Towards Optimizing Ranking in Grid-Layout for Provider-side Fairness*

Amifa Raj[1] [0000−0002−8874−6645] and Michael D. Ekstrand[2] [0000−0003−2467−0108]

[1] Microsoft, USA
[2] Dept. of Information Science, Drexel University, USA
amifaraj@u.boisestate.edu
mdekstrand@drexel.edu

Abstract. Information access systems, such as search engines and recommender systems, order and position results based on their estimated relevance. These results are then evaluated for a range of concerns, including provider-side fairness: whether exposure to users is fairly distributed among items and the people who created them. Several fairness-aware ranking and re-ranking techniques have been proposed to ensure fair exposure for providers, but this work focuses almost exclusively on linear layouts in which items are displayed in single ranked list. Many widely-used systems use other layouts, such as the grid views common in streaming platforms, image search, and other applications. Providing fair exposure to providers in such layouts is not well-studied. We seek to fill this gap by providing a grid-aware re-ranking algorithm to optimize layouts for provider-side fairness by adapting existing re-ranking techniques to grid-aware browsing models, and an analysis of the effect of grid-specific factors such as device size on the resulting fairness optimization. Our work provides a starting point and identifies open gaps in ensuring provider-side fairness in grid-based layouts.

1 Introduction

Information access systems (IAS) — search engines, recommender systems, and similar — provide utility to their users, by retrieving relevant results, but also to the *providers* of the items (authors, artists, etc.) by exposing them to users who may read, purchase, or otherwise consume their creations. Providers receive both economic and reputational benefit from this exposure; however, a

* This paper reports work supported by the National Science Foundation under Grant 17-51278 and primarily conducted while the authors were at Boise State University. This version of the contribution has been accepted for publication, after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is published by Springer in *Proceedings of the 46th European Conference on Information Retrieval* (https://doi.org/10.1007/978-3-031-56069-9_7). Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use: https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms

system may not always fairly allocate exposure to items, as some items may receive less exposure than others of similar relevance [26, 38]. This *disparate exposure* [38] may lead to unfair outcomes for item providers on either individual or group bases. Item providers or producers are often associated with sensitive or *protected* group attributes such as race, gender, religion, age, and other demographic attributes and items with similar relevance may receive unfair exposure in ranking based on their group membership [24, 47]. Provider-side fairness in ranking seeks to correct this imbalance and ensure the fair allocation of exposure across item producers and providers in system results. Several re-ranking techniques have been proposed to improve result fairness [25, 27, 33, 34, 38], and the TREC Fair Ranking track [32, 49] provided multiple tasks for which participants optimized their systems for both fairness and relevance. However, these efforts are limited to linear ranked lists, while many production IAS display results in grid-based layouts. The problem of optimizing ranking in grid layouts for fairness has received limited attention so far. Chen *et al.* [46] proposed one of the few grid-based re-ranking techniques, but did not consider fairness; re-ranking technique suitable for optimizing provider-side group fairness in grid layouts are still unknown.

Moreover, fair ranking metric scores vary depending on user browsing behavior and user browsing behavior varies across ranking layouts [37, 48, 52]. Hence, for the same set of ranked items, user attention can vary for the items depending on how they are displayed to users, resulting in different degrees of fair exposure to their providers. There is limited research on user browsing behavior in grid layouts, but the browsing models that are available have not yet been incorporated into ranking or layout strategies for fairness. Raj & Ekstrand [52] showed that a ranking that is optimized for fairness in linear layout may not preserve its fairness when rearranged into grid layout. Moreover, the geometry of grid layouts change depending on the user's device, as the number of columns changes with screen size. Therefore, fair grid layouts need to consider both suitable browsing models and the specific layout geometry to boe used. Our work helps fill this gap by providing the first re-ranking technique to optimize provider-side group fairness in grid layouts.

We adapt a commonly used re-ranking techniques from linear layouts and modify it for grid layout by incorporating grid-aware browsing models. Since designing fairness-aware re-ranking techniques for ranking in grid layouts depends on ranking design, user browsing behavior, and column size or user device, we study the impact of column sizes and browsing models on our method.

Our experimentals addrss the following research questions:

- **RQ1.** Does incorporating grid-aware browsing models to existing re-ranking technique improve fairness for results in grid layouts?
- **RQ2.** Does a ranking in a grid layout optimized for fairness on one device remain fair for other devices?
- **RQ3.** How can we optimize ranking in grid layouts for various screen sizes?

Our simple and re-configurable re-ranking approach for grid layout advances provider-side group fairness in IAS beyond simple linear ranked interfaces. As more specific user attention models are developed in the future for ranking in grid layouts, they can be plugged into our proposed method to provide more accurate fairness optimization. Our analyses also provide an initial guidance for practitioners to design more fine-tuned re-ranking approaches for grid layouts that may consider item metadata, tasks, and domain, and open several future research directions towards fairness concerns in additional layouts and variants.

2 Background and Related Work

This section provides background on grid-layout suitable user browsing models and fairness-aware re-ranking techniques in IAS.

