Federated Learning for Privacy Preservation and Energy Efficiency Optimization in IoT End Devices

Hongyu Mao *

School of Electronic and Information Engineering, Southwest University, Chongqing, 400715, China

* Corresponding Author Email: outlook_F51C95D78588B7F9@outlook.com

Abstract. Due to the large-scale deploying of the Internet of Things (IoT) end devices, data collection and transmission are suffering from serious privacy leakage threats, while the limited computing power and battery capability of terminal devices are also causing energy efficiency bottleneck. Due to this, as a distributed machine learning paradigm where data "do not leave the local area", Federated Learning (FL) has application significance in solving the aforementioned issues. This paper primarily carries out research on the design of FL privacy preserving mechanism and energy efficiency optimization mechanism in IoT end-device. It explains theoretical models of FL, IoT end-device, and privacy-preserving technology, and discusses four optimization directions of FL privacy preserving from the aspects of model lightening, terminal-edge joint training, privacy preservation energy efficiency optimization and dynamic resource scheduling. The work gives theory guidance for safe and efficient IoT end devices' running, so that these end devices can be more practical implemented with a large number.

Keywords: FL, IoT End Devices, Privacy Preservation, Energy Efficiency Optimization, Lightweight Model.

1. Introduction

With the rapid development of the IoT, end devices like smart sensors and wearable devices have become the bridge for data collection and intelligent interaction in various fields in large scale. Meanwhile, the massive data produced by the IoT terminals contains a lot of sensitive information, and the traditional centralized data processing mode suffers from high privacy leakage risks in data transmission and storage [1]. Simultaneously, the IoT end devices are usually restricted with limited computation capabilities, storage capacity and power supply of battery, which leads to significant energy efficiency bottleneck when executing heavy-weight machine learning models. Such problems greatly constrain the green sustainable development for the IoT ecosystem.

FL has become a promising approach because it allows model training in a distributed manner that the distributed data does not need to share directly, and, in particular, it could further protect the privacy of data while alleviating communication overhead due to huge amount of data transfer. Nevertheless, FL in IoT terminal devices still has challenges to be solved: traditional FL models with high parameter size cannot fit into the available computation resources at the terminals, and iterative parameter aggregation incurs huge amounts of energy consumption [2]. Thus, studying privacy preserving schemes and energy efficiency optimization on FL in the IoT terminal devices is very meaningful in reality.

The present paper is divided into two major purposes: privacy-preserving approaches which are adequate to deploy in IoT agents and energy optimal approaches for FL deployment. From the underlying theory of associated technologies and the exploration of common cases, an ideal technical framework for privacy-preserving and energy efficient IoT FL is set to facilitate the IoT industry.

2. Theoretical Foundation Analysis

2.1. FL

FL is a distributed machine learning paradigm with a training mode of "no data flows from local to remote", whose central idea is an iterative process: IoT end devices locally train their own local models on their own data and upload these trained model parameters to a server. The server merges these local model parameters. Consequently, this research updates the global model by aggregating the trained local models across all terminals and we broadcast it to all terminals for further local learning [3].

The operation logic is based on cooperative communication between the cloud center and terminal machines. FedAvg and FedProx are two classic algorithms, the former of which adopts a primitive weighted average method for parameters aggregation and the latter of which adds a proximal term to deal with devices' data heterogeneity. The cooperative mode ensures local devices to store sensitive data and the training of a high-quality global model.

2.2. IoT End Devices

IoT end device is a lightweight device with sensing, computing and communication capabilities, like smart senor, wearable equipment, smart household equipment. Their working principle is hierarchical: environment or user data captured by sensors are collected in the data collection layer, simple data cleaning and feature extraction are implemented by edge preprocessing layer locally, and data processed or model parameters are transferred to edge nodes or central servers by network transmission layer [4].

IoT terminal operation logic is low-power operation: Because of hardware limitations, IoT terminals run low-power during operation because of a limited battery capacity and computing resources. They have to determine the coordination of performance and energy during operation because of battery capacity and computing resources limitations. At the same time, multi-device coexistence is adopted during data operation to implement data interaction. IoT devices present a distributed IoT network to execute some relatively complex sensing and computing tasks.

