

Fairness and Discrimination in Recommendation and Retrieval

Michael D. Ekstrand
People & Information Research Team
Boise State University
Boise, Idaho
michaelekstrand@boisestate.edu

Robin Burke
Dept. of Information Science
University of Colorado
Boulder, Colorado
robin.burke@colorado.edu

Fernando Diaz
Microsoft Research
Montréal, Quebec
diazf@acm.org

ABSTRACT

Fairness and related concerns have become of increasing importance in a variety of AI and machine learning contexts. They are also highly relevant to recommender systems and related problems such as information retrieval, as evidenced by the growing literature in RecSys, FAT*, SIGIR, and special sessions such as the FATREC and FACTS-IR workshops and the Fairness track at TREC 2019; however, translating algorithmic fairness constructs from classification, scoring, and even many ranking settings into recommendation and other information access scenarios is not a straightforward task. This tutorial will help orient RecSys researchers to algorithmic fairness, understand how concepts do and do not translate from other settings, and provide an introduction to the growing literature on this topic.

CCS CONCEPTS

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

KEYWORDS

fairness, discrimination, bias, social effects

ACM Reference Format:

Michael D. Ekstrand, Robin Burke, and Fernando Diaz. 2019. Fairness and Discrimination in Recommendation and Retrieval. In *Thirteenth ACM Conference on Recommender Systems (RecSys '19)*, September 16–20, 2019, Copenhagen, Denmark. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3298689.3346964>

1 MOTIVATION

Recommender systems, search engines, and other algorithmic information access systems mediate much of the information experiences of members of society. They have, however, a number of potential problems in the view they present of the world or information space. Many of these issues result from a failure to consider the social context of the design, testing, and deployment of information access systems. As a result, undiagnosed problems in these systems can produce unintended societal consequences, as Noble [14] highlights.

As information access systems continue to be employed in an increasing variety of domains, it becomes crucial both for researchers

and engineers to understand how these systems interact with society in general, including the various biases — some benign, some connected to historical patterns of discrimination — in their underlying data and in the responses of their users [6]. Indeed, Belkin and Robertson [1] stress the need for considering social implications of information retrieval research when they write, “the development of theory must depend not only on the internal constraints of the science but also upon its external constraints.”

The issues of fairness, accountability, transparency, bias, discrimination, justice, and ethics that are seeing increased attention in many areas of computing also have significant relevance to the information retrieval community [3, 8, 9, 12]. There is a substantial and rapidly-growing research literature studying fairness, bias, and discrimination in general machine learning contexts [5]. While some of this work, particularly work on fair ranking [3, 15], translates easily into recommender and information retrieval systems, other issues such as the multisided nature of information discovery platforms [4] and the extreme sparsity of relevance judgments make it more difficult to apply fairness results from other fields to retrieval and recommendation settings.

The purpose of this tutorial is to provide recommender systems researchers and practitioners interested in issues of fairness, bias, and discrimination with a starting point for carrying out that work. To that end, we cover core concepts in algorithmic fairness with pointers to relevant literature, survey the problem space and existing research on fairness in recommendation and information retrieval, and explain in greater detail the methods and metrics currently developed for evaluating and providing fair rankings and recommendations along with the limitations of these methods that should drive further research. We devote particular attention on the study of fairness in applied settings [2, 10, 12, 13].

2 OBJECTIVES

It is our goal that participants in this tutorial will be able to do the following:

- Understand key concepts of algorithmic fairness, including group vs. individual fairness, disparate treatment vs. disparate impact, allocational vs. representational harms, and key results on the measurement and relationships of these constructs.
- Identify possible sources of unfairness in data, algorithms, and applications in recommender systems.
- Identify the stakeholders who may have fairness concerns in a given retrieval or recommendation application, and articulate how the system may have adverse impacts on them.
- Assess the applicability of existing metrics and experimental protocols to assessing fairness in particular problem settings.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

RecSys '19, September 16–20, 2019, Copenhagen, Denmark

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6243-6/19/09.

<https://doi.org/10.1145/3298689.3346964>

- Engage with existing research on fairness, apply it to recommendation problems, and identify new research questions on the fairness of information access systems.

3 RELEVANCE

To our knowledge, this (along with its companion tutorial at SIGIR) is the first tutorial specifically on the state of research and the challenges in applying ideas of fairness to recommendation and information retrieval. This tutorial will complement the FATREC workshops at RecSys 2017 and 2018 [7, 11], the FACTS-IR workshop at SIGIR 2019, and related RecSys 2019 workshops such as ImpactRS and RMSE.

Participants who have previously attended the *Limits of Social Data* tutorial given by Alexandra Olteanu, Emre Kıcıman, Carlos Castillo, and Fernando Diaz at WWW'18, KDD'17, and several other conferences will find this to be complementary, building on ideas there and digging deeper into their particular application to information retrieval and recommender systems.

