Asymmetric Tri-training for Debiasing Missing-Not-At-Random Explicit Feedback

Yuta Saito
Tokyo Institute of Technology
saito.y.bj@m.titech.ac.jp

ABSTRACT

In most real-world recommender systems, the observed rating data are subject to selection bias, and the data are thus *missing-not-at-random*. Developing a method to facilitate the learning of a recommender with biased feedback is one of the most challenging problems, as it is widely known that naive approaches under selection bias often lead to suboptimal results. A well-established solution for the problem is using propensity scoring techniques. The propensity score is the probability of each data being observed, and unbiased performance estimation is possible by weighting each data by the inverse of its propensity. However, the performance of the propensity-based unbiased estimation approach is often affected by choice of the propensity estimation model or the high variance problem. To overcome these limitations, we propose a model-agnostic meta-learning method inspired by the *asymmetric tri-training* framework for unsupervised domain adaptation. The proposed method utilizes two predictors to generate data with reliable pseudo-ratings and another predictor to make the final predictions. In a theoretical analysis, a propensity-independent upper bound of the true performance metric is derived, and it is demonstrated that the proposed method can minimize this bound. We conduct comprehensive experiments using public real-world datasets. The results suggest that the previous propensity-based methods are largely affected by the choice of propensity models and the variance problem caused by the inverse propensity weighting. Moreover, we show that the proposed meta-learning method is robust to these issues and can facilitate in developing effective recommendations from biased explicit feedback.

KEYWORDS

recommender systems, missing-not-at-random, selection bias, explicit feedback, matrix factorization, unsupervised domain adaptation

ACM Reference Format:
Yuta Saito. 2018. Asymmetric Tri-training for Debiasing Missing-Not-At-Random Explicit Feedback. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/1122445.1122456

Woodstock '18, June 03–05, 2018, Woodstock, NY
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The goal of recommender systems is to recommend items that users will prefer. To achieve this, recommendation algorithms predict the potential preference 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 Etsy) or the user experience in interactive systems (e.g., Spotify and Netflix). Therefore, the field of personalized recommendation has been widely studied in both the academia and industry.

Within the area of recommender systems, most existing studies assume that the observed rating data are *missing-completely-at-random* (MCAR). Generally, this assumption does not hold because real-world recommender systems are subject to *selection bias*. Selection bias occurs primarily due to the following two reasons [36, 41]. 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 observing ratings of such popular items may be large. This leads to the non-uniform missing mechanism, and the MCAR assumption is violated. Second, user self-selection happens, as 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 [29]. Another example is a song recommender system, in which users tend to rate songs that they like or dislike and seldom rate songs they feel neutral about [27]. These findings suggest that most rating datasets collected through real-world recommender systems are *missing-not-at-random* (MNAR). Several studies have theoretically and empirically indicated that the conventional methods of naively using observed ratings lead to suboptimal prediction models, as the observed ratings are not representative data of the target population [36, 37, 44]. Thus, developing a recommendation algorithm and a debiasing method that can achieve a high prediction accuracy using MNAR feedback 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 approaches are propensity-based debiasing methods such as *inverse propensity score* (IPS) and *doubly robust* (DR) estimations. These methods have been established in fields such as causal inference and missing data analysis [31, 32] and have been proposed to be utilized for debiasing learning and evaluation of the MNAR recommendation [25, 36]. IPS estimation relies on the propensity score, which is the probability of observing each rating. By weighting each sample by the inverse of its propensity score, one can unbiasedly estimate the loss function of interest using the biased rating feedback. The benefits of these propensity-based methods are theoretically principled and empirically outperform naive methods based on the unrealistic MCAR assumption [25, 36].

To ensure effectiveness of such propensity-based methods, accurately estimating the propensity score is critical. This is because the unbiasedness of the performance estimator is guaranteed only when the true propensities are available; the IPS estimator still has a bias depending on the propensity estimation bias [36, 41]. However, correctly estimating the propensity score is almost impossible, and model misspecification often occurs in real-world settings [34]. Moreover, propensity-based methods generally suffer from high variance, which can lead to suboptimal estimation when the item popularity or user activeness is highly diverse [12, 39, 41]. Improving the robustness to the choice of propensity estimator and the high variance problem of the propensity weighting technique are the important and unsolved issues.

To address the limitations of the propensity-based recommendation methods, in this work, we propose a model-agnostic meta-learning method inspired by the *asymmetric tri-training* framework in unsupervised domain adaptation [33]. Similar to causal inference, unsupervised domain adaptation addresses problem settings in which the data-generating distributions are different between the training and test sets [22, 23]. Moreover, it relies on the upper bound minimization approach, which minimizes the propensity-independent upper bound of the loss function of interest. Thus, this approach is considered to be useful to overcome the issues related to the propensity weighting technique.

In the theoretical analysis, we establish the new upper bound of the ideal loss function and demonstrate that the proposed meta-learning method attempts to minimize this bound. In contrast to the bounds presented in previous studies [36, 41], the upper bound minimized by the proposed method is independent of the propensity score; thus, issues related to the inverse propensity weighting are expected to be solved. Finally, we conduct extensive experiments using public real-world datasets. In particular, we demonstrate that the performance of the previous propensity-based methods is badly affected by the choice of propensity estimation models and the variance of the estimator. We also show that the proposed method significantly improves the recommendation quality, especially for situations where the propensity score is hard to estimate.

The contributions of this paper can be summarized as follows:

- We propose a model-agnostic meta-learning method. The proposed method is the first method that minimizes the propensity-independent upper bound of the ideal loss function to address the selection bias of recommender systems.
- We empirically show that the performance of the propensity-based recommendation is largely affected by choice of propensity estimators and the variance caused by the inverse propensity weighting.
- We demonstrate that the proposed method stably improve the recommendation quality, especially when it is difficult to estimate the propensity score.

2 RELATED WORK

In this section, we review existing related studies.

