Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/116/2025.20430

User Behavior Analysis and Prediction Based on Differential Evolution Algorithm Optimized Transformer Combined with Bidirectional Long Short-Term Memory Neural Network

Yuxin Liu[1,a,*], Yuhan Zhang[2], Minxuan Hu[3], Yuming Tu[4], Xinqi Dong[5]

[1]Faculty of Engineering, University of Sydney, Camperdown NSW 2050, USA.
[2]Master of Science in Finance, Washington University in St. Louis (washU), 1 Brookings Dr, Saint Louis, MO 63130, USA.
[3]Cornell Bowers College of Computing and Information Science, Cornell University, New York, Ithaca, 14853, USA.
[4]Independent researcher, New Jersey, USA.
[5]Management Information Systems, Department of Computer Science, University of Maine at Presque Isle, ME, USA.
a. yliu4153@uni.sydney.edu.au
**corresponding author*

Abstract: This paper discusses a method combining differential evolution algorithm and bidirectional long short-term memory neural network (BiLSTM) to optimize Transformer model, aiming to improve the accuracy of bank customer credit analysis and prediction. By optimizing the parameters of Transformer model through differential evolution algorithm and combining with the powerful time series analysis capability of BiLSTM, an efficient credit prediction model is constructed. In the process of model training, with the increase of the number of iterations, the correct rate of the model steadily improves and eventually becomes stable. Meanwhile, the value of the loss function also gradually decreases, indicating that the performance of the model on the training data is gradually enhanced. In addition, through confusion matrix analysis, we found that the prediction accuracy of the training set reached 86.15%, and that of the test set reached 82.91%, which showed that the model not only performed well on the training data, but also had strong generalization ability on the unseen data. In addition, the F1 score and DAUC values of the model are both positive, which further confirms the superiority of the model's prediction effect. The research results of this paper show that the Transformer model optimized by differential evolution algorithm and BiLSTM has high accuracy and good generalization ability in the analysis and prediction of bank customer credit. This optimization method not only improves the prediction performance of the model, but also enhances the stability and reliability of the model in practical application by reducing overfitting. Through this optimization strategy, banks can more accurately assess the credit risk of customers and make more informed credit decisions. This research provides a new technical means for credit risk assessment in the financial field, which is helpful to improve the efficiency and security of banking business.

Keywords: Swarm intelligence optimization, Transformer, American social media

DOI: 10.54254/2755-2721/116/2025.20430

1. Introduction

With the rapid development of the global economy and the continuous expansion of the financial market, the banking industry is facing more and more competitive pressure. As one of the core businesses of banks, credit business is crucial to the profitability and risk control ability of banks [1]. However, the credit decision-making process usually involves a lot of uncertainty and complexity, and how to accurately assess the credit risk of customers is an important challenge for bank managers. Especially in the context of financial crisis, banks need to control credit risk more strictly to avoid the increase of non-performing loans. This demand has prompted the research on credit forecasting, aiming to find out the potential factors of customer default through data analysis and modeling techniques, so as to optimize credit decisions [2].

In traditional credit evaluation methods, credit scoring models rely on artificially set rules and experience, resulting in limited prediction accuracy [3]. With the advent of the big data era, banks are able to obtain and store massive amounts of customer data, including personal information, consumption behavior, and past loan records. This provides a rich data base for the research of credit forecasting. In this context, researchers began to explore more efficient credit forecasting models to improve the accuracy and reliability of forecasts [4]. Machine learning algorithm, as a powerful data analysis tool, can process complex multidimensional data and continuously optimize the model through self-learning, so it is widely used in the field of credit forecasting.

The importance of machine learning algorithms in banking customer credit forecasting can be seen in several ways. First, machine learning can effectively extract hidden patterns from historical data, which makes credit evaluation not only rely on simple rules, but on a data-driven foundation. Compared with traditional statistical methods, machine learning can better capture the nonlinear relationship of customer credit risk and improve the prediction ability of the model. Through in-depth analysis of customer behavior, banks can better identify high-risk customers and adopt more scientific strategies in credit granting decisions [5]. Second, machine learning algorithms are able to handle high-dimensional data and cope with the diversity of data. Customer's credit risk is not only affected by personal economic conditions, but also by external economic environment, industry background and other multiple factors. Traditional models tend to perform poorly in the face of complex data structures, while advanced machine learning models such as Random Forest [6], XGBoost, and neural networks are able to process this data efficiently and provide more accurate credit scores. Through these models, banks can dynamically track the credit status of customers, adjust risk strategies in time, and reduce the risk of overdue credit and default. This paper optimizes Transformer algorithm based on differential evolution algorithm and bidirectional long short term memory neural network, which can be used for bank customer credit analysis and prediction.

