy example as follows. Suppose the inductive biases of the $\tilde{r}_{u,i}$ for user u on items i_1, i_2, and i_3 are $\tilde{r}_{u,i_1} - r_{u,i_1} = 1$, $\tilde{r}_{u,i_2} - r_{u,i_2} = 2$, and $\tilde{r}_{u,i_3} - r_{u,i_3} = 3$, respectively. Then the UDR estimator are able to cancel a user-specific constant $f(b_u)$ (e.g., $f(b_u) = 2$) to make the inductive biases become $1 - f(b_u)$, $2 - f(b_u)$, and $3 - f(b_u)$, which leads to smaller biases. We would like to emphasize that it is not necessary for our method to obtain a better-imputed rating than the previous DR-based methods, i.e., we do not need to figure out what the $f(b_u)$ is in the previous example. The point is that we can achieve much lower bias on the imputation side even if the imputed ratings for our method and other DR methods are the same due to the learned propensities of our methods satisfying the constraints.

One may argue that we reduce the bias of the imputation-side at the expense of the accuracy of the propensity-side. In fact, Imai and Ratkovic (2014) and Li et al. (2023c) point out that directly learning a propensity with the simple cross entropy loss is not sufficient for the propensity learning and a high-quality propensity should have the following covariate balancing property:

$$\mathbb{E}\left[\frac{o_{u,i}\phi(x_{u,i})}{p_{u,i}}\right] = \mathbb{E}\left[\frac{(1 - o_{u,i})\phi(x_{u,i})}{1 - p_{u,i}}\right] = \mathbb{E}[\phi(x_{u,i})],$$

where $\phi(\cdot)$ is an arbitrary function. Propensity constraints will make the learned propensity have the balancing property, which leads to a higher quality of learned propensities.

3.2. Further Theoretical Analysis on Variance

The proposed estimators greatly enhance the robustness of DR to inaccurate pseudo-labelings. A further question is whether such unbiasedness comes at the cost of increased variance. Impressively, the variances are highly controllable and manageable as shown below (see Appendix B for proof).

Theorem 2. *If* $1/L \leq \hat{p}_{u,i}^2 / \tilde{p}_{u,i}^2 \leq L$ *for a constant* $L \geq 1$,

$$\frac{1}{L} \cdot \mathbb{V}(\mathcal{L}_{\mathrm{DR}}(\theta)) \leq \mathbb{V}(\mathcal{L}_{\mathrm{UDR}}(\theta)) \leq L \cdot \mathbb{V}(\mathcal{L}_{\mathrm{DR}}(\theta)).$$

Theorem 2 shows that the variance of the User-DR estimator[1] can be controlled by the distance between the base propensities $\hat{p}_{u,i}$ and the learned constrained propensities $\tilde{p}_{u,i}$, which is essentially a bias-variance trade-off compared with the previous DR estimators. This motivates us to further propose a propensity reconstruction learning approach to meet the constraints in Eq. (1) and Eq. (2) with minimal changes to the original propensities $\hat{p}_{u,i}$ in the following.

3.3. Propensity reconstruction learning

We next propose a propensity reconstruction learning approach that adaptively updates the constraint weights to meet

[1] Theorem 2 also holds for Item-DR and User-Item-DR.

Algorithm 1 Propensity Reconstruction Learning

```
Input: observed ratings $\mathbf{R}^o$ and learned propensities $\hat{\mathbf{P}}$
while stopping criteria is not satisfied do
    for number of steps training the propensity model do
        Sample a batch of user-item pairs from $\mathcal{D}$
        Update $\tilde{\mathbf{P}}$: $\alpha \leftarrow \alpha - \eta \nabla_{\alpha} \mathcal{L}_p(\alpha; \theta, \beta \mid \hat{\mathbf{P}})$
    end for
    for number of steps training the imputation model do
        Sample a batch of user-item pairs from $\mathcal{O}$
        Update $\hat{\mathbf{E}}$: $\beta \leftarrow \beta - \eta \nabla_{\beta} \mathcal{L}_e(\beta; \alpha, \theta \mid \tilde{\mathbf{P}})$
    end for
    for number of steps training the prediction model do
        Sample a batch of user-item pairs from $\mathcal{D}$
        Update $\hat{\mathbf{R}}$: $\theta \leftarrow \theta - \eta \nabla_{\theta} \mathcal{L}_r(\theta; \alpha, \beta \mid \tilde{\mathbf{P}})$
    end for
end while
```

the constraints of the proposed User-Item-DR estimator[2]. The proposed algorithm alternately trains a reconstructed propensity model, a pseudo-labeling model for imputing the prediction errors $\hat{\mathbf{E}} = \{\hat{\delta}_{u,i} | (u,i) \in \mathcal{D}\}$, and a rating prediction model $\hat{\mathbf{R}} = \{\hat{r}_{u,i} | (u,i) \in \mathcal{D}\}$.

Step 1. Propensity Reconstruction $\hat{\mathbf{P}} \rightarrow \tilde{\mathbf{P}}$. Given the learned propensities $\hat{\mathbf{P}} = \{\hat{p}_{u,i} | (u,i) \in \mathcal{D}\}$ without constraints, Theorem 2 states the distance between $\hat{\mathbf{P}}$ and $\tilde{\mathbf{P}} = \{\tilde{p}_{u,i} | (u,i) \in \mathcal{D}\}$ can provide an upper bound on the variance of the proposed User-Item-DR estimator. Therefore, a natural idea is to reconstruct $\hat{\mathbf{P}}$ to the nearest $\tilde{\mathbf{P}}$ that satisfies the constraints in Eq. (1) and Eq. (2) in the User-Item-DR estimator. The optimization problem is

$$
\begin{aligned}
\min_{\tilde{p}} & \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}} \left(\frac{1}{\hat{p}_{u,i}} - \frac{1}{\tilde{p}_{u,i}} \right)^2, \\
\text{s.t. } & \tilde{p}_{u,i} > 0, \quad (u,i) \in \mathcal{D}, \\
& \sum_{i \in \mathcal{I}} \left(\frac{o_{u,i}}{\tilde{p}_{u,i}} - 1 \right) \left(\delta^{(1)}(\hat{r}_{u,i}) - \delta^{(0)}(\hat{r}_{u,i}) \right) = 0, \ u \in \mathcal{U}, \\
& \sum_{u \in \mathcal{U}} \left(\frac{o_{u,i}}{\tilde{p}_{u,i}} - 1 \right) \left(\delta^{(1)}(\hat{r}_{u,i}) - \delta^{(0)}(\hat{r}_{u,i}) \right) = 0, \ i \in \mathcal{I},
\end{aligned}
$$

which is a convex optimization problem with respect to $1/\tilde{p}$. The following states the rationality of reconstructing the inverse of the propensities rather than the propensities themselves: first, the former leads to a convex optimization, so that gradient-based algorithms can efficiently find globally optimal solutions; second, from the theoretical analysis of DR estimators (Wang et al., 2019; Guo et al., 2021; Dai et al., 2022), the bias and variance of the DR estimators are proportional to the inverse propensities and squared inverse propensities, respectively, therefore providing more theoretical guarantees. In practice, the optimization problem can be solved by minimizing the reconstruction loss with the constraints as the regularizations that

