Fairness in Algorithmic Decision Making

Abhijnan Chakraborty
Max Planck Institute for Software Systems

Krishna P. Gummadi
Max Planck Institute for Software Systems

ABSTRACT

Algorithmic (data-driven) decision making is increasingly being used to assist or replace human decision making in domains with high societal impact, such as banking (estimating creditworthiness), recruiting (ranking applicants), judiciary (offender profiling) and journalism (recommending news-stories). Consequently, in recent times, multiple research works have attempted to identify (measure) bias or unfairness in algorithmic decisions and propose mechanisms to control (mitigate) such biases. In this tutorial, we introduce the related literature to the cods-comad community. Moreover, going over the more prevalent works on fairness in classification or regression tasks, we explore fairness issues in decision making scenarios, where we need to account for preferences of multiple stakeholders. Specifically, we cover our own past and ongoing works on fairness in recommendation and matching systems. We discuss the notions of fairness in these contexts and propose techniques to achieve them. Additionally, we briefly touch upon the possibility of utilizing user interface of platforms (choice architecture) to achieve fair outcomes in certain scenarios. We conclude the tutorial with a list of open questions and directions for future work.

ACM Reference Format:
Abhijnan Chakraborty and Krishna P. Gummadi. 2020. Fairness in Algorithmic Decision Making. In *7th ACM IKDD CoDS and 25th COMAD (CoDS COMAD 2020), January 5–7, 2020, Hyderabad, India.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3371158.3371234

Type of the Tutorial: Introductory

1 MOTIVATION AND GOALS

Algorithmic decision making systems are increasingly being used to assist or replace human decision making in multiple socially important domains, such as credit lending [14], employment [12], judiciary [1], journalism [3] and healthcare [11]. There are concerns that automated decisions made in these domains can have long-lasting impact and may adversely affect certain individuals or groups [1, 11, 14]. Consequently, a new active area of research has emerged in recent times to quantify and incorporate fairness for machine learning algorithms [9, 10, 16, 17]. As concerns over algorithmic unfairness and discrimination continue to grow, it is both timely and critical for data scientists and practitioners to be accustomed with the models and mechanisms to quantify and tackle algorithmic unfairness in their respective application domains. This tutorial aims to give a new perspective on algorithm design keeping in mind the ethical implications of the resultant decision making context. The main objectives of this tutorial are:

- to overview the growing literature on fairness for machine learning, with the emphasis on classification algorithms;
- to discuss the potential for unfairness in algorithms beyond classification, e.g., ranking, recommendation and matching;
- to cast fairness objectives of such algorithms through the lens of social choice, to guide the development of fair(er) mechanisms; and finally,
- to explore the potential of choice architectures to achieve fairness in decision outcomes.

2 DESCRIPTION

We begin with an overview of the growing line of research on algorithmic fairness. We introduce *group-level notions of fairness* which require that given a decision making scenario, a certain fairness metric is equal across all protected groups. Such fairness requirement may vary from equality of opportunity [10], impact [17] or mistreatment [16]. We also discuss *individual-level notions of fairness* which requires that two individuals who are similar with respect to the task at hand should receive similar decision outcomes [9]. Most of the research works covering both these notions have focused on classification algorithms. We summarize some of the major themes and findings in fair classification.

Going beyond classification: After covering the major works in fair classification, we highlight the potential for unfairness in other decision making scenario such as ranking [2, 4, 18], recommendation [5, 7, 13], summarization [8] or matching [15]. We then discuss approaches to mitigate unfairness in such scenarios.

Fairness in ranking: Rankings of people, hotels, or songs are at the heart of selection, matchmaking and recommender systems on a variety of platforms, starting from entertainment and dating all the way to employment and income. To be successful on these platforms, ranked subjects need to gain the attention of searchers. However, searchers are susceptible to position bias, which makes them pay most of their attention to the top-ranked subjects. As a result, lower-ranked subjects often receive disproportionately less attention than they deserve according to the ranking relevance. In this tutorial, we discuss measures and mechanism which ensure that, for all subjects in a platform, the received attention approximately equals the deserved attention, while preserving ranking quality [2].

