Fairness-aware Methods in Rankings and Recommenders

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 recommender 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. In this tutorial, we aim at presenting a toolkit of methods used for ensuring fairness in rankings and recommendations. Our objectives are two-fold: (*a*) to present related methods of this novel, quickly evolving and impactful domain, and put them into perspective, and (*b*) to highlight open challenges and research paths for future work.

I. Introduction

Algorithmic systems, driven by large amounts of data, are increasingly being used in all aspects of society to assist people in forming opinions and taking decisions. Such algorithmic systems offer enormous opportunities, since they accelerate scientific discovery in various domains, including personalized medicine, smart weather forecasting and many other fields. They can also automate tasks regarding simple personal decisions, and help in improving our daily life through personal assistants and recommendations, like where to eat and what are the news. Moving forward, they have the potential of transforming society through open government and many more benefits.

Often, such systems are used to assist, or, even replace human decision making in diverse domains. Examples include software systems used in school admissions, housing, pricing of goods and services, credit score estimation, job applicant selection, and sentencing decisions in courts and surveillance. This automation raises concerns about how much can or should we trust such systems.

There are many reports and studies questioning the output of such decision support systems. For example, what images do people choose to represent careers, like for instance, in image search, when the query is about doctors or nurses, what is the percentage of images portraying women that we get in the result. Few years ago, [1] has found evidence for stereotype exaggeration and systematic underrepresentation of women when compared with the actual percentage, as estimated by the US Bureau of labor and statistics. Two interesting conclusions from the study were that people prefer and rate search results higher when these results are consistent with stereotypes. Another interesting result is that if you shift the representation of gender in image search results then the people's perception about real world distribution tend to shift as well.

Another well-known example is the COMPAS system, which is a commercial tool that uses a risk assessment algorithm to predict some categories of future crime. Specifically, this tool is used in courts in the US to assist bail and sentencing decisions, and it was found that the false positive rate, that is the people who were labeled by the tool as high risk but did not re-offend, was nearly twice as high for African-American as for white defendants [2]. This means that many times the ubiquitous use of decision support systems may create possible threats of economic loss, social stigmatization, or even loss of liberty. There are many more case studies, like the above ones. For example, names that are used by men and women of color are much more likely to generate ads related to arrest records [3]. Also using a tool called Adfisher, it was found that if you set the gender to female, this will result in getting ads for less high paid jobs[1]. Or, in the case of word embeddings the vector that represent computer programming is closer to men than to women.

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 [4]. Thus, to a great extent, such systems play a central role in shaping our experiences and influencing our perception of the world. Motivated by this, it is important to understand what causes the bias of the previous examples. For example, bias may come from the data, meaning that data may be incorrect, incomplete, may be poorly selected and outdated, or may reflect and promote historical biases. Differently, a system may learn a majority, which may result in errors concentrated in the minority class. In other cases, bias may come from the algorithms that are often seen as black boxes, making unrealistic assumptions and with output models that are hard to understand. All these reasons create a bias reinforcement cycle.

[1] https://fairlyaccountable.org/adfisher/

Focus of this tutorial. In this tutorial, we pay special attention to the concept of fairness in rankings and recommender systems [5]. 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 models of fairness have been proposed, as well as several algorithmic approaches for fair rankings and recommendations, making the landscape very convoluted. 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, we follow a systematic approach to explain the various sides of and approaches to fairness. We start by presenting models for rankings and recommendations. We organize them in a taxonomy and highlight their differences and commonalities. We distinguish between *individual* and *group* fairness, *consumer* and *producer* fairness, and fairness for *single* and *multiple* outputs.

We pay special attention on describing solutions for fair rankings and recommendations. We organize them into *pre-processing approaches*, that aim at transforming the data to remove any underlying bias or discrimination, *in-processing approaches*, that aim at modifying existing or introducing new algorithms that result in fair rankings and recommendations, and *post-processing approaches*, that modify the output of the algorithm. Within each category, we further classify approaches along several dimensions.

Finally, we discuss other cases where a system needs to make decisions and where fairness is also important, and present open research challenges pertaining to fairness in the broader context of data management.

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. Fairness in Rankings and Recommenders

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 [6], [7], and *group fairness*, i.e., treating different groups equally (e.g., nondiscrimination of sensitive groups) [8], [9].

