FairRankTune

RankTune Data Generation Technical Report

August 31, 2023

Contents

1 RankTune Data Generator **2**

1.1 Notation . . . 2

1.2 Underlying Core Idea . . . 2

1.3 Description of RankTune Data Generation Algorithm . . . 2

1.4 Assumptions underpinning RankTune . . . 3

1.5 Demonstration of RankTune Algorithm . . . 4

1.5.1 Fairness Metrics . . . 4

1.5.2 Empirical Demonstration . . . 4

1 RankTune Data Generator

A primary component of FairRankTune is the fairness-tunable ranked data generator RANKTUNE.

1.1 Notation

We use the symbol $X = x_1, x_2, ..., x_n$ to represent items to be ordered in a ranked list τ. $\tau(x_i)$ denotes the ordinal position of item x_i in the ranking τ. Items, sometimes called documents, candidates, or providers, belong to a group defined by a shared protected attribute value, such as, gender = "woman". We represent the groups associated with the items as $G = g_1, g_2, ..., g_m$. We use $D_X = (p_{X:g_1}, ..., p_{X:g_m})$ to represent the distribution of groups in the item set X, where the proportion of each group is $p_{X:g_m} = |g_m|/|X|$. For example, $D_X = (0.2, 0.3, 0.5)$ indicates that g_1 corresponds to 20% of items in X, and g_2 and g_3 to 30% and 50% of X, respectively.

1.2 Underlying Core Idea

RANKTUNE is a *fairness-tunable* ranked data generation method. It constructs a ranking τ by placing items into τ from top to bottom. The idea behind RANKTUNE is that to construct a "fair" ranking, each time we place an item in the generated ranking, the likelihood of placing an item $\in g_i$ should be equal to g_i's proportion of the items (i.e., if $p_{x:g_i} = 0.2$, meaning g_i is 20% of X then g_i should have a 20% chance of being placed). Then, on the other side of the spectrum, if we want a completely "unfair" ranking, we should place items into τ such that groups are ordered by increasing size from small to large. In this way, smaller groups would get bigger proportions of favorable positions, which maximally violates statistical parity fairness [1]. These are the two ends of the statistical parity spectrum.

To generate rankings along this spectrum, RANKTUNE samples a random number r in the $[0, 1]$ interval each time it places an item. We design this interval to have "regions" that map to groups. In this way, the unfairness tuning parameter ϕ controls representativeness, i.e., how fairly each group is represented in the ranking. Specifically, when $\phi = 0$, then each group is fairly represented. Thus each group's region is equal to the group's proportion of X (fair). As ϕ increases, the fair representation of each group degrades because regions are distorted in such a way that smaller groups have larger regions compared to their proportion of X (unfair). The fairness tuning parameter ϕ is used to create the regions prior to placing any items into ranking τ.

1.3 Description of RankTune Data Generation Algorithm

Algorithm 1 displays the pseudocode for the core algorithmic strategy of the RANKTUNE tool. This algorithm operates in three stages: (1) *creating group regions stage*, the assignment of groups to regions in the $[0, 1]$ interval using ϕ; (2) *pseduo-random item placement stage*, the repeated sampling of the $[0, 1]$ interval to place items into the to-be-constructed ranking and (3) *filling-out stage*, once a group has no items remaining, the rest of the items are placed.

Creating Group Regions: This begins by identifying the smallest group g_{min}, or in the case of multiple such groups, it chooses a random group among those of the smallest size. Next, ϕ is scaled to a new value ϕ_{scaled} which represents g_{min}'s artificially adjusted new proportion of the item set (line 4). This new proportion can be the same (case of $\phi = 0$, i.e., fair) or larger all the way to $\phi = 0$ resulting in $\phi_{scaled} = 1$ meaning the smallest group is certainly at the top of the ranking. Then the original distribution of groups D_x is adjusted to a new group distribution D_{target}. That is, all other groups are proportionally scaled down to accommodate g_{min}'s new proportion of the item set ϕ_{scaled}. Then we take the target group distribution and map each group's proportion of X into representing a corresponding region of the $[0, 1]$ space. Each group's proportion of the $[0, 1]$ interval is represented via a lower bound ($lowBound$) and upper bound ($upperBound$). This stage sets the fairness of the placement procedure.