Fairness in recommendation: To help their users to discover important items at a particular time,major websites like Twitter, Yelp, TripAdvisor or NYTimes provide Top-K recommendations (e.g., 10 Trending Topics, Top 5 Hotels in Paris or 10 Most Viewed News Stories), which rely on crowd-sourced popularity signals to select the items. However, different sections of a crowd may have different preferences. To fairly aggregate the preferences of all users while recommending top-K items, we discuss ideas from

CoDS COMAD 2020, January 5–7, 2020, Hyderabad, India

ACM ISBN 978-1-4503-7738-6/20/01.
https://doi.org/10.1145/3371158.3371234

social choice theory, and explore the fairness properties in top-K item (s)elections and approaches to achieve them [7].

Fairness in matching: Ride hailing platforms, such as Uber, Lyft, Ola or DiDi, have traditionally focused on the satisfaction of the passengers, or on boosting successful business transactions. However, recent studies provide a multitude of reasons to worry about the drivers, where the concerns range from bad working conditions to discrimination against minorities. With more and more drivers financially depending on online platforms, it is pertinent to ask what a fair distribution of income on such platforms is and what power and means the platform has in shaping these distributions. In this tutorial, we present a framework to think about fairness in the matching mechanisms of ride hailing platforms. Specifically, our notion of fairness relies on the idea that, spread over time, all drivers should receive benefits proportional to the amount of time they are active in the platform. We also explore the means of achieving two-sided fairness, their caveats and side-effects [15].

Fairness through design of choice architecture: Often biased outcomes in online platforms seem to be the sole artifact of biases in user preferences. Drawing on a rich body of work in behavioral economics, we argue that the platform can indeed play an important role in shaping users' behavior without limiting their choices. We further discuss how can utilize such choice architectures to get equitable and fair outcomes [6].

3 TARGET AUDIENCE

The tutorial is targeted towards data science practitioners and researchers working in the broad areas of machine learning, data mining, information retrieval and social computing. We plan to make the material understandable to people with no specific prerequisite. However, having basic knowledge about classical machine learning techniques would be a plus.

4 PRESENTERS

Dr. Abhijnan Chakraborty is a Post-doctoral Researcher at the Max Planck Institute for Software Systems (MPI-SWS), Germany. He obtained his Ph.D. from the Indian Institute of Technology Kharagpur under the supervision of Prof. Niloy Ganguly (IIT Kharagpur) and Prof. Krishna P. Gummadi (MPI-SWS). During Ph.D., he was awarded the Google India PhD Fellowship and the Prime Minister's Fellowship for Doctoral Research. Prior to joining Ph.D., he spent two years at Microsoft Research India. His research interests span social computing and fairness in machine learning. He has authored several papers in top-tier computer science conferences including WWW, KDD, CSCW, ICWSM, MobiCom. His research works have won the best paper award at ASONAM'16 and best poster award at ECIR'19. He is one of the recipients of the highly competitive research grant from Data Transparency Lab to advance his research on fairness and transparency in algorithmic systems.

