Article

Federated Unlearning in Financial Applications

Cassandra Lindstrom *

Posted Date: 24 September 2024

doi: 10.20944/preprints202409.1816.v1

Keywords: federated unlearning; graph neural network; digital asset; cryptocurrency; financial market

Article

Federated Unlearning in Financial Applications

Cassandra Lindstrom

Independent Researcher; cli1194@bloomberg.net

Abstract: Federated unlearning represents a sophisticated evolution in the domain of machine learning, particularly within federated learning frameworks. In financial applications, where data privacy and security are paramount, federated unlearning allows institutions to selectively remove or "unlearn" specific data from trained models without compromising their overall performance and accuracy. This capability is essential for ensuring that models remain adaptable, secure, and compliant with regulatory requirements, while minimizing the need for expensive retraining. In this paper, we explore various financial applications where federated unlearning can have a significant impact, including fraud detection, portfolio management, and credit risk modeling. By allowing targeted removal of outdated, erroneous, or sensitive data, federated unlearning enhances the agility of financial models, enabling institutions to keep pace with the dynamic financial landscape. Using practical numerical examples, we demonstrate how unlearning improves model accuracy and decision-making while maintaining data privacy across distributed systems. This paper underscores the critical role of federated unlearning in addressing the challenges of modern financial institutions, offering insights into its practical applications and future potential.

Kyeowrds: federated unlearning; graph neural network; digital asset; cryptocurrency; financial market

1. Introduction

Federated learning (FL) has emerged as a transformative approach for collaborative machine learning across distributed data silos without the need to centralize data. This decentralized model-training approach enables financial institutions, such as banks and insurance companies, to build robust models while maintaining compliance with stringent data privacy laws. However, in light of new data privacy regulations like the General Data Protection Regulation (GDPR), ensuring the right to be forgotten—which mandates complete data removal—is crucial. In this context, federated unlearning represents a significant advancement in the broader domain of machine unlearning, allowing data to be removed from models without necessitating a complete retraining process. This is particularly relevant in finance, where data privacy is paramount, and models must adapt dynamically to fast-evolving market conditions.

In financial applications, federated unlearning must address several core challenges. Primarily, the sensitive nature of financial data, the compliance landscape, and the distributed nature of FL systems present unique difficulties in effectively and securely unlearning specific data contributions. This paper aims to explore the principles, mechanisms, challenges, and practical use cases of federated unlearning, particularly in the financial sector, and propose future directions for research and development in this area.

2. Literature Review

Previous research in federated unlearning provided us with several useful frameworks. [1] has provided a summary of the most relevant research on federated unlearning. [2] presents a framework that manages data forgetting and model accuracy in DTMN through a multi-loss training approach. [3] has a FedRecovery framework that utilizes differential privacy to erase a client's data influence without retraining, ensuring statistical indistinguishability between models. This would a good framework for crypto [4] and treasury trading [5] presented in previous research effort.

 doi:10.20944/preprints202409.1816.v

[6] has a Federated Clusters method that accelerates the unlearning process, providing a significant speed-up compared to retraining. [7] uses SFU to perform gradient ascent in orthogonal space of input gradient spaces, negates a target client's contribution without additional storage. [8] use federated variational inference solutions, offering an efficient unlearning mechanism via local free energy minimization. [9] leverages on a particle-based Bayesian unlearning method, providing a non-parametric strategy for federated unlearning. [10] enables verification of the unlearning effect, ensures both execution and accuracy when removal of a participant's gradients from the global model. [11] introduces 2F2L that facilitates the certified data removal by employing a linear approximation.

Graph neural network introduced by [12] is an efficient approach for this kind of unlearning problem. [13] [14] [15] [16] have provided direct examples of applying federated unlearning in blockchain context.

3. Federated Unlearning: Concept and Mechanism

Federated unlearning refers to the process of removing specific data contributions from models in a federated learning setup. In traditional machine learning, when data is removed, the model is typically retrained from scratch. However, in a federated learning context, where data is distributed across multiple clients, the need for unlearning poses additional complexities. Federated unlearning aims to minimize these challenges by offering selective data removal while retaining the model's general utility.

