Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems | School of Computing and Information Systems

2-2024

Recommendations with minimum exposure guarantees: A post-processing framework

Ramon LOPES

Rodrigo ALVES

Antoine LEDENT
Singapore Management University, aledent@smu.edu.sg

Rodrygo L. T. SANTOS

Marius KLOFT

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Databases and Information Systems Commons, Numerical Analysis and Scientific Computing Commons, and the Theory and Algorithms Commons

Citation

LOPES, Ramon; ALVES, Rodrigo; LEDENT, Antoine; SANTOS, Rodrygo L. T.; and KLOFT, Marius. Recommendations with minimum exposure guarantees: A post-processing framework. (2024). *Expert Systems with Applications*. 236, 1-9.

Available at: https://ink.library.smu.edu.sg/sis_research/8182

This Journal Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Recommendations with Minimum Exposure Guarantees: A Post-processing Framework

Ramon Lopes[a] (rlopes@ufrb.edu.br), Rodrigo Alves[b] (rodrigo.alves@fit.cvut.cz), Antoine Ledent[c] (aledent@smu.edu.sg), Rodrygo L. T. Santos[d] (rodrygo@dcc.ufmg.br), Marius Kloft[e] (kloft@cs.uni-kl.de)

[a] Universidade Federal do Recôncavo da Bahia, Cruz das Almas, Brazil

[b] Czech Technical University in Prague, Prague, Czech Republic

[c] Singapore Management University, Singapore

[d] Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

[e] TU Kaiserslautern, Kaiserslautern, Rlp, Germany

Corresponding Author:
Ramon Lopes
Universidade Federal do Recôncavo da Bahia, Cruz das Almas, Brazil
Tel: +55 71 99281 1292
Email: rlopes@ufrb.edu.br

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

Recommendations with Minimum Exposure Guarantees: A Post-processing Framework

Ramon Lopes[a,*], Rodrigo Alves[b], Antoine Ledent[c], Rodrygo L. T. Santos[d], Marius Kloft[e]

[a] *Universidade Federal do Recôncavo da Bahia, Cruz das Almas, Brazil*
[b] *Czech Technical University in Prague, Prague, Czech Republic*
[c] *Singapore Management University, Singapore*
[d] *Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil*
[e] *TU Kaiserslautern, Kaiserslautern, Rlp, Germany*

Abstract

Relevance-based ranking is a popular ingredient in recommenders, but it frequently struggles to meet fairness criteria because social and cultural norms may favor some item groups over others. For instance, some items might receive lower ratings due to some sort of bias (e.g. gender bias). A fair ranking should balance the exposure of items from advantaged and disadvantaged groups. To this end, we propose a novel post-processing framework to produce fair, exposure-aware recommendations. Our approach is based on an integer linear programming model maximizing the expected utility while satisfying a minimum exposure constraint. The model has fewer variables than previous work and thus can be deployed to larger datasets and allows the organization to define a minimum level of exposure for groups of items. We conduct an extensive empirical evaluation indicating that our new framework can increase the exposure of items from disadvantaged groups at a small cost of recommendation accuracy.

Keywords: recommender systems, fairness, exposure, integer linear programming

*Corresponding author.
Email addresses: `rlopes@ufrb.edu.br` (Ramon Lopes), `rodrigo.alves@fit.cvut.cz` (Rodrigo Alves), `aledent@smu.edu.sg` (Antoine Ledent), `rodrygo@dcc.ufmg.br` (Rodrygo L. T. Santos), `kloft@cs.uni-kl.de` (Marius Kloft)

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

1. Introduction

The main objective of Recommender Systems (RS) is to guide users in a personalized way to items maximizing user satisfaction while providing business value (e.g. increased revenue, engagement) for the recommendation provider. To this end, a popular strategy is to predict the top-K most relevant items (Yang et al., 2012; Zehlike et al., 2017). For instance, by using prior user interactions, the RS (1) estimates the relevance of non-interacted items, (2) constructs a list by allocating items according to their predicted relevance and then (3) picks the top K items from this list. The top K items are then presented to the user *ordered according to expected relevance.* The user may then consume one of the items and rate it with a probability decreasing with the suggestion rank. In fact, the user behavior is affected by the *position bias* (Craswell et al., 2008) where the probability of an item being clicked depends not only on its relevance, but also on its position in the recommendation list. Finally, after observing user feedback, the algorithm updates its mechanisms to improve the recommendation in subsequent requests.

Although such relevance-based ranking has been used in a wide variety of RS domains (Aggarwal et al., 2016), it frequently fails to meet fairness criteria. A well-known issue is that an advantaged group can be heavily prioritized in the top ranking positions despite having no (or very slight) superiority in terms of utility, leaving only a few slots for the disadvantaged group (Abdollahpouri et al., 2019b). Such a problem can happen due to a variety of reasons, including the presence of filter bubbles (Nguyen et al., 2014), the polarity of thoughts (Ziani et al., 2017; Liu et al., 2022), or the discrepancy in consumption popularity between items (Park & Tuzhilin, 2008; Zhu et al., 2021; Abdollahpouri et al., 2019a). For instance, by analyzing public recommendation datasets enriched with gender information, Ferraro et al. (2021) expose the gender imbalance in music recommendation: only 25.6% of the recommended music are from female artists. As a result, they posit that female artists may not receive

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

an adequate level of exposure. Thus, measures must be taken to produce fair recommendations.

To provide a more equitable ranking, several methods have been proposed by the recommender systems community (Pitoura et al., 2021; Zehlike et al., 2021; Wang et al., 2022). Compared to purely accuracy-oriented algorithms, these strategies significantly improve the proportion of non-privileged items added to the list of recommendations presented to the users, resulting in a fairer outcome. There are various ways to improve fairness in ranking-based recommender systems, ranging from preprocessing approaches, which modify the data to correcting inherent biases and prejudice (Martínez et al., 2016), to post-processing approaches (Zehlike et al., 2017; Biega et al., 2018; Singh & Joachims, 2018), which rerank the recommendation list produced by a ranker to equalize the representation of all groups. One of the main advantages of post-processing approaches is that they are model-agnostic (i.e. they treat the ranking model as a black box) and can thus be used in conjunction with any ranking model; this might be a desirable property for organizations that already have a baseline ranking model in production.

Unfortunately, a serious shortcoming of most existing approaches is that they merely focus on increasing the total number of recommendations for items belonging to the disadvantaged group, without concern for their specific recommendation rank in the display list. It is well known, however, that users tend to observe and interact with top-ranked items much more often, and an item's expected exposure decreases when moving down from the top recommendation (Craswell et al., 2008; Abdollahpouri et al., 2019b; Singh & Joachims, 2018). Adding items from the disadvantaged group at the end of the display list may thus have no practical influence on exposure fairness because the user will most likely not interact with them. Thus, a fair ranking should not only provide an equitable number of recommendations, but balance the expected realized exposure among items from the advantaged group and disadvantaged groups. Doing so is a delicate task that requires an explicit model of the expected exposure (the probability of an item being consumed, as a function of

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

the position in the list of suggestions). Last but not least important, to the best of our knowledge, there is no approach in the literature that guarantees a minimum level of exposure for groups of items in the final recommendation list; for example, this might occur in platforms that must to provide a minimum level of exposure to item providers as a service level agreement.

