Lecture Note 12: Demographic Fairness in Advertising Auctions

Christian Kroer*

March 29, 2022

1 Introduction

This lecture note will take a look at internet advertising auctions from the perspective of demographic fairness. A serious concern with internet advertising auctions and recommender systems is that the increased ability to target users based on features could lead to harmful effects on subsets of the population, such as gender or race-based biases in the types of ads or content being shown. This lecture note will take the form of a series of case studies that we will use to then think about each issue, with some initial thoughts on how to address it. However, this lecture note will mostly be focused on identifying problems. The right way to address these problems is not necessarily understood yet.

1.1 Age Discrimination in Job Ads

ProPublica reported in 2017 that many companies were using age as part of their targeting criteria for job ads they were placing on Facebook [2]. The federal *Age Discrimination in Employment Act* of 1967 prohibits bias against people aged 40 or older both in hiring and employment. It is not completely clear whether Facebook or the companies in question are violating this act, since it was written long before the internet era.

1.2 Targeting Housing Ads along Racial Boundaries

ProPublica also reported in 2016 that advertisers were able to run ads for housing while choosing to exclude certain "ethnic affinities" such as "hispanic affinity" or "african-american affinity" on Facebook [1]. Since Facebook does not ask users about race, these affinity categories are stand-in estimates based on user interests and behavior. On the benign side, these features can be used to test for example how an ad in Spanish versus English will perform in a hispanic population.

However, when it comes to topics such as housing, the *Fair Housing Act* from 1968 made it illegal

> "to make, print, or publish, or cause to be made, printed, or published any notice, statement, or advertisement, with respect to the sale or rental of a dwelling that indicates any preference, limitation, or discrimination based on race, color, religion, sex, handicap, familial status, or national origin."

*Department of Industrial Engineering and Operations Research, Columbia University. Email: christian.kroer@columbia.edu.

For the remainder of the lecture, we will operate under the assumption that we wish to ensure various demographic properties of how ads are shown, for ads that are viewed as sensitive. Beyond employment and housing, another category of ads that are viewed as sensitive are credit opportunities, which are also not allowed to discriminate based on demographic properties.

2 Disallowing Targeting

If we wish to prohibit the potential discrimination described above, we could introduce a category of "sensitive ads," where we do not allow age, gender, or racial features to be used as a feature. One might naively think that this would work, but unfortunately there are many ways to perform indirect targeting of these categories. For example, zip code can often be a strong proxy for race, and thus care is needed in order to ensure that we do not allow proxy-based targeting of these sensitive features.

Facebook took such an approach in 2019 [5], based on a settlement with various civil rights organizations. In that approach, they disallow targeting on age, gender, zip code, and "cultural affinities" for what they categorize as sensitive ads. That categorization includes housing, employment, and credit opportunities.

While this approach ensures that a certain type of discrimination cannot occur, it does not necessarily rule out other forms of biases in how ads are serverd.

3 Demographic Fairness Measures

We will next study explicit quantifiable measures of fairness. These can potentially be used to audit whether a given ad or system contains biases, or as guiding measures for how to adaptively change the allocation system in order to ensure unbiasedness.

To make things concrete, suppose we have m users, and a single sensitive ad i. We will assume that each user j is associated with a non-sensitive feature vector w_j, and each user also belongs to one of two demographic groups, A or B, which is considered a sensitive attribute; let g_j denote this group. We let G_A and G_B be the set of all indices denoting users in group A or group B, respectively. As usual, we will use $x_{ij} \in [0, 1]$ to denote the probability that the ad i is shown to user j.

Statistical Parity . This notion of demographic fairness asks that ad i is shown at an equal rate across the two groups, in the following sense:

$$\frac{1}{|G_A|} \sum_{j \in G_A} x_{ij} = \frac{1}{|G_B|} \sum_{j \in G_B} x_{ij}$$

This guarantees that, in aggregate, the groups are being shown the ad at an equal rate.

