Mitigating Unfairness and Bias in Cold Start Recommenders

By Team 17
Kumsetty Nikhil Venkat - 181IT224
Amith Bhat Nekkare - 181IT105

ABSTRACT

In this project we investigated the recommendation fairness among new items. Previous efforts in this field have studied fairness in recommender systems in scenarios where unfairness arises due to biased prior user-feedback history (like clicks or views). However, to date, the problem of ensuring fairness has not been studied in a system without feedback history.

In this project, we implemented a learnable post-processing framework as a blueprint for enhancing fairness, with which we create two concrete models. Experiments over the ML1M dataset shows the effectiveness of these models for enhancing fairness while also preserving recommendation utility.

INTRODUCTION

- Recommender systems are important for connecting users to the right items. These systems are used in many places, such as Amazon's product recommender system, Netflix's movie recommender system, etc.

- However, a challenge associated with recommender systems is ensuring the fairness of the recommendations. For example, in a a recruiting recommender that recommends job candidates , are candidates of different genders treated equally? Unfairness in recommendation may lead to potential negative impacts to user satisfaction, the recommendation platform itself, etc.

INTRODUCTION

- Previous works have revealed that widely used recommendation algorithms can indeed produce unfair recommendations toward different items. In this scenario, they show that the main driver of unfairness is the data bias in historical feedback (like clicks or views), and recommendation algorithms unaware of this bias can inherit and amplify this bias to produce unfair recommendations.

- What if there is no historical feedback?

INTRODUCTION

- We know that the data bias can be transferred from warm start items to new items through item content features due to the nature of machine learning algorithms.

- This fairness gap can be especially problematic since unfairness introduced by cold start recommenders will be perpetuated and accumulated through the entire life cycle of an item, resulting in growing difficulty for mitigating unfairness as the life cycle goes on.

INTRODUCTION

- One significant obstacle we face is the issue of formally defining fairness. How exactly should we quantify the concept of fairness? In this work, we follow two well-known concepts – **equal opportunity** and **Rawlsian Max-Min fairness** principle of distributive justice – to introduce the ***Max-Min Opportunity Fairness*** in the context of cold start scenarios.

- In a classification task, **equal opportunity** requires a model to produce the same true positive rate (TPR) for all individuals or groups. **Rawlsian Max-Min fairness** requires a model to maximize the minimum utility of individuals or groups so that no subject is underserved by the model.

Motivation

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

LITERATURE REVIEW

Paper & Author	Work Done
Burke et al. [1] (2017)	Mostly concentrated on the rating prediction tasks. 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.
Kamishima et al. [2] (2017)	Evolved into using a score concept of fairness among items using regularization based methods of evaluation of fairness in recommenders.
Prost et al. [3] (2019)	Proposed a new algorithm which investigated fairness in item rankings directly than on the predicted scores which were more intermediary in nature.
Ziwei et al. [4] (2021)	Introduced the innovation of cold start recommender systems.

OUTCOME OF LITERATURE REVIEW

In our literature review, we found that most papers in the domain of recommender systems focussed on improving fairness in scenarios where unfairness arises due to biased prior user-feedback history (like clicks or views). Hence, unfairness in systems which do not have any user data was not explored till now.

PROBLEM STATEMENT

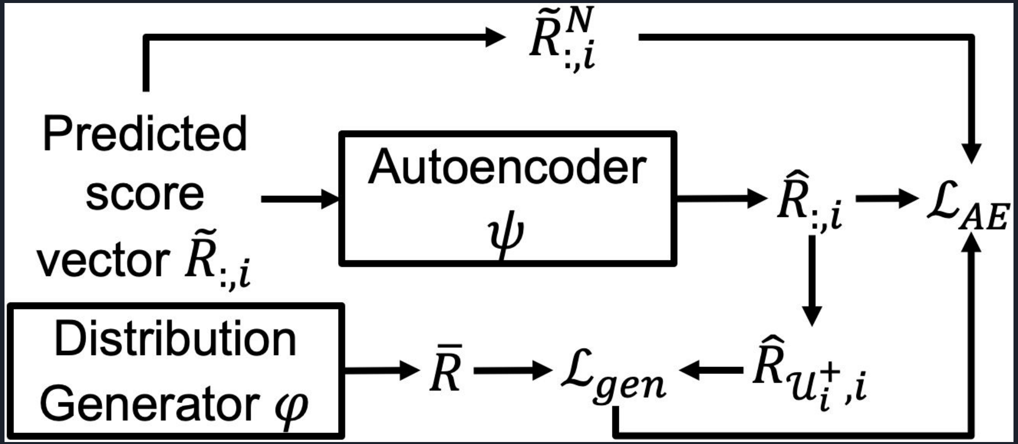
To show the prevalence of unfairness in cold start recommender systems and develop a learnable post-processing model for enhancing fairness of the recommender system, especially in the absence of user feedback data.

