# UDInfoLab at the NTCIR-17 FairWeb-1 Task

Fumian Chen[1] Hui Fang [1]

[1]Institute for Financial Services Analytics

University of Delaware, DE, USA

# Our Goal at FairWeb-1

We tested the distribution-based fair learning (DLF) framework [1] that:

- based on re-ranking and can merge with any relevance model.
- does not require gold fairness labels.
- leverages under-exploited contextual features.

The DLF achieved promising fairness performance on the NTCIR dataset, even though it does not cope with ordinal fairness groups

[1] Chen, F., & Fang, H. (2023, August). Learn to be Fair without Labels: A Distribution-based Learning Framework for Fair Ranking. In Proceedings of the 2023 ACM SIGIR International Conference on Theory of Information Retrieval (pp. 23-32).

# Distribution-based Learning Framework (DLF)

## Fairness-aware loss and learning without labels

$$\theta^* = \arg\min FL(\pi) = \sum_{i=1}^{m} w_i * KL(\epsilon_i(\pi), \epsilon_i^*)$$

$$= \arg\min \sum_{i=1}^{m} w_i * KL(\sum_{k=1}^{n} P_{\text{fair}}(s_k) * GM_{ik}, \sum_{k=1}^{n} P_{\text{fair}}(s_k^*) * GM_{ik})$$

$$= \arg\min \sum_{i=1}^{m} w_i * KL(\sum_{k=1}^{n} P_{\text{fair}}(f(X_k, \theta)) * GM_{ik}, \epsilon_i^*)$$

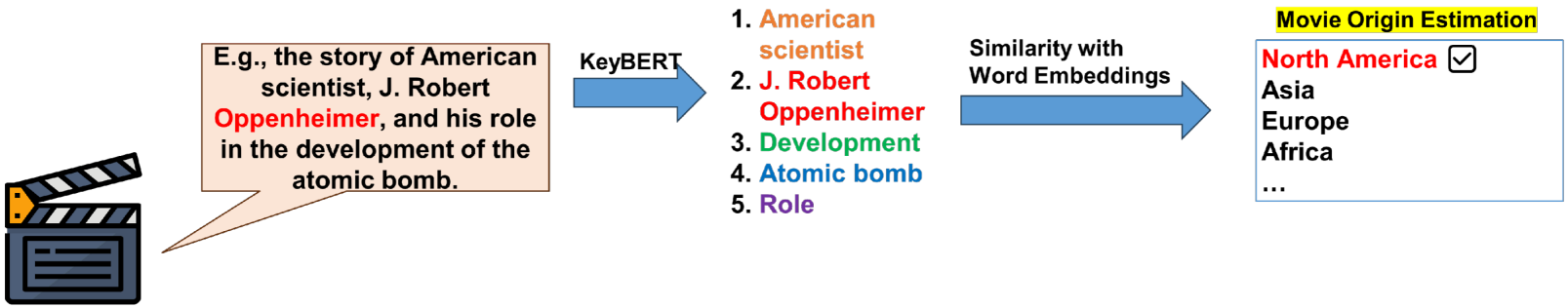
- $P_{\text{fair}}$ is top one probability, $s^*$ is the unavailable ground truth label, and $\epsilon^*$ is the target exposure distribution.
- $\textit{Final Score} = \alpha * f(X_k, \theta^*) + (1 - \alpha) *$ Relevance to make final rankings fair and relevant.

# Group Membership Estimation

## Group Membership Estimation

Since GM is not available, the group membership is obtained by calculating the cosine similarities between text keyword embeddings and fairness annotation (e.g., male and female) keyword embeddings.

- *KeyBert* to extract keywords from cleaned HTML text.
- Sentence-BERT with the pre-trained model 'all-mpnet-base-v2' for words embedding.



