Trustworthy Recommendation and Search: Introduction to the Special Issue - Part 1

HONGZHI YIN, The University of Queensland, Australia
YIZHOU SUN, University of California, Los Angeles, USA
GUANDONG XU, University of Technology Sydney, Australia
EVANGELOS KANOULAS, University of Amsterdam, Netherlands

CCS Concepts: • **Information systems** → **Retrieval models and ranking**; **Recommender systems;**

Additional Key Words and Phrases: Recommender systems, information retrieval, trustworthiness, robustness, interpretability, fairness, privacy, security

ACM Reference format:
Hongzhi Yin, Yizhou Sun, Guandong Xu, and Evangelos Kanoulas. 2023. Trustworthy Recommendation and Search: Introduction to the Special Issue - Part 1. *ACM Trans. Inf. Syst.* 41, 3, Article 51 (February 2023), 5 pages.
https://doi.org/10.1145/3579995

1 INTRODUCTION

Recommendation and search systems have already become indispensable means for helping web users identify the most relevant information/services in the era of information overload. The applications of such systems are multi-faceted, including targeted advertising, intelligent medical assistant, and e-commerce, and are bringing immense convenience to people's daily lives. However, despite rapid advances in recommendation and search, the increasing public awareness of the trustworthiness of relevant recommendation and search applications has introduced higher expectations on relevant research. Firstly, the unprecedentedly growing heterogeneity of use cases has been challenging the adaptivity of contemporary algorithms to various settings, e.g., dynamic user interests [Chen et al. 2019], highly sparse interaction records [Chen et al. 2020b], and limited computing resources [Imran et al. 2022; Long et al. 2022]. Secondly, in a broader sense, a trustworthy recommendation/search approach should also be robust, interpretable, secure, privacy-preserving, and fair across different use cases. Specifically, robustness evaluates a model's performance consistency under various operating conditions like noisy data [Zhang et al. 2020]; interpretability and fairness respectively evaluate if a model can make its decision processes transparent [Chen et al. 2021, 2020c; Cui et al. 2022; Lyu et al. 2021; Ren et al. 2021] and the decision outcomes unbiased [Chen et al. 2020a; Li et al. 2021; Yin et al. 2012]; while security and privacy respectively

Authors' addresses: H. Yin, The University of Queensland, St. Lucia, Brisbane, QLD, Australia, 4072; email: h.yin1@uq.edu.au; Y. Sun, University of California, Los Angeles, Westwood, Los Angeles, CA, USA, 90095; email: yzsun@cs.ucla.edu; G. Xu, University of Technology Sydney, Ultimo, Sydney, NSW, Australia, 2007; email: guandong.xu@uts.edu.au; E. Kanoulas, University of Amsterdam, Science Park, Amsterdam, XH, Netherlands, 1098; email: e.kanoulas@uva.nl.

1046-8188/2023/02-ART51 $15.00
https://doi.org/10.1145/3579995

emphasize a model's ability to handle cyber-attacks [Zhang et al. 2021b, 2022] and to prevent personal information leakage [Wang et al. 2022b; Yuan et al. 2023; Zhang et al. 2021a; Zhang and Yin 2022; Zhang et al. 2021c]. Consequently, trustworthiness is becoming a key performance indicator for state-of-the-art recommendation and search approaches. In light of these emerging challenges, this special section focuses on novel research in this field with the notion of trustworthiness. The articles presented in this special issue will further promote responsible AI applications, thus better universalizing the advanced techniques to a wider range of the common public.

2 OVERVIEW OF ARTICLES

The submission deadline of the special issue was 15th June, 2022 and we received 41 valid submissions in total. In Part 1 of this issue, we present 10 accepted papers and leave the remaining accepted papers to the upcoming Part 2. This issue covers a variety of topics related to trustworthy recommendation and search, including four papers on fairness/bias issues in recommendation and ranking, three papers on adversarial attacks and/or their countermeasures for recommender systems, two papers on privacy-aware recommendation, and one paper on the interpretability of recommender systems. In what follows, we provide an overview on these accepted articles.

