Fairness in Rankings and Recommenders: Models, Methods and Research Directions

Evaggelia Pitoura
University of Ioannina, Greece
pitoura@cs.uoi.gr

Kostas Stefanidis
Tampere University, Finland
konstantinos.stefanidis@tuni.fi

Georgia Koutrika
Athena Research Center, Greece
georgia@athenarc.gr

ABSTRACT

We increasingly depend on a variety of data-driven algorithmic systems to assist us in many aspects of life. Search engines and recommendation systems amongst others are used as sources of information and to help us in making all sort of decisions from selecting restaurants and books, to choosing friends and careers. This has given rise to important concerns regarding the fairness of such systems. This tutorial aims at presenting a toolkit of definitions, models and methods used for ensuring fairness in rankings and recommendations. Our objectives are three-fold: (*a*) to provide a solid framework on a novel, quickly evolving, and impactful domain, (*b*) to present related methods and put them into perspective, and (*c*) to highlight challenges and research paths for researchers and practitioners that work in data management and applications.

I. Introduction

Algorithmic systems driven by large amounts of data are increasingly being used in all aspects of society. Such systems offer enormous opportunities. They accelerate scientific discovery in all domains, including personalized medicine and smart weather forecasting, they automate tasks, they help in improving our life through personal assistants and recommendations, they have the potential of transforming society through open government, to name just a few of their benefits.

Often, such systems are used to assist, or, even replace human decision making in diverse domains, including school admissions, housing, pricing of goods, credit score estimation, and job applicant selection. A prominent case is the COMPAS software used in courts in the US to assist bail and sentencing decisions through a risk assessment algorithm that predicts future crime. The ubiquitous use of such systems may create possible threats of economic loss, social stigmatization, or even loss of liberty. For instance, a known study by ProPublica found that in COMPAS, the false positive rate for African American defendants, namely people labelled "high-risk" who did not re-offend, was nearly twice as high as that for white defendants [1]. Another well-known study shows that names used mostly by men and women of colour are much more likely to generate ads related to arrest records [2].

Data-driven systems are also being employed by search and recommendation engines, social media tools, and news outlets, among others. Recent studies report that social media has become the main source of online news with more than 2.4 billion internet users, of which nearly 64.5% receive breaking news from social media instead of traditional sources [3]. Thus, to a great extent, such systems play a central role in shaping our experiences and influencing our perception of the world. Again, there are many reports questioning the output of such systems. For instance, a known study on search results showed evidence for stereotype exaggeration in images returned when people search for professional careers [4].

Fairness in rankings and recommenders. In this tutorial, we pay special attention to the concept of fairness in rankings and in recommendation systems. By fairness, we typically mean lack of discrimination (bias). Bias may come from the algorithm, reflecting, for example, commercial or other preferences of its designers, or even from the actual data, for example, if a survey contains biased questions, or, if some specific population is misrepresented in the input data.

As fairness is an elusive concept, an abundance of definitions and models of fairness as well as several algorithmic approaches for fair rankings and recommendations have been proposed, thus making the landscape very convoluted. We believe that in order to make real progress in building fairness-aware systems, we need to de-mystify what has been done, understand how and when each model and approach can be used, and, finally, distinguish the research challenges ahead of us. Therefore, in this tutorial, we follow a systematic and structured approach to explain the various sides of and approaches to fairness. First, we lay the ground by presenting general fairness definitions. Then, we zoom in on models for rankings and recommendations. We organize them in a taxonomy and highlight their differences and commonalities as well as the opportunities for cross-domain transfer. We move on to describing solutions for fair rankings and recommendations. We organize them into pre-, in- and post-processing approaches. Within each category, we further classify approaches along several dimensions. Based on this structure, we discuss open research challenges pertaining to fairness in the broader context of data management and on designing, building, managing, and evaluating fair data systems and applications.

