2024 IEEE 40th International Conference on Data Engineering (ICDE)

Modeling User Attention in Music Recommendation

Sunhao Dai†, Ninglu Shao†, Jieming Zhu‡, Xiao Zhang†,
Zhenhua Dong‡, Jun Xu* †, Quanyu Dai* ‡, Ji-Rong Wen†
†Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China
‡Noah's Ark Lab, Huawei, Shenzhen, China
sunhaodai@ruc.edu.cn, ninglu_shao@ruc.edu.cn, jamie.zhu@huawei.com, zhangx89@ruc.edu.cn,
dongzhenhua@huawei.com, junxu@ruc.edu.cn, daiquanyu@huawei.com, jrwen@ruc.edu.cn

2024 IEEE 40th International Conference on Data Engineering (ICDE) | 979-8-3503-1715-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDE60146.2024.00064

***Abstract*—With the popularity of online music services, personalized music recommendation has garnered much research interest. Recommendation models are typically trained on datasets constructed from user feedback, which includes both the active feedback (*e.g.*, clicking the *Like* or *Skip* buttons) and passive feedback (*e.g.*, auto-play), with passive feedback comprising the majority. Due to the unavailability of user attention, the massive amount of passive feedback is unreliable, significantly compromising the quality of the training data. How to estimate the user's attention on the target music has become a critical problem in music recommendation. Heuristic methods such as exponential decay and negative sampling have been proposed. However, they either neglect the sequential dependencies between feedback actions or utilize only a small fraction of passive samples, leading to inaccurate and biased attention estimation. In this paper, we naturally propose modeling user attention prediction as a positive-unlabeled (PU) learning problem, where active feedback is treated as positive samples and passive feedback is treated as unlabeled samples, as we can only ensure that the user's attention is focused when she provides active feedback. Then we propose an extended PU-learning model with sequential dependencies, called UAE, which contains an unbiased user attention estimator and an unbiased propensity estimator. Subsequently, a joint learning algorithm is developed in which the attention and propensity estimators are optimized in alternating fashion. Theoretical analysis shows the unbiasedness and variance of the attention estimator and the propensity estimator. Extensive experiments on two large-scale datasets demonstrate the proposed UAE's effectiveness and generality in enhancing downstream music recommendation. One week online A/B testing on Huawei Music App manifests that UAE can significantly increase the users' play count and time over 2%, further demonstrating the effectiveness of UAE in real-world music recommendation products.**

***Index Terms*—music recommendation, user attention modeling, PU-learning, unbiased prediction**

I. Introduction

Recently, music recommendation has attracted more and more research interest from both academia and industry [1], [2], [3], [4], [5]. Great amounts of research efforts have been devoted to the model architecture design [6], [7], [8]. Meanwhile, little attention has been paid to the fine-grained analysis and construction of the training data, even though its quality plays an essential role in recommendation performance.

As shown in Figure 1, when a user accesses a music recommender system, the system recommends a song playlist.

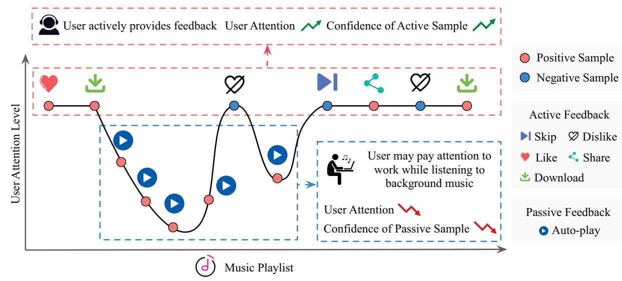


Fig. 1. Illustrating our motivation on why modeling user attention is pivotal for music recommendation. When users actively provide feedback, their attention is fully engaged, resulting in high-confidence active samples. However, when users offer passive feedback, such as listening to music in the background while focusing on other tasks, the reliability of this feedback diminishes significantly.

The user may give multiple types of *active feedback* during her interaction with the recommended song list, including clicking the "Like", "Share", "Download", "Skip", and "Dislike" buttons. Meanwhile, since the music apps are designed to automatically play the next songs, the system may also receive *passive feedback* such as "Auto-play". Currently, the training data in music recommendation is constructed based on user feedback with simple rules [3], [5], [9], [10]. For instance, the "Like", "Share", "Download", and auto-played songs are directly considered as positive samples, while the "Skip" and "Dislike" songs are negative samples. Though the active feedback clearly reflects the user preference, the reliability of passive feedback is much lower. The major reason is that the user's attention may not be on the target song when giving passive feedback. For example, the user may listen to the songs as background music or even not listen, while the songs are still auto-playing.

As a consequence, directly using passive feedback actions to define sample labels is unreasonable due to the unavailability of the user attention. The naively defined sample labels pose great challenges to the recommendation performance with existing models. How to accurately model the user attention on the target music when she gave the passive feedback has become a critical problem in industries and other data engineering areas. The key difficulty lies in that we only have access to a proportion of positive samples for modeling user attention. That is, when a user gives active feedback, we are sure that her attention is on the target music. When a user gives passive feedback, however, we are uncertain whether her attention is

*Corresponding author. Work partially done at Engineering Research Center of Next-Generation Intelligent Search and Recommendation, Ministry of Education.

2375-026X/24/$31.00 ©2024 IEEE
DOI 10.1109/ICDE60146.2024.00064

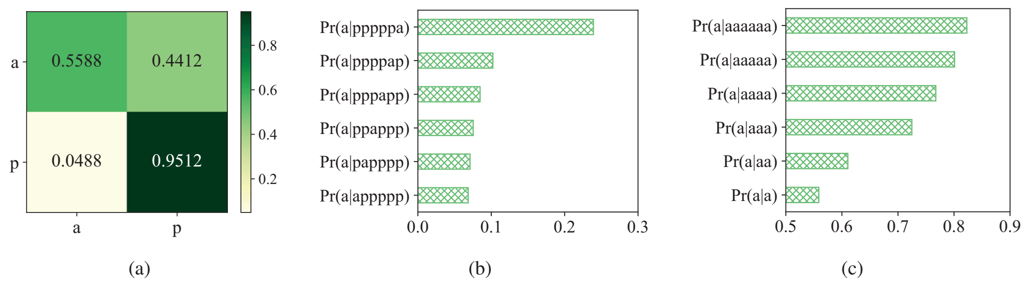


Fig. 2. Statistics on the transition probabilities of user feedback types on Huawei Music App. 'a' denotes an active action and 'p' denotes a passive action. (a): x-axis is the type of next feedback, and y-axis is the type of current feedback. The overall probabilities of appearing active and passive actions are 0.0876 and 0.9124, respectively. The number in each cell is the corresponding transition probability. (b): the probability of giving an active action w.r.t. different near history feedback sequences of length 6; (c): the probability of giving an active action w.r.t. different lengths of near history active feedback sequences.

on the target music. These make attention prediction a natural positive-unlabeled (PU) learning problem.

To model user attention, existing methods have predominantly design heuristics to circumvent the unlabeled data problem. For example, Spotify [11] assumed that user attention exponentially decreases with time and first proposed an exponential decay function to predict user attention. Zhang *et al.* [12] proposed a learning-based model with a heuristic negative sampling method for attention prediction. The sampling heuristic is, for example, a negative attention label is sampled only when a user continues the passive actions for more than 10 songs. However, in real scenarios, users may lose their attention at any time during the listening events [13]. Consequently, these existing methods often suffer from bias, and their performance may be suboptimal due to overlooking the positive-unlabeled nature of user attention in music streaming services.

On the other hand, classical PU-learning method can not be directly applied, since the uniform assumption[1] does not hold in the attention prediction task. We use some statistics from Huawei Music App for illustration. Figure 2(a) shows the transition probability of the active and passive user feedback. We can observe that a user has a marginal probability of 8.76% to give an active action. However, the probability dramatically increases to 55.88% if her last action is active, and decreases to 4.88% if the last is passive. Similar patterns can be found among the passive feedback actions. Moreover, Figure 2(b) and 2(c) further illustrate the sequential dependencies that the probability of giving an active action will increase if more active actions occurred in the near history feedback sequence. This phenomenon is caused by the fact that users are more likely to focus on the music if they have done some active actions on the music App in the past few songs. Therefore, user feedback has complicated sequential dependencies. However, classical PU-learning methods cannot characterize such fine-grained sequential dependencies of the user feedback action, and thus are unable to perform accurate user attention estimation.

In this paper, we aim to achieve an unbiased attention estimation, and apply it to improve the performance of the downstream music recommendation tasks. We naturally propose to formulate the attention prediction as a problem of sequential PU-learning, where the active feedback actions are labeled positive samples and the passive feedback actions are unlabeled samples. In this task, the occurrence of active feedback is not only influenced by the local features but also the history of features and feedback actions. Thus, we propose an Unbiased Attention Estimator (denoted as **UAE**) by extending the Empirical-Risk-Minimization (ERM) based PU-learning methods with sequential dependencies, which re-weights the active actions (positive samples) with sequential propensities and treats all the passive actions (unlabeled samples) as negative. The remaining challenge is how to obtain sequential propensity scores, as the true propensities are unknown in this task. From a dual perspective, we further propose an unbiased propensity estimator. To realize the above ideas, we design a GRU-based architecture with two output logits to respectively estimate the sequential propensities and user attention probabilities, and develop an alternating optimization algorithm to learn the model parameters. The learned attention model can be used to quantify the reliability of the passive training samples for the downstream music recommendation task, and thus can be leveraged to improve a series of recommendation models.

The main contributions of the paper are concluded as follows:

(1) This is the first work that demonstrates the user attention estimation problem from the viewpoint of PU-learning and analyzes the sequential dependencies of user attention.

(2) We propose UAE, which is the first unbiased framework for attention estimation based on ERM-based PU-learning. We further design an alternating learning algorithm for optimizing the two unbiased estimators in UAE.

(3) We provide a theoretical analysis of our proposed UAE from expectation, variance, and bias perspectives.

(4) Extensive offline experiments on both large-scale industrial and public datasets show that UAE significantly improves SOTA recommender models and outperforms the baselines.

(5) Online A/B testing on Huawei Music App shows that UAE can significantly enhance the user engagement in terms of both play count and play time.

[1]The uniform assumption is that the labeling mechanism of each instance (*i.e.*, the propensity of selecting a labeled positive example from the complete set of positive examples) is a uniform distribution or only depends on its local features [14], [15], [16], [17].

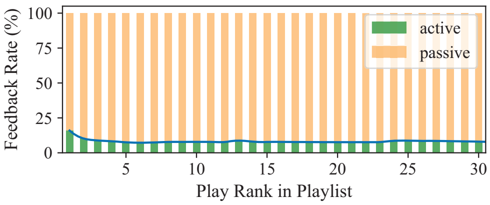


Fig. 3. User feedback rates w.r.t. the play rank of the recommended playlist. Statistics are conducted on a dataset collected from Huawei Music App.

