CS 6241: Project Report

Anthonia Carter (aoc29), Erik Louie (etl43)

Loss of Utility in Recommender Systems with Fairness Constraints

May 21, 2020

1 Introduction

In this project, we investigate the cost of fairness constraints on ranking utility for online recommenders. Online recommenders use ranking systems to present relevant results to a user where an item's position in the ranking primarily determines its level of exposure. AirBnB, LinkedIn, and Uber all use rankings in their systems to display relevant items to their users [Biega et al., 2018]. In some cases, a ranking system may be perceived as biased in its treatment of items to be ranked. For example, news aggregators could bias toward a news authority, leading to filter bubbles. One way to address filter bubbles, for example, is to distribute exposure more evenly among articles. Generalized second price auctions, where advertisers bid for the placements of advertisements, is a multibillion-dollar industry purchasing the exposure of content in ranked search results [Edelman et al., 2007], motivating the need for fair exposure across advertisers in rankings. Furthermore, recent literature has put a spotlight on unfair rankings.

Recent studies have shown bias in ranking results. Kay et al. [2015] found that image search results reflected stereotypes and showed systematic under representation of women for image search results. In another example, positions in rankings have been shown to affect click through rate, which perpetuates the problem of exposure by reinforcing the rankings Craswell et al. [2008]. Ranking positions in search results and last clicked position are shown to effectively predict which links a user will click next [Dupret and Piwowarski, 2008]. Furthermore, multiple clicks on a page can effectively predict a user's next clicks [Guo et al., 2009]. These results show the need to apply fairness to the exposure or attention given documents in search results. Granka [2010] champions the idea of diversity in online search as a way to increase exposure on a per-query level. They describe two factors of diversity on this level: site ownership and information content. Diversity in site ownership captures the comparisons of online and offline media ownership, and diversity in information content provides stronger indicators of the utility and information value provided by different sources. One way to address position bias is to impose fairness constraints on ranking systems.

We use Singh and Joachims [2018]'s computational framework to compute utility-maximizing ranking while imposing fairness constraints. Both gradient methods and convex optimization covered in class could be applied to this problem. However, as convex optimization guarantees a global solution for convex problems, we use this technique to solve a linear program with fairness constraints. We focus on group notions of fairness and study how demographic parity, disparate treatment, and disparate impact influence ranking utility. Each aims to allocate exposure among items across all groups fairly. For each, we ask: how much loss in utility is incurred by utilizing fairness constraints?

We also explore the loss of utility when relaxing the fairness constraints under the α-discriminatory approximation to fairness. We apply a variant of the α-discriminatory approximation of fairness proposed by Woodworth et al. [2017]. The α-discrimination approximation relaxes the strict equality between performance metrics between groups and instead requires that the difference be less than a constant α. Our notion of relaxed constraints is motivated by the desire to balance utility with fairness. Relaxed constraints can be use to apply corrective fairness toward one group due to historical discrimination. For example, a company might have had historically biased hiring practices and desire to highlight a discriminated group, maintain fairness, and optimize utility. We explore the impact of fairness constraints on two real-world datasets: the Yow news recommendation dataset and the MovieLens 1M Dataset [Zhang, 2005, Harper and Konstan, 2015].

Our contribution is the addition of approximate fairness constraints to rankings. We develop a general method to apply approximations of fairness constraints with the framework proposed by Singh and Joachims [2018]. We follow by empirically quantifying the price of fairness under exact

and approximate constraints.

2 Related works

2.1 Fairness in Ranking Systems

With the growing attention of fairness in machine learning algorithms, several attempts exist to address fairness in ranking systems. Singh and Joachims [2018] introduced a framework to add fairness constraints to rankings by first formulating rankings per document as probabilities and demonstrating several common notions of fairness, such as demographic parity. Biega et al. [2018] formulated an individual notion of fairness in parallel with Singh and Joachims [2018], framing the problem with a specific focus on the attention received to individuals over multiple queries, accounting for the amortized equity of attention received.

Others have focused on solving the top-k fair ranking problem to determine a subset of k items from an extensive list of $n >> k$ items that maximize utility and are subject to group fairness constraints. Zehlike et al. [2017] proposes a post-processing method to correct for systematic bias and apply a statistical test on the proportion of the protected group on every prefix of the ranking. Celis et al. [2017] construct a constrained ranking optimization problem by bounding the number of items of a particular type, based on a sensitive attribute that appears in the top k positions of the ranking. The bounded mechanism ensures that no item of a specific type dominates the top rankings.

2.2 Approximations of Fairness Constraints

From the best of our knowledge, there do not exist examples of relaxations in fair rankings. There are approximations to fair reinforcement learning and clustering that result in relaxed outcomes as well as relaxations of fairness in clustering.

Approximating fairness has largely been motivated by improving computational time incurred by adding fairness constraints. In reinforcement learning, Jabbari et al. [2017] apply two relaxations to Joseph et al. [2016]'s fairness definition, which states that an action's quality is its potential long-term discounted reward. The first relaxation is approximate-choice fairness, which requires that an algorithm never selects a worse action with probability much higher than better actions, and the second relaxation is approximate-action fairness, where an algorithm never favors an action with a drastically lower quality compared to better actions. Jabbari et al. [2017] find that under approximate-choice fairness a learning algorithm runs in polynomial time instead of exponential time for the exact fairness constraint.

