Research on Prediction of the Cash Usage in Banks Based on LSTM of Improved Grey Wolf Optimizer

Jingfeng Rong*, Di Wang
Liaoning University of engineering science, Fuxin 125105, Liaoning, China

Corresponding author’s e-mail: 823099577@qq.com

Abstract. In the real production and operation, it is impossible to predict the amount of cash in daily use. Therefore, the prediction model of improved LSTM neural network is proposed to cope with the problem for preparing excessive cash. Hence, the improved Grey Wolf Optimizer is most effective in searching for the optimal solution by optimizing the impact factors of Grey Wolf Optimizer. Combining the improved Grey Wolf Optimizer with LSTM neural network, the neural networking learning rate parameters are set reasonably by optimizing the algorithm to reduce the impact of inappropriate parameters on the prediction results of either over-fitting or under-fitting. What’s more, the neural network topological structure, weighing the number of LSTM network layers and the number of the neural units in each layer, determines the neural network’s description of data. If the network topology is too simple, the prediction results may not be enough to describe the real data. However, if it is too complex, it will not only waste the computing resources, but also make the prediction results over-fitting with only good description of training data. Therefore, it avoids the problem of large errors in predicting results caused by the parameters of neural network and realizes the prediction the daily cash usage. Finally, the test is completed on the data of a sub-branch network of bank with mean square error (MSE) 0.016. Compared with the traditional time series model ARAM and the unimproved LSTM, the improved LSTM predicts cash usage more accurately and efficiently.

Keywords. time series, finance, LSTM, Long Short-Term Memory, Gray Wolf Optimizer, particle swarm, in-depth study

1. Introduction
The stability of bank business makes great contribution to the harmony of society, which is mainly reflected in providing sufficient cash for depositors to cash freely and having sufficient scale to provide loan support services to enterprises, promoting the development of local economy. Since bank cash usage is closed related to the normal operation of each branch, it has become a decisive task that needs to be solved urgently to predict the usage, not only effectively ensuring normal operation, but also make reasonable use of cash. Currently, there is still little research of algorithm on predicting the cash usage. After analysis, it is found that the cash usage of sub-branch network of bank is a kind of time series data. According to this characteristic, the main predicting algorithm is still concentrated on using ARAM and SVM algorithm. For example, Guo Erfeng uses BP neural network to predict ATM cash flow in this paper. Both can predict the development trend of data for a period of time in the future. However, due
to its won limitations, the ARAM model has relatively large errors in long-term predictions. The grey neural networks is weak in generalization ability and in processing nonlinear information.

In recent year, with the continuous promotion of in-depth learning algorithms, the LSTM algorithm is prominent in predicting time series. Data with time series features is gradually using LSTM algorithm for prediction. For example, in the aviation operation system, Shi Qingyan adopts it to predict the flight path, reducing the instability and improving the utilization of airspace resources, hence the safety of aviation operations can be guaranteed. Chen Liang adopts it to predict the temperature and humidity of medicines, monitoring the medical storage and transportation environment, so as to make warning for unqualified pharmacies in advance.

Research results in recent years have shown that LSTM has advantages in predicting time series. Therefore, this paper aims at the problem of insufficient cash reserves and actual cash demand exposed by sub-branch network of bank in real operations. Taking a sub-branch network of bank in Fuxin area as the research object, the LSTM research is carried out on the basis of its real production data. It is found in the research that the data using traditional LSTM prediction will cause problems such as slow convergence of results and large errors, therefore, an improved LSTM algorithm is proposed, including network structure of input layer, hidden layer, and output gate. The hidden layer contains multiple LSTM neural unit structures, like forget gate, input gate and output gate. Compared with the former one, the improved GWO-U has better performance in the early search range and the later result convergence and it is more efficient and more accurate. The improved particle swarm-GWO-U optimize the initial parameters of the LSTM algorithm to avoid the prediction results because of the incorrect configuration of parameters with improved prediction accuracy and reduced error between the predicted value and the actual one. In the end, the results prove that the LSTM algorithm based on the improved GWO-U has more advantages after verifying its accuracy and performance by using the real data collected. The result is closer to the true value and provide algorithmic support for the cash prediction of sub-branch network of bank.

