A PSO-based Wavelet-core ELM for Abnormal Flow Detection

Yueyang Su[1], Jing Wan[1], and Junkai Yi[2]
(Corresponding author: Jing Wan)
College of Information Science and Technology, Beijing University of Chemical Technology[1]
North Third Ring Road 15, Chaoyang District, Beijing, China
College of Information Science and Technology, Beijing Information Science and Technology University[2]
North Fourth Ring Road 35, Chaoyang District, Beijing, China
(Email: suyy1225@163.com)
(Received July 31, 2018; Revised and Accepted Jan. 9, 2019; First Online July 16, 2019)

Abstract

Abnormal flow detection is an effective approach to discover the covert data during the transmission process of mass data. However, there exist some issues to tackle such as the high complexity of network traffic data, Low detection efficiency and low accuracy. To solve these problems, we proposes an improved wavelet-core extreme learning machine based on particle swarm optimization. First, the particle swarm optimization algorithm is applied to determine the input weights and bias thresholds of the extreme learning machine, which effectively reduces the number of hidden layer nodes. Furthermore, wavelet kernel function is proposed to be the kernel function of kernel extreme learning machine. Then the topology of the KELM can be established, and can be applied to classify the abnormal traffic. We introduce overall-accuracy and F-measure for performance measure in abnormal flow detection. To verify the effectiveness of our work, we compare the approach with the representative algorithms, and experimental results show that the improved wavelet-core extreme learning machine based on particle swarm optimization has better detection performance.

Keywords: Abnormal Flow Detection; KELM; Particle Swarm Optimization Algorithm; Wavelet Kernel Function

1 Introduction

In the distributed Internet environment, the discovery and analysis of covert data in the process of mass data transmission is a serious problem to be solved, and it is also a main guarantee for the healthy development of the virtual economy in the future. In recent years, the data transmission technology based on covert channel has developed rapidly. The covert data transmission of massive multi-modal information based on blockchain [4, 12] has been brought to our attention. Meanwhile, the security issues has been increasingly significant, the discovery and analysis of covert data has become an important requirement for new network applications.

Abnormal traffic detection is an important technology for the discovery of covert data. The detection of network abnormal flow is to analyze network flow data via statistical analysis, data mining and machine learning with the intention to discover abnormal information of network data.

Statistical analysis is an early method for anomaly detection of traffic data. First, count the number of network traffic packets, the length of packets and other characteristic information. Then, discover the characteristics rules of the traffic data. Finally, in order to detect the abnormal flow information, establish the normal behavior profile of traffic data as the standard of judgment of the traffic data to be detected. Hoang *et al.* applied the principal component analysis method and the wavelet transform to make the model, which combined with the spatio-temporal correlation of the feature matrix of traffic data [6]. They proposed an PCA-based network anomaly detection algorithm. Although this method has a good effect, it is difficult to detect during network transmission and achieve real-time abnormal traffic detection. Bhuyan *et al.* proposed a DDOS attack detection method based on the characteristics of traffic and extended entropy metric, which effectively reduced the computational complexity and detection time [3]. However, the process of establishing the traffic model is still complex and it is also difficult in the practical application. Although the statistical analysis method has a good detection effect, it is sensitive to the change of the threshold value, and at the same time, it cannot reflect the autocorrelation of the abnormal behavior in time.

Data mining is a major approaches for the detection of anomaly traffic data, which aims to establish awareness model of anomalous traffic by analyzing the mass flow

data [5]. Clustering algorithm is an important method of data mining. Unsupervised learning method can be used to classify the heterogeneous and high-dimensional massive traffic data. The density peak clustering algorithm [11] based on the assumptions as following: Within clusters, the local density of the clustering center points is the highest; Among clusters, the clustering center is away from other clusters'. The algorithm has a good effect on various data distributions, but it is sensitive to the global abnormalities in some special cases and has poor results. Ahmed *et al.* established a collective anomaly detection framework via partition clustering technology to detect DoS attacks and improve the detection accuracy [2]. However, the algorithm has limitations and works weakly in the detection of other attacks. Although the clustering algorithm can be used with unclassified sample data, the speed and accuracy is still far away from application.