There have been several approximation results in fair clustering of data. In the fairlet paper, Chierichetti et al. [2017] introduced a $(k + 1 + \sqrt{3} + \epsilon)$-approximation algorithm for k-means to obtain perfectly balanced clusters in order quadratic time. They showed how to extend this to produce relaxed fair clusterings. However, due to quadratic time complexity, their algorithm is prohibitive for high-dimensional large-scale datasets. Bercea et al. [2018] improve these results by obtaining a constant 5-approximation algorithm to the fairlet problem. In order to improve the time complexity requirements, Backurs et al. [2019] use a notion of red-blue clustering that is $(r, b) - fair$ if the cluster achieves a balance$(S) >= \frac{r}{b}$, where balance is a measure of the bias between groups. On the surface, balance appears similar to the relaxation of disparate impact that we apply, but they use hierarchically well-separated trees to generate clusters, which is specific to geometric interpretations of fairness. They were also able to show that given an α-approximation

algorithm to fair clustering, they can achieve a near-linear algorithm with minimal cost to the approximation.

3 Ranking utility under fairness constraints

3.1 Ranking utility and probabilistic rankings

Most utility measures for ranking evaluation are derived from the relevance of the individual items to be ranked. The aim is to maximize the utility of a ranking r of a set of documents $\mathcal{D} = \{d_1, d_2, d_3, \ldots, d_N\}$ for a query q while guaranteeing that r is fair. A general way to express utility measures for information retrieval is

$$U(r|q) = \sum_{u \in \mathcal{U}} P(u|q) \sum_{d \in \mathcal{D}} v(\text{rank}(d|r)) \lambda(\text{rel}(d|u, q)), \tag{1}$$

where $\text{rel}(d|u, q)$ denotes the binary relevance of the document d to a user u, $v(\text{rank}(d|r))$ indicates how much attention document d gets at rank $\text{rank}(d|r)$, and λ maps the relevance of the document for a user to its utility. The relevance of a document could also represent a Likert scale as seen in movie ratings or a real-valued score. The position bias, denoted as v, is the fraction of users who examine the document shown at a specific position out of the total number of users who issue the query q [Singh and Joachims, 2018]. Since rankings are combinatorial objects, naively searching the space of all ranking would take time that is exponential in $|\mathcal{D}|$. Therefore, we consider probabilistic rankings. A probabilistic ranking R can be used instead of a single deterministic ranking r:

$$U(R|q) = \sum_{r} R(r) \sum_{d \in \mathcal{D}} v(\text{rank}(d|r)) u(d|q) \tag{2}$$

Here,

$$u(d|q) = \sum_{u \in \mathcal{U}} \lambda(\text{rel}(d|u, q)) P(u|q) \tag{3}$$

is the expected utility of a document d for query q. Let $\mathbf{P}_{ij}$ be the probability that R places document d_i at rank j. Matrix $\mathbf{P}$ forms a doubly stochastic matrix of size $N \times N$. The sum of each row (probabilities for each document) and column (probabilities for each position) is 1. The utility of the ranking can be written as the matrix product $U(\mathbf{P}|q) = \mathbf{u}^T \mathbf{P} \mathbf{v}$, where $\mathbf{u}$ and $\mathbf{v}$ are column vectors of size N with $\mathbf{u}_i = u(d_i|q)$ and $\mathbf{v}_j = v(j)$.

3.2 Fair ranking optimization

Using Singh and Joachims [2018]'s model for optimizing fair rankings via linear programming, our problem is to find a doubly stochastic matrix $\mathbf{P}$ that maximizes the expected utility subject to a fairness constraint. More formally,

$$\mathbf{P} = \underset{\mathbf{P}}{\text{argmax}} \ \mathbf{u}^T \mathbf{P} \mathbf{v} \tag{4}$$

$$\text{s.t.} \quad \mathbb{1}^T \mathbf{P} = \mathbb{1}^T \ \wedge \ \mathbf{P}\mathbb{1} = \mathbb{1} \ \wedge \ 0 \leq \mathbf{P}_{i,j} \leq 1 \ \wedge \ \mathbf{P} \text{ is fair.} \tag{5}$$

The fairness constraint can be expressed in the linear form of

$$\mathbf{f}^T \mathbf{P} \mathbf{g} = h. \tag{6}$$

The vectors $\mathbf{f}$, $\mathbf{g}$, and scalar h can be chosen for a range of different fairness conditions. The vector $\mathbf{f}$ might encode group identity and the relevance of each document, and $\mathbf{g}$ can reflect the importance of each position.

In this paper, we will be concerned with the cost of fairness. Specifically, cost of fairness can be defined as the loss of expected utility, or

$$\text{CoF} = \mathbf{u}^T(\mathbf{P}^* - \mathbf{P})\mathbf{v}. \tag{7}$$

Fairness is inversely related to utility. In the extreme case, having all of the utility held by one group and the second group's size going to infinity, the fairness constraint would reduce the expected utility toward zero. We hypothesize, however, that in real datasets the effect of fairness will be less insidious on expected utility, as most datasets are not imbalanced in this manner. We explore the consequences in the experimental results.

