

# Examining the Impact of Multi-Objective Recommender Systems on Providers Bias

Reza Shafiloo<sup>1,\*</sup>, Kostas Stefanidis<sup>1</sup>

<sup>1</sup>Tampere University, Finland

## Abstract

Recommender systems are designed to help customers in finding their personalized content. However, biases in recommender systems can potentially exacerbate over time. Multi-objective recommender system (MORS) algorithms aim to alleviate bias while maintaining the accuracy of recommendation lists. While these algorithms effectively address item-side fairness, provider-side fairness often remains neglected. This study investigates the impact of MORS algorithms, leveraging evolutionary techniques to mitigate popularity bias on the item-side, on providers' fairness. Our findings reveal that baseline algorithms can adversely affect providers' fairness. Moreover, it is demonstrated that evolutionary algorithms, specifically those introducing less popular items to the initial population of their algorithms, exhibit superior performance compared to other MORS algorithms in enhancing providers' fairness. This research sheds light on the crucial role MORS algorithms, particularly those employing evolutionary approaches, can play in mitigating bias and promoting fairness for both users and providers in recommender systems.

## Keywords

Recommender systems, Items-side fairness, Producer-side fairness

## 1. Introduction

These days, with the increasing amount of information on the web, content providers need systems to personalize content for end-users. As a result, users can efficiently access their favorite content, leading to user satisfaction [1]. Recommender systems (RS) provide personalized content for users based on their historical interactions with systems, such as ratings or clicks on items. Despite being a crucial and valuable tool for users, RS has been identified as amplifying various biases. These biases can significantly impact the outcomes of RS, particularly concerning factors such as gender, age, race, and other characteristics. One such bias is popularity bias, where certain items typically receive a substantial number of ratings, leading to them being recommended more frequently than others.

Fairness-aware recommender systems aim to address algorithmic bias in various ways, ensuring the system's recommendations are unbiased [2]. Fairness-aware recommender systems can take into account various attributes to offer equitable recommendations. The concept involves evaluating how a recommender system treats or affects individuals or groups based on the values of specific attributes. Methods for ensuring fairness in RS can be categorized into pre-processing, which involves modifying input data [3]; in-processing, which constrains learning algorithms for fair recommendations [4]; and post-processing, which modifies the output of the baseline algorithm [5].

In RS, various stakeholders play crucial roles, with two primary groups being consumers of items and providers of items [6]. However, numerous fairness-aware RS focus on addressing consumer or producer-sided fairness, often neglecting comprehensive, all-sided multi-stakeholder fairness. While numerous studies concentrate on one-sided fairness in RS, it is essential to explore how addressing fairness for one group might impact the fairness of other stakeholders.

Using Multi-objective Recommender Systems (MORS) as

a post-processing approach offers a potential solution for achieving fairness in RS outputs [7]. Some existing MORS specifically address fairness for the item side. These approaches aim to maintain the accuracy of RS for consumer satisfaction while also creating opportunities for recommending less popular items, thereby mitigating popularity bias [8, 9, 10]. While preserving accuracy and enhancing fairness among items is valuable, it is crucial to investigate fairness among providers of items.

In this study, our objective is to investigate the behavior of MORS algorithms in mitigating item popularity bias and its impact on providers' fairness. While existing research has shown the trade-off between mitigating popularity bias and maintaining recommendation accuracy on the item side, it is crucial to delve deeper into how existing work objectives can influence providers' fairness. Prior research has yet to be conducted in this area, and our study aims to address this gap [2].

Furthermore, we aim to explore which specific objectives may have a trade-off with providers' fairness to provide a more comprehensive understanding of the issue. We have chosen MORS algorithms that benefit from evolutionary algorithms to solve a multi-objective optimization to achieve this. While evolutionary algorithms may not be the swiftest, their superiority in addressing multi-objective problems arises from their capability to tackle complex and non-linear optimization problems [8].

