Towards Secure and Robust Recommender Systems: Recent Advances and Future Prospectives

Zongwei Wang[1], Junliang Yu[2], Tong Chen[2], Hongzhi Yin[2], Shazia Sadiq[2], Min Gao[1,*]
[1]Chongqing University, China, [2]The University of Queenslan, Australia
{zongwei,gaomin}@cqu.edu.cn,{jl.yu,tong.chen,h.yin1}@uq.edu.au,shazia@eecs.uq.edu.au

ABSTRACT

Recommender systems (RS) have proven to be vulnerable to malicious activities, where harmful data is injected into training datasets or model gradients are manipulated. These attacks compromise system functionality and distort RS objectives, allowing exploitation for malicious purposes. While significant research has been dedicated to understanding attack methods and their weaknesses, defensive strategies to protect RS from these vulnerabilities remain underexplored. This tutorial addresses secure and robust recommendations by exploring three critical research questions: (1) From the secure recommender system perspective, understanding the factors involved in executing successful attacks and exploring techniques to enhance system security; (2) From the robust recommender system perspective, addressing the challenges posed by maliciously injected and noisy data, and examining methods such as data denoising and model enhancement to improve system resilience; (3) Identifying research limitations and future opportunities, particularly the role of large language models in creating new vulnerabilities for attackers and offering novel defensive strategies for defenders. Additionally, the tutorial will culminate in the release of a toolkit designed to support empirical analysis and foster future research innovations in secure and robust recommendations.

COVER SHEET

Tutorial Length. The duration of the tutorial is 3 hours.

Intended Audience. This tutorial is designed for a broad audience, including academic researchers, postgraduate students, and industry professionals from the recommendation community and related areas. The attendee will gain a solid understanding of the methodologies for building secure and robust recommender systems.

Relevant Tutorials. There is only one relevant prior tutorial on trustworthy recommender systems[1], which broadens its scope to provide an introduction and future outlook on the entire landscape of trustworthy recommender systems. However, its coverage of safety and robustness remains relatively superficial. In contrast, our tutorial focuses specifically on secure and robust recommender systems, offering an in-depth analysis. This approach allows for a more comprehensive and profound understanding of the field.

Brief Biography of Presenters:

- **Prof. Hongzhi Yin** is an ARC Future Fellow, Full Professor, and the Director of the Responsible Big Data Intelligence Lab at The University of Queensland. He has published 300+ papers with an H-index of 77, making notable contributions to recommender systems, graph learning, decentralized learning, and edge intelligence. He has rich lecture experience and taught five relevant courses, such as information retrieval and web search, social media analytics, and responsible data science. Additionally, he has delivered 20+ keynotes and tutorials at the top-tier conferences like WWW'17,22,24, DASFAA'23, and KDD'17.
- **Mr. Zongwei Wang** is currently pursuing his Ph.D. at Chongqing University and is a visiting student at The University of Queensland. His research work has been published on top data mining conferences such as KDD, TIST, PAKDD, WSDM, etc. He has ample experience tutoring relevant courses and has presented his work at multiple top-tier conferences, such as KDD and PAKDD.
- **Dr. Junliang Yu** is an ARC DECRA Fellow at the University of Queensland. His research interests include recommender systems, data-centric AI, and graph learning. With over 30 publications in premier venues, he is a recognized contributor in his field. He has delivered multiple lectures at summer schools and taught courses on recommender systems and social media analytics. He organized two well-received tutorials on self-supervised recommender systems at WWW'22 and DASFAA'23, and presented his work at multiple top-tier conferences.
- **Dr. Tong Chen** is a Senior Lecturer and ARC DECRA Fellow at The University of Queensland. His research on lightweight, on-device, and trustworthy recommender systems has been published on top-tier international venues such as KDD, SIGIR, WWW, TKDE, WSDM, TNNLS, TOIS, and CIKM. He has ample track records in lecturing, witnessed by his course design and delivery experience in business analytics, teaching experience in data science, as well as invited tutorials on cutting-edge recommender systems at the WWW'22,24, and DASFAA'23.
- **Prof. Shazia Sadiq** is a Full Professor at The University of Queensland. Her research focuses on responsible data management and aims to reduce the socio-technical barriers to data driven transformation. She is the Fellow of the Australian Academy of Technological Sciences and Engineering. Throughout her 25 year career, she has received numerous invitations to speak at prestigious conferences, academic institutions, and industry forums, delivering over 20 tutorials, talks, panels and keynotes. One notable example is the Keynote talk at SIGMOD'23.
- **Prof. Min Gao** works as a Full Professor at Chongqing University, China. She has published 100+ papers, making notable contributions to recommender systems and data mining. She has been SPC or PC for many top conferences, such as WWW, IJCAI, AAAI, KDD, WSDM, and CIKM. Prof. Min Gao has rich lecture experience and taught three relevant courses, such as advanced machine learning, computer networks, and advanced database, and has presented her work at multiple top-tier conferences.

