Network Security Situation Prediction Based on Long Short-Term Memory Network

1st Li Shang
State Grid Hebei Electric Power Co.,Ltd., information and telecommunication company, China
Shijiazhuang, China
1786340489@qq.com

2nd Wei Zhao
State Grid Hebei Electric Power Co.,Ltd., information and telecommunication company, China
Shijiazhuang, China
734234467@qq.com

3rd Jiaju Zhang
State Grid Hebei Electric Power Co.,Ltd., information and telecommunication company, China
Shijiazhuang, China
1422876461@qq.com

4th Qiang Fu
State Grid Hebei Electric Power Co.,Ltd., information and telecommunication company, China
Shijiazhuang, China
1850524995 @qq.com

5th Qian Zhao
State Key Laboratory of Networking and Switching Technology Beijing University of Posts and Telecommunications
Beijing, China
934501954@qq.com

6th Yang Yang*
State Key Laboratory of Networking and Switching Technology Beijing University of Posts and Telecommunications
Beijing, China
yyang@bupt.edu.cn

***Abstract*—Due to the rapid development of the network, the network security situation is increasingly severe. The network security situation forecast analyzes the past network data and predicts the network situation to the warning of possible network threats in the future. Network security situation prediction can play an important role in network defense, network security warning and network resource allocation. We chose to predict network data first and then evaluate the network situation. We proposed a network security situation prediction method based on LSTM-XGBoost model. We built an improved LSTM neural network model to predict network security data and then used the XGBoost model to conduct situation assessment on the predicted data. The results of comparative experiments show that the model proposed in this paper can complete the task of network security situation prediction more efficiently and accurately.**

Keywords—Situation assessment, Situation prediction, LSTM Neural network, XGBoost

I. INTRODUCTION

The Network has been used in all kinds of life and industrial fields, and the bad influence of the network security problem is getting significant. In order to cope with the severe network security situation, network security situational awareness technology has been put forward and received extensive attention. Network security situation prediction is based on network security situational awareness further forecast and evaluate the future possible network status, over a period of time to be able to predict in advance to some extent network security attacks, can help the network administrator has more time and preparation to cope with the possible arrival of threats, the reasonable allocation of network resources, adopt preventive measures against the network.

In recent years, researchers have used various machine learning and neural network models to predict network security situation. However, there are some problems in these methods, such as the loss of network data information caused by situation assessment and the low prediction accuracy of the neural network model used for the situation prediction. In order to improve the accuracy of network complete situation prediction, this paper adopts the order of network data prediction before situation assessment to build a network security situation prediction model based on Long Short-Term Memory Network (LSTM) and XGBoost. The main work of this paper is as follows:

- We designed a LSTM-XGBoost model based on the order of network data prediction first and then security situation assessment. The order of first predicting data and then evaluating the network situation can largely retain network security data information and improve prediction accuracy.
- We simplify the LSTM gate structure based on LSTM. We built the Bidirectional LSTM network based on the simplified LSTM unit in this paper, so as to enhance the accuracy.

II. RELATED WORK

Bass et al. first put forward the concept of the network security situation and caused widespread concern [1]. Olabelurin et al. proposed a prediction framework based on entropy clustering for real-time detection of DDoS attack phase and active defense against attack events [2]. Xingzhu et al. proposed an improved IPSO-RB network intrusion detection model based on the relationship between RBF neural network feature subset and parameters [3]. Such situation assessment has the problems of the complex model and low universality. In the aspect of situation assessment, network situation assessment can be regarded as a classification problem based on a large amount of marked data. The XGBoost algorithm proposed by Chen et al. is a very excellent classification method in recent years [4]. This paper introduces this method to situation assessment.

