Tutorial on Fairness of Machine Learning in Recommender Systems

Yunqi Li
Department of Computer Science
Rutgers University, NJ, US
yunqi.li@rutgers.edu

Yingqiang Ge
Department of Computer Science
Rutgers University, NJ, US
yingqiang.ge@rutgers.edu

Yongfeng Zhang
Department of Computer Science
Rutgers University, NJ, US
yongfeng.zhang@rutgers.edu

ABSTRACT

Recently, there has been growing attention on fairness considerations in machine learning. As one of the most pervasive applications of machine learning, recommender systems are gaining increasing and critical impacts on human and society since a growing number of users use them for information seeking and decision making. Therefore, it is crucial to address the potential unfairness problems in recommendation, which may hurt users' or providers' satisfaction in recommender systems as well as the interests of the platforms. The tutorial focuses on the foundations and algorithms for fairness in recommendation. It also presents a brief introduction about fairness in basic machine learning tasks such as classification and ranking. The tutorial will introduce the taxonomies of current fairness definitions and evaluation metrics for fairness concerns. We will introduce previous works about fairness in recommendation and also put forward future fairness research directions. The tutorial aims at introducing and communicating fairness in recommendation methods to the community, as well as gathering researchers and practitioners interested in this research direction for discussions, idea communications, and research promotions.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Computing methodologies** → **Artificial intelligence**;

KEYWORDS

Recommender Systems; Machine Learning; Fairness; AI Ethics

ACM Reference Format:
Yunqi Li, Yingqiang Ge, and Yongfeng Zhang. 2021. Tutorial on Fairness of Machine Learning in Recommender Systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3404835.3462814

1 MOTIVATION

Recommender systems are playing an important role on assisting human decision making. The satisfaction of users and the interests of platforms are closely related to the quality of generated recommendation results. However, as a highly data-driven system, recommender systems could be affected by data or algorithmic bias, thus generating unfair results, which could weaken the reliance of the systems. As a result, it is necessary and important to consider fairness issues in recommendation settings.

Fairness has attracted rapidly-growing attention in machine learning research communities, which include two basic tasks—fair classification [19, 43, 51, 63] and fair ranking [8, 48, 49, 62]. Though fair ranking models could be borrowed to recommendation settings in some cases, the algorithms need to be carefully designed to satisfy the requirements for recommendation. This is because the concept of fairness in recommender systems has been extended to multiple stakeholders [9], i.e., the unfairness issue should be considered not only from a single aspect such as item or provider side, but also from user-side or multi-sided perspectives. Such problems together with other challenges originated from recommendation settings, such as extreme data sparsity, make it challenging to directly apply the techniques in fairness ranking to recommendation scenarios.

This tutorial gives a retrospect of fairness research works in recommender systems, and provides the audience an intuitive understanding of fairness issues, evaluation strategies, and challenges under recommendation settings. This would help and encourage researchers, practitioners and even those new to fairness but interested in recommendation fairness to start their research work.

2 OBJECTIVES

This tutorial will help the audience to achieve the following goals:

- Get the background knowledge of fairness research works in general machine learning.
- Understand the challenges of fairness in recommender systems compared to the conventional fair ranking works.
- Understand the taxonomy of fairness concepts in recommendation, such as group fairness vs. individual fairness, single-sided fairness vs. multi-sided fairness, static fairness vs. dynamic fairness, etc.
- Understand the existing metrics and evaluation protocols to assessing fairness in particular problem settings.

3 TUTORIAL OUTLINE

3.1 Introduction

First, to address the importance and necessity for considering unfairness issues in recommendation, we will provide several examples to show how recommender systems will result in unfair results for users or items without fairness considerations, which will hurt the satisfaction of users or providers, as well as hurt the interest of the platform. Next, we will retrospect the considerations for fairness in

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8037-9/21/07…$15.00
https://doi.org/10.1145/3404835.3462814

machine learning areas in order to help participants better understand algorithmic fairness, as well as introduce the more complex situations and challenges that need to be considered when studying fairness in recommender systems.

