ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0007

A Cross-Layer Feature-Fusion TCN-BiGRU based Network Intrusion Detection System

Kun Wei*, Jiquan Shen

College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 450000, China

Abstract

Aiming at the problem that the multi classification accuracy of current intrusion detection algorithms for data class imbalance and incomplete feature learning of network intrusion detection, in view of network intrusion data has the characteristics of time series. A method combining TCN and BiGRU (bi-directional gated cyclic unit) is proposed. Network intrusion detection method of gated recurrent unit (BiGRU). In order to solve the problem that the original attack data is widely distributed and dis-crete. For the strong problem, firstly, the data are individually encoded and normalized, and the algorithm completes the balance processing of the data set, The model is composed of TCN and Bi-GRU components connected in the form of a cross-layer; the TCN component recognizes regional features to obtain the full global features, while the BiGRU obtain periodic features by memory function. Finally, the two types of features are aggregated to get comprehensive features, which can more accurately identify intrusion information. Experiments based on the is based on CSE-CIC-IDS2018. The results show that the accuracy of the proposed method for the above data sets can reach 99.62%, and the balanced data sets is better than the un-processed data set. The recognition accuracy of a few classes is significantly im-proved, and the model is superior to other existing methods.

Keywords

IDS; Bi-GRU; TCN; Network Intrusion Detection; Machine Learning.

1. Introduction

As an active security protection technology, intrusion detection system (IDS) can effectively perceive the attack behavior in the network through real-time monitoring [1], so as to facilitate the security managers to make corresponding decisions in time and ensure the stable operation of the network. However, with the development and re-search progress, deep learning technology has been widely used in the field of intrusion detection and achieved good detection results.

The network intrusion detection method based on traditional machine learning can't deal with a large number of nonlinear high-dimensional data [2], and it is difficult to adapt to the increasingly complex and diverse attack environment of network data. Deep neural network model directly learns features from a large amount of data without relying on Feature Engineering. It effectively solves the problems of low precision, unable to deal with complex data and poor classification effect in traditional machine learning. Therefore, anomaly detection based on deep learning has become a research hot spot. The proposed model is composed of TCN and Bi-GRU components by cross-layer feature-fusion. The TCN component can get global features, while the Bi-GRU com-ponent can get period features. Finally, the two types of features are fused by a feature fusion component, multi-scale and multi-domain abnormal information can be detected. This model combines the advantages of both the TCN and Bi-GRU, so it shows the good performance in than else intrusion detection methods.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0007

The main contributions of this research are given as follows:

1) Propose a cross-layer feature-fusion TCN-BiGRU intrusion detection model with a higher accuracy compared with other models.

2) Introduce TCN into the model to extract the spatial features of the data and allowing the model to learn deeper features, Bi-GRU extract the temporal features of the data, the fusion features of temporal and spatial are used to represent the multi domain features of data, which avoids the limitations of the model and reflects the advantages.

3) Use MDI (mean decrease impurity) feature selection algorithm to clean the data set, and the feature optimization and dimensionality reduction, and SMOTE-Tomeka algorithm is used to balance the data set at the data level, avoiding the problems of over fitting samples and fuzzy boundaries generated by general oversampling methods.

4) The proposed model was evaluated on the CSE-CIC-IDS2018 dataset, which contains the latest network attacks and meets all the criteria for real-world attacks.

2. Related Work

IDS has been widely used since it was first proposed., however, with the development of AI (artificial intelligence technology), mainly include from machine learning and deep learning have been applied. General deep learning methods include MLP (multi-layer perceptron), CNN (convolutional neural networks), and RNN include of LSTM (long short-term memory) and GRU (gated recurrent unit). Liu [3] put forward a detection model based on MLP, this model has significantly improved the accuracy of comparison with support vector machine, but it is usually difficult to get good training results in the environment with a large number of features and high dimensions. Li [4] analyzed the IDS with CNN model, achieving a higher accuracy rate than other single models. Li [5] aiming at the problem that a few categories in the data set are difficult to detect, according to the timing characteristics of network intrusion behavior, a network intrusion detection method based on Bi-LSTM is proposed. The detection effect of this method for web attacks is better than other methods, but the processing effect for other attack types is general. Sun [6] combines LSTM and RNN and proposes an intrusion detection method based on LSTM-CNN hybrid model. The model is studied based on KDD99 data set, and the detection effect is good, but the data set is not persuasive. According to the time sequence characteristics of network intrusion data, literature [7] proposed an intrusion detection method based on improved time sequence analysis method, which can effectively make up for the lack of active defense ability of intrusion detection system, but the prediction accuracy needs to be further improved. The accuracy of second classification in the above literature is generally high, but in the face of multi classification problems, the effect is not ideal.