TABLE I
RELATION AMONG THE VARIABLES IN THE PRESENCE OF MODELING USER ATTENTION IN MUSIC RECOMMENDATION.

User feedback	Feedback type e	User attention a	Feedback label y
Skip Dislike	1 (Active) 1 (Active)	1 (Positive)	0 (Negative)
Like Share Download	1 (Active) 1 (Active) 1 (Active)	1 (Positive)	1 (Positive)
Auto-play	0 (Passive)	? (Unknown)	1 **(Postive ?)**

II. Related Work

A. User Modeling in Music Recommendation

With the rise of online streaming services [18], [19], music recommender systems (MRS) have garnered significant interest from both research and industry communities [1], [2], [4], [5]. Extensive research has been dedicated to user modeling in music recommendation, with a focus on analyzing and modeling user behaviors within music services [20], [21], [22], [23], [24]. Moreover, much research efforts have been made to incorporate contextual information into user modeling, encompassing user-related contexts, such as activity, emotional state, and social relationships, as well as sensor data from devices, including location, current time, weather, and temperature, to further enhance recommendation performance [25], [26], [27], [28]. Recently, researchers have highlighted that in MRS, listening to music does not demand constant user attention, which underscores the significance of modeling user attention [13]. Spotify [11] was among the first to investigate user attention levels in music streaming sessions, employing an exponential decay attention model. Reza Aditya Permadi [29] also addressed the issue of label noise caused by the loss of user attention and proposed a heuristic cut-off method that considers played songs after an active user action as positive training samples. Dai *et al.* [10] proposed an adaptive label correction method to mitigate the noise and bias in music recommendation. Zhang *et al.* [12] introduced NDB, which aims to tackle biased online bandit feedback in music recommendation by modeling user attention. However, NDB relies on strong assumptions and results in biased attention estimation. In this paper, we aim to achieve an unbiased attention estimator and further use the estimated attention score as confidence for unreliable passive data to enhance the downstream music recommendation.

B. ERM-based PU-learning

Positive-unlabeled (PU) learning is a crucial branch of semi-supervised learning that focuses on training binary classifiers using only positive and unlabeled data, where the unlabeled samples could belong to either the positive or negative class [30], [31], [32], [33], [34]. Within PU learning, Empirical-Risk-Minimization (ERM) based methods play a significant role, aiming to obtain an unbiased empirical risk by incorporating appropriate weighting strategies [14], [15], [35], [36]. A series of works have been proposed under the ERM-based PU-learning framework, each making different assumptions. A common assumption is the Selected Completely At Random (SCAR) setting, where labeled examples are assumed to be selected completely at random and independently from their local features [15], [37]. Another assumption is the Selected At Random (SAR) setting, as proposed by Bekker *et al.*[38], which considers labeled examples with propensity depending on their local features. Recently, Saito *et al.*[39] extended PU-learning to address the missing-not-at-random problem for implicit feedback in the recommender system. Chang *et al.* [40] applied PU-learning to capture the temporal dynamics of negative samples from unlabeled data for link prediction. However, these existing assumptions may not be suitable for our attention estimation task, especially when dealing with sequential dependencies. Despite the power and extensive research on PU learning under various assumptions, how to adapt it to attention estimation in music recommendation is still an unsolved challenge.

III. PRELIMINARIES

As illustrated in Figure 1, we are confident about the user preference conveyed by the active feedback. For passive feedback, however, the confidence is very low because of the shift in user attention. Figure 3 shows the number of active/passive feedback w.r.t. the play rank. We can see: (1) the rate of active feedback decreases with the increase of the rank, showing that the users gradually lose their attention on the music App with time; (2) there are much more passive feedback actions than those active ones at all of the time. To realize the full potential of the large amounts of passive samples in music recommendation, it is necessary to model user attention.

A. Problem Formulation

Formally, suppose we have a set of users $\mathcal{U}$ and a set of songs $\mathcal{V}$. The interactions between the users and songs are collected and represented as a set $\mathcal{S} = \{(\mathbf{x}_i^1, e_i^1, y_i^1), \cdots, (\mathbf{x}_i^t, e_i^t, y_i^t), \cdots, (\mathbf{x}_i^{l_i}, e_i^{l_i}, y_i^{l_i})\}_{i=1}^N$, where each element is a chronologically ordered interaction session for a user and l_i is the length of the i-th session. $(\mathbf{x}_i^t, e_i^t, y_i^t)$ denotes the t-th listening event for the i-th session, $\mathbf{x}_i^t \in \mathcal{X} \subset \mathbb{R}^d$ denotes the feature vector for each event, including features from user u_i, song v_i^t and contexts. $e_i^t \in \{0, 1\}$ is an observable random variable indicating the u_i's feedback type on v_i^t, where $e_i^t = 0$ means passive feedback and $e_i^t = 1$ is active feedback. The feedback label $y_i^t \in \{0, 1\}$, where $y_i^t = 1$ means u_i gave positive feedback on song v_i^t (including passive feedback of auto-play and active feedback of like, share,

TABLE II

A SUMMARY OF NOTATIONS USED IN THIS PAPER.

Symbol	Description
N	Number of sessions in the dataset $\mathcal{S}$
l_i	The length of the i-th session
$\mathbf{x}_i^t$	The feature vector of a sample, containing user information, item information and context information
X_i^t	$X_i^t := [\mathbf{x}_i^1, \mathbf{x}_i^2, \cdots, \mathbf{x}_i^t]$ is the sequence of history and current features
e_i^t	$e_i^t \in \{0, 1\}$ is an observable random variable indicating user's feedback type, where $e_i^t = 0$ means passive feedback and 1 active
E_i^{t-1}	$E_i^{t-1} := [e_i^1, e_i^2, \cdots, e_i^{t-1}]$ is the sequence of user history observed feedback
y_i^t	$y_i^t \in \{0, 1\}$ is feedback label where $y_i^t = 1$ means user gave positive feedback on the recommended song, and $y_i^t = 0$ negative
α_i^t	$\alpha_i^t := \Pr(a_i^t = 1 \mid X_i^t)$ is the true user attention
p_i^t	$p_i^t := \Pr(e_i^t = 1 \mid X_i^t, E_i^{t-1}, a_i^t = 1)$ is the sequential propensity score
$\hat{\cdot}$	An estimate for $\cdot$

and download), and $y_i^t = 0$ means u_i gave negative feedback (including active feedback of skip and disklike). Please note that here we abstract multiple types of feedback with two binary variables e_i^t and y_i^t. Table I shows the feedback actions and the corresponding values of e_i^t and y_i^t.

One crucial problem with the feedback label y_i^t is that the observed value is unreliable when $e_i^t = 0$ (bold '?' in the far right column of Table I). The reason is that u_i's attention may not be on v_i^t if $e_i^t = 0$. We denote whether the user attention is on the music as a binary random variable $a_i^t \in \{0, 1\}$ where $a_i^t = 1$ means u_i is focus on v_i^t, and 0 otherwise. Please note that a_i^t is the true attention indicator of user attention. Obviously, $e_i^t = 1$ means $a_i^t = 1$, as the active feedback provided by the user certainly indicates a focus of attention on the music. However, we don't know the value of a_i^t when $e_i^t = 0$, *i.e.*, a_i^t is *partially observable*. The third column of Table I shows that a_i^t is *unknown* (marked with '?') when passive feedback is given.

Thus, the task of user attention prediction becomes estimating the probability that the user focuses on the t-th song:

$$\alpha_i^t := \Pr(a_i^t = 1 \mid X_i^t), \tag{1}$$

where $X_i^t = [\mathbf{x}_i^1, \mathbf{x}_i^2, \cdots, \mathbf{x}_i^t]$ is the sequence of history and current features. We can see that probability not only depends on the current information $\mathbf{x}_i^t$, but also depends on the previous interacted information X_i^{t-1}. It can be estimated with a neural network g with parameters $\mathbf{\Theta}_g$:

$$\hat{\alpha}_i^t := g(X_i^t; \mathbf{\Theta}_g). \tag{2}$$

For clear presentation, Table II lists the notations and their explanations used in this paper.

B. Ideal Risk for Attention Estimator

The focus of this study is to obtain an accurate attention predictor based on the dataset $\mathcal{S}$, the ideal empirical risk measure is defined as

$$\mathcal{R}_{ideal}^{Att}(g) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[\alpha_i^t \ell_g^+ + \left(1 - \alpha_i^t\right) \ell_g^- \right], \tag{3}$$

where ℓ_g^+ and ℓ_g^- respectively denote the simplified notations for $\ell^+(g(X_i^t; \mathbf{\Theta}_g))$ and $\ell^-(g(X_i^t; \mathbf{\Theta}_g))$. For instance, ℓ_g^+ and ℓ_g^- are the local losses for positive (with user attention) and negative (without user attention) examples, respectively. One popular loss function definition is log loss: $\ell_g^+ := -\log(g(X_i^t; \mathbf{\Theta}_g))$ and $\ell_g^- := -\log(1 - g(X_i^t; \mathbf{\Theta}_g))$.

Note that the ideal empirical risk defined in Eq. (3) uses the true but unobservable attention level α_i^t, making it infeasible to minimize Eq. (3) directly. In real applications, it is necessary to surrogate it with observed variables.

C. Are Existing Methods Unbiased against Ideal Risk?

The simplest estimator for the ideal loss is called PN (*i.e.*, ordinary supervised learning) [41], which naively treats all unlabeled data as negative samples. Its empirical risk is

$$\widehat{\mathcal{R}}_{PN}^{Att}(g) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[e_i^t \ell_g^+ + \left(1 - e_i^t\right) \ell_g^- \right]. \tag{4}$$

However, for user attention prediction, each unlabeled data could belong to either the positive or negative sample ($e_i^t = 0 \Rightarrow a_i^t = 0$ or $a_i^t = 1$) and Eq. (4) is biased against the ideal risk in Eq. (3) (*i.e.*, $\mathbb{E}[\widehat{\mathcal{R}}_{PN}^{Att}] \neq \mathcal{R}_{ideal}^{Att}$).

Another latest work NDB [12] uses a simple negative sampling heuristic for attention prediction. The rule is: the attention label is 0 only when a user continues for more than ten songs without any active feedback signals. From the viewpoint of PU-learning, NDB surrogates Eq. (3) with

$$\widehat{\mathcal{R}}_{NDB}^{Att}(g) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[e_i^t \ell_g^+ + d_i^t \left(1 - e_i^t\right) \ell_g^- \right], \tag{5}$$

where d_i^t is a passive sample mask variable. $d_i^t = 1$ if $e_i^{t-10} = e_i^{t-9} = \cdots = e_i^{t-1} = 0$, and $d_i^t = 0$ otherwise.