Specifically, the first endeavor to achieve fairness in the community is to consider fairness in classification tasks, which design algorithms that are compatible with fairness constraints [51, 63]. For binary classification, fairness metrics can be expressed by rate constraints, which regularize the classifier's positive or negative rates over different protected groups [19, 43]. To achieve fairness, the training objective is usually optimized together with such constraints over fairness metrics [5, 31]. What's more, some recent works have also considered the fairness of ranking tasks. Some works directly learn a ranking model from scratch [44, 49, 62], while others consider re-ranking or post-processing algorithms for fair ranking [8, 13]. The fairness metrics for ranking tasks are usually defined over the exposure of items that belong to different protected groups, and such metrics include both unsupervised criteria and supervised criteria [44].

Recommendation algorithms can usually be considered as a type of ranking algorithm. However, the ranking problem usually only considers fairness issue from the perspective of items, while the concept of fairness in recommender systems has been extended to multiple stakeholders [9]. Besides, since recommender systems are complex with usually multiple models and multiple goals to balance, studying fairness in recommender systems present unique challenges. The problem of extreme sparsity and numerous dynamics in recommender systems also bring additional challenges in improving recommendation fairness.

3.2 Taxonomy for Fairness in Recommendation

Next, we introduce taxonomies for fairness considerations in recommender systems. In particular, we can see fairness in recommendations from various perspectives, including group vs. individual fairness; single- vs. multi-sided fairness; static and dynamic fairness; associative vs. causal fairness, etc. The details are as follows:

Group vs. Individual Fairness: In recent studies on algorithmic fairness, there are two basic frameworks: group fairness and individual fairness. Group fairness demands that protected groups should be treated similarly to the advantaged group or the populations as a whole [47]. The group fairness perspective for supervised learning usually implies constraints such as equalized odds [32, 60] and demographic parity [11]. Individual fairness requires that similar treatment should be received by each similar individuals, which is hard to define precisely due to the lack of agreement on task-specific similarity metrics for individuals [20]. There are some works about considering group fairness in recommendations. Li et al. [38] consider the active and inactive user groups be treated similarly; Fu et al. [23] require to impair the group unfairness problem in the context of explainable recommendation over knowledge graphs with a fairness constrained approach; Lin et al. [40] provide an optimization framework for fairness-aware group recommendation from the perspective of Pareto Efficiency, and further study the fairness of measure trade-off in recommendations under a Pareto optimization framework. Besides, Patro et al. [45] view individual fairness from both producers and customers sides, and response to the question of the long-term sustainability of two-sided platforms.

User vs. Item Fairness: Fairness requirements in recommender systems may come from different objects, including users or products/providers. Therefore, fairness in recommendations can be considered from both the user side or the item side. There are some researches considering fairness in recommendations on user side. Examples include [37], which quantify the user unfairness in post-processing algorithms with the original goal of improving recommendation diversity, and [38], which study the unfair performance between different groups of users. There are more works considering fairness in recommendations on item side. For example, Beutel et al. [7] show how to measure item fairness based on pairwise comparisons, and improve fairness by adding a regularizer when training recommendation models. What's more, there are lots of work concerning the popularity bias problem in recommendations, i.e., the less popular items will get less exposure than those frequently rated ones. This problem are often solved by increasing the number of recommended unpopular items (long-tail items) or otherwise the overall catalog coverage in these researches [2–4, 33].

Single-sided vs. Multi-sided Fairness: Only considering user or item side fairness in a recommendation system can be seen as dealing with single fairness while recommender systems are often considered as multi-stakeholders systems, which attempt to generate recommendations that satisfy the needs of both the end users and other parties or stakeholders. As a result, the concept of fairness in recommender systems has also been extended to multiple stakeholders [9] to meet the fairness requirements for users, items/providers, or multi-stakeholders. There have been a few works related to multi-sided fairness in multi-stakeholder recommendation systems [1, 10, 24, 41]. In [10], a regularization approach is applied to balance the weightings of different groups when generating recommendations. Mehrotra et al. [41] consider the trade-off between the consumer side and the supplier side and measured their impact on consumer satisfaction. Abdollahpouri and Burke [1] show the close connection between multi-stakeholder recommendation and multi-sided fairness.

