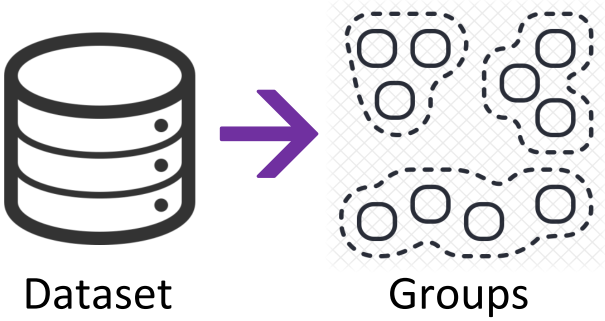
Towards Fairness-Aware Ranking by Defining Latent Groups Using Inferred Features

Yunhe Feng[¶], Daniel Saelid[¶], Ke Li[¶], Ruoyuan Gao[†], Chirag Shah[¶]

[¶]University of Washington
[†]Rutgers University

Group-level Fairness of Exposure in Rankings

RQ1: How to infer features to construct groups for fair ranking?



RQ2: How to strike a balance between relevance and fairness in rankings?

Why Fair Ranking Matters in Academic Search?

VS

Example: Authors in the groups of small institutes and big universities.
Authors in small intuitions have limited media outlets and resources. Their research work should also be treated equally to get its deserved exposures in search systems.

Challenges of Fair Ranking in Academic Search

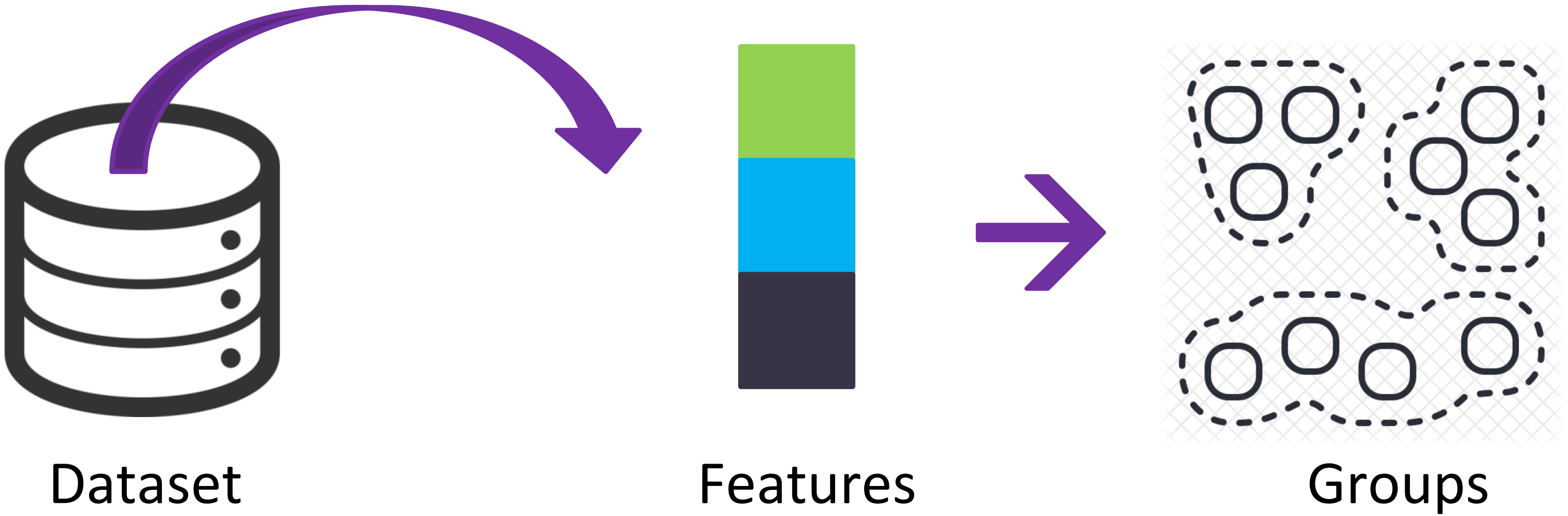
- Openness and complexity of defining the author group.
- Fair ranking algorithms’ robustness across applications.
- Trade-off between relevance and fairness.

