Course Project Report

Mitigating Unfairness and Bias in Cold Start Recommenders

Submitted By

Kumsetty Nikhil Venkat (181IT224)
Amith Bhat (181IT105)

as part of the requirements of the course

Information Retrieval (IT458) [Jul - Nov 2021]

in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology in Information Technology

under the guidance of

Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

undergone at

DEPARTMENT OF INFORMATION TECHNOLOGY

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL

JUL-NOV 2021

DEPARTMENT OF INFORMATION TECHNOLOGY

National Institute of Technology Karnataka, Surathkal

C E R T I F I C A T E

This is to certify that the Course project Work Report entitled **"Mitigating Unfairness and Bias in Cold Start Recommenders"** is submitted by the group mentioned below -

Details of Project Group

Name of the Student	Register No.	Signature with Date
Kumsetty Nikhil Venkat	181IT224	Nikhil
Amith Bhat	181IT105	Amith

this report is a record of the work carried out by them as part of the course **Information Retrieval (IT458)** during the semester **Jul - Nov 2021**. It is accepted as the Course Project Report submission in the partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Information Technology.**

(Name and Signature of Course Instructor)
Dr. Sowmya Kamath S

D E C L A R A T I O N

We hereby declare that the project report entitled **"Mitigating Unfairness and Bias in Cold Start Recommenders"** submitted by us for the course **Information Retrieval (IT458)** during the semester **Jul-Nov 2021**, as part of the partial course requirements for the award of the degree of Bachelor of Technology in Information Technology at NITK Surathkal is our original work. We declare that the project has not formed the basis for the award of any degree, associateship, fellowship or any other similar titles elsewhere.

Details of Project Group

Name of the Student	Register No.	Signature with Date
1. Kumsetty Nikhil Venkat	181IT224	
2. Amith Bhat	181IT105	

Place: NITK, Surathkal
Date is: November 24, 2021

Mitigating Unfairness and Bias in Cold Start Recommenders

Nikhil Venkat Kumsetty[1], Amith Bhat[2]

Abstract— **Till date, attempts to study bias and fairness in recommender systems have focused on improving fairness and mitigating bias only in situations and for items where a history of the user profile already exists. In this project, we explore the bias against new items without any feedback history which are added to recommender systems. For this purpose, we implement two models to improve fairness and mitigate bias in cold start recommenders, by using fairness measures such as equal opportunity and Rawlsian Min-Max scores and incorporating them into our models. We train and test our model over the MLM1 dataset, and achieve favourable results over the base paper.**

Keywords: **Information retrieval , Recommendation systems, bias , cold start, fairness.**

I. INTRODUCTION

Broadly, recommender systems can be defined as algorithms aimed at suggesting relevant items to users as per their requirements (as perceived by the algorithm). Nevertheless, achieving fairness among the different items recommended by the system remains a challenge. For example, in a (scientific) paper recommender system , it is important to ensure there is no bias between papers published in a famous or non-famous university. Further information on the importance of mitigating bias in recommendation systems has be recognized in [1],[2], [3], [4], [5]. It may also lead to legal repercussions to companies employing these systems, and users not getting the relevant recommendations, which decreases satisfaction.

The papers [1], [3], [5], [6], [7], indicate that it is possible for even the popular recommendation systems to give biased recommendations towards different items. However, these papers consider bias and fairness only well after the item has been added into the system, that is, the "warm start" situation. It has been indicated by previous works in this field, that the single-largest cause of unfairness in this scenario would be the data bias in user-history, which is considered as the feedback for the recommender system.Examples include user movements such as views, clicks, etc. . This bias can be inherited and propagated by recommendation algorithms, and produce overall unfair and unsatisfactory results.

However, in the case of the new items, when there is no historical feedback, things are a bit different. An extension of the argument in the previous paragraph would imply that the bias from the warm start items will be transferred to any newly added items which in the recommendation system. This information is often the item content features learnt by the machine learning algorithm, and induces the unfair recommendations. The problem with the new items facing an inherent bias against themselves is that this bias will only be accumulated and perpetuated through the time the item stays in the recommender, thus making it harder to mitigate this bias the longer the time the item spends in the system.