3.3 Group fairness constraints

To implement the constraint $\mathbf{P}$ is fair in the linear program above, Singh and Joachims [2018] formulate three fairness constraints assuming binary valued sensitive attributes (G_0 and G_1): demographic parity, disparate treatment, and disparate impact. They define exposure for a document d_i under a probabilistic ranking $\mathbf{P}$ as

$$\text{Exposure}(d_i|\mathbf{P}) = \sum_{j=1}^{N} \mathbf{P}_{i,j}\mathbf{v}_j. \tag{8}$$

Here, $\mathbf{v}_j$ represents the importance of position j, which is the fraction of users that examine the item at this position. The average exposure of documents for a group is

$$\text{Exposure}(G_k|\mathbf{P}) = \frac{1}{|G_k|} \sum_{d_i \in G_k} \text{Exposure}(d_i|\mathbf{P}). \tag{9}$$

This leads to the definition of demographic parity as equality in average exposure between groups:

$$\text{Exposure}(G_0|\mathbf{P}) = \text{Exposure}(G_1|\mathbf{P}) \tag{10}$$

$$\Leftrightarrow \quad \text{Exposure}(G_0|\mathbf{P}) - \text{Exposure}(G_1|\mathbf{P}) = 0. \tag{11}$$

Singh and Joachims [2018] show that this constraint is equivalent to $\mathbf{f}^T\mathbf{P}\mathbf{v} = 0$ with $\mathbf{f}_i = \left(\frac{\mathbb{1}_{d_i \in G_0}}{|G_0|} - \frac{\mathbb{1}_{d_i \in G_1}}{|G_1|}\right)$.

In some instances, it is appropriate for relevance to be proportional to exposure, which motivates the use of disparate treatment or disparate impact as a fairness constraint. Sometimes, small differences in relevance might lead to significant differences in exposure. Instead, we consider the average utility of a group defined as

$$U(G_k|q) = \frac{1}{|G_k|} \sum_{d_i \in G_k} \mathbf{u}_i. \tag{12}$$

Disparate treatment constraint requires the exposure of the two groups to be proportional to their average utility:

$$\frac{\text{Exposure}(G_0|P)}{U(G_0|q)} = \frac{\text{Exposure}(G_1|P)}{U(G_1|q)}, \tag{13}$$

which is equivalent to $\mathbf{f}^T\mathbf{P}\mathbf{v} = 0$ with $\mathbf{f}_i = \left(\frac{\mathbb{1}_{d_i \in G_0}}{|G_0|U(G_0|q)} - \frac{\mathbb{1}_{d_i \in G_1}}{|G_1|U(G_1|q)}\right)$ [Singh and Joachims, 2018]. In other scenarios, it may be appropriate to set a constraint on the impact, such as expected click though rate, which more directly reflects the economic impact of the ranking. Richardson et al. [2007] defines the probability of a document d_i being clicked as:

$$P(\text{click on document } i) = P(\text{examining } i) \times P(i \text{ is relevant}) \tag{14}$$

$$= \text{Exposure}(d_i|\mathbf{P}) \times P(i \text{ is relevant}) \tag{15}$$

$$= \left(\sum_{j=1}^{N} \mathbf{P}_{ij} v_j\right) \times u_i \tag{16}$$

This probability is generalized as the average click through rate of documents in group G_k:

$$\text{CTR}(G_k|\mathbf{P}) = \frac{1}{|G_k|} \sum_{i \in G_k} \sum_{j=1}^{N} \mathbf{P}_{i,j} \mathbf{u}_j \mathbf{v}_j. \tag{17}$$

Singh and Joachims [2018] define the disparate impact constraint as enforcing the expected click through rate of each group to be proportional to its average utility:

$$\frac{\text{CTR}(G_0|\mathbf{P})}{U(G_0|q)} = \frac{\text{CTR}(G_1|\mathbf{P})}{U(G_1|q)}, \tag{18}$$

which is equivalent to $\mathbf{f}^T\mathbf{P}\mathbf{v} = 0$ with $\mathbf{f}_i = \left(\frac{\mathbb{1}_{d_i \in G_0}}{|G_0|U(G_0|q)} - \frac{\mathbb{1}_{d_i \in G_1}}{|G_1|U(G_1|q)}\right) \mathbf{u}_i$.

