Unbiased Recommender Learning from Missing-Not-At-Random Implicit Feedback

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

Suguru Yaginuma
SMN Corporation
suguru_yaginuma@so-netmedia.jp

Yuta Nishino
SMN Corporation
yuta_nishino@so-netmedia.jp

Hayato Sakata
SMN Corporation
hayato_sakata@so-netmedia.jp

Kazuhide Nakata
Tokyo Institute of Technology
nakata.k.ac@m.titech.ac.jp

ABSTRACT

Recommender systems widely use implicit feedback such as click data because of its general availability. Although the presence of clicks signals the users' preference to some extent, the lack of such clicks does not necessarily indicate a negative response from the users, as it is possible that the users were not exposed to the items (positive-unlabeled problem). This leads to a difficulty in predicting the users' preferences from implicit feedback. Previous studies addressed the positive-unlabeled problem by uniformly upweighting the loss for the positive feedback data or estimating the confidence of each data having relevance information via the EM-algorithm. However, these methods failed to address the missing-not-at-random problem in which popular or frequently recommended items are more likely to be clicked than other items even if a user does not have a considerable interest in them. To overcome these limitations, we first define an ideal loss function to be optimized to realize recommendations that maximize the relevance and propose an unbiased estimator for the ideal loss. Subsequently, we analyze the variance of the proposed unbiased estimator and further propose a clipped estimator that includes the unbiased estimator as a special case. We demonstrate that the clipped estimator is expected to improve the performance of the recommender system, by considering the bias-variance trade-off. We conduct semi-synthetic and real-world experiments and demonstrate that the proposed method largely outperforms the baselines. In particular, the proposed method works better for rare items that are less frequently observed in the training data. The findings indicate that the proposed method can better achieve the objective of recommending items with the highest relevance.

CCS CONCEPTS

• **Information systems** → **Collaborative filtering**; • **Computing methodologies** → **Learning from implicit feedback**.

KEYWORDS

Implicit Feedback; Missing-Not-At-Random; Inverse Propensity Weighting; Positive-Unlabeled Learning; Matrix Factorization.

1 INTRODUCTION

Recommender systems are widely used in industries such as personalized advertising and video streaming [3, 9, 10]. These services utilize recommender systems to predict and provide users with items that they may be interested in to improve user experience. To provide such recommendations, predicting the items that users like is an important task [7, 9, 16].

Generally, one can use two types of data for the recommendation. One includes users' ratings on items, and the other corresponds to users' clicks (e.g., purchases, views) [9, 17]. Rating data is called explicit data, as they represent the preferences explicitly via positive or negative feedback. Typically, collecting sufficient explicit data for recommendation is difficult because it requires users to actively provide ratings [9]. In contrast, click data, which is called implicit data, are easy to collect because they represent the behavior logs of the users [10]. Any services, even those without the users' ratings, can utilize the data, and many of the actual recommender systems are based on implicit data [9, 19]. As in the information retrieval field [11, 12, 17, 27], one should estimate the relevance of an item to a user from implicit feedback to improve user experience, because one can recommend items that users are genuinely interested in by accurately predicting its relevance. However, as the implicit data is not the explicit feedback of the users' preferences, one cannot know whether unclicked feedback is negative feedback or unlabeled positive feedback; this corresponds to the positive-unlabeled problem [2, 5, 9].

To predict the relevance, [7] proposed weighted matrix factorization (WMF), which assigns unclicked items less weight to incorporate the idea that those items correspond to less confidence in prediction than clicked items. However, in some cases, one might be more confident in predicting the relevance of some unclicked items than for the other unclicked ones. Unclicked items that have been recommended several times indicate that they are less relevant for the users. Exposure matrix factorization (ExpoMF) fully utilizes this

exposure information [17]. The authors introduced exposure variables and the latent probabilistic model, in which the probability of a click is the product of the probability of the exposure and relevance level. Under this probabilistic model, when a user has been exposed to an item, clicked and unclicked information between the pair can be regarded as relevance information. Thus, ExpoMF upweights the loss of data with high exposure probability because the exposure probability is regarded as the confidence of how much relevance information each data includes. Although ExpoMF tackles the positive-unlabeled problem by introducing the exposure variables, it does not address another critical issue of implicit feedback recommendation, namely the missing-not-at-random (MNAR) problem [18, 23, 28]. This is because, by upweighting the loss of data with high exposure probability (mostly popular items), it can lead to poor prediction accuracy for rare items. Thus, ExpoMF suffers from bias and cannot achieve the objective of recommender systems to provide users with items that are relevant to them.

In this study, to establish a method to predict items with the highest relevance rather than the highest click probability, we first define the ideal loss function that should be optimized to maximize the relevance. We theoretically demonstrate that the existing methods (e.g., WMF, ExpoMF) have biases in their loss functions toward the ideal loss. Next, we simultaneously solve the positive-unlabeled problem and the MNAR problem inspired by the estimation technique for causal inference [8, 20, 21] and derive an unbiased estimator for the ideal loss estimated only from the observable feedback. Further, we analyze the variance of the proposed unbiased estimator and highlight that the variance could be large under recommendation settings. Moreover, we propose a clipped estimator that could achieve a better bias-variance trade-off than that of the unbiased estimator and investigate its theoretical characteristics. Finally, in the experiments, we verify the effectiveness of the proposed method on both semi-synthetic and real-world datasets.

The contributions of this research are as follows.

- We define an ideal loss function to be optimized to maximize relevance. We propose an unbiased estimator for the loss function of interest for the first time.
- We perform theoretical analyses for the statistical properties of the proposed estimator and indicate that its variance could be large. We address the variance problem by using a variance reduction estimator.
- We conduct experiments using semi-synthetic and real-world datasets. Compared to the baselines, the proposed method largely improves the ranking metrics. In addition, it improves the prediction for the less popular items, which comprise a major proportion of the total items.

2 NOTATION AND PROBLEM FORMULATION

In this section, we introduce the basic notation and formulate the implicit feedback recommendation.

2.1 Notation

Let $u \in \mathcal{U}$ be a user ($|\mathcal{U}| = m$), $i \in \mathcal{I}$ be an item ($|\mathcal{I}| = n$), and $\mathcal{D} = \mathcal{U} \times \mathcal{I}$ be the set of all user–item pairs. $Y \in \{0, 1\}^{m \times n}$ is a click matrix, and each entry $Y_{u,i}$ is a Bernoulli random variable representing a click between the user u and item i. If the feedback

of (u, i) is observed, then $Y_{u,i} = 1$; otherwise, $Y_{u,i} = 0$. In implicit feedback recommendation, $Y_{u,i} = 1$ indicates a positive feedback, but $Y_{u,i} = 0$ is either a negative or unlabeled positive feedback. To precisely formulate this implicit feedback setting, we introduce two other matrices. $R \in \{0, 1\}^{m \times n}$ is a relevance matrix, and each entry $R_{u,i}$ is a Bernoulli random variable representing the relevance of user u and item i. $R_{u,i} = 1$ means that u and i are relevant; however, $R_{u,i} = 0$ means that u and i are irrelevant. The other matrix is called the exposure matrix, denoted as $O \in \{0, 1\}^{m \times n}$. Each entry of this matrix $O_{u,i}$ is a random variable representing whether user u has been exposed to item i. It should be noted that in implicit feedback recommendation, both the relevance and exposure random variables are **unobserved**, and only click random variables are observable.

