Network security situation prediction method based on strengthened LSTM neural network

Dong Zhaowei, Su Xiaoyu, Sun Lihui, Xu Kuikui

1 Computer Software and Theory, Hebei University of Economics and Business, Shijiazhuang, Hebei Province, 050700, China
2 Computer Software and Theory, Hebei University of Economics and Business, Shijiazhuang, Hebei Province, 050700, China
a email: sunlihui@heuet.edu.cn

Abstract: As an important part of network security situation awareness, network security situation prediction describes the dynamic changes of security situation over time, and predicts future situation values based on historical situation values. In order to improve the accuracy of network security situation prediction, a long- and short-term memory network security situation prediction model based on the Sigmoid weighted reinforcement mechanism is proposed. Firstly, LSTM neural network is used to mine the temporal correlation of network security situation data. Sigmoid weighted linear element is introduced to deal with the gradient problem in the back propagation, and the input value is multiplied by Sigmoid activation function, so as to strengthen the structure of LSTM neural network and improve the accuracy of prediction. Then, the cuckoo search algorithm was used to optimize the super parameters to improve the training time. Finally, the public data set CICIDS2017 was used to verify the model. The simulation experiment results show that the model has a faster convergence rate and smaller errors.

1. Introduction

For increasingly complicated and diversified network attack, the traditional protection measures such as firewall, intrusion detection and access control, vulnerability scanning from two aspects of active defense and passive defensive protection, although these security devices with functions of security incidents and security log records, but independent among equipment, safety information is scattered, not to be Shared. It is difficult for the network security administrator to monitor the global network condition and make appropriate decisions when an attack occurs. In order to solve the above problems, T.Bass[1] introduced the concept of situational awareness into the field of network security for the first time. As the main research part of situational awareness, situational prediction builds a prediction model based on the analysis of historical situational data, obtains the rules between situational data, and predicts the future changes of the network.

The change of network security situation is characterized by complexity, nonlinear and time-varying characteristics, and the neural network technology has higher fault tolerance, stronger nonlinear mapping and generalization ability for complex systems. Therefore, some scholars apply neural network technology to the field of network security situation prediction. Literature [2] proposed a network security situation prediction method based on LSTM. LSTM neural network is an improvement of recursive neural network (RNN) and has strong performance in processing time series data. Literature [3] proposed the network traffic prediction method based on LSTM, and introduced the particle filter...
constraint algorithm to optimize the network parameters. In the process of network training based on LSTM neural network, the slow convergence speed affects the training cost. Literature [4~10] proposes to adopt intelligent optimization algorithm to improve the convergence speed of LSTM neural network model. Literature [11] proposed a prediction method based on two-layer recursive neural network LSTM and GRU, which combined two kinds of improved recursive neural network. Although the prediction accuracy was improved, the complexity of the model increased, and the time of training the model was also prolonged. In this paper, a security situation prediction method for LSTM network based on Sigmoid weighted reinforcement mechanism is proposed, and cuckoo search algorithm is introduced to automatically optimize the hyperparameter to accelerate the training time of the network.

2. Network security situation prediction based on improved LSTM

2.1. Activation function in the LSTM neural network

As a special recursive neural network (RNN), LSTM network is different from the traditional RNN recursive structure. LSTM has a stable and powerful ability in solving long-term and short-term dependence problems. Memory cells replace the hidden layer of traditional neurons and are the core of LSTM network. Since there are three gates, the input gate, the output gate, and the forgetting gate, the LSTM network can add or remove information to the cell state. The prediction performance of LSTM neural network mainly depends on the activation function, the input vector of the activation function includes the current input and the previous state, and then predicts the output according to the results of the hidden layer.