3.4 α-discriminatory relaxations on fairness constraints

In our project, we relax fairness constraints by applying a variant of the α-discriminatory approximation of predictor fairness proposed by Woodworth et al. [2017]. The α-discrimination approximation relaxes the strict equality between performance rates between groups and instead requires that the difference be less than a constant α. By using a approximation, we can limit the loss of utility while balancing fairness. To define the α-discriminatory fairness constraints, we will use the demographic parity ratio (DPR), disparate treatment ratio (DTR), and disparate impact ratio (DIR). Each ratio measures how much their respective fairness constraints are violated.

$$\text{DPR}(G_0, G_1|\mathbf{P}, q) = \frac{\text{Exposure}(G_0|\mathbf{P})}{\text{Exposure}(G_1|\mathbf{P})} \tag{19}$$

$$\text{DTR}(G_0, G_1|\mathbf{P}, q) = \frac{\text{Exposure}(G_0|\mathbf{P})/U(G_0|q)}{\text{Exposure}(G_1|\mathbf{P})/U(G_1|q)}, \tag{20}$$

$$\text{DIR}(G_0, G_1|\mathbf{P}, q) = \frac{\text{CTR}(G_0|\mathbf{P})/U(G_0|q)}{\text{CTR}(G_1|\mathbf{P})/U(G_1|q)} \tag{21}$$

Notice that the ratio is 1 when the groups are fairly treated, otherwise the ratio indicates which group receives preferential treatment. Intuitively, the exposure or average clickthrough rate from one group to another should not be greater than some margin from another group. While we could add a constant to the original fairness constraint, this lacks interpretability and generalizability. Instead, we adopt the notion used in the real-world usage of DIR to identify disparate impact between one group and another by the 80% rule [Feldman et al., 2015]. In this case, we could set $\alpha = 0.80$. We derive the α-discriminatory approximations for all fairness constraints below.

3.4.1 Demographic parity α-discriminatory approximation

We can require that:

$$\frac{\text{Exposure}(G_0|\mathbf{P})}{\text{Exposure}(G_1|\mathbf{P})} \geq \alpha \tag{22}$$

$$\Leftrightarrow \quad \text{Exposure}(G_0|\mathbf{P}) \geq \alpha \cdot \text{Exposure}(G_1|\mathbf{P}) \tag{23}$$

$$\Leftrightarrow \quad \text{Exposure}(G_0|\mathbf{P}) \; - \; \alpha \cdot \text{Exposure}(G_1|\mathbf{P}) \geq 0, \tag{24}$$

and rewrite this inequality as:

$$\frac{1}{|G_0|} \sum_{d_i \in G_0} \sum_{j=1}^{N} \mathbf{P}_{ij} v_j - \alpha \cdot \frac{1}{|G_1|} \sum_{d_i \in G_1} \sum_{j=1}^{N} \mathbf{P}_{ij} v_j \geq 0 \tag{25}$$

$$\Rightarrow \quad \sum_{d_i \in \mathcal{D}} \sum_{j=1}^{N} \left(\frac{\mathbb{1}_{d_i} \in G_0}{|G_0|} - \alpha \cdot \frac{\mathbb{1}_{d_i} \in G_1}{|G_1|} \right) \mathbf{P}_{ij} v_j \geq 0 \tag{26}$$

$$\Rightarrow \quad {\mathbf{f_1}}^T \mathbf{P}\mathbf{v} \geq 0 \tag{27}$$

where $\mathbf{f}_{1_i} = \left(\frac{\mathbb{1}_{d_i} \in G_0}{|G_0|} - \alpha \cdot \frac{\mathbb{1}_{d_i} \in G_1}{|G_1|} \right)$. Additionally, we can require

$$\text{Exposure}(G_1|\mathbf{P}) - \alpha \cdot \text{Exposure}(G_0|\mathbf{P}) \geq 0, \tag{28}$$

which yields an additional constraint

$${\mathbf{f_0}}^T \mathbf{P}\mathbf{v} \geq 0 \tag{29}$$

where $\mathbf{f}_{0_i} = \left(\frac{\mathbb{1}_{d_i} \in G_1}{|G_1|} - \alpha \cdot \frac{\mathbb{1}_{d_i} \in G_0}{|G_0|} \right)$.

We define (29) and (27) as demographic parity α-discriminatory constraints. These constraints give interpretability to the results: no group should receive less than $\alpha\%$ of the exposure of any other group.

3.4.2 Disparate treatment α-discriminatory approximation

We define measures for the disparate treatment α-discriminatory approximation as:

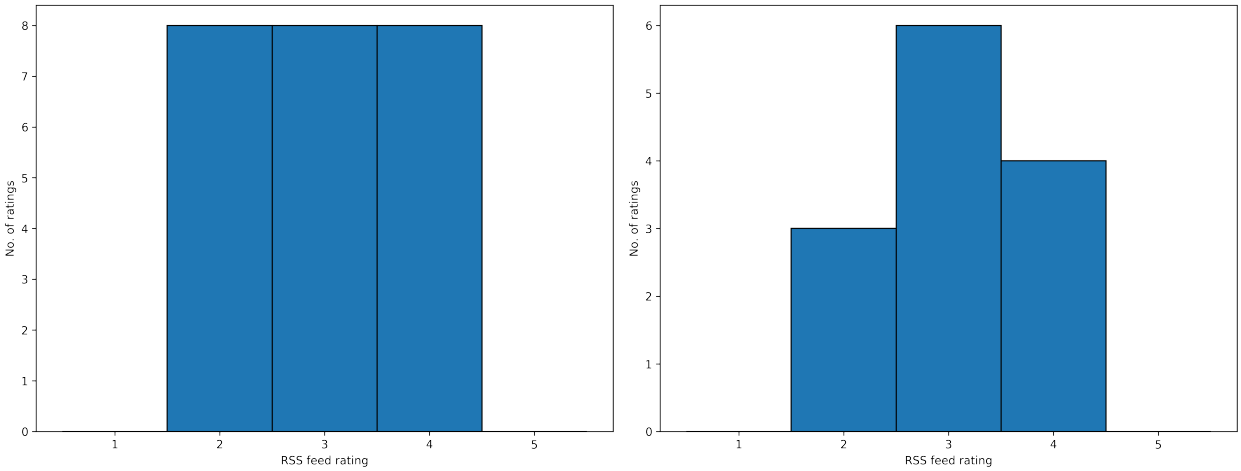
$$\text{Exposure}(G_0|\mathbf{P})/U(G_0|q) - \alpha \cdot \text{Exposure}(G_1|\mathbf{P})/U(G_1|q) \geq 0 \tag{30}$$