The classification algorithm is a supervised machine learning method. Hua [7] applied the K-means algorithm to improve the traditional KNN and divided the process of anomaly detection into two parts: off-line preprocess and on-line classification, which improved the efficiency and classification accuracy. But the feature redundancy and dimensionality disasters is still the most serious problem of the algorithm. Ma [9] and his partners applied the Naïve Bayesian network to construct classifiers for traffic classification, which limited by the fixed assumptions. Roy [10] detected and analyzed attack behaviors via deep neural networks, which improved the efficiency and performance of anomaly detection. However, the low iteration speed is still the most serious problem and it is also easy to fall into local convergence.

The extreme learning machine is a single-hidden layer feedforward neural networks. The model randomly selects the hidden layer nodes and replaces the iterative process for adjusting parameters by analyzing to get the weight matrix between the hidden layer and the output layer. With the great learning efficiency and self-adaptive ability, many researchers have devoted to the study of improving extreme learning machines. Kumari *et al.* proposed a semi-supervised support vector machine with fuzzy c-means clustering, which greatly reduced the computational complexity and improved the classification efficiency [8].

There are many abnormal traffic detection approaches with some problems around. Some existed approaches are simple to cope with the weight of traffic statistical features, and the process with equal weight may cause the loss of information. Otherwise some methods also use the whole traffic as analysis objects, and the amount of data is enormous, which can lead to low accuracy rate and efficiency. In this paper, we propose an abnormal traffic detection approach based on Particle Swarm Optimization (PSO) and wavelet kernel Extreme Learning Machine (ELM). PSO is an algorithm of global searching optimal solutions. The algorithm can be introduced to set the optimal input weight and the bias threshold of kernel ELM, which eliminates redundant nodes of hidden layer.

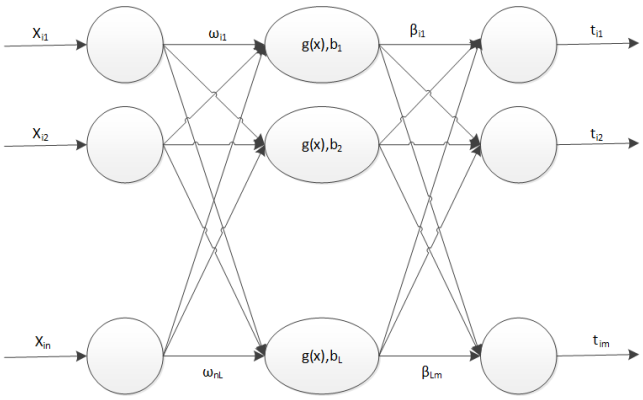

Figure 1: The extreme learning machine network model

With the application of PSO, the classification accuracy and learning efficiency is greatly improved and it is not sensitive to the number of training samples and hidden layer nodes as before. We choose wavelet kernel function as the kernel function of ELM, which improves the ability of nonlinear approximation and generalization. The experiments shows that the model we proposed has great robustness and better detection performance.

2 The PSO-based Wavelet-Core Extreme Learning Machine

2.1 Extreme Learning Machine

Extreme learning machine is a single hidden layer neural network. With random hidden layer nodes, ELM effectively reduces the training time and improves the generalization ability.

Similar to the traditional network model, the model of extreme learning machine is divided into three layers: input layer, hidden layer, and output layer. The specific structure is shown in Figure 1.

Given N sets of training data (x_i, t_i), $x_i = [x_{i1}, x_{i2}, \cdots, x_{in}]^T \in R^N$, $t_i = [t_{i1}, t_{i2}, \cdots, t_{im}]^T \in R^m$, The mathematical model of SLFN with L hidden layer nodes can be described as Equation (1):

$$y_j = \sum_{i=1}^{L} \beta_i g_i(x_i) = \sum_{i=1}^{L} \beta_i g(w_i x_j + b_i) = T_j, \quad (1)$$

where $g(x)$ is the activation function, β_i is the connection weight vector between the hidden layer and the output layer, w_i is the connection weight vector between the hidden layer and the input layer. b_i is the threshold of the ith node in the hidden layer. The above formula can also be simplified as Equation (2):

$$H\beta = T, \quad (2)$$

where H is the output matrix of the hidden layer node, β is the weight vector of the output layer, T is the expected

output matrix of the sample.