Our work shows that MORS algorithms perform better in ensuring providers' fairness than baseline algorithms. MORS algorithms enable providers to showcase their items more effectively than baseline algorithms. Although it is noteworthy that, among all MORS algorithms, there is no significant difference in covering providers' fairness, those algorithms that add less popular items to their initial population of evolutionary algorithms show better performance than other MORS algorithms.

The remainder of this paper is structured as follows. Section 2 reviews the related fairness in recommender systems. Section 3 describes the algorithms we use in our study and the measures we utilize to compare the algorithms' performance. Section 4 presents some results of our framework on the MovieLens and IMDB datasets. Section 5 concludes this work.

*Published in the Proceedings of the Workshops of the EDBT/ICDT 2024 Joint Conference (March 25-28, 2024), Paestum, Italy*

\*Corresponding author.

✉ shafiloo.reza@tuni.fi (R. Shafiloo); konstantinos.stefanidis@tuni.fi (K. Stefanidis)

🌐 <https://www.tuni.fi/fi/konstantinos-stefanidis> (K. Stefanidis)

📄 0000-0003-1317-8062 (K. Stefanidis)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## 2. Related Work

The fundamental RS aims to forecast ratings for unknown items among users, employing diverse algorithms for this task. Approaches like User-based and Item-based collaborative filtering algorithms, as explored by Adomavicius et al. [11] and Yue et al. [12], entail the identification of similar users or items to predict item ratings. CF algorithms can be used in many post-processing algorithms as baseline algorithms, from neural networks [13] to multi-objective evolutionary algorithms [9, 8, 10].

The investigation for an optimal balance between accuracy and bias mitigation has garnered significant attention in RS. Malekzadeh and Kaedi propose a strategy that simultaneously personalizes recommended items to maintain accuracy as discussed in their work [8]. Similarly, Wang et al. [10] address the long-tail problem by employing multi-objective evolutionary optimization algorithms, focusing on improving recommendation list accuracy and reducing the dominance of popular items. Shafiloo et al. [9] present a framework to alleviate popularity bias in recommender systems by incorporating users' dynamic preferences. Cai et al. [14] proposed a framework based on multi-objective algorithms designed to concurrently optimize accuracy, diversity, and coverage within recommendation lists. Utilizing multi-objective algorithms reflects their commitment to addressing multiple dimensions of recommendation quality, aiming to enhance the overall user experience. Jain et al. introduced a novel similarity metric tailored for baseline algorithms[15]. Their approach involved modifying fundamental functions of genetic algorithms, precisely the crossover operation, to effectively manage the trade-off between accuracy and diversity of recommended items. Pang et al. introduced a framework based on genetic algorithms, where accuracy and coverage serve as objective functions [16]. This innovative approach is designed to tackle popularity bias in recommendation lists, emphasizing a dual focus on improving accuracy and coverage for a more comprehensive and unbiased recommendation system.

Fairness-aware recommender systems try to tackle the algorithmic bias issue in different ways and ensure that the recommendations made by the system are unbiased [17]. However, many approaches consider tackling only one-sided fairness issues but abandon all-sided multi-stakeholder fairness [18]. In the realm of multi-stakeholder recommender systems (MS-RS), where numerous users participate in the recommendation process from multiple perspectives, as noted by Cornacchia et al. [19], there should be studies on how items side fairness how can affect another side of fairness.

## 3. Methods

In this section, our initial focus is to introduce the algorithms employed in our study. Subsequently, we will delve into the evaluation metrics utilized for comparing results. Our objective is to comprehensively explore the impact of item bias mitigation on the producers' side fairness and understand how it influences the outcomes of the recommendation systems.

### 3.1. Baseline algorithms

We have selected two baseline algorithms, item-based and user-based collaborative filtering, where no post-processing

has been applied to the output. These algorithms serve as our baseline models for evaluating bias mitigation strategies and their impact on the producers' side in subsequent analyses.

