Causal Intervention for Deconfounding and Leveraging Popularity Bias in the Recommender System

Ziling Gong
zgong@ucsd.edu

Yihan Xue
y6xue@ucsd.edu

Jiawei Wang
jiw076@ucsd.edu

Biwei Huang
bih007@ucsd.edu

Babak Salimi
bsalimi@ucsd.edu

Abstract

In today's digital landscape, recommender systems play a pivotal role in guiding user choices and facilitating content discovery, with the problem of popularity bias lying in the hidden algorithm. It makes the recommender disproportionately favor already popular items while overlooking lesser-known ones. Popularity bias serves as a confounder factor in the relationship between items and users' interaction. As the recommender continues to favor popular items, the popular items are promoted to the users which leads to more possible interaction from the users. Thus, the bias is exacerbated. Existing models either neglect the bias or eliminate the bias effect with propensity-based unbiased learning by Gao et al. (2022). In this paper, we will reproduce the model and aim to replicate the results in the paper Zhang et al. (2021), where we hope to remove the popularity bias to check the legality of the de-confounding approach by comparing the accuracy score with conventional models and alternatively utilize popular bias to improve recommendation accuracy. We implemented Popularity-bias De-confounding (PD) and Popularity-bias De-confounding and Adjusting (PDA) models on the Douban Movie dataset. We conclude that the popularity adjustment model improves 95% in recall from the baseline model (BPRMF) on average, though PDA performs approximately similarly to the PD model in this dataset.

Code: https://github.com/bettygong/DSC180A-proj1-PD.git

1 Introduction

In recent years, the field of machine learning and artificial intelligence has witnessed a profound evolution, marked by a shift in emphasis from merely uncovering correlations within data to delving into the intricate realm of underlying causal relationships. As highlighted in (Spirtes and Zhang 2016), this transformative journey underscores the critical role played by causality in scientific domains, emerging as a cornerstone in the analysis of human decision-making and behaviors. Theoretically, causal inference serves as the compass guiding us to discern the intricate relationships between variables, unraveling how one variable intricately influences another. This paradigm shift towards causal analysis not only elevates model interpretability through the use of graphical representations but also proves indispensable for a more profound understanding of intricate and complex systems.

Among all the fields, we decide to delve into the algorithms embedded within recommendation systems. In today's digital landscape, recommendation systems play a crucial role in enhancing users' access to information across various online platforms—ranging from personalized shopping suggestions to streaming services and social media advertisements—all with the goal of aligning users with potential preferences. These systems typically fall into two categories: collaborative filtering (CF) and content-based recommendation (CTR) (Gao et al. 2022). CF, for instance, extracts user preferences by analyzing historical behaviors, drawing on algorithms that discern similarities between users and items based on past interactions. On the other hand, CTR prediction utilizes high-order features related to users, items, or context, feeding them into neural networks. Despite the sophisticated deep learning techniques employed in these traditional systems, their primary focus lies in deciphering correlations within data, largely overlooking the realm of causality.

This project introduces a novel approach by integrating causality into recommendation systems, aiming to not only enhance the accuracy of personalized suggestions but also improve their interpretability. Causal reasoning, deeply rooted in user decision-making processes, provides a valuable framework for comprehending the intricacies of human behaviors. Thus, by incorporating causality, our project embarks on a more profound exploration, specifically targeting the understanding and leveraging of popularity bias and the underlying factors influencing users' choices. More concretely, we compare the performance of two models: the deconfounding-popularity-bias model (PD) and the adjusting-popularity-bias model (PDA), where we control the strength of popularity drift as a key parameter.

2 Dataset Description

The dataset utilized in our study is sourced from the **Douban** movie dataset, encompassing movie and book ratings collected from a Chinese social networking platform. These datasets encompass essential fields such as user_id, item_id, timestamp, and users' ratings spanning the period from 2005 to 2017. Ratings within this dataset range from 1 to 5, with an empty rating indicating a user's simple interaction with the item, such as clicking. To optimize our model's efficiency, we considered all rating records as positive samples and applied a filter to include data only up to 2010.

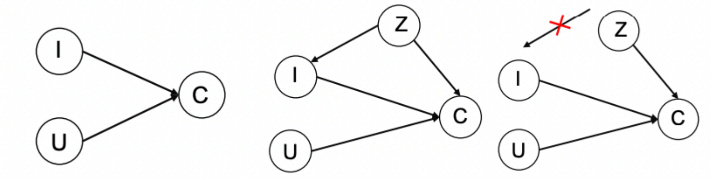
This dataset proves suitable for Machine Factorization (MF), a foundational method in recommender systems due to its incorporation of user-item interactions. MF involves the factorization of the interaction matrix into user and item matrices, both comprising latent vectors containing valuable information. Following the data cleaning phase, our dataset retained 7,174,218 interactions involving 47,890 users and 26,047 items, providing ample data for the effective training of a recommendation system.

