Unveiling the Impact of Multi-modal Content in Multi-modal Recommender Systems

Guipeng Xv*
xuguipeng@stu.xmu.edu.cn
School of Informatics, Xiamen University
Xiamen, China

Xinyu Li*
xinyuli@stu.xmu.edu.cn
School of Informatics, Xiamen University
Xiamen, China

Yi Liu
yisulphur@stu.xmu.edu.cn
Institute of Artificial Intelligence, Xiamen University
Xiamen, China

Chen Lin†
chenlin@xmu.edu.cn
School of Informatics, Xiamen University
Xiamen, China

Xiaoli Wang
xlwang@xmu.edu.cn
School of Informatics, Xiamen University
Xiamen, China

Abstract

Multi-modal recommender systems (MRSs) have emerged as critical multi-modal technologies on online platforms, but *do we truly leverage multi-modal content properly*? Through an empirical study of four diverse, real-world datasets spanning various recommendation scenarios, we observe that MRSs exhibit a stronger tendency to recommend items with high similarity to users' past interactions in terms of multi-modal content than conventional RSs. While this tendency improves the recommendation accuracy, it introduces a previously unexplored bias that significantly impacts user experience. We define this bias as **`User-side Content Bias`**: *users who prefer items similar to their historical choices receive higher quality recommendations than those seeking diverse options*. We show that `User-side Content Bias` is unrelated to the activity of users, indicating a fundamental limitation in current MRSs. We propose **ISOLATOR**: ut**I**lizing u**S**er-side c**O**ntent simi**L**arity via a model-**A**gnos**T**ic framew**OR**k to leverage multi-modal content more properly. ISOLATOR estimates the impact of `User-side Content Similarity` and proposes two intervention strategies to meet the needs for more accurate and unbiased recommendations. Extensive evaluations on several widely-used datasets demonstrate that ISOLATOR consistently improves various state-of-the-art MRSs and effectively addresses the `User-side Content Bias`. We provide our code at anonymous link.

CCS Concepts

• **Information systems** → **Multimedia and multimodal retrieval**; **Recommender systems**.

Keywords

Multi-modal Recommender System, User-side Content Bias, Causal Inference

1 Introduction

Multi-modal Recommender Systems (MRSs) have garnered considerable interest in recent years [14, 20, 38, 46]. It is well regarded [37, 43, 47] that multi-modal content enhances the overall understanding of item and improves user preference prediction compared with recommender systems using only user feedback or single-modal content. Therefore, as critical multi-modal technologies, MRSs are widely employed on online platforms, such as E-commerce [12, 18, 19, 35] and video sharing [1, 2, 29, 41] platforms.

Current MRSs primarily focus on leveraging advanced techniques to integrate multi-modal and behavioral features, with the goal of enhancing recommendation accuracy [14, 26, 31, 36]. However, there is a critical question that remains insufficiently addressed —*Do we truly leverage multi-modal content properly*?

To answer this question, we conduct a cross-domain empirical study on four real-world, widely-used datasets from E-commerce [22] and short-form video [23] platforms. We find that existing MRSs incorporate multi-modal content **with positive and negative impacts**. (1) On the one hand, compared with conventional RSs, MRSs recommend items more similar to users' past interactions regarding multi-modal content. We observe that the overall recommendation performance (i.e., recall rate) positively correlates with the multi-modal content similarity between items in the recommendation list and the user's historical interactions (Sec. 3.3). (2) On the other hand, including multi-modal content threatens the fairness of recommendations from the user's perspective. We create user groups, i.e., `Content-Consistent Users` (who prefer items similar to their past interactions) and `Content-Diverse Users` (who prefer items distinct from their past interactions). We observe that MRSs amplify the performance gap between `Content-Consistent Users` and `Content-Diverse Users`, referred to as `User-side Content Bias` (Sec. 3.4). This phenomenon brings a new fairness issue, i.e., `User-side Content Bias` is independent of the revealed unfairness on inactive users [13]. With the same number of interactions, `Content-Consistent Users` still receive higher quality recommendations than `Content-Diverse Users` (Sec. 3.5).

Given the widespread presence of MRSs of different model architectures, it is crucial to develop a generally applicable framework to balance the positive and negative impacts of multi-modal content, considering the specific goals and context of the RSs. For example, when prioritizing equal experience for all users, `User-side Content Similarity` should be discarded completely to avoid

*The authors contribute equally.
†Corresponding author.

`User-side Content Bias`. On the contrary, when recommendation accuracy is pivotal in user satisfaction and engagement, `User-side Content Similarity` should be selectively utilized to grasp the item relations and improve overall accuracy.

To properly leverage multi-modal content, we construct a causal graph [25] to analyze how the `User-side Content Similarity` influences the recommendations, and propose **ISOLATOR**: ut**I**lizing u**S**er-side c**O**ntent Simi**L**arity via a Model-**A**gnos**T**ic Framew**OR**k. In the training stage, ISOLATOR estimates the impact of `User-side Content Similarity` on recommendation accuracy and employs a *do-calculus* [25] for each user-item pair. In the inference stage, ISOLATOR proposes two strategies tailored for different recommendation scenarios: a Debiasing intervention strategy designed to *eliminate the impact* for all users to mitigate the `User-side Content Bias`; and a User-specific intervention strategy, which *leverages the impact* by applying personalized interventions for different users to enhance accuracy.

Our main contributions can be summarized as follows:

(1) We find that multi-modal content leads the MRSs to recommend items more similar to users' past interactions, resulting in amplifying `User-side Content Bias`. To our knowledge, **this is the first work to investigate `User-side Content Bias` in MRSs**.
(2) We propose ISOLATOR to use multi-modal content properly. This model-agnostic framework encompasses two strategies, one for debiasing and the other for performance enhancement, to meet different recommendation requirements. To our knowledge, **this is the first work to mitigate `User-side Content Bias` in MRSs**.
(3) Extensive experiments on several real-world datasets and widely-used MRSs verify that ISOLATOR not only improves the recommendation performance but also effectively mitigates the bias.

2 Related Work[1]

Multi-modal Recommender Systems. Multi-Modal Recommender Systems (MRSs) predict user preferences by fusing multi-modal and behavioral features, primarily categorized into two frameworks: (1) Matrix Factorization (MF)-based methods [3, 8], which integrate multi-modal item characteristics into MF training. (2) Graph Collaborative Filtering (GCF)-based methods, which can be further categorized into two subtypes. The first subtype integrates multi-modal features into the user-item graph [31, 36, 37], while the second leverages multi-modal content to construct item-item similarity graphs for embedding refinement [17, 26, 38, 46].

Bias in Multi-modal Recommender Systems. Biases have been widely studied in recommender systems, particularly those arising from feedback data and conventional recommendation algorithm [4, 11, 16, 21, 39]. However, **biases introduced by multi-modal content are more critical in MRSs**, which can be categorized into *inter-modality bias* and *item-side bias*. Inter-modality bias arises from challenges in integrating diverse modalities, such as MRSs overly rely on dominant modalities [20], leading to suboptimal recommendations. Item-side bias, on the other hand, occurs when multi-modal content causes unfair phenomena to items, such as over-recommendation of items with specific content [28] or increasing item-side popularity bias [21].

Remarks. ISOLATOR systematically explores the impact of multi-modal content and **discovers a new type of bias**, i.e., **`User-side Content Bias`**. Different from current debiasing efforts in recommender systems, ISOLATOR considers **the bias from multi-modal content in user-side** due to varying user preferences for multi-modal content. To our knowledge, ISOLATOR **is the first work** that explores the impact of multi-modal content on the user side and **mitigates the `User-side Content Bias` in MRSs**.

Table 1: Statistics of the datasets. $|\mathcal{D}|$, $|\mathcal{U}|$, and $|\mathcal{I}|$ represent the number of interactions, users, and items. $\overline{S^v}$ and $\overline{S^t}$ represents the average visual and textual similarity.

Datasets	$\lvert\mathcal{D}\rvert$	$\lvert\mathcal{U}\rvert$	$\lvert\mathcal{I}\rvert$	$\overline{S^v}$	$\overline{S^t}$	Sparsity
Baby	160,792	19,445	7,050	0.2239	0.2626	99.88%
Sports	296,337	35,598	18,357	0.2183	0.2084	99.95%
Clothing	278,677	39,387	23,033	0.2239	0.3880	99.97%
Microlens	705,174	98,129	17,228	0.5078	0.3822	99.96%

3 Empirical Study

3.1 Preliminaries

Let $\mathcal{U} = \{u_1, u_2, ..., u_{|\mathcal{U}|}\}$ and $\mathcal{I} = \{i_1, i_2, ..., i_{|\mathcal{I}|}\}$ denote the set of users and items, respectively. $|\mathcal{U}|$ and $|\mathcal{I}|$ is the number of users and items. For each user u, let $\mathcal{I}_u \subseteq \mathcal{I}$ be the set of items that u has interacted with in the training set, and let $|\mathcal{I}_u|$ denote its size. Each item i has multi-modal content features $\boldsymbol{e}_i^m \in \mathbb{R}^{d_m}$, where d is the dimension of the features, $m \in \mathcal{M}$ is the modality[2], and $\mathcal{M}$ is the set of modalities. To facilitate similarity computations, these content feature vectors are typically L_2-normalized, i.e., $||\boldsymbol{e}||_2 = 1$.

The goal of multi-modal recommender systems (MRSs) is to generate a ranked list of potential recommendations that each user u may prefer by predicting the user-item preference score $\widehat{r_{u,i}}$ using both behavior and multi-modal content. Formally,

$$\widehat{\mathcal{I}_u} = \text{Top@}k_{i \in \{\mathcal{I} \setminus \mathcal{I}_u\}} \widehat{r_{u,i}}, \tag{1}$$

where $\widehat{\mathcal{I}_u}$ denotes the recommendation list for user u, and the length of recommendation list $|\widehat{\mathcal{I}_u}| = k$.

