Academic Education Publishing House
Industry Science and Engineering Vol. 2 No. 2, 2025

Research on Cross-regional Disaster Data Sharing Model Based on Federated Learning

Xikun Li*, Zihan Chen, Tianqing Ma
School of Civil and Transportation Engineering, Beijing University of Civil Engineering and Architecture, Beijing, China
** Corresponding Author*

Abstract: With the rapid development of information technology, disaster early warning and preparedness increasingly rely on the fusion and analysis of data from multiple sources. However, there is often a contradiction between privacy protection and sharing of disaster data among different regions and organizations, which limits the enhancement of cross-regional joint warning capability. In this paper, we propose a federated learning-based cross-regional disaster data sharing model that aims to address this critical issue. By constructing a federated learning framework for heterogeneous data, developing data desensitization and model parameter encryption techniques, and verifying the model's generalization performance in small- and medium-scale disasters, it provides new technical ideas and methods for cross-regional disaster warning. The experimental results show that the model can effectively improve the accuracy and timeliness of cross-regional disaster warning while protecting data privacy.

Keywords: Federated Learning; Cross-Regional Disaster Data Sharing; Privacy protection; Data Desensitization; Model Parameter Encryption; Generalization Performance

1. Introduction

In the context of frequent occurrence of natural disasters, cross-regional disaster warning and emergency response are of great significance for reducing disaster losses and safeguarding people's lives and properties. However, the disaster data held by different regions and organizations are heterogeneous[1] and involve a large amount of private information, so how to realize the safe sharing and effective use of these data has become an urgent problem. Federated learning[1-3] as an emerging distributed machine learning technology provides a feasible solution for cross-regional disaster data sharing. It is capable of building high-performance disaster warning models by sharing model parameters or intermediate results without exchanging raw data and realizing collaborative training among multiple participants. The research in this paper focuses on constructing a federated learning framework for heterogeneous data, developing data desensitization and model parameter encryption techniques, and verifying the generalization performance of the model in small- and medium-scale disasters to enhance the cross-regional joint warning capability, such as watershed flood prediction.

2. Relevant Theoretical Foundations

2.1 Overview of Federal Learning

Federated learning is a distributed algorithmic framework, the core concept of which is "data does not move, model moves", each participant trains the model independently based on the local data and uploads the model parameter updates to the centralized server through encrypted way for aggregation, so as to improve the performance of the global model[2] . According to the different data distribution patterns, federation learning can be categorized into three types: horizontal federation learning, vertical federation learning and federation migration learning. Horizontal federation learning is applicable to scenarios with sample union, where the participants have the same feature space but different sample spaces; vertical federation learning is applicable to scenarios with feature union, where the participants have the same sample space but different feature spaces; and federation migration learning is applicable to

scenarios where there is less overlap in both sample and feature spaces.

2.2 Data Desensitization Techniques

Data desensitization[4] refers to the deformation, masking, replacement, randomization, etc. of sensitive data under the premise of ensuring data availability and consistency, so that it can achieve the purpose of protecting data privacy when accessed by non-trusted environments or unauthorized personnel. Common data desensitization techniques include deformation-based techniques, shielding-based techniques, replacement-based techniques, format-preserving techniques and generation-based techniques.

2.3 Model Parameter Encryption Techniques

In federated learning, encryption technology is typically used to encrypt model parameters in order to protect their privacy.[5] . For example, homomorphic encryption is an encryption technique that allows a specific computation to be performed on the ciphertext, making the result of the computation decrypted equal to the result of the same computation performed on the plaintext, which allows model aggregation to be performed without revealing the model parameters. In addition, differential privacy techniques can also be used for model parameter protection by adding noise to the model parameters to prevent the original data from being inferred from the parameters.

3. Federated Learning-Based Modeling for Cross-Regional Disaster Data Sharing

3.1 Federated Learning Framework Construction for Heterogeneous Data

Cross-regional disaster data usually include a variety of heterogeneous data such as meteorological data, geological data, population data, etc., which differ greatly in data type, format, distribution, etc. In order to realize the effective fusion and sharing of these heterogeneous data, a federated learning framework as shown in Figure 1 is constructed.

