Balancing Unobserved Confounding with a Few Unbiased Ratings in Debiased Recommendations

Haoxuan Li, Yanghao Xiao, Chunyuan Zheng and Peng Wu

April 25, 2024

Balancing Unobserved Confounding with a Few Unbiased Ratings in Debiased Recommendations

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

Yanghao Xiao
University of Chinese Academy of Sciences
China
xiaoyanghao22@mails.ucas.ac.cn

Chunyuan Zheng
University of California, San Diego
USA
czheng@ucsd.edu

Peng Wu*
Beijing Technology and Business University
China
pengwu@btbu.edu.cn

ABSTRACT

Recommender systems are seen as an effective tool to address information overload, but it is widely known that the presence of various biases makes direct training on large-scale observational data result in sub-optimal prediction performance. In contrast, unbiased ratings obtained from randomized controlled trials or A/B tests are considered to be the golden standard, but are costly and small in scale in reality. To exploit both types of data, recent works proposed to use unbiased ratings to correct the parameters of the propensity or imputation models trained on the biased dataset. However, the existing methods fail to obtain accurate predictions in the presence of unobserved confounding or model misspecification. In this paper, we propose a theoretically guaranteed model-agnostic balancing approach that can be applied to any existing debiasing method with the aim of combating unobserved confounding and model misspecification. The proposed approach makes full use of unbiased data by alternatively correcting model parameters learned with biased data, and adaptively learning balance coefficients of biased samples for further debiasing. Extensive real-world experiments are conducted along with the deployment of our proposal on four representative debiasing methods to demonstrate the effectiveness.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Recommender Systems; Bias; Debias; Unobserved Confounding

ACM Reference Format:
Haoxuan Li, Yanghao Xiao, Chunyuan Zheng, and Peng Wu. 2023. Balancing Unobserved Confounding with a Few Unbiased Ratings in Debiased Recommendations. In *Proceedings of the ACM Web Conference 2023 (WWW '23), April 30–May 04, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3543507.3583495

*Corresponding author.

WWW '23, April 30–May 04, 2023, Austin, TX, USA

ACM ISBN 978-1-4503-9416-1/23/04…$15.00
https://doi.org/10.1145/3543507.3583495

1 INTRODUCTION

Recommender systems (RS) are designed to accurately predict users' preferences and make personalized recommendations. In recent years, many studies have focused on deep learning for rating predictions, aiming to fit the collected data using proper deep model structures [7, 13, 15, 40]. Despite the ease of collection and large scale of observed ratings, it is known that such data always contain various biases and fail to reflect the true preferences of users [6, 22, 44]. For instance, users always choose the desired items to rate, which causes the collected ratings to be missing not at random, and training directly on those would leads to long-tail effects [1] and bias amplification [41].

To perform debiasing directly from the biased ratings, previous studies can be summarized into three categories:

- Inferring missing and biased ratings, then replacing them using pseudo-labels [31, 38]. However, the data sparsity in RS and unobserved features of users and items make it difficult to estimate those missing values accurately.
- Estimating the probability of a rating being observed, called propensity, then reweighting the observed data using the inverse propensity [35–37, 42]. However, the unobserved confounding, affecting both the missing mechanism and the ratings, makes it fail to completely eliminate the biases.
- Modeling missing mechanisms and data generating process using generative models [26]. However, it may leads to violation of model specification and data generating assumptions in the presence of unobserved variables, resulting in biased estimates.

It can be summarized that these methods would lead to biased estimates in the presence of unobserved confounding or model misspecification. To mitigate the effects of unobserved confounding, Robust Deconfounder (RD) proposes an adversarial learning that uses only biased ratings [10]. Specifically, RD assumes that the true propensity fluctuates around the nominal propensity and uses sensitivity analysis to quantify the potential impact of unobserved confounding. However, the assumption cannot be empirically verified from a data-driven way, and it is essential to relax the assumptions while reducing the bias due to the unobserved confounding and model misspecification.

In contrast to observational ratings, uniform ratings are considered the golden standard and can be obtained from A/B tests or randomized controlled trials (RCTs), but harm users' experience and are costly and time-consuming [11, 12]. Due to its small scale property, it is impractical to train prediction models directly on unbiased ratings. Recent studies propose to use a few unbiased ratings for the parameter selection of the propensity and imputation models using bi-level optimization, which has a more favorable debiasing performance compared with the RCT-free debiasing methods [5, 43]. However, we show that using unbiased ratings only to correct propensity and imputation model parameters still leads to biased predictions, in the presence of unobserved confounding or model misspecification. This motivates a more sufficient use of the unbiased ratings to combat the effects of unobserved confounding.

In this paper, we propose a model-agnostic approach to balance unobserved confounding with a few unbiased ratings. Different from the previous debiasing methods, our approach enlarges the model hypothesis space to include the unbiased ideal loss. The training objective of the balancing weights is formalized as a convex optimization problem, with balancing the loss estimation between biased and unbiased ratings as constraints. Through theoretical analysis, we prove the existence of the global optimal solution. Then, we propose an efficient training algorithm to achieve the training objectives, where the balancing weights are reparameterized and updated alternatively with the prediction model. Remarkably, the proposed balancing algorithm can be applied to any exsiting debiased recommendation methods. The main contributions of this paper are summarized as follows.

- We propose a principled balancing training objective with a few unbiased ratings for combating unmeaseured confounding in debiased recommendations.
- To optimize the objectives, we propose an efficient model-agnostic learning algorithm that alternatively updates the balancing weights and rating predictions.
- Extensive experiments are conducted on two real-world datasets to demonstrate the effectiveness of our proposal.

2 PRELIMINARIES

Let $\{u_1, u_2, \dots, u_M\}$ be a set of M users, $\{i_1, i_2, \dots, i_N\}$ be the set of N items, and $\mathcal{D} = \{(u_m, i_n) \mid m = 1, \dots, M; n = 1, \dots, N\}$ be the set of all user-item pairs. Denote $\mathbf{R} = \{r_{u,i} \mid (u,i) \in \mathcal{D}\} \in \mathbb{R}^{|\mathcal{D}|}$ be a true rating matrix, where $r_{u,i}$ is the rating of item i by user u. However, users always selectively rate items based on their interests, resulting in observed ratings, denoted as $\mathbf{R}^{\mathcal{B}} \in \mathbb{R}^{|\mathcal{B}|} (\mathcal{B} \subseteq \mathcal{D})$, are missing not at random and thus biased. For a given user-item pair (u, i), let $x_{u,i}$ be the feature vector of user u and item i, such as user gender, age, and item attributes, etc. Let $o_{u,i}$ be the binary variable indicating whether $r_{u,i}$ is observed $o_{u,i} = 1$ or missing $o_{u,i} = 0$. Given the biased ratings $\mathbf{R}^{\mathcal{B}}$, the prediction model $\hat{r}_{u,i} = f(x_{u,i}; \theta)$ in the debiased recommendation aims to predict all true ratings accurately. Ideally, it can be trained by minimizing the prediction error between the predicted rating matrix $\hat{\mathbf{R}} = \{\hat{r}_{u,i} \mid (u,i) \in \mathcal{D}\} \in \mathbb{R}^{|\mathcal{D}|}$ and the true rating matrix $\mathbf{R}$, and is given by

$$\mathcal{L}_{ideal}(\theta) = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \delta(r_{u,i}, \hat{r}_{u,i}) = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} e_{u,i}, \quad (1)$$

where $\delta(\cdot, \cdot)$ is a pre-specified loss, and $e_{u,i}$ is the prediction error, such as the squared loss $e_{u,i} = (\hat{r}_{u,i} - r_{u,i})^2$.