Pseudo-random Item Placement: Subsequently the ranking τ is constructed from top to bottom (i.e. appending), by repeatedly sampling a random number r in the uniform $[0, 1]$ interval. In the prior stage, groups were mapped to regions of the interval. Thus, when r is sampled, we know what group to place (g_{2p}). Here, g_{2p} is the group for which $lowBound[g_{2p}] < r < upperBound[g_{2p}]$.

Filling Out Stage: Once RANKTUNE has placed an entire group the remaining positions in τ are filled by placing groups according to increasing group size. RANKTUNE then returns the ranking τ.

1.4 Assumptions underpinning RankTune

The underlying assumption of RANKTUNE is that in order to generate rankings satisfying statistical parity fairness, the likelihood of a group receiving a positive outcome should be equal to that group's proportion of the candidate pool. Then unfairness can be added by distorting this proportional relationship between the likelihood of the group receiving the positive outcome and its proportion of the of candidate pool. We view the positive outcome to be the placement of a candidate into the generated ranking and each placement is made by sampling a random number in the uniform $[0, 1]$ interval. In the case of perfect statistical parity, the likelihood of placing a candidate from each group is equal to that group's proportion of all candidates. RANKTUNE is driven by a stochastic process (using a random number generator to sample the $[0, 1]$ interval), thus we are unable to formally prove that the approach always generates a ranking of a desired degree of fairness.

Algorithm 1: RANKTUNE Data Generation Algorithm

Data: Item set X and groups G, representativeness parameter ϕ
Result: Ranking τ

$\tau \leftarrow []$, $D_X \leftarrow (p_{x:g_1}, ..., p_{x:g_m})$
$g_{min} \leftarrow$ group id of smallest group;
$\phi_{scaled} \leftarrow (1-\phi) * (1 - p_{x:g_{min}}) + p_{x:g_{min}}$;
$D_{target} \leftarrow (D_X/(1 - p_{x:g_{min}}) * (1 - \phi_{scaled})$;
$D_{target}(g_{min}) \leftarrow \phi_{scaled}$; /* so $\sum D_{target} = 1$ */
$lowCount \leftarrow 0$
$lowBound$ & $upperBound \leftarrow$ empty arrays of size $|G|$
forall $g_i \in G$ **do**
    $lowBound[g_i] \leftarrow lowCount$;
    $upperBound[g_i] \leftarrow lowCount + D_{target}(g_i)$;
    $lowCount \leftarrow lowCount + D_{target}(g_i)$;
while *each group has unplaced items* **do**
    $r \leftarrow random([0, 1])$
    $g_{2p} \leftarrow$ s.t. $lowBound[g_{2p}] < r < upperBound[g_{2p}]$
    $\tau.append$(next item from group g_{2p})
$G_{rem} \leftarrow$ sorted (smallest to largest) groups with unplaced items
forall $g_i \in G_{rem}$; /* place items by increasing group size */
do
    forall $x_j \in g_i$ **do**
        $\tau.append(x_j)$
Return τ

1.5 Demonstration of RankTune Algorithm

1.5.1 Fairness Metrics

In this section we employ measures of statistical parity with specific formulations. For brevity, they are referenced by their abbreviation. Table 1 describes the specific formulation we use here.

Abbreviation	Formulation	Range	More Fair
ER	Min-max ratio of avg. group exposures as in [2]	$(0, 1]$	1
EE-D	L_2 norm of avg. group exposure as in [3]	[0,∞)	0
AWRF	Min-max ratio avg. group attention as in [4]	$(0, 1]$	1
NDKL	Proportional group representation across rank prefixes as in [5]	[0,∞)	0
ARP	Max-min difference avg. mixed pairs as in [6]	[0,1]	0

Table 1: Overview of fair ranking metrics used in this work.

1.5.2 Empirical Demonstration

We now demonstrate the capabilities of FairRankTune's RANKTUNE using a variety of group memberships. Table 2 depicts the group distributions we utilize in our demonstration.