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_L \cdot x_1 + b_L) \\ \vdots & \dots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_L \cdot x_N + b_L) \end{bmatrix}_{N \times L} \tag{3}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_m^T \end{bmatrix}_{N \times m} \tag{4}$$

The ELM can be trained without adjusting the weights and bias threshold of the input layer, the output matrix of the hidden layer is only determined by the random w_i and b_i. Therefore, the training of extreme learning machines can be transformed into the process of deriving the output weights β according to the $H\beta = T$. β can be expressed as Equation (5):

$$\beta = H^{\dagger}T, \tag{5}$$

where $H^{\dagger}$ is the Moore-Penrose generalized inverse matrix.

2.2 Kernel-ELM

In order to improve the generalization ability of the extreme learning machine, a kernel function was introduced and then the kernel-ELM (KELM) was proposed.

For the traditional extreme learning machine, the output matrix of hidden layer can be expressed as Equation (6):

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} \tag{6}$$

where $h(x_i)$ can be regarded as a non-linear mapping of x_i, if the mapping is unknown, a kernel function M can be constructed instead of HH^T. According to Mercer, the kernel matrix can be defined as:

$$HH^T, m_{ij} = h(x_i)h(x_j) = k(x_i, x_j), \tag{7}$$

where $i, j \in (1, 2, \cdots, N)$

$$h(x)H^T = \begin{bmatrix} \mathrm{k}(x, x_1) \\ \vdots \\ k(x, x_N) \end{bmatrix} \tag{8}$$

where $k(x_i, x_j)$ is the kernel function. Then the output function $f(x)$ of KELM can be expressed as Equation (9):

$$f(x) = [k(x, x_1), \cdots, k(x, x_N)] \left[\frac{1}{C} + M\right]^{-1} T. \tag{9}$$

2.3 The Morlet Wavelet Function

The kernel function which satisfies the premise of Mercer's theorem can be used as the kernel function of the kernel-ELM. Linear kernel is the simplest kernel function; polynomial kernel is a kind of non-standard kernel function which is suitable for orthogonal normalized data; Gaussian kernel function is widely used in image processing and has a great anti-jamming capability of noise in data. For the complexity of network traffic data, the Morlet wavelet function is introduced to be the kernel function of the kernel-ELM in this paper, which has great classification effect in the space without training data.

In general, the wavelet basis function can be expressed as Equation (10):

$$h_{a,b}(x) = \sqrt{a}\Phi(\frac{x-b}{a}), \tag{10}$$

where $h(x)$ is the mother wavelet function, a is the scaling factor, b is the balance factor, According to the tensor product theory, any multidimensional wavelet function can be expressed as a tensor product of multiple one-dimensional wavelet functions as Equation (11):

$$h(x) = \prod_{i=1}^{n} h(x_i). \tag{11}$$

Construct the translation-invariant kernel function via Equation (10) :

$$K(x, x')=K(Kx - x') = \prod_{i=1}^{n} \Phi(\frac{x_i - x_i{}'}{a}). \tag{12}$$

For sample x, $x' \in R$, the Morlet wavelet function $h(x) = \cos(1.75x)\exp(-\frac{x^2}{2})$, construct the kernel-ELM with the corresponding wavelet kernel function. The specific formula is as Equation (13):

$$\begin{aligned} & Waveletkernel(x, x') \\ = & \prod_{i=1}^{n} [\cos(1.75(\frac{x_i - x_i{}'}{a}))\exp(-\frac{(x_i - x_i{}')^2}{2a^2})] \end{aligned} \tag{13}$$

2.4 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a strategy proposed by Eberhart and Kennedy to solve optimization problems inspired by the feeding behavior of birds. In the POS, the solution to each optimization problem is considered as a location in the search space, called "particles." The particle velocity determines the direction and distance of the particle's flight. It can also track its own optimal position in the iterative process and the best position of all particles in the entire particle group with their own privacy memory. As a basis, it can update the speed and location.

The individual extremum is $R_i^b(t)$, the global extremum is $R_g^b(t)$. Then the motion equation of particle is as Equation (14):

$$\begin{aligned} v_i(t+1) &= \omega v_i(t) + c_1 R_1 \left[R_i^b(t) - x_i(t)\right] \\ &\quad + c_2 R_2 \left[R_g^b(t) - x_i(t)\right], \end{aligned} \tag{14}$$

$$x_i(t+1) = x_i(t) + \phi v_i(t+1), \tag{15}$$

where $v_i(t)$ and $x_i(t)$ are the velocity and position of the ith particle at the tth iteration; c_1, c_2 are the learning factors, R_1 and R_2 are the random variable which are evenly distributed over the interval $[0, 1]$, ϕ is the contraction factor.

