Facets of Fairness in Search and Recommendation

Sahil Verma*, *Ruoyuan Gao and Chirag Shah*
University of Washington and Rutgers University

April 14, 2020

Search and Recommendations

- A universal way to retrieve information
- Search engines are an integral part of all the work we accomplish
- Store petabytes of data from which relevant information is retrieved for each query
- Search engines also serve as a platform for showing advertisements

Different dimensions of search

Information retrieval literature has developed a few metrics which are used to quantify quality of search results

- **Relevance** - Search results which are pertinent to the query are considered relevant
- **Diversity** - Search results should be gathered from diverse topics related to a query
- **Novelty** - Each search result should bring in new information. This helps avoid overlapping content

Bias in Search and Recommendations

- With the widespread use of machine learning, search engines learn using the feedback provided from its users
- It also draws correlations from historical search results
- This can cause bias in the results and adversely affect opinions and knowledge

Fairness as a new dimension

- The new emerging dimension in search and recommendations is fairness
- Fairness is important for equity. We do not want to propagate biases in the future world.
- Fairness can be enforced at various stages in a machine learning system
- In this paper, we would limit ourselves to talking about various fairness metrics literature has come up in different settings

Major Recommendation settings

In this paper, we talk about five major settings which have been focused in the literature

- General Non-Personalized Recommendation Settings - *accuracy* and *error* based metrics
- Crowd-Sourced Non-Personalized Recommendation Settings
- Personalized Recommendation Settings
- Advertisement Settings
- Marketplace Settings

Accuracy based metrics in non-personalized recommendation setting

Some representative metrics:

- *Demographic parity* - In a ranked list of candidates, it proposes to have a proportional representation of all demographic groups according to the underlying distribution
- *Top-k fairness* - Proposes to have some required proportion of protected group members in the top-k ranking.
- Top-k fairness becomes equivalent to demographic parity if the required proportion is the same as the underlying proportion

Error based metrics in non-personalized recommendation setting

These metrics assume the existence of the ground-truth ranking of each candidate, which is seldom available. They have roots in error based fairness metrics in classification tasks.

- *Rank parity Error* - It has roots in demographic parity. Rank parity error is the number of times a member of the protected group is falsely ranked lower than a member of another group. The score is summed over each such inverted pair.

Metrics in Crowd-sourced non-personalized recommendation setting

An example of this setting is the top-10 trending stories on Twitter or Yelp. This could be highly affected by a hyperactive group. Relevant counteractions are:

- *Equality of Voice* - Each user gets to vote only once.
- *Anti-Plurality* - An item that has been disliked by a majority of users doesn't make it to the list. This is to counteract the division of votes among similar candidates

Metrics in personalized recommendation setting

In this setting, only a subset of relevant results is also part of personalization. A recommendation would ideally consist of items that a user would click

- *Pairwise fairness* - Probability of a clicked candidate to be ranked higher than an unclicked candidate is same across all demographic groups
- It doesn't eliminate systematic bias against a group. Enforcing inter and intra group pairwise fairness helps.

Metrics in advertisement recommendation setting

Until now fairness metrics have considered the perspective of a ranked candidate, now we see metrics from the perspective of users

- *Inter-category envy-freeness* - Each user gets to specify their choice, and it requires each user to be shown the same amount of ads from each preferred category
- *Total Variation Fairness* - It requires equal distribution of ads within each category to be shown to users, e.g., it promotes high-paying jobs to women.

Metrics in marketplace recommendation setting

In a marketplace, concerns related to consumers, suppliers, and side-stakeholders need to be addressed. It is a compromise between all the players

- *Consumer fairness* - If a service doesn't cause any disparate impact on its consumers, e.g., preferences of all users is weighed equally
- *Provider Fairness* - All providers have an equal chance of exposure to consumers. This helps evade the super-star economic situation.

Conclusions

- We collected and explained 25 fairness definitions used in various recommendation settings
- Wherever possible, we exemplified the relation between them and the relation with fairness metrics used in classification
- Applicability of several definitions depend upon the existence of ground truth which requires to be addressed
- Consensus on metrics does not exist. A lot of work required to unify them