A Re-ranking Approach for By-directional Fairness on Recommender Systems

Yaowei Peng
School of Artificial Intelligence and Computer Science, Jiangnan University, Jiangsu, China
featsccc@yahoo.com
Corresponding Author: Yaowei Peng Email: featsccc@yahoo.com

***Abstract*—Filter bubble problem has long constrained users of recommender systems from using it freely. Both side recommenders' users, the contents consumer and the contents provider, are disturbed by the meaningless repeating of few high frequency contents. While the majority of previous work is concerning the fairness issue of recommenders from one side, in this paper I provide a new light weight approach through re-ranking method increasing fairness for both sides. Experiments on 2 datasets and 4 existing models is conducted demonstrating that my proposed algorithm is capable of reducing unfair without harming overall accuracy.**

Keywords—recommendation system, fairness, re-ranking, information retrival

I. INTRODUCTION

Since both the supply and requirement of information over the web has outnumbered the capability a single person could handle, recommender system helps all of us screening for the critical information we need. During the last decade, the topic is fully discussed and related skills is well developed, as we are all facing it on Facebook or Amazon by a daily bias. Yet in recent 5 years, the awareness of filter bubble problem [1] has arisen. Specifically, recommender systems in one way or another, calculates similarity between user and user or item and item, which let the preference of highly active users who hold more records could be calculated flows into recommendation of inactive users whose record is sparser, and such flowing does not always lead to where inactive users want. Besides, recommenders also push items with the same tag once a user shows interest on a labeled item. These seem solid from common sense though, according to a survey [2] by Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher finds out recommenders implementing above 2 methods make 7 to 15 % more frequent recommendations for top 10 popular items than the actual time it was clicked. Alongside the homogenization consumers feel, such concentration causes other damage. The rest of providers make use of the recommender system will find their chance to be in the frontpage is not matched with their potential clicks thus they may consider moving to another platform, and they are the many. And this will harm the long-term operation for platforms who owns the recommender [3]. Reference 2 did not mention but I have to remind that long tail providers

have one tenth more clicks than recommendation plots meaning the one tenth interaction relies on search item name precisely where recommender system do not work.

Consider that the practice of recommendation system is mature and by-directional approaches towards its fairness are rather rare, I propose a new way to chase better experience for all stakeholders with recommenders through post process measure which takes no risk of changing the system already running. In section Ⅱ a little summary of related work is provide. In section Ⅲ I give a clear definition of the problem and my choice of metrics in terms of fairness and describes the main branch of the optimization target, its solution with an analysis of complexity. Experiment detail and an interpretation comes section Ⅳ, and section Ⅴ concludes. An average 0.16 increase of item exposure and 0.08 increase in user NDCG is observed by 4 replicate experiments on 2 datasets with 4 different models.

II. RELATED WORK

A. The Variation of Fair Recommendation Target

As the survey [4] reveals, from 2017 to early 2022, over 60 papers focusing on fair recommendation had been publish on well-known conferences like SIGIR. The initial interest regard to fairness is risen by the prejudice towards minority groups on critical platforms such as job matching websites. The algorithm & training data there producing systematically bias towards minorities derived the concept of protected groups. Later the concept expands to the active users' supremacy [5] over inactive users. In this paper I focus on the unfair triggered by data sparsity, [6] discuss the former issue.

A recommender system must have someone put contents on it for selection, and another population to safari its results. It is a two-side platform in nature, however by the introduction of [10], less than one tenth approaches chose to solve the problem from both sides simultaneously. Reference 10 gives a nice example of by-directional fairness, decent methods of consumer side fairness and provider side fairness are listed in [11][12]. The ratio of works heading for consumer fairness, provider fairness or fairness in both sides is demonstrated in Fig. 1.

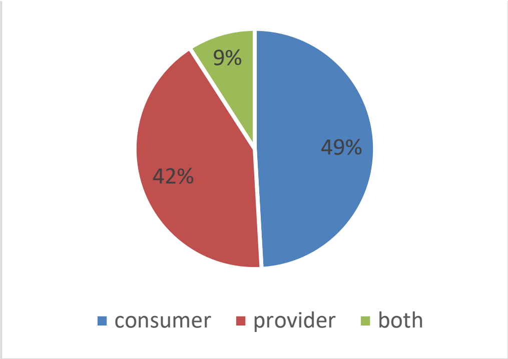

Fig. 1. The percentage of fair models from different aspect.