The pocess of PSO algorithm is as follows:

Step 1: Initialize the particle group, each particle is set with a random position and velocity;

Step 2: Evaluate the fitness of each particle;

Step 3: For each particle, compare the fitness value and its historical best position pbest, if better, update the pbest;

Step 4: For each particle, compare the fitness value with its gbest, if better, update the gbest;

Step 5: Adjust the speed and position of the particles according to (2) and (3);

Step 6: If it does not satisfied the ending condition, go to Step 5.

The ending condition of iteration depends on the specific problem. In general, it can be ended when the optimal position of the particle swarm satisfies the predetermined minimum adaptive threshold or the times of iterations reaches the maximum number.

Particle Swarm Optimization (PSO) is a global optimization algorithm via randomly searching. The cooperation mechanism between groups is introduced to reach the optimal solution. It has been widely applied to engineering optimization with the good robustness, simple operation, and free from constraints.

2.5 The Wavelet-core ELM Based On The Particle Swarm Optimization (PW-ELM)

In general, the wavelet-core extreme learning machine has good performance in abnormal flow detection, but the accuracy of the ELM is affected by many factors, such as the number of hidden layer nodes, bias threshold and so on. The number of nodes in the hidden layer has a great influence on the generalization ability and learning speed of the ELM. Too many nodes may lead to the increasing of the network complexity and overfitting. The value of the bias threshold and connection weight can also affect the training process of the ELM because of the direct relationship with the output weight. When they are both zero, some nodes of hidden layer will be invalid. In order to improve the learning process of the ELM and optimize the connection weights and thresholds, the particle swarm optimization algorithm is introduced.

In the wavelet-core ELM based on the particle swarm optimization, we abstract the input weights and bias thresholds into particles in the particle swarm, and take the root mean square error of the particles as the fitness function. The particle length(L) is determined by the number of hidden layer nodes (k) and the number of input layer nodes (m), as Equation (16):

$$L = k(m + 1). \tag{16}$$

The process of algorithm is as follows:

Step 1: Select the training data, set the input vector and the expected output vector;

Step 2: Set the topological structure of the wavelet-core ELM, initialize the number of neurons in the input layer, hidden layer, and output layer;

Step 3: Create a particle swarm based on the input vector and bias threshold of the ELM. Set the initial speed, position of the particles, and the optimize space;

Step 4: Set a suitable fitness function, the root mean square error of the particle is choosed for our model . Set the maximum number of iterations = 600, learning factor $c_1 = c_2 = 1.5$, population size $M = 25$ and the particle dimension D;

Step 5: Calculate the fitness of the particle based on the training set, find the individual extreme value and the global extreme value;

Step 6: Update the position and speed of the particles;

Step 7: If it does not satisfied the ending condition, go to Step 5;

Step 8: Establish the wavelet-core ELM with the input weights and hidden layer bias thresholds generated by the particle swarm optimization algorithm.

3 Experimental Results And Analysis

3.1 Date Collection

The wavelet-core extreme learning machine based on particle swarm optimization has good generalization ability and classification accuracy. In order to verify the performance of the improved kernel-ELM, the KDD 99 dataset was selected as the analysis object.

The KDD 99 dataset is a competition dataset used by the International Data Mining and Knowledge Discover competition in 1999. The dataset was established based on the Intrusion Detection Evaluation Project of US Department of Defense Advanced Planning Agency (DARPA) in 1998, which collected data from the simulated military network in the Lincoln Lab. The collection of data lasted for two and a half months, including different network traffic and attack methods. The aims of competition is to detect the network intrusion and achieve the abnormal classification of network connections.

A network connection consists of a sequence of TCP packets from the beginning to the end in a certain period

of time. During this period of time, the data transfers between the original address and the destination address based on a predefined protocol. The network connection record contains a status bit to mark it normal or attack. The types of exceptions can be categorized as: Remote-to-Login Attack (R2L), Denial of Service Attack (DoS), Probing Attack (PROBING) and User-to-Root Attack (U2R).