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* [10], *discounted cumulative fairness* [11], *fairness of exposure* [12], and *equity of attention* [13], as well as approaches based on pair-wise comparisons [14]. We also look into fair ranking in graphs, e.g., [15].

Then, we look at how definitions of algorithmic fairness and fair ranking have been *adopted in recommender systems* (e.g., [16], [17]). We distinguish between the multiple sides that fairness can have in recommendation systems, namely (*a*) fairness for the recommended items (e.g., [16]), (*b*) fairness for the users (e.g., [18], [19]), (*c*) fairness for groups of users (e.g., [20]–[22]) and (*d*) fairness for the item providers, and the recommendation platform (e.g., [23]). We also investigate the notion of gradual fairness in sequential and multi-round recommenders [23]–[25], where the goal is to ensure fairness in a number of interactions between the users and the system.

C. 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: Often, bias can exist in the underlying data on which systems are trained [26], and it can take two forms. *Bias in the rows* of the data exists when there are not enough representative individuals from minority (sub)groups. For example, according to a Reuters article [27], Amazon's experimental automated system to review job applicants' resumes showed a significant gender bias towards male candidates over females that was due to historical discrimination in the training data.

Bias in the columns is when features are biased (correlated) with sensitive attributes. For example, zip code tends to predict race due to a history of segregation [28]. Direct discrimination occurs when protected attributes are used explicitly in making decisions (i.e., *disparate treatment*). More pervasive nowadays is indirect discrimination, in which protected attributes are not used but reliance on variables correlated with them leads to significantly different outcomes for different groups, also known as *disparate impact*.

To address bias and avoid discrimination, several methods have been proposed for pre-processing data, for example: by adding more data to the input (e.g., [16]), by performing *database repair* [29], or by appropriate *sampling* (e.g., [30]).

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 [11], [31]. Another technique is *adding regularization terms* to the loss function of the learning system to capture fairness constraints [32]. 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 [33].

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* [19], [34], *multi-armed bandits* [35], [36] and *deep learning recommenders* (e.g., [24], [37], [38]). 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 [34]. Alternatively, we can augment the learning objective in matrix factorization by adding a smoothed variation of a fairness metric [19]. Another approach is to mitigate bias by incorporating randomness in variational autoencoders (e.g., [24]).

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 [11], [39]. We will also present works based on *constraint optimization formulations* of the problem [40], targeting at relevance maximization in terms of exposure allocation, and also works on *amortized fairness* [41], 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., [42]). 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 [22], [43]–[45].

D. Open Issues and Research Directions

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.

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.

- *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.*

Previous editions of this tutorial include a shorter 1-hour version in EDBT 2020 [46] which placed more emphasis on models than on methods and a longer 3-hour version to be presented in the upcoming ICDE 2021 conference [47]. In this edition, the focus will be on methods for achieving fairness.

IV. Presenters

Evaggelia Pitoura is a Professor at the Univ. of Ioannina, Greece, where she also leads the Distributed Management of Data Laboratory. 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).

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.

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.

REFERENCES

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

[2] J. A. et al, "Machine bias," *ProPublica*, 2016. [Online]. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

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

[4] 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

[5] E. Pitoura, K. Stefanidis, and G. Koutrika, "Fairness in rankings and recommendations: An overview," *CoRR*, vol. abs/2104.05994, 2021. [Online]. Available: https://arxiv.org/abs/2104.05994

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

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

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

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

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

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

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

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

[14] 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.

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

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

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

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

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

[20] 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.

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

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

[23] 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.

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

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

[26] T. Calders and I. Zliobaite, "Why unbiased computational processes can lead to discriminative decision procedures," in *Discrimination and Privacy in the Information Society - Data Mining and Profiling in Large Databases*, ser. Studies in Applied Philosophy, Epistemology and Rational Ethics. Springer, 2013, vol. 3, pp. 43–57.

[27] J. Dastin, "Rpt-insight-amazon scraps secret ai recruiting tool that showed bias against women," *Reuters*, 2018.

[28] D. Ingold and S. Soper, "Amazon doesn't consider the race of its customers. should it?" *Bloomberg*, 2016.

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

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

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

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

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

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

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

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

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

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

[39] 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.

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

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

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

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

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

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

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

[47] E. Pitoura, K. Stefanidis, and G. Koutrika, "Fairness in rankings and recommenders: Models, methods and research directions," in *ICDE*, 2021.