3.2 Empirical Study Protocol

3.2.1 Datasets. We conduct experiments on four widely-used public datasets following [14, 17, 26, 38, 42, 46], including three Amazon collections (namely Baby, Sports, and Clothing) [22] and the Microlens [23] dataset. These datasets were chosen due to several advantageous characteristics: (1) First, they undergo a public unified preprocessing pipeline that *standardizes both the dataset structure and the modal content*[3]; (2) Second, they originate from E-commerce platforms and short-form video platforms, which *are two representative multi-modal recommendation scenarios*; (3) Third, they offer *diverse characteristics* in terms of size, sparsity, and multi-modal properties. These advantages enable **more robust and fair comparisons**. The datasets are filtered using a 5-core criterion to

[1] Due to limited space, the detailed related work is provided in Section A of the supplementary materials.

[2] We use textual and visual modalities in this work, but it can also be extended to other modalities.

[3] https://github.com/enoche/MMRec/tree/master/data

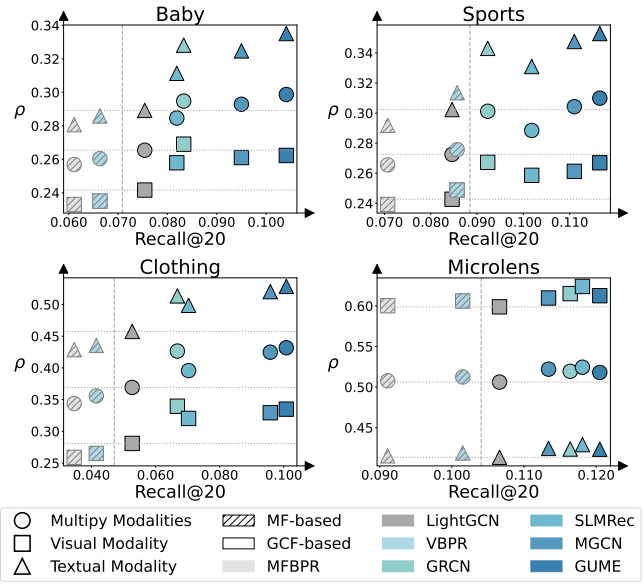


Figure 1: Relationship Between *Recall*@20 and the average content similarity of recommendation lists to users' past interactions ρ in different modalities.

ensure sufficient feedback for each user and item and are split into training, validation, and test sets in an 8:1:1 ratio. We compute the average pairwise cosine similarity across all items in each modality, denoted as $\overline{S^v}$ for images and $\overline{S^t}$ for text. Table 1 lists the statistics of the four datasets.

3.2.2 Evaluation metrics. We employ *Recall*@20 to measure the recommendation accuracy [8, 17, 26, 31, 36]. *Recall*@20 computes the ratio of relevant items for a user in the testing set appearing in the top twenty recommendations made by the baseline. Higher values of *Recall*@20 indicate more accurate recommendation results.

3.2.3 Baselines. Multi-modal recommender systems (MRSs) typically integrate multi-modal content into conventional RSs based on two primary backbone architectures: matrix factorization (MF) and graph collaborative filtering (GCF). We use two conventional RSs MFBPR [27] and LightGCN [9], and five MRSs VBPR [8], GRCN [36], MGCN [26], SLMRec [31] and GUME [17] as backbone since they are widely used[4]. It should be noted that *MFBPR and VBPR are MF architectures, while the remaining are GCF architectures.* To ensure optimal performance, we use publicly available code[5] [45] with default parameter settings. We train the models using the training set, save the best model based on *Recall*@20 on the validation set, and use the models to deliver recommendations on the testing set.

3.3 Positive Impact of Content Similarity

`User-side Content Similarity`. Essentially, multi-modal recommender systems (MRSs) match items with a user by taking into account the multi-modal content similarity between the candidate item and the items that the user previously interacted with, which we refer to as `User-side Content Similarity`.

[4]Due to limited space, the detailed baseline introductions are provided in Section B.1 of the supplementary materials.
[5]https://github.com/enoche/MMRec

Thus, to investigate the impact of multi-modal content, we monitor the relation between `User-side Content Similarity` and recommendation performance on different baselines. We use the average content similarity of the recommendation list to users' past interactions (referred to as `Average Content Similarity` ρ) to reflect the overall recommendations of the model.

$$\rho = \frac{\sum_{u\in\mathcal{U}}\left(\sum_{m\in\mathcal{M}}\left(\sum_{i\in\mathcal{I}_u}\sum_{j\in\widehat{\mathcal{I}_u}}\left(\text{sim}(\boldsymbol{e}_i^m,\boldsymbol{e}_j^m)\right)\right)\right)}{|\mathcal{U}|\times|\mathcal{M}|\times|\mathcal{I}_u|\times|\widehat{\mathcal{I}_u}|}, \tag{2}$$

where $\text{sim}(\cdot)$ is the pairwise cosine similarity, $\boldsymbol{e}^m$ is the raw features of modality m. ρ is averaged over all users, all modalities, and all items in the recommendation list for a user. A higher ρ signifies that the recommended items are closer to the user's past interactions regarding multi-modal content, meaning that the baseline is more reliant on multi-modal content similarity.

In Fig. 1, the blue markers represent multi-modal recommender systems (MRSs) while the gray markers represent conventional recommender systems. We have the following observations:

- Comparing methods based on the same backbone architecture, MRSs tend to recommend items with higher content similarity than conventional RSs. For example, the blue markers (MRSs) are typically positioned higher than the gray markers (conventional RSs) of the same shape in Fig. 1, showing that MRSs focus on utilizing content similarity more. This observation is consistent across all tested baselines and datasets, and the similarity can be measured based on either single or multiple modalities.
- MRSs consistently exhibit higher recommendation accuracy than conventional RSs using the same backbone architecture. The blue markers (MRSs) are generally positioned to the right of the gray markers (conventional RSs) of the same shape in Fig. 1, showing that MRSs outperform conventional RSs regarding the *Recall*@20 across all datasets.
- Higher values of `Average Content Similarity` generally lead to improved recommendation accuracy. In the Baby, Sports, and Clothes datasets, there is a positive correlation between accuracy (*Recall*@20) and `Average Content Similarity` (ρ). In the Microlens dataset, the largest ρ is observed on the best-performing baselines (i.e., MGCN and GUME) and the smallest ρ on the worst-performing baseline (i.e., MFBPR).
- Multi-modal content can more stably reflect content similarity than single-modal content. Some datasets show greater similarities in textual contents, while in other datasets, the image contents are more similar. Specifically, textual similarity (triangular markers in Fig. 1) is relatively high in e-commerce scenarios but lower in short-form video scenarios. Therefore, using single-modal content similarity is inaccurate for certain datasets, and we should use multi-modal content similarity as a domain-robust assessment.

3.4 Negative Impact of Content Similarity

We have shown that the overall recommendation performance is improved by recommending items that have a higher multi-modal content similarity with the user's past interactions. However, users show significantly different preference patterns. Some users prefer items with highly similar content, such as those who favor consistent vintage styles in clothing, accessories, and home goods. In

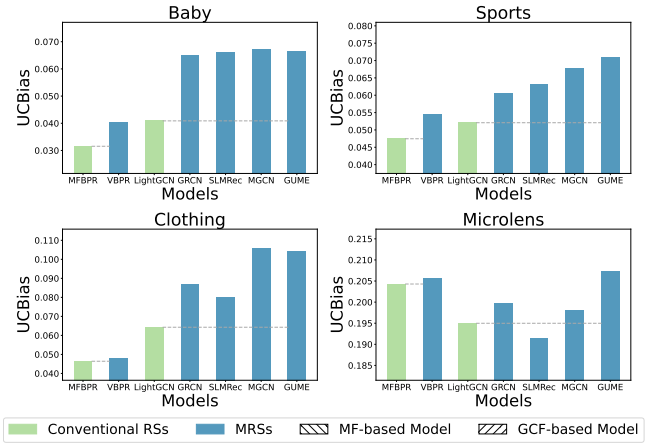


Figure 2: User-side Content Bias between `Content-Consistent Users` and `Content-Diverse Users`.

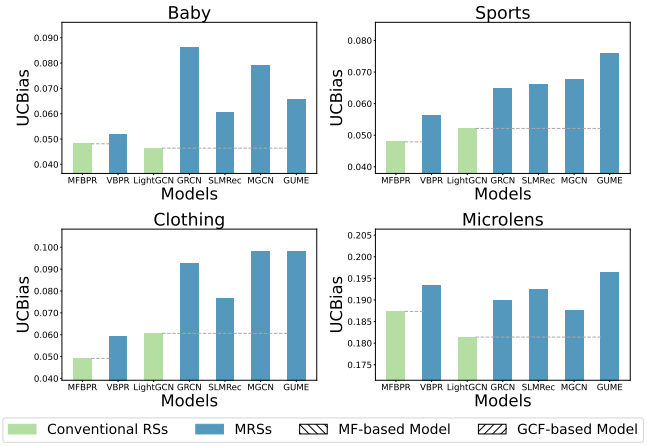


Figure 3: User-side Content Bias between `Content-Consistent Users` and `Content-Diverse Users` with the same number of interactions.

contrast, some users prefer exploring items with diverse content, such as those who enjoy transcending different styles and combining various items that may seem unrelated at first glance. In this subsection, we investigate the negative impact of `User-side Content Similarity`, i.e., whether it will cause biased performance on different users.

User Grouping. We divide users according to their preference patterns. Firstly, we compute the average pairwise content similarity of interacted items for each user (denoted as c_u):

$$c_u = \frac{\sum_{m\in\mathcal{M}}\sum_{i\in\mathcal{I}_u}\sum_{j\in\mathcal{I}_u, i\neq j}\left(\text{sim}(\boldsymbol{e}_i^m, \boldsymbol{e}_j^m)\right)}{|\mathcal{M}|\times|\mathcal{I}_u|\times(|\mathcal{I}_u|-1)}, \tag{3}$$

where $\text{sim}(\cdot)$ is the pairwise cosine similarity.

