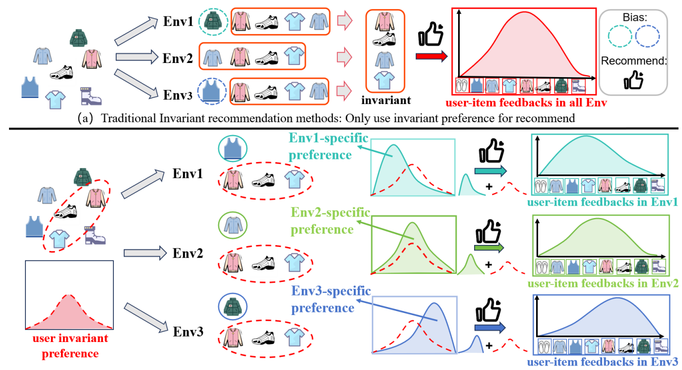
Invariant and Environment-specific Preference Learning with Auxiliary Information for Unbiased Recommendation

1st Ting Bi, 2nd Hangtong Xu, 3rd* Yuanbo Xu
Data Mining in Mobile Intelligent Computing
Jilin University
Changchun, China
Email: {biting23@mails, xuht24@mails, yuanbox@*}.jlu.edu.cn

***Abstract*—Invariant user preference learning is a core task in recommender systems. Accurately modeling user preferences is crucial, as it directly impacts the quality of recommendations. However, heterogeneous user preferences in feedback data often exhibit mixture distributions, which obscure invariant preferences and introduce bias. Existing methods typically address this issue by partitioning feedback data into multiple environments and learning invariant preferences across them. Nonetheless, these approaches often lack theoretical guarantees for environment construction and fail to capture the dynamic nature of user preferences across different environments. Along these lines, we propose a novel framework, Invariant and Environment-specific Preferences for unbiased recommendation (IEPref). IEPref leverages auxiliary information as a reliable signal to guide the environment classifier in partitioning the environment, thereby enabling the learning of more stable and generalizable user preference with theoretical guarantees. Additionally, we design environment-specific proxy modules to capture context-dependent preference patterns unique to each environment. The environment classifier assigns each user-item interaction to its corresponding latent environment, and both invariant and environment-specific preferences are integrated for recommendations. Extensive experiments on five real-world datasets demonstrate that IEPref achieves superior performance over existing baselines, effectively mitigating recommendation bias while preserving personalized modeling capabilities.**

***Index Terms*—Recommender Systems, Causal Inference, Invariant Learning, Debiasing.**



(a) Traditional Invariant recommendation methods: Only use invariant preference for recommend

(b) Our recommendation method: Reliable environment partition and dynamic environment preferences

Fig. 1: The figure illustrates the comparison between traditional invariant recommendation methods and our proposed approach: (a) Traditional methods typically assume that the user's true preference is entirely invariant and treat variations in preferences as noise induced by the environment; (b) Our approach, while accurately modeling invariant preferences, further incorporates environment-specific preferences to better capture the dynamic nature of user preferences.

I. Introduction

With the rapid growth of online information, recommender systems (RSs) have become essential for alleviating information overload on platforms such as e-commerce, entertainment, and social media [1]. By analyzing users' historical interactions [2], RSs identify and prioritize relevant content to support efficient information access. Despite significant advances in content-based recommendation, collaborative filtering, and hybrid approaches, most existing methods rely on the assumption that user feedback data is generated from a single, unified distribution. However, this assumption rarely holds in real-world scenarios. In practice, due to spatial [3] and temporal heterogeneity, as well as the operational mechanisms of recommender systems (e.g., seasonal changes, regional differences, climate variation), user interactions are typically drawn from a mixture of multiple latent distributions. This mixture structure introduces biases that hinder the accurate modeling of users' true preferences.

Recent studies have attempted to address this challenge by learning invariant preferences across multiple environments. Notable examples include CausPref [4] and InvPref [5], which adopt causal invariant learning medthod to capture stable user preferences that remain robust across varying environments. However, these approaches face several limitations. First, the partitioning of data into environments is often heuristic and lacks theoretical guarantees, resulting in potentially unreliable representations. Second, focusing solely on invariant preferences overlooks environment-specific user behaviors that are crucial for personalized recommendation. As shown in fig 1, conventional approaches (a) treat preference variation as noise, lacking the ability to dynamically capture context-

* Corresponding author.

aware behaviors. In contrast, our approach (b) interprets preference variation as environment-specific drift and assumes that user preferences consist of both invariant and environment-dependent components.

In this paper, we analyze user-item interactions from the perspective of causal graphs, which provide guidance for disentangling latent user preferences into invariant and environment-specific components. Motivated by recent advances in invariant representation learning, we propose IEPref (Invariant and Environment-Specific Preferences), a novel learning framework. IEPref consists of three main modules: (1) an Environment Classifier that leverages auxiliary information (e.g., timestamps) to infer the latent environment of each interaction, (2) an Invariant Preference (IPref) module that captures stable user-item preferences shared across environments, and (3) an Environment-Specific Proxy (EProxy) module that learns context-dependent patterns unique to each environment via specialized proxy modules. By combining both invariant and environment-specific modeling, IEPref enhances recommendation accuracy and robustness.

The main contributions of this work are as follows:

- We analyze the generation process of user-item interactions from the perspective of causal graphs and and disentangle user preferences into invariant and environment-specific preferences.
- We propose a novel invariant and environment-specific learning method (IEPref) that collaboratively utilizes auxiliary information to learn both invariant preferences and environment-specific preferences from training data, and we provide the corresponding theoretical proofs.
- We conduct extensive experiments on five real-world datasets, and the results consistently show that our method significantly outperforms state-of-the-art baselines in recommendation performance.

The remainder of this paper is organized as follows: Section II reviews the related work. Section III introduces the preliminaries, problem formulation, and the proposed IEPref framework. Section IV presents the experimental results and analysis. Section V concludes the paper and discusses future research directions.

II. Related Work

In this section, we comprehensively review invariant learning and unbiased recommendation.