$$\text{Exposure}(G_1|\mathbf{P})/U(G_1|q) - \alpha \cdot \text{Exposure}(G_0|\mathbf{P})/U(G_0|q) \geq 0. \tag{31}$$

The two inequalities produce two constraints, namely

$${\mathbf{f_1}}^T \mathbf{P}\mathbf{v} \geq 0 \quad \text{with } \mathbf{f}_{1_i} = \left(\frac{\mathbb{1}_{d_i} \in G_0}{|G_0|U(G_0|q)} - \alpha \cdot \frac{\mathbb{1}_{d_i} \in G_1}{|G_1|U(G_1|q)} \right) \quad \text{and} \tag{32}$$

$${\mathbf{f_0}}^T \mathbf{P}\mathbf{v} \geq 0 \quad \text{with } \mathbf{f}_{0_i} = \left(\frac{\mathbb{1}_{d_i} \in G_1}{|G_1|U(G_1|q)} - \alpha \cdot \frac{\mathbb{1}_{d_i} \in G_0}{|G_0|U(G_0|q)} \right). \tag{33}$$



Distribution of RSS1 feed ratings

Distribution of RSS2 feed ratings

Figure 1: Distribution of ratings for news articles.

3.4.3 Disparate impact α-discriminatory approximation

We define measures for the disparate impact α-discriminatory approximation as:

$$\mathrm{CTR}(G_0|\mathbf{P})/U(G_0|q) - \alpha \cdot \mathrm{CTR}(G_1|\mathbf{P})/U(G_1|q) \geq 0 \tag{34}$$

$$\mathrm{CTR}(G_1|\mathbf{P})/U(G_1|q) - \alpha \cdot \mathrm{CTR}(G_0|\mathbf{P})/U(G_0|q) \geq 0. \tag{35}$$

The two inequalities produce two constraints, namely

$$\mathbf{f_1}^T\mathbf{P}\mathbf{v} \geq 0 \quad \text{with } \mathbf{f}_{1_i} = \left(\frac{\mathbb{1}_{d_i} \in G_0}{|G_0|U(G_0|q)} - \alpha \cdot \frac{\mathbb{1}_{d_i} \in G_1}{|G_1|U(G_1|q)} \right) \mathbf{u}_i \quad \text{and} \tag{36}$$

$$\mathbf{f_0}^T\mathbf{P}\mathbf{v} \geq 0 \quad \text{with } \mathbf{f}_{0_i} = \left(\frac{\mathbb{1}_{d_i} \in G_1}{|G_1|U(G_1|q)} - \alpha \cdot \frac{\mathbb{1}_{d_i} \in G_0}{|G_0|U(G_0|q)} \right) \mathbf{u}_i. \tag{37}$$

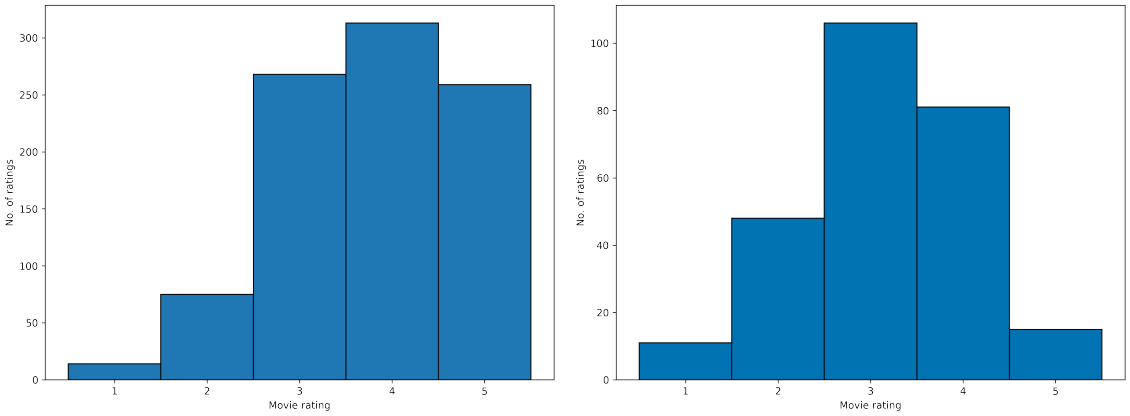
Note that α need not be equivalent for both classes. ie. we could define α_i as the permitted discrimination toward group i. For example, a user may desire to see more relevant articles from one news source or another, but still maintaining a balance from both results.

