A Generalized Doubly Robust Learning Framework for Debiasing Post-Click Conversion Rate Prediction

Quanyu Dai
Huawei Noah's Ark Lab
daiquanyu@huawei.com

Haoxuan Li
Peking University
hxli@stu.pku.edu.cn

Peng Wu*
Beijing Technology and Business University
wupeng@bicmr.pku.edu.cn

Zhenhua Dong
Huawei Noah's Ark Lab
dongzhenhua@huawei.com

Xiao-Hua Zhou*
Peking University
azhou@bicmr.pku.edu.cn

Rui Zhang
www.ruizhang.info
rayteam@yeah.net

Rui Zhang
Huawei Hong Kong Theory Lab
zhangrui191@huawei.com

Jie Sun
Huawei Hong Kong Theory Lab
j.sun@huawei.com

ABSTRACT

Post-click conversion rate (CVR) prediction is an essential task for discovering user interests and increasing platform revenues in a range of industrial applications. One of the most challenging problems of this task is the existence of severe selection bias caused by the inherent self-selection behavior of users and the item selection process of systems. Currently, doubly robust (DR) learning approaches achieve the state-of-the-art performance for debiasing CVR prediction. However, in this paper, by theoretically analyzing the bias, variance and generalization bounds of DR methods, we find that existing DR approaches may have poor generalization caused by inaccurate estimation of propensity scores and imputation errors, which often occur in practice. Motivated by such analysis, we propose a generalized learning framework that not only unifies existing DR methods, but also provides a valuable opportunity to develop a series of new debiasing techniques to accommodate different application scenarios. Based on the framework, we propose two new DR methods, namely DR-BIAS and DR-MSE. DR-BIAS directly controls the bias of DR loss, while DR-MSE balances the bias and variance flexibly, which achieves better generalization performance. In addition, we propose a novel tri-level joint learning optimization method for DR-MSE in CVR prediction, and an efficient training algorithm correspondingly. We conduct extensive experiments on both real-world and semi-synthetic datasets, which validate the effectiveness of our proposed methods.

CCS CONCEPTS

• **Information systems** → **Recommender systems**.

KEYWORDS

Recommender Systems; Post-click Conversion Rate; Selection Bias; Doubly Robust Learning

ACM Reference Format:
Quanyu Dai, Haoxuan Li, Peng Wu*, Zhenhua Dong, Xiao-Hua Zhou*, Rui Zhang, Rui Zhang, and Jie Sun. 2022. A Generalized Doubly Robust Learning Framework for Debiasing Post-Click Conversion Rate Prediction. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22), August 14–18, 2022, Washington, DC, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3534678.3539270

1 INTRODUCTION

The post-click conversion rate (CVR) prediction has gained much attention in modern recommender systems [10, 19, 24, 37, 38], as post-click conversion feedback contains strong signals of user preference and directly contributes to the gross merchandise volume (GMV). In many industrial applications, CVR prediction is commonly regarded as the central task for discovering user interests and increasing platform revenues. For a user-item pair, CVR represents the probability of the user consuming the item after he/she clicks it. Essentially, the task of CVR prediction is a **counterfactual** problem. This is because what we want to know during inference is intrinsically the conversion rates of all user-item pairs under the assumption that all items are clicked by all users, which is a hypothetical situation that contradicts reality.

Most of the literature treats CVR prediction as a missing data problem in which the conversion labels are observed in clicked events and missing in unclicked events. A conventional and natural strategy is to train the CVR model only based on clicked events and then predict CVR for all the events [18, 29]. However, this estimator is biased and often obtains a sub-optimal result due to the existence of severe selection bias [10, 19]. In addition, the data sparsity issue, namely, the sample size of clicked events being much smaller than that of unclicked events, will amplify the difference between these two types of events and thus aggravate the selection bias issue.

Several approaches have been proposed to derive unbiased estimators of CVR by dealing with selection bias. Error imputation [4] and inverse propensity score (IPS) weighting [27, 39] are two main

[*] Peng Wu and Xiao-Hua Zhou are the corresponding authors.

KDD '22, August 14–18, 2022, Washington, DC, USA.
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9385-0/22/08…$15.00
https://doi.org/10.1145/3534678.3539270

strategies for debiasing CVR prediction. In addition, Doubly robust (DR) estimators can be constructed by combining EIB and IPS approaches [24, 37, 39]. A DR estimator enjoys the property of double robustness, which guarantees the unbiased estimation of CVR if either the imputed errors or propensity scores are accurate. Compared with EIB and IPS methods, the DR method has a better performance in general [32].

There are still some concerns for DR methods, even though they usually compare favorably with EIB and IPS estimators. Theoretical analysis of DR estimators in Section 3.1 shows that the bias, variance and generalization bounds all depend on the error deviation of the imputation model weighted by the inverse of propensity score. This is a worrying result, because the inverse of propensity score tends to be large in unclicked events and error deviations of the imputation model are most likely to be inaccurate in unclicked events due to the selection bias and data sparsity. It indicates that the bias, variance and generalization bounds may still be large under inaccurate imputed errors in unclicked events. Recently, several approaches, mainly including doubly robust joint learning (DR-JL) [32] and more robust doubly robust (MRDR) [10], have been designed to alleviate this problem. MRDR aims to reduce the variance of DR loss to enhance model robustness, but it may still have poor generalization when the bias is large. DR-JL attempts to reduce the error deviation of the imputation model in order to obtain a more accurate estimator of CVR, but this method does not control the bias and variance directly. Therefore, it would be helpful if we could find a more effective way to control the bias and variance directly.

In this paper, we reveal the counterfactual issues behind the CVR prediction task and give a formal and strict causal definition of CVR. Then, by analyzing the bias, variance and generalization bound of the DR estimator, we derive a novel generalized learning framework that can accommodate a wide range of CVR estimators through specifying different metrics of loss functions. This framework unifies various existing doubly robust methods for debiasing CVR prediction, such as DR-JL and MRDR. Most importantly, it provides key insights for designing new estimators to accommodate different application scenarios in CVR prediction. Based on this framework, from a perspective of bias-variance trade-off, we propose two new doubly robust estimators, called **DR-BIAS** and **DR-MSE**, which are designed to more flexibly control the bias and mean squared error (MSE) of DR loss function, respectively. DR-MSE achieves better generalization performance based on our analysis compared with existing DR based methods. In addition, we propose a novel tri-level joint learning optimization method for flexible DR-MSE in CVR prediction, and an efficient training algorithm correspondingly. Extensive experiments are carried out to validate the advantages of the proposed methods compared with state-of-the-art techniques. DR-MSE outperforms them up to 3.22% in DCG@2 in our experiments.

The main contributions of this paper can be summarized as follows: (1) We propose a generalized framework of doubly robust learning, which not only unifies the existing DR methods, but also provides key insights for designing new estimators with different requirements to accommodate different application scenarios. (2) Based on the proposed framework, we design two new doubly robust methods, called DR-BIAS and DR-MSE, which can better control the bias and mean squared error, compared with existing methods. (3) For the bias-variance tradeoff parameter of DR-MSE, we propose a tri-level DR-MSE joint learning optimization for the CVR prediction task, and an efficient training algorithm correspondingly. (4) Experimental results on both **real-world** and **semi-synthetic** datasets show that the two proposed doubly robust methods outperform the state-of-the-art methods significantly. Especially, both datasets with missing-at-random ratings and large industrial dataset are used for comprehensive evaluation.

2 PRELIMINARIES

In this section, we uncover the counterfactual feature of CVR prediction task within the potential outcome framework [11, 23], and discuss some existing approaches for CVR prediction.

2.1 Causal Problem Definition

Notation is described as follows. Let $\mathcal{U} = \{1, 2, ..., m\}$ and $\mathcal{I} = \{1, 2, ..., n\}$ be the sets of m users and n items, respectively, and $\mathcal{D} = \mathcal{U} \times \mathcal{I}$ be the set of all user-item pairs. Let $x_{u,i}$ be the feature vector of user u and item i, and $r_{u,i} \in \{0, 1\}$ be the indicator of the observed conversion label. Let $o_{u,i}$ be the indicator of a click event, i.e., $o_{u,i} = 1$ if user u clicks item i, $o_{u,i} = 0$ otherwise. Then, $\mathcal{O} = \{(u, i) \mid (u, i) \in \mathcal{D}, o_{u,i} = 1\}$ denotes all the clicked events.

For any user-item pair (u, i), we are interested in predicting the CVR **if** user u had clicked item i. Notice in particular that the word "**if**" is **counterfactual**. Specifically, in the real world, each user clicks only some items and many items have never been clicked by some users, but what we want to know is the conversion rates of all the user-item pairs when each user clicks all items, which is a hypothetical situation that contradicts reality.

Potential outcome is a basic tool to delineate counterfactual quantity in causal inference [11]. Through it, the task of predicting CVR can be defined formally. Concretely, we treat $o_{u,i}$ as a treatment (or an intervention) and define the potential conversion label $r_{u,i}(1)$, which represents the conversion label of a user u on an item i if the item is clicked by the user. Correspondingly, $r_{u,i}(0)$ is defined as the conversion label if the user u did not click the item i. Then the CVR can be fundamentally defined as

$$\mathbb{P}(r_{u,i}(1) = 1 \mid X_{u,i} = x_{u,i}), \tag{1}$$

which is a causal definition and it is coherent and consistent with the practical implications of CVR in recommender systems. In comparison, the conventional definition of CVR (see [19]), defined by $\mathbb{P}(r_{u,i} = 1 \mid X_{u,i} = x_{u,i}, o_{u,i} = 1)$, is based on association (or correlation) and lost the meaning of "counterfactual".

For estimating CVR in Equation (1), a fundamental challenge is that only one of the potential outcome $(r_{u,i}(1), r_{u,i}(0))$ is observable. By consistency assumption, $r_{u,i}(1)$ is observed when $o_{u,i} = 1$, missing otherwise. Therefore, the goal of estimating CVR can be recast into a missing data problem.