2.1 Propensity-based Recommendation

Propensity-based methods aim to accurately estimate the loss function of interest using only biased rating feedback [25, 36, 42]. The probability of observing each entry of a rating matrix is defined as the propensity score, and the unbiased estimator for the metric of interest can be derived by weighting each sample by the inverse of its propensity [15, 31, 32]. The method of matrix factorization with IPS (MF-IPS) [25, 36] has been demonstrated to outperform naive matrix factorization and probabilistic generative models [13, 28] under MNAR settings. Moreover, the DR estimation, used in the off-policy evaluation of the bandit algorithms [6, 17], has also been applied to the MNAR recommendation [41]. The DR estimation combines the propensity score estimation and the error imputation model in a theoretically sophisticated manner and improves the statistical properties of the IPS estimator. The error imputation model is the model of predicted errors for the missing ratings, and the performance of the DR estimator has been proven to be dependent on the accuracy of the propensity score estimation and the error imputation model [41].

All the methods stated above are based on explicit feedback. For recommendations using MNAR implicit feedback, [35] is the first work to construct an unbiased estimator for the loss function of interest using only biased implicit feedback. The proposed estimator is a combination of the IPS estimation and positive-unlabeled learning [3, 8]. Although generalization error analysis is not conducted in this work, it is shown that the variance of the propensity-based estimator depends on the inverse of the propensity score, and this leads to a severe variance problem, especially when there exists severe selection bias.

These propensity-based algorithms utilize the unbiased loss function; however, the performance of these methods largely depends on the propensity score estimation. Ensuring the accuracy of the propensity score estimation is difficult for real-world recommenders [44], as the analysts cannot control the missing mechanism. Thus, methods to improve the robustness of propensity-based approaches are highly desired.

2.2 Off-policy Evaluation and Learning

Off-policy evaluation aims to accurately evaluate the performance of contextual bandit policies in offline settings [6, 9, 12]. Most existing off-policy estimators utilize the Direct Method (DM) or the IPS estimation technique [6, 15, 31, 32]. DM predict the reward function using the logged bandit feedback with arbitrary machine learning algorithms, and then, use these predictions to estimate the performance of a given policy. In contrast, the IPS estimation approach uses the propensity score and corrects the distributional shift between the past policy and the new policy that is to be evaluated. It is widely known that the DM approach is subject to the bias problem, and the IPS approach is subject to the variance problem [6]. To explore the best bias-variance trade-off for off-policy evaluation, several combinations of the DM and the IPS approach have been proposed, including DR [6], a more robust doubly robust [9], or SWITCH estimator [43].

In contrast, off-policy learning aims to obtain a well-performing action policy offline using only logged bandit feedback [18, 38]. The fundamental work for developing the off-policy optimization procedure was carried out by [38]; in their paper, they propose the counterfactual risk minimization framework and a corresponding algorithm called POEM. It optimizes the lower bound of the performance of action policies, and this lower bound consists of the mean and variance of the IPS estimator. The other promising approach for off-policy learning is the DACPOL procedure proposed in [2], where the propensity-independent lower bound is optimized via adversarial learning. The derived lower bound consists of the empirical policy performance based on observational data and a distance measure for the distributional shift between the randomized and observational data. The DACPOL procedure empirically outperforms the propensity-based POEM algorithm in situations where past treatment policies (propensities) are unknown [2].

The propensity-based recommendation methods summarized in Section 2.1 are similar to the off-policy evaluation, as both aim to unbiasedly estimate the metric of interest, for example, by using the propensity weighting estimator. However, the upper bound minimization approach, such as the DACPOL framework for off-policy learning, is also theoretically sound and has shown its strength empirically in learning situations when the propensity score is unknown. A method that uses the upper bound minimization approach has not yet been proposed for MNAR recommendation settings, despite that propensity estimation is difficult in real-life recommender systems due to several confounding factors [19, 24, 40].

2.3 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 [33]. 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 to the target domain and measuring the difference between the two domains is critical [23]. Some discrepancy measures to measure this difference have been proposed. Among them, $\mathcal{H}$-divergence and $\mathcal{H}\Delta\mathcal{H}$-divergence [4, 5] have been used to construct many prediction methods. For example, the domain adversarial neural network simultaneously minimizes source empirical errors and the $\mathcal{H}$- divergence between the source and target domains in an adversarial manner [10, 11]. The asymmetric tri-training framework trains three networks asymmetrically and is interpreted as minimizing the $\mathcal{H}$-divergence during training [33].

Our proposed method is based on the upper bound minimization framework, which has shown its effectiveness in unsupervised domain adaptation settings. This work is the first to extend the upper bound minimization approach to MNAR recommendation, and we demonstrate its advantages over the unbiased estimation approach in Section 5.

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 of interest that should be minimized to derive the predictions as

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

where $\ell(\cdot, \cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ is an arbitrary loss function. For example, when $\ell(x, y) = (x - y)^2$, Eq. (1) is the mean-squared-error (MSE).

In reality, it is impossible to calculate the ideal loss function, as most of the true ratings are 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 $\mathcal{O} = \{(u, i) \mid O_{u,i} = 1\}$. When the missing mechanism is MNAR, 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:

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

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 MCAR, the naive estimator is unbiased against the ideal loss function. However, in the case of MNAR datasets, the naive estimator is biased [36, 37], i.e.,

$$\mathbb{E}_{O}\left[\widehat{\mathcal{L}}^{\ell}_{naive}\left(\widehat{\boldsymbol{R}}\right)\right] \neq \mathcal{L}_{ideal}\left(\widehat{\boldsymbol{R}}\right) \tag{2}$$

for some given $\widehat{\boldsymbol{R}}$. Thus, one has to use an estimator that can address this bias issue alternative to using the naive one

3.3 Inverse Propensity Score Estimator

In [25, 36], the authors applied the IPS estimation to address the bias 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 [15, 31, 32]. The basic idea of this estimator is to create a pseudo-MCAR dataset by weighting 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

[1] mapping from feature space to outcome space

propensity score, the unbiased estimator for the ideal loss function can be derived as follows:

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

This estimator is unbiased against the ideal loss function, i.e.,

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

for any given $\widehat{\boldsymbol{R}}$, and thus considered to be more desirable than the naive estimator.

