Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 1 (2023)

Federated Learning; Privacy Preserving Machine Learning for Decentralized Data

[1] Jitendra Singh Chouhan, [2] Amit Kumar Bhatt, [3] Nitin Anand

[1] Asst. Professor
Information Technology
Arya Institute of Engineering and Technology, Jaipur
[2] Asst. Professor
Computer Science Engineering
Arya Institute of Engineering, Technology and Management, Jaipur
[3] Research Paper
Computer Science Engineering
Arya Institute of Engineering and Technology, Jaipur

Abstract: Federated learning represents a compelling solution for tackling the privacy challenges inherent in decentralized and distributed environments when it comes to machine learning. This scholarly paper delves deep into the realm of federated learning, encompassing its applications and the latest privacy-preserving techniques used for training machine learning models in a decentralized manner. We explore the reasons behind the adoption of federated learning, highlight its advantages over conventional centralized approaches, and examine the diverse methods employed to safeguard privacy within this framework. Furthermore, we scrutinize the current obstacles, unresolved research queries, and the prospective directions within this rapidly developing field

1. Introduction:

"The field of machine learning and artificial intelligence has seen significant expansion and utilization in various fields. However, the widespread adoption of these technologies has been impeded by substantial concerns regarding the privacy and security of data. Traditional centralized machine learning models typically involve the collection of substantial amounts of sensitive data from multiple origins, resulting in worries about privacy, data breaches, and regulatory issues. In contrast, federated learning provides a decentralized and privacy-conscious method for training machine learning models. It allows the training process to occur across dispersed devices or data sources while maintaining the data's local nature."

2. Motivation for Federated Learning:

Privacy Preservation: "Federated learning is primarily driven by the goal of safeguarding data privacy. It achieves this by permitting individual devices to perform local model updates, reducing the necessity to share raw data, all while facilitating enhancements to the global model."

Decentralization: "Federated learning is particularly suitable for situations in which data originates from various dispersed devices or entities. It minimizes the requirement to transmit data to a central server, a process that is not only inefficient but also poses potential security vulnerabilities."

3. Key Concepts of Federated Learning:

"Model Aggregation in federated learning entails the process of training models on individual devices and then periodically combining the updates to create a global model. Secure and privacy-preserving aggregation methods, including Federated Averaging and Secure Multi-Party Computation (SMPC), are employed for this purpose."

"Differential Privacy is employed in federated learning to improve privacy by introducing random noise to the model updates. This safeguards against the possibility of extracting an individual user's data from the combined updates."

4. Privacy-Preserving Techniques in Federated Learning:

- "Differential Privacy is utilized in federated learning to enhance privacy by adding random noise to the model updates, preventing the potential extraction of an individual user's data from the aggregated updates."
- "Secure Aggregation utilizes Secure Multi-Party Computation (SMPC) protocols to combine model updates in a way that guarantees the confidentiality of both participants' raw data and their individual model updates."
- "Federated Learning incorporating Differential Privacy involves introducing noise to the model updates during the aggregation process, thereby ensuring a specific degree of privacy safeguard."

5. Applications of Federated Learning:

- "Federated learning is employed in collaborative healthcare research, allowing model training using patient data while safeguarding individual medical records from exposure."
- "Within the realm of Financial Services, banks and financial organizations learning to enhance fraud detection models while upholding the privacy of their customers."
- "In the realm of IoT devices, federated learning is implemented directly on edge devices and IoT sensors to optimize the process of updating models, all while avoiding the need to transmit the raw sensor data."

6. Challenges and Future Directions:

- "Scalability is a concern in federated learning when it comes to accommodating a substantial number of devices or participants while still upholding privacy and efficiency."
- "Enhancing the effectiveness of secure aggregation methods is essential for their practical implementation."
- "Ongoing research is focused on enhancing the resilience of federated learning models against adversarial attacks and promoting fairness in model outcomes."

7. Conclusion:

"Federated learning stands as a noteworthy development in the realm of privacy-focused machine learning for decentralized models. It presents a hopeful approach to tackling the issues associated with safeguarding data privacy while allowing the training of machine learning models using data from multiple distributed sources. The ongoing refinement of federated learning is crucial for its extensive utilization across diverse applications, but certain obstacles still need to be overcome."

References:

[1] Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492.

[2] Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., ... & Ray, J. (2017). Towards federated learning at scale: System design. arXiv preprint arXiv:1610.06733.

[3] Shokri, R., Stronati, M., Song, C., Shmatikov, V., & Witchel, E. (2017). Membership inference attacks against machine learning models. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 3-18). IEEE.

[4] Hard, A., Rao, K., Mathews, R., Ramaswamy, S., Beaufays, F., Augenstein, S., & Eichner, H. (2018). Federated learning for mobile keyboard prediction. arXiv preprint arXiv:1811.03604.

[5] Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC conference on computer and communications security (pp. 1310-1321).

[6] McMahan, H. B., Ramage, D., Talwar, K., & Zhang, L. (2017). Learning differentially private recurrent language models. arXiv preprint arXiv:1710.06963.

[7] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1-19.

[8] Carlini, N., & Wagner, D. (2018). Audio adversarial examples: Targeted attacks on speech-to-text. In 2018 IEEE Security and Privacy (SP) (pp. 76-93). IEEE.

[9] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (pp. 308-318).

[10] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Transactions on Big Data, 7(3), 1656-1674.

[11] Mohassel, P., & Zhang, Y. (2017). SecureML: A system for scalable privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 521-538).

[12] Li, J., Yu, L., Zhang, X., Zeng, Y., & Yang, Q. (2020). Hybrid federated learning for edge devices. arXiv preprint arXiv:2002.08502.

[13] Papernot, N., Song, S., Mironov, I., Raghunathan, A., Talwar, K., & Erlingsson, Ú. (2018). Scalable private learning with PATE. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (pp. 886-901).

[14] Hard, A., Recht, B., & Singer, Y. (2018). Train on in-domain data, and generalize from source domain data: A benchmark. In Proceedings of the 35th International Conference on Machine Learning (ICML) (Vol. 80, pp. 1982-1991).

[15] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Ramage, D. (2019). Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977.