2. Data from data analysis

This paper downloads data set from a public website to analyze and forecast the credit of bank customers. The data set contains multiple indicators, including age, employment address, income, debt, credit and debt default situation. In debt default, 0 represents no default and 1 represents default. In order to display the data intuitively, we selected a part of the data set for display, as shown in Table 1.

Table 1: Some of the data

age	ed	employ	address	income	creddebt	othdebt	default
41	3	17	12	176	11.359392	5.008	1
27	1	10	6	31	1.362202	4.000	0
40	1	15	14	55	0.856075	2.168	0

Table 1: (continued).

41	1	15	14	120	2.65872	0.82	0
24	2	2	0	28	1.787436	3.056	1
41	2	5	5	25	0.3927	2.1	0
39	1	20	9	67	3.833874	16.668	0
43	1	12	11	38	0.128592	1.239	0
24	1	3	4	19	1.358348	3.277	1
36	1	0	13	25	2.7777	2.14	0
27	1	0	1	16	0.182512	0.089	0
25	1	4	0	23	0.252356	0.943	0
52	1	24	14	64	3.9296	2.47	0
37	1	6	9	29	1.715901	3.011	0
48	1	22	15	100	3.7037	5.3963	0
36	2	9	6	49	0.817516	3.396	1
36	2	13	6	41	2.918216	3.805	1
43	1	23	19	72	1.181952	4.290	0

3. Method

3.1. Differential evolution algorithm

Differential evolution (DE) is a population-based optimization algorithm that is inspired by biological evolution processes in nature, specifically how populations adapt to environmental changes through heredity and variation. DE algorithm is mainly used to solve complex optimization problems, covering many fields such as function optimization and combinatorial optimization. Compared with traditional optimization algorithms, differential evolution algorithm has stronger global search ability and faster convergence speed, and is suitable for dealing with high and discontinuous optimization problems [7]. The schematic diagram of differential evolution algorithm is shown in Figure 1.

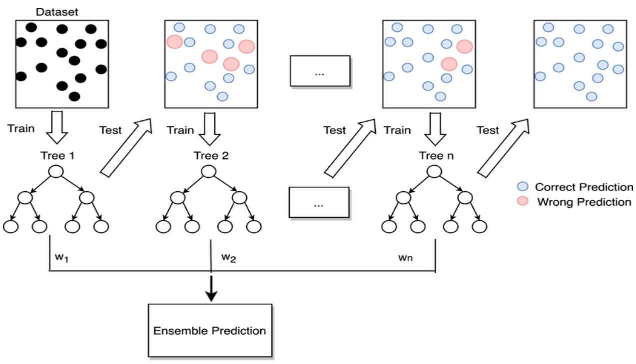

Figure 1: The schematic diagram of differential evolution algorithm.

The core idea of differential evolution algorithm is to generate new candidate solutions through three operations: variation, crossover and selection. In the initialization phase, the algorithm randomly generates a population of individuals, each representing a possible solution. Then in each iteration, the DE algorithm randomly selects three different individuals from the current population and uses the differences between them to generate new individuals. This process is called "mutation." Specifically, for a target individual, differential evolution generates a new candidate solution by adding the difference vector of two individuals to the third individual after weighted scaling. The new

DOI: 10.54254/2755-2721/116/2025.20430

solution and the original individual are combined through cross-operation to form a survival competition of candidate solutions, and finally select the solution that makes the objective function reach the best as the next generation population.

