Eliminating Bias in Recommender Systems via Pseudo-Labeling

Yuta Saito

Tokyo Institute of Technology

saito.y.bj@m.titech.ac.jp

Abstract

Addressing the non-uniform missing mechanism of rating feedback is critical to recommending items users prefer from biased real-world datasets. To tackle the challenging issue, we first define an ideal loss function that should be optimized to achieve the goal of recommendation. Then, we derive the generalization error bound of the ideal loss that alleviates the variance and the misspecification problems of the previous causal-based methods. We further propose a meta-learning method minimizing the bound. Empirical evaluation using real-world datasets validates the theoretical findings and demonstrates the practical advantages of the proposed method.

1 Introduction

The goal of recommender systems is to recommend a new item that users will prefer. To achieve this, recommendation algorithms predict the potential preferences or relevance of non-interacted user-item pairs by using sparse observed ratings. Developing effective recommendation algorithms is critical to improving the profit margin of marketing platforms (e.g., Amazon and eBay) or the user experience in user-interactive systems (e.g., Spotify and Netflix). Therefore, the field of personalized recommendation has been widely studied in both academia and industry.

Within the area of recommender systems, most of the existing studies assume that the observed rating data is missing-at-random (MAR). Generally, this assumption does not hold because real-world recommender systems are subject to selection biases. The selection biases are primarily due to the following two reasons: First, the probability of observing each rating is highly dependent on a past recommendation policy. For example, if the observed rating dataset is collected under the most popular policy, a policy that always recommends some of the popular items to all users, the probability of the observing ratings of such popular items may be large. This leads to the non-uniform missing mechanism, and the MAR assumption is violated. The second cause is user self-selection. This type of selection happens because users are free to choose the items that they wish to rate. For example, in a movie recommender system, users usually watch and rate movies that they like and rarely rate movies that they do not like [1]. Another example is a song recommender system, in which users tend to rate songs that they like or dislike and seldom rate songs that they have a neutral preference [2]. These findings suggest that most of the rating datasets collected through real-world user-interactive systems are missing-not-at-random (MNAR). Several studies have theoretically and empirically indicated that the conventional methods of using only observed ratings lead to suboptimal prediction models because the observed ratings are not the representative data of the target population [3]. Thus, developing a recommendation algorithm and a debiasing method that can lead to high prediction accuracy under the MNAR assumption is essential to achieve the goal of recommender systems in the real world.

Several related approaches directly address the MNAR problem. Among these, the most promising approach is causal-based debiasing methods involving, for example, the inverse propensity score (IPS) or doubly robust (DR) estimations. These methods have been established in fields such as causal inference and missing data analysis [4, 5], and have been proposed to be utilized for debiasing learning and evaluation of MNAR recommendations [6, 7]. IPS estimation relies on the propensity score, which is the probability of observing each rating. By reweighting each sample by the inverse of its propensity score, one can generate a pseudo dataset for which the users and items are uniformly distributed. Based on this reweighting method, IPS provides an unbiased estimation of the actual loss function. The benefits of these causal-based methods are theoretically principled and empirically outperform naive methods based on the unrealistic MAR assumption [6, 7].

For such causal-based methods to succeed, accurately estimating the propensity score is critical. This is because the unbiasedness of the loss function is guaranteed when the true propensities are available; an IPS estimator may still have a bias depending on the propensity estimation bias [6, 8]. However, correctly estimating the propensity score is almost impossible, and model misspecification often occurs in real-world settings [9]. Moreover, propensity-based methods such as the IPS approach generally suffer from a high variance and can lead to sub-optimal estimation when the item popularity or user activeness are highly diverse [8, 10, 11]. Improving the robustness to the propensity misspecification and the high variance problem of the causal-based methods is an important and unsolved problem.

To improve the robustness and prediction accuracy of arbitrary recommendation algorithms on MNAR ratings, we propose a theoretically principled meta-learning method called the asymmetric tri-training for missing-not-at-random recommendation. It utilizes two rating predictors to generate a reliable missing-completely-at-random (MCAR) data with pseudo ratings. The proposed method is based on the asymmetric tri-training framework proposed in unsupervised domain adaptation [12]. Similar to causal inference, unsupervised domain adaptation addresses the problem setting in which the data-generating distributions are different between the training and test datasets [13, 14]. Moreover, unsupervised domain adaptation does not rely on the propensity score. Thus, this approach is considered to be useful to overcome the problems related to propensity reweighting used in MNAR recommendation algorithms. However, the methods of unsupervised domain adaptation have not yet been applied to the MNAR recommendation, and this is the first work to bridge the two closely related areas.

In our theoretical analysis, we establish the generalization error bound of the ideal loss function and demonstrate that our meta-learning method minimizes this bound. In contrast to the bounds presented in previous studies [6, 8], the generalization error bound expected to be minimized by the proposed method is independent of the propensity score, and thus, the problems related to propensity scoring such as the high variance or propensity misspecification are solved. Finally, in the experimental evaluation, we conducted extensive experiments using standard benchmark datasets to demonstrate that the proposed method is robust to the propensity misspecification and the variance problems.

The contributions of this paper can be summarized as follows.

- We propose a meta-learning method based on a method proposed for unsupervised domain adaptation. The proposed method is general and can be applied to any existing algorithm to improve the prediction accuracy under the MNAR setting.

- We theoretically analyze the recommendation using MNAR ratings and show that the proposed meta-learning method is interpreted as minimizing the generalization error bound of the ideal loss function.

- We demonstrate the effectiveness of our method on popular benchmark datasets. In particular, the

2

proposed method successfully solves the propensity misspecification and the variance problems of the previous causal-based algorithms.

The remaining paper is organized as follows. In Section 2, we review the previous related works on both causal-based recommendation and unsupervised domain adaptation. Section 3 describes the formulation of the MNAR problem. In Section 4, we propose and theoretically analyze the asymmetric tri-training for missing-not-at-random recommendation. The experimental setup and results are described in Section 5. Section 6 presents the conclusions and future research directions.