For unbiased estimates of the ideal loss in Eq. (1), previous studies proposed to model the missing mechanism of the biased ratings $\mathbf{R}^{\mathcal{B}}$. Formally, the probability $p_{u,i} = \Pr(o_{u,i} = 1 | x_{u,i})$ of a user u rating an item i is called propensity. The inverse probability scoring (IPS) estimator [37] is given as

$$\mathcal{L}_{IPS}(\theta) = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \frac{o_{u,i} e_{u,i}}{\hat{p}_{u,i}},$$

where $\hat{p}_{u,i} = \pi(x_{u,i}; \phi_p)$ is an estimate of the propensity $p_{u,i}$, and the IPS estimator is unbiased when $\hat{p}_{u,i} = p_{u,i}$. The doubly robust (DR) estimator [35, 42] is given as

$$\mathcal{L}_{DR}(\theta) = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i} - \hat{e}_{u,i})}{\hat{p}_{u,i}} \right],$$

where $\hat{e}_{u,i} = m(x_{u,i}; \phi_e)$ fits the prediction error $e_{u,i}$ using $x_{u,i}$, i.e., it estimates $g_{u,i} = \mathbb{E}\left[e_{u,i} \mid x_{u,i}\right]$, and DR has double robustness, i.e., it is unbiased when either $\hat{e}_{u,i} = g_{u,i}$ or $\hat{p}_{u,i} = p_{u,i}$.

In industrial scenarios, randomized controlled trials or A/B tests are considered to be the golden standard, and users might be asked to rate randomly selected items to collect unbiased ratings, denoted as $\mathbf{R}^{\mathcal{U}} \in \mathbb{R}^{|\mathcal{U}|} (\mathcal{U} \subseteq \mathcal{D})$. The ideal loss can be estimated unbiasedly by simply taking the average of the prediction errors over the unbiased ratings

$$\mathcal{L}_{\mathcal{U}}(\theta) = \frac{1}{|\mathcal{U}|} \sum_{(u,i) \in \mathcal{U}} e_{u,i} \approx \mathcal{L}_{ideal}(\theta).$$

However, unbiased ratings are costly and small in scale in reality. To exploit both types of data, recent works proposed to use unbiased ratings to correct the parameters of the propensity or imputation models trained on the biased dataset. Learning to debias (LTD) [43] and AutoDebias [5] propose to use bi-level optimization, using unbiased ratings $\mathbf{R}^{\mathcal{U}}$ to correct the propensity and imputation model parameters, and then the prediction model is trained by minimizing the IPS or DR loss estimated on the biased ratings $\mathbf{R}^{\mathcal{B}}$. Formally, this goal can be formulated as

$$\phi^* = \arg\min_{\phi} \mathcal{L}_{\mathcal{U}}\left(\theta^*(\phi); \mathcal{U}\right) \quad (2)$$

$$\text{s.t. } \theta^*(\phi) = \arg\min_{\theta} \mathcal{L}_{\mathcal{B}}(\theta, \phi; \mathcal{B}), \quad (3)$$

where $\mathcal{L}_{\mathcal{B}}$ is a pre-defined loss on the biased ratings, such as IPS with $\phi = \{\phi_p\}$, DR with $\phi = \{\phi_p, \phi_e\}$, and AutoDebias with an extra propensity that $\phi = \{\phi_{p1}, \phi_{p2}, \phi_e\}$. The bi-level optimization first performs an assumed update of $\theta(\phi)$ by Eq. (3), then updates the propensity and imputation model parameters ϕ by Eq. (2), and finally updates the prediction model parameters θ by Eq. (3).

3 PROPOSED APPROACH

We study debiased recommendations given biased ratings with a few unbiased ratings. Different from previous studies [5, 9, 14, 22–24, 37, 42, 43], we consider there may be unmesaured confounding in the biased ratings, making the unconfoundedness assumption no longer hold. In Section 3.1, we show that simply using unbiased ratings to perform model selection of propensity and imputation does not eliminate the bias from unobserved confounding and model misspecification. In Section 3.2, we propose a balancing training objective to combat the unobserved confounding and model misspecification by further exploiting unbiased ratings. In Section 3.3, we propose an efficient model-agnostic algorithm to achieve the training objective.

3.1 Motivation

First, the unbiasedness of IPS and DR requires not only that learned propensities or imputed errors are accurate, but also the unconfoundedness assumption holds, i.e., $o_{u,i} \perp\!\!\!\perp e_{u,i} \mid x_{u,i}$. However, there may exist unobserved confounding h, making $o_{u,i} \not\perp\!\!\!\perp e_{u,i} \mid x_{u,i}$ and $o_{u,i} \perp\!\!\!\perp e_{u,i} \mid (x_{u,i}, h_{u,i})$. Let $\tilde{p}_{u,i} = \Pr(o_{u,i} = 1 \mid x_{u,i}, h_{u,i})$ be the true propensity, then the nominal propensity $p_{u,i} \neq \tilde{p}_{u,i}$, and Lemma 1 states that the existing IPS and DR on $\mathbf{R}^{\mathcal{B}}$ are biased estimates of the ideal loss in the presence of unobserved confounding.

Lemma 1. *The IPS and DR estimators are biased in the presence of unobserved confounding, even the learned propensities and imputed errors are accurate, i.e., $\hat{p}_{u,i} = p_{u,i}$, $\hat{e}_{u,i} = g_{u,i}$, then*

$$\mathbb{E}[\mathcal{L}_{IPS}(\theta)] - \mathbb{E}[\mathcal{L}_{ideal}(\theta)] = \mathrm{Cov}\left(\frac{o_{u,i} - p_{u,i}}{p_{u,i}}, e_{u,i}\right) \neq 0,$$

and

$$\mathbb{E}[\mathcal{L}_{DR}(\theta)] - \mathbb{E}[\mathcal{L}_{ideal}(\theta)] = \mathrm{Cov}\left(\frac{o_{u,i} - p_{u,i}}{p_{u,i}}, e_{u,i} - g_{u,i}\right) \neq 0.$$

Proof. For DR estimator, if $\hat{p}_{u,i} = p_{u,i}, \hat{e}_{u,i} = g_{u,i}$, we have

$$\begin{aligned}\mathbb{E}[\mathcal{L}_{DR}(\theta)] &= \mathbb{E}\left[e_{u,i} + \frac{o_{u,i} - p_{u,i}}{p_{u,i}}\left(e_{u,i} - g_{u,i}\right)\right] \\ &= \mathbb{E}[\mathcal{L}_{ideal}(\theta)] + \mathbb{E}\left[\frac{o_{u,i} - p_{u,i}}{p_{u,i}}\left(e_{u,i} - g_{u,i}\right)\right] \\ &= \mathbb{E}[\mathcal{L}_{ideal}(\theta)] + \mathrm{Cov}\left(\frac{o_{u,i} - p_{u,i}}{p_{u,i}}, e_{u,i} - g_{u,i}\right).\end{aligned}$$

The last equation follows by noting that

$$\mathbb{E}\left[\frac{o_{u,i} - p_{u,i}}{p_{u,i}}\right] = \mathbb{E}\left[\mathbb{E}\left\{\frac{o_{u,i} - p_{u,i}}{p_{u,i}} \mid x_{u,i}\right\}\right] = 0,$$

and $\mathbb{E}[e_{u,i} - g_{u,i}] = 0$. In the presence of hidden confounding, $\mathrm{Cov}((o_{u,i} - p_{u,i})/p_{u,i}, e_{u,i} - g_{u,i}) \neq 0$. The conclusions of the IPS estimator can be obtained directly from taking $g_{u,i} = 0$ in DR. □

In addition, the existing methods using bi-level optimization, as shown in Eq. (2) and Eq. (3), simply uses unbiased ratings for parameter tuning of the propensity and imputation models. It follows that the prediction models in hypothesis space $\mathcal{H}_{\phi} = \{\mathcal{L}_{\mathcal{B}}(\theta, \phi) \mid \phi \in \Phi\}$ are as a subset of DR, where Φ is the parameter space of ϕ. Though the unbiased ratings correct partial bias, in the presence of unobserved confounding or model misspecification, i.e., $\mathcal{L}_{ideal} \notin \mathcal{H}_{\phi}$, it is still biased due to the limited $\mathcal{H}_{\phi}$.

