ISSN: 2664-9640

Federated Learning for Handwritten Digit Recognition with Differential Privacy

Yi Lin*

School of Electronic and Information Engineering, Wuyi University, Jiangmen 529020, China

*Corresponding Author: liny31668@gmail.com

Abstract

With the rapid development of Internet technology, Federated Learning (FL) has attracted widespread attention in the field of machine learning as an emerging technology. FL allows model training to be performed on multiple decentralized devices or servers without sharing raw data, which is crucial for protecting user privacy. This study focuses on Horizontal Federated Learning (HFL), which is suitable for scenarios where user data exhibits highly overlapping features. We propose a FL-based deep learning approach to achieve handwritten digit recognition while preserving user privacy and data security. By conducting experiments on the MINIST public dataset, we compare the advantages and disadvantages of local models, centralized models, and FL methods based on HFL. The experimental results show that our approach not only effectively protects user privacy but also exhibits good model generalization performance, providing valuable insights for advancing privacy protection and data security.

Keywords

Federated Learning; Privacy Protection; Data Security.

1. Introduction

With the rapid development of Internet technology, people around the world are closely connected through virtual platforms, providing vast space for the exchange and dissemination of information. In this era of information explosion, individuals have become the primary providers of data. After preprocessing, this data is available for analysis and processing to meet various needs. With the continuous advancement of computer technology, digital patterns such as text, handwriting, and signatures are transmitted via the Internet in fields such as finance and manufacturing systems. Therefore, ensuring both personal privacy and identification accuracy becomes particularly important during data transmission.

This study explores the integration of privacy protection with handwritten recognition. Federated Learning [1-3] is an emerging architecture in computer technology aimed at safeguarding the security of data exchange between different users while training global machine learning models. Depending on the distribution characteristics of the data, Federated Learning is divided into Horizontal Federated Learning (HFL) and Vertical Federated Learning (VFL). HFL focuses on all users sharing the same feature space and is suitable for scenarios with significant overlap in data feature dimensions. On the other hand, VFL emphasizes training global models using different feature spaces. Due to its privacy-preserving advantages during machine learning model training, Federated Learning has extensive applications in fields such as finance, manufacturing systems, telecommunications, and healthcare.

The application of privacy protection methods in Federated Learning is becoming increasingly important. Mainstream privacy protection methods are primarily based on Differential Privacy (DP) [4-5] and Homomorphic Encryption (HE) [6-8]. Differential Privacy protects user privacy

ISSN: 2664-9640

by introducing noise during the training process. On the other hand, Homomorphic Encryption processes data through complex mathematical computations, thus not exposing raw data during the training process. The choice of privacy protection methods lies in balancing the relationship between privacy and model accuracy. Handwritten digit recognition has extensive applications in finance and industrial fields. However, the privacy of information, accuracy of recognition, and limitations of the proposed data make it challenging. Due to the widespread application of Federated Learning, federated frameworks become an ideal choice to overcome these challenges. This study focuses on Federated Learning-based handwritten digit recognition, including three key technologies: Horizontal Federated Learning, privacy protection, and digital classification techniques. Our goal is to establish a Horizontal Federated Machine Learning framework to promote information security while performing digital classification tasks.

2. Federated Learning

Federated Learning (FL) is a machine learning framework, also known as federated machine learning or collaborative learning. It allows multiple institutions to use data and machine learning modeling while meeting requirements for user privacy protection, data security, and government regulations. As a distributed machine learning paradigm, FL effectively addresses the data silo problem, enabling participants to collaboratively model without sharing data, thus technically breaking down data silos and enabling AI collaboration. Reza Shokri and Vitaly Shmatikov designed, implemented, and evaluated a practical collaborative deep learning system that can improve model accuracy using other users' data without sharing private data. Parameter sharing during model training is effective, as the stochastic gradient descent algorithm can run in parallel and asynchronously [9]. Experimental results show that the system achieves an accuracy of 99.14% on the MNIST dataset and 93.12% on the SVHN dataset. This system can be used for classification and recognition of various complex data, addressing privacy concerns in commercial or industrial enterprises.

Yuncheng Wu et al. proposed a method called "Pivot for Vertical Federated Learning," which is not influenced by third parties and ensures no data leakage during training and prediction [10]. Pivot utilizes threshold partially homomorphic encryption and multiparty computation as core algorithms. Experimental results show that models based on Pivot achieve accuracy and efficiency comparable to non-private models, effectively reducing data leakage issues in collaborative projects between financial institutions and technology companies.

