

A Comprehensive Investigation of Federated Unlearning: Challenges, Methods and Future Prospects in Privacy-Sensitive Applications

Wei Zhang^a

Electrical Engineering, Chongqing University, Chongqing, China

Keywords: Federated Unlearning, Privacy Preservation, Data Removal.


Abstract: This paper reviews a range of federated unlearning techniques, with a focus on their applications, limitations, and potential benefits. Federated unlearning addresses privacy concerns by enabling the removal of specific data from machine learning models without requiring full retraining. This is particularly relevant in complying with legal regulations like General Data Protection Regulation (GDPR). Methods like FedEraser and FedCIO provide effective data removal by partitioning and clustering data, making them suitable for handling complex, non-independent and identically distributed (Non-IID) data. FedRecovery offers high precision by storing and rolling back model gradient updates, while other approximate methods such as F2UL optimize computational efficiency through differential privacy, striking a balance between privacy and performance. The analysis reveals the trade-offs between these exact and approximate methods, with the former ensuring better data removal precision but at a higher computational cost, and the latter being more resource-efficient but involving potential privacy risks. It can be concluded that future research should focus on developing standardized evaluation metrics, improving computational efficiency, and enhancing the adaptability of federated unlearning techniques to better manage Non-IID data in real-world applications. This research aims to guide advancements in federated unlearning, promoting its application in dynamically adaptive, privacy-sensitive machine learning scenarios.

1 INTRODUCTION

Federated Learning (FL) is a privacy-preserving distributed machine learning framework that enables multiple participants to collaboratively train models in local clients without sharing raw data. With the intensification of the "Right to be Forgotten" legal requirements, removing specific data from FL models has become increasingly critical. Federated Unlearning (FU), one of the key techniques addressing this challenge, was first explored in 2021 by Liu et al. through the FedEraser method, which efficiently removes specific clients' data without requiring a full retraining process (Liu et al., 2021). This approach significantly enhances the scalability of federated systems. However, challenges remain, particularly in maintaining the model's utility while effectively eliminating specific user data (Liu et al., 2024).

Most existing unlearning research focuses on centralized systems that rely on rapid retraining or approximate unlearning to remove data. When applied to FL, however, unique challenges arise, including data privacy concerns, high computational costs, and the complexity of efficiently eliminating contributions from individual clients in distributed environments (Nguyen et al., 2024; Wu et al., 2022). Additionally, handling non-independent and identically distributed (Non-IID) data, which is common in FL, further complicates the process of federated unlearning (Wang et al., 2022; Zhao et al., 2024).

Federated unlearning techniques not only comply with regulations like the General Data Protection Regulation (GDPR) but also reduce the computational and time costs of retraining models (Wu et al., 2022; Gong et al., 2023). Moreover, these techniques enhance model adaptability, proving valuable in applications requiring frequent updates,

^a <https://orcid.org/0009-0007-1988-2874>

such as medical diagnostics and financial risk management (Su et al., 2024; Halimi et al., 2023).

Current federated unlearning methods include rapid retraining (e.g., FedEraser) (Qiu et al., 2023), approximate unlearning (Wang et al., 2022), and particle-based Bayesian methods (Gong et al., 2023). Each method offers trade-offs between precision, computational cost, and privacy protection. While rapid retraining preserves model utility, it often incurs high costs. Approximate methods like differential privacy reduce resource consumption but may leave residual data traces (Zhao et al., 2024). Particle-based methods, though promising in balancing performance and efficiency, are more suited for approximate rather than exact unlearning (Nguyen et al., 2024).

In conclusion, this paper systematically reviews federated unlearning techniques, evaluating both exact and approximate methods such as FedCIO, FedRecovery, and F2UL. These insights aim to guide future research, promoting advancements in federated unlearning and supporting its broader application in privacy-sensitive, dynamically adaptive machine learning scenarios (Zhang et al., 2023; Su et al., 2024).

2 METHODS

2.1 Exact Federated Unlearning

2.1.1 Sharding and Isolation

By partitioning the dataset and training sub-models for each shard, it ensures that the global model no longer retains the influence of removed data. The FedCIO method (Qiu et al., 2023) combines clustering and one-shot aggregation to completely eliminate the influence of target data, especially in scenarios involving non-independent and identically distributed (Non-IID) data, making it well-suited for federated learning systems with complex data distributions.

2.1.2 Model Parameter Updates

Methods based on model parameter updates aim to precisely remove the influence of target data by selectively updating model parameters. FedRecovery (Zhang et al., 2023) achieves precise data removal by storing gradient updates and rolling back related parameters, while also incorporating differential privacy to protect data privacy. This approach is

particularly suitable for applications that require high-precision data removal and privacy protection.