To fill this gap, we propose a post-processing framework to produce recommendations which are fair at the *exposure level* whilst also having reasonable time complexity. Inspired by recent attempts (Biega et al., 2018; Singh & Joachims, 2018) to produce fair rankings based on optimization approaches, we propose an Integer Linear Programming (ILP) model that maximizes the expected utility while satisfying exposure constraints chosen by the organization. The list of recommendations output by our model takes fairness considerations not only in terms of which items to include, but also in terms of the *order* in which they are presented. This fact is illustrated in Figure 1 (cf. also the comprehensive description on the caption), which for illustration purposes (as in Ferraro et al. (2021)) takes women and men as the protected and non-protected groups, respectively.

In order to cover different scenarios that might arise in real world scenarios, we propose two set of constraints so that the organization can define minimum levels of exposure for groups of items. The Minimum Exposure (ME) constraint guarantees that each group must have a minimum level of exposure in the final recommendation ranking. Thus, under this constraint, the recommendation list is fair in the sense that each group of items receives a minimum level of exposure. In turn, the Relative Minimum Exposure (RE) constraint ensures that the exposure each protected group receives is at least a factor of the exposure received by a non-protected group. Hence, under the RE constraint, the model guarantees a relative minimum level of exposure among groups. In this way, organizations might produce recommendations that meet some service level agreement w.r.t exposure, for example, in the context of multiple item providers.

There are very few other post-processing methods that address this task. Singh & Joachims (2018) propose a framework for stochastic rankings that

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

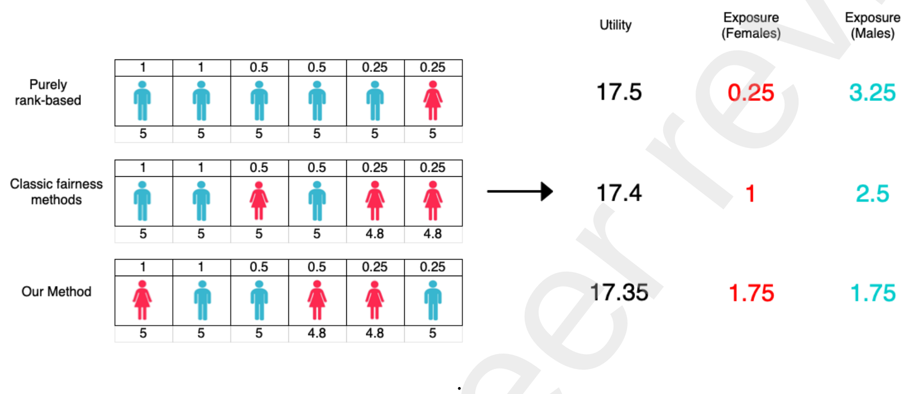


Figure 1: In this example, the protected and non-protected groups consist of female and male artists respectively, as described in Ferraro et al. (2021). The numbers written below each individual correspond to the expected utility (rating). For simplicity of illustration, we used $(1, 1, 0.5, 0.5, 0.25, 0.25)$ as a list of discount factors (represented at the top of the ranking positions). One can see that the "purely ranking based" method fills the recommendation list almost exclusively with male artists, despite the fact that the men have only very slightly higher predicted utility (this phenomenon being caused by other sources of gender bias). In the second line, we can see a classic approach, which adds members of the protected group to the list of recommendations, without regard to their precise ranking. This results in increased fairness but female artists still lack *exposure*. At the last line, we can see an illustration of the results of our method, where female artists are well-represented throughout all ranking positions. Note that the rankings are *predicted* and the difference between both groups can be partially explained by a form of popularity bias, and would become even more attenuated after the accumulation of more feedback when using our method

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

can be instantiated to guarantee certain exposure constraints. However, that method requires solving an ILP problem involving $|I|^2$ binary variables, as well as solving a decomposition problem with $O(|I|^2)$ time-complexity, where $|I|$ is the number of items in the catalog. Likewise, Biega et al. (2018) propose an ILP model with $|I|^2$ binary variables where fairness is achieved across a series of rankings. Thus, the applicability of these methods might be reduced in larger datasets.

Our model presents several advantages. First, it involves no more than $K \cdot |I|$ binary variables where K is the number of items in the recommendation list, which makes it less computationally expensive compared to other ILP-based approaches in the literature. Second, the proposed exposure constraints are flexible in the sense that organizations are able to define a minimum level of exposure that must be met w.r.t. groups. As a result, our model tackles the important trade-off between fairness and accuracy in RSs.

In summary, the main contributions of this paper are as follows:

- We propose a novel ILP model to produce fair recommendation lists. Unlike most existing work, our model not only balances the representation of various groups within the list of recommendations, but carefully adapts the order of the recommendations to equalize the exposure and relevance of group of items in the recommendation.

- We introduce and analyze two fairness raking-based constraints. These fairness constraints model some sort of service level agreement constraints found in practice by organizations.

- By considering well-established real-world datasets, we present a comprehensive empirical evaluation of our framework w.r.t. the proposed constraints in different scenarios.

The remainder of this paper is organized as follows. In Section 2, we review the related literature. In Section 3, we introduce our framework. We discuss the setup and the results of our empirical evaluation in Sections 4 and 5, respectively.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

Finally, in Section 6, we present concluding remarks.

2. Related Work

In the past few years, fairness requirements in ranking and recommendation have been the subject of intense research (Zehlike et al., 2021; Pitoura et al., 2021). Next, we briefly present the most prominent approaches found in the literature. Xiao et al. (2017) provide a multi-objective optimization framework for fairness-aware group recommendation on the basis of Pareto efficiency. Sürer et al. (2018) propose an ILP-based post-processing method to balance the distribution of recommendations across retailers in the context of multi-sided platforms, but they do not consider the impact of position bias in the final ranking. Celis et al. (2018) propose a method that produces rankings that meet lower and upper bound constraints w.r.t. the number of items with each attribute that are allowed to appear in the top-K positions of the ranking. Mehrotra et al. (2018) propose recommendation policies that balance the objectives of consumer relevance and supplier fairness in the context of two-sided platforms. Singh & Joachims (2019) propose a framework to learn fair policies in the context of LTR. Zehlike & Castillo (2020) propose an in-processing method in the context of Learning-To-Rank (LTR) that introduces a regularization term that encourages rankings where the exposure of the protected group is not less than the exposure of the non-protected group, but the model does provide any guarantees. Li et al. (2021) study the problem of fairness in recommendation from the user perspective in the sense that a fair RS should provide the same recommendation quality for different groups of users. Ge et al. (2021) explore the problem of long-term fairness in recommendation so that the model can dynamically adjust its recommendation policy to meet fairness constraints when the environment changes. Mansoury et al. (2021) propose a method based on network flow optimization to improve exposure fairness for suppliers. Marras et al. (2022) study the problem of fairness in the context of item providers where they propose an in-processing method based on the Hellinger distance. Next,

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

we discuss the approaches most related to ours.