Though $\widehat{\mathcal{R}}_{NDB}^{Att}$ can be directly optimized, it is difficult to guarantee its unbiasedness: the definition of d_i^t implies that the user attention will decay to zero only after ten songs. However, a user may lose attention at any time during the streaming listening events [13], and thus NDB is also biased (*i.e.*, $\mathbb{E}[\widehat{\mathcal{R}}_{NDB}^{Att}] \neq \mathcal{R}_{ideal}^{Att}$). In fact, all the active feedback ($e_i^t = 1$) means $a_i^t = 1$ but the passive feedback ($e_i^t = 0$) does not always mean $a_i^t = 0$, which can be viewed as a positive-unlabeled problem.

Remark 1. *According to the analysis above, we have demonstrated that the loss functions of existing methods are all biased* *against the ideal loss defined in Eq.* (3)*. Therefore, to derive a desirable unbiased estimator, it is essential to consider the PU-learning nature and the sequential dependencies.*

D. A PU-learning View of Attention Estimation

From the viewpoint of positive-unlabeled learning (PU-learning) [14], [17], for the attention estimation task, the type e_i^t plays as the *observed indicator variable* that the example $\mathbf{x}_i^t$ is labeled, and attention a_i^t plays as the *true indicator variable* that the example is positive (*i.e.*, user attention is focused). They have the following relationship:

$$\Pr(a_i^t = 1 | e_i^t = 1) = 1, \tag{6}$$

$$\Pr(a_i^t = 0 | e_i^t = 0) < 1. \tag{7}$$

Specifically, from the viewpoint of predicting the user attention a_i^t, Eq. (6) correspond to the labeled positive data in PU-learning (*i.e.*, $e_i^t = 1 \Rightarrow a_i^t = 1$); and Eq. (7) corresponds to the unlabeled data in PU-learning (*i.e.*, a_i^t is unknown or unlabeled if $e_i^t = 0$ and can be either 1 or 0).

The key of ERM-based PU-learning is to define the propensity score that re-weights the labeled positive samples. In the attention prediction for music streaming services, the probability of observing active feedback (*i.e.*, labeled positive sample) depends on both the history user feedback sequences and the features, as shown in Figure 2. Therefore, we define the *sequential propensity score* as follows:

Definition 1. *The sequential propensity score of the user u_i on the t-th song of the i-th session is*

$$p_i^t := \Pr(e_i^t = 1 | X_i^t, E_i^{t-1}, a_i^t = 1), \tag{8}$$

where $E_i^{t-1} = [e_i^1, e_i^2, \cdots, e_i^{t-1}]$ is the user history observed feedback of the i-th session.

Intuitively, the sequential propensity score p_i^t can be viewed as the probability of a positive sample (*i.e.*, $a_i^t = 1$) being labeled or selected as an active feedback sample (*i.e.*, $e_i^t = 1$).

IV. Our Approach: UAE

In this section, we first naturally formulate the task of attention prediction as a problem of PU-learning, and then propose an Unbiased Attention Estimator (UAE) by extending the traditional Empirical-Risk-Minimization (ERM) PU-learning methods to handle sequential dependencies of user attention in music recommendation. Finally, we apply the proposed UAE to the downstream music recommendation tasks.

A. Unbiased Risks in Sequential PU-learning

Given the sequential propensity p_i^t defined in Eq. 8, we further derive the following proposition that connects the observed feedback type e_i^t with the true attention α_i^t:

Proposition 1. *Given the sequential propensity score p_i^t and the true attention level α_i^t, the expectation of observing feedback variable e_i^t is*

$$\mathbb{E}\left[e_i^t\right] = p_i^t \cdot \alpha_i^t, \tag{9}$$

for all i and t, where α_i^t is the true attention probability defined in Eq. (1).

Proof. The expectation of observed feedback variable e_i^t is $\mathbb{E}(e_i^t) = \Pr(e_i^t = 1 | X_i^t, E_i^{t-1})$ as e_i^t is a Bernoulli random variable. Thus, we have

$$\begin{aligned}
&\Pr(e_i^t = 1 | X_i^t, E_i^{t-1}) \\
&= \Pr(e_i^t = 1 | X_i^t, E_i^{t-1}, a_i^t = 1) \Pr(a_i^t = 1 | X_i^t, E_i^{t-1}) \\
&\quad + \Pr(e_i^t = 1 | X_i^t, E_i^{t-1}, a_i^t = 0) \Pr(a_i^t = 0 | X_i^t, E_i^{t-1}) \\
&= \Pr(e_i^t = 1 | X_i^t, E_i^{t-1}, a_i^t = 1) \Pr(a_i^t = 1 | X_i^t, E_i^{t-1}) \\
&= \Pr(e_i^t = 1 | X_i^t, E_i^{t-1}, a_i^t = 1) \Pr(a_i^t = 1 | X_i^t) \\
&= p_i^t \cdot \alpha_i^t,
\end{aligned}$$

where the second equation holds because if the user attention a_i^t is 0, then we must not observe active feedback (*i.e.*, $e_i^t = 1$). □

Unbiased Risk for Attention Estimation. Proposition 1 states that the propensity of any observed active feedback action can be decomposed as the product of the sequential propensity p_i^t and the true attention level α_i^t. Based on this, we further propose an unbiased sequential PU-learning attention estimator by extending the ERM-based PU-learning:

$$\begin{aligned}
&\widehat{\mathcal{R}}_{unbiased}^{Att}(g) \\
=&\frac{1}{|\mathcal{S}|}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[e_i^t\left(\frac{1}{p_i^t}\ell_g^+ + \left(1-\frac{1}{p_i^t}\right)\ell_g^-\right) + \left(1-e_i^t\right)\ell_g^-\right] \\
=&\frac{1}{|\mathcal{S}|}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[\frac{e_i^t}{p_i^t}\ell_g^+ + \left(1-\frac{e_i^t}{p_i^t}\right)\ell_g^-\right].
\end{aligned} \tag{10}$$

From the first equation of Eq. (10), we can see that to create a new unbiased data distribution, the proposed unbiased sequential PU-learning estimator treats all unlabeled (passive) examples as negative with weight 1, and treats all labeled (active) examples as both positive and negative with weights of $\frac{1}{p_i^t}$ and $\left(1-\frac{1}{p_i^t}\right)$ respectively. We can theoretically prove that Eq. (10) is an unbiased risk estimator, as shown in the following theorem:

Theorem 1 (Unbiased Sequential PU-learning Risk Estimation)**.** *Eq. (10) is unbiased in terms of the ideal risk in Eq. (3):*

$$\begin{aligned}
&\mathbb{E}[\widehat{\mathcal{R}}_{unbiased}^{Att}(g)] \\
&= \mathbb{E}\left[\frac{1}{|\mathcal{S}|}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[\frac{e_i^t}{p_i^t}\ell_g^+ + \left(1-\frac{e_i^t}{p_i^t}\right)\ell_g^-\right]\right] \\
&= \frac{1}{|\mathcal{S}|}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[\frac{\mathbb{E}[e_i^t]}{p_i^t}\ell_g^+ + \left(1-\frac{\mathbb{E}[e_i^t]}{p_i^t}\right)\ell_g^-\right] \\
&= \frac{1}{|\mathcal{S}|}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[\alpha_i^t\ell_g^+ + \left(1-\alpha_i^t\right)\ell_g^-\right] = \mathcal{R}_{ideal}^{Att}(g),
\end{aligned} \tag{11}$$

where the third equation in Eq. (11) *can be obtained from $\mathbb{E}[e_i^t] = p_i^t \cdot \alpha_i^t$ in the Proposition 1.*

Theorem 1 verifies the unbiasedness of the proposed attention estimator in Eq. (10).

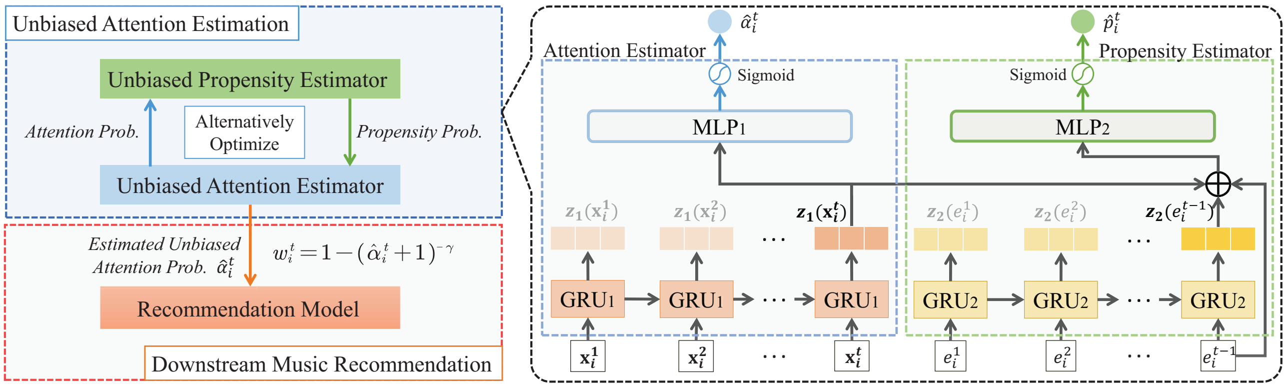


Fig. 4. An overview of our proposed UAE framework for modeling user attention in music recommendation. The left part demonstrates the full pipeline of this work. The right part is the specific architecture of the proposed GRU-based implementation.

Unbiased Risk for Propensity Estimation. The unbiased risk defined in Eq. (10) needs sequential propensity scores p_i^t, but the real value of p_i^t is always unavailable. In this paper, we aim to estimate the propensity at the t-th song with another neural network h:

$$\hat{p}_i^t := h(X_i^t, E_i^{t-1}; \mathbf{\Theta}_h), \tag{12}$$

where $E_i^{t-1} = [e_i^1, e_i^2, \cdots, e_i^{t-1}]$ is short for history observed attention sequence before t and $\mathbf{\Theta}_h$ is the parameters of the propensity estimation neural network h.

The ideal empirical risk for propensity estimation is

$$\mathcal{R}_{ideal}^{Pro}(h) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[p_i^t \ell_h^+ + \left(1 - p_i^t\right) \ell_h^- \right], \tag{13}$$

where ℓ_h^+ and ℓ_h^- are the simplified notations for $\ell^+(h(X_i^t, E_i^{t-1}; \mathbf{\Theta}_h))$ and $\ell^-(h(X_i^t, E_i^{t-1}; \mathbf{\Theta}_h))$, respectively.