The effectiveness of DE algorithm lies in its adaptability and global searching ability. Through well-designed variation and crossover schemes, DE is able to maintain diversity well and avoid falling into local optimal solutions. In addition, the algorithm usually adopts an adaptive control strategy to dynamically adjust variation and crossover parameters to improve the exploration ability and convergence speed of the algorithm. At the same time, DE algorithm has strong robustness and can adapt to various types of problems, which makes it perform well in practical applications. In a variety of fields, such as image processing, machine learning, engineering design, the application of differential evolution algorithms continues to expand, showing its broad potential as an optimization tool.

3.2. Bidirectional long short-term memory network

Bidirectional Long Short-term memory network (BiLSTM) is an improved recurrent neural network (RNN) designed to solve some of the limitations faced by traditional RNNS in processing sequence data, especially the problem of long-term dependence. LSTM networks, proposed by Hochreiter and Schmidhuber in 1997, can effectively store and transmit information by introducing a gating mechanism to avoid the problems of gradient disappearance and explosion. By considering both forward and reverse information of the sequence, bidirectional LSTM improves the ability to understand and capture contextual information, making it widely used in speech recognition, natural language processing and time series prediction [8]. The schematic diagram of the bidirectional long short-term memory network is shown in Figure 2.

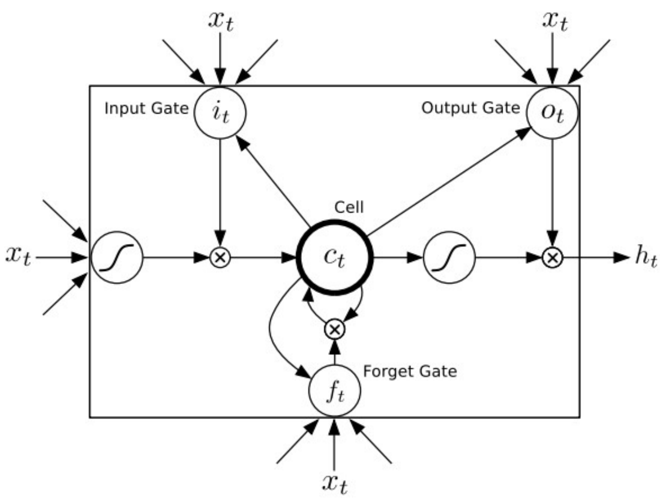

Figure 2: The schematic diagram of the bidirectional long short-term memory network.

The structure of bidirectional LSTM network consists of two parts: forward LSTM and backward LSTM. These two parts are separate LSTM networks, each processing the same input sequence, but in opposite directions. The forward LSTM reads data from the beginning of the sequence and gradually passes information to the end of the sequence, and then the forward LSTM reads data from the end of the sequence and gradually passes information back to the starting position. This bidirectional approach enables the model to consider the context information of the current time step at the same time, thus generating a richer representation of context features. When processing the output of a specific time step, BiLSTM is able to combine the output of the forward and backward LSTM to produce more accurate and comprehensive results.

DOI: 10.54254/2755-2721/116/2025.20430

3.3. Transformer

Transformer is a deep learning model based on a self-attention mechanism, first proposed by Vaswani et al., in 2017. Unlike traditional recurrent neural networks (RNNS) and convolutional neural networks (CNNS), Transformer relies entirely on self-attention mechanisms to process sequence data, making it excellent at parallelizing training and long distance dependent capture. Transformer's architecture is mainly composed of encoders and decoders, which are widely used in natural language processing (NLP) tasks, such as machine translation, text generation and semantic understanding [9]. The structure diagram of Transformer is shown in Figure 3.

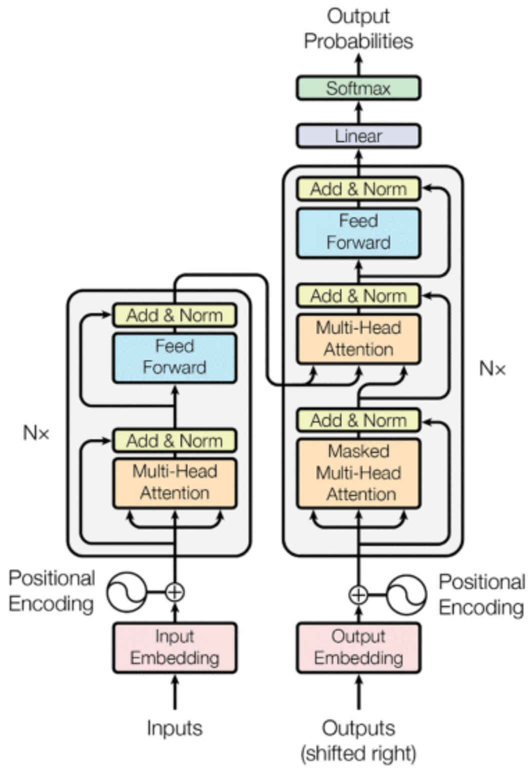

Figure 3: The structure diagram of Transformer.