Abdollahpouri et al. (2017) propose an in-processing Matrix Factorization (MF) based method for controlling popularity bias in RS. In particular, they introduce a regularization term in the loss function based on the Laplacian matrix of a co-membership matrix whose dimension is $|I| \times |I|$, which might be impractical as the number of items in the catalog increases. Experiments on two real-world datasets show that method is able to improve long-tail coverage at the cost of a reduction of recommendation quality. Since the proposed method is an in-processing approach, its applicability is limited as organizations have to make modifications into their ranking systems.

Singh & Joachims (2018) propose a post-processing ILP-based framework to guarantee exposure fairness constraints in the context of stochastic rankings. However, there are two issues that might limit its applicability in real world scenarios. First, the proposed ILP model has $|I|^2$ binary variables. Second, after solving the optimization model, its solution has to be decomposed using the Birkhoff-von Neumann decomposition (Birkhoff, 1940), which additionally incurs in $O(|I|^2)$ time-complexity. The authors propose three exposure allocation constraints: i) the average exposure of the documents in both groups is equal, ii) the exposure of the two groups is proportional to their average utility, iii) the expected click-through rate of each group is proportional to its average utility. Experiments on synthetic and small real-world datasets containing up to 25 documents demonstrate that the method is highly effective whilst only reducing recommendation quality by a small amount.

Biega et al. (2018) propose a post-processing framework to mitigate unfairness w.r.t. position bias so that the attention ranked documents receive is proportional do their relevance. In that context, fairness is achieved via a series of rankings instead of a single ranking. To this end, they propose an ILP model that has $|I|^2$ binary variables, which might be impractical in larger scale scenarios. The model is based on the concept of equity of amortized attention, where each item receives attention proportional to its cumulative relevance, and has a constraint that imposes a lower bound in the ranking quality w.r.t. the

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

ideal ranking. Experiments on synthetic and real-world datasets attest the effectiveness of the proposed method at the cost of lower quality rankings.

The works presented in this section differ from ours in several aspects. First, we aim at producing deterministic rankings instead of stochastic rankings. Second, we want to guarantee exposure fairness constraints in every single ranking produced instead of achieving fairness on the average of rankings produced. Third, we address group fairness instead of individual fairness; a careful reader should notice that individual fairness is a particular case of group fairness. Fourth, we propose a post-processing approach, which turns out to be agnostic to ranking model in production. Last but not least, we want models with a lower number of variables which can thus be deployed in real world scenarios. To the best of our knowledge, there is no approach in the literature that allows the organization to explicitly define constraints w.r.t the minimum exposure levels for groups of items and have guarantees that these requirements are met in every single ranking produced.

3. Modeling exposure constraints

In this section, we present our post-processing approach to produce recommendations with exposure guarantees. Our idea amounts to introducing constraints in an ILP to enforce that each group receives a minimum level of exposure in the final ranking. To this end, we model the exposure with a classic logarithmic discount factor model (Wang et al., 2013; Järvelin & Kekäläinen, 2002; Abdollahpouri et al., 2019a; Singh & Joachims, 2018). Based on this principle, we propose two set of constraints: i) Minimum Exposure (ME) constraints guarantee that each group receives a minimum level of exposure and ii) Relative Minimum Exposure (RE) constraints enforce that exposure of the protected group is at least a factor of that received by the non-protected group. In this way, our approach guarantees a minimum exposure in every recommendation. Last but not least, our model allows the organization to balance the level of exposure by means of a parameter in the model and thus calibrate the

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

trade-off between exposure and recommendation quality. Next, we introduce the mathematical notation required to describe our two exposure constraints.

Main notation: Let U be the set of users; I be the set of items; $G = \{g : g \subseteq I\}$ be a set of groups of items; K be the number of items in the recommendation ranking. We will write $\hat{y}_{ui}$ for the relevance estimated by a baseline recommender model for the combination of user u and item i. This will be one of the main inputs to our post-processing model, which aims to rerank the ranking provided by $\hat{y}_{ui}$ to increase its fairness. $x_{ik} \in \{0, 1\}$ will be a variable that assumes the value 1 if item i occupies the k-th position in the *final* recommendation ranking provided by our model and 0 otherwise; thus, x_{ik} is the main variable to learn. To describe group memberships, $I_i(g)$ will denote an indicator variable that assumes the value 1 if item i is a member of group $g \in G$ and 0 otherwise. We will write $p(k)$ for the position bias at position k (Craswell et al., 2008), i.e. the estimated probability of a user observing position k in a recommendation ranking; and $\alpha \in [0, 1]$ to denote the level of exposure. In fact, $p(k)$ is a quantity that must be modeled its estimate must be fed to our method. In experiments, we will focus on an inverse logarithmic which was originally suggested in (Järvelin & Kekäläinen, 2002) but has been extensively used and studied since then (Abdollahpouri et al., 2019a; Singh & Joachims, 2018; Wang et al., 2013). The idea is that the exposure probability $p(k)$ scales like $1/\log(1 + k)$. For the sake of readability, we summarize our notation in Table 1.

We present an ILP model for the problem of computing an optimal K-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

Notation	Explanation
α	level of exposure
G, g	set of groups of items, a group of items
G_p, G_n	set of protected groups, set of non protected groups
I, i	set of items, an item
$I_i(g)$	indicator variable if item i is a member of group g
K	number of items in the recommendation list
$p(k)$	bias at position k
$\rho\%$	percentage of most popular items according to the number of interactions in training
U, u	set of users, a user
x_{ik}	binary variable indicating if item i occupies k-the position
$\hat{y}_{ui}$	estimated relevance for user u and item i

Table 1: Main Notation.

ranking for a user u given estimated relevances $\hat{y}_{ui}$:

$$\max \sum_{k=1}^{K} \sum_{i \in I} p(k) \cdot \hat{y}_{ui} \cdot x_{ik} \tag{1}$$

$$\sum_{i \in I} x_{ik} = 1, \ k = 1 \dots K \tag{2}$$

$$\sum_{k=1}^{K} x_{ik} \leq 1, \ \forall i \in I \tag{3}$$

$$x_{ui} \in \{0, 1\}, \ \forall u \in U, i \in I \tag{4}$$

Eq (1) maximizes the utility considering the position bias. In fact, the utility of having item i in position k takes into account the interaction between the position bias $p(k)$ and the estimated relevance $\hat{y}_{ui}$. As a result, the higher the predicted relevance of an item the better its position in the final ranking. Eq (2) ensures that each position in the final ranking is be occupied by one and only one item. Eq (3) is a knapsack constraint that ensures that an item can occupy at most one position in the final ranking. Finally, Eq (4) ensures the variables are binary. One can clearly see that the optimal solution to this problem amounts to ranking the items in decreasing order of estimated relevance, which complies

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

with the Probability Ranking Principle (Robertson, 1997).

