RESEARCH-ARTICLE

Improving Item-side Fairness of Multimodal Recommendation via Modality Debiasing

YU SHANG, Tsinghua University, Beijing, China

CHEN GAO, Beijing National Research Center for Information Science and Technology, Beijing, China

JIANSHENG CHEN, University of Science and Technology Beijing, Beijing, China

DEPENG JIN, Tsinghua University, Beijing, China

YONG LI, Tsinghua University, Beijing, China

Published: 13 May 2024

Citation in BibTeX format

WWW '24: The ACM Web Conference 2024
May 13 - 17, 2024
Singapore, Singapore

Conference Sponsors:
SIGWEB

WWW '24: Proceedings of the ACM Web Conference 2024 (May 2024)
https://doi.org/10.1145/3589334.3648156
ISBN: 9798400701719

Improving Item-side Fairness of Multimodal Recommendation via Modality Debiasing

Yu Shang
Department of Electronic Engineering, BNRist, Tsinghua University
Beijing, China
shangy21@mails.tsinghua.edu.cn

Chen Gao
BNRist, Tsinghua University
Beijing, China
chgao96@gmail.com

Jiansheng Chen*
School of Computer & Communication Engineering, University of Science and Technology Beijing
Beijing, China
jschen@ustb.edu.cn

Depeng Jin
Department of Electronic Engineering, BNRist, Tsinghua University
Beijing, China
jindp@tsinghua.edu.cn

Yong Li
Department of Electronic Engineering, BNRist, Tsinghua University
Beijing, China
liyong07@tsinghua.edu.cn

ABSTRACT

Multimodal recommender systems have acquired applications in broad web scenarios such as e-commerce businesses and short-video platforms. Existing multimodal recommendation methods generally boost performance by introducing item-side multimodal content as supplement information. However, the common training paradigm, *i.e.,* encoding unimodal content respectively and fusing them to fit user preference scores, makes the model biased towards items with prevailing modality content under non-uniform training data. This results in a serious item-side unfairness issue, *i.e.,* some items with prevailing modality content are over-recommended while a large number of items don't receive adequate recommendation opportunities, leaving corresponding content providers at great disadvantage. Aiming to eliminate such modality bias and promote item-side fairness, we propose a fairness-aware modality debiasing framework based on counterfactual inference. In the training stage, we additionally introduce unimodal prediction branches to capture the modality bias. In the inference stage, we conduct a fairness-aware counterfactual inference to adaptively eliminate the modality bias. The proposed framework is model-agnostic and flexible to be implemented in various multimodal recommendation models. Extensive experiments on two datasets demonstrate that the proposed method can significantly enhance item-side fairness while providing competitive recommendation accuracy. Our proposed framework is expected to help mitigate the unfair treatment experienced by vulnerable content providers on multimedia web platforms. Codes are available in https://github.com/tsinghua-fib-lab/WWW2024-Modality-Debiasing.

*The corresponding authors.

WWW '24, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0171-9/24/05.
https://doi.org/10.1145/3589334.3648156

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Item-side Fairness, Multimodal Recommendation, Modality Debiasing

ACM Reference Format:
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 Web Conference 2024 (WWW '24), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3589334.3648156

1 INTRODUCTION

Multimodal recommendation has been widely used in many online services such as e-commerce [38, 41] and micro-video platforms [14, 15]. Distinct from the traditional recommendation largely relying on historical interactions to model user preference [3, 9, 11, 27], multimodal recommendation incorporates item-side multimodal content (*e.g.*, visual and textual features) to acquire more comprehensive user and item representations [26, 33]. Currently, plenty of effective multimodal recommendation methods have been proposed, including supervised-learning methods [2, 26, 32, 33] and self-supervised learning methods [23, 36, 41]. Despite the varying design details, these models share a generic recommendation pipeline, *i.e.,* first encoding unimodal content respectively and then fusing them to fit the training data.

However, considering that the modality content distribution in training data is usually uneven, the above-mentioned paradigm will inevitably introduce modality bias into model learning. Specifically, if there's some prevailing modality content in the historical interactions for training, the model will learn a spurious shortcut between such modality content and the final preference prediction. Taking clothing recommendation as an example, if the T-shirt in white frequently occurs in users' historical interactions, the trained model will rely on such visual clues to make predictions and tend

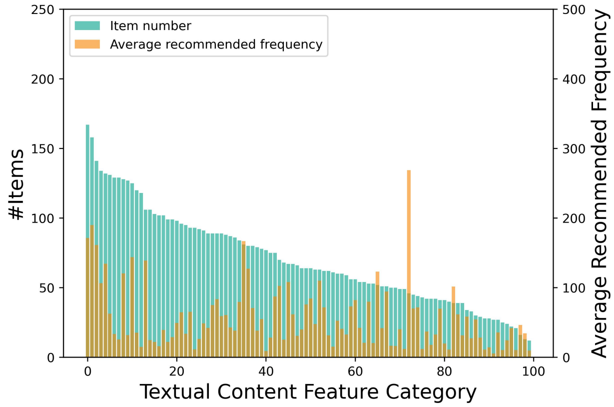

Figure 1: Illustration of item-side unfairness w.r.t. textual content in multimodal recommendation models (results from SLMRec on Baby dataset). Items are classified into 100 groups by clustering the textual features with K-means.

to blindly consider that clothes with such visual content are more likely to be preferred by users. This leads to a severe item-side fairness issue, i.e., some items are more frequently recommended while a large proportion of items don't obtain enough attention. Figure 1 illustrates a real case of items-side unfairness w.r.t. textual content feature categories. It can be seen that some categories with few items (*e.g.,* Category 72) obtain very high recommendation frequency, while some categories with many more items receive scarce recommendations (*e.g.,* Category 8), exhibiting severe bias towards modality content. In the long run, such item-side unfairness will harm both user experience and the rights of content providers, hindering the development of businesses and platforms. Therefore, it's urgently expected to improve the item-side fairness of multimodal recommendation models, making the item distribution in the recommendation list more fair and even.

Currently, there have been many works to promote the item-side fairness of recommender systems, mainly including ranking methods [1, 28] and re-ranking methods [16, 17]. Ranking methods usually inject the fairness constraint into the objective function of recommendation algorithms. Re-ranking methods work after the training phase, taking the biased recommendation list as input and re-ranking items according to the fairness requirement, *e.g.*, enforcing the minimum item coverage. Besides, some debiasing methods can also facilitate item-side fairness [22, 30]. However, all the above methods lack consideration of the unique paradigm of multimodal recommendation and overlook the special unfairness cause in these models, resulting in poor recommendation fairness.