In this work, we make the following two assumptions for all user–item pairs.

$$Y_{u,i} = O_{u,i} \cdot R_{u,i} \tag{1}$$

$$P(Y_{u,i} = 1) = \theta_{u,i} \cdot \gamma_{u,i} \tag{2}$$

We define $\theta_{u,i} = P(O_{u,i} = 1)$ and $\gamma_{u,i} = P(R_{u,i} = 1)$ as the exposure and relevance parameters, respectively and assume that both parameters are non-zero for all user–item pairs.

Eq. (1) assumes that item i is clicked by user u if i has been exposed to u and they are relevant (i.e., $Y_{u,i} = 1 \Leftrightarrow O_{u,i} = 1 = 1$ & $R_{u,i} = 1$). The position-based model in unbiased learning-to-rank [12, 27] makes the same assumption in the click generative process. This assumption precisely represents the implicit feedback setting, in which a click does not always signify relevance.

In contrast, Eq. (2) assumes that the click probability is decomposed into the exposure probability ($\theta_{u,i}$) and relevance level ($\gamma_{u,i}$)[1]. Given this assumption, the exposure probability $\theta_{u,i}$ can take different values among user–item pairs, and it can model the MNAR setting in which the click probability and relevance level are not proportional.

2.2 True Performance Metric

This section describes the objective of this study. To evaluate the recommendation policy with implicit feedback data, the top-N recommendation metrics such as the mean average precision (MAP) and recall are commonly used [28]. In general, these metrics can be defined in the following manner [28].

$$\mathcal{R}_{click}\left(\widehat{Z}\right) = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}} \underbrace{P\left(Y_{u,i} = 1\right)}_{\text{click probability}} \cdot c\left(\widehat{Z}_{u,i}\right) \tag{3}$$

where $\widehat{Z} = \{\widehat{Z}_{u,i}\}_{(u,i) \in \mathcal{D}}$ is the predicted ranking of item i for user u, and the function $c(\cdot)$ characterizes a top-N scoring metric. For example, for DCG@K, the function is defined as $c(\widehat{Z}_{u,i}) = \mathbb{I}\{\widehat{Z}_{u,i} \leq K\}/\log(\widehat{Z}_{u,i} + 1)$.

The problem is that click ($Y_{u,i}$) does not directly signify relevance ($R_{u,i}$), and thus, the top-N recommendation metrics defined as in Eq. (3) are not appropriate to measure the improvement in the user experience. Therefore, we use the following top-N recommendation

---

[1]This assumption is similar to the unconfoundedness assumption in causal inference [8, 20, 21] and can also be represented as $O \perp R | u, i$.

metric defined using the relevance level as the performance metric.

$$\mathcal{R}_{rel}\left(\widehat{Z}\right) = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \sum_{i \in I} \underbrace{P\left(R_{u,i} = 1\right)}_{\text{relevance level}} \cdot c\left(\widehat{Z}_{u,i}\right) \qquad (4)$$

The focus of this study is to optimize the performance metric in Eq. (4). To achieve this goal, we follow the basic pointwise approach and aim to optimize the following loss function of interest.

$$\mathcal{L}_{ideal}\left(\widehat{R}\right) = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[\gamma_{u,i} \delta^{(1)}\left(\widehat{R}_{u,i}\right) + \left(1 - \gamma_{u,i}\right) \delta^{(0)}\left(\widehat{R}_{u,i}\right) \right]$$

$$(5)$$

where $\widehat{R}$ is a prediction matrix, and $\delta^{(R)}, (R \in \{0, 1\})$ denotes the local loss for user–item pair (u, i). For example, when $\delta^{(R)}(\widehat{R}) = -(R \log(\widehat{R}) + (1 - R) \log(1 - \widehat{R}))$, then $\mathcal{L}_{ideal}(\widehat{R})$ is called the log loss. In the following sections, we denote $\delta^{(R)}(\widehat{R}_{u,i})$ simply as $\delta_{u,i}^{(R)}$.

A prediction matrix $\widehat{R}$ minimizing the ideal loss defined using the relevance level in Eq. (5) is expected to lead to the desired values of the top-N recommendation metric in Eq. (4).

3 ANALYSIS ON EXISTING BASELINES

In this section, we describe the standard baselines and theoretically analyze the loss functions used in these methods. In particular, we demonstrate that the loss functions are biased against the ideal loss.

3.1 Weighted Matrix Factorization

WMF is the most basic latent factor model for implicit feedback recommendation [7]. The WMF uses a simple heuristic in which all the clicked data are equally upweighted compared with the unclicked data [7, 17]. This model optimizes the following estimator for the ideal loss:

$$\widehat{\mathcal{L}}_{WMF}\left(\widehat{R}\right) = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[cY_{u,i} \delta_{u,i}^{(1)} + \left(1 - Y_{u,i}\right) \delta_{u,i}^{(0)} \right] \quad (6)$$

where $c \geq 1$ is a hyperparameter representing the confidence of the clicked data relative to that of the unclicked data. When no side information is available, c is uniform among all the clicked data. In the following proposition, we demonstrate that the estimator used in the WMF is biased.

PROPOSITION 3.1. *(Bias of WMF estimator) The bias of the estimator used in WMF is represented as follows.*

$$\left| \mathbb{E}\left[\widehat{\mathcal{L}}_{WMF}\left(\widehat{R}\right) \right] - \mathcal{L}_{ideal}\left(\widehat{R}\right) \right|$$

$$= \frac{1}{|\mathcal{D}|} \left| \sum_{(u,i) \in \mathcal{D}} \left[(c\theta_{u,i} - 1)\gamma_{u,i}\delta^{(1)} + \gamma_{u,i}(1 - \theta_{u,i})\delta_{u,i}^{(0)} \right] \right|$$

PROOF.

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{WMF}\left(\widehat{R}\right) \right] = \mathbb{E}\left[\frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[cY_{u,i} \delta_{u,i}^{(1)} + \left(1 - Y_{u,i}\right) \delta_{u,i}^{(0)} \right] \right]$$

$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[c\mathbb{E}\left[Y_{u,i}\right] \delta_{u,i}^{(1)} + \left(1 - \mathbb{E}\left[Y_{u,i}\right]\right) \delta_{u,i}^{(0)} \right]$$

$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[c\theta_{u,i}\gamma_{u,i}\delta_{u,i}^{(1)} + \left(1 - \theta_{u,i}\gamma_{u,i}\right) \delta_{u,i}^{(0)} \right]$$

Thus,

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{WMF}\left(\widehat{R}\right) \right] - \mathcal{L}_{ideal}\left(\widehat{R}\right)$$

$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[c\theta_{u,i}\gamma_{u,i}\delta_{u,i}^{(1)} + \left(1 - \theta_{u,i}\gamma_{u,i}\right) \delta_{u,i}^{(0)} \right]$$

$$- \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[\gamma_{u,i}\delta_{u,i}^{(1)} + \left(1 - \gamma_{u,i}\right) \delta_{u,i}^{(0)} \right]$$

$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \left[(c\theta_{u,i} - 1)\gamma_{u,i}\delta^{(1)} + \gamma_{u,i}(1 - \theta_{u,i})\delta_{u,i}^{(0)} \right]$$

$$\square$$

For $\widehat{\mathcal{L}}_{WMF}\left(\widehat{R}\right)$ to be theoretically unbiased, $c\theta_{u,i} - 1 = 0 \Rightarrow \theta_{u,i} = 1/c$, and $1 - \theta_{u,i} = 0 \Rightarrow \theta_{u,i} = 1$ need to be satisfied for all pairs from the last equation. However, $\theta_{u,i}$ can take different values among user–item pairs in our setting; thus, these conditions are not always satisfied, and the loss function of WMF does not satisfy the unbiasedness. This is because WMF does not address the positive-unlabeled problem, in which $Y = 0$ does not always signify $R = 0$. Thus, WMF is unsuitable for optimizing our metric of interest defined in Eq. (4).

