User Privacy in Recommender Systems

Peter Müllner(✉)

Know-Center GmbH, Graz, Austria
pmuellner@know-center.at

Abstract. Recommender systems process abundances of user data to generate recommendations that fit well to each individual user. This utilization of user data can pose severe threats to user privacy, e.g., the inadvertent leakage of user data to untrusted parties or other users. Moreover, this data can be used to reveal a user's identity, or to infer very private information as, e.g., gender. Instead of the plain application of privacy-enhancing techniques, which could lead to decreased accuracy, we tackle the problem itself, i.e., the utilization of user data. With this, we aim to equip recommender systems with means to provide high-quality recommendations that respect users' privacy.

Keywords: Recommender systems · Differential privacy · Data minimization · Neighborhood reuse

1 Motivation

Recommender systems are quintessential tools that help users navigating through the overload of information prevalent in many applications. Typically, historic user interaction data is utilized to generate personalized recommendations. This, however, poses a privacy threat, since through the utilization of their data, users' private information could be disclosed to untrusted parties or other users.

To hinder the disclosure of private information, legal and technical solutions have been implemented. For example, the General Data Protection Regulation's principle of data minimization requires that only the necessary data shall be processed [24]. However, requiring the recommender system to use a minimal amount of user data leads to a drop in recommendation accuracy [5]. Therefore, we study *RQ1: How can recommender systems use fewer data to generate meaningful recommendations?* Also, homomorphic encryption [10,15], federated learning [17,27], or differential privacy [6] have been applied to ensure privacy. Especially differential privacy became widely-used in a broad body of information retrieval applications, e.g., recommender systems [9,31]. However, differential privacy negatively affects recommendation accuracy, since it adds random noise to the user data [4,20]. Thus, we study *RQ2: How can we improve the recommendation accuracy of differentially private recommender systems?* and *RQ3: In which ways does differential privacy impact personalized recommendations?*

J. Kamps et al. (Eds.): ECIR 2023, LNCS 13982, pp. 456–461, 2023.
https://doi.org/10.1007/978-3-031-28241-6_52

Overall, we investigate how recommender systems can process user data in a more responsible way and also, we explore how we can make the negative impacts, differential privacy can have on users, less impactful.

2 Related Work

Due to the increasing awareness of the potential disclosure of private information, numerous works identify several ways in which recommender systems jeopardize users' privacy [1,2,14,26]. For example, Strucks et al. [25] can infer users' gender based on parts of the user's rating data. Also, the utilization of users' rating data to generate recommendations poses a severe privacy threat [23]. Through the generated recommendations, the recommender system could leak information about the users, whose data has been used. Similarly, Zhang et al. [30] illustrate how to identify users, whose rating data was used in the recommendation process.

To make these privacy risks less serious, for example, Biega et al. [5] minimize a user's data that the recommender system is allowed to use and find that the accuracy loss is less serious than expected. Plus, the severity of the accuracy loss is different for different users (see our work on *RQ1* in Sect. 3).

Besides applying homomorphic encryption [15] and federated learning [18], especially differential privacy [6] has been used in many research works [8,9,13,20]. For example, Zhu et al. [31] use differential privacy to privately select users, whose data is used to generate recommendations. Plus, the data itself is protected as well via the addition of noise. Through fine-tuning the level of noise, Zhu et al. can make the accuracy-privacy trade-off less serious. Similarly, also Liu et al. [19] vary the level of noise to ensure meaningful recommendations. Xin and Jaakola [28] protect only a subset of users with differential privacy, and this way, recommendation accuracy benefits from the data of unprotected users (see our work on *RQ2* in Sect. 3).