Ensuring fairness and mitigating algorithmic bias are attracting much attention within the recommendation and search community. By providing a comprehensive survey, Wang et al. [2022a] have summarized various fairness challenges in the recommendation context, reviewed key recommendation datasets and measurements in fairness studies, and provided a well-structured taxonomy of fairness-aware recommendation methods. In this survey, an insightful list of suggestions on future research on the recommendation fairness is also provided, which will benefit subsequent studies in this area. Meanwhile, three specific biases are respectively investigated in the other three papers. He et al. [2022] formulate and address the problem of items' confounding features from the causal perspective, where the proposed solution performs intervened inference with do-calculus. In addition, the authors further propose a mixture-of-experts model architecture that models each value of confounding feature with a separate expert module, so as to significantly lower the computational complexity for evaluating do-calculus. The designed solution thus represents a strong synergy between model expressiveness (i.e., accuracy) and efficiency for casual recommendation. In Liu et al. [2022], the authors study the popularity bias in recommender systems, and unlike most existing studies that only consider fairness at either the user side or item side, their proposed solution jointly mitigates the popularity bias for both users and items for recommendation. Besides, the solution is able to dynamically adapt to different input users and items to handle the differences in their popularity bias, thus contributing to an outstanding debiasing efficacy. In the context of **learning-to-rank (LTR)** algorithms, Oosterhuis [2022] presents a novel solution to the long-standing position bias issues via the lens of counterfactual doubly-robust estimation. As the first doubly-robust solution to position bias in LTR, Oosterhuis proposes to counteract the unobservable treatment (i.e., user examination) by using the expected treatment per rank instead of the actual treatment. The designed estimator is more robust and performant than existing approaches based on inverse propensity scoring, and it further provides the most robust theoretical guarantees of all known LTR estimators.

The need for privacy protection is surging in various web services, and it is no exception for recommender systems as they heavily rely on the sensitive user data to facilitate personalized recommendation. On the one hand, from the perspective of service providers, Chen et al. [2022] present a new take on the common practice of recommendation platforms – offering users binary choice on data disclosure. By designing a privacy-aware recommendation framework that gives users fine-grained control over their data, the authors perform comprehensive experiments in a simulated real-world environment to uncover how different privacy levels impact users'

information disclosure willingness and the platform's revenue. An important message from this paper is that privacy mechanisms with finer split granularity and more unrestrained disclosure strategy can bring better results for both consumers and platforms than the "all or nothing" mechanism adopted by most real-world applications. On the other hand, from the perspective of users, Xin et al. [2022] investigate the problem of user behavior leakage in recommender systems. The authors show that the sensitive historical interaction of a user can be inferred from the currently observed system exposure for this user, a.k.a. membership inference. As a remedy, a privacy protection mechanism is proposed to perturb a subset of exposed items. Because the wide access to exposure data put users' interaction history at a high risk of leakage, this paper will open up an important research topic in privacy-aware recommendation.

Another important topic often discussed in conjunction with privacy in recommendation and search is the adversarial attacks and their countermeasures. In the context of hashing retrieval, given that most existing adversarial attack models assume the impractical white-box setting and are inefficient to train, Zhu et al. [2022] propose an efficient black-box attack model against deep cross-modal hashing retrieval. Specifically, the solution proposes a multi-modal knockoff-driven adversarial generation framework to achieve efficient adversarial example generation. This allows the attacker to efficiently generate quality adversarial examples by forward-propagation with only given benign images under the black-box setting. Meanwhile, Nguyen et al. [2022] have studied poisoning attack on state-of-the-art recommenders based on **graph neural networks (GNNs)**. Attacking GNN-based recommenders is more challenging than attacking a plain GNN due to the heterogeneity of network structure and the entanglement between users and items. As such, the authors propose to surrogate a recommendation model, as well as to generate fake users and user-item interactions while preserving the user-item correlations for recommendation accuracy. The proposed solution is also resistant to various protection mechanisms, shedding light on future research on attack-resistant recommender systems. As a possible countermeasure to such attacks, Ye et al. [2022] propose to improve the robustness of GNN-based recommenders by jointly denoising the structure space and perturbing the embedding space. Notably, in the embedding space, an in-distribution perturbation method is designed to simulate adversarial attacks, thus providing a boost in the recommender's robustness to noisy interactions in the training data.