As for the prediction model of the neural network, the performance of the recurrent neural network(RNN) is outstanding. In particular, Graves et al. proposed the use of the gate structure to solve the gradient disappearance problem and built the LSTM [5]. Cho K et al. further optimized the gate structure on the basis of LSTM and built Gated Recurrent Unit neural network (GRU) to make the network more simple and efficient [6]. But the structure of the network remains complex. In order to better express complex data, Schuster et al. have proposed Bidirectional Recurrent Neural Networks (BiRNN). BiRNN has been very hot in recent years and is mainly used in natural language processing [7]. Mnih et al. introduced the attention mechanism into the RNN and took a very good effect in image processing [8]. This paper introduces Bidirectional

RNN and attention mechanism into multi-feature data prediction.

III. The Improved GRU-XGBoost Model for Network Security Situation Prediction

A. Network intrusion prediction model

The network security situation prediction model in this paper is mainly composed of two parts. The first part predicts the data through the improved Bidirectional LSTM neural network. The second part is based on the XGBoost algorithm training situation assessment model. Finally, put the data predicted by the LSTM model into the XGBoost model and get the future network situation. The network security situation prediction method in this paper is as follows:

1) Data collection: Collect network data.

2) Data processing: Process the data, including adding network situation signs and converting character-type features into numerical features.

3) Train L model and Train XGBoost model : Divide the processed data into training set and test set, and use training set data to train LSTM model and XGBoost model.

4) Predict network data: Use the trained LSTM model to predict the network data in the future.

5) Network intrusion prediction: Use the predicted data to conduct situation assessment with the trained XGBoost model.

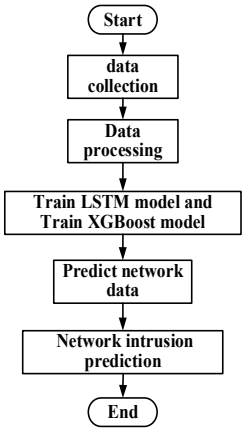

Fig. 1.Network intrusion prediction model flow chart.

B. Improved LSTM network

The main part of the network situation prediction model in this paper is the network data prediction model. Because the network data is difficult to express and the amount of data is large, this paper chooses to improve on the basis of the LSTM network. We first simplified the gate structure of the LSTM, and then built the Bidirectional LSTM model network based on the new LSTM cell.

LSTM network is a kind of RNN. The forget gate f_t mainly determines which information is to be forgotten with the output value h_{t-1} at the t-1 time step and the input value x_t at the current time. The input gate i_t mainly determines the value to be updated, updates the memory cell unit state $\tilde{C}_t$. The memory cell unit state C_t and the output gate o_t determines which part of the information can be output, and finally get the output value h_t.

In this paper, the output gate is deleted on the basis of LSTM, which makes the network structure simpler and requires fewer training parameters, such that the bidirectional network LSTM subsequent build more simple and efficient. For convenience, we refer to this structure as SLSTM. Fig.2 shows the basic structure of the LSTM neural network unit without output gate.

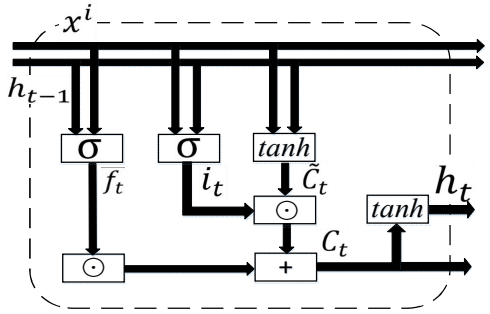

Fig. 2.The structure of SLSTM neuron.

The algorithm formula of SLSTM neural network unit is as follows:

$$f_t = \sigma(W_{hf} * h_{t-1} + W_{xf} * x_t + b_f) \quad (1)$$

$$i_t = \sigma(W_{hi} * h_{t-1} + W_{xi} * x_t + b_i) \quad (2)$$

$$\tilde{C}_t = tanh(W_{hc} * h_{t-1} + W_{xc} * x_t + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

$$h_t = \tanh(C_t) \quad (5)$$

From the SLSTM unit structure on the graph and the formulas, the simplified gate structure recurrent neural network still chooses the gate structure for selective memory and forgetting. It retains the functions and features of LSTM door structure, while trying to reduce the number of door structure to achieve the reduction of parameters in the recurrent neural network training and learning process, hoping to effect in the construction of bidirectional neural network model.

