Addressing Popularity Bias with Popularity-Conscious Alignment and Contrastive Learning

Abstract

Collaborative Filtering (CF) often encounters substantial difficulties with popularity bias because of the skewed distribution of items in real-world datasets. This tendency creates a notable difference in accuracy between items that are popular and those that are not. This discrepancy impedes the accurate comprehension of user preferences and intensifies the Matthew effect within recommendation systems. To counter popularity bias, current methods concentrate on highlighting less popular items or on differentiating the correlation between item representations and their popularity. Despite their effectiveness, current approaches continue to grapple with two significant issues: firstly, the extraction of shared supervisory signals from popular items to enhance the representations of less popular items, and secondly, the reduction of representation separation caused by popularity bias. In this study, we present an empirical examination of popularity bias and introduce a method called Popularity-Aware Alignment and Contrast (PAAC) to tackle these two problems. Specifically, we utilize the common supervisory signals found in popular item representations and introduce an innovative popularity-aware supervised alignment module to improve the learning of representations for unpopular items. Furthermore, we propose adjusting the weights in the contrastive learning loss to decrease the separation of representations by focusing on popularity. We confirm the efficacy and logic of PAAC in reducing popularity bias through thorough experiments on three real-world datasets.

1 Introduction

Contemporary recommender systems are essential in reducing information overload. Personalized recommendations frequently employ collaborative filtering (CF) to assist users in discovering items that may interest them. CF-based techniques primarily learn user preferences and item attributes by matching the representations of users with the items they engage with. Despite their achievements, CF-based methods frequently encounter the issue of popularity bias, which leads to considerable disparities in accuracy between items that are popular and those that are not. Popularity bias occurs because there are limited supervisory signals for items that are not popular, which results in overfitting during the training phase and decreased effectiveness on the test set. This hinders the precise comprehension of user preferences, thereby diminishing the variety of recommendations. Furthermore, popularity bias can worsen the Matthew effect, where items that are already popular gain even more popularity because they are recommended more frequently.

Two significant challenges are presented when mitigating popularity bias in recommendation systems. The first challenge is the inadequate representation of unpopular items during training, which results in overfitting and limited generalization ability. The second challenge, known as representation separation, happens when popular and unpopular items are categorized into distinct semantic spaces, thereby intensifying the bias and diminishing the precision of recommendations.

2 Methodology

To overcome the current difficulties in reducing popularity bias, we introduce the Popularity-Aware Alignment and Contrast (PAAC) method. We utilize the common supervisory signals present in popular item representations to direct the learning of unpopular representations, and we present a popularity-aware supervised alignment module. Moreover, we incorporate a re-weighting system in the contrastive learning module to deal with representation separation by considering popularity.

2.1 Supervised Alignment Module

During the training process, the alignment of representations usually emphasizes users and items that have interacted, often causing items to be closer to interacted users than non-interacted ones in the representation space. However, because unpopular items have limited interactions, they are usually modeled based on a small group of users. This limited focus can result in overfitting, as the representations of unpopular items might not fully capture their features.

The disparity in the quantity of supervisory signals is essential for learning representations of both popular and unpopular items. Specifically, popular items gain from a wealth of supervisory signals during the alignment process, which helps in effectively learning their representations. On the other hand, unpopular items, which have a limited number of users providing supervision, are more susceptible to overfitting. This is because there is insufficient representation learning for unpopular items, emphasizing the effect of supervisory signal distribution on the quality of representation. Intuitively, items interacted with by the same user have some similar characteristics. In this section, we utilize common supervisory signals in popular item representations and suggest a popularity-aware supervised alignment method to improve the representations of unpopular items.

We initially filter items with similar characteristics based on the user's interests. For any user, we define the set of items they interact with. We count the frequency of each item appearing in the training dataset as its popularity. Subsequently, we group items based on their relative popularity. We divide items into two groups: the popular item group and the unpopular item group. The popularity of each item in the popular group is higher than that of any item in the unpopular group. This indicates that popular items receive more supervisory information than unpopular items, resulting in poorer recommendation performance for unpopular items.

