

Equality of Opportunity in Rankings

Extended Abstract

Ashudeep Singh
Cornell University
Ithaca, NY
ashudeep@cs.cornell.edu

Thorsten Joachims
Cornell University
Ithaca, NY
tj@cs.cornell.edu

ABSTRACT

With the ubiquity of algorithmic methods for ranking in systems like search engines, recommendations, and advertisements, it becomes essential to ensure that these systems do not impact different groups at different rates. In this work, we focus on defining the notions of fairness in such ranking scenarios in terms of the amount of exposure (opportunity) each item gets in its position while keeping in mind the protected groups. We formulate these definitions for protected groups of documents, queries as well as users.

1 INTRODUCTION

Recently, Google was fined by the EU for skewing its ranking and the claim that it “denied other companies the chance to compete on the merits”¹. But what exactly does it mean to compete on merit when it comes to rankings and what makes a ranking fair? Even if one can argue that Google had no bad intention to disadvantage these companies through search result rankings on their shopping website, those companies probably would have gotten more business, if they had been presented higher in the ranking. This connects to concept of *Fair Equality of Opportunity* [9] defined by philosopher John Rawls, who states that “the social positions be formally open and meritocratically allocated, but, in addition, each individual is to have a fair chance to attain these positions”². In our case, most ranking algorithms are based on historic click through rates [4], but whether something gets clicked is itself a function of whether the historic rankings were fair. Hence, we may need to actively enforce some notion of fairness to gives equal opportunity (or exposure) to items in protected groups so that they can accumulate clicks.

In this work, we take inspiration from the following definitions of fairness in the classification settings to define fairness in rankings. In the following definitions, Y is the true class of an instance, \hat{Y} is the class predicted by the model, and each instance belongs to either one of the two groups $A = 0$ or $A = 1$ e.g. males and females, blacks and whites etc.

- **Demographic Parity for classification** [2]: $P(\hat{Y} = 1|A = 1) = P(\hat{Y} = 1|A = 0)$: This states that the rate of success be equal for the two groups.
- **Equalized Odds for classification** [3]: $P(\hat{Y} = 1|Y = 1, A = 1) = P(\hat{Y} = 1|Y = 1, A = 0)$ and $P(\hat{Y} = 1|Y = 0, A = 1) = P(\hat{Y} = 1|Y = 0, A = 0)$: This states that the probability of

being correctly classified be independent of group membership.

- **Equal Opportunity for classification** [3]: $P(\hat{Y} = 1|Y = 1, A = 1) = P(\hat{Y} = 1|Y = 1, A = 0)$: This states that the probability of classifying positive examples $Y = 1$ correctly be independent of the group membership.

We propose how to extend these notions of fairness to the case of rankings in the next section. Interestingly, fairness considerations can be made both from the point of the user population, as well as from the point of the items that are being ranked.

2 FAIR RANKING FOR PROTECTED GROUPS OF ITEMS

Consider a rental website that displays results as a ranking of properties. Because of position bias, products at higher positions in the ranking get more views (more exposure) and hence probably sell better. It is therefore important that the ranking algorithm ranks properties by merit, and it should not unfairly disadvantage properties belonging to a protected group (say A_1 or having $A = 1$). How can we define fairness in this case?

For now, consider a fixed query and let each document be either relevant or irrelevant to the query i.e. $y_i \in \{0, 1\}$. Let w_k be the fraction of users that visit a search result page and view/examine the document at position k , where $k \in \{1 \dots K\}$ called the position bias [5] e.g. $\{w_1 = 1.0, w_2 = 0.7, w_3 = 0.51, w_4 = 0.39 \dots\}$. Let A_0 and A_1 be two groups with different values of the sensitive attribute. Each document d either belongs to either A_0 or A_1 .

