Recommendation Fairness: From Static to Dynamic

Dell Zhang†‡
dell.z@ieee.org
ByteDance AI Lab
London, UK

Jun Wang∗
jun.wang@cs.ucl.ac.uk
University College London
London, UK

ABSTRACT
Driven by the need to capture users’ evolving interests and optimize their long-term experiences, more and more recommender systems have started to model recommendation as a Markov decision process and employ reinforcement learning to address the problem. Shouldn’t research on the fairness of recommender systems follow the same trend from static evaluation and one-shot intervention to dynamic monitoring and non-stop control? In this paper, we portray the recent developments in recommender systems first and then discuss how fairness could be baked into the reinforcement learning techniques for recommendation. Moreover, we argue in order to make further progress in recommendation fairness, we may want to consider multi-agent (game-theoretic) optimization, multi-objective (Pareto) optimization, and simulation-based optimization, in the general framework of stochastic games.

CCS CONCEPTS
• Information systems → Recommender systems; • Computing methodologies → Reinforcement learning.

KEYWORDS
recommender systems, reinforcement learning, machine learning fairness

1 INTRODUCTION
Fairness in recommendation is one of the most important aspects of evaluating and providing socially responsible recommender systems. In this short position paper, we briefly review the research about recommendation fairness over the last five years and present our opinions about where this area should go next. The central idea is that the paradigm of recommendation fairness will be shifting from static evaluation and one-shot intervention to dynamic monitoring and non-stop control.

2 THE RISE OF REINFORCEMENT LEARNING FOR RECOMMENDATION
The mainstream approach to building recommender systems is to formulate recommendation as matrix completion [39], i.e., given a matrix of users by items where the value at cell \((i, j)\) represents the user’s rating of the item, predict the missing cell values based on the existing cell values. Such a matrix completion problem is often solved by matrix factorization [45, 69] algorithms. Recently, the nonlinear neural variants of matrix factorization [15, 34, 35] for recommendation have attracted a lot of attention, though some researchers have their reservations [56, 70].

Today, modern recommender systems underpinning various web or mobile apps are expected to become more personalized and interactive so as to better serve the users. Consequently, traditional recommendation techniques based on matrix completion which assume users’ preferences being static and aim to maximize their immediate satisfaction would no longer work well.

Reinforcement learning (RL) [80] — an area of machine learning which is concerned with optimal decision making over time in a dynamic environment — offers a promising approach to tackling the problems of personalization and interactivity by capturing users’ evolving interests and optimizing their long-term experiences [99]. Inspired by the great successes of reinforcement learning, particularly when it is combined with deep learning [33, 48] such as in AlphaGo [74], reinforcement learning based recommender systems [46, 95, 96] have just started to gain popularity in the last couple of years.

2.1 Recommendation as an MDP
To apply reinforcement learning to recommender systems, we need to model recommendation as a Markov decision process (MDP) [80]. For example, in a video recommender system [11, 57], the state space \(S\) describes the users each accompanied with their contextual status (e.g., the time when the recommendation is made and the query text entered by the user), and the action space \(A\) consists of all possible video items available for recommendation. The state representing each user evolves as the user interacts with the recommender system; different users will have different states. Using reinforcement learning, we seek a policy \(\pi(a|s)\) which returns a distribution of video items \(a \in A\) for each given user state \(s \in S\). The objective is that the learned policy \(\pi\) can maximize the expected discounted cumulative reward over potentially infinite time horizon, where the immediate reward for taking action \(a\) at state \(s\) (i.e., recommending item \(a\) to user \(s\)) is defined by a reward function \(R : S \times A \rightarrow \mathbb{R}\).