Secondly, based on c_u, we identify two user groups:

- **`Content-Consistent Users`**: the top $N\%$ users with the greatest c_u. Their past interactions show greater consistency in content, indicating a more focused user interest.
- **`Content-Diverse Users`**: the bottom $N\%$ users with the lowest c_u. Their past interactions show greater diversity in content, indicating a more varied user interest.

We use $N = 20$ in this paper, but we consistently observe a similar situation as N varies across different coverage ranges, extending from 5% to 25%.

`User-side Content Bias`. Following [6, 16], we propose a metric to quantify the biased treatment for different user groups: `User-side Content Bias` (denoted as *UCBias*) is defined as the gap between recommendation performance on `Content-Consistent Users` and `Content-Diverse Users`, which is calculated as

$$UCBias = E(U^C) - E(U^D), \tag{4}$$

where $E(\cdot)$ is denoted as the evaluation metrics. We use *Recall*@20 for $E(\cdot)$ in this paper. U^C and U^D are the `Content-Consistent Users` and `Content-Diverse Users` groups, respectively.

We can observe from Fig. 2 that

- Regardless of the type of recommendation models, whether they are shallow or deep, multi-modal or non-multi-modal, and irrespective of the context or distribution of the datasets, **there exists a positive performance gap, indicating that `Content-Consistent Users` receive higher quality recommendations than `Content-Diverse Users`.**
- When comparing methods within the same backbone architecture, the blue bars (represent MRSs) are typically higher than the green bars (represent conventional RSs), indicating that multi-modal content amplifies the performance gap.

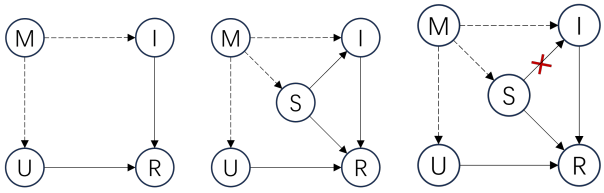
`User-side Content Bias` brings a serious unfairness issue. It makes users see more items similar to their past interactions. For `Content-Consistent Users`, this means better matches and a great experience. But for `Content-Diverse Users`, it limits diversity, worsens the filter bubble [30], and results in poor recommendations. This unfairness can drive `Content-Diverse Users` away, harming the overall performance and reputation of the recommendation systems. Furthermore, since `Content-Diverse Users` has the same size as `Content-Consistent Users` in our study, the `User-side Content Bias` affects a significant amount of users.

Algorithm 1: ISOLATOR

Input: Dataset $\mathcal{D}$; Hyper-parameter α, β, γ, ι; Inference intervention strategy: *D-ISOLATOR* or *U-ISOLATOR*;
Output: Recommendation list;
`// Training Stage.`
Calculate `User-side Content Similarity` $s_{u,i}$ by Equ. 9;
Estimate the probability of the interaction by Equ. 7;
Update parameters Θ by optimizing Equ. 5;
`// Inference Stage.`
if D-ISOLATOR **then**
  Estimate the probability of the interaction by Equ. 10;
  Generate recommendation lists by Equ. 1;
else
  Estimate the impact of `User-side Content Similarity` that users require $\widehat{s_{u,i}}$ by Equ. 11;
  Estimate the probability of the interaction by Equ. 13;
  Generate recommendation lists by Equ. 1;

3.5 Uniqueness of `User-side Content Bias`

Our next question is whether `User-side Content Bias` is a new type of bias. Since `User-side Content Bias` is evaluated from the



(a) Conventional MRSs. (b) The proposed causal graph. (c) Cutting off the confounder between S and I.

Figure 4: Causal graphs to describe the recommendation process. U: user, I: item, M: content, R: interaction probability, S: `User-side Content Similarity`.

user's perspective, we need to distinguish it from the ϵ-fairness, i.e., inactive users receive unfair treatment [13].

To show that `User-side Content Bias` is unrelated to ϵ-fairness [13], we first compute the median number of interactions per user in each dataset (denoted as a) to avoid skewing the analysis. Then, we randomly sample the same number of users who have interacted with exactly a items from both `Content-Consistent Users` and `Content-Diverse Users` groups. We report the results within these subgroups.

From Fig. 3, we can observe that `User-side Content Bias` *consistently exists regardless of the dataset used or the method applied.* Moreover, *MRSs consistently exhibit a larger* `User-side Content Bias`. The results support that `User-side Content Bias` is independent of the number of users' past interactions.

4 Methodology

To effectively leverage the multi-modal content in MRSs, we first analyze the impact of `User-side Content Similarity` on multi-modal recommender systems (MRSs) from a causal view. Then, we propose **ISOLATOR**: ut**I**lizing u**S**er-side c**O**ntent simi**L**arity via a model-**A**gnos**T**ic framew**OR**k to achieve more accurate and unbiased recommendations. ISOLATOR estimates and disentangles the impact of `User-side Content Similarity` in the training stage and employs two innovative intervention strategies in the inference stage to address its side effects. These strategies focus on debiasing and performance enhancement, respectively, to cater to various recommendation requirements. The learning program of ISOLATOR is summarized in Algorithm 1.

4.1 Causal View of MRSs

4.1.1 Causal View of Conventional MRSs. As shown in Fig. 4(a), conventional multi-modal recommender systems (MRSs) typically use four variables: user (U), exposed item (I, including exposure and features attributes), multi-modal content (M), and interaction label (R). $R = 1$ represents that the user had interacted with the item. The edges represent the causal relations between variables.

- Edges $M \rightarrow I$ and $M \rightarrow U$ signify that multi-modal content forms the feature of items and users.
- Edge $\{U, I\} \rightarrow R$ denotes that matching between user preferences U and item attributes I dictates user-item interaction R.

Based on above insights, conventional MRSs model the probability function $P(R|U, I)$[6] by the parameterized function $f_\Theta(u, i)$, which means given a user-item pair $U = u, I = i$, how likely the user u will interact with the item i. Specifically, they optimize the model parameters Θ of RSs on dataset $\mathcal{D}$ via the widely-used Bayesian Personalized Ranking [27] (BPR) loss during the training stage:

$$\mathcal{L}_{BPR} = \sum_{\langle u,i,j\rangle \in \mathcal{D}} ln\sigma\big(P_\Theta\,(r=1|u,i) - P_\Theta\,(r=1|u,j)\,\big) + \lambda_\Theta ||\Theta||_2, \quad (5)$$

where i denotes the positive sample for user u, j denotes the negative sample for user u, $\sigma(\cdot)$ is the sigmoid function, the $||\cdot||_2$ is the L2-regularization, and λ_Θ is the regularization coefficient.

4.1.2 Causal View of ISOLATOR. Unlike conventional MRSs, our method incorporates a new latent variable `User-side Content Similarity` (S) in Fig. 4(b), which includes new causal relations between variables.

- Edge $M \rightarrow S$ denotes that multi-modal content can calculate the `User-side Content Similarity`.
- Edge $\{U, I, S\} \rightarrow R$ denotes that an interaction label R is influenced by the the user U, the item I, and the `User-side Content Similarity` S. We introduce a cause node S to capture users' preferences for `User-side Content Similarity`, which is because many users interact with items solely based on `User-side Content Similarity`. Thus, the occurrence of interaction is a synergistic result of U, I, and S.
- Edge $S \rightarrow I$ denotes that `User-side Content Similarity` influences the exposure of the items. For example, we show that MRSs recommend items with greater `User-side Content Similarity` to users' past interactions in Sec. 3.3.

According to the causal theory, since `User-side Content Similarity` S affects both item I and observed interactions R, it acts as a *confounder* [25], leading to spurious associations if not properly controlled [25]. There are two causal paths: $S \rightarrow R$, and $S \rightarrow I \rightarrow R$. The first path clearly shows such similarity *directly impacts* the interaction probabilities. The second path shows such similarity **indirectly impacts** the interaction probabilities by impacting the exposure likelihood of items, making the recommendations more similar to the user's past interactions. However, **this effect can amplify the `User-side Content Bias`**. As discussed in Sec. 3.4, *not all users are interested in items similar to their past preferences.* This effect caters better to the preferences of users who seek similar items (`Content-Consistent Users`) but neglects the needs of users who seek diverse items (`Content-Diverse Users`). Consequently, it exacerbates the performance gap between `Content-Consistent Users` and `Content-Diverse Users` in MRSs, resulting in unfair recommendations. In conclusion, the path $S \rightarrow I$ has adverse effects and should be removed from MRSs.

To properly leverage multi-modal content and achieve unbiased and high-quality recommendations, it is critical to accurately estimate the causal effect of similarity (`User-side Content Similarity`) while controlling for potential confounding effect. Inspired by [33, 44], ISOLATOR leverages *do-calculus* [25] to estimate the causal effect of `User-side Content Similarity` directly from observational data. *Do-calculus* leverage the causal graph to **adjust for**

[6]For simplicity, we will not explicitly emphasize the paths $M \rightarrow I$, $M \rightarrow U$ and $M \rightarrow S$ in our subsequent formulas.

confounding variables and simulate interventions by **fixing certain variables** (e.g., setting `User-side Content Similarity`) while allowing others to vary naturally, estimating the causal effect. ISOLATOR employs the *do-calculus* to cut off the path $S \rightarrow I$ during training stage, while intervening the path $S \rightarrow R$ during inference stage.

4.2 Training Stage

ISOLATOR performs a *do-calculus* [25] to eliminate the confounding effect of `User-side Content Similarity` S on item attributes I. This can be understood as cutting off the edge $S \rightarrow I$ in the causal graph, thereby blocking the influence of S on I as shown in Fig. 4(c). We formulate the predictive model as

$$\begin{aligned} P(R|do(U,I)) &\overset{(1)}{=} P(R|U,do(I)) \\ &\overset{(2)}{=} \sum_s P(R|U,do(I),s)P(s|U,do(I)) \\ &\overset{(3)}{=} \sum_s P(R|U,I,s)P(s), \end{aligned} \tag{6}$$

where Equ. 6 (1) is because there is no path between U and R in our causal graph that requires intervention.; Equ. 6 (2) is because of Bayes' theorem; Equ. 6 (3) is because that $do(I)$ cuts off $S \rightarrow I$, so U and I are independent with S.