However, the true propensity cannot be directly learned in real-world music streaming service scenarios. According to Proposition 1 and similar to the unbiased attention estimator defined in Eq. (10), we propose an unbiased propensity estimator from a dual perspective:

$$\widehat{\mathcal{R}}_{unbiased}^{Pro}(h) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[\frac{e_i^t}{\alpha_i^t} \ell_h^+ + \left(1 - \frac{e_i^t}{\alpha_i^t}\right) \ell_h^- \right]. \tag{14}$$

It can be proved that Eq. (14) is an unbiased risk function for the ideal risk in Eq. (13) in the following theorem:

Theorem 2 (Unbiased Sequential Propensity Estimation)**.** *Eq.* (14) *is unbiased in terms of the ideal risk in Eq. (13):*

$$\begin{aligned} &\mathbb{E}[\widehat{\mathcal{R}}_{unbiased}^{Pro}(h)] \\ &= \mathbb{E}\left[\frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[\frac{e_i^t}{\alpha_i^t} \ell_h^+ + \left(1 - \frac{e_i^t}{\alpha_i^t}\right) \ell_h^- \right] \right] \\ &= \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[\frac{\mathbb{E}[e_i^t]}{\alpha_i^t} \ell_h^+ + \left(1 - \frac{\mathbb{E}[e_i^t]}{\alpha_i^t}\right) \ell_h^- \right] \\ &= \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[p_i^t \ell_h^+ + \left(1 - p_i^t\right) \ell_h^- \right] = \mathcal{R}_{ideal}^{Pro}(h), \end{aligned} \tag{15}$$

where the third equation in Eq. (15) *can be obtained from* $\mathbb{E}[e_i^t] = p_i^t \cdot \alpha_i^t$ *in the Proposition 1.*

Given true attention level α_i^t, Theorem 2 shows that we can get unbiased sequential propensity estimation through minimizing the unbiased risk in Eq. (14), by only using the observable user feedback variable e_i^t.

B. Model Specification and Learning Algorithm

Based on the proposed two dual unbiased estimators, we further implement our framework with GRU [42], and optimize their parameters via alternating optimization.

Model Specification. As illustrated in the right part of Figure 4, our implementation of UAE framework consists of two GRUs and MLPs for predicting user attention and propensity score, respectively.

For constructing the attention prediction network g, we utilize the first GRU (called GRU_1) for the sequential feature modeling. The input at each step consists of all the sparse features and dense features, including user features (*e.g.*, gender, age, and country), song features (*e.g.*, artist, album and genre), and user-item cross features (*e.g.*, the number of songs played in the last 30 day). The hidden state at each step is recurrently computed by:

$$\boldsymbol{z}_1(\mathbf{x}_i^t), \boldsymbol{h}_i^t = \text{GRU}_1(\mathbf{x}_i^t, \boldsymbol{h}_i^{t-1}),$$

where $\boldsymbol{h}_i^t$ and $\boldsymbol{h}_i^{t-1}$ are the hidden vectors at the t-th and $(t-1)$-th steps, respectively. The output representation $\boldsymbol{z}_1(\mathbf{x}_i^t)$ encodes the feature information from the current time step and the history time steps. Then we employ a MLP for conducting the attention prediction:

$$\hat{\alpha}_i^t = \sigma(\text{MLP}_1(\boldsymbol{z}_1(\mathbf{x}_i^t))),$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the element-wise Sigmoid function.

As for the propensity estimation network h, we use another GRU (denoted as GRU_2) to model the user sequential history feedback. The output $\boldsymbol{z}_2(e_i^{t-1})$ and hidden state $\boldsymbol{h}_i^{\prime t-1}$ at each step t are recurrently computed as

$$\boldsymbol{z}_2(e_i^{t-1}), \boldsymbol{h}_i^{\prime t-1} = \text{GRU}_2(e_i^{t-1}, \boldsymbol{h}_i^{\prime t-2}).$$

Algorithm 1: Learning Algorithm for UAE

Input: Training dataset $\mathcal{S}$; number of training epochs N_e; number of iteration steps N_a, N_p

Output: Parameter $\mathbf{\Theta}_g$ for attention model g

1 Initialize the parameters $\mathbf{\Theta}_g, \mathbf{\Theta}_h$

2 **for** $n_e = 1, \cdots, N_e$ **do**

3   `// Unbiased Attention Risk Minimizer`

4   **for** $n_a = 1, \cdots, N_a$ **do**

5         $\hat{\mathcal{P}} \leftarrow \{\{h(X_i^t, E_i^{t-1}; \mathbf{\Theta}_h)\}_{t=1}^{l_i}\}_{i=1}^{N}$

6         $\mathbf{\Theta}_g \leftarrow \arg\min_g \widehat{\mathcal{R}}_{UAE}^{Att}(g \mid \hat{\mathcal{P}})$ {Eq. (16)}

7   **end**

8   `// Unbiased Propensity Risk Minimizer`

9   **for** $n_p = 1, \cdots, N_p$ **do**

10         $\hat{\mathcal{A}} \leftarrow \{\{g(X_i^t; \mathbf{\Theta}_g)\}_{t=1}^{l_i}\}_{i=1}^{N}$

11         $\mathbf{\Theta}_h \leftarrow \arg\min_h \widehat{\mathcal{R}}_{UAE}^{Pro}(h \mid \hat{\mathcal{A}})$ {Eq. (17)}

12   **end**

13 **end**

14 **return** $\mathbf{\Theta}_g$

After that and according to the propensity defined in Eq. (12), we can get the predicted sequential propensity with another MLP (denoted as MLP_2) which takes $\boldsymbol{z}_1(\mathbf{x}_i^t)$, $\boldsymbol{z}_2(e_i^{t-1})$, and e_i^{t-1} as its inputs:

$$\hat{p}_i^t = \sigma(\mathrm{MLP}_2(\boldsymbol{z}_1(\mathbf{x}_i^t) \oplus \boldsymbol{z}_2(e_i^{t-1}) \oplus e_i^{t-1})),$$

where '$\oplus$' is an operation to concatenate two vectors.

Alternating Optimization. The proposed UAE model has several parameters to learn, including $\mathbf{\Theta}_g$ in the attention model g (*i.e.*, parameters in GRU_1, MLP_1) and $\mathbf{\Theta}_h$ in the propensity model h (*i.e.*, parameters in GRU_2 and MLP_2). They can be determined by optimizing the following empirical risks.

First, given all of the estimated propensity $\hat{\mathcal{P}} = \{\hat{p}_i^t\}$ for $i = 1, \cdots, N; t = 1, \cdots, l_i$, and Eq. (10), the empirical risk for deriving the attention model parameters $\mathbf{\Theta}_g$ is defined as

$$\widehat{\mathcal{R}}_{UAE}^{Att}\left(g \mid \hat{\mathcal{P}}\right) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[\frac{e_i^t}{\hat{p}_i^t} \ell_g^+ + \left(1 - \frac{e_i^t}{\hat{p}_i^t}\right) \ell_g^- \right]. \tag{16}$$

Similarly, given all of the estimated attention $\hat{\mathcal{A}} = \{\hat{\alpha}_i^t\}$ for $i = 1, \cdots, N; t = 1, \cdots, l_i$, and Eq. (14), the empirical risk for deriving the propensity model parameters $\mathbf{\Theta}_h$ is defined as

$$\widehat{\mathcal{R}}_{UAE}^{Pro}\left(h \mid \hat{\mathcal{A}}\right) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[\frac{e_i^t}{\hat{\alpha}_i^t} \ell_h^+ + \left(1 - \frac{e_i^t}{\hat{\alpha}_i^t}\right) \ell_h^- \right]. \tag{17}$$

Lastly, we design a learning algorithm by minimizing the above two empirical risks where the attention network g and the propensity network h are updated in alternating fashion. As shown in Algorithm 1, each optimization iteration consists of two phases: the first phase updates $\mathbf{\Theta}_g$ by minimizing $\widehat{\mathcal{R}}_{UAE}^{Att}$ with current $\mathbf{\Theta}_h$ (*line* 4-7). Then, the second phase updates $\mathbf{\Theta}_h$ by minimizing $\widehat{\mathcal{R}}_{UAE}^{Pro}$ with the updated $\mathbf{\Theta}_g$ (*line* 9-12).

Remark 2. *Discussion of the time complexity. For our GRU-based implementation UAE, the most time complexity comes from the GRU to capture sequential pattern characteristics, which is $\mathcal{O}(n \cdot d^2)$ per layer, where n is the sequence length and d is the representation dimension.*

C. Applying to Downstream Music Recommendation

After we get the learned parameter $\mathbf{\Theta}_g$ for the attention estimator based on the Algorithm 1, we can directly use the attention estimator to predict the attention for each data sample without the propensity estimator. As illustrated in the left part of Figure 4, the estimated user attention can be used to improve the downstream music recommendation models, through estimating the confidence of passive feedback samples with the attention scores:

$$\begin{aligned} &\widehat{\mathcal{R}}_{UAE}^{Rec}(f) \\ =&\frac{1}{|\mathcal{S}|} \sum_{(\mathbf{x}_i^t, e_i^t, y_i^t) \in \mathcal{S}} \left[e_i^t \delta\left(f(\mathbf{x}_i^t)\right) + w_i^t \cdot \left(1 - e_i^t\right) \delta\left(f(\mathbf{x}_i^t)\right) \right], \end{aligned} \tag{18}$$

where f can be any downstream music recommendation models, and $\delta\left(f(\mathbf{x}_i^t)\right) = -(y_i^t \log(f(\mathbf{x}_i^t)) + (1 - y_i^t) \log(1 - f(\mathbf{x}_i^t)))$ is the local binary cross-entropy loss. Following the experiences in [39], [43], the confidence of the passive feedback data w_i^t is calculated based on the predicted user attention $\hat{\alpha}_i^t$ ($0 < \hat{\alpha}_i^t < 1$), and defined as a re-weighting function that conforms to a power-law distribution[2]:

$$w_i^t = 1 - (\hat{\alpha}_i^t + 1)^{-\gamma}, \tag{19}$$

where $\gamma > 0$ is a hyper-parameter. The re-weighting function guarantees that $0 \leq w_i^t < 1$ and increases with the increasing of $\hat{\alpha}_i^t$ (*i.e.*, $w_i^t \propto \hat{\alpha}_i^t$). According to Eq. (19), a passive sample in the training set is more reliable if the user has more potential attention on it, consequently leading to a higher weight.