B. The Different Ways to Inplement Fairness

There is also a taxonomy dividing fair recommendation into case-by-case fairness [8] or fairness for the protected as a whole [7]. Both set a mission to reach a threshold of certain statistical target though, individual fairness in my viewpoint is in shortage of flexibility. For example, the Envy Free metric [9] is often presented at a compromise that only half individuals could reach default target, making it kind of new metric names EF 1/2. Thus, I stick on group fairness to have my mission come closer and, provide another advantage which is the division of whether an individual belongs to the protected group could be decided as fast as in an hourly schedule pushing the fairness into a more reactive level.

Nowadays desensitization of key features like age and gender is much common, and the publish of pre-process shares only 8% of total publishes by mid-2022 [13]. Also known from [13] 58% of papers chose to build a new ranking system with the idea of fairness from the scratch, while the remaining 34% trying to find a better order given by existing models. Since recommenders have been widely deployed, re-ranking is more feasible compared with brand new models, however it may also be constrained by the base methods re-ranking algorithms have to rely on.

III. THE RE-RANING ALGORITHM

Here I give a definition of fairness within recommendation systems from the perspective of both consumers and providers then introduce my algorithm trying to optimize the linear combination of user nDCG, item exposure and recommending scores given by base methods of course. And my optimizer is a practice of PYTHON-MIP [14]. Source file of my algorithm can be obtained from https://github.com/PehnYao/zhizhu

A. Problem Formulation

Set U and I being the set includes users and items respectively, with size c and d and the recommender has already generated a results in a result vector r_u for each u in U making the full recommending result for U from I a matrix $R^{c\times a}$. Construct binary unknown variables' matrix $W^{c\times a}$ with elements either 0 or 1, and element-wise multiply it with $R^{c\times a}$ we get a new matrix $N^{c\times a}$. The problem is finding N that for each row in N note as n_u, there is and only is b none-zero elements for all n_u. That means b elements is refined from the base result r_u with a elements and $b < a$. From all N fits above condition, find the best N with the highest fairness metrics scores. Such refine gives a space $c\times(a\text{-}b)$ making increase the experience of both consumer and provider simultaneously possible, and after cleaning zeros in N we get, there is b elements in each row of N as mentioned, the actual answer occupies a shape of $c\times b$.

B. Fairness Definition

First, although you may have noticed that many writings draw a line like last 0.8 or other arbitrary number, I define the protected user group to be users with interactions lower than the mean interaction times on the whole set. Consider the change of mean value is rather slow thus it is unnecessary to re-calculate at a high frequency. In large scale practice the right side can be replaced with a stored constant. Similarly, items with a below average interactions belongs to protected item group. Let S be the interaction record in real world,

$$\mathrm{count}(\mathrm{u_p})\ in\ S < \frac{\sum_{u\in U} count(u)\ in\ S}{c} \tag{1}$$

$$\mathrm{count}(\mathrm{i_p})\ in\ S < \frac{\sum_{i\in I} count(i)\ in\ S}{d} \tag{2}$$

In my viewpoint the relative order of a few elements which are few enough to be presented in the frontpage at the same time is trivial. On other side, it is unconvincing that an arbitrary divider like log(2+x) could measure the subtle difference between them thus I chose group calculative gain to represent user fairness and it looks like

$$UCG = \frac{\sum_{u\in(U-U_p)} n_u}{len(U-U_p)} - \frac{\sum_{u_p\in U_p} n_{u_p}}{len(U_p)} \tag{3}$$

I didn't make it normalized because nDCG stands for a good measurement though, the large uncertainty of ideal order due to the sparsity of the data itself, makes normalization quite suspicious as an optimizing target, let along the deviation of CG between users is a reflection of activeness between users. And I define ICG (not an abbreviation for anything) for fairness of provider side as such

$$ICG = \sum_{i\in(I-I_p)} \frac{count(i)\ in\ N}{count(i)\ in\ S} - \sum_{i_p\in I_p} \frac{count(i_p)\ in\ N}{count(i_p)\ in\ S} \tag{4}$$

The number of unique items in all recommendation results divide by total number of unique items.