A. Invariant learning

Contrary to the traditional Independent and Identically Distributed (IID) assumption, invariant learning [6] operates under the premise that real-world data often originates from multiple heterogeneous environments, each with distinct underlying distributions. The primary objective of invariant learning is to identify stable, causal features that remain consistent across these varying environments, thereby enhancing model robustness and generalization under distributional shifts. A foundational approach in this field is Invariant Risk Minimization (IRM), introduced by [7], which seeks to learn a single predictor that performs optimally across all environments by enforcing invariance in feature representations.

The theoretical underpinnings of IRM have been rigorously analyzed by [8], who established the conditions under which optimal invariant solutions exist, both in linear and nonlinear settings. Their findings highlight a critical limitation of IRM: its effectiveness heavily depends on the number and diversity of available training environments. Specifically, if the number of environments is insufficient or lacks meaningful heterogeneity, IRM may fail to identify truly invariant features, leading to degraded generalization performance.

To enhance robustness in extreme or limited-data scenarios, recent work has explored hybrid approaches that integrate invariant learning with other principles. For example, [9] combines IRM with the Information Bottleneck (IB) principle, leveraging IB's ability to compress noisy or irrelevant features while preserving predictive information. This hybrid method demonstrates strong performance even when invariant features alone do not fully capture label-relevant information, offering a more flexible solution for real-world applications.

A practical challenge in invariant learning is the need for predefined environment labels, which are often costly or infeasible to obtain. Addressing this, [10] proposes an automated framework that infers optimal environment partitions directly from data, enabling domain-invariant learning without manual annotation. Meanwhile, [6] approaches the problem from a heterogeneity perspective, introducing a novel clustering method to group data into meaningful environments that facilitate invariant feature extraction.

B. Unbiased recommendation

The observational nature of user behavior data inherently introduces various biases [11], including but not limited to selection bias, popularity bias, and exposure bias. These biases arise from the fact that observed interactions are influenced by external factors such as system recommendation [12], [13], item visibility, and user self-selection, rather than reflecting users' true preferences. To address these challenges, debiasing methods aim to bridge the gap between the biased observational data and the underlying unbiased data distribution, ensuring that recommendation models learn more accurate and fair representations of user preferences.

One prominent approach to debiasing is based on Inverse Propensity Scoring (IPS), which adjusts the weights of observed interactions to align with the unbiased distribution. For instance, [14] and [15] leverage IPS to correct selection bias by reweighting training samples according to their propensity scores. Similarly, [16] employ inverse probability weighting to mitigate popularity bias and positivity bias, ensuring that less popular or less frequently exposed items receive appropriate consideration. [17] further extend this idea by introducing a dual-IPS framework, where two propensity-weighted predictors generate pseudo-labels to refine the final recommendation model.

Beyond propensity-based methods, another line of research explores debiasing through information-theoretic principles.

[18] propose a framework based on the Information Bottleneck, which distills essential user preference signals while filtering out spurious correlations induced by bias. [19] further enhance this approach by imposing information-theoretic constraints to balance predictions between biased and unbiased data, promoting robustness against confounding factors.

Recently, invariant learning has emerged as a promising direction for unbiased recommendation, focusing on learning user preferences that remain stable across different environments. For example, [20] leverage invariance and disentanglement principles to isolate true user preferences from popularity-driven interactions, effectively filtering out bias. Similarly, [5] adopt the heterogeneous risk minimization framework. [6] partitioning data into multiple environments to learn invariant representations that generalize across different bias conditions. However, while these methods improve recommendation fairness, they often lack explicit constraints to account for the dynamic nature of user preferences.

III. Problem Statement and Methodology

Considering the dynamics of preferences, we first analyze the data generation process in the recommendation scenario. We then elaborate on the proposed framework of invariant and environment-dynamic preferences in detail.

A. Preliminary

Let $U = \{u_1, u_2, \ldots, u_{|U|}\}$ and $V = \{v_1, v_2, \ldots, v_{|V|}\}$ denote the set of users and items, respectively. The user-item interaction data is represented by a binary matrix $y \in \mathbb{R}^{|U|\times|V|}$, where each entry $y_{u,v} = 1$ indicates that user u has interacted with item v, and $y_{u,v} = 0$ otherwise. For brevity, we use u, v, and $y_{u,v}$ to denote the user, item, and interaction indicator, respectively.

B. Causal View of Recommendtion Task

The causal graph [21] that generates the observed user-item interactions can be represented in fig 2, where the edges indicate the direction of causal influences. The details are introduced as follows.

- **U** denotes the user, which contains the user's behavior history and profile (e.g. age and occupation).
- **I** denotes the item, which contains the history of feedback from users and the profile (e.g. price and category).
- $\mathbf{P_I}$ denotes the user's invariant preference, which is the user's lowest level of preference and does not change with the environment.
- $\mathbf{P_E}$ denotes the user's environment-specific preferences, and different environments have different environmental preferences.
- **E** denotes a set of environments, the one that really matters is the one to which the user and the item belong.
- **Y** denotes the observed user-item interaction, which is determined by invariant preferences $\mathrm{P_I}$ and environment-specific preferences $\mathrm{P_E}$.

The interaction Y between the user and the project is determined by invariant preferences $\mathrm{P_I}$ and environment-specific preferences $\mathrm{P_E}$. The dynamics of $\mathrm{P_E}$ is reflected in the number of environments E, which is an environment set $\{e_1, e_2, ..., e_3\}$, and the effect of E to $\mathrm{P_E}$ is actually the effect of e_i on $\mathrm{P_E}$ in the environment set.

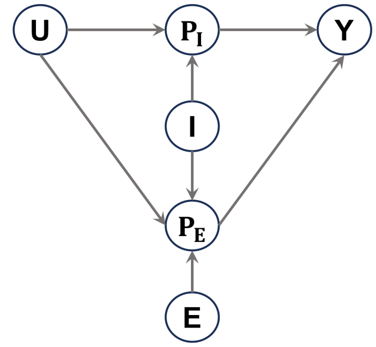


Fig. 2: Illustrates causal graph of data generation in recommendation. U denote the user, I denote the item, and E denote the environment. $\mathrm{P_I}$ denotes the user's invariant preference, and $\mathrm{P_E}$ denotes the user's environment-specific preference. The user-item interaction Y is jointly determined by $\mathrm{P_I}$ and $\mathrm{P_E}$.