2.3. Privacy-Preserving Technologies

Privacy-Preserving Technologies is a technical system providing the guaranteed non-access by an unauthorized party to the data during its processing. These privacy-preserving technologies are mainly based on cryptographic techniques, differential privacy, homomorphic encryption, etc. and using these techniques to build privacy barriers. Take, for instance, differential privacy that injects very small noise into data or model parameters so as to prevent the leakage of individual's sensitive data, and homomorphic encryption that enables computation of encrypted data in the absence of decryption.

In FL, the privacy-preserving technologies work out the operation logic, by incorporating the privacy protecting modules into the training of FL. These modules complement local train and parameter aggregation stages: in the local training, differential privacy can be leveraged to inject noise to the model updates, in parameter transmission, homomorphic encryption can be used to encrypt the parameters, so that the data remains inaccessible even if it is captured [5].

2.4. Energy efficiency

Energy efficiency is an indispensable issue in the FL process. In terms of computing, communication energy, and endurance capability, there is a substantial demand for energy. However, when the available energy is confined within a certain range, it becomes necessary to reduce ineffective energy consumption—such as cutting down on invalid computation and excessively frequent communication [6]. By reducing inefficiencies either in terminal operations or computational workload, high energy efficiency can be achieved.

3. Optimization Strategy Analysis

3.1. FL Model Lightweight Optimization

The key factor for this optimization is the low computing power of the IoT end device; conventional FL has big model size, too long training latency time and consuming enormous amount of energy. Multiple research teams have made practical solutions in this area.

Google's Group proposed FedLite algorithm, which has a purpose of decreasing the quantity of terminal model parameters via model pruning. The experiments of smartwatch data indicate that the energy consumption has been decreased by 35%. while the model performance can be held at a controllable level. Knowledge distillation has been used in Huawei Research Institute to construct lightweight FL sub-models. Compared with the original, the training speed is 40% faster in smart home devices [7].

Adaptive quantization strategy is developed at Stanford University. The model weights are quantized to 8-bit which can lower the energy consumption of IoT sensor for 28% without obvious accuracy degradation. Alibaba DAMO Academy discloses the hierarchical model architecture based on Federated Learning. Only the shallow network is trained on terminals, reducing the network delay of logistics tracking terminal devices by 50% [8].

For fault diagnosis of CNC machine tools, based on redundant computational structures, this research constructs a lightweight one-dimensional convolutional neural network (1D-CNN) model. Improve the traditional FedAvg algorithm based on industrial data heterogeneity and node quality difference to ensure the stability of diagnosis accuracy at 87.5%, higher than comparably ILDCNN (77.40-86.08%). Convergence speed is accelerated 66% to 75% compared to the classic FedAvg(121 rounds) and Fedora (83 rounds). Overall training time is 1525.2s, while the average time per round is 49.2 seconds which is orders of magnitude less than that of classic algorithms [9].

3.2. Terminal-Edge Collaborative Federated Training

Due to the insufficient computing power of individual IoT terminals, edge nodes can assist in undertaking part of the computing tasks, achieving a balance between privacy preservation and energy efficiency. This collaborative training mode reduces the computational burden on terminals and shortens training time.

Microsoft Research Asia proposed the Edge-FL framework, where edge nodes are responsible for intermediate feature computation, and terminals only perform parameter updates. Application in industrial IoT devices resulted in a 42% reduction in energy consumption. Baidu Smart Cloud constructed a two-level federated architecture of "terminal-edge", where edge nodes aggregate parameters from local terminals, reducing cross-regional communication volume and lowering communication energy consumption by 38% in urban traffic sensor networks [10].

3.3. Privacy-Enhanced Energy Efficiency Optimization

Nevertheless, even if FL makes the data "not out of local area", the parameter transmission still has a risk of leakage of privacy. Thus, we should consider the trade-off between privacy protection and energy efficiency, and make sure that our privacy protection will not raise an overly large energy consumption.

University of California integrated differential privacy and model compression, injecting small noise on terminal parameters and pruning redundant parameters. For applications in health monitoring devices, it satisfied the privacy budget requirements from GDPR and also resulted 30% of energy savings. In Chinese Academy of Sciences Institute of Computing Technology, in the case that the terminal performs a lot of complex decrypted operations, a lightweight improved homomorphism encryption scheme, which let the edge nodes to make parameters addition in cipher texts was applied. Accordingly, it reduces the operational energy by 33%, in the financial IoT terminals [11].

