Proceedings of Business and Economic Studies, 2025, Volume 8, Issue 2
http://ojs.bbwpublisher.com/index.php/PBES
Online ISSN: 2209-265X
Print ISSN: 2209-2641

The Technological Progress, Applications, and Challenges of Federated Learning

Yanling Liu*, Yun Li

Hainan Vocational University of Science and Technology, Haikou 570100, Hainan, China

**Author to whom correspondence should be addressed.*

Abstract: With the advent of the era of big data, the exponential growth of data generation has provided unprecedented opportunities for innovation and insight in various fields. However, increasing privacy and security concerns and the existence of the phenomenon of "data silos" limit the collaborative utilization of data. This paper systematically discusses the technological progress of federated learning, including its basic framework, model optimization, communication efficiency improvement, privacy protection mechanism, and integration with other technologies. It then analyzes the broad applications of federated learning in healthcare, the Internet of Things, Internet of Vehicles, smart cities, and financial services, and summarizes its challenges in data heterogeneity, communication overhead, privacy protection, scalability, and security. Finally, this paper looks forward to the future development direction of federated learning and proposes potential research paths in efficient algorithm design, privacy protection mechanism optimization, heterogeneous data processing, and cross-industry collaboration.

Keywords: Federated learning; Data privacy; Distributed machine learning; Heterogeneous data

Online publication: April 28, 2025

1. Introduction

In the era of big data, the exponential growth of data generation has brought unprecedented opportunities for innovation and insight across all fields. However, this surge in data has also heightened privacy and security concerns. Many organizations face the problem of "data silos," where data is dispersed and stored across different entities or jurisdictions, limiting its value for use in collaborative analytics. This data silos often stem from regulatory frameworks such as the General Data Protection Regulation (GDPR) and growing concerns about individual privacy rights, which collectively limit the free exchange [1,2] of sensitive information. In this context, how to realize the efficient use of data while protecting privacy has become an important direction of current research.

Federated learning emerged as a transformative approach to machine learning. It is a distributed machine learning framework designed to train models collaboratively through local datasets on multiple decentralized devices or servers. Unlike traditional centralized approaches, federated learning ensures that the raw data always remains on the local device, with only model updates (such as gradients or parameters) shared to a central server for aggregation. This architecture minimizes privacy risks while enabling efficient model training [3]. In addition, federated learning incorporates secure aggregation technology that enables model updates to be merged without exposing the contributions of a single participant [4].

This article aims to explore the technological advances, application scenarios, and challenges of federated learning. By analyzing the applications of federated learning in healthcare, recommendation systems, IoT environments, and generative modeling, this paper demonstrates the potential of federated learning to drive data-driven innovation while protecting privacy. In addition, the unsolved challenges of federated learning in system scalability, communication efficiency, and vulnerability to adversarial attacks will be discussed [1–3].

2. Technical advances in federated learning

2.1. Basic framework of federal learning

Federated learning is a distributed machine learning paradigm that aims to train models collaboratively through decentralized devices while protecting data privacy. The typical architecture of federated learning revolves around the federated averaging (FedAvg) algorithm proposed by Google, and its main flow consists of the following steps:

Local training: Each client trains the local model on its private data set based on global model parameters received from a central server.

Model upload: After local training is complete, the client sends its model updates (such as gradients or weights) to a central server.

Aggregate updates: The central server aggregates model updates from the clients through techniques such as weighted averaging and updates the global model, which is then redistributed to all clients.

This iterative process continues until the model converges. The framework ensures that the original data always remains on the local device, thus effectively addressing privacy concerns and data silos.

2.2. Direction of technology development

2.2.1. Model optimization

Improvements to FedAvg algorithm:

Given the limitations of FedAvg on non-independent co-distributed (non-IID) data, improved algorithms such as FedProx and FedUB are proposed. By introducing update bias into the loss function, FedUB makes the local and global optimal solutions more consistent, thus improving convergence and generalization [5,6].

Techniques such as adaptive data sampling (such as FAST) better approximate global optimal solutions by adjusting local training strategies and accelerate convergence in heterogeneous environments [7].

Optimization for heterogeneous data:

Methods such as HeteroFair introduce fairness constraints into the loss function and mitigate bias caused by non-independently homo distribution data by reweighting aggregation [6].

Non-aggregative methods such as FedAF avoid client drift by leveraging peer knowledge between clients rather than direct aggregation and perform better in data environments with skewed labels or features [8].

2.2.2. Communication efficiency

Model compression techniques:

Techniques such as pruning, quantification, and knowledge distillation are widely used to reduce communication overhead while maintaining model performance [9]. For example, FedSQ combines sparsity and quantization techniques and introduces an error compensation mechanism to maintain model performance while achieving high compressibility [10].

Gradient compression:

Gradient sparsity and coding methods such as Wyner-Ziv coding, are used to reduce bandwidth requirements during the model update process while striking a balance between accuracy and communication costs [11]. Over-the-air federated learning, combined with gradient compression, has also emerged as a promising solution in bandwidth-constrained environments such as wireless networks [12].

