Towards Resolving Propensity Contradiction
in Offline Recommender Learning

Yuta Saito[1] , **Masahiro Nomura**[2]

[1]Cornell University
[2]CyberAgent, Inc.
ys552@cornell.edu, nomura_masahiro@cyberagent.co.jp

Abstract

We study offline recommender learning from explicit rating feedback in the presence of *selection bias*. A current promising solution for dealing with the bias is the *inverse propensity score* (IPS) estimation. However, the existing propensity-based methods can suffer significantly from the propensity estimation bias. In fact, most of the previous IPS-based methods require some amount of *missing-completely-at-random* (MCAR) data to accurately estimate the propensity. This leads to a critical ***self-contradiction***; IPS is ineffective without MCAR data, even though it originally aims to learn recommenders from only *missing-not-at-random* feedback. To resolve this ***propensity contradiction***, we derive a ***propensity-independent generalization error bound*** and propose a novel algorithm to minimize the theoretical bound via adversarial learning. Our theory and algorithm do not require a propensity estimation procedure, thereby leading to a well-performing rating predictor without the true propensity information. Extensive experiments on real-world datasets demonstrate that the proposed algorithm is superior to a range of existing methods both in rating prediction and ranking metrics in practical settings without MCAR data. Full version of the paper (including the appendix) is available at: https://arxiv.org/abs/1910.07295.

1 Introduction

In industrial recommender systems, it is essential to obtain a well-performing rating predictor from sparse rating feedback to recommend items relevant to users. An important challenge is that most of the missing mechanisms of real-world explicit rating data are *missing-not-at-random* (*MNAR*) [Marlin and Zemel, 2009; Schnabel *et al.*, 2016; Wang *et al.*, 2019; Chen *et al.*, 2020]. There are two main causes behind the MNAR mechanism. The first is the past recommendation policy. Suppose that we rely on a policy that recommends popular items with high probabilities. Then, the observed ratings under that policy include more data of popular items [Bonner and Vasile, 2018]. The second is self-selection by users. It is well-known that the users

tend to rate items for which they exhibit positive preferences, and the ratings for items with negative preferences are more likely to be missing [Marlin and Zemel, 2009; Schnabel *et al.*, 2016].

1.1 Open Problem: Propensity Contradiction

Selection bias makes it difficult to learn rating predictors, as naive methods typically result in suboptimal and biased recommendations with MNAR data [Schnabel *et al.*, 2016; Steck, 2010; Wang *et al.*, 2019]. One of the most established solutions to this problem is the propensity-based approach. It defines the probability of each feedback being observed as a *propensity score* and obtains an unbiased estimator for the true metric via inverse propensity weighting [Liang *et al.*, 2016; Schnabel *et al.*, 2016; Wang *et al.*, 2019]. In general, unbiasedness of IPS is desirable, however, it is valid only when the true propensity score is available. Previous studies utilize some amount of *missing-completely-at-random* (*MCAR*) test data to estimate the propensity score and ensure their empirical performance [Schnabel *et al.*, 2016; Wang *et al.*, 2019]. In most real-world recommender systems, however, the true propensity score is unknown, and MCAR data are unavailable as well, resulting in a severe bias in the estimation of the loss function of interest. Thus, there remains a critical ***contradiction***: the performance of the IPS variants relies heavily on mostly inaccessible MCAR data, although they originally aim at training recommenders using only MNAR data.

Two previous methods tackle the challenge of those propensity-based methods. The first one is *causal embeddings* (CausE) [Bonner and Vasile, 2018]. It introduces a domain regularization term to address the bias. However, CausE requires some amount of MCAR data by its design; it cannot be generalized to a realistic setting with only MNAR training data. Moreover, its domain regularization technique is a heuristic approach and lacks a theoretical guarantee. The other method is to use *1-bit matrix completion* (1BitMC) [Ma and Chen, 2019] to estimate propensity scores using only MNAR training data, along with a theoretical guarantee. However, the problem is that these methods presuppose the debiasing procedure with **inverse** propensity weighting and thus cannot be used when there is a user–item pair with zero observed probability. Furthermore, the experiments on 1BitMC were conducted using only extremely small datasets (Coat and MovieLens 100k) and

prediction accuracy measures (MSE); accordingly, its performance on moderate size benchmark datasets (e.g., Yahoo! R3 [Mnih and Salakhutdinov, 2008]) and on a ranking task remain unknown.

1.2 Contributions

To overcome the limitations of the existing methods, we establish a new theory for offline recommender learning inspired by the theoretical framework of *unsupervised domain adaptation* (UDA) [Ben-David *et al.*, 2010]. UDA aims to obtain a good predictor in settings where the feature distributions between the training and test sets are different. To this end, UDA utilizes distance metrics that measure the dissimilarity between probability distributions and does not depend on the propensity score [Ganin *et al.*, 2016; Saito *et al.*, 2017; Saito, 2020a]. Thus, the framework is usable even when the true propensity score is unknown and is expected to alleviate the issues caused by the propensity estimation bias in the absence of MCAR data. Moreover, the method is valid even when there is a user–item pair with zero observed probability.