2 Related Works

In this section, we review the existing related studies.

2.1 Causal-based Recommendation

To handle MNAR rating datasets, some works consider the missing data model and rating model and estimate the parameters via the EM algorithms [2, 15]. However, these methods are highly complex and do not perform well on real-world datasets [6].

The causal-based recommendation is a different approach to tackle the MNAR problem [6, 7, 16]. In this approach, the IPS estimation [4, 5, 17] established in causal inference is applied to the rating prediction methods based on matrix factorization. The probability of observing each entry of the rating matrix is defined as the propensity score, and the unbiased estimator for the true metric can be derived by weighting each instance by the inverse of its propensity. The method of matrix factorization with IPS (MF-IPS) has been demonstrated to outperform the naive matrix factorization and the probabilistic generative model [15] under MNAR settings. Moreover, the DR estimation, used in the off-policy evaluation of the bandit algorithm and reinforcement learning [18, 19], has also been applied to the MNAR recommendation [8]. DR estimation effectively combines the propensity score estimation and the error imputation model and improves the statistical properties of the IPS estimator. The error imputation model is the model of the predicted errors for the missing ratings, and the performance of the DR estimator has been proved to be dependent on the accuracy of the propensity score estimation and the error imputation model [8]. These causal-based algorithms utilize the unbiased loss function; however, the performance of these methods largely depends on the propensity score estimation and error imputation model. Ensuring the accuracy of the propensity score estimation and error imputation model is challenging in real-world settings [9], and thus, methods to improve the robustness of the causal-based approaches are highly desired.

2.2 Unsupervised Domain Adaptation

Unsupervised domain adaptation aims to train a predictor that works well on a target domain by using only labeled source samples and unlabeled target samples during training [12]. One difficulty is that the feature distributions and the labeling functions[1] are different between the source and target domains. Therefore, a model trained on the source domain does not generalize well on the target domain, and measuring the difference between the two domains is critical in unsupervised domain adaptation [14]. Some discrepancy measures to measure this difference have been proposed. Among them, $\mathcal{H}$-divergence and $\mathcal{H}\Delta\mathcal{H}$-divergence [20, 21] have been used to construct many prediction methods. For example, domain adversarial neural

---

[1]mapping from feature space to outcome space

network (DANN) simultaneously minimizes the source empirical errors and the $\mathcal{H}$-divergence between the source and target domains in an adversarial manner [22, 23]. The asymmetric tri-training framework trains three networks asymmetrically and is considered to minimize the $\mathcal{H}$-divergence during training [12]. Our proposed method is based on this framework, and it is the first extension of the unsupervised domain adaptation method to alleviate the bias of the MNAR recommendation.

3 Preliminaries

In this section, we introduce the basic notation and formulation of the MNAR explicit recommendation.

3.1 Problem Formulation

Let $\mathcal{U}$ be a set of users ($|\mathcal{U}| = m$), and $\mathcal{I}$ be a set of items ($|\mathcal{I}| = n$). We denote the set of all user and item pairs as $\mathcal{D} = \mathcal{U} \times \mathcal{I}$. Let $\boldsymbol{R} \in \mathbb{R}^{m \times n}$ be a true rating matrix; each entry $R_{u,i}$ is the true rating of user u to item i.

The focus of this study is to establish an algorithm to obtain an optimal predicted rating matrix denoted as $\widehat{\boldsymbol{R}}$. Each entry $\widehat{R}_{u,i}$ is the predicted rating for the user-item pair (u, i). To achieve this goal, we formally define the ideal loss function that should be minimized to derive the predictions $\widehat{\boldsymbol{R}}$ as

$$\mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}\right) = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \ell\left(R_{u,i}, \widehat{R}_{u,i}\right) \tag{1}$$

where $\ell(\cdot, \cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ is an arbitrary loss function. For example, when $\ell(x, y) = (x - y)^2$, $\mathcal{L}_{ideal}(\widehat{\boldsymbol{R}})$ is the mean-squared-error (MSE).

In reality, it is impossible to calculate the ideal loss function, as most of the true ratings are often missing. To formulate the missing mechanism of the true ratings, we introduce another matrix $\boldsymbol{O} \in \{0, 1\}^{m \times n}$ called the indicator matrix, and each entry $O_{u,i}$ is a Bernoulli random variable representing whether the true rating of (u, i) is observed. If $O_{u,i} = 1$, then $R_{u,i}$ is observed; otherwise, $R_{u,i}$ is unobserved. Using indicator variables, we can denote the set of user-item pairs for the observed ratings as $\mathcal{O} = \{(u, i) \,|\, O_{u,i} = 1\}$. Accurately estimating the ideal loss function using the observed dataset $\mathcal{O}$ is essential to derive an effective recommender.

3.2 Naive Estimator

The simplest estimator for the ideal loss function is called the naive estimator, which is defined as follows:

$$\widehat{\mathcal{L}}_{naive}^{\ell}\left(\widehat{\boldsymbol{R}}\right) = \frac{1}{|\mathcal{O}|} \sum_{(u,i) \in \mathcal{O}} \ell\left(R_{u,i}, \widehat{R}_{u,i}\right) \tag{2}$$

This estimator calculates the average loss function over the observed ratings, and most existing methods are based on this simple estimator. If the missing ratings are MAR, the naive estimator is unbiased against the ideal loss function. However, in the case of MNAR datasets, the naive estimator is biased [3, 6].

$$\mathbb{E}_{\boldsymbol{O}}\left[\widehat{\mathcal{L}}_{naive}^{\ell}\left(\widehat{\boldsymbol{R}}\right)\right] \neq \mathcal{L}_{ideal}\left(\widehat{\boldsymbol{R}}\right)$$

Thus, the naive estimator should be debiased when learning a recommender using MNAR rating datasets.