$$
\begin{aligned}
\mathcal{L}(\tilde{p} \mid \hat{p}) = & \frac{1}{2} \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}} \left(\frac{1}{\hat{p}_{u,i}} - \frac{1}{\tilde{p}_{u,i}} \right)^2 \\
& + \frac{\gamma}{2} \sum_{u \in \mathcal{U}} \lambda_u \Big[\sum_{i \in \mathcal{I}} \left(\frac{o_{u,i}}{\tilde{p}_{u,i}} - 1 \right) \left(\delta^{(1)}(\hat{r}_{u,i}) - \delta^{(0)}(\hat{r}_{u,i}) \right) \Big]^2 \\
& + \frac{\gamma}{2} \sum_{i \in \mathcal{I}} \lambda_i \Big[\sum_{u \in \mathcal{U}} \left(\frac{o_{u,i}}{\tilde{p}_{u,i}} - 1 \right) \left(\delta^{(1)}(\hat{r}_{u,i}) - \delta^{(0)}(\hat{r}_{u,i}) \right) \Big]^2,
\end{aligned}
$$

where $\tilde{p}_{u,i} = \pi(x_{u,i}; \alpha)$ is the reconstructed propensity model, λ_u and λ_i are Lagrange multipliers reflecting the constraint strength, γ is a trade-off hyper-parameter. Nevertheless, there are $O(|\mathcal{U}| + |\mathcal{I}|)$ constraints as well as the Lagrange multipliers in Eq. (1) and Eq. (2). Therefore, the dual optimization will not lead to faster efficiency. To address this problem, we propose an attention mechanism for collaborative filtering that adaptively learns the constraint strength (which is also considered to be the role of Lagrange multipliers), thus reducing the number of parameters of the dual problem. Specifically, let $\boldsymbol{s}_u$ and $\boldsymbol{t}_i$ be the latent vectors of user u and item i, we propose to use an attention mechanism to learn λ_u and λ_i, which can be formalized as

$$
\lambda_u = \frac{\sum_{i \in \mathcal{I}} \exp(\tilde{\boldsymbol{s}}_u^\top \boldsymbol{t}_i)}{\sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}} \exp(\tilde{\boldsymbol{s}}_u^\top \boldsymbol{t}_i)}, \text{ and } \tilde{\boldsymbol{s}}_u = \tanh(\boldsymbol{A}\boldsymbol{s}_u + \boldsymbol{b}),
$$

where $\boldsymbol{A}$ is the connection weight matrix and $\boldsymbol{b}$ is the bias, and λ_i can be obtained from a similar way. We further empirically explored other selections of λ_u in Section 5, such as the constant weights $\lambda_u = \lambda = 1/|\mathcal{U}|$, or obtain the weights via a multilayer perceptron.

Step 2. Training Pseudo-labelling with $\tilde{\mathbf{P}}$. The pseudo-labeling model can be learned by minimizing the weighted average loss of the prediction error and the imputed error of the observed samples

$$
\mathcal{L}_e(\beta; \alpha, \theta \mid \tilde{\mathbf{P}}) = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \frac{o_{u,i}(\delta_{u,i} - \hat{\delta}_{u,i})^2}{\tilde{p}_{u,i}},
$$

where β is the parameter of the pseudo-labeling model, $\delta_{u,i} = r_{u,i}\delta^{(1)}(\hat{r}_{u,i}) + (1 - r_{u,i})\, \delta^{(0)}(\hat{r}_{u,i})$ is the prediction error, and $\hat{\delta}_{u,i} = \tilde{r}_{u,i}\delta^{(1)}(\hat{r}_{u,i}) + (1 - \tilde{r}_{u,i})\, \delta^{(0)}(\hat{r}_{u,i})$ is the imputed error.

Step 3. Training Prediction Model with $\tilde{\mathbf{P}}$. Given the reconstructed propensities $\tilde{p}_{u,i}$ obtained in Step 1, the prediction model can be learned by minimizing the proposed User-Item-DR loss

$$
\mathcal{L}_r(\theta; \alpha, \beta \mid \tilde{\mathbf{P}}) = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \Big[\hat{\delta}_{u,i} + \frac{o_{u,i}(\delta_{u,i} - \hat{\delta}_{u,i})}{\tilde{p}_{u,i}} \Big].
$$

[2] Without loss of generality, we use User-Item-DR estimator in Section 3.3 for illustration purpose.

Table 2. Relative errors on ML-100K dataset with user-specific and item-specific inductive bias.

U-BIAS	Naive	EIB	IPS	SNIPS	DR	UDR	IDR	UIDR
ONE	0.068 ± 0.003	0.223 ± 0.004	0.035 ± 0.004	$\underline{0.034 \pm 0.004}$	0.047 ± 0.005	$\mathbf{0.006} \pm \mathbf{0.006}^*$	$\mathbf{0.016} \pm \mathbf{0.011}^*$	$\mathbf{0.003} \pm \mathbf{0.002}^*$
THREE	0.078 ± 0.004	0.234 ± 0.004	0.040 ± 0.004	$\underline{0.039 \pm 0.005}$	0.049 ± 0.005	$\mathbf{0.005} \pm \mathbf{0.003}^*$	$\mathbf{0.017} \pm \mathbf{0.028}^*$	$\mathbf{0.002} \pm \mathbf{0.001}^*$
FIVE	0.100 ± 0.004	0.247 ± 0.004	$\underline{0.050 \pm 0.004}$	0.050 ± 0.005	0.054 ± 0.005	$\mathbf{0.009} \pm \mathbf{0.009}^*$	$\mathbf{0.028} \pm \mathbf{0.025}^*$	$\mathbf{0.010} \pm \mathbf{0.006}^*$
ROTATE	0.137 ± 0.002	0.036 ± 0.001	0.068 ± 0.004	0.069 ± 0.002	$\underline{0.008 \pm 0.002}$	$\mathbf{0.001} \pm \mathbf{0.001}^*$	$\mathbf{0.003} \pm \mathbf{0.003}^*$	$\mathbf{0.002} \pm \mathbf{0.001}^*$
SKEW	0.025 ± 0.002	0.108 ± 0.002	0.012 ± 0.002	$\underline{\mathbf{0.012} \pm \mathbf{0.002}}$	0.028 ± 0.003	$\mathbf{0.003} \pm \mathbf{0.002}^*$	0.014 ± 0.015	$\mathbf{0.002} \pm \mathbf{0.001}^*$
CRS	0.105 ± 0.003	0.216 ± 0.004	0.053 ± 0.003	0.052 ± 0.004	$\underline{0.024 \pm 0.003}$	$\mathbf{0.004} \pm \mathbf{0.005}^*$	$\mathbf{0.012} \pm \mathbf{0.013}^*$	$\mathbf{0.003} \pm \mathbf{0.000}^*$
I-BIAS	Naive	EIB	IPS	SNIPS	DR	UDR	IDR	UIDR
ONE	0.069 ± 0.004	0.222 ± 0.003	$\underline{0.034 \pm 0.004}$	0.035 ± 0.005	0.049 ± 0.005	$\mathbf{0.031} \pm \mathbf{0.012}$	$\mathbf{0.010} \pm \mathbf{0.007}^*$	$\mathbf{0.004} \pm \mathbf{0.002}^*$
THREE	0.078 ± 0.003	0.234 ± 0.004	$\underline{0.038 \pm 0.003}$	0.039 ± 0.004	0.050 ± 0.004	$\mathbf{0.017} \pm \mathbf{0.013}^*$	$\mathbf{0.012} \pm \mathbf{0.020}^*$	$\mathbf{0.006} \pm \mathbf{0.003}^*$
FIVE	0.103 ± 0.004	0.245 ± 0.005	$\underline{0.050 \pm 0.004}$	0.052 ± 0.004	0.057 ± 0.004	$\mathbf{0.013} \pm \mathbf{0.012}^*$	$\mathbf{0.012} \pm \mathbf{0.008}^*$	$\mathbf{0.007} \pm \mathbf{0.003}^*$
ROTATE	0.138 ± 0.002	0.035 ± 0.001	0.070 ± 0.004	0.069 ± 0.003	$\underline{0.008 \pm 0.001}$	$\mathbf{0.002} \pm \mathbf{0.002}^*$	$\mathbf{0.001} \pm \mathbf{0.000}^*$	$\mathbf{0.002} \pm \mathbf{0.001}^*$
SKEW	0.025 ± 0.003	0.106 ± 0.001	$\underline{0.011 \pm 0.002}$	0.012 ± 0.003	0.028 ± 0.002	$\mathbf{0.009} \pm \mathbf{0.006}^*$	$\mathbf{0.009} \pm \mathbf{0.003}^*$	$\mathbf{0.004} \pm \mathbf{0.001}^*$
CRS	0.105 ± 0.004	0.216 ± 0.002	0.051 ± 0.003	0.052 ± 0.004	$\underline{\mathbf{0.024} \pm \mathbf{0.004}}$	0.031 ± 0.019	$\mathbf{0.008} \pm \mathbf{0.006}^*$	$\mathbf{0.006} \pm \mathbf{0.003}^*$
UI-BIAS	Naive	EIB	IPS	SNIPS	DR	UDR	IDR	UIDR
ONE	0.066 ± 0.001	0.445 ± 0.006	$\underline{\mathbf{0.031} \pm \mathbf{0.002}}$	0.032 ± 0.002	0.094 ± 0.003	0.062 ± 0.011	$\mathbf{0.025} \pm \mathbf{0.023}$	$\mathbf{0.007} \pm \mathbf{0.007}^*$
THREE	0.076 ± 0.002	0.470 ± 0.004	$\underline{\mathbf{0.036} \pm \mathbf{0.003}}$	$\mathbf{0.037} \pm \mathbf{0.003}$	0.099 ± 0.003	0.050 ± 0.032	0.062 ± 0.063	$\mathbf{0.009} \pm \mathbf{0.003}^*$
FIVE	0.099 ± 0.001	0.488 ± 0.003	$\underline{0.047 \pm 0.002}$	0.048 ± 0.002	0.109 ± 0.001	$\mathbf{0.037} \pm \mathbf{0.028}^*$	$\mathbf{0.013} \pm \mathbf{0.013}^*$	$\mathbf{0.009} \pm \mathbf{0.006}$
ROTATE	0.138 ± 0.001	0.071 ± 0.002	0.070 ± 0.002	0.069 ± 0.002	$\underline{0.015 \pm 0.001}$	$\mathbf{0.004} \pm \mathbf{0.004}^*$	$\mathbf{0.007} \pm \mathbf{0.010}^*$	$\mathbf{0.002} \pm \mathbf{0.001}^*$
SKEW	0.025 ± 0.001	0.214 ± 0.003	$\underline{\mathbf{0.011} \pm \mathbf{0.001}}$	$\mathbf{0.012} \pm \mathbf{0.001}$	0.056 ± 0.002	0.015 ± 0.011	0.027 ± 0.030	$\mathbf{0.012} \pm \mathbf{0.004}$
CRS	0.105 ± 0.003	0.432 ± 0.003	0.051 ± 0.003	0.052 ± 0.003	$\underline{0.048 \pm 0.002}$	$\mathbf{0.048} \pm \mathbf{0.035}$	$\mathbf{0.013} \pm \mathbf{0.007}^*$	$\mathbf{0.009} \pm \mathbf{0.008}^*$