Static vs. Dynamic Fairness: Static fairness is the one that does not consider the changes in the recommendation environment, such as the changes in item utility or attributes, therefore dynamic fairness has been studied recently, which considers the dynamic factors in the environment and learns a strategy that accommodates such dynamics. One research direction focuses on the changing utility of items, and works on it include [57] and [42], which incorporate user feedback in the learning process and dynamically adjust to the changing utility with fairness constraints. On the other hand, another type of dynamics, where group labels can be dynamic due to the nature of recommendations being an interactive process, has been explored by Ge et al. [26], which propose a fairness constrained policy optimization framework to deal with the changing group labels of items.

Associative vs. Causal Fairness: The research community firstly achieve fairness in machine learning by developing association-based (or correlation-based) fairness notions, with the aim to find the discrepancy of statistical metrics between individuals or sub-populations. For example, in binary classification, fairness metrics can be represented by regularizing the classifier's positive or negative rates over different protected groups [19, 43]. Recently, researchers have found that fairness cannot be well studied based only on association [34, 36, 64, 65]. The reason is that they cannot reason about the causal relations between features. However, unfair treatment may result from a causal relation between the sensitive features (e.g. gender) and model decisions (e.g. admission). Therefore, researchers have proposed causal-based fairness notions [36, 52], which are mostly defined on counterfactual reasoning or interventions [46]. In specific, counterfactual considers a hypothetical world beyond the real world, while intervention can be achieved by simulated random experiments. Li et al. [39] achieve personalized counterfactual fairness in recommender systems, while most of the previous works about fairness in recommendations consider the association-based fairness notions. We will show how causal-based considerations will open up new challenges and opportunities for studying fairness in recommendations.

Fairness Measures: Several works investigated fairness as a set of threshold-based constraints [12, 13, 61]. Recently, more works attempt to propose fairness metrics based on various constraints, such as distance and ratio between the proportion of a protected attribute and the overall attribute [58], pairwise comparisons regarding utility and prediction errors [7, 35, 59], as well as presented exposure distributions against the desired distribution [30, 58].

4 AUDIENCE AND RELEVANCE

The tutorial will be mainly targeting on information retrieval and recommendation system researchers and practitioners since we will mainly introduce the knowledge and research works about fairness in recommendations. The tutorial will also attract researchers who work in broader AI/ML communities especially AI Ethics such as fairness of machine learning, since we will briefly introduce fairness in other machine leaning tasks including the two basic tasks: classification and ranking. What's more, the tutorial will also attract industry researchers and practitioners from different areas, since fairness has attracted more and more attention in the industry because of the need for legitimacy and the promotion of commercial interests. This tutorial is closely connected to the fairness works at past SIGIR and related conferences such as WWW, KDD, RecSys. Previous work has aroused people's attention of fairness in recommender systems, and put forward the idea of how to formalize and achieve fairness under different recommendation scenarios. In this tutorial, we will provide an introduction to the growing literature on this topic, and extend those ideas by opening up new challenges and opportunities for studying fairness in recommendations.

There have been some tutorials about fairness concerns in search and recommendation including [21] (RecSys'19), [22] (SIGIR'19) and [25] (RecSys'20). The key difference between the previous tutorials and ours is that they consider fairness mainly from user study and evaluation perspective, while our tutorial will focus on fairness in recommendations from the AI and machine learning perspectives.

5 BRIEF BIO OF ORGANIZERS

Yunqi Li[1] is a Ph.D. student in the Department of Computer Science at Rutgers University advised by Prof. Yongfeng Zhang. Her research interests lie in the intersection of Machine Learning and Information Retrieval. Her recent researches focus on AI Ethics including bringing fairness [26, 38, 39] and interpretability [14, 56] to machine learning algorithms, as well as causal inference in machine learning [55, 56]. Her works have appeared in premier IR and AI/ML conferences such as WWW, SIGIR, WSDM, AAAI, etc.

