Impact of Meta Learning for Privacy-Preserving Recommender Systems

Peter Müllner, Elisabeth Lex, Dominik Kowald

The Responsible AI Forum 2021
AI ethics in practice: Some applications

Motivation

- Many large-scale recommender systems
- No “one size fits all” recommendations
- Personalization to increase recommendation accuracy
- Inclusion of more user data, e.g., interaction data, personal data, ...

Privacy Issues

- Utilization of enormous amounts of user data
- Sensitive attributes could get leaked
  - Data breach
  - Infer private attributes
- Also state of the art could leak data, e.g., federated learning [1, 7]

Prior Work

Robustness of Meta Matrix Factorization Against Strict Privacy Constraints [3]

- Users have different inclinations to reveal information
- Reason: No data disclosure if there is no data

Prior Work

Robustness of Meta Matrix Factorization Against Strict Privacy Constraints [3]

- Users have different inclinations to reveal information
- Reason: No data disclosure if there is no data
- **Privacy Budget** is the fraction of data a user reveals to the recommender system
- Small privacy budget ⇒ high privacy, but bad recommendations
- Large privacy budget ⇒ low privacy, but good recommendations

Prior Work

Robustness of Meta Matrix Factorization Against Strict Privacy Constraints [3]

- Users have different inclinations to reveal information
- Reason: No data disclosure if there is no data
- **Privacy Budget** is the fraction of data a user reveals to the recommender system
- Small privacy budget $\Rightarrow$ high privacy, but bad recommendations
- Large privacy budget $\Rightarrow$ low privacy, but good recommendations

How can we have high privacy and high accuracy?

Meta Learning

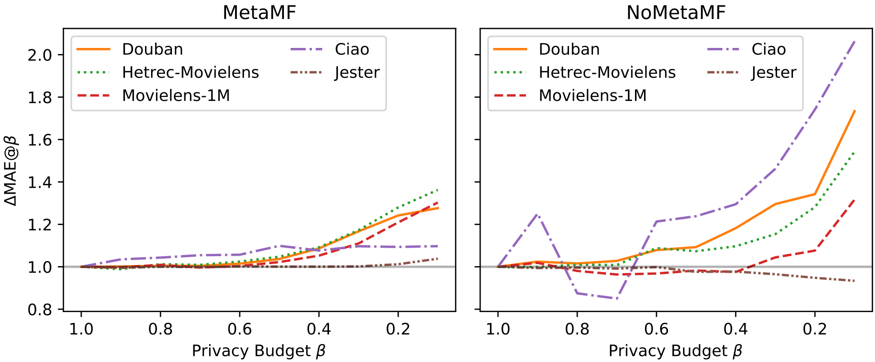
- Meta Learning has been used for problems related to privacy, e.g., cold-start problem [6], few-shot learning [5], ...
- Learn model for group of users and not individual users
- Meta Matrix Factorization (MetaMF) developed by Lin et al. [2]

Approach

- Underline the impact of meta learning on privacy-preserving recommendender systems
- Side-by-side comparison between MetaMF and **NoMetaMF**, i.e., MetaMF with no meta learning
- *RQ1: What is the benefit of meta learning for small privacy budgets?*
- *RQ2: Do small privacy budgets impact users differently?*