Based on the formula above, we first estimate $P(R|U,I,S)$. We focus on disentangling the user-item matching and `User-side Content Similarity` in the training stage, which simplifies the application across MRSs and avoids re-evaluate the entire model in the inference stage. We ensure that the model outputs are monotonically increasing without restricting them to $[0,1]$, which reduces unnecessary computational costs while avoiding impact item rankings. We design it as:

$$P_\Theta(R|U=u,I=i,S=s_{u,i}) = ELU'\big(f_\Theta(u,i)\big) \times \big(\sigma(s_{u,i})\big)^\gamma, \tag{7}$$

where $f_\Theta(u,i)$ represents any MRSs. Following [44], we use a variant of the Exponential Linear Unit [5] activation function $ELU'(\cdot)$ to ensure the matching score is positive. The bigger $f_\Theta(u,i)$ is, the bigger $ELU'\big(f_\Theta(u,i)\big)$ is:

$$ELU'(x) = \begin{cases} e^x, & \text{if } x \le 0 \\ x+1, & \text{else} \end{cases} \tag{8}$$

Hyper-parameter $\gamma > 0$ is used to adjust the `User-side Content Similarity` $s_{u,i}$, a larger value increases its impact. The `User-side Content Similarity` $s_{u,i}$ is calculated by

$$s_{u,i} = \frac{\sum_{m\in\mathcal{M}} \sum_{j\in\widehat{\mathcal{I}_u}} \big(\text{sim}(\boldsymbol{e}_i^m, \boldsymbol{e}_j^m)\big)}{|\mathcal{M}| \times |\widehat{\mathcal{I}_u}|}, \tag{9}$$

where $\boldsymbol{e}$ is multi-modal content features that remain unchanged during training. Therefore, $s_{u,i}$ can be precomputed and stored before training to reduce computational cost. $\sigma(\cdot)$ is the sigmoid activation function to ensure such content similarity is positive. The bigger $\sigma(s)$ is, the bigger $\big(\sigma(s)\big)^\gamma$ is.

Since the $s_{u,i}$ is an unique constant for each user-item pair, we can estimate $P(R|do(U=u,I=i))$ by $P_\Theta(R|U=u,I=i,S=s_{u,i})$. ISOLATOR also employs the Bayesian Personalized Ranking [27] loss in Equ. 5 to train the model.

4.3 Inference Stage

After disentangling the impact of `User-side Content Similarity` in the training stage, we eliminate such impact by doing the intervention $P\big(R|do(U,I),do(S)\big)$ in the inference stage. We set the value of `User-side Content Similarity` S while allowing others to vary naturally [25]. We propose two strategies to meet different recommendation scenarios.

4.3.1 Debiasing intervention strategy. When we strive to **reduce bias as much as possible while maintaining performance**, we can adopt a Debiasing intervention strategy (short for *D-ISOLATOR*). This strategy applies the same intervention to all users to completely eliminate the impact of `User-side Content Similarity`. The formula of the Debiasing intervention strategy can be expressed as:

$$P\big(R|do(U,I),do(S)\big) = ELU'\big(f_\Theta(u,i)\big). \tag{10}$$

Since we disentangle user-item matching and `User-side Content Similarity` in the training stage, we can efficiently remove such impact without re-training the entire model.

4.3.2 User-specific intervention strategy. The *D-ISOLATOR* eliminates the impact of `User-side Content Similarity` and applies the same intervention to all users. However, as discussed in Sec. 3.4, users have significantly different preferences for multi-modal content. Thus, when we pursue the goal of **enhancing performance without significantly increasing bias**, we can adopt an User-specific intervention strategy (short for*U-ISOLATOR*) to personalize content similarity for each user.

Since each user has a different preference for multi-modal content, we use the difference between user u's average pairwise content similarity of past interactions c_u and the `User-side Content Similarity` $s_{u,i}$ to estimate the impact of `User-side Content Similarity` that users require.

$$\widehat{s_{u,i}} = g\big(s_{u,i} - c_u\big). \tag{11}$$

Specifically, we use the average pairwise similarity of past interactions c_u to reflect the user's expected `User-side Content Similarity`. There are two different scenarios: (1) When $s_{u,i} - c_u > 0$, it indicates that the current item is similar to past interactions. The smaller the difference, the more closely the current item meets the user's expected `User-side Content Similarity`. (2) When $s_{u,i} - c_u < 0$, it indicates the current item is different from the past interactions. The larger the difference, the more the current item deviates from the user's expected `User-side Content Similarity`. These two scenarios should be treated separately because *a positive and negative difference, even with the same magnitude, have different impacts on recommendation accuracy.* Therefore, we construct a generating function $g(\cdot)$ as follows:

$$g(x) = \begin{cases} e^{-\alpha x}, & \text{if } x \ge 0 \\ \frac{4}{1+e^{-\beta x}} - 1, & \text{else} \end{cases} \tag{12}$$

where $\alpha > 0$ and $\beta > 0$ are two hyper-parameters to control the impact; the larger the α, the larger the β, the smaller the impact on the user. When $x = 0$, $g(x)$ reaches its maximum value, which is $g(0) = 1$.

Finally, the formula of the User-specific intervention strategy can be expressed as:

$$P\big(R|do(U,I),do(S)\big) = ELU'\Big(f_\Theta(u,i)\Big) \times \Big(\sigma\big(\widehat{s_{u,i}}\big)\Big)^\iota, \tag{13}$$

where $\iota > 0$ is a hyper-parameter to adjust the $\widehat{s_{u,i}}$.

Table 2: Performance of MRSs with and without ISOLATOR. The best and second results are marked with Bold and Underline. * indicates that the p-value is less than 0.05.

Datasets	Baby				Sports				Clothing				Microlens			
Models	R@10	R@20	N@10	N@20	R@10	R@20	N@10	N@20	R@10	R@20	N@10	N@20	R@10	R@20	N@10	N@20
MFBPR	0.0388	0.0611	0.0213	0.0271	0.0475	0.0714	0.0256	0.0317	0.0245	0.0353	0.0139	0.0166	0.0591	0.0911	0.0310	0.0393
LightGCN	0.0476	0.0751	0.0256	0.0327	0.0556	0.0848	0.0307	0.0382	0.0353	0.0535	0.0193	0.0240	0.0704	0.1066	0.0367	0.0461
VBPR	0.0414	0.0654	0.0219	0.0282	0.0543	0.0837	0.0295	0.0371	0.0296	0.0441	0.0164	0.0201	0.0660	0.0992	0.0344	0.0430
+*D-ISOLATOR*	0.0424*	0.0682*	0.0226*	0.0292*	0.0556*	0.0847*	0.0302*	0.0377*	0.0305*	0.0468*	0.0170*	0.0208*	0.0674*	0.1024*	0.0352*	0.0440*
+*U-ISOLATOR*	**0.0453***	**0.0715***	**0.0243***	**0.0311***	**0.0559***	**0.0858***	**0.0305***	**0.0382***	**0.0322***	**0.0485***	**0.0175***	**0.0216***	**0.0684***	**0.1038***	**0.0356***	**0.0448***
GRCN	0.0526	0.0827	0.0284	0.0362	0.0582	0.0890	0.0313	0.0393	0.0431	0.0657	0.0228	0.0286	0.0765	0.1160	0.0399	0.0501
+*D-ISOLATOR*	0.0531*	0.0839*	0.0287*	0.0367*	**0.0605***	**0.0925***	**0.0328***	**0.0411***	0.0440*	0.0665*	0.0233*	0.0289*	0.0772*	0.1168*	**0.0403***	**0.0505***
+*U-ISOLATOR*	**0.0536***	**0.0851***	**0.0289***	**0.0369***	0.0602*	0.0924*	0.0325*	0.0408*	**0.0442***	**0.0673***	**0.0235***	**0.0294***	**0.0773***	**0.1174***	**0.0403***	**0.0505***
SLMRec	0.0525	0.0799	0.0285	0.0356	0.0671	0.1010	0.0368	0.0456	0.0461	0.0691	0.0251	0.0309	0.0784	0.1189	0.0406	0.0510
+*D-ISOLATOR*	0.0545*	0.0851*	0.0295*	0.0373*	0.0674*	0.1013*	0.0371*	0.0459*	0.0466*	0.0696*	0.0253*	0.0312*	0.0786*	0.1193*	0.0407*	0.0511*
+*U-ISOLATOR*	**0.0548***	**0.0852***	**0.0296***	**0.0374***	**0.0680***	**0.1020***	**0.0373***	**0.0460***	**0.0473***	**0.0706***	**0.0257***	**0.0316***	**0.0793***	**0.1199***	**0.0411***	**0.0515***
MGCN	0.0619	0.0958	0.0334	0.0421	0.0744	0.1122	0.0409	0.0506	0.0656	0.0955	0.0361	0.0437	0.0752	0.1138	0.0387	0.0486
+*D-ISOLATOR*	0.0631*	0.0969*	0.0339*	0.0426*	0.0748*	0.1128*	**0.0412***	0.0510*	0.0653	0.0959*	0.0359	0.0438*	0.0756*	0.1143*	0.0390*	0.0489*
+*U-ISOLATOR*	**0.0633***	**0.0980***	**0.0342***	**0.0431***	**0.0752***	**0.1134***	**0.0412***	**0.0511***	**0.0662***	**0.0967***	**0.0364***	**0.0441***	**0.0759***	**0.1150***	**0.0391***	**0.0492***
GUME	0.0682	0.1041	0.0368	0.0460	0.0776	0.1168	0.0425	0.0526	0.0688	0.1008	0.0373	0.0454	0.0813	0.1204	0.0426	0.0526
+*D-ISOLATOR*	0.0690*	0.1062*	0.0368	0.0464*	0.0778*	0.1171*	**0.0427***	**0.0528***	0.0690*	0.1010*	0.0374*	0.0455*	0.0817*	0.1207*	**0.0428***	**0.0528***
+*U-ISOLATOR*	**0.0695***	**0.1063***	**0.0371***	**0.0465***	**0.0779***	**0.1173***	**0.0427***	**0.0528***	**0.0694***	**0.1019***	**0.0376***	**0.0458***	**0.0819***	**0.1211***	**0.0428***	**0.0528***