C. Bidirectional SLSTM model

In this paper, the Bidirectional LSTM model is built on the basis of the SLSTM and the Bidirectional LSTM. The Bidirectional LSTM network is mainly used to express contextual semantics. Since this paper directly predicts multi-feature data, each group of data has a connection, so we try to apply this idea to network data prediction. Fig. 3 shows the structure of Bidirectional LSTM network.

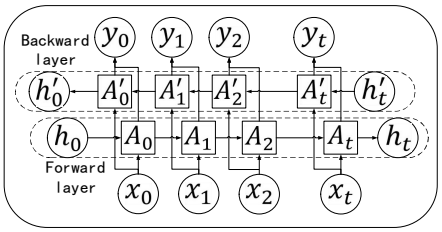

Fig. 3.The structure of Bidirectional LSTM network

In the Forward layer, the forward calculation is performed from time 1 to time t, and the output of the forward hidden layer at each time is obtained and saved. In the Backward

layer, the calculation is reversed along time t to time 1, and the output of the backward hidden layer at each time is obtained and saved. Finally, at each moment, the final output is obtained by combining the results of the corresponding time of the Forward layer and the Backward layer. The mathematical expression is as follows:

$$h_t = f(w_1 x_t + w_2 h_{t-1}) \quad (6)$$

$$h'_t = f(w_3 x_t + w_4 h'_{t+1}) \quad (7)$$

$$o_t = g(w_5 h_t + w_6 h'_t) \quad (8)$$

D. Situation assessment based on XGBoost

XGBoost algorithm is a kind of improved GBDT algorithm, it based on decision tree (CART) warrant a search for the base of learning Gradient boosting algorithm. XGBoost expands and improves GDBT, and the XGBoost algorithm is faster and relatively.

The objective function of the XGBoost algorithm is as follows:

$$\text{obJ}(f_t) = \sum_{i=1}^{n} L\left(y, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + \text{con} \quad (9)$$

In Where $L\left(y, \hat{y}_i^{(t-1)}\right)$ is the training error of the model, $\Omega(f_t)$ is the regularization term, and con represents the structure of the former t-1 tree, which is a constant.

Then let $g_i$and h_i respectively represent the first derivative and the second derivative of the prediction error for the current model, and the current model iterates toward the direction in which the prediction error decreases.

$$g_i = \frac{\partial L(y, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}} \quad (10)$$

$$h_i = \frac{\partial^2 L(y, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}} \quad (11)$$

The decision tree complexity calculation formula is as follows:

$$\Omega(f_t) = \gamma \cdot T_t + \lambda \frac{1}{2} \sum_{j=1}^{T} w_j^2 \quad (12)$$

Where T_t is the number of leaf nodes of the t-th tree, w is the leaf node vector, γ represents the difficulty of node segmentation, and λ represents the L2 regularization coefficient.

Solve the objective function and find the best w and the corresponding objective function optimal value. The two results correspond to the following:

$$w_j^* = \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (13)$$

$$\text{obJ}^*(f_t) = -\frac{1}{2} \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma \cdot T \quad (14)$$

$\text{obJ}^*(f_t)$ is the scoring function. It is a standard for measuring the structure of a tree. The smaller the value, the better the structure is represented. The scoring function is used to select the best segmentation point to construct the CART tree.

IV. EXPERIMENT

In this experiment, we compared the model BiSLSTM proposed by this paper with the other two prediction models, GRU and LSTM. We conducted experiments on the kddcup99 dataset [9] downloaded from the Internet. We experimented with a TITAN Xp graphics card and built four models with Python+Tensorflow+Keras.We used sklearn to implement the XGBoost algorithm, trained the situation assessment model, and evaluated the predicted data of the four models to obtain the final situation prediction results.