Note: * means (p-value ≤ 0.05) using the paired-t-test compared with the best baseline. We bold the best three results and underline the best baseline result.

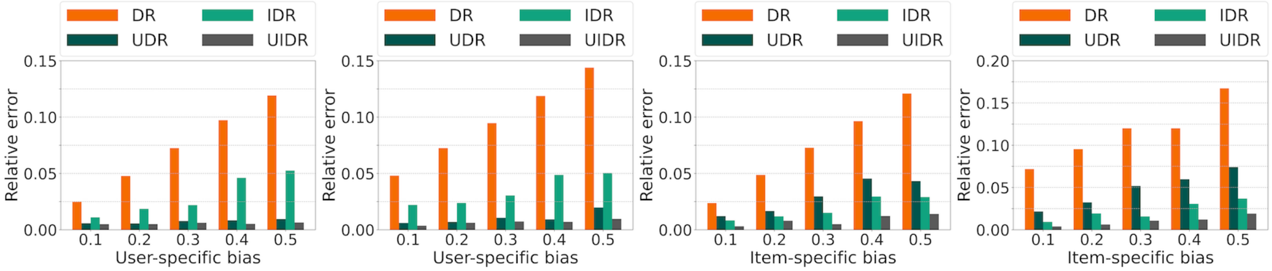


Figure 2. Relative error with varying inductive bias size. The left (right) two figures are user (item)-specific inductive bias scenarios with the item (user)-specific inductive bias fixed as 0 and 0.1.

By alternately implementing the above steps, the prediction model can achieve debiased learning under inaccurate pseudo-labelings. We summarize the alternating training process in the Algorithm 1.

4. Semi-Synthetic Experiments

Experiment Setup. We conduct semi-synthetic experiments using MOVIELENS 100K (ML-100K) dataset, which contains 100,000 ratings from 943 users for 1,682 items. We focus on two research questions below: (1) whether the proposed estimators are unbiased with the user-specific and item-specific inductive bias; (2) how the varying bias level affects the estimators' performance.

Experimental Details. We first generate the ground truth probability matrix $\mathbf{R}$, ground truth propensity matrix $\mathbf{P}$ and observation matrix $\mathbf{O}$ following the previous studies (Schnabel et al., 2016; Wang et al., 2019; Guo et al., 2021) (see Appendix C for the detailed generation process). To verify the effectiveness of the proposed estimators, we generate several $\hat{\mathbf{R}}$ based on $\mathbf{R}$ as follows:

- **ONE**: The predicted matrix $\hat{\mathbf{R}}$ is identical to the true matrix $\mathbf{R}$, except that randomly select $r_{u,i} = 0.1$ with total amount $|\{(u,i) \mid r_{u,i} = 0.9\}|$ are flipped to 0.9.
- **THREE**: Same as **ONE**, but flipping $r_{u,i} = 0.3$ instead.
- **FIVE**: Same as **ONE**, but flipping $r_{u,i} = 0.5$ instead.
- **ROTATE**: $\hat{r}_{u,i} = r_{u,i} - 0.2$ when $r_{u,i} \geq 0.3$, and $\hat{r}_{u,i} = 0.9$ when $r_{u,i} = 0.1$.
- **SKEW**: Predicted $\hat{r}_{u,i}$ are sampled from the Gaussian distribution $\mathcal{N}(\mu = r_{u,i}, \sigma = (1 - r_{u,i})/2)$, and clipped to the interval $[0.1, 0.9]$.
- **CRS**: $\hat{r}_{u,i} = 0.2$ if $r_{u,i} \leq 0.6$. Otherwise, $\hat{r}_{u,i} = 0.6$.

Following the previous studies (Guo et al., 2021; Dai et al., 2022), we estimate the inverse propensity by $1/\hat{\mathbf{P}} = (1 - \rho)/\mathbf{P} + \rho/p_e$, where $p_{u,i}$ is the ground truth propensity, $p_e = |\mathcal{D}|^{-1} \sum_{(u,i) \in \mathcal{D}} o_{u,i}$, and ρ is randomly sampled from the uniform distribution $U(0, 1)$ to introduce noises. Next, we simulate the biased pseudo-labelings $\tilde{r}_{u,i}$ for EIB, DR and the proposed estimators in three ways: (1) $\tilde{r}_{u,i} =$

Table 3. Performance on AUC, NDCG@K, and F1@K on the unbiased test set of Coat, Music and KuaiRec.