5 Experiments

In this section, we mainly answer the following questions: (1) **RQ1:** How does ISOLATOR perform in terms of recommendation accuracy? (Sec. 5.1) (2) **RQ2:** How does ISOLATOR perform in mitigating the `User-side Content Bias`? (Sec. 5.2) (3) **RQ3:** How does ISOLATOR perform in single-modal scenarios? (Sec. 5.3) (4) **RQ4:** How does each strategy affect the performance of ISOLATOR? (Sec. 5.4)

We use the datasets described in Sec. 3.2. We implement our method in PyTorch [24]. The embedding dimension d is fixed to 64 for all models to ensure a fair comparison. We optimize all models with the Adam [10] optimizer, where the batch size is fixed at 2,048. We use the Xavier initializer [7] to initialize the model parameters. We train the model using the same learning rate as the backbone model. The optimal hyper-parameters are determined via grid search on the validation set: the γ in Equation 7 and the ι in Equation 13 are tuned amongst {0.001, 0.01, 0.1, 1}, the α in Equation 12 is tuned amongst {2, 4, 6, 8, 10}, the β in Equation 12 is tuned amongst {1, 5, 10, 15, 20}. For convergence consideration, the early stopping and total epochs are fixed at 25 and 1,000, respectively. **We make our code available online to ease reproducibility**[7].

5.1 Impacts on Recommendation Accuracy

To explore the impact of ISOLATOR on recommendation accuracy, we apply ISOLATOR to various multi-modal recommender systems (MRSs) mentioned in Sec. 3.2.3. We use *Recall@K* and *NDCG@K* (abbreviated as R and N) as evaluation metrics following prior works [8, 17, 26, 31, 36]. Higher values of *Recall* and *NDCG* indicate more accurate recommendation results. Here we set $K = 10$ and $K = 20$. We have following observations from Table 2:

- ISOLATOR can be applied to various MRSs and consistently improve the recommendation accuracy. We observe consistent improvements across all backbone models in terms of *Recall* and *NDCG* on all datasets. This highlights its ability to **effectively capture user preferences for multi-modal content by intervening in the impact of `User-side Content Similarity`**.
- *U-ISOLATOR* improves recommendation performance by leveraging `User-side Content Similarity` based on individual user needs, outperforming *D-ISOLATOR* in most cases with superior *Recall* and *NDCG*. This demonstrates that **leverage the impact of `User-side Content Similarity` properly can boost recommendation accuracy**.
- The performance of MRSs typically surpasses that of conventional RSs using the same backbone architecture. This indicates that **incorporating multi-modal content to comprehend the similarity between items can enhance accuracy**.

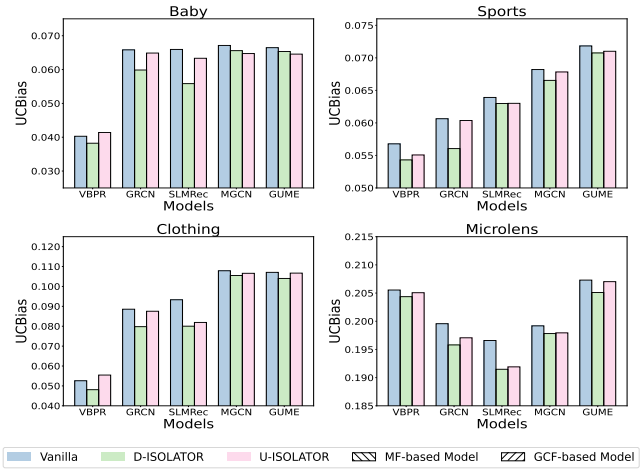


Figure 5: `User-side Content Bias` in MRSs with and without ISOLATOR. Lower *UCBias* for better de-biasing effectiveness.

5.2 Impacts on `User-side Content Bias`

We investigate the effects of ISOLATOR on mitigating `User-side Content Bias` in this subsection. We use the top and bottom 20% users as `Content-Consistent Users` and `Content-Diverse`

[7] https://github.com/11Lixinlv20/ISOLATOR.git

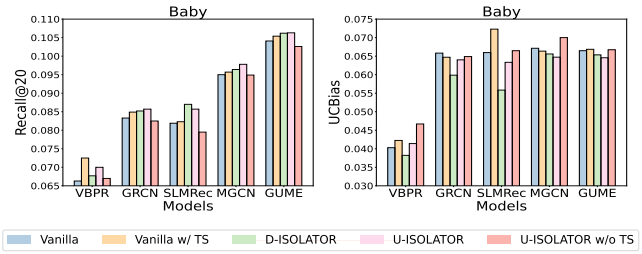


Figure 6: Performance and `User-side Content Bias` of Inference Strategies in ISOLATOR on the Baby Dataset: Higher *Recall*@20 and Lower *UCbias* indicate Better Results.

`Users` following Sec. 3.4. The details are described in Sec. 3.4. We use the *Recall*@20 to calculate the `User-side Content Bias` *UCBias*. A lower *UCBias* suggests a smaller bias.

As shown in Fig. 5, we have following observations:

- *D-ISOLATOR* can effectively mitigate `User-side Content Bias`, which consistently achieves the lowest *UCBias* across all models and datasets, reducing the bias by up to 15.33% compared with the vanilla models. *D-ISOLATOR* completely removes the impact of `User-side Content Similarity` for all users, highlighting that **eliminating this impact can effectively reduce the `User-side Content Bias`**.
- *U-ISOLATOR* effectively enhances recommendation performance while controling the `User-side Content Bias` in most cases. *U-ISOLATOR* improves accuracy and reduces the `User-side Content Bias` for all GCF-based MRSs. Despite a slight increase in `User-side Content Bias` for VBPR on the Baby and Clothing datasets, it achieves an average accuracy improvement of 7.10% on them. These highlights that **strategically using the impact of `User-side Content Similarity` is crucial**.

5.3 Impacts of Modality

To explore the effects of different modalities on ISOLATOR, we compare the results of GUME on the Baby dataset across different modalities because GUME performed the best among the backbones we used. We employ the same evaluation metrics in Sec. 5.1 and Sec. 5.2 to validate the accuracy and debiasing capability, respectively. From Table 3, we can observe that:

- Different modalities do not diminish the effectiveness of ISOLATOR. ISOLATOR consistently boosts recommendation performance (higher *Recall* and *NDCG*) while reducing `User-side Content Bias` (lower *UCBias*) in single-modal settings, mirroring the trends seen in multi-modal settings. This highlights **its robust generalization and broad adaptability across various modality environments**.
- Multi-modal content outperforms single-modal content (higher *Recall* and *NDCG*) by leveraging diverse modalities for more comprehensive information. This is also why we employ multi-modal content in this work.
- The textual modality provides more accurate recommendations than visual modality due to higher explicitness and information density but amplifies `User-side Content Bias`.
- `User-side Content Bias` exists across all modalities, even in single-modality settings, highlighting its **pervasive nature irrespective of the modality used**.

Table 3: Performance and `User-side Content Bias` of GUME with and without ISOLATOR in each modality on Baby dataset. The best and second results are marked with Bold and Underline.

Modality	Model	*R*@10↑	*R*@20↑	*N*@10↑	*N*@20↑	*UCBias*↓
Visual Modality	GUME	0.0536	0.0844	0.0291	0.0370	0.0564
	+*D-ISOLATOR*	0.0549	0.0877	0.0301	0.0385	0.0553
	+*U-ISOLATOR*	**0.0552**	**0.0885**	**0.0302**	**0.0388**	**0.0529**
Textual Modality	GUME	0.0675	0.1024	0.0362	0.0451	0.0717
	+*D-ISOLATOR*	**0.0684**	0.1034	0.0366	0.0456	**0.0709**
	+*U-ISOLATOR*	**0.0684**	**0.1051**	**0.0372**	**0.0466**	0.0781
Multi-modalities	GUME	0.0682	0.1041	0.0368	0.0460	0.0664
	+*D-ISOLATOR*	0.0690	0.1062	0.0368	0.0464	0.0653
	+*U-ISOLATOR*	**0.0695**	**0.1063**	**0.0371**	**0.0465**	**0.0646**

5.4 Impacts of Strategies

ISOLATOR employs three strategies: one training strategy *TS* and two inference strategies *D-ISOLATOR* and *U-ISOLATOR*. To assess their effectiveness, we create these variants[8]: (1) *Vanilla*: no strategy in training or inference; (2) *Vanilla w/ TS*: *TS* in training, no strategy in inference; (3) *D-ISOLATOR*: *TS* in training, Debiasing intervention strategy in inference; (4) *U-ISOLATOR*: *TS* in training, User-specific intervention strategy in inference; (5) *U-ISOLATOR w/o TS*: no *TS* in training, User-specific intervention strategy in inference. We test these variants on the Baby dataset using the MRSs from Sec. 3.2.3 as backbones. We use the same evaluation metrics in Sec. 3.4. Higher *Recall*@20 and lower *UCBias* indicate better results. From Fig. 6, we can observe that:

- **Combining training and inference strategies are the best.** *D-ISOLATOR* and *U-ISOLATOR* achieve higher *Recall*@20 and lower *UCBias* than other variants across most backbones. Generally, ***D-ISOLATOR* keeps `User-side Content Bias` low, while *U-ISOLATOR* boosts accuracy**, enabling flexible strategy selection for accurate and unbiased recommendations.
- Training strategy boosts accuracy but also raises `User-side Content Bias`. All three variants using this strategy (i.e., *D-ISOLATOR*, *U-ISOLATOR*, *Vanilla w/ TS*) surpass the *vanilla* in *Recall*@20, yet *Vanilla w/ TS* gets a larger *UCBias*. This shows that training strategy can **model both the positive and negative impacts of `User-side Content Similarity`** but requires inference strategies to balance the trade-offs.
- Inference strategy can not be solely used. U-ISOLATOR *w/o TS* performs worse than others. Since the impact of `User-side Content Similarity` is not distinguished without training strategy, the inference strategy might overemphasize it, leading to poorer performance and increased bias.