Ge [8] combined CNN and LSTM to analyze network intrusion is based on the global features by CNN; Then the periodic features by LSTM, which is obtained a high accuracy and a low false positive rate, but the model of the LSTM network is processed by the CNN which will lead to the loss of some periodic features [9.10]. To solve these problems, this model proposes an intrusion detection model based on the cross-layer feature fusion of TCN and Bi-GRU components. However, how to build a model without triggering the performance degradation issue and address the class imbalance problem in the training set are two major challenges.

3. System Components

In the normal network environment, the network characteristic information is very obvious periodicity [11], and regular [12] and has while abnormal network information does not have these characteristics. So, this paper proposes a cross-layer feature fusion TCN-BiGRU intrusion detection model, the structural system of which is illustrated in Figure 1. The algorithm is mainly composed of data preprocessing, TCN, BiGRU, and feature fusion components. In the data preprocessing model, the input is numerically processed, normalized, data cleaning and sample equalization to meet the requirements of the neural network. Use the latest intrusion detection data set a to train and test the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0007

model, the dataset contains the latest network attacks. Fusion neural network can extract local parallel features and analyze the influence of the in-formation before and after each feature point on the feature point. The TCN com-ponent can extract globe features and detect whether the feature distribution is nor-mal. The BiGRU component was composed of several GRU cells, and it is mainly used to detect the periodicity of network information by its memory function. The characteristic fusion component is composed of MLP (multi-layer perceptions), and it is mainly used to fuse the features extracted from TCN and BiGRU components, and normalize the classification probability to obtain the final result.

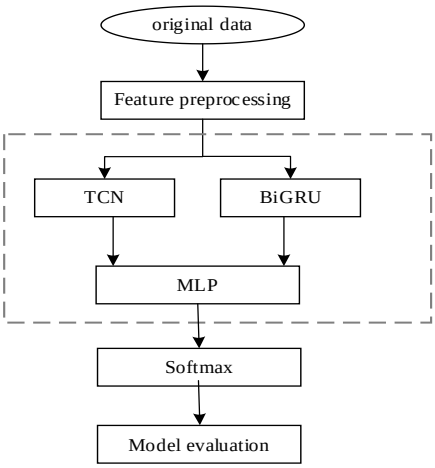

Fig.1 Cross-layer feature-fusion TCN-BiGRU intrusion detection model

3.1 Data Preprocessing Module Component

The data preprocessing part mainly includes the cleaning of the original data set, feature selection, data set balance and other operations. Intrusion detection system should classify according to the behavior characteristics of network information to avoid bias towards the information attached to specific identification. Therefore, it is necessary to delete the features related to specific network identification, and delete the feature columns with the same data value under the same feature by analyzing the data.

In order to solve the sample intrusion problem and fuzzy boundary problem caused by smote algorithm, the combination of Tomek links algorithm and SMOTE algorithm is called smote Tomek algorithm to solve the problem of data imbalance. Firstly, smote algorithm is used to over sample the original data to synthesize a few samples, and then Tomek links algorithm is used to clean the data processed by smote algorithm and remove the Tomek links pairs in the data set, so as to filter out the noise data and overlapping samples between the two categories.