At the heart of the Transformer model is the Self-Attention mechanism, which allows the model to process each element of the input sequence while taking into account information from all other elements in the sequence. In Transformer, the input sequence is first mapped to three different vector Spaces called Query, Key, and Value. By calculating the similarity of the query to the key, the model is able to determine the relationships and importance between different locations, and then perform weighted summations of the values based on those relationships to generate context-sensitive representations. This mechanism gives Transformer the flexibility to capture complex relationships in input structures. In addition, each layer of Transformer contains Multi-Head Attention mechanisms to learn information representation in parallel through different shadow casting Spaces, thus enhancing information capture capabilities [10].

3.4. Optimization of Transformer based on differential evolution algorithm and bidirectional long short term memory network

The Transformer model has shown remarkable success in natural language processing (NLP) and other sequential tasks, but its performance and efficiency often depend on the setting of hyperparameters. This makes optimizing Transformer hyperparameters a key step in improving model performance. Differential evolution algorithm is a powerful global optimization algorithm that can be used to automatically adjust these hyperparameters. By combining BiLSTM with differential evolution algorithms, you can effectively discover task-based hyperparameter configurations to further enhance Transformer's performance.

This combination method based on differential evolution algorithm and bidirectional LSTM optimization Transformer takes full advantage of the powerful and adaptive optimization strategy of the self-attention mechanism to form a robust optimization framework. In practice, with this optimization, the Transformer model can better adapt to data characteristics for specific tasks, improving performance while reducing the risk of overfitting.

4. Result

The software used in this experiment is Matlab, the graphics card is 3080, and the memory is 32GB. In terms of parameter Settings, the optimizer selects Adam, the maximum training epoch is set to 200, the batch size is set to 256, the initial learning rate is set to 0.0001, the learning rate reduction factor is set to 0.1, and the gradient clipping threshold is set to 10.

First, the changes of loss and accuracy of the training set in the training process are output, and the results are shown in Figure 4.

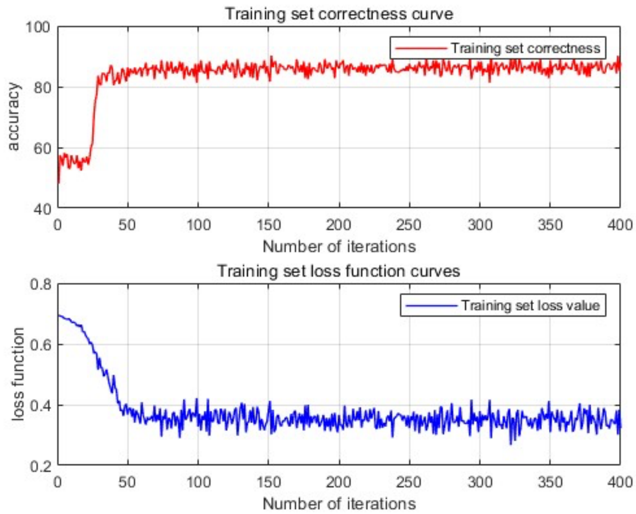

Figure 4: The changes of loss and accuracy of the training set in the training process.

As can be seen from the change diagram of the accuracy curve of the training set, the accuracy of the model gradually increases and becomes stable with the increase of the number of iterations. This shows that the model gradually ADAPTS to the data during the learning process and achieves good performance on the training set. Ideally, we want to see a smooth upward curve, which means that the model is constantly learning and improving its predictive power. At the same time, the loss curve gradually decreases and tends to be stable, which further verifies that the model has a good performance in forecasting.

Output a predictive confusion matrix for the training set, as shown in Figure 5. Output the confusion matrix predicted by the test set, as shown in Figure 6.