II. Tutorial Outline

A. Motivation and Background

In this tutorial, we start by presenting motivating examples for the need for fair rankings and recommendations from several domains, including justice, ads, image search and others. We highlight possible causes of unfairness, such as biased or incomplete data, and algorithmic inefficiencies. We point out potential harms, such as filter bubbles, polarization, loss of opportunity, and discrimination.

B. General Fairness Definitions

Fairness is a general term and coming up with a single definition or model is tricky. We start this part of the tutorial by reviewing definitions of fairness which, in general, ask for nondiscrimination of users or items, based on the values of one or more sensitive or protected attributes, such as gender or race. We organize the definitions with respect to the notions of *individual fairness*, i.e., treating similar individuals similarly [5], [6], and *group fairness*, i.e., treating different groups equally (e.g., nondiscrimination of sensitive groups) [7], [8].

Most work so far has focused on classification algorithms used in decision making. Since many approaches to the fair ranking and recommendation tasks build on definitions of fairness in classification, we will present a short survey of this work for completeness [9], [10], including: (a) *Demographic (or statistical) parity* (e.g., [8]), stating that the proportion of each part of a protected class (e.g., gender) receiving a positive outcome should follow that in the general population, (b) *Calibration-based fairness* (e.g., [11]), stating that if a group receives a predicted probability p, at least a fraction p of its members should belong to the predicted class, (c) *Counterfactual fairness* (e.g., [6]), stating that a decision for an individual is fair, if it is the same in both the actual world and a counterfactual world where the individual belongs to a different demographic group, and (d) *Conditional statistical parity* (e.g., [9]), which defines statistical parity given a set of legitimate factors.

C. Fairness in Rankings and Recommenders

When it comes to ranking and recommender systems, we define three dimensions to classify fairness models: *level* (individual vs group), *side* (producer vs consumer), and *graduality* (single vs sequential output) of fairness. We review fairness models for ranked outputs and recommendations and we classify them using the dimensions of our taxonomy.

A central issue in ranking is position bias, i.e., the fact that items ranked at the top positions tend to attract most of the user attention. We review a variety of related models, including *fairness constraints* [12], *discounted cumulative fairness* [13], *fairness of exposure* [14], and *equity of attention* [15], as well as approaches based on pair-wise comparisons [16]. We also look into fair ranking in graphs, e.g., [17].

Then, we look at how definitions of algorithmic fairness and fair ranking have been *adopted in recommender systems* (e.g., [18], [19]). We distinguish between the multiple sides that fairness can have in recommendation systems, namely (*a*) fairness for the recommended items (e.g., [18]), (*b*) fairness for the users (e.g., [20], [21]), (*c*) fairness for groups of users (e.g., [22]–[24]) and (*d*) fairness for the item providers, and the recommendation platform (e.g., [25]). We also investigate the notion of gradual fairness in sequential and multi-round recommenders [25]–[27], where the goal is to ensure fairness in a number of interactions between the users and the system.

D. Methods

We first discuss the trade-offs among fairness, personalization and accuracy. Taking a cross-type view, we present approaches divided into three categories:

1) Pre-processing methods: We present pre-processing approaches that modify the input to the system so that any underlying bias or discrimination is removed, for example: by appropriate *sampling* (e.g., [28]), by adding more data to the input (e.g., [18]), or by performing *database repair* [29].

2) In-processing methods: These methods target at modifying existing or introducing new algorithms that result in fair rankings and recommendations.

In-processing approaches for fair rankings modify the result generation process to allow the systematic control of the degree of unfairness in the output. One family of approaches targets *learning to rank*. One technique to achieve fairness is by introducing an intermediate level between the input and the output of the learning system that constitutes a fair representation of the input [13], [30]. Another technique is *adding regularization terms* to the loss function of the learning system to capture fairness constraints [31]. Another line of research considers linear ranking function where the score of an item is a weighted sum of some of the feature of the item. The goal in this case is to *adjust the weights* so as to achieve fairness [32].