A ranking $r : \mathcal{D} \rightarrow \{1 \dots K\}$ is said to be fair according to the following definitions³ if it satisfies the constraints:

DEFINITION 1 (DEMOGRAPHIC PARITY FOR RANKINGS).

$$\mathbb{E}_d[w_{r(d)}|d \in A_0] = \mathbb{E}_d[w_{r(d)}|d \in A_1]$$

where $r(d)$ is the rank of document d and w_i is the position bias of position i (fraction of users that view the i th position). We can also write the same as:

$$\frac{1}{|A_0|} \sum_{d \in A_0} w_{r(d)} = \frac{1}{|A_1|} \sum_{d \in A_1} w_{r(d)}$$

Since w_k is the fraction of users who open the page and view the document ranked at position k , this definition is equivalent to saying that the fraction of views that the documents in A_0 gets is equal to that of A_1 .

DEFINITION 2 (EQUAL OPPORTUNITY FOR RANKINGS).

$$\mathbb{E}_d[w_{r(d)}|d \in A_0, Y(d) = 1] = \mathbb{E}_d[w_{r(d)}|d \in A_1, Y(d) = 1]$$

¹<http://money.cnn.com/2017/06/27/technology/business/google-eu-antitrust-fine/index.html>

²<https://edeq.stanford.edu/sections/fair-equality-opportunity>

³named similar to the classification fairness definitions described in section 1

where Y denotes whether the document is relevant to the query or not. We can also write it as:

$$\frac{1}{|A_0 \cap Y = 1|} \sum_{d \in A_0, y(d)=1} w_{r(d)} = \frac{1}{|A_1 \cap Y = 1|} \sum_{d \in A_1, y(d)=1} w_{r(d)}$$

where $r(d)$ is the rank of document d .

This definition is equivalent to saying that the fraction of views that the relevant documents form the set A_0 and the relevant documents from the set A_1 .

Since these constraints might not be satisfiable because we have finite number of documents and rank positions under consideration, we can apply the above definitions to distribution over rankings [8]. Distribution over rankings is defined as $\mathbb{P}(r(d_j) = i) \forall i, j$ such that $\mathbb{P} : \{1 \dots D\} \times \{1 \dots K\} \rightarrow [0, 1]$ and is valid i.e. $\forall j \sum_{i=1}^K P(r(d_j) = i) = 1$ and $\forall i \sum_{j=1}^D P(r(d_j) = i) = 1$. Now we can define the above definitions with respect to this distribution defined by \mathbb{P} :

DEFINITION 3 (EQUAL OPPORTUNITY FOR RANKING DISTRIBUTIONS).

$$\mathbb{E}_{r \sim P}[\mathbb{E}_d[w_{r(d)} | d \in A_0, Y(d) = 1]] = \mathbb{E}_{r \sim P}[\mathbb{E}_d[w_{r(d)} | d \in A_1, Y(d) = 1]]$$

where P is the ranking distribution. We can also write it as:

$$\frac{1}{|A_0 \cap Y = 1|} \sum_{\substack{d \in A_0 \\ Y(d)=1}} \sum_{k=1}^K w_k P(r(d) = k) = \frac{1}{|A_1 \cap Y = 1|} \sum_{\substack{d \in A_1 \\ Y(d)=1}} \sum_{k=1}^K w_k P(r(d) = k)$$

where K is the total number of ranks under consideration.

This definition means that the expected fraction of views for the relevant documents form the set A_0 and the relevant documents from the set A_1 is equal.

3 FAIR RANKING FOR PROTECTED GROUPS OF USERS

There is also an alternate case of fairness in rankings beyond fairness to the items being ranked, but instead considering fairness of the ranking to the user population. For convenience, we will define fairness only for users but it is the same for queries as well since each user can be thought of as a (user, query) pair. This relates to works, like [7], that audit search engines to see if there is a differential satisfaction of different demographics.

With respect to a user population, we would like our fairness constraints to convey that a user belonging to a protected group, on average, should not have to spend more effort than a user from outside the protected group. For example, a man looking for a movie he likes should not systematically need to do more or less work than a woman.