To figure out a solution to the propensity contradiction, we first define a novel discrepancy metric to quantify the similarity between two missing mechanisms of rating feedback. Building on our discrepancy measure, we derive a ***propensity independent generalization error bound*** for the loss function of interest. We further propose ***domain adversarial matrix factorization*** (DAMF), which minimizes the derived theoretical bound in an adversarial manner. Our theoretical bound and algorithm are independent of the propensity score and thus address the contradiction of the previous propensity-based solutions. Finally, we conduct extensive experiments using public real-world datasets. In particular, we demonstrate that the proposed approach outperforms existing propensity-based methods in terms of rating prediction and ranking performance under a realistic situation where the true propensity score is inaccessible.

2 Preliminaries

Let $u \in [m]$ be a user and $i \in [n]$ be an item in a recommender system. We also define $\mathcal{D} := [m] \times [n]$ as the set of all user–item pairs. Let $\boldsymbol{R} \in \mathbb{R}^{m \times n}$ denote a (deterministic) true rating matrix, where each entry $R_{u,i}$ represents the real-valued true rating of user u for item i.

Our goal is to develop an algorithm to obtain a better predicted rating matrix (or a *hypothesis*) $\widehat{\boldsymbol{R}}$, where each entry $\hat{R}_{u,i}$ denotes a predicted rating value for (u,i). To this end, we formally define ***the ideal loss function of interest***, which should ideally be minimized to obtain a rating predictor as follows:

$$\mathcal{L}_{ideal}^{\ell}(\widehat{\boldsymbol{R}}) := \frac{1}{mn} \sum_{(u,i) \in \mathcal{D}} \ell(R_{u,i}, \hat{R}_{u,i}), \qquad (1)$$

where $\ell(\cdot, \cdot) : \mathbb{R} \times \mathbb{R} \to [0, \Delta]$ is a loss function bounded by Δ. For example, when $\ell(x, y) = (x - y)^2$, Eq. (1) represents the *mean-squared-error* (MSE).

In real-life recommender systems, one cannot directly calculate the ideal loss function, as most rating data are missing

in nature. To precisely formulate this missing mechanism, we introduce two additional matrices. The first one is the *propensity matrix*, denoted as $\boldsymbol{P} \in \mathcal{P}$, where $\mathcal{P}$ represents the space of probability distributions over $\mathcal{D}$. Each of its entries $P_{u,i} \in [0, 1]$ is the *propensity score* of (u, i), and it represents the probability of the rating feedback being observed. Next, let $\boldsymbol{O} \in \{0, 1\}^{m \times n}$ be an *observation matrix* where each entry $O_{u,i} \in \{0, 1\}$ is a Bernoulli random variable with its expectation $\mathbb{E}[O_{u,i}] = P_{u,i}$. If $O_{u,i} = 1$, the rating of the pair is observed, otherwise, it is unobserved. Finally, we use $\boldsymbol{O} \sim \boldsymbol{P}$ when the entries of $\boldsymbol{O}$ are the realizations of Bernoulli distributions defined by the entries of $\boldsymbol{P}$. For simplicity and without loss of generality, we assume that $M := \sum_{(u,i) \in \mathcal{D}} O_{u,i}$ hereinafter.

2.1 Existing Methods

In our formulation, it is essential to approximate the ideal loss function using only observable feedback. Here, we summarize the existing methods to estimate the ideal loss function and discuss their limitations.

Naive Estimator. Given a set of observed rating feedback, the most basic estimator for the ideal loss is the naive estimator, which is defined as follows:

$$\widehat{\mathcal{L}}_{naive}^{\ell}(\widehat{\boldsymbol{R}}; \boldsymbol{O}) := \frac{1}{M} \sum_{(u,i) \in \mathcal{D}} O_{u,i} \cdot \ell(R_{u,i}, \hat{R}_{u,i}). \qquad (2)$$

The naive estimator simply averages the loss of user-item pairs with the observed rating feedback (i.e., $O_{u,i} = 1$). This is valid when the missing mechanism is MCAR[1], as it is unbiased against the ideal loss function with MCAR data [Schnabel *et al.*, 2016]. However, several previous studies indicate that this estimator exhibits a bias under general MNAR settings (i.e., $\mathbb{E}[\widehat{\mathcal{L}}_{naive}^{\ell}] \neq \mathcal{L}_{ideal}^{\ell}$ for some $\widehat{\boldsymbol{R}}$). This means that the naive estimator does not converge to the ideal loss, even with infinite data. Therefore, one should use an estimator that addresses the bias as an alternative to the naive one [Schnabel *et al.*, 2016].

