International Journal of Neural Networks 15 (1), 2025
DOI: 11.012294j/a223ae.2025.08972

Differentially Private Lookalike Audience Expansion via Federated Representation Learning

Samuel Okafor, School of Computing and Information Systems, University of Melbourne, Parkville, VIC 3010, Australia

Copyright © 2025 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License.

Abstract

Lookalike audience expansion extrapolates from seed
users to larger targetable segments but risks privacy
leakage through high-dimensional behavioral profiles.
We present a federated representation learning
framework in which local encoders map user histories to low-dimensional embeddings updated via secure aggregation, with Gaussian noise enforcing user-level differential privacy. A Bayesian matching model then selects expansion candidates using only embeddings and coarse metadata. Experiments on a real social platform dataset show that our approach achieves near-parity on downstream campaign lift while reducing re-identification risk, enabling compliant audience expansion.

References

1. Dwork, Cynthia. 2008. “Differential Privacy: A Survey of Results.” In Theory and Applications of Models of Computation, 1–19. Berlin: Springer.
2. Hitaj, Briland, Giuseppe Ateniese, and Fernando Pérez-Cruz. 2017. “Deep Models under the GAN: Information Leakage from Collaborative Deep Learning.” In Proceedings of CCS 2017, 603–18.

3. Kairouz, Peter, et al. 2019. “Advances and Open Problems in Federated Learning.” arXiv:1912.04977.
4. McMahan, H. Brendan, et al. 2018. “Learning Differentially Private Recurrent Language Models.” In ICLR 2018.
5. Papernot, Nicolas, et al. 2018. “Scalable Private Learning with PATE.” In ICLR 2018.
6. Shokri, Reza, et al. 2017. “Membership Inference Attacks against Machine Learning Models.” In Proceedings of IEEE S&P 2017, 3– 18.

7. Song, Shuang, et al. 2019. “Privacy Risks of Model Explanations.” In Proceedings of CCS 2019, 241–57.
8. Zhu, Ligeng, Zhijian Liu, and Song Han. 2019. “Deep Leakage from Gradients.” In NeurIPS 2019, 14747–56.
9. Rao, Aditi, and Florian Schaub. 2020. “Engineering Privacy in Online Advertising Systems.” Communications of the ACM 63 (12): 68–77.
10. Voigt, Paul, and Axel von dem Bussche. 2017. The EU General Data Protection Regulation (GDPR). Cham: Springer.
11. Li, Xiaotong, et al. 2020. “Modeling Delayed Feedback in Display Advertising.” IEEE Transactions on Knowledge and Data Engineering 32 (3): 372–85.
12. Papadimitriou, Panagiotis, et al. 2019. “Constrained Optimization for Attribution Modeling.” In Proceedings of WWW 2019, 1891–1902.
13. Hardt, Moritz, and Eric Price. 2016. “Disparate Impact in Learning.” In Proceedings of the 29th Conference on Neural Information Processing Systems, 3315–23.
14. Kairouz, Peter, et al. 2019. “Advances and Open Problems in Federated Learning.” arXiv:1912.04977.
15. McMahan, H. Brendan, et al. 2017. “Communication-Efficient Learning of Deep Networks from Decentralized Data.” In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 1273–82.
16. Rogers, Ryan, et al. 2020. “LinkedIn’s Audience Expansion: A Look under the Hood.” In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 3183–91. New York: ACM.
17. Dudík, Miroslav, John Langford, and Lihong Li. 2011. “Doubly Robust Policy Evaluation

and Learning.” In Proceedings of the 28th International Conference on Machine Learning, 1097–1104.

18. Dwork, Cynthia, Aaron Roth, and Guy Rothblum. 2010. “Boosting and Differential Privacy.” In Proceedings of the 51st IEEE Symposium on Foundations of Computer Science, 51–60.
19. Jiang, Nan, and Lihong Li. 2016. “Doubly Robust Off-Policy Value Evaluation for Reinforcement Learning.” In Proceedings of the 33rd International Conference on Machine Learning, 652–61.
20. Kallus, Nathan. 2018. “Balanced Policy Evaluation and Learning.” Advances in Neural Information Processing Systems 31: 8895–902.
21. Athey, Susan, and Guido W. Imbens. 2016. “Recursive Partitioning for Heterogeneous Causal Effects.” Proceedings of the National Academy of Sciences 113 (27): 7353–60.
22. Balle, Borja, and Yu-Xiang Wang. 2018. “Improving the Gaussian Mechanism for Differential Privacy: Analytical Calibration and Optimal Denoising.” In Proceedings of the 35th International Conference on Machine Learning, 394–403.

23. Dwork, Cynthia, and Aaron Roth. 2014. “The Algorithmic Foundations of Differential Privacy.” Foundations and Trends in Theoretical Computer Science 9 (3–4): 211–407.
24. Mehta, D. B. (2023). Privacy-Preserving Machine Learning Architecture for Cross-Channel Advertising Optimization in a Large-Scale Social Media Platform. International Journal of Communication Networks and Information Security (IJCNIS), 15(8), 23–32. https://www.ijcnis.org/index.php/ijcnis/article/v iew/8396
25. Farronato, Chiara, Andrey Fradkin, and Catherine Tucker. 2020. “Platform Design and Advertising Auctions.” Journal of Industrial Economics 68 (4): 800–834.
26. McMahan, H. Brendan, Daniel Ramage, Kunal Talwar, and Li Zhang. 2018. “Learning Differentially Private Recurrent Language Models.” In 6th International Conference on Learning Representations (ICLR).
27. Narayanan, Arvind, and Vitaly Shmatikov. 2008. “Robust De-anonymization of Large Sparse Datasets.” In IEEE Symposium on Security and Privacy, 111–25. Los Alamitos, CA: IEEE.
28. Rogers, Ryan, Aaron Roth, Jonathan Ullman, and Salil Vadhan. 2016. “Privacy Odometers and Filters: Pay-as-You-Go Composition.” In Advances in Neural Information Processing Systems 29, 1921–29. Red Hook, NY: Curran.
29. Shokri, Reza, and Vitaly Shmatikov. 2015. “Privacy-Preserving Deep Learning.” In

Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 1310–21. New York: ACM.
30. Sweeney, Latanya. 2002. “k-Anonymity: A Model for Protecting Privacy.” International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (5): 557–70.
31. Tang, Liang, Yexi Jiang, Lijun Zhang, and Tao Li. 2014. “Personalized Recommendation via Parameter-Free Contextual Bandits.” In Proceedings of the 37th International ACM SIGIR Conference
on Research and Development in Information Retrieval, 323–32. New York: ACM.

32. Thakurta, Abhradeep, Andrew Thakurta, and Om Thakkar. 2020. “Differentially Private Learning Needs Better Features (or Much More Data).” In Proceedings of the 37th International Conference on Machine Learning, 9334–45.