4

3.3 Inverse Propensity Score Estimator

In [6, 7], the authors applied the IPS estimation to debias the naive estimator under the MNAR mechanism. The propensity scoring method has been previously proposed in the context of causal inference to estimate treatment effects using observational data [4, 5, 17]. The basic idea of this estimator is to create a pseudo-MAR dataset by reweighting the observed ratings by the inverse of its propensity score.

In this work, the propensity score of user-item pair (u, i) is formally defined as $P_{u,i} = \mathbb{P}\left(O_{u,i} = 1\right) = \mathbb{E}\left[O_{u,i}\right]$. By using the propensity score, the unbiased estimator for the ideal loss function can be derived as follows:

$$\widehat{\mathcal{L}}^{\ell}_{IPS}\left(\widehat{\boldsymbol{R}}\right) = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{O}} \frac{\ell\left(R_{u,i}, \widehat{R}_{u,i}\right)}{P_{u,i}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} O_{u,i} \cdot \frac{\ell\left(R_{u,i}, \widehat{R}_{u,i}\right)}{P_{u,i}} \tag{3}$$

This IPS estimator is unbiased against the ideal loss function, and thus considered to be more desirable than the naive estimator.

$$\mathbb{E}_{\boldsymbol{O}}\left[\widehat{\mathcal{L}}^{\ell}_{IPS}\left(\widehat{\boldsymbol{R}}\right)\right] = \mathcal{L}^{\ell}_{ideal}\left(\widehat{\boldsymbol{R}}\right)$$

As theoretically and empirically stated in [6], unbiasedness of the IPS estimator is desirable; however, this property depends on the true propensity score. In reality, the true propensity score is not observable and thus has to be estimated using the naive Bayes, logistic regression, or Poisson factorization [6, 7]. If the propensity estimation model is misspecified, the IPS estimator is no longer an unbiased estimator. Moreover, the IPS estimator often suffers from a high variance because the inverse of the propensities might be large [9, 18]. Therefore, developing learning methods that are robust to the propensity misspecification and the variance of the estimator is critical to apply the methods to real-world MNAR problems.

4 Method

This section presents the proposed meta-learning method called asymmetric tri-training for missing-not-at-random recommendation. Asymmetric tri-training has been previously established in the field of unsupervised domain adaptation [12] and considered to be useful for constructing a learning method that is robust to the selection bias of the MNAR problem. This is because unsupervised domain adaptation, including asymmetric tri-training, does not rely on the propensity score, and thus, the estimation bias caused by the propensity misspecification and high variance of propensity-based methods are eliminated.

4.1 Meta-learning procedure

The objective here is to construct a rating predictor that can obtain a smaller value of the ideal loss function. To realize this, the asymmetric tri-training framework utilizes three rating predictors asymmetrically. First, two of the three predictors are trained to assign pseudo-ratings to generate a pseudo-MCAR dataset. After generating a reliable pseudo-MCAR dataset, the other predictor is trained on that pseudo dataset. We can use any recommendation algorithm, such as matrix factorization [24, 25], MF-IPS [6, 7], factorization machines [26], and neural network matrix factorization [27], for the three predictors. Thus, the proposed method is highly general and can be used to improve the prediction accuracy of methods proposed in the future.

The asymmetric tri-training framework consists of the three steps. First, in the **pre-training step**, we pre-train the three selected recommendation algorithms A_1, A_2, and A_3 using the observed rating data $\mathcal{O}$. Next, we generate an MCAR dataset with pseudo-ratings by randomly sampling user-item pairs, which is

Algorithm 1 Asymmetric Tri-training for Missing-Not-At-Random Recommendation

---

Input: observed rating dataset $\mathcal{O}$, three predictors A_1, A_2, A_3, set of hyperparameters { ε, *iter*, *number of steps* }

Output: predicted rating matrix $\widehat{\boldsymbol{R}}$ by A_3

1: Pre-train A_1, A_2, A_3 using the observed rating dataset $\mathcal{O}$
2: randomly sample user-item pairs from $\mathcal{D}$ to generate $\mathcal{D}'$
3: generate pseudo-MCAR dataset $\widetilde{\mathcal{D}}$ with $A_1, A_2, \mathcal{D}', \varepsilon$
4: **for** i = 1 to *number of steps* **do**
5:     **for** j = 1 to *iter* **do**
6:         Update A_1 and A_2 with mini-batch from $\widetilde{\mathcal{D}}$
7:         Update A_3 with mini-batch from $\widetilde{\mathcal{D}}$
8:     **end for**
9:     randomly sample user-item pairs from $\mathcal{D}$ to generate $\mathcal{D}'$
10:     generate pseudo-MCAR dataset $\widetilde{\mathcal{D}}$ with $A_1, A_2, \mathcal{D}', \varepsilon$
11: **end for**
12: **return** $\widehat{\boldsymbol{R}}$ by A_3

---

considered as an unlabeled MCAR dataset[2], denoted as $\mathcal{D}'$. Then, we predict the ratings of the unlabeled dataset $\mathcal{D}'$ using two of the three algorithms A_1, A_2. The predicted rating for $(u, i) \in \mathcal{D}'$ by A_1 and A_2 is denoted as $\widehat{R}_{u,i}^{(1)}$ and $\widehat{R}_{u,i}^{(2)}$, respectively. We regard one of the two predicted values (here, $\widehat{R}_{u,i}^{(1)}$ or $\widehat{R}_{u,i}^{(2)}$) as the pseudo-rating for (u, i) if the two predicted values are sufficiently similar. The resulting pseudo MCAR dataset is denoted as

$$\widetilde{\mathcal{D}} = \left\{ \left(u, i, \widehat{R}_{u,i}^{(1)} \right) : (u, i) \in \mathcal{D}', \left| \widehat{R}_{u,i}^{(1)} - \widehat{R}_{u,i}^{(2)} \right| \leq \varepsilon \right\} \tag{4}$$

where $\varepsilon > 0$ is a hyperparameter. This step is the **pseudo labeling step**. Finally, we train the remaining predictor A_3 by minimizing the following loss function.