3.2 TCN and BiGRU Component

TCN is a simply network combining one-dimensional full convolution network and causal convolution [20]. For TCN, its convolution layers have the same length, and zero filling ensures that the higher layer has the same length as the previous layer. In order to obtain long-term effective historical information, a very deep network model is needed to extract the features from historical time series. In order to solve this problem, TCN introduces extended convolution, which greatly reduces the depth and complexity of the network while ensuring the acquisition of long enough historical effective information. TCN introduces extended convolution to expand the receptive field exponentially, so as to solve the problem that ordinary causal convolution can’t deal with time series efficiently. When training deep network, residual connection can transfer information across layers, so as to improve the training speed of the network and avoid the problem of gradient disappearance.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0007

The residual connection module of TCN includes two layers of convolution and nonlinear identity mapping.

GRU solves the long-distance transmission information loss, while BiGRU contains two ordinary GRU units, and GRU can capture the characteristic information ignored by BiGRU, so it can improve the accuracy of long-time series data, the expression forms of these steps are, respectively, as shown in the following formula.

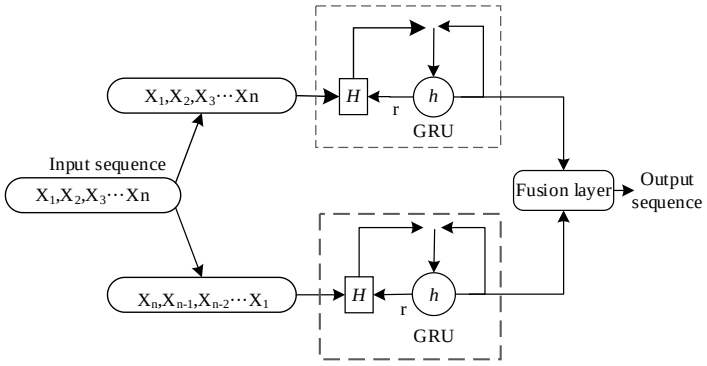

Fig.2 BiGRU architecture

$$\vec{h}_t = GRU(x_t, \vec{h}_{t-1}) \tag{1}$$

$$\overleftarrow{h}_t = GRU(x_t, \overleftarrow{h}_{t-1}) \tag{2}$$

$$h_t = W_t \vec{h}_t + W_v \overleftarrow{h}_t \tag{3}$$

Where W_t and W_v respectively represent the weight matrix corresponding to the forward hidden state and feature fusion processing component.

MLP is usually composed of input layer, output layer and several hidden layers. As shown in Figure 5, the adjacent layers are fully connected, and the corresponding activation function is set to realize the nonlinear transformation. The features extracted by TCN and BIGRU components are combined into comprehensive features with multi domain features after flatten processing. The integrated features enter the MLP module through the input layer and carry out nonlinear mapping in the hidden layer; Finally, the output layer outputs the predicted classification results with SoftMax function, as shown in the following formula (4).

$$s(x) = \frac{1}{1 + e^{x_j - x_i}} \tag{4}$$

After com-paring the model with the actual results, the parameters of each layer are modified by back propagation of loss function, as shown in the following formula (5),

$$H(p,q) = -\sum_1^n p(x_i) log(q(x_i)) \tag{5}$$

Then the training is completed after multiple parameter modifications.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0007

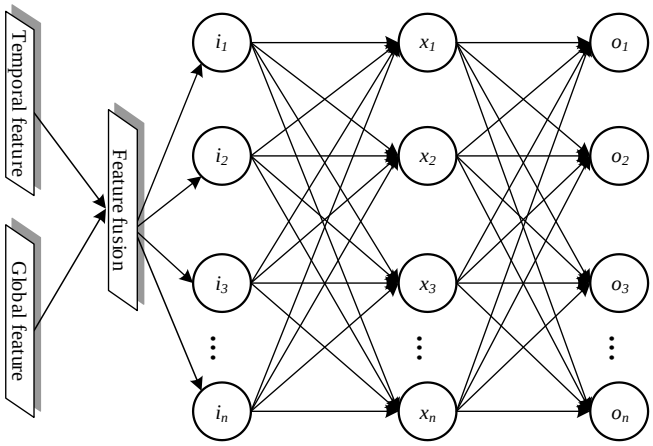

Fig. 3 Feature fusion component architecture

4. Experiments

4.1 Experiment Environment

Experimental environment of this research was shown in Table 1.

Table 1. Experimental environment.