The process involves three main components: the client, the server, and the global model. Each client trains the model locally using their data, and the global model is updated based on aggregated client updates. In the event that a client requests unlearning, federated unlearning ensures that the contributions from the specific data points or entire clients are effectively removed from both local and global models without necessitating a complete retraining. This selective removal preserves computational resources and minimizes disruptions to model accuracy.

3.1. Challenges of Federated Unlearning in Financial Applications

The application of federated unlearning in financial sectors presents distinct challenges. Firstly, financial data is highly sensitive, comprising transaction records, personal identifiable information (PII), and corporate strategies. Unauthorized access or misuse of such data can lead to severe consequences, including financial loss, regulatory penalties, and reputational damage. Federated unlearning must guarantee that once data is removed, it can no longer influence future model updates or predictions.

Furthermore, compliance with data privacy regulations such as GDPR imposes strict requirements on how financial institutions manage and process data. The *right to be forgotten* mandates data controllers to remove personal data upon request. Implementing this in a federated learning context, where data is dispersed across several entities, makes federated unlearning essential yet complex. Additionally, the distributed nature of financial data adds another layer of difficulty, as banks or financial entities operate on heterogenous datasets with varying structures and constraints.

3.2. Federated Unlearning Framework for Financial Data

To implement federated unlearning in financial applications, a clear system architecture is needed. The architecture must support selective data removal from clients while maintaining the integrity of the global model. Federated unlearning frameworks for financial data generally involve:

1. **Client-Specific Models:** Since financial institutions operate under varying risk parameters and market strategies, each client may have different models tailored to their specific needs. Federated unlearning ensures that data unlearning occurs at the client level while synchronizing updates to the global model.

 doi:10.20944/preprints202409.1816.v

2. **Selective Data Removal:** Unlearning in financial applications may focus on removing specific data points (such as an individual's loan repayment history) or entire datasets. This is especially critical when clients, such as retail banks, want to remove specific transactions or user data from their models without affecting the global model's robustness.

3. **Global Model Updates:** After performing local unlearning at the client level, the global model must be updated in a way that reflects the unlearning operation. Techniques like model pruning, reweighting of model parameters, and differential privacy mechanisms are often employed to ensure that unlearned data does not continue influencing the global model.

3.3. Privacy and Security Considerations

Ensuring data privacy and model security is crucial in federated unlearning, especially in financial applications where sensitive information is prevalent. One of the main threats is *privacy attacks*, where adversaries try to extract private information from trained models. In federated unlearning, this risk increases when attempting to reverse the contributions of removed data. Techniques such as *differential privacy* and *secure aggregation* are critical in protecting the privacy of individual clients while enabling the secure removal of their data contributions from the global model.

Additionally, the possibility of *malicious clients* manipulating the federated unlearning process to degrade model performance or inject erroneous updates presents another challenge. Solutions include employing cryptographic methods and robust anomaly detection mechanisms to ensure the integrity of the unlearning process.

4. Optimization Strategies for Federated Unlearning

The trade-off between the computational cost of unlearning and the overall performance of the model is a significant factor in federated unlearning. Unlike retraining, federated unlearning offers optimization strategies that prioritize efficiency. These include:

1. **Model Repair vs. Retraining:** Instead of retraining models from scratch, federated unlearning uses model repair techniques to selectively remove unwanted data. This minimizes the time and resources needed to restore the model to its optimal state.

2. **Communication Efficiency:** Since federated learning involves periodic communication between clients and the server, unlearning can be optimized by reducing the amount of communication overhead involved in sending updates. Compression techniques, such as quantization and sparsification, are often employed to streamline communication in large-scale networks.

Federated unlearning holds significant promise in several key financial applications:

4.1. Fraud Detection

Federated unlearning can help improve fraud detection systems by ensuring that outdated or erroneous fraud patterns are effectively removed from the model without needing complete retraining. This allows financial institutions to adapt to new fraud tactics rapidly. Fraud detection is a critical application in the financial industry, where institutions must continuously monitor transactions for signs of fraudulent activity. The dynamic nature of fraud patterns, which evolve rapidly as fraudsters find new ways to exploit systems, requires machine learning models to remain adaptable and up-to-date. Federated learning allows multiple financial institutions to collaboratively train fraud detection models without sharing sensitive customer data. However, as fraud detection models are continuously updated with new transaction data, there may be cases where erroneous or

 doi:10.20944/preprints202409.1816.v

outdated fraud patterns need to be removed from the model. This is where federated unlearning becomes crucial.