4 Datasets

We explore the impact of fairness constaints on two real-world datasets: the Yow news recommendation dataset and the MovieLens 1M Dataset.

4.1 News recommendation dataset

Singh and Joachims [2018] use the Yow news recommendation dataset [Zhang, 2005] in their experimental results. The dataset includes users, news authorities, and relevance of RSS feeds to the user. We replicate Singh and Joachims [2018]'s study by randomly sampling a subset of news articles in the "people" topic from the top two sources based on the RSS feed identifier. We use the "relevant" field as measure of relevance. Relevance is given as a rating from 1 to 5, so we normalize by dividing by 5 and resolve ties by adding Gaussian noise with $\epsilon = 0.05$. In total, there are 37



Drama ratings Action ratings

Figure 2: User distribution of ratings for movies divided by genre.

articles rated from the top two new sources. Figure (1) gives the distribution of ratings for the top two news sources. The mean of the ratings for RSS1 is 3 and the variance is 0.6667. The mean of the rating for RSS2 is 3.0769 and the variance is 0.5325. The ratings are slightly biased towards RSS2. New articles from RSS1 might rank slightly lower in the unfair ranking. We do not expect the utility of the fair rankings to be significantly lower than the unfair ranking. Given that the two datasets exhibit a similar distribution of ratings, a limitation of this dataset is that fair and unfair rankings are unlikely to be substantially different.

4.2 Movie ratings dataset

The MovieLens 1M Dataset [Harper and Konstan, 2015] is a recommendation dataset from the University of Minnesota that includes over 1 million users ratings for movies. We use the ratings (from 1 to 5) as the relevance feature. We included the ratings of the user with the highest number of submitted ratings. We selected the genre with the highest number of submitted ratings, dramas, as well as the genre with the largest number of submitted ratings with a significantly different distribution, action films. The user has 2314 ratings, drama has 929 ratings and action has 261 ratings. Figure (2) gives the distribution of ratings for the selected genres. The ratings are biased in favor of dramas with a mean of 3.784 and variance 0.979 and biased against action films with a mean 3.157 and variance 0.868. Action films will likely rank lower in the unfair ranking, since the relevance is lower. As such, we expect that the utility of a fair ranking to be significantly lower than the unfair ranking. A problem arises in this dataset as genres exist that include both "action" and "drama" leading to ambiguity in resolving group membership. Given that these could theoretically improve the rankings of one genre or another and vastly change the unfair rankings, we chose to remove ambiguous results from the dataset. It would be a worthy followup to explore how to handle ambiguous group membership.

5 Experimental results

5.1 Cost of Fairness

As mentioned in Section 3.2, fairness constraints typically inversely affect expected utility, measured as cost of fairness (CoF). It is expected that real-world datasets will typically exhibit low CoF,

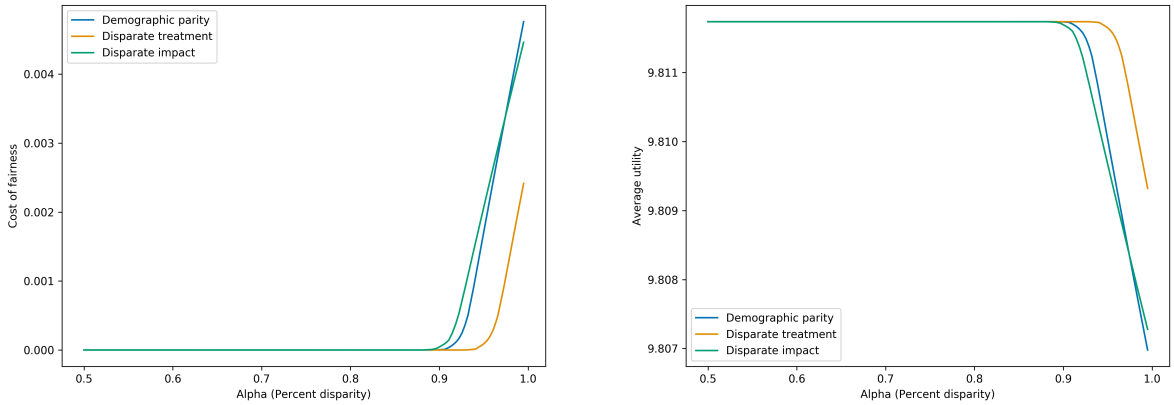


Figure 3: Measuring the cost of fairness and average utility of each fairness constraint with the Yow dataset

