2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)

Network Log Anomaly Detection Based on GRU and SVDD

Shirong Liu*[§], Xiong Chen*, Xingxiong Peng[†],Ruliang Xiao*[†§]
*College of Mathematics and Informatics, Fujian Normal University, Fuzhou , China
[†]Fujian Provincial Key Laboratory of Network Security and Cryptology, Fuzhou , China
[§]Digit Fujian Internet-of-Things Laboratory of Environmental Monitoring
Fuzhou , China
Email: xiaoruliang@fjnu.edu.cn

***Abstract*—Using machine learning to detect anomalies in network logs has become a research hotspot in the field of industrial Internet of Things security. In the era of large data, it is inefficient using traditional methods to detect anomalies under the environment of high-dimensional data and extensive data association. How to detect anomalies efficiently and accurately is a challenge. This paper presents a novel method of anomaly detection for network logs based on GRU and SVDD. First, PCA is used to reduce the dimension of high-dimensional datasets and extract effective attributes. Then, the processed datasets are used to train the GRU-SVDD classifier model. Finally, the actual logs to be detected are input into the GRU-SVDD comparator to detect the day. Experiments on classical KDD Cup99 datasets show that our method is superior to classical GRU-MLP and LSTM algorithms.**

Keywords—anomaly detection, GRU, SVDD

I. Introduction

With the rapid development of big data, Internet of Things and other technologies, the traditional relatively closed industrial control system has become more complex and open, while bringing many conveniences to the people, the network attacks against the industrial Internet of Things platform have become more common, and the problem of information security has become more serious. From the point of view of previous literatures, it has shown great potential to effectively guarantee the reliability and security of the system and to analyze and utilize large-scale historical log data. How to effectively utilize these large-scale network log data sets to efficiently construct anomaly detection model and apply it to anomaly detection analysis of industrial Internet of Things. It is of great significance.

Research on anomaly detection has become one of the most popular research topics in the field of Internet and Internet of Things security[1], and has a wide range of applications in various fields, such as medical drug research [2], credit card fraud detection [3], anomaly detection from Web logs [4], anomaly detection in financial loans [5], and the Federation of Industry and Commerce. Network anomaly detection [6]. The most common method to detect intrusions is to analyze user activities [7,8]. Academia and industry often use machine learning to discover hidden rules in large-scale historical network log data of server systems.

Some progress has been made in the application of existing methods to large-scale network log datasets, but there are still two problems as follows:

- The traditional anomaly detection algorithm does not reduce the dimensionality of these high-dimensional feature attribute data very well. The better method is to extract some features randomly for detection. The detection efficiency of the algorithm is low, and it needs a lot of manpower and material resources.
- Most of the existing algorithms for anomaly detection of large-scale network log data using machine learning methods directly classify the existing normal and multiple types of attacks, while the detection ability of unknown anomaly types is weak and the detection accuracy is low.

In order to solve these two problems, this paper proposes an improved GRU-based anomaly detection model, which can be efficiently applied to the network log data on the industrial Internet of Things. The main contributions of this paper are as follows:

- In view of the highly correlated network log data sets between high-dimensional and feature attributes, the main highlight of this paper is that the principal component analysis (PCA) method is used to pre-process the original high-dimensional data in advance, eliminating redundant attributes.
- In the construction of classification algorithm framework, the main highlight of this paper is the combination of optimized GRU and SVDD algorithm, GRU(Gated Recurrent Unit) algorithm can extract feature attributes very well, and SVDD(Support Vector Domain Description) algorithm is an efficient single classification algorithm, as an output layer classifier, it can well model large-scale normal network logs in this paper environment. Unknown network attacks can be detected.

We have carried out a lot of comparative experiments, using KDD CUP99, and sampled data sets for many times to generate the experimental data set in this paper, using two kinds of comparative forms: (1) Comparing the effect of using PCA to reduce the dimension of data and choosing different new principal components on the detection efficiency of the algorithm; (2) Compared with several classical algorithms such as GRU-MLP and LSTM, Experiments show that the algorithm proposed in this paper is more efficient when applied to large-scale network log data sets, and has good scalability for data size.