3.1. Minimum Exposure (ME) Model

This model is suitable for scenarios where an organization is required to guarantee a minimum level of exposure for certain groups of items in the final recommendation ranking. For instance, a business rule might require that each one out of ten genres of books should receive at least 10% of the total exposure in every recommendation list in order to attain minimum exposure levels. To this end, we propose the ME model that is initially defined by Eq (1)–(4) and the following set of constraint:

$$\sum_{k=1}^{K}\sum_{i\in I} p(k)\cdot I_i(g)\cdot x_{ik} \geq \alpha \cdot \sum_{k=1}^{K} p(k),\ \forall g \in G \tag{5}$$

Eq (5) ensures that each group $g \in G$ receives at least a fraction $\alpha \in [0, 1]$ of the total exposure in the final ranking. One should note that a necessary condition for the set of feasible solutions to be nonempty is that $\alpha \cdot |G| \leq 1$.

3.2. Relative Minimum Exposure (RE) Model

In turn, this model is applicable in scenarios where an organization should guarantee a minimum level of exposure for protected items relative to that of non-protected items. For instance, a company might require that female director's movies should receive at least 90% of the exposure received by those of male directors. In this context, we propose the RE model whose motivation is that each protected group receives a minimum level of exposure compared to that received by a non-protected group. RE is defined by Eq (1)–(4) and the following set of constraints:

$$\sum_{k=1}^{K}\sum_{i\in g_p} p(k)\cdot x_{ik} \geq \alpha \cdot \sum_{k=1}^{K}\sum_{i\in g_n} p(k)\cdot x_{ik},\ (g_p, g_n) \in G_r \tag{6}$$

Eq (6) ensures that the exposure received by the protected group g_p is at least $\alpha \in [0, 1]$ times the exposure received by the non-protected group g_n, where

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

$G_r = \{(g_p, g_n)|g_p \in G_p, g_n \in G_n\}$ is the set of pairs of protected and non-protected groups for which we should impose relative exposure constraints (here G_p (resp. G_n) denotes the set of protected (resp. non protected) groups).

4. Experimental Setup

To gain a better understanding of the merits of the proposed framework, we design experiments to answer the following research questions:

RQ1 How effective are our post-processing models at increasing the exposure of protected items?

RQ2 How do our post-processing models affect recommendation quality?

RQ3 How effective are our post-processing models compared to current fairness-aware state-of-the-art?

To answer these questions we make use of several public datasets in our experiments. In the following, we describe the setup of our empirical investigations.

Datasets: the MovieLens datasets (Harper & Konstan, 2015) are widely used in the RS literature and include several datasets of various sizes. Each dataset consists of users' preferences for movies expressed on a five-star rating scale. In this work, we consider the MovieLens 1M (ML1M for short) dataset that comprises 1 million ratings from 6,000 users on 4,000 movies. The Amazon dataset (McAuley et al., 2015) contains 142.8 million product reviews spanning May 1996 - July 2014 and product metadata from Amazon. Ratings are given by users on a five-star rating scale. In this work, we select two representative product categories from that dataset, namely, Electronics and Kindle. Since we do not focus on cold start issues, we keep only users and items with at least 5 ratings, i.e., we perform a k-core preprocessing to make sure each user/item has sufficient feedback. Meng et al. (2020) show the splitting strategy employed is an important factor when evaluating RS. In our experiments, we employ the temporal global splitting (Meng et al., 2020), where any interactions after a

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

fixed time point are used for testing; this strategy arguably represents a more realistic scenario. Thus, we split the datasets into training (80%), validation (10%) and testing (10%) subsets in this chronological way.

Baselines: we consider two strong baselines in our experiments that are most related to our approach. Abdollahpouri et al. (2017) propose an in-processing fairness-aware framework based on MF. In turn, Biega et al. (2018) propose a post-processing optimization model to produce fair rankings that takes into account equity-of-attention fairness.

Recommender Models and Parameter Tuning: we evaluate our post-processing models over a biased matrix factorization recommender (Koren et al., 2009), which was implemented in PyTorch, and an item-item recommender (Deshpande & Karypis, 2004), whose implementation is provided by Lenskit (Ekstrand, 2020). The choice of these recommendation models is motivated by findings of recent works (Ferrari Dacrema et al., 2019, 2021; Ludewig et al., 2019; Latifi et al., 2022) that show that well-tuned MF and item-based models provide competitive results compared to neural-based recommenders. We performed hyperparameter tuning so that the models provide the best possible results for the sake of a fair comparison. For the biased matrix factorization, we kept the number of latent factors fixed at 30 while experimenting with values of the learning rate in the range $\{10^{-3}, 10^{-2}, 10^{-1}\}$ and values of the regularization parameter in the range $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$. We used the popular Adam optimizer (Kingma & Ba, 2014). In turn, for the item-item recommender, we experimented with values of neighbors in $\{5, 10, 50\}$ and a minimum number of neighbors for scoring an item in the range $\{1, 5, 10\}$. Thus, we report the results for these methods on the basis of the best parameter configuration found on the validation set. Finally, optimization problems are solved to optimality using Gurobi 9.51. [1] By the time of publication of this article, we will made our code available in public repository.

Evaluation Methodology: we employ a binary scale of relevance where items

[1]https://www.gurobi.com/

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

are relevant if their rating is above the user average rating are not relevant otherwise. Based on the properties of the logarithmic discount factor (Wang et al., 2013; Järvelin & Kekäläinen, 2002), we assume the position bias for position k is given by $p(k) = 1/\log_2(k+1)$. Thus, we define the total exposure received by a group of items in a recommendation list as the sum of their corresponding $p(k)$. We report our results using Normalized Discounted Cumulative Gain (nDCG) (Järvelin & Kekäläinen, 2002), Mean Reciprocal Rank (MRR) (Shi et al., 2012), total exposure of the protected group and number of items from the protected group in the final ranking.

