Alleviating Dual Biases in Recommendation (Student Abstract)

Sijin Lu, Fangyuan Luo, Jun Wu*

MoE Key Lab of Big Data & Artifcial Intelligence in Transportation, Beijing Jiaotong University, Beijing 100044, China
{lusijin, fangyuanluo, wuj}@bjtu.edu.cn

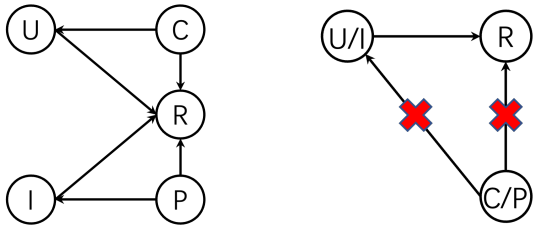
Abstract

Causal Inference (CI) plays a crucial role in building unbiased recommender systems. However, most current CI-based debiasing methods only pay attention on either popularity bias or conformity bias. This paper presents a Disentangled Counterfactual Reasoning framework to alleviate dual biases in recommendation, so called DCR. Concretely, we consider the impact of both item popularity and user conformity during training, and separate their indirect effects by disentangling user and item embeddings into biased and unbiased components. In the inference stage, we perform counterfactual reasoning to simultaneously mitigate the indirect and direct effects of bias factors. Experimental results demonstrate the effectiveness of our DCR.

Introduction

Recommender systems provide users with personalized contents by learning from historical user-item interactions. As the user-item interactions are observational rather than experimental, biases are easily introduced into the observed data. Currently, Causal Inference (CI) (Pearl 2009) provides key insights for addressing biases in recommender systems from a causal view. Most CI-based debiasing methods only focus on either popularity bias or conformity bias. However, little is known on a more comprehensive framework that is able to addresses dual biases simultaneously. Although Wei et al. (Wei et al. 2021) aimed to eliminate the effects of both item popularity and user conformity, they conflated the biased and unbiased representations of users and items, overlooking the indirect effects of bias factors. Therefore, to achieve more comprehensive debiasing, it is necessary not only to consider the impact of both item popularity and user conformity but also to simultaneously mitigate their indirect and direct effects, as depicted in the causal graphs in Figure 1a and 1b.

To this end, we propose a Disentangled Counterfactual Reasoning (DCR) framework, which incorporates both item popularity and user conformity into the recommendation process. To be specific, we learn disentangled biased and unbiased embeddings on both user and item sides to separate the indirect effects of bias factors. In the inference phase, we conduct counterfactual inference, suppressing biased embeddings in the user-item matching to mitigate the indirect effects of bias factors. Similarly, we imagine a counterfactual world where only the bias factors are considered, and accordingly adjust the actual prediction score to mitigate the direct effects.



(a) Rec with popularity and conformity.

(b) Mitigating indirect and direct effects.

Figure 1: Causal graphs for recommendation process. U: user, I: item, R: prediction score, P: item popularity, C: user conformity.

Method

To separate the indirect effects of bias factors, we use Mutual Information (MI) to disentangle the embeddings of users and items. For all $u \in U$ and $i \in I$, the biased and unbiased embeddings of u and i should not contain MI. Taking the user side as an example, we minimize $\frac{1}{|U|}\sum_{u\in U}\mathcal{I}(e_u^b; e_u^n)$, where e_u^b and e_u^n represent the user biased and unbiased embeddings, respectively. For each positive sample $(u, i) \in R^+$, we aim to maximize the MI between user and item biased embeddings, i.e., maximize $\frac{1}{|R^+|}\sum_{(u,i)\in R^+}\mathcal{I}(e_u^b; e_i^b)$, as well as between user and item unbiased embeddings. To distinguish between biased and unbiased embeddings of users and items, e_u^b and e_i^b should be optimized under two objectives: (1) e_u^b can accurately estimate the user conformity; (2) e_i^b can accurately estimate the item popularity.