	COAT			MUSIC			KUAIREC		
Method	AUC	N@5	F1@5	AUC	N@5	F1@5	AUC	N@50	F1@50
MF	$0.680_{\pm0.006}$	$0.616_{\pm0.011}$	$0.470_{\pm0.006}$	$0.651_{\pm0.005}$	$0.626_{\pm0.001}$	$0.300_{\pm0.001}$	$0.741_{\pm0.003}$	$0.724_{\pm0.003}$	$0.566_{\pm0.002}$
IPS	$0.710_{\pm0.003}$	$0.603_{\pm0.009}$	$0.450_{\pm0.008}$	$0.656_{\pm0.002}$	$0.633_{\pm0.001}$	$0.308_{\pm0.001}$	$0.750_{\pm0.003}$	$0.734_{\pm0.003}$	$0.572_{\pm0.002}$
ASIPS	$0.712_{\pm0.008}$	$0.627_{\pm0.010}$	$0.470_{\pm0.007}$	$0.661_{\pm0.003}$	$0.641_{\pm0.004}$	$0.322_{\pm0.003}$	$0.746_{\pm0.009}$	$0.733_{\pm0.004}$	$0.585_{\pm0.006}$
DR	$0.710_{\pm0.006}$	$0.632_{\pm0.003}$	$0.471_{\pm0.003}$	$0.656_{\pm0.009}$	$0.669_{\pm0.007}$	$0.330_{\pm0.005}$	$0.745_{\pm0.004}$	$0.718_{\pm0.003}$	$0.574_{\pm0.003}$
DR-JL	$0.714_{\pm0.007}$	$0.646_{\pm0.009}$	$0.486_{\pm0.006}$	$0.682_{\pm0.001}$	$0.660_{\pm0.002}$	$0.326_{\pm0.001}$	$0.759_{\pm0.002}$	$0.757_{\pm0.004}$	$0.582_{\pm0.005}$
MRDR-JL	$0.715_{\pm0.004}$	$0.653_{\pm0.006}$	$0.492_{\pm0.005}$	$0.684_{\pm0.001}$	$0.645_{\pm0.001}$	$0.315_{\pm0.001}$	$0.762_{\pm0.003}$	$0.751_{\pm0.002}$	$0.579_{\pm0.003}$
CVIB	$0.718_{\pm0.004}$	$0.640_{\pm0.008}$	$0.486_{\pm0.008}$	$0.685_{\pm0.001}$	$0.647_{\pm0.003}$	$0.316_{\pm0.002}$	$0.758_{\pm0.001}$	$0.752_{\pm0.001}$	$0.575_{\pm0.001}$
DIB	$0.726_{\pm0.003}$	$0.628_{\pm0.008}$	$0.469_{\pm0.007}$	$0.690_{\pm0.002}$	$0.653_{\pm0.002}$	$0.320_{\pm0.001}$	$0.775_{\pm0.001}$	$0.760_{\pm0.001}$	$0.592_{\pm0.001}$
DR-MSE	$0.715_{\pm0.001}$	$0.630_{\pm0.009}$	$0.475_{\pm0.008}$	$0.685_{\pm0.001}$	$0.648_{\pm0.002}$	$0.316_{\pm0.002}$	$0.779_{\pm0.002}$	$0.773_{\pm0.002}$	$0.589_{\pm0.002}$
MR	$0.728_{\pm0.006}$	$0.655_{\pm0.009}$	$0.489_{\pm0.007}$	$0.698_{\pm0.004}$	$\underline{0.680}_{\pm0.004}$	$0.323_{\pm0.005}$	$0.776_{\pm0.002}$	$0.793_{\pm0.001}$	$0.599_{\pm0.002}$
TDR	$0.730_{\pm0.008}$	$0.651_{\pm0.013}$	$0.487_{\pm0.010}$	$0.694_{\pm0.002}$	$0.667_{\pm0.003}$	$\underline{0.337_{\pm0.002}}$	$0.792_{\pm0.004}$	$0.799_{\pm0.003}$	$0.604_{\pm0.003}$
TDR-JL	$0.729_{\pm0.005}$	$\underline{0.656}_{\pm0.011}$	$\underline{0.493}_{\pm0.010}$	$\underline{\mathbf{0.702}_{\pm0.002}}$	$0.672_{\pm0.004}$	$0.332_{\pm0.003}$	$\underline{0.793}_{\pm0.004}$	$\underline{0.799}_{\pm0.004}$	$0.603_{\pm0.003}$
Stable-DR	$0.719_{\pm0.006}$	$0.631_{\pm0.008}$	$0.475_{\pm0.006}$	$0.687_{\pm0.001}$	$0.650_{\pm0.003}$	$0.316_{\pm0.002}$	$0.764_{\pm0.003}$	$0.791_{\pm0.003}$	$0.595_{\pm0.002}$
ESMM	$0.686_{\pm0.004}$	$0.638_{\pm0.005}$	$0.485_{\pm0.004}$	$0.601_{\pm0.002}$	$0.665_{\pm0.003}$	$0.328_{\pm0.001}$	$0.721_{\pm0.006}$	$0.764_{\pm0.006}$	$0.576_{\pm0.004}$
Multi-IPS	$0.711_{\pm0.005}$	$0.604_{\pm0.005}$	$0.463_{\pm0.008}$	$0.651_{\pm0.005}$	$0.667_{\pm0.006}$	$0.331_{\pm0.004}$	$0.748_{\pm0.005}$	$0.738_{\pm0.008}$	$0.579_{\pm0.004}$
Multi-DR	$0.719_{\pm0.004}$	$0.634_{\pm0.009}$	$0.480_{\pm0.011}$	$0.686_{\pm0.002}$	$0.660_{\pm0.003}$	$0.323_{\pm0.002}$	$0.752_{\pm0.014}$	$0.767_{\pm0.012}$	$0.581_{\pm0.008}$
ESCM^2-IPS	$0.721_{\pm0.005}$	$0.645_{\pm0.006}$	$0.490_{\pm0.005}$	$0.680_{\pm0.002}$	$0.653_{\pm0.002}$	$0.322_{\pm0.002}$	$0.779_{\pm0.001}$	$0.767_{\pm0.001}$	$0.592_{\pm0.002}$
ESCM^2-DR	$\underline{\mathbf{0.730}}_{\pm0.009}$	$0.642_{\pm0.010}$	$0.489_{\pm0.009}$	$0.688_{\pm0.001}$	$0.669_{\pm0.002}$	$0.326_{\pm0.001}$	$0.788_{\pm0.001}$	$0.796_{\pm0.001}$	$\underline{0.606}_{\pm0.001}$
UDR (ours)	$\mathbf{0.739}^{*}{}_{\pm0.004}$	$\mathbf{0.676}^{*}{}_{\pm0.003}$	$\mathbf{0.521}^{*}{}_{\pm0.003}$	$\mathbf{0.705}^{*}{}_{\pm0.001}$	$\mathbf{0.766}^{*}{}_{\pm0.002}$	$\mathbf{0.389}^{*}{}_{\pm0.001}$	$\mathbf{0.802}^{*}{}_{\pm0.003}$	$\mathbf{0.804}^{*}{}_{\pm0.002}$	$\mathbf{0.610}^{*}{}_{\pm0.002}$
IDR (ours)	$0.721_{\pm0.002}$	$\mathbf{0.681}^{*}{}_{\pm0.003}$	$\mathbf{0.529}^{*}{}_{\pm0.005}$	$0.694_{\pm0.003}$	$\mathbf{0.747}^{*}{}_{\pm0.002}$	$\mathbf{0.378}^{*}{}_{\pm0.001}$	$\mathbf{0.801}^{*}{}_{\pm0.001}$	$\mathbf{0.803}^{*}{}_{\pm0.002}$	$\mathbf{0.607}_{\pm0.002}$
UIDR (ours)	$\mathbf{0.740}^{*}{}_{\pm0.006}$	$\mathbf{0.722}^{*}{}_{\pm0.006}$	$\mathbf{0.539}^{*}{}_{\pm0.005}$	$\mathbf{0.713}^{*}{}_{\pm0.001}$	$\mathbf{0.752}^{*}{}_{\pm0.001}$	$\mathbf{0.382}^{*}{}_{\pm0.001}$	$\mathbf{0.804}^{*}{}_{\pm0.004}$	$\mathbf{0.804}^{*}{}_{\pm0.004}$	$\mathbf{0.610}^{*}{}_{\pm0.003}$