Proposition 2. *The IPS and DR estimators are biased, in the presence of (a) unobserved confounding or (b) model misspecification.*

Proposition 2 concludes the biased property of IPS and DR in the presence of unobserved confounding or model misspecification.

3.2 Training Objective

To combat unobserved confounding and model misspecification on biased ratings, we propose a balancing approach to fully leverage the unbiased ratings for debiased recommendations. First, when there is no unobserved confounding, we have

$$\mathbb{E}[\mathcal{L}_{\mathcal{B}}(\theta, \phi; \mathcal{B})] = \mathbb{E}[\mathcal{L}_{\mathcal{U}}(\theta(\phi); \mathcal{U})].$$

To obtain unbiased estimates in the presence of unmeasured confounding or model misspecification, we propose to enlarge the hypothesis space to include the ideal loss, from $\mathcal{H}_{\phi}$ to $\mathcal{H}_{Bal} = \{\boldsymbol{w}^T \mathcal{L}_{\mathcal{B}}(\boldsymbol{x}; \theta, \phi) \mid \phi \in \Phi, \boldsymbol{w} \in \mathbb{R}^{|\mathcal{D}|}\}$, where $\mathcal{L}_{\mathcal{B}}(\boldsymbol{x}; \theta, \phi) \in \mathbb{R}^{|\mathcal{D}|}$ consists of the contribution of (u, i) to $\mathcal{L}_{\mathcal{B}}$. The effects of the unobserved confounding and model misspecification can be balanced through introducing the coefficients $w_{u,i}$ for each (u, i), by making

$$\mathbb{E}[\boldsymbol{w}^T \mathcal{L}_{\mathcal{B}}(\boldsymbol{x}; \theta, \phi)] = \mathbb{E}[\mathcal{L}_{\mathcal{U}}(\theta(\phi); \mathcal{U})] = \mathbb{E}[\mathcal{L}_{ideal}(\theta)]. \quad (4)$$

Proposition 3 is the empirical version of Eq. (4) in terms of the balanced IPS, DR, and AutoDebias loss.

Proposition 3. *(a) There exsits $w_{u,i} > 0, (u, i) \in \mathcal{B}$ such that*

$$\sum_{(u,i)\in\mathcal{B}} w_{u,i} \frac{e_{u,i}}{\hat{p}_{u,i}} = \frac{1}{|\mathcal{U}|} \sum_{(u,i)\in\mathcal{U}} e_{u,i}.$$

(b) There exsits $w_{u,i,1} > 0, (u, i) \in \mathcal{D}$ and $w_{u,i,2} > 0, (u, i) \in \mathcal{B}$ such that

$$\sum_{(u,i)\in\mathcal{D}} w_{u,i,1}\hat{e}_{u,i} + \sum_{(u,i)\in\mathcal{B}} w_{u,i,2} \frac{e_{u,i} - \hat{e}_{u,i}}{\hat{p}_{u,i}} = \frac{1}{|\mathcal{U}|} \sum_{(u,i)\in\mathcal{U}} e_{u,i}.$$

(c) There exsits $w_{u,i,1} > 0, (u, i) \in \mathcal{D}$ and $w_{u,i,2} > 0, (u, i) \in \mathcal{B}$ such that

$$\sum_{(u,i)\in\mathcal{D}} w_{u,i,1} \frac{\hat{e}_{u,i}}{\hat{p}_{u,i,1}} + \sum_{(u,i)\in\mathcal{B}} w_{u,i,2} \frac{e_{u,i}}{\hat{p}_{u,i,2}} = \frac{1}{|\mathcal{U}|} \sum_{(u,i)\in\mathcal{U}} e_{u,i}.$$

From Proposition 3(a), when $w_{u,i} \equiv |\mathcal{D}|^{-1}$, the left-hand side (LFS) degenerates to the standard IPS with maximal entropy of the balancing weights. The training objectives of the balanced IPS are

$$\max_{\boldsymbol{w}\in\mathbb{R}^{|\mathcal{B}|}} \sum_{(u,i)\in\mathcal{B}} w_{u,i} \log(w_{u,i}) \quad (5)$$

$$\text{s.t. } w_{u,i} > 0, \quad (u, i) \in \mathcal{B} \quad (6)$$

$$\frac{1}{|\mathcal{B}|} \sum_{(u,i)\in\mathcal{B}} w_{u,i} = \frac{1}{|\mathcal{D}|} \quad (7)$$

$$\sum_{(u,i)\in\mathcal{B}} w_{u,i} \frac{e_{u,i}}{\hat{p}_{u,i}} = \frac{1}{|\mathcal{U}|} \sum_{(u,i)\in\mathcal{U}} e_{u,i}, \quad (8)$$

where the training objective in Eq. (5) is to maximize the empirical entropy of the balancing weights and to be able to prevent extreme weights. The positivity and normality of the balancing weights are guaranteed by Eq. (6) and Eq. (7), respectively, and the influence of unobserved confounding and model misspecification is balanced out by reweighting the IPS estimates on biased ratings in Eq. (8).

Similarly, for balanced DR and AutoDebias in Proposition 3(b) and 3(c), the estimators are re-weighted by $w_{u,i,1}$ and $w_{u,i,2}$ on the entire and biased user-item pairs, respectively, to combat unobserved confounding and model misspecification. The training objectives of the balanced DR are

$$\max_{\boldsymbol{w}_1, \boldsymbol{w}_2} \sum_{(u,i)\in\mathcal{D}} w_{u,i,1}\log(w_{u,i,1}) + \sum_{(u,i)\in\mathcal{B}} w_{u,i,2}\log(w_{u,i,2}) \tag{9}$$

$$\text{s.t. } w_{u,i,1} > 0, \quad (u,i)\in\mathcal{D}, \qquad w_{u,i,2} > 0, \quad (u,i)\in\mathcal{B} \tag{10}$$

$$\sum_{(u,i)\in\mathcal{D}} w_{u,i,1} = 1, \quad \frac{1}{|\mathcal{B}|}\sum_{(u,i)\in\mathcal{B}} w_{u,i,2} = \frac{1}{|\mathcal{D}|} \tag{11}$$

$$\sum_{(u,i)\in\mathcal{D}} w_{u,i,1}\hat{e}_{u,i} + \sum_{(u,i)\in\mathcal{B}} w_{u,i,2}\frac{e_{u,i}-\hat{e}_{u,i}}{\hat{p}_{u,i}} = \frac{1}{|\mathcal{U}|}\sum_{(u,i)\in\mathcal{U}} e_{u,i}, \tag{12}$$

where $\boldsymbol{w}_1 = [w_{u,i,1} \mid (u,i)\in\mathcal{D}]$, $\boldsymbol{w}_2 = [w_{u,i,2} \mid (u,i)\in\mathcal{B}]$, and the difference in balanced AutoDebias is that Eq. (12) comes to

$$\sum_{(u,i)\in\mathcal{D}} w_{u,i,1}\frac{\hat{e}_{u,i}}{\hat{p}_{u,i,1}} + \sum_{(u,i)\in\mathcal{B}} w_{u,i,2}\frac{e_{u,i}}{\hat{p}_{u,i,2}} = \frac{1}{|\mathcal{U}|}\sum_{(u,i)\in\mathcal{U}} e_{u,i}, \tag{13}$$

where the LFS of Eq. (12) and Eq. (13) degenetates to standard DR and AutoDebias, respectively, when $w_{u,i,1} \equiv |\mathcal{D}|^{-1}$ on $\mathcal{D}$ and $w_{u,i,1} \equiv |\mathcal{D}|^{-1}$ on $\mathcal{B}$. Theorem 4 proves the existence of global optimal solutions corresponding to the proposed balanced IPS, DR and AutoDebias using Karush-Kuhn-Tucker conditions.