As a corollary, ensuring minimal or no bias against new/cold start items would lead to increased fairness in the recommendations. Furthermore, this information would propagate, increasing the chances of training fairer models down the line. Therefore, we explore the possibilities of mitigating bias and increasing fairness in the cold start scenario (ie, adding new items to the system) at the start of the items' tenure in the recommender.

A problem which we encounter here is the issue of formally defining fairness. How exactly should we quantify the concept of fairness? The base paper considers two popular ideas – Rawlsian Max-Min fairness principle [8] and the equal opportunity principle [9] – and incorporated it into the models, so that the fairness is measured inherently. We wish to give recommendations that increases the true positive rate(TPR) of the least favoured items as much as possible. The TPR is defined as the fraction of number of accurate recommendations made of items to users who will like it, to the total number of recommendations made.

The equal opportunity principle leads to us to score fairness by using TPR as the measure, hence leading fairness to be directly proportional to the feedback received and thus user satisfaction. The Rawlsian Min-Max fairness principle allows us to accept inequalities, hence not requiring us to decrease the score of popular items just because they are popular, ie, we do not penalize items for being popular. Hence, without losing user satisfaction, we can improve item fairness.

II. LITERATURE REVIEW

A. Related Work

Biased recommendations for items in recommender systems decreases the satisfaction of the user, who does not get the best and most relevant item recommended, and the system owners, who do not receive as much profit [1],[2], [3], [4], [5]. Early studies into fairness and bias in recommender systems mostly concentrated on the rating prediction tasks. These works also explored fairness and unfairness in the ranking of an item by calculating the differences in the rating distributions predicted by recommenders across item groups [4], [6], [10], [11], [12]. Later, works evolved past this rating and score concept of fairness among items to a latent factor

manipulation [4] and regularization based [6], [10], [11], [12] methods of evaluation of fairness in recommenders.

Also, with the rise of recommender systems based on ranking, there were proposals of new variations of algorithms [1], [3], [5], [7], [13] that investigated fairness in item rankings directly than on the predicted scores which were more intermediary in nature. Other works [1], [3], [13], [4] propose equal opportunity based fairness measures. These measures require an equal TPR across the group of items. They were, in part, influenced by the fairness-aware classification of items in recommenders. To improve these proposed algorithms, many variants employing techniques such as adversarial learning [5], regularization [1], [13], re-ranking [3] , etc. have also been put forward.

We also refer to our base paper, [14]. Here, they recognize the limitations of the work in their field, where fairness of items recommended was not discussed for the new items context. Here, they propose, develop and implement two models to give fair recommendations, for new items.

B. Outcome of Literature Review

In our literature review, we found that most papers in the domain of recommender systems focused on improving fairness in the warm start scenario. Hence, unfairness in systems which do not have any user data was not explored till now. Also, we try to improve the rankings produced by the models proposed in the base paper.

C. Motivation

To create an equitable recommender system which gives fair recommendations to users even when there is no input about the users' preferences.

D. Problem Statement

To implement two machine learning models to improve the fairness in the ranking of and mitigate bias in cold start recommenders among new items added to the recommender.

III. METHODOLOGY

A. The Joint-learning generative model

The detailed structure and architectural design of the joint-learning generative model is portrayed in figure 1.

The joint-learning generative model consists are 2 components: a distribution generator ϕ which helps in generating a target distribution $\overline{P}$. It is built with the help of a multi-layer perceptron model consisting of 1-dimension input and output layers. These layers take an input of S random seeds from a standard normal distribution. and the model will generate an output of S samples $\overline{R} \in \mathbb{R}^S$ to represent the target distribution.
and an auto-encoder ψ. For every iteration in the training

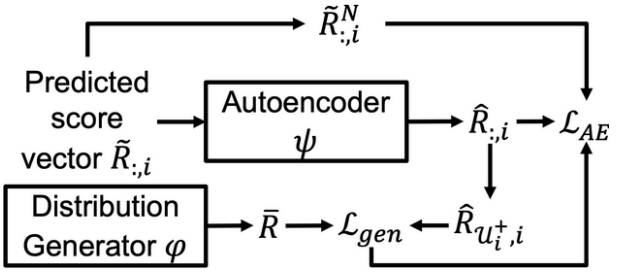

Fig. 1: Architecture of Joint-learning generative method

of the model the distribution ϕ component creates a temporary target distribution $\overline{P}$, then the auto-encoder ψ component of the model is updated so that it enhances the items that were previously under-recommended in the last epoch. This helps in improving the computed matched-user score distribution which is denoted by $P(\hat{R_{U^+}}_i, i)$. This distribution score is make to converge on to the target distribution $\overline{P}$.