Note: * means (p-value ≤ 0.05) using the paired-t-test compared with the best baseline. We bold the best three results and underline the best baseline result.

$r_{u,i} + b_u$ (user-specific inductive bias); (2) $\tilde{r}_{u,i} = r_{u,i} + b_i$ (item-specific inductive bias); (3) $\tilde{r}_{u,i} = r_{u,i} + b_u + b_i$ (user-item inductive bias), where b_u and b_i are randomly sampled from the uniform distribution $U(0, \nu)$, leading to the inaccurate pseudo-labelings. Finally, we use $r_{u,i} \in [0, 1]$ as the positive sample probabilities for Bernoulli sampling to obtain the binary outcome matrix. The absolute relative error (RE) is used for evaluation, which is defined as $\mathrm{RE}(\mathcal{L}_{\mathrm{est}}) = |\mathcal{L}_{\mathrm{ideal}}(\hat{\mathbf{R}}) - \mathcal{L}_{\mathrm{est}}(\hat{\mathbf{R}})|/\mathcal{L}_{\mathrm{ideal}}(\hat{\mathbf{R}})$, where $\mathcal{L}_{\mathrm{est}}$ denotes the loss of estimator to be compared. RE evaluates the accuracy of the estimated loss, the smaller the RE, the more accurate the estimation.

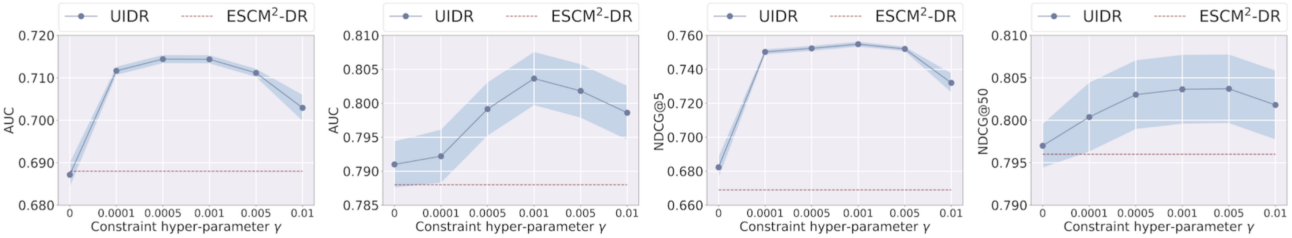
Performance Comparison. The experiment results with bias level $\nu = 0.2$ are shown in Table 2. First, the proposed estimators significantly outperform the baselines in all scenarios. Impressively, the proposed User-DR (Item-DR) estimator is still able to outperform the previous DR estimator in item-specific (user-specific) inductive bias scenarios. Second, the proposed User-DR (Item-DR) outperforms in the case of user-specific (item-specific) inductive bias, and the User-Item-DR estimator shows the most competing performance in all three scenarios, which further validates the ability of the proposed methods for eliminating the inductive bias. Meanwhile, since the IPS and SNIPS estimators are not affected by pseudo-labelings, they show competitive performance when increasing the inductive biases.

In addition, Figure 2 shows the experiment results with varying inductive bias levels by taking **ONE** for illustration, and similar results can be found in the other five prediction matrices. In the user-specific inductive bias scenario, as the bias level increases, both User-DR and User-Item-DR estimators demonstrate relatively stable performance. Meanwhile, the Item-DR estimator exhibits a slow increase in terms of RE compared with the DR estimator. Similar results are observed in the item-specific inductive bias scenario.

5. Real-World Experiments

Datasets and Evaluation Metrics. Three widely-used real-world datasets are adopted in our experiments, which are COAT, MUSIC, and KUAIREC (Gao et al., 2022). COAT contains ratings from 290 users to 300 items, with 6,960 biased ratings and 4,640 unbiased ratings in total. MUSIC contains ratings from 15,400 users to 1,000 items, with 311,704 biased ratings and 54,000 unbiased ratings. KUAIREC contains a fully exposed industrial dataset which contains 4,676,570 video watching ratio records from 1,411 users to 3,327 videos. Three widely-used evaluation metrics, namely AUC, NDCG@K, and F1@K, are used to evaluate performance, where K is set to 5 for COAT and MUSIC and K is set to 50 for KUAIREC (see Appendix D for more details).

Baselines. We take matrix factorization **(MF)** (Koren et al., 2009) as the base model, and compare with the following baselines: **IPS** (Schnabel et al., 2016; Saito et al., 2020), **ASIPS** (Saito, 2020), **DR** (Saito, 2020), **CVIB** (Wang et al., 2020b), **DIB** (Liu et al., 2021), **TDR** (Li et al., 2023b),



(a) AUC on Music (b) AUC on KuaiRec (c) N@5 on Music (d) N@50 on KuaiRec

Figure 3. Effects of varying hyper-parameter γ in UIDR loss.

Table 4. Ablation study on KuaiRec.

Method	$\mathcal{L}_{re}$	γ	AUC	N@50	F1@50
DR-JL	×	×	0.759	0.757	0.582
UDR w/o γ	✓	×	0.754	0.754	0.589
UDR w/o $\mathcal{L}_{re}$	×	✓	0.787	0.789	0.599
UDR	✓	✓	**0.802**	**0.804**	**0.610**

Table 5. In-depth Analysis for λ_u.

	Music			KuaiRec		
Method	AUC	N@5	F1@5	AUC	N@50	F1@50
Constant	0.695	0.739	0.373	0.797	0.796	0.605
MLP	0.696	0.743	0.376	0.803	0.801	0.605
Attention	**0.713**	**0.752**	**0.382**	**0.804**	**0.804**	**0.610**

DR-BIAS (Dai et al., 2022), **DR-MSE** (Dai et al., 2022), **Stable-DR** (Li et al., 2023d), and **MR** (Li et al., 2023a). We also consider the following baselines based on joint learning and multi-task learning: **DR-JL** (Wang et al., 2019), **MRDR-JL** (Guo et al., 2021), **TDR-JL** (Li et al., 2023b), **ESMM** (Ma et al., 2018), **Multi-IPS** (Zhang et al., 2020), **Multi-DR** (Zhang et al., 2020), **ESCM^2-IPS** (Wang et al., 2022a) and **ESCM^2-DR** (Wang et al., 2022a).