Traditional debiasing methods [14] exploit correlation to fit data. The correlation between user-item and interaction have two sources: static causal relationships ($\mathrm{P_I}$ to Y) and dynamic causal relationships ($\mathrm{P_E}$ to Y). While traditional debiasing methods may also learn dynamic relationships, they do not take the influence of the environment E into account, ignoring biases caused by different environments. Recent invariant learning methods [5], [22], although they consider the influence of the environment E, only account for static causal relationships ($\mathrm{P_I}$ to Y), neglecting dynamic causal relationships ($\mathrm{P_E}$ to Y). Compared to traditional debiasing methods, they may remove some biases due to different environments to a certain extent. However, because they do not consider dynamic causal relationships ($\mathrm{P_E}$ to Y), their predictions lose accuracy. Additionally, their methods of partition environment [23] lack constraints. Based on this, we propose a new invariant and environment-specific preferences framework according to the defined causal graph to address the debiasing problem. We define the debiasing problem as follows:

Problem 1. *(Recommendation Debiasing Problem). Given the observational training data* $D_{train} = \{(y_j^i, u_j^i, i_j^i)\}_{j=1}^{n_e}$ *and test data* $D_{test} = \{(y_j^*, u_j^*, i_j^*)\}_{j=1}^{n_*}$*, where* D_{train} *is collected from multiple environments* $e_{train} \in \varepsilon$*, and the test sample* $\{(y_j^*, u_j^*, i_j^*)\}$ *comes from the environment* e_{test} *following the distribution* $P_{e_{test}}(Y, U, I)$ *of environment* e_{test}*, where* $e_{test} \in e_{train}$*. For all* $e_1, e_2 \in \varepsilon$*,* $e_1 \neq e_2 \iff P_{e_1}(Y, U, I) \neq P_{e_2}(Y, U, I)$*. The goal of recommendation is to infer the user's preference under the environment of the test data based on the training data, and leverage this information to accurately predict the user's feedback on items.*

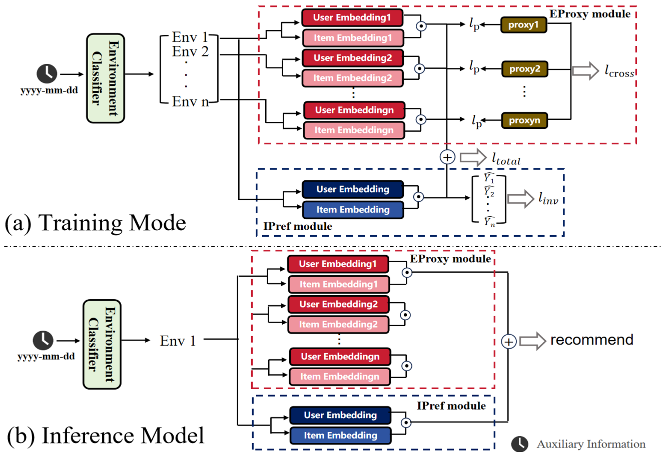


Fig. 3: illustrates the overall framework of IEPref, which consists of three main modules: the Environment Classifier leverages auxiliary information (e.g., time, timestamp) to infer the latent environment of each interaction; the Invariant Preference (IPref) module models user preferences that remain stable across different environments; and the Environment-specific Proxy (EProxy) module captures user preferences specific to each environment. Finally, the model integrates both types of preferences to make more accurate predictions of user-item interactions.

C. Proposed Framework

In order to solve the above debiasing problem in recommendation, we propose a debiasing framework named IEPref, which consists of three key modules.

1) Environment Classifier module: We consider that there exists additional auxiliary information $\boldsymbol{A} \in \mathbb{R}^{d_a}$ is associated with (U, I, Y), we propose a Classifier that leverages auxiliary information $\boldsymbol{A}$ for environmental inference, enabling invariant learning from the heterogeneous dataset D without environment partition. Furthermore, in the subsequent subsections, we establish the necessary and sufficient conditions for identifying invariance and provide the theoretical foundation for utilizing auxiliary information to perform environment partitioning.

We aim to learn a function $h(\cdot) : \mathbb{R}^{d_a} \rightarrow \mathbb{R}^K$ that assigns data to K environments.let $h^{(k)}(\cdot)$ denote the k-th entry of $h(\cdot)$, with $h(\boldsymbol{A}) \in [0, 1]^k$ and $\sum_k h^{(k)}(\boldsymbol{A})$ = 1.We set up two linear layers in the classifier to learn the environment to which the samples belong:

$$h(\boldsymbol{A}) = \sigma(h_2(h_1(\boldsymbol{A}))), \tag{1}$$

where h_1,h_2 are linear layers, σ is a softmax function.

2) Invariant Preference module: In order to model the invariant preferences of users across heterogeneous environments, we introduce an Invariant Preference module. This module captures invariant preferences based on the heterogeneous environments classified by the previously discussed environment classifier. Specifically, We use $\mathbf{u}^{inv} \in \mathbb{R}^L$ and $\mathbf{i}^{inv} \in \mathbb{R}^L$ to denote the invariant embeddings of user u and item i, respectively. Given user u and item i, the invariant preference $\mathbf{m}_{u,i}^{inv}$ of u for item i is modeled as Hadamard product of embedding $\mathbf{u}^{inv}$ and $\mathbf{i}^{inv}$:

$$\mathbf{m}_{u,i}^{inv} = \mathbf{u}^{inv} \odot \mathbf{i}^{inv}, \tag{2}$$

the invariant preference $\mathbf{m}_{u,i}^{inv}$ captures the underlying user preferences across different environments, characterizing the consistent interaction between user u and item v. From this, the feedback $y_{u,i}^{inv}$ is computed, which solely depends on the invariant preference, reflecting the stable preference of user u towards item i:

$$y_{u,i}^{inv} = \varphi(\mathbf{m}_{u,i}^{inv}), \tag{3}$$

where φ is the feedback function. then we define a cross-entropy loss function $\mathcal{L}_{inv}$ for $m_{u,i}^{inv}$ to capture invariant preferences:

$$\mathcal{L}_{inv} = \ell(y_{u,i}^{inv}, y_{u,i}). \tag{4}$$

To effectively capture the invariant user preference that is invariant across different environments, we need to ensure that the invariant preference $\mathbf{M}^{inv}$ maximally capture invariance, how can we measure this? Inspired by IRM, assuming the environments have been given by a classifier $h(\cdot)$, the measurement of the invariance of $\mathbf{M}^{inv}$ transforms into ensuring that the invariant preference $\mathbf{M}^{inv}$ is optimal in all environments. To measure the optimality of $\mathbf{M}^{inv}$ in the k-th environment, we can fit an environment-specific optimal preference $\mathbf{M}_k^{inv}$on the data from that environment. If $\mathbf{M}_k^{inv}$ achieves a smaller value, then we know that $\mathbf{M}^{inv}$ is not optimal in this environment. We can further train a set of environment-specific optimal preferences $\{\mathbf{M}^{inv}\}_{k=1}^{K}$ to measure whether the invariant preference $\mathbf{M}_k^{inv}$ is simultaneously optimal in all environments. Thus,we give its invariant penalty term as follows:

$$\begin{aligned}\mathcal{L}_{inv-penalty} = \sum_{k=1}^{K} [&h^{(k)}(\boldsymbol{A})\ell(\varphi(\mathbf{M}^{inv}), Y) \\ &- h^{(k)}(\boldsymbol{A})\ell(\varphi(\mathbf{M}_k^{inv}), Y)],\end{aligned} \tag{5}$$

next, we consider how to learn the partition function $h(\cdot)$. A good partition function should generate environments among which the environment-specific prefenece exhibit instability, so that there is a large penalty if $\mathbf{M}^{inv}$ extracts environment-specific prefenece. Thus, we seek for an environment partition that maximizes the invariance penalty. The overall framework is provided below:

$$\min_{\varphi, \mathbf{M}^{inv}} \max_{h, \{\mathbf{M}_1^{inv}, \ldots, \mathbf{M}_k^{inv}\}} \mathcal{L}(\varphi, \mathbf{M}^{inv}, \mathbf{M}_1^{inv}, \ldots, \mathbf{M}_k^{inv}, h), \tag{6}$$

since the above formula is challenging, we replace the invariant penalty term with its first order approximation. Specifically we replace it with the following formula:

$$\mathcal{L}_{inv-penalty} = \sum_{k=1}^{K} ||\nabla_{\mathbf{M}^{inv}} h^{(k)}(\boldsymbol{A})\ell(\varphi(\mathbf{M}^{inv}), Y)||^2. \tag{7}$$

3) Environment-specific Proxy module: In the IPref module we got the invariant preferences $\mathbf{P}_{\mathrm{E}}$ in the causal graph, in this module we need to get their environment-specific preferences $\mathbf{P}_{\mathrm{E}}$. Similar to the IPref module, we use $\mathbf{u}_k^{env} \in \mathbb{R}^L$ and $\mathbf{i}_k^{env} \in \mathbb{R}^L$ to denote the k-th denote the environment-specific embeddings of user u and item i for the k-th environment, respectively. Similarly we define the environment-specific preference $\mathbf{m}_k^{env}$ for the k-th environment we assign an environment preference to each environment individually:

$$\mathbf{m}_k^{env} = \mathbf{u}_k^{env} \odot \mathbf{i}_k^{env}, \tag{8}$$

$$\mathbf{M}^{env} = \{\mathbf{m}_k^{env}\}_{k=1}^{K}, \tag{9}$$

the environment-specific preference M^{env} captures preferences specific to each environment through data obtained from the classifier for each environment, then we can compute final feedback $y_{u,i}^{e}$ of the user u on item i, that is obtained from a combination of invariant preferences and environment-specific preferences:

$$y_{u,i}^{e} = \varphi(\mathbf{m}_{u,i}^{inv} \cdot \mathbf{m}_k^{env}), \tag{10}$$

where φ is the feedback function.

To achieve precise modeling of environment-specific preferences and facilitate environmental classification, we introduce a set of dedicated preference proxies $\{\boldsymbol{p}_i\}_{i=1}^{k}$, where each proxy $\boldsymbol{p}_i$ corresponds to the i-th environment. The proxies serve as reference points that guide the learning process for each environment, ensuring that preferences are effectively captured. We encourage samples from each environment to move closer to their corresponding proxy while being repelled from proxies associated with other environments, thereby ensuring that preferences within each environment converge toward their designated proxy. To formalize this approach, the metric loss is defined as:

$$\mathcal{L}_{\text{cross}} = -\log\left(\frac{\exp(\gamma \cdot s_{ik})}{\exp(\gamma \cdot s_{ik}) + \sum_{j\neq i}\exp(\gamma \cdot s_{ij})}\right), \tag{11}$$

where s_{ik} denotes the similarity between the environmental data points in environment i and their corresponding preference proxy, γ denotes the temperature coefficient. Given the training samples $\{\mathbf{u}_k^{env}, \mathbf{i}_k^{env}\}$ in k-th environment, along with their environment-specific preferences $\boldsymbol{m}_k^{env}$, we optimize the environment proxy $\boldsymbol{p}_i$ via the following objective. This optimization ensures that the learned proxy closely represents the preference structure of each environment:

$$\mathcal{L}_{\text{p}} = -\frac{1}{N_k}\sum_{i=1}^{N_k}\log\left(\frac{\exp\left(sim(\mathbf{m}_k^{env}, \boldsymbol{p}_k)\right)}{\sum_{j=1}^{k}\exp\left(sim(\mathbf{m}_k^{env}, \boldsymbol{p}_j)\right)}\right), \tag{12}$$

where N_k denotes the number of samples in the k th environment, and sim denotes cosine similarity.

To capture both invariant and environment-specific preferences, we define the overall loss function $\mathcal{L}_{total}$ as:

$$\mathcal{L}_{total} = \ell(y_{u,i}^{e}, y_{u,i}) + \mathcal{L}_{\text{cross}} + \mathcal{L}_{\text{p}}, \tag{13}$$

where ℓ is the cross-entropy loss between the predicted feedback $y_{u,i}^{e}$ and the observed feedback $y_{u,i}$. Finally, our objective is to minimize the following overall loss, which combines the contributions from invariant learning, environment-specific preference modeling, and the regularization of proxy learning:

$$min\mathcal{L} = \alpha\mathcal{L}_{total} + \beta\mathcal{L}_{inv} + \lambda\mathcal{L}_{inv-penalty}, \tag{14}$$

where α, β, λ are hyperparameters controlling the trade-offs among different components of the loss.

