Unbiased Pairwise Learning from Implicit Feedback

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

Abstract

Implicit feedback is prevalent in the real-world and widely used to construct recommender systems. However, utilizing the implicit feedback data is considerably more complicated than the case with its explicit counterpart. This is because implicit feedback provides only positive feedback, and we cannot know whether the non-interacted feedback is positive or negative. Furthermore, positive feedback for popular items is more frequently observed than rare ones; the relevance of such popular items is often overestimated. Existing solutions to the challenges have shown to be subject to a bias towards the ideal losses of interest or accepts a simple pointwise approach, which is inappropriate for the top-N recommendation task. In this work, we first define an ideal pairwise loss function that should be used to optimize ranking metrics and propose an unbiased estimator for this ideal pairwise loss. Then, we propose a corresponding algorithm called *Unbiased Bayesian Personalized Ranking*. The pairwise algorithm addressing the two major difficulties has not yet been investigated, and the proposed algorithm is the first pairwise method to solve the two major challenges in a theoretically principal way. Through theoretical analysis, we provide critical statistical properties of the proposed unbiased estimator and a variance reduction technique. The empirical evaluations using semi-synthetic and real-world datasets demonstrate the practical strength of our approach.

1 Introduction

In the literature concerning recommender systems, collaborative filtering is one of the most basic approaches for achieving well-performing top-N recommendations [27]. Conventionally, there are two types of feedback data used in collaborative filtering systems. The first one is called **explicit feedback data**. In collaborative filtering based on explicit feedback, one has access to the users' preferences on items, and the goal is to predict preferences or ratings of the unrated user–item pairs by using the sparse observed feedback. Explicit feedback data usually contains both positive and negative feedback, and thus, this type of feedback is desirable to obtain a recommender. However, collecting sufficient explicit feedback takes time and cost; the use of this type of data is limited in real-world systems. The other one is called **implicit feedback data**. This feedback is collected through a users' natural behavior such as clicking or viewing and is more prevalent than the other. However, there are two major challenges to make recommendations using implicit feedback. First, implicit feedback contains only positive feedback, and negative feedback is unobserved. In other words, we cannot know whether a non-interacted item in the user's history is irrelevant to the user or has simply not yet been exposed. Addressing this positive-unlabeled nature of implicit feedback is essential to recommend items that are highly relevant to users. The second challenge is that the missing mechanism of feedback is missing-not-at-random (MNAR) [26]. For example, users are more likely to interact with popular items than tail items, and recommender systems are also more likely to recommend popular items than tail ones [26]. It is widely known that these biased feedback data lead to sub-optimal and severely biased recommendations [26, 19, 21, 24].

NewInML co-located with NeurIPS 2019, Vancouver, Canada.

In collaborative filtering using implicit feedback, latent factor models (LFM) have been widely used [6]. Weighted matrix factorization (WMF) is one of the most popular methods among them [6, 13]. It addresses the positive-unlabeled problem by upweighting the prediction loss of interacted items because they are always considered to be positive. This kind of prediction method aims to predict the relevance levels of the user–item pairs directly and is known as the pointwise approach. On the other hand, Bayesian personalized ranking (BPR) learns the scoring function that correctly ranks users' preference levels over items [15, 20]. This method uses a loss function over two items per user and is known as the pairwise approach. This approach is more suitable than its pointwise counterpart for the top-N recommendation settings [15, 20, 10].

These conventional methods have shown their effectiveness empirically; however, they do not directly address the problems of implicit feedback. Both pointwise and pairwise approaches minimize loss functions by regarding interacted data as positive feedback and non-interacted data as negative feedback. However, in implicit feedback data, non-interacted data does not always signify negative feedback, and thus, the loss functions that conventional methods optimize are considered to be biased.

On the other hand, few studies directly address the positive-unlabeled problem. For example, latent probabilistic models using exposure variables have been proposed [13, 3, 22]. The exposure variables represent whether a user has been exposed to an item. Once a user has been exposed to an item, the interaction between the user–item pair represents their relevance. Exposure matrix factorization (ExpoMF) is the most basic method based on this probabilistic model and EM-algorithm [13]. In the E-step, it estimates the exposure probability of each item, and in the M-step, it updates user–item latent factors by minimizing the loss that upweights the data with high estimated exposure probability. This method is the first one to model the positive-unlabeled mechanism using exposure variables. However, it does not solve the other problem, i.e., the MNAR problem. This is because ExpoMF upweights the loss of data with high exposure probability (mostly popular items); therefore, the prediction accuracy for tail items will be degraded.

The work that is most related to ours is [18]. They proposed the unbiased estimator for the loss function of interest that can be estimated from only the MNAR implicit feedback. The matrix factorization model that utilized this unbiased loss function is called Rel-MF and empirically outperformed WMF and ExpoMF. However, Rel-MF (and also WMF and ExpoMF) is based on the pointwise approach. In the top-N recommendation settings, the pairwise approach has been considered to be desirable and exhibits promising empirical results [23, 20]. However, the pairwise approach directly addressing two major challenges of implicit feedback has not yet been fully investigated, even though it is widely acknowledged that the pairwise algorithm generally outperforms the pointwise counterpart with respect to the top-N recommendation quality.

In this work, we first define the ideal pairwise loss function defined using the relevance parameter. Then, we propose an unbiased estimator for the ideal pairwise loss and a corresponding pairwise algorithm called *Unbiased Bayesian Personalized Ranking*. To the best of our knowledge, the proposed method is the first pairwise algorithm that theoretically solves the positive-unlabeled and MNAR problems of implicit feedback simultaneously. Moreover, we provide several theoretical properties of the proposed unbiased estimator and a variance reduction technique that further improves the statistical quality of the estimator. Finally, we conduct extensive empirical comparisons using semi-synthetic and real-world datasets. The results demonstrate that the proposed algorithm consistently outperforms the baseline algorithms under the situation where the observable interaction logs are severely biased.