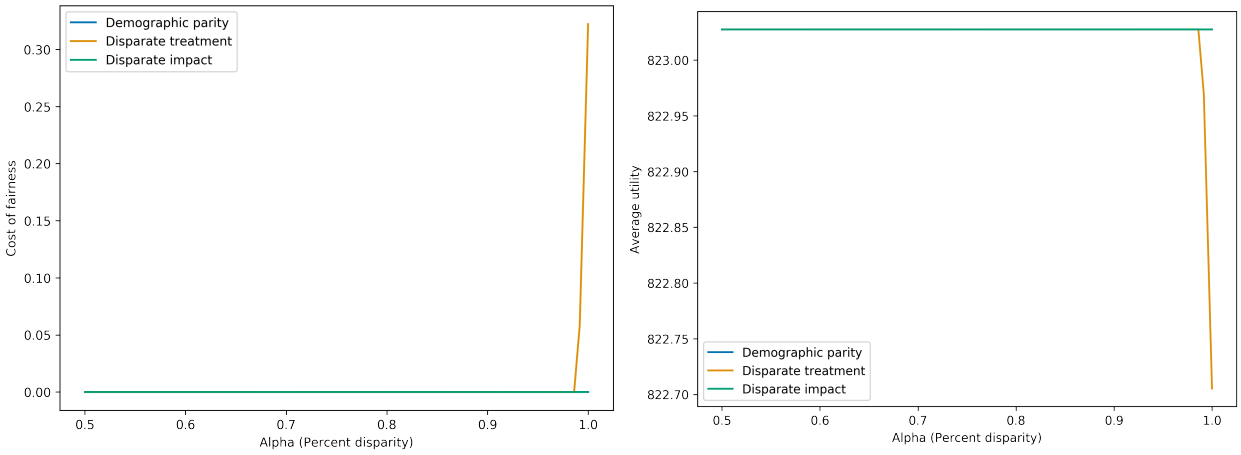


Figure 4: Measuring the cost of fairness and average utility of each fairness constraint with the MovieLens 1M dataset between drama and action.

since they do not reflect worst-case conditions. We measured this against the Yow dataset, as demonstrated in Figure (3). For all strict fairness constraints ($\alpha = 1$), there is a marginal cost of fairness, which matches the expected behavior. When $\alpha < 0.9$ the fairness constraints have no impact on the CoF, within some margin of error. However, beginning with disparate impact, each relaxed fairness constraint increases in CoF when $\alpha > 0.9$.

Generally, the effect of fairness on utility is low, matching the hypothesis in Section 3.2. Items are largely distributed evenly, so rankings are unlikely to change drastically and increase CoF. For example, in the 1M dataset, the distributions of both datasets roughly matches a normal curve, so the worst-case divide of high-utility in one group and a large low-utility second group is not reflected in the data, leading to a reduced impact on utility.

After the initial results giving low difference in utility with fair constraints on the 1M dataset, we posited that fair constraints lead to lower changes in utility for large, well-distributed datasets, but would be more impacting toward smaller datasets. There are two reasons for this. Firstly, we compute a probability distribution over a larger number of rankings, which reduces the magnitude of changes in expected utility. Secondly, the relevance in these datasets are discrete over 1 through 5, which causes most changes of relevance to be within the same relevance class and leading to reduced changes in utility. If an item from one group is 3 relevance and another group is also 3 relevance, then there is no change in utility if they swap positions, but their exposure is fairer. This leads to cost of fairness to be negligible for large datasets with few classes of relevance. We expect smaller datasets will lead to more discrete relevance class changes and a higher utility cost to fairness-based changes. We formulated a second experiment where we sampled uniformly at random 20 data points each from the drama and action ratings to roughly match the distribution of the original dataset. Action included several 3's and a single 5, whereas drama included several 5's, so the likelihood of lower relevance exchanging rank with a more relevant item given a fair constraint is high, which should increase the impact on cost of fairness. The results in Figure (5) matches the hypothesis, with all fairness constraints exhibiting cost of fairness.

We see in Figure (3, 4, 5) that after a threshold, all fairness constraints initially follow a curve and after another threshold become linearly increasing in the cost of fairness. This result suggests that it may be possible to estimate the line of the cost of fairness by using at least three points: the unrelaxed fair constraint, a relaxed fair constraint, and the unfair optimization utility. Given these, we could determine an optimal balance between the cost of fairness and the unfair constraint in utility. That is, estimating the cost of relaxed fairness is a constant order more time complexity than obtaining the initial unrelaxed fair constraint.

The linear program works over a large solution space. Even with a few thousand data points for a single user can be impractical to compute in real time, whereas companies have datasets that include millions of points per user. Biega et al. [2018] found similar challenges in their implementation of amortized fairness and suggested LP relaxations or greedy solutions. A clearer path to optimization of the approximate fairness as proposed here, aside from the aforementioned possible linear estimation, would be to use a parametric linear program solver, such as proposed by Yu and Monniaux [2019]. Parametric linear program solvers attempt to find the optimal value for an uncertain constraint, such as α above. This would reduce the problem from a search space over possible values of alpha to solving a single parametric linear program.