2.2.3. Privacy protection

Differential privacy (DP):

Differential privacy technology ensures privacy protection while maintaining model utility by adding calibration noise before model updates. These methods are already widely used in fields such as healthcare and the Internet of Things [13].

Homomorphic encryption (HE) and secure multi-party computation (SMPC):

Homomorphic encryption allows computations to be performed directly on encrypted data without decryption; Secure multi-party computing ensures that collaborative computing between multiple parties is secure. These techniques are increasingly used in privacy-sensitive fields such as medical diagnostics and financial analysis [13].

2.2.4. Personalized federal learning

Personalized model training:

Personalized federated learning frameworks such as SPIDER optimized the neural network structure for each client to accommodate the data distribution of heterogeneous clients through Neural Architecture Search (NAS) methods [14].

Metrics-based fuzzification techniques, such as d-privacy, enhance personalization in diverse user populations while protecting privacy [15].

2.3. Integration with other technologies

Blockchain integration: Blockchain-based frameworks such as BFLPP decentralize verification of updates through smart contracts and committee consensus mechanisms, enhancing the credibility of federated learning and enabling secure aggregation without a central server [14].

Reinforcement learning (RL): Reinforcement learning is integrated into federated learning to optimize resource allocation, especially in energy-constrained environments such as IoT networks and satellite systems, improving energy efficiency and model performance [15].

3. Applications and challenges of federated learning

3.1. Applications of federated learning

Federated learning has been widely used in many fields because of its ability to realize collaborative model training while protecting data privacy. Here are some of the key application areas:

(1) Medical field

Federal learning is widely used in healthcare to address data privacy regulations and data silos. For example, Federated learning enables hospitals and institutions to collaborate on training models for disease prediction, diagnosis, and treatment planning without sharing sensitive patient data. A case in point is the use of the federal network HONEUR, which supports clinical data analysis across multiple hospitals while ensuring local data governance [1].

(2) Internet of Vehicles (IoV)

In the field of IoV, federated learning is advancing the development of intelligent transportation systems by supporting collaborative training models between vehicles and edge devices. These models can be used for traffic prediction, autonomous driving, and safety applications while avoiding exposure of raw sensor data.

(3) Smart cities and Internet of Things (IoT) systems

Federal learning plays an important role in smart city infrastructure, supporting decentralized learning for applications such as energy management (such as smart grids), public safety (such as surveillance systems), and urban planning (such as traffic flow optimization). These systems take advantage of federated learning's ability to process locally generated data on edge devices while maintaining privacy. In an IoT environment, federated learning reduces network overhead and enhances system scalability by aggregating knowledge of distributed sensors or devices.

(4) Financial services

Federal learning has seen increasing use in fraud detection, credit scoring, and personalized financial services. It enables banks and financial institutions to collaborate on training models without sharing sensitive customer data.

3.2. Challenges of federal learning

Although federated learning shows great potential in multiple areas, it still faces numerous challenges in technology and practice:

(1) Data heterogeneity

A major challenge in federated learning is dealing with data that is not independently co-distributed (non-IID) between clients. Differences in data distribution can lead to model bias or slower convergence. To address this, advanced optimization techniques such as adaptive aggregation or personalized federated learning strategies for individual client data need to be employed.

(2) Communication overhead

The iterative nature of federated learning results in a frequent exchange of model updates between clients and central servers, introducing significant communication costs. Techniques such as model compression, gradient sparsity, and over-the-air computing have been proposed to mitigate this problem but remain an active area of research.

(3) Privacy concerns

Although federated learning improves privacy by keeping raw data on local devices, it remains vulnerable to threats such as model inversion attacks or member inference attacks. Techniques such as differential privacy, homomorphic encryption, and secure multi-party computing can enhance privacy protection but often require trade-offs between model accuracy and computational efficiency [1].

(4) Scalability

When extending federated learning to a large number of clients with different computing power, there are problems such as resource allocation, fault tolerance, and efficient aggregation. Techniques such as hierarchical federated learning (such as using edge servers as intermediaries) are being explored but still need to be further optimized.

(5) Security threats

Federated learning systems are vulnerable to adversarial attacks such as poisoning attacks (malicious updates) or Byzantine failures (unreliable clients). Robust aggregation methods such as Krum or Trimmed Mean are designed to mitigate these risks but generally add computational complexity.

(6) Have regulatory compliance

While federal learning addresses many privacy concerns, there are still regional differences in regulatory frameworks to contend with when deployed globally. Ensuring compliance with laws and regulations such as GDPR or HIPAA requires incorporating legal considerations into the design of federal systems.