Performance Analysis. The results of proposed methods and baselines on Coat, Music and KuaiRec are shown in Table 3. First, almost all debiasing methods perform better than the base model (MF), which shows the necessity of debiasing in CF. Second, all three proposed methods significantly outperform all the baselines with a p-value less than 0.05, which is attributed to the more relaxed and realistic unbiasedness assumptions compared to the DR baselines. Finally, among the three proposed methods, Item-DR performs slightly worse than User-DR and User-Item-DR, which may be attributed to the inductive bias on the item side is less than that on the user side in practice.

Ablation Studies. We conduct the ablation study for User-DR on the large-scale industrial dataset KuaiRec with the results shown in Table 4. When only the reconstruction loss is retained, our propensity model will be the same as the base propensity model. Therefore, User-DR degenerates to the DR-JL. When only the constraint losses are retained, the reconstructed propensities will be much more distinct from the base propensities, leading to an increasing variance as shown in Theorem 2, which harms the performance.

In-Depth Analysis. We investigate the effect of hyper-parameter γ on the performance of User-Item-DR on the Music and KuaiRec datasets, and the results are shown in Figure 3. We can see that moderate constraints are the most helpful to trade-off between propensity estimation and estimator robustness and thus improve the debiasing performance. Meanwhile, the weight of each constraint is crucial throughout the learning process. Table 5 shows the impact of different models when learning λ_u on User-DR prediction performance. When we use the constant model to generalize all the λ_u, each user's constraint is equally important. As a result, the model will not utilize the user information effectively, which harms the performance. In addition, although MLP has a good fitting capacity, it is easy to allocate a higher weight to some specific constraints and to ignore other constraints, which also harms the performance. Therefore, when the attention mechanism is used for model training, it allows the model to adaptively trade off the fitting capacity and flexibility, which results in the desirable debiasing performance.

6. Conclusion

In this paper, we proposed the User-DR, Item-DR, and User-Item-DR estimators that can achieve unbiased learning even under inaccurate pseudo-labelings. The proposed estimators greatly relax the unbiasedness condition and improve the robustness of existing DR estimators to inaccurate pseudo-labelings. Our theoretical analysis shows that the variances of the proposed estimators are highly controllable and manageable. We further propose a propensity reconstruction learning to be compatible with the theory of the proposed estimators, which uses an attention mechanism that adaptively updates the weights of the proposed constraints. Extensive experiments are conducted to verify the validity of our proposal. A limitation of this work is the proposed learning method cannot guarantee that all the constraints in Eq. (1) and Eq. (2) hold strictly, due to the computational burden.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (623B2002, 62102420, 12301370).

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Aman Agarwal, Ivan Zaitsev, Xuanhui Wang, Cheng Li, and Thorsten Joachims. Estimating position bias without intrusive interventions. In *WSDM*, 2019.

Qingyao Ai, Keping Bi, Cheng Luo, Jiafeng Guo, and W Bruce Croft. Unbiased learning to rank with unbiased propensity estimation. In *SIGIR*, 2018.

Jiawei Chen, Can Wang, Martin Ester, Qihao Shi, Yan Feng, and Chun Chen. Social recommendation with missing not at random data. In *ICDM*, 2018.

Jiawei Chen, Hande Dong, Yang Qiu, Xiangnan He, Xin Xin, Liang Chen, Guli Lin, and Keping Yang. AutoDebias: Learning to debias for recommendation. In *SIGIR*, 2021.

Quanyu Dai, Haoxuan Li, Peng Wu, Zhenhua Dong, Xiao-Hua Zhou, Rui Zhang, Xiuqiang He, Rui Zhang, and Jie Sun. A generalized doubly robust learning framework for debiasing post-click conversion rate prediction. In *KDD*, 2022.

Chongming Gao, Shijun Li, Wenqiang Lei, Jiawei Chen, Biao Li, Peng Jiang, Xiangnan He, Jiaxin Mao, and Tat-Seng Chua. KuaiRec: A fully-observed dataset and insights for evaluating recommender systems. In *CIKM*, 2022.

Siyuan Guo, Lixin Zou, Yiding Liu, Wenwen Ye, Suqi Cheng, Shuaiqiang Wang, Hechang Chen, Dawei Yin, and Yi Chang. Enhanced doubly robust learning for debiasing post-click conversion rate estimation. In *SIGIR*, 2021.

José Miguel Hernández-Lobato, Neil Houlsby, and Zoubin Ghahramani. Probabilistic matrix factorization with non-random missing data. In *ICML*, 2014.

Shanshan Huang, Haoxuan Li, Qingsong Li, Chunyuan Zheng, and Li Liu. Pareto invariant representation learning for multimedia recommendation. In *ACM-MM*, 2023.

Kosuke Imai and Marc Ratkovic. Covariate balancing propensity score. *Journal of the Royal Statistical Society (Series B)*, 76(1):243–263, 2014.

Joseph D.Y. Kang and Joseph L. Schafer. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22:523–539, 2007.

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

Karl Krauth, Yixin Wang, and Michael I Jordan. Breaking feedback loops in recommender systems with causal inference. *arXiv preprint arXiv:2207.01616*, 2022.

Chen Li, Yang Cao, Ye Zhu, Debo Cheng, Chengyuan Li, and Yasuhiko Morimoto. Ripple knowledge graph convolutional networks for recommendation systems. *Machine Intelligence Research*, pages 1–14, 2024.

Haoxuan Li, Quanyu Dai, Yuru Li, Yan Lyu, Zhenhua Dong, Xiao-Hua Zhou, and Peng Wu. Multiple robust learning for recommendation. In *AAAI*, 2023a.

Haoxuan Li, Yan Lyu, Chunyuan Zheng, and Peng Wu. TDR-CL: Targeted doubly robust collaborative learning for debiased recommendations. In *ICLR*, 2023b.

Haoxuan Li, Yanghao Xiao, Chunyuan Zheng, Peng Wu, and Peng Cui. Propensity matters: Measuring and enhancing balancing for recommendation. In *ICML*, 2023c.

Haoxuan Li, Chunyuan Zheng, and Peng Wu. StableDR: Stabilized doubly robust learning for recommendation on data missing not at random. In *ICLR*, 2023d.

Dugang Liu, Pengxiang Cheng, Hong Zhu, Zhenhua Dong, Xiuqiang He, Weike Pan, and Zhong Ming. Mitigating confounding bias in recommendation via information bottleneck. In *RecSys*, 2021.

Yiming Liu, Xuezhi Cao, and Yong Yu. Are you influenced by others when rating? improve rating prediction by conformity modeling. In *RecSys*, 2016.

Jinwei Luo, Dugang Liu, Weike Pan, and Zhong Ming. Unbiased recommendation model based on improved propensity score estimation. *Journal of Computer Applications*, 2021.

Zheqi Lv, Wenqiao Zhang, Shengyu Zhang, Kun Kuang, Feng Wang, Yongwei Wang, Zhengyu Chen, Tao Shen, Hongxia Yang, Beng Chin Ooi, et al. Duet: A tuning-free device-cloud collaborative parameters generation framework for efficient device model generalization. In *WWW*, 2023.

Zheqi Lv, Wenqiao Zhang, Zhengyu Chen, Shengyu Zhang, and Kun Kuang. Intelligent model update strategy for sequential recommendation. In *WWW*, 2024.

Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu, Xiaoqiang Zhu, and Kun Gai. Entire space multi-task model: An effective approach for estimating post-click conversion rate. In *SIGIR*, 2018.

Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, and Bamshad Mobasher. Feedback loop and bias amplification in recommender systems. In *WWW*, 2020.

Benjamin Marlin, Richard S Zemel, Sam Roweis, and Malcolm Slaney. Collaborative filtering and the missing at random assumption. *UAI*, 2007.