2.1 User Browsing Models

Users do not provide equal attention to every position in ranked results [26], item at the lower position of a ranked list will not receive similar attention as the item at the top position. Since user attention varies across positions in ranked results, the position weight for each position in ranking depends on how users browse the displayed ranked results.

There are several user browsing models to demonstrate user browsing behavior in linear ranked lists. *Cascade* [11] and *geometric* [12] are two popularly used user browsing models to infer the probability of user visiting an item in a particular position in ranking. These models differ in their underlying components and parameter settings but can be cast as different configurations of the same model. In both geometric and cascade models, user attention or position weight decays exponentially with ranking positions but in cascade browsing model, user selection probability is a function of item relevance.

To implement grid layout-aware evaluation metrics, it is important to understand how users provide attention to items in grid layout or how user attention changes across items when they are displayed in a grid layout. Tatler [9] observed the *central fixation* tendency where users provide more attention at the center of the page but Djamasbi *et al.* [17] and Zhao *et al.* [22] found that users usually show an *F-shaped* reading pattern by focusing on the results located at the top left-hand side. The viewing pattern is dependent on task, content, and complexity of the web pages [8]. Xie *et al.* [23] showed various user browsing behaviors in grid layout in e-commerce search results and they showed that users show *row-skipping*, *slower-decay*, and *middle-bias* while browsing items in grid layout. Users often skip rows while browsing ranked results in a grid layout and they tend to show higher attention to the middle position of columns in ranking. Moreover, user attention decays slowly across items in grid layout than linear list. Raj & Ekstrand [52] provided modified versions of geometric and cascade browsing models incorporating grid layout-suitable *row-skipping* and *slower-decay* user browsing behaviors and they implemented fair ranking metrics in grid layouts by incorporating grid-aware browsing models.

2.2 Re-Ranking Techniques

Several re-ranking and *learning to rank* (LTR) approaches have been proposed to optimize ranking for utility [3, 5–7, 13, 14, 16, 27, 34, 35]. LTR methods learn to rank based on scoring functions which is used to determine an optimized ranking; individual items, list, or pair of ranked items are considered to measure loss function against ideal ranking. Depending on the design of the loss function, the LTR approaches are categorized into pairwise, point-wise, and list-wise approaches [30, 35]. Pairwise approaches are often based on the change in ranking quality with the swap of each pair of items in ranking [15]. In RankNet [16] and LambdaMART [5], ranking quality is optimized by predicting an optimal ordering for each pair of items in ranked list before generating the final ranking. In point-wise optimization approaches, the ranking model is trained to minimize loss function determined from each individual item score [7]. In this approach, each of the item in candidate set is scored independently based on the target quality. Unlike previous two approaches, list-wise approaches consider the entire ranked list and the ranking function is trained on the entire list based on the minimization of the loss function [6, 10, 21].

Fairness-aware Re-ranking Techniques Fairness optimization in ranking often involves trade-offs between utility and fairness score [19, 28, 31, 38], where fairness-aware LTR and re-ranking approaches aim to improve fairness with minimum utility loss. Approaches to improve the fairness of algorithms, including IAS rankings, are often categorized into *pre-processing*, *in-processing*, and *post-processing* (typically re-ranking) techniques [44]. In pre-processing approaches, the potential bias in datasets or training labels are investigated in order to identify and mitigate bias in ranking [41, 45]. In in-processing approaches, the IAS algorithms or models are adjusted to optimize for fairness or a combination of fairness and utility in the training phase [31, 42]. Post-processing approaches take already-ranked results and reorder them to improve or optimize a fairness objective [24]. Constraint optimization approaches have also been proposed to re-rank results [28, 34, 38]; the optimization constraints often include both user satisfaction metrics and fair ranking metrics to preserve a balance between fairness and utility.

Various fair ranking metrics are used to measure fairness in ranking and to determine the target fairness score. Provider-side fairness in ranking is often measured by the discrepancy in between the expected exposure and the exposure providers receive from ranking [26, 28, 38, 48]. Hence, fairness-aware re-ranking techniques consider the optimization of the fairness score derived from the fairness metrics. Liu *et al.* [34] proposed a personalized fairness-aware re-ranking algorithm for micro-lending recommendations where each item from the initial ranking will be assigned to a position in the displayed ranking based on the optimization or maximization of personalization and group fairness. Singh & Joachims [28] and Diaz *et al.* [38] considered exposure of provider-side in ranking in their fairness-aware ranking optimization techniques. However, all the ap-

Table 1. Summary of notation.

$d \in D$	document or item
$q \in Q$	request (user or context)
L	ranked results of N items from D
$L(\mathrm{i})$	the item in position i of linear (1-column) layout
$L^{-1}(d)$	rank of item d in linear layout
$L(\mathrm{k}, \cdot)$	items in kth row in grid layout
$L(\mathrm{k}, \mathrm{c})$	items in row k and column c in grid layout
$y(d\|q)$	relevance of d to q
E_i	event: user examines the item at position i
S_i	event: user selects the item at position i
A_i	event: user abandons the process after examining the item at position i
K_k	event: user skipping the kth row.

proaches discussed above are proposed and implemented in linear ranked results when items are displayed in single-column list.