Due to space constraints, we conduct **parameter experiments in Section B.2 of the supplementary materials**. γ and ι control the impact of `User-side Content Similarity` in training and inference stages, we find that **over-reliance on this impact leads to suboptimal recommendations** (decreasing *Recall*@20) and **introduce unwanted bias** (increasing *UCBias*). α and β regulate the impact of `User-side Content Similarity` that users require in *U-ISOLATOR*. *U-ISOLATOR* achieves **stable accuracy**

[8] *D-ISOLATOR w/o TS* equals *Vanilla*, as Debiasing intervention strategy doesn't consider the impact of `User-side Content Similarity`.

(*Recall*@20) while enabling adjustments to balance bias (*UCBias*), with the optimal values identified as $\alpha = 6$ and $\beta = 10$.

Furthermore, we present a **case study in Section B.3 of the supplementary materials** to show how ISOLATOR enhances recommendations for both `Content-Consistent Users` and `Content-Diverse Users`. For `Content-Consistent Users`, it identifies preferred items among high-similarity candidates, while for `Content-Diverse Users`, it recommends less similar yet more relevant items. These results highlight **ISOLATOR's effectiveness in improving accuracy, diversity, and meeting user needs**.

6 Conclusion

In this paper, we uncover an **unexplored `User-side Content Bias`** in multi-modal recommender systems, and introduce an effective framework, ISOLATOR, to utilize multi-modal content more properly. Our empirical studies demonstrate the existence of `User-side Content Bias`, and our experiments validate the effectiveness of ISOLATOR. We provide insights into several key perspectives: (1) Utilizing `User-side Content Similarity` increases the exposure of items similar to the user's past interactions, thereby improving the accuracy but amplifying `User-side Content Bias`. (2) Completely eliminating the impact of `User-side Content Similarity` can reduce `User-side Content Bias`, and strategically using the impact of `User-side Content Similarity` can enhance accuracy. We plan to investigate `User-side Content Bias` and debiasing methods in other content-based recommender systems in the future, such as using multi-modal large language models for recommender systems.

References

[1] Desheng Cai, Shengsheng Qian, Quan Fang, Jun Hu, and Changsheng Xu. 2022. Adaptive Anti-Bottleneck Multi-Modal Graph Learning Network for Personalized Micro-video Recommendation. In *MM '22: The 30th ACM International Conference on Multimedia, Lisboa, Portugal, October 10 - 14, 2022*. ACM, 581–590. https://doi.org/10.1145/3503161.3548420

[2] Gaode Chen, Ruina Sun, Yuezihan Jiang, Jiangxia Cao, Qi Zhang, Jingjian Lin, Han Li, Kun Gai, and Xinghua Zhang. 2024. A Multi-modal Modeling Framework for Cold-start Short-video Recommendation. In *Proceedings of the 18th ACM Conference on Recommender Systems, RecSys 2024, Bari, Italy, October 14-18, 2024*. ACM, 391–400. https://doi.org/10.1145/3640457.3688098

[3] Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, and Tat-Seng Chua. 2017. Attentive Collaborative Filtering: Multimedia Recommendation with Item- and Component-Level Attention. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*. ACM, 335–344. https://doi.org/10.1145/3077136.3080797

[4] Yewang Chen, Weiyao Ye, Guipeng Xv, Chen Lin, and Xiaomin Zhu. 2023. TCCM: Time and Content-Aware Causal Model for Unbiased News Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*. ACM, 3778–3782. https://doi.org/10.1145/3583780.3615272

[5] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. 2016. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). In *4th International Conference on Learning Representations*.

[6] Yashar Deldjoo, Vito Walter Anelli, Hamed Zamani, Alejandro Bellogín, and Tommaso Di Noia. 2021. A flexible framework for evaluating user and item fairness in recommender systems. *User Model. User Adapt. Interact.* 31, 3 (2021), 457–511. https://doi.org/10.1007/S11257-020-09285-1

[7] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2010, Chia Laguna Resort, Sardinia, Italy, May 13-15, 2010 (JMLR Proceedings, Vol. 9)*. 249–256.

[8] Ruining He and Julian J. McAuley. 2016. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*. AAAI Press, 144–150. https://doi.org/10.1609/AAAI.V30I1.9973

[9] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*. 639–648.

[10] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

[11] Dominik Kowald and Emanuel Lacic. 2022. Popularity Bias in Collaborative Filtering-Based Multimedia Recommender Systems. In *Advances in Bias and Fairness in Information Retrieval - Third International Workshop, BIAS 2022, Stavanger, Norway, April 10, 2022, Revised Selected Papers (Communications in Computer and Information Science, Vol. 1610)*. Springer, 1–11. https://doi.org/10.1007/978-3-031-09316-6_1

[12] Kushal Kumar, Tarik Arici, Tal Neiman, Jinyu Yang, Shioulin Sam, Yi Xu, Hakan Ferhatosmanoglu, and Ismail B. Tutar. 2023. Unsupervised Multi-Modal Representation Learning for High Quality Retrieval of Similar Products at E-commerce Scale. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*. ACM, 4667–4673. https://doi.org/10.1145/3583780.3615504

[13] Yunqi Li, Hanxiong Chen, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2021. User-oriented Fairness in Recommendation. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 624–632. https://doi.org/10.1145/3442381.3449866

[14] Yang Li, Qi'ao Zhao, Chen Lin, Jinsong Su, and Zhilin Zhang. 2024. Who To Align With: Feedback-Oriented Multi-Modal Alignment in Recommendation Systems. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*. ACM, 667–676. https://doi.org/10.1145/3626772.3657701

[15] Zihan Liao, Xiaodong Wu, Shuo Shang, Jun Wang, and Wei Zhang. 2024. Modeling Dynamic Item Tendency Bias in Sequential Recommendation With Causal Intervention. *IEEE Trans. Knowl. Data Eng.* 36, 12 (2024), 8814–8828. https://doi.org/10.1109/TKDE.2024.3427719

[16] Chen Lin, Xinyi Liu, Guipeng Xv, and Hui Li. 2021. Mitigating Sentiment Bias for Recommender Systems. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*. ACM, 31–40. https://doi.org/10.1145/3404835.3462943

[17] Guojiao Lin, Zhen Meng, Dongjie Wang, Qingqing Long, Yuanchun Zhou, and Meng Xiao. 2024. GUME: Graphs and User Modalities Enhancement for Long-Tail Multimodal Recommendation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, CIKM 2024, Boise, ID, USA, October 21-25, 2024*. ACM, 1400–1409. https://doi.org/10.1145/3627673.3679620

[18] Chang Liu, Peng Hou, Anxiang Zeng, and Han Yu. 2024. Transformer-empowered multi-modal item embedding for enhanced image search in e-commerce. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'24/IAAI'24/EAAI'24)*. AAAI Press, Article 2588, 9 pages. https://doi.org/10.1609/aaai.v38i21.30311

[19] Xinyi Liu, Wanxian Guan, Lianyun Li, Hui Li, Chen Lin, Xubin Li, Si Chen, Jian Xu, Hongbo Deng, and Bo Zheng. 2022. Pretraining Representations of Multi-modal Multi-query E-commerce Search. In *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*. ACM, 3429–3437. https://doi.org/10.1145/3534678.3539200

[20] Xiaohao Liu, Zhulin Tao, Jiahong Shao, Lifang Yang, and Xianglin Huang. 2022. EliMRec: Eliminating Single-modal Bias in Multimedia Recommendation. In *MM '22: The 30th ACM International Conference on Multimedia, Lisboa, Portugal, October 10 - 14, 2022*. ACM, 687–695. https://doi.org/10.1145/3503161.3548404

[21] Daniele Malitesta, Giandomenico Cornacchia, Claudio Pomo, and Tommaso Di Noia. 2023. On Popularity Bias of Multimodal-aware Recommender Systems: A Modalities-driven Analysis. In *Proceedings of the 1st International Workshop on Deep Multimodal Learning for Information Retrieval, MMIR 2023, Ottawa ON, Canada, 2 November 2023*. ACM, 59–68. https://doi.org/10.1145/3606040.3617441

[22] Jianmo Ni, Jiacheng Li, and Julian J. McAuley. 2019. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*. Association for Computational Linguistics, 188–197. https://doi.org/10.18653/V1/D19-1018

[23] Yongxin Ni, Yu Cheng, Xiangyan Liu, Junchen Fu, Youhua Li, Xiangnan He, Yongfeng Zhang, and Fajie Yuan. 2023. A Content-Driven Micro-Video Recommendation Dataset at Scale. *arXiv preprint arXiv:2309.15379* (2023).

[24] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Z. Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. 8024–8035.

[25] Judea Pearl. 2009. *Causality*. Cambridge University Press.

[26] YU Peng, Zhiyi Tan, Guanming Lu, and Bingkun Bao. 2023. Multi-View Graph Convolutional Network for Multimedia Recommendation. *Proceedings of the 31st ACM International Conference on Multimedia* (2023). https://api.semanticscholar.org/CorpusID:260681274

[27] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009*. 452–461.

[28] Yu Shang, Chen Gao, Jiansheng Chen, Depeng Jin, and Yong Li. 2024. Improving Item-side Fairness of Multimodal Recommendation via Modality Debiasing. In *Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024*. ACM, 4697–4705. https://doi.org/10.1145/3589334.3648156

[29] Yu Shang, Chen Gao, Jiansheng Chen, Depeng Jin, Meng Wang, and Yong Li. 2023. Learning Fine-grained User Interests for Micro-video Recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*. 433–442.