As theoretically and empirically stated in [36], unbiasedness of the IPS estimator is desirable; however, this property depends on the true propensity score. In reality, the true propensity score is unobservable and thus has to be estimated using the naive Bayes, logistic regression, or Poisson factorization [25, 36]. 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, as the inverse of the propensities might be large [6, 34].

These problems can also be theoretically explained.

THEOREM 3.1. *(Theorem 5.2 of [36]) Suppose that the loss function is bounded above by a positive constant Δ. Then, for any finite hypothesis space of predictions $\mathcal{H} = \{\widehat{\boldsymbol{R}}_1, \dots, \widehat{\boldsymbol{R}}_{|\mathcal{H}|}\}$ and for any $\delta \in (0,1)$, the following inequality holds with a probability of at least $1-\delta$.*

$$\mathcal{L}_{ideal}\left(\widehat{\boldsymbol{R}}_{ERM}\right) \le \widehat{\mathcal{L}}_{IPS}\left(\widehat{\boldsymbol{R}}_{ERM}\mid \boldsymbol{O}\right) + \underbrace{\frac{\Delta}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left|1-\frac{P_{u,i}}{\widehat{P}_{u,i}}\right|}_{bias\ term}$$
$$+ \underbrace{\frac{\Delta}{|\mathcal{D}|}\sqrt{\frac{1}{2}\log\frac{2|\mathcal{H}|}{\delta}}\sqrt{\sum_{(u,i)\in\mathcal{D}}\frac{1}{\hat{P}^2_{u,i}}}}_{variance\ term}$$

where $\widehat{P}_{u,i}$ is an estimated value for $P_{u,i}$, and

$$\widehat{\boldsymbol{R}}_{ERM} = \underset{\widehat{\boldsymbol{R}}\in\mathcal{H}}{\arg\min}\ \widehat{\mathcal{L}}_{IPS}(\widehat{\boldsymbol{R}}\mid\boldsymbol{O})$$

is the empirical risk minimizer.

The generalization error bound of the empirical risk minimizer in Theorem 3.1 depends on both the bias and variance terms. When the estimation error of the propensity estimator is large, the bias term can also be large. Moreover, the variance term depends on the inverse of the estimated propensity scores; the variance problem results in a loose generalization upper bound.

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

4.1 Meta-learning procedure

To realize the objective with only biased rating feedback, we propose the *asymmetric tri-training* framework that utilizes three rating predictors asymmetrically. First, two of the three predictors are trained to generate a reliable dataset with pseudo ratings. Then, the other predictor is trained on that pseudo-ratings. We can use any recommendation algorithm, such as matrix factorization [21, 28], MF-IPS [25, 36], factorization machines [30], and neural network matrix factorization [7], 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.

Algorithm 1 Asymmetric tri-training procedure for missing-not-at-random explicit feedback

```
Input: observed rating dataset $O$, three predictors $A_1, A_2, A_3$, set
    of hyperparameters $\{\epsilon$, number of iterations, number of steps$\}$
Output: predicted rating matrix $\widehat{\boldsymbol{R}}$ by $A_3$
Pre-train $A_1, A_2, A_3$ using the observed rating dataset $O$
randomly sample user-item pairs from $\mathcal{D}$ to generate $\mathcal{D}'$
generate a dataset with pseudo-ratings $\widetilde{\mathcal{D}}$ by Eq. (4)
for $i = 1$ to number of iterations do
    for $j = 1$ to number of steps do
        Update $A_1$ and $A_2$ with mini-batch data from $\widetilde{\mathcal{D}}$
        Update $A_3$ with mini-batch data from $\widetilde{\mathcal{D}}$
    end for
    randomly sample user-item pairs from $\mathcal{D}$ to generate $\mathcal{D}'$
    generate a dataset with pseudo-ratings $\widetilde{\mathcal{D}}$ by Eq. (4)
end for
return $\widehat{\boldsymbol{R}}$ by $A_3$
```

The asymmetric tri-training framework consists of 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 O. Next, we randomly sample user-item pairs[2], denoted as $\mathcal{D}'$. Then, we predict the ratings of the unlabeled dataset $\mathcal{D}'$ using two of the three algorithms A_1 and A_2. The predicted rating for $(u,i)\in\mathcal{D}'$ by A_1 and A_2 is denoted as $\widehat{R}^{(1)}_{u,i}$ and $\widehat{R}^{(2)}_{u,i}$, respectively. We regard one of the two predicted values as the pseudo-rating for (u,i) if the two predicted values are sufficiently similar. By doing this, we can construct a dataset with reliable pseudo-ratings. The resulting dataset is denoted as

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

where $\epsilon > 0$ is a hyperparameter and should be tuned via a parameter tuning procedure. This step is the **pseudo labeling step**.

Finally, we train the remaining predictor A_3 by minimizing the following loss function.

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

where $\{\widehat{R}^{(1)}_{u,i}\}$ 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 created pseudo-labeled dataset $\widetilde{\mathcal{D}}$. This step is called the **final prediction step**.

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

In the algorithm, we iterate the pseudo-labeling step several times to generate a reliable pseudo ratings. Algorithm 1 describes the complete learning procedure of the asymmetric tri-training.

4.2 Theoretical Analysis

In this subsection, we theoretically analyze the MNAR recommendation. Specifically, we drive the propensity-independent upper bound of the ideal loss function and demonstrate that the proposed asymmetric tri-training framework attempts to minimize the part of the upper bound while keeping it informative during training.