Remark 3. *When incorporating the estimated attention score into the downstream task, our UAE model is solely utilized to infer the user attention score for re-weighting the unreliable passive feedback during the training phase of the downstream music recommendation model. Therefore, UAE will not increase the inference cost of the downstream recommendation model. Moreover, similar to standard re-weighting methods, our proposed UAE does not disrupt the convergence of the downstream recommendation model; instead, it provides a more accurate estimation of user attention and helps the convergence to a better solution. We will conduct an experimental analysis of the convergence of the downstream music recommendation models with and without our proposed UAE in Section VI-C.*

V. Discussion

A. Variance Analysis

Theorem 1 and Theorem 2 verify the unbiasedness of the attention estimator and propensity estimator through re-

[2]The design of w_i^t is still a open problem and need more explore in the future work. Here we only use a simple but effective empirical formula.

weighting with propensities, respectively. However, propensity-based re-weighting methods often suffer from high variance [39], [44], [45], [46], [47]. The following two theorems analyze the variances of the estimators in UAE.

Theorem 3 (Variance of unbiased attention estimator)**.** *The variance of the proposed unbiased attention risk estimator defined in Eq.* (10) *is*

$$\mathbb{V}\left[\widehat{\mathcal{R}}^{Att}_{unbiased}(g)\right] = \frac{1}{|\mathcal{S}|^2}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[\alpha_i^t\left(\frac{1}{p_i^t}-\alpha_i^t\right)\left(\ell_g^+-\ell_g^-\right)^2\right].$$

Theorem 4 (Variance of unbiased propensity estimator)**.** *The variance of the proposed unbiased propensity risk estimator defined in Eq.* (14) *is*

$$\mathbb{V}\left[\widehat{\mathcal{R}}^{Pro}_{unbiased}(h)\right] = \frac{1}{|\mathcal{S}|^2}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[p_i^t\left(\frac{1}{\alpha_i^t}-p_i^t\right)\left(\ell_h^+-\ell_h^-\right)^2\right].$$

Due to space limitations, we only present the proof of Theorem 3, as the proof of Theorem 4 can be derived in a similar way.

Proof. First, let $S_i^t = \frac{e_i^t}{p_i^t}\ell_g^+ + \left(1-\frac{e_i^t}{p_i^t}\right)\ell_g^-$. Then, $\mathbb{V}[S_i^t]$ can be represented as

$$\mathbb{V}[S_i^t] = \mathbb{E}[(S_i^t)^2] - \mathbb{E}^2[S_i^t].$$

Specifically, we have

$$\begin{aligned}
&\mathbb{E}[(S_i^t)^2] = \mathbb{E}\left[\left(\frac{e_i^t}{p_i^t}\ell_g^+ + \left(1-\frac{e_i^t}{p_i^t}\right)\ell_g^-\right)^2\right]\\
&= \mathbb{E}\left[\left(\frac{e_i^t}{p_i^t}\right)^2(\ell_g^+)^2 + 2\frac{e_i^t}{p_i^t}\left(1-\frac{e_i^t}{p_i^t}\right)\ell_g^+\ell_g^- + \left(1-\frac{e_i^t}{p_i^t}\right)^2(\ell_g^-)^2\right]\\
&= \frac{\mathbb{E}[(e_i^t)^2]}{(p_i^t)^2}(\ell_g^+)^2 + 2\left(\frac{\mathbb{E}[e_i^t]}{p_i^t} - \frac{\mathbb{E}[(e_i^t)^2]}{(p_i^t)^2}\right)\ell_g^+\ell_g^-\\
&\quad + \left(1-2\frac{\mathbb{E}[e_i^t]}{p_i^t} + \frac{\mathbb{E}[(e_i^t)^2]}{(p_i^t)^2}\right)(\ell_g^-)^2\\
&= \frac{\mathbb{E}[e_i^t]}{(p_i^t)^2}(\ell_g^+)^2 + 2\left(\frac{\mathbb{E}[e_i^t]}{p_i^t} - \frac{\mathbb{E}[e_i^t]}{(p_i^t)^2}\right)\ell_g^+\ell_g^-\\
&\quad + \left(1-2\frac{\mathbb{E}[e_i^t]}{p_i^t} + \frac{\mathbb{E}[e_i^t]}{(p_i^t)^2}\right)(\ell_g^-)^2\\
&= \frac{\alpha_i^t}{p_i^t}(\ell_g^+)^2 + 2\left(\alpha_i^t - \frac{\alpha_i^t}{p_i^t}\right)\ell_g^+\ell_g^- + \left(1-2\alpha_i^t + \frac{\alpha_i^t}{p_i^t}\right)(\ell_g^-)^2,
\end{aligned}$$

where the fourth equation holds because e_i^t is a Bernoulli random variable (*i.e.*, $\mathbb{E}[(e_i^t)^2] = \mathbb{E}[e_i^t]$), and the fifth equation comes from $\mathbb{E}[e_i^t] = p_i^t \cdot \alpha_i^t$ in Proposition 1.

Next, according to Theorem 1, $\mathbb{E}^2[S_i^t]$ can be calculated as

$$\begin{aligned}
&\mathbb{E}^2[S_i^t] = \left(\alpha_i^t\ell_g^+ + \left(1-\alpha_i^t\right)\ell_g^-\right)^2\\
&= (\alpha_i^t)^2(\ell_g^+)^2 + 2\alpha_i^t(1-\alpha_i^t)\ell_g^+\ell_g^- + (1-\alpha_i^t)^2(\ell_g^-)^2.
\end{aligned}$$

Finally, we obtain

$$\begin{aligned}
\mathbb{V}[S_i^t] &= \mathbb{E}[(S_i^t)^2] - \mathbb{E}^2[S_i^t]\\
&= \alpha_i^t(\frac{1}{p_i^t} - \alpha_i^t)(\ell_g^+ - \ell_g^-)^2,
\end{aligned}$$

from which we have

$$\begin{aligned}
\mathbb{V}[\widehat{\mathcal{R}}^{Att}_{unbiased}(g)] &= \frac{1}{|\mathcal{S}|^2}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\mathbb{V}[S_i^t]\\
&= \frac{1}{|\mathcal{S}|^2}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[\alpha_i^t(\frac{1}{p_i^t} - \alpha_i^t)(\ell_g^+ - \ell_g^-)^2\right].
\end{aligned}$$

□

These two theorems show that the variance of the two unbiased estimators depends on their inverse of the propensity scores (p_i^t for $\widehat{\mathcal{R}}^{Att}_{unbiased}$ and α_i^t for $\widehat{\mathcal{R}}^{Pro}_{unbiased}$), respectively. They also indicate that overestimating the propensity scores will reduce the effect of high variance problem for the two unbiased estimators in UAE, which confirms the clipping technique [39], [44], [46] for controlling the variance. However, overestimating the propensity inevitably leads to a rise in bias. We will analyze the phenomenon in the next section.

B. Bias Analysis

Though Theorem 1 shows the unbiasedness of attention estimation with true propensities, the estimated sequential propensity scores with Eq. (17) can still be inaccurate and introduce bias to the attention estimation. We show the bias of the attention estimator with the estimated propensities:

Theorem 5 (Bias of Attention Estimator with Estimated Propensities)**.** *Let $\hat{p}_i^t$ and p_i^t be the estimated sequential propensity score and true sequential propensity score, respectively. The bias of the attention estimator in UAE using $\hat{p}_i^t$ is*

$$\left|\frac{1}{|\mathcal{S}|}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[\left(\frac{p_i^t}{\hat{p}_i^t}-1\right)\alpha_i^t\left(\ell_g^+-\ell_g^-\right)\right]\right|.$$

Similarly, we give the bias of propensity estimator with the estimated attention from Eq. (16) as follows:

Theorem 6 (Bias of Propensity Estimator with Estimated Attention Scores)**.** *Let $\hat{\alpha}_i^t$ and α_i^t be the estimated attention score and true attention score, respectively. The bias of the propensity estimator in UAE using $\hat{\alpha}_i^t$ is*

$$\left|\frac{1}{|\mathcal{S}|}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[\left(\frac{\alpha_i^t}{\hat{\alpha}_i^t}-1\right)p_i^t\left(\ell_h^+-\ell_h^-\right)\right]\right|.$$

The proofs of Theorem 5 and Theorem 6 are similar and we only show the proof of Theorem 5 due to the page limitation.

Proof. First, given all estimated propensities $\hat{\mathcal{P}} = \{\hat{p}_i^t\}$ for $i = 1, \cdots, N; t = 1, \cdots, l_i$ the bias of our unbiased attention estimator is defined as

$$\mathrm{Bias}\left[\widehat{\mathcal{R}}^{Att}_{UAE}(g \mid \hat{\mathcal{P}})\right] := \left|\mathbb{E}\left[\widehat{\mathcal{R}}^{Att}_{UAE}(g \mid \hat{\mathcal{P}})\right] - \mathcal{R}^{Att}_{ideal}(g)\right|. \quad (20)$$

Then, for $\mathbb{E}\left[\widehat{\mathcal{R}}^{Att}_{UAE}(g \mid \hat{\mathcal{P}})\right]$, we have

$$\begin{aligned}
\mathbb{E}\left[\widehat{\mathcal{R}}^{Att}_{UAE}(g \mid \hat{\mathcal{P}})\right] &= \frac{1}{|\mathcal{S}|}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[\frac{\mathbb{E}[e_i^t]}{\hat{p}_i^t}\ell_g^+ + \left(1-\frac{\mathbb{E}[e_i^t]}{\hat{p}_i^t}\right)\ell_g^-\right]\\
&= \frac{1}{|\mathcal{S}|}\sum_{i=1}^{N}\sum_{t=1}^{l_i}\left[\frac{p_i^t\alpha_i^t}{\hat{p}_i^t}\ell_g^+ + \left(1-\frac{p_i^t\alpha_i^t}{\hat{p}_i^t}\right)\ell_g^-\right],
\end{aligned} \quad (21)$$

where the second equation comes from $\mathbb{E}[e_i^t] = p_i^t \cdot \alpha_i^t$ in the Proposition 1.

For $\mathcal{R}_{ideal}^{Att}(g)$, we have

$$\mathcal{R}_{ideal}^{Att}(g) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[\alpha_i^t \ell_g^+ + \left(1 - \alpha_i^t\right) \ell_g^- \right]. \quad (22)$$

Thus, we can complete the proof by substituting Eq. (21) and Eq. (22) into Eq. (20):

$$\begin{aligned} &\text{Bias}\left[\widehat{\mathcal{R}}_{UAE}^{Att}(g \mid \hat{\mathcal{P}})\right] \\ &= \left| \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left\{ \left[\frac{p_i^t \alpha_i^t}{\hat{p}_i^t} \ell_g^+ + \left(1 - \frac{p_i^t \alpha_i^t}{\hat{p}_i^t}\right) \ell_g^- \right] \right. \right. \\ &\quad \left. - \left[\alpha_i^t \ell_g^+ + \left(1 - \alpha_i^t\right) \ell_g^- \right] \right\} \Big| \\ &= \left| \frac{1}{|\mathcal{S}|} \sum_{i=1}^{N} \sum_{t=1}^{l_i} \left[\left(\frac{p_i^t}{\hat{p}_i^t} - 1 \right) \alpha_i^t \left(\ell_g^+ - \ell_g^- \right) \right] \right|. \end{aligned}$$

□

Theorem 5 indicates that a better estimated propensity score $\hat{p}_i^t$ (more close to true propensity score p_i^t) can reduce bias during the learning process. It also indicate that underestimate the propensity will result in a higher bias. Theorem 6 tells us the similar conclusion from the dual perspective.