For ease of presentation, we denote $\mathbf{R} \in \{0, 1\}^{m \times n}$ as the full potential conversion label matrix with each element being $r_{u,i}(1)$, and let $\mathbf{R}^o = \{r_{u,i}(1) \mid (u, i) \in \mathcal{O}\} = \{r_{u,i} \mid (u, i) \in \mathcal{O}\}$ be the set consisting of potential conversion labels $r_{u,i}(1)$ in clicked events. Let $\hat{\mathbf{R}} \in [0, 1]^{m \times n}$ be the predicted conversion rate matrix, where each entry $\hat{r}_{u,i}(1) \in [0, 1]$ denotes the predicted conversion rate obtained by a model $f_\phi(x_{u,i})$ with parameters ϕ. If the full potential conversion label matrix $\mathbf{R}$ was observed, the ideal loss function is

$$\mathcal{L}_{ideal}(\hat{\mathbf{R}}, \mathbf{R}) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} e_{u,i}, \tag{2}$$

where $e_{u,i}$ is the prediction error. In this paper, we employ the cross entropy loss $e_{u,i} = -r_{u,i}(1)\log\{\hat{r}_{u,i}(1)\} - \{1 - r_{u,i}(1)\}\log\{1 - \hat{r}_{u,i}(1)\}$. $\mathcal{L}_{ideal}(\hat{\mathbf{R}}, \mathbf{R})$ can be regarded as a benchmark of unbiased loss function theoretically, even though it is infeasible due to the inaccessibility of $\mathbf{R}$ practically.

2.2 Existing Methods

A direct method is to use the following loss function $\mathcal{L}_{naive}(\hat{\mathbf{R}}, \mathbf{R}^o) = |\mathcal{O}|^{-1}\sum_{(u,i)\in\mathcal{O}} e_{u,i}$ based on the observed conversion labels $\mathbf{R}^o$. $\mathcal{L}_{naive}(\hat{\mathbf{R}}, \mathbf{R}^o)$ is not an unbiased estimate of $\mathcal{L}_{ideal}(\hat{\mathbf{R}}, \mathbf{R})$. Next, we will briefly review some typical and latest methods for addressing the selection bias issue.

2.2.1 Error Imputation Based Estimator. The error imputation based (EIB) estimator can be derived by introducing an error imputation model $\hat{e}_{u,i} = g_\theta(x_{u,i})$ to fit the prediction error $e_{u,i}$. Given the imputed errors, the loss function of EIB method is given as $\mathcal{L}_{EIB}(\hat{\mathbf{R}}, \mathbf{R}^o) = |\mathcal{D}|^{-1}\sum_{(u,i)\in\mathcal{D}}[o_{u,i}e_{u,i} + (1 - o_{u,i})\hat{e}_{u,i}]$.

2.2.2 Inverse Propensity Score Estimator. The inverse propensity score (IPS) approach [27] aims to recover the distribution of all events by weighting the clicked events with $1/p_{u,i}$, where $p_{u,i} = \mathbb{P}(o_{u,i} = 1) = \mathbb{E}[o_{u,i}]$ is the propensity score [22]. Given the estimate of $p_{u,i}$, denoted as $\hat{p}_{u,i}$, the loss function of IPS estimator is presented as $\mathcal{L}_{IPS}(\hat{\mathbf{R}}, \mathbf{R}^o) = |\mathcal{D}|^{-1}\sum_{(u,i)\in\mathcal{D}} o_{u,i}e_{u,i}/\hat{p}_{u,i}$.

2.2.3 Doubly Robust Joint Learning Estimator. Doubly robust (DR) estimator can be constructed in the augmented IPS form [2, 32] by combining EIB and IPS methods. Given the learned propensities $\hat{\mathbf{P}} = \{\hat{p}_{u,i} \mid (u,i) \in \mathcal{D}\}$ and imputed errors $\hat{\mathbf{E}} = \{\hat{e}_{u,i} \mid (u,i) \in \mathcal{D}\}$, its loss function is formulated as

$$\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o) = \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\Big[\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i} - \hat{e}_{u,i})}{\hat{p}_{u,i}}\Big]. \tag{3}$$

$\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)$ involves the conversion rate model $\hat{r}_{u,i}(1) = f_\phi(x_{u,i})$ and error imputation model $\hat{e}_{u,i} = g_\theta(x_{u,i})$. Doubly robust joint learning (DR-JL) approach [32] estimates them alternately: given $\hat{\theta}$, ϕ is updated by minimizing (3); given $\hat{\phi}$, θ is updated by minimizing

$$\mathcal{L}_e^{DR-JL}(\theta) = \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}(\hat{e}_{u,i} - e_{u,i})^2}{\hat{p}_{u,i}}. \tag{4}$$

2.2.4 More Robust Doubly Robust Estimator. Recently, the more robust doubly robust (MRDR) method [10] enhances the robustness of DR-JL by optimizing the variance of the DR estimator with the imputation model. Specifically, MRDR keeps the loss of the CVR prediction model in (3) unchanged, while replacing the loss of the imputation model in (4) with the following loss

$$\mathcal{L}_e^{MRDR}(\theta) = \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}(\hat{e}_{u,i} - e_{u,i})^2}{\hat{p}_{u,i}} \cdot \frac{1 - \hat{p}_{u,i}}{\hat{p}_{u,i}}. \tag{5}$$

This substitution can help reduce the variance of $\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)$ and hence a more robust estimator might be obtained.

2.2.5 Bias, Variance and Generalization Bound of DR Estimator. Given a hypothesis space $\mathcal{H}$ of CVR prediction matrix $\hat{\mathbf{R}}$, we define the optimal $\hat{\mathbf{R}}^*$ as $\hat{\mathbf{R}}^* = \arg\min_{\hat{\mathbf{R}}\in\mathcal{H}} \mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)$. Given imputed errors $\hat{\mathbf{E}}$ and learned propensities $\hat{\mathbf{P}}$, The following Lemmas 1 and 2 present the existing theoretical results of DR estimator [28, 32].

Lemma 1 (Bias and Variance). *The bias and variance of DR estimator are given as*

$$Bias[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)] = \frac{1}{|\mathcal{D}|}\Big|\sum_{(u,i)\in D}(p_{u,i} - \hat{p}_{u,i})\frac{(e_{u,i} - \hat{e}_{u,i})}{\hat{p}_{u,i}}\Big|,$$

$$\mathbb{V}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)] = \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}} p_{u,i}(1 - p_{u,i})\frac{(\hat{e}_{u,i} - e_{u,i})^2}{\hat{p}_{u,i}^2}.$$

Lemma 2 (Generalization Bound). *For any finite hypothesis space $\mathcal{H}$ of prediction matrices, then with probability $1 - \eta$,*

$$\mathcal{L}_{ideal}(\hat{\mathbf{R}}^*, \mathbf{R}) \le \mathcal{L}_{DR}(\hat{\mathbf{R}}^*, \mathbf{R}^o) + \underbrace{\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\frac{|p_{u,i} - \hat{p}_{u,i}|}{\hat{p}_{u,i}}|e_{u,i} - \hat{e}^*_{u,i}|}_{\text{Bias term}}$$

$$+ \underbrace{\sqrt{\frac{\log(2|\mathcal{H}|/\eta)}{2|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\Big(\frac{e_{u,i} - \hat{e}^\dagger_{u,i}}{\hat{p}_{u,i}}\Big)^2}}_{\text{Variance term}},$$

where $\hat{e}^*_{u,i}$ is the prediction error associated with $\hat{\mathbf{R}}^*$, $\hat{e}^\dagger_{u,i}$ is the prediction error corresponding to the prediction matrix $\hat{\mathbf{R}}^\dagger = \arg\max_{\hat{\mathbf{R}}^h\in\mathcal{H}}\sum_{(u,i)\in\mathcal{D}}(e_{u,i} - \hat{e}^h_{u,i})^2/\hat{p}^2_{u,i}$.

3 PROPOSED METHODS

3.1 Motivation

We reveal some worrying features of DR method, which provides an initial motivation. Lemma 1 formally gives the bias and variance of the DR estimator. According to the lemma, $Bias[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)] \approx 0$, if either $(\hat{e}_{u,i} - e_{u,i}) \approx 0$ or $(\hat{p}_{u,i} - p_{u,i}) \approx 0$, which is the property of double robustness. Nonetheless, both the bias and variance terms still have some issues. Specifically, the bias consists of the product of the errors of the propensity score model and imputation model weighted by $1/\hat{p}_{u,i}$. The term $(e_{u,i} - \hat{e}_{u,i})/\hat{p}_{u,i}$ is worrisome, as $1/\hat{p}_{u,i}$ tends to be large in unclicked events and inaccurate estimates of $e_{u,i}$ are most likely to occur in these events. Analogously, $(\hat{e}_{u,i} - e_{u,i})^2/\hat{p}^2_{u,i}$ in the variance term is also likely to be problematic.

It can be seen that both the bias and variance are correlated with the term of error deviation $|\hat{e}_{u,i} - e_{u,i}|$. Thus, it may be helpful to reduce them if the magnitude of error deviation is small. This is the basic idea of DR-JL approach that tries to reduce the error deviations of all events by optimizing the loss function (4). Further, the MRDR method [10] proposed replacing $\mathcal{L}_e^{DR-JL}(\theta)$ in (4) with $\mathcal{L}_e^{MRDR}(\theta)$ in (5) to deal with the large variance term. The idea behind Equation (5) is the truth that

$$\mathbb{V}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)] = \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\mathbb{E}_O\Big[\frac{o_{u,i}(1 - p_{u,i})(\hat{e}_{u,i} - e_{u,i})^2}{\hat{p}^2_{u,i}}\Big],$$

namely, the expectation of $\mathcal{L}_e^{MRDR}(\theta)$ equals to $\mathbb{V}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)]$.