In recommenders, we first study fairness in systems that produce recommendations for individuals, which comprise the majority of existing recommenders. We will present algorithms for *fair matrix factorization* [21], [33], *multi-armed bandits* [34], [35] and *deep learning recommenders* (e.g., [26], [36], [37]). For instance, we show that when fairness with respect to both consumers and to item providers is important, variants of the well-known sparse linear method (SLIM) can be used to negotiate the trade-off between fairness and accuracy and improve the balance of user and item neighborhoods [33]. Alternatively, we can augment the learning objective in matrix factorization by adding a smoothed variation of a fairness metric [21]. Another approach is to mitigate bias by incorporating randomness in variational autoencoders (e.g., [26]).

3) Post-processing methods: These methods treat the algorithms for producing rankings and recommendations as black boxes, without changing their inner workings. To ensure fairness, they modify the output of the algorithm.

For rankings, we will present a generative process for producing fair rankings that aims at satisfying *statistical tests of representativeness* when ranking items in a certain order [13], [38]. We will also present works based on *constraint optimization formulations* of the problem [39], targeting at relevance maximization in terms of exposure allocation, and also works on *amortized fairness* [40], which consider that the accumulated attention across a series of rankings should be proportional to accumulated relevance, as indicating long term ranking fairness.

Finally, we present post-processing approaches that modify the output of the recommenders to ensure fairness (e.g., [41]). Moving from individuals to groups, group recommendations have attracted significant research efforts for their importance in benefiting a group of users. However, maximizing the satisfaction of each group member while minimizing the unfairness between them is very challenging. We study different fair-aware algorithms for group recommenders [24], [42]–[44].

E. Open Issues and Research Directions

Taking a broader look at algorithmic fairness, we discuss algorithmic fairness as a question about programs and their properties [45], [46]. Is a program fair, under some definition of fairness? Can we quantify how fair it is, in some way? We will discuss two approaches: (*a*) verifying if a program is fair [47], and (*b*) having fairness as a first-class citizen in programming [48].

We present a critical comparison of the existing work on ensuring fair rankings and recommendations, and the lessons learnt in these areas. We discuss open research challenges pertaining to fairness in the broader context of data management and on designing, building, managing, and evaluating fair data systems and applications.

Finally, while the potential benefits of fairness are well-accepted nowadays, we need to study the actual impact of fairness-enhancing algorithms. By testing people's perception of different fairness definitions, we can find definitions that are appropriate for particular contexts [49]–[51]. Extensive studies are needed to evaluate the level of acceptance of the fairness-enhanced results by the users and the long term effect of these results on their own perceptions and preferences. The relation of fairness with other requirements in designing socially-aware data-driven systems such as diversity and transparency [52] should also be investigated deeper as well as the connections of fairness with explainability and personalization.

III. Related Tutorials

The following three tutorials have a stricter focus than ours, the first one focusing on concepts and metrics of fairness and the challenges in applying these to recommendation and information retrieval while the latter two focusing on scoring methods. On the other hand, our tutorial has a much wider coverage and depth, presenting a structured survey and comparison of methods and models for ensuring fairness in rankings and recommendations, based on our extensive survey of the area currently under submission.

- *M. D. Ekstrand, R. Burke, F. Diaz. Fairness and Discrimination in Recommendation and Retrieval. RecSys 2019.*
- *A. Asudeh, H. V. Jagadish. Fairly Evaluating and Scoring Items in a Data Set. PVLDB, 2020.*
- *H. Oosterhuis, R. Jagerman, M. de Rijke. Unbiased Learning to Rank: Counterfactual and Online Approaches. WWW 2020.*

The following tutorials focus on fairness issues especially in the context of machine learning and data mining.