2 Preliminaries

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

2.1 Notation

Let $u \in \mathcal{U}$ denote a user and $i \in \mathcal{I}$ denote an item. $\mathcal{D}_{\text{point}} = \mathcal{U} \times \mathcal{I}$ and $\mathcal{D}_{\text{pair}} = \mathcal{U} \times \mathcal{I} \times \mathcal{I}$ be the set of all possible data for pointwise and pairwise algorithms, respectively. $Y_{u,i}$ denotes a binary random variable representing interactions between user u and item i. If the interaction of (u, i) is observed, then $Y_{u,i} = 1$ else, $Y_{u,i} = 0$. Note that, in collaborative filtering with implicit feedback, $Y_{u,i} = 1$ indicates positive feedback, on the other hand, $Y_{u,i} = 0$ is either a negative or unlabeled positive feedback. To precisely formulate this implicit nature, we introduce relevance and exposure variables. First, the relevance variable for (u, i) is denoted as $R_{u,i}$, and this is a binary random variable representing relevance between user u and item i. $R_{u,i} = 1$ means u and i are relevant, in contrast, $R_{u,i} = 0$ suggests u and i are irrelevant. The other is the exposure variable denoted as $O_{u,i}$, which is a random variable representing whether user u has been exposed to the item i. One difficulty of the implicit feedback recommendation is that both the relevance and exposure random variables are **unobserved**; only interaction variables are observable in nature.

We now model the implicit feedback recommendation problems as follows.

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

$$\begin{aligned} P(Y_{u,i} = 1) &= P\left(O_{u,i} = 1\right) \cdot P\left(R_{u,i} = 1\right) \\ &= \theta_{u,i} \cdot \gamma_{u,i} \\ & \theta_{u,i} > 0,\ \gamma_{u,i} > 0, \quad \forall (u,i) \in \mathcal{D}_{\text{point}} \end{aligned} \tag{2}$$

where $\theta_{u,i} = P\left(O_{u,i} = 1\right)$ and $\gamma_{u,i} = P\left(R_{u,i} = 1\right)$ are defined as exposure and relevance parameters, respectively.

Eq. (1) assumes that interaction between the item i and user u is observed if i has been exposed to u and they are relevant (i.e., $Y_{u,i} = 1 \Leftrightarrow O_{u,i} = 1 = 1 \;\&\; R_{u,i} = 1$). Position-based model, an established click generative model in unbiased learning-to-rank [9, 25], makes the same assumption. This model precisely formulates the implicit feedback setting where an interaction does not always signify a relevance signal.

On the other hand, Eq. (2) assumes that interaction probability can be represented as the product of the exposure and relevance parameters[1]. Under this assumption, the MNAR problem is interpreted as the situation where interaction probability and relevance level are not proportional due to the non-uniform exposure probabilities.

2.2 Performance Metric of Interest

Top-N scoring metrics such as mean average precision and discounted cumulative gain are often used to evaluate recommendation policies with implicit feedback [26, 13]. In general, these metrics are defined using interaction probability; however, it is undesirable to measure the quality of recommender systems with respect to user experience because click does not always signify relevance in our model. This motivates us to consider the following quality measure defined using relevance level as the performance metric.

$$\mathcal{R}_{rel}\left(\widehat{\boldsymbol{Z}}\right) = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}} \underbrace{P\left(R_{u,i} = 1\right)}_{\text{relevance level}} \cdot c\left(\widehat{Z}_{u,i}\right) \tag{3}$$

where $\widehat{\boldsymbol{Z}} = \{\widehat{Z}_{u,i}\}_{(u,i) \in \mathcal{D}}$ is the predicted ranking of item i for user u, and the function $c(\cdot)$ characterizes a top-N scoring metric [26].

The focus of this study is to optimize the performance metric defined as in Eq. (3) using only the observed interaction logs.

2.3 Ideal Loss Functions of Interest

Here, we define ideal pointwise and pairwise loss functions that should be optimized to maximize the ranking metric on the relevance level as in Eq. (3).

***Definiton* 1.** *The ideal* **pointwise** *loss function is defined as*

$$\mathcal{L}_{ideal}^{point}\left(\widehat{\boldsymbol{R}}\right) = \frac{1}{|\mathcal{D}_{point}|} \sum_{(u,i) \in \mathcal{D}_{point}} \gamma_{u,i} \delta^{(1)}\left(\widehat{R}_{u,i}\right) + \left(1 - \gamma_{u,i}\right) \delta^{(0)}\left(\widehat{R}_{u,i}\right) \tag{4}$$

where $\widehat{R}_{u,i}$ is the prediction for the relevance and $\delta^{(R)},\ (R \in \{0,1\})$ denotes local loss for user–item pairs (u,i). For example, when $\delta^{(R)}(\widehat{R}) = -R\log(\widehat{R}) - (1-R)\log(1-\widehat{R})$, then Eq. (4) is called the binary cross entropy loss.

[1] This assumption comprises the same structure as that of the no-hidden confounder assumption in causal inference [17, 16, 8] and can also be represented as $E \perp R | u, i$.

Definiton 2. *The ideal* **pairwise** *loss function is defined as*

$$\mathcal{L}_{ideal}^{pair}\left(\widehat{\boldsymbol{X}}\right) = \frac{1}{|\mathcal{D}_{pair}|} \sum_{(u,i,j)\in\mathcal{D}_{pair}} \gamma_{u,i}\left(1-\gamma_{u,j}\right)\ell\left(\widehat{X}_{uij}\right) \tag{5}$$

where $\widehat{X}_{uij}$ is the difference between the predicted scores of items i and j, and ℓ denotes the local loss for the triplet (u,i,j). We describe the formal definition of the loss function of BPR in Section 3.

In the following sections, we denote $\delta^{(R)}(\widehat{R}_{u,i})$ as $\delta_{u,i}^{(R)}$ and $\ell(\widehat{X}_{uij})$ as ℓ_{uij} for simplicity.