Groups: our models can be adapted to a wide variety of possible protected and non protected group. Following (Abdollahpouri et al., 2019a; Liu & Zheng, 2020), we focus on the simple and informative case where the non-protected group is defined to be the subset of items that are among the ρ% of most popular items according to the number of interactions they received in the training set, i.e., so called "short-head items". In turn, the protected group is composed of those items that are not in the non-protected group, i.e., so called "long-tail items". For instance, if we set ρ to 30%, the non-protected group is composed of the 30% of most popular items in the training dataset. In our experiments, we evaluate two different scenarios w.r.t to the choice of ρ: i) $\rho = 30\%$ represents the usual scenario where a reduced percentage of items accounts for the most interactions in the dataset and so most items are under exposed (Abdollahpouri et al., 2019a) and ii) $\rho = 90\%$ represents an extreme scenario where only items in the very end of the long tail should have their exposure improved. Our motivation for this setting lies in the well known fact that items with fewer known interactions tend to be ranked lower, even when they do not actually have lower utility (Zhu et al., 2021). Thus, we can evaluate the effectiveness of our post-processing models by contrasting short-head and long-tail items in two different scenarios.

In the next section, we report our results as an average of metric figures across all users in the test. Statistically significant differences are verified using a two-tailed paired t-test. In particular, we use the symbols $\bullet$ and $\circ$ to denote

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

respectively significant difference at the $p < 0.01$ and $p < 0.05$ levels from a given baseline; ↑ and ↓ indicate whether larger or smaller values are better, respectively.

5. Results

In the following, we address each of the research questions posed in Section 4 in turn. We adopt a specific nomenclature to make clear which method is under consideration. Namely, MF stands for the biased matrix factorization recommender and IB for the item-based recommender, ME-α and RE-α for our post-processing strategies with fairness ratio value α, MF-T for the baseline proposed in (Abdollahpouri et al., 2017) and EA for that proposed in (Biega et al., 2018). Results are displayed in Tables 2–4 for the MF recommender and Tables 5–7 for the IB counterpart. A careful reader should keep in mind that the optimization models are solved to optimality and so they ensure that every ranking produced attains the fairness requirements imposed.

5.1. Results for RQ1

To address RQ1, first we assess the performance of our post-processing models over the MF and IB recommenders. Overall, Tables 2–7 show that the larger the value of α, the larger the exposure and the number of protected items recommended for different configurations of ρ (recalling that ρ denotes the percentage threshold to define "most popular" items by number of interactions in training, whilst α – from Equations (5) and (6) – denotes the threshold in the fairness constraint). However, from Tables 5 and 6, one can see that for our model ME, the larger the value of α, the lesser the exposure and the number of protected items recommended for $\rho = 30\%$. In these settings, there is an over exposure of protected items compared to level α defined in the experiments. Thus, this finding suggests that our model ME is able to increase the exposure of protected group in cases of under exposure as well as to reduce its exposure in case of over exposure. Finally, one should note that IB produces recommendations where

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

the number of protected items in the final recommendation is higher compared to those provided by MF. For example, in the extreme scenario where $\rho = 90\%$, the MF method can barely recommend items from the protected group. This result suggests that MF-based recommenders might benefit from our methods to give exposure to long-tail items.

5.2. Results for RQ2

To address RQ2, we compare the recommendation quality provided by our models over the MF and IB recommenders to those provided by the original recommenders, i.e., no post-processing is applied. Overall, Tables 2–4 show that the original MF recommender systematically produces better recommendations compared to our post-processing models except in the Kindle dataset where the differences displayed have no statistical significance.

Next, we compare our models against the IB recommender. Table 7 shows that our models do not affect recommendation quality in the Kindle dataset; in the Electronics dataset (Table 6), recommendation quality is significantly reduced in two configurations (ME-0.4 and RE-0.9). Finally, in the ML1M dataset (Table 5), ME produces better recommendations when $\rho = 0.3$; on the other hand, our models presents a reduced recommendation quality when $\rho = 0.9$. To better understand the impact on recommendation quality, we next assess the robustness of our methods when recommending for users with various levels of sparsity.

Figure 2 provides performance breakdowns (with respect to nDCG) for our methods in the ML1M dataset. Each subfigure shows four bins representing user groups that interacted with an increasing amount of items organized according to the first (Q1), second (Q2) and third (Q3) quartiles of the corresponding training dataset. The first and second subgraphs from the first row represent respectively the results over MF for ME and RE, and one can conclude that the larger the user's history the lesser the degradation of recommendation quality. In turn, the first and second graphs from the second row show respectively the results over IB for ME and RE; if on the one hand ME improves recommendation

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

quality over all levels of user history, on the other hand RE produces worse recommendations but one can note the larger the user's history the lesser the reduction in recommendation quality.

In order to better understand these findings, we present box plots for item popularity according to the users history size in Figure 3. One can conclude that the larger the user history size the lower the median popularity of interacted items, which suggests that low profile users tend to interact with popular items while very active users not only explore the catalog but also interact with long-tail items. This might explain why the larger the user's history the lesser the reduction of recommendation quality in our experiments. In fact, our models tend to promote long-tail items since we defined groups with respect to item popularity. Thus, since low profile users tend to interact with popular items, recommending long-tail items might decrease recommendation quality. In contrast, active users benefit from recommendations that promote long-tail items.

Recalling research question RQ2, we can conclude that there is a trade-off between recommendation quality and the exposure of protected items in the recommendation list; this trade-off has been pointed out in several works in the literature (Zehlike & Castillo, 2020; Biega et al., 2018; Singh & Joachims, 2018). As a matter of fact, in most scenarios, the recommendation quality decreases significantly, especially over the MF recommender in the ML1M dataset. However, one should note that the larger the exposure ratio α the greater the reduction in recommendation quality. Thus, our models are flexible in the sense that the exposure ratio α should be fine-tuned to balance the desired minimum exposure and recommendation quality.

5.3. Results for RQ3

To address RQ3, we compare our models against MF-T and AE. From Tables 2–4, one can see that overall our models outperform MF-T w.r.t both recommendation quality and protected group exposure, specially in extreme cases (e.g. $\rho = 90\%$). This finding suggests the importance of taking into account the

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

position bias to produce fair recommendations.

Next, we compare our models against AE: i) ML1M dataset (Tables 2 and 5): our models systematically show lower exposure when $\rho = 30\%$ but they provide better recommendation quality, ME-0.4 and RE-0.9 outperform EA w.r.t protected group exposure when $\rho = 90\%$, ii) Electronics dataset (Tables 3 and 6): overall, our models show improved protected group exposure, and iii) Kindle dataset (Tables 4 and 7): our models show lower exposure of protected group when $\rho = 30\%$, but they improve protected group exposure when $\rho = 90\%$. These findings suggest that our methods are able to provide improved protected group exposure, specially for items in the very long tail (i.e. scenario where $\rho = 90\%$).