* Min Gao is the contact person.
[1] https://advanced-recommender-systems.github.io/trustworthiness-tutorial/

1 TOPIC AND MOTIVATION

Advanced recommender systems (RS) have been shown to be vulnerable to malicious data or noisy data. These inputs can corrupt training datasets or interfere with model gradients, thereby compromising system performance and distorting the intended outcomes of RS. Recent studies have explored the multifaceted nature of adversarial threats to secure and robust recommender systems, with a focus on the impact of various attack methods [1, 3, 5, 8, 19, 21] and noisy data [4, 7, 23, 26, 27]. In parallel, significant efforts have been dedicated to developing and refining defensive strategies that mitigate these vulnerabilities to safeguard the integrity and functionality of recommender systems[2, 9–11, 16, 18].

The tutorial will introduce secure and robust recommendations by focusing on three key aspects:

(1) Secure recommender system: An attacker must account for numerous factors throughout the entire process of executing a successful attack, while various defensive measures are currently employed to enhance system security and robustness. In this tutorial, we will introduce a systematic framework that encompasses the four key stages of a poisoning attack: attack goals, attacker capabilities, victim architecture, and poisoning strategies. Additionally, we will categorize defensive strategies into two primary approaches: filtering poisoned data and implementing robust training techniques.

(2) Robust recommender system. Recommender systems often face challenges from both maliciously injected data and unintentionally generated noisy data, which can degrade model performance and increase vulnerability to attacks. These issues can be mitigated through data denoising techniques, while the model itself can be strengthened using various enhancement technologies. In this tutorial, we will provide a detailed exploration of the robustness challenges that recommender systems encounter, presenting existing robust methods from two perspectives: those that operate without auxiliary information and those that incorporate auxiliary information to improve system robustness.

(3) Research limitations and future opportunities. Current research on secure recommender systems faces several limitations, leaving significant room for further exploration. We will address the gaps in existing studies and propose innovative directions for advancing the field of secure recommender systems. In particular, we will examine how the integration of large language models introduces both new vulnerabilities that attackers can exploit and novel defensive strategies for defenders, offering a promising avenue for enhancing system security and robustness.

While existing research on secure and robust recommendations has shown promising results, these critical questions have not been thoroughly or systematically explored. There is an urgent need to reveal the answers to these questions in order to further promote this line of research. Additionally, by releasing our toolkit for a secure and robust recommendation, we expect to facilitate empirical comparisons and future methodological development of secure and robust recommendation methods.

2 RELEVANT PUBLICATIONS BY THE ORGANIZERS

- Secure recommender system [6, 12, 15, 17, 20, 22, 24, 28].
- Robust recommender system [13, 25].
- A survey paper on the secure recommendation [14].
- A toolkit for secure recommendation[2].

3 FORMAT AND DETAILED SCHEDULE

The tutorial is delivered as a lecture-style tutorial (3 hours in duration). The schedule and detailed organization of the topics are as follows.