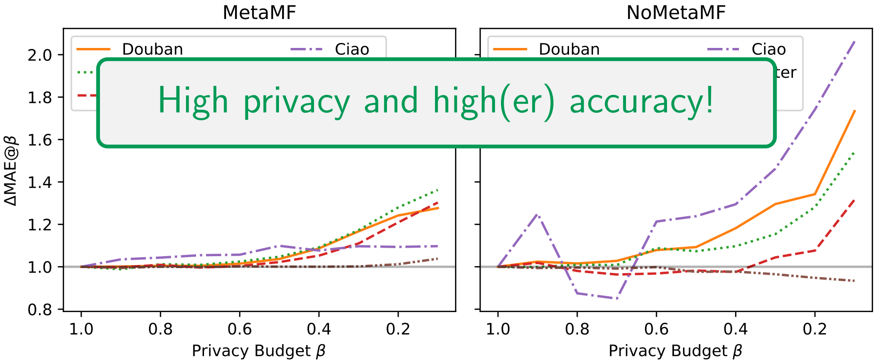
RQ1: Benefit of meta learning

- **MetaMF** with meta learning, **NoMetaMF** without meta learning
- Privacy budget β is the fraction of a user's revealed data
- Relative accuracy loss $\Delta\mathrm{MAE}@\beta = \mathrm{MAE}@\beta\ /\ \mathrm{MAE}@1.0$



RQ1: Benefit of meta learning

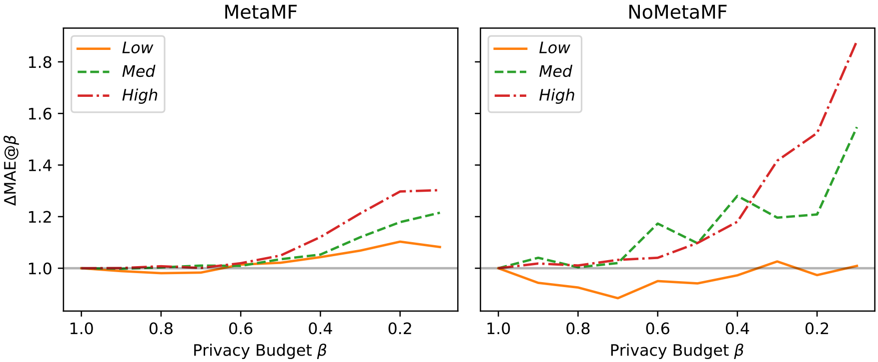
- **MetaMF** with meta learning, **NoMetaMF** without meta learning
- Privacy budget β is the fraction of a user's revealed data
- Relative accuracy loss $\Delta\mathrm{MAE}@\beta = \mathrm{MAE}@\beta \;/\; \mathrm{MAE}@1.0$



High privacy and high(er) accuracy!

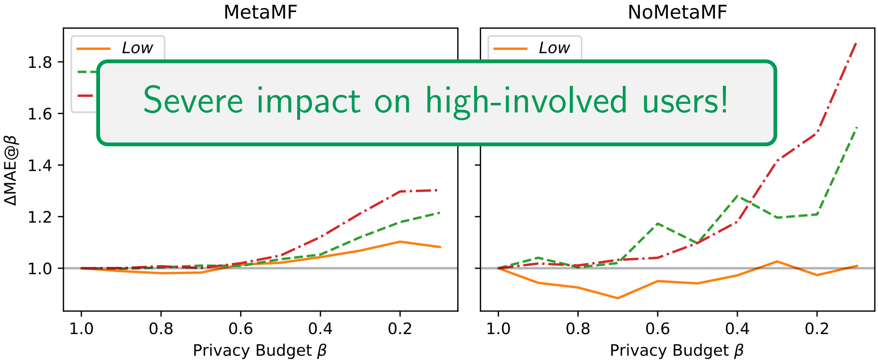
RQ2: Effect on different user types (Douban)

- High-involved users, medium-involved users, low-involved users
- Identify user types based on number of ratings
- Train on all users, evaluate on *Low*, *Med*, and *High*



RQ2: Effect on different user types (Douban)

- High-involved users, medium-involved users, low-involved users
- Identify user types based on number of ratings
- Train on all users, evaluate on *Low*, *Med*, and *High*



Conclusions & Future Work

Conclusion

- Revealing $\geq 50\%$ of data only marginally decreases accuracy
- Revealing $< 50\%$ of data requires meta learning
- Low-involved users can afford privacy
- High-involved users have to sacrifice a lot of accuracy for privacy
- Individual accuracy-privacy trade-off!

Future Work

- Consider that there is data with different sensitivities
- What data should be shared based on data characteristics to sustain accuracy of recommendations
- Not revealing data could impact recommendations of other users ($\rightarrow$ See our conceptual approach for privacy-aware simulations [4])

Thank you!