Theorem 4. *There exists global optimal solutions to the optimization problem in balanced IPS, DR and AutoDebias.*

Proof. Note that the empirical entropy as the optimization objectives in Eq. (5) and Eq. (9) are strictly convex. The inequality constraints in Eq. (6) and Eq. (10) are strictly feasible, i.e., there exists $w_{u,i}$ in $\mathcal{D}$ such that $w_{u,i} > 0$. The equality constraints are affine in Eq. (7), Eq. (8), Eq. (11), and Eq. (12). By the Karush-Kuhn-Tucker condition, there exist global optimal solutions. □

Theoretically, due to the convexity of the objective function, its local optimal solution is same as the global optimal solution. The generalized Lagrange multiplier method can be used to solve the primal and the dual problem, and such balancing weights can effectively combat the unobserved confounding as in Proposition 3.

3.3 Training Algorithm

Next, we propose an efficient mode-agnostic training algorithm to achieve the training objective in Section 3.2. The algorithm consists of three parts: first, training the propensity and imputation models using a bi-level optimization, *but without updating the prediction model;* then, reparameterizing and updating the gradients of the balancing weights to combat the effects of unobserved confounding and model misspecification; and finally, *minimizing the estimated balancing loss,* named Bal-IPS, Bal-DR, or Bal-AutoDebias, and updating the prediction model to achieve unbiased learning.

3.3.1 Propensity and Imputation Model Training. Different from LTD and AutoDebias that use bi-level optimization to update the prediction model, we only perform assumed updates of the prediction model parameters $\theta(\phi)$ using bi-level optimization by Eq. (3), and updates of the propensity and imputation model parameters ϕ by Eq. (2). Since there may exist unobserved confounding or model misspecification, we postpone the true update of the prediction model parameters θ to Section 3.3.3, after performing the balancing steps in Section 3.3.2. We summarized the propensity and imputation model training algorithm in Alg. 1.

Algorithm 1: Propensity and Imputation Model Training

Input: S, $\mathbf{R}^{\mathcal{B}}$, $\mathbf{R}^{\mathcal{U}}$, ϕ_0, θ_0, η
for $s = 0, \ldots, S-1$ **do**
  Sample mini-batches $\mathcal{B}_s \subseteq \mathcal{B}$ and $\mathcal{U}_s \subseteq \mathcal{U}$;
  Compute the lower loss in Eq. (3) on $\mathcal{B}_s$;
  Compute an assumed update $\theta_{s+1}(\phi_s) \leftarrow \theta_s - \eta\nabla_{\theta_s}\mathcal{L}_{\mathcal{B}}(\theta,\phi;\mathcal{B}_s)$;
  Compute the upper loss in Eq. (2) on $\mathcal{U}_s$;
  Update the propensity and imputation model $\phi_{s+1} \leftarrow \phi_s - \eta\nabla_{\phi_s}\mathcal{L}_{\mathcal{U}}(\theta_{s+1}(\phi);\mathcal{B}_s)$;
end
Output: ϕ_S

3.3.2 Balancing Unobserved Confounding Training. One challenge in solving the balancing optimization problem is that as the number of user-item pairs increases, the number of balancing weights also increases, resulting in a significant increase in solution time for large-scale datasets. To address this issue, we propose to *reparameterize* $w_{u,i}$ in the balanced IPS, i.e., $w_{u,i} = g(x_{u,i};\xi)$, where ξ is the balancing model parameter. To satisfy the optimization constraints Eq. (6) and Eq. (7), the last layer of $g(x_{u,i};\xi)$ uses Sigmoid as the activation function to guarantee positivity and batch normalization to guarantee normality. The balancing weights in the balanced IPS are trained by minimizing the negative empirical entropy with the violation of the balanced constraint Eq. (8) as regularization

$$\mathcal{L}_{W-IPS}(\xi) = -\sum_{(u,i)\in\mathcal{B}} w_{u,i}\log(w_{u,i}) + \lambda\left[\sum_{(u,i)\in\mathcal{B}} w_{u,i}\frac{e_{u,i}}{\hat{p}_{u,i}} - \frac{1}{|\mathcal{U}|}\sum_{(u,i)\in\mathcal{U}} e_{u,i}\right]^2,$$

where $\lambda > 0$ is a hyper-parameter, for trade-off the original loss estimation with the correction due to the unobserved confounding.

Similarly, $w_{u,i,1}$ and $w_{u,i,2}$ in the balanced DR and balanced AutoDebias are also reparameterized as $w_{u,i,1} = g(x_{u,i};\xi_1)$ and $w_{u,i,2} = g(x_{u,i};\xi_2)$. The balancing weights in the balanced DR and balanced AutoDebias are trained by minimizing

$$\mathcal{L}_{W-DR}(\xi) = -\sum_{(u,i)\in\mathcal{D}} w_{u,i,1}\log(w_{u,i,1}) - \sum_{(u,i)\in\mathcal{B}} w_{u,i,2}\log(w_{u,i,2}) + \lambda\left[\sum_{(u,i)\in\mathcal{D}} w_{u,i,1}\hat{e}_{u,i} + \sum_{(u,i)\in\mathcal{B}} w_{u,i,2}\frac{e_{u,i}-\hat{e}_{u,i}}{\hat{p}_{u,i}} - \frac{1}{|\mathcal{U}|}\sum_{(u,i)\in\mathcal{U}} e_{u,i}\right]^2,$$

and

$$\mathcal{L}_{W-Auto}(\xi) = -\sum_{(u,i)\in\mathcal{D}} w_{u,i,1}\log(w_{u,i,1}) - \sum_{(u,i)\in\mathcal{B}} w_{u,i,2}\log(w_{u,i,2}) + \lambda\left[\sum_{(u,i)\in\mathcal{D}} w_{u,i,1}\frac{\hat{e}_{u,i}}{\hat{p}_{u,i,1}} + \sum_{(u,i)\in\mathcal{B}} w_{u,i,2}\frac{e_{u,i}}{\hat{p}_{u,i,2}} - \frac{1}{|\mathcal{U}|}\sum_{(u,i)\in\mathcal{U}} e_{u,i}\right]^2,$$

where $\lambda > 0$ is a hyper-parameter, and $\xi \equiv \{\xi_1, \xi_2\}$ are the parameters of the balancing model.

Algorithm 2: Balancing Unobserved Confounding Training

Input: T, S, $\mathbf{R}^{\mathcal{B}}$, $\mathbf{R}^{\mathcal{U}}$, ϕ_0, θ_0, ξ_0, η, λ

for $t = 0, \dots, T-1$ **do**

  Call Alg. 1 by $\phi_{t+1} \leftarrow$ Alg. 1$(S, \mathbf{R}^{\mathcal{B}}, \mathbf{R}^{\mathcal{U}}, \phi_t, \theta_t, \eta)$;

  for $s = 0, \dots, S-1$ **do**

    Sample mini-batches $\mathcal{D}_t^s \subseteq \mathcal{D}$, $\mathcal{B}_t^s \subseteq \mathcal{B}$ and $\mathcal{U}_t^s \subseteq \mathcal{U}$;

    Compute unmeasured confounding balancing loss;

    Update the balancing weight $\xi_t^{s+1} \leftarrow \xi_t^s - \eta \nabla_{\xi_t^s} \mathcal{L}_W(\xi)$;

    Compute the balanced prediction error loss;

