Unbiased Pairwise Learning from Biased Implicit Feedback

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

ABSTRACT

Implicit feedback is prevalent in real-world scenarios and is widely used in the construction of recommender systems. However, the application of implicit feedback data is much more complicated than its explicit counterpart because it provides only positive feedback, and we cannot know whether the non-interacted feedback is positive or negative. Furthermore, positive feedback for rare items is observed less frequently than popular items. The relevance of such rare items is often underestimated. Existing solutions to such challenges are subject to bias toward the ideal loss function of interest or accept a simple pointwise approach, which is inappropriate for a ranking task. In this study, we first define an ideal pairwise loss function defined using the ground-truth relevance parameters that should be used to optimize the ranking metrics. Subsequently, we propose a theoretically grounded unbiased estimator for this ideal pairwise loss and a corresponding algorithm, *Unbiased Bayesian Personalized Ranking*. A pairwise algorithm addressing the two major difficulties in using implicit feedback has yet to be investigated, and the proposed algorithm is the first pairwise method for solving these challenges in a theoretically principal manner. Through theoretical analysis, we provide the critical statistical properties of the proposed unbiased estimator and a practical variance reduction technique. Empirical evaluations using real-world datasets demonstrate the practical strength of our approach.

CCS CONCEPTS

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

KEYWORDS

implicit feedback, recommender systems, missing-not-at-random, inverse propensity score, positive-unlabeled learning, Bayesian personalized ranking.

ACM Reference Format:
Yuta Saito. 2018. Unbiased Pairwise Learning from Biased Implicit Feedback. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/1122445.1122456

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

In the literature on recommender systems, collaborative filtering has been one of the most basic approaches for achieving well-performing top-N recommendations [24]. Two types of feedback data are conventionally used in collaborative filtering systems. The first is called *explicit feedback* data. In collaborative filtering based on such feedback, one has access to user preferences such as in the form of ratings on items; the goal is to predict the ratings of unrated user–item pairs by using the observed sparse feedback. Explicit feedback data usually contain both positive and negative feedback and are therefore desirable when training a recommender. However, collecting an adequate amount of explicit feedback requires time and money, and the use of such data is limited in real-world systems. The second type is called *implicit feedback* data. Such feedback is collected through natural behaviors of the users such as clicking or viewing, and is more prevalent than explicit feedback. 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 simply, has not yet been exposed. Addressing this *positive-unlabeled* nature of implicit feedback is essential to recommending items that are highly relevant to the users. The second challenge is that the missing mechanism of feedback is *missing-not-at-random* (MNAR) [15, 23]. For example, users are more likely to interact with popular items, and recommender systems are also more likely to recommend popular items rather than tail items [23]. It is widely known that such biased feedback data lead to sub-optimal recommendations [16, 18, 22, 23].

Related Work. In collaborative filtering using implicit feedback, latent factor models have been widely used [4]. Among such models, weighted matrix factorization (WMF) is one of the most popular methods [4, 11], which addresses the positive-unlabeled problem by upweighting the prediction loss of interacted items because they are always considered to be positive. Such a prediction method aims to predict the relevance levels of the user–item pairs directly and is called a pointwise approach. In contrast, Bayesian personalized ranking (BPR) learns the scoring function that correctly ranks the users' preference levels among different items [12, 17]. This approach is more suitable than its pointwise counterpart for top-N recommendation settings [8, 12, 17]. These conventional methods have shown their effectiveness empirically [4, 12]. However, they do not directly address the challenges of implicit feedback. Both pointwise and pairwise approaches minimize the loss functions by regarding the interacted data as positive feedback and the non-interacted data as negative feedback. It is worth noting that, in implicit feedback data, non-interacted data do not always signify negative feedback; thus, the loss functions optimized by conventional methods are considered to be biased.

There are only a few studies directly addressing the positive-unlabeled problem. For example, latent probabilistic models using exposure variables have been proposed [3, 11, 19]. The exposure variables represent whether a user has been exposed to an item. Once a user has been exposed, an interaction between a user and an item represents their relevance. Exposure matrix factorization (ExpoMF) is the most basic method based on this probabilistic model and the EM-algorithm [11]. In the E-step, ExpoMF estimates the exposure probability of each item, and in the M-step, it updates the latent user–item factors by minimizing the loss that upweights the data with a high estimated exposure probability. This method is the first to model a 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 a high exposure probability (mostly popular items), and the prediction accuracy for the tail items is degraded.

The first study solving these two challenging issues is [15]. The authors proposed an unbiased estimator for the loss function of interest, which can be estimated from only the MNAR implicit feedback. A matrix factorization model utilizing this unbiased loss function is called Rel-MF, which empirically outperforms WMF and ExpoMF. However, Rel-MF (as well as WMF and ExpoMF) is based on the pointwise approach. In the top-N recommendation settings, the pairwise approach has been considered to be more desirable and provides promising empirical results [17, 20]. Although it is widely acknowledged that the pairwise algorithm generally outperforms the pointwise approach with respect to the top-N recommendation quality, a pairwise approach directly addressing the two major challenges of implicit feedback has not yet been completely investigated.