Inverse Propensity Score (IPS) Estimator. To improve the naive estimator, several previous studies apply the IPS estimation to alleviate the bias of MNAR rating feedback [Liang *et al.*, 2016; Schnabel *et al.*, 2016; Saito *et al.*, 2020; Saito, 2020b; Saito *et al.*, 2021]. In causal inference, propensity scoring estimators are widely used to estimate the causal effects of a treatment from observational data [Imbens and Rubin, 2015]. In our formulation, we can derive an unbiased estimator for the loss function of interest by using the true propensity score as follows:

$$\widehat{\mathcal{L}}_{IPS}^{\ell}(\widehat{\boldsymbol{R}}; \boldsymbol{O}) := \frac{1}{mn} \sum_{(u,i) \in \mathcal{D}} O_{u,i} \cdot \frac{\ell(R_{u,i}, \hat{R}_{u,i})}{P_{u,i}}. \qquad (3)$$

This estimator is unbiased for the ideal loss (i.e., $\mathbb{E}[\widehat{\mathcal{L}}_{IPS}^{\ell}] = \mathcal{L}_{ideal}^{\ell}$ for any $\widehat{\boldsymbol{R}}$) and is more desirable than the naive estimator in terms of bias. However, its unbiasedness is ensured only

---

[1]A missing mechanism is said to be MCAR if the propenisty score is constant, i.e., $P_{u,i} = C, \forall (u, i) \in \mathcal{D}$. Ensuring that the data is MCAR is extremely difficult, because the users' self-selection as to which items they provide rating feedback is out of our control.

when the true propensity score is available, and it can have a bias with an inaccurate propensity estimator (see Lemma 5.1 of Schnabel et al. [Schnabel *et al.*, 2016]). The bias of IPS can arise in most real-world recommender systems, because the missing mechanism of rating feedback can depend on user self-selection, which cannot be controlled by analysts and is difficult to estimate [Marlin and Zemel, 2009; Schnabel *et al.*, 2016]. Indeed, most previous studies estimate the propensity score by using some amount of MCAR test data to ensure empirical performance [Schnabel *et al.*, 2016; Wang *et al.*, 2019]. However, this is infeasible owing to the costly annotation process [Joachims *et al.*, 2017]. In addition, this method is valid only when the ratings for some user–item pairs are observed with non-zero probabilities (i.e., $P_{u,i} \in (0, 1], \forall (u, i) \in \mathcal{D}$), which is a criterion difficult to verify in practice [Ma and Chen, 2019]. Therefore, we explore a theory and algorithm independent of the propensity score and MCAR data, aiming to alleviate the concurrent issues of the propensity-based methods.

Doubly Robust (DR) Estimator. The DR estimator utilizes the error imputation model and propensity score to decrease the variance of the IPS approach. MF-DR [Wang *et al.*, 2019] optimizes the following DR estimator of the ideal loss.

$$\widehat{\mathcal{L}}_{DR}^{\ell}(\widehat{\boldsymbol{R}}; \boldsymbol{O})$$
$$:= \frac{1}{mn} \sum_{(u,i) \in \mathcal{D}} \left\{ \hat{\ell}_{u,i} + O_{u,i} \frac{\ell(R_{u,i}, \hat{R}_{u,i}) - \hat{\ell}_{u,i}}{P_{u,i}} \right\}, \quad (4)$$

where $\hat{\ell}_{u,i}$ is the *imputation model*, which estimates $\ell(R_{u,i}, \hat{R}_{u,i})$. As discussed in [Wang *et al.*, 2019], this estimator satisfies the unbiasedness with the true propensity (i.e., $\mathbb{E}[\widehat{\mathcal{L}}_{DR}^{\ell}] = \mathcal{L}_{ideal}^{\ell}$) and often achieves a tighter estimation error bound compared to IPS. MF-DR obtains the final prediction ($\widehat{\boldsymbol{R}}$) and the imputation model simultaneously via a joint learning procedure.

However, the proposed joint learning algorithm still requires a pre-estimated propensity score [Wang *et al.*, 2019]. Furthermore, the estimation performance of the DR estimator is significantly degraded when the error imputation model and the propensity model are both misspecified [Dudík *et al.*, 2011]. In fact, in the empirical evaluations of [Wang *et al.*, 2019], MCAR test data are used to estimate the propensity score. In the experiments, we validate the performance of the DR estimator in a more practical setting where only MNAR data are available (without any MCAR data).

Causal Embeddings (CausE) The work most closely related to ours is [Bonner and Vasile, 2018], which proposed CausE, a domain adaptation-inspired method to address the selection bias in rating feedback. The loss function of CausE is given as follows:

$$\widehat{\mathcal{L}}_{CausE}^{\ell}(\widehat{\boldsymbol{R}}_{MCAR}, \widehat{\boldsymbol{R}}_{MNAR}; \boldsymbol{O}_{MCAR}, \boldsymbol{O}_{MNAR})$$
$$:= \widehat{\mathcal{L}}_{naive}^{\ell}(\widehat{\boldsymbol{R}}_{MCAR}; \boldsymbol{O}_{MCAR})$$
$$+ \widehat{\mathcal{L}}_{naive}^{\ell}(\widehat{\boldsymbol{R}}_{MNAR}; \boldsymbol{O}_{MNAR})$$
$$+ \beta \cdot \Omega_{domain}(\widehat{\boldsymbol{R}}_{MCAR}, \widehat{\boldsymbol{R}}_{MNAR}), \quad (5)$$

where $\beta \geq 0$ is a trade-off hyper-parameter. Two prediction matrices ($\widehat{\boldsymbol{R}}_{MCAR}$ and $\widehat{\boldsymbol{R}}_{MNAR}$) are for the MCAR and MNAR data, respectively, where $\widehat{\boldsymbol{R}}_{MNAR}$ is used to make the final predictions. The last term represents the *regularizer between tasks* and penalizes the divergence between the predictions for the MCAR and MNAR data.