I. Introduction (20 mins)

(1) Overview of Recommendation (5 mins)
(2) Overview of Secure and Robust Recommendation (15 mins)

II. Secure Recommendation (50 mins)

(1) Poisoning Attacks against Recommendation (40 mins)
(2) Defense Against Poisoning Attacks in Recommendation (30 mins)

III. Robust Recommendation (50 mins)

(1) Robustness Problem in Recommendation (10 mins)
(2) Robust Methods Against Noisy Data (40 mins)

IV. Limitations and Future Research Trends (30 mins)

(1) Extension of Existing Research Questions (15 mins)
(2) Large Language Models in Secure and Robust Recommendation (15 mins)

V. Open-source Toolkit for Robust and Safe Recommendation (10 mins)

4 TARGET AUDIENCE

This tutorial is designed for a broad audience, including academic and industrial researchers, graduate students, and practitioners from the recommendation field and related areas. By the end of the tutorial, participants will have a solid understanding of basic poisoning attacks and defensive strategies to enhance the robustness and security of recommendation systems. Additionally, they will gain hands-on experience using an open-source toolkit. While prior knowledge of recommendation systems is preferred, the tutorial will also cover foundational concepts to ensure better engagement and accessibility for all attendees.

5 TUTORIAL MATERIALS

The tutorial materials, including the slides and video recordings, will be made available online in advance for attendees via GitHub. To avoid the potential occurrence of technical problems, a pre-recorded lecture will be uploaded to YouTube beforehand.

REFERENCES

[1] Wenqi Fan, Tyler Derr, Xiangyu Zhao, Yao Ma, Hui Liu, Jianping Wang, Jiliang Tang, and Qing Li. 2021. Attacking black-box recommendations via copying cross-domain user profiles. In *ICDE*.