2.1.3 Reverse Gradients

Reverse gradient methods use stochastic gradient ascent (SGA) optimization to precisely remove the influence of specific data from the model. The framework proposed by Wu et al. (Wu et al., 2022) calculates gradients and applies reverse optimization to eliminate the contribution of target data. The innovative aspect lies in the application of reverse optimization to achieve complete data removal, making it suitable for scenarios that require high precision and seamless unlearning.

2.1.4 Snapshot and Reconstruction

By saving snapshots of the model at different stages during training and reconstructing the model during unlearning, this method enables rapid removal of specific data influence. FedEraser (Liu et al., 2021) preserves multiple model snapshots during training and rolls back these snapshots upon receiving an unlearning request to ensure precision in unlearning. This approach is particularly suitable for systems that require fast responses to unlearning requests. FedRecovery (Zhang et al., 2023) also uses a similar approach by storing and rolling back gradient updates to ensure that the model behaves as if it had never seen the target data.

2.2 Approximate Federated Unlearning

2.2.1 Privacy Preservation

Differential privacy is used to protect user privacy by adding noise during the unlearning process, providing an approximate solution for data removal. The "Privacy-Preserving Federated Unlearning" (Liu et al., 2024) introduces noise during unlearning to protect user privacy while reducing the need for retraining, making it suitable for scenarios with high privacy requirements and limited computational resources. The "Ferrari" method by Gu et al. (Gu et al., 2024) combines feature sensitivity optimization within a differential privacy framework to remove the influence of target data features efficiently, simplifying preprocessing and ensuring effective unlearning. This approach is ideal for scenarios requiring fast unlearning and high feature privacy.

2.2.2 Efficient Parameter Updates

By selectively updating certain model parameters, computational costs associated with full retraining can be reduced. The “Update Selective Parameters” method proposed by Xu et al. (Xu et al., 2024) leverages model interpretation to selectively update parameters for data removal. The innovation lies in the combination of model interpretation with selective updates, ensuring a balance between efficiency and effectiveness. This approach is particularly useful for resource-constrained applications requiring efficient unlearning.

2.2.3 Model Architectural Adjustments

This method adjusts the model architecture, such as pruning techniques, to reduce the dependency on specific data, thereby achieving approximate unlearning. The “Federated Unlearning via Class-Discriminative Pruning” (Wang et al., 2022) removes the model’s dependency on specific class data by pruning, with the workflow including pruning and parameter adjustment, significantly reducing the computational cost of retraining. It is particularly suitable for scenarios that require rapid removal of the influence of specific class data.

3 DISCUSSIONS

This study introduces various Federated Unlearning (FU) methods, categorized into exact unlearning and approximate unlearning. Exact unlearning methods, such as FedCIO and FedRecovery, focus on thoroughly removing the influence of specific data from the model, which is crucial in scenarios requiring strict privacy protection. On the other hand, approximate unlearning methods (such as those based on differential privacy and model architecture adjustments) strike a balance between computational efficiency and privacy protection, making them more suitable for scenarios with relatively lower privacy requirements (Su et al., 2024).

3.1 Advantages and Disadvantages of Existing Methods

3.1.1 Precision and Privacy Protection

Exact unlearning methods, such as FedCIO and FedRecovery, can completely remove the influence of specific data, rendering the resulting model statistically indistinguishable from one that has never

seen the data. This is crucial in complying with privacy regulations such as the General Data Protection Regulation (GDPR). Additionally, FAST enhances system security by quickly removing malicious terminals at the server side, making it particularly advantageous in privacy-sensitive sectors like finance and healthcare (Huynh et al., 2024; Guo et al., 2024; Wang et al., 2022). However, the downside of exact unlearning methods is their high computational and storage costs. Methods like FedRecovery require storing and rolling back model updates, which significantly limits scalability as the model and the number of participants grow. This issue becomes even more pronounced in federated learning environments that handle non-independent and identically distributed (Non-IID) data, where the computational burden increases considerably. Furthermore, the resource demands make it challenging to maintain efficiency in real-time applications.

3.1.2 Computational Efficiency and Applicability

Approximate unlearning methods reduce computational complexity by optimizing model updates or reducing reliance on the global model (Xu et al., 2024). For example, F2UL optimizes feature sensitivity combined with differential privacy, enabling the unlearning process to respond quickly with minimal computational cost (Su et al., 2024). Similarly, the Update Selective Parameters method, as discussed by Xu et al. (Xu et al., 2024), further lowers the cost of retraining by selectively updating model parameters based on their contributions to model performance, which enhances both computational efficiency and real-time responsiveness (Xu et al., 2024). However, approximate unlearning methods pose certain risks to privacy protection. Although these methods are computationally efficient, they may not completely eliminate data traces, making them vulnerable to inference attacks (Nguyen et al., 2024; Liu et al., 2022). For instance, Momentum Degradation performs well in terms of computational efficiency, but in some cases, residual data traces may still exist, posing privacy risks (Halimi et al., 2023; Dinsdale et al., 2022).