To tackle this issue, we propose a counterfactual inference-based debiasing framework tailored for general multimodal recommendation models to promote item-side fairness. Specifically, we first formulate the causal graph of the generic multimodal recommendation pipeline. From the causal view, the final preference prediction indeed includes two parts: multimodal feature-based prediction and unimodal feature-based prediction. Existing works are generally aware of the former part while overlooking the latter one, which represents the cause-effect pathway of modality bias. To remove the influence of such modality bias, during the training phase, we add extra unimodal prediction branches in the original training paradigm to better estimate the effect of modality bias. During the inference phase, for each modality (*e.g.,* vision, text), we imagine a counterfactual world where the unimodal feature only directly influences the final prediction and has no impact on the fused item representation. Then we conduct a counterfactual inference, deducting the preference score obtained in the counterfactual world from the overall preference score. Besides, to achieve more precise debiasing, we incorporate a fairness-aware debiasing strength design into the debiasing process based on the relation between item preference score ranking and modality content frequency ranking. Based on the above designs, we establish an adaptive modality debiasing framework driven by the item-side fairness objective. Notably, our whole framework is model-agnostic and can be conveniently implemented on existing mainstream multimodal recommendation models. We believe that the proposed framework can help mitigate the unfair treatment experienced by vulnerable content providers on multimedia web platforms.

The contributions of this work are summarized as follows:

- We point out that a critical cause of the unfairness issue in multimodal recommendation models is modality bias, and formulate the causal graph to analyze its cause-effect pathway.
- We propose a novel fairness-aware modality debiasing framework to improve the item-side fairness of multimodal recommendation models. The method is model-agnostic and can be implemented in various multimodal recommendation models.
- We conduct extensive experiments on two datasets, which confirm the effectiveness of our framework in enhancing item-side fairness while providing competitive recommendation accuracy.

2 RELATED WORK

2.1 Multimodal Recommendation

Multimodal recommendation methods generally incorporate item-side multimodal content with traditional collaborative filtering signals in model learning, aiming to enrich the semantics of user and item representations. Some early works like VBPR [8] adopt matrix factorization to encode the combination of multimodal information and ID embeddings. With the blooming of deep learning, more advanced techniques have been utilized to build multimodal recommendation models, *e.g.,* variational autoencoder [24, 37] and graph neural networks [5, 10, 31, 33]. MMGCN [33] is the first to utilize the graph convolutional network to learn modality representations separately and then conduct multimodal fusion to obtain the final representation. Besides, self-supervised learning methods have also been utilized to model the relation between different modalities [23, 36]. However, existing multimodal recommendation methods commonly focus on modality representation learning for better accuracy while overlooking the potential fairness issue.

2.2 Fairness of Recommender Systems

The fairness issue of recommender systems has received great attention from academia and industry for a long period, including both user-side [12, 13, 19] and item-side fairness [25, 28, 43] (the focus of this work). According to the working phase, there are three approaches to improve recommendation fairness: data-oriented

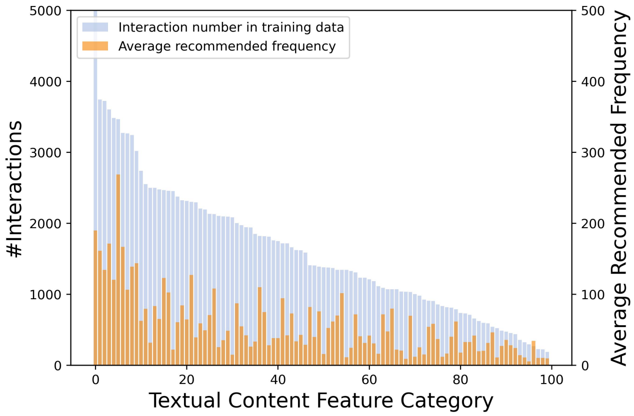

Figure 2: Distribution of interaction frequency and average recommended frequency of different textual content groups (results from SLMRec on Baby dataset).

methods, ranking methods and re-ranking methods [29]. Data-oriented methods usually aim at making the training data unbiased to alleviate the biased model [4, 20]. Ranking methods try to modify the model design or optimization goal to generate more fair results [34, 42]. For instance, Zhu *et al.* [42] design a regularization term to force user (or item) representations and fairness-related attribute vectors to be orthogonal. Re-ranking methods work after model training and adjust the item ranking according to fairness-related metrics [17, 35]. For example, Patro *et al.* [17] design a two-stage re-ranking method ensuring the minimum item exposure to promote recommendation fairness. Apart from the three ways mentioned above, some debiasing methods [30, 39] also help improve item-side fairness, *e.g.,* popularity debiasing [30]. However, all these methods are developed under traditional recommendation scenarios, achieving limited performance in the multimodal recommendation scenario which has a unique recommendation pipeline.

3 METHODOLOGY

In this section, we first analyze the cause of modality bias in the multimodal recommendation scenario (Section 3.1), then we detail how modality bias affects the multimodal recommendation pipeline from a causal view (Section 3.2), based on which we finally detail our proposed fairness-aware modality debiasing framework to enhance item-side fairness (Section 3.3).

3.1 Cause of modality bias

In this part, we elaborate on the reason why modality bias exists in the multimodal recommendation scenario. Existing multimodal recommendation models are commonly optimized to fit the training data in an end-to-end manner. When the modality content distribution shows clear unevenness (*e.g.,* long-tail distribution), models will naturally overfit to the prevailing modality content, resulting in a significant bias towards items with such modality content in the preference prediction. We verify this through an empirical analysis (taking the result from SLMRec on Baby dataset as an example).

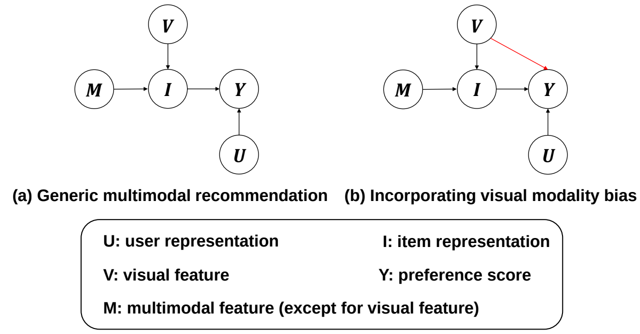

Figure 3: Causal graph for (a) the generic multimodal recommendation pipeline and (b) the version incorporating modality bias (taking visual modality bias as an example).

As shown in Figure 2, it can be found that the distribution of interaction frequency and average recommended frequency across textual content groups have very similar patterns, *e.g.,* the textual content group with more interactions also tends to have a higher recommended frequency. Through the above analysis, we conclude that the cause of modality bias is learning from biased training data with uneven modality content distribution.