In this work, we modify a pairwise swap re-ranking technique to optimize ranking in grid layout for provider-side group fairness.

3 Problem Formulation and Proposed Approach

In this work, we consider a recommender system that recommends n items $d_1, d_2, \ldots, d_n \in D$ in response to information requests from users $q_1, q_2, \ldots, q_m \in Q$ based on their relevance to the request $y(d|q)$ and presents the results in a wrapped grid layout L (notation summarized in Table 1). Items are associated with producers or providers who in turn can be associated with demographic attributes identifying them with one or more of g groups. We model group membership of documents with group alignment vector $\mathcal{G}(d) \in [0, 1]^g$ (s.t. $\|\mathcal{G}(d)\|_1 = 1$) forming a distribution over groups; this allows for mixed, partial, or uncertain membership in an arbitrary number of groups.

Wrapped grid layouts — in which a single ranking is laid out in a grid by filling each row before wrapping to the next — are not the only grid-based layout [52]; many systems such as streaming video platforms use *multi-list* grid layouts where each row is a separate list of recommendations, possibly produced by a different algorithm. We focus on wrapped layouts in this paper because there has not yet been sufficient research on user browsing behavior in multi-list layouts to produce the browsing models needed for fairness-aware re-ranking.

The ranking will be optimized for provider-side group fairness while preserving a balance between utility and fairness. The purpose of this work is to provide a preliminary approach to develop fairness-aware re-ranking techniques for ranking in grid layout, so we focus on fairness optimization in a single-ranking setting, leaving fair grid layouts in stochastic settings for future work. We use the same grid-layout suitable browsing models to measure utility and fairness for consistency.

3.1 Layout Objective

Our layouts strive to provide both fairness and utility; both are measured with a browsing model that accounts for grid-specific browsing behaviors instead of the simple linear models typically used in fairness and utility metrics.

User Browsing Model In this work, we use the modified versions of geometric browsing models incorporating grid layout-suitable *row-skipping* and *slower-decay* user browsing behaviors provided by Raj & Ekstrand [52]. We use the generalized and configurable framework [51] of user browsing models that adapts *row-skipping* and *slower-decay* user behavior in *geometric* browsing model to measure position weight in grid layout. For a given ranking in grid layout, the visiting probability of item d in geometric-based row-skipping model is:

$$P_{RS(\text{geometric})}[V_d] = \left[\prod_{k=0}^{\text{row}(d)} (1-\gamma) \prod_{i \in L(\text{k},\cdot)} (1-\psi) + \prod_{k=0}^{\text{row}(d)} \gamma \right] \prod_{i \in \text{row}(d)} (1-\psi) \quad (1)$$

and the geometric visiting probability of item d with slower decay is:

$$P_{SD(\text{geometric})}[V_d] = \min(\beta^{\text{row}(d)} \prod_{i=[0,L^{-1}(d)]} (1-\psi), 1) \quad (2)$$

Target Fairness To measure provide-side group fairness in single ranking layout, we follow recommendations from the comprehensive analysis of fair ranking metrics in [48] and use AWRF. Sapiezynski *et al.* [36] proposed attention-weighted rank fairness or AWRF which measures the difference between group exposure and configurable target distribution $\hat{\mathbf{p}}$ which represents the ideal exposure distribution over groups. Attention vector and the group alignment matrix is used to derive group exposure $\boldsymbol{\epsilon}_L$ ($\boldsymbol{\epsilon}_\text{L} = \mathcal{G}(\text{L})^\text{T}\mathbf{a}_\text{L}$) by aggregating the attention given to items of each group in proportion to their group membership as represented by the alignment vector. Since our distribution difference function in bounded by $[0, 1]$, we invert it so that AWRF $= 1$ at maximal fairness to be more directly comparable to the effectiveness metrics:

$$\text{AWRF}(L) = 1 - \Delta(\boldsymbol{\epsilon}_L, \hat{\mathbf{p}}) \quad (3)$$

Target Utility To measure utility in ranking, we consider an effectiveness metric that consider items position weights in measurement. Moffat & Zobel [12] proposed *rank-biased precision* (RBP) which combined a geometric browsing model with binary relevance to measure the overall effectiveness of a ranking in a manner similar to nDCG, but with a re-configurable browsing model. The source of relevance can be the actual relevance judgement which generates RBP or system estimated relevance which generates $\hat{\text{RBP}}$. For a given ranking L, the rank-biased precision metric score is

$$\hat{\text{RBP}} = \psi \sum_{i=[0,L^{-1}(d)]} y(L(i)|q)\dot{(}1-\psi)^{i-1} \tag{4}$$

where $y(L(i)|q)$ is the systems estimated relevance score for the item in position i and the stopping probability ψ is decaying exponentially with ranking position. This metric can be adapted to measure $\hat{\text{RBP}}$ in grid layout by incorporating grid layout suitable browsing behavior. Thus, we modify the attention model used in this metric by considering geometric-based row-skipping model (equation 1) and geometric-based slower-decay (equation 2).