D. Proof

We explain why we need to use auxiliary information from a theoretical point of view. We start with a simple but general setting: $P = [P_I, P_E]$, $\boldsymbol{M}$ is the preference filter, and φ is a general non-linear function $H(\cdot,\cdot)$ is the cross-entropy loss.

Assumption 1. *For a given invariant preference model and any constant $\varepsilon > 0$, there exists $\varphi \in \Phi$ such that $\mathbb{E}[H(\varphi(\boldsymbol{M}(u,i), Y)] \leq H(Y|\boldsymbol{M}(u,i)) + \varepsilon$.*

Assumption 2. *If the invariant preference violates the invariance constraint, adding another environment-specific preference would not make the penalty vanish, i.e., there exists a constant $\delta > 0$ so that for environment-specific preference $P_{\mathrm{e}_1} \subset P_E$ and any preference $P_2 \subset P$, $H(Y|P_{e_1}, P_2) - H(Y|h(\boldsymbol{A}), P_{e_1}, P_2) \geq \delta(H(Y|P_{e_1}) - H(Y|h(\boldsymbol{A}), P_{e_1}))$.*

Assumption 3. *For any distinct preference P_1,P_2, we have the conditional entropy that $H(Y|P_1, P_2) \leq H(Y|P_1) - \gamma$ with fixed $\gamma > 0$.*

We next present our sufficient conditions to identify invariant preference.

Condition 1. *(Invariance Preserving Condition) Given invariant preference P_I and any function $h(\cdot)$, it holds that $H(Y|h(\boldsymbol{A})) = H(Y|P_I)$.*

Condition 2. *(Non-invariance Distinguishing Condition) For any environment-specific preference $P_{e_i} \in P_E$, there exists a function $h(\cdot)$ and a constant $C > 0$ such that $H(Y|P_{e_i}) - H(Y|P_{e_i}, h(\boldsymbol{A})) \geq C$.*

Condition 1 requires that invariant preference should remain invariant with respect to any environment partition induced by $h(\boldsymbol{A})$. Otherwise, if there exists a partition where an invariant preference becomes non-invariant, then this preference would induce a positive penalty. Condition 2 implies that for each environment-specific preference, there exists at least one partition so that this preference is non-invariant in the split environments. If a environment-specific preference does not incur any invariance penalty in all possible environment partitions, we can never distinguish it from true invariant preference.

Theorem 1 (Identifiability of Invariant Preference). *With the Assumptions 1-3 and the Conditions 1-2, if $\varepsilon < \frac{C\gamma\delta}{4\gamma+2C\delta H(Y)}$ and $\lambda \in [\frac{H(Y)+1/2\delta C}{\delta C-4\varepsilon}]$, then we have $\mathcal{L}(\hat{\boldsymbol{M}}_k) < \mathcal{L}(\hat{\boldsymbol{M}})$, where H(Y) denotes the entropy of Y. Thus, the solution to Equation 6 identifies invariant preference.*

IV. EXPERIMENTS

In this section, we conduct extensive experiments to answer the following questions:

- RQ1: How does the IEPref perform in comparison to other debiasing strategies?
- RQ2: To what extent do the different modules contribute to the effectiveness of the IEPref?
- RQ3: Does the IEPref learn true environmental partitioning?

A. Experimental Settings

we detail datasets used and baselines compared as follows:

1) Dataset: We evaluate the effectiveness of the proposed model on three widely-used real-world datasets: the KuaiRec dataset [24], the MovieLens dataset[1] (including MovieLens-1M, MovieLens-100K, and MovieLens-LatestSmall), and the MIND dataset[2].

For the KuaiRec dataset, the test set adopts a full-exposure setting, where all users are assumed to be exposed to the entire item pool during evaluation. For the MovieLens datasets, we follow the unbiased evaluation protocol proposed in [25], where the test data is constructed to simulate a realistic exposure scenario and accounts for 20 $\%$ of the total data volume. Both the KuaiRec and MovieLens datasets contain not only explicit user ratings on a 1-to-5 scale, but also interaction timestamps, which we leverage as auxiliary features for modeling environment context. Following standard practice, we consider user-item interactions with ratings ≥ 4 as positive samples, while the rest are either ignored or treated as unlabeled data.

The MIND dataset is a large-scale news recommendation benchmark, consisting of news articles and corresponding user click labels. It records both impression logs and click behaviors, making it suitable for modeling both exposure and interaction. During the testing phase, we adopt the public candidate set of news articles as the item pool for each user.

2) Baselines: We compared the following mainstream recommendation debiasing methods on the above two datasets:

- **IPS** [14]: A matrix factorization-based method that applies Inverse Propensity Scoring (IPS) to adjust for selection bias. It reweights observed interactions to approximate the unbiased user preference distribution.
- **SNIPS** [26]: An improved IPS variant that incorporates multiplicative control variates, which reduces the variance of IPS estimates and mitigates overfitting in the learned propensity models.
- **CVIB** [19]: A debiasing method that separates factual and counterfactual mutual information in the Information Bottleneck, introducing contrastive loss and confidence penalty to enhance generalization.
- **InvPref** [5]: A framework that decomposes user preferences into invariant and variant components across different environments. Only the invariant part is used for recommendation to ensure robustness to environmental shifts.
- **Fair** [27]: A regularization-based method designed to reduce popularity bias in learning-to-rank models. It encourages fairness by penalizing over-reliance on item popularity signals during training.
- **MACR** [25]: A multi-task learning framework that integrates causal estimation into the training process. It identifies and neutralizes popularity-induced biases through counterfactual reasoning during both training and inference.
- **WMF** [28]: The Weighted Matrix Factorization model, a classical method that interprets implicit user behavior as a combination of preference and confidence, and optimizes recommendations through alternating least squares.
- **EXMF** [29]: An exposure-aware matrix factorization model that treats user exposure as a latent variable, which is jointly inferred with user preferences to improve the overall recommendation quality.