A. Data set processing

We selected 10% kddcup99 dataset for the experiment. Our network status of the data set to the identifier, the network state is divided into five categories: network normal, Dos (denial of service attacks), the Probe (scanning attack) and R2L (unauthorized access from remote host), U2R (unauthorized local super user privileges access), in this paper, the network situation is set to 0, 1, 2, 3, 4.

TABLE I. NETWORK INTRUSION LABEL

Network intrusion	Meaning	Label
Normal	Normal network status	0
Dos	denial-of-service	1
Probe	surveillance and probing	2
R2L	unauthorized access from a remote host	3
U2R	unauthorized access to local superuser privileges by a local unpivileged user	4

We also converted other characteristic data in the data set into numerical data, and finally set the data set size to 100,000, 120,000, and 150,000 to do experiments. Performance comparison.

In the network data prediction part, we directly predict the first 40-dimensional features in the data set. For each data set, we select the first 80% of the data as the training set, and the last 20% as the test set for data prediction.

B. Model comparison

This simulation experiment mainly selected three criteria: epoch loss, Root mean squared error (RMSE), and Accuracy.

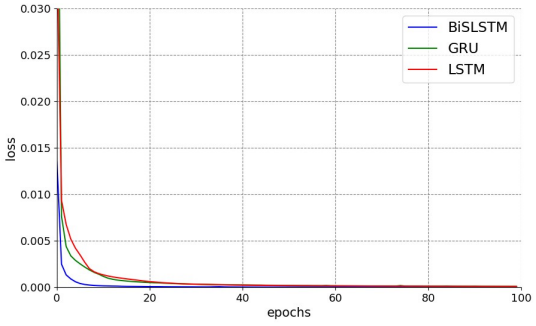

Fig. 4. The epoch loss compare of 150,000 dataset.

Fig. 4 show the loss of each epoch training on the dataset of 150,000 size. All three models converge quickly and smoothly. On the 150,000 dataset, the BiSLSTM reached an inflection point about 4 epochs faster than the GRU, reaching

an inflection point 6 epochs faster than the LSTM. The loss value of BiSLSTM is smaller than the other two models.

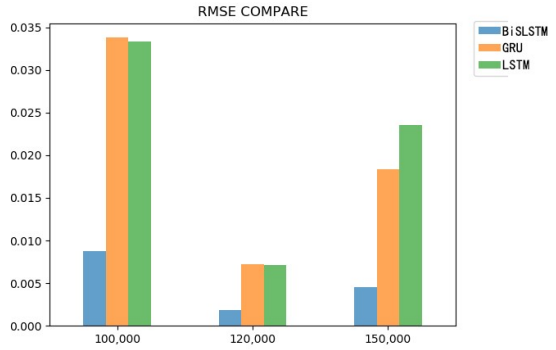

Fig. 5.The RMSE compare

Fig.5 shows respectively the RMSE of three models on different size datasets. The three model all have a low RMSE. The BiSLSTM model of this paper has a smaller RMSE than the other two models on different datasets. This shows that the performance of this model is better than the other two models in this prediction of multi-feature data tasks. However, we can also see that the RMSE fluctuations are still relatively large, because the data distribution is uneven, resulting in some data errors may be concentrated. This may because we directly predict data which has 40 features. For the range of values for each feature is different, the magnitude of the difference between RMSE on different features will be different.

In the part of evaluating network situation with XGBoost, our dataset still uses 80% as the training set, and the remaining 20% is used to verify the network data predicted by the neural network model. In the 80% training set, we divided 20% as a validation set to facilitate XGBoost to verify its classification effect.