To tackle the issue of insufficient representation learning for unpopular items, we utilize the concept that items interacted with by the same user share some similar characteristics. Specifically, we use similar supervisory signals in popular item representations to improve the representations of unpopular items. We align the representations of items to provide more supervisory information to unpopular items and improve their representation, as follows:

$$L_{SA} = \sum_{u \in U} \frac{1}{|I_u|} \sum_{i \in I^u_{pop}, j \in I^u_{unpop}} ||f(i) - f(j)||_2, \tag{1}$$

where $f(\cdot)$ is a recommendation encoder and $h_i = f(i)$. By efficiently using the inherent information in the data, we provide more supervisory signals for unpopular items without introducing additional side information. This module enhances the representation of unpopular items, mitigating the overfitting issue.

2.2 Re-weighting Contrast Module

Recent research has indicated that popularity bias frequently leads to a noticeable separation in the representation of item embeddings. Although methods based on contrastive learning aim to enhance overall uniformity by distancing negative samples, their current sampling methods might unintentionally worsen this separation. When negative samples follow the popularity distribution, which is dominated by popular items, prioritizing unpopular items as positive samples widens the gap between popular and unpopular items in the representation space. Conversely, when negative samples follow a uniform distribution, focusing on popular items separates them from most unpopular ones, thus worsening the representation gap. Existing studies use the same weights for positive and negative samples in the contrastive loss function, without considering differences in item popularity. However, in real-world recommendation datasets, the impact of items varies due to dataset characteristics and interaction distributions. Neglecting this aspect could lead to suboptimal results and exacerbate representation separation.

We propose to identify different influences by re-weighting different popularity items. To this end, we introduce re-weighting different positive and negative samples to mitigate representation separation from a popularity-centric perspective. We incorporate this approach into contrastive learning to better optimize the consistency of representations. Specifically, we aim to reduce the risk of pushing items with varying popularity further apart. For example, when using a popular item as a positive sample, our goal is to avoid pushing unpopular items too far away. Thus, we introduce two hyperparameters to control the weights when items are considered positive and negative samples.

To ensure balanced and equitable representations of items within our model, we first propose a dynamic strategy to categorize items into popular and unpopular groups for each mini-batch. Instead of relying on a fixed global threshold, which often leads to the overrepresentation of popular items across various batches, we implement a hyperparameter x. This hyperparameter readjusts the classification of items within the current batch. By adjusting the hyperparameter x, we maintain a balance between different item popularity levels. This enhances the model's ability to generalize across diverse item sets by accurately reflecting the popularity distribution in the current training context. Specifically, we denote the set of items within each batch as I_B. And then we divide I_B into a popular group I_{pop} and an unpopular group I_{unpop} based on their respective popularity levels, classifying the top $x\%$ of items as I_{pop}:

$$I_B = I_{pop} \cup I_{unpop}, \forall i \in I_{pop} \land j \in I_{unpop}, p(i) > p(j), \tag{2}$$

where $I_{pop} \in I_B$ and $I_{unpop} \in I_B$ are disjoint, with I_{pop} consisting of the top $x\%$ of items in the batch. In this work, we dynamically divided items into popular and unpopular groups within each mini-batch based on their popularity, assigning the top 50% as popular items and the bottom 50% as unpopular items. This radio not only ensures equal representation of both groups in our contrastive learning but also allows items to be classified adaptively based on the batch's current composition.

After that, we use InfoNCE to optimize the uniformity of item representations. Unlike traditional CL-based methods, we calculate the loss for different item groups. Specifically, we introduce the hyperparameter α to control the positive sample weights between popular and unpopular items, adapting to varying item distributions in different datasets:

$$L_{item}^{CL} = \alpha \times L_{pop}^{CL} + (1-\alpha) \times L_{unpop}^{CL}, \tag{3}$$

where L_{pop}^{CL} represents the contrastive loss when popular items are considered as positive samples, and L_{unpop}^{CL} represents the contrastive loss when unpopular items are considered as positive samples. The value of α ranges from 0 to 1, where $\alpha = 0$ means exclusive emphasis on the loss of unpopular items L_{unpop}^{CL}, and $\alpha = 1$ means exclusive emphasis on the loss of popular items L_{pop}^{CL}. By adjusting α, we can effectively balance the impact of positive samples from both popular and unpopular items, allowing adaptability to varying item distributions in different datasets.