3) Evaluation metrics: We measure the performance of the model using the following metrics. NDCG (Normalized Discounted Cumulative Gain) is a widely used metric in information retrieval and recommendation systems for evaluating the performance of ranking algorithms. NDCG@K specifically refers to the NDCG value calculated for the top K results, indicating the relevance of the top K recommendations. NDCG combines relevance and rank position, providing a comprehensive measure of the ranking effectiveness. The calculation of NDCG@K involves two main steps: computing DCG (Discounted Cumulative Gain) and IDCG (Ideal Discounted Cumulative Gain).

$$\text{NDCG@}K = \frac{\text{DCG@}K}{\text{IDCG@}K},$$

$$\text{DCG@}K = \sum_{i=1}^{K} \frac{rel_i}{\log_2(i+1)},$$

where DCG measures the cumulative gain of a result list, IDCG is the DCG value in the best possible case where all relevant results are perfectly ranked and NDCG normalizes DCG to a [0,1] range, eliminating the effects of varying result list lengths and total relevance.

Recall@K is a commonly used metric in information retrieval and recommendation systems to evaluate the performance of an algorithm. It measures the proportion of relevant items that are successfully retrieved within the top K recommendations.

$$\text{Recall@K}_i = \frac{|\text{Rel}_i \cap \text{Rec}_{i(K)}|}{|\text{Rel}_i|},$$

where Rel_i denotes the set of items related to user i, $\text{Rec}_{i(K)}$ denotes the set of recommended items of user i in the first K positions.

[1]https://movielens.org/

TABLE I: Comparative analysis of debias performance across different models on five benchmark datasets: MovieLens-1M, 100k, LatestSmall, MIND, and Kuairec.

Method	MovieLens-1M		MovieLens-100k		LatestSmall		MIND		Kuairec	
	NDCG@20	Recall@20	NDCG@20	Recall@20	NDCG@20	Recall@20	NDCG@20	Recall@20	NDCG@20	Recall@20
MF	0.087	0.081	0.032	0.074	0.041	0.102	0.239	0.479	0.536	0.522
FAIR	0.091	0.083	0.033	0.076	0.046	0.108	0.240	0.480	0.538	0.524
MACR	0.103	0.095	0.036	0.082	0.053	0.113	0.230	0.470	0.539	0.525
WMF	0.101	0.091	0.034	0.076	0.044	0.106	**0.254**	0.501	0.539	0.523
EXMF	0.093	0.082	0.032	0.072	0.041	0.102	0.246	0.481	0.537	0.522
IPS	0.102	0.082	0.033	0.073	0.046	0.106	0.240	0.482	0.543	0.526
SNIPS	0.104	0.086	0.034	0.075	0.048	0.109	0.241	0.485	0.545	0.525
CVIB	0.085	0.084	0.031	0.072	0.040	0.104	0.240	0.480	0.549	0.528
InvPref	0.089	0.085	0.035	0.077	0.051	0.110	0.253	0.509	0.553	0.539
IEPref	**0.108**	**0.118**	**0.040**	**0.086**	**0.054**	**0.115**	0.252	**0.514**	**0.554**	**0.542**
Improvement	+3.8%	+24.2%	+11.1%	+4.9%	+1.9%	+1.8%	-0.4%	+1.0%	+0.2%	+0.6%

B. Performance Comparison(RQ1)

We first compare the ranking performance of IEPref with a suite of representative baseline models. Table I reports the NDCG@20 and Recall@20 scores on five widely used benchmark datasets: MovieLens-1M, MovieLens-100K, MovieLens-LatestSmall, MIND, and KuaiRec. The best result for each metric is highlighted in bold, and the second-best is underlined. Our key observations are as follows:

- The result demonstrates that the IEPref model consistently outperforms the baselines regarding Recall and NDCG across various datasets and evaluation metrics, indicating its superior ability to recommend relevant next items to users. Remarkably, IEPref substantially improves Recall and NDCG compared to the baselines.
- Compared to traditional matrix factorization models such as MF, WMF, and EXMF, IEPref demonstrates substantial improvements. This is especially apparent on MovieLens-100K and LatestSmall.
- Traditional debiasing recommendation algorithms, such as IPS and SNIPS, often outperform basic matrix factorization (MF) methods, but their effectiveness tends to be limited to specific datasets. In contrast, invariant learning-based approaches like InvPref demonstrate strong performance across multiple datasets by introducing invariant preferences.
- IEPref further advances this line of work by incorporating an Environment-specific Proxy module, which enables the model to capture dynamic environment-specific preferences while preserving invariant preferences. Compared to traditional invariant recommendation methods, IEPref achieves superior recommendation performance.

C. Ablation Study (RQ2)

Table II shows the ablation study of IEPref. We investigate the impact of both the environment classifier and invariant preferences (I) on the effectiveness of IEPref. Table II presents the ablation study of IEPref, where we compare two important variants: (1) IEPref w/o I (removal of invariant preferences) and (2) IEPref w/o E (removal of environmental preferences). Our key observations from this study are as follows:

- The ablation results highlight that when environmental preferences are removed (IEPref w/o E), the model continues to outperform InvPref, showcasing the effectiveness of the invariant learning method that incorporates additional information. However, the overall performance is reduced, underscoring the importance of including environmental dynamics into the model. This indicates that environmental preferences play a significant role in capturing dynamic user behaviors and improving prediction accuracy.
- In the case where invariant preferences are removed (IEPref w/o I), the model's performance significantly deteriorates. This demonstrates that invariant preferences, which capture stable, long-term user preferences, are crucial for enhancing the recommendation system. The results confirm our analysis that invariant preferences contribute significantly to the overall performance improvement, and without them, the model loses a key component of its predictive power.
- In addition to the ablation study, we further evaluate the impact of invariant and environmental preferences by adjusting key hyperparameters: λ, α, and β. These parameters respectively control the strength of change preferences, constant preferences, and environmental preferences. Our analysis of model performance, specifically with respect to NDCG@10, shows that the optimal values for α and β are at 10, while the optimal value of λ is 1000. This suggests that the strengths of constant preferences and environmental preferences are not the same, and the model achieves the best performance by balancing between them. This balance ensures that both stable (invariant) preferences and dynamic (environmental) preferences are captured effectively, leading to a more robust and accurate recommendation system.

D. In-depth Analysis (RQ3)

We conducted a sensitivity analysis regarding the number of environments, as shown in Fig 5 (a). The blue line corresponds to the full model (IEPref), and the red line represents the variant without environment modeling (IEPref w/o E). As the number of environments increases, the performance of IEPref

TABLE II: Ablation Study on MovieLens and Kuairec datasets.