With those in mind, the optimization target is

$$maxmize \sum_{x=0}^{c} \sum_{y=0}^{a} N_{xy} - \alpha * UCG - \beta * ICG$$

$$s.t. \sum_{y=0}^{a} W_{xy} == b, \{x \in Z | 0 \leq x < c\} \quad (5)$$

Where alpha and beta are hyperparameters. And please note *N* is the element-wise product of variables' matrix *W* and score matrix *R*. By now you may have found that (5) looks familiar just like the 0-1 knapsack problem, with limited non-repeatable items can take and a score attached on each item. That feels right, thus it is an NP-hard problem. In fact, these problems have a name "Integer Linear Programming". And thanks to [14] some heuristic function is available to conquer ILP in feasible time. And here I also provide a simple approach to solve (5).

C. Problem Solver

The main idea is set all elements in *W* equals zero, and for each time change a row of *W* from 0 to 1, then for each element in the row with 1, for each time change an element from 1 to 0, calculate (5), and then change it to 1 again, calculate (5) and record the difference.

Algorithm Simple Solver:

```
1 Input W=0, S,R,I_p,U_p ,A=0
2 for x in [0, c]
3   W_x=1, list(a)
4   for y in [0, a]
5       m = fair_score
6       W_xy=0
7       n = fair_score
8       list.append(m - n)
9   list.sort()
10  A_x=list[:top 10], W_x=0
11 return A
```

Finally, I discuss the complexity of above method. Line 1 creates the largest variables W, and R both consumes O(c×a) space. Since a is a constant the space complexity will be O(c) that is linear to input size.

Line 2-10 is a double nested c*a loop. Line 3 and10 creates and alter elements at O(1), both line 5 and 7 calculates formula (5) with a complexity of O(c×a), while Line 5,7 finishes at the cost O(1). Therefore, the time complexity is also O(c), linear to input size.

IV. Experiments

I made 4 repeat experiments on Movie Lens 100K and Movie Lens 1M dataset, each with the following base methods: Variational Autoencoders for Collaborative Encoding (VAECF)[15], Neural Matrix Factorization (NeuMF)[16], Collaborative Filtering for Implicit Feedback Datasets (CFIF)[17], and BPR[18]. The comparison is done by refine 10 items out of 50 items the base method provides, to compare with the top 10 items ranked by base method.

Table 1. statistics of datasets

Dataset	Users	Items	Ratings	Sparsity
ML100K	1000	1700	100000	94.1%
ML1M	6000	4000	1000000	95.8%

Table 2. Results on ML100K

Model	NDCG@10	Precision @10	Recall@10	Exposure
NeuMF	0.1448	0.1235	0.1108	0.0719
NeuMF-R	0.2061	0.4580	0.4209	0.3521
BPR	0.1103	0.0989	0.0783	0.0082
BPR-R	0.0393	0.3334	0.3026	0.2324
CFIF	0.0906	0.0838	0.0616	0.0467
CFIF-R	0.1127	0.2867	0.2674	0.2365
VAECF	0.1408	0.1215	0.1075	0.1105
VAECF-R	0.2832	0.4586	0.4202	0.3741

Table 3. Results on ML1M

Model	NDCG@10	Precision @10	Recall@10	Exposure
NeuMF	0.1339	0.1229	0.0717	0.2406
NeuMF-R	0.4819	0.4550	0.4265	0.2621
BPR	0.1105	0.1015	0.0520	0.0170
BPR-R	0.0621	0.3400	0.3156	0.1698
CFIF	0.0948	0.0868	0.0463	0.0152
CFIF-R	0.0492	0.3050	0.2780	0.1648
VAECF	0.1299	0.1186	0.0691	0.1841
VAECF-R	0.3829	0.4358	0.4015	0.2464

The user nDCG rise for 8% and item exposure for 16% is observed after the refining from 50 to 10 items. Achieves the goal to increase fairness at the same time. However, maybe due to the algorithm does not take relative position into calculation, the performance user nDCG looks very unstable.