Contributions. In this study, we first define an ideal pairwise loss function using the ground-truth relevance parameter. Subsequently, we propose a theoretically refined unbiased estimator for the ideal loss and a corresponding pairwise algorithm called *Unbiased Bayesian Personalized Ranking*. To our knowledge, the proposed method is the first pairwise algorithm that theoretically solves the positive-unlabeled and MNAR problems of implicit feedback. In addition, we provide several theoretical properties of the proposed unbiased estimator as well as a variance reduction technique that further improves the statistical quality of the estimator. Finally, we conduct empirical comparisons using two real-world datasets. The results demonstrate that the proposed algorithm consistently outperforms the baseline algorithms when the observable interaction logs are severely biased.

Our contributions can be summarized as follows.

- We propose the first pairwise algorithm addressing both the positive-unlabeled and MNAR problems of an implicit feedback recommendation.
- We theoretically analyze our unbiased estimator and provide important statistical properties. In addition, we introduce a practical variance reduction technique.
- We conduct comprehensive experiments on two real-world datasets. The results demonstrate that the proposed method significantly outperforms the other baselines, particularly for cold-start users and tail items.

2 PRELIMINARIES

In this section, we introduce the basic notation of the paper 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. In addition, let $\mathcal{D}_{point} = \mathcal{U} \times \mathcal{I}$ be the set of all possible data. Here, $Y_{u,i}$ denotes a binary random variable representing implicit interactions between user u and item i. If the interaction of (u, i) is observed, then $Y_{u,i} = 1$; otherwise, $Y_{u,i} = 0$. Note that, in collaborative filtering with implicit feedback, $Y_{u,i} = 1$ indicates positive feedback, whereas $Y_{u,i} = 0$ indicates either a negative or unlabeled positive feedback. To precisely formulate this implicit nature, we introduce relevance and exposure variables. The relevance variable for (u, i) is denoted as $R_{u,i}$ and 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 that u and i are irrelevant. The exposure variable, denoted as $O_{u,i}$, is a random variable representing whether u has been exposed to i. One particular difficulty in implicit feedback recommendation is that both the relevance and exposure random variables are **unobserved**, and only the interaction variables are observable in nature.

We now model the implicit feedback recommendation as follows.

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

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

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

Eq. (1) assumes that the interaction between item i and user u is observed if u has been exposed to i and both are relevant (i.e., $Y_{u,i} = 1 \Leftrightarrow O_{u,i} = 1 = 1 \ \& \ R_{u,i} = 1$). A position-based model, an established click generative model in information retrieval [6, 21], makes a similar assumption. This model precisely formulates the implicit feedback setting where an interaction does not always signify a relevance signal.

By contrast, Eq. (2) assumes that the interaction probability can be represented as the product of the exposure and relevance parameters[1]. Under this assumption, the MNAR problem is interpreted as a situation in which the interaction probability and relevance level are disproportional, which is due to the non-uniform exposure probabilities.

2.2 Performance Metric of Interest

Top-N scoring metrics such as the *mean average precision* (MAP) and *discounted cumulative gain* (DCG) are often used to evaluate recommendation policies with implicit feedback [11, 23]. In general, these metrics are defined using the interaction probability (i.e., $P(Y_{u,i} = 1)$). However, it is undesirable to measure the quality of a recommender system with respect to the user experience because a click does not always signify relevance with our model. This

[1]This assumption comprises the same structure as that of a no-hidden confounder assumption in causal inference [5, 13, 14], and can also be represented as $O \perp R | u, i$.

motivated us to consider the following quality measure defined using the relevance level as a performance metric [15].

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

Herein, $\widehat{Z} = \{\widehat{Z}_{u,i}\}_{(u,i) \in \mathcal{D}}$ is the predicted ranking of i for u, and the function $c(\cdot)$ characterizes a top-N scoring metric [23]. For example, when we use $c\left(\widehat{Z}_{u,i}\right) = \mathbb{I}\{\widehat{Z}_{u,i} \leq K\}/K$, $\mathcal{R}(\widehat{Z})$ is the *Precision@K* metric[2].

The focus of this study is to optimize the performance metric defined in Eq. (3) by using only the observed implicit feedback.

2.3 Ideal Loss Functions of Interest

Herein, we define ideal pointwise and pairwise loss functions, which should ideally be optimized to maximize the metric in Eq. (3).

Definition 2.1. The ideal ***pointwise*** *loss function is defined as*

$$\begin{aligned} &\mathcal{L}_{ideal}^{point}(f) \\ &= \frac{1}{|\mathcal{D}_{point}|} \sum_{(u,i) \in \mathcal{D}_{point}} \gamma_{u,i} \delta^{(1)}(f(u,i)) + \left(1 - \gamma_{u,i}\right) \delta^{(0)}(f(u,i)) \end{aligned} \tag{4}$$

where $f : \mathcal{U} \times \mathcal{I} \to [0,1]$ is a relevance level predictor and $\delta^{(R)}(\cdot)$ denotes the local loss for user–item pairs (u,i). For example, when $\delta^{(R)}(f) = -R\log(f(u,i)) - (1-R)\log(1-f(u,i))$, then Eq. (4) is called the binary cross-entropy loss. For simplicity, we denote $\delta^{(R)}(f(u,i))$ as $\delta_{u,i}^{(R)}$ hereafter

.