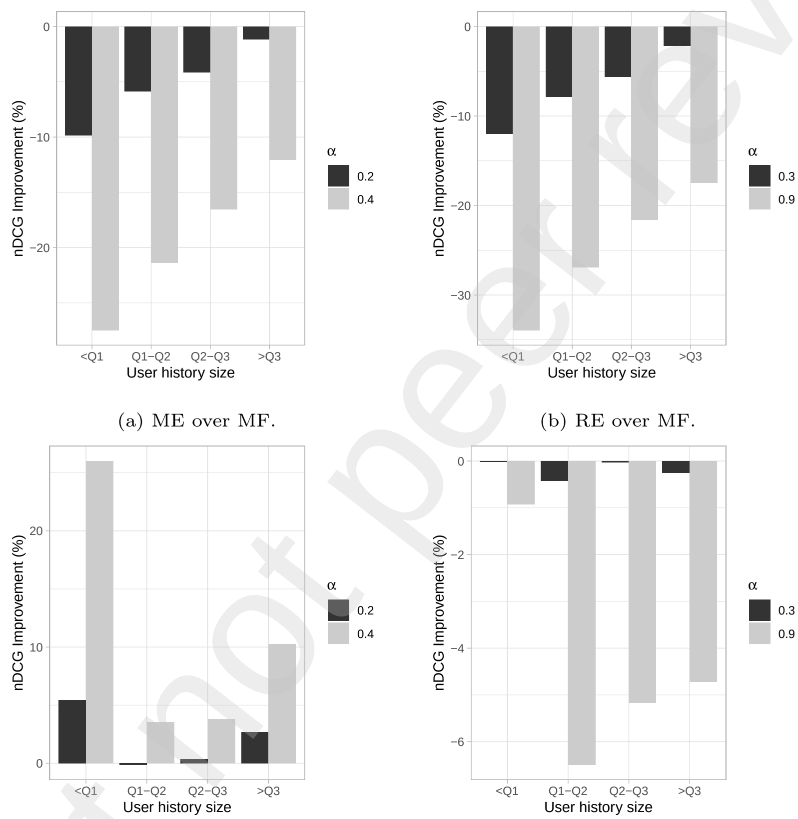
Thus, we conclude that our models provide higher quality recommendations and increase the exposure of items from the protected group in all scenarios but one when compared to the MF-T baseline. When compared to the AE baseline, our models are able to provide better recommendations w.r.t protected exposure or recommendation quality in four out of six configurations evaluated, which attest the benefits of our method. Recalling, our model has a parameter α that should be carefully tuned to balance the trade-off between increasing exposure and degrading recommendation quality.

6. Conclusion

In this work, we studied the problem of fairness of exposure in recommendations. In practice, items placed near the top of the recommendation list tend to have a higher probability of being consumed or rated by the user; thus, one should take this fact into account when defining and enforcing fairness conditions. To this end, we proposed a framework based on an ILP model that provides recommendations with minimum exposure guarantees: each group of items receives a minimum level of total expected exposure. Namely, the ME constraint guarantees that each group receives a minimum proportion of the total (expected) exposure while the RE constraint ensures that each item in the

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780



(a) ME over MF.

(b) RE over MF.

(c) ME over IB.

(d) RE over IB.

Figure 2: nDCG breakdown according to the users' history size in ML1M when $\rho = 0.3$.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

		MF	MF-T	EA	ME-0.2	ME-0.4	RE-0.3	RE-0.9
$\rho = 30\%$	nDCG↑	0.1528	•0.0832	•0.0317	•0.1442	•0.1221	•0.1416	•0.1135
	MRR↑	0.3060	•0.1360	•0.1843	•0.2854	•0.2326	•0.2789	•0.2174
	Exposure↑	1.1554	•1.7499	•5.8071	•1.6803	•2.8748	•1.8443	•3.3758
	Count↑	3.4560	•5.2141	•18.2937	•4.8956	•8.0114	•5.3216	•9.2751
$\rho = 90\%$	nDCG↑	0.1528	•0.0875	•0.0317	•0.1290	•0.0989	•0.1243	•0.0886
	MRR↑	0.3060	•0.1566	•0.1843	•0.2872	•0.2133	•0.2743	•0.1866
	Exposure↑	0.0003	0.0001	•1.8952	•1.4114	•2.8178	•1.6264	•3.3365
	Count↑	0.0010	0.0001	•6.5832	•4.4126	•7.6546	•4.8356	•8.7487

Table 2: Results for MF recommendation model: ML1M dataset

		MF	MF-T	EA	ME-0.2	ME-0.4	RE-0.3	RE-0.9
$\rho = 30\%$	nDCG↑	0.0042	•0.0018	•0.0077	○0.0042	0.0037	0.0041	•0.0031
	MRR↑	0.0024	0.0012	•0.0058	•0.0024	•0.0019	○0.0023	•0.0016
	Exposure↑	2.2275	•3.5924	•0.0001	•2.3384	•3.0202	•2.4004	•3.4570
	Count↑	6.6893	•10.0576	•0.0001	•7.0497	•8.9882	•7.2442	•10.1615
$\rho = 90\%$	nDCG↑	0.0042	•0.0010	•0.0077	•0.0032	•0.0025	•0.0032	•0.0023
	MRR↑	0.0024	○0.0009	•0.0058	•0.0019	•0.0015	•0.0018	•0.0013
	Exposure↑	0.1584	•1.2476	•0.0001	•1.4171	•2.8195	•1.6292	•3.3379
	Count↑	0.4732	•3.5121	•0.0001	•4.2260	•7.8129	•4.7302	•9.1645

Table 3: Results for MF recommendation model: Electronics dataset.

non-protected group receives a minimum proportion of the exposure relative to that received by a non-protected group. Compared to previous works, our

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

		MF	MF-T	EA	ME-0.2	ME-0.4	RE-0.3	RE-0.9
$\rho = 30\%$	nDCG↑	0.0066	0.0047	•0.0010	0.0064	0.0061	0.0064	0.0060
	MRR↑	0.0046	0.0043	•0.0014	0.0045	0.0040	0.0043	0.0039
	Exposure↑	0.6495	•1.3791	•5.5459	•1.4691	•2.8249	•1.6665	•3.3410
	Count↑	2.1167	•4.9126	•17.1753	•4.7983	•8.7187	•5.3944	•10.1155
$\rho = 90\%$	nDCG↑	0.0066	0.0048	•0.0010	0.0065	0.0058	0.0068	0.0056
	MRR↑	0.0046	0.0045	•0.0014	0.0041	0.0042	0.0044	0.0041
	Exposure↑	0.0567	•0.0262	•0.5553	•1.4142	•2.8183	•1.6270	•3.3369
	Count↑	0.1436	○0.0993	•1.4919	•4.3082	•8.0987	•4.8677	•9.4919

Table 4: Results for MF recommendation model: Kindle dataset.