V. Conclusion

In this paper, I invented a new measurement of provider fairness (4), proposed a new way to judge by-directional fairness (5) and shared a new approximate approach to let the problem solved. The results are satisfying for the considerable upgrade of fairness metrics while did no harm to overall precision of recommendation. There may be more comparison with other re-ranking methods, and I will also try to stabilize the performance on position related metrics.

ACKNOWLEDGMENT

This work was supported by Professor Xuezhong Qian in Jiangnan University.

REFERENCES

[1] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In Proceedings of the 23rd international conference on World wide web (WWW '14). Association for Computing Machinery,NewYork,NY,USA,677–686. https://doi.org/10.1145/2566486.2568012

[2] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher. 2020. The Connection Between Popularity Bias, Calibration, and Fairness in Recommendation. In Proceedings of the 14th ACM Conference on Recommender Systems (RecSys '20). Association for Computing Machinery, NewYork,NY,USA,726–731. https://doi.org/10.1145/3383313.3418487

[3] Himan Abdollahpouri and Robin Burke. 2019. Multi-stakeholder Recommendation and its Connection to Multi-sided Fairness. In Proceedings of RMSE@RecSys'19. ACM, Copenhagen, Denmark, 6 pages.

[4] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2023. A Survey on the Fairness of Recommender Systems. ACM Trans. Inf. Syst. 41, 3, Article 52 (July 2023), 43 pages. https://doi.org/10.1145/3547333

[5] Ricardo Baeza-Yates. 2018. Bias on the web. Commun. ACM 61, 6 (June 2018), 54–61. https://doi.org/10.1145/3209581

[6] Harini Suresh and John Guttag. 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '21), October 5–9, 2021, --, NY, USA. ACM, New York, NY, USA 9 Pages. https://doi.org/10.1145/3465416.3483305

[7] Zuohui Fu, Yikun Xian, Ruoyuan Gao, Jieyu Zhao, Qiaoying Huang, Yingqiang Ge, Shuyuan Xu, Shijie Geng, Chirag Shah, Yongfeng Zhang, et al. 2020. Fairness-aware explainable recommendation over knowledge graphs. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 69–78.

[8] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through Awareness. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (Cambridge, Massachusetts) (ITCS '12). Association for Computing Machinery, New York, NY, USA, 214–226. https://doi.org/10.1145/2090236.2090255

[9] Dimitris Serbos, Shuyao Qi, Nikos Mamoulis, Evaggelia Pitoura, and Panayiotis Tsaparas. 2017. Fairness in Package-to-Group Recommendations. In Proceedings of the 26th International Conference on World Wide Web (Perth, Australia) (WWW '17). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 371–379. https://doi.org/10.1145/3038912.3052612

[10] Mohammadmehdi Naghiaei, Hossein A. Rahmani, and Yashar Deldjoo. 2022. CPFair: Personalized Consumer and Producer Fairness Re-ranking for Recommender Systems. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3477495.3531959

[11] Y. Deldjoo, V. W. Anelli, H. Zamani, A. Bellogin, T. Di Noia, A flexible framework for evaluating user and item fairness in recommender systems, User Modeling and User-Adapted Interaction (2021) 1–47.

[12] Z. Zhu, J. Kim, T. Nguyen, A. Fenton, J. Caverlee, Fairness among new items in cold start recommender systems, in: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '21, 2021, pp. 767–776.

[13] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Difonzo, Dario Zanzonelli, A Survey of Research on Fair Recommender Systems. arXiv:2205.11127. unpublished.

[14] Andersen, G. Cornuéjols and Y. Li, Reduce-and-Split Cuts: Improving the Performance of Mixed-Integer Gomory Cuts. Management Science, 51, 1593-1732, 2005.

[15] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. In Proceedings of The 2018 Web Conference (WWW 2018). ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3178876.318615

[16] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In Proceedings of the 26th International Conference on World Wide Web (WWW '17). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva,CHE,173–182. https://doi.org/10.1145/3038912.3052569

[17] Y. Hu, Y. Koren and C. Volinsky, "Collaborative Filtering for Implicit Feedback Datasets," 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 263-272, doi: 10.1109/ICDM.2008.22.

[18] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI '09). AUAI Press, Arlington, Virginia, USA, 452–461