Group Definition

- When defining author groups, we considered **genders** and **countries** of authors because the two demographic features are general enough for different applications.

- Re-ranking algorithms based on such group definitions are more likely to demonstrate strong robustness in various scenarios.

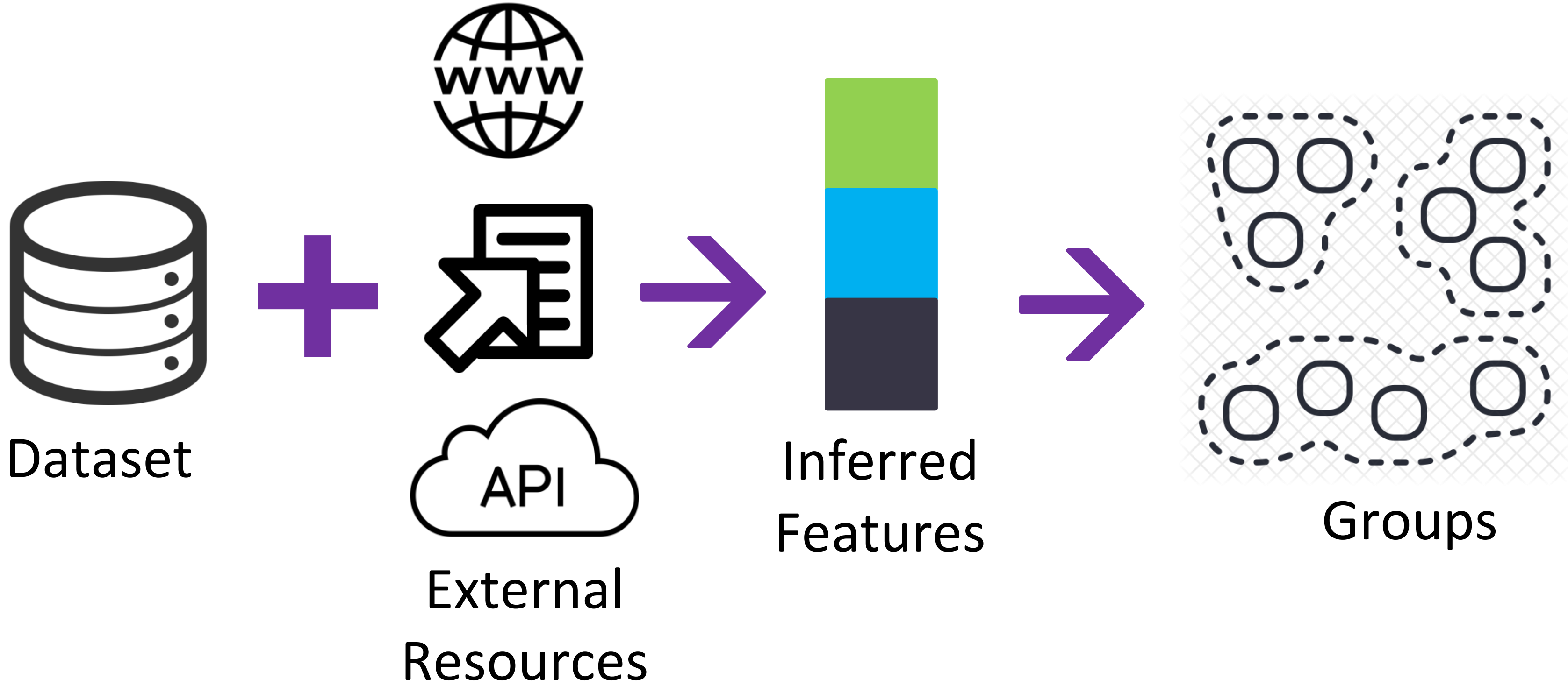
When Features Are Self-Contained



No Gender and Country Info in S2ORC Dataset

- Semantic Scholar Research Open Corpus (S2ORC), released by Allen Institute for AI [1] and used in TREC 2020 Fair Ranking Track [2].
- Features:
  - Paper title
  - Abstract
  - Author name
  - Number of citations
  - ……

Define Latent Groups Using Inferred Features



Gender Inference Using genderize.io API [3]

The genderize.io API is powered by a large dataset that maps first names to binary genders.