TABLE II. XGBOOST CLASSIFICATION ACCURACY

Dataset	**100,000**	**120,000**	**150,000**
Accuracy	0.990187	0.991927	0.999917

It shows the accuracy of network situation assessment on the validation set based on XGBoost. It can be seen that XGBoost is very suitable for this task for it has a high accuracy.

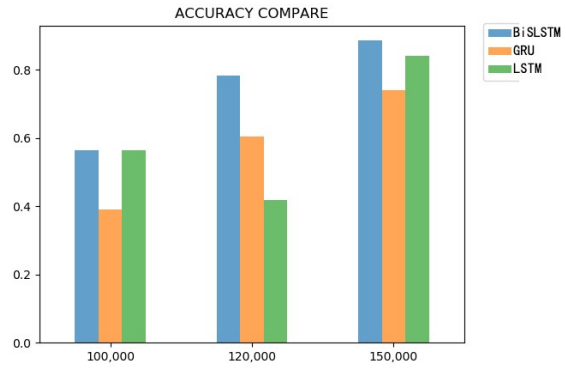

Fig. 6.The Accuracy compare

Fig.6 shows the accuracy of the data obtained by the three prediction models after the XGBoost model network situation assessment. On the 100,000 dataset BiSLSTM and LSTM achieved almost the same accuracy while the BiSLSTM is 0.17 higher than GRU. On the 120,000 dataset, BiSLSTM is 0.17 higher than GRU, 0.36 higher than LSTM. On the 150,000 dataset, BiSLSTM is 0.14 higher than GRU and 0.04 higher than LSTM. Overall, the larger the data set, the higher the accuracy. BiSLSTM achieves high accuracy on all three data sets, followed by LSTM and finally GRU. The reason for the large fluctuation may be that the data distribution in the data set is not uniform, resulting in data over-fitting, and the deviation of the prediction data may be concentrated on certain features, resulting in the final misjudgment. This is also because we have directly predicted all the data in the previous steps, and did not do data feature analysis and feature screening.

V. CONCLUSION

We used the method of predicting network data and then evaluating the situation to predict the network situation, built an improved Bidirectional LSTM model. The situation assessment introduced the XGBoost algorithm.

Our improved LSTM model simplifies the gate structure, then establishes a Bidirectional LSTM model to achieve lower RMSE than the other two models in the task of multi-feature data prediction. Our situation assessment part uses the XGBoost algorithm to make a certain degree of evaluation. We use the training set to train XGBoost for situation assessment and then we input the data predicted by the three models into XGBoost for situation assessment. Our network situation prediction model also has a higher accuracy rate than other models.

ACKNOWLEDGMENT

The work presented in this paper was supported by State Grid Hebei Electric Power Co.,Ltd., information and telecommunication company “Data network security assessment strategy and status analysis research Information communication and security technology project ” (SGHEXT00DDJS1800196)

REFERENCES

[1] Bass T. Gruber D. A glimpse into the future of id [J]. The Magazine of USENIX & SAGE, 1999, 24(3) : 40-49.

[2] Olabelurin A, Veluru S, Healing A, et al. Entropy clustering approach for improving forecasting in DDos attacks[C]//2015 IEEE 12th International Conference on Networking. Sensing and Control (ICNSC). Taipei: IEEE, 2015:315-320.

[3] Xingzhu W . Network Intrusion Prediction Model based on RBF Features Classification[J]. International Journal of Security & Its Applications, 2016, 10(4):241-248.

[4] Chen T , Guestrin C . XGBoost: A Scalable Tree Boosting System[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.

[5] Graves A , Jürgen Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks, 2005, 18(5-6):602-610.

[6] Cho K , Van Merrienboer B , Gulcehre C , et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. Computer Science, 2014.

[7] Schuster M, K.K. Paliwal. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing, 2002, 45(11):2673-2681.

[8] Mnih V , Heess N , Graves A , et al. Recurrent Models of Visual Attention[J]. Advances in neural information processing systems, 2014.

[9] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html