3.2 Effect of modality bias from a causal view

For a better understanding of how the modality bias affects preference prediction and recommendation fairness, in this part, we resort to causal graphs, which are directed acyclic graphs describing casual relationships between variables. We construct the causal graph of the generic multimodal recommendation pipeline and the version incorporating modality bias, which is shown in Figure 3. Here we analyze visual modality bias as the example, and there are five variables in this scenario: user representation (U), item representation (I), visual feature (V), multimodal feature except for visual feature (M), and preference score (Y). In the traditional generic multimodal recommendation pipeline shown in Figure 3(a), there are two important causal relations:

- $U \& I \longrightarrow Y$: The fundamental user-item matching, *i.e.,* the preference score is calculated from user and item representations.
- $M \& V \longrightarrow I \longrightarrow Y$: Unimodal features are fused to get the item representation and then affect the final preference score.

However, when considering modality bias (*e.g.,* visual modality bias), there will be a direct effect from the visual feature (V) to the predicted preference score (Y) as shown in Figure 3(b):

- $V \longrightarrow Y$: The mentioned visual modality bias results in the preference prediction to be directly affected by the visual feature, not only based on fused item representation. This relation represents the cause-effect pathway of visual modality bias, which is indeed a learned spurious shortcut to be removed.

As for the effect of modality bias for the multimodal recommendation, we elaborate it through an intuitive clothing recommendation example: if the white T-shirt occurs very frequently in historical interactions for training, the trained model will simply assume clothes with such visual characteristics are highly likely to be preferred by users while neglecting the textual information (*e.g.,* clothing material). As a result, the final recommendation list will be filled

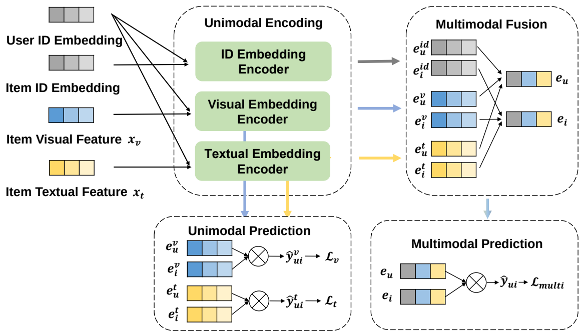

Figure 4: The training pipeline of modality debiasing framework. The first line shows the generic multimodal recommendation pipeline. The second line shows the unimodal and multimodal prediction branches for multi-task learning.

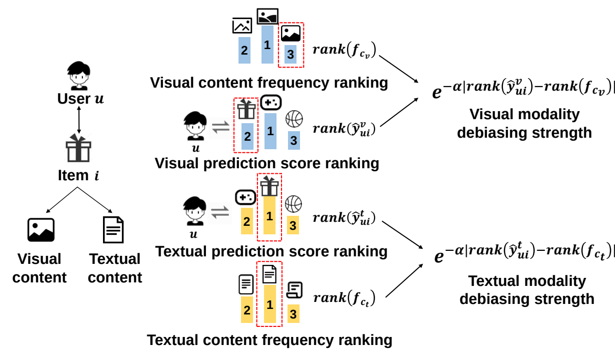

Figure 5: Illustration of fairness-oriented modality debiasing strength calculation.

with homogeneous white T-shirts without other diverse clothes. This issue not only hides the actual preference signal but also leads to unfair item distribution in recommendation results.

3.3 Fairness-aware modality debiasing

Aiming to remove the effect of modality bias and promote item-side fairness, we design a fairness-aware modality debiasing framework based on counterfactual inference techniques. Next, we first introduce counterfactual analysis and then detail the framework design.

3.3.1 Counterfactual analysis. Still taking visual modality bias as an example, counterfactual analysis [18] tries to answer such a question: how the visual feature will affect the preference score when the item representation is fixed. As shown in Figure 3(b), the causal effect (*i.e.,* Total Effect, TE) of V on Y includes two parts: Total Indirect Effect (TIE) through $V \longrightarrow I \longrightarrow Y$ and Natural Direct Effect (NDE) through $V \longrightarrow Y$, which is formulated as follows:

$$TE = TIE + NDE. \tag{1}$$

TE represents the difference between two situations $V = v$ and $V = v^*$ (typically set as null):

$$TE = Y_{u,i,v} - Y_{u,i^*,v^*}, \tag{2}$$

where i^* is a constant value of I when $V = v^*$.

NDE means the value change of Y with V changing from v^* to v through the direct path $V \longrightarrow Y$, while the mediator I is set to the value when $V = v^*$:

$$NDE = Y_{u,i^*,v} - Y_{u,i^*,v^*}. \tag{3}$$

Finally, TIE can be calculated by subtracting NDE from TE:

$$TIE = TE - NDE = Y_{u,i,v} - Y_{u,i^*,v}, \tag{4}$$

which eliminates the negative effect (*i.e., NDE*) caused by the path $V \longrightarrow Y$. Using TIE as the final preference score is expected to mitigate the modality bias.

3.3.2 Multi-task learning training scheme. According to the causal graph in Figure 3, the final preference prediction includes two parts: multimodal feature-based prediction and unimodal feature-based prediction. Taking the scenario with visual and textual modality as an example, the preference score is formulated as:

$$\hat{y}_{u,i,v,t} = \hat{y}_{ui} * \sigma(\hat{y}^v_{ui}) * \sigma(\hat{y}^t_{ui}), \tag{5}$$

where $\sigma(\cdot)$ is the sigmoid function used for scaling unimodal prediction scores to probabilities and adjusting the relying extent upon user-item matching score $\hat{y}_{ui}$, following the design of some existing works [15, 30].

To better capture the direct effect of unimodal features on the preference score, we introduce unimodal prediction branches in the generic multimodal recommendation pipeline and adopt a multi-task learning scheme for model training. The whole training pipeline is shown in Figure 4. The traditional multimodal feature-based prediction branch is supervised by the BPR loss [21], which is formulated as:

$$\mathcal{L}_{multi} = -\frac{1}{N}\sum_{i,j}(\hat{y}_{ui} - \hat{y}_{uj}). \tag{6}$$

Similarly, visual and textual feature-based prediction branches are learned with the following objectives:

$$\mathcal{L}_v = -\frac{1}{N}\sum_{i,j}(\hat{y}^v_{ui} - \hat{y}^v_{uj}), \mathcal{L}_t = -\frac{1}{N}\sum_{i,j}(\hat{y}^t_{ui} - \hat{y}^t_{uj}). \tag{7}$$

We aggregate the above objectives and get the final loss function $\mathcal{L}$ as follows:

$$\mathcal{L} = \mathcal{L}_{multi} + \mathcal{L}_v + \mathcal{L}_t. \tag{8}$$

Here we grant each prediction task identical weights for simplicity, avoiding intensive hyper-parameter tuning.