For computing the similarity between two users, we have:

$$sim(u, v) = \frac{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u) \cdot (r_{vk} - \mu_v)}{\sqrt{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u)^2} \cdot \sqrt{\sum_{k \in I_u \cap I_v} (r_{vk} - \mu_v)^2}} \quad (1)$$

Equation 1 defines the similarity measure between two users,  $u$  and  $v$ , calculated based on the items they have both rated. Here,  $I_u$  represents the subset of items rated by user  $u$ ,  $r_{uk}$  denotes the rating given by user  $u$  to item  $k$ , and  $\mu_u$  is the average rating provided by user  $u$ .

$$\hat{r}_{uj} = \mu_u + \frac{\sum_{v \in P_u(j)} sim(u, v) \cdot (r_{vj} - \mu_v)}{\sum_{v \in P_u(j)} |sim(u, v)|} \quad (2)$$

Equation 2 outlines the predicted rating ( $\hat{r}_{uj}$ ) of user  $u$  for item  $j$ . It incorporates the average rating  $\mu_u$  and calculates the predicted rating by considering the similarity between user  $u$  and other users ( $v$ ) who have rated the same item ( $j$ ). The set  $P_u(j)$  represents the group of nearest users to  $u$  who have provided ratings for item  $j$ . The item-based collaborative filtering is similar to the user-based collaborative filtering.

### 3.2. Multi-objective algorithms

In this section, we introduce algorithms that leverage the outputs of baseline algorithms, implementing reranking strategies to achieve specific objectives. Each algorithm is characterized by an objective function to mitigate item popularity bias.

Malekzadeh and Kaedi [8] employ the simulated annealing algorithm to address the long-tail problem in recommender systems. Their approach begins with applying a collaborative filtering algorithm to generate initial recommendation lists. Subsequently, an evolutionary algorithm is employed to optimize the combination of items in these lists, focusing on satisfying three defined objective functions. These functions encompass considerations for personalized diversification, accuracy, and increased participation of long-tail items, aiming to enhance recommendations' overall quality. The objective functions are:

1. **Diversity:** The Shannon entropy is used for diversity which the entropy  $H_a(u)$  for attribute  $a$  of user  $u$  is defined using the formula:

$$H_a(u) = - \sum_{i=1}^k p_i \cdot \log_k p_i \quad (3)$$

In this Equation:

$H_a(u)$  is the entropy for attribute  $a$  of user  $u$ .  $k$  is the number of possible values for attribute  $a$ .  $p_i$  represents the ratio of the number of ratings given by user  $u$  to items with attribute  $a$  having the value  $i$ , divided by the total number of user's ratings.

Essentially, this formula calculates the entropy of the distribution of ratings given by a user  $u$  across different values of attribute  $a$ .

The attribute-based diversity measurement in this study is determined using an equation to assess a recommendation list's diversity. The formula is expressed as:

$$\text{Diversity}_a(c_1, \dots, c_n) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=i}^n (1 - \text{similarity}_a(c_i, c_j)) \quad (4)$$

In this context,  $n$  signifies the number of items within the recommendation list, and  $c_1, \dots, c_n$  represents the items recommended. The term  $\text{similarity}_a(c_i, c_j)$  denotes the measure of similarity between two items  $c_i$  and  $c_j$  based on the attribute  $a$ .

Equation 3 illustrates the ideal diversity for a specific user, capturing an optimal scenario. Subsequently, the deviation between this ideal diversity and the actual diversity computed from Equation 4 for the recommendation list is measured. The disparity for each item attribute is quantified through Equation (7):

$$\text{Personalized Diversity} = |H_a - \text{Diversity}_a| \quad (5)$$

In this expression,  $\text{Diversity}_a$  denotes the diversity of the recommendation list based on attribute  $a$ , while  $H_a$  signifies the entropy of user preferences related to attribute  $a$ . This metric, termed *Personalized Diversity*, serves to quantify the difference between the ideal and actual diversity in the recommendation list for a given user, explicitly considering the preferences associated with a particular attribute.