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

Proposition 4.1. *Suppose that the loss function ℓ obeys the triangle inequality. Then, for any given predicted rating matrices $\widehat{\boldsymbol{R}}^{(1)}$, $\widehat{\boldsymbol{R}}^{(2)}$, and $\widehat{\boldsymbol{R}}$, the following inequality holds.*

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

Proof. We apply the triangle inequality twice:

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

□

We further analyze the propensity-independent upper bound of the ideal loss function.

Lemma 4.2. *(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 $\eta > 0$.*

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

See Theorem 2 in [14] for the proof.

Theorem 4.3. *(Propensity-independent generalization error bound) Suppose that 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 Δ. 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 $1 - \delta$.*

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

where

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

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

$$\begin{aligned}\mathcal{L}^{\ell}_{ideal}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) &\leq \widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) \\ &+ bias\left(\widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right) + \frac{\Delta}{|\widetilde{\mathcal{D}}|}\sqrt{\frac{|\mathcal{D}|}{2}\log\left(\frac{2\,|\mathcal{H}|}{\delta}\right)}\end{aligned} \tag{5}$$

First, the following equation holds:

$$\begin{aligned}&\mathcal{L}^{\ell}_{ideal}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) \\ &= \mathcal{L}^{\ell}_{ideal}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right) - \mathbb{E}\left[\widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right] + \mathbb{E}\left[\widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right] \\ &= \mathbb{E}\left[\widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right] + bias\left(\widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right)\end{aligned} \tag{6}$$

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

$$\begin{aligned}&\mathbb{P}\left(\left|\mathbb{E}\left[\widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right] - \widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right| \geq \eta\right) \\ &\leq \mathbb{P}\left(\max_{\widehat{\boldsymbol{R}}' \in \mathcal{H}}\left|\mathbb{E}\left[\widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}', \widehat{\boldsymbol{R}}^{(1)}\right)\right] - \widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}', \widehat{\boldsymbol{R}}^{(1)}\right)\right| \geq \eta\right) \\ &\leq \mathbb{P}\left(\bigvee_{\widehat{\boldsymbol{R}}' \in \mathcal{H}}\left|\mathbb{E}\left[\widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}', \widehat{\boldsymbol{R}}^{(1)}\right)\right] - \widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}', \widehat{\boldsymbol{R}}^{(1)}\right)\right| \geq \eta\right) \\ &\leq \sum_{\widehat{\boldsymbol{R}}' \in \mathcal{H}} 2\exp\left(\frac{-2|\widetilde{\mathcal{D}}|^2\eta^2}{|\mathcal{D}|\Delta^2}\right) \leq 2\,|\mathcal{H}|\exp\left(\frac{-2|\widetilde{\mathcal{D}}|^2\eta^2}{|\mathcal{D}|\Delta^2}\right)\end{aligned}$$

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

$$\mathbb{P}\left(\left|\mathbb{E}\left[\widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right] - \widehat{\mathcal{L}}^{\ell}_{pseudo}\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{R}}^{(1)}\right)\right| \leq t\right) \geq 1 - \delta \tag{7}$$

where

$$t = \frac{\Delta}{|\widetilde{\mathcal{D}}|}\sqrt{\frac{|\mathcal{D}|}{2}\log\left(\frac{2\,|\mathcal{H}|}{\delta}\right)}$$

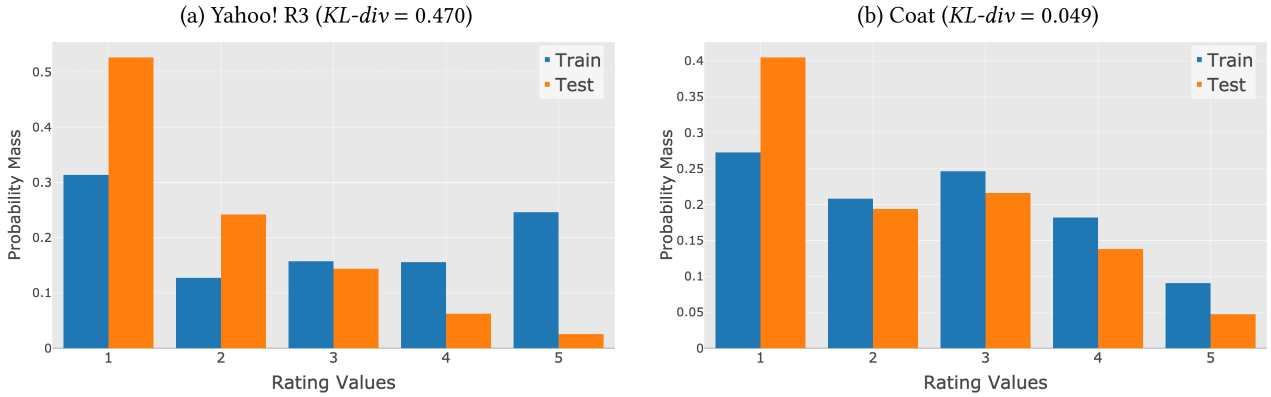
. By combining Eq. (6) and Eq. (7), Eq. (5) is obtained. Finally, combining Proposition 4.1 and Eq. (5) completes the proof. □

As suggested in Theorem 4.3, the following three factors 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.

Note that the derived upper 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 learning procedure do not use the propensity scoring technique, and thus, the high variance and the propensity misspecification problems are avoided in our theoretical bound.

It should also be noted that the asymmetric tri-training framework is interpreted as a method that attempts to minimize the part of the upper bound of the ideal loss in Theorem 4.3. As described



(a) Yahoo! R3 (*KL-div* = 0.470)

(b) Coat (*KL-div* = 0.049)

Figure 1: Comparing rating distributions of training and test sets for Yahoo! R3 and Coat datasets

Notes: The rating distributions are significantly different between the training and test sets for both datasets. Note that *KL-div* is the Kullback–Leibler divergence of the rating distributions between training and test sets. Therefore, the distributional shift of Yahoo! R3 dataset is relatively large compared to that of the Coat dataset.