Through analyzing and studying the function of activation function, it is found that each activation function has its own advantages and disadvantages. The sigmoid function maps any input value range between 0 and 1. The function compresses the large input value into a small output range. In the process of back propagation, the gradient of the function drops exponentially, resulting in a small weight adjustment in the initial layer of the neural network. The acceptable input range of the tanh function is between -1 and 1, so that the output of negative input can be calculated. Therefore, the vanishing gradient of the tanh function is smaller than that of the sigmoid, but this problem still exists. Later, in order to find a better activation function to alleviate the gradient problem, research scholars studied the corrected linear unit (ReLU) activation function. The ReLU function only contains two simple constraints, as shown in formula (3). The ReLU function maps all negative values to 0, and other values to \( x \), so even small input values reflect significant changes in output. However, there is a problem with the ReLU activation function. In the derivative of the ReLU activation function, all negative values of \( x \) are mapped to zero, making the network stop learning, which will cause the neuron to stop learning after a certain number of continuous inputs without gradients.

2.2. Neural network model based on SiLU-LSTM

This paper proposes multiplying the input value and the sigmoid function of the input gate, and multiplying the input value and the tanh function in the candidate vector to reduce the impact of the vanishing gradient problem, so that LSTM has a more complex structure to capture the input layer and the hidden layer the recursive relationship between.

The structure diagram of SiLU-LSTM proposed in this paper is shown in Figure 1.
When deciding to keep the current input, the network learns the new information carried by the current input by comparing the current input with the previous state. The previous state vector $h_{t-1}$ contains the contents of the previous input and output changes. On this basis, the range of the sigmoid function in the first input gate is a value between 0 and 1, which describes the range of new information available in the current input, and the obtained value is multiplied by the input value, as shown in formula (1). Now the range of the input gate value is $[0, +\infty)$, avoiding the vanishing gradient of the same value range in the input vector, so that the network can learn from each input.

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

For the candidate vector $C_{t}$, the range of the tanh function is a value between -1 and +1. The obtained value is multiplied by the input value, and the input produces a candidate vector between $(-\infty, +\infty)$, as shown in formula (2). This in turn improves the ability to learn and update the new information in the cell state vector, and makes the time series change smaller.

$$C_{t} = x_t \ast tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$  \hspace{1cm} (2)

The memory cell output of LSTM is $h_t$ (current hidden state) and $C_t$ (current storage state). The output gate and hidden state of LSTM neural network are shown in formulas (3) and (4).

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

$$h_t = o_t \ast tanh(C_t)$$  \hspace{1cm} (4)

3. Cuckoo search algorithm optimizes model hyperparameters

3.1. Cuckoo search algorithm
Cuckoo Search (CS), also known as Cuckoo Search, is an algorithm inspired by competitive breeding patterns. Standard CS uses three simplified rules:(1) Each cuckoo lays an egg at a time, and randomly selects a parasitic nest to hatch it, and each egg corresponds to a solution vector; (2) In a randomly selected set of parasitic nests, the best parasitic nest (solution) will be retained to the next generation; (3) The number of available parasitic nests is fixed, and the probability that the owner of a parasitic nest can find a foreign bird egg is $P_a$. In this case, the host can choose to remove the egg, or choose to abandon the nest and build a new nest in a new place.

In this simplified scenario, each nest corresponds to an egg, and this egg also represents a cuckoo. Mathematically speaking, cuckoo search uses a combination of local random walk and global random walk, controlled by probability $P_a$. Local random walk, as shown in formula (5).

$$x_t^{f+1} = x_t^f + \beta s \bigotimes H(p_a - \varepsilon) \bigotimes (x^f - x_k^f)$$  \hspace{1cm} (5)

Among them, $x_t^f$ and $x_k^f$ are two different solutions selected by random permutation. $H(u)$ is the Heaviside function (unit step function). $\varepsilon$ is a random number extracted from a uniform distribution and $s$ refers to the step length. In addition, $\bigotimes$ represents the dot product of two vectors and $\beta$ represents a small scale factor.
The global random walk using \( \text{Levy} \) flight is shown in formula (6).

\[
x^{i+1} = x^i + \alpha \otimes L(s, \lambda)
\]

\[
L(s, \lambda) \sim \frac{\Delta^2(\lambda) \sin(\pi \lambda / 2)}{\sqrt{s}} (s \gg 0)
\]

Among them, \( \alpha \) is the step-size scale factor, which is related to the scale of the problem to be solved. "\( \sim \)" emphasizes that the steps of searching according to random numbers \( L(s, \lambda) \) should be drawn from the distribution \( \text{Levy} \). Formula (7) represents an exponential idempotent law distribution approximate to the distribution. The use of flight search makes the algorithm more traversal and explorable, and the algorithm is easier to jump out of any local optimal solution.