3.2 Exposure Matrix Factorization

In contrast to the WMF, the ExpoMF is a prediction method considering the exposure matrix (O), and it is based on the following latent probabilistic model [17]:

$$U \sim \mathcal{N}\left(0, \lambda_U^{-1}I_K\right), \quad V \sim \mathcal{N}\left(0, \lambda_V^{-1}I_K\right)$$

$$O_{u,i} \sim Bern\left(\mu_i\right), \quad Y_{u,i} \mid O_{u,i} = 1 \sim \mathcal{N}\left(U_u^{\top}V_i, \lambda_y^{-1}\right)$$

where λ_U, λ_V, and λ_y are the hyperparameters denoting the inverse variance of the prior distributions, and $Bern(\cdot)$ is the Bernoulli distribution. Following the probabilistic model defined above, $P(Y_{u,i} = 0 \mid O_{u,i} = 0) = 1$, which is consistent with our assumptions.

The log-likelihood to derive the parameters (i.e., μ_i, U_u, and V_i) can be written as[2]

$$\log\left(P\left(o_{u,i}, y_{u,i} \mid \mu_{u,i}, U_u, V_i, \lambda_y^{-1}\right)\right)$$

$$= \log Bern\left(o_{u,i} \mid \mu_{u,i}\right) + \underbrace{o_{u,i} \cdot \log \mathcal{N}\left(y_{u,i} \mid U_u^{\top}V_i, \lambda_y^{-1}\right)}_{(a)}$$

$$(7)$$

In Eq. (7), the loss to derive the user and item matrices (U, V) is (a), and it can be defined in the following manner:

$$(a) = \begin{cases} \delta^{(1)}\left(U_u^{\top}V_i\right) & (y_{u,i} = 1, o_{u,i} = 1) \\ \delta^{(0)}\left(U_u^{\top}V_i\right) & (y_{u,i} = 0, o_{u,i} = 1) \\ 0 & \text{otherwise } (o_{u,i} = 0) \end{cases}$$

Therefore, the loss function of the ExpoMF is designed to consider the local loss of user–item pairs when the item has been exposed to the user (i.e., $o_{u,i} = 1$). This is because if an item has been exposed, the click variable is considered to represent the relevance information (i.e., $O_{u,i} = 1 \Rightarrow Y_{u,i} = R_{u,i}$).

---

[2]Equation (2) of [17]. The third term is always 0 and is thus omitted here.

However, the realizations of exposure variables $\{o_{u,i}\}$ are unobserved. Therefore, ExpoMF uses an EM-like iterative algorithm to derive the user–item matrices. In the E-step, the posterior exposure probability is estimated, and in the M-step, the model parameters are updated to optimize the log-likelihood[3]. Given the true posterior exposure probabilities, the M-step of ExpoMF is interpreted to minimize the following weighted loss function.

$$\widehat{\mathcal{L}}_{ExpoMF}\left(\widehat{R}\right) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \theta'_{u,i} \left[Y_{u,i}\delta_{u,i}^{(1)} + \left(1 - Y_{u,i}\right)\delta_{u,i}^{(0)} \right] \quad (8)$$

where $\theta'_{u,i} = \mathbb{E}\left[O_{u,i} \mid Y_{u,i}\right]$ is the posterior exposure probability. For example, $\mathbb{E}\left[O_{u,i} \mid Y_{u,i} = 1\right] = 1$ because $Y_{u,i} = 1 \Rightarrow O_{u,i} = 1$. This posterior probability represents the confidence of the amount of relevance information contained in click indicator $Y_{u,i}$.

ExpoMF utilizes the posterior probability of the exposure to reweight data and improves the WMF. However, the following proposition suggests that the loss function in Eq. (8) optimized in the M-step of the ExpoMF is also biased against the ideal loss.

PROPOSITION 3.2. *(Bias of ExpoMF) When the posterior exposure probabilities are given, then, the bias of the estimator used in ExpoMF is represented as follows.*

$$\left| \mathbb{E}\left[\widehat{\mathcal{L}}_{ExpoMF}\left(\widehat{R}\right)\right] - \mathcal{L}_{ideal}\left(\widehat{R}\right) \right|$$
$$= \frac{1}{|\mathcal{D}|} \Bigg| \sum_{(u,i)\in\mathcal{D}} \gamma_{u,i}(\theta'_{u,i}\theta_{u,i} - 1)\delta_{u,i}^{(1)}$$
$$+ \left\{\theta'_{u,i} - 1 - \gamma_{u,i}(\theta_{u,i}\theta'_{u,i} - 1)\right\}\delta_{u,i}^{(0)} \Bigg|$$

PROOF.
$$\mathbb{E}\left[\widehat{\mathcal{L}}_{ExpoMF}\left(\widehat{R}\right)\right]$$
$$= \mathbb{E}\left[\frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \theta'_{u,i}\left[Y_{u,i}\delta_{u,i}^{(1)} + \left(1 - Y_{u,i}\right)\delta_{u,i}^{(0)}\right]\right]$$
$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \theta'_{u,i}\left[\mathbb{E}\left[Y_{u,i}\right]\delta_{u,i}^{(1)} + \left(1 - \mathbb{E}\left[Y_{u,i}\right]\right)\delta_{u,i}^{(0)}\right]$$
$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[\theta'_{u,i}\theta_{u,i}\gamma_{u,i}\delta_{u,i}^{(1)} + \theta'_{u,i}\left(1 - \theta_{u,i}\gamma_{u,i}\right)\delta_{u,i}^{(0)}\right]$$

Thus we obtain:
$$\mathbb{E}\left[\widehat{\mathcal{L}}_{ExpoMF}\left(\widehat{R}\right)\right] - \mathcal{L}_{ideal}\left(\widehat{R}\right)$$
$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[\theta'_{u,i}\theta_{u,i}\gamma_{u,i}\delta_{u,i}^{(1)} + \theta'_{u,i}\left(1 - \theta_{u,i}\gamma_{u,i}\right)\delta_{u,i}^{(0)}\right]$$
$$- \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[\gamma_{u,i}\delta_{u,i}^{(1)} + \left(1 - \gamma_{u,i}\right)\delta_{u,i}^{(0)}\right]$$
$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[\gamma_{u,i}(\theta'_{u,i}\theta_{u,i}-1)\delta_{u,i}^{(1)} + \left\{\theta'_{u,i}-1-\gamma_{u,i}(\theta_{u,i}\theta'_{u,i}-1)\right\}\delta_{u,i}^{(0)}\right]$$