3.2 Re-Ranking Algorithm

Pairwise swapping re-ranking is a commonly used post-processing approach that we adapt to optimize ranking in grid layout for provider-side group fairness. For a given initial ranking L, we optimize the ranking by considering alternative ranking position for each pair of ranked items and finally generate a fairness-aware ranked result L'. Starting from the top of the list, for each position i, we consider each potential swap with positions $j > i$, items swap their position and temporarily generate a new ranking $L_{i\longleftrightarrow j}$ keeping all the other items at the same place. Then we measure the lift in fairness as $\Delta\text{AWRF}(L, L_{i\longleftrightarrow j})$ and the loss in utility as $\Delta\text{RBP}(L, L_{i\longleftrightarrow j})$.

$$\Delta\text{RBP}(L, L_{i\longleftrightarrow j}) = \text{RBP}(L_{i\longleftrightarrow j}) - \text{RBP}(L) \tag{5}$$

$$\Delta\text{AWRF}(L, L_{i\longleftrightarrow j}) = \text{AWRF}(L_{i\longleftrightarrow j}) - \text{AWRF}(L) \tag{6}$$

Thus for each of the position i, we select the best swap by solving the maximization of lift function, $F(i \leftrightarrow j | i, j \in \{1, ..., N\}, i < j)$:

$$F(i \leftrightarrow j) = \arg\max_{j \in i,\ldots,N} \{\Lambda\Delta\text{AWRF}(L, L_{i\longleftrightarrow j}) \cdot (1-\Lambda)\,(1 - \Delta\text{RBP}(L, L_{i\longleftrightarrow j}))\} \tag{7}$$

Algorithm 1 shows the formal algorithm for optimizing grid-ranking for provider-side group fairness. In each iteration, the item in position i is temporarily swapped with items that are in higher position than i and for each respective swap, it measures the AWRF improvement and inverse RBP loss. The swap that gives the maximum lift in fairness score with minimum utility loss is selected to generate a new ranking. Λ is used as a configurable balancing factor between fairness and utility.

4 Experimental Setup

We now present an experiment with real-world IAS dataset to observe whether and how the provider-side group fairness improves with our modified re-ranking techniques, addressing the research questions laid out in the introduction.

Algorithm 1 Fairness-Aware Re-ranking for Grid Ranking

Require: initial ranking L, user q, estimated relevance score $y(L|u)$, balancing factor Λ
Ensure: Re-ranked L'

procedure RE-RANK(L)
  $L' \leftarrow L$
  measure AWRF(L)
  measure RBP(L)
  for $i \in 1, ..., N$ **do**
    for $j \in i, ..., N$ **do**
      swap items in position i and j to generate $L_{i \leftrightarrow j}$
      measure ΔRBP$(L, L_{i \leftrightarrow j})$
      measure ΔAWRF$(L, L_{i \leftrightarrow j})$
    end for
    $i' = \arg\max_{j \in i,...,N} \{\Lambda \Delta \mathrm{AWRF}(L, L_{i \leftrightarrow j}) \cdot (1 - \Lambda)\,(1 - \Delta RBP(L, L_{i \leftrightarrow j}))\}$
    if $i' \neq i$ **then**
      $L^{'} \leftarrow L_{i \leftrightarrow i'}$
      AWRF$(L) \leftarrow$ AWRF$(L_{i \leftrightarrow i'})$
      RBP$(L) \leftarrow$ RBP$(L_{i \leftrightarrow i'})$
    end if
  end for
  return L'
end procedure

4.1 Data and Algorithms

In this work, we use *GoodReads* [29] book dataset integrated with the PIReT Book Data Tools [40] to obtain author metadata. This data records interactions from 870K users with 1.1M books. Consistent with the prior research using this data set [40, 43], we used LensKit [39] to generate 1000 personalized book recommendations for 5000 test users with four implicit-feedback collaborative filtering (CF) algorithms as configured by Ekstrand & Kluver [40]: user-based CF (UU [2]), item-based CF (II [4]), matrix factorization (WRLS [18]), and Bayesian Personalized Ranking (BPR [15]). The fairness goal is to be fair to the book authors of different genders;[3] the data contains 177K books by women and 283K books by men, with other books having unknown author gender.

4.2 Experiment Design

We optimize provider-side group fairness in grid layout using the modified re-ranking technique considering two types of user browsing models. We also observe the affect of column sizes on fairness optimization in grid layout.

[3] Due to limitations of the underlying data set[40], we are only able to consider binary gender. We understand the potential harm of misrepresentation of gender in research [50]; our methods in this paper are extensible to non-binary gender or other attributes when suitable data is available.

RQ1. *Improvement of Fairness in Grid Layout* To observe the group fairness score improvement for provide-side fairness in grid layout,

- We implement the fair ranking metric AWRF to measure fairness in single ranking. We use distribution of male and female authors in book dataset to compute target distribution $\hat{\mathbf{p}}$. We compare the improvement of AWRF score in the re-ranked grid ranking where 1 is the highest score of fairness.
- To measure utility, we implement effectiveness metric RBP.
- Both AWRF and RBP are implemented with grid-layout suitable browsing models, *row-skipping* and *slower-decay* with column size 5.
- We use 0.5 as the default value of the fairness-utility balancing parameter Λ.

