Addressing Bias and Fairness in Search Systems

Ruoyuan Gao*
gaoruoyuan@gmail.com
Rutgers University, Piscataway, NJ, USA

Chirag Shah
chirags@uw.edu
University of Washington, Seattle, WA, USA

ABSTRACT

Search systems have unprecedented influence on how and what information people access. These gateways to information on the one hand create an easy and universal access to online information, and on the other hand create biases that have shown to cause knowledge disparity and ill-decisions for information seekers. Most of the algorithms for indexing, retrieval, and ranking are heavily driven by the underlying data that itself is biased. In addition, orderings of the search results create position bias and exposure bias due to their considerable focus on relevance and user satisfaction. These and other forms of biases that are implicitly and sometimes explicitly woven in search systems are becoming increasing threats to information seeking and sense-making processes. In this tutorial, we will introduce the issues of biases in data, in algorithms, and overall in search processes and show how we could think about and create systems that are fairer, with increasing diversity and transparency. Specifically, the tutorial will present several fundamental concepts such as relevance, novelty, diversity, bias, and fairness using socio-technical terminologies taken from various communities, and dive deeper into metrics and frameworks that allow us to understand, extract, and materialize them. The tutorial will cover some of the most recent works in this area and show how this interdisciplinary research has opened up new challenges and opportunities for communities such as SIGIR.

CCS CONCEPTS

• **Information systems** → **Information retrieval**; • **Social and professional topics** → **User characteristics**.

KEYWORDS

bias and fairness; diversity and novelty; ranking; evaluation; social effects

ACM Reference Format:
Ruoyuan Gao and Chirag Shah. 2021. Addressing Bias and Fairness in Search Systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3404835.3462807

*This work was done prior to joining Amazon.

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8037-9/21/07…$15.00
https://doi.org/10.1145/3404835.3462807

1 MOTIVATION

Biased search results can lead to unfair distribution of opportunities and resources. Our reliance on search systems for all kinds of information needs – small or critical – and how the system results are presented make it hard or even impossible for an average user to get an unbiased exposure to information. Biased results can affect the user's credibility judgment, selection making, belief and attitude shaping of information, and even decision making. People often digest the presented results without being aware of the existence of bias in such systems. And yet, we lack tools and techniques for addressing bias and bringing fairness to search systems. At the same time, those systems and organizations behind them lack enough motivation to solve this problem in a direct, explicit, and effective manner. This is a huge socio-technical issue and we need interdisciplinary research to address it. The proposed tutorial will cover the problem as well as several underlying concepts, theories, and methods to address the issue.

2 OBJECTIVES

The learning outcomes of this tutorial are to enable the participants:

(1) to know and understand different concepts in fairness in information retrieval;
(2) to be familiar with the evaluation metrics and frameworks for fairness-aware systems;
(3) to use the fairness framework to assist analysis and system design;
(4) to be cognizant of the challenges and opportunities in the fairness area.

3 TARGET AUDIENCE AND RELEVANCE

This tutorial is intended for graduate students and researchers to be more aware of the issue of bias in designing and implementing search systems. Participants will learn about the differences between bias and other concepts in the field, where the bias may occur, how to address such bias, basic techniques to measure bias and fairness, strategies to improve fairness, and frameworks for analyzing and evaluating the fairness and its effect on the systems and users. The topics of fairness and transparency have become quite relevant in the recent years with issues of misinformation and biases becoming prevalent. This tutorial will be closely connected to the fairness works at past SIGIR conferences. Previous works have brought the attention and awareness of the fairness issue in search systems, and proposed ideas of how to materialize or translate from other domains the fairness notion in the search setting. This tutorial will extend those ideas with data-driven fairness-aware analysis, methods for evaluating fairness-aware search systems, and the impact of encoded fairness from the user perspective.

4 TUTORIAL OUTLINE

This half-day tutorial will have six modules. Below, they are described in detail.

4.1 Introduction

First, we will introduce the general area of research and provide several examples to demonstrate how search systems are biased and what a fair result may look like. This will be contextualized within the larger area of inquiry pertaining to bias and fairness in machine learning.

4.2 Metrics and Frameworks

Next, we will quickly introduce various measures for assessing relevance, including MAP, NDCG, and MRR. Many in the audience may already be familiar with these. Then we will provide details on some of the measures used in quantifying diversity and novelty, such as α-nDCG, MAP-IA, and NRBP. We will focus more on metrics for quantifying bias and fairness. We will cover several metrics and frameworks proposed in literature, categorized in two types.

The first type is to measure the bias and fairness directly on the search results. Fairness is often measured indirectly via the measure of bias. Bias measures are either based on the equity and equality of the results. The key to fairness is to ensure the appropriate exposure of the protected attributes such as demographic properties (gender, location, etc.) and diversity popularity (item genres, supplier, etc.). Generally, there are two types of metrics of quantifying bias on search results: distance-based, and error-based. We will explain each type of metrics in detail and discuss their advantages and drawbacks in terms of application.