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 dataset generated by pre-trained A_1 and A_2. In the pseudo-labeling step, A_1 and A_2 are repeatedly updated with same pseudo-ratings, and expected to be similar as the iterations progress. Thus, the value of (b) in the RHS of the upper bound is kept small during the pseudo-labeling step (not minimized), which makes the upper bound informative during training. In addition, the other predictor A_3 is trained using $\widetilde{\mathcal{D}}$, and this minimizes the value of (a). Note that the value of (c) depends on the performance of A_2, and thus, the upper bound can be loose when A_2 performs poorly. Nonetheless, in the experiments, we empirically demonstrate that our method actually minimizes the sum of two terms in the bound (i.e., $(a) + (b)$). We also show that minimizing the upper bound of the ideal loss function is an effective approach to further improve the recommendation quality on the test set.

5 EXPERIMENTAL RESULTS

We conducted comprehensive experiments using benchmark real-world datasets. The code for reproducing the results can be found at **https://github.com/usaito/asymmetric-tri-rec-real**

5.1 Experimental Setup

5.1.1 Datasets and Preprocessing. We used the following real-world datasets.

- MovieLens (ML) 100K dataset[3]: It contains five-star movie ratings collected from a movie recommendation service, and the ratings are MNAR. This dataset involves approximately 100,000 ratings from 943 users and 1,682 movies. In the experiments, we kept movies that had been rated by at least ***min_items*** users, and the values of ***min_items*** varied with respect to the experimental settings.
- Yahoo! R3 dataset[4]: It contains five-star user-song ratings. The training set consists of approximately 300,000 MNAR ratings of 1,000 songs from 15,400 users, and the test set is collected by asking a subset of 5,400 users to rate ten randomly selected songs. Thus, the test set is regarded as an MCAR dataset.
- Coat dataset[5]: It contains five-star user-coat ratings from 290 Amazon Mechanical Turk workers on an inventory of 300 coats. The training set contains 6,500 MNAR ratings collected through self-selections by the Turk workers. In contrast, the test set is MCAR collected by asking the Turk workers to rate 16 randomly selected coats.

For the ML 100K dataset, we created a test set with a different **item distribution** from the original one. We created it by first sampling a test set with 50% of the original dataset, and then, re-sampling data from the test set based on the **inverse** of the relative item probabilities in Eq. (8). This creates a test set, such that each item has a uniform observed probability.

$$P_{*,i} = \frac{\sum_{u\in\mathcal{U}} O_{u,i}}{\max_{i\in\mathcal{I}} \sum_{u\in\mathcal{U}} O_{u,i}} \tag{8}$$

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. Figure 1 shows the rating distributions of training and test sets for the Yahoo! R3 and Coat datasets. The rating distributions are completely different between the training and test sets, which introduces a severe bias when training a recommendation algorithm.

5.1.2 Compared methods and propensity estimators. Here, we describe the baselines and the proposed methods compared in the experiments. We implemented all methods in the *Tensorflow* environment.

Matrix Factorization with Inverse Propensity Score (MF-IPS): MF-IPS is based on the MF model [21]. It predicts each rating by

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

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

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

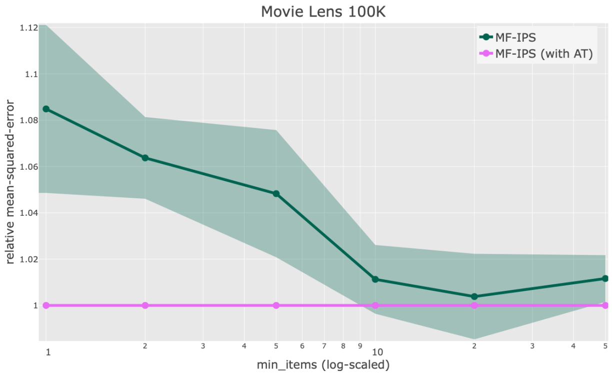


Figure 2: Comparing robustness to the variance issue of recommenders with and without *asymmetric tri-training* (AT)

Notes: The figure reports relative prediction accuracies and their standard errors of MF-IPS with and without AT on a different value of *min_items*. Both methods were trained with the specified propensity model. MF-IPS with AT significantly outperforms that without AT, especially when a large skewness of the propensity score distribution is present (with a small value of *min_items*).

$\widehat{R}_{u,i} = \theta_u^\top \beta_i + b_u + b_i + b$, where $\{\theta_u\}$ and $\{\beta_i\}$ are user and item latent factors, respectively. b_u and b_i are the user and item bias terms. b is the global bias. 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 MF with naive loss in Eq. (2) for A_3. Thus, the final training step is guaranteed to be independent of the propensity score.

For both the baseline and the proposed methods, we tested the following propensity estimators (NB represents *naive Bayes*).

$$\begin{aligned}
\textit{uniform propensity} &: \widehat{P}_{*,*} = \frac{\sum_{u,i \in \mathcal{D}} O_{u,i}}{|\mathcal{D}|} \\
\textit{user propensity} &: \widehat{P}_{u,*} = \frac{\sum_{i \in \mathcal{I}} O_{u,i}}{\max_{u \in U} \sum_{i \in \mathcal{I}} O_{u,i}} \\
\textit{item propensity} &: \widehat{P}_{*,i} = \frac{\sum_{u \in \mathcal{U}} O_{u,i}}{\max_{i \in I} \sum_{u \in \mathcal{U}} O_{u,i}} \\
\textit{user-item propensity} &: \widehat{P}_{u,i} = \widehat{P}_{u,*} \cdot \widehat{P}_{*,i} \\
\textit{NB (uniform)} &: \widehat{P}_{u,i} = \mathbb{P}(R = R_{u,i} \mid O = 1)\mathbb{P}(O = 1) \\
\textit{NB (true)} &: \widehat{P}_{u,i} = \frac{\mathbb{P}(R = R_{u,i} \mid O = 1)\mathbb{P}(O = 1)}{\mathbb{P}(R = R_{u,i})}
\end{aligned}$$

where $R_{u,i} \in \{1, 2, 3, 4, 5\}$ is a realized rating for (u, i). Note that when uniform propensity is used, the MF-IPS is identical to the MF with the naive loss function [21]. NB (true) is often used as a propensity in previous works [36, 41]. However, this estimator cannot be used in most real-world problems, as it requires the MCAR explicit feedback to estimate the prior rating distribution (the numerator); we report the results with the this propensity estimator, just for reference.