    Update the prediction model $\theta_t^{s+1} \leftarrow \theta_t^s - \eta \nabla_{\theta_t^s} \mathcal{L}_{Bal}(\theta)$;

  end

  Copy the balancing model's parameters $\xi_{t+1}^0 \leftarrow \xi_t^S$;

  Copy the prediction model's parameters $\theta_{t+1}^0 \leftarrow \theta_t^S$;

end

Output: θ_T

3.3.3 Prediction Model Training. Since the optimization of the balancing weights aims to balance the prediction errors on the biased and unbiased ratings, which also depends on the prediction model, we propose to update the balancing model and the prediction model alternatively. Specifically, given the balancing weights of IPS, the prediction model is trained by minimizing the balanced IPS (Bal-IPS)

$$\mathcal{L}_{Bal-IPS}(\theta) = \sum_{(u,i)\in\mathcal{B}} w_{u,i} \frac{e_{u,i}}{\hat{p}_{u,i}}. \tag{14}$$

Similarly, for balanced DR (Bal-DR) or balanced AutoDebias (Bal-AutoDebias), the prediction model is trained by minimizing

$$\mathcal{L}_{Bal-DR}(\theta) = \sum_{(u,i)\in\mathcal{D}} w_{u,i,1}\hat{e}_{u,i} + \sum_{(u,i)\in\mathcal{B}} w_{u,i,2} \frac{e_{u,i} - \hat{e}_{u,i}}{\hat{p}_{u,i}}, \tag{15}$$

or

$$\mathcal{L}_{Bal-Auto}(\theta) = \sum_{(u,i)\in\mathcal{D}} w_{u,i,1} \frac{\hat{e}_{u,i}}{\hat{p}_{u,i,1}} + \sum_{(u,i)\in\mathcal{B}} w_{u,i,2} \frac{e_{u,i}}{\hat{p}_{u,i,2}}. \tag{16}$$

Next, given the prediction model, the balancing weights are updated again as described in Section 3.3.2. The balancing weights and the prediction model are updated alternately, allowing a more adequate use of unbiased ratings, resulting in unbiased learning of the prediction model.

The main difference compared with LTD [43] and AutoDebias [5] is that we do not only use unbiased ratings to select the parameters of the propensity and imputation models, and then use standard IPS or DR for the prediction model update. Instead, we combat the effects of unobserved confounding by introducing a balancing model, and then perform prediction model updates based on the balanced losses. Remarkably, the proposed method is model-agnostic and can be applied to any of the debiased recommendation methods. Here we use IPS, DR and AutoDebias for illustration. We summarized the whole training algorithm in Alg. 2.

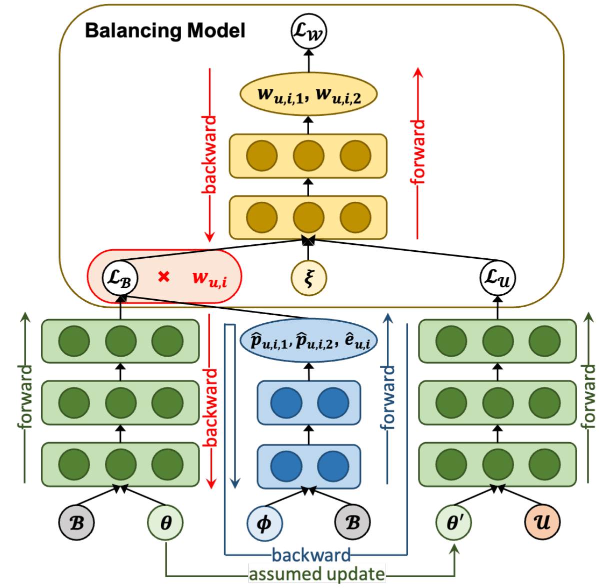


Figure 1: The proposed workflow for balancing unobserved confounding consists of four steps: (1) assumed updating the prediction model parameters from $\theta(\phi)$ to $\theta'(\phi)$ using $\mathbf{R}^{\mathcal{B}}$ (green arrow); (2) updating the propensity and imputation model parameters ϕ using $\mathbf{R}^{\mathcal{U}}$ (blue arrow); (3) updating the balancing model parameters ϕ using both $\mathbf{R}^{\mathcal{B}}$ and $\mathbf{R}^{\mathcal{U}}$ (red arrow); (4) actually updating the prediction model parameters θ using the balanced loss $w^T\mathcal{L}_{\mathcal{B}}$ (red arrow).

Table 1: Summary of the datasets.

	Users	Items	Training	Uniform	Validation	Test
Music	15,400	1,000	311,704	2,700	2,700	48,600
Coat	290	300	6,960	232	232	4,176

3.3.4 Training Efficiency. The proposed workflow for balancing the unobserved confounding is shown in Figure 1. In Section 3.3.1, our algorithm performs two forward and backward passes for the prediction model on $\mathbf{R}^{\mathcal{B}}$ and $\mathbf{R}^{\mathcal{U}}$, respectively, and one forward and backward pass for the propensity and imputation model on $\mathbf{R}^{\mathcal{B}}$. The backward-on-backward pass is used to obtain the gradients of the propensity and imputation models. In Section 3.3.2, one forward and one reverse pass are performed for the balancing model. In Section 3.3.3, a backward pass is used to actually update the prediction model. We refer to [33, 43] that the running time of a backward-on-backward pass and a forward pass are about the same. As a result, the training time of the proposed algorithm does not exceed 3x learning time compared to two-stage learning and about 1.5x learning time compared to LTD and AutoDebias.

4 REAL-WORLD EXPERIMENTS

In this section, We conduct extensive experiments on two real-world datasets to answer the following research questions (RQs):

RQ1. Do the proposed Bal-methods improve the debiasing performance compared with the existing methods?

Table 2: Performance comparison in terms of AUC, NDCG@5, and NDCG@10. The best results to each base method are bolded.

Method	Music						Coat					
	AUC	RI	NDCG@5	RI	NDCG@10	RI	AUC	RI	NDCG@5	RI	NDCG@10	RI
CausE	0.731	-	0.551	-	0.656	-	0.761	-	0.500	-	0.605	-
KD-Label	0.740	-	0.580	-	0.680	-	0.750	-	0.504	-	0.610	-
MF (biased)	0.727	-	0.550	-	0.655	-	0.747	-	0.500	-	0.606	-
MF (uniform)	0.573	-	0.449	-	0.591	-	0.579	-	0.358	-	0.482	-
MF (combine)	0.730	-	0.554	-	0.659	-	0.750	-	0.503	-	0.611	-
Bal-MF	**0.739**	1.23%	**0.579**	4.51%	**0.679**	3.03%	**0.761**	1.47%	**0.511**	1.59%	**0.620**	1.47%
IPS	0.723	-	0.549	-	0.656	-	0.760	-	0.509	-	0.613	-
Bal-IPS	**0.727**	0.55%	**0.564**	2.73%	**0.668**	1.83%	**0.771**	1.45%	**0.521**	2.36%	**0.628**	2.45%
DR	0.724	-	0.550	-	0.656	-	0.765	-	0.521	-	0.620	-
Bal-DR	**0.731**	0.97%	**0.569**	3.45%	**0.669**	1.98%	**0.770**	0.65%	**0.523**	0.38%	**0.628**	1.29%
AutoDebias	0.741	-	0.645	-	0.725	-	0.766	-	0.522	-	0.621	-
Bal-AutoDebias	**0.749**	1.08%	**0.670**	3.88%	**0.744**	2.62%	**0.772**	0.78%	**0.544**	4.21%	**0.640**	3.06%

Note: RI refers to the relative improvement of Bal-methods over the corresponding baseline.

RQ2. Do our methods stably perform well with different initializations of the prediction model?
RQ3. How does the balancing model affect the performance of our methods?
RQ4. What factors influence the effectiveness of our methods?