$$\widehat{\mathcal{L}}_{pseudo}^{\ell} \left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)} \right) = \frac{1}{|\widetilde{\mathcal{D}}|} \sum_{(u,i) \in \widetilde{\mathcal{D}}} \ell \left(\widehat{R}_{u,i}, \widehat{R}_{u,i}^{(1)} \right) = \frac{1}{|\widetilde{\mathcal{D}}|} \sum_{(u,i) \in \mathcal{D}} O'_{u,i} \cdot \ell \left(\widehat{R}_{u,i}, \widehat{R}_{u,i}^{(1)} \right) \tag{5}$$

where $\{\widehat{R}_{u,i}^{(1)}\}$ are the pseudo-ratings provided by A_1, $\{\widehat{R}_{u,i}\}$ are the predicted ratings provided by A_3, and $\{O'_{u,i}\}$ are other indicator variables representing whether the user-item pairs are in the pseudo-MCAR dataset $\widetilde{\mathcal{D}}$.

In the actual algorithm, we iterate the pseudo-labeling step several times to generate a reliable pseudo-MCAR dataset. Algorithm 1 describes the complete learning procedure of the asymmetric tri-training for missing-not-at-random recommendation.

4.2 Theoretical Analysis

In this subsection, we theoretically analyze the MNAR recommendation based on the theoretical framework of unsupervised domain adaptation. In particular, we drive the generalization error bound of the ideal loss function and demonstrate that the proposed asymmetric tri-training framework can minimize the bound.

In the following proposition, we first derive a simple upper bound of the ideal loss function based on the triangle inequality.

---

[2]This sampling is optional; one can simply use $\mathcal{D}$ as the unlabeled MCAR dataset.

Proposition 1. Assume that the loss function ℓ obeys the triangle inequality. Then, for any predicted rating matrices $\widehat{\boldsymbol{R}}^{(1)}$, $\widehat{\boldsymbol{R}}^{(2)}$, and $\widehat{\boldsymbol{R}}$, the following inequality holds.

$$\mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}, \boldsymbol{R}\right) \leq \mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) + \mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}^{(1)}, \widehat{\boldsymbol{R}}^{(2)}\right) + \mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}^{(2)}, \boldsymbol{R}\right) \tag{6}$$

Proof. We apply the triangle inequality twice:

$$
\begin{aligned}
\mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}, \boldsymbol{R}\right) &\leq \mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) + \mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}^{(1)}, \boldsymbol{R}\right) \\
&\leq \mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) + \mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}^{(1)}, \widehat{\boldsymbol{R}}^{(2)}\right) + \mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}^{(2)}, \boldsymbol{R}\right)
\end{aligned}
$$

$\square$

Based on the inequality derived in Proposition 4.1, we further analyze the generalization error bound of the ideal loss function that is independent of the propensity score.

Lemma 1. (Hoeffding's Inequality) Independent bounded random variables $Z_1, ..., Z_n$ that take values in intervals of sizes $\zeta_1, ..., \zeta_n$ satisfy the following inequality for any $\epsilon > 0$.

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} Z_i - \mathbb{E}\left[\sum_{i=1}^{n} Z_i\right]\right| \geq \epsilon\right) \leq 2 \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^{n} \zeta_i^2}\right) \tag{7}$$

See Theorem 2 in [28] for the proof.

Theorem 1. (Propensity-agnostic generalization error bound) A pseudo-labeled dataset $\widetilde{D}$, and two predicted matrices $\widehat{\boldsymbol{R}}^{(1)}$ and $\widehat{\boldsymbol{R}}^{(2)}$ are given. In addition, a loss function ℓ obeys the triangle inequality and is bounded above by a positive constant M. Then, for any $\widehat{\boldsymbol{R}} \in \mathcal{H}$ where $\mathcal{H} = \{\widehat{\boldsymbol{R}}_1, \ldots, \widehat{\boldsymbol{R}}_{|\mathcal{H}|}\}$ is a given finite hypothesis space, and for any $\delta \in (0, 1)$, the following inequality holds with a probability of at least $1 - \delta$.

$$
\begin{aligned}
\mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}, \boldsymbol{R}\right) \leq \ & \underbrace{\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)}_{(a)} + bias\left(\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right) \\
&+ \underbrace{\mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}^{(1)}, \widehat{\boldsymbol{R}}^{(2)}\right)}_{(b)} + \underbrace{\mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}^{(2)}, \boldsymbol{R}\right)}_{(c)} + \frac{M}{|\widetilde{\mathcal{D}}|}\sqrt{\frac{|\mathcal{D}|}{2}\log\left(\frac{2\,|\mathcal{H}|}{\delta}\right)}
\end{aligned} \tag{8}
$$

where

$$bias\left(\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right) = \mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) - \mathbb{E}\left[\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right]$$

Proof. We prove that the following inequality holds with a probability of at least $1 - \delta$:

$$\mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) \leq \widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) + bias\left(\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right) + M\sqrt{\frac{1}{2|\widetilde{\mathcal{D}}|}\log\left(\frac{2\,|\mathcal{H}|}{\delta}\right)} \tag{9}$$

7

First, the following equation holds:

$$\mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) = \mathcal{L}_{ideal}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) - \mathbb{E}\left[\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right] + \mathbb{E}\left[\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right]$$

$$= \mathbb{E}\left[\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right] + bias\left(\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right) \qquad (10)$$