3.1.3 Interpretability and Transparency

Some methods improve interpretability and transparency, which enhances users’ trust in the system. For example, Backdoor Unlearning can clearly identify and remove malicious patterns,

ensuring a highly transparent unlearning process, which is particularly crucial in security-sensitive applications (Dhasade et al., 2023). This interpretability is especially important in sectors like healthcare and finance, where data sensitivity is high. However, despite the higher interpretability of some methods, many federated unlearning techniques lack standardized evaluation metrics, making it difficult to systematically assess the completeness of the unlearning process and the residual influence of data (Tang et al., 2024; Ding et al., 2024). This lack of evaluation standards limits the broader application of these methods across different scenarios.

3.2 Challenges and Limitations

Federated unlearning faces several challenges in practical applications. First, methods like FedRecovery exhibit poor scalability due to the need to store and manage a large volume of historical gradient updates, leading to significant resource consumption in large-scale federated networks (Huynh et al., 2024; Dinsdale et al., 2022). Second, existing methods struggle with handling Non-IID data. Most approaches assume independent and identically distributed data, but in real-world scenarios, client data often vary significantly in both quantity and distribution (Guo et al., 2024; Dinsdale et al., 2022). For example, Fast-FedUL offers a training-free rapid federated unlearning solution specifically designed to address challenges associated with uneven data distribution. Moreover, there is a lack of unified evaluation standards for assessing the effectiveness of federated unlearning methods. Current evaluation methods often rely on heuristics, making it difficult to ensure consistency and fairness across different application scenarios (Zuo et al., 2024).

3.3 Future Prospects

Future research should focus on several key areas. First, unified evaluation metrics should be developed to assess the success and completeness of federated unlearning processes (Nguyen et al., 2024). Standardized metrics would allow researchers to systematically compare the performance of different methods and enhance their practical application. Second, improving the computational efficiency of both exact and approximate unlearning methods should be a research priority. By introducing more efficient model update strategies and storage optimization techniques, resource consumption associated with exact unlearning can be significantly

reduced (Wang et al., 2022). Lastly, enhancing the adaptability of federated unlearning techniques to handle Non-IID data is critical. By incorporating techniques such as transfer learning, domain adaptation, and clustering, future federated unlearning methods will be better equipped to handle the complex data distributions found in real-world scenarios (Zuo et al., 2024).

4 CONCLUSIONS

This paper has reviewed various federated unlearning techniques, analyzing their advantages, limitations, and potential applications. Exact unlearning methods, such as FedCIO, effectively remove data influence by partitioning and clustering data, making them suitable for handling complex Non-IID data distributions. FedRecovery, by storing and rolling back model gradient updates, offers a high-precision data removal solution, particularly applicable in privacy-sensitive domains. However, the computational complexity and scalability of these methods remain significant challenges. In contrast, the Update Selective Parameters method reduces computational costs by selectively updating model parameters, making it ideal for resource-constrained environments. Additionally, F2UL combines feature sensitivity optimization with differential privacy, enhancing computational efficiency without compromising privacy protection. While these approximate methods improve efficiency, they may involve trade-offs in terms of privacy protection.

Future research should focus on the following areas: first, developing unified evaluation metrics to assess the success and completeness of federated unlearning processes. Second, improving the computational efficiency of both exact and approximate unlearning methods by optimizing model update strategies and storage mechanisms. Lastly, enhancing federated unlearning techniques to better handle Non-IID data by incorporating transfer learning and clustering techniques to address uneven data distributions in practical scenarios.

REFERENCES

- Dhasade, A., Ding, Y., Guo, S., Kermarrec, A., De Vos, M., & Wu, L. 2023. QuickDrop: Efficient Federated Unlearning by Integrated Dataset Distillation. arXiv Preprint.
- Ding, N., Wei, E., & Berry, R. 2024. Strategic Data Revocation in Federated Unlearning. In IEEE