Definition 2.2. The ideal ***pairwise*** *loss function is defined as*

$$\begin{aligned} &\mathcal{L}_{ideal}^{pair}(f) \\ &= \frac{1}{|\mathcal{D}_{pair}|} \sum_{(u,i,j) \in \mathcal{D}_{pair}} \mathbb{E}\left[R_{u,i}\left(1 - R_{u,j}\right) \ell(f(u,i) - f(u,j))\right] \\ &= \frac{1}{|\mathcal{D}_{pair}|} \sum_{(u,i,j) \in \mathcal{D}_{pair}} \gamma_{u,i}\left(1 - \gamma_{u,j}\right) \ell(f(u,i) - f(u,j)) \end{aligned} \tag{5}$$

where $f : \mathcal{U} \times \mathcal{I} \to \mathbb{R}$ is a scoring function, and $\mathcal{D}_{pair} = \{(u,i,j) \mid u \in \mathcal{U}, (i,j) \in \mathcal{I} \times \mathcal{I}, i \neq j\}$ is the set of all possible (user, item, item) triplets for pairwise learning. In addition, $\ell(\cdot)$ denotes the local loss for the triplet (u,i,j). We describe the formal definition of the loss function of the BPR in Section 2.4.4. In the following sections, we denote $\ell(f(u,i) - f(u,j))$ as ℓ_{uij} for simplicity.

A relevance predictor or scoring function, which minimizes 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 a statistical estimation problem and aim to estimate the ideal loss functions using only the biased interaction feedback.

2.4 Summary of Existing Methods

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

[2] $\mathbb{I}\{\cdot\}$ is the indicator function.

2.4.1 *Weighted Matrix Factorization (WMF).*

WMF is a commonly used method for an implicit feedback recommendation [4] and relies on the following estimator, which uniformly downweights the non-interacted data:

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

where $c \leq 1$ is a hyperparameter determining the weight of the non-interacted data. When user or item features are unavailable, a positive constant c is set for all non-interacted data, and these weights are tuned using a cross-validation. Although this is the standard baseline model for implicit feedback [11], [15] showed that the loss function used in WMF has a bias against the ideal pointwise loss, i.e., for a given f,

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

As stated above, WMF optimizes a biased loss function because it treats non-interacted items as low-confidence negative feedback. However, non-interacted feedback does not always imply a negative feedback, and thus the loss function of WMF is unsuitable for optimizing the metric of interest in Eq. (3).

2.4.2 *Exposure Matrix Factorization (ExpoMF).*

To address the positive-unlabeled problem of implicit feedback, ExpoMF adopts a loss function that differs from that of the WMF model. The method is constructed based on the following latent probabilistic model [11, 19].

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

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

, where $Bern(\cdot)$ is a Bernoulli distribution. Here, λ_U, λ_V, and λ_y are hyperparameters for prior distributions. The model also assumes that

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

which is consistent with our formulation.

ExpoMF utilizes an EM-based iterative algorithm to derive user–item matrices. During the E-step, ExpoMF estimates the posterior exposure probability $\theta'_{u,i} = \mathbb{E}\left[O_{u,i} \mid Y_{u,i}\right]$, and during the M-step, the model parameters are updated to maximize the log-likelihood[3]. When the true posterior exposure probabilities are given, the M-step can be viewed as optimizing the following weighted loss function:

$$\widehat{\mathcal{L}}_{ExpoMF}(f) = \frac{1}{|\mathcal{D}_{point}|} \sum_{(u,i) \in \mathcal{D}_{point}} \theta'_{u,i}\left(Y_{u,i}\delta_{u,i}^{(1)} + \left(1 - Y_{u,i}\right)\delta_{u,i}^{(0)}\right) \tag{7}$$

In Eq. (7), the posterior exposure probability represents the confidence regarding the amount of relevance information included in an interaction indicator $Y_{u,i}$. Therefore, the loss function of the ExpoMF model was designed to consider the local loss of the user–item pairs where the user has seen the item (i.e., $O_{u,i} = 1$) because if the user has been aware of the item, an interaction can be viewed as representing the relevance information ($O = 1 \Rightarrow Y = R$). Thus, this approach aims to solve the positive-unlabeled problem by selecting high confidence negative samples from numerous unlabeled samples.

[3] The detailed procedure of this is described in Section 3.3 of [11].