3.3.3 Fairness-aware counterfactual inference. According to the analysis in Section 3.3.1, we conclude that the modality bias can be mitigated by taking TIE as the final preference score. Considering that different modalities might contain different levels of bias, we choose to conduct separate modality debiasing for each modality. Taking visual modality debiasing as an example, according to Eq. (4), the debiased preference score is as follows:

$$\hat{y}_{ui} * \sigma(\hat{y}^v_{ui}) * \sigma(\hat{y}^t_{ui}) - \hat{y}_{ui^*} * \sigma(\hat{y}^v_{ui}) * \sigma(\hat{y}^{t^*}_{ui}), \tag{9}$$

where i^* and t^* mean the fused item representation and textual feature are set as constant values. Similarly, for the textual modality, we can get the debiased preference score as follows:

$$\hat{y}_{ui} * \sigma(\hat{y}^v_{ui}) * \sigma(\hat{y}^t_{ui}) - \hat{y}_{ui^*} * \sigma(\hat{y}^{v^*}_{ui}) * \sigma(\hat{y}^t_{ui}), \tag{10}$$

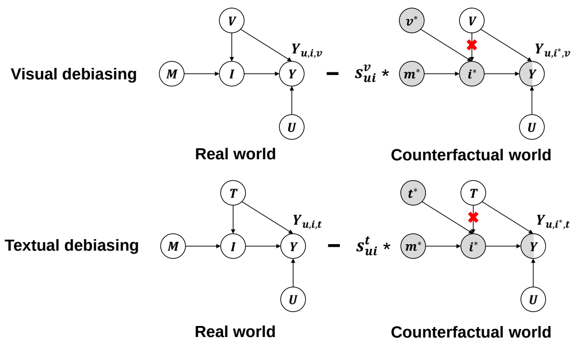

Figure 6: Illustration of the fairness-aware counterfactual inference-based modality debiasing framework.

where i^* and v^* mean the fused item representation and visual feature are set as constant values.

Based on the above standard counterfactual inference process, we propose to introduce fairness-oriented modality debiasing strength for more precise score adjustment. We scrutinize the relationship between two factors: item preference score ranking and modality content frequency ranking. Intuitively, when an item with prevailing modality content ranks high, it indicates that the modality bias produces a clear effect on worsening the unfairness issue, making such items more likely to be recommended. In this situation, the modality bias should be strongly eliminated. On the contrary, if an item with prevailing modality content has a low ranking, it means that the modality bias shows no significant impact on fairness, in which case a weak debiasing strength is enough.

Following this idea, we design an effective method to generate the fairness-oriented modality debiasing strength as shown in Figure 5. Taking visual modality debiasing as an example, given an instance (u, i, v) related to user u, item i and item visual feature v, we first extract the visual content frequency ranking $rank(f_{c_v})$, where f_{c_v} is the interaction frequency of the visual content group c_v which visual feature v belongs to. Next, we extract the visual preference score ranking of item i, denoted as $rank(\hat{y}^v_{ui})$. Finally we define the debiasing strength s^v_{ui} as follows:

$$s^v_{ui} = e^{-\alpha|rank(\hat{y}^v_{ui})-rank(f_{c_v})|}, \tag{11}$$

where α is a hyper-parameter adjusting the strength distribution. In this way, the larger gap between the two rankings produces smaller debaising strength. Here we try various decaying functions and find that the exponential function works slightly better so we finally adopt it. Similarly for the textual modality, the debiasing strength s^t_{ui} is defined as:

$$s^t_{ui} = e^{-\alpha|rank(\hat{y}^t_{ui})-rank(f_{c_t})|}. \tag{12}$$

Up to now, we can derive the final modality debiasing framework, as shown in Figure 6. The fairness-aware visual debiasing is formulated as follows:

$$TIE(visual) = \hat{y}_{ui} * \sigma(\hat{y}^v_{ui}) * \sigma(\hat{y}^t_{ui}) - s^v_{ui} * \hat{y}_{ui^*} * \sigma(\hat{y}^v_{ui}) * \sigma(\hat{y}^{t^*}_{ui}). \tag{13}$$

Table 1: Statistics of two datasets for experiments.

Dataset	#Users	#Items	#Interactions	Sparsity
Baby	19,445	7,050	160,792	99.88%
Clothing	39,387	23,033	278,677	99.97%

Similarly, for the textual modality, we can get the fairness-aware debiased preference score as follows:

$$TIE(text) = \hat{y}_{ui} * \sigma(\hat{y}^v_{ui}) * \sigma(\hat{y}^t_{ui}) - s^t_{ui} * \hat{y}_{ui^*} * \sigma(\hat{y}^{v^*}_{ui}) * \sigma(\hat{y}^t_{ui}). \tag{14}$$

We take the average of the above two terms as the final prediction score, which eliminates both the visual and textual modality bias:

$$TIE(final) = \frac{TIE(visual) + TIE(text)}{2} \tag{15}$$

4 EXPERIMENTS

To validate the effectiveness of our proposed modality debiasing framework, in this section, we conduct experiments to answer the following research questions:

- **RQ1:** Does the proposed modality debiasing bring greater fairness improvement than other methods for multimodal recommendation models?
- **RQ2:** How do different designs impact the performance of modality debiasing?
- **RQ3:** Does modality debiasing effectively mitigate the unfairness issue caused by modality bias?

4.1 Experiment Settings

4.1.1 Datasets. We utilize Amazon product review dataset [7] to conduct experiments, containing visual and textual modality content features of products. We select two subcategories, *i.e.,* Baby and Clothing datasets for experiments, which have been widely used in previous works [8, 38, 41]. The statistics of the two datasets are shown in Table 1. The interactions of each user are randomly split into training, validation and testing sets with the ratio 8:1:1 following previous works [8, 38, 41].

4.1.2 Baselines. We mainly consider three kinds of methods helpful to item-side fairness for comparison, including ranking methods, re-ranking methods and debiasing methods. The three specific baselines are as follows:

- **FairRec** [17]: This is a re-ranking method adjusting the recommended item distribution to ensure the minimum item coverage in final recommendation lists.
- **MACR** [30]: This is a debiasing method to eliminate the popularity bias, which also helps for item-side fairness.
- **Embedding re-ranking** [28]: This method introduces a fairness-related constraint called (α, β)-fairness in the training objective to improve item-side fairness.

4.1.3 Evaluation metrics. We evaluate both the recommendation accuracy and item-side fairness on our method and baselines. All metrics are calculated with the top K (set as 10 and 20) items following [38, 40]. For the accuracy metrics, we use top-K Recall (R@K) and Normalized Discounted Cumulative Gain (N@K). For the item-side fairness metrics, we use Gini index, Entropy and Coverage

Table 2: Performance of different fairness-improving methods on five base models on Baby dataset. For each base model, the best and second-best results in each column are highlighted in bold and underline, respectively.