DOI: 10.54254/2755-2721/116/2025.20430

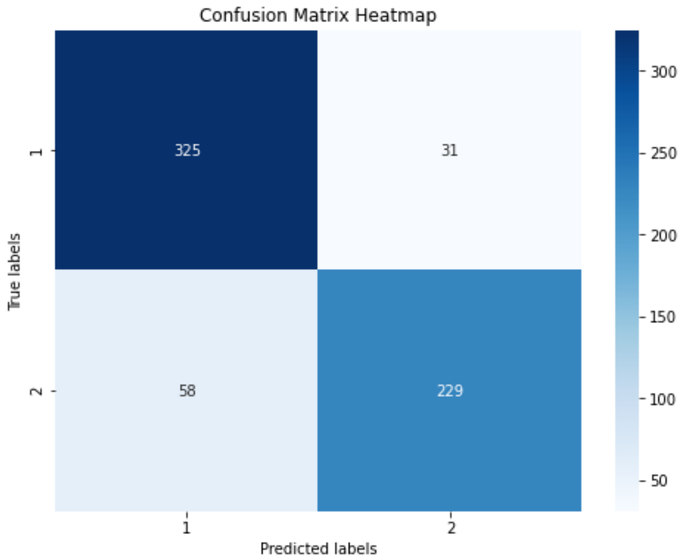

Figure 5: The confusion matrix.

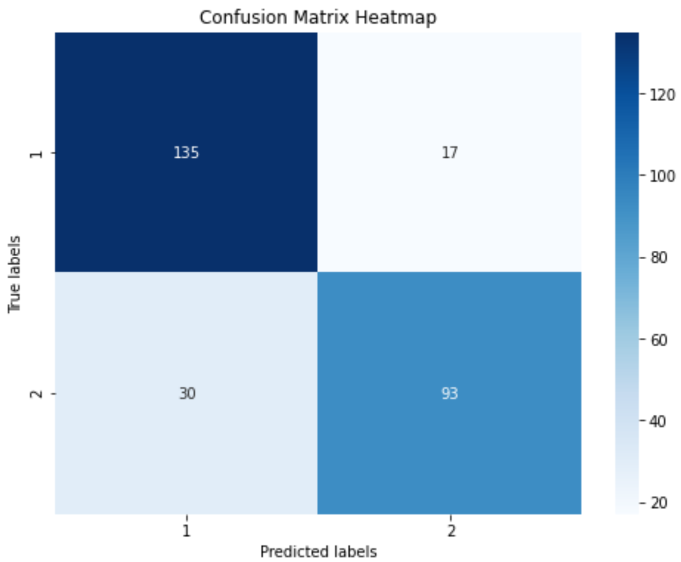

Figure 6: The confusion matrix.

The numbers in the confusion matrix represent the relationship between the actual class and the class predicted by the model. The numbers on the diagonal (325 and 229) represent the number of correctly classified samples, not the numbers on the diagonal (31 and 58) represent the number of incorrectly classified samples. From these figures, it can be seen that the prediction accuracy of the training set reached 86.15%, and the prediction accuracy of the test set reached 82.91%. The prediction accuracy of the model is high, it can predict the credit situation of bank customers well, and the generalization ability of the model is also good.

DOI: 10.54254/2755-2721/116/2025.20430

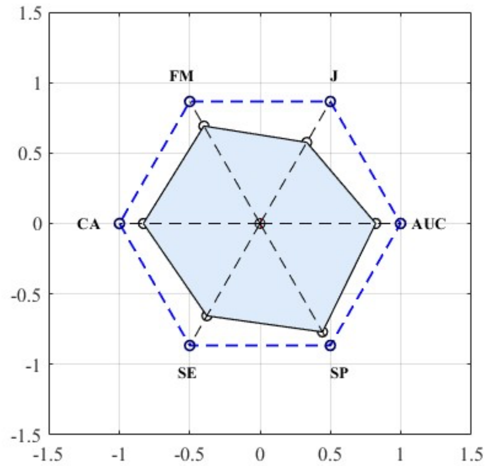

Figure 7: Other parameter index.