Given a name, genderize.io will return *male* if there are more instances of the name associated with men, and it will return *female* otherwise.

Table 1: The distribution of inferred genders by genderize.io

Gender	Count	Percentage
Male	18810	58.8%
Female	6235	19.5%
Unidentified	6930	21.7%
Total	31975	100%

Country Inference - Methods

- Search for the author by name in Google Scholar using the Scholarly API [4]
    1. Parse the email extension for a country code (e.g. .uk -> the United Kingdom).
    2. Parse the affiliation for a university name, then return the country in which that university is located.
    3. Parse the affiliation for a city name, then return that city's country.
    4. Search author name, author affiliation on Google, scrape the first URL, then parse for country code.
    5. Call Google Places API with author affiliation, then return associated country.

- Search author name + *homepage* on Google, scrape the first URL, then parse for country code.

Country Inference - Results

Once all authors had been processed, we mapped each author’s affiliated country to ***advanced economy*** or ***developing economy*** based on the IMF’s October 2019 World Economic Outlook report [5].

Table 2: The economy distribution of inferred locations

Locations	Count	Percentage
Advanced	15106	47.2%
Developing	3926	12.3%
Unidentified	12933	40.5%
Total	31975	100%

Balance Relevance & Fairness - Cost Function

$$C(d, \mathbf{w}, \mathbf{R}, \mathcal{D}', q) = w_r * F(d, \mathcal{D}', q) + \sum_{v \in \{g,c\}} w_v * KL(p(v, \mathbf{R} + \{d\}) \parallel p(v, \mathcal{D}'))$$

where d : the document to be added into ranked list $\mathbf{R}$;

$\mathbf{w} = \{w_r, w_g, w_c\}; w_r + w_g + w_c = 1$: the cost weight;

$\mathcal{D}'$: the candidate corpus;

q : the query;

$F(d, \mathcal{D}', q)$: the reversed normalized BM25 score;

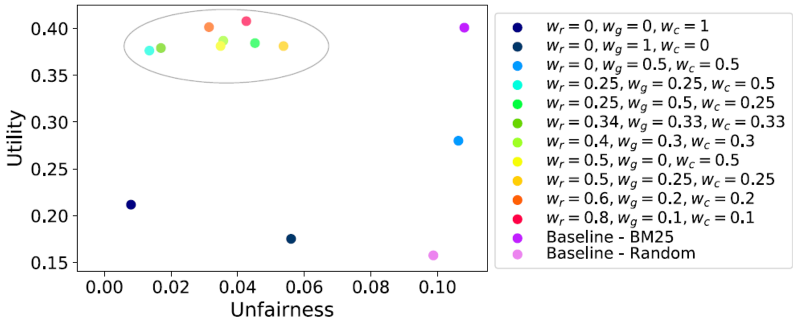
$KL(p(v, \mathbf{R} + \{d\}) || p(v, \mathcal{D}'))$: the Kullback-Leibler divergence regarding group v between the updated $\mathbf{R}$ and $\mathcal{D}'$.