Consider a scenario where a federated learning model is used to detect fraudulent credit card transactions across multiple banks. Initially, the model is trained on a dataset containing thousands of past transactions from various clients, including flagged fraudulent activities. Over time, a particular pattern of transactions is identified as fraudulent and incorporated into the model, but later it is discovered that the pattern belonged to legitimate transactions mistakenly marked as fraud. To rectify this, the model must unlearn these erroneous patterns.

For example, assume the model identifies fraudulent transactions based on specific characteristics such as transaction amount, geographic location, and the time of day. One bank reports that 100 transactions amounting to $10,000 in a particular region were flagged as fraudulent based on the model's current behavior. However, after further investigation, the bank realizes that these transactions were legitimate and must be removed from the model's fraud detection logic.

In a traditional setting, removing these incorrect patterns would require retraining the entire model, which is time-consuming and computationally expensive. However, using federated unlearning, only the contributions from these 100 transactions can be selectively removed from both the local and global models. For instance, if the model was using parameters like P(transaction = fraud ∣ amount > $5000, region = X) to flag transactions, the unlearning process would adjust these parameters so that the model no longer falsely associates these conditions with fraudulent activity.

By unlearning these erroneous transactions, the model is able to update its fraud detection criteria in a more targeted manner. For example, suppose the original fraud detection rate was 95%, with a false positive rate of 2%. After removing the erroneous data through federated unlearning, the false positive rate decreases to 1.5%, while maintaining the same overall fraud detection rate. This improvement not only enhances the accuracy of the model but also ensures that legitimate transactions are not incorrectly flagged, leading to fewer customer complaints and better trust in the financial institution. Federated unlearning thus plays a pivotal role in maintaining the performance and integrity of fraud detection systems, ensuring that models can adapt quickly to new information and correct past mistakes without a full retraining process.

4.2. Portfolio Management

In portfolio management, federated unlearning allows models to dynamically adjust their risk profiles by removing data from certain assets or financial instruments that are no longer relevant, ensuring more accurate and up-to-date decision-making. In the field of portfolio management, investors and financial institutions strive to optimize returns while managing risk by allocating investments across various asset classes, such as stocks, bonds, and cryptocurrencies. Machine learning models play an essential role in this process by analyzing historical performance, predicting future returns, and adjusting portfolio weights based on market trends. In a federated learning framework, multiple financial entities can collaborate on improving portfolio management models without sharing sensitive data. However, these models must adapt to rapidly changing financial environments. When outdated or irrelevant data, such as past performance of underperforming assets, skews model predictions, federated unlearning allows the selective removal of such data, ensuring the model remains accurate and up-to-date without the need for full retraining.

For instance, imagine a portfolio management model that allocates investments across three asset classes: stocks, bonds, and cryptocurrency. Initially, the model assigns 60% of the portfolio weight to Stock A, 30% to Bonds, and 10% to Cryptocurrency, based on recent historical returns. In the previous quarter, Stock A showed a 12% return, Bonds yielded 4%, and Cryptocurrency delivered an 18% return. Using these returns, the model calculates the expected overall portfolio return as:

Expected Portfolio Return = (60% × 12%) + (30% × 4%) + (10% × 18%) = 9.6%

Expected Portfolio Return= (60%×12%) + (30%×4%) + (10%×18%) = 9.6%

This suggests a relatively high return due to the strong past performance of Stock A.

However, after the initial training, new information emerges indicating that Stock A is expected to underperform due to economic factors such as rising interest rates and declining industry growth.

 doi:10.20944/preprints202409.1816.v

To prevent the model from over-allocating to a now-risky asset, federated unlearning is applied to remove the contribution of the outdated data associated with Stock A. Once unlearned, the model adjusts its portfolio allocation to reflect the new market conditions, lowering the weight of Stock A to 30%, increasing the bond allocation to 50%, and the cryptocurrency weight to 20%. Additionally, the return expectations for Stock A are updated to a more modest 3%.

As a result of these adjustments, the expected portfolio return is recalculated:

Updated Portfolio Return = (30% × 3%) + (50% × 4%) + (20% × 18%) = 6.9 %

Updated Portfolio Return=(30%×3%)+(50%×4%)+(20%×18%)=6.9%

This decreases from 9.6% to 6.9% reflects the more realistic future performance of the assets. By using federated unlearning, the portfolio management model can adapt to evolving market conditions and avoid overexposure to risky assets like Stock A, ensuring more accurate investment decisions.