We will not be assuming any prior familiarity with algorithmic fairness or its legal and social foundations, and will only be assuming exposure to the fundamentals of recommender systems, not familiarity with specific lines of current research. Thus, the tutorial will be accessible to early-stage researchers, but will also contain useful information for intermediate and experienced IR researchers looking to expand their research and teaching activities to include fairness. We will also connect our presentation to production concerns, leaning on the work of Holstein et al. [10], to make this tutorial useful for industrial practitioners as well.

4 FORMAT AND SCHEDULE

This is a half-day tutorial in lecture format, with topics organized as follows:

4.1 Session 1: Foundations and Problems

- Welcome and Intro
- Some Motivating Examples
- Introduction to Fairness Problems and Concepts
- Survey of Algorithmic Fairness Concepts, Metrics, and Results
- What's Different about Information Access?

4.2 Session 2: Metrics and (Partial) Solutions

- Fair for Who? Multisided Nature of Information Access Fairness
- Fair How? Personalization, Relevance, and Other Problems with Fairness
- Consumer Fairness - Who Is It Good For?
- Provider Fairness - Who Gets Exposure?
- Feedback Loops
- Fairness in Production
- Some Open Problems
- Questions

5 SUPPORT MATERIALS

Support materials for this tutorial, including slides and a bibliography, are available at <https://fair-ia.ekstrand.net>.

ACKNOWLEDGMENTS

This tutorial is partially based on work supported by NSF grant IIS 17-51278.

REFERENCES

- [1] N. J. Belkin and S. E. Robertson. 1976. Some ethical and political implications of theoretical research in information science. In *Proceedings of the ASIS Annual Meeting*.
- [2] Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Allison Woodruff, Christine Luu, Pierre Kreitmann, Jonathan Bischof, and Ed H. Chi. 2019. Putting Fairness Principles into Practice: Challenges, Metrics, and Improvements. *CoRR* abs/1901.04562 (2019).
- [3] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of Attention: Amortizing Individual Fairness in Rankings. In *Proc. SIGIR '18*. ACM, 405–414. <https://doi.org/10.1145/3209978.3210063>
- [4] Robin Burke. 2017. Multisided Fairness for Recommendation. (July 2017). arXiv:cs.CY/1707.00093 <http://arxiv.org/abs/1707.00093>
- [5] Alexandra Chouldechova and Aaron Roth. 2018. The Frontiers of Fairness in Machine Learning. (Oct. 2018). arXiv:cs.LG/1810.08810 <http://arxiv.org/abs/1810.08810>
- [6] Fernando Diaz. 2016. Worst Practices for Designing Production Information Access Systems. *SIGIR Forum* 50, 1 (June 2016), 2–11.
- [7] Michael D Ekstrand and Amit Sharma. 2017. FATREC Workshop on Responsible Recommendation. In *Proc. ACM RecSys '18*. ACM, 382–383. <https://doi.org/10.1145/3109859.3109960>
- [8] Michael D Ekstrand, Mucun Tian, Ion Madrazo Azpiazu, Jennifer D Ekstrand, Oghenemaro Anuyah, David McNeill, Pera, and Maria Soledad. 2018. All The Cool Kids, How Do They Fit In?: Popularity and Demographic Biases in Recommender Evaluation and Effectiveness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (PMLR)*, Vol. 81. 172a–186. <http://proceedings.mlr.press/v81/ekstrand18b.html>
- [9] Michael D Ekstrand, Mucun Tian, Mohammed R Imran Kazi, Hoda Mehrpouyan, and Daniel Kluver. 2018. Exploring Author Gender in Book Rating and Recommendation. In *Proc. ACM RecSys '18*. ACM. <https://doi.org/10.1145/3240323.3240373>
- [10] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudik, and Hanna Wallach. 2019. Improving fairness in machine learning systems: What do industry practitioners need?. In *Proc. CHI 2019*.
- [11] Toshihiro Kamishima, Pierre-Nicolas Schwab, and Michael D Ekstrand. 2018. 2nd FATREC workshop: responsible recommendation. In *Proc. ACM RecSys '18*. ACM, 516–516. <https://doi.org/10.1145/3240323.3240335>
- [12] Rishabh Mehrotra, Ashton Anderson, Fernando Diaz, Amit Sharma, Hanna Wallach, and Emine Yilmaz. 2017. Auditing Search Engines for Differential Satisfaction Across Demographics. In *WWW '17 Companion*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 626–633. <https://doi.org/10.1145/3041021.3054197>
- [13] 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 and Satisfaction in Recommendation Systems. In *Proc. CIKM '18*.
- [14] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.
- [15] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of Exposure in Rankings. In *Proc. KDD '18 (KDD '18)*. ACM, New York, NY, USA, 2219–2228. <https://doi.org/10.1145/3219819.3220088>