2. **The participation of long tail items:** The long tail metric is computed using the formula:

$$\text{Long Tail} = \sum_{item=1}^k \text{Popularity}(item) \quad (6)$$

In this Equation,  $k$  signifies the size of the recommendation list, representing the total number of items included in the recommendation. A lower value obtained from this calculation indicates a higher likelihood of incorporating less popular items in the recommendation list. This suggests a greater emphasis on the inclusion of long-tail items in the recommendations, reflecting a preference for diversity and coverage beyond just popular items.

3. **Accuracy:** The Accuracy metric is evaluated using the following Equation:

$$\text{Accuracy} = \frac{1}{\sum_{item=1}^k \text{PredictedRate}(item)} \quad (7)$$

In this Equation,  $\text{PredictedRate}(item)$  denotes the predicted rating assigned to the item. The formula computes the inverse of the sum of the predicted ratings for all recommended items, offering a metric to assess the accuracy of the recommendation system. A lower value in the Accuracy metric suggests a higher overall accuracy in the predicted ratings for the recommended items.

Wang et al. [10] address the long-tail problem by defining two objective functions. The first function assesses the accuracy of recommendation lists, while the second aims to reduce the dominance of popular items. The objectives formula can be expressed as follows:

1. **Accuracy:** The primary objective function for assessing accuracy, labeled as  $F1$ , is formulated as:

$$F1 = \sum_{i=1}^k \hat{r}_{u,i} \quad (8)$$

In this expression,  $k$  denotes the length of the recommendation list. A higher  $F1$  value signifies increased popularity of the items within the list.

2. **Long tail recommendation:** Items with higher ratings might be prioritized higher on the ranking list for all users, and popular items often receive similar ratings, resulting in low variance. To measure the unpopularity in terms of the mean and variance of item ratings, Tamas et al. proposed a value for an item  $i$ :

$$m_i = \frac{1}{\mu_i(\sigma_i + 1)^2} \quad (9)$$

Here,  $\mu_i$  and  $\sigma_i$  represent the mean and variance of ratings for item  $i$  across all users. To prevent division by zero, a value of one is added to the variance. The reciprocal of this mean-variance combination yields the value  $m_i$ , where a smaller value indicates a more popular item.

Motivated by this concept, an objective function  $F2$  is introduced to calculate the unpopularity of the recommendation result:

$$F2 = \sum_{i=1}^k \frac{1}{\mu_i(\sigma_i + 1)^2} \quad (10)$$

This function quantifies the unpopularity of the recommended items, with lower values indicating more popular items in the list.

They employ a genetic algorithm to achieve these objectives, seeking optimal combinations of items within recommendation lists that satisfy the defined criteria. This approach aims to enhance accuracy and mitigate popularity bias for more balanced and practical recommendations.

Shafiloo et al. [9] introduced a framework to alleviate popularity bias in recommender systems by incorporating users' dynamic preferences. Their approach employs a memetic algorithm, creating opportunities to include unpopular items in recommendation lists. They define two objective functions within their framework, aiming to simultaneously preserve accuracy and mitigate popularity bias. This innovative solution focuses on providing more diverse and unbiased recommendations, catering to the dynamic preferences of users. The objectives to be achieved are:

1. **Accuracy:** In their research, they employ accuracy as expressed in formula 7.
2. **Long tail participation:** They utilize long tail participation as described in formula 6.

Additionally, in their research, they modified the memetic algorithm. Instead of randomly adding items to the initial population, as is common in other genetic algorithms, they introduced a higher possibility of including items from the long tail and a lower possibility of including popular items in the initial population.

## 4. Experimental Evaluation

This section presents the datasets employed for evaluating the proposed method. Subsequently, we outline the evaluation criteria and comprehensively represent the comparison result.