4.1 Experimental Setup

Dataset and preprocessing. Following the previous studies [5, 35, 42, 43], we conduct extensive experiments on the two widely used real-world datasets with both missing-not-at-random (MNAR) and missing-at-random (MAR) ratings: **Music**[1] and **Coat**[2]. In particular, Music dataset contains 15,400 users and 1,000 items with 54,000 MAR and 311,704 MNAR ratings. Coat dataset contains 290 users and 300 items with 4,640 MAR and 6,960 MNAR ratings. Following [5, 28], we take all the biased data as the training set and randomly split the uniform data as three parts: 5% for balancing the unobserved confounding, 5% for validation set and 90% for test set. We summarize the datasets and splitting details in Table 1.

Baselines. In our experiments, we compare the proposed Bal-methods with the following baselines:

• **Base Model** [21]: the Matrix Factorization (MF) model is trained on biased data, uniform data and both of them respectively, denoted as MF (biased), MF (uniform) and MF (combine).
• **Inverse Propensity Scoring (IPS)** [37]: a reweighting method using inverse propensity scores to weight the observed events.
• **Doubly Robust (DR)** [35, 42]: an efficient method combining imputations and inverse propensities with double robustness.
• **CausE** [28]: a sample-based knowledge distillation approach to reduce computational complexity.
• **KD-Label** [28]: an efficient framework for knowledge distillation to transfer unbiased information to teacher model and guide the training of student model.
• **AutoDebias** [5]: a meta-learning based method using few unbiased data to further mitigate the selection bias.

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

Experimental protocols and details. Following [5, 43], AUC, NDCG@5 and NDCG@10 are adopted as the evaluation metrics to measure the debiasing performance. Formally,

$$AUC = \frac{\sum_{(u,i)\in\mathcal{U}^+} \hat{Z}_{u,i} - |\mathcal{U}^+| \cdot (|\mathcal{U}^+| + 1)/2}{|\mathcal{U}^+| \cdot (|\mathcal{U}| - |\mathcal{U}^+|)},$$

and NDCG@k measures the quality of ranking list as

$$DCG_u@k = \sum_{(u,i)\in\mathcal{U}} \frac{\mathbb{I}(\hat{Z}_{u,i} \le k)}{\log(\hat{Z}_{u,i} + 1)}, \quad NDCG@k = \frac{1}{M}\sum_{m=1}^{M} \frac{DCG_{u_m}@k}{IDCG_{u_m}@k},$$

where $\mathcal{U}^+ \subseteq \mathcal{U}$ denotes the positive ratings in the uniform dataset, $\hat{Z}_{u,i}$ is the rank position of (u,i) given by the rating predictions, and $IDCG_{u_m}@k$ is the ideal $DCG_{u_m}@k$.

All the methods are implemented on PyTorch. Throughout, Adam optimizer is utilized for propensity and imputation model with learning rate and weight decay in [1e-4, 1e-2]. SGD optimizer is utilized for prediction model and balancing model with learning rate in [1e-7, 1] and weight decay in [1e-4, 1]. We tune the regularization hyper-parameter λ in $\{0, 2^{-9}, 2^{-6}, 2^{-3}, 1\}$. All hyper-parameters are tuned based on the performance on the validation set.

4.2 Real World Performance Comparison (RQ1)

Table 2 compares the prediction performance of the various methods on two real-world datasets Music and Coat. We find that the proposed model-agnostic Bal-methods have significantly improved performance when applied to MF, IPS, DR and AutoDebias with respect to all metrics. Overall, Bal-AutoDebias exhibits the best performance. Impressively, although AutoDebias hardly improves the performance on Coat compared with DR as reported in [5], the proposed Bal-AutoDebias improves 4.21% and 3.06% on NDCG@5 and NDCG@10 compared with the best baseline, respectively, validating the effectiveness of the proposed balancing approach.

In addition, MF using only uniform data exhibits the worst performance, due to its small size which causes unavoidable overfitting. Directly combining the biased and unbiased ratings increases the MF performance slightly and insignificantly. As in [5], AutoDebias

Table 3: Performance of the Bal-methods under different prediction models as initializations on Music and Coat.

	Initial Method	Initial with IPS			Initial with DR			Initial with AutoDebias		
Dataset	Method	AUC	NDCG@5	NDCG@10	AUC	NDCG@5	NDCG@10	AUC	NDCG@5	NDCG@10
Music	Baseline	0.723	0.549	0.656	0.724	0.550	0.656	0.741	0.645	0.725
	Bal-IPS	$0.726_{0.4\%\uparrow}$	$0.561_{2.2\%\uparrow}$	$0.666_{1.5\%\uparrow}$	$0.726_{0.3\%\uparrow}$	$0.562_{2.2\%\uparrow}$	$0.666_{1.5\%\uparrow}$	$0.747_{0.8\%\uparrow}$	$0.656_{1.7\%\uparrow}$	$0.733_{1.1\%\uparrow}$
	Bal-DR	$0.725_{0.3\%\uparrow}$	$0.556_{1.3\%\uparrow}$	$0.665_{1.4\%\uparrow}$	$0.726_{0.3\%\uparrow}$	$0.559_{1.6\%\uparrow}$	$0.667_{1.7\%\uparrow}$	$0.748_{0.9\%\uparrow}$	$0.658_{2.0\%\uparrow}$	$0.734_{1.2\%\uparrow}$
	Bal-AutoDebias	$\mathbf{0.739}_{2.2\%\uparrow}$	$\mathbf{0.584}_{6.4\%\uparrow}$	$\mathbf{0.683}_{4.1\%\uparrow}$	$\mathbf{0.740}_{2.2\%\uparrow}$	$\mathbf{0.586}_{6.5\%\uparrow}$	$\mathbf{0.684}_{4.3\%\uparrow}$	$\mathbf{0.749}_{1.1\%\uparrow}$	$\mathbf{0.670}_{3.9\%\uparrow}$	$\mathbf{0.744}_{2.6\%\uparrow}$
Coat	Baseline	0.760	0.509	0.613	0.765	0.521	0.620	0.766	0.522	0.621
	Bal-IPS	$\mathbf{0.771}_{1.4\%\uparrow}$	$0.521_{2.4\%\uparrow}$	$0.628_{2.4\%\uparrow}$	$0.770_{0.7\%\uparrow}$	$0.523_{0.4\%\uparrow}$	$0.627_{1.1\%\uparrow}$	$0.770_{0.5\%\uparrow}$	$0.523_{0.2\%\uparrow}$	$0.629_{1.3\%\uparrow}$
	Bal-DR	$0.770_{1.3\%\uparrow}$	$0.523_{2.8\%\uparrow}$	$0.628_{2.4\%\uparrow}$	$0.771_{0.8\%\uparrow}$	$0.522_{0.2\%\uparrow}$	$0.629_{1.5\%\uparrow}$	$0.770_{0.5\%\uparrow}$	$0.523_{0.2\%\uparrow}$	$0.629_{1.3\%\uparrow}$
	Bal-AutoDebias	$\mathbf{0.771}_{1.4\%\uparrow}$	$\mathbf{0.531}_{4.3\%\uparrow}$	$\mathbf{0.632}_{3.1\%\uparrow}$	$\mathbf{0.772}_{0.9\%\uparrow}$	$\mathbf{0.539}_{3.5\%\uparrow}$	$\mathbf{0.637}_{2.7\%\uparrow}$	$\mathbf{0.772}_{0.8\%\uparrow}$	$\mathbf{0.544}_{4.2\%\uparrow}$	$\mathbf{0.640}_{3.1\%\uparrow}$

Table 4: Effects of balancing models on Bal-AutoDebias.