Model	MovieLens				Kuairec			
	NDCG@20	NDCG@30	Recall@20	Recall@30	NDCG@20	NDCG@30	Recall@20	Recall@30
IEPref w/o I	0.096	0.101	0.103	0.119	0.526	0.587	0.520	0.641
IEPref w/o E	0.101	0.106	0.112	0.123	0.538	0.593	0.517	0.650
IEPref	**0.108**	**0.112**	**0.118**	**0.132**	**0.554**	**0.609**	**0.542**	**0.675**
Improvement	+6.9%	+5.7%	+5.4%	+7.3%	+3.0%	+2.7%	+4.8%	+3.8%

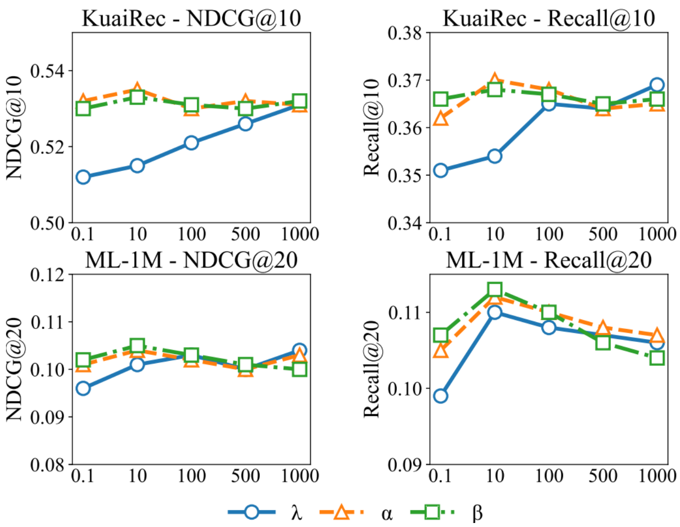


Fig. 4: Hyperparametric sensitivity analysis.

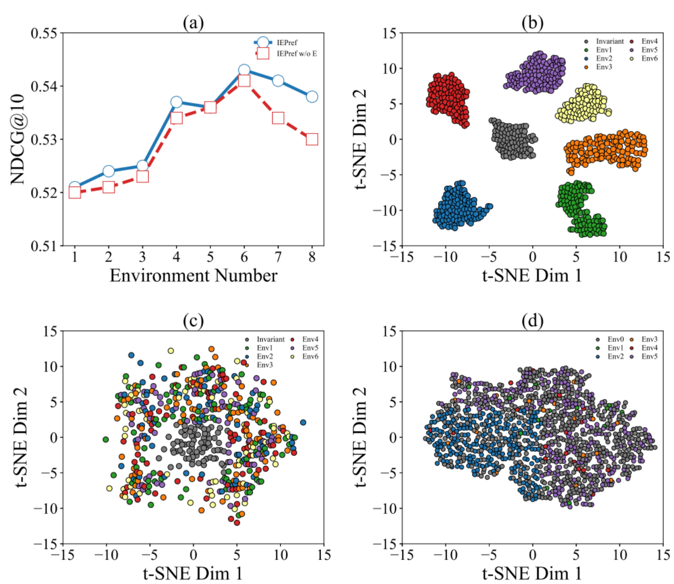


Fig. 5: (a) shows the sensitivity analysis with respect to the number of environmental variables. (b), (c), and (d) present t-SNE visualizations of preference representations on the Kuairec dataset. Specifically, (b) corresponds to our proposed method, (c) shows the result after removing environmental proxies, and (d) illustrates the visualization of the InvPref method.

w/o E improves, but it begins to degrade when the number of environments becomes excessively large. While the trends for IEPref and IEPref w/o E are generally similar, a significant divergence occurs when the number of environments is 4 or 5: IEPref w/o E continues to improve, while IEPref starts to decline. This suggests that the benefit of modeling preference variation is not linearly scalable with the number of environments, indicating the need for an optimal balance in the number of environments.

To further illustrate how environment modeling affects user preference representations, we provide a set of visualizations on the KuaiRec dataset under different settings:

- When using our proposed IEPref model with environment modeling enabled (number of environments set to 6 in Fig 5 (b)), we observe a clear separation between environment-specific preferences in the latent space. This demonstrates that the model effectively learns and differentiates user preferences corresponding to different environments.
- In the absence of the environment-specific proxy(Fig 5 (c)), the representation of user preferences becomes noticeably more entangled. This suggests that removing the proxy module weakens the model's ability to isolate and model environment-dependent variations in user behavior.
- The visualization from InvPref(Fig 5 (d)), which captures only invariant preferences, shows that all environments are collapsed into a shared invariant preference space. The invariant preference vector lies approximately at the center of all clusters, reflecting its theoretical invariance. Interestingly, environment 0 appears to encompass the entire invariant space, while environments 1–5 effectively merge into environment 0. This indicates that relying solely on invariant preferences fails to adapt to the nuanced differences among environment-specific behaviors.

V. Conclusion

In this paper, we proposed the Invariant and Environment-speciffc Preferences (IEPref) framework aimed at addressing the challenges of unbiased recommendation in real-world scenarios. Traditional recommendation methods often face diffffculties when dealing with data originating from mixed or shifting distributions, which hinders their ability to accurately capture users' true underlying preferences. To overcome this limitation, our approach employs causal graph techniques to explicitly distinguish between invariant user preferences that remain stable across different contexts and environment-speciffc preferences that vary depending on dynamic factors.

By leveraging auxiliary information such as timestamps and contextual signals, IEPref effectively partitions user interactions into distinct environments, allowing the model to learn both stable and dynamic components of user preferences. This dual modeling strategy enables more precise and personalized recommendations, as the model can adapt to preference changes induced by varying environmental conditions without losing sight of core user interests.

Looking ahead, future work could focus on extending the IEPref framework in several promising directions. One potential avenue is to integrate IEPref with other advanced recommendation paradigms, such as sequential recommendation and session-based recommendation systems, where user preferences evolve more rapidly and exhibit more complex temporal dependencies. Additionally, exploring the application of IEPref in cross-domain recommendation or multi-modal recommendation settings could further enhance its generalizability and practical impact. Finally, investigating more sophisticated environment partitioning strategies and causal inference techniques may lead to even more accurate disentanglement of user preferences and improved recommendation quality.