A prediction matrix $\widehat{\boldsymbol{R}}$ or a scoring set $\widehat{\boldsymbol{X}}$ minimizing the ideal losses defined using relevance levels is expected to lead to the desired values of the top-N recommendation metrics in Eq. (3). Thus, we see the implicit feedback problem as the statistical estimation problem and aim to estimate the ideal loss functions using only the biased interaction feedback.

2.4 Summary of Existing Estimators and Algorithms

In this section, we describe standard baseline algorithms (WMF, ExpoMF, Rel-MF, and BPR) and estimators used in these methods.

2.4.1 Weighted Matrix Factorization (WMF)

WMF is a commonly used model for implicit feedback recommendation [6]. It relies on the following estimator.

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

where $c \geq 1$ is a hyperparameter determining the weight of interacted data relative to non-interacted ones. When user or item features are unavailable, a positive constant is set as c for all interacted data and this weight are tuned via cross-validation. Although this is the standard baseline model for implicit feedback [13], [18] showed that the loss function used in the WMF has a bias against the ideal pointwise loss as follows (see Proposition 3.1 of [18] for detail).

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{\text{WMF}}\left(\widehat{\boldsymbol{R}}\right)\right] \neq \mathcal{L}_{\text{ideal}}^{\text{point}}\left(\widehat{\boldsymbol{R}}\right)$$

As stated above, WMF actually optimizes a biased loss function. This is because WMF treats non-interacted items as low confidence negative feedback. However as stated in the previous Section, non-interacted feedback does not always a negative feedback. Thus, the loss function of WMF is considered to be unsuitable for optimizing the metric of interest in Eq. (3).

2.4.2 Exposure Matrix Factorization (ExpoMF)

ExpoMF adopts a loss function different from that of the WMF model to address the positive-unlabeled problem of implicit feedback. The method is constructed based on the latent probabilistic model as follows [13, 22].

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

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

where λ_U, λ_V, and λ_y are hyperparameters for prior distributions. The model also assumes

$$O_{u,i}=0 \Rightarrow Y_{u,i}=0 \tag{7}$$

which is consistent with our formulation.

The ExpoMF utilizes an EM-based iterative algorithm to derive user–item matrices. In the E-step, it estimates posterior exposure probability $\theta'_{u,i} = \mathbb{E}\left[O_{u,i} \,|\, Y_{u,i}\right]$, and in the M-step, model parameters are updated to maximize the log-likelihood[2]. As described in [18], given the true posterior exposure probabilities, the M-step can be seen as optimizing the following weigted loss function [18].

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

[2]The detailed procedure is described in Section 3.3 of [13].

In this loss function, the posterior probability represents the confidence of how much relevance information an interaction indicator $Y_{u,i}$ includes. Therefore, the loss function of the ExpoMF model is designed to consider the local loss of user–item pairs where the user has seen the item (i.e., $O_{u,i} = 1$). This is because if an item has been exposed, an interaction can be viewed as representing relevance information ($O = 1 \Rightarrow Y = R$). Thus, this approach aims to solve the positive-unlabeled problem by selecting high confidence negative feedback from numerous unlabeled feedbacks.

However, as discussed in [18], the loss function in Eq. (8) optimized in the M-step of the ExpoMF algorithm is also biased against the ideal pointwise loss.

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{\text{ExpoMF}}\left(\widehat{\boldsymbol{R}}\right)\right] \neq \mathcal{L}_{\text{ideal}}^{\text{point}}\left(\widehat{\boldsymbol{R}}\right)$$

This is because ExpoMF upweights the local loss of data data having high exposure probability. This upweighting leads to poor prediction accuracy for data having low exposure probability such as tail items. Therefore, it will fail to achieve the goal of recommender systems; recommending relevant items from non-interacted ones.

2.4.3 Relevance Matrix Factorization (Rel-MF)

Rel-MF is currently the only method that utilizes an unbiased estimator for the ideal pointwise loss as its loss function. The loss function of Rel-MF is defined as follows.

$$\widehat{\mathcal{L}}_{\text{Rel-MF}}\left(\widehat{\boldsymbol{R}}\right) = \frac{1}{|\mathcal{D}_{\text{point}}|} \sum_{(u,i)\in\mathcal{D}_{\text{point}}} \frac{Y_{u,i}}{\theta_{u,i}} \delta_{u,i}^{(1)} + \left(1 - \frac{Y_{u,i}}{\theta_{u,i}}\right) \delta_{u,i}^{(0)} \tag{9}$$

As shown in Proposition 4.3 of [18], the estimator defined in Eq. (9) satisfies the unbiasedness toward the ideal pointwise loss as:

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{\text{Rel-MF}}\left(\widehat{\boldsymbol{R}}\right)\right] = \mathcal{L}_{\text{ideal}}^{\text{point}}\left(\widehat{\boldsymbol{R}}\right)$$

This unbiasedness is deserible to optimize the pointwise metric in Eq. (4). However, in the top-N recommendation literature, it is widely known that the pairwise approach that considers the relevance orders of given pair of items is suitable for the task and has empirically outperformed the pointwise counterpart [15, 10, 20]. Moreover, the pairwise algorithms such as LambdaMART are also preferred because of its practical high performance in the Learning-to-Rank literature []. Therefore, the unbiased pairwise loss function should be developed as well as the pointwise loss for debiasing the implicit feedback to improve the recommendation quality from biased user feedback. In addition, empirical comparisons of pointwise and pairwise approaches in the MNAR implicit feedback recommendation setting is needed.

2.4.4 Bayesian Personalized Ranking (BPR)

BPR is a well-established pairwise algorithm for the top-N recommendations based on implicit feedback. It models a user's preference over two items, where the interaction of one item is observed, and that of the other is not. BPR assumes that interacted items should be ranked higher than all the other non-interacted items and optimizes the following loss function to obtain latent factors.

$$\widehat{\mathcal{L}}_{\text{BPR}}\left(\widehat{\boldsymbol{X}}\right) = \frac{1}{|\mathcal{D}_{\text{pair}}|} \sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} Y_{u,i}(1 - Y_{u,j})\ell\left(\widehat{X}_{uij}\right) \tag{10}$$

where $\ell(\cdot) = -\ln(\sigma(\cdot))$ is generally used[3], and $\widehat{X}_{uij} = U_u^T V_i - U_u^T V_j$ is the difference of predicted scores, where U_u, V_j are low-dimensional user and item factors, respectively.