2. Modeling and theory

2.1. ARAM model

Form a statistical point of view, time series analysis is an important branch of statistics, and a significant research field for application based on random theory and statistics. Time series can be divided into stationary and nonstationary series according to their statistical characteristics. The Box-Jenkins method, first systematically proposed by British statistician G.M. Jenkins in 1970, is a relatively complete statistical prediction method with its function providing analysis and prediction for time series data in the real work and identification, assessment and diagnose for ARMA model. The advantage is that after an accurate model is established and parameters are determined, the trend can be predicted based on the existing data set, among which the ARMA model is commonly used for stationary time series.

2.2. LSTM neural networks

Long-Short Term Memory (LSTM) paper was first published in 1997. Due to its unique design structure, LSTM is suitable for processing and predicting some important events with very long intervals and delays in time series. There are three gain gate structures inside LSTM.

Forget gate: This gate mainly focuses on selectively forgetting the memory state $C_{t-1}$ passed from the previous node, namely forgetting the unimportant and remember the important. To be specific, through the sigmoid function, the calculated $Z^f$ ($f$ stands for forget) is used as the forget gate control unit to decide what needs to be left or forgotten for the previous state $C_{t-1}$.

$$f_t = \sigma(W_f [h_{t-1}, X_t] + b_f)$$  \hspace{1cm} (1)
Input gate: This gate controls input and selectively memorizes it, mainly for input $X^t$ with necessary information being recorded and the unnecessary less being memorized. The current input is represented by the previously calculated $Z$ and the selected gate signals is controlled by $Z^i$ ( $i$ stands for information).

Then the results obtained in the above two steps are added to get $C^{t-1}$ transferring to the next state, that is, the first formula in the figure above.

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

Output gate: This gate determines which will be output as the current state, being controlled mainly by $Z^0$ and scaling $C^0$ obtained in the previous state. (Changes through a tanh activation function)

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o)$$

$$C_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c)$$

$$C_t = f_t \ast C_{t-1} + i_t \ast \tilde{C}_t$$

$$h_t = O_t \ast \tanh(C_t)$$

2.3. Particle swarm-grey wolf optimizer

Grey Wolf Optimizer (GWO) is a swarm intelligence optimizer proposed by Mirjalili and other scholars at Griffith University in Australia in 2014, which is illuminated by the prey hunting activities of gray wolves and developed an optimized searching method with strong convergence, few parameters and easy implementation. Gray wolves, strictly abiding by a hierarchy of social domination, belong to the canines living in groups and are top-end creature in the food chain shown in Figure 1.

![Grey Wolves Hierarchy Structure](image)

**Figure 1.** Grey Wolves Hierarchy Structure

The first level: The first wolf in the wolf group is recorded as $\alpha$ who is responsible for making decisions while others obey his orders called dominant wolves.

The second level: The wolf $\beta$, dominating wolves on other social levels, obeys the wolf $\alpha$ and he assists the Wolf $\alpha$ in making decisions.

The third level: The wolf $\delta$ obeys the wolf $\alpha$ and $\beta$ while dominating the remaining levels of wolves.

The fourth level: The wolf $\omega$ usually need to obey the wolves on other social levels.

The GWO optimization process includes the following steps like social hierarchy of grey wolves,
tracking, encircling and attacking prey.

1) Social Hierarchy: When designing GWO, social hierarchy model of gray wolves should first be constructed with the ranks from high to low as $\alpha$, $\beta$, $\delta$, $\omega$.

2) Encircling Prey: When gray wolves seek its prey, it will gradually approach and encircle it. The mathematical model of this behavior is shown in the following formula.

$$D = C \circ X_p(t) - X(t)$$ (8)

$$X(t + 1) = X_p - A \circ D$$ (9)

$$A = 2a \circ r_1 - a$$ (10)

$$C = 2r_2$$ (11)

**Formula: The Mathematical Model of Encircling Prey**

3) Hunting

Gray wolves have the ability to identify the location of potential prey (optimal solution). The searching process is mainly completed by the guidance of the wolf $\alpha$, $\beta$ and $\delta$. The mathematical model of this behavior is shown in the following formula.