Interestingly, Lemma 2 shows that the generalization bound depends on a weighted sum of the bias term and square root of the variance term in addition to the empirical loss, which fully reflects the feature of bias-variance trade-off. Since DR-JL does not control the bias and variance directly and MRDR pays no attention to the bias, both of them may still have poor generalization performance.

3.2 A Generalized DR Learning Framework

The difference between DR-JL and MRDR lies in the loss function of the error imputation model. As presented in Section 2.2, the alternating algorithm of DR-JL implies that its underlying loss is $\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o) + \mathcal{L}_e^{DR-JL}(\theta)$. Similarly, the real loss of MRDR is $\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o) + \mathcal{L}_e^{MRDR}(\theta)$. Note that the real loss functions of DR-JL and MRDR share a similar structure, so they can be discussed within a generalized framework. The real loss function of this framework has the following form

$$\mathcal{L}(\hat{\mathbf{R}}, \mathbf{R}^o) + Metric\{\mathcal{L}(\hat{\mathbf{R}}, \mathbf{R}^o)\}, \tag{6}$$

where $\mathcal{L}(\hat{\mathbf{R}}, \mathbf{R}^o)$ is an arbitrary unbiased loss function for training CVR prediction model, such as $\mathcal{L}_{IPS}(\hat{\mathbf{R}}, \mathbf{R}^o)$, $\mathcal{L}_{EIB}(\hat{\mathbf{R}}, \mathbf{R}^o)$ and $\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)$. $Metric\{\mathcal{L}(\hat{\mathbf{R}}, \mathbf{R}^o)\}$ is a pre-specified metric that reflects some features of $\mathcal{L}(\hat{\mathbf{R}}, \mathbf{R}^o)$ and is usually applied to learn the error imputation model. For example, the MRDR chooses $\mathbb{V}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)]$ as the metric, and DR-JL uses $\sum_{(u,i)\in\mathcal{D}}(\hat{e}_{u,i} - e_{u,i})^2$. Table 2 summarizes the metrics and ideas of existing doubly robust methods and our proposed methods, DR-BIAS and DR-MSE, which will be detailedly illustrated in Section 3.3.

Table 1: Generalized framework of various DR methods

Method	Metric	Goal
DR-JL	$\sum_{(u,i)\in\mathcal{D}}(\hat{e}_{u,i} - e_{u,i})^2$	Control error of imputation.
MRDR	$\mathbb{V}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)]$	Control variance.
DR-BIAS	$Bias[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)]$	Further reduce bias.
DR-MSE	$MSE[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)]$	Bias-variance trade-off.

It is noteworthy that due to the missing $r_{u,i}(1)$, optimizing $Metric\{\mathcal{L}(\hat{\mathbf{R}}, \mathbf{R}^o)\}$ directly is sometimes not feasible. In this case, one can use an approximation of $Metric\{\mathcal{L}(\hat{\mathbf{R}}, \mathbf{R}^o)\}$. For example, DR-JL adopts the feasible loss function (4) to approximate the infeasible $\sum_{(u,i)\in\mathcal{D}}(\hat{e}_{u,i} - e_{u,i})^2$, and MRDR employs (5) to substitute $\mathbb{V}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)]$.

Importantly, the proposed framework provides a valuable opportunity to develop a series of new unbiased CVR estimators with different characteristics to accommodate different application scenarios. In Section 3.3, we will develop two new DR approaches based on this framework.

3.3 Two New DR Methods

As discussed in Section 3.1, MRDR aims to reduce the variance of $\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)$, and is expected to achieve a more robust performance. However, this strategy works well only when $Bias[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)]$ is small enough as suggested by the generalization bound presented in Lemma 2. Reducing variance is less effective when the bias is large. DR-JL attempts to lower both the bias and variance by reducing the error deviation of the imputation model. Nevertheless, it does not directly control the bias and variance of $\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)$. To alleviate these limitations, we propose two new DR methods, DR-BIAS and DR-MSE, which are designed to further reduce bias and achieve better bias-variance trade-off, respectively.

3.3.1 DR-BIAS. DR-BIAS aims at further reducing the bias of the typical DR method through the optimization of the imputation model, since an accurate CVR prediction means that the bias should be small enough. Based on Lemmas 1 and 2, we design a variant of the bias of DR method as the metric to achieve this goal, given by

$$\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in D}\frac{(o_{u,i} - \hat{p}_{u,i})^2}{\hat{p}_{u,i}^2}(e_{u,i} - \hat{e}_{u,i})^2.$$

However, the above metric is infeasible due to the missing of $e_{u,i}$ in unclicked events. We make an approximation of it and define the loss of the imputation model of DR-BIAS as follows

$$\mathcal{L}_e^{DR-BIAS}(\theta) = \sum_{(u,i)\in\mathcal{D}}\frac{o_{u,i}(\hat{e}_{u,i} - e_{u,i})^2}{\hat{p}_{u,i}}\cdot\frac{(o_{u,i} - \hat{p}_{u,i})^2}{\hat{p}_{u,i}^2}. \tag{7}$$

By a comparison between Equation (5) and (7), we find that (7) just substitutes the weight $(1 - \hat{p}_{u,i})/\hat{p}_{u,i}$ with $(1 - \hat{p}_{u,i})^2/\hat{p}_{u,i}^2$ in clicked events. Also note that

$$\begin{cases} (1 - \hat{p}_{u,i})/\hat{p}_{u,i} > 1, \text{ if } \hat{p}_{u,i} < 1/2, \\ (1 - \hat{p}_{u,i})/\hat{p}_{u,i} < 1, \text{ if } \hat{p}_{u,i} > 1/2, \end{cases}$$

which means that DR-BIAS further magnifies the penalty of the clicked events with low propensity, and minifies those with high propensity. This leads to a desired effect: in the clicked events that the propensity model performs poorly, the amplified weights force the error imputation model to perform well. In other words, error imputation model complements the inaccurate part of the propensity score model. Thus, DR-BIAS would have smaller bias than other methods.

3.3.2 DR-MSE. Lemma 2 indicates that pursuing the bias reduction or variance reduction alone cannot fully control the generalization error. Seeking a better balance between the bias and variance appears to be a more effective way to improve the prediction accuracy. Therefore, we design a new model, namely DR-MSE, to achieve this goal. Specifically, a generalized Mean Squared Error (MSE) metric for DR-MSE method is defined as

$$\mathcal{L}_e^{DR-MSE}(\theta) = \lambda\mathcal{L}_e^{DR-BIAS}(\theta) + (1-\lambda)\mathcal{L}_e^{MRDR}(\theta), \tag{8}$$

where λ is a hyper-parameter for controlling the strength of the bias term and the variance term. When $\lambda = 1$, DR-MSE is reduced to DR-BIAS; when $\lambda = 0$, DR-MSE is reduced to MRDR; when $\lambda = 0.5$, DR-MSE optimizes the MSE of $\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)$ scaled by 0.5 through the imputation model.

However, simply using a hyper-parameter λ for all samples is not flexible enough due to the different characteristics and popularities of users and items. Specifically, different samples suffer from different issues during training, i.e., some might have higher variance while others might have worse bias. Thus, it is necessary to adopt different bias-variance tradeoff strategies for different user-item pairs. To achieve this goal, λ can be computed through a function $\lambda_\xi(x_{u,i})$ parameterized by ξ, such as a neural network, which enables personalized values for different user-item pairs. The

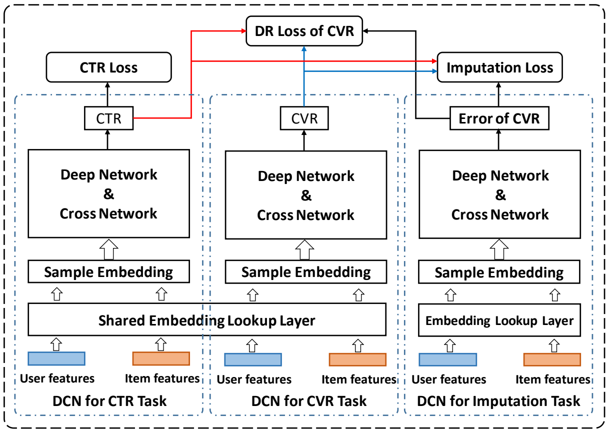

Figure 1: Model architecture of DR-BIAS and DR-MSE for experiments on large-scale industrial dataset. DCN is used as the base model for feature interaction learning for illustration only, and it can be readily replaced with other models such as FM [20], Wide&Deep [7] and DeepFM [9].

improved loss of DR-MSE is as follows

$$\mathcal{L}_e^{DR-MSE}\left(\theta, \lambda_\xi\right) = \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}\lambda_\xi(x_{u,i})(\hat{e}_{u,i}-e_{u,i})^2}{\hat{p}_{u,i}} \cdot \frac{(o_{u,i}-\hat{p}_{u,i})^2}{\hat{p}_{u,i}^2} + \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}(1-\lambda_\xi(x_{u,i}))(\hat{e}_{u,i}-e_{u,i})^2}{\hat{p}_{u,i}} \cdot \frac{1-\hat{p}_{u,i}}{\hat{p}_{u,i}}. \tag{9}$$

Essentially, the generalization bound of DR methods contains a weighted sum of the bias term and square root of the variance term, which can be flexibly tradeoff via the proposed generalized MSE metric in (8). Thus, it is expected that DR-MSE can obtain a better prediction performance under the tighter generalization bound.

4 PROPOSED TRAINING APPROACH

4.1 Model Architecture and Training Objective

Figure 1 shows the architecture of DR-BIAS and DR-MSE for experiments on real industrial scenarios. It is a multi-task learning framework with three DCN networks for the predication of post-view click-through rate (CTR), CVR, and error imputation, respectively. The embedding lookup layers of the DCN models for the CTR and CVR tasks are shared to tackle data sparsity issue, while the DCN model for the error imputation has its own embedding lookup layer. Note that DCN can be readily replaced with other models such as FM [20], Wide&Deep [7] and DeepFM [9]. We evaluate our proposed methods with both FM and DCN in our experiments.