### 4.1. Dataset

In our experimental evaluation, we use 2 real-world datasets, namely MovieLens and IMDB. The MovieLens dataset is a commonly employed dataset for evaluating methods addressing long-tail problems in various studies. Specifically, we utilize the MovieLens 1M dataset that features 6040 users and 1 million ratings for 3883 items. The IMDB Dataset is also employed to enhance information about movie providers, and director information is extracted. In this study, movie directors are considered providers, and the dataset includes information on 2208 movie directors.

### 4.2. Evaluation metric

The study evaluates methods addressing the long-tail problem using three criteria for comparison. The first criterion is accuracy, measured through the precision metric defined as:

$$Precision = \frac{N_{rs}}{N_s} \quad (11)$$

Here,  $N_s$  represents the total number of items recommended to the user, and  $N_{rs}$  denotes the relevant items suggested to the user. Relevant items are those with ratings higher than the user's average ratings, as outlined by Wang et al. [10].

The second criterion, aggregate diversity (AG) (Equation 12), counts the number of distinct items offered to users, particularly focusing on long-tail items in recommendation lists [8].

$$Aggregate\ diversity = \left| \bigcup_{u \in U} L_n(u) \right| \quad (12)$$

Equation 12 introduces the aggregate diversity criterion, where  $u$  represents a specific user from the set of users  $U$ , and  $L_n(u)$  is the list of items recommended to the user  $u$ . The equation 12 is normalized by the number of items. This equation is used to measure popularity bias on the item and provider sides.

The third criterion is Novelty, calculated as:

$$Novelty = \frac{1}{\sum_{all\ recommended\ items} Popularity(items)} \quad (13)$$

This Equation indicates that the novelty of the recommendation list decreases as the popularity of items increases, emphasizing a preference for less popular items. The study employs these criteria to compare and evaluate the results of different methods addressing the long-tail problem in recommender systems.

Equation 14 introduces a measurement for intra-user diversity proposed by Zou et al. [20]. This measurement, denoted as  $D_u(k)$ , is defined for a specific user  $u$  and is calculated as follows:

$$D_u(k) = \frac{1}{k(k-1)} \sum_{p \neq q} Sim(i_p, i_q) \quad (14)$$

Here,  $k$  represents the length of the recommendation lists for user  $u$ , and  $Sim(i_p, i_q)$  calculates the similarity between two items  $i_p$  and  $i_q$  based on a similarity metric defined in Equation 1. The purpose of  $D_u(k)$  is to quantify the similarity of items within user  $u$ 's recommendation list.

The intra-user diversity for all users is then defined as:

$$D_{all\ users}(k) = \frac{1}{m} \sum_{u \in U} D_u(k) \quad (15)$$

Here,  $m$  denotes the number of users in the set  $U$ . This Equation provides a measure of intra-user diversity considering all users in the study.

Equation 16 introduces the Normalized Discounted Cumulative Gain (NDCG) measurement, a widely used metric for evaluating the quality of recommendations. This measurement is defined as:

$$NDCG@k(u) = \frac{DCG@k(u)}{IDCG@k(u)} \quad (16)$$

Here,  $IDCG@k(u)$  represents the ideal  $DCG@k(u)$  for user  $u$ , where the ideal scenario assumes that all relevant items in the user's recommendation list appear at the top rank, resulting in the maximum possible  $DCG@k(u)$ .

The discounted cumulative gain at position  $k$  for user  $u$ , denoted as  $DCG@k(u)$ , is calculated using the formula:

$$DCG@k(u) = \sum_{i=1}^k \frac{rel(i)}{\log_2(i+1)} \quad (17)$$

In this Equation,  $rel(i)$  is an indicator function that determines if item  $i$  is relevant to user  $u$ . A value of 1 indicates that item  $i$  is relevant, while a 0 indicates that item  $i$  is irrelevant.

NDCG provides a normalized measure of the effectiveness of a recommendation list by considering both relevance and the position of items within the list.