We also have an accepted article focusing on explainable recommendation. Motivated by the fact that exploring users' fine-grained preferences as well as the relationships among those preferences could improve the recommendation performance, Dong et al. [2022] propose a dual preference distribution learning framework to jointly learn a user's general preference to items and the user's specific preference to item attributes. To support interpretability, a preferred attribute profile is summarized for each user, where the explanation for each recommended item can be generated by checking the overlap between its own and the user's preferred attributes.

3 CONCLUSIONS

In summary, these accepted articles are a strong reflection of the depth and breadth of current research on trustworthy recommendation and search. A wide range of research questions have been discussed in Part 1 of this special issue, including how to achieve a sensible privacy-utility trade-off, how to mitigate algorithmic bias, how to understand and prevent possible adversarial attacks, and how to preserve explainability of a complex algorithm. In the meantime, there are still open challenges to be addressed in this research sector, such as a recommendation/retrieval model's awareness to data veracity, capability of performing reasoning when interacting with users, and ability to handle high throughput data streams. In Part 2 of the special issue, we will introduce more work in those areas.

ACKNOWLEDGMENTS

We thank all researchers who submitted their work to the special issue and all reviewers who spent a significant amount of time and effort to help all authors improve their manuscripts with constructive feedback. We also thank Prof. Min Zhang, the Editor-in-Chief of the journal for the guidance and support provided.

REFERENCES

Hongxu Chen, Yicong Li, Xiangguo Sun, Guandong Xu, and Hongzhi Yin. 2021. Temporal meta-path guided explainable recommendation. In *WSDM*. 1056–1064.

Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2020a. Bias and debias in recommender system: A survey and future directions. *arXiv preprint arXiv:2010.03240* (2020).

Tong Chen, Hongzhi Yin, Hongxu Chen, Rui Yan, Quoc Viet Hung Nguyen, and Xue Li. 2019. AIR: Attentional intention-aware recommender systems. In *ICDE*. 304–315.

Tong Chen, Hongzhi Yin, Quoc Viet Hung Nguyen, Wen-Chih Peng, Xue Li, and Xiaofang Zhou. 2020b. Sequence-aware factorization machines for temporal predictive analytics. *ICDE* (2020).

Tong Chen, Hongzhi Yin, Guanhua Ye, Zi Huang, Yang Wang, and Meng Wang. 2020c. Try this instead: Personalized and interpretable substitute recommendation. In *SIGIR*. 891–900.

Ziqian Chen, Fei Sun, Yifan Tang, Haokun Chen, Jinyang Gao, and Bolin Ding. 2022. Studying the impact of data disclosure mechanism in recommender systems via simulation. *ACM Transactions on Information Systems* (2022).

Zhihong Cui, Hongxu Chen, Lizhen Cui, Shijun Liu, Xueyan Liu, Guandong Xu, and Hongzhi Yin. 2022. Reinforced KGs reasoning for explainable sequential recommendation. *World Wide Web Journal* 25, 2 (2022), 631–654.

Xue Dong, Xuemeng Song, Na Zheng, Yinwei Wei, and Zhongzhou Zhao. 2022. Dual preference distribution learning for item recommendation. *ACM Transactions on Information Systems* (2022).

Xiangnan He, Yang Zhang, Fuli Feng, Chonggang Song, Lingling Yi, Guohui Ling, and Yongdong Zhang. 2022. Addressing confounding feature issue for causal recommendation. *ACM Transactions on Information Systems* (2022).

Mubashir Imran, Hongzhi Yin, Tong Chen, Nguyen Quoc Viet Hung, Alexander Zhou, and Kai Zheng. 2022. ReFRS: Resource-efficient federated recommender system for dynamic and diversified user preferences. *ACM Transactions on Information Systems* (2022).

Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, and Yongfeng Zhang. 2021. Towards personalized fairness based on causal notion. In *SIGIR*. 1054–1063.

Zhongzhou Liu, Yuan Fang, and Min Wu. 2022. Mitigating popularity bias for users and items with fairness-centric adaptive recommendation. *ACM Transactions on Information Systems* (2022).

