Addressing Bias in Recommendation Systems: A Debiased Deep Cross Network Approach

Mustapha N*
CSIR NGEI
Pretoria, South Africa
nmustapha@csir.co.za

Ndlovu N
CSIR NGEI
Pretoria, South Africa
nndlovu2@csir.co.za

Ngcobo S
CSIR NGEI
Pretoria, South Africa
sngcobo2@csir.co.za

Khoza N
CSIR NGEI
Pretoria, South Africa
nkhoza1@csir.co.za

July 27, 2023

Abstract

With the increasing prevalence of personalization of recommendation systems in various domains[2], this research paper examines the presence of bias associated with such systems. The study focuses on the detection and mitigation of bias through the development of a baseline hybrid model that predicts ratings and generates user recommendations. Bias detection is achieved by analyzing disparities in the model's performance across different population subgroups, while the calculation of the eigenvector centrality of nodes aids in identifying influential nodes within the recommendation system's network. To mitigate bias, a regularization technique is employed,adjusting the impact of user ratings based on movie popularity. The effectiveness of the regularization technique is demonstrated by the low root mean squared error (RMSE) scores, highlighting its success in addressing bias within personalized recommendation systems.

Keywords: personalized recommendation systems, bias, detection, mitigation, eigenvector centrality, influential nodes, regularization technique,

*https://www.linkedin.com/in/nadiera-mustapha-27230a180/

1 Introduction

In today's digital era, personalized recommendations have revolutionized the online experience. These systems have become integral to various industries, including e-commerce, and music/video streaming services, enhancing user experience and aligning recommendations with individual interests to achieve this level of customisation relies on user data, which predominantly observational rather than experimental [1].

Biases stem from various sources in the data such as the user demographics, historical preferences, and patterns of interaction within the system. Some of the biases can cause the system to suffer from but not limited to selection bias, position bias, exposure bias, and popularity bias[1].

As these systems expand into diverse domains, such as healthcare and e-learning, and the increasing reliance on personalized recommendations, it becomes crucial to understanding and addressing the biases within these systems and ensure fairness and transparency. This research aims to advocate for fairness and transparency in recommendation systems by identifying and mitigating bias.The paper consist of three main sections:Baseline model development,bias detection and mitigation

2 Baseline Model development

The first step in detecting and mitigating bias was to develop a baseline model that can act as a point of comparison for evaluating the effectiveness of subsequent bias techniques.

The MovieLens 1M dataset was ultized,to enhance personalization user and movie features were used. To capture both low and high-order feature interactions effectively,as well as learn intricate and complex representations, a deep cross network (DCN) architecture was employed. Previous research has demonstrated that incorporating these features significantly improves the expressiveness of models[3].

2.1 Results

During the training process,it could be noted that there was a gradual decrease in both the Root Mean Square Error(RMSE) and training loss values,indicating effective learning and generalization by avoiding over-fitting.A final RMSE of **0.8499** for the training set was reached,accompanied by a loss value of **0.7191**.

Subsequently,the test set evaluation yielded an RMSE of **0.9542** with a corresponding loss of **0.90213** , slightly higher than the training set. Although this demonstrates a slightly increased deviation, the performance remains reasonable.

3 Bias Detection

Two approaches were used to explore bias in the model:

1. Subset Datasets Based on User Demographics and Ratings: To analyze bias related to user demographics, subsets were created for specific population groups, such as young users, middle-aged users, male users, and female users. Separate datasets were formed for positive examples (ratings of 4 or higher) and negative examples (ratings below 4). This allowed for the examination of potential biases associated with different user groups and rating patterns.

2. Calculation of Eigenvector Centrality Using Ratings as Weights: Eigenvector centrality of nodes in a graph were computed with ratings as weights. This approach aimed to understand the influence of users or movies in shaping recommendations. The top 10 nodes with the highest centrality values were analyzed to gain insights into the influential factors shaping the recommendations.

3.1 Results

Subset	**RMSE Score**
Young Users Subset	0.82
Middle-Aged Users Subset	0.69
Male Users Subset	0.79
Female Users Subset	0.71

Table 1: Table 1: RMSE Scores for Different Subsets

Node	**Centrality**
b'2858'	0.1426
b'260'	0.1198
b'1196'	0.1170
b'593'	0.1081
b'2028'	0.1068
b'1210'	0.1064
b'608'	0.1061
b'2571'	0.1049
b'589'	0.1027
b'1198'	0.1008

Table 2: Top 10 Nodes by Centrality

Table 1 displays RMSE values for various user groups, revealing variations and potential bias in recommendations. Meanwhile, Table 2 highlights the top 10 Nodes by Centrality, representing influential users and popular movies shaping the system's recommendations.