However, as discussed in [15], the loss function in Eq. (7), optimized through the M-step of the ExpoMF algorithm, is also biased against the ideal pointwise loss, i.e., for a given f,

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

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

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 [15]. The unbiased loss function is defined as follows:

$$\widehat{\mathcal{L}}_{\text{Rel-MF}}(f) = \frac{1}{|\mathcal{D}_{point}|} \sum_{(u,i) \in \mathcal{D}_{point}} \frac{Y_{u,i}}{\theta_{u,i}} \delta_{u,i}^{(1)} + \left(1 - \frac{Y_{u,i}}{\theta_{u,i}}\right) \delta_{u,i}^{(0)} \tag{8}$$

As shown in Proposition 4.3 of [15], the estimator defined in Eq. (8) satisfies the unbiasedness toward the ideal pointwise loss, i.e., for any given f,

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

This unbiasedness is desirable to optimize the pointwise metric in Eq. (4). However, in studies on top-N recommendations, a pairwise approach that considers the relevance orders of a given pair of items was found to be more suitable for the task, and empirically outperforms the pointwise approach [8, 12, 17]. Moreover, pairwise algorithms such as LambdaMART are also preferred because of their practical performance achieved in studies on learning-to-rank approaches [7]. Therefore, an unbiased pairwise learning method should be developed to further improve the recommendation quality from biased user feedback. In addition, empirical studies on evaluating pointwise and pairwise approaches with biased implicit feedback are needed to compare their empirical performance under MNAR situations.

2.4.4 Bayesian Personalized Ranking (BPR). BPR is a well-established pairwise algorithm for top-N recommendations based on implicit feedback. It models users' preferences over two items, wherein the interaction of one item is observed and that of the other is not. BPR assumes that the interacted items should be ranked higher than all non-interacted items and optimizes the following loss function to obtain the latent factors:

$$\widehat{\mathcal{L}}_{BPR}(f) = \frac{1}{|\mathcal{D}_{pair}|} \sum_{(u,i,j) \in \mathcal{D}_{pair}} Y_{u,i}(1 - Y_{u,j}) \ell_{uij} \tag{9}$$

where $\ell(\cdot) = -\ln(\sigma(\cdot))$ is generally used[4] In addition, $f(u,i) = U_u^T V_i$ is the predicted score for (u,i) where U_u and V_j are low-dimensional user and item factors, respectively.

The BPR algorithm has shown promising empirical results on ranking tasks [12]. However, the following proposition states that the loss function of BPR is in reality biased against the ideal pairwise loss function.

[4]Here, $\sigma(\cdot)$ is the sigmoid function

Proposition 2.3. *(Bias of BPR) The estimator optimized in BPR is biased against the ideal pairwise loss in Eq. (5), i.e., for some given f,*

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

Proof.

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

Thus, we obtain the following:

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

For $\widehat{\mathcal{L}}_{BPR}(f)$ to be theoretically unbiased, $\theta_{u,i} - 1 = 0 \Rightarrow \theta_{u,i} = 1$ and $1 - \theta_{u,i}\theta_{u,j} = 0 \Rightarrow \theta_{u,i}\theta_{u,j} = 1$ must be satisfied for all pairs from the last equation. However, in our setting, $\theta_{u,i}$ and $\theta_{u,j}$ can take different values among user–item pairs, and thus these conditions are not always satisfied. Therefore, the loss function of BPR is biased toward the ideal pairwise loss function. □

As stated in Proposition 2.3, the loss function of the BPR model is biased toward the ideal pairwise loss 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 a novel estimator for the ideal pairwise loss, which is inspired by the *inverse propensity score* estimator in causal inference [5, 13, 14]. We then provide the essential statistical properties of our estimator and describe the resulting *Unbiased Bayesian Personalized Ranking* algorithm.

3.1 Proposed Estimator

We start by formally defining the parameter called the *propensity score* which palys a critical role in causal inference [5, 13, 14].

Definition 3.1. (Propensity score) Propensity score of user–item pair (u,i) is

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

The proposed estimator is then defined using the propensity score.

Definition 3.2. (Unbiased estimator) When propensity scores are given, the unbiased pairwise loss function is defined as:

$$\widehat{\mathcal{L}}_{UB}^{pair}(f) = \frac{1}{|\mathcal{D}_{pair}|} \sum_{(u,i,j) \in \mathcal{D}_{pair}} \frac{Y_{u,i}}{\theta_{u,i}} \left(1 - \frac{Y_{u,j}}{\theta_{u,j}}\right) \ell_{uij} \tag{10}$$

where UB stands for Unbiased.

The proposed estimator weights each data by the inverse of their propensity, and can also be represented in the following form:

$$\frac{1}{|\mathcal{D}_{pair}|} \sum_{(u,i,j)\in\mathcal{D}_{pair}} \frac{Y_{u,i}}{\theta_{u,i}} \left(Y_{u,j} \left(1 - \frac{1}{\theta_{u,j}} \right) + \left(1 - Y_{u,j} \right) \right) \ell_{uij} \quad (11)$$

Eq. (11) can be interpreted 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 truly unbiased against the ideal pairwise loss:

PROPOSITION 3.3. *The unbiased pairwise loss in Eq. (10) is statistically unbiased against the ideal pairwise loss in Eq. (5), i.e., fo any given f:*

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

PROOF.

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

□

Proposition 3.3 shows that the proposed unbiased pairwise loss function derived via inverse propensity weighting is valid for debiasing the MNAR implicit feedback.