$$D_\alpha = C_1 \circ X_\alpha - X$$ (12)

$$D_\beta = C_2 \circ X_\beta - X$$ (13)

$$D_\delta = C_3 \circ X_\delta - X$$ (14)

$$X_1 = X_\alpha - A_1 \circ D_\alpha$$ (15)

$$X_2 = X_\beta - A_2 \circ D_\beta$$ (16)

$$X_3 = X_\delta - A_3 \circ D_\delta$$ (17)

$$X_{(t+1)} = \frac{x_1 + x_2 + x_3}{3}$$ (18)

**Formula: The Mathematical Model of Hunting**

4) Attacking Prey: In the process of constructing the attacking prey model, according to the formula in Figure 2, the value A will fluctuate with the decrease of the value a. In other words, A is a random vector on the interval $[-a, a]$, where a decreases linearly during the iteration. When A is in the interval $[-1, 1]$, the next position of the Search Agent can be anywhere between the current grey wolf and the prey.

5) Search for Prey: Grey wolves mainly rely on the information from the wolf $\alpha$, $\beta$ and $\delta$ to search for prey. They scatter to search for it and then gather to attack the prey.

2.4. Improved Grey Wolf Optimizer

Aiming at the problem that the parameter a of GWO (Equation 19) decreases linearly in the iterative process, we assume that when max is 100, the image of a is as follows.

$$a = 2 - 2 \left( \frac{t}{max} \right)$$ (19)

The Formula of Parameter a
Parameter A will change with parameter a and when $|A| > 1$ makes the search agent far away from the prey, this search method enables GWO to perform a global search. Therefore, appropriately adjusting the curve of a make the change of A accordingly, which enhances the ability of the function to search globally in the early stage, and local search in the later stage. From the image below, the value range of parameter A of the original function is in the interval $[0,2]$.

Optimize the parameter a in this formula 4 to make the data slowly decrease in the early stage, rapidly decrease in the middle stage and then slowly decrease again in the final stage as it is shown in Figure 3.

$$\varphi(t) = \exp\left[-\frac{t^2}{(2\delta)}\right]$$  \hspace{1cm} (20)  
$$\delta = \frac{T}{3}$$  \hspace{1cm} (21)

**Figure 2.** (a) Parameter a when max is 100; (b) The Image of A When Max is 100 and Parameter is a
Figure 3. (a) The curve after Optimization of Parameter $a$; (b) The Image of A When Max is 100 and Parameter is $a$

It can be seen that the image is more in line with our requirements, which can enhance the global searchability of the wolf $\alpha$, $\beta$ and $\delta$ in the early stage of GWO and better search the target direction so as to avoid falling into local traps, hence, the quickly local search can be achieved in the later stage. Combine the GWO in the standard test function: multi-modal standard test function.

$$f(x) = x^2 - 10 \cos(2\pi x) + 10$$ (22)

The original formula GWO and the improved GWO-U are used respectively to get solution of minimum value of function under the standard test in groups. It is verified in the experiments that when the maximum number of iterations is set to 300, the test results are shown in Table 1.

Table 1. The Solution to Multi-modal Function
From the result of the text, the optimized GWO is more effective than original GWO when solving the minimum value of the multi-modal function.

| Multi-modal Function | Number of iterations in the result | The value of X (Minimum function value) | True value |
|---------------------|-----------------------------------|---------------------------------------|------------|
| GWO                 | 300                               | 1.7053025658242404e-13                | 0          |
| GWO-U               | 168                               | 0                                     | 0          |

2.5. The algorithm based on GWO-LSTM

The main idea of the improved algorithm is to take advantage of GWO and its ability to quickly find solution to the minimum value of the function to search for the network parameters of LSTM. Based on the loss value of LSTM, more suitable network parameters can be found. Specific steps are as follows.

Step 1: Initialize the parameter range such as the population size of the GWO and create an initial range for input number, neural unit, network layer number and learning factors of optimal target LSTM.

Step 2: Determine the fitness function of GWO, and take the mean square error of the LOSS function after training as the evaluation function, such as the formula.

\[
\text{loss}(x, y) = \sum_{i=0}^{n} \sum_{j=0}^{m} (a_{ij} - b_{ij})^2
\]

Formula: The function of LOSS

Step 3: Initialize the positions of the wolf \(\alpha, \beta, \delta\) and \(\omega\) within the range of the optimized parameters and complete the social grading.