3.2 Discussion

The observed variations in RMSE values across different population subsets in Table 1 indicate performance discrepancies based on user demographics, suggesting potential bias in the recommendation system. Higher RMSE values for

the young and female user subsets imply larger prediction deviations for these groups, highlighting a possible bias in the system's recommendations for specific user segments.
Additionally, the eigenvector centrality analysis provides insights into the influential users and popular movies in the recommendation system. Users with higher centrality values shown in Table 2, such as Node b'2858', b'260', and b'1196', have a more significant impact in shaping the recommendations. Similarly, movies with higher centrality values, like Node b'593', b'2028', and b'1210', are more influential and popular within the system. These findings further underscore the presence of bias in the system.

4 Bias Mitigation

The core approach taken to mitigate popularity bias was implementing a regularization technique that adjusted the impact of the user ratings on movie recommendations based on the relative popularity of movies. This method calculates the popularity weight based on the number of ratings a movie has received. Movies with a larger number of ratings are considered popular and are given a higher popularity weight. Non-popular movies are given a popularity weight that is the reciprocal of the number of ratings they have received. The popularity weight is multiplied by the user rating, resulting in the regularized rating.
The regularization technique effectively balanced the influence of user preferences and movie popularity by reducing the impact of user ratings on more popular movies, ensuring that their ratings contribute less to the overall recommendation process. Conversely, it increases the impact of user ratings on less popular movies, giving them more weight in the recommendations.

4.1 Results

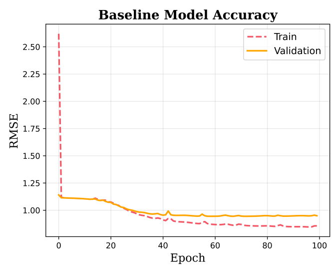
In contrast to the baseline model,the debiased model yielded significantly better results as seen in Figure 1 and results from the p-test in Table 3.

An independent t-test was performed as an additional measure to compare the performance of the model with the popularity weight regularization against the model without the popularity weight regularization. The t-test results are as follows:

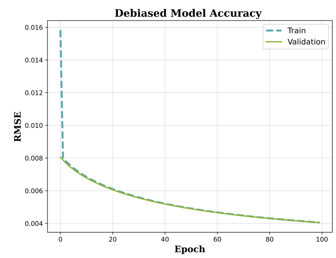
Table 3: Results of Independent Two-Sample t-test

t-statistic	-13.54
p-value	0.0054

With a t-statistic of **-13.54** and corresponding p-value of **0.0054** which is less than the significance level (typically 0.05) the null hypothesis which states there is no difference between the two models performance can be rejected and the alternative hypothesis can be accepted **Accuracy Plots:**



(a) Baseline Model Accuracy



(b) Debiased Model Accuracy

Figure 1: Comparison of Model Accuracy's

Based on these results, it can be inferred that the DCN model with popularity weight regularization performs significantly better than the DCN model without popularity weight regularization.

These findings highlight the importance of incorporating popularity weight regularization in the DCN model, as it leads to improved performance in terms of RMSE.

4.2 Discussion

The debiasing model, aimed at mitigating bias, shows improved performance on the evaluation dataset. The RMSE values for this model are substantially lower, with values around **2.2873e-04** and **2.3162e-04**. These near-zero values indicate significantly reduced prediction errors, highlighting the effectiveness of the debiasing model in addressing bias.

5 Conclusion

In conclusion, this research paper focused on bias detection and mitigation in personalized recommendation systems. Through the creation of subset datasets based on user demographics and ratings potential biases associated with different user groups and rating patterns were identified.The calculation of eigenvector centrality further shed light on the influential users and popular movies have, revealing potential sources of bias within the recommendation system.The developed debiased model showcases remarkable performance with low RMSE values, measuring around **2.2873e-04** and **2.3162e-04**,ultimately contributing to more equitable and accurate recommendations across diverse user subgroups and creating more fair and transparent recommnedation systems.

References

[1] Jiawei Chen et al. "Bias and Debias in Recommender System: A Survey and Future Directions". In: *ACM Trans. Inf. Syst.* 41.3 (Feb. 2023). ISSN: 1046-8188. DOI: 10.1145/3564284. URL: https://doi.org/10.1145/3564284.

[2] "Personalized and Location Aware Online Product Recommendation System". In: *IJSR* 6 (7 2017), pp. 650–654. DOI: 10.21275/art20175182.

[3] Ruoxi Wang et al. "Deep & Cross Network for Ad Click Predictions". In: *Proceedings of the ADKDD'17*. ADKDD'17. Halifax, NS, Canada: Association for Computing Machinery, 2017. ISBN: 9781450351942. DOI: 10.1145/3124749.3124754. URL: https://doi.org/10.1145/3124749.3124754.