The rest of this paper is organized as follows: the second part introduces the related work of anomaly detection, the third part details the motivation and background of the proposed method, the fourth part gives the proposed method, the fifth part analyses the experiment and experimental results, and the sixth part summarizes the whole paper.

978-1-7281-4328-6/19/$31.00 ©2019 IEEE
DOI 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00177

II. Related Work

Using machine learning methods to detect anomalies in industrial Internet of Things is a hot research field of anomaly detection [9,10,11], It can efficiently analyze and detect anomalies in data. At present, a large number of anomaly detection methods based on machine learning have been proposed in academia. After summarizing, they can be divided into three categories: anomaly detection based on association rule learning [12,13], anomaly detection based on clustering [14,15] and anomaly detection based on classification [16,17].

1) An anomaly detection method based on association rule learning. This method extracts hidden rules from data samples, and then extracts the access behavior patterns of most users. Using the data to be measured to match the sequence rules, if it can not match more completely, it indicates that there is an anomaly. Xiaochu Yun and Yongzheng Zhang et al. of Chinese Academy of Sciences In document [13] proposed an anomaly technology model based on session anomaly degree, which can detect and locate anomalies in data sets efficiently by mining behavior model using frequent sequential pattern mining algorithm. However, the algorithm uses a prefix number of frequent pattern mining, and updates and maintainers the algorithm later. Because of its large workload, it can not be well applied to systems with high real-time requirements.

2) Anomaly detection method based on clustering analysis. Clustering can be defined as dividing data into a group of similar objects. It is an unsupervised learning method, in which each cluster is composed of objects similar to each other and different from those in other clusters [15]. Clustering for anomaly detection research can directly use historical data sets without consuming huge amounts of physical resources to label the categories. But when using clustering for anomaly detection of large-scale data sets, the time efficiency of the algorithm is high and the efficiency will be affected by the number of clusters.

3) Anomaly detection method based on classification analysis. Classification can be defined as the problem of identifying sample categories in test data sets by training classifiers in training datasets of known class attribute members. The class attribute members of sample are also called class tags. In the field of machine learning, the most widely used method of unsupervised learning is classification. In document [7], Sung et al. proposed a method of anomaly detection of network log by using support vector machine (SVM) and artificial neural network (ANN). In data mining, SVM can use hyperplanes to separate data points in different classes. Neural network is a computational model that simulates the human brain. It represents the transmission of information from one neuron to another. Alalshekmubarak et al. proposed a method to classify time series by combining neural network with SVM in document [16]. The algorithm in this paper combines echo state network (ESN), a variant of recurrent neural network (RNN) with SVM. In document [18], Nga Nguyen Thi et al. proposed a new method to classify time series better by using the structure of long single memory neural network (LSTM), and achieved good results. However, the training time of these methods is long and the application scenarios are limited.

In view of the above problems, this paper proposes an anomaly detection method combining GRU and SVDD, which can effectively overcome the above shortcomings. A neural network anomaly detection model combining PCA method and GRU-SVDD is added between the remote client and server of enterprise management. The anomaly in the system is detected by using the network log data characteristics in the communication network between the client and the system. The system architecture diagram is shown.

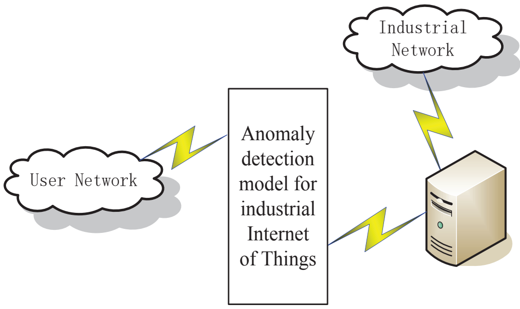

Fig. 1. System architecture model

The combination of GRU and SVDD anomaly detection algorithm is an efficient single classification model based on normal network log data sets, in which principal component analysis is used to pre-process the original data sets and reduce the correlation between attributes in the original data sets. Then an efficient neural network model GRU is used to extract the feature attributes after pretreatment. The output layer is replaced by a single classification model named as SVDD. Through the final experimental comparison, it can be seen that the proposed algorithm can be well applied to anomaly detection on large data platform.