		IB	EA	ME-0.2	ME-0.4	RE-0.3	RE-0.9
$\rho = 30\%$	nDCG↑	0.0656	•0.0063	•0.0668	•0.0720	0.0655	•0.0626
	MRR↑	0.1099	•0.0393	•0.1118	•0.1209	0.1099	•0.1036
	Exposure↑	4.3110	•6.4217	•4.2312	•3.8461	•4.3197	•4.4707
	Count↑	11.4281	•18.9483	•11.1603	•9.9607	•11.4519	•11.8304
$\rho = 90\%$	nDCG↑	0.0656	•0.0063	•0.0636	•0.0539	•0.0628	•0.0483
	MRR↑	0.1099	•0.0393	•0.1053	•0.0800	•0.1017	•0.0675
	Exposure↑	2.0752	2.1156	•2.2073	•2.9247	•2.2636	•3.3799
	Count↑	4.3495	•5.9938	•4.6525	•6.3671	•4.7549	•7.6319

Table 5: Results for IB recommendation model: ML1M dataset.

model shows two striking advantages: i) it has a smaller number of variables, which makes it much more computationally tractable and ii) it allows an organi-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

		IB	EA	ME-0.2	ME-0.4	RE-0.3	RE-0.9
$\rho = 30\%$	nDCG↑	0.0011	0.0008	0.0012	0.0015	0.0011	0.0011
	MRR↑	0.0007	0.0006	0.0008	0.0012	0.0007	0.0007
	Exposure↑	5.3814	5.3685	•5.1240	•4.1500	•5.3843	•5.3951
	Count↑	15.0970	•15.5032	•14.3599	•12.0364	15.0975	•15.1231
$\rho = 90\%$	nDCG↑	0.0011	0.0008	○0.0011	○0.0007	○0.0011	•0.0005
	MRR↑	0.0007	0.0006	0.0007	•0.0005	0.0006	•0.0003
	Exposure↑	0.7497	•0.6724	•1.4833	•2.8247	•1.6722	•3.3412
	Count↑	2.1645	•1.8764	•3.9414	•6.8498	•4.2930	•8.1581

Table 6: Results for IB recommendation model: Electronics dataset.

		IB	EA	ME-0.2	ME-0.4	RE-0.3	RE-0.9
$\rho = 30\%$	nDCG↑	0.0014	0.0006	0.0014	0.0019	0.0015	0.0016
	MRR↑	0.0012	0.0010	0.0011	0.0022	0.0012	0.0015
	Exposure↑	4.9404	•5.4900	•4.7727	•4.0344	•4.9431	•4.9989
	Count↑	13.9120	•15.8270	•13.4381	•11.6768	13.9114	•14.0347
$\rho = 90\%$	nDCG↑	0.0014	0.0006	0.0013	0.0014	0.0013	0.0015
	MRR↑	0.0012	0.0010	0.0010	0.0012	0.0009	0.0011
	Exposure↑	0.6162	•0.8658	•1.4608	•2.8231	•1.6557	•3.3400
	Count↑	1.7834	•2.4853	•3.7834	•6.7271	•4.1293	•8.0557

Table 7: Results for IB recommendation model: Kindle dataset.

zation to explicitly define minimum exposure levels for groups of items and have guarantees that these requirements are met in every single ranking produced.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

Experiments in several public datasets attest the effectiveness of our methods in giving exposure to items in protected groups. In some scenarios, there is a clear reduction in recommendation quality. However, a breakdown analysis shows that the larger the user history in training, the smaller the degradation in recommendation quality. In conclusion, our post-processing approach is able to produce rankings with minimum exposure guarantees and so might be used by organizations that have to meet some service level agreement w.r.t exposure of group of items.

References

Abdollahpouri, H., Burke, R., & Mobasher, B. (2017). Controlling popularity bias in learning-to-rank recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* RecSys '17 (p. 42–46). New York, NY, USA: Association for Computing Machinery.

Abdollahpouri, H., Burke, R., & Mobasher, B. (2019a). Managing popularity bias in recommender systems with personalized re-ranking. In *The thirty-second international flairs conference*.

Abdollahpouri, H., Mansoury, M., Burke, R., & Mobasher, B. (2019b).

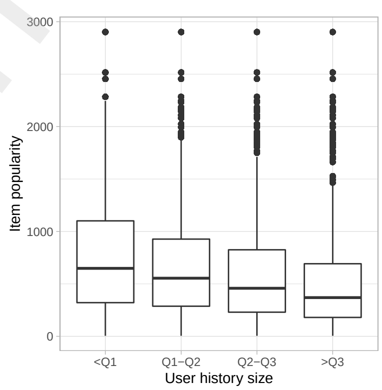


Figure 3: Popularity of items in ML1M when $\rho = 0.3$ according to the users' history size.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

The unfairness of popularity bias in recommendation. *arXiv preprint arXiv:1907.13286*, .

Aggarwal, C. C. et al. (2016). *Recommender systems* volume 1. Springer.

Biega, A. J., Gummadi, K. P., & Weikum, G. (2018). Equity of attention: Amortizing individual fairness in rankings. In *The 41st International ACM SIGIR Conference on Research amp; Development in Information Retrieval* SIGIR '18 (p. 405–414). New York, NY, USA: Association for Computing Machinery.

Birkhoff, G. (1940). *Lattice Theory*. Number v. 25;v. 1940 in American Mathematical Society: Colloquium publications. American Mathematical Society.

Celis, L. E., Straszak, D., & Vishnoi, N. K. (2018). Ranking with fairness constraints. In *Proceedings of the 45th International Colloquium on Automata, Languages, and Programming*.

Craswell, N., Zoeter, O., Taylor, M., & Ramsey, B. (2008). An experimental comparison of click position-bias models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* WSDM '08 (p. 87–94). New York, NY, USA: Association for Computing Machinery.

Deshpande, M., & Karypis, G. (2004). Item-based top-n recommendation algorithms. *ACM Transactions of Information Systems*, *22*, 143–177.

Ekstrand, M. D. (2020). Lenskit for python: Next-generation software for recommender systems experiments. In *Proceedings of the 29th ACM International Conference on Information amp; Knowledge Management* CIKM '20 (p. 2999–3006). New York, NY, USA: Association for Computing Machinery.

Ferrari Dacrema, M., Boglio, S., Cremonesi, P., & Jannach, D. (2021). A troubling analysis of reproducibility and progress in recommender systems research. *ACM Transactions on Information Systems*, *39*.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

Ferrari Dacrema, M., Cremonesi, P., & Jannach, D. (2019). Are we really making much progress? a worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems* RecSys '19 (p. 101–109). New York, NY, USA: Association for Computing Machinery.

Ferraro, A., Serra, X., & Bauer, C. (2021). Break the loop: Gender imbalance in music recommenders. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* CHIIR '21 (p. 249–254). New York, NY, USA: Association for Computing Machinery.