□

For $\widehat{\mathcal{L}}_{ExpoMF}\left(\widehat{R}\right)$ to be theoretically unbiased, $\theta_{u,i}\theta'_{u,i} - 1 = 0 \Rightarrow \theta_{u,i}\theta'_{u,i} = 1$, and $\theta'_{u,i} - 1 = 0 \Rightarrow \theta'_{u,i} = 1$ need to be satisfied for all pairs from the last equation. However, $\theta_{u,i}$ and $\theta'_{u,j}$ can take different values among user–item pairs in our setting, and thus,

[3]The detailed learning procedure of ExpoMF is described in Section 3.3 of [17].

these conditions are not always satisfied, and the loss function of the ExpoMF does not satisfy the unbiasedness. This is because the ExpoMF upweights the local loss of data that is frequently observed in the training data (i.e., data having high exposure probability). This upweighting leads to the poor prediction accuracy for data having low exposure probability such as tail items, and thus, it fails to achieve the goal of recommender systems. Therefore, dealing with the MNAR problem as well as the unlabeled nature of implicit feedback is essential to derive a desirable estimator.

4 PROPOSED METHOD

In this section, we propose an unbiased estimator for the ideal loss to overcome the limitations described in the previous section. The proposed unbiased estimator is an extension of the Inverse Propensity Score (IPS) in causal inference [8, 20, 21] and an estimator in positive-unlabeled learning [2, 15]. In our theoretical analysis, we prove that our estimator is valid in the implicit recommendation setting. Moreover, we analyze the variance of the unbiased estimator and indicate that it can suffer from a high variance. Finally, we provide and analyze a technique to address the variance problem.

4.1 Proposed Estimator

First, we define the propensity score to deal with the MNAR problem of implicit feedback as follows.

Definition 4.1. (Propensity Score) The propensity score of user–item pair (u, i) is

$$\theta_{u,i} = P(O_{u,i} = 1) = P\left(Y_{u,i} = 1 \mid R_{u,i} = 1\right)$$

Next, our proposed estimator is defined using the propensity score.

Definition 4.2. (Unbiased Estimator) When the propensity scores are given, the unbiased estimator is defined as

$$\widehat{\mathcal{L}}_{unbiased}\left(\widehat{R}\right) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[\frac{Y_{u,i}}{\theta_{u,i}}\delta_{u,i}^{(1)} + \left(1 - \frac{Y_{u,i}}{\theta_{u,i}}\right)\delta_{u,i}^{(0)}\right] \quad (9)$$

Note that this unbiased estimator is not the standard form of the IPS estimator in the explicit feedback setting [23] because one cannot use the exposure indicator in the implicit feedback setting. The proposed estimator can also be represented in the following form.

$$\frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[Y_{u,i}\left(\frac{1}{\theta_{u,i}}\delta_{u,i}^{(1)} + \left(1 - \frac{1}{\theta_{u,i}}\delta_{u,i}^{(0)}\right)\right) + (1 - Y_{u,i})\delta_{u,i}^{(0)}\right]$$

Thus, the unbiased estimator applies both positive loss ($\delta^{(1)}$) and negative loss ($\delta^{(0)}$) for the clicked data ($Y_{u,i} = 1$). In the following proposition, we show that our unbiased estimator is unbiased against the ideal loss.

PROPOSITION 4.3. *The unbiased estimator defined in Eq. (9) is unbiased against the ideal loss in Eq. (5).*

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{unbiased}\left(\widehat{R}\right)\right] = \mathcal{L}_{ideal}\left(\widehat{R}\right)$$

PROOF.

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{unbiased}\left(\widehat{R}\right)\right]$$

$$= \mathbb{E}\left[\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left[\frac{Y_{u,i}}{\theta_{u,i}}\delta_{u,i}^{(1)} + \left(1 - \frac{Y_{u,i}}{\theta_{u,i}}\right)\delta_{u,i}^{(0)}\right]\right]$$

$$= \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left[\frac{\mathbb{E}[Y_{u,i}]}{\theta_{u,i}}\delta_{u,i}^{(1)} + \left(1 - \frac{\mathbb{E}[Y_{u,i}]}{\theta_{u,i}}\right)\delta_{u,i}^{(0)}\right]$$

$$= \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left[\gamma_{u,i}\delta_{u,i}^{(1)} + \left(1 - \gamma_{u,i}\right)\delta_{u,i}^{(0)}\right] = \mathcal{L}_{ideal}\left(\widehat{R}\right)$$

□

Proposition 4.3 validates that de-biasing using propensity reweighting is valid in MNAR implicit recommendation. However, propensity-based estimators often suffer from a high variance [4, 22, 25]. Moreover, variance analysis was not conducted in a previous study on positive-unlabeled learning [2, 15]. Thus, in the following theorem, we provide the variance of the unbiased estimator.

THEOREM 4.4. *(Variance of the unbiased estimator) Given sets of independent random variables* $\{Y_{u,i}\}$, $\{O_{u,i}\}$, *and* $\{R_{u,i}\}$, *propensity scores* $\{\theta_{u,i}\}$, *and predicted matrix* $\widehat{R}$, *the variance of the unbiased estimator is*

$$\mathbb{V}\left(\widehat{\mathcal{L}}_{unbiased}\left(\widehat{R}\right)\right) = \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\gamma_{u,i}\left(\frac{1}{\theta_{u,i}} - \gamma_{u,i}\right)\left(\delta_{u,i}^{(1)} - \delta_{u,i}^{(0)}\right)^2$$

PROOF. First, we define $X_{u,i} = \frac{Y_{u,i}}{\theta_{u,i}}\delta_{u,i}^{(1)} + \left(1 - \frac{Y_{u,i}}{\theta_{u,i}}\right)\delta_{u,i}^{(0)}$. Subsequently, $\mathbb{V}\left(X_{u,i}\right)$ can be written as

$$\mathbb{V}\left(X_{u,i}\right) = \underbrace{\mathbb{E}\left[(X_{u,i})^2\right]}_{(b)} - \underbrace{\left(\mathbb{E}\left[X_{u,i}\right]\right)^2}_{(c)}$$

By Proposition 4.3,

$$(c) = (\gamma_{u,i}\delta_{u,i}^{(1)} + (1 - \gamma_{u,i})\delta_{u,i}^{(0)})^2$$
$$= \gamma_{u,i}^2(\delta_{u,i}^{(1)})^2 + 2\gamma_{u,i}(1 - \gamma_{u,i})\delta_{u,i}^{(1)}\delta_{u,i}^{(0)} + (1 - \gamma_{u,i})^2(\delta_{u,i}^{(0)})^2$$

