Algorithmic Bias in Rankings

Carlos Castillo
chato@acm.org
Universitat Pompeu Fabra
Barcelona, Catalonia, Spain

ABSTRACT

In this talk we refer to bias in its everyday sense, as a prejudice against a person or a group, and ask whether an algorithm, particularly a ranking algorithm, can be biased. We begin by defining under which conditions this can happen. Next, we describe key results from research on algorithmic fairness, much of which studies automatic classification by a supervised learning method. Finally, we attempt to map these concepts to rankings and to introduce new, ranking-specific ways of looking at algorithmic bias.

CCS CONCEPTS

• **Information systems → Information retrieval**; *Data mining*.

KEYWORDS

bias, discrimination, ranking

ACM Reference Format:

Carlos Castillo. 2019. Algorithmic Bias in Rankings. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3308560.3316451

1 OUTLINE

Concerns about biased (in the sense of prejudiced) algorithms have been motivated by research that has exposed "racist" [1] and "sexist" [3, 8] algorithms. In this talk, we will start by defining specifically under which circumstances an algorithm can be considered to engage in discrimination, borrowing a definition by Lippert-Rasmussen [10].

Next, we will discuss foundational research on algorithmic bias that took place within the Data Mining (DM) community [9] and later among Machine Learning (ML) researchers. While ensuring fairness, accountability, and transparency have been considered key strategic elements in Information Retrieval [7], algorithmic bias concerns have been much less studied than in DM and ML. Studying biased ranking algorithms requires to shift the perspective from the people issuing a search, to the organizations and people that are represented by the items being searched. Specifically, we will seek for a *sufficient presence*, a *consistent treatment*, and a *proper representation* of items, particularly those belonging to protected or disadvantaged groups [5].

In the main part of the talk, we will overview some key results on fair rankings introducing methods to measure fairness in a ranking [13, 14] concepts such as fairness of exposure [12], and algorithms to ensure rankings are fair [2, 6], among others. We will also describe new methods for transparency and explainability in ranking.

Finally, we will describe ways of moving forward, avoiding the paralysis of having multiple competing fair ranking definitions [11], and show examples of detection and mitigation of biased rankings in practice.

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3316451

ACKNOWLEDGMENTS

The author thanks La Caixa project LCF/PR/PR16/11110009 for partial support.

This talk is partially based on a keynote at the CIKM 2018 Workshop on Data and Bias [5] and a talk at the HUMAINT 2018 Workshop [4], both by the same author.

REFERENCES

[1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias. *ProPublica* 23 (May 2016).
[2] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of Attention: Amortizing Individual Fairness in Rankings. *arXiv:1805.01788 (pre-print)* (2018).
[3] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in neural information processing systems.* 4349–4357.
[4] Carlos Castillo. 2018. Algorithmic Discrimination at 1st HUMAINT Workshop. In *Assessing the impact of machine intelligence on human behaviour.* Joint Research Centre, 23–26.
[5] Carlos Castillo. 2018. Fairness and Transparency in Ranking. *ACM SIGIR Forum* 52, 2 (Dec 2018), 64–71. http://chato.cl/slides/DAB_2019_fairness_transparency.pdf.
[6] L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. 2018. Ranking with fairness constraints. In *Proc. of ICALP.* 28:1–28:15.
[7] J. Shane Culpepper, Fernando Diaz, and Mark D. Smucker (editors). 2018. Report from the Third Strategic Workshop on Information Retrieval(SWIRL). (2018).
[8] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated experiments on ad privacy settings. *Proceedings on privacy enhancing technologies* 2015, 1 (2015), 92–112.
[9] Sara Hajian, Francesco Bonchi, and Carlos Castillo. 2016. Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proc. of SIGKDD (Tutorials).* ACM, 2125–2126. http://francescobonchi.com/algorithmic_bias_tutorial.html.
[10] Kasper Lippert-Rasmussen. 2014. *Born free and equal?: a philosophical inquiry into the nature of discrimination.* Oxford University Press.
[11] Arvind Narayanan. 2018. Tutorial: 21 fairness definitions and their politics. https://www.youtube.com/watch?v=jIXIuYdnyyk.
[12] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of Exposure in Rankings. In *Proc. of KDD.* 2219–2228.
[13] Ke Yang and Julia Stoyanovich. 2017. Measuring Fairness in Ranked Outputs. In *Proc. of SSDB.* ACM, 22.
[14] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Mohamed Hajian, Saraand Megahed, and Ricardo Baeza-Yates. 2017. FA*IR: A fair top-k ranking algorithm. (2017), 1569–1578.