Baby Dataset		Accuracy-related metrics				Fairness-related metrics					
Model	Method	R@10 (↑)	R@20 (↑)	N@10 (↑)	N@20 (↑)	G@10 (↓)	G@20 (↓)	E@10 (↑)	E@20 (↑)	C@10 (↑)	C@20 (↑)
MMGCN	Standard	0.0354	0.0584	0.0189	0.0248	0.888	0.876	4.392	4.964	37.0%	52.1%
	FairRec	0.0328	0.0539	0.0172	0.0224	0.866	0.858	4.995	5.364	40.7%	59.2%
	MACR	**0.0363**	**0.0589**	**0.0195**	**0.0251**	0.847	0.848	5.268	5.685	42.6%	59.5%
	Embedding Re-ranking	0.0348	0.0556	0.0179	0.0232	0.845	0.844	5.545	5.876	46.1%	61.6%
	Modality Debiasing	0.0339	0.0568	0.0178	0.0238	**0.835**	**0.832**	**6.330**	**6.771**	**53.7%**	**67.8%**
VBPR	Standard	0.0414	0.0650	0.0220	0.0281	0.897	0.899	5.807	6.095	60.4%	76.3%
	FairRec	0.0357	0.0622	0.0201	0.0258	0.865	0.863	5.921	6.233	65.6%	79.8%
	MACR	0.0426	0.0656	0.0229	0.0288	0.867	0.864	5.984	6.188	64.1%	77.4%
	Embedding Re-ranking	0.0384	0.0632	0.0204	0.0265	0.852	0.859	6.203	6.375	72.2%	83.5%
	Modality Debiasing	**0.0437**	**0.0683**	**0.0242**	**0.0305**	**0.825**	**0.821**	**6.701**	**6.952**	**80.3%**	**87.5%**
GRCN	Standard	0.0476	0.0754	0.0262	0.0326	0.822	0.817	6.564	6.621	83.6%	85.9%
	FairRec	0.0432	0.0735	0.0234	0.0310	0.803	0.796	6.961	7.026	87.5%	90.4%
	MACR	0.0464	0.0752	0.0249	0.0323	0.812	0.809	6.885	6.903	85.8%	87.6%
	Embedding Re-ranking	0.0445	0.0741	0.0237	0.0317	0.808	0.803	6.922	7.061	88.7%	90.9%
	Modality Debiasing	**0.0489**	**0.0765**	**0.0273**	**0.0348**	**0.795**	**0.787**	**7.061**	**7.158**	**90.8%**	**96.8%**
SLMRec	Standard	0.0483	0.0727	0.0272	0.0330	0.833	0.816	6.546	7.011	48.2%	68.2%
	FairRec	0.0466	0.0705	0.0260	0.0318	0.805	0.794	6.785	7.148	55.6%	73.5%
	MACR	0.0485	0.0720	0.0270	0.0329	0.808	0.802	6.723	7.167	54.2%	73.8%
	Embedding Re-ranking	0.0469	0.0711	0.0265	0.0323	0.803	0.799	6.846	7.215	58.6%	76.2%
	Modality Debiasing	**0.0493**	**0.0729**	**0.0274**	**0.0333**	**0.782**	**0.753**	**7.084**	**7.506**	**64.5%**	**82.1%**
MMGCL	Standard	0.0484	**0.0754**	0.0262	0.0331	0.835	0.831	5.854	6.319	24.2%	36.8%
	FairRec	0.0471	0.0738	0.0252	0.0320	0.805	0.804	5.968	6.487	28.8%	43.5%
	MACR	**0.0486**	0.0753	0.0264	**0.0338**	0.825	0.819	5.926	6.397	29.6%	45.6%
	Embedding Re-ranking	0.0473	0.0741	0.0253	0.0324	0.801	0.798	6.034	6.465	32.5%	47.5%
	Modality Debiasing	0.0482	0.0743	**0.0265**	0.0334	**0.795**	**0.781**	**6.176**	**6.638**	**37.2%**	**54.9%**

to measure the uniformity of item distribution in all users' recommendation lists widely used in previous works [6, 16]. Here we specifically detail fairness-related metrics as follows:

- **Gini index@K (G@K)**: The measure of how uniformly items are distributed in the whole recommendation list, ranging between 0 and 1. The lower value means more uniform and fair. Given all users' recommendation lists L and the proportion of the k-th least recommended item $p(i_k|L)$, Gini index is formulated as:

$$Gini\,index(L) = \frac{1}{|I|-1}\sum_{k=1}^{|I|}(2k-|I|-1)p(i_k|L), \tag{16}$$

  where $|I|$ is the total number of items.
- **Entropy@K (E@K)**: Another measure of how uniformly items are distributed in the whole recommendation list. A larger value means a more uniform item distribution, which is formulated as:

$$Entropy(L) = -\sum_{i\in I} p(i|L)\log p(i|L), \tag{17}$$

  where $p(i|L)$ is the occurring proportion of item i in the recommendation list L.
- **Coverage@K (C@K)**: The proportion of the items that occur at least once in the recommendation lists to all items.

4.1.4 Implementation Details. We test our method and baselines on five mainstream multimodal recommendation models, including MMGCN [33], VBPR [8], GRCN [32], SLMRec [23] and MMGCL [36], covering supervised and self-supervised methods. We set the embedding size as 64 for all models following previous works [38, 41]. We set other parameters following the original papers. For the modality debiasing, we search the hyper-parameter α in $\{0.001, 0.01, 0.1, 1, 10\}$, wide enough to find the optimal value.

4.2 Performance Comparison (RQ1)

We incorporate our modality debiasing framework and baseline methods into the five aforementioned multimodal recommendation models, then compare both accuracy-related and fairness-related metrics. The overall results on two datasets are summarized in Table 2 and 3, from which we have the following observations:

- **In terms of item-side fairness, our modality debiasing method stably outperforms other fairness-improving baselines in all cases.** It can be found that in the multimodal recommendation scenario, the modality debiasing gets a higher fairness gain than various types of baselines. Statistically, on Baby dataset, our method achieves a 2.36% improvement in Gini index, 6.18% improvement in Entropy and 9.93% improvement in Coverage. On Clothing dataset, our method achieves a 3.22% improvement on Gini index, 4.00% improvement in Entropy and 10.97% improvement in Coverage. This demonstrates the superiority of our method in improving item-side fairness.
- **Modality debiasing achieves the most competitive recommendation accuracy in general.** For example, on Baby dataset, compared with the standard model, our framework can still have a positive accuracy gain, improving 1.87% on Recall and 3.76% on NDCG. On the contrary, other methods are harmful to the accuracy in general, *e.g.,* Embedding Re-ranking leads to a 3.38% decrease in Recall and a 4.48% decrease in NDCG compared with the standard model. Therefore, our method achieves the best trade-off between recommendation accuracy and fairness.