- *S. Bird, B. Hutchinson, K. Kenthapadi, E. Kiciman, M. Mitchell. Fairness-aware Machine Learning: Practical Challenges and Lessons Learned. KDD2019, WWW2019, WSDM 2019.*
- *S Barocas, M. Hardt. Fairness in Machine Learning. NIPS 2017.*
- *F. Bonchi, C. Castillo, S. Hajia. Algorithmic bias: from discrimination discovery to fairness-aware data mining. KDD 2016.*

Our earlier, EDBT 2020 tutorial on "Fairness in Rankings and Recommenders" [53] was a short, 1-hour, introduction. The current proposal of 3 hours will provide a deeper and more systematic coverage of the existing methods for ensuring fairness in recommendations and rankings.

IV. Presenters

Evaggelia Pitoura is a Professor at the Univ. of Ioannina, Greece, where she also leads the Distributed Management of Data Laboratory. She received her PhD degree from Purdue Univ., USA. Her research interests are in data management systems with a recent emphasis on social networks and responsible data management. Her publications include more than 150 articles in international journals (including TODS, TKDE, PVLDB) and conferences (including SIGMOD, ICDE, WWW) and a highly-cited book on mobile computing. Her research has been funded by the EC and national sources. She has served or serves on the editorial board of ACM TODS, VLDBJ, TKDE, DAPD and as a group leader, senior PC member, or co-chair of many international conferences (including PC chair of EDBT 2016 and ICDE 2012). She has more than 20 years experience in teaching. **Prior tutorials:** Temporal Graphs [eBISS'17], Social Graphs [BigDat'15], Data Graphs [SummerSOC'14], Personalization [ICDE'10], Mobile Computing [ICDE'03], Pervasive Computing [ICDE'00].

Kostas Stefanidis is an Assoc. Professor on Data Science at the Tampere University, Finland. He got his PhD in personalized data management from the Univ. of Ioannina, Greece. His research interests lie in the intersection of databases, information retrieval, data mining and the Web, and include personalization and recommender systems, large-scale entity resolution and information integration, and query and data exploration paradigms. His publications include more than 80 papers in peer-reviewed conferences and journals, including SIGMOD, ICDE, and ACM TODS, and a book on entity resolution in the Web of data. He has 8 years experience in teaching. **Prior tutorials:** Recommender Systems [MUMIA Training School'14], Personalization [ICDE'10], Entity Resolution [ICDE'17, ESWC'16, WWW'14, CIKM'13].

Georgia Koutrika is a Research Director at Athena Research Center in Greece. She has more than 15 years of experience in multiple roles at HP Labs, IBM Almaden, and Stanford. Her work focuses on data exploration, recommendations, and data analytics, and has been incorporated in commercial products, described in 14 granted patents and 26 patent applications in the US and worldwide, and published in more than 90 papers in top-tier conferences and journals. She is Editor-in-chief for VLDB Journal, PC chair for VLDB 2023, associate editor for TKDE, and an ACM Distinguished Speaker. She has served or serves as PC member or co-chair of many conferences. **Prior tutorials:** Fairness in Rankings and Recommenders [EDBT20], Recommender Systems [SIGMOD'18, EDBT'18, ICDE'15], Personalization [ICDE'10, ICDE'07, VLDB'05].

References

[1] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine bias," *ProPublica*, 2016. [Online]. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[2] L. Sweeney, "Discrimination in online ad delivery," *Commun. ACM*, vol. 56, no. 5, pp. 44–54, 2013.

[3] N. Martin, "How social media has changed how we consume news," *Forbes*, 2018. [Online]. Available: https://www.forbes.com/sites/nicolemartin1/2018/11/30/how-social-media-has-changed-how-we-consume-news/18ae4c093c3c

[4] M. Kay, C. Matuszek, and S. A. Munson, "Unequal representation and gender stereotypes in image search results for occupations," in *CHI*, 2015.

[5] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. S. Zemel, "Fairness through awareness," in *Innovations in Theoretical Computer Science*, 2012.