VI. Acknowledgment

This work is supported by the Natural Science Foundation of China No. 62472196, Jilin Science and Technology Research Project 20230101067JC, National Key R&D Program of China under Grant No. 2021ZD0112501 and 2021ZD0112502, National Natural Science Foundation of China under Grant No. 62272193, National Key R&D Program of China under Grant Nos. 2022YFB3103700 and 2022YFB3103702.

References

[1] W. Guo, F. Zhuang, X. Zhang, Y. Tong, and J. Dong, "A comprehensive survey of federated transfer learning: Challenges, methods and applications," *Frontiers Comput. Sci.*, vol. 18, no. 6, p. 186 356, 2024.

[2] Y. Xu, E. Wang, Y. Yang, and Y. Chang, "A unified collaborative representation learning for neural-network based recommender systems," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 11, pp. 5126–5139, 2022.

[3] Y. Jiang, Y. Yang, Y. Xu, and E. Wang, "Spatial-temporal interval aware individual future trajectory prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 10, pp. 5374–5387, 2024.

[4] Y. He, Z. Wang, P. Cui, *et al.*, "Causpref: Causal preference learning for out-of-distribution recommendation," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 410–421.

[5] Z. Wang, Y. He, J. Liu, W. Zou, P. S. Yu, and P. Cui, "Invariant preference learning for general debiasing in recommendation," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 1969–1978.

[6] J. Liu, Z. Hu, P. Cui, B. Li, and Z. Shen, *Heterogeneous risk minimization*, 2021. arXiv: 2105.03818 [cs.LG].

[7] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, *Invariant risk minimization*, 2020. arXiv: 1907.02893 [stat.ML].

[8] E. Rosenfeld, P. Ravikumar, and A. Risteski, "The risks of invariant risk minimization," *arXiv preprint arXiv:2010.05761*, 2020.

[9] K. Ahuja, E. Caballero, D. Zhang, *et al.*, "Invariance principle meets information bottleneck for out-of-distribution generalization," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, Curran Associates, Inc., 2021, pp. 3438–3450.

[10] E. Creager, J.-H. Jacobsen, and R. Zemel, "Environment inference for invariant learning," in *International Conference on Machine Learning*, PMLR, 2021, pp. 2189–2200.

[11] H. Xu, Y. Xu, and Y. Yang, "Separating and learning latent confounders to enhancing user preferences modeling," in *Database Systems for Advanced Applications - 29th International Conference, DASFAA 2024, Gifu, Japan, July 2-5, 2024, Proceedings, Part III*, M. Onizuka, J. Lee, Y. Tong, *et al.*, Eds., ser. Lecture Notes in Computer Science, vol. 14852, Springer, 2024, pp. 67–82.

[12] H. Xu, Y. Xu, C. Li, and F. Zhuang, "Causal structure representation learning of unobserved confounders in latent space for recommendation," *ACM Transactions on Information Systems*, 2025.

[13] H. Xu, Y. Xu, H. Liu, and E. Wang, "Flow-based time-aware causal structure learning for sequential recommendation,"

[14] T. Schnabel, A. Swaminathan, A. Singh, N. Chandak, and T. Joachims, "Recommendations as treatments: Debiasing learning and evaluation," in *Proceedings of The 33rd International Conference on Machine Learning*, M. F. Balcan and K. Q. Weinberger, Eds., ser. Proceedings of Machine Learning Research, vol. 48, New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 1670–1679.

[15] 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.

[16] J. Huang, H. Oosterhuis, M. De Rijke, and H. Van Hoof, "Keeping dataset biases out of the simulation: A debiased simulator for reinforcement learning based recommender systems," in *Proceedings of the 14th ACM conference on recommender systems*, 2020, pp. 190–199.

[17] Y. Saito, "Asymmetric tri-training for debiasing missing-not-at-random explicit feedback," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 309–318.

[18] D. Liu, P. Cheng, Z. Dong, X. He, W. Pan, and Z. Ming, "A general knowledge distillation framework for counterfactual recommendation via uniform data," in *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, 2020, pp. 831–840.

[19] Z. Wang, X. Chen, R. Wen, S.-L. Huang, E. Kuruoglu, and Y. Zheng, "Information theoretic counterfactual learning from missing-not-at-random feedback," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1854–1864, 2020.

[20] A. Zhang, J. Zheng, X. Wang, Y. Yuan, and T.-S. Chua, "Invariant collaborative filtering to popularity distribution shift," in *Proceedings of the ACM Web Conference 2023*, ser. WWW '23, Austin, TX, USA: Association for Computing Machinery, 2023, pp. 1240–1251, ISBN: 9781450394161.

[21] H. Luo, F. Zhuang, R. Xie, H. Zhu, and D. Wang, "A survey on causal inference for recommendation," *CoRR*, vol. abs/2303.11666, 2023.

[22] H. Pan, J. Chen, F. Feng, W. Shi, J. Wu, and X. He, "Discriminative-invariant representation learning for unbiased recommendation," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, E. Elkind, Ed., Main Track, International Joint Conferences on Artificial Intelligence Organization, Aug. 2023, pp. 2270–2278.

[23] Y. Lin, S. Zhu, L. Tan, and P. Cui, "Zin: When and how to learn invariance without environment partition?" In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., Curran Associates, Inc., pp. 24 529–24 542.

[24] W. Kweon and H. Yu, "Doubly calibrated estimator for recommendation on data missing not at random," in *Proceedings of the ACM on Web Conference 2024*, 2024, pp. 3810–3820.

[25] T. Wei, F. Feng, J. Chen, Z. Wu, J. Yi, and X. He, "Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 1791–1800.

[26] A. Swaminathan and T. Joachims, "The self-normalized estimator for counterfactual learning," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28, Curran Associates, Inc., 2015.

[27] H. Abdollahpouri, R. Burke, and B. Mobasher, "Controlling popularity bias in learning-to-rank recommendation," in *Proceedings of the eleventh ACM conference on recommender systems*, 2017, pp. 42–46.

[28] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," in *2008 Eighth IEEE international conference on data mining*, Ieee, 2008, pp. 263–272.

[29] D. Liang, L. Charlin, J. McInerney, and D. M. Blei, "Modeling user exposure in recommendation," in *Proceedings of the 25th international conference on World Wide Web*, 2016, pp. 951–961.