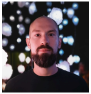
Source Code
github.com/pmuellner/RobustnessOfMetaMF

User Groups Data
zenodo.org/record/4031011

Contacts
{pmuellner, dkowald}@know-center.at
elisabeth.lex@tugraz.at

Peter Müllner
Know-Center GmbH

Elisabeth Lex
Graz University of Technology



Dominik Kowald
Know-Center GmbH

References I

E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, *How to backdoor federated learning*, in International Conference on Artificial Intelligence and Statistics, PMLR, 2020, pp. 2938–2948.

Y. Lin, P. Ren, Z. Chen, Z. Ren, D. Yu, J. Ma, M. d. Rijke, and X. Cheng, *Meta matrix factorization for federated rating predictions*, in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 981–990.

P. Muellner, D. Kowald, and E. Lex, *Robustness of meta matrix factorization against strict privacy constraints*, in Advances in Information Retrieval, Cham, 2021, Springer International Publishing, pp. 107–119.

References II

P. Müllner, E. Lex, and D. Kowald, *Position paper on simulating privacy dynamics in recommender systems*, arXiv preprint arXiv:2109.06473, (2021).

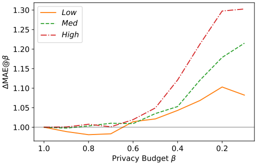
J. Snell, K. Swersky, and R. S. Zemel, *Prototypical networks for few-shot learning*, arXiv preprint arXiv:1703.05175, (2017).

M. Vartak, A. Thiagarajan, C. Miranda, J. Bratman, and H. Larochelle, *A meta-learning perspective on cold-start recommendations for items*, (2017).

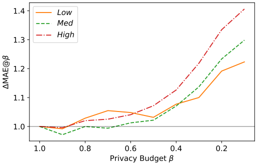
L. Zhu and S. Han, *Deep leakage from gradients*, in Federated learning, Springer, 2020, pp. 17–31.

RQ2: Effect on different user types

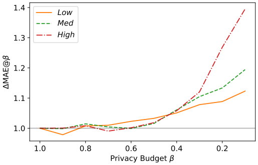
MetaMF - with meta learning



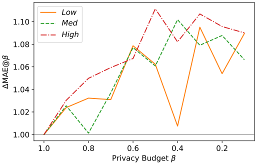
(a) Douban



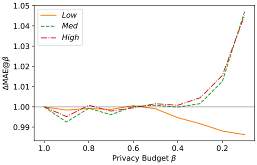
(b) Hetrec-MovieLens



(c) MovieLens 1M



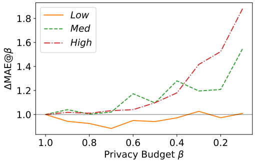
(d) Ciao



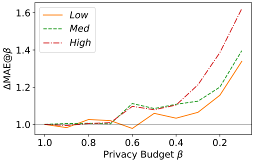
(e) Jester

RQ2: Effect on different user types

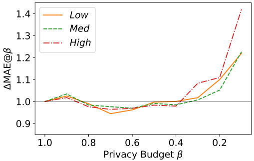
NoMetaMF - without meta learning



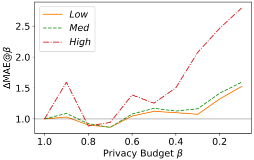
(f) Douban



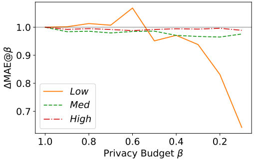
(g) Hetrec-MovieLens



(h) MovieLens 1M



(i) Ciao



(j) Jester

Dataset statistics

Dataset	$\|U\|$	$\|I\|$	$\|R\|$	User Group Size
Douban	2,509	39,576	893,575	125
Hetrec-MovieLens	2,113	10,109	855,598	106
MovieLens 1M	6,040	3,706	1,000,209	302
Ciao	7,373	105,096	282,619	369
Jester	73,321	100	4,136,360	2,671