III. Motivation And Background

A. Motivation and Background of Using Neural Network GRU

The most common way to detect anomalies in network log data is to analyze the user's behavior patterns. This method requires a lot of computing and material resources. Therefore, it is imperative to use machine learning method. Neural network has high efficiency in analyzing network log and great research and development value.

Traditional neural network is used for network log analysis, and full connection is used between adjacent layers of the neural network model, but each node in each layer is not connected. The algorithm is independent in processing input data, so it can not do a good research on log sequence, which is a time series type data. Considering the time-series characteristics of network log data, cyclic neural networks are usually used to analyze and predict sequence data. Document [19] proposes a data classification method for network logs based on RNN and SVM, which can classify network logs well when facing a sequence type data, but the algorithm uses a large number of neurons and consumes a lot of software resources.

GRU is a special structure of RNN model [20]. In traditional RNN, due to the multiplication of activation function in the process of back propagation, it is easy to cause gradient disappearance and gradient explosion, which makes it difficult for the later step to learn the information of the previous step under the long time-steps. By adding some links between hidden layer nodes, GRU can handle the input of the relationship with time series very well, and uses a GRU unit to control the output of data, In forward computing, it is deliberately chosen to remember a certain proportion of past information and a certain proportion of present information, which can effectively express the changes in time series. Compared with other RNN models, the length dependence of the algorithm can be expressed by resetting the activity of the gate. From the algorithm structure, we can see that GRU (Gated Recurrent Unit) can extract useful features of time series data and solve the "long dependence" problem in practical application.

B. Motivation and Background of Using SVDD Algorithms

This paper deals with the analysis of network log anomalies on large data platforms. These data are very imbalanced. Most offline network log files collected are normal data. It is more valuable to select a good classification model to analyze this kind of network log.

The traditional neural network uses the output layer to classify the network logs by using Softmax. The derivative of the Sigmoid function is between 0 and 1, as shown in Fig.2. The output value is suitable for binary classification, Softmax function calculates multiple classifications of probability values based on sigmoid ,With the deepening of the network layer, the influence of the error of the output layer becomes smaller, that is, the gradient decreases, and at the same time the phenomenon of "gradient disappears" is brought about. This is the error score of the traditional model. Document [19] proposes a data classification method based on RNN and SVM to classify network logs, which improves the classification effect. However, the original data need to be tagged, and the characteristics of data sets in the practical application environment of this paper are not well applied.

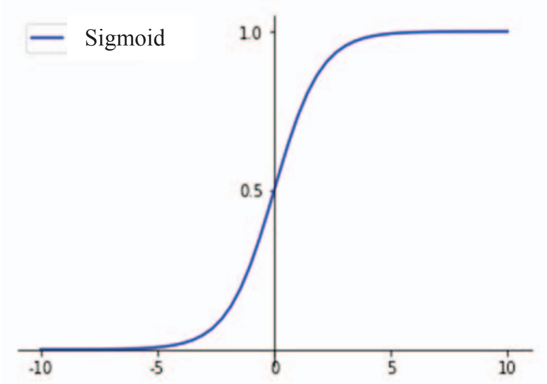

Fig.2. Sigmoid function image

In 1999, Tax and Duin. R P[21] proposed a new single classification method, Support Vector Domain Description (SVDD), based on the theory of Support Vector Machine (SVM). It is one of the most famous support vector learning methods. The algorithm tries to distinguish a group of normal data from all other possible abnormal objects by using the ball defined in the feature space. If there is only one class to be judged, SVDD can include the log data of this kind by finding the most suitable hypersphere. If the data falls into this hypersphere, it can be determined that the data belongs to this class, otherwise, the data does not belong to this class [22]. Therefore, SVDD can be better applied to the output layer of neural network.