Here $\{O'_{u,i}\}$ are independent, and we apply Hoeffding's inequality in Lemma 4.2 to $\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)$, which yields:

$$\mathbb{P}\left(\left|\mathbb{E}\left[\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right] - \widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right| \geq \epsilon\right) \leq 2\exp\left(\frac{-2|\widetilde{\mathcal{D}}|^2\epsilon^2}{|\mathcal{D}|M^2}\right)$$

$$(11)$$

We obtain Eq. (11) because the loss function takes value in the interval $[0, M]$ by assumption. Furthermore,

$$\mathbb{P}\left(\left|\mathbb{E}\left[\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right] - \widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right| \geq \epsilon\right)$$

$$\leq \mathbb{P}\left(\max_{\widehat{\boldsymbol{R}}' \in \mathcal{H}}\left|\mathbb{E}\left[\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}', \widehat{\boldsymbol{R}}^{(1)}\right)\right] - \widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}', \widehat{\boldsymbol{R}}^{(1)}\right)\right| \geq \epsilon\right)$$

$$\leq \mathbb{P}\left(\bigcup_{\widehat{\boldsymbol{R}}' \in \mathcal{H}}\left|\mathbb{E}\left[\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}', \widehat{\boldsymbol{R}}^{(1)}\right)\right] - \widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}', \widehat{\boldsymbol{R}}^{(1)}\right)\right| \geq \epsilon\right)$$

$$\leq \sum_{\widehat{\boldsymbol{R}}' \in \mathcal{H}} 2\exp\left(\frac{-2|\widetilde{\mathcal{D}}|^2\epsilon^2}{|\mathcal{D}|M^2}\right) \leq 2\,|\mathcal{H}|\exp\left(\frac{-2|\widetilde{\mathcal{D}}|^2\epsilon^2}{|\mathcal{D}|M^2}\right)$$

We set $\delta = 2\,|\mathcal{H}|\exp\left(\frac{-2|\widetilde{\mathcal{D}}|^2\epsilon^2}{|\mathcal{D}|M^2}\right)$, and solving it for ϵ yields:

$$\mathbb{P}\left(\left|\mathbb{E}\left[\widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right] - \widehat{\mathcal{L}}_{pseudo}^{\ell}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right| \leq \frac{M}{|\widetilde{\mathcal{D}}|}\sqrt{\frac{|\mathcal{D}|}{2}\log\left(\frac{2\,|\mathcal{H}|}{\delta}\right)}\right) \geq 1 - \delta \qquad (12)$$

By combining Eq. (10) and Eq. (12), Eq. (9) is obtained. Finally, combining Eq. (6) and Eq. (9) completes the proof. $\square$

As suggested in Theorem 4.3, the following three conditions are essential to achieve a small ideal loss: (a) the loss with respect to the pseudo-ratings. (b) the similarity of the predicted values by A_1 and A_2. (c) the ideal loss of A_2 with respect to the true ratings.

In the experiments, we used the MF-IPS as A_2, because it generally achieves a smaller ideal loss than that of other simple factorization models in MNAR settings [6, 7] and is thus expected to provide better pseudo-ratings. Note that the derived generalization error bound is independent of the propensity score, even if we use propensity-based algorithms for A_1 or A_2. This is because the pseudo-labeling step and the final prediction step of the proposed algorithm do not use propensity scoring, and thus, the high variance and the propensity misspecification problems are avoided.

It should also be noted that the asymmetric tri-training framework is interpreted as minimizing the generalization error bound derived in Theorem 4.3[3]. As described in Algorithm 1, two of the three predictors A_1 and A_2 are trained independently using the observed rating dataset $\mathcal{O}$. Subsequently, the other predictor A_3 is trained using the pseudo-MCAR dataset generated by pre-trained A_1 and A_2. In the pseudo-labeling step, the predicted value of A_1 (i.e., $\widehat{R}_{u,i}^{(1)}$) is regarded as the pseudo rating for the pair (u, i) if the two predicted values are sufficiently similar. Thus, the value of (b) in the RHS of the generalization error bound is kept small during the pseudo-labeling step. In addition, the other predictor A_3 is trained using the pseudo-MCAR dataset $\widetilde{\mathcal{D}}$, and this minimizes the value of (a) in the RHS of the generalization error bound. Therefore, the asymmetric tri-training is a theoretically principled approach because it minimizes the generalization error bound of the ideal loss function.

In the experiments, we empirically demonstrate that the proposed method minimizes the sums of two terms ((a) and (b)) of the generalization error bound in Eq. (8) during training, and one can further improve the prediction accuracy by minimizing the bound.

5   Experimental Results

We conducted two comprehensive experiments to evaluate the effectiveness of the proposed meta-learning method. First, we investigated the robustness of the proposed method using a movie rating dataset. Then, we compared our methods with the other standard matrix factorization models on benchmark real-world datasets.

5.1   Datasets

We used the following benchmark datasets.

- Movie Lens 100K and 1M datasets[4]: They contain five-star movie ratings collected from a movie recommendation service, and the ratings are MNAR. The 100K dataset involves approximately 100,000 ratings from 943 users and 1,682 movies, in contrast, the 1M dataset involves approximately 1 million ratings from 6,040 users and 3,706 movies. In the experiments, we kept movies that had been rated by at least *min_items*, and the values of *min_items* varied with respect to the experimental settings.

- Yahoo! R3 dataset[5]: It contains five-star user-song ratings. The training data contains approximately 300,000 MNAR ratings from 154,000 users against 1,000 songs, and the test data is collected by asking a subset of 54,000 users to rate 10 randomly selected songs. Thus, the test data is regarded as a MAR dataset.

- Coat dataset[6]: It contains five-star user-coat ratings from 290 Amazon Mechanical Turk workers on an inventory of 300 coats. The training data contains 6,500 MNAR ratings collected through self-selections by the Turk workers. On the other hand, the test data is MAR collected by asking the Turk workers to rate 16 randomly selected coats.

---

[3]The analogous discussion can be found in [14]; however, their discussion is based on the unsupervised domain adaptation setting and complexity of deep neural networks, which is different from the MNAR recommendation setting.