This method empirically outperforms propensity-based methods in a binary rating feedback setting [Bonner and Vasile, 2018]. A problem is that this algorithm requires some MCAR data, which are generally unavailable. Moreover, it uses the idea of domain adaptation in a heuristic manner by myopically adding $\Omega_{domain}(\cdot)$; there is no theoretical guarantee for its loss function. Therefore, we explore an algorithm that is preferable to CausE in the following ways: (i) Our algorithm does not use any MCAR data during training, and thus is feasible in a realistic situation having no MCAR training data, (ii) Our proposed method is theoretically refined in the sense that it minimizes the propensity-independent upper bound of the ideal loss function.

3 Proposed Framework and Algorithm

This section derives a propensity-independent generalization error bound for the ideal loss and an algorithm to minimize the bound using only MNAR data.

3.1 Theoretical Bound

First, we define a discrepancy measure to quantify the similarity between two different propensity matrices.

Definition 1. *(Propensity Matrix Divergence (PMD)) Let $\mathcal{H}$ be a set of real-valued predicted rating matrices and $\widehat{\boldsymbol{R}} \in \mathcal{H}$ be a specific prediction. The PMD between any two given propensity matrices $\boldsymbol{P}$ and $\boldsymbol{P}'$ is defined as follows:*

$$\psi_{\widehat{\boldsymbol{R}}, \mathcal{H}}(\boldsymbol{P}, \boldsymbol{P}') := \sup_{\widehat{\boldsymbol{R}}' \in \mathcal{H}} \{\mathcal{L}^{\ell}(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}'; \boldsymbol{P}) - \mathcal{L}^{\ell}(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}'; \boldsymbol{P}')\},$$

where $\mathcal{L}^{\ell}(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}'; \boldsymbol{P}) := \mathbb{E}_{\boldsymbol{O} \sim \boldsymbol{P}}[\widehat{\mathcal{L}}_{naive}^{\ell}(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}'; \boldsymbol{O})] = \frac{1}{M} \sum_{(u,i) \in \mathcal{D}} P_{u,i} \cdot \ell(\hat{R}_{u,i}, \hat{R}'_{u,i}).$

Note that the PMD is well defined because $\psi_{\widehat{\boldsymbol{R}}, \mathcal{H}}(\boldsymbol{P}, \boldsymbol{P}) = 0, \forall \boldsymbol{P} \in \mathcal{P}$, and it satisfies nonnegativity and subadditivity. Moreover, it is independent of the true rating matrices and can be calculated for any given pair of propensity matrices without true rating information.

We now use $\psi_{\widehat{\boldsymbol{R}}, \mathcal{H}}$ and derive a propensity-independent generalization error bound for the ideal loss function.[2]

Theorem 1. *(Propensity Independent Generalization Error Bound) Suppose two observation matrices with MCAR and MNAR mechanisms ($\boldsymbol{O}_{MCAR} \sim \boldsymbol{P}_{MCAR}$ and $\boldsymbol{O}_{MNAR} \sim \boldsymbol{P}_{MNAR}$) are given. For any prediction matrix $\widehat{\boldsymbol{R}} \in \mathcal{H}$ and*

---

[2]Note that the theoretical bounds of all existing studies depend on the cardinality of the hypothesis space $|\mathcal{H}|$ [Schnabel *et al.*, 2016; Wang *et al.*, 2019]. This is unrealistic, as $|\mathcal{H}|$ is infinite in almost all cases (as the parameter space is continuous). Under a realistic setting with infinite $\mathcal{H}$, therefore, the existing analysis becomes invalid. Thus, we use the Rademacher complexity to consider a realistic situation with infinite hypothesis spaces.

Algorithm 1 Domain Adversarial Matrix Factorization

Input: MNAR observation matrix $\boldsymbol{O}_{MNAR}$; set of user-item pairs $\mathcal{D}$; trade-off parameter β; regularization parameter λ; batch size B; number of steps k

Output: learnt prediction matrix $\widehat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{V}^\top$

1: Randomly initialize $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{U}'$, and $\boldsymbol{V}'$
2: **repeat**
3:     Sample mini-batch data of size B from $\boldsymbol{O}_{MNAR}$
4:     **for** $1, \ldots, k$ **do**
5:         Update $\boldsymbol{U}$ and $\boldsymbol{V}$ by gradient descent according to Eq. (9) with fixed $\widehat{\boldsymbol{R}}^* = \boldsymbol{U}'(\boldsymbol{V}')^\top$
6:     **end for**
7:     Uniformly sample user–item pairs of size B from $\mathcal{D}$ to construct $\boldsymbol{O}_{MCAR}$
8:     **for** $1, \ldots, k$ **do**
9:         Update $\boldsymbol{U}'$ and $\boldsymbol{V}'$ by gradient ascent according to Eq. (8) with fixed $\widehat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{V}^\top$
10:     **end for**
11: **until** convergence;