The Second type is to evaluate the fairness-aware algorithms. Given the definitions of fairness and bias on search results, we are interested in evaluating the effectiveness of the fairness-aware algorithms. This is different from measuring bias directly on the results because an algorithm performs differently for different queries. We need to evaluate the algorithm's ability on achieving fairness which considers all queries, and its trade-offs between bringing fairness and retaining system utilities which are dependent on the choices of utility metrics. We will describe the concept of error-based metrics and feasibility-based metrics.

4.3 Recent Works

In order to provide a sense of what the current state of research is, we will briefly cover some of the most recent and relevant works for both defining and addressing bias and fairness in search. Some of these are listed below.

Fairness definition. In social impact, fairness is often concerned in terms of *individual fairness* [3, 20, 21] and *group fairness* [6, 22, 30]. Individual fairness treats similar individuals similarly; as such, statistical parity treats the demographics of those being exposed in the search results the same as the demographics of the population as a whole. Singh and Joachims [30] proposed to view group fairness from a perspective of exposure. Mehrotra et al. [24] studied the problem of fairness from the perspective of satisfaction levels across different demographics. Mansoury [23], Mehrotra et al. [25], Patro et al. [27] addressed the problem of supplier fairness in recommendation platforms.

Quantifying fairness. Kulshrestha et al. [19] developed a framework that quantified the amount of bias that arise from the data source and from the ranking system. Beutel et al. [2], Kuhlman et al. [18], Yao and Huang [35] proposed pairwise comparisons regarding utility and prediction errors. Geyik et al. [16], Yang and Stoyanovich [34] proposed distance-based fairness measures for ranked outputs based on statistical parity. Gao [10] proposed a unified evaluation metric for fairness-aware ranking algorithms. Diaz et al. [7] proposed fairness metrics for stochastic rankings. There have been several studies that compare existing fairness metrics [14, 28, 29, 32].

Fairness-aware algorithms Mehrotra et al. [25] proposed randomized algorithms for jointly optimizing fairness and user utility in search systems. Singh and Joachims [30] proposed a conceptual and computational framework for fairness ranking which maximized the utility for the user while satisfying some fairness constraints. Geyik et al. [16] developed a fairness-aware framework that improved the fairness for individuals without affecting the business metrics. Celis et al. [4] framed the problem of controlling polarization in personalization as a multi-armed bandit problem and proposed constrained and unconstrained ε-greedy algorithms to optimize for relevance. Cerrato et al. [5], Morik et al. [26], Singh and Joachims [31], Yadav et al. [33] particularly focused on ensuring fairness in learning-to-rank.

4.4 Strategies for Fairness-Aware Searching

The tutorial will then engage the audience in thinking through different strategies for eliminating bias and improving the fairness in search, targeting different causes of bias. Bias may arise from the source data, algorithmic or system bias, and cognitive bias. Algorithms that learn from and mirror real world statistics may unavoidably carry social bias from the original data to the search system [1]. To eliminate bias and develop a fairness-aware search system, the first step is to minimize the bias in the data. This includes fairness-aware sampling methodologies in the data collection process. However, we often do not have access to the data collection process, but are given the dataset without knowing the existence of the potential bias. This tutorial will introduce ideas of how to address fairness in such cases, such as encoding fairness in the training and learning algorithm, post-processing the results and modifying the presentations of the results. We will introduce three types post-processing optimization strategies which are the most popular in current state-of-the-art works in the search field.

4.5 A Framework for Fairness Analysis

Given that we have various optimization policies to incorporate fairness constraints into the algorithmic framework and a search system, which one should we choose? Ideally, researchers try to develop general algorithms and frameworks that are intended to work for all kinds of data. But in reality, adopting an algorithm or framework independent of data may be both impractical and unreasonable. Therefore, whether the objective is clear or not, it is generally a good idea to first associate with the data and application before selecting a policy. This tutorial will introduce a pareto-optimal based theoretical framework [11] as well as a fairness-aware evaluation framework [10] to teach and inspire the participants how to connect their problem with data and evaluate their proposed fairness-aware algorithms.

We will first provide examples of data of different characteristics, and teach how to use the theoretical framework to estimate the boundaries of customized fairness and utility scores, visualize the results, and conduct a series analyses and answer research questions such as: what is the effect of introducing fairness into the system, what are the trade-offs between fairness and utility and what are the best results we can achieve on a given data, what fairness-aware optimization strategies should we choose when designing a fair system. We will then teach how to design a unified evaluation metric, and develop optimization algorithms accordingly for fairness-aware search systems. We will demonstrate the use of the framework on a recently collected dataset, and discuss the impact of fairness on the utility metrics and user's awareness.