[4]http://grouplens.org/datasets/movielens/

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

[6]https://www.cs.cornell.edu/~schnabts/mnar/

5.2 Baselines

Here we describe the baseline and the proposed methods compared in the experiments. We implemented all the methods in the Tensorflow environment.

- **Probabilistic Matrix Factorization (PMF)** [24]: PMF is one of the most basic latent factor models and predicts the rating of (u,i) by $\widehat{R}_{u,i} = \theta_u^\top \beta_i$, where $\{\theta_u\}$ and $\{\beta_i\}$ are user and item latent factors. PMF optimizes its parameters by minimizing the naive loss in Eq. (2) with regularization terms.

- **Matrix Factorization** [25]: MF is one of the most popular prediction models predicting the rating of (u,i) by $\widehat{R}_{u,i} = \theta_u^\top \beta_i + b_u + b_i + b$, where b_u and b_i are the user and item bias terms, respectively. b is the global bias. MF optimizes its parameters by minimizing the naive loss in Eq. (2) with regularization terms.

- **Matrix Factorization with Inverse Propensity Score** [6]: MF-IPS is based on the matrix factorization models. It predicts each rating in the same way as the MF, but it optimizes its parameters by minimizing the IPS loss in Eq. (3) with regularization terms.

- **MF-IPS with asymmetric-tri training (MF-IPS with AT)**: We used MF-IPS with different initializations for A_1 and A_2, and the vanilla MF for A_3. Thus, the final training step is guaranteed to be independent of the propensity score.

5.3 Experimental Setup

- **train/validation/test split**: For the Movie Lens 100K datasets, we followed the procedure used in [7, 29] and created a test set having different item distribution from the original one. We created it by first sampling a test set with 50% of the original dataset, then, resampling data from the test set based on the **inverse** of the relative item probabilities in Eq. (13) with $\gamma = 1.0$. This creates a test set, such that each item has uniform observed probability.

$$P_{u,i} = \left(\frac{\sum_{u \in \mathcal{U}} O_{u,i}}{\max_{i \in \mathcal{I}} \sum_{u \in \mathcal{U}} O_{u,i}} \right)^{\gamma} \tag{13}$$

  For the Yahoo! R3 and Coat datasets, the original datasets were divided into training and test sets. We randomly selected 10% of the original training set for the validation set.

- **hyperparameters**: For all the baselines, the tuning of the regularization hyperparameter was performed in the range of $\{10^{-2}, \ldots, 10^{-6}\}$ and that of the the dimensions of the latent factors was performed in the range of $\{5, 10, 15, 20, 25, 30\}$ for all the datasets. For the proposed method, we used the same combination of the hyperparameters with the MF for the base algorithms (A_1, A_2, A_3) and tuned ε in the range of $\{0.1, 0.001, 0.001\}$. For all the methods, we conducted mini-batch optimization with a batch size of 512 using the Adam optimizer [30] with a learning_rate of 0.01.

5.4 Results

Here, we present the experimental results in detail.

(Left) Robustness to the variance problem   (Right) Robustness to the misspecification
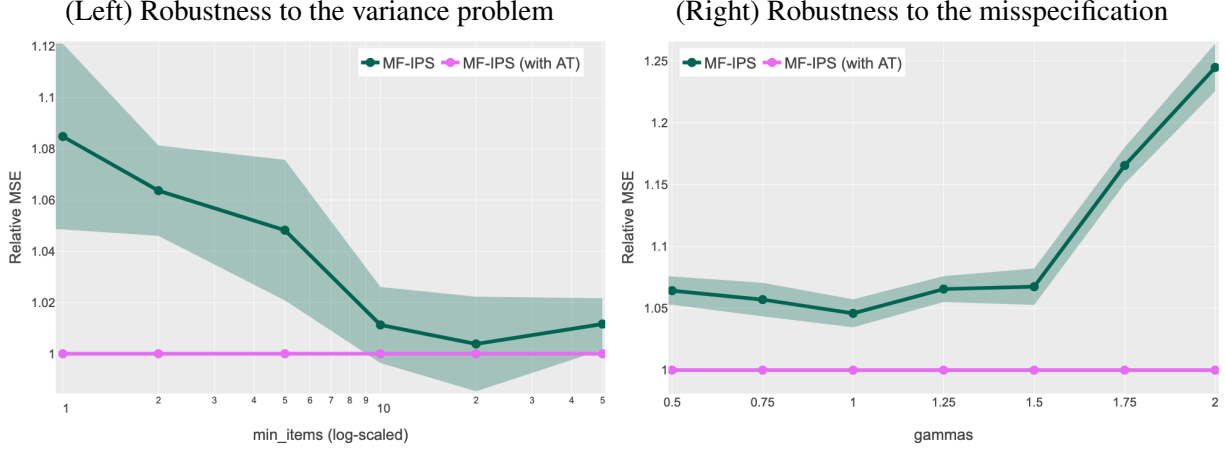
Figure 1: Relative prediction accuracies and their standard errors of MF-IPS and MF-IPS with asymmetric tri-training (MF-IPS with AT) on a different value of *min_items*. (Left) Both methods were trained with the **specified** propensity model. The proposed asymmetric tri-training framework significantly outperforms the vanilla MF-IPS, especially when a large skewness of the propensity score distribution is present (with a small value of *min_items*). (Right) Both methods were trained with the **misspecified** propensity model. The proposed method outperformes the other, when the propensity misspecification is severe.