4. Future directions

Federated learning's technical development direction is mainly focused on improving algorithm efficiency, enhancing privacy protection mechanisms, and better dealing with heterogeneous data problems. Efficient algorithm design is the focus of future research, especially in resource-constrained environments such as edge computing and the Internet of Things. Lightweight design and energy saving (LDES) algorithms significantly reduce energy consumption through sparse or binary neural networks, while methods such as selective model aggregation (SAM) strike a balance between communication efficiency and model performance by probabilistically selecting clients to participate in model updates. In terms of privacy protection, although technologies such as differential privacy (DP) and homomorphic encryption (HE) have been widely used, the risk of data leakage in model updating still needs to be further addressed. Future research should develop hybrid privacy protection frameworks that integrate techniques such as HE, DP, and secure multi-party computing (SMPC) to address diverse threat models while improving their efficiency in real-time applications.

In the direction of application development, the potential of federated learning is not limited to existing fields such as healthcare, the Internet of Things, and autonomous driving but can be further extended to scenarios such as video analytics, intrusion detection systems (IDS), and 6G networks. In addition, cross-industry collaboration brings new opportunities for federated learning, such as collaborative models in healthcare and finance that can use financial data to predict health risks or personalize financial planning through health data. A federal learning framework incorporating blockchain technology also enables trust-free collaboration within the IoT ecosystem, ensuring secure model aggregation.

5. Conclusion

This paper systematically reviews the technological progress, application scenarios, and challenges of federated learning. At the technical level, federated learning provides innovative solutions for distributed machine learning through efficient algorithm design, communication optimization, and privacy protection mechanisms while showing potential in dealing with non-IID data and heterogeneous environments. At the application level, federated learning has been widely used in the fields of healthcare, the Internet of Things, the Internet of Vehicles, etc., promoting data-driven innovation. However, federated learning still faces issues such as data heterogeneity, communication overhead, privacy protection tradeoffs, and security threats. Future research should focus on efficient algorithm design, development of hybrid privacy protection frameworks, and optimization of dynamic aggregation strategies while exploring cross-industry collaboration and emerging application scenarios.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Shin H, Ryu K, Kim J, et al., 2024, Application of Privacy Protection Technology to Healthcare Big Data. *Digital Health*, 10.

[2] ElZemity A, Arief B, 2024, Privacy Threats and Countermeasures in Federated Learning for Internet of Things: A Systematic Review. 2024 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics, Copenhagen, Denmark, 331–338.

[3] Song Y, Wu Y, Xue Y, 2024, University Major Information Recommendation based on Federated Learning. *Applied and Computational Engineering, 40: 280–287*.

[4] Gargary AV, De C, 2024, A Systematic Review of Federated Generative Models. *arXiv. https://doi.org/10.48550/arXiv.2405.16682*

[5] Zhang H, Zhang P, Hu M, et al., 2024, FedUB: Federated Learning Algorithm Based on Update Bias. *Mathematics*, 12(10): 1601.

[6] Li Y, Zhang J, Zhao Y, et al., 2024, Fairness Aware Federated Learning Framework on Heterogeneous Data Distributions. *ICC 2024 IEEE International Conference on Communications*, 728–733.

[7] Wang Z, Xu H, Xu Y, et al., 2024, FAST: Enhancing Federated Learning Through Adaptive Data Sampling and Local Training. *IEEE Transactions on Parallel and Distributed Systems*, 35: 221–236.

[8] Wang Y, Fu H, Kanagavelu R, et al., 2024, An Aggregation Free Federated Learning for Tackling Data Heterogeneity, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 26233–26242.

[9] Yan S, 2024, Optimizing Federated Learning Efficiency: A Comparative Analysis of Model Compression Techniques for Communication Reduction. *Applied and Computational Engineering, 107: 107–117*.

[10] Long Z, Chen Y, Dou H, et al., 2024, FedSQ: Sparse Quantized Federated Learning for Communication Efficiency. *IEEE Transactions on Consumer Electronics*, 70: 4050–4061.

[11] Sharma M, Kaur P, 2023, An Empirical Study of Gradient Compression Techniques for Federated Learning. *2023 Second International Conference on Informatics (ICI)*, 1–4.

[12] Ruan M, Li Y, Zhang W, et al., 2024, Optimal Power Control for Over-the-Air Federated Learning with Gradient Compression. *2024 IEEE 30th International Conference on Parallel and Distributed Systems (ICPADS),* 326–333.

[13] Dasari J, Joshith TS, Daya Lokesh, et al., 2023, Privacy Preserving Sensitive Data on Medical Diagnosis Using Federated Learning and Homomorphic Re-encryption. *2023 3rd International Conference on Intelligent Technologies (CONIT),* 1–7.

[14] Zhang S, Zhu J, 2023, Privacy Protection Federated Learning Framework Based on Blockchain and Committee Consensus in IoT Devices. *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, 627–636.

[15] Zhou A, Wang Y, Zhang Q, 2024, Energy-Efficient Resource Management for Federated Learning in LEO Satellite IoT. *2024 IEEE Wireless Communications and Networking Conference (WCNC),* 1–6.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.