6 Conclusion

In this project we applied approximate fairness constraints to Singh and Joachims [2018]'s framework for fair ranking problems through the lens of exposure rates between groups. We measured

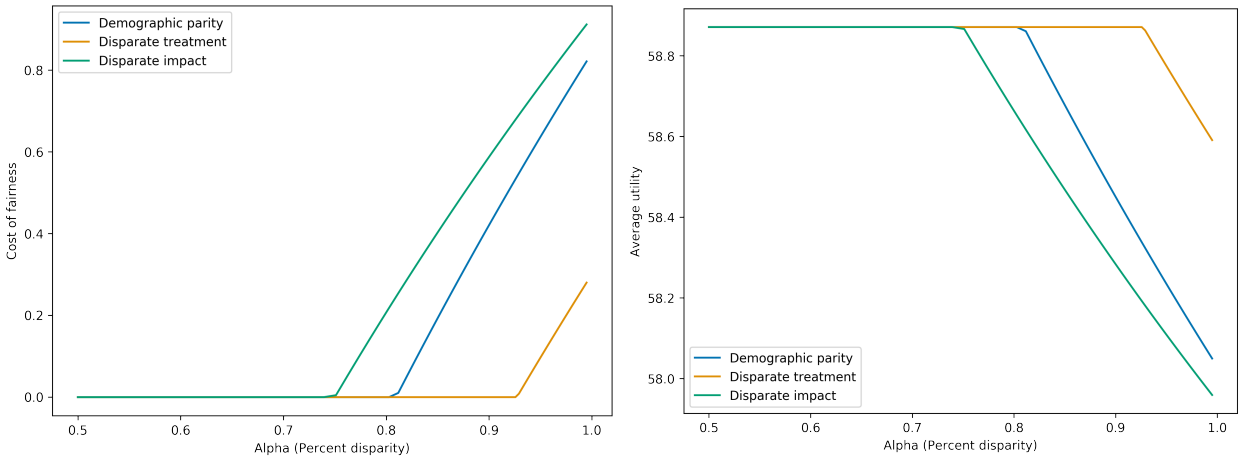


Figure 5: Measuring the cost of fairness and average utility of each fairness constraint with the MovieLens 1M dataset between drama and action using 20 review from each.

and compared the cost of exact and approximate fairness constraints on ranking utility. We discovered that fairness constraints marginally affect ranking utility. Empirically, we find that estimating the cost of relaxed fairness is a constant order more time complexity than the unrelaxed fairness constraint. Some possible future directions for our work are:

- Using an approximation algorithm to offset the time complexity of large-scale rankings (ie. using parametric linear programming to test many alphas).
- Investigating the possibility of a linear estimator for the approximate fairness.
- Apply the techniques used here to datasets that include items over a continuous relevance to determine the impact of discrete vs continuous relevance over CoF.
- Determine the impact of alpha-approximation on multigroup fairness (ie. perfect equality requires constraints linear in the number of groups, but alpha-approximation requires more).
- Generalized results on worst-case or average-case analysis on the impact of alpha-approximation.

References

Arturs Backurs, Piotr Indyk, Krzysztof Onak, Baruch Schieber, Ali Vakilian, and Tal Wagner. Scalable fair clustering. *arXiv preprint arXiv:1902.03519*, 2019.

Ioana O Bercea, Martin Groß, Samir Khuller, Aounon Kumar, Clemens Rösner, Daniel R Schmidt, and Melanie Schmidt. On the cost of essentially fair clusterings. *arXiv preprint arXiv:1811.10319*, 2018.

Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*, pages 405–414, 2018.

L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. Ranking with fairness constraints. *arXiv preprint arXiv:1704.06840*, 2017.

Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. In *Advances in Neural Information Processing Systems*, pages 5029–5037, 2017.

Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the 2008 international conference on web search and data mining*, pages 87–94, 2008.

Georges E Dupret and Benjamin Piwowarski. A user browsing model to predict search engine click data from past observations. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 331–338, 2008.

Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American economic review*, 97(1):242–259, 2007.

Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, page 259–268, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336642. doi: 10.1145/2783258.2783311. URL https://doi.org/10.1145/2783258.2783311.

Laura A Granka. The politics of search: A decade retrospective. *The Information Society*, 26(5): 364–374, 2010.

Fan Guo, Chao Liu, and Yi Min Wang. Efficient multiple-click models in web search. In *Proceedings of the second acm international conference on web search and data mining*, pages 124–131, 2009.

F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context, 2015.

Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. Fairness in reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1617–1626. JMLR. org, 2017.

Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*, pages 325–333, 2016.

Matthew Kay, Cynthia Matuszek, and Sean A Munson. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3819–3828, 2015.

Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting clicks: Estimating the click-through rate for new ads. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, page 521–530, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595936547. doi: 10.1145/1242572.1242643. URL https://doi.org/10.1145/1242572.1242643.

Ashudeep Singh and Thorsten Joachims. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, page 2219–2228, New York, NY, USA, 2018. Association for Computing Machinery. doi: 10.1145/3219819.3220088. URL https://doi.org/10.1145/3219819.3220088.

Blake Woodworth, Suriya Gunasekar, Mesrob I Ohannessian, and Nathan Srebro. Learning non-discriminatory predictors. *arXiv preprint arXiv:1702.06081*, 2017.

Hang Yu and David Monniaux. An efficient parametric linear programming solver and application to polyhedral projection. In *International Static Analysis Symposium*, pages 203–224. Springer, 2019.

Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. Fa*ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, page 1569–1578, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450349185. doi: 10.1145/3132847.3132938. URL https://doi.org/10.1145/3132847.3132938.

Yi Zhang. Bayesian graphical models for adaptive information filtering, 2005. URL https://users.soe.ucsc.edu/~yiz/papers/data/YOWStudy/.