Prof. Krishna P. Gummadi is a Scientific Director at the Max Planck Institute for Software Systems (MPI-SWS). He received his Ph.D. (2005) and B.Tech. (2000) degrees in Computer Science and Engineering from the University of Washington and the Indian Institute of Technology, Madras, respectively. Krishna's research interests are in the measurement, analysis, design, and evaluation of complex Internet-scale systems. His current projects focus on understanding and building social computing systems. Specifically, they tackle the challenges associated with (i) assessing the credibility of information shared by anonymous online crowds, (ii) understanding and controlling privacy risks for users sharing data on online forums, (iii) understanding, predicting and influencing human behaviors on social media sites (e.g., viral information diffusion), and (iv) enhancing fairness and transparency of machine (data-driven) decision making in social computing systems. Krishna's work on online social networks, Internet access networks, and peer-to-peer systems has been widely cited and his papers have received numerous awards, including SIGCOMM Test of Time, IW3C2 WWW Best Paper Honorable Mention, and Best Papers at NIPS ML & Law Symposium, ACM COSN, ACM/Usenix SOUPS, AAAI ICWSM, Usenix OSDI, ACM SIGCOMM IMC, ACM SIGCOMM CCR, and SPIE MMCN. He has also co-chaired AAAI's ICWSM 2016, IW3C2 WWW 2015, ACM COSN 2014, and ACM IMC 2013 conferences. He received an ERC Advanced Grant in 2017 to investigate "Foundations for Fair Social Computing".

REFERENCES

[1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias: there's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica* (2016).
[2] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. In *ACM SIGIR*.
[3] Abhijnan Chakraborty, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2015. Can trending news stories create coverage bias? on the impact of high content churn in online news media. In *Computation and Journalism Symposium*.
[4] Abhijnan Chakraborty, Aniko Hannak, Asia J Biega, and Krishna P Gummadi. 2017. Fair sharing for sharing economy platforms. In *FATREC*.
[5] Abhijnan Chakraborty, Johnnatan Messias, Fabricio Benevenuto, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2017. Who makes trends? understanding demographic biases in crowdsourced recommendations. In *AAAI ICWSM*.
[6] Abhijnan Chakraborty, Nuno Mota, Asia J Biega, Krishna P Gummadi, and Hoda Heidari. 2019. On the Impact of Choice Architectures on Inequality in Online Donation Platforms. In *WWW*.
[7] Abhijnan Chakraborty, Gourab K Patro, Niloy Ganguly, Krishna P Gummadi, and Patrick Loiseau. 2019. Equality of voice: Towards fair representation in crowdsourced top-k recommendations. In *ACM FAT**.
[8] Abhisek Dash, Anurag Shandilya, Arindam Biswas, Kripabandhu Ghosh, Saptarshi Ghosh, and Abhijnan Chakraborty. 2019. Summarizing User-generated Textual Content: Motivation and Methods for Fairness in Algorithmic Summaries. *Proceedings of ACM Human Computer Interaction* 3, CSCW (2019).
[9] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *ACM ITCS*.
[10] Moritz Hardt, Eric Price, Nati Srebro, et al. 2016. Equality of opportunity in supervised learning. In *NeurIPS*.
[11] Robert David Hart. 2017. If you're not a white male, artificial intelligence's use in healthcare could be dangerous. *Quartz* (2017).
[12] Claire Cain Miller. 2015. Can an algorithm hire better than a human. *The New York Times* 25 (2015).
[13] Gourab K Patro, Abhijnan Chakraborty, Niloy Ganguly, and Krishna P Gummadi. 2019. Incremental Fairness in Two-Sided Market Platforms: On Updating Recommendations Fairly. *arXiv preprint arXiv:1909.10005* (2019).
[14] Kevin Petrasic, Benjamin Saul, James Greig, Matthew Bornfreund, and Katherine Lamberth. 2017. Algorithms and bias: What lenders need to know. *White & Case* (2017).
[15] Tom Sühr, Asia J Biega, Meike Zehlike, Krishna P Gummadi, and Abhijnan Chakraborty. 2019. Two-Sided Fairness for Repeated Matchings in Two-Sided Markets: A Case Study of a Ride-Hailing Platform. In *ACM KDD*.
[16] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. 2017. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *WWW*.
[17] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rogriguez, and Krishna P Gummadi. 2017. Fairness Constraints: Mechanisms for Fair Classification. In *AISTATS*.
[18] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. Fa* ir: A fair top-k ranking algorithm. In *ACM CIKM*.