4. Network security situation prediction model construction process

Step 1 The data set of network security situation value is divided into training set and test set. Input selects the current sequence input \( M \) and the previous \( M-1 \) situation value, and the output of the model is the situation value at the time of \( m+1 \).

Step 2 Initialize the LSTM neural network, with \( m \) input layer nodes and 1 output layer node, randomly generate the number of iterations, learning rate, and the number of hidden layer nodes, and initialize and optimize the change interval of hyperparameters.

Step 2.1 Set the number of nests \( \text{Nest} \) to 10, and the probability of bird eggs being found by the owner of the nest is \( \text{Pa} \) to 0.25. Randomly generate the position of the bird's nest. The position of each bird's nest contains 4 parameters. Train according to the various parameters of the bird's nest initialization, calculate the predicted value of each bird's nest, find the bird's nest position with the smallest error according to the prediction error, and save it to the next generation.

Step 2.2 Update the bird's nest according to the position and path update formula (5), calculate the prediction error through the LSTM neural network, and compare it with other bird's nests except the smallest bird's nest in the previous step. The bird's nest with the smallest error is obtained instead of the largest error Bird's nest, and get the current optimal bird's nest position.

Step 2.3 Get the optimal bird's nest, if the accuracy requirement is met, return to the previous step to continue searching, otherwise output the current optimal value.

Step 3 Input training data into the neural network model for training. If the prediction accuracy is not reached or exceeds the iteration range, continue learning; if the number of iterations or prediction accuracy is reached, the neural network learning is stopped and the current optimal hyperparameter combination is saved.

Step 4 Construct the forget gate, and initialize its offset to 1, which is used to reduce the excessive information that is forgotten in the initial stage of training.

Step 5 Construct the input gate and candidate vector. The input value \( x_t \) at the current moment and the output value \( h_{t-1} \) at the previous moment pass through the input gate and candidate vector, and update the cell state \( C_t \) through the formula (5) and (6).

Step 6 The process of constructing the output gate is consistent with the traditional model, and the output value at the current moment is obtained through \( O_t \) and \( C_t \).

Step 7 After calculating the output value, in order to verify the accuracy of the algorithm, input the test sample data set into the prediction model to obtain the predicted situation value as the output result of the model, construct the mean square error function MSE of the prediction model as the objective function of the model, and use Adam the algorithm updates the weight and bias of the neural network until the training error of the model reaches the preset goal and saves the model.

5. Simulation experiments and analysis

5.1 Selection of experimental data and model parameters

This article focuses on the CIC-IDS2017 data set proposed by the Canadian Institute of Cyber Security. The data set contains normal data traffic and the latest attack traffic. All traffic is collected in a real environment and has strong reliability. According to the network security situation assessment method
Based on attack detection, the network security situation value is obtained, and the situation value curve is drawn, as shown in Figure 2.

![Network security situation value](image)

**Figure 2 Network security situation value**

The basis of network security situation prediction is network security situation assessment. In order to verify the effectiveness of the model proposed in this article, this article composes the prediction sample set of this chapter through the network security situation values obtained, as shown in Table 1.

| input data | output data |
|------------|-------------|
| 0.43 0.42 0.35 0.45 0.4 0.41 0.42 | 0.42 |
| 0.42 0.35 0.45 0.4 0.41 0.42 0.5 | 0.5 |
| 0.35 0.45 0.4 0.41 0.42 0.5 0.48 | 0.48 |
| 0.45 0.4 0.41 0.42 0.5 0.48 0.49 | 0.49 |
| 0.4 0.41 0.42 0.5 0.48 0.49 0.45 | 0.45 |

Table 1 Partial sample set of prediction model

In the LSTM neural network, the hyperparameters of the model play a vital role in the performance of the model. According to the Sigmoid weighted reinforcement LSTM neural network proposed in this paper, the activation functions of the input layer and candidate vectors are set to x*sigmoid and x*tanh, and the cuckoo search algorithm is used to optimize the structure parameters of the LSTM neural network, including the number of iterations and the learning rate, the number of nodes in the two hidden layers. The parameter settings for cuckoo search include setting the number of nests Nest to 10, and the probability of bird eggs being found by the nest owner is Pa to 0.25. According to the situation data obtained from the network security situation assessment, in view of the periodic characteristics of the situation time series, the sliding window is set to 6, the first 6 values are selected as the input data, and the data at the next moment as the output data, thereby determining the input neuron It is 6 and the output neuron is 1.