---

for any $\delta \in (0,1)$, the following inequality holds with a probability of at least $1 - \delta$:

$$
\mathcal{L}^\ell_{ideal}(\widehat{\boldsymbol{R}})
$$
$$
\leq \widehat{\mathcal{L}}^\ell_{naive}(\widehat{\boldsymbol{R}}; \boldsymbol{O}_{MNAR}) + \psi_{\widehat{\boldsymbol{R}}, \mathcal{H}}(\boldsymbol{O}_{MCAR}, \boldsymbol{O}_{MNAR})
$$
$$
+ 2L(3\mathfrak{R}_{\boldsymbol{P}, M}(\mathcal{H}) + 2\mathfrak{R}_{\boldsymbol{P}', M}(\mathcal{H})) + 3\Delta\sqrt{\frac{\log(6/\delta)}{2M}}, \tag{6}
$$

where L is the Lipschitz constant of the loss function and $\mathfrak{R}_{\boldsymbol{P}, M}(\mathcal{H})$ is the Rademacher complexity of $\mathcal{H}$, which is defined in Appendix A. See also Appendix A for the proof.

The theoretical bound consists of the following four factors: (i) Naive loss on MNAR data, (ii) Empirical PMD, (iii) Complexity of the hypothesis class, (iv) Confidence term that depends on δ. Note that (i) and (ii) can be optimized with only observable data, and (iii) can be controlled by adding a regularization term, as described in the next section. Moreover (iv) converges to zero as M increases. We will empirically show that the bound is informative in the sense that optimizing (i) and (ii) results in a desired value of the ideal loss function.

3.2 Algorithm

Here, we describe the proposed algorithm. Building on the theoretical bound derived in Theorem 1, we aim at minimizing the following objective:

$$
\min_{\widehat{\boldsymbol{R}} \in \mathcal{H}} \underbrace{\widehat{\mathcal{L}}^\ell_{naive}(\widehat{\boldsymbol{R}}; \boldsymbol{O}_{MNAR})}_{\text{naive loss on MNAR training data}}
$$
$$
+ \beta \cdot \underbrace{\psi_{\widehat{\boldsymbol{R}}, \mathcal{H}}(\boldsymbol{O}_{MCAR}, \boldsymbol{O}_{MNAR})}_{\text{discrepancy}} + \lambda \cdot \underbrace{\Omega(\widehat{\boldsymbol{R}})}_{\text{regularization}}, \tag{7}
$$

where $\beta \geq 0$ is the trade-off hyper-parameter between the naive loss and discrepancy measure. $\Omega(\cdot)$ is a regularization function that penalizes the complexity of $\widehat{\boldsymbol{R}}$, and $\lambda \geq 0$ is

the regularization hyper-parameter. In addition, we can obtain $\boldsymbol{O}_{MCAR}$ by uniformly sampling unlabeled user–item pairs from $\mathcal{D}$. The objective in Eq. (7) builds on the two controllable terms of the theoretical bound in Eq. (6). Thus, it is independent of the propensity score, and we need ***not*** estimate the propensity score to optimize this objective. Note that we add the regularization term to our objective to control the complexity of the hypothesis class $\mathcal{H}$, as it is impossible to minimize the Rademacher complexity directly.

To solve Eq. (7), we empirically approximate $\psi_{\widehat{\boldsymbol{R}}, \mathcal{H}}$ by solving

$$
\max_{\widehat{\boldsymbol{R}}' \in \mathcal{H}} \widehat{\mathcal{L}}^\ell_{naive}(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}'; \boldsymbol{O}_{MCAR}) - \widehat{\mathcal{L}}^\ell_{naive}(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}'; \boldsymbol{O}_{MNAR}). \tag{8}
$$

This optimization corresponds to accurately estimating PMD from observable data.

Given the aforementioned details, Eq. (7) is now reduced to:

$$
\min_{\widehat{\boldsymbol{R}} \in \mathcal{H}} \underbrace{\frac{1}{M} \sum_{(u,i):O_{u,i}=1} \ell(R_{u,i}, \hat{R}_{u,i})}_{\text{empirical loss on MNAR training data}}
$$
$$
+ \beta \cdot \underbrace{\{\hat{L}^\ell(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^*; \boldsymbol{O}_{MCAR}) - \hat{L}^\ell(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^*; \boldsymbol{O}_{MNAR})\}}_{\text{approximated PMD between MCAR and MNAR}}
$$
$$
+ \lambda \cdot \underbrace{\Omega(\widehat{\boldsymbol{R}})}_{\text{regularization}}, \tag{9}
$$

where $\widehat{\boldsymbol{R}}^*$ is the solution of Eq. (8).

Algorithm 1 describes the detailed procedure of our DAMF, which is based on the matrix factorization (MF) model, where a prediction matrix is defined as the product of user–item latent factors $\boldsymbol{U} \in \mathbb{R}^{m \times d}$ and $\boldsymbol{V} \in \mathbb{R}^{n \times d}$. Given that the predictions are modelled by MF, a possible regularization is $\Omega(\widehat{\boldsymbol{R}}) := \|\boldsymbol{U}\|_2^2 + \|\boldsymbol{V}\|_2^2$, where $\|\cdot\|_2$ is the L2 norm.