We also derive the variance of the unbiased estimator as follows:

THEOREM 3.4. *(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 a scoring function f, the variance of the unbiased pairwise loss is:*

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

where

$$\begin{aligned}
v_{uij} &= \gamma_{u,i}(1 - 2\gamma_{u,j}) \left(\frac{1}{\theta_{u,i}} - \gamma_{u,i} \right) + \gamma_{u,i}\gamma_{u,j} \left(\frac{1}{\theta_{u,i}\theta_{u,j}} - \gamma_{u,i}\gamma_{u,j} \right) \\
w_{uijk} &= \left(\frac{1}{\theta_{u,i}} - \gamma_{u,i} \right) \gamma_{u,i}(1 - \gamma_{u,j})(1 - \gamma_{u,k})
\end{aligned}$$

We omit the proof due to space constraints.

The RHS of the variance depends on the inverse of the product of two propensity scores. Thus, it can be loose, especially for tail items having low exposure probability. Moreover, the unbiased estimator in Eq. (10) can take large negative values by its definition, which causes the severe variance issue in some cases. Based on these implications, we propose to utilize the following *non-negative* estimator inspired by the work in positive-unlabeled learning [10] as a practical variance control technique.

Algorithm 1 Unbiased Bayesian Personalized Ranking (UBPR)

Input: observed interaction data $\{Y_{u,i}\}$; learning_rate μ; mini-batch size M; regularization parameter λ.
Output: learned model parameters $\boldsymbol{U}, \boldsymbol{V}$

```
Initialize U, V, and estimate propensity scores {θ_{u,i}}
repeat
    Sample a size M of mini-batch data from D
    Compute loss = L̂_UB^pair(f) + λ(‖U‖_2^2 + ‖V‖_2^2)
    Update user latent factor by U ← U + μ · ∂loss/∂U
    Update item latent factor by V ← V + μ · ∂loss/∂V
until convergence;
return U, V
```

Definition 3.5. (Non-negative estimator) When the positive constant $\beta \geq 0$ is given, then, the non-negative estimator is defined as

$$\widehat{\mathcal{L}}_{non\text{-}neg}^{pair}(f) = \frac{1}{|\mathcal{D}_{pair}|} \sum_{(u,i,j)\in\mathcal{D}_{pair}} \max\left\{ \frac{Y_{u,i}}{\theta_{u,i}} \left(1 - \frac{Y_{u,j}}{\theta_{u,j}} \right) \ell_{uij}, -\beta \right\} \quad (12)$$

The non-negative estimator reduces the variance of the estimator at the cost of introducing some bias. β clips a large negative value and controls the bias-variance trade-off. We use this non-negative form of estimator in the experimental part as an alternative to using the unbiased estimator and tune the value of β using the validation set.

3.2 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 the ideal pairwise loss function:

$$\boldsymbol{U}, \boldsymbol{V} = \underset{\boldsymbol{U},\boldsymbol{V}}{\arg\min}\ \widehat{\mathcal{L}}_{UB}^{pair}(f) + \lambda \left(\|\boldsymbol{U}\|_2^2 + \|\boldsymbol{V}\|_2^2 \right)$$

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. Note that our learning procedure is general and can be used in combination with any complex pairwise algorithm such as *neural collaborative ranking* [17] or *dual neural personalized ranking* [8].

Table 1: Statistics of the datasets after preprocessing.

Datasets	#users	#items	#interactions	sparsity
Yahoo! R3	15,400	1,000	123,812	99.10%
Coat	290	300	2,159	97.24%

Note: **#interactions** is the sum of positive interactions, i.e., $\sum_{u,i} Y_{u,i}$. **sparsity** is the percentage of positive interactions among all user-item pairs, i.e., $\sum_{u,i} Y_{u,i} / |\mathcal{D}_{point}|$.

4 EMPIRICAL EVALUATIONS

In this section, we demonstrate the effectiveness of the proposed UBPR and investigate its empirical properties. [5]

4.1 Experimental Setup

Datasets and preprocessing. We used Yahoo! R3[6] and Coat[7]. These are the explicit feedback data with five-star ratings, and we can consequently utilize the ground truth relevance information in the test set. Besides, they contain training and test sets with different user-item distributions, therefore include the MNAR problem inherently. These two datasets enable the evaluation of recommenders with the ground-truth ranking metric in Eq. (3) in the test set while creating a biased implicit feedback setting in the training set.

For both datasets, we employed the following preprocessing procedure. Table 1 shows the dataset statistics after the preprocessing.

(1) Transform five-star ratings into relevance parameter using the following methodology in information retrieval [2].

$$\gamma_{u,i} = \epsilon + (1 - \epsilon) \frac{2^r - 1}{2^{r_{max}} - 1}$$

where $r \in \{1, 2, 3, 4, 5\}$ denotes a five-star rating and r_{max} is the maximum possible rating, which is 5 in our case. $\epsilon \in [0, 1]$ controls the noise level in the grade information. We apply $\epsilon = 0.1$ for the training datasets and $\epsilon = 0$ for the test datasets.

(2) Sample the binary relevance variable $R_{u,i}$ by performing the following Bernoulli sampling:

$$R_{u,i} \sim Bern(\gamma_{u,i}), \ \forall (u, i) \in \mathcal{D}_{point}$$

where $Bern(\cdot)$ denotes a Bernoulli distribution.