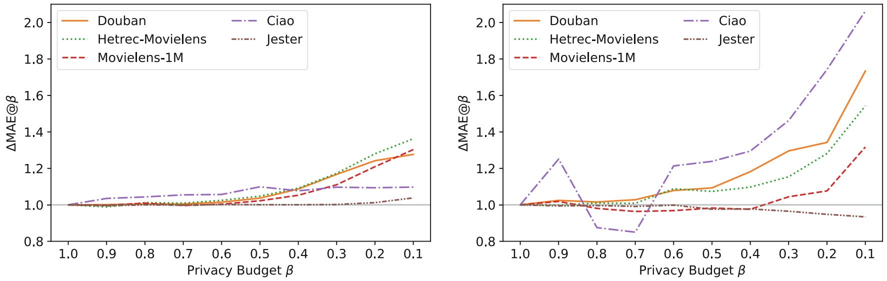
Moreover, Zemel et al. [29] and Ekstrand et al. [7] discuss the impact of differential privacy on fairness, which is an example of the different ways in which differential privacy might impact users (see our work on *RQ3* in Sect. 3).

3 Research Questions and Preliminary Results

RQ1: How can recommender systems use fewer data to generate meaningful recommendations?

Data minimization can aid user privacy, however, it results in a drop in recommendation accuracy [5]. Thus, we study how strong the amount of data can be minimized and how serious this impacts recommendation accuracy. Plus, we explore if meta learning [11] can help to make this accuracy drop less severe.

Our Work. In our work [21], we experiment with MetaMF [18] and test its robustness against small privacy budgets. We measure robustness via the relative accuracy loss ΔMAE, and privacy budget is the fraction β of each user's



(a) MetaMF

(b) NoMetaMF

Fig. 1. Relative accuracy drop of MetaMF and NoMetaMF for different privacy budgets β, i.e., the fraction of a user's data that the recommender system can use. $\beta \geq 0.5$ is sufficient to keep recommendation accuracy. In case $\beta < 0.5$, meta learning is required to keep the accuracy loss at a moderate level.

data that the recommender system is allowed to use. Furthermore, we isolate the impact meta learning has on the recommendation accuracy via evaluating NoMetaMF (i.e., a variant without meta learning).

We find that for most datasets, approximately 50% of each user's data is sufficient to keep recommendation accuracy (see Fig. 1). In case fewer data is available, the relative accuracy loss increases much slower for MetaMF than for NoMetaMF. This shows that for small privacy budgets, meta learning helps to keep recommendation accuracy. Moreover, highly active users with lots of data experience a more severe accuracy loss than users that are less active, which is a sign of different accuracy-privacy trade-offs.

Open Issues. In our existing work, we assume that every piece of data is equally sensitive, i.e., how much it puts a user's privacy at risk when disclosed. In reality, there exists high-sensitive data, e.g., gender, as well as low-sensitive data, e.g., favorite color. Similarly, the recommender system can generate more or less accurate recommendations depending on the data it can use. Thus, we will study how to quantify the privacy-sensitivity of data and how important the data is for the recommender system to generate meaningful recommendations.

RQ2: How can we improve the recommendation accuracy of differentially private recommender systems?

Differential privacy [6] typically leads to a decrease in recommendation accuracy due to the addition of noise to the rating data [3]. Our idea is to minimize the set of users that need to be protected with differential privacy. This way, we limit the addition of noise, which leads to a better accuracy-privacy trade-off.

Our Work. In *KNN* recommender systems, neighbors' data is used to generate recommendations, which poses a privacy risk for these neighbors [23]. In our work [22], we develop the *ReuseKNN* recommender system, which reuses the same neighbors for many recommendations. This way, only a few users are used as neighbors and need to be protected with differential privacy, while most users do not need to be protected, since they are only rarely used as neighbors.

We find that *ReuseKNN* can substantially decrease the number of users that need to be protected with differential privacy. Depending on the dataset, down to 24% of users need to be protected, compared to 80% for traditional *UserKNN* [12,31]. Also, recommendation accuracy can be preserved, and in many cases improved over *UserKNN*, while the users' privacy risk decreases. Plus, we find that *ReuseKNN* does not exacerbate an existing popularity bias.

Open Issues. In our work, a user's privacy risk quantifies how often the user's data is utilized in the recommendation process. However, similar to our idea for future work on *RQ1*, this does not consider that the data of two users could differ in their sensitivity. Also, users with high-sensitive data could be protected differently than users with low-sensitive data.

RQ3: In which ways does differential privacy impact personalized recommendations?