It is obvious from the loss function in Eq. (10), BPR compares the interacted item ($Y_{u,i} = 1$) and non-interacted ones ($1 - Y_{u,j} = 1 \Rightarrow Y_{u,j} = 0$) for each user. The BPR algorithm has shown promising empirical results on ranking tasks [15]. However, the following proposition states that the loss function of BPR is, in reality, biased against the ideal pairwise loss function.

***Proposition* 1.** *(Bias of BPR) The estimator optimized in BPR is biased against the ideal pairwise loss in Eq. (5).*

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{BPR}\left(\widehat{\boldsymbol{X}}\right)\right] \neq \mathcal{L}_{ideal}^{pair}\left(\widehat{\boldsymbol{X}}\right)$$

[3] $\sigma(\cdot)$ is the sigmoid function

As stated in Proposition 1, the loss function of the BPR model is biased towards the ideal pairwise loss. This is because BPR treats all non-interacted feedback as negative feedback and does not deal with the positive-unlabeled problem; therefore, it may underestimate the relevance of non-interacted pairs.

3 Proposed Method

In this section, we present an unbiased estimator for the ideal pairwise loss inspired by the inverse propensity score (IPS) estimator in the context of causal inference [17, 16, 8]. We then provide the learning algorithm of *Unbiased Bayesian Personalized Ranking (UBPR)* and the essential theoretical properties of the proposed unbiased estimator.

3.1 Proposed Estimator

First, we formally define the propensity score to deal with the MNAR problem of implicit feedback. Propensity score is often used to estimate causal effects of treatments from observational data [8, 16, 17]. The propensity score in the implicit recommendation setting is defined as follows.

***Definiton* 3.** *(Propensity Score) Propensity score of user–item pair (u, i) is*

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

Then, the proposed estimator is defined using the propensity score.

***Definiton* 4.** *(Unbiased estimator for the ideal pairwise loss) When propensity scores are given, then the unbiased pairwise loss function is defined as*

$$\widehat{\mathcal{L}}_{unbiased}\left(\widehat{\boldsymbol{X}}\right) = \frac{1}{|\mathcal{D}_{pair}|} \sum_{(u,i,j)\in\mathcal{D}_{pair}} \frac{Y_{u,i}}{\theta_{u,i}} \left(1 - \frac{Y_{u,j}}{\theta_{u,j}}\right) \ell_{uij} \tag{11}$$

The proposed estimator weights each data by the inverse of the propensity and can also be represented as

$$\frac{1}{|\mathcal{D}_{\text{pair}}|} \sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} \frac{Y_{u,i}}{\theta_{u,i}} \left(Y_{u,j}\left(1 - \frac{1}{\theta_{u,j}}\right) + (1 - Y_{u,j})\right) \ell_{uij} \tag{12}$$

From the expression in Eq. (12), it can be seen as utilizing user-item pairs where the interaction of both items are observed ($Y_{ui} = Y_{uj} = 1$) as well as the interacted and non-interacted pair of items ($Y_{ui} = 1 \& Y_{uj} = 0$). In the following proposition, we show that the unbiased pairwise loss is truely unbiased against the ideal pairwise loss.

***Proposition* 2.** *The unbiased pairwise loss in Eq. (11) is unbiased against the ideal pairwise loss in Eq. (5).*

$$\mathbb{E}\left[\widehat{\mathcal{L}}_{unbiased}\left(\widehat{\boldsymbol{X}}\right)\right] = \mathcal{L}_{ideal}^{pair}\left(\widehat{\boldsymbol{X}}\right)$$

Proposition 2 shows that the proposed unbiased pairwise loss function derived via propensity weighting is valid for debiasing the MNAR implicit feedback.

We also present the variance the proposed unbiased estimator.

***Theorem* 3.** *(Variance of the unbiased pairwise loss) Given sets of independent random variables $\{(Y_{u,i}, O_{u,i}, R_{u,i})\}$, propensity scores $\{\theta_{u,i}\}$, and predicted scoring set $\widehat{\boldsymbol{X}}$, the variance of the unbiased pairwise loss is*

$$\mathbb{V}\left(\widehat{\mathcal{L}}_{unbiased}\left(\widehat{\boldsymbol{X}}\right)\right) = \frac{1}{|\mathcal{D}_{pair}|^2}\left(\sum_{(u,i,j)\in\mathcal{D}_{pair}} v_{uij}\ell_{uij} + \sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{I}}\sum_{\substack{(j,k)\in\mathcal{I}\times\mathcal{I}\\ j\neq k}} w_{uijk}\ell_{uij}\ell_{uik}\right)$$

where

$$v_{uij} = \gamma_{u,i}(1 - 2\gamma_{u,j})\left(\frac{1}{\theta_{u,i}} - \gamma_{u,i}\right) + \gamma_{u,i}\gamma_{u,j}\left(\frac{1}{\theta_{u,i}\theta_{u,j}} - \gamma_{u,i}\gamma_{u,j}\right)$$

$$w_{uijk} = \left(\frac{1}{\theta_{u,i}} - \gamma_{u,i}\right)\gamma_{u,i}(1 - \gamma_{u,j})(1 - \gamma_{u,k})\ell_{uik}\ell_{uij}$$