Jing Long, Tong Chen, Nguyen Quoc Viet Hung, and Hongzhi Yin. 2022. Decentralized collaborative learning framework for next POI recommendation. *ACM Transactions on Information Systems* (2022).

Yanzhang Lyu, Hongzhi Yin, Jun Liu, Mengyue Liu, Huan Liu, and Shizhuo Deng. 2021. Reliable recommendation with review-level explanations. In *ICDE*. 1548–1558.

Thanh Toan Nguyen, Khang Nguyen Duc Quach, Thanh Tam Nguyen, Thanh Trung Huynh, Viet Hung Vu, Phi Le Nguyen, Jun Jo, and Quoc Viet Hung Nguyen. 2022. Poisoning GNN-based recommender systems with generative surrogate-based attacks. *ACM Transactions on Information Systems* (2022).

Harrie Oosterhuis. 2022. Doubly-robust estimation for correcting position-bias in click feedback for unbiased learning to rank. *ACM Transactions on Information Systems* (2022).

Xuhui Ren, Hongzhi Yin, Tong Chen, Hao Wang, Zi Huang, and Kai Zheng. 2021. Learning to ask appropriate questions in conversational recommendation. In *SIGIR*. 808–817.

Qinyong Wang, Hongzhi Yin, Tong Chen, Junliang Yu, Alexander Zhou, and Xiangliang Zhang. 2022b. Fast-adapting and privacy-preserving federated recommender system. *The VLDB Journal* 31, 5 (2022), 877–896.

Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2022a. A survey on the fairness of recommender systems. *ACM Transactions on Information Systems* (2022).

Xin Xin, Jiyuan Yang, Hanbing Wang, Jun Ma, Pengjie Ren, Hengliang Luo, Xinlei Shi, Zhumin Chen, and Zhaochun Ren. 2022. On the user behavior leakage from recommender system exposure. *ACM Transactions on Information Systems* (2022).

Haibo Ye, Xinjie Li, Yuan Yao, and Hanghang Tong. 2022. Towards robust neural graph collaborative filtering via structure denoising and embedding perturbation. *ACM Transactions on Information Systems* (2022).

Hongzhi Yin, Bin Cui, Jing Li, Junjie Yao, and Chen Chen. 2012. Challenging the long tail recommendation. *VLDB Endowment* 5, 9 (2012).

Wei Yuan, Hongzhi Yin, Fangzhao Wu, Shijie Zhang, Tieke He, and Hao Wang. 2023. Federated unlearning for on-device recommendation. *WSDM* (2023).

Minxing Zhang, Zhaochun Ren, Zihan Wang, Pengjie Ren, Zhumin Chen, Pengfei Hu, and Yang Zhang. 2021b. Membership inference attacks against recommender systems. *CCS* (2021).

Peng-Fei Zhang, Yang Li, Zi Huang, and Hongzhi Yin. 2021a. Privacy protection in deep multi-modal retrieval. In *SIGIR*. 634–643.

Shijie Zhang and Hongzhi Yin. 2022. Comprehensive privacy analysis on federated recommender system against attribute inference attacks. *arXiv preprint arXiv:2205.11857* (2022).

Shijie Zhang, Hongzhi Yin, Tong Chen, Zi Huang, Lizhen Cui, and Xiangliang Zhang. 2021c. Graph embedding for recommendation against attribute inference attacks. In *The Web Conference*. 3002–3014.

Shijie Zhang, Hongzhi Yin, Tong Chen, Zi Huang, Quoc Viet Hung Nguyen, and Lizhen Cui. 2022. PipAttack: Poisoning federated recommender systems for manipulating item promotion. *WSDM* (2022).

Shijie Zhang, Hongzhi Yin, Tong Chen, Quoc Viet Nguyen Hung, Zi Huang, and Lizhen Cui. 2020. GCN-based user representation learning for unifying robust recommendation and fraudster detection. In *SIGIR*. 689–698.

Lei Zhu, Tianshi Wang, Jingjing Li, Zheng Zhang, Jialie Shen, and Xinhua Wang. 2022. Efficient query-based black-box attack against cross-modal hashing retrieval. *ACM Transactions on Information Systems* (2022).

Received 5 January 2023; accepted 9 January 2023