IV. Anomaly Detection Algorithm For Network Log Based On GRU And SVDD

This paper presents an efficient anomaly detection algorithm for network log with combining GRU and SVDD algorithms to detect anomalies in network log data sets. Traditional machine learning methods can model network log data sets very well. The GRU selected in this paper is a better algorithm than traditional LSTM, It can extract the effective features from the original log data set very well , GRU can modulate the content of previous memory by reset gate, which can help to more effectively memorize long sequence dependence. The original GRU algorithm can effectively detect anomaly in network logs, but it also has two shortcomings [22]. Firstly, the data analyzed are high-dimensional and the correlation between attributes is strong, which affects the efficiency of anomaly detection. Secondly, the traditional GRU classification of network log data sets can only detect known anomalies, and it is inefficient.

But in this article, an efficient statistical method of Principal Component Analysis (PCA) is used to reduce the dimensionality of the original network log data set with high correlation between high latitude and feature attributes so that The attributes in the network log data set are transformed into a new principal component with uncorrelated attributes. This method can effectively reduce the problem of high false alarm and missed alarm rate. And the GRU is used to extract the attributes of data after dimension reduction pretreatment. Finally, an efficient single classification method of SVDD will be used to replace the output layer of GRU to construct the anomaly classification model for detecting unknown anomalies. The specific flow of the anomaly detection algorithm combined with GRU and SVDD is shown in Fig. 3.

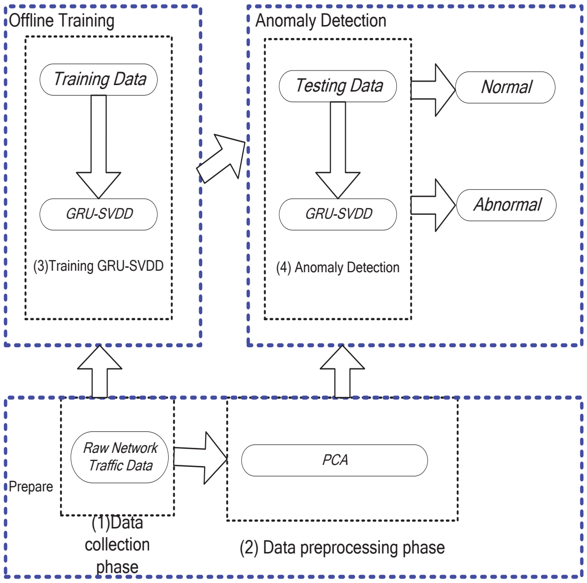

Fig.3. Anomaly detection structure

The process of the algorithm is shown below:

```
Algorithm: Anomaly detection algorithm based on GRU and SVDD
    Input: Original Network Log Data Set (D)
    Output: Anomalous users
1. Preprocesses the original network log data set.
2. Uses principal component analysis to reduce the dimension of the original network log data set and extract the principal component.
3. Divides the data processed by PCA into two parts, one is training sample and the other is testing sample.
4. Inputs the trained data features after processing into GRU model
5. Calculates the cell state in GRU and updates the parameter matrix to initialize the parameters weights and biases.
6. Calculates the results of neural network prediction by SVDD single-class decision function.
7. Uses the Adam optimizer algorithm to optimize the loss function and adjust the parameter matrix weights and biases.
8. Trains the model GRU-SVDD classification model.
9. Enters test data sets containing abnormal user data and returns user abnormalities.
```

Firstly, the original network log data set is preprocessed, including three non-numerical attributes. In the experiment, we first standardize the attributes. Principal Component Analysis (PCA) is used to extract new principal components from the original network log data set. These new principal components reduce redundancy between attributes in the original data set and the interaction between attributes. The processed data sets represented by new components are input into a single classification model of GRU-SVDD, in which GRU extracts more effective feature attributes, replaces the output of GRU neural network with SVDD, and trains a large number of parameters in the neural network with back-propagation neural network. Finally, a high-quality single classification model of normal network log can be trained. It can detect and locate anomalies in test data sets very well.