Then,

$$X_{u,i}^2 = \frac{Y_{u,i}}{\theta_{u,i}^2}(\delta_{u,i}^{(1)})^2 + 2\frac{Y_{u,i}}{\theta_{u,i}}\left(1 - \frac{Y_{u,i}}{\theta_{u,i}}\right)\delta_{u,i}^{(1)}\delta_{u,i}^{(0)} + \left(1 - \frac{Y_{u,i}}{\theta_{u,i}}\right)^2(\delta_{u,i}^{(1)})^2$$

$$= \frac{Y_{u,i}}{\theta_{u,i}^2}(\delta_{u,i}^{(1)})^2 + 2\left(\frac{Y_{u,i}}{\theta_{u,i}} - \frac{Y_{u,i}}{\theta_{u,i}^2}\right)\delta_{u,i}^{(1)}\delta_{u,i}^{(0)}$$

$$+ \left(1 - 2\frac{Y_{u,i}}{\theta_{u,i}} + \frac{Y_{u,i}}{\theta_{u,i}^2}\right)(\delta_{u,i}^{(0)})^2$$

where $Y_{u,i}^2 = Y_{u,i}$ and $(1 - Y_{u,i})^2 = (1 - Y_{u,i})$. Next, (b) is calculated as

$$(b) = \frac{\gamma_{u,i}}{\theta_{u,i}}(\delta_{u,i}^{(1)})^2 + 2\left(\gamma - \frac{\gamma}{\theta_{u,i}}\right)\delta_{u,i}^{(1)}\delta_{u,i}^{(0)} + \left(1 - 2\gamma + \frac{\gamma}{\theta_{u,i}}\right)(\delta_{u,i}^{(0)})^2$$

Therefore,

$$\mathbb{V}\left(X_{u,i}\right) = (b) - (c) = \gamma_{u,i}\left(\frac{1}{\theta_{u,i}} - \gamma_{u,i}\right)\left(\delta_{u,i}^{(1)} - \delta_{u,i}^{(0)}\right)^2$$

From the assumptions, $\{X_{u,i}\}$ is a set of independent random variables. Thus,

$$\mathbb{V}\left(\widehat{\mathcal{L}}_{unbiased}\left(\widehat{R}\right)\right) = \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\mathbb{V}\left(X_{u,i}\right)$$

$$= \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\gamma_{u,i}\left(\frac{1}{\theta_{u,i}} - \gamma_{u,i}\right)\left(\delta_{u,i}^{(1)} - \delta_{u,i}^{(0)}\right)^2$$

□

As shown in Theorem 4.4, the variance of the unbiased estimator depends on the inverse of the propensity score. The propensity score is defined as the exposure probability, and thus it can be small, especially for tail items [28]. Therefore, the variance of the proposed estimator can be large in the implicit recommendation setting.

4.2 Variance Reduction Technique

In the theoretical analysis, we demonstrated that the proposed unbiased estimator is unbiased against the ideal loss. However, Theorem 4.4 suggests that the unbiased estimator can suffer from high variance. In this subsection, we apply the propensity clipping technique [4, 6, 24] to our estimator and analyze the bias-variance trade-off of an estimator with clipping.

First, we introduce the propensity score with the clipping technique.

Definition 4.5. (Clipped Propensity Score) M is a positive constant that takes values in the interval $[0, 1]$. Then, the clipped propensity score is defined as:

$$\bar{\theta}_{u,i} = \max\{\theta_{u,i}, M\} \tag{10}$$

The clipped propensity score clips a small value of θ by M. We further define the clipped estimator for the ideal loss by using the clipped propensity score.

Definition 4.6. (Clipped Estimator) The clipped estimator for the ideal loss is defined as

$$\widehat{\mathcal{L}}_{clipped}\left(\widehat{R}\right) = \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left[\frac{Y_{u,i}}{\bar{\theta}_{u,i}}\delta_{u,i}^{(1)} + \left(1 - \frac{Y_{u,i}}{\bar{\theta}_{u,i}}\right)\delta_{u,i}^{(0)}\right] \tag{11}$$

As $M \to 1$, the clipped estimator approaches the naive estimator (estimator of WMF with $c = 1$); in contrast, as $M \to 0$, it approaches the unbiased estimator. Thus, the clipped estimator is a general form of the two estimators. By definition, the inverse of the clipped propensity score is always smaller than that of the propensity score. Thus, using the clipped estimator reduces the effect of the variance problem of the unbiased estimator. However, it introduces some bias because the clipped propensity score is not always equal to the true propensity. We provide the bias and variance of the clipped estimator as follows.

PROPOSITION 4.7. *(Bias of the clipped estimator) Given a constant* $M \in [0, 1]$, *the bias of the clipped estimator is*

$$\left|\mathbb{E}\left[\widehat{\mathcal{L}}_{clipped}\left(\widehat{R}\right)\right] - \mathcal{L}_{ideal}\left(\widehat{R}\right)\right|$$

$$= \left|\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\mathbb{I}\{\theta_{u,i} \leq M\}\gamma_{u,i}\left(\frac{\theta_{u,i}}{M} - 1\right)\left(\delta_{u,i}^{(1)} - \delta_{u,i}^{(0)}\right)\right|$$

PROOF. First, we define $X_{u,i} = \frac{Y_{u,i}}{\theta_{u,i}} \delta_{u,i}^{(1)} + \left(1 - \frac{Y_{u,i}}{\theta_{u,i}}\right) \delta_{u,i}^{(0)}$, then,

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{clipped}\left(\widehat{R}\right)\right]$$
$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \mathbb{I}\{\theta_{u,i} > M\}\mathbb{E}[X_{u,i}]$$
$$+ \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \mathbb{I}\{\theta_{u,i} \le M\} \underbrace{\mathbb{E}\left[\frac{Y_{u,i}}{M}\delta_{u,i}^{(1)} + \left(1 - \frac{Y_{u,i}}{M}\right)\delta_{u,i}^{(0)}\right]}_{(d)}$$

By using $\mathbb{E}[Y_{u,i}] = \theta_{u,i}\gamma_{u,i}$, (d) is calculated as

$$(d) = \frac{\theta_{u,i}\gamma_{u,i}}{M}\delta_{u,i}^{(1)} + \left(1 - \frac{\theta_{u,i}\gamma_{u,i}}{M}\right)\delta_{u,i}^{(0)}$$

By Proposition 4.3, $\mathbb{E}\left[X_{u,i}\right] = \gamma_{u,i}\delta_{u,i}^{(1)} + (1 - \gamma_{u,i})\delta_{u,i}^{(0)}$. Then, the ideal loss can also be represented as

$$\mathcal{L}_{ideal}\left(\widehat{R}\right) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \mathbb{I}\{\theta_{u,i} > M\}\mathbb{E}[X_{u,i}] + \mathbb{I}\{\theta_{u,i} \le M\}\mathbb{E}[X_{u,i}]$$

Therefore,

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{clipped}\left(\widehat{R}\right)\right] - \mathcal{L}_{ideal}\left(\widehat{R}\right)$$
$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \mathbb{I}\{\theta_{u,i} \le M\}\left((d) - \mathbb{E}[X_{u,i}]\right)$$
$$= \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[\mathbb{I}\{\theta_{u,i} \le M\}\gamma_{u,i}\left(\frac{\theta_{u,i}}{M} - 1\right)\left(\delta_{u,i}^{(1)} - \delta_{u,i}^{(0)}\right)\right]$$

□

COROLLARY 4.8. *(Variance of the clipped estimator) Under the same condition as in Theorem 4.4, the variance of the clipped estimator is*

$$\mathbb{V}\left(\widehat{\mathcal{L}}_{clipped}\left(\widehat{R}\right)\right) = \frac{1}{|\mathcal{D}|^2} \sum_{(u,i)\in\mathcal{D}} \gamma_{u,i}\left(\frac{1}{\bar{\theta}_{u,i}} - \gamma_{u,i}\right)\left(\delta_{u,i}^{(1)} - \delta_{u,i}^{(0)}\right)^2$$
$$\le \mathbb{V}\left(\widehat{\mathcal{L}}_{unbiased}\left(\widehat{R}\right)\right) \because \frac{1}{\bar{\theta}_{u,i}} \le \frac{1}{\theta_{u,i}}$$

PROOF. Propensity scores are not random variables. Thus, we obtain the variance by replacing $\theta_{u,i}$ in Theorem 4.4 by $\bar{\theta}_{u,i}$. □

As shown above, the clipped estimator always reduces the variance of the unbiased estimator but introduces some bias depending on the value of M. In the experimental part, we tuned the clipping constant M as a hyper-parameter using a validation set to address the variance problem.