- INFOCOM 2024 - IEEE Conference on Computer Communications (pp. 1151–1160). IEEE.
- Dinsdale, N. K., Jenkinson, M., & Namburete, A. I. L. 2022. FedHarmony: Unlearning Scanner Bias with Distributed Data. In L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, & S. Li (Eds.), *Medical Image Computing and Computer Assisted Intervention (MICCAI 2022)*, Part VIII (Vol. 13438, pp. 695–704). Springer.
- Gong, J., Simeone, O., & Kang, J. 2023. Compressed Particle-Based Federated Bayesian Learning and Unlearning. *IEEE Communications Letters*, 27(2), 556–560.
- Gu, H., Ong, W., Chan, C. S., & Fan, L. 2024. Ferrari: Federated Feature Unlearning via Optimizing Feature Sensitivity. *arXiv Preprint*.
- Guo, X., Wang, P., Qiu, S., Song, W., Zhang, Q., Wei, X., & Zhou, D. 2024. FAST: Adopting Federated Unlearning to Eliminating Malicious Terminals at Server Side. *IEEE Transactions on Network Science and Engineering*, 11(2), 2289–2302.
- Halimi, A., Kadhe, S., Rawat, A., & Baracaldo, N. 2023. Federated Unlearning: How to Efficiently Erase a Client in FL?. *arXiv Preprint*. <https://arxiv.org/abs/2207.05521>
- Huynh, T. T., Nguyen, T. B., Nguyen, P. L., Nguyen, T. T., Weidlich, M., Nguyen, Q. V. H., & Aberer, K. 2024. Fast-FedUL: A Training-Free Federated Unlearning with Provable Skew Resilience. In *Proceedings of Machine Learning and Knowledge Discovery in Databases. Research Track* (pp. 55–72). Springer Nature Switzerland.
- Liu, G., Ma, X., Yang, Y., Wang, C., & Liu, J. 2021. FedEraser: Enabling Efficient Client-Level Data Removal from Federated Learning Models. In *Proceedings of the 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQoS)* (pp. 1–10). IEEE.
- Liu, Y., Xu, L., Yuan, X., Wang, C., & Li, B. 2022. The Right to be Forgotten in Federated Learning: An Efficient Realization with Rapid Retraining. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications* (pp. 1749–1758). IEEE.
- Liu, Z., Jiang, Y., Shen, J., Peng, M., Lam, K.-Y., Yuan, X., & Liu, X. 2024. A Survey on Federated Unlearning: Challenges, Methods, and Future Directions. *ACM Computing Surveys*, Article 3679014.
- Liu, Z., Ye, H., Jiang, Y., Shen, J., Guo, J., Tjuawinata, I., & Lam, K.-Y. 2024. Privacy-Preserving Federated Unlearning with Certified Client Removal. *arXiv Preprint*.
- Nguyen, T.-H., Vu, H.-P., Nguyen, D. T., Nguyen, T. M., Doan, K. D., & Wong, K.-S. 2024. Empirical Study of Federated Unlearning: Efficiency and Effectiveness. In *Proceedings of the 15th Asian Conference on Machine Learning* (pp. 959–974).
- Qiu, H., Wang, Y., Xu, Y., Cui, L., & Shen, Z. 2023. FedCIO: Efficient Exact Federated Unlearning with Clustering, Isolation, and One-shot Aggregation. In *Proceedings of the 2023 IEEE International Conference on Big Data* (pp. 5559–5568). IEEE.
- Su, W., Kang, B., Zhao, X., & Zhang, Y. 2024. F2UL: Fairness-Aware Federated Unlearning for Data Trading. *IEEE Transactions on Mobile Computing*, 1–16.
- Tang, Y., Zhao, S., Chen, H., Li, C., Zhai, J., & Zhou, Q. 2024. Fuzzy Rough Unlearning Model for Feature Selection. *International Journal of Approximate Reasoning*, 165, 109102.
- Wang, J., Song, G., Xin, X., & Heng, Q. 2022. Federated Unlearning via Class-Discriminative Pruning. In *Proceedings of the ACM Web Conference 2022* (pp. 622–632). ACM.
- Wu, L., Guo, S., Wang, J., Hong, Z., Zhang, J., & Ding, Y. 2022. Federated Unlearning: Guarantee the Right of Clients to Forget. *IEEE Network*, 36(5), 129–135.
- Xu, H., Zhu, T., Zhang, L., Zhou, W., & Yu, P. S. 2024. Update Selective Parameters: Federated Machine Unlearning Based on Model Explanation. *IEEE Transactions on Big Data*, 1–16.
- Zhang, L., Zhu, T., Zhang, H., Xiong, P., & Zhou, W. 2023. FedRecovery: Differentially Private Machine Unlearning for Federated Learning Frameworks. *IEEE Transactions on Information Forensics and Security*, 18, 4732–4746.
- Zhao, Y., Wang, P., Qi, H., Huang, J., Wei, Z., & Zhang, Q. 2024. Federated Unlearning with Momentum Degradation. *IEEE Internet of Things Journal*, 11(5), 8860–8870.
- Zuo, X., Wang, M., Zhu, T., Zhang, L., Yu, S., & Zhou, W. 2024. Federated Learning with Blockchain-Enhanced Machine Unlearning: A Trustworthy Approach. *arXiv Preprint*.