5.4.1 Robustness to the skewness of propensity score

First, we evaluated the influence of the skewness of the propensity score distribution on the performance of the MF-IPS and MF-IPS with the asymmetric tri-training (MF-IPS with AT) using the Movie Lens 100K dataset. To evaluate the effects of skewness, we investigated the performance corresponding to varying values of the *min_items*[7]. A smaller value of *min_items* introduces a large skewness of the propensity score distribution because the minimum value of the propensity score in Eq. (13) also becomes small. For example, when *min_items* is 1, the minimum relative propensity is 0.0017, on the other hand, when *min_items* is 50, the minimum relative propensity is 0.0859. Note that each model was trained with the true value of $\gamma = 1.0$ (the propensity model is specified) to evaluate the pure effect of the skewness.

Figure 1 (Left) shows the effect of the skewness of the propensity score distribution on the performance of the vanilla MF-IPS and MF-IPS with AT. The result shows that the vanilla MF-IPS is severely affected by the skewness of the propensity distribution, its performance is worsened for a smaller *min_items*. This is because the IPS approach generally suffers from the skewness and variance of the loss function based on the propensity score. In contrast, the MF-IPS with AT relatively performed well, especially when the skewness of the propensity score distribution was large. This is because the final prediction step of our asymmetric tri-training does not rely on the propensity score and thus does not suffer from the variance problem. The result empirically shows that the proposed meta-learning method is robust to the skewness of the propensity score distribution.

5.4.2 Robustness to the propensity score misspecification

Subsequently, we evaluated the influence of the misspecification of the propensity score estimation model on the performance of the vanilla MF-IPS and MF-IPS with AT using the Movie Lens 100K dataset. To evaluate

---

[7]The values of *min_items* were set as 1, 2, 5, 10, 20, and 50

the effects of misspecification, we investigated the performance of the methods trained from the propensity score estimates with varying values of $\gamma \in [0.5, 2.0]$. The misspecification occurred because the values of γ used in the model training and test set creation were different.

Figure 1 (Right) shows the influence of the misspecification of the propensity score estimation on the performance of the vanilla MF-IPS and MF-IPS with AT. The results suggest that the vanilla MF-IPS is severely affected by the propensity model misspecification because it demonstrates worse performance, especially when the value of γ used in the model evaluation is considerably different from 1.0. In contrast, the MF-IPS with AT is not substantially affected by the propensity model misspecification and largely outperforms the vanilla MF-IPS under a larger value of γ. This result shows that the proposed asymmetric tri-training is robust to the misspecification of the propensity estimation model.

Datasets	Movie Lens 1M	Yahoo! R3	Coat
Methods / Metrics	MSE ($\pm$ SE)	MSE ($\pm$ SE)	MSE ($\pm$ SE)
PMF	1.547 ($\pm$ 0.020)	1.308 ($\pm$ 0.017)	1.462 ($\pm$ 0.142)
MF	1.106 ($\pm$ 0.008)	1.905 ($\pm$ 0.050)	1.181 ($\pm$ 0.004)
MF-IPS	1.087 ($\pm$ 0.007)	1.075 ($\pm$ 0.001)	1.229 ($\pm$ 0.020)
MF-IPS (with AT)	**1.053 ($\pm$ 0.009)**	**1.055 ($\pm$ 0.001)**	**1.157 ($\pm$ 0.005)**

Table 1: Results on Movie Lens 1M, Yahoo! R3, and Coat datasets. For all the methods, the average results with their standard errors (SE) over five different simulations are reported. The MF-IPS with AT is the combination of the proposed asymmetric tri-training framework and the MF-IPS. The proposed method significantly outperforms the other methods on all datasets.

5.4.3 Performance on real-world benchmark datasets

Finally, we compared the performance of the proposed method with the baseline methods on the Movie Lens 1M ($min\_items = 10$), the Yahoo! R3, and the Coat datasets. We estimated the propensity score for training the MF-IPS by the relative item popularity as in Eq. (13) with $\gamma = 1.0$[8].

First, Table 1 indicates that MF-IPS with AT significantly outperforms the other methods. This result validates that the proposed method is useful to improve the prediction accuracy of matrix-factorization-based models on biased real-world recommender systems.

Moreover, Figure 2 shows that the values of the loss on pseudo-labels, (a), the similarity between A_1 and A_2, (b), and their summations, $(a) + (b)$ during the pseudo-labeling step of the proposed method for the three datasets. It can be noted that the proposed meta-learning method minimizes the sum of (a) and (b) of the generalization error bound in Eq. (8). In addition, Figure 3 suggests that the MSE on the test sets considerably decreases during the pseudo-labeling step, and thus, minimizing the generalization error bound is empirically justified to be an effective way to improve the prediction accuracy.

6   Conclusion

In this study, we explored the problem of learning rating predictors from missing-not-at-random explicit rating datasets. First, we proposed a meta-learning method called the asymmetric tri-training for missing-not-at-random recommendation, which is based on a method for unsupervised domain adaptation. Next, we

---

[8]In contrast to previous studies [6, 8], we did not use any data in the test set for the propensity estimation to imitate the real-world situation.