5 SEMI-SYNTHETIC EXPERIMENT

In this section, we conduct an experiment with semi-synthetic datasets[4] and investigate the following research questions (RQs).

RQ1. How does the level of exposure bias affect the performance of the MF-Naive and MF-Unbiased models?

RQ2. Does optimizing the ideal pointwise loss in Eq. (5) yield a better value of the ranking metric in Eq. (4)?

---

[4]The implementation for the semi-synthetic experiment is available at https://github.com/usaito/unbiased-implicit-rec.

5.1 Experimental Setup

5.1.1 Dataset. We used the MovieLens (ML) 100K dataset[5]. This dataset contains five-star movie ratings collected from a movie recommendation service, and the ratings are MNAR. To facilitate ground-truth evaluation against a fully known relevance and exposure parameters, we created these parameters as follows.

1. Using rating-based matrix factorization [14], we found an approximation of the true ratings as

$$R_{u,i} \approx \widehat{R}_{u,i}, \forall(u,i) \in \mathcal{D}$$

where $\widehat{R}_{u,i} \in [1, 5]$.

2. Using logistic matrix factorization [13], we found an approximation of the true observations as

$$O_{u,i} \approx \widehat{O}_{u,i}, \forall(u,i) \in \mathcal{D}$$

where $O_{u,i}$ is a binary variable representing whether the rating of (u, i) is observed. If the rating of (u, i) is observed, $O_{u,i} = 1$; otherwise, $O_{u,i} = 0$. Thus, $\widehat{O}_{u,i} \in (0, 1)$ is the estimated probability of observing the rating of (u, i).

3. We generate the ground-truth relevance and exposure parameters as follows.

$$\gamma_{u,i} = \sigma(\widehat{R}_{u,i} - \epsilon), \ \theta_{u,i} = (\widehat{O}_{u,i})^p, \ \forall(u,i) \in \mathcal{D}$$

where $\sigma(\cdot)$ is the sigmoid function, ϵ controls the overall relevance level, and p controls the skewness of the distribution of the exposure parameter. When a large value of p is used, then a huge exposure bias is introduced. In the experiment, we set $\epsilon = 5$ and $p = 0.5, 1, 2, 3, 4$.

4. Following the probabilistic model described in Eq. (1) and Eq. (2), we generated click variables as follows.

$$O_{u,i} \sim Bern(\theta_{u,i}), R_{u,i} \sim Bern(\gamma_{u,i})$$
$$Y_{u,i} = O_{u,i} \cdot R_{u,i}, \forall(u,i) \in \mathcal{D}$$

where $Bern(\cdot)$ is the Bernoulli distribution.

Note that one can evaluate the ground-truth performance of the methods with the true relevance and exposure parameters by using semi-synthetic datasets.

5.1.2 Baselines. We used and compared the following models.

- MF-Oracle: The matrix factorization model trained using the ground truth relevance information. Thus, the performance of this method is the best achievable performance.
- MF-Naive: The matrix factorization model trained using only the observable click information. As shown in Proposition 3.1, this estimator has a bias, and this bias problem can be severe, particularly when a large value of p is used.
- MF-Unbiased: The matrix factorization model trained with the unbiased loss function defined in Eq. (9). The variance of the unbiased loss function might be large when a large value of p is used.

5.1.3 Evaluation Metrics. We used the log-loss and discounted cumulative gain (DCG) on test sets to evaluate the relevance prediction and the raking performance, respectively. Note that both evaluation metrics are calculated using the true relevance information, which is inaccessible when using a real-world dataset.

---

[5]http://grouplens.org/datasets/movielens/
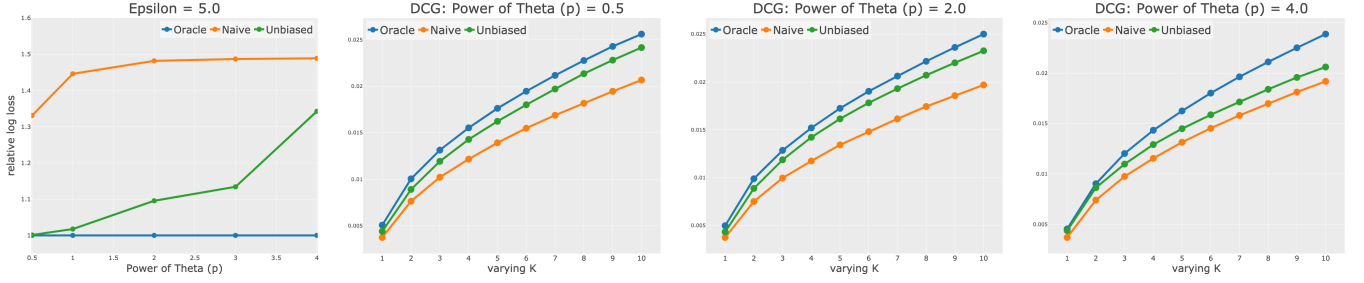
Figure 1: (Left) Log-loss on test sets of the MF-Naive and MF-Unbiased models relative to that of the MF-Oracle with different values of p (x-axis). **(Others)** DCG@K on test sets of the MF-Oracle, MF-Naive, and MF-Unbiased models with different levels of exposure bias.

5.2 Results & Discussions

5.2.1 RQ1. How does the level of exposure bias affect the performance of the MF-Naive and MF-Unbiased models? Here, we investigate the performance of the MF-Naive and MF-Unbiased models with different levels of exposure bias (different values of p). Figure 1 (left) shows the log-loss on test sets of the MF-Naive and MF-Unbiased models relative to that of the MF-Oracle model. We report the results for varying values of p (x-axis).

The figure reveals that the performance of both the MF-Naive and MF-Unbiased models declines with larger values of p. The poor performance of the MF-Naive model with severe exposure bias is due to the bias problem (stated in Proposition 3.1). On the other hand, the poor performance of the MF-Unbiased model arises from the variance problem of the propensity weighting technique (stated in Theorem 4.4). Although the benefit of using the unbiased estimator is relatively small when a severe exposure bias exists, it consistently outperforms the naive estimator in all settings. The results demonstrate the effectiveness of the unbiased estimation approach to the different levels of exposure bias.