During optimization, the CTR, CVR, and error imputation models are updated alternatively with stochastic gradient descent. Specifically, with the parameters of both CTR and CVR models fixed, the error imputation model is updated first by optimizing (9). With model parameters of the error imputation model fixed, the CTR and CVR models are optimized jointly through the sum of CVR loss and CTR loss

$$\mathcal{L}_{\text{CTCVR}}\left(\phi, \zeta, \theta(\lambda_\xi)\right) = \mathcal{L}_{DR}(\phi, \theta(\lambda_\xi)) + \mathcal{L}_{CTR}(\zeta),$$

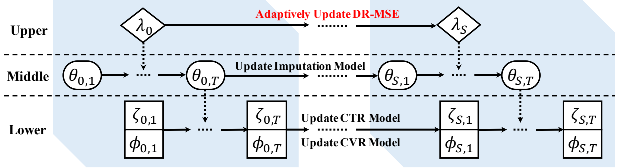

Figure 2: The proposed tri-level DR-MSE joint learning optimization method updates the λ_s in DR-MSE adaptively in upper level, while the existing bias-variance tradeoff approach uses a fixed λ_0. Middle and lower levels are for the joint training between the CTR&CVR and error imputation models.

where $\mathcal{L}_{DR}(\phi, \theta(\lambda_\xi)) = \sum_{(u,i)\in\mathcal{D}}[\hat{e}_{u,i} + o_{u,i}(e_{u,i} - \hat{e}_{u,i})/\hat{p}_{u,i}]$, $\mathcal{L}_{CTR}(\zeta) = -\sum_{(u,i)\in\mathcal{D}}[o_{u,i} \cdot \log(\hat{p}_{u,i}) + (1 - o_{u,i}) \cdot \log(1 - \hat{p}_{u,i})]$, $\hat{p}_{u,i}$ is the predicted CTR value, and used as the estimated propensity for unbiased CVR estimation. This joint learning process continues until the model converges. For DR-MSE, the optimization process involves updating $\lambda_\xi(\cdot)$, which makes it more challenging. In Section 4.2, we formally formulate the optimization problem and propose an effective training algorithm.

4.2 Tri-Level DR-MSE Joint Learning (JL) Optimization and Training Algorithm

We propose the tri-level optimization DR-MSE JL approach shown in Figure 2. Compared to the existing joint learning methods, our approach allows adaptively updating the λ_ξ in DR-MSE. This goal can be formalized as the following tri-level optimization problem

$$\xi^* = \arg\min_{\xi} \mathcal{L}_{DR}\left(\phi^*(\theta^*(\lambda_\xi)), \zeta^*(\theta^*(\lambda_\xi))\right)$$

$$\text{s.t. } \phi^*(\theta^*(\lambda_\xi)), \zeta^*(\theta^*(\lambda_\xi)) = \arg\min_{\phi,\zeta} \mathcal{L}_{\text{CTCVR}}\left(\phi, \zeta, \theta^*(\lambda_\xi)\right)$$

$$\text{s.t. } \theta^*(\lambda_\xi) = \arg\min_{\theta} \mathcal{L}_e^{DR-MSE}\left(\theta, \lambda_\xi\right)$$

There are two challenges for solving the above problem. Firstly, it is computationally expensive to search for the optimal DR-MSE by minimizing the upper loss in the tri-level DR-MSE JL optimization method. Secondly, the DR-MSE parameter λ_ξ of the upper model is difficult to be minimized as there is no closed-form solution. To address them, we further propose a training algorithm for this tri-level optimization problem as shown in Alg. 1.

For illustration purposes, the relevant parameters in Alg. 1 are updated using vanilla SGD. In practice, both SGD and its variants can be used for iterative updates. Specifically, for the error imputation optimization problem, $\mathcal{L}_e^{DR-MSE}$ is differentiable w.r.t. the parameter θ of the error imputation model. Given λ_s, one can compute the value of $\theta_{s+1}(\lambda_s)$ after a single vinalla SGD. It should be noted that this value is not directly used for the update of the error imputation model parameter θ_{s+1}. Moreover, the reason for using single-step SGD is that multi-step SGD here does not result in better performance, but rather increases the computational complexity [12].

Similarly, $\mathcal{L}_{\text{CTCVR}}$ is differentiable w.r.t. both CTR model parameter ζ and CVR model parameter ϕ. Given the pseudo-updated $\theta_{s+1}(\lambda_s)$, one can compute the value of $\zeta_{s+1}(\theta_{s+1}(\lambda_s))$ and $\phi_{s+1}(\theta_{s+1}(\lambda_s))$ after a single vanilla SGD. Both of the values are

Algorithm 1: Tri-Level DR-MSE JL Optimization Training

```
Input: $S$, observed ratings $\mathbf{R}^o$, learned propensities $\hat{\mathbf{P}}$,
for $O_s^l, O_s^u \subset O$ and $\mathcal{D}_s \subset \mathcal{D}$ $(s \in \{0, 1, \cdots, S-1\})$ do
    Compute an update function based on $O_s^l$
        $\theta_{s+1}(\lambda_s) \leftarrow \theta_s - \eta\nabla_{\theta_s}\mathcal{L}_e^{DR-MSE}(\theta, \lambda_s)$;
    Compute an update function based on $\mathcal{D}_s$
        $\phi_{s+1}(\theta_{s+1}(\lambda_s)) \leftarrow \phi_s - \eta\nabla_{\phi_s}\mathcal{L}_{\text{CTCVR}}(\phi, \zeta_s, \theta_{s+1}(\lambda_s))$;
    Compute an update function
        $\zeta_{s+1}(\theta_{s+1}(\lambda_s)) \leftarrow \zeta_s - \eta\nabla_{\zeta_s}\mathcal{L}_{\text{CTCVR}}(\phi_s, \zeta, \theta_{s+1}(\lambda_s))$;
    Compute the upper loss based on $O_s^u$
    $\mathcal{L}_{DR}(\phi_{s+1}(\theta_{s+1}(\lambda_s)), \zeta_{s+1}(\theta_{s+1}(\lambda_s)))$;
    Update the bias-variance trade off parameter
    $\xi_{s+1} \leftarrow \xi_s - \eta\nabla_{\xi_s}\mathcal{L}_{DR}\left(\phi_{s+1}(\theta_{s+1}(\lambda_\xi)), \zeta_{s+1}(\theta_{s+1}(\lambda_\xi))\right)$;
    Update the bias-variance trade off model $\lambda_{s+1} \leftarrow \lambda_{\xi_{s+1}}$;
    for $O_{s,t}^l \subset O$ $(t \in \{0, 1, \cdots, T-1\})$ do
        Update the imputation model based on $O_{s,t}^l$
            $\theta_{s,t+1} \leftarrow \theta_{s,t} - \eta\nabla_{\theta_{s,t}}\mathcal{L}_e^{DR-MSE}(\theta, \lambda_{s+1})$;
    end
    for $\mathcal{D}_{s,t} \subset \mathcal{D}$ $(t \in \{0, 1, \cdots, T-1\})$ do
        Update the propensity model based on $\mathcal{D}_{s,t}$
            $\zeta_{s,t+1} \leftarrow \zeta_{s,t} - \eta\nabla_{\zeta_{s,t}}\mathcal{L}_{\text{CTCVR}}(\phi_{s,t}, \zeta, \theta_{s,T})$;
        Update the predication model based on $\mathcal{D}_{s,t}$
            $\phi_{s,t+1} \leftarrow \phi_{s,t} - \eta\nabla_{\phi_{s,t}}\mathcal{L}_{\text{CTCVR}}(\phi, \zeta_{s,t}, \theta_{s,T})$;
    end
    Copy the model parameter $\theta_{s+1,0} \leftarrow \theta_{s,T}$;
    Copy the propensity model's parameter $\zeta_{s+1,0} \leftarrow \zeta_{s,T}$;
    Copy the predication model's parameter $\phi_{s+1,0} \leftarrow \phi_{s,T}$.
end
```

not directly used for the update of the CTR and CVR model as well. After that, with given $\zeta_{s+1}(\theta_{s+1}(\lambda_s))$ and $\phi_{s+1}(\theta_{s+1}(\lambda_s))$, we update the bias-variance tradeoff parameter in DR-MSE from λ_s to λ_{s+1} via a single vanilla SGD. Finally, based on the updated λ_{s+1}, we take the idea of joint learning to update the error imputation model parameter θ_{s+1} and the CTR&CVR model parameters ζ_{s+1}, ϕ_{s+1}, in which the classical multi-step SGD is used until the stopping criteria is satisfied.

5 REAL-WORLD EXPERIMENTS

In this section, we evaluate the proposed methods by conducting experiments on three real-world datasets, including two benchmark datasets with missing-at-random (MAR) ratings and one large-scale industrial product dataset. We aim to answer the following two research questions (RQ): (1) How do our methods compare with state-of-the-art models in terms of debiasing performance in practice? (2) How do the bias-variance tradeoff and the modeling of unobserved data affect the performance of the proposed methods in practice?

5.1 Experimental Setup

5.1.1 Datasets with MAR Ratings. A MAR testing set is important for assessing the performance of an unbiased recommender. Thus, we follow existing studies [10, 24] to use **Coat Shopping**[1] and **Yahoo! R3**[2] for the evaluation of CVR prediction model. To make the two datasets consistent with the CVR prediction task, we further preprocess them following previous studies [10, 24]. The detailed descriptions of these two datasets and the corresponding data preprocessing method are provided in Appendix B.1.

Table 2: Statistics of the advertising dataset Product

Dataset	#Impression	#Click	#Conversion	#User	#Item
Training	739.66M	3.73M	1.90M	524K	68K
Testing	99.73M	519K	268K	283K	52K

Note: "M" means million, and "K" means thousand.