As before, considering a fixed query, let the users belong to group A_0 and A_1 and each document is either relevant to the user or not ($Y = 0$ or 1). Let z_k be the amount of work required to inspect the document ranked at position k . For example, z_k could be the time it take to get to position k , or it can be interpreted as the fraction of people who drop out of viewing the ranking

before position k (i.e $z_k = 1 - w_k$, where w_k is the position bias). A ranking $r : \mathcal{D} \rightarrow \{1 \dots K\}$ is said to be fair if it satisfies the following definition:

DEFINITION 4 (EQUAL OPPORTUNITY FOR RANKINGS IN PRESENCE OF PROTECTED USERS).

$$\mathbb{E}_u[\mathbb{E}_d[w_{r(d)} | u \in A_0, Y(d, u) = 1]] = \mathbb{E}_u[\mathbb{E}_d[w_{r(d)} | u \in A_1, Y(d, u) = 1]]$$

where Y denotes whether the document is relevant to the user and query pair or not. We can also write the same as:

$$\frac{1}{|A_0|} \sum_{u \in A_0} \frac{1}{|\{d : y(d, u) = 1\}|} \sum_{d, y(d, u)=1} z_{r(d)} = \frac{1}{|A_1|} \sum_{u \in A_1} \frac{1}{|\{d : y(d, u) = 1\}|} \sum_{d, y(d, u)=1} z_{r(d)}$$

where $r(d)$ is the rank of document d .

4 DISCUSSION

For both the user-centric and the item-centric view on fairness of a ranking, fairness potentially comes at reduction of ranking quality as defined by conventional ranking metrics like DCG. We are currently investigating what the price of fairness is, depending on the size of the protected class and the difference in relevance between classes.

Knowledge of true Y is typically unknown in practical ranking systems (as is in classification). However, equal opportunity relies on the fact that there is a ground truth value Y for each document and query pair and asks for exposing these documents at equal rates. It is therefore interesting to think about how to efficiently measure fairness in real-world systems through targeted interventions. This has already proven feasible for inferring position bias [6].

The next step is to design new or adapt the existing machine learning algorithms that can satisfy (albeit approximately) these constraints while maximizing utility of the rankings. Related works such as [10] calculate set-wise parity at a series of cut-off points in ranking to measure fairness. This is very close to the demographic parity definition in our case. Meanwhile [1] solves the ranking problem as a maximum bipartite matching and inducing a similar constraint. Zehlike et al.[11] also satisfy a demographic parity like constraint while maximizing the utility of the model that is defined in terms of relevance. However similar, our definitions are an easy extension of their classification counterparts while having an interpretation in terms of the exposure. Hence it will be interesting to see if the two constraints are compatible with each other as well as the machine learning methods that can find feasible solutions satisfying them.

The applications of fairness in ranking are numerous and it comes from the way we define our protected groups. For example, one such consequence of defining groups as differing ideologies or opinions. Fairness constraints allow the algorithm to maintain diversity of ideas, hence avoid polarization.

REFERENCES

- [1] L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. 2017. Ranking with Fairness Constraints. *arXiv preprint arXiv:1704.06840* (2017).
- [2] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ACM, 214–226.
- [3] Moritz Hardt, Eric Price, Nati Srebro, et al. 2016. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems*. 3315–3323.
- [4] Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 133–142.
- [5] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. 2007. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems (TOIS)* 25, 2 (2007), 7.
- [6] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased learning-to-rank with biased feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 781–789.
- [7] Rishabh Mehrotra, Ashton Anderson, Fernando Diaz, Amit Sharma, Hanna Wallach, and Emine Yilmaz. 2017. Auditing Search Engines for Differential Satisfaction Across Demographics. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 626–633.
- [8] Shuzi Niu, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. 2013. Stochastic rank aggregation. *arXiv preprint arXiv:1309.6852* (2013).
- [9] John Rawls. 2009. *A theory of justice*. Harvard university press.
- [10] Ke Yang and Julia Stoyanovich. 2016. Measuring fairness in ranked outputs. *arXiv preprint arXiv:1610.08559* (2016).
- [11] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. FA* IR: A Fair Top-k Ranking Algorithm. *arXiv preprint arXiv:1706.06368* (2017).