[30] Anne Shelley. 2012. Book review of Eli Pariser's *The filter bubble: What the Internet is hiding from you. First Monday* 17, 6 (2012). https://doi.org/10.5210/FM.V17I6.4100

[31] Zhulin Tao, Xiaohao Liu, Yewei Xia, Xiang Wang, Lifang Yang, Xianglin Huang, and Tat-Seng Chua. 2023. Self-Supervised Learning for Multimedia Recommendation. *IEEE Trans. Multim.* 25 (2023), 5107–5116. https://doi.org/10.1109/TMM.2022.3187556

[32] Lei Wang, Chen Ma, Xian Wu, Zhaopeng Qiu, Yefeng Zheng, and Xu Chen. 2024. Causally Debiased Time-aware Recommendation. In *Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024*. ACM, 3331–3342. https://doi.org/10.1145/3589334.3645400

[33] Wenjie Wang, Fuli Feng, Xiangnan He, Xiang Wang, and Tat-Seng Chua. 2021. Deconfounded Recommendation for Alleviating Bias Amplification. In *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021*. ACM, 1717–1725. https://doi.org/10.1145/3447548.3467249

[34] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019*. ACM, 165–174. https://doi.org/10.1145/3331184.3331267

[35] Zongyi Wang, Yanyan Zou, Anyu Dai, Linfang Hou, Nan Qiao, Luobao Zou, Mian Ma, Zhuoye Ding, and Sulong Xu. 2023. An Industrial Framework for Personalized Serendipitous Recommendation in E-commerce. In *Proceedings of the 17th ACM Conference on Recommender Systems, RecSys 2023, Singapore, Singapore, September 18-22, 2023*. 1015–1018.

[36] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, and Tat-Seng Chua. 2020. Graph-Refined Convolutional Network for Multimedia Recommendation with Implicit Feedback. In *MM '20: The 28th ACM International Conference on Multimedia, Virtual Event / Seattle, WA, USA, October 12-16, 2020*. ACM, 3541–3549. https://doi.org/10.1145/3394171.3413556

[37] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and Tat-Seng Chua. 2019. MMGCN: Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video. In *Proceedings of the 27th ACM International Conference on Multimedia, MM 2019, Nice, France, October 21-25, 2019*. ACM, 1437–1445. https://doi.org/10.1145/3343031.3351034

[38] Guipeng Xv, Xinyu Li, Ruobing Xie, Chen Lin, Chong Liu, Feng Xia, Zhanhui Kang, and Leyu Lin. 2024. Improving Multi-modal Recommender Systems by Denoising and Aligning Multi-modal Content and User Feedback. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*. ACM, 3645–3656. https://doi.org/10.1145/3637528.3671703

[39] Guipeng Xv, Chen Lin, Hui Li, Jinsong Su, Weiyao Ye, and Yewang Chen. 2022. Neutralizing Popularity Bias in Recommendation Models. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*. ACM, 2623–2628. https://doi.org/10.1145/3477495.3531907

[40] Wei Yang, Zhengru Fang, Tianle Zhang, Shiguang Wu, and Chi Lu. 2023. Modal-aware Bias Constrained Contrastive Learning for Multimodal Recommendation. In *Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023*. ACM, 6369–6378. https://doi.org/10.1145/3581783.3612568

[41] Zixuan Yi, Xi Wang, Iadh Ounis, and Craig MacDonald. 2022. Multi-modal Graph Contrastive Learning for Micro-video Recommendation. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*. ACM, 1807–1811. https://doi.org/10.1145/3477495.3532027

[42] Jinghao Zhang, Yanqiao Zhu, Qiang Liu, Shu Wu, Shuhui Wang, and Liang Wang. 2021. Mining Latent Structures for Multimedia Recommendation. In *MM '21: ACM Multimedia Conference, Virtual Event, China, October 20 - 24, 2021*. ACM, 3872–3880. https://doi.org/10.1145/3474085.3475259

[43] Jinghao Zhang, Yanqiao Zhu, Qiang Liu, Mengqi Zhang, Shu Wu, and Liang Wang. 2023. Latent Structure Mining With Contrastive Modality Fusion for Multimedia Recommendation. *IEEE Trans. Knowl. Data Eng.* 35, 9 (2023), 9154–9167. https://doi.org/10.1109/TKDE.2022.3221949

[44] Yang Zhang, Fuli Feng, Xiangnan He, Tianxin Wei, Chonggang Song, Guohui Ling, and Yongdong Zhang. 2021. Causal Intervention for Leveraging Popularity Bias in Recommendation. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*. ACM, 11–20. https://doi.org/10.1145/3404835.3462875

[45] Xin Zhou. 2023. MMRec: Simplifying Multimodal Recommendation. In *ACM Multimedia Asia Workshops, MMAsia 2023, Tainan, Taiwan, December 6-8, 2023*, Wen-Huang Cheng, Wei-Ta Chu, Min-Chun Hu, Jiaying Liu, Munchurl Kim, and Wei Zhang (Eds.). ACM, 6:1–6:2. https://doi.org/10.1145/3611380.3628561

[46] Xin Zhou and Zhiqi Shen. 2023. A Tale of Two Graphs: Freezing and Denoising Graph Structures for Multimodal Recommendation. In *Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023*. ACM, 935–943. https://doi.org/10.1145/3581783.3611943

[47] Xin Zhou, Hongyu Zhou, Yong Liu, Zhiwei Zeng, Chunyan Miao, Pengwei Wang, Yuan You, and Feijun Jiang. 2023. Bootstrap Latent Representations for Multi-modal Recommendation. In *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023*. ACM, 845–854. https://doi.org/10.1145/3543507.3583251

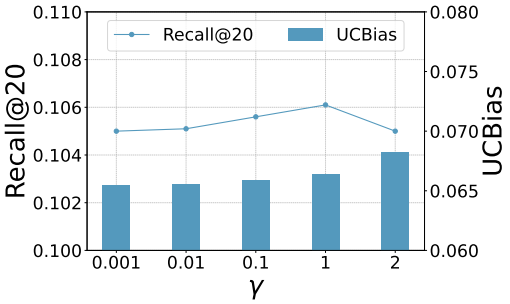


Figure 7: Effect of γ under GUME+*D-ISOLATOR* on Baby dataset.

A Related work

A.1 Multi-modal Recommender Systems

Multi-modal recommender systems leverage both multi-modal features and behavioral features to provide a comprehensive prediction of user preferences. They can be divided into two categories based on the adopted backbone architecture: **Matrix Factorization (MF) architecture** [3, 8] and **Graph Collaborative Filtering (GCF) architecture** [17, 26, 31, 34, 36–38]. Previous methods primarily used MF as the backbone, which decomposes a user-item interaction matrix into two lower-dimensional matrices. For example, VBPR [8] incorporates visual features of items as part of their characteristics and employs an MF structure for model training. However, due to the difficulty of MF in fully leveraging modal information and its suboptimal recommendation performance, recent works focus on Graph Collaborative Filtering (GCF) architecture, which exploit the graph structure to capture intricate relationships. GCF-based methods can be primarily categorized into two approaches. (1) The first approach involves propagating and updating modal features within a user-item bipartite graph. For example, MMGCN [37] and GRCN [36] build modality-specific user-item bipartite graphs to learn user and item features and concatenates them for prediction. SLMRec [31] augments item multi-modal features into two views, extracts embeddings from each via graph convolution on a user–item bipartite graph, and aligns them with contrastive learning to capture latent patterns. (2) The second approach focuses on leveraging the similarity of modal features to construct a homogeneous item-item graph, thereby enhancing the learning process of the recommender systems. For example, MGCN [26] constructs an item similarity graph and uses a behavior-aware fuser to weigh the significance of features from different modalities; GUME [17] enhances the user-item graph by utilizing cross-modal item similarities and extracts meaningful representations for enhanced recommendation performance.

A.2 Bias in Multi-modal Recommender Systems

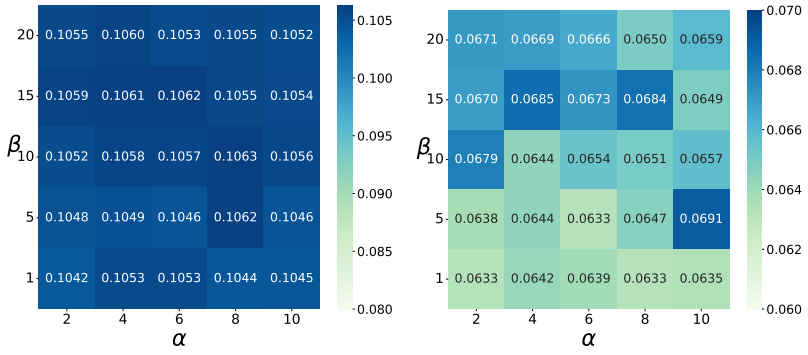
Biases and de-biasing methods have been extensively studied in conventional RSs, primarily focusing on biases resulting from feedback data [4, 11, 16, 21, 39]. However, **the biases introduced by multi-modal contents are of greater concern in MRSs**, primarily focusing on the **inter-modality bias** and **item-side bias**. Liu *et al.* [20] expose the *Single-modal bias*, wherein single-modal characteristics inherently skew the generated multi-modal features, representing an *inter-modal bias*. Shang *et al.* [28] reveal *Modality bias*, wherein non-uniform modal content distribution in training data leads to over-recommendation of certain items , constituting an *item-side bias*. Yang *et al.* [40] expose *Modal-aware Bias*, wherein self-supervised learning in current multi-modal recommender systems can generate biased augmentations, causing information loss and noise. Malitesta *etal.* [21] discover that the use of multi-modal content can indeed exacerbate *item-side popularity bias*.