Geert Molenberghs, Garrett Fitzmaurice, Michael G. Kenward, Anastasios Tsiatis, and Geert Verbeke. *Handbook of Missing Data Methodology*. Chapman & Hall/CRC, 2015.

Harrie Oosterhuis. Doubly robust estimation for correcting position bias in click feedback for unbiased learning to rank. *ACM Transactions on Information Systems*, 2023.

Yong Rui, Vicente Ivan Sanchez Carmona, Mohsen Pourvali, Yun Xing, Wei-Wen Yi, Hui-Bin Ruan, and Yu Zhang. Knowledge mining: A cross-disciplinary survey. *Machine Intelligence Research*, 19(2):89–114, 2022.

Yuta Saito. Asymmetric tri-training for debiasing missing-not-at-random explicit feedback. In *SIGIR*, 2020.

Yuta Saito. Doubly robust estimator for ranking metrics with post-click conversions. In *RecSys*, 2020.

Yuta Saito and Masahiro Nomura. Towards resolving propensity contradiction in offline recommender learning. In *IJCAI*, 2022.

Yuta Saito, Hayato Sakata, and Kazuhide Nakata. Doubly robust prediction and evaluation methods improve uplift modeling for observational data. In *SIAM*, 2019.

Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. Unbiased recommender learning from missing-not-at-random implicit feedback. In *WSDM*, 2020.

Masahiro Sato, Janmajay Singh, Sho Takemori, Takashi Sonoda, Qian Zhang, and Tomoko Ohkuma. Uplift-based evaluation and optimization of recommenders. In *RecSys*, 2019.

Masahiro Sato, Sho Takemori, Janmajay Singh, and Tomoko Ohkuma. Unbiased learning for the causal effect of recommendation. In *RecSys*, 2020.

Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. Recommendations as treatments: Debiasing learning and evaluation. In *ICML*, 2016.

Shaun R. Seaman and Stijn Vansteelandt. Introduction to double robust methods for incomplete data. *Statistical Science*, 33:184–197, 2018.

Zijie Song, Jiawei Chen, Sheng Zhou, Qihao Shi, Yan Feng, Chun Chen, and Can Wang. CDR: Conservative doubly robust learning for debiased recommendation. In *CIKM*, 2023.

Harald Steck. Evaluation of recommendations: rating-prediction and ranking. In *RecSys*, 2013.

Zhiqiang Tan. Comment: understanding OR, PS and DR. *Statistical Science*, 22:560–568, 2007.

Karel Vermeulen and Stijn Vansteelandt. Bias-reduced doubly robust estimation. *Journal of the American Statistical Association*, 110:1024–1036, 2015.

Hao Wang, Tai-Wei Chang, Tianqiao Liu, Jianmin Huang, Zhichao Chen, Chao Yu, Ruopeng Li, and Wei Chu. ESCM^2: Entire space counterfactual multi-task model for post-click conversion rate estimation. In *SIGIR*, 2022a.

Hao Wang, Jiajun Fan, Zhichao Chen, Haoxuan Li, Weiming Liu, Tianqiao Liu, Quanyu Dai, Yichao Wang, Zhenhua Dong, and Ruiming Tang. Optimal transport for treatment effect estimation. In *NeurIPS*, 2023a.

Haotian Wang, Wenjing Yang, Longqi Yang, Anpeng Wu, Liyang Xu, Jing Ren, Fei Wu, and Kun Kuang. Estimating individualized causal effect with confounded instruments. In *KDD*, 2022b.

Haotian Wang, Kun Kuang, Haoang Chi, Longqi Yang, Mingyang Geng, Wanrong Huang, and Wenjing Yang. Treatment effect estimation with adjustment feature selection. In *KDD*, 2023b.

Haotian Wang, Kun Kuang, Long Lan, Zige Wang, Wanrong Huang, Fei Wu, and Wenjing Yang. Out-of-distribution generalization with causal feature separation. *IEEE Transactions on Knowledge and Data Engineering*, 36(4):1758–1772, 2024.

Jun Wang, Haoxuan Li, Chi Zhang, Dongxu Liang, Enyun Yu, Wenwu Ou, and Wenjia Wang. CounterCLR: Counterfactual contrastive learning with non-random missing data in recommendation. In *ICDM*, 2023c.

Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. Doubly robust joint learning for recommendation on data missing not at random. In *ICML*, 2019.

Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. Combating selection biases in recommender systems with a few unbiased ratings. In *WSDM*, 2021.

Yixin Wang, Dawen Liang, Laurent Charlin, and David M. Blei. Causal inference for recommender systems. In *RecSys*, 2020a.

Zifeng Wang, Xi Chen, Rui Wen, Shao-Lun Huang, Ercan E Kuruoglu, and Yefeng Zheng. Information theoretic counterfactual learning from missing-not-at-random feedback. In *NeurIPS*, 2020b.

Tianxin Wei, Fuli Feng, Jiawei Chen, Ziwei Wu, Jinfeng Yi, and Xiangnan He. Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system. In *SIGKDD*, 2021.

Hongyi Wen, Xinyang Yi, Tiansheng Yao, Jiaxi Tang, Lichan Hong, and Ed H. Chi. Distributionally-robust recommendations for improving worst-case user experience. In *WWW*, 2022.

Mengyue Yang, Quanyu Dai, Zhenhua Dong, Xu Chen, Xiuqiang He, and Jun Wang. Top-n recommendation with counterfactual user preference simulation. In *CIKM*, 2021.

Mengyue Yang, Guohao Cai, Furui Liu, Jiarui Jin, Zhenhua Dong, Xiuqiang He, Jianye Hao, Weiqi Shao, Jun Wang, and Xu Chen. Debiased recommendation with user feature balancing. *ACM Transactions on Information Systems*, 41(4):1–25, 2023.

Honglei Zhang, Shuyi Wang, Haoxuan Li, Chunyuan Zheng, Xu Chen, Li Liu, Shanshan Luo, and Peng Wu. Uncovering the propensity identification problem in debiased recommendations. In *ICDE*, 2024.

Min Zhang, Junkun Yuan, Yue He, Wenbin Li, Zhengyu Chen, and Kun Kuang. Map: Towards balanced generalization of iid and ood through model-agnostic adapters. In *ICCV*, 2023.

Wenhao Zhang, Wentian Bao, Xiao-Yang Liu, Keping Yang, Quan Lin, Hong Wen, and Ramin Ramezani. Large-scale causal approaches to debiasing post-click conversion rate estimation with multi-task learning. In *WWW*, 2020.

Yang Zhang, Fuli Feng, Xiangnan He, Tianxin Wei, Chonggang Song, Guohui Ling, and Yongdong Zhang. Causal intervention for leveraging popularity bias in recommendation. In *SIGIR*, 2021.

Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Depeng Jin, and Yong Li. Disentangling user interest and conformity for recommendation with causal embedding. In *WWW*, 2021.

A. Related Works

Collaborative filtering (CF) plays an important role in today's digital and informative world (Chen et al., 2018; Huang et al., 2023; Lv et al., 2023; 2024). However, the collected data is observational rather than experimental, leading to the selection bias, which causes the distribution of the training data to be different from the distribution of the test data, thus making it challenging to achieve unbiased estimation and learning (Wang et al., 2022b; 2023b; Zhang et al., 2023; Wang et al., 2023a; 2024). Error imputation-based (EIB) methods first learn pseudo-labelings for missing events and train the prediction model with both pseudo-labelings and observed labels (Marlin et al., 2007; Steck, 2013; Hernández-Lobato et al., 2014). However, the unbiasedness EIB requires that pseudo-labelings are accurate for all user-item pairs. By introducing an additional propensity model, doubly robust (DR) method is proposed to improve EIB (Wang et al., 2019), with many enhanced DR approaches developed (Guo et al., 2021; Zhang et al., 2020; Chen et al., 2021; Wang et al., 2021; 2022a; Oosterhuis, 2023). The advantage of DR estimators is attributed to the property of double robustness, i.e., it is unbiased if either the pseudo-labelings or the learned propensities are accurate.