Algorithm 1 Unbiased Bayesian Personalized Ranking

Input: observed interaction data $\{Y_{u,i}\}$, learning_rate μ, regularization parameter λ.
Output: learned model parameters $\boldsymbol{U}, \boldsymbol{V}$
1: Initialize $\boldsymbol{U}, \boldsymbol{V}$, and estimate propensity scores $\{\theta_{u,i}\}$
2: **repeat**
3: Sample mini-batch samples $\mathcal{D}_{\text{mini}}$ from $\mathcal{D}$
4: Compute $loss = \widehat{\mathcal{L}}_{\text{unbiased}}\left(\widehat{\boldsymbol{X}}\right) + \lambda\left(\|\boldsymbol{U}\|_2^2 + \|\boldsymbol{V}\|_2^2\right)$ with $\mathcal{D}_{\text{mini}}$
5: Update parameters $\boldsymbol{U} \leftarrow \boldsymbol{U} + \mu\left(\frac{\partial loss}{\partial \boldsymbol{U}} - \lambda \boldsymbol{U}\right), \boldsymbol{V} \leftarrow \boldsymbol{V} + \mu\left(\frac{\partial loss}{\partial \boldsymbol{V}} - \lambda \boldsymbol{V}\right)$
6: **until** convergence;
7: **return** $\boldsymbol{U}, \boldsymbol{V}$

3.2 Variance Reduction Technique

The RHS of both the variance depend on the product of two propensity scores. Thus, these bounds can be loose, especially for tail items having low exposure probability. From these implications, we also propose to utilize the following non-negative estimator inspired by the work in positive-unlabeled learning [11] as follows:

Definiton 5. *(Non-negative estimator) When propensity scores and a constant $\beta \geq 0$ are given, then the non-negative estimator is defined as*

$$\widehat{\mathcal{L}}_{\textit{non-neg}}\left(\widehat{\boldsymbol{X}}\right) = \frac{1}{|\mathcal{D}_{\textit{pair}}|} \sum_{(u,i,j)\in\mathcal{D}_{\textit{pair}}} \max\left\{\ell_{\textit{unbiased}}\left(\widehat{X}_{uij}\right) = \frac{Y_{u,i}}{\theta_{u,i}}\left(1 - \frac{Y_{u,j}}{\theta_{u,j}}\right)\ell_{uij}\left(\widehat{X}_{uij}\right), 0\right\} \tag{13}$$

The non-negative estimator reduces the variance of the estimator at the cost of introducing some bias.

3.3 Learning Algorithm

Here, we formally describe the proposed UBPR algorithm. It obtains its model parameters by optimizing the following loss function based on the unbiased pairwise loss for an ideal pairwise loss function.

$$\boldsymbol{U}, \boldsymbol{V} = \arg\min_{\boldsymbol{U},\boldsymbol{V}} \widehat{\mathcal{L}}_{\text{unbiased}}\left(\widehat{\boldsymbol{X}}\right) + \lambda\left(\|\boldsymbol{U}\|_2^2 + \|\boldsymbol{V}\|_2^2\right) \tag{14}$$

where the second term is the L2-regularization for the latent factors and λ is a hyperparameter for the regularization.

We summarize the whole learning procedure of UBPR in Algorithm 1.

4 Experimental Results

In this section, we empirically demonstrate the effectiveness of the proposed UBPR.

4.1 Experimental Setup

We used the Yahoo! R3 dataset[4]. This is an explicit feedback dataset collected from a song recommendation service. As described in [26], it contains users' ratings towards randomly selected sets of music as a test set and thus can be used to measure recommenders' true performances. In the experiment, we treated items rated greater than or equal to 4 as relevant, and the others were considered irrelevant.

We compared WMF[6], ExpoMF [13], Rel-MF [18], BPR [15], and UBPR[5]. For all the methods, the size of latent dimensions was tuned within the range of $\{50, 55, \ldots, 150\}$ using a validation set. The

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

[5]We used the implementation provided at https://github.com/dawenl/expo-mf for ExpoMF.

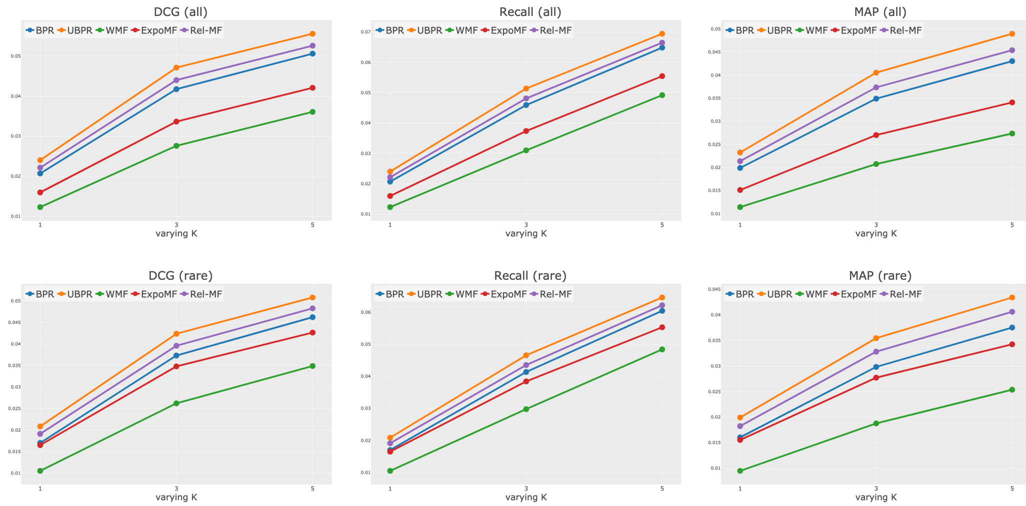

Figure 1: Ranking metrics on the Yahoo! R3 dataset with different values of K. UBPR consistently outperforms the other baselines for both all and rare items.

L2 reguralization hyper-parameter λ was set to 10^{-4} for all the method. As for the proposed method and the Rel-MF, we estimated the propensity score by the following relative item popularity.

$$\widehat{\theta}_{*,i} = \left(\frac{\sum_{u \in \mathcal{U}} Y_{u,i}}{\max_{i \in \mathcal{I}} \sum_{u \in \mathcal{U}} Y_{u,i}} \right)^{\eta} \tag{15}$$

In our assumption, interaction probability depends on both exposure probability and relevance level, thus, we set $\eta = 0.5$.