Next, let's see an example of how statistical parity could be broken even though targeting by demographic is disallow. Suppose that a sensitive ad (say a job ad) wishes to target users in either demographic, and has a value of \$1 per click, with a click-through rate that depends only on w_j and not g_j. Secondly, there's another ad which is not sensitive, which has a value per click of \$2, and click-through rates of 0.1 and 0.6 for groups A and B respectively. Now, the sensitive ad will never be able to win any slots for group B since even with a CTR of 1, their bid will be lower than $0.6 \cdot 2 = 1.2$. As a result, the sensitive ad will be shown only to group A. A concrete example of how this competition-driven form of bias might occur is when the non-sensitive ad is some form of female-focused product such as clothing or make-up.

A potential criticism of this fairness measure is that it does not require the ad to be shown to equally interested users in both groups. Thus, one could for example worry that the ad might end up buying highly relevant slots among one group, and cheap irrelevant slots in the other group in order to satisfy the constraint.

Similar Treatment Similar treatment (ST) asks for an individual-level fairness guarantee: if two users i and k have the same non-sensitive feature vector $w_j = w_k$, then they should be treated similarly regardless of the value of g_j and g_k. A simple version of this principle for ad auctions could be that we require $x_{ij} = x_{ik}$ whenever $w_j = w_k$. However, if the feature space is large, some features are continuous, or we just want this to hold even when users are similar in terms of w_j and w_k, then we need a slightly more complicated constraint. Suppose we have a measure $d(w_j, w_k)$ that measures similarity between feature vectors. Then, ST can be defined as

$$|x_{ij} - x_{ik}| \leq d(w_j, w_k).$$

With this definition, we are asking for more than just equality when $w_j = w_k$; instead we also ask that the difference between x_{ij} and x_{ik} should decrease smoothly as the non-sensitive feature vectors get closer to each other, as measured by d.

4 Implementing Fairness Measures on a Per-Ad Basis

In this section we highlight some difficulties in applying these fairness notions straightforwardly in ad auction markets. We will focus on statistical parity; similar treatment seems even more difficult to implement.

If we consider the hindsight optimization problem faced by an individual ad, we could add a constraint that the ad's allocation satisfies statistical parity.

$$\max_{x_i \in [0,1]^T} \sum_{t=1}^{m} (v_{it} - p_{it}) x_{it} \tag{1}$$

$$\text{s.t.} \sum_{t=1}^{T} p_{it} x_{it} \leq B_i \tag{2}$$

$$\frac{1}{|G_A|} \sum_{j \in G_A} x_{ij} = \frac{1}{|G_B|} \sum_{j \in G_B} x_{ij} \tag{3}$$

However, this constraint is not easy to implement as part of an online allocation procedure, for two reasons. The first is that equality constraints such as this one are harder to handle as part of an online learning procedure, than the simpler "packing constraint" needed for the budgets (a less-than-or-equals constraints with only positive coefficients). The second reason is that we do not know the normalizing factors until the end.

5 Historical Notes

The field of "algorithmic fairness" pioneered a lot of the fairness considerations that we considered in this note, in the context of machine learning. Dwork et al. [4] introduced similar treatment in the context of machine learning classification, and the notion that we use here for ad auction allocation is an adaptation of their definitions. They also study statistical parity in the classification context.

A book-level treatment of fairness in machine learning is given by Barocas et al. [3]. Many of these fairness notions were also previously known in the education testing and psychometrics literature. See the biographical notes in Barocas et al. [3] for an overview of these older works.

References

[1] Julia Angwin and Terry Parris Jr. Facebook lets advertisers exclude users by race. *ProPublica*, 2016. URL `https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race`.

[2] Julia Angwin, Noam Scheiber, and Ariana Tobin. Dozens of companies are using facebook to exclude older workers from job ads. *ProPublica*, 2016. URL `https://www.propublica.org/article/facebook-ads-age-discrimination-targeting`.

[3] Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning*. fairmlbook.org, 2019. `http://www.fairmlbook.org`.

[4] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226. ACM, 2012.

[5] Sheryl Sandberg. Doing more to protect against discrimination in housing, employment and credit advertising. *Facebook*, 2019. URL `https://about.fb.com/news/2019/03/protecting-against-discrimination-in-ads/`.