Method		Music			Coat		
$w_{u,i,1}$	$w_{u,i,2}$	AUC	NDCG@5	NDCG@10	AUC	NDCG@5	NDCG@10
MF	MF	0.749	0.670	0.744	0.772	0.544	0.640
MF	NCF	0.745	0.667	0.742	0.769	0.539	0.635
NCF	MF	**0.762**	**0.675**	**0.748**	**0.774**	**0.548**	**0.646**
NCF	NCF	0.749	0.671	0.745	0.771	0.545	0.639

has the most competitive performance among the existing methods, due to the use of unbiased ratings for the parameter selection of the propensity and imputation models. However, as discussed in previous sections, the previous methods were unable to combat the potential unobserved confounding in the biased data. The proposed Bal-methods address this issue by further utilizing unbiased ratings to balance the loss estimates from biased ratings.

4.3 In-depth Analysis (RQ2)

We further conduct an in-depth analysis by using the pre-trained prediction model parameters given by IPS, DR and AutoDebias as initialization in Alg. 2, respectively, to verify that the proposed Bal-methods can be effectively applied to any existing debiasing methods. The results are presented in Table 3. We find that all Bal-methods show significant performance improvement in all metrics compared to the pre-trained prediction models. Notably, applying the Bal-methods to any initialized predictions can stably boost the performance compared with AutoDebias on Coat, which can be explained by the possible presence of unobserved confounding and model misspecification in the biased data, while our method can mitigate the potential bias via a model-agnostic manner.

4.4 Ablation Study (RQ3)

To explore the impact of the proposed balancing model on the debiasing performance, we conduct ablation experiments using varying regularization hyperparameters λ for trade-offs between the original loss estimation and the correction due to the unobserved confounding. Note that when $\lambda = 0$, the globally optimal balancing weights equal to $1/|\mathcal{D}|$ with maximum entropy, degenerating to the standard IPS, DR and AutoDebias. We tune λ in {0, 2^{-9}, 2^{-6}, 2^{-3}, 1} on Bal-IPS, Bal-DR and Bal-AutoDebias, and the results are shown in Figure 2, where the black dashed line is used as the most competitive baseline for reference. We find that the AUC and NDCG@K of all methods first increase and then decrease

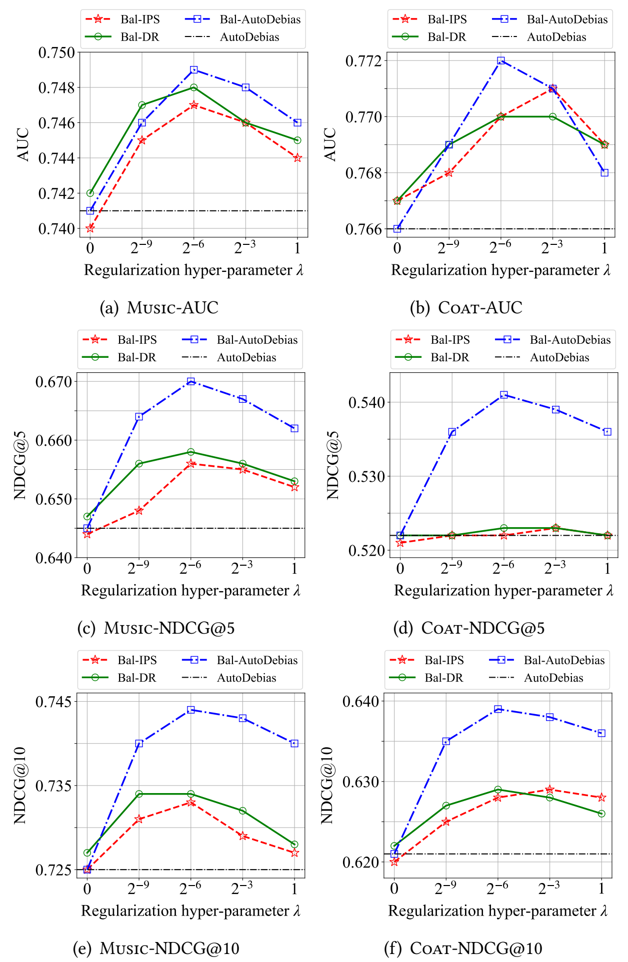
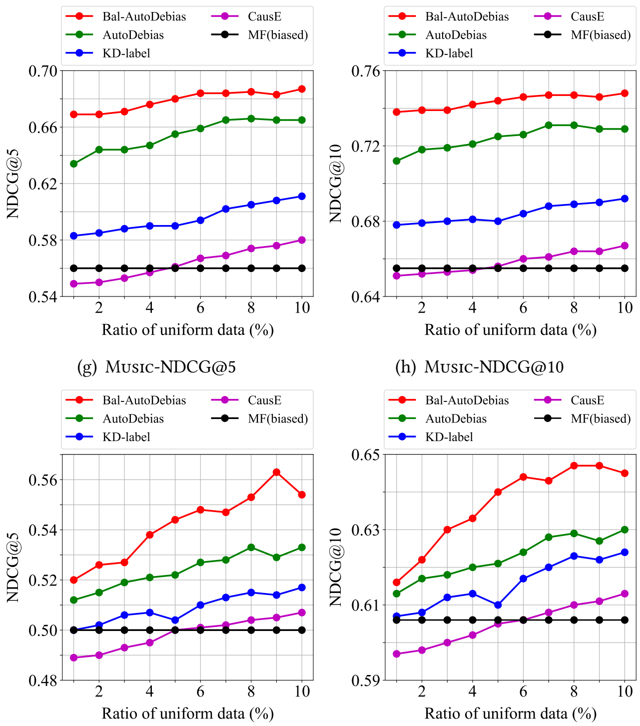


Figure 2: Effect of regularization strength λ on Music and Coat, degenerating to standard AutoDebias when $\lambda = 0$.

with the increasing constraint strength, with optimal performance around $\lambda = 2^{-6}$. This is interpreted as the best tradeoff between estimated loss and unobserved confounding. All methods using $\lambda > 0$ stably outperform the standard AutoDebias and the case without considering unobserved confounding, i.e., $\lambda = 0$, so it can be concluded that the proposed balancing model plays an important role in the debiasing.



(g) Music-NDCG@5 (h) Music-NDCG@10

(i) Coat-NDCG@5 (j) Coat-NDCG@10

Figure 3: Effect of varying size of uniform data.

4.5 Exploratory Analysis (RQ4)

Effect of balancing model selections. We further explore the effect of model selections on the balanced weights to the debiasing performance. Specifically, we take different combinations of MF and NCF as balancing models for $w_{u,i,1}$ on $\mathcal{D}$ and $w_{u,i,2}$ on $\mathcal{B}$, and the results are shown in Table 4. The performance can be significantly improved when NCF and MF are used to model $w_{u,i,1}$ and $w_{u,i,2}$, respectively. We argue that the main reason is that $|\mathcal{D}| \gg |\mathcal{B}|$, leading to a reasonable reparameterization of $w_{u,i,1}$ using deep models (e.g., NCF), and $w_{u,i,2}$ using simple models (e.g., MF).

Effect of uniform data size. Figure 3 shows the sensitivity of the debiasing methods to the size of the uniform data ranging from 1% to 10%. We find that the proposed Bal-AutoDebias stably outperforms the existing methods for varying sizes of unbiased ratings. For the previous methods, AutoDebias has a more competitive performance compared with KD-label and CausE. When providing with a small size (e.g., 1%) of the unbiased ratings, CausE performs even worse than the biased MF, while Bal-AutoDebias achieves the optimal performance. Compared with AutoDebias, our methods make significant improvements on both NDCG@5 and NDCG@10, validating the effectiveness of the proposed balancing learning.