4 Experiments

This section empirically evaluates the proposed method on public real-world datasets. Appendix B describes detailed experimental settings such as dataset description, evaluation metrics, and hyper-parameter tuning. Our code is available at https://github.com/usaito/ijcai2022-adversarial-mf.

4.1 Experimental Setups

Datasets. We use Yahoo! R3[3] and Coat[4], which separately contain MNAR and MCAR datasets. We randomly selected 10% of the MNAR training set as the validation set, which is used to perform hyper-parameter tuning.

Baseline Methods and Propensity Estimators. We compare our proposed algorithm with the following baselines:

- **MF** [Koren *et al.*, 2009], which optimizes its model parameters by minimizing the naive loss in Eq. (2) and does not depend on the propensity score.

---

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

Datasets	Methods	MSE ($\pm$ StdDev)
Yahoo! R3	MF	1.7343 ($\pm$ 0.0309)
	MF-IPS	1.7320 ($\pm$ 0.0311)
	MF-DR	1.7445 ($\pm$ 0.0189)
	DAMF (ours)	**1.2787 ($\pm$ 0.0126)**
	CausE	1.7390 ($\pm$ 0.0283)
	MF-IPS (true)	1.1281 ($\pm$ 0.0161)
	MF-DR (true)	1.0435 ($\pm$ 0.0166)
Coat	MF	1.2166 ($\pm$ 0.0007)
	MF-IPS	1.2044 ($\pm$ 0.0012)
	MF-DR	1.2108 ($\pm$ 0.0016)
	DAMF (ours)	**1.1371 ($\pm$ 0.0025)**
	CausE	1.2801 ($\pm$ 0.0043)
	MF-IPS (true)	1.0675 ($\pm$ 0.0043)
	MF-DR (true)	1.0760 ($\pm$ 0.0037)

Table 1: Rating Prediction Performance

Note: The bold font indicates the best performance in each metric and dataset among methods using only MNAR data.

- **MF-IPS** [Schnabel *et al.*, 2016], which optimizes its model parameters by minimizing the IPS loss in Eq. (3).

- **MF-DR** [Wang *et al.*, 2019], which optimizes its model parameters by minimizing the DR loss in Eq. (4).

- **CausE** [Bonner and Vasile, 2018], which minimizes the sum of the naive loss and the domain regularization term $\Omega_{domain}(\cdot)$, which measures the divergence between the MNAR and MCAR data. To calculate its loss function, we sample 5% of the MCAR test data. MCAR data are in general unavailable in real-world settings, and thus we report the results of CausE as a reference.

For MF-IPS and MF-DR, we use *user propensity*, *item propensity*, *user–item propensity*, and *1BitMC* [Ma and Chen, 2019] as the variants of the propensity estimator and report the results with the best estimator for each dataset. Note that these four propensity estimators (formally defined in the Appendix) are usable in real-world recommender systems because they use only MNAR data.

We also report the results of MF-IPS and MF-DR with the *true propensity score*. We follow previous studies [Schnabel *et al.*, 2016; Wang *et al.*, 2019] and calculate this true propensity using 5% of the MCAR test data. Note that the true propensity is unavailable in real-world situations, as it requires MCAR explicit feedback. Therefore, we report the results of MF-IPS and MF-DR with the true propensity score as a reference.

4.2 Results and Discussion

Here, we report and discuss the experimental results with respect to four research questions (RQ1-RQ4). We run the experiments 10 times with different random seeds and report the averaged results in Tables 1-4.

RQ1. How well does the proposed algorithm perform on the rating prediction task? We evaluate and compare the prediction performance of DAMF and the baselines on Yahoo! R3 and Coat. Table 3.2 provides the MSEs and their standard deviations (StdDev). For Yahoo! R3, DAMF reveals the best rating prediction performance among the methods using only MNAR data. Specifically, it outperforms MF-IPS by 26.3% and MF-DR by 26.7% in terms of MSE. For Coat, DAMF outperforms MF-IPS by 5.59% and MF-DR by 6.08%, even though the distributional shift of Coat is significantly smaller than that of Yahoo! R3. Note that it is reasonable to assume that there exists a user–item pair with zero observed probability on Coat. This is because the training set was collected via workers' self-selection, and male workers did not provide the ratings of women's coats and vice versa. Thus, the results also suggest the stability and adaptability of the proposed method to the data with a user–item pair with $P_{u,i} = 0$. In contrast, the performance of the propensity-based methods on Coat is not theoretically grounded, as the training and test sets may not overlap.

RQ2. How do propensity-based methods perform with different propensity estimators? Here, we evaluate the sensitivity of MF-IPS and MF-DR to the choice of the propensity estimator. First, consistent with previous studies [Schnabel *et al.*, 2016; Wang *et al.*, 2019], Table 3.2 shows that MF-IPS and MF-DR with the true propensity exhibit better rating prediction performance compared to the other methods. However, Table 3.2 shows that they perform poorly with other propensity estimators, including *1BitMC*, and especially for Yahoo! R3. In some cases, they even underperform MF, which is based on the naive loss. Therefore, although propensity-based methods are potentially high-performing with the true propensity, their prediction performances are highly sensitive to the choice of propensity estimator and negatively affected by the propensity estimation bias. These empirical observations imply the **propensity contradiction** of the previous methods.