RQ2. *Consistency of Optimized Fairness Across Devices* Based on user devices, column size of grid layout changes. For example, *Goodreads* shows book recommendations in grid layout and the column size changes across devices; books are displayed in 5 columns on laptop, 2 columns on phone, and 9 columns on iPad. Hence, the system can display the same set of items in various column sizes depending on user device. Re-ranking the items by taking device size into consideration can help to preserve fairness across devices because optimizing the ranked results in grid layout for a particular device may not remain fair for other devices.

- We observe if and how the optimized fairness score from a re-ranked grid layout of column size n changes in other columns sizes.
- We optimize the grid-based ranked results with column size of 5 and use that fairness-aware re-ranked results to measure provider-side group fairness by changing column size to 2, 3, 4, 7, and 9.

RQ3. *Preserve Fairness Across Devices* Since item exposure varies across column sizes in grid layout which affect the fairness score for provider groups, we want to preserve provider-side fairness across devices. With that goal,

- We implement the grid-aware re-ranking technique for multiple column sizes to maintain group fairness across user devices and observe the change in fairness optimization with the change of column sizes.
- We implement the grid-aware re-ranking algorithm for grid ranked results with common columns sizes of 2, 3, 4, 5, 7, and 9.

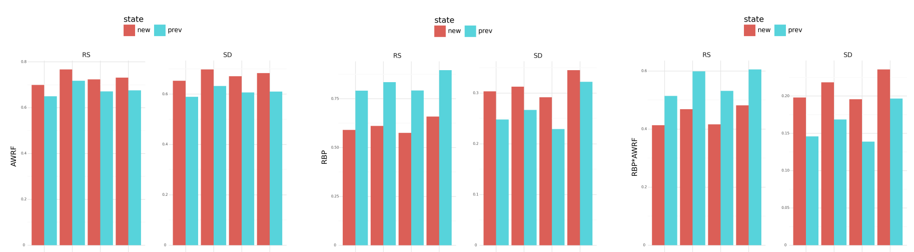
To observe the impact of browsing models on fairness optimization in grid layout, we implement both group fairness metric and effectiveness metric incorporating grid-layout suitable *row-skipping* and *slower-decay* browsing models with their default parameter settings.

4.3 Results

This section provide the results from our experiments.

RQ1 *Does incorporating grid-aware browsing models to existing re-ranking technique improve fairness for ranked results in grid layout?*

Figure 1(a) shows that the AWRF score increases in all the recommendation algorithms for both *row-skipping* and *slower-decay* browsing models. We do *paired t-test* [20] to observe the significance of this fairness improvement and find that for the algorithms in both browsing models, the AWRF score improvement is statistically significant with $p_{val} < 10^{-20}$. We round up the p-values at $\alpha = 0.05$ with Benjamini-Hochberg correction [1]. In both browsing models, the fairness score varies across recommendation algorithms during both pre and post-optimization showing the same patterns. For all the recommendation algorithms, the fairness scores improves significantly for ranking in grid layout when we consider grid-layout suitable browsing models. Figure 1(b) shows the RBP score and Figure 1(c) shows the RBP $*$ AWRF score differences in between pre and post-optimization. For the *slower-decay* browsing model, the combined score improves in all the algorithms and the utility score improves after re-ranking. The improvement does not hold for *row-skipping* browsing model. This observation emphasizes the importance of using grid-aware re-ranking technique while optimizing ranked results displayed in grid layout.



(a) AWRF score difference (b) RBP score difference (c) Fairness-Utility trade-off

Fig. 1. Pre and post-optimization fairness and utility scores in grid layout with column size 5

RQ2. *Does a ranking in grid layout optimized for fairness in a device remain fair for other devices?*

RQ2 shows the impact of column sizes on fairness optimization in grid layout. Figure 2 shows how fairness score for an optimized ranking changes with column sizes. A fairness-aware re-ranked 5-column grid layout does not remain fair when the column size is different and this pattern is true for all the algorithms. The pattern is more notable in *row-skipping* browsing model for all the algorithms. This result implies the need of considering appropriate column size to preserve fairness for the same set of ranked items across devices.