Fairness-aware Re-ranking Algorithm

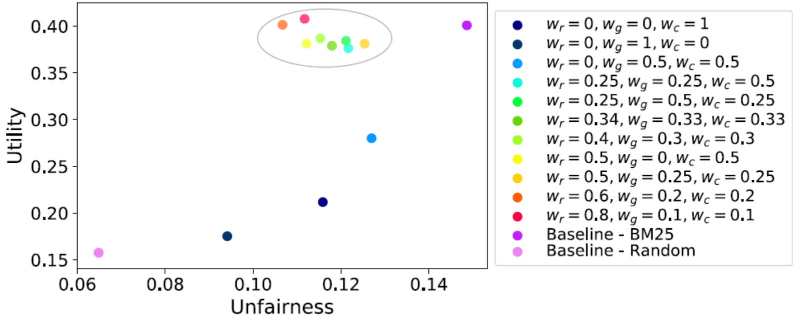
Algorithm 1: Fairness-aware Re-ranking Algorithm

```
Input: 𝒟: document corpus; q: query of interest; l: length of expected ranked
  list ; w: component weight vector
Output: R: re-ranked list of relevant documents
R ← Ø ; // initialize the ranked list as empty
𝒟′, 𝒟″ ← Retrieve relevant document candidates from 𝒟 for query q ;
  // document candidate corpus for q
for i = 1 → l do
    c_min ← A Large Integer; // initialize the minimal cost
    d_min ← None ; // initialize the document with the minimal cost
    for d ∈ 𝒟″ do
        Calculate the cost C(d, w, R, 𝒟′, q) according to Equation 1 ;
          // calculate the cost of adding d into R
        if C(d, w, R, 𝒟′, q) < c_min then
            d_min ← d ; // update the document with the minimal cost
            c_min ← C(d, w, R, 𝒟′, q) ; // update the minimal cost
        end
    end
    append d_min to R ; // add the document with the minimal cost into
      the re-ranked list R
    𝒟″ ← 𝒟″ − {d_min} ; // remove the added document d_min from 𝒟″
end
return R
```

Fairness-aware Re-ranking Results



(a) The group of country



(b) The group of gender

Fig. 1. Utility versus unfairness with different group definitions. The utility and unfairness scores were calculated based on Equation (7) and Equation (6) in the TREC 2019 Fairness Ranking Track [6] respectively.

- BM 25 – relatively high utility score but low fairness score.
- Random Ranking – low fairness on the country group, but the highest fairness on the gender group.
- Our methods' utility drops greatly when excluding BM25 scores ($w_r = 0$).
- When $w_r > 0$, the performance of our methods with different combinations of w_r, w_g, w_c are comparable on both country and gender groups.

Limitation and Future Work

- When inferring the gender, we treated it as a binary attribute guessed through the first name.
  - ✓ In the future, we will undertake fine-grained gender detection and utilize multimodal data (e.g., profile photos and gender pronouns) to infer the gender attribute more inclusively and robustly.

- 40.5% of countries failed to be detected, leading to potentially inaccurate group classifications.
  - ✓ In the future, we will explore more public personal location information, such as Twitter profile locations.

Conclusion

- **RQ1**: How to infer features to construct groups for fair ranking?
  - ✓ We explored non-self-contained features of gender and country by external resources, to construct latent groups for fairness purposes.
- **RQ2**: How to strike a balance between relevance and fairness?
  - ✓ We proposed a weighted fairness-aware re-ranking algorithm to strike a balance between relevance and fairness.

Reference

1. Lo Kyle, Lucy Lu Wang Mark Neumann, Kinney Rodney, and Weld Daniel S.. 2020. S2orc: The semantic scholar open research corpus. In Proceedings of ACL.
2. Biega, A.J., Diaz, F., Ekstrand, M.D., Kohlmeier, S.: The TREC 2020 Fairness Track (2020), https://fair-trec.github.io
3. ApS, D.: genderize.io (2020), https://genderize.io/
4. Cholewiak, S.A., Ipeirotis, P., Revision, V.S.: scholarly: Simple access to Google Scholar authors and citations (2020), https://pypi.org/project/scholarly/
5. Dept., I.M.F.R.: World economic outlook. World Economic Outlook, INTERNATIONAL MONETARY FUND (2019). https://doi.org/http://dx.doi.org/10.5089/9781513508214.081
6. Biega, A.J., Diaz, F., Ekstrand, M.D., Kohlmeier, S.: Overview of the trec 2019 fair ranking track. In: The Twenty-Eighth Text REtrieval Conference (TREC 2019) Proceedings (2019)

Thank you!
Q&A