### 4.3. Results and discussion

In this section, the study compares and analyzes the results obtained from various methods using the criteria introduced in the section above. For comparison, we use a real life scenario where the length of recommendation lists in all algorithms is considered to be 10.

The results in Table 1 indicate that MORS algorithms outperform baseline algorithms in addressing the long-tail problem. These algorithms demonstrate superior performance in diversifying items in recommendation lists, effectively increasing the participation of unpopular items. Notably, the study highlights that MORS algorithms achieve this diversification without compromising the accuracy of the recommendation lists. Therefore, the MORS algorithms are successful in preserving accuracy while simultaneously enhancing the inclusion of less popular items in the recommendations, addressing the long-tail problem in recommender systems.

**Table 1**

Comparison of Results: Evaluation measures and outcomes for various algorithms. Accuracy measures include precision, NDCG and diversity and fairness measures include Novelty, AG-Items, and AG-providers.

Algorithms	Evaluation metrics					
	Precision	Novelty	AG-Items	Diversity	NDCG	AG-Providers
CF-Item	0.7163	3.23E-05	0.5318	0.3873	0.5906	0.5067
CF-Users	0.6622	3.26E-05	0.5449	0.4051	0.7938	0.5230
Malekzadeh [8] (MORS)	<b>0.8338</b>	3.72E-05	0.6651	0.6221	0.9381	0.5711
Wang [10] (MORS)	0.6968	3.61E-05	0.6422	0.6200	0.9173	0.5697
Shafiloo [9] (MORS)	0.7989	<b>4.02E-05</b>	<b>0.6930</b>	<b>0.7396</b>	<b>0.9599</b>	<b>0.6059</b>

The comparison table suggests that while MORS algorithms effectively mitigate popularity bias in recommendation lists, there is not a significant difference in the diversity of providers between baseline algorithms and MORS algorithms. For example, CF-User has a value of 0.5230 in AG-providers, while Malekzadeh and Wang show 0.5697 and 0.5711, respectively. Although MORS algorithms, aided by item diversifying objectives, offer providers a better chance to present their items, the disparity in aggregate diversity is more noticeable on the item side than on the provider side when comparing MORS algorithms with baseline algorithms. Moreover, the comparison indicates that Baseline algorithms with higher accuracy than MORS algorithms exhibit poor performance in aggregate diversity, suggesting that recommendation list accuracy can negatively impact provider-side fairness. Specifically, CF-items achieve an accuracy of 0.7163, whereas CF-Users attain 0.6622. However, AG-providers exhibit respective values of 0.5067 and 0.5230.

Also, Table 1 indicates that among MORS algorithms, Malekzadeh’s work outperforms Shafiloo and Wang’s work in terms of the precision metric. However, this superiority adversely impacts aggregate diversity on both the provider and item sides. Specifically, Shafiloo’s work exhibits a precision of 0.7989, with aggregate provider diversity at 0.6059 and aggregate item diversity at 0.6930. In contrast, Malekzadeh’s work achieves a precision of 0.8338, but the aggregate provider diversity decreases to 0.5711, and the aggregate item diversity is 0.6651.

Furthermore, in Figure 1, we present the provider frequency using a bucketing technique. Specifically, in this figure, providers are assigned to a bucket based on the number of items belonging to that specific provider that are represented in all recommendation lists generated by the algorithm. For instance, a provider is placed in bucket one if only one item from all items associated with that provider is present in all recommendation lists.

This figure shows that baseline algorithms exhibit a weakness in recommending items from providers who lack popularity. This is illustrated in the initial buckets, where baseline algorithms struggle to include more items from less famous providers. Conversely, the first part of the buckets shows that MORS algorithms provide a more significant opportunity for less-known providers to showcase their items in the recommendation lists, offering more visibility.