Table 3: Performance of different fairness-improving methods on five base models on Clothing dataset. For each base model, the best and second-best results in each column are highlighted in bold and underline, respectively.

Clothing Dataset		Accuracy-related metrics				Fairness-related metrics					
Model	**Method**	**R@10 (↑)**	**R@20 (↑)**	**N@10 (↑)**	**N@20 (↑)**	**G@10 (↓)**	**G@20 (↓)**	**E@10 (↑)**	**E@20 (↑)**	**C@10 (↑)**	**C@20 (↑)**
MMGCN	Standard	**0.0245**	**0.0384**	**0.0122**	**0.0158**	0.860	0.858	6.597	7.042	19.8%	29.2%
	FairRec	0.0187	0.0306	0.0101	0.0122	0.838	0.836	7.045	7.387	27.8%	39.6%
	MACR	0.0217	0.0358	0.0116	0.0143	0.849	0.845	6.789	7.146	26.6%	37.5%
	Embedding Re-ranking	0.0213	0.0345	0.0109	0.0138	0.831	0.828	7.215	7.476	33.4%	45.8%
	Modality Debiasing	0.0201	0.0323	0.0104	0.0135	**0.795**	**0.789**	**7.699**	**8.037**	**42.1%**	**55.5%**
VBPR	Standard	0.0246	0.0363	0.0135	0.0163	0.845	0.841	7.403	7.784	51.1%	67.2%
	FairRec	0.0231	0.0352	0.0122	0.0154	0.822	0.818	7.785	7.969	60.1%	74.5%
	MACR	0.0250	0.0364	0.0132	0.0161	0.809	0.802	7.938	8.146	63.7%	77.9%
	Embedding Re-ranking	0.0236	0.0354	0.0125	0.0155	0.786	0.782	8.067	8.301	68.1%	80.6%
	Modality Debiasing	**0.0256**	**0.0395**	**0.0137**	**0.0170**	**0.750**	**0.741**	**8.424**	**8.688**	**71.6%**	**84.5%**
GRCN	Standard	0.0385	0.0605	0.0201	0.0257	0.664	0.663	9.018	9.112	87.1%	93.5%
	FairRec	0.0376	0.0596	0.0194	0.0244	0.652	0.650	9.037	9.125	87.6%	93.9%
	MACR	0.0390	0.0603	0.0203	0.0254	0.650	0.648	9.065	9.139	87.9%	94.4%
	Embedding Re-ranking	0.0376	0.0593	0.0189	0.0238	0.660	0.658	9.026	9.136	87.4%	93.8%
	Modality Debiasing	**0.0395**	**0.0613**	**0.0207**	**0.0262**	**0.645**	**0.646**	**9.105**	**9.185**	**89.7%**	**95.7%**
SLMRec	Standard	**0.0440**	**0.0661**	**0.0236**	**0.0293**	0.863	0.861	6.794	7.309	24.5%	40.0%
	FairRec	0.0422	0.0643	0.0228	0.0286	0.845	0.840	6.987	7.568	32.6%	49.5%
	MACR	0.0420	0.0639	0.0223	0.0281	0.859	0.852	6.936	7.425	34.6%	48.8%
	Embedding Re-ranking	0.0413	0.0630	0.0218	0.0276	0.840	0.838	7.054	7.592	37.6%	52.8%
	Modality Debiasing	0.0427	0.0654	0.0231	0.0291	**0.825**	**0.814**	**7.448**	**7.965**	**43.2%**	**60.1%**
MMGCL	Standard	**0.0430**	**0.0643**	**0.0231**	**0.0281**	0.870	0.869	6.324	6.880	15.3%	25.1%
	FairRec	0.0415	0.0618	0.0215	0.0262	0.853	0.847	6.430	6.969	16.2%	27.3%
	MACR	0.0420	0.0621	0.0221	0.0268	0.857	0.852	6.419	6.934	17.1%	27.8%
	Embedding Re-ranking	0.0411	0.0614	0.0212	0.0256	0.843	0.841	6.467	6.951	17.8%	28.8%
	Modality Debiasing	0.0422	0.0634	0.0225	0.0274	**0.829**	**0.825**	**6.621**	**7.152**	**19.7%**	**31.6%**

Table 4: Ablation study on the effect of fairness-oriented debiasing strength with Gini index@20 as the fairness metric.

Dataset	Method	MMGCN	VBPR	GRCN	SLMRec	MMGCL
Baby	Standard	0.876	0.899	0.817	0.816	0.831
	Full debiasing	**0.832**	**0.821**	**0.787**	**0.753**	**0.781**
	w/o debiasing strength	0.853	0.859	0.803	0.787	0.804
Clothing	Standard	0.858	0.841	0.663	0.861	0.869
	Full debiasing	**0.789**	**0.741**	**0.646**	**0.814**	**0.825**
	w/o debiasing strength	0.795	0.816	0.650	0.844	0.839

Table 5: Results of debiasing for each modality with Gini index@20 as the fairness metric.

Dataset	Method	MMGCN	VBPR	GRCN	SLMRec	MMGCL
Baby	Full debiasing	**0.832**	**0.821**	**0.787**	**0.753**	**0.781**
	Visual debiasing only	0.848	0.839	0.795	0.768	0.801
	Textual debiasing only	0.863	0.867	0.814	0.796	0.812
Clothing	Full debiasing	**0.789**	**0.741**	**0.646**	**0.814**	**0.825**
	Visual debiasing only	0.802	0.765	0.658	0.830	0.837
	Textual debiasing only	0.818	0.791	0.672	0.849	0.845

- **Modality bias has a more significant impact on the fairness issue than popularity bias.** We compare our method with MACR, a popularity debiasing framework, and find that modality debiasing is more effective in enhancing item-side fairness. Considering that multimodal recommendation models regard modality content as the most important information, the bias contained in the modality content will be more significant than the popularity bias. To sum up, in the multimodal recommendation scenario, the modality bias is more to blame for the unfairness.

4.3 In-depth Analysis (RQ2)

In this section, we study how some designs impact the performance of modality debiasing, including the effect of introducing fairness-oriented modality debiasing strength and a hyper-parameter α for adjusting debiasing strength. Besides, we explore the effect of debiasing for different modalities.

4.3.1 Effect of fairness-oriented modality debiasing strength. In order to verify the effectiveness of the proposed fairness-oriented modality debiasing strength, we conduct an ablation study and compare the fairness-related metric (focusing on Gini index@20) obtained by standard models, models equipped with the full debiasing framework, and models equipped with the debiasing framework without debiasing strength. The results are shown in Table 4, from which it can be seen that the fairness gain shrinks when discarding the debiasing strength design from the full framework, verifying its contribution to the whole performance.