5.2.2 RQ2. Does the optimization of the ideal pointwise loss actually lead to a better value of the ranking metric? Next, we compared the ranking performance of the MF with that of the different estimators. Figure 1 shows the DCG@K with different values of K ($K \in \{1, 2, \ldots, 10\}$). We report the results with $p = 0.5, 2.0, 4$. As shown in the figure, MF-Unbiased outperforms MF-Naive in all settings. In particular, with $p = 0.5$, the MF-Unbiased model largely outperforms the MF-Naive model and its performance is close to that of the MF-Oracle model. However, when a large exposure bias exists ($p = 4$), the benefit of using the unbiased estimator is relatively small. These results are well correlated with those for the relevance level prediction task, as discussed in Section 5.2.1. Therefore, the results validate that optimizing the ideal pointwise loss is a valid approach to improving recommendation performance with respect to the ranking metrics.

6 REAL-WORLD EXPERIMENT

In this section, we compare the prediction methods based on the proposed estimator and the existing baselines by using a standard real-world dataset. In particular, we investigate the following research question.

RQ3. How does the proposed unbiased estimator perform compared with other existing methods?
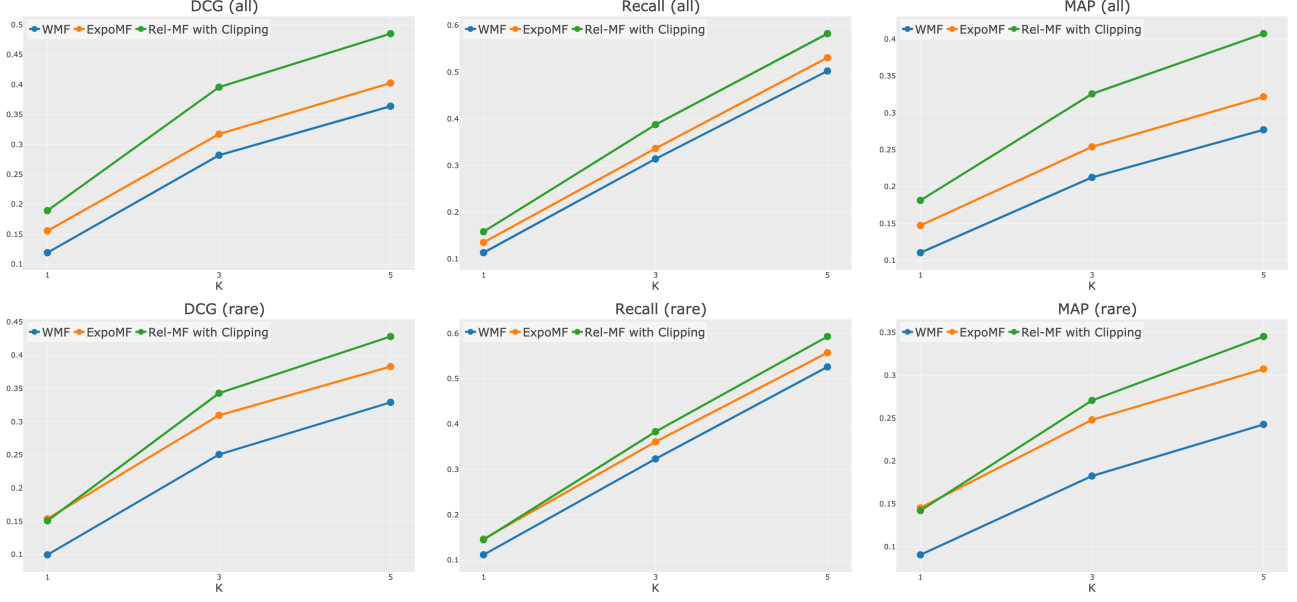
6.1 Experimental Setup

6.1.1 Dataset. We used the Yahoo! R3 dataset[6]. This is an explicit feedback dataset collected from a song recommendation service. As described in [28], it contains users' ratings for randomly selected sets of music as a test set, and thus it can be used to measure the recommenders' true performances. The dataset includes explicit feedback data; we treated items rated greater than or equal to 4 as relevant, and the others were considered irrelevant. We used this dataset because it contains the training and test sets with different item and user distributions. Moreover, it contains the explicit feedback data, thus allowing the utilization of the ground truth relevance information in the test set. Therefore, this dataset is appropriate for the evaluation of recommender in the situation where both positive-unlabeled and MNAR problems exist. To the best of our knowledge, this dataset is the only one that satisfies these properties.

6.1.2 Baselines and the proposed model. We describe the existing baselines and the proposed model used in the real-world experiment.

- Weighted Matrix Factorization (WMF) [7]: The WMF is a basic baseline model for implicit recommendation and is described in Section 3.1.
- Exposure Matrix Factorization (ExpoMF) [17]: ExpoMF is based on the latent probabilistic model using exposure variables, and it is described in Section 3.2. We used the implementation provided at https://github.com/dawenl/expo-mf.
- Relevance Matrix Factorization (Rel-MF): Rel-MF is the proposed model and is based on the same latent factor model as the MF. It updates its user–item factors by minimizing the clipped version of the proposed estimator in Eq. (11). We estimated the propensity score by using the following relative item popularity.

$$\widehat{\theta}_{*,i} = \left(\frac{\sum_{u \in \mathcal{U}} Y_{u,i}}{\max_{i \in \mathcal{I}} \sum_{u \in \mathcal{U}} Y_{u,i}} \right)^{\eta} \tag{12}$$

---

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

Figure 2: Ranking metrics on the Yahoo! R3 data. Rel-MF consistently outperforms the other baselines in almost all situations.

In our assumption, the click probability depends on both the exposure probability and relevance level. Thus, we estimated the propensity score using the relative click probability with a parameter $\eta \leq 1$ and we set $\eta = 0.5$ in this experiment.

6.1.3 Evaluation Metrics. We used the DCG, Recall, and MAP to evaluate the ranking performance. We report the results with varying values of $K \in \{1, 3, 5\}$ for all the metrics. These ranking metrics are defined as follows.

$$DCG@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}_u^{test}:R_{u,i}=1} \frac{\mathbb{I}\left\{\widehat{Z}_{u,i} \leq K\right\}}{\log(\widehat{Z}_{u,i} + 1)}$$

$$Recall@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}_u^{test}:R_{u,i}=1} \frac{\mathbb{I}\left\{\widehat{Z}_{u,i} \leq K\right\}}{\sum_{i \in \mathcal{I}_u^{test}} R_{u,i}}$$

$$MAP@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}_u^{test}:R_{u,i}=1} \sum_{k=1}^{K} \frac{\mathbb{I}\left\{\widehat{Z}_{u,i} \leq k\right\}}{k}$$

where $\mathcal{I}_u^{test}$ is a set of items rated by user u in the test set and $R_{u,i}$ takes 1 if the user rated greater than or equal to 4.