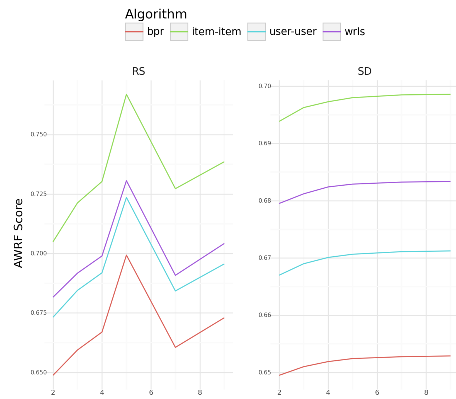


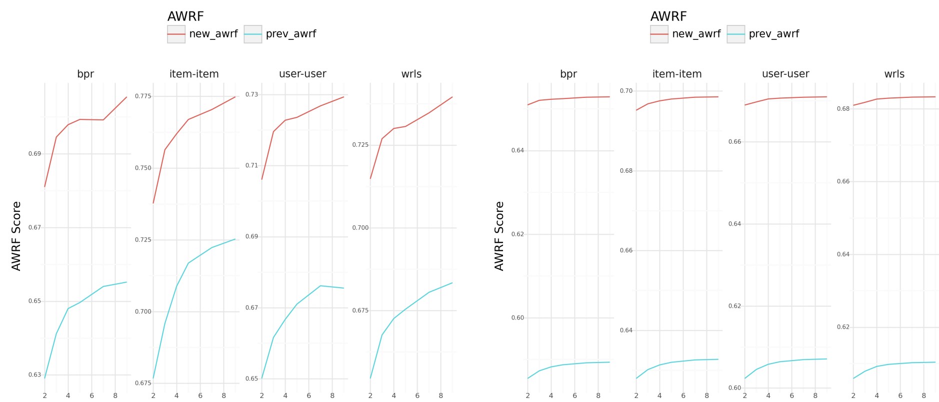
Fig. 2. An optimized grid layout with column size 5 is not fair for other column sizes.

RQ.3 *How can we optimize ranking in grid layout for various screen sizes?*

Figure 3 shows the improvement in fairness scores after optimizing ranking in the grid layout for various column sizes and the result shows a consistency in fairness improvement across column sizes. For all the considered column sizes, AWRF score improves significantly in all the recommendation algorithms ($p_{val} < 0.0001$ rounded at $\alpha = 0.05$ with Benjamini-Hochberg correction) after optimizing ranking in grid layout using grid-aware browsing models. By looking at figure 3, we can see that fairness score varies with the change of column sizes and this pattern remains consistent even after optimization in all the algorithms for both browsing models. This result shows that, fairness optimization of a given grid layout of column size n should consider the same column size while measuring position weight using browsing models to improve fairness in that ranking.

Discussion Through our experiments we provide insights on the impact of device sizes and browsing models on fairness optimization in grid layout. We have made following observations from our analysis.

- It is possible to improve fairness in grid layout if we can make re-ranking techniques grid-aware by incorporating grid-layout suitable browsing models. However, the improvement in fairness score can vary depending on user browsing models. This observation highlight the importance of considering suitable browsing models while measuring and optimizing group fairness in grid layout. Understanding how users browse grid layout and identifying various browsing tendencies can help to develop more accurate fairness optimization technique for grid layout.
- Device size is an important factor in improving fairness in grid layout. Optimizing provider-side group fairness in ranking in grid layout with a particular column size will not remain fair with the change of column sizes. Hence, a



(a) *Row-skipping* browsing model (b) *Slower-decay* browsing model

Fig. 3. Improvement in fairness across column sizes in grid-aware browsing models

ranked result which is optimized for fairness while displaying in phone will not remain fair while displaying in a laptop. Therefore, to preserve fairness across devices, a retrieved results displayed in a particular device needs to be re-ranked considering the appropriate column size while displaying in another device.

– The consistency of fairness score across column sizes varies based on browsing models. For *row-skipping* browsing model, the fairness score varies notably across column sizes but for *slower-decay*, the fairness scores are more consistent across column sizes. This observation emphasizes the need of selecting suitable browsing model and column size while optimizing ranking in grid layout for provider-side group fairness.

5 Conclusion

In this paper, we work towards filling a gap in the research area of provider-side group fairness in ranking in IAS by studying fairness improvement in grid layout. We modify a widely used fairness-aware re-ranking technique to make it grid-aware by incorporating grid-layout suitable user browsing models. We implement the modified grid-aware re-ranking technique in real-world IAS dataset to observe the fairness improvement in ranking in grid layout. Our analysis shows that device size and user browsing models are crucial factors in designing fairness-aware re-ranking technique to optimize provider-side group fairness in grid layout in IAS.

This work opens up several potential research directions in improving provider-side fairness in grid layout. Our work shows the importance of using accurate user browsing models in fairness optimization for grid layout. User browsing behavior in ranking in grid layout has not received much attention yet, hence, further research work on understanding user browsing behavior in grid layout will help ensuring fairness in grid layout with minimum utility loss.

Moreover, in this work, we do not consider multi-list grid layout where items are displayed in multiple categories. Re-ranking technique designed for wrapped grid-layout may not work for multi-list grid layout because in multi-list grid, each rows represents different genre or categories. Moreover, same item can appear in multiple rows. Hence, future work is needed to optimize multi-list grid ranking for fairness by considering unique features and suitable user browsing models for multi-list ranking.

We believe this work will provide researcher and practitioners an guideline on what to expect while designing an optimization technique for fairness in grid layout and what factors to consider carefully.

