Title: Fair Ranking with Noisy Protected Attributes

Year: 2022

Venue: NeurIPS

Main Problem:

- **Ranking Problem:** Generally, when we have m items and we have to choose n of them (n<=m) and return a permutation of those, the permutation is formally called ranking. When we want to solve a ranking problem, the relevance of the items must be known. Formally, we have an m x n matrix W, and placing the ith item in the jth position will be generating the utility of W_{ij}. We want to have a ranking that maximizes the overall utility. For the final part of the formulation we need to define the assignment marices R^{mxn}. R_{ij} will be 1 if item i appears in the position j and 0 shows the opposite. In this notation, the utility of ranking (which we tend to maximize) is: <R, W>: = $\sum_{i=1}^{n} \sum_{j=1}^{n} R_{ij}W_{ij}$
- **Fair Ranking Problem: Formally, we consider that we have p groups,** $G_1, G_2, \dots G_p \subseteq [m]$ and each of these m items belong to 1 or more of this socially salient groups. Our goal is to come-up with a ranking that maximizes the fairness while satisfying certain fairness criteria. **"**The appropriate notion of fairness is context dependent, and to capture different fairness criteria numerous fairness constraints have been proposed." To clarify what the fairness constraint is, it can be applying a constraint to an algorithm to ensure one or more definitions of fairness are satisfied. There are three common ways to do that (this work uses the third one based on my understanding):
    1. Post-processing your model's output.
    2. Altering the loss function to incorporate a penalty for violating a fairness metric.
    3. Directly adding a mathematical constraint to an optimization problem.

Drawbacks of Previous Works:

All of the mentioned related works in this paper require access to the socially-salient attributes or their correct values and distribution. Usually in real life situations, these attributes have noises, are missing, or inaccessible.

Proposed Method and Contributions:

This work presents a fair ranking framework that guarantees given fairness criteria is satisfied when the socially-salient attributes are assumed to follow a specific probabilistic noise model. Instead of sampling the attribute values and applying the constraint on them they apply it on the relaxed-fairness criteria to the expected number of items from each group that appear in the first k positions.

The noise model that they use is previously appeared in [1][2][3]

Informally, my understanding was that they borrowed the mentioned noise model, did some justification, mathematically proved it guarantees the fairness and used that in a common ranking formulation explained above.

Datasets:

1. Synthetic Dataset (generated by the code from [3])
2. Occupations dataset [9]
3. Chess Ranking Data [11]

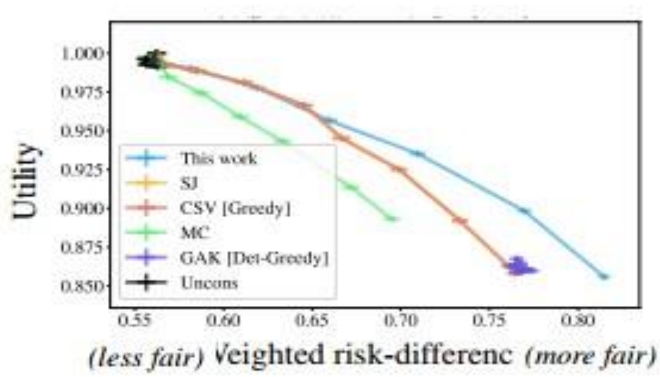
Baselines and Simulations Summary:

For simulation, the correct choice for the metric is dependent on the context of our data and problem. In this paper they used weighted risk-difference (RD) which is a position-weighted version of standard risk-difference metric [4]. Simply, it measures the extent that ranking violates equal representation.

They compared their framework namely, NResilient against CSV [5], SJ [6], GAK (DetGreedy) [7], MC [8] and Uncons which makes the ranking without fairness consideration.

They made considerable amount of simulations on synthetic and real-world datasets and they were available in the supplementary material of their work. The following is a sample of their simulation on Occupations dataset [9] "which contains the top 100 Google Image results for 96 occupation-related queries. For each image, the data has its position in search results, gender (coded as male/female) of the individual depicted in the image, collected via MTurk. We use the (true) gender labels in the data to compute RD and to estimate Phat, but do not provide them to algorithms."

Note that estimation of Phat (probabilistic info about protected attributes) is context dependent. For this example, they used a CNN-based gender classifier from [10] to predict (apparent) gender of the person depicted in each image.



The figure for RD and utilities (NDCG) "averaged over 1000 iterations. We observe that NResilient achieves the best RD (≈0.81) and has a better RD-utility trade-off than the other baselines. In contrast, CSV, SJ, and GAK, achieve a worse RD (≤0.77). MC achieves the worst

RD ($\leq$0.70) and a worst RD-utility trade-off. In particular, NResilient's RD-utility trade-off strictly dominates all baselines for RD $\geq$ 0.66. This value of RD can arise in practice."

To put it in a nutshell, in all the simulations, it is observed that this method outperforms all the baselines mentioned in this paper in terms of fairness. And regarding fairness-utility trade-off, whether it's better or similar in worse case.

Limitations:

Compared to existing fair-ranking frameworks, this framework does not need accurate socially salient attributes (or protected attributes), but assumes that errors in these attributes are **random** and **independent**. When these assumptions do not hold, the framework may not satisfy its guarantees and a careful assessment of this on application-specific data would be important to avoid any (unintended) negative social impact

Codebase:

https://github.com/AnayMehrotra/FairRankingWithNoisyAttributes

References:

[1] Seyed A. Esmaeili, Brian Brubach, Leonidas Tsepenekas, and John Dickerson. Probabilistic Fair Clustering. In NeurIPS, 2020.

[2] Alexandre Louis Lamy and Ziyuan Zhong. Noise-Tolerant Fair Classification. In NeurIPS, pages 294–305, 2019

[3] Anay Mehrotra and L. Elisa Celis. Mitigating Bias in Set Selection with Noisy Protected Attributes. In FAccT, pages 237–248. ACM, 2021

[4] Toon Calders and Sicco Verwer. Three Naive Bayes Approaches for Discrimination-Free Classification. Data Min. Knowl. Discov., 21(2):277–292, 2010

[5] L. Elisa Celis, Damian Straszak, and Nisheeth K. Vishnoi. Ranking with Fairness Constraints. In ICALP, volume 107 of LIPIcs, pages 28:1–28:15. Schloss Dagstuhl – Leibniz Zentrum fuer Informatik, 2018

[6] Ashudeep Singh and Thorsten Joachims. Fairness of Exposure in Rankings. In KDD, pages 2219–2228. ACM, 2018.

[7] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search. In KDD, pages 2221–2231. ACM, 2019

[8] Anay Mehrotra and L. Elisa Celis. Mitigating Bias in Set Selection with Noisy Protected Attributes. In FAccT, pages 237–248. ACM, 2021.

[9] L. Elisa Celis and Vijay Keswani. Implicit Diversity in Image Summarization. Proc. ACM Hum. Comput. Interact., 4(CSCW2):139:1–139:28, 2020

[10] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep Expectation of Real and Apparent Age from a Single Image without Facial Landmarks. International Journal of Computer Vision, 126(2-4):144–157, 2018.

[11] Avijit Ghosh, Ritam Dutt, and Christo Wilson. When Fair Ranking Meets Uncertain Inference. In SIGIR, pages 1033–1043. ACM, 2021