3 Methods

3.1 Popularity Bias Causal Graph

Popularity bias, a prevalent issue frequently afflicting Collaborative Filtering (CF) methods, tends to exhibit a disproportionate preference for already popular items while neglecting less-known ones. Illustrated in 3.1(a), conventional recommendation methods consider user information (U node) and item information (I node) to explain the interaction, specifically the click (C node), known as the collider effect. However, in real-world scenarios, a third factor, popularity (Z node), influences both the interaction (C node) and the item information (I node), introducing a confounding element depicted in 3.1(b).

On one hand, individuals often exhibit a conformity mentality, leaning towards popular items in their purchasing decisions (Z -> C). Consequently, the more popular items become, the higher the likelihood of user interaction. On the other hand, recommender models tend to perpetuate biases present in the data, disproportionately highlighting popular items (Z -> I -> C), thus exacerbating the popularity bias (Zhang et al. 2021). The popularity factor in the second path serves as a bias amplification, becoming the target of the removal process.



(a) Causal graph of traditional methods.

(b) Causal graph that considers item popularity.

(c) We cut off $Z \rightarrow I$ for model training.

Figure 1: Causal graphs to describe the recommendation process. U: user, I: exposed item, C: interaction probability, Z: item popularity. We identify Z as the confounder between I and C, and propose deconfounded training with $P(C|do(U,I))$ as the interest matching.

To eliminate the Z -> I path, we employ do-calculus, as advocated by (Zhang et al. 2021), intervening in the recommended item I to render it impervious to the effects of popularity Z. This approach, rooted in causal inference, serves as a remedy, effectively debiasing the data and mitigating the impact of popularity. Consequently, recommendation systems can furnish more equitable and unbiased suggestions to users.

3.2 Derive Predictive Model P(C|do(U, I))

Denote G as Figure 3.1 (b) and G' be figure 3.1(c).
P(C|do(U, I)) means how likely the user clicks (C) the item by fixing variables in nodes U and I. The reason for do(U, I) instead of do(I) is that the recommender system takes both U and I as input, and without losing generality, the backdoor path I <- Z ->C in G is blocked by do(U, I).

$$\begin{aligned} P(C \mid do(U, I)) &\overset{(1)}{=} P_{G'}(C \mid U, I) \\ &\overset{(2)}{=} \sum_{z} P_{G'}(C \mid U, I, z) P_{G'}(z \mid U, I) \\ &\overset{(3)}{=} \sum_{z} P_{G'}(C \mid U, I, z) P_{G'}(z) \\ &\overset{(4)}{=} \sum_{\sim} P(C \mid U, I, z) P(z), \end{aligned}$$

3.2.1 Step 1: Estimating P(C|do(U, I))

Let the parameters of the conditional probability function be θ, and we parameterize U=u, I=i, and popularity $Z = m_i^t$. We use L2 to optimize the pairwise BPR objective function on historical data D.

$$\max_{\Theta} \sum_{(u,i,j) \in \mathscr{D}} \log \sigma \left(P_{\Theta} \left(c = 1 \mid u, i, m_i^t \right) - P_{\Theta} \left(c = 1 \mid u, j, m_j^t \right) \right) \tag{1}$$

Here m_i^t represents the local popularity of item i on the stage t, $m_i^t = D_t^i / \sum D_j^t$, where we divided data D into T stages and D_i^t is the number of observed interactions for item i at stage t. According to (Zhang et al. 2021), the usage of local popularity is because the most recent data has a larger impact on the system's exposure mechanism.
Subsequently, we parametrize $P_\theta(c = 1 | u, i, m_i^t) = \mathrm{ELU}'(f_\theta(u, i)) \cdot (m_i^t)^\gamma$.
The function $f_\theta(u, i)$ represents any user-item matching model, with Machine Factorization (MF) being our choice in this context. The parameter γ governs the intensity of the conformity effect; for instance, $\gamma = 0$ characterizes our PD model, which is devoid of popularity bias. In contrast, in the PDA model, a higher γ corresponds to a more pronounced impact.