Following this, we fine-tune the weighting of negative samples in the contrastive learning framework using the hyperparameter β. This parameter controls how samples from different popularity groups contribute as negative samples. Specifically, we prioritize re-weighting items with popularity opposite to the positive samples, mitigating the risk of excessively pushing negative samples away and reducing representation separation. Simultaneously, this approach ensures the optimization of intra-group consistency. For instance, when dealing with popular items as positive samples, we separately calculate the impact of popular and unpopular items as negative samples. The hyperparameter β is then used to control the degree to which unpopular items are pushed away. This is formalized as follows:

$$L_{pop}^{'} = \sum_{i \in I_{pop}} \log \frac{\exp(h_i^{'} h_i / \tau)}{\sum_{j \in I_{pop}} \exp(h_i^{'} h_j / \tau) + \beta \sum_{j \in I_{unpop}} \exp(h_i^{'} h_j / \tau)}, \tag{4}$$

similarly, the contrastive loss for unpopular items is defined as:

$$L_{unpop}^{'} = \sum_{i \in I_{unpop}} \log \frac{\exp(h_i^{'} h_i / \tau)}{\sum_{j \in I_{unpop}} \exp(h_i^{'} h_j / \tau) + \beta \sum_{j \in I_{pop}} \exp(h_i^{'} h_j / \tau)}, \tag{5}$$

where the parameter β ranges from 0 to 1, controlling the negative sample weighting in the contrastive loss. When $\beta = 0$, it means that only intra-group uniformity optimization is performed. Conversely, when $\beta = 1$, it means equal treatment of both popular and unpopular items in terms of their impact on positive samples. The setting of β allows for a flexible adjustment between prioritizing intra-group uniformity and considering the impact of different popularity levels in the training. We prefer to push away items within the same group to optimize uniformity. This setup helps prevent over-optimizing the uniformity of different groups, thereby mitigating representation separation.

The final re-weighting contrastive objective is the weighted sum of the user objective and the item objective:

$$L_{CL} = \frac{1}{2} \times (L_{item}^{CL} + L_{user}^{CL}). \tag{6}$$

In this way, we not only achieved consistency in representation but also reduced the risk of further separating items with similar characteristics into different representation spaces, thereby alleviating the issue of representation separation caused by popularity bias.

2.3 Model Optimization

To reduce popularity bias in collaborative filtering tasks, we employ a multi-task training strategy to jointly optimize the classic recommendation loss (L_{REC}), supervised alignment loss (L_{SA}), and re-weighting contrast loss (L_{CL}).

$$L = L_{REC} + \lambda_1 L_{SA} + \lambda_2 L_{CL} + \lambda_3 ||\Theta||^2, \tag{7}$$

where Θ is the set of model parameters in L_{REC} as we do not introduce additional parameters, λ_1 and λ_2 are hyperparameters that control the strengths of the popularity-aware supervised alignment loss and the re-weighting contrastive learning loss respectively, and λ_3 is the L_2 regularization coefficient. After completing the model training process, we use the dot product to predict unknown preferences for recommendations.

3 Experiments

In this section, we assess the efficacy of PAAC through comprehensive experiments, aiming to address the following research questions:

- How does PAAC compare to existing debiasing methods?
- How do different designed components play roles in our proposed PAAC?

- How does PAAC alleviate the popularity bias?
- How do different hyper-parameters affect the PAAC recommendation performance?

3.1 Experiments Settings

3.1.1 Datasets

In our experiments, we use three widely public datasets: Amazon-book, Yelp2018, and Gowalla. We retained users and items with a minimum of 10 interactions.

3.1.2 Baselines and Evaluation Metrics

We implement the state-of-the-art LightGCN to instantiate PAAC, aiming to investigate how it alleviates popularity bias. We compare PAAC with several debiased baselines, including re-weighting-based models, decorrelation-based models, and contrastive learning-based models.

We utilize three widely used metrics, namely Recall@K, HR@K, and NDCG@K, to evaluate the performance of Top-K recommendation. Recall@K and HR@K assess the number of target items retrieved in the recommendation results, emphasizing coverage. In contrast, NDCG@K evaluates the positions of target items in the ranking list, with a focus on their positions in the list. We use the full ranking strategy, considering all non-interacted items as candidate items to avoid selection bias during the test stage. We repeated each experiment five times with different random seeds and reported the average scores.

3.2 Overall Performance

As shown in Table 1, we compare our model with several baselines across three datasets. The best performance for each metric is highlighted in bold, while the second best is underlined. Our model consistently outperforms all compared methods across all metrics in every dataset.