V. Experiments And Analysis

In order to verify the feasibility of combining GRU with SVDD anomaly detection method, we experimentally implemented in Ubuntu operating system using Python language. The experimental development environment is desktop computer, 4-core CPU, 8G memory, and the development tool is PYCHARM, TENSORFLOW 1.6, an open source machine learning library, is used to construct the available GRU neural network model.

A. Data Set Description

This paper chooses KDD CUP 99 data set [24], in which is constructed for the application of network log anomaly detection method. Its single record has 42 available statistical features, and its 42th features are class markers, marking a variety of abnormal and normal types. This paper studies the four most common types of attacks, DOS, R2L, U2R, PROBE. In this experiment, 10% of the KDD CUP99 data set is selected to train the model. In addition, the experimental data set KDD CUP99 is randomly selected and divided into five groups to make effective comparisons in many aspects.

B. Experimental Steps

In this experiment, detection rate (DR) and false alarm rate (FAR) are used to evaluate the anomaly detection effect of the model. The detection rate indicates the proportion of correctly detected anomalies (TP+TN) to all abnormal users (P+N), while the false alarm rate is the proportion of normal samples that are wrongly detected as abnormal users (FP+FN) in all normal samples (P+N).

The 2nd, 3rd and 4th dimension data of KDD CUP99 are non-numerical and can not be directly input into the anomaly detection algorithm model. Before the experiment, the data of these three dimensions will be pre-processed numerically, specifically, the number of occurrences in these three attributes will be counted first, and then it can be used. Sort them alphabetically, and the serial number is used to replace the original content.

The first group of experiments in this section is to select suitable number of principal components to improve the efficiency of anomaly detection. By selecting different number of new principal components, when the efficiency of anomaly detection is the highest, the number of new principal components is the best choice.

The second group of experiments in this section is to compare the detection rate and false alarm rate of this algorithm. The first comparison is the accuracy and false alarm rate of this method compared with classical GRU-MLP [25], LSTM [24], PCA-SVM [26] and LSTM-RNN [27]. GRU-MLP, LSTM-RNN and LSTM are classical anomaly detection methods. They can detect anomaly quickly in the network, but they are inefficient when applied to large-scale data sets. PCA-SVM is an efficient method to improve SVM, but this method simply reduces the dimension of large-scale data sets. Moreover, these algorithms are a method for marking all data sets and can only detect known anomalies. The algorithm proposed in this section combines GRU with SVDD. It not only uses PCA to reduce the dimension of data in the pretreatment part, but also uses an efficient SVDD single classification algorithm to replace the output of GRU, so that The algorithm not only has higher computational efficiency, but also can find new abnormal cases in real time.

C. Experiments Results and Analysis

This experiment also applies to a large number of parameters. In the research based on document [23], the parameters of the algorithm including multiple neural networks in the algorithm are adjusted manually, but not automatically optimized by super-parameters. As shown in Table I.

TABLE I.

Hyper-Parameteres In Neural Networks

Hyper-parameter	GRU-SVDD
Batch Size	256
Cell Size	256
Dropout Rate	0.85
Epochs	5
Learning Rate	1*10-5

In this section, the algorithm uses normal network logs to train the single classification model. The training assembly randomly selects 30,000 log data from the part labeled

Normal. Five groups of test data will be selected to test the detection efficiency of the algorithm, as shown in Table II.

TABLE II. Data Grouping

Data group	Normal	DOS	R2L	U2R	PROBE	Total
Group 1	2000	2000	1000	500	1500	6000
Group 2	2000	3000	0	0	0	5000
Group 3	2000	0	1000	0	0	3000
Group 4	2000	0	0	500	0	2500
Group 5	2000	0	0	0	1500	3500

The experiment involves the number of principal components selected by attributes after processing data by principal component analysis. The first set of data sets, including all kinds of abnormal samples, can better show the effect of increasing principal component analysis processing data on detection rate. Fig. 4, horizontal ordinate represents the number of principal components selected by principal component analysis, vertical coordinates represents the detection rate of the algorithm.