The KDD 99 dataset is divided into two subset, the one is the training dataset which contains about 5,000,000 records, and another is the test dataset including about 2,000,000 records. The distribution of samples is shown in Table 1.

3.2 Extract Features Of Network Flow

There are various network attacks divided into 4 categories. The Denial of Service is the most common one including UDP floods, Land attacks, e-mail bombs, *etc.*; The other is Exploitable Attacks which contains Password Guessing, Trojans, Buffer Overflows, *etc.*; The information-gathering Attacks is used to obtain the useful information including Address Scanning, Port Scanning and DNS Domain Conversion; The last one is False-messaging Attacks which mainly contains DNS Cache Pollution and Fake Emails, *etc.* [1]. Different network attack methods are different in the abnormal behavior of traffic data. And it is important to select the appropriate statistics features of data flow. The number of features is also important for the classification accuracy. In general, the larger the number of feature values is, the higher the classification accuracy will be. However, when the number is too large, the overall performance of the classifier would be worse [13].

We selected 16 representative characteristics of data flow in this paper including the network service type of the target host, the number of urgent packets, the number of error segments, the connection status (normal or error), and transmission protocol, *etc.* The specific information is in Table 2.

3.3 Data Preprocessing

For the complexity of the sample data about network connection, the input data of classifier needs to be normalized to reduce the classification error and accelerate the convergence speed.

$$X = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{17}$$

where, $x_{\max}$ is the maximum in the dataset, $x_{\min}$ is the minimum.

3.4 Evaluation Function

In order to analyze the experimental results, the $Overall - accuracy$ and $F - measure$ were used to evaluate the classification performance of different methods. TP is the number of samples which is correctly classified, FP is the number of other classes' samples which are erroneously divided into this class, FN is the number of the a certain class's samples which are misclassified.

$$precision(i) = \frac{TP_i}{TP_i + FP_i} \tag{18}$$

$$recall(i) = \frac{TP_i}{TP_i + FN_i} \tag{19}$$

The $Overall - accuracy$ used in this paper is the ratio of the model's correct prediction to the total number on all the test sets, the specific formula is as Equation (20):

$$Overall - accuracy(i) = \frac{\sum_{i=1}^{m} TP_i}{\sum_{i=1}^{m} TP_i + FN_i} \tag{20}$$

There are sometimes contradictions between the *precision* indicator and the *recall* indicator. In order to consider them comprehensively, $F - measure$ is introduced. It is a reconciliation measure between *recall* and *precision*. The specific formula is as Equation (21):

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \tag{21}$$

3.5 Experimental Results

In the experiment, we selected 10% training subsets and 10% test subsets from the KDD 99 dataset. The distribution of sample is shown in Table 3.

First, the features should be numerically normalized to convert the data to the standard input data of ELM. After 10 experiments, we find the average of the 10 accuracy as the final result of the experimental accuracy.

In order to verify the performance of the wavelet-core extreme learning machine based on particle swarm optimization in anomaly detection, we selected the non-kernel ELM, Gaussian kernel ELM (Gauss-kernel ELM), and Gaussian kernel support vector machine (Gauss-kernel SVM) as contrast on the KDD 99 dataset. The parameters of each classifier are shown in Table 4.

After 10 experiments, we find the $Overall - accuracy$ and $F - measure$ for analysis, the results and average of $Overall - accuracy$ in the 10 experiments are shown in Table 5 and Table 6.

According to the Table 5 and Table 6, we can realize that the $Overall - accuracy$ of PW-ELM, Gauss-kernel SVM, Gauss-kernel ELM and ELM on the KDD 99 dataset are: 94.746%, 90.205%, 83.207% and 74.942%. The $Overall - accuracy$ of PW-ELM is obviously higher than the other algorithms, approaching 95%. The performance of Gauss-kernel SVM is higher than the ELM and Gauss-kernel ELM. In summary, the performance of the PW-ELM achieve an ideal $Overall - accuracy$ in abnormal flow detection.