## 5. Conclusions

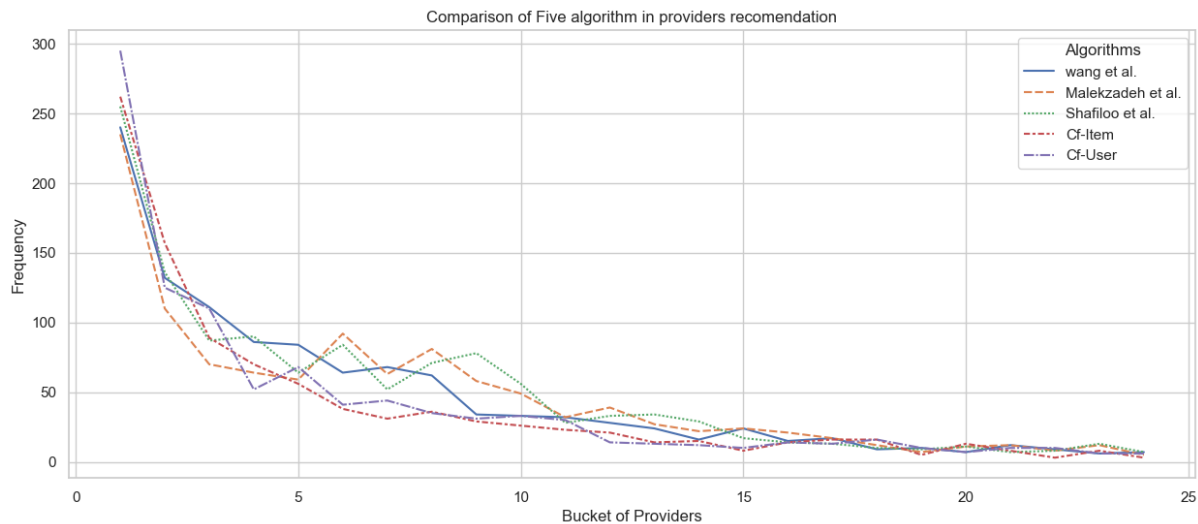
In conclusion, our study highlights the significance of MORS algorithms in addressing the issue of bias in recommender systems and promoting fairness for both items and providers. Our findings reveal that while baseline algorithms can negatively impact the provider’s fairness, MORS algorithms,

particularly those leveraging evolutionary techniques and introducing less popular items to the initial population of their algorithms, can effectively mitigate popularity bias and enhance the provider’s fairness. This emphasizes the importance of considering provider-side fairness in the development of recommender systems, as it is often neglected in current research.

Overall, our research contributes to the growing body of work on fairness and bias in recommender systems and emphasizes the crucial role of MORS algorithms, particularly those employing evolutionary approaches, in mitigating bias and promoting fairness for both items and providers. Our study provides insights into how existing work objectives can influence provider fairness. It highlights the need for future research to delve deeper into this issue to provide a more comprehensive understanding of the problem. The effectiveness of MORS algorithms for providers could be further enhanced if a specific objective function were dedicated to mitigating provider bias. The absence of such an objective function might limit the algorithms’ ability to address biases related to the popularity of providers in the recommendation process.

## References

- [1] M. Stratigi, E. Pitoura, K. Stefanidis, Squirrel: A framework for sequential group recommendations through reinforcement learning, *Information Systems* 112 (2023) 102128. URL: <https://www.sciencedirect.com/science/article/pii/S0306437922001065>. doi:<https://doi.org/10.1016/j.is.2022.102128>.
- [2] E. Pitoura, K. Stefanidis, G. Koutrika, Fairness in rankings and recommendations: an overview, *The VLDB Journal* (2022) 1–28.
- [3] B. Salimi, L. Rodriguez, B. Howe, D. Suci, Interventional fairness: Causal database repair for algorithmic fairness, in: *Proceedings of the 2019 International Conference on Management of Data*, 2019, pp. 793–810.
- [4] Z. Zhu, X. Hu, J. Caverlee, Fairness-aware tensor-based recommendation, in: *Proceedings of the 27th ACM international conference on information and knowledge management*, 2018, pp. 1153–1162.
- [5] T. Kamishima, S. Akaho, H. Asoh, J. Sakuma, Recommendation independence, in: *Conference on fairness, accountability and transparency*, PMLR, 2018, pp. 187–201.
- [6] G. Giannopoulos, G. Papastefanatos, D. Sacharidis, K. Stefanidis, Interactivity, fairness and explanations in recommendations, in: *ACM UMAP*, 2021, pp. 157–161.
- [7] L. Xiao, Z. Min, Z. Yongfeng, G. Zhaoquan, L. Yiqun,