Name	Distribution - D_X
Dist A	$(0.2, 0.3, 0.5)$
Dist B	$(0.1, 0.3, 0.6)$
Dist C	$(0.2, 0.3, 0.1, 0.05, 0.03)$
Dist D	$(0.2, 0.2, 0.2, 0.2, 0.2)$
Dist E	$(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.05, 0.05)$
Dist F	$(0.6, 0.08, 0.02, 0.15, 0.1, 0.05)$

Table 2: Group distributions employed in our empirical analysis and case study. We use $X = 1,000$ items. For example, *Dist A* has three groups: the first has 200 members, the second has 300 members, and the third has 500 members.

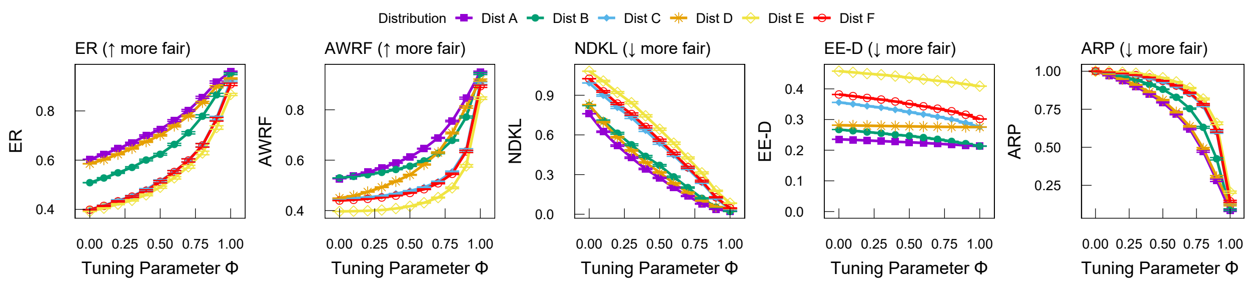

Figure 1: Average metric values (with 95% confidence intervals) from 200 RANKTUNE constructed rankings with the distributions in Table 2 for $1,000$ items. As ϕ increases, RANKTUNE outputs increasingly fairer rankings. ER and AWRF are more fair at 1 (thus upward slopes), and NDKL, EE-D, and ARP are more fair at 0 (thus downward slopes).

Figure 1, displays the results of generating 200 rankings for different ϕ representativeness values for the multiple groups distributions from Table 2 (*Dist A - F*. Specifically, we generate 200 rankings for ϕ ranging from 0 to 1 in increments of 0.1.

We observe that across all fairness metrics and all group distributions (colored lines), RANKTUNE outputs progressively fairer rankings as ϕ is adjusted from 0 to 1. The metrics follow the same overall trends, but vary in their ranges and sensitivity to different group distributions as seen by how different metrics relatively order *Dist A - F* when $\phi = 1$ or other ϕ values. Critically, for each distribution and metric, we consistently see fairness increase as ϕ increases. Robustness is seen through the small 95% confidence intervals over 200 trials. Thus, we illustrate that while RANKTUNE is agnostic to particular formulations of group fairness, it is generally applicable to settings in which statistical parity is studied.

References

[1] Caitlin Kuhlman, Walter Gerych, and Elke Rundensteiner. Measuring group advantage: A comparative study of fair ranking metrics. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 674–682, 2021.

[2] Meike Zehlike and Carlos Castillo. Reducing disparate exposure in ranking: A learning to rank approach. In *Proceedings of The Web Conference 2020*, pages 2849–2855, 2020.

[3] Fernando Diaz, Bhaskar Mitra, Michael D. Ekstrand, Asia J. Biega, and Ben Carterette. Evaluating stochastic rankings with expected exposure. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020.

[4] Avijit Ghosh, Ritam Dutt, and Christo Wilson. When fair ranking meets uncertain inference. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021.

[5] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 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*, pages 2221–2231, 2019.

[6] Kathleen Cachel, Elke Rundensteiner, and Lane Harrison. Mani-rank: Multiple attribute and intersectional group fairness for consensus ranking. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 2022.