C. Differences from Classical PU-learning

UAE is a PU-learning model adapted for modeling user attention in music recommendation. In that scene, it is similar to existing PU-learning models, but it also has striking differences.

First, classical PU-learning usually assumes that the labeling mechanism of each instance is uniform, or that the labeling mechanism only depends on its local features [14], [15], [16], [17]. That means the propensity of selecting a labeled positive example from the complete set of positive examples can be defined as $\Pr(e_i^t = 1 | \mathbf{x}_i^t, a_i^t = 1)$. However, the assumption often does not hold in real-world scenarios [17]. In music streaming recommendation, the active feedback actions (*i.e.*, labeled as positive samples in attention prediction task) have sequential dependency nature, which goes beyond the classical PU-learning assumption. Thus, UAE defines the sequential propensity as $\Pr(e_i^t = 1 | X_i^t, E_i^{t-1}, a_i^t = 1)$ which extends current PU-learning assumption by taking both the current and historical user-song interaction features and historical user feedback actions into consideration, which is more realistic.

Second, existing PU-learning methods often assume that the propensity scores are known (*e.g.*, from domain knowledge or previous studies), or directly treat them as hyper-parameters and tune on a small fully labeled data [15], [39], [40], [48], [49]. UAE, on the contrary, employs an alternating optimization algorithm to learn the attention estimator and propensity estimator simultaneously.

VI. Experiments

To understand the effectiveness of the proposed UAE, we conduct offline experiments on a public dataset and a product dataset collected from Huawei Music App, and online A/B testing for consecutive 7 days on Huawei Music.

TABLE III
Statistics of the experimental datasets in this paper.

Dataset	#Sessions	#Users	#Songs	#Features	#Feedback Types
`30-Music`	455 K	5.5 K	1.99 M	12	3
`Product`	8.47 M	3.75 M	1.73 M	44	6

Note: "M" means million and "K" means thousand.

A. Experimental Settings

Datasets. We evaluate the effectiveness and efficiency of the proposed UAE on a public dataset called 30-Music and a real-world dataset collected from Huawei Music App.

`30-Music`[3]: 30-Music dataset contains over 1-year time (Jan 2014 - Jan 2015) user listening events from Last.fm API. We filter the sessions with fewer than 10 interactions and split the data into a training set, a validation set, and a testing set, with ratios of $8:1:1$.

`Product`: We collect consecutive 9 days of user listening events from the Huawei Music App. The product dataset contains more feedback types, including active feedback (*i.e.*, "Like", "Share", "Download", "Skip", and "Dislike") and passive feedback (*i.e.*, "Auto-play"). The data collected in the first 7 days are used as the training set, the next 1 day as the validation set, and the final day as the testing set.

Table III summarizes the statistics of the two datasets.

Baselines. To verify the effectiveness and generality of UAE, we test the performances of several popular and state-of-the-art recommendation models (called base models) [4]. Meanwhile, the performances of the base models equipped with the user attention predicted by UAE are also tested for comparisons. The base models include factorization machines (**FM**) [50], wide & deep learning model (**Wide&Deep**) [51], **DeepFM** [6], **YoutubeNet** [52], deep & cross network (**DCN**) [53], automatic feature interaction learning network (**AutoInt**) [7], and the improved Deep & Cross Network (**DCN-V2**) [8]. In the experiments, we denote the base models equipped with UAE as "model name + UAE". For example, DCN-V2 equipped with UAE is denoted as "DCN-V2 + UAE".

We also compare UAE with existing user attention models and classical PU-learning models: **EDM** [11] is a heuristic method that assumes user attention exponentially decays with time until another active feedback happens or plunges to zero. **NDB** [12] directly learns user attention from observed user feedback using a heuristic negative sampling, which is biased according to our analysis in Section III-C. **PN** [41] is a baseline for PU-learning that treats the attention of all unlabeled (passive) samples as zero. **SAR** [38] is a PU-learning method that assumes the labeling mechanism only depends on the local features.

Metrics. Two widely used metrics were used to evaluate

[3]https://recsys.deib.polimi.it/datasets/

[4]It is infeasible to evaluate the accuracy of user attention prediction directly, owing to the challenges in obtaining the ground-truth of user attention.

TABLE IV

Overall performance of seven base recommendation models trained with and without UAE on `30-Music` and `Product` datasets. All experiments are conducted five times with different random seeds and the averaged results are reported as percentage values with "%" omitted. '$*$' indicates the improvements over base models are statistically significant (t-test, p-value< 0.05).

Dataset	Metric		Base Model						
			FM	Wide&Deep	DeepFM	YoutubeNet	DCN	AutoInt	DCN-V2
`30-Music`	AUC	Base	74.90	73.84	73.62	73.73	73.78	73.91	73.95
		+UAE (Ours)	**75.02***	**73.95***	**73.64**	**73.96***	**74.08***	**74.17***	**74.11***
		RelaImpr	0.48	0.46	0.08	0.97	1.26	1.09	0.67
	GAUC	Base	59.65	60.10	60.01	60.14	60.16	59.57	60.13
		+UAE (Ours)	**59.78***	**60.14**	**60.16***	**60.18**	**60.17**	**60.03***	**60.20**
		RelaImpr	1.35	0.40	1.50	0.39	0.10	4.81	0.69
`Product`	AUC	Base	76.64	78.78	78.75	79.30	79.33	79.39	79.42
		+UAE (Ours)	**76.77***	**78.93***	**78.93***	**79.45***	**79.46***	**79.50***	**79.52***
		RelaImpr	0.49	0.52	0.63	0.51	0.44	0.37	0.34
	GAUC	Base	58.44	59.73	59.66	60.12	60.30	60.27	60.37
		+UAE (Ours)	**58.79***	**59.94***	**59.95***	**60.37***	**60.43***	**60.50***	**60.60***
		RelaImpr	4.15	2.16	3.00	2.47	1.26	2.24	2.22

the recommendation accuracy in offline experiments, *i.e.*, area under ROC curve (AUC) [54] and group AUC (GAUC) [55]:

$$\text{AUC} = \frac{1}{|\mathcal{P}||\mathcal{N}|} \sum_{p \in \mathcal{P}} \sum_{n \in \mathcal{N}} \mathbb{I}\left(f(p) > f(n)\right),$$

$$\text{GAUC} = \frac{\sum_{u_i \in \mathcal{U}} w_{u_i} * \text{AUC}_{u_i}}{\sum_{u_i \in \mathcal{U}} w_{u_i}},$$

where $\mathcal{P}$ and $\mathcal{N}$ denote the positive sample set and negative sample set, respectively. $f(\cdot)$ is the recommendation model and $\mathbb{I}$ is the indicator function. The weight w_{u_i} in GAUC is the number of clicks of user u_i.

Moreover, following [56], [57], [58], [59], we further introduce the RelaImpr [60] metric to show the relative improvements more clearly. Since AUC and GAUC are 0.5 for a random strategy, RelaImpr is defined as:

$$\text{RelaImpr} = \left(\frac{Metric(\text{evaluated model}) - 0.5}{Metric(\text{base model}) - 0.5} - 1\right) \times 100\%,$$

where $Metric$ could be AUC or GAUC. In the online A/B testing, we calculate the relative improvements in terms of user play time and play count, which are also widely used in industrial music streaming services products.

Implementation details. For all of the base models and UAE, we set the embedding size as 8 and the hidden layers of MLPs as $(256, 128, 64)$. We set the batch size as 4096 for all models for fair comparisons. The hidden dimensions of the two GRUs in UAE were tuned among $\{64, 128, 256\}$. We utilize Adam[61] as the optimizer and tuned the initial learning rate for attention model g and propensity model h among $\{1e^{-2}, 1e^{-3}, 1e^{-4}, 1e^{-5}\}$. For the iteration steps for attention risk minimization and propensity risk minimization in Algorithm 1, we set $N_a = 1$ and $N_p = 2$ as we find that the attention estimator has faster convergence than the propensity estimator in our experiments. Following the practices in [37], [62], we adopt a risk-clipped technique when optimization to get non-negative risks, which can reduce the variances.

B. Experimental Results on Offline Datasets

We first test the performances of the 7 base models with and without UAE on `30-Music` and `Product`. The averaged results under five times experiments with different random seeds are reported in Table IV. The line "+UAE" is our approach (*i.e.*, equipping the corresponding base model with UAE). We can see that, after equipping our proposed UAE, all of the base models were enhanced in terms of AUC and GAUC on both datasets. Most of the improvements are statistically significant with p-value< 0.05 under the t-test. We attribute the better performance to the predicted user attention by UAE helped to remove the impact of unreliable passive feedback samples for the recommendation task. These results verify UAE's effectiveness in accurately predicting user attention and thereby can be applied to various downstream music recommendation models to enhance the recommendation performance.

In particular, the improvements in terms of GAUC are more significant than that of AUC on both datasets. Specifically, on `Product` dataset, the average gains and the average relative improvement of GAUC are 0.24% and 2.50%, respectively. Note that such improvements are remarkable in the real-world industrial scenario because GAUC usually reflects the online real performance improvements well, and even 0.1% offline GAUC gains can lead to significant online promotions [8], [47], [63], [64]. Based on these promising results, we can conclude that UAE has the potential to achieve real online improvements, which will be further verified in the online A/B testing in the following Section VI-D.

To compare UAE with existing user attention models and PU-learning models, we select the best-performing base models, AutoInt and DCN-V2 [8], as the base models. And then we equip them with the attention predicted by EDM [11], NDB [12], PN [41], and SAR [38], achieving five baselines for each base model and denoted as "Base", "+EDM", "+NDB", "+PN", and "+SAR", respectively. Our approach is denoted as "+UAE". Table V reports the experimental results on both the

TABLE V

PERFORMANCE COMPARISONS OF AUTOINT/DCN-V2 EQUIPPED WITH DIFFERENT ATTENTION PREDICTION MODELS ON `30-Music` AND `Product` DATASETS. ALL EXPERIMENTS ARE CONDUCTED FIVE TIMES WITH DIFFERENT RANDOM SEEDS AND THE AVERAGED RESULTS ARE REPORTED AS PERCENTAGE VALUES WITH "%" OMITTED. '$*$' INDICATES THE IMPROVEMENTS OVER THE BEST BASELINES ARE SIGNIFICANT (t-TEST, p-VALUE< 0.05).