Yingqiang Ge[2] is a PhD student at the Computer Science Department of Rutgers University supervised by Prof. Yongfeng Zhang. His research interests broadly lie in IR and machine learning, including economic recommendation [27, 28], explainable recommendation [54] and fairness in recommendation [23, 26, 29, 38], etc. His recent work on fairness includes fairness in explainable recommendation [23], long-term fairness in recommendation [26], user-oriented fairness [38], and fairness-aware IR evaluation. He has served as PC member/reviewer in top computer science conferences or journals such as SIGIR, IJCAI, AAAI, RecSys and ACM TOIS.

Yongfeng Zhang[3] is an Assistant Professor in the Department of Computer Science at Rutgers University (The State University of New Jersey). His research interest is in Information Retrieval, Economics of Data Science, Explainable AI, Fairness in AI, and AI Ethics. In the previous he was a postdoc advised by Prof. W. Bruce Croft in the Center for Intelligent Information Retrieval (CIIR) at UMass Amherst, and did his PhD and BE in Computer Science at Tsinghua University, with a BS in Economics at Peking University. He is a Siebel Scholar of the class 2015, and a Baidu Scholar of the class 2014. Together with coauthors, he has been consistently working on explainable search and recommendation models [6, 15–18, 53, 66–73], fairness-aware recommendation [23, 26, 38, 40, 50], echo chambers in IR systems [29], as well as causal/counterfactual models for information retrieval [55, 56]. His recent research on fairness in recommendation include long-term fairness [26], user-oriented fairness [38], group fairness [40], explainable fairness [23], Pareto fairness [50] and fairness/diversity in echo chambers [29]. He has served as PC members or senior PC members in various Web&IR related conferences such as SIGIR, WWW, CIKM, WSDM, ICTIR and CHIIR, and he is serving as the associate editor for ACM Transactions on Information Systems (TOIS). He has presented the WWW'19/SIGIR'19/ICTIR'19 Tutorial on Explainable Recommendation and Search, and the RecSys'20/WSDM'21 Tutorial on Conversational Recommendation.

6 AVAILABILITY OF MATERIALS

The tutorial materials such as the slides and video recordings are publicly available on the internet[4].

REFERENCES

[1] Himan Abdollahpouri and Robin Burke. 2019. Multi-stakeholder Recommendation and its Connection to Multi-sided Fairness. arXiv:cs.IR/1907.13158

[2] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2017. Controlling popularity bias in learning-to-rank recommendation. In *RecSys*. 42–46.

[3] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2019. Managing popularity bias in recommender systems with personalized re-ranking. *arXiv preprint arXiv:1901.07555* (2019).

[4] Gediminas Adomavicius and YoungOk Kwon. 2011. Improving aggregate recommendation diversity using ranking-based techniques. *TKDE* (2011), 896–911.

[5] A. Agarwal, A. Beygelzimer, M. Dudík, J. Langford, and H. Wallach. 2018. A reductions approach to fair classification. *arXiv preprint arXiv:1803.02453* (2018).

[6] Q. Ai, V. Azizi, X. Chen, and Y. Zhang. 2018. Learning Heterogenous Knowledge base Embeddings for Explainable Recommendation. *Algorithms* (2018).

[1] https://yunqi-li.github.io/

[2] https://yingqiangge.github.io/

[3] http://yongfeng.me

[4] http://fairness-tutorial.github.io/

[7] Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Li Wei, Yi Wu, Lukasz Heldt, Zhe Zhao, Lichan Hong, Ed H Chi, et al. 2019. Fairness in recommendation ranking through pairwise comparisons. In *SIGKDD*. 2212–2220.

[8] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. In *SIGIR*. 405–414.

[9] Robin Burke. 2017. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093* (2017).