(3) Define the exposure variable for a user–item pair as follows:

$$O_{u,i} = \begin{cases} 1 & \text{(if item } i \text{ is rated by user } u\text{)} \\ 0 & \text{(if item } i \text{ is not rated by user } u\text{)} \end{cases}$$

(4) The training set is now given as $\{(u, i, Y_{u,i})\}_{(u,i) \in \mathcal{D}_{point}}$ where $Y_{u,i} = O_{u,i} \cdot R_{u,i}$ is the observed implicit interaction for (u, i). Note that $R_{u,i}$ and $O_{u,i}$ were unobservable in our setting and were not used for training the recommenders.

Baselines and the proposed method. Herein, we describe the existing baselines and proposed methods compared in the real-world experiment.

- **Item popularity model (ItemPop)**: This method always recommends the k-most clicked items in the training set for all users; thus, it is not personalized.
- **Weighted matrix factorization (WMF) [4]:** WMF is a basic baseline pointwise recommendation model for implicit recommendation and is described in Section 2.4.1.
- **Exposure matrix factorization (ExpoMF) [11]:** ExpoMF is based on the latent probabilistic model using exposure variables and described in Section 2.4.2.
- **Relevance matrix factorization (Rel-MF) [15]:** Rel-MF is based on the same latent factor model as ExpoMF, but it updates its user–item factors by minimizing the unbiased estimator for the ideal pointwise loss in Eq. (8). We used the clipping estimator introduced in [15] as a variance reduction technique for this method.
- **Bayesian personalized ranking (BPR) [12]:** BPR is the most basic pairwise recommendation model used with implicit feedback and is described in Section 2.4.4.
- **Unbiased Bayesian personalized ranking (UBPR):** UBPR is our proposed unbiased pairwise recommendation algorithm with implicit feedback. It updates its user–item factors by minimizing the non-negative estimator in Eq. (12) to address the positive-unlabeled and MNAR issues with implicit feedback.

For Rel-MF and UBPR, we estimated the propensity score based on the following relative item popularity, which is used in the unbiased evaluation method developed in [23]:

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

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

Hyperparameter tuning. For both datasets, the originals are divided into the training and test sets. We randomly sampled 10% of the training set as the validation set. Using the validation set, we tuned the dimensions of the user-item latent factors within the range of $\{100, 120, \dots, 300\}$ and L2 regularization hyperparameter within the range of $[10^{-7}, 10^{-3}]$ for all methods. For UBPR, we tuned the trade-off hyperparameter β within the range of $[0.01, 10]$. For the baselines and the proposed method, we determined the best set of hyperparameters using *Optuna* [1] with an adaptive hyperparameter search algorithm. The model parameters of all methods are optimized using the *Adam* optimizer [9] with an initial learning rate of 0.001 and a mini-batch size of 256. During mini-batch training, we randomly sampled 1 unlabeled item for each positive item for WMF and Rel-MF. Moreover, we randomly sampled 10 unlabeled items for each positive item for BPR and UBPR.

Evaluation metrics. We used DCG, recall, and MAP to evaluate the ranking performance. We report results with varying values of $K \in \{3, 5, 8\}$. The definitions of the ranking metrics are as follows[8]:

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

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

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

[5] Our code is available at ***https://github.com/usaito/unbiased-pairwise-rec***.
[6] http://webscope.sandbox.yahoo.com/
[7] https://www.cs.cornell.edu/ schnabts/mnar/
[8] We follow the definitions of [15, 23]

Table 2: Comparing ranking performance of alternative methods on Yahoo! R3 and Coat datasets.

		Ranking Metrics								
		DCG@K			*Recall@K*			*MAP@K*		
Datasets	Methods	$K=3$	$K=5$	$K=8$	$K=3$	$K=5$	$K=8$	$K=3$	$K=5$	$K=8$
Yahoo! R3	ItemPop	0.04569	0.05904	0.07633	0.05158	0.08056	0.12869	0.03411	0.04183	0.05115
	WMF	0.08353	0.09687	0.10821	0.09039	0.11890	0.15036	0.06827	0.07675	0.08354
	ExpoMF	0.09375	0.10488	0.11539	0.09977	0.12356	0.15264	0.07832	0.08584	0.09231
	Rel-MF	0.08576	0.09861	0.10998	0.09249	0.11993	0.15150	0.07065	0.07887	0.08574
	BPR	0.09020	0.10361	0.11367	0.09709	0.12539	0.15331	0.07409	0.08267	0.08881
	UBPR (ours)	**0.09458**	**0.10762**	**0.11720**	**0.10191**	**0.12956**	**0.15612**	**0.07862**	**0.08723**	**0.09326**
Coat	ItemPop	0.06488	0.08563	0.11539	0.06863	0.11304	0.19545	0.05255	0.06721	0.08445
	WMF	0.09938	0.12026	0.14233	0.10917	0.15392	0.21560	0.08351	0.09778	0.11177
	ExpoMF	0.09101	0.10430	0.13157	0.09714	0.12508	0.20153	0.08348	0.09168	0.10742
	Rel-MF	0.10391	0.12395	0.14733	0.11495	0.15811	**0.22327**	0.08540	0.09911	0.11417
	BPR	0.10017	0.12208	0.14394	0.10893	0.15538	0.21584	0.08447	0.09927	0.11280
	UBPR (ours)	**0.10736**	**0.12679**	**0.14937**	**0.11868**	**0.16041**	0.22316	**0.08821**	**0.10115**	**0.11535**

Note: The table reports the ranking metrics averaged over 10 different initializations. The bold font indicates the best performing method for each metric and dataset.