The $F - measure$ of the four algorithms is shown in Figure 3. On the KDD 99 dataset, PW-ELM has

Table 1: The distribution of KDD 99 dataset

Category	**Normal**	Abnormal			
	Normal	Dos	U2R	R2L	PROBE
Training Dataset	0.1969	0.7924	0.0001	0.0022	0.0083
Test Dataset	0.1975	0.7490	0.0007	0.0528	0.0136

Table 2: The representative characteristics of data flow

characteristics	**description**	**amount**
Network Connection	network service type of the target host, the number of expedited packets,the number of error segments, connection status (normal or error), transmission protocol	5
Package	the number of packages	1
Bytes	the bytes of data from the source host to the destination host, The bytes of data from the target host to the source host	2
Packet size	the average, maximum, minimum, standard deviation of packet size	4
Connection time	the average, maximum, minimum, standard deviation of connection time	4
total		16

Table 3: The distribution of training subsets and test subsets

Category	**Normal**	Abnormal			
	Normal	Dos	U2R	R2L	PROBE
Training Dataset	97278	391458	52	1126	4107
Test Dataset	60593	229853	228	16189	4166

Table 4: The parameters of each classifier

Algorithm / Parameters	**ELM**	**Gauss-kernel ELM**	**Gauss-kernel SVM**	**PW-ELM**
Penalty factor(C)	1000	1000	1000	1000
Kernel parameters(a)		2.5	1.8	2.0
The number of hidden layer nodes(L)	800			

Table 5: The $Overall-accuracy$ of classifiers

$Overall-accuracy$	**1**	**2**	**3**	**4**	**5**
ELM	74.273	74.611	75.902	75.059	74.385
Gauss-kernel ELM	81.925	82.328	83.667	83.516	82.739
Gauss-kernel SVM	90.325	90.816	89.884	89.857	90.251
PW-ELM	94.561	95.092	93.829	94.966	95.362

Table 6: The $Overall-accuracy$ of classifiers

$Overall-accuracy$	**6**	**7**	**8**	**9**	**10**	**Average**
ELM	75.109	75.433	74.927	74.658	75.062	74.942
Gauss-kernel ELM	84.051	83.152	82.694	84.973	83.049	83.207
Gauss-kernel SVM	90.537	90.032	89.334	90.238	90.776	90.205
PW-ELM	94.466	94.752	95.093	95.372	93.964	94.746

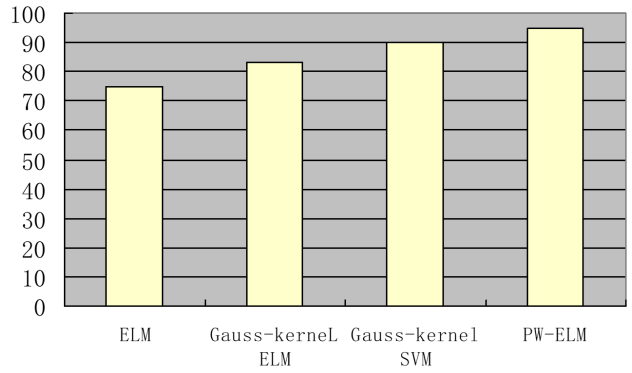

Figure 2: The $Overall - accuracy$ on the KDD 99

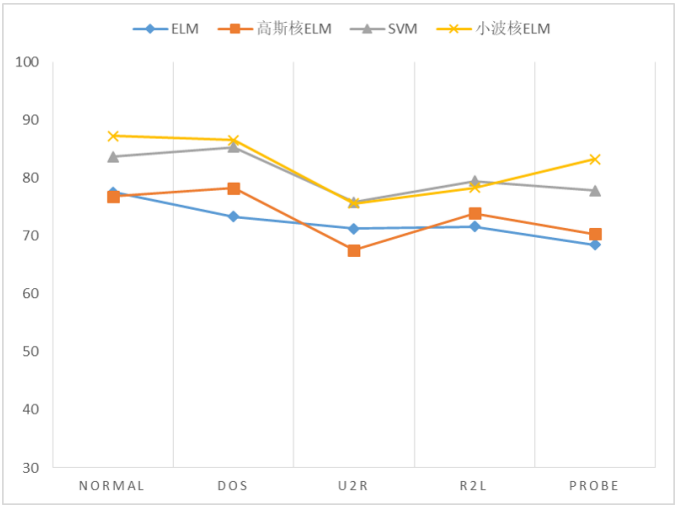

Figure 3: The $F - measure$ on the KDD 99

a good classification effect on Dos and PROBE. Gauss-kernel SVM is slightly better than PW-ELM in detecting R2L. However, in conclusion, compared with ELM, Gauss-kernel ELM and Gauss-kernel SVM, PW-ELM has more advantages and can be widely used in application.