METHODOLOGY

- Develop a model to learn a transformation from item content feature of warm start items to user-item interactions between warm start users and items during training.
- Apply the learned transformation process to the content feature of a new item to predict the possible interactions between users and new items as a recommendation during testing.
- For this purpose we will use mainly 2 types of approaches
    - Joint-learning Generative Method
    - Score Scaling Method

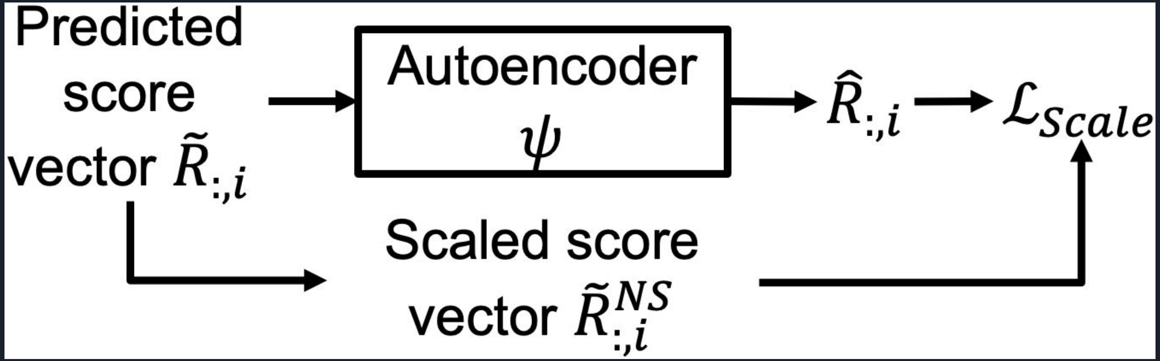
METHODOLOGY

- Joint-learning Generative Method:



METHODOLOGY

- Score Scaling Method:



WORK DONE

- We pre-processed the ML1M dataset.

- We have implemented the joint-learning generative, and the score-scaling models. It was run on the preprocessed ML1M dataset.

- Analysed the obtained results and optimized the gen and scale model to produce optimal results.

Results and Analysis - gen method

- 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.
- We also observe that the joint-learning generative method greatly helped in improving the fairness of the recommendation of new items.

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

RESULTS OF JOINT-LEARNING GENERATIVE METHOD

Results and Analysis - scale method

- The item-view utility improvement in score scaling method is due to that items originally under-served by base models receive more utility from the two proposed models, leading to the improvement of overall item-view utility even though the best-served items receive lower utility.

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

RESULTS OF SCORE SCALING METHOD

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. Hence we get fairly better results than the base paper.

- 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.

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

RESULTS OF BASE PAPER

CONCLUSION

In this work, we investigate the fairness among new items in cold start recommendation systems. First, we experimentally prove the presence of unfairness in existing cold start systems.

We implemented two concrete models – Scale and Gen. Lastly, we perform experiments to show the effectiveness of the two proposed models for enhancing fairness and preserving recommendation utility.

FUTURE WORK

- Explore the recommendation fairness between cold and warm items in a unified recommendation scenario, so as to apply this model in real-world context.

- Implement this model on other datasets such as CiteULike (dataset recording user preferences toward scientific articles), etc.

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

Individual Contributions

- ***Amith Bhat***: Data Preprocessing on the ML1M dataset, implementing joint-learning generative model.

- ***Kumsetty Nikhil Venkat***: Data Preprocessing on the ML1M dataset, implementing score-scaling model.

CITATIONS

[1] R. Burke, “Multisided fairness for recommendation,”CoRR, vol.abs/1707.00093, 2017. [Online]. Available: http://arxiv.org/abs/1707.00093

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

[3] 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

[4] **(BASE PAPER)** Zhu, J. Kim, T. Nguyen, A. Fenton, and J. Caverlee,“Fairness among new items in cold start recommender systems,”inProceedings 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

Thank You