In this context, federated unlearning enhances the flexibility and responsiveness of portfolio management models by allowing them to swiftly adjust to new information. This capability is especially valuable in fast-moving markets, where institutions must continuously fine-tune their strategies to remain competitive while also complying with data privacy regulations and removing outdated or erroneous information from their decision-making processes.

Predictive Modeling in Credit Risk

Credit risk is a critical area of focus for financial institutions, where predictive modeling is used to assess the likelihood of a borrower defaulting on a loan. These models analyze a variety of factors such as income, credit history, debt levels, and economic conditions to assign a credit score or risk level to each borrower. Federated learning enables multiple banks and financial entities to collaboratively improve their predictive models without sharing sensitive customer data. However, as credit conditions evolve or specific borrowers' financial situations change, there may be a need to remove outdated or erroneous data to ensure the accuracy of predictions. This is where federated unlearning becomes valuable, allowing selective removal of certain data contributions without retraining the entire model from scratch.

For example, consider a credit risk model that uses data from several banks to predict the probability of loan default for different borrowers. Initially, the model incorporates factors like a borrower's credit score, monthly income, outstanding debt, and recent payment history to determine their credit risk. Suppose the model was trained on a dataset where Borrower A had a credit score of 750, a monthly income of $5,000, and an outstanding debt of $20,000. Based on these factors, the model assigns BorrowerA a low default probability of 2%.

The model may predict default probabilities as follows:

Default Probability for Borrower A=2%

Default Probability for Borrower B=5%

Default Probability for Borrower C=12%

Over time, new information becomes available indicating that Borrower A recently lost their job, drastically affecting their ability to repay the loan. This new financial situation is not yet reflected in the data used by the model, and the bank wants to update the model by unlearning the previous favorable data for Borrower A to avoid assigning them a low credit risk. In a traditional setting, the entire model would need to be retrained with updated data, but with federated unlearning, the contributions from Borrower A's previous data can be removed efficiently.

After unlearning the outdated data and incorporating the updated financial information (e.g., loss of income), the model recalculates Borrower A's credit risk: Default Probability for Borrower A=15%

This new prediction indicates a significantly higher risk of default based on the borrower's current financial condition. Similarly, the model can adjust predictions for other borrowers as new data becomes available, allowing the institution to more accurately assess and manage credit risk across its portfolio.

By using federated unlearning, the model is able to remove inaccurate or outdated data, such as old income information for Borrower A, without compromising the overall integrity of the predictive

 doi:10.20944/preprints202409.1816.v

model. This targeted unlearning process ensures that the model remains up-to-date and reflects the latest risk factors without the need for complete retraining. As a result, financial institutions can better manage credit risk, improving loan underwriting decisions, and reducing the likelihood of defaults, all while maintaining compliance with privacy regulations by keeping sensitive borrower data secure.

In conclusion, federated unlearning enhances the adaptability of credit risk models by allowing financial institutions to selectively update or remove information. This capability enables them to respond quickly to changes in borrowers' financial situations, improving the precision of credit risk assessments and ensuring that outdated or incorrect data does not skew predictions.

5. Future Directions

The future of federated unlearning will likely see the integration of emerging technologies such as *artificial intelligence* and *blockchain.* AI can enhance federated unlearning by providing more advanced algorithms for data removal and model repair, while blockchain offers immutable audit trails to ensure compliance with data privacy regulations. As federated learning continues to evolve, federated unlearning will play a critical role in ensuring that machine learning systems remain adaptive, secure, and privacy compliant. Federated unlearning can be combined with feature extraction and model optimization in deep learning as show in [17] and [18]. Self-supervised learning would further enhance the federated learning with proper optimization techniques [18].

6. Conclusions

Federated unlearning presents a crucial advancement in the domain of federated learning, particularly for financial applications where data privacy and security are paramount. By enabling the precise removal of data from distributed models, federated unlearning ensures compliance with regulatory mandates such as GDPR while maintaining the performance and integrity of financial models. However, significant challenges remain, including optimizing unlearning efficiency, safeguarding against malicious attacks, and navigating the complexities of distributed data. Future research must continue to explore innovations in this field, particularly in the areas of AI-driven unlearning and secure, scalable implementations tailored for financial use cases.

