UNIVERSITY OF AMSTERDAM

INFORMATICA — UNIVERSITEIT VAN AMSTERDAM

Creating an experimental environment for participating in TREC Fair Ranking Track

René Emmaneel

November 12, 2020

Supervisor(s): Fatemeh Sarvi

Signed:

Abstract

The TREC Fair Ranking Track is an annual academic search task with the stated goal of developing a benchmark for evaluating retrieval systems in terms of fairness. To participate we created a framework. [1]

Contents

Introduction

Ranking systems allow items to be seen by the users, where the exposure each item gets is significantly determined by the position in the ranking. Traditionally a ranking system would order items based on their relevance so that the system is of maximum utility to the user. However, modern information access systems often influence both the consumer and the producer of the content that they serve. Examples are hiring platforms where it is important that the system has a high utility for both the employer and the job-seeker. But also more traditional environments such as music, book or video recommendations can be considered two-sided because the indirect matching of users to content creators.

The ranking system allows exposure to the producers, which exposes risks to companies such as Google that hold a lot of power, and has already faced legal challenges such as the European Union antitrust violation fine [2]. While it is unlikely that a universal definition of fairness can be created, there are various attempts to measure unfairness and devise various fair ranking algorithms to ensure that the received attention of a given subject is approximately equal to its deserved attention [1].

To evaluate different fair ranking algorithms, the TREC Fair Ranking track was created. The stated goal was to develop a benchmark for evaluating retrieval systems in terms of fairness, as well as releasing a dataset for benchmarking fair ranking algorithms. The TREC 2020 Track was an academic search task, where a set of academic article abstract and queries were submitted to an adademic search engine. The central goal is to provide fair exposure to different groups of authors, where the group definitions can be arbitrary. Various sets of data has been provided, as well as a detailed evaluation protocol. The data of the 2019 track includes the Semantic Scolar Open Corpus, which consists of data describing various papers. It also includes query data, and query sequences.

To participate in the TREC fair ranking track, an environment needs to be build for the entire process. This process includes handling of the input data, implementing the evaluation algorithm, making it modular for various uses, testing of the framework and implementing learning to rank machine learning algorithms. To design and implement the framework the following research question will be answered:

How can a modular framework be constructed to create an experimental environment for participating in the TREC fair ranking track.

TODO: Structure van project

Theoretical background

2.1 Ethical aspect

2.2 Measuring fairness

Fairness can be subjective and differ for various use cases. Before describing the fairness definition we have to describe the way exposure can influence the discoverability for a certain group.

2.2.1 Utility of a ranking

2.2.2

2.3 TREC Fair Ranking Track definition

2.4

2.5

Creating the framework

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetuer odio sem sed wisi.

Experiments

Conclusions

Bibliography

[1] Asia J. Biega, Krishna P. Gummadi, and Gerhard Weikum. Equity of attention: Amortizing individual fairness in rankings. *CoRR*, abs/1805.01788, 2018.

[2] Mark Scott. Google fined record $2.7 billion in e.u. antitrust ruling. *New York Times*, 2017.