5 RELATED WORK

Debiased Recommendation. Recommender algorithms are often trained based on the historical interactions. However, the historical data cannot fully represent the user's true preference [6, 44], because user behavior is affected by various factors, such as conformity [29] and item popularity [46], etc. Many methods were developed for achieving unbiased learning, aiming to capture the true user preferences with biased data. For example, [37] noticed the missing data problem in RS and recommended using the IPS strategy to remove the bias, [42] designed a doubly robust (DR) loss and suggested adopting the joint learning method for model training. Subsequently, several approaches enhanced the DR method by pursuing a better bias-variance trade-off [9, 14], leveraging parameter sharing and multi-task learning technique [30, 39, 45], combing a small uniform dataset [4, 5, 28, 43], addressing the problem of small propensities and weakening the reliance on extrapolation [24], and reducing bias and variance simultaneously when the imputed errors are less accurate [23]. In addition, [22] proposed a multiple robust learning method that allows the use of multiple candidate propensity and imputation models and is unbiased when any of the propensity or imputation models is accurate. [6, 44] reviewed the recent progress in debiased recommendation. To mitigate the effects of unobserved confounding, [10] proposed an adversarial learning method that uses only biased ratings. Unlike the existing methods, this paper combats the effect of unmeasured confounding with a small uniform dataset to achieve exact unbiasedness.

Causal Inference under Unmeasured Confounding. Unmeasured confounding is a difficult problem in causal inference and the main strategies for addressing it can be divided into two classes [3, 17, 19, 25]. One is the sensitivity analysis [8, 20, 34] that seeks bounds for the true causal effects with datasets suffering from unmeasured confounders. The other class methods aim to obtain unbiased causal effect estimators by leveraging some auxiliary information, such as instrument variable methods [2, 16], front door adjustment [32], and negative control [27]. In general, finding a reliable instrument variable or a mediator that satisfies the front door criterion [16, 18] is a challenging task in practice. Different from these methods based on an observational dataset, this paper considers a more practical scenario in debiased recommendations, i.e., addressing unmeasured confounding by fully exploiting the unbiasedness property of a small uniform dataset.

6 CONCLUSION

This paper develops a method for balancing unobserved confounding with few unbiased ratings. We first show theoretically that previous methods that simply using unbiased ratings to select propensity and imputation model parameters is not sufficient to combat the effects of unobserved confounding and model misspecification. We then propose a balancing optimization training objective, and further propose a model-agnostic training algorithm to achieve the training objective using reparameterization techniques. The balancing model is alternately updated with the prediction model to combat the effect of unobserved confounding. We conduct extensive experiments on two real-world datasets to demonstrate the superiority of the proposed approach. To the best of our knowledge, this is the first paper using a few unbiased ratings to combat the effects of unobserved confounding in debiased recommendations. For future works, we will derive theoretical generalization error bounds for the balancing approaches, as well as explore more effective ways to leverage the unbiased ratings to enhance the debiasing performance of the prediction models.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (No. 2018YFB1701500 and No. 2018YFB1701503).

REFERENCES

[1] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2017. Controlling popularity bias in learning-to-rank recommendation. In *RecSys*.

[2] Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. 1996. Identification of Causal Effects Using Instrumental Variables. *J. Amer. Statist. Assoc.* 91 (1996), 444–455.

[3] Elias Bareinboim and Judea Pearl. 2016. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences* 113, 27 (2016), 7345–7352.

[4] Stephen Bonner and Flavian Vasile. 2018. Causal embeddings for recommendation. In *RecSys*.

[5] 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*.

[6] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2020. Bias and Debias in Recommender System: A Survey and Future Directions. *arXiv:2010.03240* (2020).

[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] H Christopher Frey and Sumeet R Patil. 2002. Identification and review of sensitivity analysis methods. *Risk analysis* (2002), 553–578.

[9] Quanyu Dai, Haoxuan Li, Peng Wu, Zhenhua Dong, Xiao-Hua Zhou, Rui Zhang, Xiuqiang He, Rui Zhang, and Jie Sun. 2022. A Generalized Doubly Robust Learning Framework for Debiasing Post-Click Conversion Rate Prediction. In *KDD*.

[10] Sihao Ding, Peng Wu, Fuli Feng, Xiangnan He, Yitong Wang, Yong Liao, and Yongdong Zhang. 2022. Addressing Unmeasured Confounder for Recommendation with Sensitivity Analysis. In *KDD*.

[11] Alexandre Gilotte, Clément Calauzènes, Thomas Nedelec, Alexandre Abraham, and Simon Dollé. 2018. Offline A/B Testing for Recommender Systems. In *WSDM*.

[12] Alois Gruson, Praveen Chandar, Christophe Charbuillet, James McInerney, Samantha Hansen, Damien Tardieu, and Ben Carterette. 2019. Offline Evaluation to Make Decisions About PlaylistRecommendation Algorithms. In *WSDM*.

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

[14] 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*.

[15] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *WWW*.

[16] Miguel A. Hernán and James M. Robins. 2020. *Causal Inference: What If.* Boca Raton: Chapman and Hall/CRC.

[17] Paul Hünermund and Elias Bareinboim. 2019. Causal inference and data fusion in econometrics. *arXiv:1912.09104* (2019).

[18] Guido W Imbens. 2020. Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature* (2020), 1129–79.

[19] Nathan Kallus, Aahlad Manas Puli, and Uri Shalit. 2018. Removing hidden confounding by experimental grounding. In *NeurIPS*.

[20] Nathan Kallus and Angela Zhou. 2018. Confounding-Robust Policy Improvement. In *NeurIPS*.

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

[22] Haoxuan Li, Quanyu Dai, Yuru Li, Yan Lyu, Zhenhua Dong, Xiao-Hua Zhou, and Peng Wu. 2023. Multiple Robust Learning for Recommendation. In *AAAI*.

[23] Haoxuan Li, Yan Lyu, Chunyuan Zheng, and Peng Wu. 2023. TDR-CL: Targeted Doubly Robust Collaborative Learning for Debiased Recommendations. In *ICLR*.

[24] Haoxuan Li, Chunyuan Zheng, and Peng Wu. 2023. StableDR: Stabilized Doubly Robust Learning for Recommendation on Data Missing Not at Random. In *ICLR*.

[25] Xinyu Li, Wang Miao, Fang Lu, and Xiao-Hua Zhou. 2021. Improving efficiency of inference in clinical trials with external control data. *Biometrics* (2021).

[26] Dawen Liang, Laurent Charlin, James McInerney, and David M Blei. 2016. Modeling user exposure in recommendation. In *WWW*.

[27] Marc Lipsitch, Eric Tchetgen Tchetgen, and Ted Cohen. 2011. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* (2011).

[28] 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*.

[29] Yiming Liu, Xuezhi Cao, and Yong Yu. 2016. Are You Influenced by Others When Rating? Improve Rating Prediction by Conformity Modeling. In *RecSys*.

[30] 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*.

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

[32] Judea Pearl. 2009. *Causality: Models, Reasoning, and Inference* (second ed.). Cambridge University Press.

[33] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. 2018. Learning to Reweight Examples for Robust Deep Learning. In *ICML*.

[34] Paul R. Rosenbaum. 2020. *Design of Observational Studies* (second ed.). Springer Nature Switzerland AG.

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

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

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

[38] Harald Steck. 2010. Training and testing of recommender systems on data missing not at random. In *KDD*.

[39] Hao Wang, Tai-Wei Chang, Tianqiao Liu, Jianmin Huang, Zhichao Chen, Chao Yu, Ruopeng Li, and Wei Chu. 2022. ESCM^2: Entire Space Counterfactual Multi-Task Model for Post-Click Conversion Rate Estimation. *arXiv:2204.05125* (2022).

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

[41] Wenjie Wang, Fuli Feng, Xiangnan He, Xiang Wang, and Tat-Seng Chua. 2021. Deconfounded recommendation for alleviating bias amplification. In *KDD*.

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

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

[44] 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. In *IJ-CAI*.

[45] 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*.

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