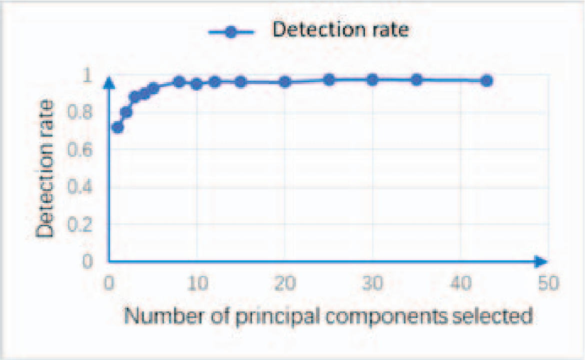

Fig.4. Principal components analysis of data sets

From Fig. 4, it can be seen that the selection of different principal components has an impact on the algorithm of anomaly detection model. When the number selected is small, the effective information extracted by the model is not enough to support the algorithm to detect abnormal information. From the graph, 30 principal components are selected, and the effect is better. The experimental results also show that using principal components to preprocess KDD CUP 99 network log data sets with high correlation between high dimensions and attributes can improve the efficiency of algorithm detection very well. The reason is that KDD CUP 99 data sets have 43 attributes originally, and the dimensions are large, and the parts of these 43 attributes are highly correlated. It will hinder the anomaly detection model. The algorithm proposed in this section is to pre-process the original network log data set by using principal component analysis, and construct new principal component attributes which is easier to express than the original data set.

In order to compare the abnormal cases detected by the exception detection model algorithm against the four most common attack types, the experiments choose the classical BGRU-MLP [25], LSTM [24], PCA-SVM [26] and LSTM-RNN [27] to compare with the improved GRU algorithm in this section. The experiment compares the detection rate and false alarm rate of the four algorithms on the basis of the above five groups of experiments.

TABLE III. Detection Of Five Types Of Anomalies By Four Algorithms

type		mixed	DOS	R2L	U2R	PROBE
GRU-SVDD	DR	98.7%	99.6%	98.5%	56.3%	96.7%
	FAR	2.4%	0.01%	5.2%	0.09%	0.05%
BGRU-MLP	DR	98.2%	99.8%	50.40%	53.84%	95.3%
	FAR	0.82%	0.08%	0.03%	0.05%	0.30%
LSTM-RNN	DR	97.6%	99.2%	3.06%	4.9%	53.2%
	FAR	6.5%	0.9%	0.03%	0.012%	0.14%
LSTM	DR	77.7%	99.5%	92.9%	60.0%	84.7%
	FAR	7.2%	0.7%	0.04%	0.02%	0.2%
PCA-SVM	DR	96.9%	100%	97.50%	17.54%	95.60%
	FAR	5.88%	0.18%	14.73%	0.40%	0.20%

From the experimental results in Table III, we can see that the improved GRU anomaly detection model proposed in this section can be well applied to large-scale high-dimensional network log data sets. Compared with the classical GRU-MLP, LSTM-MLP, LSTM and PCA-SVM algorithms, it can be seen that the detection rate of this section algorithm is better for four kinds of abnormal situations, and the detection efficiency of the other two algorithms is lower for such abnormal situations as U2R. The proposed algorithm is obviously better than them.

There are two reasons for the better accuracy. The first reason is that the algorithm uses PCA to reduce the dimensionality of the attributes of the network log data set, which reduces the redundant attributes of the original record. Secondly, this paper uses an efficient single classification model of SVDD as the output of GRU neural network, and only uses the normal network log to train the model, so it has a good ability to check the abnormal situation such as U2R. And it can also be used to detect new attack types.

