10.18686/aitr.v2i3.4407

Research on Credit Card Fraud Detection System Based on Federated Learning

Ya Wen

Xinjiang Bank CO., LTD, Xinjiang, 830026

Abstract: With the rapid development of the mobile Internet, financial technology, especially artificial intelligence and blockchain technology, has profoundly changed our consumer behavior and the development model of the traditional financial industry. However, the accompanying risks are also being transmitted with unprecedented speed and complexity. Credit card fraud, as an important form of financial fraud, has caused huge economic losses and a crisis of trust for both financial institutions and consumers. Therefore, the establishment of an efficient and secure credit card fraud detection system has become an inevitable requirement for the continuous development of financial technology in the new era. The purpose of this paper is to study and develop a credit card fraud detection system based on federated learning, which can balance the positive and negative samples in credit card transaction data, protect the privacy of bank data, and effectively improve the accuracy and efficiency of fraud detection.

Keywords: Federated learning; Privacy protection; Fraud detection

Introductory

Federated Learning (FL) is an emerging machine learning framework that allows multiple participants (e.g., banks and financial institutions) to collaborate in training a shared model without sharing raw data. This approach not only protects users' privacy, but also improves the performance of the model without violating data privacy regulations. Therefore, federated learning based credit card fraud detection system becomes an effective way to solve the above problems.

1. Advantages of Federal Learning

1.1 Effective protection of data privacy

Federated learning revolutionizes data privacy by enabling collaborative model training without centralized data storage, reducing the risk of breaches. It allows participants to train models locally and share updates, not raw data, ensuring control stays with the data owners. Techniques like encryption and differential privacy enhance security during updates.

1.2 Ability to overcome data silos and improve generalization of models

in sectors like finance where data sharing is restricted, federated learning bridges silos by enabling organizations to collaborate on model training without sharing data. This integration of diverse data sources leads to more robust, generalized models with improved predictive accuracy. It fosters industry-wide knowledge sharing and advancements, offering a holistic approach to solving issues like credit card fraud detection.

2. Challenges to Credit Card Fraud Detection

2.1 Data imbalance and rapid evolution of fraudulent behavior

Credit card fraud detection faces significant challenges, including data imbalance and evolving fraudulent techniques. Data scarcity on fraud instances hampers machine learning model training, leading to overfitting on normal patterns. Moreover, the dynamic nature of fraud demands systems to quickly adapt to emerging tactics.

2.2 Privacy Protection and Cross-Agency Data Sharing Barriers

privacy concerns and data sharing barriers hinder collaboration among institutions for improved fraud detection. Breaking silos while safeguarding user data is crucial for enhancing model accuracy and generalization.

2.3 Conflict between real-time requirements and accuracy guarantees

balancing real-time response requirements with detection accuracy is vital. Ensuring prompt responses without compromising precision poses a critical challenge in designing effective fraud detection systems. Striking the right balance between speed and accuracy is key to combating fraud effectively.

3. Design of fraud detection system based on federated learning

3.1 System Architecture Design

In building such a system, consideration needs to be given to how the framework of federated learning can be combined with the specific needs of fraud detection. The federated learning architecture is shown in the following figure:

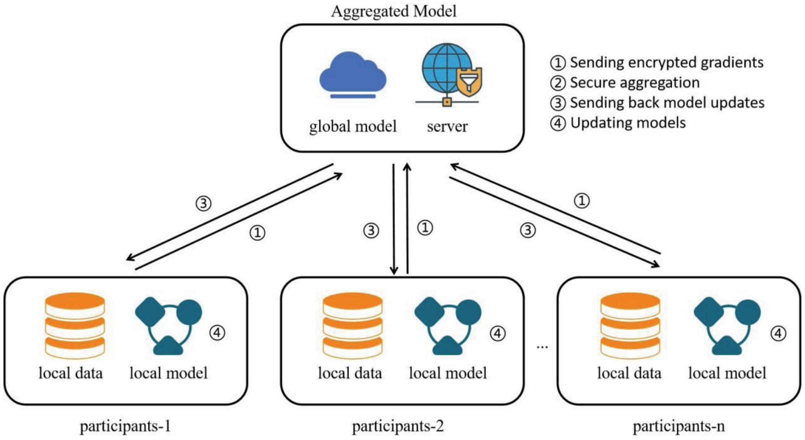

Figure 1: The proposed federated learning framework.

The system should include multiple participants, each representing a financial institution that holds its own transaction data locally. These participants interact with a central server (or coordinator) through a secure communication network. The central server is responsible for coordinating the model training process of each participant, collecting model updates, and performing model aggregation. The system architecture also needs to take into account data privacy and security requirements, such as the use of encryption to secure data during transmission and differential privacy techniques to prevent model updates from revealing sensitive information. In addition, the system should have the flexibility to adapt to the computing power and network conditions of different participants to ensure the efficiency and stability of the entire federated learning process.

3.2 Model selection and optimization

In federated learning, choosing the right machine learning model is crucial for accurate fraud detection. Due to the characteristics of credit card fraud data like class imbalance and feature diversity, deep learning models or integrated methods might be necessary for better detection. To prevent overfitting and excessive resource use in a multi-participant setting, the model complexity should be moderate. Local model updates and global aggregation can optimize model parameters iteratively for improved detection. Regularization techniques and data enhancements can boost model generalization during training.

For the task of credit card fraud detection, Long Short-Term Memory (LSTM) network with temporal information can be used as the model for training data. Long Short-Term Memory networks are a type of recurrent neural network (RNN) architecture designed to overcome the vanishing gradient problem in traditional RNNs, allowing them to effectively capture long-term dependencies in sequential data.

LSTMs have a more complex structure compared to standard RNNs, incorporating a memory cell that can maintain information over long periods of time. They achieve this by using multiple gating mechanisms (input gate, forget gate, and output gate) that regulate the flow of information into and out of the memory cell.

The key equations governing an LSTM cell are as follows:

Input Gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{1}$$

Forget Gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

Output Gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3}$$

Memory Cell:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \qquad (3)$$

Hidden State:

$$h_t = o_t \tanh(c_t) \tag{4}$$

where i_t is the input gate value at time step t. f_t is the forget gate value at time step t. o_t is the output gate value at time step t. c_t is the cell state at time step t. h_t is the hidden state at time step t. x_t is the input at time step t. W_i, W_f, W_o, W_c are weight matrices. b_i, b_f, b_o, b_c are bias vectors. σ represents the sigmoid activation function. tanh represents the hyperbolic tangent activation function. The structure of LSTM is shown in the following figure:

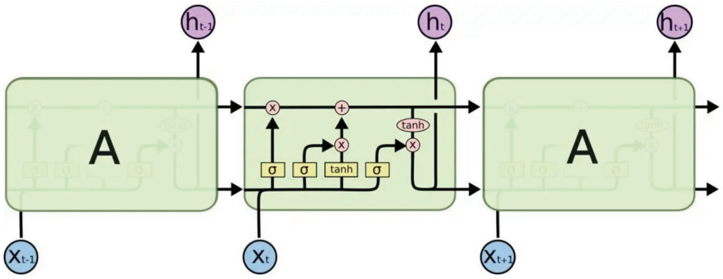

Figure 2: LSTM Network Architecture

3.3 Privacy protection mechanisms

Privacy-preserving mechanisms are also a central element in the design of federated learning-based fraud detection systems. In federated learning, protecting the data privacy of the participants is a top priority. To this end, a variety of technical tools can be used to enhance privacy protection.

For example, the use of homomorphic encryption techniques allows the computation of encrypted data without decryption, thus protecting the privacy of data during transmission and aggregation. The figure below shows a comparison between a machine learning inference structure using homomorphic encryption and plaintext processing.

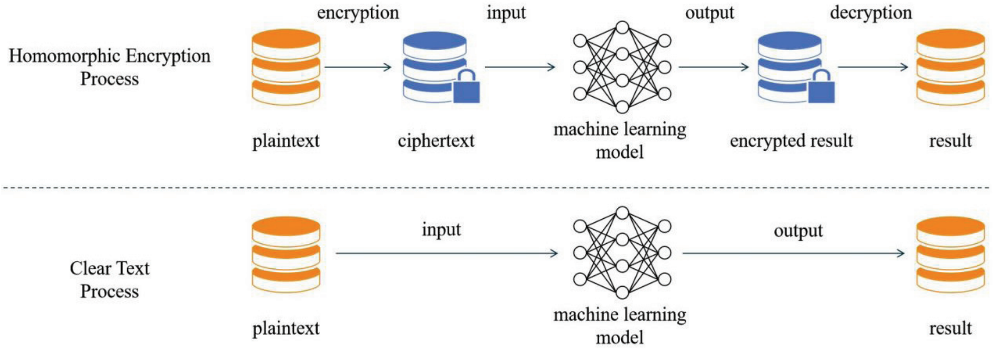

Figure 3: comparison between homomorphic encryption and plaintext

Differential privacy techniques, on the other hand, can protect the privacy of individual data by adding noise to model updates to prevent individual data points from overly influencing the model. In addition, the level of privacy protection in the system can be ensured by setting a privacy budget and monitoring the risk of privacy leakage.

The definition of differential privacy techniques is as follows:

If for any adjacent datasets (i.e., datasets that differ by only one individual's data) and for any possible output result set, the following inequality holds:

$$\frac{Pr[Alg(D_1 \in S)]}{Pr[Alg(D_2 \in S)]} \le e^{\varepsilon} \tag{5}$$

where D_1, D_2 are two datasets differing by only one data point, Alg is an algorithm, S is an output result set, S is the differential privacy parameter.

This definition indicates that in ε-differential privacy, the impact of changing a data point of an individual dataset on the output result is controlled. By limiting this impact, data privacy is preserved. In the formula, e is the base of the natural logarithm, and ε is the parameter that controls the privacy level. This formula quantitatively describes the restriction on the change in the output result when adding or removing a data point in a given dataset.

4. Concluding remarks

The research on credit card fraud detection system based on federated learning can not only improve the accuracy and efficiency of fraud detection, but also effectively protect the data privacy of users and the data security of financial institutions. With the continuous progress of technology and the expansion of application scenarios, federated learning is expected to become an important branch in the field of financial technology, providing a more reliable and secure solution to the problem of credit card fraud. Future research needs to continue to explore the application of federated learning in different scenarios and how to further optimize the model performance and privacy protection mechanisms to cope with changing fraud tactics and regulatory requirements.

References

[1] Zhu J, Ma X, Blaschko M B. Confidence-aware personalized federated learning via variational expectation maximization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 24542-24551.

[2] Zhao J C, Elkordy A R, Sharma A, et al. The resource problem of using linear layer leakage attack in federated learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 3974-3983.

[3] Zhang R, Xu Q, Yao J, et al. Federated domain generalization with generalization adjustment[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 3954-3963.

[4] Bhagoji A N, Chakraborty S, Mittal P, et al. Analyzing federated learning through an adversarial lens[C]//International conference on machine learning. PMLR, 2019: 634-643.

[5] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[6] Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network[J]. Physica D: Nonlinear Phenomena, 2020, 404: 132306.

[7] Yi X, Paulet R, Bertino E, et al. Homomorphic encryption[M]. Springer International Publishing, 2014.

[8] Acar A, Aksu H, Uluagac A S, et al. A survey on homomorphic encryption schemes: Theory and implementation[J]. ACM Computing Surveys (Csur), 2018, 51(4): 1-35.

[9] Dwork C. Differential privacy: A survey of results[C]//International conference on theory and applications of models of computation. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008: 1-19.

[10] Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy[C]//Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016: 308-318.