Dataset	Base Model	Metric	Base	+EDM	+NDB	+PN	+SAR	+UAE (Ours)
`30-Music`	AutoInt	AUC	73.91	73.80	73.89	67.51	73.97	**74.17***
		RelaImpr	0.00	−0.46	−0.0	−26.7	0.25	1.09
		GAUC	59.57	59.42	59.81	53.80	59.78	**60.03***
		RelaImpr	0.00	−1.57	2.51	−60.29	2.19	4.81
	DCN-V2	AUC	73.95	73.90	73.80	67.43	73.97	**74.11***
		RelaImpr	0.00	−0.21	−0.63	−27.22	0.08	0.67
		GAUC	60.13	60.10	60.17	53.62	60.16	**60.20**
		RelaImpr	0.00	−0.30	0.39	−64.26	0.30%	0.69
`Product`	AutoInt	AUC	79.39	79.26	79.33	54.65	79.29	**79.50***
		RelaImpr	0.00	−0.44	−0.2	−84.1	−0.34	0.37
		GAUC	60.27	60.44	60.26	52.46	60.34	**60.50***
		RelaImpr	0.00	1.66	−0.10	−76.05	0.68	2.24
	DCN-V2	AUC	79.42	79.42	79.43	54.98	79.38	**79.52***
		RelaImpr	0.00	0.00	0.03	−83.07	−0.14	0.34
		GAUC	60.37	60.40	60.42	52.39	60.45	**60.60***
		RelaImpr	0.00	0.29	0.48	−76.96	0.77	2.22

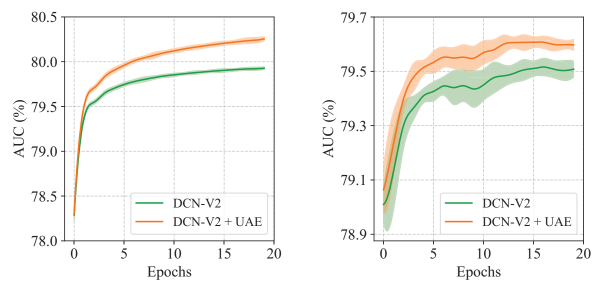
`30-Music` and `Product` datasets.

We first observe that the attention models ("EDM" and "NDB"), PU-learning models (except "PN"), and our approach, in general, demonstrate improvements over the base models. This observation underscores the essential importance of accurate user attention prediction in the context of music recommendation. However, PN performs worst because it directly discarded all the valuable information in the passive feedback. Our product experience also suggests that using solely limited data (only active feedback samples) leads to poor performance and is not a preferable approach. In contrast, compared with all baselines, the model equipped with UAE (our approach) performs the best under the same base model. Particularly, it outperforms other attention models, indicating the crucial role of utilizing all passive feedback actions and conducting unbiased estimation. Moreover, it also outperforms the classical PU-learning model of SAR, further validating the effectiveness of UAE in accurately predicting user attention by capturing sequential dependencies between feedback actions.

C. Experimental Analysis

As seen in Table IV, among all of the base models, DCN-V2 achieves the best performance when equipped with UAE. We further conduct experiments to further analyze UAE with DCN-V2 as the base model on `Product` dataset.

Convergence Analysis. To verify the convergence of our learning algorithm in Algorithm 1, we track the AUC curves of both DCN-V2 and DCN-V2 + UAE with respect to the training epochs. Both models are trained for 20 epochs, and we record the AUC scores on the training set and validation set. To ensure robustness, we conduct the experiments 10 times and show the averaged numbers and their 95% confidence intervals in Figure 5. From the results, we can conclude that



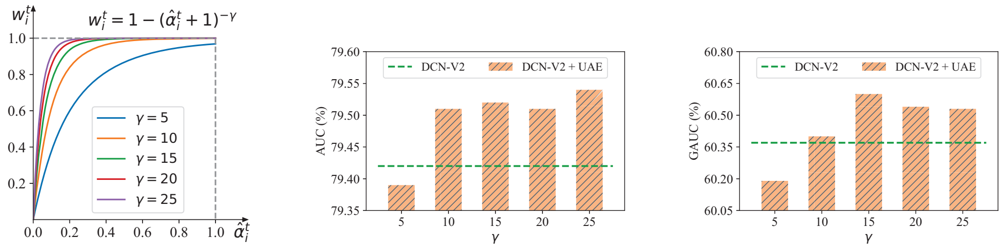
(a) Training (b) Validation

Fig. 5. Performance curves of DCN-V2 with and without UAE w.r.t. the training epochs. The shaded area indicates the 95% confidence intervals of t-distribution under the 10 experiments with different random seeds.

UAE can help the base model in terms of not only consistently aiding the base model in converging to a better solution, but also reducing the variance. The phenomenon is observed in both the training and validation set. Overall, these observations highlight UAE's potential as a valuable method for enhancing music recommendation systems by improving convergence and reducing variance, thereby enhancing recommendation performance and user experience in real-world applications.

Parameter Analysis. When applying the predicted user attention scores to the downstream music recommendation, one key issue is transforming the estimated attention scores to the weights on the unreliable passive feedback samples. Eq. (19) gives a transformation function that contains a parameter γ that needs to be tuned in the experiments. Figure 6(a) shows the curves of the re-weight function w.r.t. the predicted attention, with varying values of γ.

In the experiments, we vary γ among $\{5, 10, 15, 20, 25\}$.



(a) (b) (c)

Fig. 6. Analysis of the parameter γ in re-weight function. (a): The re-weighting function curves with different γ values. (b): AUC of DCN-V2 trained with UAE w.r.t. different γ values. (c): GAUC of DCN-V2 trained with UAE w.r.t. different γ values.

Figure 6(b) and Figure 6(c) report the AUC and GAUC of DCN-V2+UAE with different γ values. The performance of DCN-V2 is shown as dashed vertical lines. From the results, we can observe that DCN-V2+UAE, in general, outperformed DCN-V2 when γ is set to relatively large values (*e.g.*, ≥ 10), showing the robustness of Eq. (19). Optimal performance is achieved around $\gamma = 15$. We attribute this to the fact that although the unreliable passive feedback samples would damage the recommendation performance because of the noisy labels, they still contribute valuable information to the training of recommendation models. Assigning excessively small weights to the unreliable passive feedback samples (*i.e.*, setting γ to smaller values) will hurt the overall recommendation performance. Moreover, it is noteworthy that the performance curves are quite close when γ becomes large, indicating that the performance of DCN-V2+UAE is not highly sensitive to γ in this range.

D. Online A/B Testing

To evaluate UAE's effectiveness in real music recommendation products, we deploy UAE to Huawei Music App and conduct a one-week online A/B testing. Specifically, we randomly split the users into two groups, namely the control group and the experimental group, each involving hundreds of thousands of daily active users (DAU). For the control group, the users are served by a highly-optimized deep base model without UAE. For the experimental group, the users are served by the same base model equipped with UAE. Figure 7 shows the improvements over the base model after equipping it with UAE, in terms of play time and play count. The dashed horizontal lines represent the average improvements over the seven-day testing period. The results consistently demonstrate that UAE significantly enhances recommendation performance with a substantial margin, both in play time and play count. On average, the relative performance increase in users' play time and play count exceeds 2%. The significant and remarkable online improvements unequivocally showcase the effectiveness of UAE in real-world music recommendation products. The success can be attributed to UAE's superior ability in accurately predicting user attention and quantifying the reliability of passive feedback samples. Notably, Huawei Music App has millions of active users and tens of millions of daily plays, of which 10% of traffic has been served by our method.

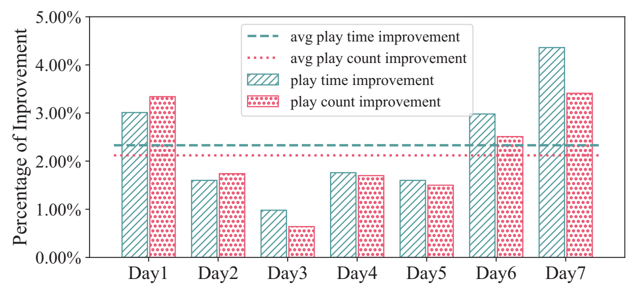


Fig. 7. Online A/B performances for consecutive 7 days on Huawei Music.

VII. Conclusion and Future Work

This paper studies the problem of modeling user attention for counteracting the unreliable passive user feedback in music recommendation. We first naturally formulate the user attention prediction as a PU-learning problem, and further propose an extended ERM-based PU-learning model called UAE to characterize the sequential dependencies of the user feedback actions. Theoretical analysis shows the unbiasedness and variance of the user attention estimator and propensity estimator. Experimental results on offline datasets and online A/B testing on Huawei Music both verify the effectiveness and robustness of our approach.

In future work, we plan to further study how to better apply user attention to enhance other downstream tasks to improve users' online experience in real-world industrial applications.

Acknowledgment

This work was funded by the National Key R&D Program of China (2023YFA1008704), National Natural Science Foundation of China (U2001212, 62377044, 62376275), Beijing Key Laboratory of Big Data Management and Analysis Methods, Major Innovation & Planning Interdisciplinary Platform for the "Double-First Class" Initiative, the Research Funds of Renmin University of China (No.22XNH027), Public Computing Cloud, Renmin University of China, and funds for building world-class universities (disciplines) of Renmin University of China. We would like to thank Yalei Lv of Tsinghua University for her early discussions during her internship at Noah's Ark Lab.

References

[1] O. Celma, "Music recommendation," in *Music recommendation and discovery: The long tail, long fail, and long play in the digital music space*. Springer, 2010, pp. 43–85.

[2] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender systems handbook*. Springer, 2011, pp. 1–35.

[3] M. Schedl, H. Zamani, C.-W. Chen, Y. Deldjoo, and M. Elahi, "Current challenges and visions in music recommender systems research," *International Journal of Multimedia Information Retrieval*, vol. 7, no. 2, pp. 95–116, 2018.

[4] N. Wlömert and D. Papies, "On-demand streaming services and music industry revenues—insights from spotify's market entry," *International Journal of Research in Marketing*, vol. 33, no. 2, pp. 314–327, 2016.

[5] M. Schedl, P. Knees, B. McFee, D. Bogdanov, and M. Kaminskas, "Music recommender systems," *Recommender systems handbook*, pp. 453–492, 2015.

[6] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, "Deepfm: A factorization-machine based neural network for ctr prediction," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017, pp. 1725–1731.

[7] W. Song, C. Shi, Z. Xiao, Z. Duan, Y. Xu, M. Zhang, and J. Tang, "Autoint: Automatic feature interaction learning via self-attentive neural networks," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 1161–1170.