5.2 Experimental results and analysis

This article composes the sample set of network security situation prediction through the situation value after network security situation assessment, and compares and analyzes the original LSTM activation function and the most widely used ReLU, Leaky ReLU and the activation function proposed in this article. The prediction performance of each method as shown in Table 2.
Table 2 Comparison of activation functions

| Model                  | RMSE   | MAE    | R2   |
|------------------------|--------|--------|------|
| CS-LSTM                | 0.0843 | 0.0549 | 0.869|
| CS-SiLU-LSTM           | 0.0577 | 0.0398 | 0.955|
| CS-RELU-LSTM           | 0.0777 | 0.0488 | 0.889|
| CS-Leaky ReLU-LSTM     | 0.0779 | 0.0491 | 0.888|

It can be seen from Table 5-5 that compared with the original LSTM activation function, the Sigmoid weighted enhanced LSTM prediction method proposed in this paper, RMSE, MAE, and R2 are respectively 0.0577, 0.0398, 0.955, so the structure of the LSTM neural network can be optimized in CS through the Sigmoid weighted enhanced LSTM. Based on the LSTM neural network, the prediction accuracy is further improved.

The four optimal hyperparameters determined by CS are shown in Figure 4 and the prediction result of CS-SiLU-LSTM is shown in Figure 4.
6. Conclusion
The network security situation sequence has the characteristics of time series and nonlinearity. Due to its good ability to deal with long-term dependent problems, the LSTM neural network has strong performance in processing time series. Aiming at the problem of LSTM neural network reducing the vanishing gradient, this paper proposes a network security situation prediction based on the strengthened LSTM neural network model, combining with the prediction model to mine the correlation law between network security situation elements and situation values, and introducing the cuckoo algorithm to automatically determine the network The hyper-parameters can prevent subjective factors from affecting network performance and accurately provide network security managers with better decisions. Experimental results show that the model in this paper improves the prediction accuracy.

Resources
[1] Bass, Tim. Intrusion detection systems and multisensor data fusion[J]. Communications of the Acm, 2000, 43(4):99-105.
[2] Li Ying, Luo Shoushan. Research on network security situation prediction model based on LSTM [EB/OL]. Beijing: China Science and Technology Paper Online [2019-11-07]. http://www.paper.edu.cn/releasepaper/content/201911-17.
[3] Li Xiaolin, Wu Teng. Efficient network traffic prediction method based on PF-LSTM network [J]. Application Research of Computers, 2019, 36(12): 3833-3836.
[4] Zhu Jiang, Chen Sen. Network security situation prediction method based on NAWL-ILSTM[J]. Computer Science, 2019, 046(010):161-166.
[5] Song Gang, Zhang Yunfeng, Bao Fangxun, Qin Chao. Stock prediction model based on particle swarm optimization LSTM[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(12): 2533-2542.
[6] Yao Jialun, Huang Gaoming, Tian Wei. Radar beam scanning behavior prediction based on Drosophila algorithm optimized LSTM[J]. Firepower and Command Control, 2020, 45(03): 122-127.
[7] Supreetha B S , Shenoy N , Nayak P , et al. Lion Algorithm-Optimized Long Short-Term Memory Network for Groundwater Level Forecasting in Udupi District, India[J]. Applied Computational Intelligence and Soft Computing, 2020, 2020(2):1-8.
[8] Cui Minghui. Research on key technologies of network security situation assessment and prediction [D]. Xidian University, 2019.
[9] WANG B W, KONG W, GUAN H, et al. Air quality forecasting based on gated recurrent long short term memory model in internet of things [J]. IEEE Access, 2019, 7: 69524-69534.