4.3.2 Effect of the selection of hyper-parameter α. In our framework, we introduce a hyper-parameter α as a coefficient on the exponential term to adjust the debiasing strength distribution. For a better understanding of its effect, we conduct a hyper-parameter study and search it in $\{0.001, 0.01, 0.1, 1, 10\}$, which is wide enough to find the optimal value. Here we take Gini index@20 and Entropy@20 as the fairness metrics and conduct experiments on two datasets. The results are presented in Figure 7, from which it can be found that VBPR, SLMRec and MMGCL achieve the best fairness when α is 0.01 at most times, MMGCN and GRCN get the best result when α is 0.1. Therefore, we follow this finding to set the value of α for different models.

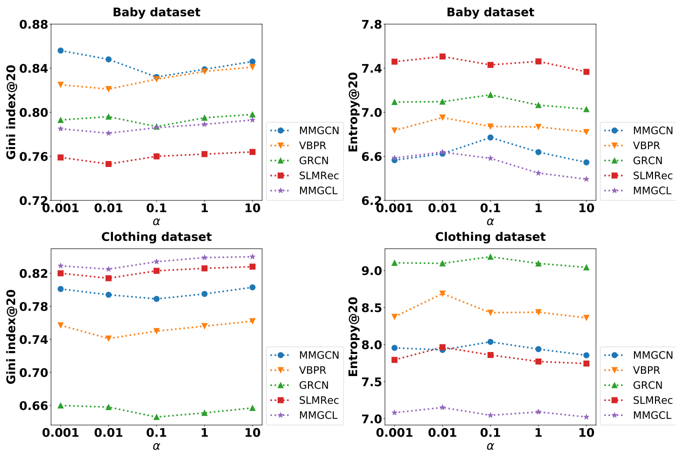

Figure 7: Study of the impact of α on the recommendation fairness on two datasets.

4.3.3 Effect of debiasing for different modalities. In this part, we try to answer the following question: *how much modality bias or how much unfairness is introduced by each modality?* To this end, we separate the whole modality debiasing process into two parts: visual debiasing only and textual debiasing only, and compare the fairness of obtained recommendation results. The comparison result is presented in Table 5, from which we can see that the visual modality debiasing contributes more to the whole improvement of fairness than the textual modality debiasing. What's more, this phenomenon is consistent in the two datasets. From this view, we conclude that the unfairness issue brought by visual modality bias is more critical than textual modality bias.

4.4 Capability of Modality Debiasing (RQ3)

In this section, we investigate whether our proposed modality debiasing framework effectively eliminates the modality bias and mitigates the unfairness issue. We compare the recommended item distribution across different modality content groups before and after modality debiasing, taking the results from SLMRec on Baby dataset as an example. As shown in Figure 8 and 9, it can be seen that the recommended frequency distribution of different visual or textual content groups becomes more uniform after modality debiasing. Remarkably, the average recommendation frequency of items with prevailing modality content is reduced because we conduct a stronger score deduction towards these items in the fairness-aware debiasing process. In this way, many previously disadvantaged items can obtain more recommended opportunities. This study demonstrates that the modality bias can be effectively eliminated with our method and the item distribution in the final recommendation list can be much more fair.

5 CONCLUSION

In this paper, we present a systematic study of the modality bias ubiquitous in the multimodal recommendation scenario and figure out its negative impact on item-side fairness. To tackle this problem, we first analyze the cause of such modality bias and point out that it is attributed to the overfitting of training data with uneven modality content distribution, making items with prevailing modality content over-recommended. Following this finding we propose a model-agnostic fairness-aware modality debiasing framework based on counterfactual inference techniques. Extensive experiments demonstrate the effectiveness of our proposed framework in improving item-side fairness with very competitive recommendation accuracy. We believe our framework can help alleviate the unfair treatment suffered by vulnerable content providers on multimedia web platforms.

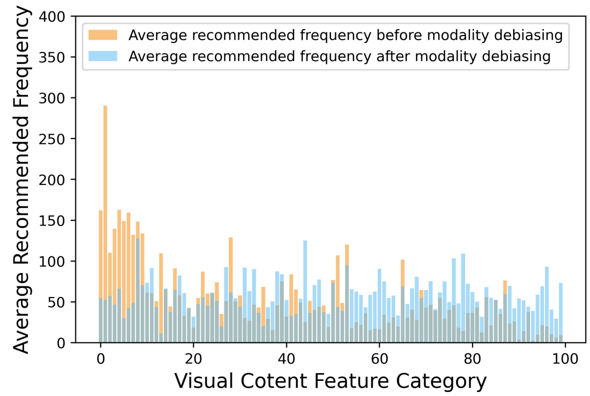

Figure 8: Distribution of the average recommended frequency of different visual content groups before and after modality debiasing (results from SLMRec on Baby dataset).

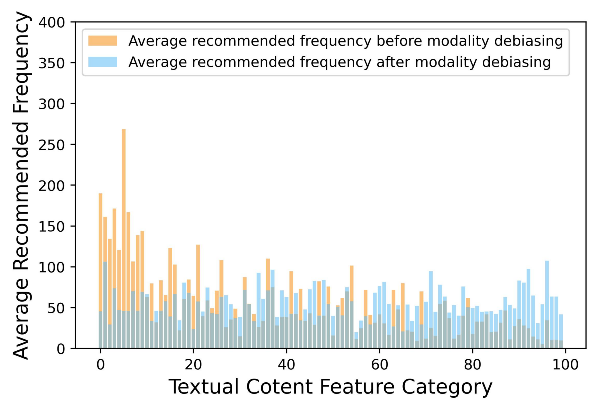

Figure 9: Distribution of the average recommended frequency of different textual content groups before and after modality debiasing (results from SLMRec on Baby dataset).

ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundation of China (No.62272262, No.U23B2030, No.62376024), the National Key Research and Development Program of China (No.2022YFB3104702), and a grant from the Guoqiang Institute, Tsinghua University (No. 2021GQG1005).

REFERENCES

[1] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2017. Controlling popularity bias in learning-to-rank recommendation. In *Proceedings of the eleventh ACM conference on recommender systems*. 42–46.

[2] Feiyu Chen, Junjie Wang, Yinwei Wei, Hai-Tao Zheng, and Jie Shao. 2022. Breaking Isolation: Multimodal Graph Fusion for Multimedia Recommendation by Edge-wise Modulation. In *Proceedings of the 30th ACM International Conference on Multimedia*. 385–394.

[3] Jingtao Ding, Yuhan Quan, Quanming Yao, Yong Li, and Depeng Jin. 2020. Simplify and robustify negative sampling for implicit collaborative filtering. *Advances in Neural Information Processing Systems* 33 (2020), 1094–1105.

[4] Michael D Ekstrand, Mucun Tian, Ion Madrazo Azpiazu, Jennifer D Ekstrand, Oghenemaro Anuyah, David McNeill, and Maria Soledad Pera. 2018. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *Conference on fairness, accountability and transparency*. PMLR, 172–186.