In order to make a comparison with more classical algorithms, this paper uses the first set of data sets of mixed cases to compare the detection rates of the algorithm with BP neural network, Semi-Supervised GHSOM [28], GRU-Sigmoid [29], BGRU-MLP [25], as shown in Table IV.

TABLE IV. Five Algorithms For Detecting Mixed Anomaly Types

Algorithm	DR
GRU-SVDD	98.7%
BP	69.5%
Semi-Supervised GHSOM[28]	91.2%
GRU-SoftMax[29]	70.75%
BGRU-MLP[25]	98.2%

From Table IV, we can see that the improved GRU-based anomaly detection algorithm performs well in anomaly detection of network log datasets with high correlation between high dimensionality and feature attributes. At the same time, four simple deep learning algorithms are compared. They also use neural network technology to detect network log datasets. But these algorithms classify all kinds of normal and abnormal attributes in the dataset without considering the unknown attack. This paper is to model the normal network log, as long as it is not in the normal range, it belongs to the abnormal situation. This is also from the big data platform. Most of the access logs collected by the big data platform are normal, and the single classification model can be updated incrementally according to the detected logs. From the detection rate, the proposed algorithm is also better than the other four algorithms.

VI. Conclusion

Although most of the existing anomaly detection algorithms are suitable for anomaly detection of network logs, the traditional algorithms are inefficient, expensive and need data labels when they use large-scale and high- dimensional historical access network log datasets collected by the Industrial Internet of Things. This paper aims at the problem of network intrusion detection in the Industrial Internet of Things. A network log anomaly detection model based on GRU and SVDD is proposed. Firstly, the high-dimensional original data sets are pre-processed by principal component analysis to eliminate redundant feature attributes and improve detection efficiency. Then, the characteristics of pre-processed network log time series data are extracted by using GRU neural network structure. In the final classified output, an efficient single-classification algorithm of SVDD is used to model the normal network log. This method is more efficient and practical. A large number of experiments on the data set and analysis of the experimental results show that the improved GRU-based algorithm is better than the traditional method in anomaly detection. The accuracy is higher.

The limitations of the current algorithm and the focus of the follow-up work are as follows: firstly, the algorithm adds a principal component analysis to pre-process the original data set, but the experiment only reduces the redundancy between attributes and improves the detection efficiency of the algorithm. The new principal component may affect the time efficiency of the algorithm. Secondly, this paper aims at large data. The existing platform environment is to collect a large number of high-dimensional network log data sets, which can be analyzed and detected by the algorithm in this paper. However, in the future, the collected data will be more complex, and the environment will be extended to the field of the Internet of Things. The use of this algorithm in preprocessing data and adjusting algorithm parameters needs further study.

References

[1] Mehta S, Kothuri P, Garcia D L. Anomaly Detection for Network Connection Logs[J]. arXiv preprint arXiv:1812.01941, 2018.

[2] Koziol J. Intrusion Detection with Snort[J]. Advanced Ids Techniques Using Snort Apache Mysql Php & Acid, 2003, 30(10):369-373.

[3] Akila S, Reddy U S. Cost Sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for Credit Card Fraud Detection[J]. Journal of Computational Science, 2018, 27:247-254.

[4] Omrani T, Dallali A, Rhaimi B C. Fusion of ANN and SVM Classifiers for Network Attack Detection[J]. CoRR abs/1801.02746 (2018)

[5] Jidiga G R, Porika S. Anomaly Detection Through Comparison of Heterogeneous Machine Learning Classifiers vs KPCA. SSCC 2015: 483-495

[6] Alnafessah A, Casale G. Anomaly Detection for Big Data Technologies[C]//2018 Imperial College Computing Student Workshop (ICCSW 2018:8:1-8.1). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[7] Mukkamala S , Janoski G , Sung A . Intrusion detection using neural networks and support vector machines[C]// International Joint Conference on Neural Networks. IEEE, 2002

[8] Alexander A. Suárez León, Carolina Varon, Rik Willems, Sabine Van Huffel, Carlos R. Vázquez-Seisdedos:T-wave end detection using neural networks and Support Vector Machines. Comp. in Bio. and Med. 96: 116-127 (2018)

[9] Meshram A, Haas C. Anomaly Detection in Industrial Networks using Machine Learning: A Roadmap[M]// Machine Learning for Cyber Physical Systems. 2017:65-72.