4.6 Challenges and Opportunities

Although we have algorithms for eliminating bias from the system side, how do we make the users aware of the bias and reduce the bias from the user side. This tutorial will discuss some challenges in breaking the loop of bias in search systems. We will discuss how each component of the search ecosystem can enhance the bias of each other and why is it hard to break the loop. We will also introduce some opportunities in bringing fairness in the SIGIR community, including improving fairness at the query level, session level, task level, and the user level (personalized experience, customized interface design).

5 MATERIALS

The authors will make the tutorial slides as well as the bibliography publicly available. We will also release our code for some of the mentioned experiments and algorithms during the tutorial for the attendees to try then or continue their explorations afterward.

6 RELATION TO PREVIOUS TUTORIAL

The authors presented a tutorial on a similar topic, titled "Counteracting Bias and Increasing Fairness in Search and Recommender Systems" at ACM RecSys conference [12]. The proposed tutorial will have a significant overlap with that tutorial with the following differences.

- The focus of this tutorial will be on search systems, with addition to works containing learning to rank.
- This tutorial will also show the dangers of bringing diversity in search without thinking through and how that could lead to more misinformation.
- More recent works will be added, especially those from recent conferences and journals in IR.
- The tutorial will also be longer, allowing the authors to have more breadth and depth.

7 TUTORIAL PRESENTERS

Ruoyuan Gao has a PhD in Computer Science from Rutgers University, with a dissertation titled "Toward a fairer information retrieval system". She has worked as a teaching assistant for undergraduate and graduate courses in AI, ML and Data Science. She is the lead author of a paper on fairness in theoretical IR at ICTIR 2019, a paper on bias in web search in the Journal of IP&M 2020, and co-authored papers on fairness-aware recommendations at SIGIR 2020 and WSDM 2021.

Chirag Shah is an Associate Professor at University of Washington in Seattle. He has taught undergraduate and graduate courses in IR, HCI, and Data Science. He has also taught several courses and tutorials on topics related to IR at different international places, including at SIGIR and SIGIR conferences, Russian Summer School on Information Retrieval (RuSSIR), and Asian Summer School in Information Access (ASSIA). He has developed Coursera course on Social Media Data Analytics, and a SAGE course on Machine Learning for Data Science. He gave the keynote address to the International Workshop on Algorithmic Bias in Search and Recommendation, held virtually at ECIR 2020. His textbook on Data Science was recently published by Cambridge University Press.

The authors have published their work on the topic of this tutorial at ACM ICTIR [11], ACM SIGIR [9], ACM WSDM [15], ECIR workshop [8, 17, 32], and Journal of Information Processing and Management [13]. They also delivered a tutorial on this topic at ACM RecSys 2020 conference [12].

REFERENCES

[1] Solon Barocas and Andrew D Selbst. 2016. Big data's disparate impact. *California Law Review* 104 (2016), 671.

[2] Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Li Wei, Yi Wu, Lukasz Heldt, Zhe Zhao, Lichan Hong, Ed H. Chi, and Cristos Goodrow. 2019. Fairness in Recommendation Ranking Through Pairwise Comparisons. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. ACM, New York, NY, USA, 2212–2220.

[3] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of Attention: Amortizing Individual Fairness in Rankings. In *The 41st International ACM SIGIR Conference on Research Development in Information Retrieval*. ACM, New York, NY, USA, 405–414.

[4] L Elisa Celis, Sayash Kapoor, Farnood Salehi, and Nisheeth Vishnoi. 2019. Controlling Polarization in Personalization: An Algorithmic Framework. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 160–169.

[5] Mattia Cerrato, Marius Köppel, Alexander Segner, Roberto Esposito, and Stefan Kramer. 2020. Fair pairwise learning to rank. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 729–738.

[6] Alexandra Chouldechova. 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data* 5, 2 (2017), 153–163.

[7] Fernando Diaz, Bhaskar Mitra, Michael D Ekstrand, Asia J Biega, and Ben Carterette. 2020. Evaluating Stochastic Rankings with Expected Exposure. *arXiv preprint arXiv:2004.13157* (2020).

[8] Yunhe Feng, Daniel Saelid, Ke Li, Ruoyuan Gao, and Chirag Shah. 2021. Towards Fairness-Aware Ranking by Defining Latent Groups Using Inferred Features. In *International Workshop on Algorithmic Bias in Search and Recommendation*. Springer.

[9] Zuohui Fu, Yikun Xian, Ruoyuan Gao, Jieyu Zhao, Qiaoying Huang, Yingqiang Ge, Shuyuan Xu, Shijie Geng, Chirag Shah, Yongfeng Zhang, et al. 2020. Fairness-Aware Explainable Recommendation over Knowledge Graphs. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Xi'an, China.

