RMIT at TREC 2021 Fair Ranking Track

SACHIN PATHIYAN CHERUMANAL, RMIT University, Australia
DAMIANO SPINA, RMIT University, Australia
FALK SCHOLER, RMIT University, Australia
W. BRUCE CROFT, University of Massachusetts Amherst, USA

This report describes the data, the assumptions, methodology, and our results involved in the participation at the TREC 2021 Fair Ranking Track. While most of the fairness-aware re-ranking techniques require explicitly defining protected attributes, we tried to leverage the implicit features of the Wikimedia articles by using an implicit diversification technique to study the impact of diversification on a fair ranking problem.

1 INTRODUCTION

Presenting the users with relevant search results while ensuring fair exposure to different groups has been a challenge in the IR research community for a while. Fairer rankings are expected to provide relevant documents while being fair to the under-represented groups. For example, a significant Wikipedia article from a comparably less popular geography like Antarctica should also be given equal treatment.

The TREC 2021 Fairness Track adopts a resource allocation task which supports Wikipedia coordinators who search for documents that needs to be improved. During such a task, it is imperative that the documents about, or that somehow represent, certain protected characteristics receive a fair exposure. For instance, one of the attributes provided by the track organisers was the geographic location. However, the organisers evaluate the fairness of the systems across an undisclosed protected attribute as well.

Some of the previous works have already investigated the relationship between fairness and diversity metrics for evaluation purposes [2, 4]. So, at TREC 2021 Fair Ranking Track, our submissions were aimed to investigate the impact of diversification techniques on a fair ranking problem. We hypothesize that, when the protected attributes are unknown (as is the case in this track due to the undisclosed protected attribute), diversification techniques using the implicit features of the Wikimedia articles can play a major role in achieving fairness.

Section 2 of this notebook describes the different systems submitted to the Task 1 of the TREC 2021 Fair Ranking Track and Section 3 discusses the results of the fairness evaluation.

2 METHODOLOGY

In this work, we reused components from the Pyserini framework [3] for indexing and retrieval. The Wikimedia articles provided by the organisers were indexed using Pyserini's Sparse Lucene Index. No prior pre-processing was done on the corpus, as we wanted all the implicit features intact and see how diversification algorithms performed with minimal intervention. We use a popular ranking algorithm i.e., BM25, that has been widely accepted across the industry and academic researchers. It predicts the relevance of a document to a user query and ranks the results accordingly. For each query, We used BM25 [5] from the Pyserini framework with the parameters $k1 = 0.8$ and $b = 0.3$ to generate a ranked list of 1000 Wikimedia articles, namely, RMITRet. This baseline ranking would further be re-ranked using an implicit diversification technique: Maximal Marginal Relevance (MMR) [1].

Authors' addresses: Sachin Pathiyan Cherumanal, RMIT University, Melbourne, Australia, sachin.pathiyan.cherumanal@student.rmit.edu.au; Damiano Spina, RMIT University, Melbourne, Australia, damiano.spina@rmit.edu.au; Falk Scholer, RMIT University, Melbourne, Australia, falk.scholer@rmit.edu.au; W. Bruce Croft, University of Massachusetts Amherst, Amherst, MA, USA, croft@cs.umass.edu.

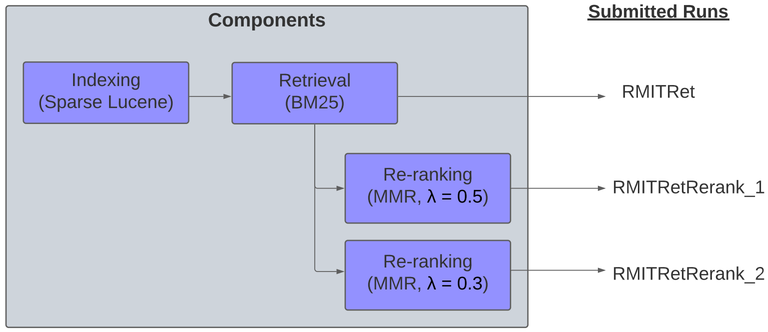

Fig. 1. This figure shows all the components that were used to generate the three runs - RMITRet, RMITRetRerank_1, and RMITRetRerank_2.