[6] M. J. Kusner, J. R. Loftus, C. Russell, and R. Silva, "Counterfactual fairness," in *NIPS*, 2017.

[7] P. Awasthi, M. Kleindessner, and J. Morgenstern, "Effectiveness of equalized odds for fair classification under imperfect group information," *CoRR*, vol. abs/1906.03284, 2019.

[8] V. Tsintzou, E. Pitoura, and P. Tsaparas, "Bias disparity in recommendation systems," in *RMSE*, 2019.

[9] S. Verma and J. Rubin, "Fairness definitions explained," in *FairWare*, 2018.

[10] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *CoRR*, vol. abs/1908.09635, 2019.

[11] G. Pleiss, M. Raghavan, F. Wu, J. Kleinberg, and K. Q. Weinberger, "On fairness and calibration," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 5680–5689.

[12] L. E. Celis, D. Straszak, and N. K. Vishnoi, "Ranking with fairness constraints," in *ICALP*, 2018.

[13] K. Yang and J. Stoyanovich, "Measuring fairness in ranked outputs," in *SSDM*, 2017.

[14] A. Singh and T. Joachims, "Fairness of exposure in rankings," in *KDD*, 2018.

[15] A. J. Biega, K. P. Gummadi, and G. Weikum, "Equity of attention: Amortizing individual fairness in rankings," in *SIGIR*, 2018.

[16] A. Beutel, J. Chen, T. Doshi, H. Qian, L. Wei, Y. Wu, L. Heldt, Z. Zhao, L. Hong, E. H. Chi, and C. Goodrow, "Fairness in recommendation ranking through pairwise comparisons," in *KDD*, 2019.

[17] S. Tsioutsiouliklis, E. Pitoura, P. Tsaparas, I. Kleftakis, and N. Mamoulis, "Fairness aware pagerank," in *WWW*, 2021.

[18] H. Steck, "Calibrated recommendations," in *RecSys*, 2018.

[19] S. Yao and B. Huang, "New fairness metrics for recommendation that embrace differences," *FAT/ML*, 2017.

[20] J. Leonhardt, A. Anand, and M. Khosla, "User fairness in recommender systems," in *WWW*, 2018.

[21] S. Yao and B. Huang, "Beyond parity: Fairness objectives for collaborative filtering," in *NIPS*, 2017.

[22] S. Amer-Yahia, S. B. Roy, A. Chawla, G. Das, and C. Yu, "Group recommendation: Semantics and efficiency," *PVLDB*, vol. 2, no. 1, pp. 754–765, 2009.

[23] E. Ntoutsi, K. Stefanidis, K. Nørvåg, and H. Kriegel, "Fast group recommendations by applying user clustering," in *ER*, 2012.

[24] D. Serbos, S. Qi, N. Mamoulis, E. Pitoura, and P. Tsaparas, "Fairness in package-to-group recommendations," in *WWW*, 2017.

[25] G. K. Patro, A. Chakraborty, N. Ganguly, and K. P. Gummadi, "Incremental fairness in two-sided market platforms: On smoothly updating recommendations," in *AAAI*, 2020.

[26] R. Borges and K. Stefanidis, "Enhancing long term fairness in recommendations with variational autoencoders," in *MEDES*, 2019.

[27] M. Stratigi, J. Nummenmaa, E. Pitoura, and K. Stefanidis, "Fair sequential group recommendations," in *SAC*, 2020.

[28] L. E. Celis, A. Deshpande, T. Kathuria, and N. K. Vishnoi, "How to be fair and diverse?" *CoRR*, vol. abs/1610.07183, 2016.

[29] B. Salimi, L. Rodriguez, B. Howe, and D. Suciu, "Interventional fairness: Causal database repair for algorithmic fairness," in *SIGMOD*, 2019.

[30] R. S. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork, "Learning fair representations," in *ICML*, 2013.