- Our proposed model PAAC consistently outperforms all baselines and significantly mitigates the popularity bias. Specifically, PAAC enhances LightGCN, achieving improvements of 282.65%, 180.79%, and 82.89% in NDCG@20 on the Yelp2018, Gowalla, and Amazon-Book datasets, respectively. Compared to the strongest baselines, PAAC delivers better performance. The most significant improvements are observed on Yelp2018, where our model achieves an 8.70% increase in Recall@20, a 10.81% increase in HR@20, and a 30.2% increase in NDCG@20. This improvement can be attributed to our use of popularity-aware supervised alignment to enhance the representation of less popular items and re-weighted contrastive learning to address representation separation from a popularity-centric perspective.
- The performance improvements of PAAC are smaller on sparser datasets. For example, on the Gowalla dataset, the improvements in Recall@20, HR@20, and NDCG@20 are 3.18%, 5.85%, and 5.47%, respectively. This may be because, in sparser datasets like Gowalla, even popular items are not well-represented due to lower data density. Aligning unpopular items with these poorly represented popular items can introduce noise into the model. Therefore, the benefits of using supervisory signals for unpopular items may be reduced in very sparse environments, leading to smaller performance improvements.
- Regarding the baselines for mitigating popularity bias, the improvement of some is relatively limited compared to the backbone model (LightGCN) and even performs worse in some cases. This may be because some are specifically designed for traditional data-splitting scenarios, where the test set still follows a long-tail distribution, leading to poor generalization. Some mitigate popularity bias by excluding item popularity information. Others use invariant learning to remove popularity information at the representation level, generally performing better than the formers. This shows the importance of addressing popularity bias at the representation level. Some outperform the other baselines, emphasizing the necessary to improve item representation consistency for mitigating popularity bias.
- Different metrics across various datasets show varying improvements in model performance. This suggests that different debiasing methods may need distinct optimization strategies for models. Additionally, we observe varying effects of PAAC across different datasets. This difference could be due to the sparser nature of the Gowalla dataset. Conversely, our model can directly provide supervisory signals for unpopular items and conduct intra-group optimization, consistently maintaining optimal performance across all metrics on the three datasets.

3.3 Ablation Study

To better understand the effectiveness of each component in PAAC, we conduct ablation studies on three datasets. Table 2 presents a comparison between PAAC and its variants on recommendation performance. Specifically, PAAC-w/o P refers to the variant where the re-weighting contrastive loss of popular items is removed, focusing instead on optimizing the consistency of representations for unpopular items. Similarly, PAAC-w/o U denotes the removal of the re-weighting contrastive loss for unpopular items. PAAC-w/o A refers to the variant without the popularity-aware supervised alignment loss. It's worth noting that PAAC-w/o A differs from

Table 1: Performance comparison on three public datasets with K = 20. The best performance is indicated in bold, while the second-best performance is underlined. The superscripts * indicate $p \leq 0.05$ for the paired t-test of PAAC vs. the best baseline (the relative improvements are denoted as Imp.).

Model	Yelp2018			Gowalla			Amazon-book		
	Recall@20	HR@20	NDCG@20	Recall@20	HR@20	NDCG@20	Recall@20	HR@20	NDCG@20
MF	0.0050	0.0109	0.0093	0.0343	0.0422	0.0280	0.0370	0.0388	0.0270
LightGCN	0.0048	0.0111	0.0098	0.0380	0.0468	0.0302	0.0421	0.0439	0.0304
IPS	0.0104	0.0183	0.0158	0.0562	0.0670	0.0444	0.0488	0.0510	0.0365
!MACR	0.0402	0.0312	0.0265	0.0908	0.1086	0.0600	0.0515	0.0609	0.0487
α-Adjnorm	0.0053	0.0088	0.0080	0.0328	0.0409	0.0267	0.0422	0.0450	0.0264
InvCF	0.0444	0.0344	0.0291	0.1001	0.1202	0.0662	0.0562	0.0665	0.0515
Adap-τ	0.0450	0.0497	0.0341	0.1182	0.1248	0.0794	0.0641	0.0678	0.0511
SimGCL	0.0449	0.0518	0.0345	0.1194	0.1228	0.0804	0.0628	0.0648	0.0525
PAAC	**0.0494***	**0.0574***	**0.0375***	**0.1232***	**0.1321***	**0.0848***	**0.0701***	**0.0724***	**0.0556***
Imp.	+9.78 %	+10.81%	+8.70%	+3.18%	+5.85%	+5.47%	+9.36%	+6.78%	5.90%

SimGCL in that we split the contrastive loss on the item side, L^{CL}_{item}, into two distinct losses: L^{CL}_{pop} and L^{CL}_{unpop}. This approach allows us to separately address the consistency of popular and unpopular item representations, thereby providing a more detailed analysis of the impact of each component on the overall performance.