Simultaneously, the auto-encoder ψ preserves the recommendation weights computed until that epoch. Finally, the distribution generator ϕ component is updated to the newly computed value, which in turn generates a new target distribution by computing the average of all the item's matched-user score distributions $P(\hat{R_{U^+}}_i, i)$, by updating the distribution generator ϕ , then sum of distribution distances is minimized between the newly generated samples $\overline{R}$ from ψ and the generated matched-user scores $P(\hat{R_{U^+}}_i, i)$, so this helps the model to minimize the underlying distribution of $\overline{R}$ in which the sum of distances to be minimized from the list of items. Then the model implements a Maximum Mean Discrepancy method that is used to compute the distribution distance by small samples from the 2 generated distributions as the resultant loss of the distribution generator.

B. The Score Scaling Method

This method helps to reduce the popularity bias in the recommendation system and enhances the fairness among new items.

The main goal of this model is to re-scale the train the models on train set with the priority to popularity parameter. This is achieved by improving the ratings for all the items with low popularity and decreasing the rating value for all the more popular rated warm-items, later the model normalizes the recommendation model and then it is trained on the newly scaled data points and hence, outputs a de-biased recommendation for the users. Along with this, the model scales a user-normalized predicted score vector $R^{\tilde{N}}{}_i$

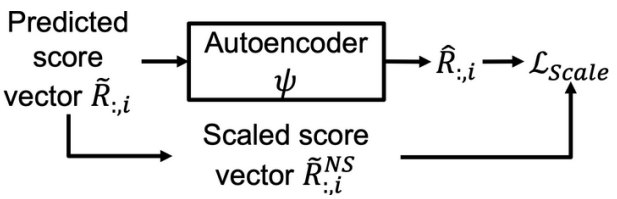

Fig. 2: Architecture of Score Scaling method

By implementing this model we achieved an enhancement in fairness rating scores for all the existing items in the dataset. The auto-encoder has been trained with the newly generated scaled results as new ranking of recommendation, this will in turn bring the much needed fair recommendations for cold items. The model also has the RMSE loss function to better learn the proposed Scale model, and improve the efficiency of the implemented model.

C. Novelty

- We implemented various approaches to mitigate specific problems faced by cold start recommendation systems, such as bias of the cold start recommendation system as a whole (taken care of in gen method), and popularity bias which is taken care by the score scaling method.
- We also run the gen and scale models with k=200 (200 cold start items in a batch). This greatly increases the values of the precision, recall and NDCG.

IV. RESULTS AND ANALYSIS

We observe that from the results obtained the score scaling model is performing better than joint-learning generative method. hence, by decreasing the popularity bias in the cold start recommendation system the recommenders ranking of the new items has improved greatly.

TABLE I: Results of Joint-learning generative models

	@15	@30	@200
NDCG	53.54%	61.85%	77.67%
Recall	33.05%	34.46%	78.27%
Precision	37.94%	50.29%	76.86%

TABLE II: Results of Score scaling method

	@15	@30	@200
NDCG	47.91%	58.72%	79.67%
Recall	47.11%	55.38%	79.76%
Precision	52.81%	61.35%	77.59%

We also observe that the joint-learning generative method greatly helped in improving the fairness of the recommendation of new items. Even though both joint-learning generative method, score scaling method are very effective at improving the fairness of cold start recommendation system, the joint-learning generative method is best used for improving the fairness and the score scaling method is best used to item-view recommendation.

The item-view parameter has greatly improved in score scaling method when compared to joint-learning generative model, because the items initially were under-rated by base recommendation models received better performance from the two implemented models, this lead to the increase in overall item-view parameter.

Based on the above discussed results, we observe that the Score Scaling model as well as Joint-learning Generative model are very effective in reducing the unfairness among cold items, and also these models are successful in conserving and improving the item-view parameter.

Also, as mentioned in the novelty section, we have obtained better results compared to the base paper. The differences between our results are detailed in the below table.