5.1.2 Industrial Product Dataset. To provide more comprehensive and reliable evaluation, we also conduct experiments on a large-scale App advertising dataset collected from a real-world system. We denote this dataset as **Product** with some statistics of it displayed in Table 2. It contains 8 consecutive days logged data from the system, with the first 7 days for training and the last day for testing. Each sample of the dataset contains features from a user, an item and the corresponding context. Although the unbiased data in CVR prediction is unobtainable in real applications since we cannot force users to randomly click the exposed items, the experiments can still provide valuable observations for the applications of debiasing CVR prediction models in real systems.

5.1.3 Baselines and Implementation. For experiments on **Coat** and **Yahoo**, we compare our methods with several competitive baselines, including Naive, IPS, DR-JL and MRDR. The base model for all methods is factorization machine. Some brief descriptions of them and implementation details are provided in Appendix B. For experiments on **Product**, we also select some state-of-the-art CVR prediction models for large datasets, including DCN [31], ESMM [19], Multi_IPW [39] and Multi_DR [39]. The base model for all methods is DCN. More details are provided in Appendix C.

5.1.4 Experimental Protocols. For experiments on **Coat** and **Yahoo**, we evaluate the ranking performance with two types of metrics, i.e., discounted cumulative gain (DCG) and recall, as prior work on debiasing CVR prediction [10, 24]. For experiments on **Product**, we evaluate our proposed methods on three important tasks, i.e., CTR, CVR, and CTCVR ($CTCVR = CTR * CVR$) predictions, with the AUC score following existing works [19, 39].

5.2 Overall Performance (RQ1)

5.2.1 Unbiased Evaluation. The experimental results on **Coat** and **Yahoo** are shown in Table 3. We have the following observations.

First, our proposed methods are effective for debiasing CVR prediction task. As shown in Table 3, both DR-MSE and DR-BIAS consistently outperform all the other ones in terms of DCG@K and Recall@K ($K = 2, 4, 6$) on the two real-world datasets, with only one exception of DR-MSE on Recall@6 of **Yahoo**. In particular, DR-MSE achieves a significant 3.22%, 2.65% and 1.87% relative improvements over MRDR on DCG@2, DCG@4 and DCG@6, respectively.

Second, it is necessary to improve the bias and variance of the typical DR method under inaccurate propensity estimation and

[1] https://www.cs.cornell.edu/~schnabts/mnar/
[2] http://webscope.sandbox.yahoo.com/

Table 3: Performance comparison based on Coat and Yahoo.

Datasets	Models	DCG@2	DCG@4	DCG@6	Recall@2	Recall@4	Recall@6
Coat	Naïve	0.7283 ± 0.0264	0.9763 ± 0.0258	1.1512 ± 0.0241	0.8474 ± 0.0310	1.3786 ± 0.0374	1.8490 ± 0.0379
	IPS	0.7102 ± 0.0220	0.9596 ± 0.0222	1.1299 ± 0.0210	0.8248 ± 0.0272	1.3596 ± 0.0360	1.8174 ± 0.0377
	DR-JL	0.7416 ± 0.0224	1.0021 ± 0.0224	1.1762 ± 0.0229	0.8645 ± 0.0264	1.4225 ± 0.0362	1.8906 ± 0.0403
	MRDR	0.7442 ± 0.0225	1.0132 ± 0.0219	1.1947 ± 0.0194	0.8736 ± 0.0273	1.4494 ± 0.0325	1.9370 ± 0.0318
	DR-BIAS	**0.7648 ± 0.0192***	**1.0353 ± 0.0169***	**1.2127 ± 0.0162***	**0.8959 ± 0.0251***	**1.4751 ± 0.0273***	**1.9517 ± 0.0324***
	DR-MSE	**0.7682 ± 0.0151***	**1.0401 ± 0.0150***	**1.2170 ± 0.0139***	**0.8997 ± 0.0194***	**1.4816 ± 0.0241***	**1.9569 ± 0.0262***
Yahoo	Naïve	0.5469 ± 0.0009	0.7466 ± 0.0008	0.8714 ± 0.0004	0.6479 ± 0.0012	1.0745 ± 0.0016	1.4098 ± 0.0013
	IPS	0.5502 ± 0.0010	0.7520 ± 0.0009	0.8751 ± 0.0009	0.6545 ± 0.0017	1.0797 ± 0.0017	**1.4168 ± 0.0019**
	DR-JL	0.5602 ± 0.0034	0.7586 ± 0.0030	0.8808 ± 0.0025	0.6615 ± 0.0042	1.0849 ± 0.0049	1.4129 ± 0.0039
	MRDR	0.5623 ± 0.0024	0.7603 ± 0.0027	0.8820 ± 0.0020	0.6646 ± 0.0033	1.0881 ± 0.0045	1.4145 ± 0.0037
	DR-BIAS	**0.5646 ± 0.0023***	**0.7624 ± 0.0021***	**0.8841 ± 0.0018***	**0.6676 ± 0.0026***	**1.0904 ± 0.0028***	**1.4169 ± 0.0020**
	DR-MSE	**0.5662 ± 0.0017***	**0.7639 ± 0.0016***	**0.8850 ± 0.0014***	**0.6670 ± 0.0026***	**1.0891 ± 0.0029**	1.4140 ± 0.0028

Note: * statistically significant results (p-value $\leq$ 0.05) using the paired-t-test compared with the best baseline.

Table 4: Performance comparison based on Product.

Models	CTR AUC (%)	CVR AUC (%)	CTCVR AUC (%)
DCN	90.763	75.691	95.254
ESMM	90.704	81.647	95.505
DR-JL	90.754	81.768	95.548
Multi_IPW	90.794	81.912	95.571
Multi_DR	90.807	81.864	95.569
MRDR	90.721	81.810	95.535
DR-BIAS	**90.913**	81.974	95.633
DR-MSE	90.825	**82.067**	**95.654**

error imputation so as to enhance its robustness and ranking performance. As shown in Table 3, IPS has worse performance on **Coat** and only comparable performance on **Yahoo** compared with the Naive method, since it suffers heavily from the high variance issue. Both DR-JL and MRDR performs better compared with IPS because of their double robustness. DR-BIAS improves over MRDR by achieving smaller bias through magnifying the penalty of the clicked events with low propensity while minifying those with high propensity as analyzed in Section 3.3.1. However, these DR methods still suffer from the high bias and/or variance issues. Our proposed DR-MSE can further achieve improvements over all other DR methods by better controlling the bias and variance.

5.2.2 Large-scale Industrial Dataset. The experimental results on **Product** are shown in Table 4. Firstly, we can observe that ESMM improves over DCN on CVR and CTCVR prediction tasks by tackling the data sparsity issue with the multi-task learning framework, but it still suffers from the selection bias issue. Secondly, the debiasing CVR models can simultaneously tackle the data sparsity and selection bias issues, thus they outperform DCN and ESMM. Thirdly, our proposed methods achieve significant improvements over existing debiasing CVR prediction models, including DR-JL, Multi_IPW, Multi_DR and MRDR, which is consistent with the observations on experiments with unbiased evaluation. It demonstrates that our proposed methods have both theoretical guarantee and great application potentials in real industrial systems.

5.3 In-depth Analysis of DR-MSE (RQ2)

We conduct an analysis of two important aspects of DR-MSE with **Coat** in this section. The experimental results are displayed in Figure 3. Note that similar results can be observed on other datasets, and we do not present them here only due to space limitations.

The loss of the imputation model of DR-MSE contains a bias term and a variance term. We conduct experiments by manually varying λ in Eq. (8) to demonstrate the necessity of conducting bias-variance tradeoff. The left part of Figure 3 presents the experimental results of DR-MSE when varying λ from 0.1 to 0.9. We can find that the performance of DR-MSE first improves with the increase of λ, and then gradually drops. It shows that an appropriate tradeoff between this two terms can improve model generalization performance.

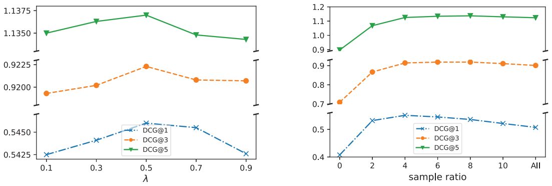

Figure 3: The effect of the coefficient λ for balancing bias and variance, and the sample ratio of unclicked events to clicked events on the ranking performance of DR-MSE.

DR methods can achieve double robustness by jointly considering clicked events and unclicked events. Here, we also study the effect of the sample ratio of unclicked events to clicked events on the performance of DR-MSE. When the sample ratio is set to "All", all the unclicked events are utilized for training; when the sample ratio is set to 0, only clicked events are utilized. As shown in Figure 3, when the sample ratio ranges from 0 to "All", the DCG@K ($K = 1, 3, 5$) scores on **Coat** show an apparent increase first, and then tend to saturate or decrease slightly. It suggests that a certain amount of unclicked events can provide useful information for improving the prediction model with the assistance of an imputation model, but further improvement is marginal when passing some threshold. In real advertising applications, the unclicked events are usually composed of the exposed but unclicked events in consideration of time efficiency. This empirical study on the sample ratio can provide some justification of the practice.

6 SEMI-SYNTHETIC DATA EXPERIMENTS

In this section, we aim to investigate the robustness of our proposed method through experiments based on semi-synthetic datasets with different levels of selection bias.

6.1 Experimental Setup

6.1.1 Datasets and Preprocessing. **MovieLens 100K**[3] **(ML-100K)** is a dataset collected from a movie recommendation service with

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

Table 5: Performance comparison on semi-synthetic datasets based on ML-100k.

