Evaluating Fairness in Argument Retrieval

Sachin Pathiyen Cherumanal, Damiano Spina, Falk Scholer, and W. Bruce Croft
sachin.pathiyen.cherumanal@student.rmit.edu.au, {damiano.spina, falk.scholer}@rmit.edu.au, croft@cs.umass.edu

Argument Retrieval

Argument retrieval is the task of retrieving relevant supporting (PRO) and attacking (CON) documents for a given query.

Motivation

While argument retrieval systems provide multiple relevant answers for both stances i.e., PRO/CON, there may exist bias in exposure of these stances in the top results.

Research Question

“How to evaluate fairness in argument retrieval?”

Data and Metrics

- **Data**: Test collection (qrels) used at Touché@CLEF 2020, documents (arguments) from args.me corpus (debate portals) manually annotated with relevance and labeled with PRO/CON stances.
- **(Un)fairness metrics**: Normalized Discounted Difference (rND), Normalized Discounted K-L Divergence (rKL), Normalized Discounted Ratio (rRD)
- **Diversity metric**: $\alpha$-nDCG

Methodology

- 21 official runs from argument retrieval task at Touché@CLEF 2020.
- This work studies the relation between, relevance, diversity, and the adapted (un)fairness metrics.

- Controlled scenario using synthetic data to characterize the behaviour of diversity and (un)fairness metrics.

Results

- The most effective systems were not necessarily the most fair.
- Fairness and diversity metrics were related but not equivalent.

Future Work

- Consider fairness of topical categories as well as stance i.e., multi-attribute fairness.
- Explore other fairness and diversity metrics.