[10] Robin Burke, Nasim Sonboli, and Aldo Ordonez-Gauger. 2018. Balanced Neighborhoods for Multi-sided Fairness in Recommendation. In *FAT*. 202–214.

[11] Toon Calders, Faisal Kamiran, and Mykola Pechenizkiy. 2009. Building classifiers with independency constraints. In *2009 ICDM Workshops*. IEEE, 13–18.

[12] L Elisa Celis, Sayash Kapoor, Farnood Salehi, and Nisheeth Vishnoi. 2019. Controlling Polarization in Personalization: An Algorithmic Framework. In *FAT*.

[13] L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. 2018. Ranking with Fairness Constraints. In *ICALP*. https://doi.org/10.4230/LIPIcs.ICALP.2018.28

[14] Hanxiong Chen, Shaoyun Shi, Yunqi Li, and Yongfeng Zhang. 2021. Neural Collaborative Reasoning. In *WWW*.

[15] Xu Chen, Hanxiong Chen, Hongteng Xu, Yongfeng Zhang, Yixin Cao, Zheng Qin, and Hongyuan Zha. 2019. Personalized fashion recommendation with visual explanations based on multimodal attention network: Towards visually explainable recommendation. In *SIGIR*. 765–774.

[16] Xu Chen, Zheng Qin, Yongfeng Zhang, and Tao Xu. 2016. Learning to rank features for recommendation over multiple categories. In *SIGIR*. ACM, 305–314.

[17] Xu Chen, Hongteng Xu, Yongfeng Zhang, Jiaxi Tang, Yixin Cao, Hongyuan Zha, and Zheng Qin. 2018. Sequential Recommendation with User Memory Networks. In *WSDM*. ACM.

[18] Xu Chen, Yongfeng Zhang, and Zheng Qin. 2019. Dynamic Explainable Recommendation based on Neural Attentive Models. *AAAI* (2019).

[19] Andrew Cotter, Heinrich Jiang, and Karthik Sridharan. 2019. Two-player games for efficient non-convex constrained optimization. In *ALT*.

[20] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *ITCS*. 214–226.

[21] Michael D Ekstrand, Robin Burke, and Fernando Diaz. 2019. Fairness and discrimination in recommendation and retrieval. In *RecSys*. 576–577.

[22] Michael D Ekstrand, Robin Burke, and Fernando Diaz. 2019. Fairness and discrimination in retrieval and recommendation. In *SIGIR*. 1403–1404.

[23] Zuohui Fu, Yikun Xian, Ruoyuan Gao, Jieyu Zhao, Qiaoying Huang, Yingqiang Ge, Shuyuan Xu, Shijie Geng, Chirag Shah, Yongfeng Zhang, et al. 2020. Fairness-Aware Explainable Recommendation over Knowledge Graphs. *SIGIR* (2020).

[24] Ruoyuan Gao and Chirag Shah. 2019. How Fair Can We Go: Detecting the Boundaries of Fairness Optimization in Information Retrieval. In *ICTIR*. 229–236.

[25] Ruoyuan Gao and Chirag Shah. 2020. Counteracting Bias and Increasing Fairness in Search and Recommender Systems. In *RecSys*. 745–747.

[26] Yingqiang Ge, Shuchang Liu, Ruoyuan Gao, Yikun Xian, Yunqi Li, Xiangyu Zhao, Changhua Pei, Fei Sun, Junfeng Ge, Wenwu Ou, and Yongfeng Zhang. 2021. Towards Long-term Fairness in Recommendation. *WSDM* (2021).

[27] Yingqiang Ge, Shuyuan Xu, Shuchang Liu, Zuohui Fu, Fei Sun, and Yongfeng Zhang. 2020. Learning Personalized Risk Preferences for Recommendation. In *SIGIR*. 409–418.

[28] Yingqiang Ge, Shuyuan Xu, Shuchang Liu, Shijie Geng, Zuohui Fu, and Yongfeng Zhang. 2019. Maximizing marginal utility per dollar for economic recommendation. In *WWW*. 2757–2763.