NUS presented federated learning's privacy-sensitive data selection and schedule to prioritize low privacy sensitive data for training, enhanced overall resource utilization performance by 25% in retail

IoT devices. Amazon AWS designed federated transfer learning technique, where terminals can recycle privacy-protected pre-trained knowledge provided by edge nodes, decreasing training energy by 36% in Agricultural IoT sensor devices. Terminal selection of edge nodes on terminals with low power consumption and high computing power to train [12].

Edge nodes pick terminals that have most contributed to model convergence (i.e., terminals with high local loss) to speed up training and indirectly the total energy consumption. For example, in the power-of-choice scheme, edge nodes permit terminals with the highest local loss to contribute in training, by reducing the number of rounds for model convergence by 40% and the total energy consumption by 30%. In addition to computing task allocation: Edge nodes allocate tasks according to the difference in computing power of the terminal [13].

3.4. Dynamic Resource Scheduling and Energy Efficiency Adaptation

The working status of IoT end devices changes dynamically, with variations in remaining battery power, network bandwidth, and computing load. Thus, FL strategies need to be adjusted adaptively to avoid energy overload and ensure stable operation.

Xiaomi, a branch of Xiaomi's AIoT team, proposed a battery-powered FL participation mechanism according to battery power threshold. When the power of the terminal is lower than 20%, the training stops, and the battery life of smart home devices could be extended to 12h [14]. Samsung Electronic introduces a network bandwidth adaptive adjustment scheme, whose parameter is uploaded with a low frequency when the low bandwidth occurs, and reduces the variability of communication energy consumption in 5G IoT terminals by 45% [15].

Tsinghua University introduced an energy expenditure forecast algorithm for Federated Learning, and according to the previous data, the batch size of the model's training phase was adaptively adjusted to optimize the energy expenditure during the training process, where the industrial sensor was more stable in terms of energy consumption [16]. The OPPO Research Institute built an "energy efficiency-accuracy" trade-off model, so that the terminal can choose a low-precision training mode according to actual needs when training based on application scenarios. Reduced energy consumption by 22% on smart bracelet health monitoring tasks by controlling the accuracy loss to less than 5% [17].

4. Discussion and Analysis

4.1. Advantages of Current Optimization Strategies

FL can provide useful benefits for privacy preservation and for energy efficiency optimization on IoT end device side. It overcomes the data silo issue by performing distributed training without using raw data sharing, hence it preserves user privacy. The proposed optimization methods target at alleviating energy efficiency bottleneck of terminals: model lightening can decrease the computation and energy consumption of terminals, terminal-edge collaboration can collaborate the edge computing resource to achieve a balance performance-energy consumption, privacy preserving optimization can make the tasks both secure and efficient work, dynamic resources scheduling can adapt to dynamic features of IoT devices.

Both of these approaches were also validated in realistic scenarios and have resulted in significant savings of energy consumption (between 22% and 42%), as well as enhancement in the training efficiency while the privacy requirement is met. They are viable approaches for large-scale deployment of FL-based IoT systems.

4.2. Limitations of Current Research

Although the existing studies have achieved remarkable advancements, there are still certain research bottlenecks. One way of optimizing lightweight model can cause a certain loss of accuracy, how to determine accuracy and energy consumption reasonably still need more investigation.

Terminal-edge collaborative training still depends too much on the coverage of edge nodes computing power, remote or undeveloped regions still may not have enough computing power of edge nodes.

There is still some more work to be optimized that privacy-preserving technologies like homomorphic encryption also bring more computational overhead, lightweight needs to be further optimized and the dynamic resource scheduling strategies do not suit for highly heterogeneous IoT devices, because all the heterogeneous IoT devices may use different types of terminals with various hardware configuration and working characteristics.

4.3. Future Research Directions

There are three future research directions. One is how to design adaptive FL frameworks for heterogeneous IoT terminals in view of the device hardware configurations and application scenarios, making the FL framework adapt to them through customized models and strategies. The other one is how to apply the AI large models for the prediction of FL energy efficiency. Some large models can be applied to the FL terminal energy efficiency prediction in order to gain a good accuracy to predict the terminal's energy consumption in advance and adaptively adjust the corresponding FL training strategies. Third, explore quantum privacy-preserving technologies utilizing quantum computing advantages for improved security and lower overhead privacy protection in FL. Furthermore, the cross-layer optimization, also bring better comprehensive performance for FL-based IoT systems.