5.1.3 *Hyperparameter Tuning.*

For all baselines, the tuning of the L2-regularization hyperparameter was performed in the range of $[10^{-6}, 1]$, and that of the dimensions of the latent factors was performed in the range of $\{5, 10, \ldots, 50\}$. For the proposed method, we used the same hyperparameter tuning procedure as with the baselines for the base algorithms (A_1, A_2, A_3) and tuned ϵ in the range of $[10^{-3}, 1]$. We searched for an optimal set of hyperparameters using an adaptive procedure implemented in *Optuna* [1]. For all methods, we conducted mini-batch optimization with a batch size of 2^{10} using the *Adam* optimizer [20] with an initial learning rate of 0.01. For the proposed method, we set *number of iterations* = 10 and *number of steps* = 10 (see Algorithm 1).

5.2 Results & Discussions

Below, we address the four research questions (RQs).

RQ1. Is the proposed method robust to the variance problem?: First, we evaluated the influence of the skewness of the propensity score distribution on the performance of MF-IPS with and without AT using the ML 100K dataset. To evaluate the effects of skewness, we investigated the performance corresponding to varying values of the *min_items*[6]. A smaller value of *min_items* introduces a large skewness of the propensity score distribution, as the minimum value of the propensity score in Eq. (8) also becomes small. For example, when *min_items* is 1, the minimum relative propensity is 0.0017, in contrast, when *min_items* is 50, the minimum relative propensity is 0.0859. Note that each model was trained with the specified propensity model in Eq. (8) to evaluate the pure effect of the variance.

Figure 2 shows the effect of the skewness of the propensity score distribution on the performance of the MF-IPS with and without AT. The result shows that the MF-IPS without AT 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 variance of the loss function based on the propensity score. In contrast, the MF-IPS with AT performed relatively 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 inverse 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 variance problem.

RQ2. Is the proposed method robust to the choice of propensity score estimator?: Subsequently, we evaluated the influence of the choice of the propensity score estimation model on the performance of the MF-IPS with and without AT. Table 1 summarizes the rating prediction performance evaluated by MSE and MAE and the ranking performance measured by *normalized discounted cumulative gain* (nDCG) [16] on Yahoo! R3 and Coat datasets.

First, for the Yahoo! R3 dataset, MF-IPS without AT is severely affected by the choice of propensity estimator; only MF-IPS with NB (true) achieves the performance reported in previous works [36, 41], and it completely fails in rating prediction with other propensity estimators. Therefore, MF-IPS is highly susceptible to

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

Table 1: Comparing prediction and ranking performances of recommenders with and without *asymmetric tri-training* (AT)

		Metrics					
		MAE		MSE		nDCG@3	
Datasets	Propensity	without AT	with AT	without AT	with AT	without AT	with AT
Yahoo! R3	uniform	1.133	**0.981**	1.907	**1.452**	0.351	**0.352**
	user	1.062	**0.947**	1.712	**1.353**	0.3523	**0.3525**
	item	1.142	**0.989**	1.940	**1.503**	0.351	**0.353**
	user-item	1.162	**0.991**	1.979	**1.513**	0.349	**0.353**
	NB (uniform)	1.170	**1.010**	1.954	**1.511**	0.351	**0.352**
	NB (true)	0.797	**0.765**	1.055	**1.014**	0.351	**0.353**
Coat	uniform	**0.873**	0.878	**1.109**	1.183	0.291	**0.293**
	user	0.873	**0.832**	**1.109**	1.115	0.291	**0.292**
	item	0.873	**0.832**	1.117	**1.115**	0.291	**0.293**
	user-item	0.874	**0.832**	1.117	**1.116**	0.291	**0.293**
	NB (uniform)	0.951	**0.920**	1.268	**1.260**	0.281	**0.289**
	NB (true)	0.852	**0.831**	**1.105**	1.121	0.284	**0.290**

Notes: For all methods, the average results over 20 different initializations and train-validation splits are reported. The results show that the performance of MF-IPS without AT is severely affected by the choice of propensity estimator. The proposed method generally improves the rating prediction (MSE and MAE) and ranking quality (nDCG@3) for both datasets. Moreover, it demonstrates the robustness to the choice of propensity estimators, especially for Yahoo! R3 data.

the propensity misspecification problem. It is difficult to address the effect of selection bias of real-world recommender systems when the MCAR data is unavailable. In contrast, MF-IPS with AT reveals a stable performance with different propensity models and outperforms MF-IPS without AT in most cases. In particular, the proposed method significantly improves the rating prediction accuracies (MSE and MAE) under the realistic situation where the true rating prior is unavailable. Thus, this result validates that the proposed asymmetric tri-training can provide robustness to the choice of the propensity estimation model and improvements of the recommendation quality on biased real-world datasets.

As for the Coat dataset, MF-IPS with and without AT show almost the same rating prediction performance. Moreover, the effect of using different propensity estimators is small. This is because the shift of rating distributions between training and test sets is small in this dataset (see Figure 1). Nonetheless, the proposed method consistently improves the ranking performance (nDCG@3) on this dataset.