References

1. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57,** 289–300 (1995).
2. Herlocker, J. L., Konstan, J. A., Borchers, A. & Riedl, J. *An Algorithmic Framework for Performing Collaborative Filtering* in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Association for Computing Machinery, New York, NY, USA, 1999), 230–237. ISBN: 1581130961.
3. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics,* 1189–1232 (2001).
4. Deshpande, M. & Karypis, G. Item-based Top-n Recommendation Algorithms. *ACM Transactions on Information Systems (TOIS)* **22,** 143–177 (2004).
5. Burges, C. *et al. Learning to rank using gradient descent* in *Proceedings of the 22nd international conference on Machine learning* (2005), 89–96.
6. Cao, Z., Qin, T., Liu, T.-Y., Tsai, M.-F. & Li, H. *Learning to rank: from pairwise approach to listwise approach* in *Proceedings of the 24th international conference on Machine learning* (2007), 129–136.
7. Li, P., Wu, Q. & Burges, C. Mcrank: Learning to rank using multiple classification and gradient boosting. *Advances in neural information processing systems* **20** (2007).
8. Shrestha, S. & Lenz, K. Eye gaze patterns while searching vs. browsing a website. *Usability News* **9,** 1–9 (2007).
9. Tatler, B. W. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of vision* **7,** 4–4 (2007).
10. Xu, J. & Li, H. *Adarank: a boosting algorithm for information retrieval* in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (2007), 391–398.

11. Craswell, N., Zoeter, O., Taylor, M. & Ramsey, B. *An experimental comparison of click position-bias models* in *Proceedings of the 2008 international conference on web search and data mining* (2008), 87–94.
12. Moffat, A. & Zobel, J. Rank-biased Precision for Measurement of Retrieval Effectiveness. *ACM Transactions on Information Systems (TOIS)* **27,** 1–27 (2008).
13. Taylor, M., Guiver, J., Robertson, S. & Minka, T. *Softrank: optimizing non-smooth rank metrics* in *Proceedings of the 2008 International Conference on Web Search and Data Mining* (2008), 77–86.
14. Xia, F., Liu, T.-Y., Wang, J., Zhang, W. & Li, H. *Listwise approach to learning to rank: theory and algorithm* in *Proceedings of the 25th international conference on Machine learning* (2008), 1192–1199.
15. Rendle, S., Freudenthaler, C., Gantner, Z. & Schmidt-Thieme, L. *BPR: Bayesian Personalized Ranking from Implicit Feedback* in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* (AUAI Press, Montreal, Quebec, Canada, 2009), 452–461. ISBN: 9780974903958.
16. Burges, C. J. From ranknet to lambdarank to lambdamart: An overview. *Learning* **11,** 81 (2010).
17. Djamasbi, S., Siegel, M. & Tullis, T. *Visual hierarchy and viewing behavior: An eye tracking study* in *International conference on human-computer interaction* (2011), 331–340.
18. Takács, G., Pilászy, I. & Tikk, D. *Applications of the Conjugate Gradient Method for Implicit Feedback Collaborative Filtering* in *Proceedings of the Fifth ACM Conference on Recommender Systems* (Association for Computing Machinery, Chicago, Illinois, USA, Oct. 2011), 297–300. ISBN: 9781450306836.
19. Dwork, C., Hardt, M., Pitassi, T., Reingold, O. & Zemel, R. *Fairness through awareness* in *Proceedings of the 3rd innovations in theoretical computer science conference* (2012), 214–226.
20. Hsu, H. & Lachenbruch, P. A. Paired t test. *Wiley StatsRef: statistics reference online* (2014).
21. Lan, Y., Zhu, Y., Guo, J., Niu, S. & Cheng, X. *Position-Aware ListMLE: A Sequential Learning Process for Ranking.* in *UAI* (2014), 449–458.
22. Zhao, Q., Chang, S., Harper, F. M. & Konstan, J. A. *Gaze prediction for recommender systems* in *Proceedings of the 10th ACM Conference on Recommender Systems* (2016), 131–138.
23. Xie, X. *et al. Investigating examination behavior of image search users* in *Proceedings of the 40th international acm sigir conference on research and development in information retrieval* (2017), 275–284.
24. Yang, K. & Stoyanovich, J. *Measuring Fairness in Ranked Outputs* in *Proceedings of the 29th International Conference on Scientific and Statistical Database Management* (2017), 1–6.
25. Zehlike, M. *et al. FA*IR: A Fair Top-k Ranking Algorithm* in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*

(Association for Computing Machinery, Singapore, Singapore, 2017), 1569–1578. ISBN: 9781450349185. https://doi.org/10.1145/3132847.3132938.