[31] M. Zehlike, G.-T. Diehn, and C. Castillo, "Reducing disparate exposure in ranking: A learning to rank approach," in *The Web Conf (WWW)*, 2020.

[32] A. Asudeh, H. V. Jagadish, J. Stoyanovich, and G. Das, "Designing fair ranking schemes," in *SIGMOD*, 2019.

[33] R. Burke, "Multisided fairness for recommendation," *CoRR*, vol. abs/1707.00093, 2017.

[34] M. Joseph, M. J. Kearns, J. H. Morgenstern, and A. Roth, "Fairness in learning: Classic and contextual bandits," in *NIPS*, 2016.

[35] Y. Liu, G. Radanovic, C. Dimitrakakis, D. Mandal, and D. C. Parkes, "Calibrated fairness in bandits," *CoRR*, vol. abs/1707.01875, 2017.

[36] Z. Zhu, X. Hu, and J. Caverlee, "Fairness-aware tensor-based recommendation," in *CIKM*, 2018.

[37] R. Borges and K. Stefanidis, "On mitigating popularity bias in recommendations via variational autoencoders," in *SAC*, 2021.

[38] M. Zehlike, F. Bonchi, C. Castillo, S. Hajian, M. Megahed, and R. Baeza-Yates, "Fa*ir: A fair top-k ranking algorithm," in *CIKM*, 2017.

[39] A. Singh and T. Joachims, "Fairness of exposure in rankings," in *KDD*, Y. Guo and F. Farooq, Eds., 2018.

[40] A. J. Biega, K. P. Gummadi, and G. Weikum, "Equity of attention: Amortizing individual fairness in rankings," in *SIGIR*, 2018.

[41] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma, "Recommendation independence," in *FAT*, 2018.

[42] X. Lin, M. Zhang, Y. Zhang, Z. Gu, Y. Liu, and S. Ma, "Fairness-aware group recommendation with pareto-efficiency," in *RecSys*, 2017.

[43] D. Sacharidis, "Top-n group recommendations with fairness," in *SAC*, 2019.

[44] M. Stratigi, H. Kondylakis, and K. Stefanidis, "Fairgrecs: Fair group recommendations by exploiting personal health information," in *DEXA*, 2018.

[45] E. C. Cortés and D. Ghosh, "An invitation to system-wide algorithmic fairness," in *AIES*, A. N. Markham, J. Powles, T. Walsh, and A. L. Washington, Eds., 2020.

[46] N. Kallus, X. Mao, and A. Zhou, "Assessing algorithmic fairness with unobserved protected class using data combination," in *FAT**, 2020.

[47] A. Albarghouthi, L. D'Antoni, S. Drews, and A. V. Nori, "Fairness as a program property," *CoRR*, vol. abs/1610.06067, 2016. [Online]. Available: http://arxiv.org/abs/1610.06067

[48] A. Albarghouthi and S. Vinitsky, "Fairness-aware programming," in *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* 2019, Atlanta, GA, USA, January 29-31, 2019*, 2019.

[49] N. Grgic-Hlaca, E. M. Redmiles, K. P. Gummadi, and A. Weller, "Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction," in *WWW*, 2018.

[50] A. C. Plane, E. M. Redmiles, M. L. Mazurek, and M. C. Tschantz, "Exploring user perceptions of discrimination in online targeted advertising," in *USENIX*, 2017.

[51] N. A. Saxena, K. Huang, E. DeFilippis, G. Radanovic, D. C. Parkes, and Y. Liu, "How do fairness definitions fare?: Examining public attitudes towards algorithmic definitions of fairness," in *AIES*, 2019.

[52] E. Pitoura, "Social-minded measures of data quality: Fairness, diversity, and lack of bias," *ACM J. Data Inf. Qual.*, vol. 12, no. 3, 2020.

[53] E. Pitoura, G. Koutrika, and K. Stefanidis, "Fairness in rankings and recommenders," in *EDBT*, 2020.