**Figure 1:** Figure illustrates provider frequency distribution using a bucketing technique. Providers are categorized based on the number of items they contribute to all recommendation lists. This visualization offers insights into provider diversity and prevalence within recommendation systems.

- M. Shaoping, Fairness-aware group recommendation with pareto-efficiency, in: *Proceedings of the eleventh ACM conference on recommender systems*, 2017, pp. 107–115.
- [8] E. M. Hamedani, M. Kaedi, Recommending the long tail items through personalized diversification, *Knowledge-Based Systems* 164 (2019) 348–357. doi:10.1016/j.knosys.2018.11.004.
- [9] R. Shafiloo, M. Kaedi, A. Pourmiri, Considering user dynamic preferences for mitigating negative effects of long tail in recommender systems, *arXiv preprint arXiv:2112.02406* (2021).
- [10] S. Wang, M. Gong, H. Li, J. Yang, Multi-objective optimization for long-tail recommendation, *Knowledge-Based Systems* 104 (2016) 145–155. doi:10.1016/j.knosys.2016.04.018.
- [11] G. Adomavicius, Y. Kwon, Improving aggregate recommendation diversity using ranking-based techniques, *IEEE Transactions on Knowledge and Data Engineering* 24 (2011) 896–911. doi:10.1109/TKDE.2011.15.
- [12] W. Yue, Z. Wang, W. Liu, B. Tian, S. Lauria, X. Liu, An optimally weighted user-and item-based collaborative filtering approach to predicting baseline data for friedreich’s ataxia patients, *Neurocomputing* 419 (2021) 287–294.
- [13] R. Borges, K. Stefanidis, Feature-blind fairness in collaborative filtering recommender systems, *Knowledge and Information Systems* 64 (2022) 943–962.
- [14] X. Cai, Z. Hu, P. Zhao, W. Zhang, J. Chen, A hybrid recommendation system with many-objective evolutionary algorithm, *Expert Systems with Applications* 159 (2020) 113648.
- [15] A. Jain, P. K. Singh, J. Dhar, Multi-objective item evaluation for diverse as well as novel item recommendations, *Expert Systems with Applications* 139 (2020) 112857.
- [16] J. Pang, J. Guo, W. Zhang, Using multi-objective optimization to solve the long tail problem in recommender system, in: *Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14-17, 2019, Proceedings, Part III* 23, Springer, 2019, pp. 302–313.
- [17] C.-T. Li, C. Hsu, Y. Zhang, Fairsr: Fairness-aware sequential recommendation through multi-task learning with preference graph embeddings, *ACM Transactions on Intelligent Systems and Technology (TIST)* 13 (2022) 1–21.
- [18] H. Wu, C. Ma, B. Mitra, F. Diaz, X. Liu, A multi-objective optimization framework for multi-stakeholder fairness-aware recommendation, *ACM Transactions on Information Systems* 41 (2022) 1–29.
- [19] G. Cornacchia, F. M. Donini, F. Narducci, C. Pomo, A. Ragone, Explanation in multi-stakeholder recommendation for enterprise decision support systems, in: *International Conference on Advanced Information Systems Engineering*, Springer, 2021, pp. 39–47.
- [20] F. Zou, D. Chen, Q. Xu, Z. Jiang, J. Kang, A two-stage personalized recommendation based on multi-objective teaching-learning-based optimization with decomposition, *Neurocomputing* 452 (2021) 716–727.