TABLE III: Results of the base paper

	@15	@30
Score Scaling	52.82%	51.35%
Joint Learning Generative	53.79%	52.06%

As we can see, we have significant improvements over the base paper models. We attribute this to the application of the various measures to mitigate popularity bias and unfairness in cold start recommenders. We also obtained significantly better results when compared to the base paper [14]. Even though the NDCG@15 is similar to base paper, the NDCG@30 values show an overall increase of 5% and also, we implemented the joint-generative model and score scaling model for obtained the values of NDCG@200, precision, and recall. This was not implement in [14].

V. CONCLUSIONS

In this work, we investigated the unfairness and bias among recommendation systems in terms of recommending new item in a cold user. We implemented two machine learning models joint-learning generative model and score scaling model. We also performed several experimentations to identify the effectiveness of the two implemented models to decrease the unfairness and preserve the recommendation feature space.

For future work, we also want to experiment of the unfairness and bias in unified recommendation systems that is caused by warm and cold items, We would also like to implement this model on other datasets such as CiteULike, etc.

We would also like to deploy the trained models as an application that can recommend items to users.

REFERENCES

[1] A. Beutel, J. Chen, T. Doshi, H. Qian, L. Wei, Y. Wu, L. Heldt, Z. Zhao, L. Hong, E. H. Chi, and C. Goodrow, "Fairness in recommendation ranking through pairwise comparisons," *CoRR*, vol. abs/1903.00780, 2019. [Online]. Available: http://arxiv.org/abs/1903.00780

[2] R. Burke, "Multisided fairness for recommendation," *CoRR*, vol. abs/1707.00093, 2017. [Online]. Available: http://arxiv.org/abs/1707.00093

[3] S. C. Geyik, S. Ambler, and K. Kenthapadi, "Fairness-aware ranking in search recommendation systems with application to linkedin talent search," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, ser. KDD '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 2221–2231. [Online]. Available: https://doi.org/10.1145/3292500.3330691

[4] Z. Zhu, X. Hu, and J. Caverlee, "Fairness-aware tensor-based recommendation," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, ser. CIKM '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1153–1162. [Online]. Available: https://doi.org/10.1145/3269206.3271795

[5] Z. Zhu, J. Wang, and J. Caverlee, "Measuring and mitigating item under-recommendation bias in personalized ranking systems," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 449–458. [Online]. Available: https://doi.org/10.1145/3397271.3401177

[6] T. Kamishima and S. Akaho, "Considerations on recommendation independence for a find-good-items task," 2017.

[7] W. Liu and R. Burke, "Personalizing fairness-aware re-ranking," *CoRR*, vol. abs/1809.02921, 2018. [Online]. Available: http://arxiv.org/abs/1809.02921

[8] C. L. Ten, *Mind*, vol. 112, no. 447, pp. 563–566, 2003. [Online]. Available: http://www.jstor.org/stable/3489212

[9] M. Hardt, E. Price, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016. [Online]. Available: https://proceedings.neurips.cc/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf

[10] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma, "Efficiency improvement of neutrality-enhanced recommendation," in *Decisions@RecSys*, 2013.

[11] ——, "Recommendation independence," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, ser. Proceedings of Machine Learning Research, S. A. Friedler and C. Wilson, Eds., vol. 81. PMLR, 23–24 Feb 2018, pp. 187–201. [Online]. Available: https://proceedings.mlr.press/v81/kamishima18a.html

[12] T. Kamishima, S. Akaho, H. Asoh, and I. Sato, "Model-based approaches for independence-enhanced recommendation," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 2016, pp. 860–867.

[13] F. Prost, H. Qian, Q. Chen, E. H. Chi, J. Chen, and A. Beutel, "Toward a better trade-off between performance and fairness with kernel-based distribution matching," *CoRR*, vol. abs/1910.11779, 2019. [Online]. Available: http://arxiv.org/abs/1910.11779

[14] Z. Zhu, J. Kim, T. Nguyen, A. Fenton, and J. Caverlee, "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*, ser. SIGIR '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 767–776. [Online]. Available: https://doi.org/10.1145/3404835.3462948

Team17_AmithB_NikhilV.pdf

ORIGINALITY REPORT

3%	0%	3%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Ziwei Zhu, Jingu Kim, Trung Nguyen, Aish Fenton, James Caverlee. "Fairness among New Items in Cold Start Recommender Systems", Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021 Publication	3%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		