References

1. N. Li, C. Zhou, Y. Gao, H. Chen, A. Fu, Z. Zhang and Y. Shui, "Machine Unlearning: Taxonomy, Metrics, Applications, Challenges, and Prospects," arXiv preprint arXiv:2403.08254, 2024.
2. H. Xia, S. Xu, J. Pei, R. Zhang, Z. Yu, W. Zou, L. Wang and C. Liu, "Fedme 2: Memory evaluation & erase promoting federated unlearning in dtmn," IEEE Journal on Selected Areas in Communications, 2023.
3. L. Zhang, T. Zhu, H. Zhang, P. Xiong and W. Zhou, "Fedrecovery: Differentially private machine unlearning for federated learning frameworks," IEEE Transactions on Information Forensics and Security, 2023.
4. Z. Li, B. Wang and Y. Chen, "A Contrastive Deep Learning Approach to Cryptocurrency Portfolio with US Treasuries," Journal of Computer Technology and Applied Mathematics, vol. 1, pp. 1-10, 2024.
5. Z. Li, B. Wang and Y. Chen, "Incorporating economic indicators and market sentiment effect into US Treasury bond yield prediction with machine learning," Journal of Infrastructure, Policy and Development, vol. 8, p. 7671, 2024.
6. C. Pan, J. Sima, S. Prakash, V. Rana and O. Milenkovic, "Machine unlearning of federated clusters," arXiv preprint arXiv:2210.16424, 2022.
7. G. Li, L. Shen, Y. Sun, Y. Hu, H. Hu and D. Tao, "Subspace based federated unlearning," arXiv preprint arXiv:2302.12448, 2023.
8. J. Gong, O. Simeone and J. Kang, "Bayesian variational federated learning and unlearning in decentralized networks," in 2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2021.
9. J. Gong, J. Kang, O. Simeone and R. Kassab, "Forget-svgd: Particle-based bayesian federated unlearning," in 2022 IEEE Data Science and Learning Workshop (DSLW), 2022.
10. X. Gao, X. Ma, J. Wang, Y. Sun, B. Li, S. Ji, P. Cheng and J. Chen, "Verifi: Towards verifiable federated unlearning," IEEE Transactions on Dependable and Secure Computing, 2024.

 doi:10.20944/preprints202409.1816.v

11. R. Jin, M. Chen, Q. Zhang and X. Li, "Forgettable federated linear learning with certified data removal," arXiv preprint arXiv:2306.02216, 2023.
12. Z. Wang, Y. Zhu, Z. Li, Z. Wang, H. Qin and X. Liu, "Graph neural network recommendation system for football formation," Applied Science and Biotechnology Journal for Advanced Research, vol. 3, no. 3, p. 33–39, 2024.
13. J. Zhu, J. Cao, D. Saxena, S. Jiang and H. Ferradi, "Blockchain-empowered federated learning: Challenges, solutions, and future directions," ACM Computing Surveys, vol. 55, p. 1–31, 2023.
14. Y. Li, C. Chen, N. Liu, H. Huang, Z. Zheng and Q. Yan, "A blockchain-based decentralized federated learning framework with committee consensus," IEEE Network, vol. 35, p. 234–241, 2020.
15. Y. Qu, M. P. Uddin, C. Gan, Y. Xiang, L. Gao and J. Yearwood, "Blockchain-enabled federated learning: A survey," ACM Computing Surveys, vol. 55, p. 1–35, 2022.
16. A. Qammar, A. Karim, H. Ning and J. Ding, "Securing federated learning with blockchain: a systematic literature review," Artificial Intelligence Review, vol. 56, p. 3951–3985, 2023.
17. Y. Wei, X. Gu, Z. Feng, Z. Li and M. Sun, "Feature Extraction and Model Optimization of Deep Learning in Stock Market Prediction," Journal of Computer Technology and Software, vol. 3, 2024.
18. H. Zhao, Y. Lou, Q. Xu, Z. Feng, Y. Wu, T. Huang, L. Tan and Z. Li, "Optimization Strategies for Self-Supervised Learning in the Use of Unlabeled Data," Journal of Theory and Practice of Engineering Science, vol. 4, p. 30–39, 2024.