26. Biega, A. J., Gummadi, K. P. & Weikum, G. *Equity of Attention: Amortizing Individual Fairness in Rankings* in *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval* (2018), 405–414.
27. Ekstrand, M. D., Tian, M., Kazi, M. R. I., Mehrpouyan, H. & Kluver, D. *Exploring Author Gender in Book Rating and Recommendation* in *Proceedings of the 12th ACM Conference on Recommender Systems* (2018), 242–250.
28. Singh, A. & Joachims, T. *Fairness of Exposure in Rankings* in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Association for Computing Machinery, London, United Kingdom, 2018), 2219–2228. ISBN: 9781450355520. https://doi.org/10.1145/3219819.3220088.
29. Wan, M. & McAuley, J. *Item Recommendation on Monotonic Behavior Chains* in *Proceedings of the 12th ACM Conference on Recommender Systems* (Association for Computing Machinery, Vancouver, British Columbia, Canada, 2018), 86–94. ISBN: 9781450359016. https://doi.org/10.1145/3240323.3240369.
30. Wu, L., Hsieh, C.-J. & Sharpnack, J. *Sql-rank: A listwise approach to collaborative ranking* in *International Conference on Machine Learning* (2018), 5315–5324.
31. Beutel, A. *et al. Fairness in Recommendation Ranking Through Pairwise Comparisons* in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019), 2212–2220.
32. Biega, A. J., Diaz, F., Ekstrand, M. D. & Kohlmeier, S. *Overview of the TREC 2019 Fair Ranking Track* in *The Twenty-Eighth Text REtrieval Conference (TREC 2019) Proceedings* (2019).
33. Geyik, S. C., Ambler, S. & Kenthapadi, K. *Fairness-aware ranking in search & recommendation systems with application to linkedin talent search* in *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining* (2019), 2221–2231.
34. Liu, W., Guo, J., Sonboli, N., Burke, R. & Zhang, S. *Personalized fairness-aware re-ranking for microlending* in *Proceedings of the 13th ACM Conference on Recommender Systems* (2019), 467–471.
35. Pei, C. *et al. Personalized re-ranking for recommendation* in *Proceedings of the 13th ACM conference on recommender systems* (2019), 3–11.
36. Sapiezynski, P., Zeng, W., E Robertson, R., Mislove, A. & Wilson, C. *Quantifying the Impact of User Attention on Fair Group Representation in Ranked Lists* in *Companion Proceedings of The 2019 World Wide Web Conference* (Association for Computing Machinery, San Francisco, USA, 2019), 553–562. ISBN: 9781450366755. https://doi.org/10.1145/3308560.3317595.
37. Xie, X. *et al. Grid-based evaluation metrics for web image search* in *The world wide web conference* (2019), 2103–2114.

38. Diaz, F., Mitra, B., Ekstrand, M. D., Biega, A. J. & Carterette, B. *Evaluating Stochastic Rankings with Expected Exposure* in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Association for Computing Machinery, Virtual Event, Ireland, 2020), 275–284. ISBN: 9781450368599. https://doi.org/10.1145/3340531.3411962.
39. Ekstrand, M. D. *LensKit for Python: Next-Generation Software for Recommender Systems Experiments* in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Association for Computing Machinery, Virtual Event, Ireland, 2020), 2999–3006. ISBN: 9781450368599. https://doi.org/10.1145/3340531.3412778.
40. Ekstrand, M. D. & Kluver, D. Exploring Author Gender in Book Rating and Recommendation. *User Modeling and User-Adapted Interaction.* https://md.ekstrandom.net/pubs/bag-extended (Feb. 2020).
41. Jiang, H. & Nachum, O. *Identifying and correcting label bias in machine learning* in *International Conference on Artificial Intelligence and Statistics* (2020), 702–712.
42. Narasimhan, H., Cotter, A., Gupta, M. R. & Wang, S. *Pairwise Fairness for Ranking and Regression.* in *AAAI* (2020), 5248–5255.
43. Raj, A., Wood, C., Montoly, A. & Ekstrand, M. D. Comparing fair ranking metrics. *arXiv preprint arXiv:2009.01311* (2020).
44. Pitoura, E., Stefanidis, K. & Koutrika, G. Fairness in rankings and recommendations: an overview. *The VLDB Journal,* 1–28 (2021).
45. Sonoda, R. A Pre-processing Method for Fairness in Ranking. *arXiv preprint arXiv:2110.15503* (2021).
46. Chen, S. *et al.* Reinforcement Re-ranking with 2D Grid-based Recommendation Panels. *arXiv preprint arXiv:2204.04954* (2022).
47. Ekstrand, M. D., Das, A., Burke, R., Diaz, F., *et al.* Fairness in information access systems. *Foundations and Trends® in Information Retrieval* **16,** 1–177 (2022).
48. Raj, A. & Ekstrand, M. D. *Measuring Fairness in Ranked Results: An Analytical and Empirical Comparison* in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2022), 726–736.
49. Ekstrand, M. D., McDonald, G., Raj, A. & Johnson, I. Overview of the TREC 2022 Fair Ranking Track. *arXiv preprint arXiv:2302.05558* (2023).
50. Pinney, C., Raj, A., Hanna, A. & Ekstrand, M. D. Much Ado About Gender: Current Practices and Future Recommendations for Appropriate Gender-Aware Information Access. *arXiv preprint arXiv:2301.04780* (2023).
51. Raj, A. & Ekstrand, M. Unified Browsing Models for Linear and Grid Layouts. *arXiv preprint arXiv:2310.12524* (2023).
52. Raj, A. & Ekstrand, M. D. Towards Measuring Fairness in Grid Layout in Recommender Systems. *arXiv preprint arXiv:2309.10271* (2023).