symmetry

MDPI

Article

Key Feature Recognition Algorithm of Network Intrusion Signal Based on Neural Network and Support Vector Machine

Kai Ye

Department of Educational Affairs, Zhengzhou Institute of Technology, Zhengzhou 450044, China; ye18135777788@163.com

Received: 14 January 2019; Accepted: 6 March 2019; Published: 14 March 2019

Abstract: When identifying the key features of the network intrusion signal based on the GA-RBF algorithm (using the genetic algorithm to optimize the radial basis) to identify the key features of the network intrusion signal, the pre-processing process of the network intrusion signal data is neglected, resulting in an increase in network signal data noise, reducing the accuracy of key feature recognition. Therefore, a key feature recognition algorithm for network intrusion signals based on neural network and support vector machine is proposed. The principal component neural network (PCNN) is used to extract the characteristics of the network intrusion signal and the support vector machine multi-classifier is constructed. The feature extraction result is input into the support vector machine classifier. Combined with PCNN and SVM (Support Vector Machine) algorithms, the key features of network intrusion signals are identified. The experimental results show that the algorithm has the advantages of high precision, low false positive rate and the recognition time of key features of R2L (it is a common way of network intrusion attack) data set is only 3.18 ms.

Keywords: neural network; support vector machine; network; intrusion signal; key features; classifier

1. Introduction

With the development of computer network technology, network security is increasingly becoming the focus of attention. Intrusion detection is the process of identifying and responding to computer and network resource intrusions. At present, the existing intrusion detection and identification methods have problems such as high false positive rate, poor system intelligence and lack of early warning function. The key to intrusion detection is how to extract representative system feature patterns from a large amount of audit data to identify the user's normal behavior and abnormal behavior. In essence, intrusion detection needs to solve the problem of attacker planning and identification in the field of cyber attack defense [1]. Intrusion detection methods are mainly divided into abuse detection and anomaly detection. Among them, anomaly detection is to establish a user's normal behavior model in advance, and detect according to the principle of whether the system state deviates from normal, so it can detect new intrusion methods [2,3]. In order to obtain ideal detection results, people seek to apply machine learning theory to anomaly detection, and carry out the research of anomaly detection algorithm based on artificial immune theory, neural network, D-S evidence theory and so on. Good results have been achieved [4]. However, most of these methods do not consider the analysis of intrusion characteristics because they are designed to improve detection accuracy and reduce false alarm rate [5,6]. In practice, although all input features can be used as the input of the classifier, irrelevant or redundant features may have a negative impact on the generalization ability of the classifier [7–9].

Current feature analysis techniques mainly include feature selection and feature extraction. Feature selection is based on a certain performance criterion to select important features from all input

features and remove secondary features, which is often conducive to shortening the detection time and discovering the intrinsic features of a certain type of attack [10]. Feature extraction is not to explicitly remove some features from the input features, but to carry out linear or nonlinear transformation of the input features, extracting from them to replace the table components, using these components instead of the original input features, so as to achieve dimensionality reduction and feature space conversion [11–13].

At present, there are few researches on intrusion detection from the perspective of feature extraction. Authors in [14] constructed a multi-layer hybrid intrusion detection model, using support vector machine and extreme learning machine to improve the efficiency of detection of known and unknown attacks; then proposed an improved k-means method, established a new small training data set representing the entire training data set, greatly shortened the training time of classifier, and improved the performance of the intrusion detection system. However this method has the problem of low accuracy. Authors in [15] constructed a wireless grid intrusion detection system based on genetic algorithm and multi-support vector machine classifier. The system chooses the information characteristics of each type of attack, rather than the common features of all attacks. The network simulator is used to simulate the intrusion data set generated by wireless mesh network. The system is evaluated with the parameters of packet transmission rate and delay. However, the system has the problem of high false alarm rate.