In addition to the impact of differential privacy on recommendation accuracy, we are also interested in how beyond-accuracy objectives are impacted, e.g., popularity bias [16], and if different users are impacted differently. Also, we investigate the longitudinal effects on the users if differential privacy is applied.

Our Work. We apply differential privacy by adding noise to the users' ratings [6] and compare recommendation lists generated with and without differential privacy to quantify how severe differential privacy can change a user's recommendations. Plus, we study in which ways differential privacy impacts the recommendation lists by monitoring recommendation accuracy and popularity bias.

Our first experiments indicate that depending on the dataset and algorithm, between 54% and 80% of users are impacted by differential privacy. On average, these users' recommendation accuracy decreases by 5% to 26%. However, for 80% to 91% of these users, differential privacy leads to the recommendation of fewer popular items, which can help to alleviate popularity bias.

Open Issues. We identify two open issues. First, specific user behavior could influence the impact of differential privacy: for example, users with diverse preferences could be impacted in different ways than users with narrow preferences. Second, the impact of differential privacy could accumulate when many recommendations are generated for a user over time. For these open issues, we believe that a simulation study is a well-suited approach.

Acknowledgements. Thanks to my supervisors Dominik Kowald and Elisabeth Lex for their feedback on this work. This work is supported by the "DDAI" COMET Module within the COMET - Competence Centers for Excellent Technologies Programme, funded by the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit), the Austrian Federal Ministry for Digital and Economic Affairs (bmdw), the

Austrian Research Promotion Agency (FFG), the province of Styria (SFG) and partners from industry and academia. The COMET Programme is managed by FFG.

References

1. Beigi, G., Liu, H.: "Identifying novel privacy issues of online users on social media platforms" by Ghazaleh Beigi and Huan Liu with Martin Vesely as coordinator. SIGWEB Newsl. (Winter) (2019). https://doi.org/10.1145/3293874.3293878
2. Beigi, G., Liu, H.: A survey on privacy in social media: identification, mitigation, and applications. ACM Trans. Data Sci. **1**(1), 1–38 (2020)
3. Berkovsky, S., Kuflik, T., Ricci, F.: The impact of data obfuscation on the accuracy of collaborative filtering. Expert Syst. Appl. **39**(5), 5033–5042 (2012)
4. Berlioz, A., Friedman, A., Kaafar, M.A., Boreli, R., Berkovsky, S.: Applying differential privacy to matrix factorization. In: Proceedings of the 9th ACM Conference on Recommender Systems, pp. 107–114 (2015)
5. Biega, A.J., Potash, P., Daumé, H., Diaz, F., Finck, M.: Operationalizing the legal principle of data minimization for personalization. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 399–408 (2020)
6. Dwork, C.: Differential privacy: a survey of results. In: Agrawal, M., Du, D., Duan, Z., Li, A. (eds.) TAMC 2008. LNCS, vol. 4978, pp. 1–19. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-79228-4_1
7. Ekstrand, M.D., Joshaghani, R., Mehrpouyan, H.: Privacy for all: ensuring fair and equitable privacy protections. In: Conference on Fairness, Accountability and Transparency, pp. 35–47. PMLR (2018)
8. Friedman, A., Berkovsky, S., Kaafar, M.A.: A differential privacy framework for matrix factorization recommender systems. User Model. User-Adap. Inter. **26**(5), 425–458 (2016). https://doi.org/10.1007/s11257-016-9177-7
9. Gao, C., Huang, C., Lin, D., Jin, D., Li, Y.: DPLCF: differentially private local collaborative filtering. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 961–970 (2020)
10. Gentry, C.: A Fully Homomorphic Encryption Scheme. Stanford University (2009)
11. Ha, D., Dai, A., Le, Q.V.: Hypernetworks. arXiv preprint arXiv:1609.09106 (2016)
12. Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J.: An algorithmic framework for performing collaborative filtering. In: Proceedings of the 22nd annual international ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 230–237 (1999)
13. Hou, D., Zhang, J., Ma, J., Zhu, X., Man, K.L.: Application of differential privacy for collaborative filtering based recommendation system: a survey. In: 2021 12th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP), pp. 97–101. IEEE (2021)
14. Jeckmans, A.J., Beye, M., Erkin, Z., Hartel, P., Lagendijk, R.L., Tang, Q.: Privacy in recommender systems. In: Ramzan, N., van Zwol, R., Lee, J.S., Clüver, K., Hua, X.S. (eds.) Social Media Retrieval, pp. 263–281. Springer, Cham (2013). https://doi.org/10.1007/978-1-4471-4555-4_12
15. Kim, S., Kim, J., Koo, D., Kim, Y., Yoon, H., Shin, J.: Efficient privacy-preserving matrix factorization via fully homomorphic encryption. In: Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security, pp. 617–628 (2016)