[10] Zhang J, Vukotic I,Anomaly detection in wide area network meshes using two machine learning algorithms. Future Generation Comp. Syst. 93: 418-426 (2019).

[11] Imamverdiyev Y, Sukhostat L. Anomaly detection in network traffic using extreme learning machine[C]// IEEE International Conference on Application of Information & Communication Technologies. 2017.848-863.

[12] Buczak A L, Berman D S, Yen S W. Using sequential pattern mining for common event format (CEF) cyber data. CISRC 2017: 2:1-2:4

[13] Jun Xiao, Xiaolei Yun, Yongzheng Zhang. Application Layer Distributed Denial of Service Attack Filtering Based on Session Abnormality Model[J]. Chinese Journal of Computer Science,2010,33(09):1713-1724

[14] Lin W C, Ke S W, Tsai C F. CANN: An intrusion detection system based on combining cluster centers and nearest neighbors[J]. Knowledge-based systems, 2015, 78: 13-21.

[15] Kwon D, Kim H, Kim J. A survey of deep learning-based network anomaly detection[J]. Cluster Computing, 2017(5):1-13.

[16] Alalshekmubarak A, Smith L S. A novel approach combining recurrent neural network and support vector machines for time series classification[C]// International Conference on Innovations in Information Technology. 2013. 85-91

[17] Zou M, Wang C, Li F. Network Phenotyping for Network Traffic Classification and Anomaly Detection[J]. CoRR abs/1803.01528 (2018)

[18] Thi N N , Cao V L , Le-Khac N A . One-Class Collective Anomaly Detection Based on LSTM-RNNs[J]. 2017.36:73-85.

[19] Innovations in Information Technology (IIT), 2013 9th International Conference on[M].

[20] Cho K , Van Merrienboer B , Gulcehre C. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. Computer Science, EMNLP 2014: 1724-1734.

[21] D M J, Duin R P W. Support vector domain description[J]. Pattern Recognit Lett, 1999, 20(11–13):1191-1199.

[22] Guo S M , Chen L C , Tsai J S H . A boundary method for outlier detection based on support vector domain description[J]. Pattern Recognition, 2009, 42(1):77-83.

[23] Tavallaee M, Bagheri E, Lu W. A detailed analysis of the KDD CUP 99 data set[C]// IEEE International Conference on Computational Intelligence for Security & Defense Applications. 2009.

[24] Zhao Y, Jie L, Shuang X. Investigating gated recurrent neural networks for acoustic modeling[C]// International Symposium on Chinese Spoken Language Processing. 2017.

[25] Xu, Congyuan, Jizhong Shen, Xin Du. "An Intrusion Detection System Using a Deep Neural Network With Gated Recurrent Units." IEEE Access 6 (2018): 48697-48707.

[26] Staudemeyer R C. Applying long short-term memory recurrent neural networks to intrusion detection[J]. South African Computer Journal, 2015, 56(1): 136-154.

[27] Kim, Jihyun, et al. "Long short term memory recurrent neural network classifier for intrusion detection." 2016 International Conference on Platform Technology and Service (PlatCon). IEEE, 2016.

[28] Shilai Y , Yahui Y , Qingni S , et al. A Method of Intrusion Detection Based on Semi-Supervised GHSOM[J]. Journal of Computer Research & Development, 2013, 50(11):2375-2382.

[29] Abien Fred Agarap:A Neural Network Architecture Combining Gated Recurrent Unit (GRU) and Support Vector Machine (SVM) for Intrusion Detection in Network Traffic Data. ICMLC 2018: 26-30.