[10] Ruoyuan Gao. 2021. *Toward a fairer information retrieval system*. Ph.D. Dissertation. Rutgers The State University of New Jersey, School of Graduate Studies.

[11] Ruoyuan Gao and Chirag Shah. 2019. How Fair Can We Go: Detecting the Boundaries of Fairness Optimization in Information Retrieval. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval (ICTIR '19)*. ACM, New York, NY, USA, 229–236.

[12] Ruoyuan Gao and Chirag Shah. 2020. Counteracting Bias and Increasing Fairness in Search and Recommender Systems. In *Fourteenth ACM Conference on Recommender Systems*. 745–747.

[13] Ruoyuan Gao and Chirag Shah. 2020. Toward creating a fairer ranking in search engine results. *Information Processing & Management* 57, 1 (2020), 102138.

[14] Pratyush Garg, John Villasenor, and Virginia Foggo. 2020. Fairness Metrics: A Comparative Analysis. *arXiv preprint arXiv:2001.07864* (2020).

[15] Yingqiang Ge, Shuchang Liu, Ruoyuan Gao, Yikun Xian, Yunqi Li, Xiangyu Zhao, Changhua Pei, Fei Sun, Junfeng Ge, Wenwu Ou, and Yongfeng Zhang.

2021. Towards Long-Term Fairness in Recommendation. In *Proceedings of WSDM (WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 445–453.
[16] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search. In *Proceedings KDD (KDD '19)*. ACM, New York, NY, USA, 2221–2231.
[17] Bin Han, Chirag Shah, and Daniel Saelid. 2021. Users' Perception of Search Engine Biases and Satisfaction. In *International Workshop on Algorithmic Bias in Search and Recommendation*. Springer.
[18] Caitlin Kuhlman, MaryAnn VanValkenburg, and Elke Rundensteiner. 2019. FARE: Diagnostics for Fair Ranking Using Pairwise Error Metrics. In *The World Wide Web Conference (WWW '19)*. ACM, New York, NY, USA, 2936–2942.
[19] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P Gummadi, and Karrie Karahalios. 2017. Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media. In *Proceedings of CSCW*. ACM, New York, NY, USA, 417–432.
[20] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual Fairness. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 4066–4076.
[21] Preethi Lahoti, Gerhard Weikum, and Krishna P Gummadi. 2018. iFair: Learning Individually Fair Data Representations for Algorithmic Decision Making. *arXiv preprint arXiv:1806.01059* (2018).
[22] Yunqi Li, Hanxiong Chen, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2021. User-oriented Fairness in Recommendation. *WWW* (2021).
[23] Masoud Mansoury. 2021. Fairness-Aware Recommendation in Multi-Sided Platforms. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 1117–1118.
[24] Rishabh Mehrotra, Ashton Anderson, Fernando Diaz, Amit Sharma, Hanna Wallach, and Emine Yilmaz. 2017. Auditing Search Engines for Differential Satisfaction Across Demographics. In *Proceedings of the 26th International Conference on World Wide Web Companion*. 626–633.
[25] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a Fair Marketplace: Counterfactual Evaluation of the Trade-off Between Relevance, Fairness & Satisfaction in Recommendation Systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, New York, NY, USA, 2243–2251.
[26] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. 2020. Controlling Fairness and Bias in Dynamic Learning-to-Rank. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*. Association for Computing Machinery, New York, NY, USA, 429–438.
[27] Gourab K Patro, Arpita Biswas, Niloy Ganguly, Krishna P. Gummadi, and Abhijnan Chakraborty. 2020. FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms. In *Proceedings of The Web Conference 2020 (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 1194–1204.
[28] Amifa Raj, Connor Wood, Ananda Montoly, and Michael D Ekstrand. 2020. Comparing Fair Ranking Metrics. *arXiv preprint arXiv:2009.01311* (2020).
[29] 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 (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 553–562.
[30] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of Exposure in Rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, 2219–2228.
[31] Ashudeep Singh and Thorsten Joachims. 2019. Policy learning for fairness in ranking. *arXiv preprint arXiv:1902.04056* (2019).
[32] Sahil Verma, Ruoyuan Gao, and Chirag Shah. 2020. Facets of Fairness in Search and Recommendation. In *International Workshop on Algorithmic Bias in Search and Recommendation*. Springer, 1–11.
[33] Himank Yadav, Zhengxiao Du, and Thorsten Joachims. 2019. Fair learning-to-rank from implicit feedback. *arXiv preprint arXiv:1911.08054* (2019).
[34] Ke Yang and Julia Stoyanovich. 2017. Measuring Fairness in Ranked Outputs. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*. ACM, New York, NY, USA, 22:1—22:6.
[35] Sirui Yao and Bert Huang. 2017. New fairness metrics for recommendation that embrace differences. *arXiv preprint arXiv:1706.09838* (2017).