16. Kowald, D., Muellner, P., Zangerle, E., Bauer, C., Schedl, M., Lex, E.: Support the underground: characteristics of beyond-mainstream music listeners. EPJ Data Sci. **10**(1), 1–26 (2021). https://doi.org/10.1140/epjds/s13688-021-00268-9
17. Li, Q., et al.: A survey on federated learning systems: vision, hype and reality for data privacy and protection. IEEE Trans. Knowl. Data Eng. (2021). https://ieeexplore.ieee.org/document/9599369
18. Lin, Y., et al.: Meta matrix factorization for federated rating predictions. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 981–990 (2020)
19. Liu, J., Hu, Y., Guo, X., Liang, T., Jin, W.: Differential privacy performance evaluation under the condition of non-uniform noise distribution. J. Inf. Secur. Appl. **71**, 103366 (2022)
20. Liu, X., et al.: When differential privacy meets randomized perturbation: a hybrid approach for privacy-preserving recommender system. In: Candan, S., Chen, L., Pedersen, T.B., Chang, L., Hua, W. (eds.) DASFAA 2017. LNCS, vol. 10177, pp. 576–591. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55753-3_36
21. Muellner, P., Kowald, D., Lex, E.: Robustness of meta matrix factorization against strict privacy constraints. In: Hiemstra, D., Moens, M.-F., Mothe, J., Perego, R., Potthast, M., Sebastiani, F. (eds.) ECIR 2021. LNCS, vol. 12657, pp. 107–119. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72240-1_8
22. Müllner, P., Lex, E., Schedl, M., Kowald, D.: ReuseKNN: neighborhood reuse for differentially-private KNN-based recommendations (2022). https://doi.org/10.48550/ARXIV.2206.11561
23. Ramakrishnan, N., Keller, B.J., Mirza, B.J., Grama, A.Y., Karypis, G.: When being weak is brave: privacy in recommender systems. IEEE Internet Comput. **5**(6), 54–62 (2001)
24. Parliament Regulation: Regulation (EU) 2016/679 of the European parliament and of the council. Regulation (EU) 679, 2016 (2016)
25. Strucks, C., Slokom, M., Larson, M.: BlurM(or)e: revisiting gender obfuscation in the user-item matrix (2019)
26. Wang, C., Zheng, Y., Jiang, J., Ren, K.: Toward privacy-preserving personalized recommendation services. Engineering **4**(1), 21–28 (2018)
27. Wang, Q., Yin, H., Chen, T., Yu, J., Zhou, A., Zhang, X.: Fast-adapting and privacy-preserving federated recommender system. VLDB J. **31**(5), 877–896 (2022)
28. Xin, Y., Jaakkola, T.: Controlling privacy in recommender systems. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS 2014, vol. 2, pp. 2618–2626. MIT Press, Cambridge, MA, USA (2014)
29. Zemel, R., Wu, Y., Swersky, K., Pitassi, T., Dwork, C.: Learning fair representations. In: International Conference on Machine Learning, pp. 325–333. PMLR (2013)
30. Zhang, M., et al.: Membership inference attacks against recommender systems. In: Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security, pp. 864–879 (2021)
31. Zhu, T., Li, G., Ren, Y., Zhou, W., Xiong, P.: Differential privacy for neighborhood-based collaborative filtering. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 752–759 (2013)