2.2 Reinforcement Learning Algorithms in Recommender Systems
Many different reinforcement learning algorithms have been employed by a variety of recommender systems in recent years. The simplest ones [5, 32] utilize contextual (multi-armed) bandits [47, 77] which solve the special case of one-step reinforcement learning; some [100–102] use value-based methods such as the Deep Q

\[ \text{ABSTRACT} \]

Driven by the need to capture users’ evolving interests and optimize their long-term experiences, more and more recommender systems have started to model recommendation as a Markov decision process and employ reinforcement learning to address the problem. Shouldn’t research on the fairness of recommender systems follow the same trend from static evaluation and one-shot intervention to dynamic monitoring and non-stop control? In this paper, we portray the recent developments in recommender systems first and then discuss how fairness could be baked into the reinforcement learning techniques for recommendation. Moreover, we argue in order to make further progress in recommendation fairness, we may want to consider multi-agent (game-theoretic) optimization, multi-objective (Pareto) optimization, and simulation-based optimization, in the general framework of stochastic games.

\[ \text{CCS CONCEPTS} \]

• Information systems → Recommender systems; • Computing methodologies → Reinforcement learning.

\[ \text{KEYWORDS} \]

recommender systems, reinforcement learning, machine learning fairness

\[ \text{1 INTRODUCTION} \]

Fairness in recommendation is one of the most important aspects of evaluating and providing socially responsible recommender systems. In this short position paper, we briefly review the research about recommendation fairness over the last five years and present our opinions about where this area should go next. The central idea is that the paradigm of recommendation fairness will be shifting from static evaluation and one-shot intervention to dynamic monitoring and non-stop control.

\[ \text{2 THE RISE OF REINFORCEMENT LEARNING FOR RECOMMENDATION} \]

The mainstream approach to building recommender systems is to formulate recommendation as matrix completion [39], i.e., given a matrix of users by items where the value at cell \((i, j)\) represents the user’s rating of the item, predict the missing cell values based on the existing cell values. Such a matrix completion problem is often solved by matrix factorization [45, 69] algorithms. Recently, the nonlinear neural variants of matrix factorization [15, 34, 35] for recommendation have attracted a lot of attention, though some researchers have their reservations [56, 70].

Today, modern recommender systems underpinning various web or mobile apps are expected to become more personalized and interactive so as to better serve the users. Consequently, traditional recommendation techniques based on matrix completion which assume users’ preferences being static and aim to maximize their immediate satisfaction would no longer work well.

Reinforcement learning (RL) [80] — an area of machine learning which is concerned with optimal decision making over time in a dynamic environment — offers a promising approach to tackling the problems of personalization and interactivity by capturing users’ evolving interests and optimizing their long-term experiences [99]. Inspired by the great successes of reinforcement learning, particularly when it is combined with deep learning [33, 48] such as in AlphaGo [74], reinforcement learning based recommender systems [46, 95, 96] have just started to gain popularity in the last couple of years.

\[ \text{2.1 Recommendation as an MDP} \]

To apply reinforcement learning to recommender systems, we need to model recommendation as a Markov decision process (MDP) [80]. For example, in a video recommender system [11, 57], the state space \(S\) describes the users each accompanied with their contextual status (e.g., the time when the recommendation is made and the query text entered by the user), and the action space \(A\) consists of all possible video items available for recommendation. The state representing each user evolves as the user interacts with the recommender system; different users will have different states. Using reinforcement learning, we seek a policy \(\pi(a|s)\) which returns a distribution of video items \(a \in A\) for each given user state \(s \in S\). The objective is that the learned policy \(\pi\) can maximize the expected discounted cumulative reward over potentially infinite time horizon, where the immediate reward for taking action \(a\) at state \(s\) (i.e., recommending item \(a\) to user \(s\)) is defined by a reward function \(R : S \times A \rightarrow \mathbb{R}\).

\[ \text{2.2 Reinforcement Learning Algorithms in Recommender Systems} \]

Many different reinforcement learning algorithms have been employed by a variety of recommender systems in recent years. The simplest ones [5, 32] utilize contextual (multi-armed) bandits [47, 77] which solve the special case of one-step reinforcement learning; some [100–102] use value-based methods such as the Deep Q
Network (DQN) [36, 63]; some [11, 57, 81] use policy-based methods such as the policy gradient algorithm REINFORCE [80, 84]; and some [12, 88] adapt actor-critic methods [44, 75]. A hot research topic is to develop offline reinforcement learning [42, 49, 90] methods for interactive recommendation [87, 105] which can make effective use of previously collected user-item interaction data without expensive online data collection.