IKai Wang et al. proposed a new task transfer framework called "Federated Transfer Learning Domain Conversion Prediction (FTL-DCP)" [11]. The goal of FTL-DCP is to train models with limited data while protecting user data privacy. Experimental results show that FTL-DCP achieves good accuracy of 97.04% and 89.05% on the source and target domains, respectively. This technology has wide applications in the industrial sector and performs well on public datasets, making it suitable for image recognition tasks. Handwritten digit recognition has extensive applications in finance and industry, but challenges such as privacy, accuracy, and data limitations exist. A federated learning-based framework is a good choice to address these challenges.

3. Privacy Protection

With the widespread application of Federated Learning (FL) in various collaborative projects, privacy protection becomes increasingly important. Currently, mainstream privacy protection methods in Federated Learning are primarily based on two technologies: Differential Privacy and Homomorphic Encryption. Differential Privacy protects user privacy by adding noise during the training process. Li et al. [12] proposed an intelligent metering system based on

ISSN: 2664-9640

Differential Privacy. In this system, each smart meter adds a certain amount of noise before sending data to the supplier. The aggregated noise is sufficient to provide privacy protection for each user while maintaining high model accuracy. Experimental results show that this method effectively hides users' actual activities. However, in situations with small amounts of data, the impact of noise is significant, leading to deviations from accurate values that render the results unusable.

Another approach is to use complex mathematical computations to process data, thus not revealing raw data during training. Phong found that local data information might be leaked to an honest but curious server [13-14]. To address this drawback, they combined asynchronous stochastic gradient descent with additive homomorphic encryption into neural networks. Results show that the proposed model can maintain deep learning accuracy without leaking privacy. Fang and Qian [15] designed a multi-parity privacy protection machine learning framework based on homomorphic encryption and federated learning, named PFMLP. This framework provides privacy protection and model security for users participating in model training. Testing PFMLP on the MNIST and metal datasets achieved accuracy of over 80% on both datasets. Due to the advantages of DP, such as ease of implementation, minimal accuracy impact, and fast speed, this paper considers it as a privacy protection method.