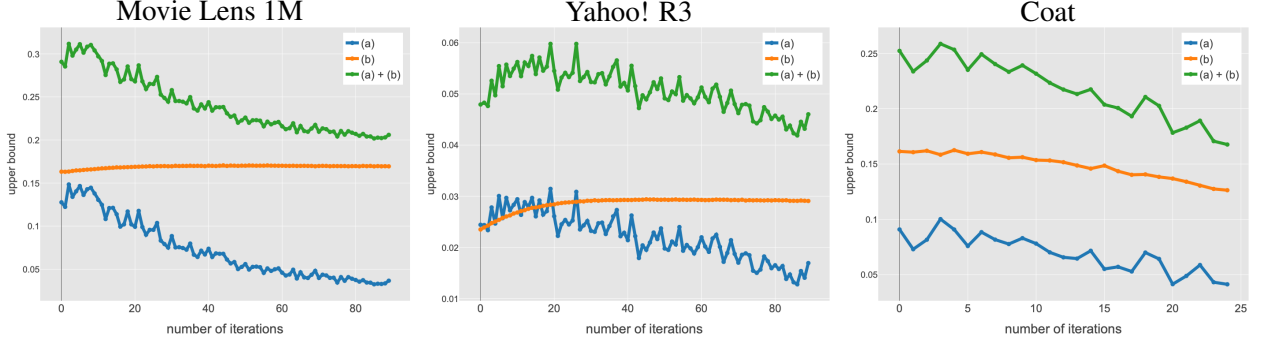Figure 2: Values of the loss on the pseudo-MCAR dataset, (a), and the similarity between A_1 and A_2, (b), in Eq. (8) during the pseudo-labeling step of the proposed method. The green line represents the sum of the two terms. The results show that asymmetric tri-training minimizes the sum of the two terms during training. Note that the pseudo-labeling step was stopped when there was no further improvement of the MSE on the validation set.
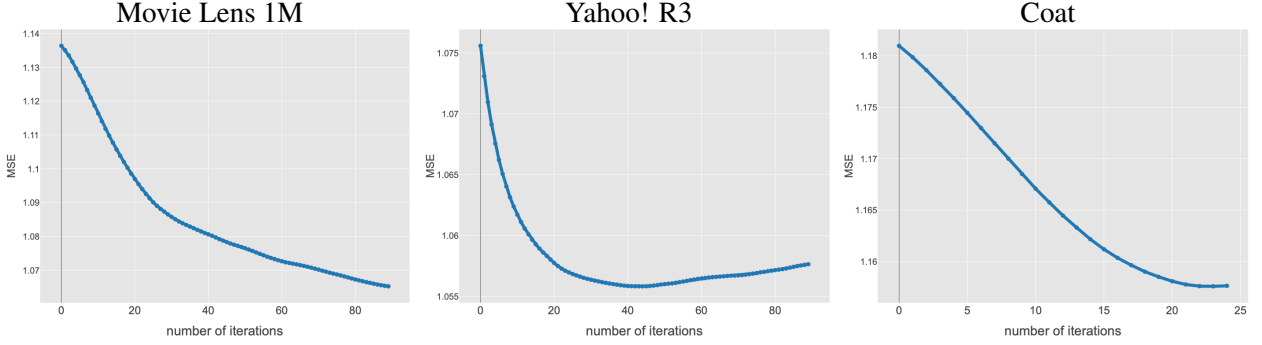

Figure 3: The MSE on the test set during the pseudo-labeleing step of the proposed method. The values almost monotonically decrease with iterations. This results suggest that minimizing the upper bound in Eq. (8) is a valid approach to improve rating predictors.

derived the propensity-agnostic generalization error bound of the ideal loss function and demonstrated that the meta-learning method minimizes the bound and is thus theoretically principled. Finally, experimental results showed that the proposed method is robust to the skewness of the propensity distribution and the misspecification of the propensity estimation model. Moreover, the proposed method effectively improved the prediction accuracy of matrix factorization models on standard benchmark datasets.

13

Bibliography

[1] Bruno Pradel, Nicolas Usunier, and Patrick Gallinari. Ranking with non-random missing ratings: influence of popularity and positivity on evaluation metrics. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 147–154. ACM, 2012.

[2] 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.

[3] 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.

[4] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[5] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[6] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. Recommendations as treatments: Debiasing learning and evaluation. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1670–1679, New York, New York, USA, 20–22 Jun 2016. PMLR.

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

[8] 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.

[9] Yuta Saito, Hayato Sakata, and Kazuhide Nakata. Doubly robust prediction and evaluation methods improve uplift modeling for observational data. In *Proceedings of the 2019 SIAM International Conference on Data Mining*, pages 468–476. SIAM, 2019.

[10] Alexandre Gilotte, Clément Calauzènes, Thomas Nedelec, Alexandre Abraham, and Simon Dollé. Offline a/b testing for recommender systems. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 198–206. ACM, 2018.

[11] Adith Swaminathan and Thorsten Joachims. The self-normalized estimator for counterfactual learning. In *advances in neural information processing systems*, pages 3231–3239, 2015.

[12] Kuniaki Saito, Yoshitaka Ushiku, and Tatsuya Harada. Asymmetric tri-training for unsupervised domain adaptation. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2988–2997, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

[13] Seiichi Kuroki, Nontawat Charonenphakdee, Han Bao, Junya Honda, Issei Sato, and Masashi Sugiyama. Unsupervised domain adaptation based on source-guided discrepancy. *arXiv preprint arXiv:1809.03839*, 2018.

[14] Jongyeong Lee, Nontawat Charoenphakdee, Seiichi Kuroki, and Masashi Sugiyama. Domain discrepancy measure using complex models in unsupervised domain adaptation. *arXiv preprint arXiv:1901.10654*, 2019.

[15] José Miguel Hernández-Lobato, Neil Houlsby, and Zoubin Ghahramani. Probabilistic matrix factorization with non-random missing data. In *International Conference on Machine Learning*, pages 1512–1520, 2014.

[16] Yixin Wang, Dawen Liang, Laurent Charlin, and David M. Blei. The deconfounded recommender: A causal inference approach to recommendation. *CoRR*, abs/1808.06581, 2018.

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

[18] Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. *CoRR*, abs/1103.4601, 2011.

[19] Nan Jiang and Lihong Li. Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*, pages 652–661, 2016.

[20] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. In *Advances in neural information processing systems*, pages 137–144, 2007.

[21] 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.

[22] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1180–1189, Lille, France, 07–09 Jul 2015. PMLR.

[23] 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.

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

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

[26] Steffen Rendle. Factorization machines. In *2010 IEEE International Conference on Data Mining*, pages 995–1000. IEEE, 2010.

[27] Gintare Karolina Dziugaite and Daniel M Roy. Neural network matrix factorization. *arXiv preprint arXiv:1511.06443*, 2015.

[28] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, pages 409–426. Springer, 1994.

[29] 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.

[30] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.