4.2 Results

Figure 2 shows the performance of the methods, corresponding to rare and all items. We defined the items that had been clicked by 300 users at most 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 5.7%, Recall@5 by 4.4%, and MAP@5 by 7.8% over the Rel-MF. Furthermore, for rare items, the proposed method improved the DCG@5 by 5.2%, Recall@5 by 3.8%, and MAP@5 by 6.8% over the best baseline.

The results indicate that the proposed algorithm significantly outperforms the other baselines and validates the effectiveness of the proposed debiasing approach on biased implicit feedback.

5 Conclusion

In this study, we explored two major challenges; the positive-unlabeled and MNAR problems of collaborative filtering with implicit feedback. To solve the challenges, we first modeled the implicit recommendation problem using relevance and exposure variables. Then, we proposed the unbiased estimator for the ideal pairwise loss function and a corresponding algorithm called *Unbiased Bayesian Personalized Ranking*. Empirical evaluation using the standard public dataset demonstrated that the proposed algorithm outperformed the current state-of-the-art baselines.

References

[1] Qingyao Ai, Keping Bi, Cheng Luo, Jiafeng Guo, and W Bruce Croft. Unbiased learning to rank with unbiased propensity estimation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 385–394. ACM, 2018.

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

[3] Jiawei Chen, Can Wang, Martin Ester, Qihao Shi, Yan Feng, and Chun Chen. Social recommendation with missing not at random data. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 29–38. IEEE, 2018.

[4] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182. International World Wide Web Conferences Steering Committee, 2017.

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

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

[7] Ziniu Hu, Yang Wang, Qu Peng, and Hang Li. Unbiased lambdamart: An unbiased pairwise learning-to-rank algorithm. In *The World Wide Web Conference*, pages 2830–2836. ACM, 2019.

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

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

[10] Seunghyeon Kim, Jongwuk Lee, and Hyunjung Shim. Dual neural personalized ranking. In *The World Wide Web Conference*, pages 863–873. ACM, 2019.

[11] Ryuichi Kiryo, Gang Niu, Marthinus C du Plessis, and Masashi Sugiyama. Positive-unlabeled learning with non-negative risk estimator. In *Advances in neural information processing systems*, pages 1675–1685, 2017.

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

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

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

[15] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452–461. AUAI Press, 2009.

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

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

[18] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. Relevance matrix factorization. *arXiv preprint arXiv:1909.03601*, 09 2019.

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

[20] Bo Song, Xin Yang, Yi Cao, and Congfu Xu. Neural collaborative ranking. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1353–1362. ACM, 2018.

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

[22] Menghan Wang, Mingming Gong, Xiaolin Zheng, and Kun Zhang. Modeling dynamic missingness of implicit feedback for recommendation. In *Advances in neural information processing systems*, pages 6669–6678, 2018.

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

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

[25] Xuanhui Wang, Nadav Golbandi, Michael Bendersky, Donald Metzler, and Marc Najork. 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*, pages 610–618. ACM, 2018.

[26] Longqi Yang, Yin Cui, Yuan Xuan, Chenyang Wang, Serge Belongie, and Deborah Estrin. Unbiased offline recommender evaluation for missing-not-at-random implicit feedback. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, pages 279–287, New York, NY, USA, 2018. ACM.

[27] Zainab Zolaktaf, Reza Babanezhad, and Rachel Pottinger. A generic top-n recommendation framework for trading-off accuracy, novelty, and coverage. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 149–160. IEEE, 2018.

Supplementary Material

A Omitted Proofs

A.1 Proof of Proposition 1

Proof.

$$\begin{aligned}
\mathbb{E}\left[\widehat{\mathcal{L}}_{\text{BPR}}\left(\widehat{\boldsymbol{X}}\right)\right] &= \mathbb{E}\left[\frac{1}{|\mathcal{D}_{\text{pair}}|}\sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} Y_{u,i}(1-Y_{u,j})\ell_{uij}\right] \\
&= \frac{1}{|\mathcal{D}_{\text{pair}}|}\sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} \mathbb{E}\left[Y_{u,i}\right](1-\mathbb{E}\left[Y_{u,j}\right])\ell_{uij} \\
&= \frac{1}{|\mathcal{D}_{\text{pair}}|}\sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} \theta_{u,i}\gamma_{u,i}(1-\theta_{u,j}\gamma_{u,j})\ell_{uij}
\end{aligned}$$

Thus we obtain,

$$\begin{aligned}
\mathbb{E}\left[\widehat{\mathcal{L}}_{\text{BPR}}\left(\widehat{\boldsymbol{X}}\right)\right] - \mathcal{L}^{\text{pair}}_{\text{ideal}}\left(\widehat{\boldsymbol{X}}\right) &= \frac{1}{|\mathcal{D}_{\text{pair}}|}\sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} \theta_{u,i}\gamma_{u,i}(1-\theta_{u,j}\gamma_{u,j})\ell_{uij} - \frac{1}{|\mathcal{D}_{\text{pair}}|}\sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} \gamma_{u,i}(1-\gamma_{u,j})\ell_{uij} \\
&= \frac{1}{|\mathcal{D}_{\text{pair}}|}\sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} \gamma_{u,i}\left((\theta_{u,i}-1)+(1-\theta_{u,i}\theta_{u,j})\gamma_{u,j}\right)\ell_{uij}
\end{aligned}$$

For $\widehat{\mathcal{L}}_{\text{BPR}}\left(\widehat{\boldsymbol{X}}\right)$ to be theoretically unbiased, $\theta_{u,i}-1=0 \Rightarrow \theta_{u,i}=1, 1-\theta_{u,i}\theta_{u,j}=0 \Rightarrow \theta_{u,i}\theta_{u,j}=1$ are 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, and thus, these conditions are not always satisfied. Thus, the loss function of BPR is biased toward the ideal pointwise loss function. □