Ge, Y., Liu, S., Gao, R., Xian, Y., Li, Y., Zhao, X., Pei, C., Sun, F., Ge, J., Ou, W., & Zhang, Y. (2021). Towards long-term fairness in recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* WSDM '21 (p. 445–453). New York, NY, USA: Association for Computing Machinery.

Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, *5*, 1–19.

Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems*, *20*, 422–446.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. URL: https://arxiv.org/abs/1412.6980.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, *42*, 30–37.

Latifi, S., Jannach, D., & Ferraro, A. (2022). Sequential recommendation: A study on transformers, nearest neighbors and sampled metrics. *Information Sciences*, *609*, 660–678.

Li, Y., Chen, H., Fu, Z., Ge, Y., & Zhang, Y. (2021). User-oriented fairness in recommendation. In *Proceedings of the Web Conference 2021* WWW '21 (p. 624–632). New York, NY, USA: Association for Computing Machinery.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

Liu, H., Guo, Y., Yin, J., Gao, Z., & Nie, L. (2022). Review polarity-wise recommender. *IEEE Transactions on Neural Networks and Learning Systems*, .

Liu, S., & Zheng, Y. (2020). Long-tail session-based recommendation. In *Proceedings of the 14th ACM Conference on Recommender Systems* RecSys '20 (p. 509–514). New York, NY, USA: Association for Computing Machinery.

Ludewig, M., Mauro, N., Latifi, S., & Jannach, D. (2019). Performance comparison of neural and non-neural approaches to session-based recommendation. In *Proceedings of the 13th ACM Conference on Recommender Systems* RecSys '19 (p. 462–466). New York, NY, USA: Association for Computing Machinery.

Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., & Burke, R. (2021). A graph-based approach for mitigating multi-sided exposure bias in recommender systems. *ACM Transactions on Information Systems*, *40*.

Marras, M., Boratto, L., Ramos, G., & Fenu, G. (2022). Regulating group exposure for item providers in recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* SIGIR '22 (p. 1839–1843). New York, NY, USA: Association for Computing Machinery.

Martínez, L., Castro, J., & Yera, R. (2016). Managing natural noise in recommender systems. In C. Martín-Vide, T. Mizuki, & M. A. Vega-Rodríguez (Eds.), *Theory and Practice of Natural Computing* (pp. 3–17). Cham: Springer International Publishing.

McAuley, J., Pandey, R., & Leskovec, J. (2015). Inferring networks of substitutable and complementary products. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* KDD '15 (p. 785–794). New York, NY, USA: Association for Computing Machinery.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

Mehrotra, R., McInerney, J., Bouchard, H., Lalmas, M., & Diaz, F. (2018). Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness and & satisfaction in recommendation systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* CIKM '18 (p. 2243–2251). New York, NY, USA: Association for Computing Machinery.

Meng, Z., McCreadie, R., Macdonald, C., & Ounis, I. (2020). Exploring data splitting strategies for the evaluation of recommendation models. In *Fourteenth ACM Conference on Recommender Systems* (p. 681–686). New York, NY, USA: Association for Computing Machinery.

Nguyen, T. T., Hui, P.-M., Harper, F. M., Terveen, L., & Konstan, J. A. (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 (p. 677–686). New York, NY, USA: Association for Computing Machinery.

Park, Y.-J., & Tuzhilin, A. (2008). The long tail of recommender systems and how to leverage it. In *Proceedings of the 2008 ACM Conference on Recommender Systems* RecSys '08 (p. 11–18). New York, NY, USA: Association for Computing Machinery.

Pitoura, E., Stefanidis, K., & Koutrika, G. (2021). Fairness in rankings and recommendations: An overview. *The VLDB Journal*, *31*, 431–458.

Robertson, S. E. (1997). The probability ranking principle in ir. In *Readings in Information Retrieval* (p. 281–286). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Shi, Y., Karatzoglou, A., Baltrunas, L., Larson, M., Oliver, N., & Hanjalic, A. (2012). Climf: Learning to maximize reciprocal rank with collaborative less-is-more filtering. In *Proceedings of the Sixth ACM Conference on Recommender Systems* RecSys '12 (pp. 139–146). New York, NY, USA: ACM.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

Singh, A., & Joachims, T. (2018). Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining* KDD '18 (p. 2219–2228). New York, NY, USA: Association for Computing Machinery.

Singh, A., & Joachims, T. (2019). Policy learning for fairness in ranking. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc.

Sürer, O., Burke, R., & Malthouse, E. C. (2018). Multistakeholder recommendation with provider constraints. In *Proceedings of the 12th ACM Conference on Recommender Systems* RecSys'18 (p. 54–62). New York, NY, USA: Association for Computing Machinery.

Wang, Y., Ma, W., Zhang*, M., Liu, Y., & Ma, S. (2022). A survey on the fairness of recommender systems. *ACM Transactions of Infation Systems*, .

Wang, Y., Wang, L., Li, Y., He, D., & Liu, T.-Y. (2013). A theoretical analysis of ndcg type ranking measures. In *Proceedings of the 26th Annual Conference on Learning Theory* (pp. 25–54). Princeton, NJ, USA: PMLR.

Xiao, L., Min, Z., Yongfeng, Z., Zhaoquan, G., Yiqun, L., & Shaoping, M. (2017). Fairness-aware group recommendation with pareto-efficiency. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* RecSys '17 (p. 107–115). New York, NY, USA: Association for Computing Machinery.

Yang, X., Steck, H., Guo, Y., & Liu, Y. (2012). On top-k recommendation using social networks. In *Proceedings of the Sixth ACM Conference on Recommender Systems* RecSys '12 (p. 67–74). New York, NY, USA: Association for Computing Machinery.

Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., & Baeza-Yates, R. (2017). Fa*ir: A fair top-k ranking algorithm. In *Proceedings of the 2017*

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780

ACM on Conference on Information and Knowledge Management CIKM '17 (p. 1569–1578). New York, NY, USA: Association for Computing Machinery.

Zehlike, M., & Castillo, C. (2020). Reducing disparate exposure in ranking: A learning to rank approach. In *Proceedings of The Web Conference 2020* WWW'20 (p. 2849–2855). New York, NY, USA: Association for Computing Machinery.

Zehlike, M., Yang, K., & Stoyanovich, J. (2021). Fairness in ranking: A survey. URL: `https://arxiv.org/abs/2103.14000`.

Zhu, Z., He, Y., Zhao, X., Zhang, Y., Wang, J., & Caverlee, J. (2021). Popularity-opportunity bias in collaborative filtering. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* WSDM '21 (p. 85–93). New York, NY, USA: Association for Computing Machinery.

Ziani, A., Azizi, N., Schwab, D., Aldwairi, M., Chekkai, N., Zenakhra, D., & Cheriguene, S. (2017). Recommender System Through Sentiment Analysis. In *2nd International Conference on Automatic Control, Telecommunications and Signals*. Annaba, Algeria.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4274780