4. System Model

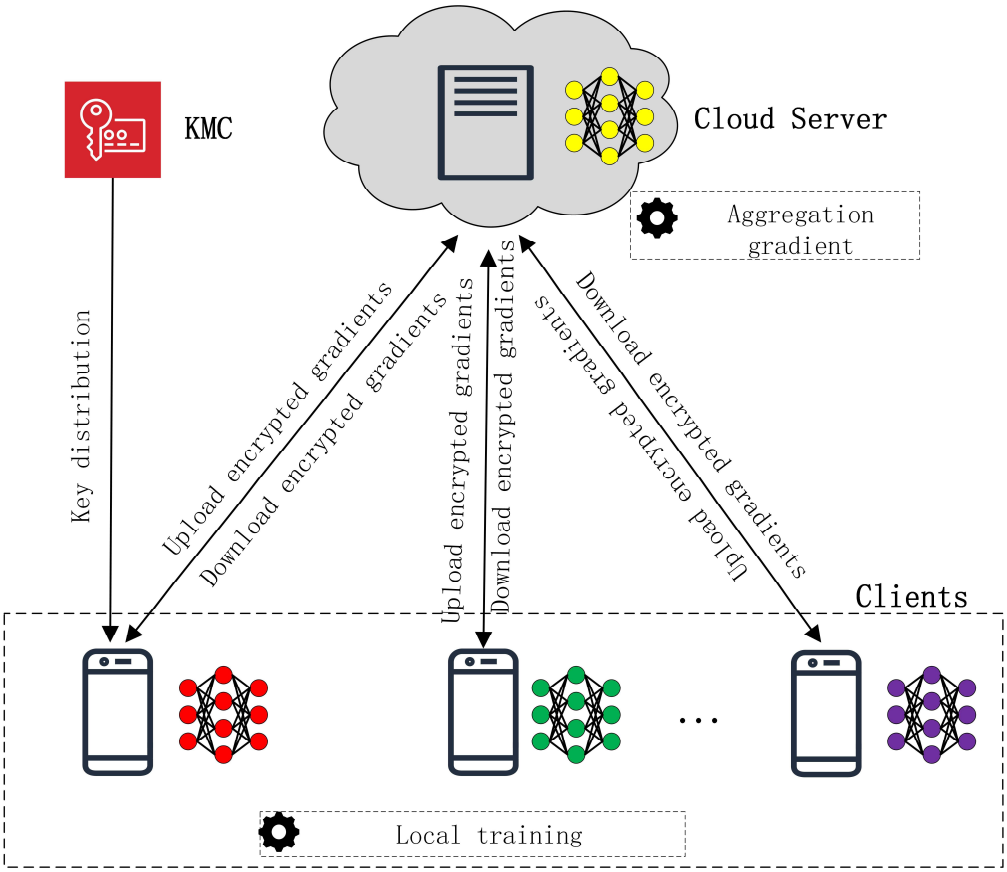

Figure 1. System model

The system model, as shown in Figure 1, is an architecture for image classification and privacy protection based on Horizontal Federated Learning. In this architecture, N users with the same data structure collaborate to learn a classification model. The general process can be summarized into five steps: gradient computation, gradient encryption, secure aggregation, update return, and model update. Detailed explanations of each step are as follows:

ISSN: 2664-9640

(1) Gradient Computation: N users compute training gradients separately based on global parameters provided by the server. No user is allowed to leak information to the server or transmit information between users.

(2) Gradient Encryption: N users mask the original gradients using Differential Privacy and send the masked results to the server. The purpose of this step is to protect information privacy without affecting model accuracy.

(3) Secure Aggregation: The server aggregates encrypted gradients received from users. Thus, the server can aggregate information from N users without knowing any user information.

(4) Update Return: The server returns the aggregated gradients to users. Before starting training, each user obtains global model parameters from the server. Therefore, each user can obtain global information without directly sharing their data.

(5) Model Update: Users decrypt the gradients aggregated by the server and update their respective models. This process loops until the global model converges. This architecture is independent of specific machine learning algorithms, such as logistic regression, artificial neural networks, etc. Finally, all users share the same model parameters.

5. Experimental

5.1. Experimental Setup

The MNIST dataset originates from the National Institute of Standards and Technology (NIST) in the United States. This dataset contains 70,000 handwritten digit images, with 60,000 images used for training and 10,000 images for testing. The training set consists of digits handwritten by 250 different individuals, with 50% from high school students and the other 50% from Census Bureau employees. The testing set also consists of handwritten digit data in the same proportions, but the authors of the testing set do not intersect with the authors of the training set. As shown in Figure 2, each image consists of a 28x28 pixel grayscale image representing a single digit from 0 to 9. The images are presented in black-on-white format, with black represented by the digit 0 and white represented by floating-point numbers between 0 and 1. The closer the value is to 1, the whiter the color.

Therefore, we conducted a series of experiments on the MINIST dataset, comparing three different methods: local model, centralized model, and Federated Learning based on Horizontal Federated Learning (HFL). Firstly, we used the local model as the baseline, which trains and tests the model on a single device. Secondly, we employed a centralized model where all data is sent to a central server for model training and testing. Finally, we proposed Federated Learning based on HFL, which conducts model training and testing across multiple devices without sharing raw data.

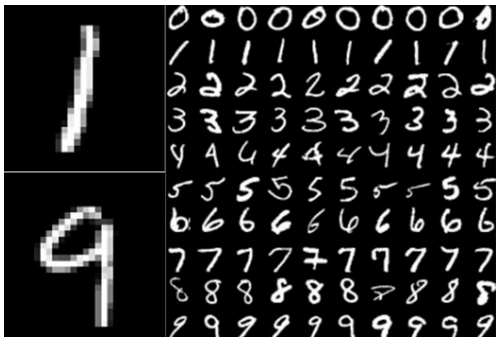

Figure 2. Handwritten digital data set

ISSN: 2664-9640

5.2. Experimental Results

Figure 3 shows the training results on three clients under the scenarios of a single device, centralized, and Federated Learning based on HFL, representing the local model, centralized model, and federated model, respectively. In this setup, the Federated Learning model had three clients, and the learning rate for all scenarios was set to 1.5e-3, with a total of 30 epochs for training. The experimental results indicate that, in terms of maintaining accuracy, the Federated Learning method based on HFL performs comparably to the local model and centralized model, eventually achieving an accuracy of 97%. Therefore, the Federated Learning method based on HFL demonstrates good performance in handwritten digit recognition tasks, preserving user privacy while maintaining model accuracy and generalization capability.