Metrics	AUC			Log-loss		
ρ	0.5	1	2	0.5	1	2
Naïve	0.7250 ± 0.0001	0.6731 ± 0.0001	0.5279 ± 0.0070	0.3178 ± 0.0000	0.3343 ± 0.0001	0.4683 ± 0.0179
IPS	0.7316 ± 0.0001	0.6648 ± 0.0028	0.5263 ± 0.0055	0.3165 ± 0.0001	0.3304 ± 0.0034	0.4789 ± 0.0132
DR-JL	0.7319 ± 0.0004	0.6673 ± 0.0035	0.5703 ± 0.0032	0.3116 ± 0.0002	0.3255 ± 0.0012	0.3607 ± 0.0014
MRDR	0.7335 ± 0.0006	0.6765 ± 0.0021	0.5563 ± 0.0082	0.3067 ± 0.0002	0.3238 ± 0.0006	0.3650 ± 0.0047
DR-BIAS	**0.7349 ± 0.0006***	**0.6916 ± 0.0009***	**0.6073 ± 0.0054***	**0.3064 ± 0.0001***	**0.3194 ± 0.0013***	**0.3494 ± 0.0058**
DR-MSE	**0.7359 ± 0.0002***	**0.6928 ± 0.0020***	**0.6084 ± 0.0168***	**0.3059 ± 0.0001***	**0.3193 ± 0.0028***	**0.3477 ± 0.0084**

Note: * statistically significant results (p-value $\leq$ 0.05) using the paired-t-test compared with the best baseline.

100,000 MNAR ratings from 943 users and 1,682 movies. We used it to generate semi-synthetic datasets for experiments with the following standard procedures as previous studies [24, 26].

(1) Obtain an approximation of the true ratings of each user on all items with rating-based matrix factorization [16]. We denote the predicted rating of a user u on an item i as $\hat{R}_{u,i}$. Then, the ground-truth CVR for conversion generation is generated as follows:

$$p_{u,i}^{cvr} = \sigma(\hat{R}_{u,i} - \epsilon), \forall (u,i) \in \mathcal{D},$$

where $\sigma(\cdot)$ is the sigmoid function, and ϵ controls the level of overall relevance; ϵ is set to 5 in experiments.

(2) Obtain an approximation of the true observations with logistic matrix factorization [13]. We denote the predicted probability of a user-item pair (u,i) being observed as $\hat{O}_{u,i}$. Then, the ground-truth CTR for generating the click events is defined as follows:

$$p_{u,i}^{ctr} = (\hat{O}_{u,i})^{\rho}, \forall (u,i) \in \mathcal{D},$$

where ρ controls the skewness of the distribution of the CTR. A large value of ρ means a huge selection bias in the clicked events and a small number of observed click and conversion events. We set ρ as 0.5, 1, and 2 in the experiments.

(3) Sample binary click and conversion events with Bernoulli sampling based on the ground-truth CTR and CVR as follows:

$$o_{u,i} \sim Bern(p_{u,i}^{ctr}),\ r_{u,i} \sim Bern(p_{u,i}^{cvr}),\ \forall (u,i) \in \mathcal{D},$$

where $Bern(\cdot)$ is the Bernoulli distribution. Then, the post-click conversions can be derived as $\{(u,i,r_{u,i}) | o_{u,i} = 1\}$.

6.1.2 Baselines and Implementation. The baseline algorithms include the Naive method, IPS [28], DR-JL [32], and MRDR [10]. The detailed descriptions of the baselines and model implementation are provided in Appendix B.

6.1.3 Evaluation Protocols. In semi-synthetic datasets, we have the ground-truth user preference information and the level of selection bias of the considered datasets, so that we can investigate model robustness through experiments. We generate the semi-synthetic datasets by setting ρ as 0.5, 1 and 2. The biased set consists of the clicked events generated by the procedure described in Section 6.1.1, which is further divided into a training set (90%) and a validation set (10%). We conduct experiments in each setting for 10 times and report the average results. Note that larger value of ρ means higher selection bias and less clicked events for training because of lower propensity. We use AUC and Log-loss on test sets to evaluate the ranking performance and the relevance prediction, respectively. The test set consists of user-item pairs randomly sampled from the unclicked ones, and we uniformly sample 50 items for each user in the experiments.

6.2 Results & Discussion

Our method DR-MSE has the best AUC scores and Log-loss results across all the considered levels of selection bias ($\rho = 0.5, 1, 2$). It demonstrates that DR-MSE can achieve better ranking performance and relevance prediction. DR-BIAS also has impressive performance and outperforms MRDR significantly, which is probably because DR-BIAS achieves smaller bias by magnifying the penalty of the clicked events with low propensity while minifying those with high propensity. With the increase of the power ρ, the performance of IPS drops dramatically, and was even worse than that of the Naive method. It shows that IPS suffers heavily from the high variance issue. Doubly robust learning approaches, including DR-JL, MRDR, DR-BIAS and DR-MSE, have better robustness against the selection bias and demonstrate better results compared with the IPS method. Our proposed DR-MSE performs the best because of its bias and variance reduction characteristics.

7 RELATED WORK

7.1 Approaches to CVR Estimation

In practice, CTR prediction models are commonly applied to CVR prediction task due to their inherent similarity. These CTR prediction approaches include logistic regression based methods [8, 21], factorization machine based methods [14, 20], deep learning based methods [7, 9, 31, 35], etc. In addition, many approaches are specially designed for CVR prediction because of several unique and critical issues of the task, such as delayed feedback [5, 29, 34], data sparsity [19, 36] and selection bias [10, 39]. In this paper, we mainly focus on tackling the selection bias issue.

Selection bias refers to the distribution drift between the train and inference data, which is widely studied recently [10, 19, 24, 39]. Some existing multi-task learning methods, such as ESMM [19] and ESM^2 [36], can alleviate the selection bias, but they are heuristic methods and lack theoretic guarantee. Further, the author in [39] tried to use DR method to debias CVR prediction and proposed a model namely Multi_DR with theoretic guarantee. But they only validated the proposed methods with the biased training and testing sets. The authors in [25] proposed a dual learning algorithm for simultaneously tackling the delayed feedback issue and the selection bias issue. MRDR [10] designs a new loss for the imputation model to reduce the variance of Multi_DR [39]. However, it might still suffer from the high bias of DR method due to the incorrect estimations of both propensity scores and imputed errors (which is common in practice). To tackle these problems, in this paper, we proposed a generalized doubly robust learning framework for debiasing CVR prediction, which enables us to propose two new DR methods with more favorable properties.

7.2 Debiasing in Recommendation Tasks

Recent years have witnessed many contributions on incorporating the causal inference idea into the recommendation domain for unbiased learning [28, 32]. For example, [28] explains the recommendation problem by a treatment-effect model, and designs an IPS based method to remove the bias in the observed data based on explicit feedback. [32] improves over the IPS based method by designing a doubly robust learning approach. In addition, several existing works [3, 6, 17, 33] design debiasing models by leveraging the available small set of unbiased data. Though these methods have achieved many successes in debiasing recommendation tasks, none of them are specially proposed for CVR prediction. How to design an unbiased learning algorithm for CVR prediction is highly important and needs to be studied further.

8 CONCLUSION

We have proposed a generic doubly robust (DR) learning framework for debiasing CVR prediction based on the theoretical analysis of the bias, variance and generalization bounds of existing DR methods. This framework enables us to develop a series of new estimators with different desired characteristics to accommodate different application scenarios in CVR prediction. In particular, based on the framework, we proposed two new DR methods, namely DR-BIAS and DR-MSE, which are designed to further reduce the bias and achieve a better bias-variance trade-off. In addition, we propose a novel tri-level optimization for DR-MSE, and the corresponding efficient training algorithm. Finally, we empirically validate the effectiveness of the proposed methods by extensive experiments on both semi-synthetic and real-world datasets.

ACKNOWLEDGEMENTS

This study was partly supported by grants from the National Science and Technology Major Project of the Ministry of Science and Technology of China (No.2021YFF0901400). We appreciate the support from Mindspore[4], which is a new deep learning computing framework.

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, et al. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.

[2] Heejung Bang and James M. Robins. 2005. Doubly robust estimation in missing data and causal inference models. *Biometrics* 61 (2005), 962–972.

[3] Stephen Bonner and Flavian Vasile. 2018. Causal embeddings for recommendation. In *RecSys*. 104–112.

[4] Yin-Wen Chang, Cho-Jui Hsieh, Kai-Wei Chang, Michael Ringgaard, and Chih-Jen Lin. 2010. Training and testing low-degree polynomial data mappings via linear SVM. *Journal of Machine Learning Research* 11, 4 (2010).

[5] Olivier Chapelle. 2014. Modeling delayed feedback in display advertising. In *KDD*. ACM, 1097–1105.

[6] Jiawei Chen, Hande Dong, Yang Qiu, Xiangnan He, Xin Xin, Liang Chen, Guli Lin, and Keping Yang. 2021. AutoDebias: Learning to Debias for Recommendation. In *SIGIR*.

[7] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *1st DLRS workshop*.

[8] Muhammad Junaid Effendi and Syed Abbas Ali. 2017. Click Through Rate Prediction for Contextual Advertisment Using Linear Regression. *IJCSIS* (2017).

[9] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. Deepfm: a factorization-machine based neural network for ctr prediction. In *IJCAI*.

[10] Siyuan Guo, Lixin Zou, Yiding Liu, Wenwen Ye, Suqi Cheng, Shuaiqiang Wang, Hechang Chen, Dawei Yin, and Yi Chang. 2021. Enhanced Doubly Robust Learning for Debiasing Post-Click Conversion Rate Estimation. In *SIGIR*.

[11] G. W. Imbens and D. B. Rubin. 2015. *Causal Inference For Statistics Social and Biomedical Science*. Cambridge University Press.

[12] Simon Jenni and Paolo Favaro. 2018. Deep bilevel learning. In *Proceedings of the European conference on computer vision (ECCV)*. 618–633.

[13] Christopher C. Johnson. 2014. Logistic Matrix Factorization for Implicit Feedback Data.

[14] Yuchin Juan, Yong Zhuang, Wei-Sheng Chin, and Chih-Jen Lin. 2016. Field-aware factorization machines for CTR prediction. In *RecSys*. 43–50.