[5] Chen Gao, Yu Zheng, Nian Li, Yinfeng Li, Yingrong Qin, Jinghua Piao, Yuhan Quan, Jianxin Chang, Depeng Jin, Xiangnan He, et al. 2023. A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems* 1, 1 (2023), 1–51.

[6] Yingqiang Ge, Shuchang Liu, Ruoyuan Gao, Yikun Xian, Yunqi Li, Xiangyu Zhao, Changhua Pei, Fei Sun, Junfeng Ge, Wenwu Ou, et al. 2021. Towards long-term fairness in recommendation. In *Proceedings of the 14th ACM international conference on web search and data mining*. 445–453.

[7] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*. 507–517.

[8] Ruining He and Julian McAuley. 2016. VBPR: visual bayesian personalized ranking from implicit feedback. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 30.

[9] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong 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*. 639–648.

[10] Hengchang Hu, Wei Guo, Yong Liu, and Min-Yen Kan. 2023. Adaptive Multi-Modalities Fusion in Sequential Recommendation Systems. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 843–853.

[11] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 426–434.

[12] Yunqi Li, Hanxiong Chen, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2021. User-oriented fairness in recommendation. In *Proceedings of the Web Conference 2021*. 624–632.

[13] Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, and Yongfeng Zhang. 2021. Towards personalized fairness based on causal notion. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1054–1063.

[14] Shang Liu, Zhenzhong Chen, Hongyi Liu, and Xinghai Hu. 2019. User-video co-attention network for personalized micro-video recommendation. In *The World Wide Web Conference*. 3020–3026.

[15] Xiaohao Liu, Zhulin Tao, Jiahong Shao, Lifang Yang, and Xianglin Huang. 2022. EliMRec: Eliminating Single-modal Bias in Multimedia Recommendation. In *Proceedings of the 30th ACM International Conference on Multimedia*. 687–695.

[16] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. 2020. Fairmatch: A graph-based approach for improving aggregate diversity in recommender systems. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization*. 154–162.

[17] Gourab K Patro, Arpita Biswas, Niloy Ganguly, Krishna P Gummadi, and Abhijnan Chakraborty. 2020. Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms. In *Proceedings of the web conference 2020*. 1194–1204.

[18] Judea Pearl. 2009. *Causality*. Cambridge university press.

[19] Hossein A Rahmani, Mohammadmehdi Naghiaei, Mahdi Dehghan, and Mohammad Aliannejadi. 2022. Experiments on generalizability of user-oriented fairness in recommender systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2755–2764.

[20] Bashir Rastegarpanah, Krishna P Gummadi, and Mark Crovella. 2019. Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems. In *Proceedings of the twelfth ACM international conference on web search and data mining*. 231–239.

[21] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. 452–461.

[22] Jiakai Tang, Shiqi Shen, Zhipeng Wang, Zhi Gong, Jingsen Zhang, and Xu Chen. 2023. When Fairness meets Bias: a Debiased Framework for Fairness aware Top-N Recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 200–210.

[23] Zhulin Tao, Xiaohao Liu, Yewei Xia, Xiang Wang, Lifang Yang, Xianglin Huang, and Tat-Seng Chua. 2022. Self-supervised learning for multimedia recommendation. *IEEE Transactions on Multimedia* (2022).

[24] Nhu-Thuat Tran and Hady W Lauw. 2022. Aligning Dual Disentangled User Representations from Ratings and Textual Content. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1798–1806.

[25] Jiayin Wang, Weizhi Ma, Jiayu Li, Hongyu Lu, Min Zhang, Biao Li, Yiqun Liu, Peng Jiang, and Shaoping Ma. 2022. Make Fairness More Fair: Fair Item Utility Estimation and Exposure Re-Distribution. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1868–1877.

[26] Qifan Wang, Yinwei Wei, Jianhua Yin, Jianlong Wu, Xuemeng Song, and Liqiang Nie. 2021. Dualgnn: Dual graph neural network for multimedia recommendation. *IEEE Transactions on Multimedia* (2021).

[27] 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*. 165–174.

[28] Xiuling Wang and Wendy Hui Wang. 2022. Providing Item-side Individual Fairness for Deep Recommender Systems. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 117–127.

[29] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2023. A survey on the fairness of recommender systems. *ACM Transactions on Information Systems* 41, 3 (2023), 1–43.

[30] Tianxin Wei, Fuli Feng, Jiawei Chen, Ziwei Wu, Jinfeng Yi, and Xiangnan He. 2021. Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 1791–1800.

[31] Yinwei Wei, Xiang Wang, Xiangnan He, Liqiang Nie, Yong Rui, and Tat-Seng Chua. 2021. Hierarchical user intent graph network for multimedia recommendation. *IEEE Transactions on Multimedia* 24 (2021), 2701–2712.

[32] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, and Tat-Seng Chua. 2020. Graph-refined convolutional network for multimedia recommendation with implicit feedback. In *Proceedings of the 28th ACM international conference on multimedia*. 3541–3549.

[33] 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*. 1437–1445.

[34] Le Wu, Lei Chen, Pengyang Shao, Richang Hong, Xiting Wang, and Meng Wang. 2021. Learning fair representations for recommendation: A graph-based perspective. In *Proceedings of the Web Conference 2021*. 2198–2208.

[35] Tao Yang and Qingyao Ai. 2021. Maximizing marginal fairness for dynamic learning to rank. In *Proceedings of the Web Conference 2021*. 137–145.

[36] Zixuan Yi, Xi Wang, Iadh Ounis, and Craig Macdonald. 2022. Multi-modal graph contrastive learning for micro-video recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1807–1811.

[37] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 353–362.

[38] Jinghao Zhang, Yanqiao Zhu, Qiang Liu, Shu Wu, Shuhui Wang, and Liang Wang. 2021. Mining latent structures for multimedia recommendation. In *Proceedings of the 29th ACM International Conference on Multimedia*. 3872–3880.

[39] 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 *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 11–20.

[40] Hongyu Zhou, Xin Zhou, Zhiwei Zeng, Lingzi Zhang, and Zhiqi Shen. 2023. A Comprehensive Survey on Multimodal Recommender Systems: Taxonomy, Evaluation, and Future Directions. *arXiv preprint arXiv:2302.04473* (2023).

[41] 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*. 845–854.

[42] Ziwei Zhu, Xia Hu, and James Caverlee. 2018. Fairness-aware tensor-based recommendation. In *Proceedings of the 27th ACM international conference on information and knowledge management*. 1153–1162.

[43] Ziwei Zhu, Jingu Kim, Trung Nguyen, Aish Fenton, and James Caverlee. 2021. Fairness among new items in cold start recommender systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 767–776.