[29] Yingqiang Ge, Shuya Zhao, Honglu Zhou, Changhua Pei, Fei Sun, Wenwu Ou, and Yongfeng Zhang. 2020. Understanding echo chambers in e-commerce recommender systems. In *SIGIR*. 2261–2270.

[30] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search. In *SIGKDD (KDD '19)*. ACM, 2221–2231.

[31] Gabriel Goh, Andrew Cotter, Maya Gupta, and Michael P Friedlander. 2016. Satisfying real-world goals with dataset constraints. In *NeurIPS*. 2415–2423.

[32] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of opportunity in supervised learning. In *NeurIPS*. 3315–3323.

[33] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. 2014. Correcting Popularity Bias by Enhancing Recommendation Neutrality. In *RecSys*.

[34] A. Khademi, S. Lee, D. Foley, and V. Honavar. 2019. Fairness in algorithmic decision making: An excursion through the lens of causality. In *WWW*.

[35] Caitlin Kuhlman, MaryAnn VanValkenburg, and Elke Rundensteiner. 2019. FARE: Diagnostics for Fair Ranking Using Pairwise Error Metrics. In *WWW*. 2936–2942.

[36] Matt Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual fairness. In *NeurIPS*. 4069–4079.

[37] Jurek Leonhardt, Avishek Anand, and Megha Khosla. 2018. User Fairness in Recommender Systems. In *Companion Proceedings of the The Web Conf.* 101–102.

[38] Yunqi Li, Hanxiong Chen, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2021. User-oriented Fairness in Recommendation. *WWW* (2021).

[39] Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, and Yongfeng Zhang. 2021. Towards Personalized Fairness based on Causal Notion. *SIGIR* (2021).

[40] Xiao Lin, M. Zhang, Y. Zhang, Z. Gu, Y. Liu, and S. Ma. 2017. Fairness-aware group recommendation with pareto-efficiency. In *RecSys*. 107–115.

[41] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a Fair Marketplace: Counterfactual Evaluation of the Trade-off Between Relevance, Fairness & Satisfaction in Recommendation Systems. In *CIKM*.

[42] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. 2020. Controlling Fairness and Bias in Dynamic Learning-to-Rank. In *SIGIR*.

[43] Harikrishna Narasimhan. 2018. Learning with complex loss functions and constraints. In *International Conference on Artificial Intelligence and Statistics*.

[44] Harikrishna Narasimhan, Andrew Cotter, Maya R Gupta, and Serena Wang. 2020. Pairwise Fairness for Ranking and Regression.. In *AAAI*. 5248–5255.

[45] Gourab K Patro, Arpita Biswas, Niloy Ganguly, Krishna P Gummadi, and Abhijnan Chakraborty. 2020. FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms. In *WWW*. 1194–1204.

[46] Judea Pearl. 2009. *Causality*. Cambridge university press.

[47] Dino Pedreschi, Salvatore Ruggieri, and Franco Turini. 2009. Measuring discrimination in socially-sensitive decision records. In *SIAM*. 581–592.

[48] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of exposure in rankings. In *SIGKDD*. 2219–2228.

[49] Ashudeep Singh and Thorsten Joachims. 2019. Policy learning for fairness in ranking. In *NeurIPS*. 5426–5436.

[50] Guang Wang, Yongfeng Zhang, Zhihan Fang, Shuai Wang, Fan Zhang, and Desheng Zhang. 2020. FairCharge: A data-driven fairness-aware charging recommendation system for large-scale electric taxi fleets. *IMWUT* (2020), 1–25.

[51] B. Woodworth, S. Gunasekar, M. I. Ohannessian, and N. Srebro. 2017. Learning non-discriminatory predictors. *arXiv preprint arXiv:1702.06081* (2017).

[52] Y. Wu, L. Zhang, X. Wu, and H. Tong. 2019. Pc-fairness: A unified framework for measuring causality-based fairness. *arXiv preprint arXiv:1910.12586* (2019).