Step 4: Complete the process of encircling prey, hunting and attacking prey.

Step 5: Complete the updating of particle location.

Step 6: Iterate steps three and four until the constraint of iteration is reached.

Step 7: Obtain the optimal network parameters, use it to train model and finally output the value predicted.

The complete steps are shown in the following figure 4.
2.6. The combined model based on GWO-U-LSTM

The iterative training of data makes it consume computer resources during training to a large extent, and

![Diagram of GWO-LSTM procedure](image-url)
the complexity of calculation leads to low efficient training. How to save resource occupied and improve the training efficiency? From the data of GWO-U model, it can be seen that GWO-U has more advantages than the original formula GWO in terms of calculation efficiency and time. The structure of the combined model is shown in Figure 5.

![Figure 5. The structure of combined model of GW)-U-LSTM](image)

3. The collection of data and pre-treatment

3.1. The collection of data

In order to meet the daily cash requirements of sub-branch banks, the general approach is to prepare a large amount of cash, that is, receiving excessive cash every morning and turning over the remaining cash to the higher-level treasury after a day of operations. In this survey and research, the data we use is from the data warehouse of a sub-branch bank of Fuxin from January 2014 to December 2019.

3.2. Pre-treatment, research and judgement of data

Traverse the acquired data, select a null value and then fill it with the monthly average value. It is displayed through two-dimensional images with the help of the drawing function of python. The vertical coordinate is the time axis in the series of time, a day as a unit and horizontal coordinate indicates the amount of cash used at the end of each day. The negative direction shows the balance of the cash, while the positive one shows the cash in use.

From the analysis of the image, it is found that during the period from 2014 to 2019, the sequence of the sub-branch bank contains the medium and long-term trend of the curve. Through the research and judgment of the trend, it is found that the trend is stable and does not have an overall upward or downward trend.

4. Simulation results and analysis

4.1. The setting of simulation condition
The training data was taken from the cash usage data of a sub-branch bank in Fuxin area from January 1, 2014 to April 30, 2019 with 1,920 pieces. The test data was taken from May 1, 2019 to November 30, 2019 with 240 pieces.

4.2. The evaluation criteria in comparison of simulation results

MSE: This statistic parameter is the mean value of the sum of squares of the errors of the predicted data and corresponding points of the original data, such as the formula below.

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} w_i (y_i - \hat{y})^2 \]  
Formula 5: MSE

RMSE: Measure the deviation between the observed value and the true value, such as the formula below.

\[ \text{RMSE}(X, h) = \left( \frac{1}{m} \sum_{i=1}^{m} (h(x_i) - y_i)^2 \right)^{1/2} \]  
Formula: RMSE

MAE: The average value of absolute error, which is actually a more general form of error average, such as the formula below.

\[ \text{MAE} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \]  
Formula: MAE

MAPE: Mean absolute percentage error, like the formula below.

\[ \text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \]  
Formula 8: MAPE

SMAPE: Symmetric mean absolute percentage error, like the formula below.

\[ \text{SMAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \right| \]  
Formula: SMAPE

4.3. Preparation before Simulation

Normalize the data and map the data to [0,1], so that the processed data will make the indicators in the same order of magnitude after the original data being processed through standardization. Then it is suitable for comprehensive comparative evaluation, and speed up the solution of the gradient descent.

4.4. LSTM analysis based on GWO-U

In order to compare the various methods, the initial parameter of LSTM input is set as 2 and the number of neural units as 64, the layers of LSTM network as 3. The prediction model is finally obtained using initial function LSTM model, LSTM-GWO model and improved LSTM-GWO-U model after 300 integrations of training. Then the prediction data of each model are finally obtained after the tests of each prediction model using test data, as it is shown in Figure 6.
The predicted values and the real value of contrast

a) Comparison of LSTM predicted data with real data

b) Comparison of LSTM-GWO predicted data with real data
c) Comparison of LSTM predicted data with real data based on GWO-U

![The predicted values and the real value of contrast](image)

The error result data of each model is obtained through simulation experiment, as it is shown in Table 2.