6.1.4 Train/Test splitting and hyperparameter tuning criteria. The original Yahoo! R3 dataset is divided into the training and test sets. We randomly sampled 10% of positive feedback in the training set as the validation set. Using the validation set, we tuned the dimensions of user–item latent factors within the range of $\{30, 40, \ldots, 200\}$. For the WMF, we tuned the weighting constant in its loss function within the range of $[1, 100]$. In addition, for the Rel-MF with clipping, the clipping hyperparameter M in Eq. (11) was tuned within the range of $[0.01, 0.1]$. For all the baselines and the proposed method, the best set of the hyperparameters was determined using the *Optuna* software [1] with a TPE sampler. For WMF and

Rel-MF, we used the squared-loss as $\delta^{(1)}$ and $\delta^{(0)}$ because it is used in the implementation of ExpoMF. Note that the item distributions between the validation and test data are different. Thus, we used the self-normalized inverse propensity score (SNIPS) estimator [28] of the DCG@5 to tune the hyperparameters. The SNIPS estimator is defined as follows.

$$\widehat{\mathcal{R}}_{SNIPS}\left(\widehat{Z}_{u,i}\right) = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{\sum_{i \in \mathcal{I}_u^{val}} \frac{Y_{u,i}}{\widehat{\theta}_{*,i}}} \sum_{i \in \mathcal{I}_u^{val}} \frac{Y_{u,i}}{\widehat{\theta}_{*,i}} \cdot c\left(\widehat{Z}_{u,i}\right)$$

where $\mathcal{I}_u^{val}$ is a set of items in the validation set for user u[7]. We used the relative item popularity in Eq. (12) as the propensity score of SNIPS estimator.

6.2 Results & Discussions

Here, we present the experimental results in detail.

6.2.1 RQ3. How does the proposed unbiased estimator perform compared with other existing methods? Figure 2 shows the performance of the methods, corresponding to rare and all items. We treated 500 items (out of 1,000) with small relative item popularity in Eq. (12) in the training set as rare. As shown in the figure, the Rel-MF significantly outperformed the other baselines on almost all the metrics, for example, improving the DCG@5 by 20.4%, Recall@5 by 9.6%, and MAP@5 by 26.6% over the ExpoMF. Furthermore, for rare items, the proposed method improved the DCG@5 by 11.8%, Recall@5 by 6.4%, and MAP@5 by 11.2% over the best baseline. However, for all metrics on rare items, the Rel-MF was outperformed by ExpoMF for $K = 1$. The propensity estimation might be the reason for this. We estimated the user independent propensity score by Eq. (12). However, in real-world recommender systems, the user activeness can be diverse, and different users can have

---

[7]We randomly sampled 100 unlabeled items for each user to construct $\mathcal{I}_u^{val}$.

different propensity scores for the same item. Thus, considering only the item dependent propensity score might be too simple.

Overall, it was observed that the ExpoMF outperformed the WMF in all settings, which is consistent with the previous experiments [17, 26]. Moreover, Rel-MF outperformed the other baseline methods in most cases. In addition, the proposed method improved the ranking metrics for less-frequently observed items in the training sets (rare items). This is because the ExpoMF downweights the prediction losses on the items having a low exposure probability; in contrast, the proposed method utilizes the theoretically principal estimation technique and solves the MNAR problem by ensuring prediction accuracy on rare items. These results suggest that the proposed method is the most suitable method for optimizing the metric of interest defined in Eq. (4) from biased implicit feedback.

7 CONCLUSION

In this study, we first defined the ideal loss function for maximizing the relevance to optimize the user experience. Subsequently, we demonstrated that the loss functions of WMF and ExpoMF are biased toward the ideal loss. Furthermore, we proposed an unbiased estimator for the ideal loss inspired by the estimation method for causal inference and positive-unlabeled learning. We also analyzed the variance of the unbiased estimator and introduced a clipped estimator, which, by introducing a small bias, could reduce the variance and achieve better performance as a result of a better bias-variance trade-off. In the experiments, the proposed method significantly outperformed the existing methods with respect to the relevance maximization. In particular, the proposed method outperformed these methods for items with low exposure probability, and this finding empirically suggests that the proposed approach can suitably maximize the user experience.

A possible next step would be the development of a sophisticated method to estimate the exposure probabilities. The proposed unbiased estimator relies on the true propensity scores for its unbiasedness. Thus, a better estimation of the exposure probabilities could lead to better prediction performances owing to the improved estimation of the loss of interest. Moreover, pairwise algorithms that address the two challenging problems of implicit feedback have not yet been proposed. Therefore, the extension of the unbiased estimator to the pairwise algorithm is another interesting theme.

REFERENCES

[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*. ACM, 2623–2631.

[2] Jessa Bekker and Jesse Davis. 2018. Beyond the selected completely at random assumption for learning from positive and unlabeled data. *arXiv preprint arXiv:1809.03207* (2018).

[3] Stephen Bonner and Flavian Vasile. 2018. Causal Embeddings for Recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18)*. ACM, New York, NY, USA, 104–112. https://doi.org/10.1145/3240323.3240360

[4] Minmin Chen, Alex Beutel, Paul Covington, Sagar Jain, Francois Belletti, and Ed H Chi. 2019. Top-k off-policy correction for a REINFORCE recommender system. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM, 456–464.

[5] 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.

[6] 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.

[7] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining*. Ieee, 263–272.

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

[9] Zanker M Jannach D., Lerche L. 2018. *Recommending Based on Implicit Feedback*. Springer.

[10] Elin Rønby Pedersen Jiahui Liu, Peter Dolan. 2010. Personalized news recommendation based on click behavior.. In *Proc. of 14th Int. Conf. on Intelligent User Interfaces (IUI)*. ACM, 31–40.

[11] Thorsten Joachims and Adith Swaminathan. 2016. Counterfactual evaluation and learning for search, recommendation and ad placement. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 1199–1201.

[12] 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.

[13] Christopher C Johnson. 2014. Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems* 27 (2014).

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

[15] Xiao-Li Li and Bing Liu. 2005. Learning from positive and unlabeled examples with different data distributions. In *European Conference on Machine Learning*. Springer, 218–229.

[16] Dawen Liang, Jaan Altosaar, Laurent Charlin, and David M Blei. 2016. Factorization meets the item embedding: Regularizing matrix factorization with item co-occurrence. In *Proceedings of the 10th ACM conference on recommender systems*. ACM, 59–66.

[17] 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.

[18] Dugang Liu, Chen Lin, Zhilin Zhang, Yanghua Xiao, and Hanghang Tong. 2019. Spiral of Silence in Recommender Systems. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM, 222–230.

[19] Yang Yang Menghan Wang, Xiaolin Zheng and Kun Zhang. 2018. Collaborative filtering with social exposure: A modular approach to social recommendation. In *The Thirty-Second AAAI Conference on Artificial Intelligence*. AAAI, 2516–2523.

[20] 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.

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

[22] 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.

[23] 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

[24] Yi Su, Lequn Wang, Michele Santacatterina, and Thorsten Joachims. 2019. CAB: Continuous Adaptive Blending for Policy Evaluation and Learning. In *International Conference on Machine Learning*. 6005–6014.

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

[26] Gong M. Zheng X. Wang, M. and K Zhang. 2018. Modeling dynamic missingness of implicit feedback for recommendation.. In *Conference on Neural Information Processing Systems*.

[27] 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.

[28] 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 (RecSys '18)*. ACM, New York, NY, USA, 279–287. https://doi.org/10.1145/3240323.3240355