It can be mentioned that dimensionality reduction in NN inputs and topology to enhance generalisation has been applied in other fields. Authors in [16] constructed a spectrum prediction method based on back propagation-training model of neural network. In order to optimize the structure of the neural network and reduce the aggressive weight, a genetic algorithm is used to avoid falling into the local optimal solution. Selection, crossover, and mutation functions are used to increase randomness, so that the population converges to a set containing global optimal solutions. Authors in [17] studied the accuracy of the combined model of artificial neural network (ANN) and singular spectrum analysis (SSA) in monthly precipitation prediction. First, the original input signal is pre-processed by SSA to provide high quality data for artificial neural network. Combined with artificial neural network and SSA, a rainfall-forecasting model was established to study the accuracy of monthly precipitation forecasting.

In this paper, a key feature recognition algorithm of network intrusion signal based on neural network and support vector machine is proposed. Generalized learning rules are used to train linear principal component neural network to extract feature components, and then support vector machine is used to do so. Classification, Support Vector Machine (SVM) is a recently developed machine learning theory. It is becoming a popular method in the field of intrusion detection due to its strong generalization ability, no local minima, and sparse representation of solutions.

2. Materials and Methods

2.1. Key Feature Extraction of Network Intrusion Signals Based on PCNN

Feature extraction is the first step in identifying the key features of network intrusion signals. The purpose is to transform the input data into new low-dimensional features by principal component neural networks. Principal component analysis is a linear feature extraction method based on second-order statistical analysis of data. It is essentially equivalent to the Karhunen-Loeve transform (KLT) in signal processing [18,19]. By calculating the eigenvectors of the sample covariance matrix, PCA linearly maps the input space into a low-dimensional eigenvector space and the new features are irrelevant.

Principal Component Neural Network (PCNN) is a kind of neural network for principal component analysis. It can be implemented in many structures, such as single-layer neural network, autocorrelation multi-layer perceptron, etc. The generalized Hebbian algorithm (GHA) is a common

learning algorithm for multi-principal component extraction. It is used to train single-layer neural networks for principal component extraction. Its structure is shown in Figure 1.

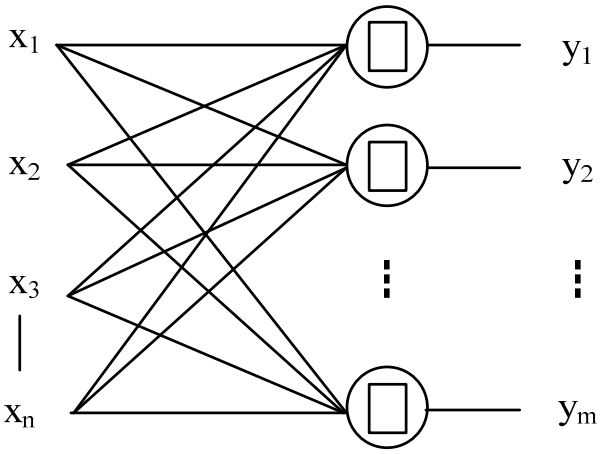

Figure 1. Structure diagram of principal component neural network.

Usually the output number of the network (extraction feature number) m is less than the n of the network. The output of the i neuron is $y_i = w_i^T x$, and the weight of neurons is changed by formula (1):

$$\Delta w_i(k) = \beta(k) y_i(k) \left[x(k) - \sum_{l=1}^{i} y_l(k) w_l(k) \right] \tag{1}$$

Among them, β is the learning factor, x is the input vector, and w_i is the synaptic weight vector of the i neuron.

2.2. Identification of Key Features of Network Intrusion Signals Based on Multi-Class Support Vector Machine

2.2.1. Linear Support Vector Machine

Support Vector Machine (SVM) is a method proposed by Vapnik et al. based on statistical learning theory to achieve structural risk minimization. Its learning strategy is to keep the empirical risk value fixed and minimize the confidence range.

(1) Linear separable condition

Support Vector Machine (SVM) is proposed from the optimal classification hyperplane with linear separability, which is mainly aimed at the binary classification problem. The goal is to find a hyperplane so that it can correctly separate the two types of data points without errors, while keeping the separated data points farthest from the classification surface [20], as shown in Figure 2.