A.3 Causal Inference in Recommender Systems

Causal inference methods [25] are extensively utilized to reveal the causal relationships underlying recommendations and reduce the biases present in recommender systems [15, 20, 32, 33, 44]. Existing causal debiasing techniques for recommender systems can be categorized into two primary groups: (1) *Confounding Elimination*, which tackles bias by regarding it as a confounding factor and applying backdoor adjustment to diminish its influence [32, 33, 44]. For example, PDA [44] identifies item popularity as a confounder between item exposure and observed interactions and addresses the impact of popularity bias based on the backdoor adjustment; DecRS [33] introduces an approximation operator for backdoor adjustment that can be seamlessly integrated into most recommender models, and devises an inference strategy to regulate backdoor adjustment based on user status dynamically; (2) *Counterfactual Inference*, which typically employs counterfactual analysis to create a hypothetical scenario, and derives the Total Indirect Effect (TIE) by examining the Total Effect (TE) in the real world and the Natural Direct Effect (NDE) in the counterfactual scenario [20, 28]. For example, EliMRec [20] constructs a counterfactual scenario permitting variation in a single modality while others are held constant, subsequently mitigating the impact of single-modal bias by subtracting counterfactual outcomes from those of the real-world model; Shang *etal.* [28] presents a fairness-aware modality debiasing framework based on counterfactual inference to eliminate modality bias and enhance item-side fairness. During training, the framework incorporates unimodal prediction branches to capture modality bias. In the inference phase, a fairness-aware counterfactual analysis is performed to eliminate modality bias adaptively.

A.4 Remarks

ISOLATOR systematically explore the impact of multi-modal content and discover a new type of bias, i.e., **`User-side Content Bias`**. Different from current debiasing efforts in MRSs, which primarily focus on the inter-modal bias or item-side bias, ISOLATOR arising from **the bias in user-side** due to varying user preferences for multi-modal content. ISOLATOR uses confounding elimination and proposes two strategies to better utilize the impact of `User-side Content Similarity` to obtain more fair or more accurate recommendations. To our knowledge, ISOLATOR **is the first work** that systematically explores the impact of multi-modal content in MRSs **on the user side** and **mitigates the `User-side Content Bias` in MRSs from a causal perspective**.



(a) *Recall*@20 w.r.t α and β (b) *UCBias* w.r.t α and β

Figure 8: Effect of α and β under GUME+*U-ISOLATOR* on Baby Dataset.

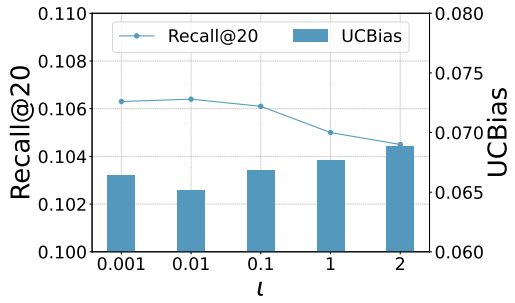


Figure 9: Effect of ι under GUME+*U-ISOLATOR* on Baby dataset.

B Experiment

B.1 Baseline

Multi-modal recommender systems (MRSs) typically integrate multi-modal content into conventional RSs based on two primary backbone architectures: matrix factorization (MF) and graph collaborative filtering (GCF). We use two conventional RSs as backbone:

- MFBPR [27], which decomposes user-item matrices into lower-dimensional latent factors to capture preferences and generate recommendations.
- LightGCN [9], which efficiently captures user preferences by aggregating neighbor embeddings in a simplified structure, enabling scalable and robust recommendations.

We have five multi-modal recommender systems as backbone:

- VBPR [8], which integrates visual embeddings with ID embeddings directly to enhance recommendation performance;
- GRCN [36], which employs graph convolutions to merge multi-modal content with ID embeddings, capturing complex relationships;
- MGCN [26], which uses a behavior-aware fuser to weigh the significance of features from different modalities adaptively;
- SLMRec [31], which aligns content across various modalities for the same item, ensuring consistency and coherence;
- GUME [17], which leverages multi-modal item similarities to refine the user-item graph and learns high-quality representations for improved recommendations.

It should be noted that *MFBPR and VBPR are Matrix Factorization architectures, while the remaining are graph architectures.* To ensure optimal performance, we use publicly available code[9] [45] with default parameter settings.

B.2 Impacts of Hyper-parameters

To investigate the impact of hyper-parameters on ISOLATOR, we conduct a series of experiments using the GUME as the backbone on the Baby dataset because GUME performed the best among the backbones we used. We focus on four key hyper-parameters: γ in Equ. 7, α and β in Equ. 12, and ι in Equ. 13. For clearer visualizations, we adjust one hyper-parameter at a time while keeping the others constant.

[9]https://github.com/enoche/MMRec

B.2.1 Impact of γ for Debiasing intervention strategy. The parameter γ controls the impact of `User-side Content Similarity` during the training phase. By varying γ across {0.001, 0.01, 0.1, 1, 2}, we can observe from Fig. 7 that:

- When $\gamma < 1$, increasing γ leads to an improvement in *Recall*@20 while the bias remains relatively stable. This highlights that **appropriately increasing the influence of `User-side Content Similarity` during training to achieve enhanced recommendation performance while maintaining a balanced `User-side Content Bias` level**.
- Specifically, moderate settings of γ (e.g., $\gamma = 1$) yield the optimal balance, achieving relatively high *Recall*@20 while maintaining a controlled bias.
- When γ is set too high (e.g., $\gamma = 2$), both *Recall*@20 and *UCBias* are negatively impacted. Specifically, this results in a decrease in *Recall*@20 and an increase in *UCBias*, which highlights that **excessive reliance on `User-side Content Similarity` during training can lead to less relevant recommendations and introduce more bias**.

B.2.2 Impact of α and β for User-specific intervention strategy. The parameters α and β control the impact of `User-side Content Similarity` that users require. We test α within the range of $\{2, 4, 6, 8, 10\}$ and β within the range of $\{1, 5, 10, 15, 20\}$. As shown in Fig. 8, we observe that:

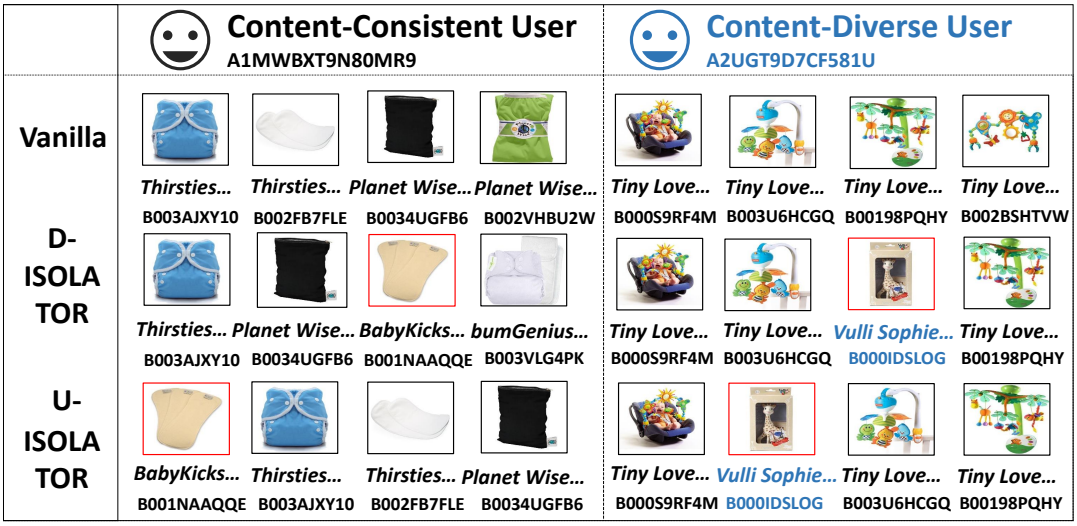


Figure 10: A case study of recommendation list on Baby dataset for `Content-Consistent Users` and `Content-Diverse Users`, where red boxes for correct recommendations, black boxes for incorrect recommendations, black text for high content similarity, and gray text for low content similarity.

- Adjusting α and β have a minimal impact on *Recall*@20, demonstrating **the stability of ISOLATOR**.
- Adjusting α and β affects *UCBias*. When α remains constant, *UCBias* initially grows slowly as β grows but subsequently decreases after reaching a certain point. Additionally, when β is fixed at 1, *UCBias* remains relatively stable at a low level despite variations in α.
- **α and β affect the trade-off between recommendation performance (*Recall*@20) and `User-side Content Bias` (*UCBias*).** The combination of $\alpha = 6$ and $\beta = 10$ achieves the highest *Recall*@20 while maintaining a relatively low `User-side Content Bias`.

B.2.3 Impact of ι for User-specific intervention strategy. The parameter ι determines the impact of `User-side Content Similarity` during the inference phase. By varying ι across {0.001, 0.01, 0.1, 1, 2}, we can analyze the resulting trends in Fig. 9

- For $\iota \leq 0.01$, increasing ι results in improved *Recall*@20 performance and reduced *UCBias*. This highlights that **a moderate adjustment of the impact of `User-side Content Similarity` during inference can enhance recommendation quality while minimizing bias**. The optimal balance is achieved at $\iota = 1$, which maintains high *Recall*@20 with minimal `User-side Content Bias`.
- When $\iota > 0.01$, both metrics deteriorate: *Recall*@20 decreases and *UCBias* increases. This suggests that **excessive emphasis on the impact of `User-side Content Similarity` during inference can lead to suboptimal recommendation outcomes and introduce unwanted biases**.

B.3 Case Study

We introduce a case study to illustrate ISOLATOR's impact on improving recommendations for both `Content-Consistent Users` and `Content-Diverse Users`. We randomly select one user from these user groups in the Baby dataset. We generate the recommendation lists using GUME with and without ISOLATOR, then visualize them.

From Fig. 10, we can observe that:

- For `Content-Consistent Users`, ISOLATOR can identify the user-preferred item among numerous high-similarity candidates. For example, the top four diapers recommended by *vanilla* are wrong, while ISOLATOR can retrieve the diaper the user likes. In comparison, *U-ISOLATOR* tends to rank the correct item higher, indicating its ability to leverage the impact of `User-side Content Similarity` to achieve better recommendations. Meanwhile, *D-ISOLATOR* eliminates the impact of similarity, offering a more diverse range of recommended items.
- For `Content-Diverse Users`, ISOLATOR can accurately identify user-preferred candidates with lower content similarity and produce a more diversified recommendation list. For example, *vanilla* recommends highly similar baby stroller toys, while ISOLATOR recommends a less similar handheld toy, precisely meeting the needs of `Content-Diverse Users`.