Environment	Value
OS	Windows10
CPU	I5-12400 H
Memory	32G
Language	Python
Tool	Anaconda-Keras

4.2 Data Preprocessing

CSE-CIC-IDS2018 intrusion detection dataset has 83 data feature fields and A1 category labels, a total of 84 labels, which are composed of normal traffic and 14 attack traffic, including 14 attack scenarios. Then delete the feature value with data value of 0, and calculate the feature importance score of the remaining 68 feature fields according to the feature importance score formula. In order to construct an 8 * 8 convolution neural network data matrix, the feature fields are sorted according to the feature importance score, and 64 feature fields are reserved in the final data set. Considering the limitation of computer performance, this experiment randomly selects traffic from normal traffic as the total sample of normal traffic, and the attack traffic samples remain unchanged (the data of other processed data sets are divided into training set, verification set and test set according to the ratio of 8:1:1). A few samples are synthesized and cleaned by smote Tomek algorithm to complete the data balance. Finally, in order to fully evaluate the performance of the model, accuracy, precision, F1 and FRP were used as the evaluation indexes of the model.

$$\text{accuracy} = \frac{\text{TP+ TN}}{\text{TP+ TN+ FP+ FN}} \tag{6}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP+ FP}} \tag{7}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0007

$$FPR = \frac{FP}{FP+TN} \tag{8}$$

$$DR = \frac{TP}{TP+FN} \tag{9}$$

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{10}$$

TP (True Positive) denotes the number of positive samples correctly classified as positive; TN (True Negative) denotes the number of negative samples correctly classified as negative; FP (False Positive) denotes the number of negative samples misclassified as positive; and FN (False Negative) denotes the number of positive samples misclassified as negative.

4.3 Hyper-Parameter Setting

The super parameters will directly affect the effect of the model. In this research, Nadam optimization algorithm would be used to replace default Adam to strengthen the constraint on learning rate [21], the optimal super parameter setting of the experiment is obtained: the number of nodes in BIGRU hidden layer is 64; the size of con-volution kernel is 3 × 3; The discard rate is 0.5; The batch size is 256; The number of iterations is 20; The learning rate is 0.005, loss function is categorical cross entropy.

4.4 Experimental Performance Evaluation

In order to verify the effectiveness of smote Tomek algorithm on data balance, the data sets processed by smote Tomek algorithm and those not processed by smote Tomek algorithm are compared and tested on the model in this paper, and the accuracy of each category before and after data set balance is obtained.

The data set processed by smote Tomek algorithm has significantly improved the recognition accuracy of a few classes compared with the data set not processed. After being processed by smote Tomek algorithm, the recognition accuracy of DoS attacks slow http test class increases from 0 increased to 39.66%, and the recognition accuracy of SQL injection class increased from 0 to 100%, UDP force and xsicforce increased by 58.5% and 34.5% respectively. The accuracy of FTP brute force class decreased by 18.62%. Although the brute force matrix can accurately distinguish the two classes, the HTTP brute force and FTP systems can be confused, and the brute force matrix can be used to deal with them all. The experimental results show the effectiveness of smote Tomek dataset balancing algorithm in improving the effect of model detection.

The performance of the classical single deep learning model is compared with the model in this paper. The comparative experiment is multi classification. The classical single depth learning model is compared with the model in this paper, and the comparative experiment is multi classification in the experiment, TCN and BIGRU single model are used to compare with the model in this paper. All models use the training set processed by smote Tomek algorithm to train the model, and compare the performance of these models on the classification of various types of traffic respective-ly, taking the accuracy rate, recall rate and F1 value as the evaluation indexes. The multi classification results are shown in Table 2. In the experiment, TCN and BIGRU single models are used to compare with the models in this paper. All models use the training set processed by smote Tomek algorithm to train the models, and compare the performance of these models on the classification of various types of traffic respectively. The accuracy, recall and F1 value are used as evaluation indicators The multi classification results are shown in Table 2.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0007