# Fairness Performance

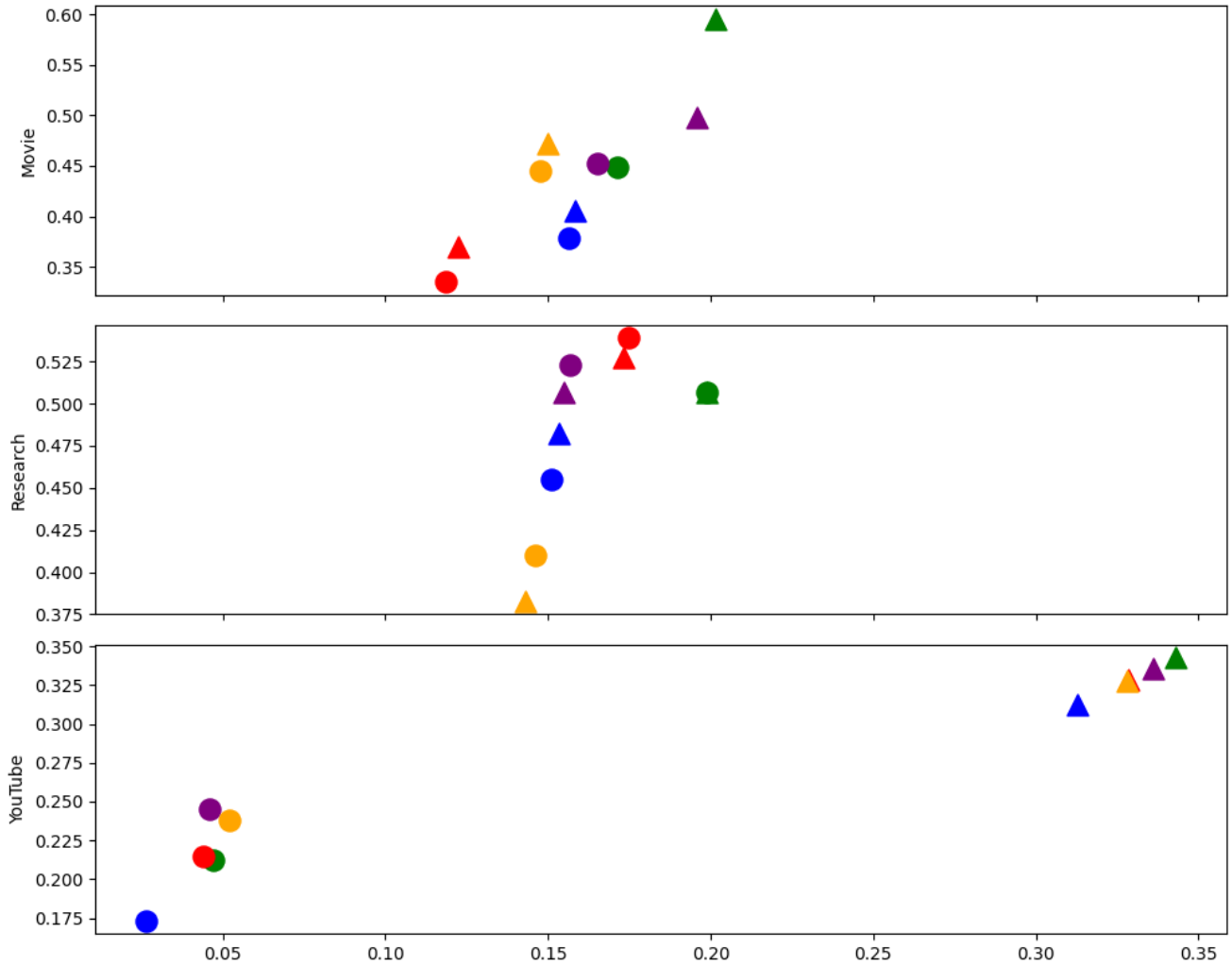
Table: Fairness improvements after re-ranking V.S. their initial rankings over the 15 topics for Movie, Researcher, and YouTube, respectively [2].

| Run Name | Movie | | | Researcher | | | Youtube | |
|---|---|---|---|---|---|---|---|---|
| | *Mean GF JSD (Origin)* | *Mean GF NMD (Ratings)* | *Mean GF RNOD (Ratings)* | *Mean GF JSD (Gender)* | *Mean GF NMD (H-index)* | *Mean GF RNOD (H-index)* | *Mean GF NMD (Subscription)* | *Mean GF RNOD (Subscription)* |
| UDinfo-D-RR-1 | 0.3672* | 0.4279* | 0.3913* | 0.4985 | 0.4682* | 0.4434* | 0.3157** | 0.3017** |
| UDinfo-Q-RR-2 | 0.4493* | 0.5132* | 0.4706* | 0.5096* | 0.4977* | 0.4605* | 0.3315** | 0.3081** |
| UDinfo-D-RR-3 | 0.3476* | 0.3876* | 0.3569* | 0.5374 | 0.5195 | 0.4866 | 0.3228** | 0.3091** |
| UDinfo-Q-RR-4 | 0.4601* | 0.5161* | 0.4750* | 0.5190 | 0.4994 | 0.4650 | 0.3309** | 0.3157** |
| UDinfo-D-RR-5 | 0.4543* | 0.4888* | 0.4488* | 0.3829 | 0.3765 | 0.3554 | 0.3279** | 0.3083** |
| bm25-depThre3-D | 0.3401 | 0.3993 | 0.3630 | 0.4694 | 0.4400 | 0.4155 | 0.1777 | 0.1731 |
| bm25-depThre3-Q | 0.4135 | 0.4623 | 0.4283 | 0.5096 | 0.4977 | 0.4605 | 0.2112 | 0.2039 |
| qld-depThre3-D | 0.3122 | 0.3507 | 0.3208 | 0.5497 | 0.5306 | 0.4975 | 0.2155 | 0.2100 |
| qld-depThre3-Q | 0.4275 | 0.4668 | 0.4351 | 0.5356 | 0.5152 | 0.4807 | 0.2451 | 0.2391 |
| qljm-depThre3-D | 0.4273 | 0.4606 | 0.4211 | 0.4120 | 0.4038 | 0.3824 | 0.2454 | 0.2329 |

[2] * indicates that re-ranking improves its baseline. ** indicates the improvements are also statistically significant (based on a randomized Tukey HSD test with $B = 5,000$ trials, $\alpha = 0.05$)

# Fairness and Relevance

Figure: GFR v.s. Relevance Scores (Mean ERR). Triangles indicate our five runs, whereas circles are their baselines. Different colors indicate different pairs.



# Conclusion and Future Work

In the FairWeb-1 task, we tested a recently proposed fair ranking framework DLF and found:

- The result shows that DLF can help initial rankings improve fairness while maintaining relevance in most cases.
- It performs poorly for longer text (e.g., research articles) that carries more information.

In the future, we plan to:

- incorporate DLF with ordinal fairness attributes.
- explore the relationship between fairness attributes and the impact on model performance.
- refine the way for keyword extraction and similarity calculation.

# Thank You!

email: fmchen@udel.edu