[15] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*, Yoshua Bengio and Yann LeCun (Eds.).

[16] Yehuda Koren, Robert Bell, Chris Volinsky, et al. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.

[17] Dugang Liu, Pengxiang Cheng, Zhenhua Dong, Xiuqiang He, Weike Pan, and Zhong Ming. 2020. A general knowledge distillation framework for counterfactual recommendation via uniform data. In *SIGIR*. 831–840.

[18] Quan Lu, Shengjun Pan, Liang Wang, Junwei Pan, Fengdan Wan, and Hongxia Yang. 2017. A Practical Framework of Conversion Rate Prediction for Online Display Advertising. In *ADKDD*. ACM, 9:1–9:9.

[19] Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu, Xiaoqiang Zhu, and Kun Gai. 2018. Entire Space Multi-Task Model: An Effective Approach for Estimating Post-Click Conversion Rate. In *SIGIR*. 1137–1140.

[20] Steffen Rendle. 2010. Factorization machines. In *ICDM*.

[21] Matthew Richardson, Ewa Dominowska, and Robert Ragno. 2007. Predicting clicks: estimating the click-through rate for new ads. In *WWW*.

[22] P. R. Rosenbaum and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal. *Biometric* 70 (1983), 41–55.

[23] D. B. Rubin. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational psychology* 66 (1974), 688–701.

[24] Yuta Saito. 2020. Doubly Robust Estimator for Ranking Metrics with Post-Click Conversions. In *RecSys*. ACM, 92–100.

[25] Yuta Saito, Gota Morishita, and Shota Yasui. 2020. Dual Learning Algorithm for Delayed Conversions. In *SIGIR*. ACM, 1849–1852.

[26] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. 2020. Unbiased Recommender Learning from Missing-Not-At-Random Implicit Feedback. In *WSDM*.

[27] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as Treatments: Debiasing Learning and Evaluation. In *ICML*. 1670–1679.

[28] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as treatments: Debiasing learning and evaluation. *arXiv preprint arXiv:1602.05352* (2016).

[29] Yumin Su, Liang Zhang, Quanyu Dai, Bo Zhang, Jinyao Yan, Dan Wang, Yongjun Bao, Sulong Xu, Yang He, and Weipeng Yan. 2020. An Attention-based Model for Conversion Rate Prediction with Delayed Feedback via Post-click Calibration. In *IJCAI*. 3522–3528.

[30] Roman Vershynin. 2018. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press.

[31] Ruoxi Wang, Bin Fu, Gang Fu, and Mingliang Wang. 2017. Deep & cross network for ad click predictions. In *ADKDD*. 1–7.

[32] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2019. Doubly Robust Joint Learning for Recommendation on Data Missing Not at Random. In *ICML*.

[33] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2021. Combating Selection Biases in Recommender Systems with a Few Unbiased Ratings. In *WSDM*.

[34] Yanshi Wang, Jie Zhang, Qing Da, and Anxiang Zeng. 2021. Delayed Feedback Modeling for the Entire Space Conversion Rate Prediction. In *AAAI*.

[35] Yikai Wang, Liang Zhang, Quanyu Dai, Fuchun Sun, Bo Zhang, Yang He, Weipeng Yan, and Yongjun Bao. 2019. Regularized Adversarial Sampling and Deep Time-aware Attention for Click-Through Rate Prediction. In *CIKM*. 349–358.

[36] Hong Wen, Jing Zhang, Yuan Wang, Fuyu Lv, Wentian Bao, Quan Lin, and Keping Yang. 2020. Entire Space Multi-Task Modeling via Post-Click Behavior Decomposition for Conversion Rate Prediction. In *SIGIR*. ACM, 2377–2386.

[37] Liang Wu, Diane Hu, Liangjie Hong, and Huan Liu. 2018. Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce. In *SIGIR*.

[38] Peng Wu, Haoxuan Li, Yuhao Deng, Wenjie Hu, Quanyu Dai, Zhenhua Dong, Jie Sun, Rui Zhang, and Xiao-Hua Zhou. 2022. On the Opportunity of Causal Learning in Recommendation Systems: Foundation, Estimation, Prediction and Challenges. *IJCAI*.

[39] Wenhao Zhang, Wentian Bao, Xiao-Yang Liu, Keping Yang, Quan Lin, Hong Wen, and Ramin Ramezani. 2020. Large-scale Causal Approaches to Debiasing Post-click Conversion Rate Estimation with Multi-task Learning. In *WWW*.

[4] https://www.mindspore.cn

Appendices

Appendix A PROOF OF LEMMAS

This supplementary material contains the proofs of Lemma 1 and Lemma 2. For ease of exposition, let $\mathcal{L}(\hat{\mathbf{R}}) = \mathcal{L}(\hat{\mathbf{R}}, \mathbf{R}^o)$.

Lemma 1 (Bias and Variance). *Given imputed errors $\hat{\mathbf{E}}$ and learned propensities $\hat{\mathbf{P}}$ with $\hat{p}_{u,i} > 0$ for all user-item pairs, the bias and variance of DR estimator are given as*

$$Bias[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)] = \frac{1}{|\mathcal{D}|}\Big|\sum_{(u,i)\in D}(p_{u,i} - \hat{p}_{u,i})\frac{(e_{u,i} - \hat{e}_{u,i})}{\hat{p}_{u,i}}\Big|,$$

$$\mathbb{V}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}, \mathbf{R}^o)] = \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}} p_{u,i}(1 - p_{u,i})\frac{(\hat{e}_{u,i} - e_{u,i})^2}{\hat{p}^2_{u,i}}.$$

Proof. According to the definition of bias,

$$\begin{aligned} Bias[\mathcal{L}_{DR}(\hat{\mathbf{R}})] &= \Big|\mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}})] - \mathcal{L}_{ideal}(\hat{\mathbf{R}}, \mathbf{R})\Big| \\ &= \Big|\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\mathbb{E}_O[\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i} - \hat{e}_{u,i})}{\hat{p}_{u,i}} - e_{u,i}]\Big| \\ &= \Big|\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}[\hat{e}_{u,i} + \frac{p_{u,i}(e_{u,i} - \hat{e}_{u,i})}{\hat{p}_{u,i}} - e_{u,i}]\Big| \\ &= \frac{1}{|\mathcal{D}|}\Big|\sum_{(u,i)\in D}\frac{p_{u,i} - \hat{p}_{u,i}}{\hat{p}_{u,i}}(e_{u,i} - \hat{e}_{u,i})\Big|. \end{aligned}$$

The variance of $\mathcal{L}_{DR}(\hat{\mathbf{R}})$ with respect to click indicator is given as

$$\begin{aligned} \mathbb{V}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}})] &= \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\mathbb{V}_O[\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i} - \hat{e}_{u,i})}{\hat{p}_{u,i}}] \\ &= \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\mathbb{V}_O[o_{u,i}]\cdot\left(\frac{e_{u,i} - \hat{e}_{u,i}}{\hat{p}_{u,i}}\right)^2 \\ &= \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\frac{p_{u,i}(1 - p_{u,i})}{\hat{p}^2_{u,i}}(\hat{e}_{u,i} - e_{u,i})^2. \end{aligned}$$

□

To show the generalization bound of doubly robust estimator, we need the Hoeffding's inequality for general bounded random variables, which is presented in Lemma 3.

Lemma 2 (Generalization Bound). *For any finite hypothesis space $\mathcal{H}$ of prediction matrices, given imputed errors $\hat{\mathbf{E}}$ and learned propensities $\hat{\mathbf{P}}$, then with probability $1 - \eta$,*

$$\mathcal{L}_{ideal}(\hat{\mathbf{R}}^*, \mathbf{R}) \le \mathcal{L}_{DR}(\hat{\mathbf{R}}^*, \mathbf{R}^o) + \underbrace{\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\frac{|p_{u,i} - \hat{p}_{u,i}|}{\hat{p}_{u,i}}|e_{u,i} - \hat{e}^*_{u,i}|}_{\textit{Bias term}} + \underbrace{\sqrt{\frac{\log(2|\mathcal{H}|/\eta)}{2|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}(\frac{e_{u,i} - \hat{e}^\dagger_{u,i}}{\hat{p}_{u,i}})^2}}_{\textit{Variance term}},$$

where $\hat{e}^\dagger_{u,i}$ is the prediction error corresponding to the prediction matrix $\hat{\mathbf{R}}^\dagger = \arg\max_{\hat{\mathbf{R}}^h\in\mathcal{H}}\sum_{(u,i)\in\mathcal{D}}(e_{u,i} - \hat{e}^h_{u,i})^2/\hat{p}^2_{u,i}$, $\hat{e}^*_{u,i}$ is the prediction error associated with $\hat{\mathbf{R}}^*$.