From Table 2, we observe that PAAC-w/o A outperforms SimGCL in most cases. This validates that re-weighting the importance of popular and unpopular items can effectively improve the model's performance in alleviating popularity bias. It also demonstrates the effectiveness of using supervision signals from popular items to enhance the representations of unpopular items, providing more opportunities for future research on mitigating popularity bias. Moreover, compared with PAAC-w/o U, PAAC-w/o P results in much worse performance. This confirms the importance of re-weighting popular items in contrastive learning for mitigating popularity bias. Finally, PAAC consistently outperforms the three variants, demonstrating the effectiveness of combining supervised alignment and re-weighting contrastive learning. Based on the above analysis, we conclude that leveraging supervisory signals from popular item representations can better optimize representations for unpopular items, and re-weighting contrastive learning allows the model to focus on more informative or critical samples, thereby improving overall performance. All the proposed modules significantly contribute to alleviating popularity bias.

Table 2: Ablation study of PAAC, highlighting the best-performing model on each dataset and metrics in bold. Specifically, PAAC-w/o P removes the re-weighting contrastive loss of popular items, PAAC-w/o U eliminates the re-weighting contrastive loss of unpopular items, and PAAC-w/o A omits the popularity-aware supervised alignment loss.

Model	Yelp2018			Gowalla			Amazon-book		
	Recall@20	HR@20	NDCG@20	Recall@20	HR@20	NDCG@20	Recall@20	HR@20	NDCG@20
SimGCL	0.0449	0.0518	0.0345	0.1194	0.1228	0.0804	0.0628	0.0648	0.0525
PAAC-w/o P	0.0443	0.0536	0.0340	0.1098	0.1191	0.0750	0.0616	0.0639	0.0458
PAAC-w/o U	0.0462	0.0545	0.0358	0.1120	0.1179	0.0752	0.0594	0.0617	0.0464
PAAC-w/o A	0.0466	0.0547	0.0360	0.1195	0.1260	0.0815	0.0687	0.0711	0.0536
PAAC	**0.0494***	**0.0574***	**0.0375***	**0.1232***	**0.1321***	**0.0848***	**0.0701***	**0.0724***	**0.0556***

3.4 Debias Ability

To further verify the effectiveness of PAAC in alleviating popularity bias, we conduct a comprehensive analysis focusing on the recommendation performance across different popularity item groups. Specifically, 20% of the most popular items are labeled 'Popular', and the rest are labeled 'Unpopular'. We compare the performance of PAAC with LightGCN, IPS, MACR, and SimGCL using the NDCG@20 metric across different popularity groups. We use Δ to denote the accuracy gap between the two groups. We draw the following conclusions:

- Improving the performance of unpopular items is crucial for enhancing overall model performance. Specially, on the Yelp2018 dataset, PAAC shows reduced accuracy in recommending popular items, with a notable decrease of 20.14% compared to SimGCL. However, despite this decrease, the overall recommendation accuracy surpasses that of SimGCL by 11.94%, primarily due to a 6.81% improvement in recommending unpopular items. This improvement highlights the importance of better recommendations for unpopular items and emphasizes their crucial role in enhancing overall model performance.

- Our proposed PAAC significantly enhances the recommendation performance for unpopular items. Specifically, we observe an improvement of 8.94% and 7.30% in NDCG@20 relative to SimGCL on the Gowalla and Yelp2018 datasets, respectively. This improvement is due to the popularity-aware alignment method, which uses supervisory signals from popular items to improve the representations of unpopular items.