Training The training process is the process of fitting the biased data. Therefore, we devise the DCR framework according to the causal graph in Figure 1a. Unlike traditional recommendation models, DCR simultaneously considers the impact of item popularity, user conformity, and user-item

*Corresponding Author

Method	Adressa		MovieLens-1M		Gowalla		Yelp2018	
	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG
UDIPS	0.1420	0.0473	0.3801	0.2325	0.1481	0.0574	0.0843	0.0327
DICE	0.1421	0.0451	0.3766	0.2305	0.1465	0.0562	0.0843	0.0326
MACR	0.1557	0.0508	0.3811	0.2320	0.1603	0.0614	0.0878	0.0344
PPAC	0.2157	0.0797	0.4023	0.2473	0.1849	0.0771	0.1005	0.0401
DCR	**0.2418**	**0.0924**	**0.4395**	**0.2756**	**0.1950**	**0.0810**	**0.1073**	**0.0429**
Impr.	12.1%	15.9%	9.2%	11.4%	5.5%	5.1%	6.8%	7.0%

Table 1: Overall performance of DCR compared with the state-of-the-art methods. We highlight the best results in bold and underline the second-best results. Impr. denotes the improvement of DCR over the best existing performance.

matching score during training. The prediction score is calculated as follows:

$$\hat{r}_{u,i} = \sigma(\hat{p}_i) * \sigma(\hat{c}_u) * f_{\text{Match}}(e_u, e_i), \tag{1}$$

where $\hat{p}_i$ and $\hat{c}_u$ are the values learned from the item biased embedding e_i^b and user biased embedding e_u^b, respectively. The function f_{Match} can be any recommendation model that requires debiasing. Additionally, e_u is obtained by concatenating the biased and unbiased embeddings of users, and e_i is obtained by concatenating the biased and unbiased embeddings of items, i.e., $e_u = [e_u^b, e_u^n]$ and $e_i = [e_i^b, e_i^n]$.

Inference In the inference phase, we no longer rely on the prediction score in the factual world for recommendation. We conduct counterfactual inference, where the biased embeddings of users and items are assigned inverse propensity weights to mitigate the indirect effects of bias factors. The modified embeddings of users and items are

$$e'_u = \left[\frac{e_u^b}{\exp(c_u)}, e_u^n\right], \quad e'_i = \left[\frac{e_i^b}{\exp(p_i)}, e_i^n\right]. \tag{2}$$

where c_u and p_i are calculated from the training set.

Next, to mitigate the direct effects, we imagine a counterfactual world where the user-item matching score is disregarded, and interactions are caused by item popularity and user conformity. We just deduct the prediction score in counterfactual world from the score obtained from the factual world. Lastly, our inference formulation is as follows:

$$\sigma(\hat{p}_i) * \sigma(\hat{c}_u) * f_{\text{Match}}(e'_u, e'_i) - \lambda * p_i * c_u, \tag{3}$$

where λ is a tunable hyperparameter.

Experiments

We conduct experiments on Adressa, MovieLens-1M, Gowalla, and Yelp2018 datasets. To evaluate the debiasing capability of DCR, we construct the unbiased test sets, ensuring that all items in the test sets receive an equal number of interactions. We use LightGCN (He et al. 2020) to implement DCR and all baselines (Zheng et al. 2021; Wei et al. 2021; Luo and Wu 2023; Ning et al. 2024).

From the quantitative results in Table 1, we have the following observations: (1) DCR alleviates both popularity bias and conformity bias, demonstrating superior performance compared to methods like UDIPS, DICE, and PPAC that focus on only one bias. (2) Although MACR focuses on dual biases, it fails to distinguish users' true interests from bias-influenced behaviors, leading to suboptimal performance.

Conclusion

In this paper, we propose a disentangled counterfactual reasoning framework to alleviate both popularity bias and conformity bias. Specifically, we simultaneously mitigate the indirect and direct effects of both item popularity and user conformity. Experimental results on four datasets demonstrate the effectiveness of our proposed method.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (2023JBZY038), and the Innovation Found from the Engineering Research Center of Integration and Application of Digital Learning Technology, Ministry of Education (1321004)

References

He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; and Wang, M. 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.

Luo, F.; and Wu, J. 2023. User-Dependent Learning to Debias for Recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2491–2495.

Ning, W.; Cheng, R.; Yan, X.; Kao, B.; Huo, N.; Haldar, N. A. H.; and Tang, B. 2024. Debiasing Recommendation with Personal Popularity. In *Proceedings of the ACM on Web Conference 2024*, 3400–3409.

Pearl, J. 2009. *Causality*. Cambridge university press.

Wei, T.; Feng, F.; Chen, J.; Wu, Z.; Yi, J.; and He, X. 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.

Zheng, Y.; Gao, C.; Li, X.; He, X.; Li, Y.; and Jin, D. 2021. Disentangling user interest and conformity for recommendation with causal embedding. In *Proceedings of the Web Conference 2021*, 2980–2991.