4 Conclusions

In this paper, we introduced a wavelet-core extreme learning machine based on particle swarm optimization to detect abnormal traffic. Experiments show that the model can achieve good performance in the detection of abnormal network traffic. Compared with the ELM, Gauss-kernel ELM and Gauss-kernel SVM, the nonlinear approximation ability and generalization ability are greatly improved. And the model also solve the problem about the redundancy of hidden layer node and inefficiency. With the better learning efficiency and classification accuracy, the wavelet-core extreme learning machine based on particle swarm optimization can be widely use in the anomaly detection of network traffic data.

Acknowledgments

This study was supported by Projects U1636208 funded by National Natural Science Foundation of China (NSFC).

References

[1] M. Ahmed, "Collective anomaly detection techniques for network traffic analysis," *Annals of Data Science*, vol. 5, no. 4, pp. 497-512, 2018.

[2] M. Ahmed and A. N. Mahmood, "Novel approach for network traffic pattern analysis using clustering-based collective anomaly detection," *Annals of Data Science*, vol. 2, no. 1, pp. 111–130, 2015.

[3] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "E-ldat: A lightweight system for ddos flooding attack detection and ip traceback using extended entropy metric," *Security and Communication Networks*, vol. 9, no. 16, pp. 3251–3270, 2016.

[4] K. Croman, C. Decker, I. Eyal, A. E. Gencer, A. Juels, A. Kosba, A. Miller, P. Saxena, E. Shi, E. G. Sirer, *et al.*, "On scaling decentralized blockchains," in *International Conference on Financial Cryptography and Data Security*, pp. 106–125, 2016.

[5] S. M. A. M. Gadal and R. A. Mokhtar, "Anomaly detection approach using hybrid algorithm of data mining technique," in *International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE'17)*, pp. 1–6, 2017.

[6] D. H. Hoang and H. D. Nguyen, "A pca-based method for iot network traffic anomaly detection," in *The 20th International Conference on Advanced Communication Technology (ICACT'18)*, pp. 381–386, 2018.

[7] H. Y. Hua, Q. M. Chen, H. Liu, Y. Zhang, and P. Q. YUAN, "Hybrid kmeans with knn for network intrusion detection algorithm," *Computer Science*, vol. 3, pp. 32, 2016.

[8] V. V. Kumari and P. R. K. Varma, "A semi-supervised intrusion detection system using active learning svm and fuzzy c-means clustering," in *International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC'17)*, pp. 481–485, 2017.

[9] Y. Ma, S. Liang, X. Chen, and C. Jia, "The approach to detect abnormal access behavior based on naive bayes algorithm," in *The 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS'16)*, pp. 313–315, 2016.

[10] S. S. Roy, A. Mallik, R. Gulati, M. S. Obaidat, and P. V. Krishna, "A deep learning based artificial neural network approach for intrusion detection," in *International Conference on Mathematics and Computing*, pp. 44–53, 2017.

[11] S. Wang, D. Wang, C. Li, Y. Li, and G. Ding, "Clustering by fast search and find of density peaks with

data field," *Chinese Journal of Electronics*, vol. 25, no. 3, pp. 397–402, 2016.

[12] A. Wright and P. D. Filippi, "Decentralized blockchain technology and the rise of lex cryptographia," *SSRN Electronic Journal*, 2015. (`https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2580664`)

[13] L. Zheng, R. Diao, and Q. Shen, "Self-adjusting harmony search-based feature selection," *Soft Computing*, vol. 19, no. 6, pp. 1567–1579, 2015.

Biography

Yueyang Su Yueyang Su is a graduate student in School of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China. His research focuses on cyberspace security,artificial intelligence.

Jing Wan Jing Wan received her Ph.D. degrees in Beijing University of Chemical Technology. She is now a Professor at School of Information Science and Technology in Beijing University of Chemical Technology, China. Her research focuses on cyberspace security, artificial intelligence, intelligent information system, Knowledge Graph.

Junkai Yi Junkai Yi received his MS and Ph.D. degrees in Computer Science from Beijing Institute of Technology in 1995 and 1998 respectively. He is now a Professor at School of Information Science and Technology in Beijing Information Science and Technology University, China. His research focuses on cyberspace security, artificial intelligence, intelligent information system, text classification, pattern recognition.