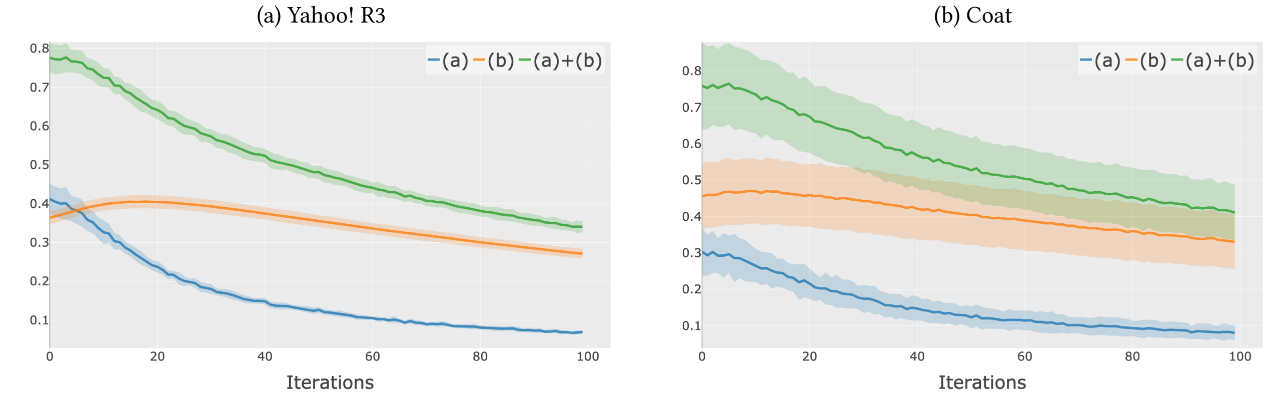
In summary, the performance of MF-IPS is substantially affected by the choice of propensity estimators. It is difficult to reveal reasonable performance in most real-world situations when the NB with true prior propensity estimator cannot be used. In addition, the proposed meta-learning method largely improves the recommendation quality especially for the Yahoo! R3 dataset and demonstrates stable performance across different levels of selection bias.

RQ3. Does the proposed method actually minimize the upper bound of the ideal loss function?: Next, we empirically show that the proposed asymmetric tri-training method can actually minimize the upper bound of the ideal loss function derived in Theorem 4.3.

Figure 3 shows the values of the loss on pseudo-labels (a), the similarity between A_1 and A_2 (b), and their summations $(a)+(b)$ from the pseudo-labeling step of the proposed method. For all datasets, it can be noted that the proposed meta-learning method successfully minimizes the sum of (a) and (b) (the green lines). Thus, as discussed in Section 4.2, the propensity-independent upper bound of the ideal loss function can be minimized effectively using the proposed asymmetric tri-training framework.

RQ4. Is the upper bound minimization approach valid for minimizing the ideal loss function of the test data?: Finally, we demonstrate that minimizing the upper bound of the ideal loss function in Theorem 4.3 is a valid approach for minimizing the ideal loss function of the test set in Eq. (1).

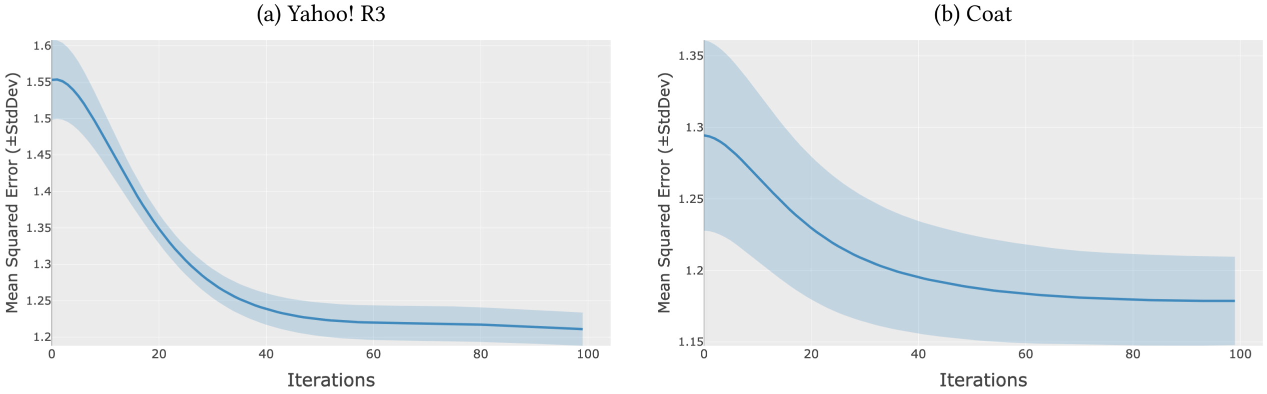
Figure 4 shows the MSE on test sets during the pseudo-labeling step of the proposed method. The results suggest that the MSE on the test sets considerably decreases during the pseudo-labeling step; thus, the upper bound minimization approach is empirically justified as an effective way to improve the prediction accuracy from biased explicit feedback.


(a) Yahoo! R3

(b) Coat

Figure 3: Upper bound minimization performance of *asymmetric tri-training*

Notes: This figure presents averaged values of the loss on generated pseudo ratings (a), and the similarity between A_1 and A_2 (b), in Theorem 4.3 and their standard deviations during the pseudo-labeling step of the proposed method. The green lines represent the sum of the two terms $(a) + (b)$. The results show that asymmetric tri-training minimizes the sum of the two terms (i.e., the upper bound of the ideal loss) during training.


(a) Yahoo! R3

(b) Coat

Figure 4: Improved performance on the test sets by *asymmetric tri-training*

Notes: This figure reports averaged MSEs on the test sets and their standard deviations (StdDev) during the pseudo-labeling step of the proposed method. The values almost monotonically decrease with iterations. These results suggest that minimizing the upper bound of the ideal loss is a valid approach to improve recommenders.