- PAAC has successfully narrowed the accuracy gap between different item groups. Specifically, PAAC achieved the smallest gap, reducing the NDCG@20 accuracy gap by 34.18% and 87.50% on the Gowalla and Yelp2018 datasets, respectively. This indicates that our method treats items from different groups fairly, effectively alleviating the impact of popularity bias. This success can be attributed to our re-weighted contrast module, which addresses representation separation from a popularity-centric perspective, resulting in more consistent recommendation results across different groups.

3.5 Hyperparameter Sensitivities

In this section, we analyze the impact of hyperparameters in PAAC. Firstly, we investigate the influence of λ_1 and λ_2, which respectively control the impact of the popularity-aware supervised alignment and re-weighting contrast loss. Additionally, in the re-weighting contrastive loss, we introduce two hyperparameters, α and β, to control the re-weighting of different popularity items as positive and negative samples. Finally, we explore the impact of the grouping ratio x on the model's performance.

3.5.1 Effect of λ_1 and λ_2

As formulated in Eq. (11), λ_1 controls the extent of providing additional supervisory signals for unpopular items, while λ_2 controls the extent of optimizing representation consistency. Horizontally, with the increase in λ_2, the performance initially increases and then decreases. This indicates that appropriate re-weighting contrastive loss effectively enhances the consistency of representation distributions, mitigating popularity bias. However, overly strong contrastive loss may lead the model to neglect recommendation accuracy. Vertically, as λ_1 increases, the performance also initially increases and then decreases. This suggests that suitable alignment can provide beneficial supervisory signals for unpopular items, while too strong an alignment may introduce more noise from popular items to unpopular ones, thereby impacting recommendation performance.

3.5.2 Effect of re-weighting coefficient α and β

To mitigate representation separation due to imbalanced positive and negative sampling, we introduce two hyperparameters into the contrastive loss. Specifically, α controls the weight difference between positive samples from popular and unpopular items, while β controls the influence of different popularity items as negative samples.

In our experiments, while keeping other hyperparameters constant, we search α and β within the range {0, 0.2, 0.4, 0.6, 0.8, 1}. As α and β increase, performance initially improves and then declines. The optimal hyperparameters for the Yelp2018 and Gowalla datasets are $\alpha = 0.8$, $\beta = 0.6$ and $\alpha = 0.2$, $\beta = 0.2$, respectively. This may be attributed to the characteristics of the datasets. The Yelp2018 dataset, with a higher average interaction frequency per item, benefits more from a higher weight α for popular items as positive samples. Conversely, the Gowalla dataset, being relatively sparse, prefers a smaller α. This indicates the importance of considering dataset characteristics when adjusting the contributions of popular and unpopular items to the model.

Notably, α and β are not highly sensitive within the range [0, 1], performing well across a broad spectrum. Performance exceeds the baseline regardless of β values when other parameters are optimal. Additionally, α values from [0.4, 1.0] on the Yelp2018 dataset and [0.2, 0.8] on the Gowalla dataset surpass the baseline, indicating less need for precise tuning. Thus, α and β achieve optimal performance without meticulous adjustments, focusing on weight coefficients to maintain model efficacy.

3.5.3 Effect of grouping ratio x

To investigate the impact of different grouping ratios on recommendation performance, we developed a flexible classification method for items within each mini-batch based on their popularity. Instead of adopting a fixed global threshold, which tends to overrepresent popular items in some mini-batches, our approach dynamically divides items in each mini-batch into popular and unpopular categories. Specifically, the top $x\%$ of items are classified as popular and the remaining (100 - x)% as unpopular, with x varying. This strategy prevents the overrepresentation typical in fixed distribution models, which could skew the learning process and degrade performance. To quantify the effects of these varying ratios, we examined various division ratios for popular items, including 20%, 40%, 60%, and 80%, as shown in Table 3. The preliminary results indicate that both extremely low and high ratios negatively affect model performance, thereby underscoring the superiority of our dynamic data partitioning approach. Moreover, within the 40%-60% range, our model's performance remained consistently robust, further validating the effectiveness of PAAC.

Table 3: Performance comparison across varying popular item ratios x on metrics.