Table 1. nDCG, AWRF, and overall score (dot product of nDCG and AWRF) for the three submitted runs.

		nDCG			**AWRF**			**Score**		
Run	System	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
RMITRet	BM25	0.2075	0.1502	0.1347	0.6413	0.6216	0.1141	0.1317	0.1109	0.0863
RMITRetRerank_1	BM25 + MMR (λ = 0.5)	0.1760	0.1358	0.1065	0.6577	0.6561	0.0974	0.1144	0.0965	0.0694
RMITRetRerank_2	BM25 + MMR (λ = 0.3)	0.1768	0.1359	0.1102	0.6582	0.6415	0.1011	0.1146	0.0955	0.0714

MMR is a diversity-based re-ranking technique that tries to maximise the novelty, diversity, and relevance of a ranked list. A document will have high marginal relevance if it is both relevant to the query and contains minimal similarity to previous documents in the re-ranked list [1].

In this work, for each topic, we take the top 100 retrieved documents and diversify using MMR (refer Figure 1). However, we made two submissions of MMR using different λ values. As per Carbonell and Goldstein [1], MMR computes a maximal diverse ranking when $\lambda = 0$ and the standard relevance-ranked list when $\lambda = 1$. We wanted to see the impact of optimising diversity and relevance, on fairness. So we make two submissions of MMR, namely, RMITRetRerank_1 and RMITRetRerank_2, with $\lambda = 0.5$ and $\lambda = 0.3$ respectively. Overall, the systems submitted to the Task 1 of the TREC 2021 Fair Ranking Track were the following:

RMITRet. BM25 for retrieval, no re-ranking.

RMITRetRerank_1. Re-ranked results using MMR diversification with $\lambda = 0.5$.

RMITRetRerank_2. Re-ranked results using MMR diversification with $\lambda = 0.3$.

Note: The RMITRetRerank_1 and RMITRetRerank_2 submissions had a bug. However, the results discussed in the following section are based on the corrected runs that are made publicly available at https://github.com/rmit-ir/FairTrec-2021-Runs.

For the Task 1 of TREC 2021 Fair Ranking Track each system submission included a single ranked list consisting of 1000 Wikimedia articles for each query. The fairness of these ranked lists was further evaluated using attention-weighted rank fairness (AWRF) [6]. The final metric of the task was the dot product of AWRF and nDCG, which was used to gauge the overall performance of the submitted systems along both fairness and relevance.

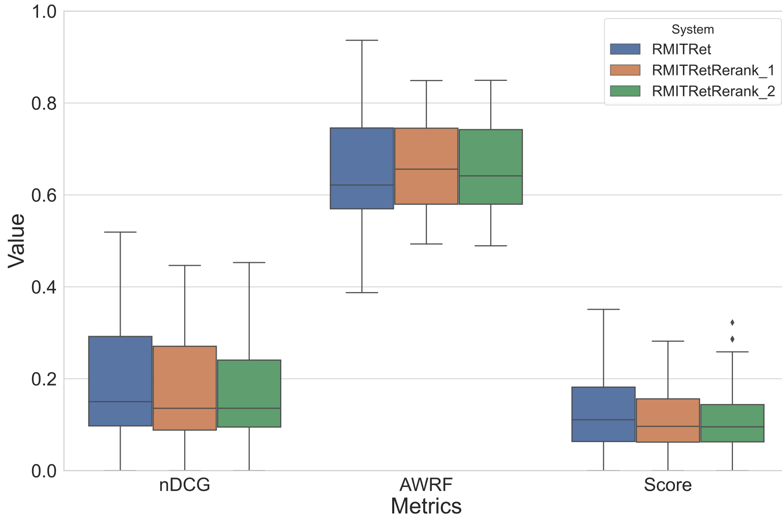

Fig. 2. The boxplots represent the distribution of different metrics across topics for systems RMITRet, RMITRetRerank_1, RMITRetRerank_2.