5. Conclusion

This paper concentrates on federated learning for IoT end devices' privacy protection and energy efficiency optimization. Through discussion on the theoretical basis of FL, IoT end devices, and privacy-preserving technologies, it thoroughly summarizes four main optimization tactics as follows: model lightening, terminal-edge joint training, privacy-enriched energy efficiency optimization, and dynamic resource scheduling.

The aforementioned approaches focus on the specific privacy problem and energy-efficient optimization problem in the process of federated learning, such as lightweight learning model, the cloud-edge transfer, the quality of terminal equipment, etc. From this point of view, this paper solves some of the main issues in federated learning and provides the basis for a deeper development. The Future work can be done in adaptive framework architecture, AI-based energy efficiency forecasting and enhance privacy preserving mechanism for further enhancing the growth of a secure and efficient IoT industry. The current work gives out the theoretical foundation, experimental supporting facts and other necessary references for the work in the future.

References

[1] Alahmari S, Alghamdi I. A Comprehensive Survey on Energy-Efficient and Privacy-Preserving Federated Learning for Edge Intelligence and IoT. Results in Engineering, 2025: 107849.

[2] Badr M, Mahmoud M E A, Fang Y, et al. Privacy-preserving and communication-efficient energy prediction scheme based on federated learning for smart grids. IEEE Internet of Things Journal, 2023, 10 (9): 7719 - 7736.

[3] Asad M, Moustafa A, Ito T. Fedopt: Towards communication efficiency and privacy preservation in federated learning. Applied Sciences, 2020, 10 (8): 2864.

[4] Fang C, Guo Y, Hu Y, et al. Privacy-preserving and communication-efficient federated learning in internet of things. Computers & Security, 2021, 103: 102199.

[5] Zhao T, Chen X, Sun Q, et al. Energy-efficient federated learning over cell-free IoT networks: Modeling and optimization. IEEE Internet of Things Journal, 2023, 10 (19): 17436 - 17449.

[6] Zhao Y, Zhao J, Jiang L, et al. Privacy-preserving blockchain-based federated learning for IoT devices. IEEE Internet of Things Journal, 2020, 8 (3): 1817 - 1829.

[7] Hou M W, Niu C. Governance of Privacy Protection for Medical Education Big Data: A Technical Review and Strategy Study. China Medical Education Technology, 2025: 1 – 12.

[8] Li W. Cross-Domain Collaborative Defense for Public Security Systems: A Two-Layer Incentive Privacy Computing Mechanism. Network Security Technology and Application, 2025 (11): 114 – 116.

[9] Lu W D, Feng K, Ding Y, et al. Research on Security, Privacy and Energy Efficiency of UAV-Assisted Federated Edge Learning Communication Systems. Journal of Electronics & Information Technology, 2025, 47 (05): 1322 – 1331.

[10] Li X Y, Yang Z H, Huang C W, et al. 6G Endogenous Intelligent Wireless Large Models: Security, Privacy, Ethics and High Energy Efficiency. Mobile Communications, 2025, 49 (01): 59 – 66.

[11] Yan K, Shu N, Wu T, et al. A Survey of Energy Efficiency Strategies for Federated Learning in Mobile Edge Computing. Frontiers of Information Technology & Electronic Engineering, 2024, 25 (05): 645 – 664.

[12] Villegas-Ch W, Gutierrez R, Navarro A M, et al. Optimizing federated learning on TinyML devices for privacy protection and energy efficiency in IoT networks. IEEE Access, 2024.

[13] Xu L Y, Lu Y, Zhao J. Research on Federated Learning for Fault Diagnosis of CNC Machine Tools Based on Edge-Cloud Collaboration. Acta Metrologica Sinica, 2024, 45 (06): 873 – 880.

[14] Gong M G, Gao Y, Wang J Q, et al. Adaptive Federated Learning Algorithm Based on Evolutionary Strategy. Scientia Sinica Informationis, 2023, 53 (03): 437 – 453.

[15] Yang S Y. Federated Learning Algorithm for User Heterogeneity and Privacy Protection. Harbin: Heilongjiang University, 2025.

[16] Awan K A, Din I U, Almgren A, et al. Privacy-preserving big data security for IoT with federated learning and cryptography. IEEE Access, 2023, 11: 120918 - 120934.

[17] Zhang J L, Guo X, Zhang H. Privacy-Preserving Federated Averaging on Heterogeneous Data. Acta Mathematica Sinica, English Series, 2025: 1 – 20. Available from: