Proceedings of the 5th International Conference on Business and Policy Studies
DOI: 10.54254/2754-1169/2026.31318

Credit Card Fraud Detection Based on Machine Learning Algorithms

Yuanke Wang

School of Business Administration, Beijing Normal-Hong Kong Baptist University, Zhuhai, China
q030006178@mail.bnbu.cn

Abstract. The wide popularity of electronic payment has driven credit cards to become the core payment tool for daily consumption and commercial transactions. However, the continuous expansion of transaction scale has provided more breeding space for fraud, making credit card fraud detection an important issue in the field of financial security. This paper first conducts data exploration through correlation analysis and violin graph analysis, and then constructs multiple machine learning algorithms for performance comparison experiments. The experimental results show that the proposed Transformer-BiGRU algorithm presents significant advantages in all performance indicators and has the most outstanding comprehensive performance. The accuracy rate, recall rate, precision rate and F1 value of this algorithm all reach 98.6%. It is the only model among all the comparison algorithms that has achieved unification and reached the highest level in these four core indicators, fully demonstrating extremely strong classification consistency and reliability. This research achievement provides an efficient and feasible technical solution for credit card fraud detection, which has significant practical value in enhancing the risk prevention and control capabilities of financial transactions and ensuring transaction security.

Keywords: Electronic payment, credit card fraud detection, Transformer, BiGRU

1. Introduction

With the popularization of electronic payment, credit cards have become the core tool for daily consumption and business transactions. The continuous expansion of transaction scale has also provided more room for fraud to breed [1]. In recent years, credit card fraud methods have been constantly upgrading, presenting characteristics of intelligence and concealment. New fraud models such as cross-platform fraud and dynamic information forgery have occurred frequently, not only causing direct economic losses to financial institutions and users, but also seriously damaging the trust system of the payment ecosystem [2]. Traditional fraud detection mainly relies on manually formulated rule engines. Although they can handle simple fraud scenarios, when dealing with complex association patterns in massive transaction data, they have problems such as lagging rule updates, high missed detection rates, and high costs of misjudgment, making it difficult to meet the demands of real-time and precise detection. Therefore, building an efficient credit card fraud detection system has become an urgent task in the field of financial security [3].

DOI: 10.54254/2754-1169/2026.31318

Machine learning algorithms, with their data-driven feature learning capabilities, provide crucial technical support for credit card fraud detection [4]. Compared with traditional rule engines, machine learning models can automatically mine hidden patterns from historical transaction data, effectively handle nonlinear relationships among transaction features, and dynamically adjust according to changes in fraud patterns, thus having stronger adaptability [5]. At present, a variety of machine learning algorithms have been applied in this field. Logistic regression can quickly construct detection baselines, random forests enhance the ability to capture feature interactions through ensemble learning, and long short-term memory networks take advantage of time series modeling to analyze the time dependence of transaction sequences. However, the existing algorithms still have limitations. Logistic regression is difficult to handle complex feature interactions, random forests have insufficient temporal correlation modeling for long sequence transactions, and long short-term memory networks are prone to vanishing gradients when capturing long-distance transaction dependencies. These problems restrict the accuracy and generalization ability of fraud detection.

To address the shortcomings of existing algorithms and enhance the model's recognition ability for complex transaction patterns and the capture effect of long-term temporal dependencies, this paper proposes the Transformer-BiGRU classification algorithm. This algorithm integrates the advantages of Transformer and bidirectional gated recurrent units. The multi-head attention mechanism of Transformer can focus on the key features in transaction data and accurately identify the differences between abnormal transactions and normal transactions in dimensions such as amount fluctuations, merchant types, and geographical locations.

2. Data sources

The dataset used in this article contains 716 transaction records and 12 feature columns, covering transaction-related information, including transaction amount, transaction hours, transaction type, merchant type, whether it is an international transaction, the number of transactions in the past 30 days, and the number of hours since the last transaction. User and card information includes the user's age, the card's usage period, and security-related indicators, including whether a PIN code is used, whether the balance is insufficient, and whether it is a fraudulent transaction.

Some datasets are selected for display, as shown in Table 1.

Table 1. The results of the comparative experiment

amount	user_age	card_age_years	transaction_type	merchant_type	is_international	transactions_last30d	hours_since_last	used_pin	insufficient_funds	is_fraud
12899.7	65	0.5	0	3	0	22	11.4	0	0	0
1737.36	22	2.2	0	2	1	27	10.9	1	0	0
8168.19	58	0.8	0	2	1	60	30.8	0	1	1
5351.69	63	5.1	2	3	0	3	42.5	0	0	1
5330.98	31	4.6	1	3	0	5	18	0	1	0

Draw the correlation heat map of each variable as shown in Figure 1.

DOI: 10.54254/2754-1169/2026.31318

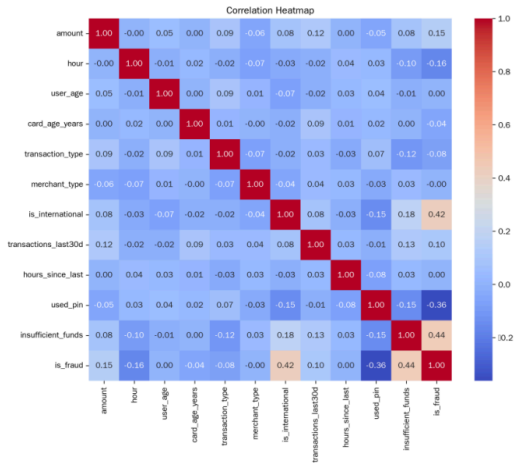

Figure 1. The correlation heat map

Draw violin plots for each variable as shown in Figure 2.

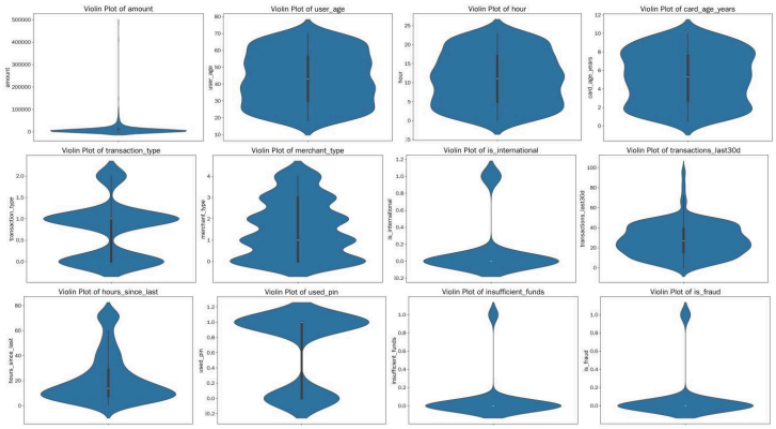

Figure 2. The violin plots for each variable

3. Method

3.1. Transformer

The Transformer algorithm is a deep learning model based on the self-attention mechanism, which completely breaks away from the dependence of recurrent neural networks on sequence order and significantly improves training efficiency through parallel computing [6]. Its core structure consists of an encoder and a decoder, both of which incorporate a multi-layer multi-head self-attention mechanism and a feedforward neural network. The multi-head self-attention mechanism, through parallel computing of multiple attention heads, can simultaneously capture the dependency relationships at different distances in the input sequence, enhancing the model's ability to model global information. To make up for the lack of position information caused by the non-cyclic structure, the Transformer introduces position encoding, embedding absolute or relative position information into the input vector, enabling the model to perceive the temporal features of the sequence. The design of layer normalization and residual connection effectively alleviates the vanishing gradient problem and helps the model train deep networks [7]. The network structure of the Transformer is shown in Figure 3.

DOI: 10.54254/2754-1169/2026.31318

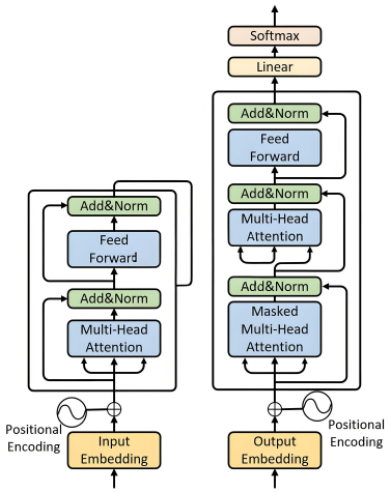

Figure 3. The network structure of the Transformer

3.2. BiGRU

The BiGRU algorithm, or Bidirectional Gated Recurrent Unit, is an extended sequence modeling algorithm based on GRU, focusing on capturing the bidirectional temporal dependencies of sequences. As a variant of recurrent neural networks, GRU dynamically regulates the transmission and forgetting of information through two core gating mechanisms: update gates and reset gates, effectively alleviating the vanishing gradient problem of traditional RNNS [8]. BiGRU sets up two GRU units in parallel, the forward unit and the reverse unit. The forward unit passes information backward from the starting position of the sequence, capturing the dependencies from the past to the present, while the reverse unit passes information forward from the end of the sequence, capturing the dependencies from the future to the present. The output vectors from the two directions are concatenated or weighted fused to form a feature representation containing complete bidirectional time series information, enabling the model to understand the sequence context more comprehensively [9]. The network structure of BiGRU is shown in Figure 4.

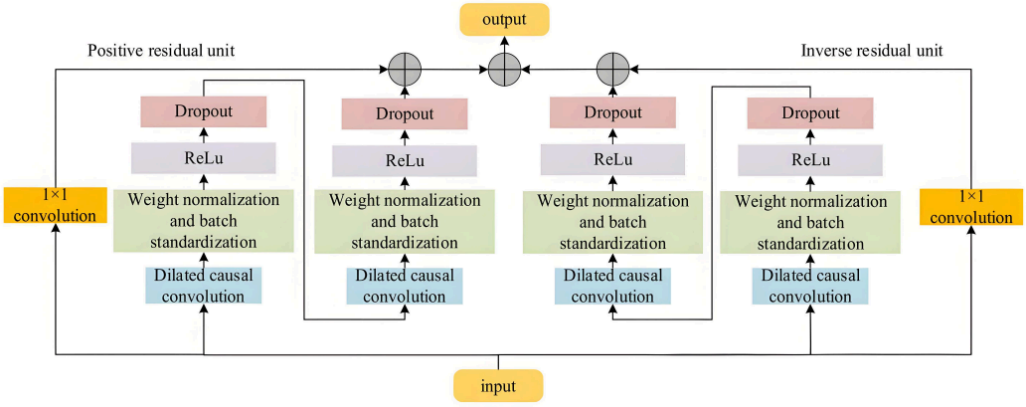

Figure 4. The network structure of BiGRU

3.3. Transformer-BiGRU

The Transformer-BigRU algorithm is a hybrid model that combines the advantages of Transformer and BiGRU, aiming to simultaneously capture the global dependencies and local temporal features

DOI: 10.54254/2754-1169/2026.31318

of sequences [10]. The core design concept is to complement the modeling capabilities of the two models: BiGRU is good at capturing local continuous temporal dependencies and can effectively extract local context information in the sequence, while Transformer is good at modeling long-distance global dependencies through its self-attention mechanism, solving the problem of insufficient global information capture of BiGRU on long sequences. Common fusion methods include series structure and parallel structure. In the series structure, local temporal features are usually extracted by BiGRU first, and then the output results are input into the Transformer for global dependency modeling. The parallel structure simultaneously utilizes two models to extract features respectively, and then fuses the features through the attention mechanism or concatenation operation.

4. Result

The parameter Settings of the project are as follows: The proportion of the training set to the dataset is 0.7. In the Transformer-BiGRU model, the number of input channels is equal to the feature dimension, the maximum position encoding is 512, the number of heads of the self-attention mechanism is 4, the number of key channels of each head is 32, the total number of key channels is 128, it contains GRU layers with 6 and 10 units respectively, and the Dropout rate is 0.01. During training, the optimizer selects Adam, with a maximum of 200 training rounds and a batch size of 256. The dataset is shuffled at each epoch. The initial learning rate is 0.01, the learning rate decline factor is 0.1, the decline period is 80, the L2 regularization coefficient is 0.001, the gradient clipping threshold is 10, and the execution environment automatically selects to prioritize the use of GPU.

In this paper, Decision tree, Random Forest, AdaBoost, GBDT, KNN, BP neural network and Naive Bayes were used as comparison models. The experimental results are shown in Table 2.

Table 2. A part of the dataset metrics

Model	Accuracy	Recall	Precision	F1	AUC
Decision tree	0.967	0.967	0.976	0.97	0.962
Random Forest	0.981	0.981	0.982	0.98	0.978
AdaBoost	0.967	0.967	0.968	0.968	0.985
GBDT	0.949	0.949	0.956	0.951	0.979
KNN	0.93	0.93	0.922	0.912	0.93
BP neural network	0.907	0.907	0.823	0.863	0.919
Naive Bayes	0.944	0.944	0.947	0.928	0.971
Transformer-BiGRU	0.986	0.986	0.986	0.986	0.982

Output the bar comparison charts of each indicator, as shown in Figure 5. The Transformer-BiGRU algorithm proposed in this paper shows significant advantages in all performance indicators and has the best overall performance. Its accuracy rate, recall rate, precision rate and F1 value all reach 98.6%, making it the only model among all algorithms that achieves a unified and highest level in these four core indicators, demonstrating extremely strong classification consistency and reliability. In terms of the AUC metric, this algorithm ranked second with a score of 98.2%, only slightly lower than AdaBoost's 98.5%. However, AdaBoost was below 97% in the other four metrics, and its overall performance was inferior to that of Transformer-BiGRU. The Random Forest algorithm performed second, with all indicators around 98%. Among them, the accuracy rate and recall rate were 98.1%, the precision rate was 98.2%, the F1 value was 98%, and the AUC was

DOI: 10.54254/2754-1169/2026.31318

97.8%. Although there was a small gap compared with the Transformer-BiGRU, However, it is still the model with the most outstanding overall performance apart from this algorithm. The accuracy and recall rates of decision trees and AdaBoost are both 96.7%, showing similar performance. However, the accuracy rate of decision trees at 97.6% is higher than that of AdaBoost at 96.8%, and the AUC advantage of AdaBoost is more prominent.

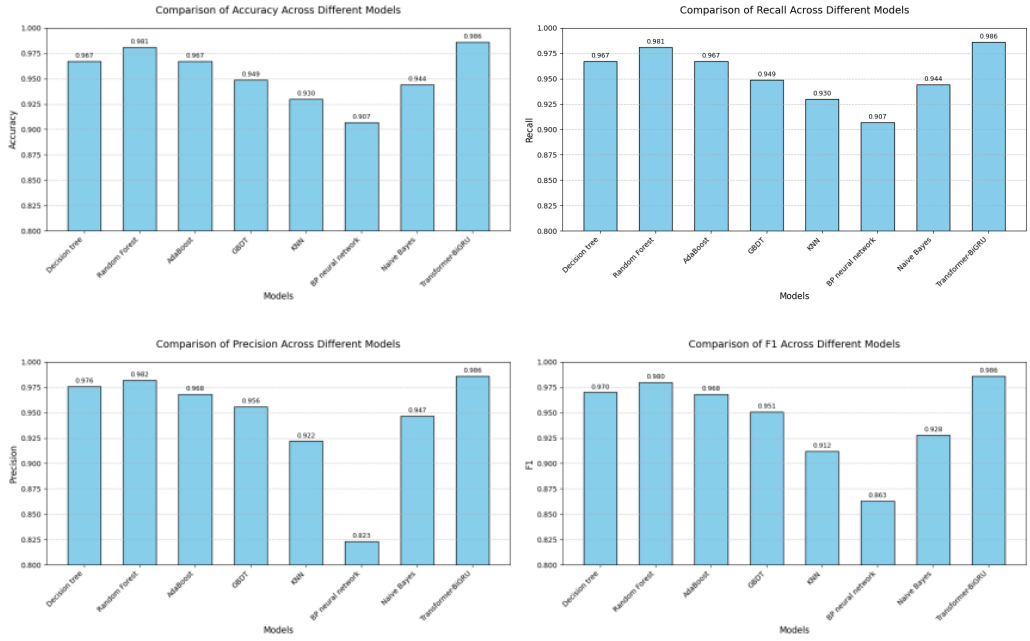

Figure 5. The bar comparison charts of each indicator

Output the test set confusion matrix of the Transformer-BiGRU model proposed in this paper, as shown in Figure 6.

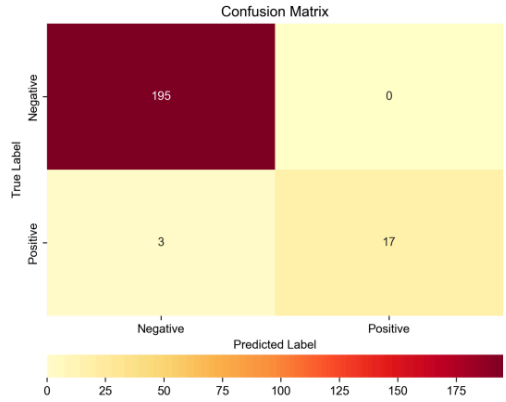

Figure 6. The test set confusion matrix of the Transformer-BiGRU model

5. Conclusion

With the wide application of electronic payment, credit cards have become an indispensable core tool in daily consumption and business transactions. The continuous growth of transaction volume has also provided more breeding ground for fraud. This study first conducted data exploration through correlation analysis and violin graph analysis, and then constructed multiple machine

DOI: 10.54254/2754-1169/2026.31318

learning algorithms for comparative experiments. The experimental results show that the proposed Transformer-BiGRU algorithm presents significant advantages in all performance indicators and has the most outstanding comprehensive performance. The accuracy rate, recall rate, precision rate and F1 value of this algorithm all reach 98.6%. It is the only model among all comparison algorithms that has achieved unification and reached the highest level in these four core indicators, fully demonstrating extremely strong classification consistency and reliability. This research achievement not only provides an efficient and feasible technical solution for credit card fraud detection, but also offers an important reference for the intelligent upgrade of risk prevention and control in the financial field, which has positive significance for ensuring transaction security and the stability of the financial market.

References

[1] Mienye, Ibomoiye Domor, and Nobert Jere. "Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions." IEEE Access (2024).
[2] Chatterjee, Pushpita, Debashis Das, and Danda B. Rawat. "Digital twin for credit card fraud detection: Opportunities, challenges, and fraud detection advancements." Future Generation Computer Systems 158 (2024): 410-426.
[3] Yu, Chang, et al. "Credit card fraud detection using advanced transformer model." 2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom). IEEE, 2024.
[4] Dastidar, Kanishka Ghosh, Olivier Caelen, and Michael Granitzer. "Machine learning methods for credit card fraud detection: A survey." IEEE Access (2024).
[5] Dornadula, Vaishnavi Nath, and Sa Geetha. "Credit card fraud detection using machine learning algorithms." Procedia computer science 165 (2019): 631-641.
[6] Chan, Philip K., et al. "Distributed data mining in credit card fraud detection." IEEE Intelligent Systems and Their Applications 14.6 (2002): 67-74.
[7] Fu, Kang, et al. "Credit card fraud detection using convolutional neural networks." International conference on neural information processing. Cham: Springer International Publishing, 2016.
[8] Xuan, Shiyang, et al. "Random forest for credit card fraud detection." 2018 IEEE 15th international conference on networking, sensing and control (ICNSC). IEEE, 2018.
[9] Oche, Agada Joseph, et al. "A systematic review of key retrieval-augmented generation (rag) systems: Progress, gaps, and future directions." arXiv preprint arXiv: 2507.18910 (2025).
[10] Zhu, Mengran, et al. "Enhancing credit card fraud detection a neural network and smote integrated approach." arXiv preprint arXiv: 2405.00026 (2024).