3 RESULTS AND DISCUSSION

From Table 1 and Figure 2, we infer that, RMITRetRerank_1 and RMITRetRerank_2 provided fairer results than the baseline RMITRet. We also notice that by increasing the diversity aspect in RMITRetRerank_2 (using $\lambda = 0.3$), we get a slightly fairer ranking when compared to RMITRetRerank_1 and RMITRet. Although this can seem to conclude that diversification can lead to fairer results in the case where protected attributes are unknown, independent t-tests on diversification-based submissions (RMITRetRerank_1 and RMITRetRerank_2) did not show any statistically significant improvement in fairness over the baseline (RMITRet).

4 FUTURE WORK

This work indexes the corpus as such, without any pre-processing. We believe the implicit diversification techniques used in this work show no statistically significant improvement over the baseline because the content of the Wikimedia articles prevails over the fields in the article which are potential candidates to be a protected characteristic. It would be worth exploring how identifying and prioritising different fields in the articles over the content can help implicit diversification techniques achieve fairer results. Although this borders using explicitly defined protected attributes, it reduces the challenge of picking one or two potential protected attributes to provide fairer results. Along the lines of improving fairness using explicitly protected attributes, we also think the category labels embedded in the Wikimedia articles could help in achieving fairer results. However, the different levels involved in the category of these articles can lead to doubts on the level of granularity that needs to be considered.

5 CONCLUSION

In this report, we provided a description about the different systems submitted to the TREC 2021 Fair Ranking Track and the hypothesis that drove our decision on using diversification techniques on the Wikimedia corpus to investigate it's impact in a fair ranking problem. We also highlight future work that could potentially achieve fairer results using the implicit features of the articles.

Acknowledgments. This work is partially supported by the Australian Research Council (DE200100064, DP190101113).

REFERENCES

[1] Jaime Carbonell and Jade Goldstein. 1998. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Melbourne, Australia) *(SIGIR '98)*. Association for Computing Machinery, New York, NY, USA, 335–336. https://doi.org/10.1145/290941.291025

[2] Ruoyuan Gao and Chirag Shah. 2020. Toward creating a fairer ranking in search engine results. *Information Processing & Management* 57, 1 (2020), 102138. https://doi.org/10.1016/j.ipm.2019.102138

[3] Jimmy Lin, Xueguang Ma, Sheng-Chieh Lin, Jheng-Hong Yang, Ronak Pradeep, and Rodrigo Nogueira. 2021. Pyserini: A Python Toolkit for Reproducible Information Retrieval Research with Sparse and Dense Representations. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, 2356–2362. https://doi.org/10.1145/3404835.3463238

[4] Sachin Pathiyan Cherumanal, Damiano Spina, Falk Scholer, and W. Bruce Croft. 2021. Evaluating Fairness in Argument Retrieval. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. Association for Computing Machinery, New York, NY, USA, 3363–3367. https://doi.org/10.1145/3459637.3482099

[5] Stephen E Robertson, Steve Walker, Susan Jones, Micheline M Hancock-Beaulieu, Mike Gatford, et al. 1995. Okapi at TREC-3. *Nist Special Publication Sp* 109 (1995), 109.

[6] Piotr Sapiezynski, Wesley Zeng, Ronald E Robertson, Alan Mislove, and Christo Wilson. 2019. Quantifying the Impact of User Attentionon Fair Group Representation in Ranked Lists. In *Companion Proceedings of The 2019 World Wide Web Conference* (San Francisco, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 553–562. https://doi.org/10.1145/3308560.3317595