[8] R. Wang, R. Shivanna, D. Cheng, S. Jain, D. Lin, L. Hong, and E. Chi, "Dcn v2: Improved deep & cross network and practical lessons for web-scale learning to rank systems," in *Proceedings of the Web Conference 2021*, ser. WWW '21, New York, NY, USA, 2021, p. 1785–1797.

[9] F. Meggetto, C. Revie, J. Levine, and Y. Moshfeghi, "On skipping behaviour types in music streaming sessions," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 3333–3337.

[10] Q. Dai, Y. Lv, J. Zhu, J. Ye, Z. Dong, R. Zhang, S.-T. Xia, and R. Tang, "Lcd: Adaptive label correction for denoising music recommendation," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 3903–3907.

[11] K. Ahmed, "Analyzing user behavior and sentiment in music streaming services," Master's thesis, KTH Royal Institute of Technology, 2016.

[12] X. Zhang, S. Dai, J. Xu, Z. Dong, Q. Dai, and J.-R. Wen, "Counteracting user attention bias in music streaming recommendation via reward modification," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 2504–2514.

[13] B. Zhang, G. Kreitz, M. Isaksson, J. Ubillos, G. Urdaneta, J. A. Pouwelse, and D. Epema, "Understanding user behavior in spotify," in *2013 Proceedings IEEE INFOCOM*. IEEE, 2013, pp. 220–224.

[14] J. Bekker and J. Davis, "Learning from positive and unlabeled data under the selected at random assumption," in *Second International Workshop on Learning with Imbalanced Domains: Theory and Applications*. PMLR, 2018, pp. 8–22.

[15] M. Du Plessis, G. Niu, and M. Sugiyama, "Convex formulation for learning from positive and unlabeled data," in *International conference on machine learning*. PMLR, 2015, pp. 1386–1394.

[16] X.-L. Li and B. Liu, "Learning from positive and unlabeled examples with different data distributions," in *European conference on machine learning*. Springer, 2005, pp. 218–229.

[17] J. Bekker and J. Davis, "Learning from positive and unlabeled data: A survey," *Machine Learning*, vol. 109, no. 4, pp. 719–760, 2020.

[18] X. Zhou, D. Qin, X. Lu, L. Chen, and Y. Zhang, "Online social media recommendation over streams," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, 2019, pp. 938–949.

[19] X. Zhou, L. Chen, Y. Zhang, L. Cao, G. Huang, and C. Wang, "Online video recommendation in sharing community," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 2015, pp. 1645–1656.

[20] L. Xia, C. Huang, Y. Xu, P. Dai, M. Lu, and L. Bo, "Multi-behavior enhanced recommendation with cross-interaction collaborative relation modeling," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, 2021, pp. 1931–1936.

[21] J. Yuan, W. Ji, D. Zhang, J. Pan, and X. Wang, "Micro-behavior encoding for session-based recommendation," in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 2022, pp. 2886–2899.

[22] Y. Song, S. Dixon, and M. Pearce, "A survey of music recommendation systems and future perspectives," in *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval*, 2012, pp. 395–410.

[23] A. van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Advances in Neural Information Processing Systems 26*, 2013, pp. 2643–2651.

[24] C. Gao, X. He, D. Gan, X. Chen, F. Feng, Y. Li, T.-S. Chua, and D. Jin, "Neural multi-task recommendation from multi-behavior data," in *2019 IEEE 35th international conference on data engineering (ICDE)*, 2019, pp. 1554–1557.

[25] M. Kaminskas and F. Ricci, "Contextual music information retrieval and recommendation: State of the art and challenges," *Computer Science Review*, vol. 6, no. 2-3, pp. 89–119, 2012.

[26] X. Wang, D. Rosenblum, and Y. Wang, "Context-aware mobile music recommendation for daily activities," in *Proceedings of the 20th ACM international conference on Multimedia*, 2012, pp. 99–108.

[27] M. Braunhofer, M. Kaminskas, and F. Ricci, "Location-aware music recommendation," *International Journal of Multimedia Information Retrieval*, vol. 2, pp. 31–44, 2013.

[28] C. Hansen, C. Hansen, L. Maystre, R. Mehrotra, B. Brost, F. Tomasi, and M. Lalmas, "Contextual and sequential user embeddings for large-scale music recommendation," in *Proceedings of the 14th ACM Conference on Recommender Systems*, 2020, pp. 53–62.

[29] R. Reza Aditya Permadi, "Improving recommender systems algorithms for personalized music video television by incorporating user consumption behaviour and multiple types of user feedback," Master's thesis, Delft University of Technology, 2018.

[30] B. Liu, Y. Dai, X. Li, W. S. Lee, and P. S. Yu, "Building text classifiers using positive and unlabeled examples," in *Third IEEE international conference on data mining*. IEEE, 2003, pp. 179–186.

[31] X. Li and B. Liu, "Learning to classify texts using positive and unlabeled data," in *IJCAI*, vol. 3, no. 2003. Citeseer, 2003, pp. 587–592.

[32] M. C. Du Plessis and M. Sugiyama, "Class prior estimation from positive and unlabeled data," *IEICE TRANSACTIONS on Information and Systems*, vol. 97, no. 5, pp. 1358–1362, 2014.

[33] K. Zhou, G.-R. Xue, Q. Yang, and Y. Yu, "Learning with positive and unlabeled examples using topic-sensitive plsa," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 1, pp. 46–58, 2009.

[34] S. Liang, Y. Zhang, and J. Ma, "Active model selection for positive unlabeled time series classification," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, 2020, pp. 361–372.

[35] C.-J. Hsieh, N. Natarajan, and I. Dhillon, "Pu learning for matrix completion," in *International conference on machine learning*. PMLR, 2015, pp. 2445–2453.

[36] W. Gerych, T. Hartvigsen, L. Buquicchio, A. Alajaji, K. Chandrasekaran, H. Mansoor, E. Rundensteiner, and E. Agu, "Positive unlabeled learning with a sequential selection bias," in *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*. SIAM, 2022, pp. 19–27.

[37] R. Kiryo, G. Niu, M. C. Du Plessis, and M. Sugiyama, "Positive-unlabeled learning with non-negative risk estimator," *Advances in neural information processing systems*, vol. 30, 2017.

[38] J. Bekker, P. Robberechts, and J. Davis, "Beyond the selected completely at random assumption for learning from positive and unlabeled data," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2019, pp. 71–85.

[39] Y. Saito, S. Yaginuma, Y. Nishino, H. Sakata, and K. Nakata, "Unbiased recommender learning from missing-not-at-random implicit feedback," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, 2020, pp. 501–509.

[40] S. Chang, Y. Zhang, J. Tang, D. Yin, Y. Chang, M. A. Hasegawa-Johnson, and T. S. Huang, "Positive-unlabeled learning in streaming networks." in *KDD*, 2016, pp. 755–764.

[41] J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset shift in machine learning*. Mit Press, 2008.

[42] K. Cho, B. van Merrienboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 1724–1734.

[43] L. Yang, Y. Cui, Y. Xuan, C. Wang, S. Belongie, and D. Estrin, "Unbiased offline recommender evaluation for missing-not-at-random implicit feedback," in *Proceedings of the 12th ACM conference on recommender systems*, 2018, pp. 279–287.

[44] M. Chen, A. Beutel, P. Covington, S. Jain, F. Belletti, and E. H. Chi, "Top-k off-policy correction for a reinforce recommender system," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 2019, pp. 456–464.
[45] X. Zhang, H. Jia, H. Su, W. Wang, J. Xu, and J. Wen, "Counterfactual reward modification for streaming recommendation with delayed feedback," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, pp. 41–50.
[46] Z. Wang, S. Shen, Z. Wang, B. Chen, X. Chen, and J.-R. Wen, "Unbiased sequential recommendation with latent confounders," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 2195–2204.
[47] S. Dai, Y. Zhou, J. Xu, and J.-R. Wen, "Dually enhanced delayed feedback modeling for streaming conversion rate prediction," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 390–399.
[48] S. Jain, J. Delano, H. Sharma, and P. Radivojac, "Class prior estimation with biased positives and unlabeled examples," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 4255–4263.
[49] J. Bekker and J. Davis, "Estimating the class prior in positive and unlabeled data through decision tree induction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
[50] S. Rendle, "Factorization machines," in *2010 IEEE International conference on data mining*. IEEE, 2010, pp. 995–1000.
[51] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir *et al.*, "Wide & deep learning for recommender systems," in *Proceedings of the 1st workshop on deep learning for recommender systems*, 2016, pp. 7–10.
[52] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for youtube recommendations," in *Proceedings of the 10th ACM conference on recommender systems*, 2016, pp. 191–198.
[53] R. Wang, B. Fu, G. Fu, and M. Wang, "Deep & cross network for ad click predictions," in *Proceedings of the ADKDD'17*, 2017, pp. 1–7.
[54] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
[55] H. Zhu, J. Jin, C. Tan, F. Pan, Y. Zeng, H. Li, and K. Gai, "Optimized cost per click in taobao display advertising," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 2191–2200.
[56] G. Zhou, X. Zhu, C. Song, Y. Fan, H. Zhu, X. Ma, Y. Yan, J. Jin, H. Li, and K. Gai, "Deep interest network for click-through rate prediction," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 1059–1068.
[57] M. Luo, F. Chen, P. Cheng, Z. Dong, X. He, J. Feng, and Z. Li, "Metaselector: Meta-learning for recommendation with user-level adaptive model selection," in *Proceedings of The Web Conference 2020*, 2020, pp. 2507–2513.
[58] R. Xie, C. Ling, Y. Wang, R. Wang, F. Xia, and L. Lin, "Deep feedback network for recommendation," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2021, pp. 2519–2525.
[59] Y. Zhu, R. Xie, F. Zhuang, K. Ge, Y. Sun, X. Zhang, L. Lin, and J. Cao, "Learning to warm up cold item embeddings for cold-start recommendation with meta scaling and shifting networks," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, pp. 1167–1176.
[60] L. Yan, W.-j. Li, G.-R. Xue, and D. Han, "Coupled group lasso for web-scale ctr prediction in display advertising," in *International Conference on Machine Learning*. PMLR, 2014, pp. 802–810.
[61] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, San Diega, CA, USA, 2015.
[62] Y. Saito, G. Morisihta, and S. Yasui, "Dual learning algorithm for delayed conversions," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 1849–1852.
[63] X. Ma, L. Zhao, G. Huang, Z. Wang, Z. Hu, X. Zhu, and K. Gai, "Entire space multi-task model: An effective approach for estimating post-click conversion rate," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 1137–1140.
[64] Y. Chen, J. Jin, H. Zhao, P. Wang, G. Liu, J. Xu, and B. Zheng, "Asymptotically unbiased estimation for delayed feedback modeling via label correction," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 369–379.