A.2 Proof of Proposition 2

Proof.

$$\begin{aligned}
\mathbb{E}\left[\widehat{\mathcal{L}}_{\text{unbiased}}\left(\widehat{\boldsymbol{X}}\right)\right] &= \mathbb{E}\left[\frac{1}{|\mathcal{D}_{\text{pair}}|}\sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} \frac{Y_{u,i}}{\theta_{u,i}}\left(1-\frac{Y_{u,j}}{\theta_{u,j}}\right)\ell_{uij}\right] \\
&= \frac{1}{|\mathcal{D}_{\text{pair}}|}\sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} \frac{\mathbb{E}\left[Y_{u,i}\right]}{\theta_{u,i}}\left(1-\frac{\mathbb{E}\left[Y_{u,j}\right]}{\theta_{u,j}}\right)\ell_{uij} \\
&= \frac{1}{|\mathcal{D}_{\text{pair}}|}\sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}} \gamma_{u,i}\left(1-\gamma_{u,j}\right)\ell_{uij} \\
&= \mathcal{L}^{\text{pair}}_{\text{ideal}}\left(\widehat{\boldsymbol{X}}\right)
\end{aligned}$$

□

A.3 Proof and Statement of a Technical Lemma

Here we state some technical lemmas that are used to prove the later theorems.

***Lemma* 4.** *(Covariance) Given a scoring set X_{uij}, Let a random variable Z_{uij} be*

$$Z_{u,i,j} = \frac{Y_{u,i}}{\theta_{u,i}}\left(1-\frac{Y_{u,j}}{\theta_{u,j}}\right)\ell_{uij}$$

Then for any user $u \in U$ and items $i \in \mathcal{I}$, $j \in \mathcal{I}$, $k \in \mathcal{I}$ where $i \neq j \neq k$, the covariance of Z_{uij} and Z_{uik} are

$$Cov(Z_{uij}, Z_{uik}) = \left(\frac{1}{\theta_{u,i}} - \gamma_{u,i}\right)\gamma_{u,i}(1-\gamma_{u,j})(1-\gamma_{u,k})\ell_{uik}\ell_{uij} \tag{16}$$

Proof. First, the covariance can be represented as

$$Cov(Z_{uij}, Z_{uik}) = \mathbb{E}[Z_{uij}Z_{uik}] - \mathbb{E}[Z_{uij}]\mathbb{E}[Z_{uik}]$$

By proposition 2, the second term of the RHS is

$$\mathbb{E}[Z_{uij}]\mathbb{E}[Z_{uik}] = \gamma_{u,i}^2(1-\gamma_{u,j})(1-\gamma_{u,k})\ell_{uij}\ell_{uik}$$

Then, we calculate the first term below.

$$\begin{aligned} Z_{uij}Z_{uik} &= \left(\frac{Y_{u,i}}{\theta_{u,i}}\left(1-\frac{Y_{u,j}}{\theta_{u,j}}\right)\ell_{uij}\right) \times \left(\frac{Y_{u,i}}{\theta_{u,i}}\left(1-\frac{Y_{u,k}}{\theta_{u,k}}\right)\ell_{uik}\right) \\ &= \frac{Y_{u,i}}{\theta_{u,i}^2}\left(1-\frac{Y_{u,j}}{\theta_{u,j}}\right)\left(1-\frac{Y_{u,k}}{\theta_{u,k}}\right)\ell_{uij}\ell_{uik} \end{aligned}$$

Thus, the expectation of $Z_{uij}Z_{uik}$ is

$$\mathbb{E}[Z_{uij}Z_{uik}] = \frac{\gamma_{u,i}}{\theta_{u,i}}(1-\gamma_{u,j})(1-\gamma_{u,k})\ell_{uij}\ell_{uik}$$

Here, we use $\mathbb{E}[Y_{u,i}] = \theta_{u,i}\gamma_{u,i}, \mathbb{E}[Y_{u,j}] = \theta_{u,j}\gamma_{u,j}, \mathbb{E}[Y_{u,k}] = \theta_{u,k}\gamma_{u,k}$ and the independence assumption. Finally, the covariance can be obtained as follows.

$$\begin{aligned} \mathbb{E}[Z_{uij}Z_{uik}] - \mathbb{E}[Z_{uij}]\mathbb{E}[Z_{uik}] &= \frac{\gamma_{u,i}}{\theta_{u,i}}(1-\gamma_{u,j})(1-\gamma_{u,k})\ell_{uij}\ell_{uik} - \gamma_{u,i}^2(1-\gamma_{u,j})(1-\gamma_{u,k})\ell_{uij}\ell_{uik} \\ &= \left(\frac{1}{\theta_{u,i}} - \gamma_{u,i}\right)\gamma_{u,i}(1-\gamma_{u,j})(1-\gamma_{u,k})\ell_{uik}\ell_{uij} \end{aligned}$$

□

A.4 Proof of Theorem 1

Proof. First, we define

$$Z_{u,i,j} = \frac{Y_{u,i}}{\theta_{u,i}}\left(1-\frac{Y_{u,j}}{\theta_{u,j}}\right)\ell_{uij}$$

Then we have $\mathbb{V}(Z_{uij})$ as

$$\mathbb{V}(Z_{uij}) = \underbrace{\mathbb{E}\left[(Z_{uij})^2\right]}_{(b)} - \underbrace{(\mathbb{E}[Z_{uij}])^2}_{(c)}$$

By Proposition 2, $(c) = (\gamma_{u,i}(1-\gamma_{u,j})\ell_{u,i,j})^2 = (\gamma_{u,i}^2 - 2\gamma_{u,i}\gamma_{u,j} + \gamma_{u,i}^2)\ell_{uij}^2$. Next,

$$Z_{uij}^2 = \frac{Y_{u,i}}{\theta_{u,i}^2}\left(1 - \frac{2Y_{u,j}}{\theta_{u,j}} + \frac{Y_{u,j}}{\theta_{u,j}^2}\right)\ell_{uij}^2 \tag{17}$$

where $Y_{u,i}^2 = Y_{u,i}$ and $Y_{u,j}^2 = Y_{u,j}$. The expectation of the RHS of Eq. (17) is

$$(b) = \frac{\gamma_{u,i}}{\theta_{u,i}}\left(1 - 2\gamma_{u,j} + \frac{\gamma_{u,j}}{\theta_{u,j}}\right)\ell_{uij}^2$$