6 CONCLUSION

In this study, we explored the problem of learning recommenders from MNAR explicit feedback. To this end, we proposed a model-agnostic meta-learning method and demonstrated that it minimizes the part of the propensity-independent upper bound of the ideal loss function, while keeping it tight during training. In the experiments, we empirically demonstrated that the previous propensity-based recommendations are subject to the propensity misspecification and variance issues. Furthermore, we showed that the proposed method is robust to the variance and the choice of the propensity estimation model.

As future work, we plan to apply other unsupervised domain adaptation methods, such as domain adversarial learning [10, 11] to the MNAR recommendation. Moreover, we plan to construct a similar learning method for implicit feedback recommendation [19, 26]. Implicit feedback is prevalent in real-world interactive systems; however, methods for debiasing the implicit feedback recommender have not yet been thoroughly investigated. Thus, we believe that the proposed method can have a significant impact on the implicit feedback recommendation.

REFERENCE

[1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2623–2631.

[2] Onur Atan, William R Zame, and Mihaela van der Schaar. 2018. Learning optimal policies from observational data. *arXiv preprint arXiv:1802.08679* (2018).

[3] Jessa Bekker and Jesse Davis. 2018. Learning from positive and unlabeled data: A survey. *arXiv preprint arXiv:1811.04820* (2018).

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

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

[6] Miroslav Dudík, John Langford, and Lihong Li. 2011. Doubly Robust Policy Evaluation and Learning. *CoRR* abs/1103.4601 (2011). arXiv:1103.4601 http://arxiv.org/abs/1103.4601

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

[8] Charles Elkan and Keith Noto. 2008. Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 213–220.

[9] Mehrdad Farajtabar, Yinlam Chow, and Mohammad Ghavamzadeh. 2018. More Robust Doubly Robust Off-policy Evaluation. In *International Conference on Machine Learning*. 1446–1455.

[10] Yaroslav Ganin and Victor Lempitsky. 2015. Unsupervised Domain Adaptation by Backpropagation. In *Proceedings of the 32nd International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Francis Bach and David Blei (Eds.), Vol. 37. PMLR, Lille, France, 1180–1189. http://proceedings.mlr.press/v37/ganin15.html

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

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

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

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

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

[16] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* 20, 4 (2002), 422–446.

[17] Nan Jiang and Lihong Li. 2016. Doubly Robust Off-policy Value Evaluation for Reinforcement Learning. In *International Conference on Machine Learning*. 652–661.

[18] Thorsten Joachims, Adith Swaminathan, and Maarten de Rijke. 2018. Deep learning with logged bandit feedback. (2018).

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

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

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

[22] Seiichi Kuroki, Nontawat Charonenphakdee, Han Bao, Junya Honda, Issei Sato, and Masashi Sugiyama. 2018. Unsupervised Domain Adaptation Based on Source-guided Discrepancy. *arXiv preprint arXiv:1809.03839* (2018).

[23] Jongyeong Lee, Nontawat Charoenphakdee, Seiichi Kuroki, and Masashi Sugiyama. 2019. Domain Discrepancy Measure Using Complex Models in Unsupervised Domain Adaptation. *arXiv preprint arXiv:1901.10654* (2019).

[24] Shuai Li, Yasin Abbasi-Yadkori, Branislav Kveton, S Muthukrishnan, Vishwa Vinay, and Zheng Wen. 2018. Offline evaluation of ranking policies with click models. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1685–1694.

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

[26] Dawen Liang, Laurent Charlin, James McInerney, and David M Blei. 2016. Modeling user exposure in recommendation. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 951–961.

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

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

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

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

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

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

[33] Kuniaki Saito, Yoshitaka Ushiku, and Tatsuya Harada. 2017. Asymmetric Tri-training for Unsupervised Domain Adaptation. In *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Doina Precup and Yee Whye Teh (Eds.), Vol. 70. PMLR, International Convention Centre, Sydney, Australia, 2988–2997. http://proceedings.mlr.press/v70/saito17a.html

[34] Yuta Saito, Hayato Sakata, and Kazuhide Nakata. 2019. Doubly Robust Prediction and Evaluation Methods Improve Uplift Modeling for Observational Data. In *Proceedings of the 2019 SIAM International Conference on Data Mining*. SIAM, 468–476.

[35] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. 2020. Unbiased Recommender Learning from Missing-Not-At-Random Implicit Feedback. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 501–509.

[36] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as Treatments: Debiasing Learning and Evaluation. In *Proceedings of The 33rd International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Maria Florina Balcan and Kilian Q. Weinberger (Eds.), Vol. 48. PMLR, New York, New York, USA, 1670–1679. http://proceedings.mlr.press/v48/schnabel16.html

[37] Harald Steck. 2010. 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*. ACM, 713–722.

[38] Adith Swaminathan and Thorsten Joachims. 2015. Counterfactual risk minimization: Learning from logged bandit feedback. In *International Conference on Machine Learning*. 814–823.

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

[40] Xuanhui Wang, Nadav Golbandi, Michael Bendersky, Donald Metzler, and Marc Najork. 2018. Position bias estimation for unbiased learning to rank in personal search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 610–618.

[41] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2019. Doubly Robust Joint Learning for Recommendation on Data Missing Not at Random. In *International Conference on Machine Learning*. 6638–6647.

[42] Yixin Wang, Dawen Liang, Laurent Charlin, and David M. Blei. 2018. The Deconfounded Recommender: A Causal Inference Approach to Recommendation. *CoRR* abs/1808.06581 (2018). arXiv:1808.06581 http://arxiv.org/abs/1808.06581

[43] Yu-Xiang Wang, Alekh Agarwal, and Miroslav Dudik. 2017. Optimal and adaptive off-policy evaluation in contextual bandits. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 3589–3597.

[44] Longqi Yang, Yin Cui, Yuan Xuan, Chenyang Wang, Serge Belongie, and Deborah Estrin. 2018. Unbiased Offline Recommender Evaluation for Missing-not-at-random Implicit Feedback. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) *(RecSys '18)*. ACM, New York, NY, USA, 279–287. https://doi.org/10.1145/3240323.3240355