In addition, the activation function ELU' is employed to guarantee positivity.

$$ELU'(x) = \begin{cases} e^{x}, & \text{if } x \leq 0 \\ x+1, & \text{else} \end{cases} \tag{2}$$

3.2.2 Step 2: Estimating $\sum P(C|U,I,z)P(z)$

$$\begin{aligned} P(C \mid do(U,I)) &= \sum_{z} P(C \mid U,I,z)P(z) \\ &= \sum_{z} ELU'\left(f_{\Theta}(u,i)\right) \times z^{\gamma} P(z) \\ &= ELU'\left(f_{\Theta}(u,i)\right) \sum_{z} z^{\gamma} P(z) \\ &= ELU'\left(f_{\Theta}(u,i)\right) E\left(Z^{\gamma}\right) \end{aligned}$$

We replace P(C | U, I, z) by the equation in step 1, and take ELU' out of summation. $\sum_{z} z^{\gamma} * P(z)$ is expectation equation of Z^{γ}. Since the expectation value is constant, we neglect the term and use $ELU(f_{\theta}(u,i))$ to estimate $P(C|do(U,I))$.

Following is pseudocode for PD and PDA, where PDA is different from PD by assigning different values of γ to the equation.

```
Algorithm 1: PD/PDA
Input: dataset $\mathscr{D} = \{(u,i,m_i^t)\}$; hyper-parameter $\gamma$; predicted popularity $\{\tilde{m}_i\}$;
       mode: PD or PDA
while stop condition is not reached do
    Update model parameters $\Theta$ by optimizing Equation (4);
    if $mode == PD$ then
        validate model with $ELU'(f_{\Theta}(u,i))$;
    else
        validate model with Equation (9) (simplify $\tilde{\gamma} = \gamma$ );
if $mode == PD$ then
    recommend items using $ELU'(f_{\Theta}(u,i))$;
else
    recommend items using Equation (9) (simplify $\tilde{\gamma} = \gamma$ );
```

4 Results

Table 1 displays the model performance for baseline model ($BPRMF$), PD, $PDA-D_i$ and $PDA-D_i^t$ by taking top-20 recommendations. Table 2 displays the model performance by taking top-50 recommendations, where

$BPRMF$ is the baseline model which optimizes the MF model with BPR loss.

PD is the popularity-bias deconfounding model. ($\gamma = 0$)

$PDA-D_i$ is the popularity-bias deconfounding and Adjusting model with last stage popularity D_i.

$PDA-D_i^t$ is the popularity-bias deconfounding and Adjusting model with linearly predicted local popularity D_i^t.

The highest rate is bolden in each column.

Table 1: Recommendation performance taking top-20 recommendations

Method	Recall	Precision	HR	NDCG
BPRMF	0.0274	0.0336	0.2888	0.0405
PD	0.0455	0.0454	0.3970	0.0607
$PDA-D_i$	0.0564	0.0557	**0.448**	**0.0746**
$PDA-D_i^t$	**0.0565**	**0.0557**	**0.448**	0.0745

Table 2: Recommendation performance taking top-50 recommendations

Method	Recall	Precision	HR	NDCG
BPRMF	0.0581	0.0291	0.4280	0.0475
PD	0.0843	0.0362	0.5271	0.0686
$PDA-D_i$	**0.1066**	**0.0437**	0.582	**0.0845**
$PDA-D_i^t$	**0.1066**	0.0436	**0.583**	0.0844

5 Conclusion

We have successfully replicated the model performance by rerunning the saved model from Zhang et al. (2021). The evaluation metrics in both top-k tables indicate that PD's performance surpasses that of the baseline model, and PDA exhibits improvement over PD. Taking recall rate as an example, PD demonstrates an average improvement of 55.2% over BPRMF, while PDA shows an average improvement of 25.3% over PD. This observation aligns with the assertion that eliminating popularity bias enhances recommendation accuracy. Moreover, the superior performance of PDA models compared to PD models suggests that leveraging popularity bias is beneficial and prevents the learning model from amplifying the popularity effect. Ultimately, the recall rate of the best PDA models increases to 95% compared to the baseline model.

Now, let's delve into the nuanced difference between $PDA-D_i$ and $PDA-D_i^t$ performance on **Douban**. As mentioned in Zhang et al. (2021), $PDA-D_i^t$, which incorporates local popularity, is theoretically expected to reflect the conformity effect more than the $PDA-D_i$ model and should perform better. The work indeed demonstrated that $PDA-D_i^t$ outperformed $PDA-D_i$ in the other two datasets in the original paper. However, in **Douban**, we observe approximately similar performance. From the original work, it is speculated that this could be attributed to **Douban** spanning over a more extended period than the other two datasets, making it more challenging to predict popularity trends accurately during the training stages, leaving potential improvement for further studies.

References

Gao, Chen, Yu Zheng, Wenjie Wang, Fuli Feng, Xiangnan He, and Yong Li. 2022. “Causal inference in recommender systems: A survey and future directions.” *arXiv preprint arXiv:2208.12397*

Spirtes, Peter, and Kun Zhang. 2016. “Causal discovery and inference: concepts and recent methodological advances.” In *Applied informatics*. SpringerOpen

Zhang, Yang, 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*.