**Table 2. Various Error Data of Experimental Results**

|       | ARAM  | LSTM  | LSTM-GWO | LSTM-GWO-U |
|-------|-------|-------|----------|-------------|
| MSE   | 0.05235 | 0.0302334 | 0.017327936 | 0.01647721 |
| RMSE  | 0.22880 | 0.17387754 | 0.13163562 | 0.12836358 |
| MAE   | 0.19269 | 0.13005461 | 0.10531602 | 0.10264468 |

The three predictions of the LSTM and the error comparison of the data in each iteration, as it is shown in Figure 7.
4.5. Horizontal comparison

Taking another sub-branch bank in Fuxin area as the research object, this comparative test collected the data from May 30, 2019 to November 30, 2019 to verify the fitting effect of the neural network LSTM based on the GWO-U in order to discover both the advantages and disadvantages of the algorithm (Table 3).

| MSE        | RMSE    | MAE       | MAPE        | SMAPE       |
|------------|---------|-----------|-------------|-------------|
| 0.18991241 | 0.4357894 | 0.3481489 | 74406.06079101562 | 162.3202323913574 |

4.6. Experiment results

Through comparing the results of various prediction models, it can be seen that the error using LSTM to predict time series data is much smaller than that of using traditional ARAM models. The neural network LSTM based on GWO-U optimal algorithm has improved to some extent in the error of the prediction results. The optimal solution can be found among a large number of parameters matching so as to reach the best parameter matching ratio of the model, solving the problem of poor prediction results of the model due to improper use of parameters. Moreover, the iterative error convergence of neural network LSTM based on the GWO-U is more stable and efficient which enables the model to better describe predicted results, closer to the true value and more practical.

5. Conclusion

The improved particle swarm GWO-U has faster solution and better convergence. Through experiment and analysis of experimental results, it is confirmed that the improved LSTM has better prediction performance and faster convergence. Compared with traditional LSTM, its error prediction result proves the effectiveness of LSTM algorithm based on GWO-U. What’s more, the prediction on cash is closer to the real value, which can provide reference for the real work.

However, the predicting method in this paper still has some shortcomings, such as inability to accurately predict the use of cash with human intervention and the subjective factors from corporate customers. The next step is to further study the customers’ behavior and habits and perform weighted calculations on the predicted data.
References

[1] WenWenBo, Thu wai. Summary of ant colony algorithm. Journal of petrochemical industry automation, no. 1, pp. 19-22, 2002.

[2] Shi-jie Zhao, Gar Mound, Dong-mei yu, ACTS. Fusion energy cyclical decline and an improved algorithm of HHO Newton local enhancement. Control and decision: 1-8 [2020-07-06]. HTTP: / / https://doi.org/10.13195/j.kzyjc.2019.0810.

[3] Heidari A., Mirjalili S., Faris H., et al., HarrisHawks Optimization, Algorithm and Applications. Journal of Future Generation Computer Systems, no. 97, pp. 849-872, 2019.

[4] Fan-hui Zeng. Numerical Analysis. Publishing house of China university of Mining, 2009: 25-30. (ZENG Fanhui. Numerical Analysis. ChinaUnivesity of Mining, and Technology Press, 2009: 25-30.)

[5] Wu Yue, Ye Hourong, Qi Zheng, Guang-xu Liu. Mathematical dictionary: higher education press, 1989-07

[6] Wang Yuqing, Ren Xincheng. Quantity order of magnitude estimate of the study. Journal of college physics, vol. 23, no. 12, pp. 30-32, 2004.

[7] Bo Wei, Xuewen Xia, Fei Yu, Yinglong Zhang, Xing Xu, Hongrun Wu, Ling Gui, Guoliang He. Multiple adaptive strategies based particle swarm optimization algorithm. Swarm and Evolutionary Computation, 2020, 57.

[8] Gomes G. F., da Cunha S. S., Ancelotti A. C., A Sunflower Optimization (SFO) Algorithm Applied to Damage Identification on Laminated Composite Plates. Engineering with Computers, 2018: 1-8.

[9] Sadollah A., Sayyaadi H., Yadav A. A DynamicMetaheuristic Optimization Model Inspired byBiological Nervous Systems: Neural Network Algorithm. Applied Soft Computing, no. 71, pp. 747-782, 2018.