RQ3. How well does the proposed algorithm perform on the ranking task? Next, we test the ranking performance of DAMF on Yahoo! R3 and Coat. Tables 3.2 and 3.2 describe the ranking metrics computed on the MCAR test data.

For Yahoo! R3, DAMF demonstrates the best ranking performance among all methods, including the methods with the true propensity score. It outperforms MF-IPS by 3.39% and MF-DR by 4.15%, MF-IPS (true) by 6.04%, and MF-DR (true) by 3.61% in NDCG. The results suggest that DAMF is powerful and useful for improving the recommendation quality and user experience with only biased rating data. Moreover, DAMF is the best in Coat, improving MF-IPS by 3.48%, MF-DR by 1.94%, MF-IPS (true) by 1.10%, and MF-DR (true) by 1.32% in NDCG. These results demonstrate that the proposed method works satisfactorily in the ranking task even when the dataset size is small and the degree of bias is not significant.

RQ.4 Does adversarial learning really minimize the ideal loss? Finally, we investigate the validity of minimizing the propensity-independent upper bound in Eq. (6) as a solution to minimize the ideal loss function in Eq. (1) using only MNAR data. Figure 1 depicts the sum of the first two terms of the upper bound (blue) and ideal loss (orange) during the training of DAMF on Yahoo! R3 and Coat. First, it is evident from the figures that our adversarial learning procedure effectively minimizes the upper bound of the ideal loss during training. Moreover, the figures suggest that the ideal loss function is

Datasets	Methods	user propensity	item propensity	user-item propensity	1BitMC
Yahoo! R3	MF-IPS	1.8216 (+61.5%)	1.8083 (+60.3%)	1.8714 (+65.9%)	1.7343 (+53.8%)
	MF-DR	1.7445 (+67.2%)	1.9074 (+82.8%)	1.8815 (+80.4%)	1.8953 (+81.6%)
Coat	MF-IPS	1.2171 (+14.0%)	1.2161 (+13.9%)	1.2044 (+12.8%)	1.2087 (+13.2%)
	MF-DR	1.2141 (+12.8%)	1.2108 (+12.5%)	1.2181 (+13.2%)	1.2195 (+13.3%)

Table 2: Rating Prediction Performance of MF-IPS and MF-DR with Different Propensity Estimators

Note: Performances relative to the true propensity score are in parentheses. The result suggests that the performances of MF-IPS and MF-DR using only MNAR data are significantly worse than the performances of those using the true propensity.

Methods	NDCG@5 ($\pm$ StdDev)	Recall@5 ($\pm$ StdDev)
MF	0.7645 ($\pm$ 0.0008)	0.5616 ($\pm$ 0.0006)
MF-IPS	0.7645 ($\pm$ 0.0008)	0.5616 ($\pm$ 0.0006)
MF-DR	0.7589 ($\pm$ 0.0007)	0.5602 ($\pm$ 0.0005)
DAMF (ours)	**0.7904 ($\pm$ 0.0011)**	**0.5737 ($\pm$ 0.0004)**
CausE	0.7645 ($\pm$ 0.0007)	0.5617 ($\pm$ 0.0006)
MF-IPS (true)	0.7454 ($\pm$ 0.0021)	0.5565 ($\pm$ 0.0011)
MF-DR (true)	0.7629 ($\pm$ 0.0032)	0.5624 ($\pm$ 0.0015)

Table 3: Ranking Performance on **Yahoo! R3**

Methods	NDCG@5 ($\pm$ StdDev)	Recall@5 ($\pm$ StdDev)
MF	0.6617 ($\pm$ 0.0029)	0.3858 ($\pm$ 0.0008)
MF-IPS	0.6717 ($\pm$ 0.0026)	0.3844 ($\pm$ 0.001)
MF-DR	0.6789 ($\pm$ 0.0018)	0.3887 ($\pm$ 0.0016)
DAMF (ours)	**0.6847 ($\pm$ 0.0034)**	**0.3901 ($\pm$ 0.0011)**
CausE	0.6332 ($\pm$ 0.0039)	0.3752 ($\pm$ 0.0013)
MF-IPS (true)	0.6772 ($\pm$ 0.002)	0.3906 ($\pm$ 0.0009)
MF-DR (true)	0.6758 ($\pm$ 0.0014)	0.3900 ($\pm$ 0.0012)

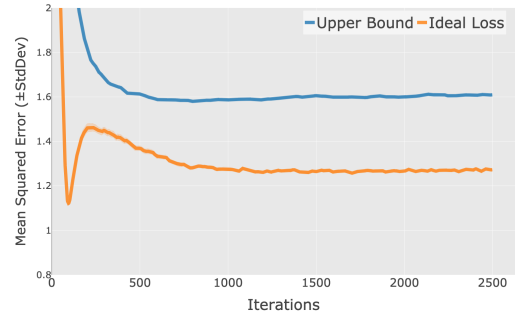Table 4: Ranking Performance on **Coat**

Note: The bold font indicates the best performance in each metric and dataset among methods using only MNAR data.