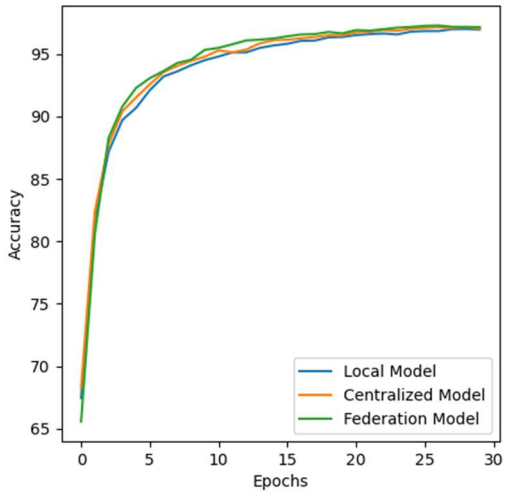

Figure 3. Comparison of the training results

6. Conclusion

In this study, we explored the task of handwritten digit recognition based on federated learning, with a focus on preserving user privacy while maintaining model performance. Through a series of experiments on the MINIST dataset, we compared local models, centralized models, and the horizontal federated learning (HFL) approach. The experimental results demonstrate that the HFL-based federated learning approach exhibits performance comparable to that of local models and centralized models in terms of accuracy. This indicates the potential of federated learning methods in the task of handwritten digit recognition and their effectiveness in addressing data privacy protection issues.

References

[1] Li T, Sahu A K, Talwalkar A, et al. Federated learning: Challenges, methods, and future directions[J]. IEEE signal processing magazine, 2020, 37(3): 50-60.

[2] Bonawitz K, Eichner H, Grieskamp W, et al. Towards federated learning at scale: System design[J]. Proceedings of machine learning and systems, 2019, 1: 374-388.

[3] Wang B, Chen Y, Jiang H, et al. PPeFL: Privacy-Preserving Edge Federated Learning with Local Differential Privacy[J]. IEEE Internet of Things Journal, 2023.

[4] McMahan H B, Ramage D, Talwar K, et al. Learning differentially private recurrent language models[J]. arxiv preprint arxiv:1710.06963, 2017.

ISSN: 2664-9640

[5] Zhou, J.; Wu, N.; Wang, Y.; Gu, S.; Cao, Z.; Dong, X.; Choo, K.-K.R. A differentially private federated learning model against poisoning attacks in edge computing. IEEE Trans. Dependable Secure. 2023, 20, 1941–1958.

[6] Aono, Y.; Hayashi, T.; Wang, L.; Moriai, S. Privacy-preserving deep learning via additively homomorphic encryption. IEEE Trans. Inf. Forensics Secur. 2017, 13, 1333–1345.

[7] Park, J.; Lim, H. Privacy-preserving federated learning using homomorphic encryption. Appl. Sci. 2022, 12, 734.

[8] Du, W.; Li, M.; Wu, L.; Han, Y.; Zhou, T.; Yang, X. A efficient and robust privacy-preserving framework for cross-device federated learning. Complex Intell. Syst. 2023.

[9] Shokri, R., & Shmatikov, V. (2015). Privacy-Preserving Deep Learning. ACM Conference on Computer and Communications Security. ACM.

[10] Wu, Y., Cai, S., Xiao, X., Chen, G., & Ooi, B. (2020). Privacy preserving vertical federated learning for tree-based models. Proceedings of the VLDB Endowment, 13, 2090-2103.

[11] I-Kai Wang, K., Zhou, X., Liang, W., Yan, Z., & She, J. (2021). Federated Transfer Learning Based Cross-Domain Prediction for Smart Manufacturing. IEEE Transactions on Industrial Informatics, 1–1.

[12] Li, Y., Zhou, Y., Jolfaei, A., Yu, D., & Zheng, X. (2020). Privacy-preserving federated learning framework based on chained secure multi-party computing. IEEE Internet of Things Journal, 99-99, 1-1.

[13] Privacy-Preserving Deep Learning. ACM Conference on Computer and CommunicationsSecurity. ACM.

[14] Le Trieu Phong, Aono, Y., Hayashi, T., Lihua Wang, & Moriai, S. (2018). Privacy-Preserving Deep Learning via Additively Homomorphic Encryption. IEEE Transactions on InformationForensics and Security, 13(5), 1333–1345.

[15] Fang, H., & Qian, Q. (2021). Privacy preserving machine learning with homomorphic encryption and federated learning. Future Internet, 13(4),94. https://doi.org/10.3390/fi13040094.