3 THE EVOLUTION OF RECOMMENDATION FAIRNESS

Significant progress has been made in the space of algorithmic fairness [59, 71] for recommender systems [18, 28, 52], as reflected by the FAccTRec workshop series since it first took place in 2017 [20, 21, 41].

3.1 Static Recommendation Fairness

It has been discovered and reported that various types of bias could exist in recommender systems, such as those with respect to user demographics (gender, age, and race, etc.) [30, 92, 103], user activity [26], and item popularity [1–3, 98]. Hence, many different metrics of fairness [9] have been proposed in order to build fairness-aware recommender systems. Such recommendation fairness metrics can be defined at two levels: individual fairness and group fairness. Generally speaking, the techniques for counteracting bias and promoting fairness in recommendation so far are largely in the form of constrained optimization [93]: either maximizing utility (which is for the most part the relevance of recommended items to users) subject to a set of fairness constraints [10, 30], or maximizing fairness subject to a lower bound of utility [103], or jointly optimizing both for an overall satisfaction [27]. Those works mostly make fairness adjustments to traditional matrix completion based recommender systems. Their concept of recommendation fairness is static in the sense that the protected groups are fixed during the recommendation process.

3.2 Dynamic Recommendation Fairness

Little research has been conducted to investigate the fairness of reinforcement learning based recommender systems where the protected groups may change over time. In dynamic environments where the distribution of population is shifting or the decisions being made have feedback effects, counter-intuitive phenomena (like the Simpson’s paradox) could occur and biases could be iteratively amplified. For example, imposing static fairness criteria myopically at every step may actually exacerbate unfairness [16, 17, 54, 65, 83, 97].

Note that although there exist a few papers talking about fairness in reinforcement learning in general, not all those fairness definitions are related to the fair treatment of different users or items (grouped by sensitive protected attributes) in recommendation. For example, the “meritocratic” fairness of reinforcement learning [38] does not seem to be very relevant to our fairness concern for recommender systems.

Liu et al. [55] have proposed a fairness-aware recommendation framework based on reinforcement learning to dynamically balance recommendation accuracy and user fairness in the long run. The constantly changing user preferences and fairness statuses are jointly represented as the states in the MDP model of recommendation. Furthermore, a two-fold reward function is designed to combine accuracy and fairness. Also on the user-side of fairness but not particularly for recommender systems, Wen et al. [82] studied reinforcement learning under group fairness constraints. They show how fairness constraints from the supervised learning setting such as demographic parity and equality of opportunity can be extended to the MDP setting. The algorithms developed by them to solve the MDP problem can ensure that the learned policy does not favor the majority sub-population over the minority sub-population. Zhang et al. [97] investigated specifically the dynamics of group qualification rates [65, 83] under the more general partially-observed MDP setting. Moreover, there also exist some studies of fairness-aware (contextual) multi-armed bandits [5, 31, 32, 47, 77] — a simplified form of reinforcement learning — where the fairness constraint is defined as a minimum rate at which a task/resource is assigned to a user [13, 14, 51, 66, 89].

Ge et al. [29] have made an attempt on dynamic recommendation fairness of not users but items. Their work focuses on the fairness to different groups of items divided according to their degrees of popularity which dynamically change during the recommendation process: popular items could become unpopular after a while and vice versa. To achieve long-term fairness in terms of item exposure, the MDP model of recommendation is augmented with a set of fairness constraints each of which is an auxiliary fairness cost function bounded by the corresponding limit. Such a constrained MDP problem can be solved by performing constrained policy optimization with an actor-critic architecture.

When a recommender system starts to utilize reinforcement learning to optimize its users’ long-term engagement, there will be a risk of the unethical phenomenon “user tampering” [8, 24] whereby the recommender system tries to actively manipulate its users’ preferences via its recommendations in order to gain maximum accumulated reward. For example, a news recommender system may be tempted to (politically) polarize its users with the early recommendations so that the users will become more engaged with the system’s later recommendations catering to such polarization. Obviously, fairness issues will arise if different users are affected by user tampering differently. How to ensure dynamic recommendation fairness while addressing user tampering is still an open question.

3.3 Looking Ahead

3.3.1 Multi-Agent (Game-Theoretic) Optimization. Since recommender systems are by their nature multi-sided platforms or marketplaces [64, 76] involving at least the consumers (customers) of items as well as the producers (providers) of items [6, 79], there have been some works on optimizing the static fairness for all stakeholders of the recommender system [7, 58, 60, 67, 78]. The usual approach to multi-sided static fairness is to use a linear interpolation of all stakeholders’ fairness metrics [67, 78, 85] as the optimization objective or constraint. That is probably not sophisticated enough to handle the intricacies of dynamic fairness in reinforcement learning.

We think that the principled approach to multi-sided dynamic fairness for recommender systems is to consider it as a multi-agent
Recommendation Fairness: From Static to Dynamic

reinforcement learning (MARL) [91] problem, as in [40, 94, 104]. This implies adopting the mathematical framework of stochastic games [50, 53, 72] which generalize MDPs to multiple interacting decision makers. Note that to maximize the total social welfare while ensuring each agent (player) get a fair share of opportunities, we may want to go beyond Nash equilibrium and embrace the more general correlated equilibrium which is also computationally more efficient. The advantages of correlated equilibrium over Nash equilibrium in terms of fairness could be illustrated by the classic “battle of the sexes” game [50, 72]. Furthermore, behavior game theory may be useful for analyzing multi-sided dynamic fairness in real-life recommender systems. As revealed by the observation of human playing the “ultimatum bargaining” game, people have the tendency to pursue fairness even when it contradicts with the subgame perfect equilibrium that maximizes their monetary payoffs [50, 72].

3.3.2 Multi-Objective (Pareto) Optimization. It is probably also time to leave the constrained optimization [27, 93] approach to recommendation fairness behind and seek the Pareto optimization [12] of multiple objectives including the utility and the fairness. Even when fairness is the only goal of our concern, researchers have rigorously proved that there are inherent conflicts among some common fairness metrics and it is often impossible to optimize them simultaneously [25, 43]. A few early works have emerged [73, 86] and more should follow the steps.

3.3.3 Simulation Environment. To fully comprehend and tackle the complexities of dynamic fairness in recommendation, it is highly desirable to develop a simulation environment for such multi-agent multi-objective recommender systems where a number of fairness metrics could be continuously monitored and optimized. Although Google has released Fairness-Gym [17] for the simulation of simple dynamic fairness tasks (loan application [54], college admission [37, 61], and attention allocation [22, 23]), a simulator dedicated to dynamic fairness in recommendation is not available, yet. This would require more fundamental research on understanding user behavior and building user models in recommender systems [4, 19, 68]. It also looks promising to incorporate fairness metrics and models into some newly emerging probabilistic simulators of multi-agent recommender systems such as RecSimNG [62].

4 CONCLUSION

The recent developments in the area of recommendation fairness exhibits a clear trend towards the dynamic view of fairness. Accordingly, the underlying mathematical framework of fair recommendation will probably move from matrix completion to Markov decision process and then to stochastic games. Such new models and algorithms for fairness may not only improve different kinds of recommender systems but also have impacts upon the broader field of Responsible AI.

ACKNOWLEDGMENTS

We are grateful to Dr Hang Li, Director of AI Lab at ByteDance, for his guidance on and support for this line of research. We thank the workshop organizers for providing this forum and the reviewers for their constructive feedback.

REFERENCES

[1] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2017. Controlling Popularity Bias in Learning-to-Rank Recommendation. In Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17). Association for Computing Machinery, New York, NY, USA, 42–46. https://doi.org/10.1145/3109859.3109912
[2] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher. 2019. The Unfairness of Popularity Bias in Recommendation. arXiv:1907.13286 [cs] (Sept. 2019). arXiv:1907.13286 [cs]
[3] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher. 2020. The Connection Between Popularity Bias, Calibration, and Fairness in Recommendation. In Fourteenth ACM Conference on Recommender Systems (RecSys '20) Association for Computing Machinery, New York, NY, USA, 726–731. https://doi.org/10.1145/3383313.3418487
[4] Krzysztof Balog, Filip Radlinski, and Shushan Arakelyan. 2019. Transparent, Scalable and Explainable User Models for Personalized Recommendation. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19). Association for Computing Machinery, New York, NY, USA, 265–274. https://doi.org/10.1145/3331184.3331211
[5] Andrea Barraza-Urba and Dorota Glowacka. 2020. Introduction to Bandits in Recommender Systems. In Fourteenth ACM Conference on Recommender Systems (RecSys '20). Association for Computing Machinery, New York, NY, USA, 748–750. https://doi.org/10.1145/3383313.3411547
[6] Robin Burke, Himan Abdollahpouri, Edward M. Malthouse, KP Thai, and Yongfeng Zhang. 2019. Recommendation in Multistakeholder Environments. In Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19). Association for Computing Machinery, New York, NY, USA, 566–567. https://doi.org/10.1145/3292868.3349793
[7] Robin Burke, Nasim Sonboli, and Aldo Ordoñez-Gauger. 2018. Balanced Neighborhoods for Multi-Sided Fairness in Recommendation. In Conference on Fairness, Accountability and Transparency, PMLR, 202–214.
[8] Micah Carroll, Dylan Hadfield-Menell, Stuart Russell, and Anca Dragan. 2021. Estimating and Penalizing Preference Shift in Recommender Systems. In Fifteenth ACM Conference on Recommender Systems (RecSys '21). Association for Computing Machinery, New York, NY, USA, 661–667. https://doi.org/10.1145/3469231.3478849
[9] Alessandro Castelvetro, Riccardo Curpi, Grete Greco, and Daniele Regoli. 2021. The Zoo of Fairness Metrics in Machine Learning. arXiv:2106.00467 [cs, stat] (June 2021). arXiv:2106.00467 [cs, stat]
[10] L. Elisa Celis, Sayash Kapoor, Farnood Salehi, and Nisheeth Vishnoi. 2019. Controlling Polarization in Personalization: An Algorithmic Framework. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAIT’19). Association for Computing Machinery, New York, NY, USA, 149–169. https://doi.org/10.1145/3287560.3287601
[11] Minmin Chen, Alex Beutel, Paul Covington, Sagar Jain, Francois Belletti, and Ed H. Chi. 2019. Top-K Off-Policy Correction for a REINFORCE Recommender System. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM ’19). Association for Computing Machinery, New York, NY, USA, 456–464. https://doi.org/10.1145/3289630.3290999
[12] Xu Chen, Yali Du, Long Xiang, and Jun Wang. 2021. Reinforcement Recommendation with User Multi-Aspect Preference. In Proceedings of the Web Conference 2021 (WWW ’21). Association for Computing Machinery, New York, NY, USA, 425–435. https://doi.org/10.1145/3442581.3449846
[13] Yifang Chen, Alex Cuelar, Haipeng Luo, Jignesh Modi, Heranz-Nemlekar, and Stefano Nikolaidis. 2020. Fair Contextual Multi-Armed Bandits: Theory and Experiments. In Conference on Uncertainty in Artificial Intelligence. PMLR, 181–190.
[14] Houston Clauer, Yifang Chen, Jignesh Modi, Malte Jung, and Stefano Nikolaidis. 2020. Multi-Armed Bandits with Fairness Constraints for Distributing Resources to Human Teammates. arXiv:1907.00313 [cs] (Dec. 2020). arXiv:1907.00313 [cs]
[15] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In Proceedings of the 10th ACM Conference on Recommender Systems. ACM, Boston Massachusetts USA, 191–198. https://doi.org/10.1145/2959100.2959190
[16] Elliot Creager, David Madras, Toniann Pitassi, and Richard Zemel. 2020. Causal Modeling for Fairness in Dynamical Systems. In International Conference on Machine Learning, PMLR, 2185–2195.
[17] Alexander D’Amour, Hansa Srinivasan, James Atwood, Pallavi Baljekar, D. Sculley, and Yoni Halpern. 2020. Fairness in Machine Learning is Not Static: Deeper Understanding of Long Term Fairness via Simulation Studies. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAIT ’20). Association for Computing Machinery, New York, NY, USA, 525–534. https://doi.org/10.1145/3351095.3372878
[96] Weinan Zhang, Xiangyu Zhao, Li Zhao, Dawei Yin, Grace Hui Yang, and Alex Beutel. 2020. Deep Reinforcement Learning for Information Retrieval: Fundamentals and Advances. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20) Association for Computing Machinery, New York, NY, USA, 2468–2471. https://doi.org/10.1145/3397271.3401467

[97] Xueru Zhang, Ruibo Tu, Yang Liu, Mingyan Liu, Hedvig Kjellstrom, Kun Zhang, and Cheng Zhang. 2020. How Do Fair Decisions Fare in Long-Term Qualification?. In Advances in Neural Information Processing Systems, Vol. 33. Curran Associates, Inc., 18457–18469.

[98] Yang Zhang, Fuli Feng, Xiangnan He, Tianxin Wei, Chonggang Song, Guohui Ling, and Yongdong Zhang. 2021. Causal Intervention for Leveraging Popularity Bias in Recommendation. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21). Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/3404835.3406287

[99] Xiangyu Zhao, Changsheng Gu, Haoshenglun Zhang, Xiaobing Liu, Xiwang Yang, and Jiliang Tang. 2019. Deep Reinforcement Learning for Online Advertising in Recommender Systems. arXiv:1909.03662 [cs] (Sept. 2019). arXiv:1909.03662 [cs]

[100] Xiangyu Zhao, Liang Zhang, Zhuoye Ding, Long Xia, Jiliang Tang, and Dawei Yin. 2018. Recommendations with Negative Feedback via Pairwise Deep Reinforcement Learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18). Association for Computing Machinery, New York, NY, USA, 1040–1048. https://doi.org/10.1145/3219819.3219886

[101] Xiangyu Zhao, Xudong Zheng, Xiwang Yang, Xiaobing Liu, and Jiliang Tang. 2020. Jointly Learning to Recommend and Advertise. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20). Association for Computing Machinery, New York, NY, USA, 3319–3327. https://doi.org/10.1145/3394486.3403384

[102] Guanjie Zheng, Fuzheng Zhang, Zihan Zheng, Yang Xiang, Nicholas Jing Yuan, Xing Xie, and Zhenhui Li. 2018. DRN: A Deep Reinforcement Learning Framework for News Recommendation. In Proceedings of the 2018 World Wide Web Conference (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 167–176. https://doi.org/10.1145/3178876.3185994

[103] Ziwei Zhu, Xia Hu, and James Caverlee. 2018. Fairness-Aware Tensor-Based Recommendation. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18). Association for Computing Machinery, New York, NY, USA, 1153–1162. https://doi.org/10.1145/3269206.3271795

[104] Matthieu Zimmer, Claire Glanois, Umer Siddique, and Paul Weng. 2021. Learning Fair Policies in Decentralized Cooperative Multi-Agent Reinforcement Learning. In International Conference on Machine Learning, PMLR, 12967–12978.

[105] Lexin Zou, Long Xia, Pan Du, Zhao Zhang, Ting Bai, Weidong Liu, Jianyun Nie, and Dawei Yin. 2020. Pseudo Dyna-Q: A Reinforcement Learning Framework for Interactive Recommendation. In Proceedings of the 13th International Conference on Web Search and Data Mining (WSDM '20). Association for Computing Machinery, New York, NY, USA, 816–824. https://doi.org/10.1145/3336191.3371801