Figure 7 provides additional indicators of model performance, including FM (F1 score) and DAUC (area under the distinguished curve). The F1 score is the harmonic average of accuracy and recall, which measures the accuracy and completeness of the model in binary classification problems. DAUC is a measure of a model's ability to distinguish between two categories. As can be seen from the figure, F1 scores and DAUC values are both positive, indicating that the model has a good prediction effect.

5. Conclusion

This paper discusses the optimization of Transformer model using differential evolution algorithm and Bi-LSTM to improve the accuracy of bank customer credit analysis and prediction. By introducing this optimization model for training, we observe that with the increase of the number of iterations, the accuracy of the model steadily improves and eventually becomes stable. This trend indicates that the model gradually learns and ADAPTS to the data features during training, thus achieving high prediction accuracy on the training set. At the same time, with the progress of iteration, the loss value of the model also showed a downward trend and finally stabilized at a low level, which further confirmed the superiority of the model in forecasting performance.

In the analysis of confusion matrix, we found that the accuracy of the training set reached 86.15%, and the accuracy of the test set reached 82.91%. This result not only shows that the model has excellent performance on the training data, but also shows that the model has good generalization ability on the unseen data. This ability to predict with high accuracy is crucial for bank credit analysis, as it helps banks to more accurately assess their customers' credit risk and thus make more informed credit decisions.

In addition, the F1 score and the area under the differentiation curve (DAUC) values of the model are both positive, which further confirms the superiority of the model prediction effect. The F1 score is a harmonic average of accuracy and recall, which takes into account the accuracy and completeness of the model, while the DAUC value measures the model's ability to distinguish between different categories. The positive values of these two indicators indicate that the optimized model can not only accurately identify defaulting customers, but also reduce misjudgments when predicting the credit situation of bank customers, which has important practical significance for banks.

To sum up, the Transformer model proposed in this paper based on differential evolution algorithm and Bi-LSTM optimization performs well in the task of bank customer credit analysis and prediction.

DOI: 10.54254/2755-2721/116/2025.20430

The model not only achieves high accuracy on the training set, but also maintains good performance on the test set, showing strong generalization ability. In addition, the positive F1 scores and DAUC values of the model further prove the reliability of the prediction effect. Therefore, we can conclude that this model provides an effective tool for bank credit analysis, which can help banks to predict customers' credit risk more accurately, so as to optimize the credit decision-making process. This research result not only has important application value for the banking industry, but also provides a new technical path for the field of credit risk management.

References

[1] Lin, T., & Cao, J. (2020). Touch Interactive System Design with Intelligent Vase of Psychotherapy for Alzheimer's Disease. Designs, 4(3), 28.

[2] Chauhan, Priyavrat, Nonita Sharma, and Geeta Sikka. "The emergence of social media data and sentiment analysis in election prediction." Journal of Ambient Intelligence and Humanized Computing 12 (2021): 2601-2627.

[3] Ji, Rongrong, et al. "Survey of visual sentiment prediction for social media analysis." Frontiers of Computer Science 10 (2016): 602-611.

[4] Nguyen, Thien Hai, Kiyoaki Shirai, and Julien Velcin. "Sentiment analysis on social media for stock movement prediction." Expert Systems with Applications 42.24 (2015): 9603-9611.

[5] Nguyen, Le T., et al. "Predicting collective sentiment dynamics from time-series social media." Proceedings of the first international workshop on issues of sentiment discovery and opinion mining. 2012.

[6] Wang, Yilin, and Baoxin Li. "Sentiment analysis for social media images." 2015 IEEE international conference on data mining workshop (ICDMW). IEEE, 2015.

[7] Wołk, Krzysztof. "Advanced social media sentiment analysis for short-term cryptocurrency price prediction." Expert Systems 37.2 (2020): e12493.

[8] Chen, Liang-Chu, Chia-Meng Lee, and Mu-Yen Chen. "Exploration of social media for sentiment analysis using deep learning." Soft Computing 24.11 (2020): 8187-8197.

[9] Nguyen, Thien Hai, and Kiyoaki Shirai. "Topic modeling based sentiment analysis on social media for stock market prediction." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015.

[10] Derakhshan, Ali, and Hamid Beigy. "Sentiment analysis on stock social media for stock price movement prediction." Engineering applications of artificial intelligence 85 (2019): 569-578.