Table 3: Comparing ranking performance of alternative methods for cold-start users and rare items on Yahoo! R3.

		Ranking Metrics								
		DCG@K			*Recall@K*			*MAP@K*		
Subsets	Methods	$K=3$	$K=5$	$K=8$	$K=3$	$K=5$	$K=8$	$K=3$	$K=5$	$K=8$
Cold-start users	WMF	0.08431	0.09676	0.10804	0.09006	0.11675	0.14805	0.07002	0.07805	0.08474
	ExpoMF	0.08524	0.09975	0.11223	0.09115	0.12182	0.15637	0.07015	0.07882	0.08633
	Rel-MF	0.08739	0.09920	0.11071	0.09358	0.11881	0.15076	0.07318	0.08078	0.08754
	BPR	0.09161	0.10395	0.11424	0.09855	0.12453	0.15317	0.07608	0.08406	0.09030
	UBPR (ours)	**0.09503**	**0.10785**	**0.11702**	**0.10209**	**0.12937**	**0.15484**	**0.08091**	**0.08941**	**0.09514**
Rare items	WMF	0.04243	0.05383	0.06033	0.04714	0.07148	0.08929	0.03206	0.03905	0.04288
	ExpoMF	0.05070	0.05860	0.06419	0.05496	0.07178	0.08708	0.04163	0.04664	0.05002
	Rel-MF	0.04456	0.05491	0.06159	0.04917	0.07132	0.08953	0.03475	0.04116	0.04509
	BPR	0.04693	0.05673	0.06271	0.05185	0.07278	0.08915	0.03679	0.04281	0.04644
	UBPR (ours)	**0.05215**	**0.06170**	**0.06629**	**0.05708**	**0.07736**	**0.08982**	**0.04121**	**0.04725**	**0.05020**

Note: The table reports the ranking metrics averaged over 10 different initializations. The cold-start users are defined as users with fewer than six clicks in the training sets. The rare items are defined as the items with less than 100 clicks in the training sets. The bold fonts represent the best performing method for each metric.

where $\mathcal{I}_u^{test}$ is a set of items rated by user u in the test set, and $\mathcal{S}_u^* = \sum_{i \in \mathcal{I}_u^{test}} R_{u,i}$, which are the special cases of the performance metric of interest in Eq. (3).

4.2 Results

Herein, we present the experimental findings.

Table 2 summarizes the ranking performance of all methods compared on both the Yahoo! R3 and Coat datasets. For both datasets, our UBPR algorithm outperforms the other baseline methods under almost all settings by utilizing biased implicit feedback in a theoretically grounded manner. The results suggest that the recommendation quality can be significantly improved using biased implicit feedback in the proper manner.

Table 3 reports the ranking performance of the compared methods for different subsets of Yahoo! R3. Specifically, we report the performances of the methods for cold-start users and rare items of the Yahoo! R3 dataset. We defined items with less than 100 clicks in the training sets as rare (approximately 70 % of all items are rare). In addition, we defined users having less than six clicks in the training sets as cold-start users (approximately 50 % of all users are cold). It should be noted that it is more difficult to recommend relevant items to cold-start users and to find relevant items among rare items because there are few positive labels for the cold-start users and rare items in the training set. We do not present the results for the Coat data in Table 3 because the user activeness and item popularity are not largely divergent.

Table 3 shows that for both cold-start users and rare items, UBPR outperforms the best baselines in all ranking metrics. The improvements achieved by the proposed method for cold-start users and rare items are greater than those observed for all data reported in Table 2 under almost all settings. The reason behind these results is the inverse propensity weighting employed by the loss function used in our method. It up-weights the loss for the (user, item, item) triplets with fewer clicks in the training data, which helps the method find positive user-item relationships among unlabeled data. In real-world scenarios, the performance for cold-start users and rare items is critical to achieving a goal personalized recommendation. Therefore, the results demonstrate the real-world applicability of our approach.

Overall, the proposed unbiased pairwise method outperforms the other baselines on both datasets in terms of the ranking performance. The results validate that theoretically refined unbiased pairwise learning is valid for addressing the positive-unlabeled and MNAR issues. Moreover, the proposed method performed the best for cold-start users and rare items based on the power of the inverse propensity weighting, which highlights its advantages in real-world applications. Given these results, we conclude that the proposed unbiased pairwise learning framework is a suitable choice for constructing recommender systems with only biased implicit feedback.