Proof. We first note that

$$\begin{aligned} &\mathcal{L}_{ideal}(\hat{\mathbf{R}}^*, \mathbf{R}) - \mathcal{L}_{DR}(\hat{\mathbf{R}}^*) \\ &= \mathcal{L}_{ideal}(\hat{\mathbf{R}}^*, \mathbf{R}) - \mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}^*)] + \mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}^*)] - \mathcal{L}_{DR}(\hat{\mathbf{R}}^*) \\ &\le Bias[\mathcal{L}_{DR}(\hat{\mathbf{R}}^*)] + \mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}^*)] - \mathcal{L}_{DR}(\hat{\mathbf{R}}^*) \\ &\le \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\frac{|p_{u,i} - \hat{p}_{u,i}|}{\hat{p}_{u,i}}|e_{u,i} - \hat{e}^*_{u,i}| + \mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}^*)] - \mathcal{L}_{DR}(\hat{\mathbf{R}}^*). \end{aligned} \tag{10}$$

Next we focus on analyzing $\mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}^*)] - \mathcal{L}_{DR}(\hat{\mathbf{R}}^*)$. By Hoeffding's inequality in Lemma 3, let $X_{u,i} = \frac{o_{u,i}(e_{u,i} - \hat{e}_{u,i})}{\hat{p}_{u,i}}$, then $M_{u,i} - m_{u,i} = \frac{|e_{u,i} - \hat{e}_{u,i}|}{\hat{p}_{u,i}}$, and for any $\epsilon > 0$, we have

$$\begin{aligned} &\mathbb{P}\Big\{\Big|\mathcal{L}_{DR}(\hat{\mathbf{R}}^*) - \mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}^*)]\Big| \le \epsilon\Big\} \\ &= 1 - \mathbb{P}\Big\{\Big|\mathcal{L}_{DR}(\hat{\mathbf{R}}^*) - \mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}^*)]\Big| > \epsilon\Big\} \\ &\ge 1 - \mathbb{P}\Big\{\sup_{\hat{\mathbf{R}}^h\in\mathcal{H}}\Big|\mathcal{L}_{DR}(\hat{\mathbf{R}}^h) - \mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}^h)]\Big| > \epsilon\Big\} \\ &\ge 1 - \sum_{h=1}^{\mathcal{H}}\mathbb{P}\Big\{\Big|\mathcal{L}_{DR}(\hat{\mathbf{R}}^h) - \mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}^h)]\Big| > \epsilon\Big\} \\ &= 1 - \sum_{h=1}^{\mathcal{H}}\mathbb{P}\{|\sum_{(u,i)\in\mathcal{D}}(\frac{o_{u,i}(e_{u,i} - \hat{e}^h_{u,i})}{\hat{p}_{u,i}} - \mathbb{E}_O(\frac{o_{u,i}(e_{u,i} - \hat{e}^h_{u,i})}{\hat{p}_{u,i}}))| > \epsilon|\mathcal{D}|\} \\ &\ge 1 - \sum_{h=1}^{\mathcal{H}}2\exp\Big\{-2\epsilon^2|\mathcal{D}|^2\Big/\sum_{(u,i)\in\mathcal{D}}(\frac{e_{u,i} - \hat{e}^h_{u,i}}{\hat{p}_{u,i}})^2\Big\} \\ &\ge 1 - 2|\mathcal{H}|\exp\Big\{-2\epsilon^2|\mathcal{D}|^2\Big/\sum_{(u,i)\in\mathcal{D}}(\frac{e_{u,i} - \hat{e}^\dagger_{u,i}}{\hat{p}_{u,i}})^2\Big\}. \end{aligned}$$

Letting $2|\mathcal{H}|\exp\Big\{-2\epsilon^2|\mathcal{D}|^2\Big/\sum_{(u,i)\in\mathcal{D}}(\frac{e_{u,i} - \hat{e}^\dagger_{u,i}}{\hat{p}_{u,i}})^2\Big\} = \eta$ yields that

$$\epsilon = \sqrt{\frac{\log(2|\mathcal{H}|/\eta)}{2|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}(\frac{e_{u,i} - \hat{e}^\dagger_{u,i}}{\hat{p}_{u,i}})^2}.$$

Then with probability $1 - \eta$, we have

$$\mathbb{E}_O[\mathcal{L}_{DR}(\hat{\mathbf{R}}^*)] - \mathcal{L}_{DR}(\hat{\mathbf{R}}^*) \le \sqrt{\frac{\log(2|\mathcal{H}|/\eta)}{2|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}(\frac{e_{u,i} - \hat{e}^\dagger_{u,i}}{\hat{p}_{u,i}})^2}. \tag{11}$$

Lemma 2 follows immediately from inequalities (10) and (11). □

Lemma 3 (Hoeffding's inequality for general bounded random variables). *Let $X_1, ..., X_N$ be independent random variables. Assume that $X_i \in [m_i, M_i]$ for every i, Then, for any $\epsilon > 0$, we have*

$$\mathbb{P}\{\Big|\sum_{i=1}^N X_i - \sum_{i=1}^N \mathbb{E}X_i\Big| > \epsilon\} \le 2\exp\Big\{-\frac{2\epsilon^2}{\sum_{i=1}^N(M_i - m_i)^2}\Big\}.$$

Proof. The proof can be found in Theorem 2.2.6 of [30]. □

Appendix B EXPERIMENTAL SETTINGS ON COAT, YAHOO, AND SEMI-SYNTHETIC DATASETS

Here, we provide more detailed experimental settings on **Coat**, **Yahoo**, and semi-synthetic datasets generated from **ML-100K**.

B.1 Datasets

- **Coat Shopping**: It contains a MNAR training set and a MAR testing set. Specifically, there are 6,960 five-star ratings from 290 Amazon Mechanical Turkers on an inventory of 300 coats in the training set. There are 4,640 ratings collected from the 290 workers on 16 randomly selected coats in the testing set.
- **Yahoo! R3**: It includes a MNAR training set with 311,704 five-star ratings from 15,400 users and 1,000 songs, and a MAR testing set with 54,000 ratings from 5,400 users on 10 randomly selected songs.

To make the two datasets consistent with the CVR prediction task, we further preprocess them following previous studies [10, 24]:

(1) The conversion label $r_{u,i}$ is defined as 1 if the rating of item i by user u is greater than or equal to 4, and 0 otherwise.
(2) The click indicator $o_{u,i}$ is defined as 1 if user u rated item i, and 0 otherwise.
(3) The sets of observed potential conversion labels $r_{u,i}(1)$ is denoted as $\mathbf{R}^o = \{r_{u,i}(1) \mid o_{u,i} = 1\} = \{r_{u,i} \mid o_{u,i} = 1\}$.

For both datasets, we split the corresponding MNAR dataset into a training (90%) and a validation (10%) sets, while all the MAR data is set to testing set. In addition, we restrict our samples to the users with at least one conversion behavior in the testing set as [10, 24].

B.2 Baselines

We compare our proposed methods with the following baselines:

- **Naive**: It directly uses the naive estimator as the loss function for CVR prediction.
- **IPS** [28]: It uses the inverse propensity reweighting approach to adjust the distribution of the biased training data.
- **DR-JL** [32]: It proposes a doubly robust learning model which jointly trains the imputation model and prediction model.
- **MRDR** [10]: It is the state-of-the-art model for debiasing CVR prediction, which reduces the variance of doubly robust learning method by designing a new loss for the imputation model.

For all considered methods, we follow prior work [10] to use factorization machine (FM) [20] for both CTR and CVR predictions in experiments of **Coat**, **Yahoo**, and the semi-synthetic datasets. The CTR prediction model is firstly learned with FM, and used to generate the CTR scores for inverse propensity weighting as [10, 32].

B.3 Model Implementation

We implement all models with TensorFlow [1] and optimize them with mini-batch Adam [15]. We determine the hyper-parameters of each model based on grid search, and the search ranges for the embedding size, batch size, learning rate, L2 regularization coefficient, and sample ratio of unclicked events to clicked events are set as {16, 32, 64, 128, 256}, {256, 512, 1024, 2048}, {5e-5, 1e-4, 5e-4, 1e-3, 5e-3, 1e-2}, {1e-5, 5e-5, 1e-4, 5e-4, 1e-3, 5e-3}, and {2, 4, 6, 8}, respectively. The best configuration for each method is determined based on the ranking performance on the validation set.

Appendix C EXPERIMENTAL SETTINGS ON DATASET PRODUCT

C.1 Baselines

We further provide some descriptions of the baselines as follows:

- **DCN** [31]: It is a widely used deep CTR prediction model with a naive estimator. It consists of a deep network and a cross network for feature interaction learning. It is the base model for building all other models.
- **ESMM** [19]: It is a multi-task learning model that jointly optimizes CTR prediction and CTCVR prediction.
- **DR-JL** [32]: This model is proposed for debiasing rating prediction by designing a doubly robust learning approach that jointly trains the error imputation model and prediction model. We adapt it for CVR prediction on large-scale dataset with the model architecture shown in Figure 1.
- **Multi_IPW** [39]: This model tackles the selection bias in CVR prediction with the inverse propensity weighting approach. It jointly optimizes the CTR loss and IPS based CVR loss.
- **Multi_DR** [39]: This model tackles the selection bias in CVR prediction with the doubly robust learning approach inspired by the DR-JL method.
- **MRDR** [10]: It is the state-of-the-art model for debiasing CVR prediction, which reduces the variance of DR method by designing a new loss for the imputation model. However, in the original paper, no experiments on large-scale datasets have been conducted. The original model implementation is not suitable for large-scale dataset, thus we adapt it for experiments on **Product** with the model architecture shown in Figure 1.

For DCN, we train two separate models for CTR and CVR predictions, respectively, and then combine the predictions of these two tasks to obtain the prediction of CTCVR. Besides, the prediction models of DR based methods, including DR-JL, MRDR, DR-BIAS and DR-MSE, are adapted into a multi-task learning framework presented in Figure 1 to jointly model CTR prediction and CVR prediction. In other words, the propensity estimation model is jointly learned with the prediction model to handle the data sparsity and selection bias issues.

C.2 Model Implementation

We implement all models with TensorFlow and optimize them with mini-batch Adam. For DCN, the embedding size, batch size, learning rate, keep probability of dropout, L2 regularization coefficient and L1 regularization coefficient are set to 150, 8000, 1.5e-4, 0.9, 1e-4, and 1e-8, respectively. The structure of deep network of DCN is set to [1024, 512, 64], and the number of cross layers is set to 3. Other models, including ESMM, DR-JL, Multi_IPW, Multi_DR, DR-BIAS and DR-MSE, are built upon DCN. They use similar settings as the baseline DCN for common hyper-parameters. Besides, IPS based loss suffers from the high variance issue. We clip the predicted CTR with $\max\{0.03, CTR\}$ to obtain propensity score for both IPS based methods and DR based methods to alleviate this issue. **Product** contains a training set and a testing set. We report the best results among all training epochs on the testing set of all methods in Table 4 for comparison.