Ratio	Yelp2018			Gowalla		
	Recall@20	HR@20	NDCG@20	Recall@20	HR@20	NDCG@20
20%	0.0467	0.0555	0.0361	0.1232	0.1319	0.0845
40%	0.0505	0.0581	0.0378	0.1239	0.1325	0.0848
50%	0.0494	0.0574	0.0375	0.1232	0.1321	0.0848
60%	0.0492	0.0569	0.0370	0.1225	0.1314	0.0843
80%	0.0467	0.0545	0.0350	0.1176	0.1270	0.0818

4 Related Work

4.1 Popularity Bias in Recommendation

Popularity bias is a prevalent problem in recommender systems, where unpopular items in the training dataset are seldom recommended. Numerous techniques have been suggested to examine and decrease performance variations between popular and unpopular items. These techniques can be broadly divided into three categories.

- Re-weighting-based methods aim to increase the training weight or scores for unpopular items, redirecting focus away from popular items during training or prediction. For instance, IPS adds compensation to unpopular items and adjusts the prediction of the user-item preference matrix, resulting in higher preference scores and improving rankings for unpopular items. α-AdjNorm enhances the focus on unpopular items by controlling the normalization strength during the neighborhood aggregation process in GCN-based models.
- Decorrelation-based methods aim to effectively remove the correlations between item representations (or prediction scores) and popularity. For instance, MACR uses counterfactual reasoning to eliminate the direct impact of popularity on item outcomes. In contrast, InvCF operates on the principle that item representations remain invariant to changes in popularity semantics, filtering out unstable or outdated popularity characteristics to learn unbiased representations.
- Contrastive-learning-based methods aim to achieve overall uniformity in item representations using InfoNCE, preserving more inherent characteristics of items to mitigate popularity bias. This approach has been demonstrated as a state-of-the-art method for alleviating popularity bias. It employs data augmentation techniques such as graph augmentation or feature augmentation to generate different views, maximizing positive pair consistency and minimizing negative pair consistency to promote more uniform representations. Specifically, Adap-τ adjusts user/item embeddings to specific values, while SimGCL integrates InfoNCE loss to enhance representation uniformity and alleviate popularity bias.

4.2 Representation Learning for CF

Representation learning is crucial in recommendation systems, especially in modern collaborative filtering (CF) techniques. It creates personalized embeddings that capture user preferences and item characteristics. The quality of these representations critically determines a recommender system's effectiveness by precisely capturing the interplay between user interests and item features. Recent studies emphasize two fundamental principles in representation learning: alignment and uniformity. The alignment principle ensures that embeddings of similar or related items (or users) are closely clustered together, improving the system's ability to recommend items that align with a user's interests. This principle is crucial when accurately reflecting user preferences through corresponding item characteristics. Conversely, the uniformity principle ensures a balanced distribution of all embeddings across the representation space. This approach prevents the over-concentration of embeddings in specific areas, enhancing recommendation diversity and improving generalization to unseen data.

In this work, we focus on aligning the representations of popular and unpopular items interacted with by the same user and re-weighting uniformity to mitigate representation separation. Our model PAAC uniquely addresses popularity bias by combining group alignment and contrastive learning, a first in the field. Unlike previous works that align positive user-item pairs or contrastive pairs, PAAC directly aligns popular and unpopular items, leveraging the rich information of popular items to enhance the representations of unpopular items and reduce overfitting. Additionally, we introduce targeted re-weighting from a popularity-centric perspective to achieve a more balanced representation.

5 Conclusion

In this study, we have examined popularity bias and put forward PAAC as a method to lessen its impact. We postulated that items engaged with by the same user exhibit common traits, and we utilized this insight to coordinate the representations of both popular and unpopular items via a popularity-conscious supervised alignment method. This strategy furnished additional supervisory data for less popular items. It is important to note that our concept of aligning and categorizing items according to user-specific preferences introduces a fresh perspective on alignment. Moreover, we tackled the problem of representation separation seen in current CL-based models by incorporating two hyperparameters to regulate the influence of items with varying popularity levels when considered as positive and negative samples. This method refined the uniformity of representations and successfully reduced separation. We validated our method, PAAC, on three publicly available datasets, demonstrating its effectiveness and underlying rationale.

In the future, we will explore deeper alignment and contrast adjustments tailored to specific tasks to further mitigate popularity bias. We aim to investigate the synergies between alignment and contrast and extend our approach to address other biases in recommendation systems.

Acknowledgments

This work was supported in part by grants from the National Key Research and Development Program of China, the National Natural Science Foundation of China, the Fundamental Research Funds for the Central Universities, and Quan Cheng Laboratory.