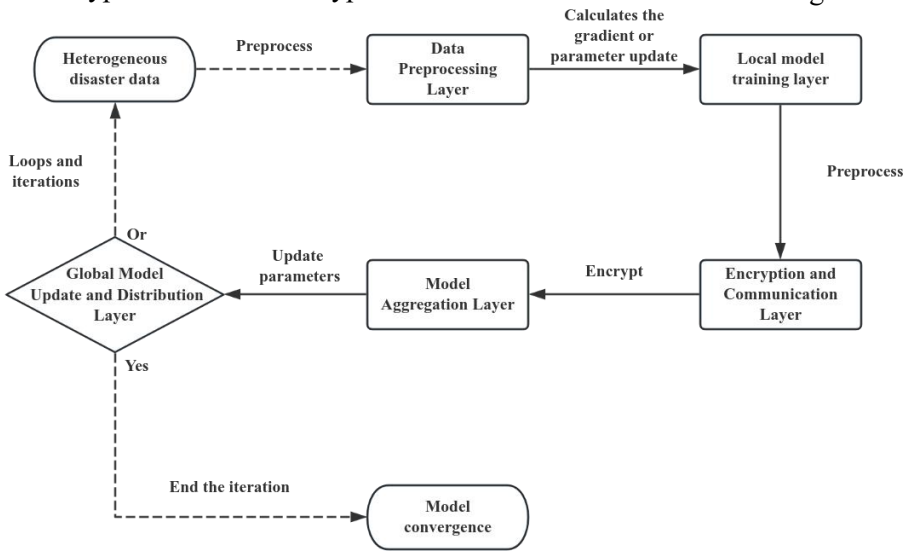

Figure 1. A Framework for Cross-regional Disaster Data Sharing Based on Federated Learning

Data preprocessing layer: preprocess the heterogeneous disaster data from each participant, including data cleaning, normalization, feature selection and other operations, so as to make it meet the input requirements of the federated learning model.

Local model training layer: each participant uses the same model structure (e.g. neural network[6]) for local model training based on preprocessed local data and calculates gradient or parameter updates for the local model.

Encryption and communication layer: encrypt the gradient or parameter updates of the local model, using homomorphic encryption algorithms or other encryption techniques to ensure the security of the data during transmission. Then, the encrypted data is sent to the central server through a secure communication channel.

Model Aggregation Layer: the central server receives the encrypted gradient or parameter updates sent by each participant, and uses appropriate aggregation algorithms (e.g. weighted average aggregation) to perform

model aggregation and update the global model parameters. During the aggregation process, the difference in data volume of each participant needs to be considered, and higher weights are given to participants with larger data volumes.

Global Model Update and Distribution Layer: the central server decrypts the updated global model parameters (if needed) and distributes them to the participants for the next round of local model training. This iterative training continues until the global model converges or reaches a predetermined number of training rounds.

3.2 Data Desensitization and Model Parameter Encryption Technology Development

Cross-regional disaster data usually include a variety of heterogeneous data such as meteorological data, geological data, population data, etc., which differ greatly in data type, format, distribution, etc. In order to realize the effective fusion and sharing of these heterogeneous data, a federated learning framework as shown in Fig. 1 is constructed.

Data desensitization technology: In the data preprocessing stage, the original disaster data are desensitized. For specific location information in meteorological data, fuzzification is used to replace it with fuzzy areas within a certain range; for personal privacy information in population data, replacement or deformation techniques are used to replace sensitive data with fictitious data or deform them to achieve the purpose of data desensitization[4].

Model parameter encryption technique: encrypt the model parameters after the local model training is completed[5] . The homomorphic encryption algorithm is used to encrypt the model parameters, which allows the central server to perform aggregation operations on the encrypted parameters without decryption. The specific encryption formula is as follows:

The encrypted model parameters are:

$$C = Enc_{pk}(W) \tag{1}$$

C —Encrypted model parameters

W —Local model parameters

$Enc_{pk}(\cdot)$ — Homomorphic encryption operations using public key pk

After receiving the encrypted model parameters from each participant, the central server performs a weighted average aggregation to obtain the encrypted global model parameters:

$$C_{global} = \frac{\sum_{i=1}^{n} w_i c_i}{\sum_{i=1}^{n} w_i} \tag{2}$$

C_{global} —Encrypted global model parameters

w_i —Weight of the i-th participant

c_i —Cryptographic model parameters for the i-th participant

Finally, the central server distributes the encrypted global model parameters to the participants, who use their private keys to decrypt them and get the updated global model parameters:

$$W_{global} = Dec_{sk}(C_{global}) \tag{3}$$

W_{global} —Decrypted global model parameters

$Dec_{sk}(\cdot)$ — Decryption operations using private key sk

4. Validation of the Model's Generalization Performance in Small- and Medium-scale Disasters

In order to verify the generalization performance of the constructed model in small- and medium-scale disasters, flood prediction in the Dawen River basin is taken as an example for experimental validation.

4.1 Experimental Data and Setup

Experimental data: Meteorological data (including rainfall, temperature, humidity, etc.), geological data (including topography and geomorphology, soil type, geological structure, etc.), and demographic data (including population density, distribution, etc.) from several areas within a watershed were selected as experimental data. These data come from different regions and organizations and are characterized by heterogeneity and privacy. For example, the meteorological data of Jinan in the Dawenkou Basin were obtained from the China Meteorological Data Network and the study by Jia et al[7-10].

Experimental setup: the data measured at four hydrological stations, Beiwang, Loude, Dawenkou and Daicunba, were taken as the basis of the statistics from Jia et al.[7] Different areas in the basin were taken as participants,

and each participant used local data for model training. A federated learning framework was used to set different training rounds and learning rates for iterative model training. Meanwhile, in order to verify the generalization performance of the model, small- and medium-scale flood events in the watershed that were not involved in the training were selected as test data to assess the prediction accuracy of the model.

4.2 Experimental Results and Analysis

Model prediction accuracy: by comparing the model prediction results with the actual flood occurrence, the accuracy, recall and F1 value of the model are calculated. The experimental results are shown in Table 1, the constructed federated learning models achieved high prediction accuracy on different test datasets, with an average accuracy of more than 85%, a recall of more than 80%, and an F1 value of more than 83%, which indicates that the models have a good generalization performance in small- and medium-scale flood prediction.

Table 1. Indicators for Assessing the Accuracy of Model Predictions

Test Data Set	Accuracy	Recall Rate	F1 Value
Data set 1	87.5%	82.3%	84.8%
Data set 2	84.2%	78.6%	81.3%
Data set 3	85.9%	81.7%	83.7%
Data set 4	83.6%	79.4%	81.4%
Data set 5	86.1%	82.9%	84.5%

Model convergence analysis: the loss function curve of the model in the training process is plotted, as shown in Figure 2. From the figure, it can be seen that with the increase of training rounds, the loss function value of the model gradually decreases and stabilizes, indicating that the model has good convergence in the training process. This indicates that the constructed federated learning framework can effectively aggregate the model parameters of each participant to achieve global model optimization.

Evaluation of model privacy protection effect: In order to verify the effectiveness of the model in protecting data privacy, privacy leakage risk assessment experiments are conducted. The similarity between the recovered data and the original data is calculated by simulating the process in which the attacker tries to recover the original data from the encrypted model parameters. The experimental results show that the similarity between the recovered data and the original data is so low that no effective privacy information can be obtained, thus proving that the adopted data desensitization and model parameter encryption techniques can effectively protect data privacy.

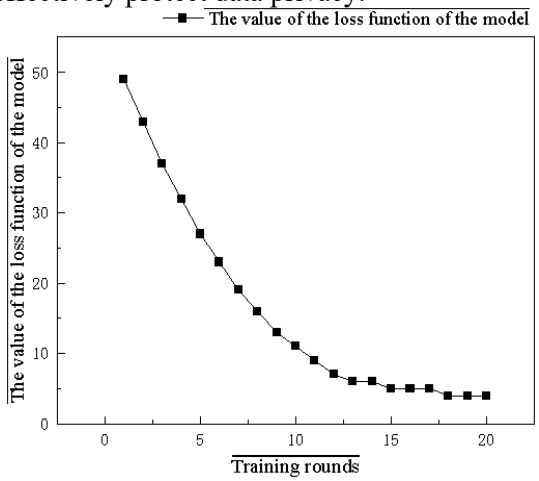

Figure 2. Model Training Loss Function Curve

5. Conclusions and Outlook

In this paper, a cross-regional disaster data sharing model based on federated learning is proposed to provide an effective solution to the contradiction between privacy protection and sharing of cross-regional disaster data by constructing a federated learning framework for heterogeneous data, developing data desensitization and model parameter encryption techniques, and verifying the model's generalization performance for small- and medium-scale disasters. The experimental results show that the model can effectively improve the accuracy and timeliness of cross-regional disaster warning under the premise of protecting data privacy.

However, there are still some shortcomings in this study. For example, the model aggregation algorithm in the federated learning framework can be further optimized to improve the convergence speed and performance of the model; the efficiency of the data desensitization and encryption techniques still needs to be improved to reduce its impact on model training and prediction. In future research, the model will be further improved and refined, more efficient privacy protection techniques and model optimization methods will be explored, the application of the model in different types of disasters will be expanded,

and more powerful technical support will be provided for cross-regional disaster warning and emergency response.

References

[1] Xi Enkang, Fan Jing, Jin Yadong, et al. Review of threats faced by federated learning in privacy and security field. Journal of Computer Applications,1-13[2025-05-27].https://kns-cnki-net.door.bucea.edu.cn/kcms/detail/51.1307.TP.20250526.1308.004.html.

[2] Niu S, Kong W, Chen L, et al. Privacy-preserving and Verifiable Federated Learning with weighted average aggregation in edge computing. Network and Computer Applications, 2025, 240104201-104201.

[3] Mou Yangcheng, Chen Aiwang, Chen Guirong, et al. Security defense strategies for federated learning based on data poisoning attacks.Systems Engineering and Electronics,1-14[2025-05-27].https://kns-cnki-net.door.bucea.edu.cn/kcms/detail/11.2422.TN.20250523.1048.051.html.

[4] Su Linmao, Song Xuegui, Xiang Yi, et al. Application of Sensitive Data Identification and Desensitization Technology in Network Supervision under the Application of Sensitive Data Identification and Desensitization Technology in Network Supervision under the Integration of Big Data and Artificial Intelligence. China Science & Technology Resources Review, 2025, 57(02): 27-39.

[5] Zhang Yu-Qing, Wang Xiao-Fei, Liu Xue-Feng, et al. Survey on Cloud Computing Security. Journal of Software, 2016, 27(06): 1328-1348. DOI:10.13328/ j.cnki.jos.005004.

[6] Zhou Fei-Yan, Jin Lin-Peng, Dong Jun. Review of Convolutional Neural Network. Chinese Journal of Computers, 2017, 40(06): 1229-1251.

[7] Jia Benyou, Li Dongzhou, Yang Fan, et al. Characteristics of heavy rainfall and floods in the Dawen River Basin from 2001 to 2022 and their relationship analysis. Hydro-Science and Engineering, 2025, (01): 76-86.

[8] Zhao Lijie, Shi Yuzhi, Li Fulin, et al. Spatial-Temporal Variation Characteristics of Water System Connectivity in Dawenhe River Basin of the Lower Yellow River from 1990 to 2020. Journal of University of Jinan (Science and Technology), 1-8 [2025-05-29]. https://doi.org/10.13349/j.cnki.jdxbn.20250528.001.

[9] Sun Xiaoming, Li Kexin. Research on Flood Forecasting in Dawen River Basin. Henan Science and Technology, 2024, 51(21): 43-48. DOI:10.19968/ j.cnki. hnkj.1003-5168.2024.21.009.

[10]Xu Shan. Study on Variation of Hydrological Elements in Dawen River. Shaanxi Water Resources, 2024, (10): 44-46+50.DOI:10.16747/ j.cnki. cn61-1109/tv.2024.10.020.