2019

Fairness and Discrimination in Recommendation and Retrieval

Michael D. Ekstrand  
*Boise State University*, michaelekstrand@boisestate.edu

Robin Burke  
*University of Colorado*

Fernando Diaz  
*Microsoft Research*

Follow this and additional works at: https://scholarworks.boisestate.edu/cs_facpubs

Part of the Computer Sciences Commons

Publication Information

This document was originally published in *RecSys '19: Proceedings of the 13th ACM Conference on Recommender Systems* by Association for Computing Machinery. Copyright restrictions may apply. doi: 10.1145/3298689.3346964
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:

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 seen 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.
• 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 Kucman, 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.ekstrandom.net.

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

REFERENCES