[2] https://github.com/CoderWZW/ARLib

[2] Yaojun Hao, Guoyan Meng, Jian Wang, and Chunmei Zong. 2023. A detection method for hybrid attacks in recommender systems. *Information Systems* 114 (2023), 102154.
[3] Chen Lin, Si Chen, Hui Li, Yanghua Xiao, Lianyun Li, and Qian Yang. 2020. Attacking recommender systems with augmented user profiles. In *CIKM*.
[4] Weilin Lin, Xiangyu Zhao, Yejing Wang, Yuanshao Zhu, and Wanyu Wang. 2023. Autodenoise: Automatic data instance denoising for recommendations. In *Proceedings of the ACM Web Conference 2023*. 1003–1011.
[5] Jing Long, Tong Chen, Quoc Viet Hung Nguyen, Guandong Xu, Kai Zheng, and Hongzhi Yin. 2023. Model-agnostic decentralized collaborative learning for on-device POI recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 423–432.
[6] Hao Ma, Min Gao, Feng Wei, Zongwei Wang, Feng Jiang, Zehua Zhao, and Zhengyi Yang. 2024. Stealthy attack on graph recommendation system. *Expert Systems with Applications* (2024), 124476.
[7] Yuhan Quan, Jingtao Ding, Chen Gao, Lingling Yi, Depeng Jin, and Yong Li. 2023. Robust preference-guided denoising for graph based social recommendation. In *Proceedings of the ACM Web Conference 2023*. 1097–1108.
[8] Qinyong Wang, Hongzhi Yin, Tong Chen, Junliang Yu, Alexander Zhou, and Xiangliang Zhang. 2022. Fast-adapting and privacy-preserving federated recommender system. *The VLDB Journal* (2022).
[9] Shilei Wang, Peng Zhang, Hui Wang, Hongtao Yu, and Fuzhi Zhang. 2022. Detecting shilling groups in online recommender systems based on graph convolutional network. *Information Processing & Management* 59, 5 (2022), 103031.
[10] Wenjie Wang, Fuli Feng, Xiangnan He, Liqiang Nie, and Tat-Seng Chua. 2021. Denoising implicit feedback for recommendation. In *Proceedings of the 14th ACM international conference on web search and data mining*. 373–381.
[11] Yu Wang, Xin Xin, Zaiqiao Meng, Joemon M Jose, Fuli Feng, and Xiangnan He. 2022. Learning robust recommenders through cross-model agreement. In *Proceedings of the ACM Web Conference 2022*. 2015–2025.
[12] Zongwei Wang, Min Gao, Jundong Li, Junwei Zhang, and Jiang Zhong. 2022. Gray-box shilling attack: an adversarial learning approach. *ACM TIST* (2022).
[13] Zongwei Wang, Min Gao, Wentao Li, Junliang Yu, Linxin Guo, and Hongzhi Yin. 2023. Efficient Bi-Level Optimization for Recommendation Denoising. In *SIGKDD*.
[14] Zongwei Wang, Junliang Yu, Min Gao, Guanhua Ye, Shazia Sadiq, and Hongzhi Yin. 2024. Poisoning Attacks and Defenses in Recommender Systems: A Survey. *arXiv preprint arXiv:2406.01022* (2024).
[15] Zongwei Wang, Junliang Yu, Min Gao, Hongzhi Yin, Bin Cui, and Shazia Sadiq. 2024. Unveiling Vulnerabilities of Contrastive Recommender Systems to Poisoning Attacks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3311–3322.
[16] Chenwang Wu, Defu Lian, Yong Ge, Zhihao Zhu, Enhong Chen, and Senchao Yuan. 2021. Fight fire with fire: towards robust recommender systems via adversarial poisoning training. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1074–1083.
[17] Fan Wu, Min Gao, Junliang Yu, Zongwei Wang, Kecheng Liu, and Xu Wang. 2021. Ready for emerging threats to recommender systems? A graph convolution-based generative shilling attack. *Information Sciences* (2021).
[18] Yishu Xu and Fuzhi Zhang. 2019. Detecting shilling attacks in social recommender systems based on time series analysis and trust features. *Knowledge-Based Systems* 178 (2019), 25–47.
[19] Liu Yang, Ben Tan, Vincent W Zheng, Kai Chen, and Qiang Yang. 2020. Federated recommendation systems. *Federated Learning: Privacy and Incentive* (2020), 225–239.
[20] Wei Yuan, Quoc Viet Hung Nguyen, Tieke He, Liang Chen, and Hongzhi Yin. 2023. Manipulating Federated Recommender Systems: Poisoning with Synthetic Users and Its Countermeasures. *arXiv* (2023).
[21] Wei Yuan, Hongzhi Yin, Fangzhao Wu, Shijie Zhang, Tieke He, and Hao Wang. 2023. Federated unlearning for on-device recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*.
[22] Wei Yuan, Shilong Yuan, Chaoqun Yang, Nguyen Quoc Viet hung, and Hongzhi Yin. 2023. Manipulating Visually Aware Federated Recommender Systems and Its Countermeasures. *ACM Transactions on Information Systems* 42, 3 (2023), 1–26.
[23] Chi Zhang, Rui Chen, Xiangyu Zhao, Qilong Han, and Li Li. 2023. Denoising and prompt-tuning for multi-behavior recommendation. In *Proceedings of the ACM Web Conference 2023*. 1355–1363.
[24] 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. In *WSDM*.
[25] 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*.
[26] Weibin Zhao, Lin Shang, Yonghong Yu, Li Zhang, Can Wang, and Jiajun Chen. 2023. Personalized tag recommendation via denoising auto-encoder. *World Wide Web* 26, 1 (2023), 95–114.
[27] Jiawei Zheng, Qianli Ma, Hao Gu, and Zhenjing Zheng. 2021. Multi-view denoising graph auto-encoders on heterogeneous information networks for cold-start recommendation. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 2338–2348.
[28] Ruiqi Zheng, Liang Qu, Tong Chen, Kai Zheng, Yuhui Shi, and Hongzhi Yin. 2024. Poisoning decentralized collaborative recommender system and its countermeasures. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1712–1721.