Table 2. Comparison with other detection methods on CSE-CIC-IDS2018 Dataset

Category	Testing metrics			
	Accuracy	Precision	Recall	F1
Decision tree	96.6	97.62	96;67	97.13
Random forest	98.21	98.58	93.40	95/92
TCN	98.52	94.81	95.77	94.86
BiGRU	97.47	94.86	95.89	95.89
TCN-BiGRU	99.62	98.02	97.39	97.69

It can be seen from table 2 that the model proposed in this paper can achieve better performance in almost all indicators, including accuracy, accuracy, recall and com-prehensive evaluation indicators. The reason is that compared with machine learning methods such as random forest and decision tree, this model uses deep neural net-work to learn the data set. Because the neural network has strong nonlinear fitting ability and can map any complex nonlinear relationship, this model has stronger feature extraction ability and higher recognition accuracy; Compared with a single model, this model adopts the method of TCN and BiGRU fusion for feature extraction, and the extracted feature information is more comprehensive. TCN overcomes the limitations of general convolution operation, and the feature extraction of current time data will not be affected by future data, so it has higher reliability and better multi classification effect.

5. Conclusion

In order to solve the problems of poor feature extraction effect and low accuracy of multi classification of general intrusion detection algorithms, a network intrusion detection method integrating TCN and BiGRU is proposed in this paper. This method combines TCN and BiGRU, and improves the data set processing. Finally, through the performance analysis experiment on the data set and whether the data set is processed or not, it is proved that the model has strong feature extraction ability, high detection accuracy and low false alarm rate when dealing with large-scale high-dimensional network data, which provides a promising forward-looking real-time application for intrusion detection system.

References

[1] LIAO H J, LIN C H R, LIN Y C, et al. Intrusion detection system: a comprehensive review[J]. Journal of network and computer applications, 2013, 36(1): 16-24.

[2] Zhai Jiqiang, Ma Wenting, Xiao Yajun. Intrusion Detection system based on Apriori-KNN algorithm alarm Filtering Mechanism [J]. Journal of microcomputers,2018,39(12):2632-2635.

[3] Liu Hui Research on intrusion detection method based on principal component analysis and multilayer perceptron neural network [J] Software engineering, 2020,23 (07): 10-12 + 9.

[4] Li Jing, Huang Jie, Zhu Guowei, Yuan Hui, Li Weijian, long Jiachuan Network intrusion detection method based on adaptive one-dimensional CNN [J / OL] Journal of Wuhan University (Engineering Edition): 1-9Ahmad I , Basheri M , Iqbal M J , et al. Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection[J]. IEEE Access, 2018, 6:33789-33795.

[5] Li Jun, XIA Song-zhu, LAN Hai-yan, Li Shou-zheng, Sun Jianguo. Network Intrusion Detection Method Based on GRU-RNN [J]. Journal of Harbin Engineering University, 201,42(06):879-884.

[6] SUN et al. Extracting features using CNN-LSTM hybrid network for intrusion detection system[J]. Security and Communication Networks, 2020.

[7] Jian Shijie, Lu Zhigang, Du Dan, Jiang Bo, Liu Baoxu Overview of network intrusion detection technology [J] Journal of information security, 2020,5 (04): 96-122.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0007

[8] Ge Jike, Liu haoyin, Li Qingxia, Chen zuqin Research on network intrusion detection model based on improved cnn-lstm [J] Software engineering, 2022,25 (01): 56-58 + 55.

[9] Zha Wenting, Liu Jie,Li Yalong, Liang Yingyu. Ultra-short-term power forecast method for the wind farm based on feature selection and temporal convolution network. [J]. ISA transactions,2022.

[10] Li Jun, XIA Song-zhu, LAN Hai-yan, Li Shou-zheng, Sun Jianguo. Network Intrusion Detection Method Based on GRU-RNN [J]. Journal of Harbin Engineering University, 201,42(06):879-884.

[11] Zheng, Z.; Yatao, Y.; Niu, X. Wide & Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids. IEEE Trans. Ind. Inform. 2017, 14, 1606–1615.

[12] Zhang, K.; Hu, Z.; Zhan, Y.; Wang, X.; Guo, K. A Smart Grid AMI Intrusion Detection Strategy Based on Extreme Learning Machine. Energies 2020, 13, 4907.