Despite the double robustness providing additional protection against inaccurate pseudo-labelings, i.e., unbiasedness holds when either the learned propensities or the pseudo-labelings are accurate for all user-item pairs, recent studies have shown that DR methods are highly sensitive to inaccurate pseudo-labelings, i.e., when the learned propensities are slightly inaccurate, the DR estimator can be severely biased with inaccurate pseudo-labelings (Kang and Schafer, 2007; Molenberghs et al., 2015; Vermeulen and Vansteelandt, 2015; Seaman and Vansteelandt, 2018). Thus, it can be summarized that the effectiveness of both the EIB and DR methods rely heavily on accurate pseudo-labelings.

Unfortunately, obtaining such accurate pseudo-labelings for all user-item pairs is usually impractical in practice, as user-item interactions are influenced by various factors, such as user self-selection (Ma et al., 2018; Luo et al., 2021), user conformity (Liu et al., 2016; Zheng et al., 2021), item popularity (Zhang et al., 2021; Wei et al., 2021), and item exposure position (Ai et al., 2018; Agarwal et al., 2019), causing inaccurate pseudo-labelings in practice. Motivated by this, we theoretically proposes several novel DR estimators and learning approach that achieve unbiasedness even with inaccurate pseudo-labelings, which greatly relaxes the unbiasedness condition of the doubly robust estimators.

B. Proof of Theorem 2

Theorem 2. *If $1/L \leq \hat{p}^2_{u,i}/\tilde{p}^2_{u,i} \leq L$ for a constant L,*

$$\frac{1}{L} \cdot \mathbb{V}(\mathcal{L}_{\mathrm{DR}}(\theta)) \leq \mathbb{V}(\mathcal{L}_{\mathrm{UDR}}(\theta)) \leq L \cdot \mathbb{V}(\mathcal{L}_{\mathrm{DR}}(\theta)).$$

Proof. As shown in Guo et al. (2021); Dai et al. (2022),

$$\mathbb{V}(\mathcal{L}_{\mathrm{DR}}(\theta)) = \frac{1}{|\mathcal{D}|^2} \sum_{(u,i)\in\mathcal{D}} \frac{p_{u,i}(1-p_{u,i})(\delta_{u,i}-\hat{\delta}_{u,i})^2}{\hat{p}^2_{u,i}}.$$

The variance of UDR has the same form of DR but replaces $\hat{p}_{u,i}$ with $\tilde{p}_{u,i}$ satisfying Eq. (1),

$$\mathbb{V}(\mathcal{L}_{\mathrm{UDR}}(\theta)) = \frac{1}{|\mathcal{D}|^2} \sum_{(u,i)\in\mathcal{D}} \frac{p_{u,i}(1-p_{u,i})(\delta_{u,i}-\hat{\delta}_{u,i})^2}{\tilde{p}^2_{u,i}}.$$

Now, it is clear that if $1/L \leq \hat{p}^2_{u,i}/\tilde{p}^2_{u,i} \leq L$ for a constant L, then we have

$$\frac{1}{L} \leq \frac{\mathbb{V}(\mathcal{L}_{\mathrm{UDR}}(\theta))}{\mathbb{V}(\mathcal{L}_{\mathrm{DR}}(\theta))} = \frac{\hat{p}^2_{u,i}}{\tilde{p}^2_{u,i}} \leq L.$$

□

C. More Details about Semi-Synthetic Experiments

Data Preprocessing. Following the previous studies (Schnabel et al., 2016; Wang et al., 2019; Guo et al., 2021), the detailed preprocessing for the ML-100K[3] dataset is shown as follows.

[3] https://grouplens.org/datasets/movielens/100k/

(1) Complete the full rating matrix using Matrix Factorization (MF) (Koren et al., 2009). Since the rating matrix completed by MF will have an unrealistic high prediction value for ratings of almost all user-item pairs, we adjust the proportion of ratings to match a more realistic rating distribution by first sorting all ratings in ascending order, then set ratings below the p_1 quantile to 1, set ratings between p_1 quantile and p_2 quantile to 2, and so on. The adjusted rating matrix contains $R_{u,i} \in \{1, 2, 3, 4, 5\}$ with proportion $[p_1, p_2, p_3, p_4, p_5]$, respectively (Schnabel et al., 2016; Guo et al., 2021).

(2) Set a propensity $p_{u,i} \in (0, 1)$ for each user-item pair with $p_{u,i} = p\alpha^{\min(4, 6-R_{u,i})}$. In our experiment, $p = 1$ and $\alpha = 0.5$ (Wang et al., 2019; Guo et al., 2021). Then we obtain the ground truth propensity matrix $\mathbf{P}$.

(3) Transfer the adjusted rating matrix to the probability matrix by replacing $R_{u,i} \in \{1, 2, 3, 4, 5\}$ with $r_{u,i} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ correspondingly. Because only binary click indicators can be observed, we sample click indicators according to the following Bernoulli distribution:

$$o_{u,i} \sim \mathrm{Bern}(p_{u,i}), \forall (u, i) \in \mathcal{D},$$

where $\mathrm{Bern}(\cdot)$ denotes the Bernoulli distribution. Then we obtain a fully observed observation matrix $\mathbf{O}$ and a ground truth probability matrix $\mathbf{R}$.

D. Dataset and Preprocessing, Experimental Protocols and Details

Dataset and Preprocessing. Following the previous studies (Wang et al., 2019; 2021; Chen et al., 2021), we conduct extensive experiments on three real-world datasets: COAT[4], MUSIC[5], and KUAIREC[6] (Gao et al., 2022). COAT dataset contains 290 users and 300 items with 6,960 biased ratings and 4,640 unbiased ratings. MUSIC dataset contains 15,400 users and 1,000 items with 311,704 biased ratings and 54,000 unbiased ratings. COAT and MUSIC are both five-scale datasets, and we binarize ratings less than three to 0, otherwise to 1. KUAIREC is a large-scale fully exposed industrial dataset collected from a short video sharing platform, which contains 4,676,570 video watching ratios from 1,411 users to 3,327 videos. For KUAIREC dataset, we binarize the video watching ratios less than one to 0, otherwise to 1.

Experimental Protocols and Details. We use three widely adopted evaluation metrics, AUC, NDCG@K, and F1@K, where K is set to 5 for COAT and MUSIC and 50 for KUAIREC. All the experiments are implemented on PyTorch with Adam as the optimizer. For all experiments, we use GeForce RTX 3090 as the computing resource. Logistic regression is used as the propensity model for all the methods with propensity. We tune the learning rate in $\{0.001, 0.005, 0.01, 0.05, 0.1\}$ and the batch size in $\{32, 64, 128, 256\}$ for COAT and $\{1024, 2048, 4096, 8192\}$ for MUSIC and KUAIREC. We tune the embedding dimension in $\{2, 4, 8, 16, 32, 64\}$ for COAT and $\{16, 32, 64, 128, 256, 512\}$ for MUSIC and KUAIREC. Moreover, we tune the hyper-parameter γ in $\{1e-6, 5e-5, 1e-5, ..., 1e-1\}$.

[4] https://www.cs.cornell.edu/˜schnabts/mnar/

[5] http://webscope.sandbox.yahoo.com/

[6] https://github.com/chongminggao/KuaiRec