[53] Yikun Xian, Zuohui Fu, S Muthukrishnan, Gerard De Melo, and Yongfeng Zhang. 2019. Reinforcement knowledge graph reasoning for explainable recommendation. In *SIGIR*. 285–294.

[54] Yikun Xian, Zuohui Fu, Handong Zhao, Yingqiang Ge, Xu Chen, Qiaoying Huang, Shijie Geng, Zhou Qin, Gerard de Melo, S. Muthukrishnan, and Yongfeng Zhang. 2020. CAFE: Coarse-to-Fine Neural Symbolic Reasoning for Explainable Recommendation. *CIKM* (2020).

[55] Shuyuan Xu, Yingqiang Ge, Yunqi Li, Zuohui Fu, Xu Chen, and Yongfeng Zhang. 2021. Causal Collaborative Filtering. *arXiv preprint arXiv:2102.01868* (2021).

[56] Shuyuan Xu, Yunqi Li, Shuchang Liu, Zuohui Fu, and Yongfeng Zhang. 2020. Learning Post-Hoc Causal Explanations for Recommendation. *arXiv preprint arXiv:2006.16977* (2020).

[57] Y. Saito, S. Yaginuma, Y. Nishino, H. Sakata, K. Nakata. 2020. Unbiased Recommender Learning from Missing-Not-At-Random Implicit Feedback. In *WSDM*.

[58] Ke Yang and Julia Stoyanovich. 2017. Measuring Fairness in Ranked Outputs. In *SSDBM*. ACM, 22:1—-22:6. https://doi.org/10.1145/3085504.3085526

[59] Sirui Yao and Bert Huang. 2017. New fairness metrics for recommendation that embrace differences. *arXiv preprint arXiv:1706.09838* (2017).

[60] 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*. 1171–1180.

[61] M. Zehlike, F. Bonchi, C. Castillo, S. Hajian, M. Megahed, and R. Baeza-Yates. 2017. FA*IR: A Fair Top-k Ranking Algorithm. In *CIKM*.

[62] Meike Zehlike and Carlos Castillo. 2020. Reducing disparate exposure in ranking: A learning to rank approach. In *WWW*. 2849–2855.

[63] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. 2013. Learning fair representations. In *ICML*. 325–333.

[64] Junzhe Zhang and Elias Bareinboim. 2018. Equality of opportunity in classification: A causal approach. In *NeurIPS*. 3675–3685.

[65] Junzhe Zhang and Elias Bareinboim. 2018. Fairness in decision-making—the causal explanation formula. In *AAAI*, Vol. 32.

[66] Yongfeng Zhang. 2015. Incorporating phrase-level sentiment analysis on textual reviews for personalized recommendation. In *WSDM*. ACM, 435–440.

[67] Yongfeng Zhang. 2017. Explainable Recommendation: Theory and Applications. *arXiv preprint arXiv:1708.06409* (2017).

[68] Yongfeng Zhang and Xu Chen. 2020. Explainable Recommendation: A Survey and New Perspectives. *Foundations and Trends in Information Retrieval* (2020).

[69] Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. 2014. Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis. *SIGIR* (2014), 83–92.

[70] Yongfeng Zhang, Haochen Zhang, Min Zhang, Yiqun Liu, et al. 2014. Do Users Rate or Review? Boost Phrase-level Sentiment Labeling with Review-level Sentiment Classification. *SIGIR* (2014), 1027–1030.

[71] Yongfeng Zhang, Min Zhang, Yi Zhang, Guokun Lai, Yiqun Liu, Honghui Zhang, and Shaoping Ma. 2015. Daily-aware personalized recommendation based on feature-level time series analysis. In *WWW*. 1373–1383.

[72] Yongfeng Zhang, Yi Zhang, and Min Zhang. 2018. Report on EARS'18: 1st International Workshop on ExplainAble Recommendation and Search. *SIGIR Forum* (2018).

[73] Yongfeng Zhang, Yi Zhang, and Min Zhang. 2018. SIGIR 2018 Workshop on ExplainAble Recommendation and Search (EARS 2018). *SIGIR* (2018).