5 CONCLUSION

In this study, we explored the *positive-unlabeled* and *missing-not-at-random* problems of implicit feedback. To solve these issues, we first modeled the recommendation problem with implicit feedback using relevance and exposure variables. Subsequently, we proposed an unbiased estimator for the ideal pairwise loss function and a corresponding algorithm, *Unbiased Bayesian Personalized Ranking*. We also provided critical theoretical properties of the proposed unbiased estimator, including unbiasedness and variance. An empirical evaluation using real-world datasets demonstrated that the proposed algorithm outperforms the current state-of-the-art baselines when both positive-unlabeled and MNAR problems are present. These theoretical and empirical findings suggest that our unbiased pairwise learning framework is a suitable choice for maximizing the relevance in interactive systems from biased implicit feedback.

ACKNOWLEDGEMENTS

We thank Minato Sato and the anonymous reviewers for their helpful comments, discussions, and advice.

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* (Anchorage, AK, USA) *(KDD '19)*. ACM, New York, NY, USA, 2623–2631. https://doi.org/10.1145/3292500.3330701

[2] Olivier Chapelle, Donald Metlzer, Ya Zhang, and Pierre Grinspan. 2009. Expected reciprocal rank for graded relevance. In *Proceedings of the 18th ACM conference on Information and knowledge management*. 621–630.

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

[4] 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.

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

[6] 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.

[7] Shubhra Kanti Karmaker Santu, Parikshit Sondhi, and ChengXiang Zhai. 2017. On application of learning to rank for e-commerce search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 475–484.

[8] Seunghyeon Kim, Jongwuk Lee, and Hyunjung Shim. 2019. Dual Neural Personalized Ranking. In *The World Wide Web Conference*. ACM, 863–873.

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

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

[11] 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.

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

[13] 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.

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

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

[16] 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

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

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

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

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

[21] 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.

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

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

[24] Zainab Zolaktaf, Reza Babanezhad, and Rachel Pottinger. 2018. A Generic Top-N Recommendation Framework For Trading-off Accuracy, Novelty, and Coverage. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*. IEEE, 149–160.

A OMITTED PROOFS

A.1 Proof and Statement of a Technical Lemma

Here we state a technical lemma used to prove the later theorem.

Lemma A.1. *(Covariance) Given a scoring set $\widehat{X}$, 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 $u \in \mathcal{U}$ 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

$$Cov(Z_{uij}, Z_{uik}) = \left(\frac{1}{\theta_{u,i}} - \gamma_{u,i}\right)\gamma_{u,i}(1-\gamma_{u,j})(1-\gamma_{u,k})\ell_{uik}\ell_{uij} \tag{13}$$

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 3.3, 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.

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

□

It should also be note that for user u and items i, j, k, l (where $i \neq j \neq k \neq l$), $Cov(Z_{uij}, Z_{ukl}) = 0$.

A.2 Proof of Theorem 3.4

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 $\mathbb{V}\left(Z_{uij}\right)$ as

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

By Proposition 3.3, $(b) = \left(\gamma_{u,i}(1-\gamma_{u,j})\ell_{u,i,j}\right)^2 = (\gamma_{u,i}^2 - 2\gamma_{u,i}^2\gamma_{u,j} + \gamma_{u,j}^2)\ell_{uij}^2$. Next,

$$Z_{uij}^2 = \frac{Y_{u,i}}{\theta_{u,i}^2}\left(1 - \frac{2Y_{u,j}}{\theta_{u,j}} + \frac{Y_{u,j}}{\theta_{u,j}^2}\right)\ell_{uij}^2 \tag{14}$$

where $Y_{u,i}^2 = Y_{u,i}$ and $Y_{u,j}^2 = Y_{u,j}$. The expectation of the RHS of Eq. (14) is

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

Therefore,

$$\begin{aligned} &(a) - (b) \\ &= \frac{\gamma_{u,i}}{\theta_{u,i}}\left(1 - 2\gamma_{u,j} + \frac{\gamma_{u,j}}{\theta_{u,j}}\right)\ell_{uij}^2 - \left(\gamma_{u,i}^2 - 2\gamma_{u,i}^2\gamma_{u,j} + \gamma_{u,j}^2\right)\ell_{uij}^2 \\ &= \left[\gamma_{u,i}(1-2\gamma_{u,j})\left(\frac{1}{\theta_{u,i}} - \gamma_{u,i}\right) + \gamma_{u,i}\gamma_{u,j}\left(\frac{1}{\theta_{u,i}\theta_{u,j}} - \gamma_{u,i}\gamma_{u,j}\right)\right]\ell_{uij}^2 \end{aligned} \tag{15}$$

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

$$\begin{aligned} \mathbb{V}\left(\widehat{\mathcal{L}}_{UB}^{pair}(f)\right) &= \frac{1}{|\mathcal{D}_{pair}|^2}\sum_{(u,i,j)\in\mathcal{D}_{pair}} \mathbb{V}\left(Z_{uij}\right) \\ &+ \frac{1}{|\mathcal{D}_{pair}|^2}\sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{I}}\sum_{\substack{(j,k)\in\mathcal{I}\times\mathcal{I} \\ i\neq j\neq k}} Cov(Z_{uij}, Z_{uik}) \end{aligned} \tag{16}$$

Combining Eq. (13), Eq. (15), and Eq. (16) completes the proof. □