Fair Ranking with Noisy Protected Attributes (NeurIPS 2022)

Anay Mehrotra and Nisheeth K. Vishnoi

Yuha Park

January 30, 2023

Seoul National University

Contents

Introduction

1. Ranking problem
2. Ranking problem with fairness constraints (Fair-ranking problem)
3. Ranking problem with fairness constraints, but groups are random variables (Ranking problem with noisy attributes)

Model of fair ranking with noisy attributes

1. Ranking problem

- **Ranking** : given m items, one has to select a subset of n items and output a *permutation* of the selected items. This permutation is said to be a ranking. We denote rankings by assignment matrices:

$$R \in \{0,1\}^{m \times n}$$

  where $R_{ij} = 1$ indicates that item i appears in position j, and $R_{ij} = 0$ indicates otherwise.

- **Utility** : given an $m \times n$ matrix W, such that placing the i-th item at the j-th position generates utility W_{ij}.
- **The utility of a ranking** is the sum of utilities generated by each item in its assigned position. In this notation, the utility of a ranking is

$$\langle R, W \rangle := \sum_{i=1}^{m} \sum_{j=1}^{n} R_{ij} W_{ij}$$

1. Ranking problem

- **Ranking problem** : to solve

$$max_{R \in R} \langle R, W \rangle$$

where $\mathcal{R}$ is the set of all assignment matrices denoting a ranking:

$$\mathcal{R} := \left\{ X \in \{0,1\}^{m \times n} : \forall i \in [m], \sum_{j=1}^{n} X_{ij} \leq 1, \ \forall j \in [n], \sum_{i=1}^{m} X_{ij} = 1 \right\}$$

Here, the constraint $\sum_{i=1}^{m} X_{ij} = 1$ ensures position j has exactly one item and the constraint $\sum_{j=1}^{n} X_{ij} \leq 1$ ensures that item i occupies at most one position.

- The algorithmic task in the ranking problem is to output a ranking with the highest utility.

2. Fair-ranking problem

- **Assumptions** : with $p \geq 2$ socially-salient groups $G_1, G_2, \cdots, G_p \subseteq [m]$ (e.g., the group of all women or all Black people) which are often protected by law. Each of the m items belongs to *one or more* of these socially-salient groups.
- **Fair-ranking problem** : to output the ranking with maximum utility subject to satisfying certain fairness criteria with respect to these groups.

Definition 3.1 (Fairness constraints). Given a matrix $U \in \mathbb{Z}_+^{n \times p}$, a ranking R satisfies the upper bound constraint if $\sum_{i \in G_\ell} \sum_{j=1}^{k} R_{ij} \leq U_{k\ell}$, for all $\ell \in [p]$ and $k \in [n]$.

Definition 3.2 (Noise model). Let $P \in [0,1]^{m \times p}$ be a known matrix. The groups $G_1, \cdots, G_p \subseteq [m]$ are random variables, such that, for each $i \in [m]$ and $\ell \in [p], Pr[G_\ell \ni i] = P_{i\ell}$. Moreover, for different items $i \neq j$ the events $G_\ell \ni i$ and $G_k \ni j$ are independent for all ℓ, $k \in [p]$.

3. Ranking problem with noisy attributes

Definition 3.4 ((ϵ, δ)-constraint). For any $\epsilon \in \mathbb{R}^n \geq 0$ and $\delta \in (0, 1]$, a ranking R is said to satisfy (ϵ, δ)-constraint if with probability at least $1 - \delta$ over the draw of $G_1, \cdots, G_p$,

$$\forall k \in [n], \forall \ell \in [p], \sum_{i \in G_\ell} \sum_{j=1}^{k} R_{ij} \leq U_{k\ell}(1 + \epsilon_k).$$

Problem 3.5 (Ranking problem with noisy attributes). Given matrices P, U, and W, find the ranking R such that, for some small ϵ and δ,

$$\max_{R \in \mathcal{R}} \langle R, W \rangle \qquad \text{s.t.} \;\; R \;\text{ satisties }\; (\epsilon, \delta) - \text{constraint.}$$

- Note that solving **Problem 3.5** is NP-hard.

Theoretical results

Optimization framework

- **Input**: Matrices $P \in [0,1]^{m \times p}, W \in \mathbb{R}_{\geq 0}^{m \times n}, U \in \mathbb{R}^{n \times p}$
- **Parameters**: Constant $c > 1$, failure probability $\delta \in (0,1]$, and $k \in [n]$, relaxation parameter

$$\gamma_k := 12 \cdot \log\left(\frac{2np}{\delta}\right) \cdot \max_{\ell \in [p]} \sqrt{\frac{1}{U_{k\ell}}} \tag{1}$$

- **Our Fair-Ranking Program** (2) :

$$\max_{R \in \mathcal{R}} \langle R, W \rangle, \quad \text{(Noise Resilient)}$$

$$\text{s.t. } \forall \ell \in [p] \quad \forall k \in [n]$$

$$\sum_{\substack{i \in [m], \\ j \in [k]}} P_{i\ell} R_{ij} \leq U_{k\ell}\left(1 + \left(1 - \frac{1}{2\sqrt{c}}\right)\gamma_k\right).$$

Fairness and utility of optimal solution

Theorem 4.1 Let $\gamma \in \mathbb{R}^n$ be as defined in Equation (1). There is an optimization program (Program (2)), parameterized by a constant c and failure probability δ, such that for any $c > 1$ and $\delta \in (0, \frac{1}{2}]$ its optimal solution satisfies $(c\gamma, \delta)$-constraint and has a utility at least as large as the utility of any ranking satisfying $((c - \sqrt{c})\gamma, \delta)$-constraint.

- Note that solving Program (2) is a polynomial complexity.

Lower bound on fairness guarantee

Theorem 4.2 There is a family of matrices $U \in \mathbb{Z}_+^{n \times p}$ such that for any U in the family and any parameters $\delta \in [0, 1)$ and $\epsilon_1, \cdots, \epsilon_n \geq 0$, if for any position $k \in [n], \epsilon_k \leq 1$ and $\epsilon_k < \max_{\ell \in [p]} \sqrt{\frac{1}{2U_{k\ell}} \log \frac{1}{4\delta}}$ then there exists a matrix $P \in [0, 1]^{m \times p}$, such that it is information theoretically impossible to output a ranking that satisfies (ϵ, δ)-constraint.

- Since γ_k is $O\left(\log\left(\frac{np}{\delta}\right) \cdot \max_\ell \sqrt{\frac{1}{U_{k\ell}}}\right)$, Theorem 4.2 shows that Theorem 4.1's fairness guarantee is optimal up to log-factors.