also minimized well by our method. These results demonstrate that minimizing the theoretical upper bound in Eq. (6) is a valid approach for improving the recommendation quality on the MCAR test set.
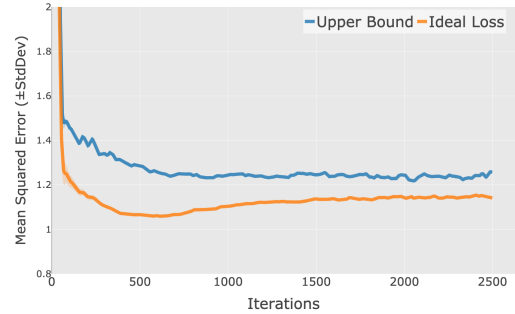
5 Conclusion

We explored the problem of offline recommender learning from *only* MNAR explicit rating feedback. To this end, we derived the *propensity independent generalization error bound* of the loss function and proposed an algorithm to minimize the bound via adversarial learning. Our algorithm enables an accurate rating prediction without estimating the propensity score, thereby resolving the *propensity contradiction* of the existing literature. Extensive experiments demonstrate the advantages of the proposed framework in terms of rating prediction and ranking measures when true propensities are inaccessible.

Our work represents an initial attempt for offline recommender learning via adversarial training, and there are still many problems to study in this direction. For example, our theory leads to a propensity-independent upper bound of the (ideal) rating prediction loss. It would be valuable to construct a corresponding theory for the ranking task, such as showing the propensity-independent upper bound of the additive

(a) Yahoo! R3

(b) Coat

Figure 1: Theoretical Upper Bound and Ideal Loss Function during DAMF Training.

ranking metrics. Moreover, testing our method in an online experiment will support the reliability of our framework. We leave the investigation of these directions for future studies.

References

[Ben-David *et al.*, 2010] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1-2):151–175, 2010.

[Bonner and Vasile, 2018] Stephen Bonner and Flavian Vasile. Causal embeddings for recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, pages 104–112, New York, NY, USA, 2018. ACM.

[Chen *et al.*, 2020] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. Bias and debias in recommender system: A survey and future directions. *arXiv preprint arXiv:2010.03240*, 2020.

[Dudík *et al.*, 2011] Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pages 1097–1104, 2011.

[Ganin *et al.*, 2016] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.

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

[Joachims *et al.*, 2017] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. Unbiased learning-to-rank with biased feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 781–789. ACM, 2017.

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

[Liang *et al.*, 2016] Dawen Liang, Laurent Charlin, and David M Blei. Causal inference for recommendation. In *Causation: Foundation to Application, Workshop at UAI*, 2016.

[Ma and Chen, 2019] Wei Ma and George H Chen. Missing not at random in matrix completion: The effectiveness of estimating missingness probabilities under a low nuclear norm assumption. In *Advances in Neural Information Processing Systems*, pages 14871–14880, 2019.

[Marlin and Zemel, 2009] Benjamin M Marlin and Richard S Zemel. Collaborative prediction and ranking with non-random missing data. In *Proceedings of the third ACM conference on Recommender systems*, pages 5–12. ACM, 2009.

[Mnih and Salakhutdinov, 2008] Andriy Mnih and Ruslan R Salakhutdinov. Probabilistic matrix factorization. In *Advances in neural information processing systems*, pages 1257–1264, 2008.

[Saito *et al.*, 2017] Kuniaki Saito, Yoshitaka Ushiku, and Tatsuya Harada. Asymmetric tri-training for unsupervised domain adaptation. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2988–2997. PMLR, 2017.

[Saito *et al.*, 2020] Yuta Saito, Yaginuma Suguru, Yuta Nishino, Hayato Sakata, and Nakata Kazuhide. Unbiased recommender learning from missing-not-at-random implicit feedback. In *Proceedings of the Thirteenth ACM International Conference on Web Search and Data Mining*. ACM, 2020.

[Saito *et al.*, 2021] Yuta Saito, Takuma Udagawa, Haruka Kiyohara, Kazuki Mogi, Yusuke Narita, and Kei Tateno. Evaluating the robustness of off-policy evaluation. In *Fifteenth ACM Conference on Recommender Systems*, pages 114–123, 2021.

[Saito, 2020a] Yuta Saito. Asymmetric tri-training for debiasing missing-not-at-random explicit feedback. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 309–318, 2020.

[Saito, 2020b] Yuta Saito. Unbiased pairwise learning from biased implicit feedback. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, pages 5–12, 2020.

[Schnabel *et al.*, 2016] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. Recommendations as treatments: Debiasing learning and evaluation. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48, pages 1670–1679. PMLR, 2016.

[Steck, 2010] Harald Steck. Training and testing of recommender systems on data missing not at random. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 713–722. ACM, 2010.

[Wang *et al.*, 2019] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. Doubly robust joint learning for recommendation on data missing not at random. In *International Conference on Machine Learning*, pages 6638–6647, 2019.