Therefore,

$$\begin{aligned} \mathbb{V}(Z_{uij}) &= (b) - (c) \\ &= \frac{\gamma_{u,i}}{\theta_{u,i}}\left(1 - 2\gamma_{u,j} + \frac{\gamma_{u,j}}{\theta_{u,j}}\right)\ell_{uij}^2 - \left(\gamma_{u,i}^2 - 2\gamma_{u,i}\gamma_{u,j} + \gamma_{u,i}^2\right)\ell_{uij}^2 \\ &= \underbrace{\left[\gamma_{u,i}(1-2\gamma_{u,j})\left(\frac{1}{\theta_{u,i}} - \gamma_{u,i}\right) + \gamma_{u,i}\gamma_{u,j}\left(\frac{1}{\theta_{u,i}\theta_{u,j}} - \gamma_{u,i}\gamma_{u,j}\right)\right]}_{v_{uij}}\ell_{uij}^2 \end{aligned} \tag{18}$$

Then, the variance of the sums of random variables $\{Z_{uij}\}$ are

$$\mathbb{V}\left(\widehat{\mathcal{L}}_{\text{unbiased}}\left(\widehat{\boldsymbol{R}}\right)\right) = \frac{1}{|\mathcal{D}_{\text{pair}}|^2}\left(\sum_{(u,i,j)\in\mathcal{D}_{\text{pair}}}\mathbb{V}(Z_{uij}) + \sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{I}}\sum_{\substack{(j,k)\in\mathcal{I}\times\mathcal{I} \\ j\neq k}} Cov(Z_{uij}, Z_{uik})\right) \tag{19}$$

Combining Eq. (16), Eq. (18), and Eq. (19) completes the proof. □

B Related Work

In this section, we review the existing related studies.

B.1 Implicit Recommendation Algorithms

The most basic prediction algorithm in the implicit recommendation is WMF [6]. It minimizes the weighted pointwise loss function (e.g., binary cross-entropy loss or mean-squared loss) and obtains user–item latent factors. Another well-known baseline is BPR [15]. This method minimizes the pairwise loss function and learns user–item latent factors that give interacted items higher scores than non-interacted ones. Several papers apply deep neural networks to these basic latent factor models and show promising results on benchmark datasets [4, 20, 10].

As discussed in the introduction, these methods do not address the two major challenges of implicit feedback recommendation. For example, for both pointwise and pairwise approaches, unlabeled feedback is regarded as negative feedback. Thus, these conventional methods often underestimate the relevance level of unlabeled data. Moreover, these methods do not deal with the MNAR problem caused by popularity, presentation, and recency biases [26]. Thus, recommendations made by these methods might be sub-optimal [14, 21, 19, 26].

B.2 Debiasing Recommender Systems

There are some studies directly addressing the MNAR problem in the explicit feedback recommendation. Some works assumed missing data model and rating prediction model and estimated the parameters of these models via the EM algorithm [14, 5]. Another approach to the MNAR problem is the causal-based recommendation [2, 19, 12]. These methods applied the Inverse Propensity Score (IPS) estimation [17, 16, 8] established in the context of causal inference or domain adaptation to deal with the MNAR problem of when learning a recommender using explicit feedback.

In comparison with the explicit counterpart, there are only a few methods directly addressing the challenging problems in implicit recommendation studies. Among them, ExpoMF is the most basic one [13, 22, 3]. It introduces the exposure variable representing whether a user has been exposed to an item and a corresponding probabilistic model assuming the generative mechanism of how interactions occur. ExpoMF updates its parameters by minimizing the weighted squared loss in the M-step, and the weights are the exposure probability estimated during the E-step. This is the important work explicitly addressing the positive-unlabeled problem of implicit feedback. However, it does not tackle the MNAR problem and might demonstrate the poor prediction accuracy on tail items. In addition, [18] proposed an unbiased estimator for the ideal pointwise loss function using estimation techniques from causal inference and positive-unlabeled learning. A corresponding prediction method called Rel-MF is also proposed. However, this method is based on the pointwise approach. In general, the pairwise algorithm practically outperforms the pointwise approach in the optimization of ranking metrics. However, a method to deal with the bias of the pairwise approach has not yet been fully investigated despite its effectiveness in the top-N recommendation problem. [15, 20].

B.3 Unbiased Learning-to-Rank

Unbiased Learning-to-Rank aims to obtain an optimal ranking function to optimally order documents to a given query using only the biased click log data [9, 25, 7, 1]. To achieve this goal, addressing the bias of the click log data is essential, which is similar to our implicit feedback recommendation setting. Most studies in this area assume the position-based model (PBM) on the click generative process[9, 25, 1]. Based on this click model, one can obtain an unbiased estimator for a listwise risk function of interest through the IPS estimation [9].

The method that is most related to ours in the context of unbiased learning-to-rank is pairwise debiasing and Unbiased LambdaMART [7]. Pairwise debiasing is the first debiasing method for the pairwise loss function in unbiased learning-to-rank. In the theoretical analysis of [7], the unbiasedness of the pairwise debiasing is shown; however, this unbiasedness is based on the unrealistic assumption that both click and unclick probabilities are proportional to relevance and irrelevance level, respectively[6]. This assumption is more strict than the simple PBM because unclick probability

[6]Eq. (16) and Eq. (17) of [7].

is not necessarily proportional to the irrelevance level in this click model. In contrast to the work of [7], our algorithm and loss function is based on looser assumptions and can be used to improve the estimation quality of the pairwise debiasing method for the unbiased learning-to-rank problem.