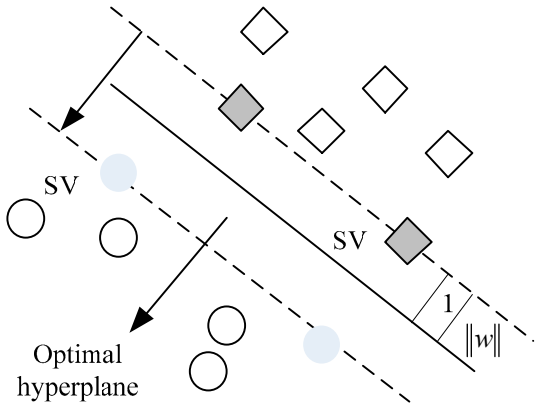

Figure 2. Optimal hyperplane in linearly separable case.

The solution is to construct a constrained quadratic programming problem, specifically, a constrained quadratic programming problem, specifically: $\min \frac{1}{2}\|w\|^2$, $sty_i(w \cdot x_i + b) \geq 1, i = 1, 2, \cdots, n$, to solve the problem and get the classifier.

According to statistical learning theory, the error-free separation ensures that the empirical risk is minimized (0); the distance between classifications is maximized, that is, the allowable empirical risk is realized by the simplest learning machine, and the confidence range of the generalization bound is minimized, thus minimizing the real risk. Therefore, the support vector machine has good generalization ability. The generalization ability refers to the adaptability of machine learning algorithm to fresh samples. The higher the generalization ability, the better the adaptability of the algorithm to fresh samples. The training samples of the nearest point from the classification plane and parallel to the optimal hyperplane in the two kinds of samples are called support vector.

(2) Linear inequalities.

For the linear inseparable case, the relaxation variable $\zeta_i \geq 0$ is introduced, $i = 1, 2, \cdots, n$, and the constrained optimization problem of the classification hyperplane is transformed into: $\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \zeta_i$, $sty_i(w \cdot x_i + b) \geq 1 - \zeta_i, \zeta_i \geq 0, i = 1, 2, \cdots, n$. Among them, C is the penalty factor, and the larger the C, the greater the penalty for erroneous classification. The Lagrange multiplier method is used to solve this quadratic programming problem with linear constraints, which can be transformed into a dual form and solved by efficient algorithms.

2.2.2. Nonlinear Support Vector Machine

For the nonlinear classification problem, the support vector machine (SVM) solution is to map the input vector x into a high-dimensional feature space by some pre-selected nonlinear mapping, and then construct the optimal classification hyperplane in this space, as shown in Figure 3. The kernel function method avoids complex operations in high-dimensional feature space.

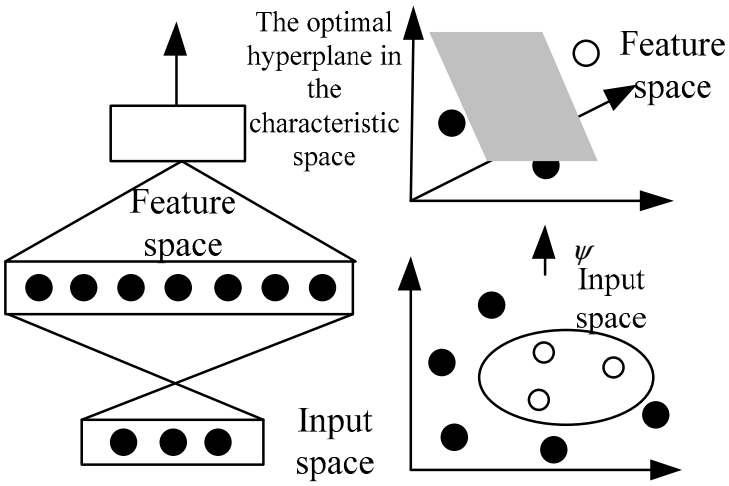

Figure 3. Mapping relationship between input space and high-dimensional feature space.

Figure 3 describes the mapping relationship between the input space and the high-dimensional feature space in detail. The selection of the kernel function $k(x_i, x_j)$ needs to satisfy the Mercer condition. The advantage of SVM is that it only needs to define the inner product operation $k(x_i, x_j) = (\psi(x_i) \square \psi(x_j))$ in the high-dimensional space, and it is unnecessary to know the specific form of the mapping ψ, so as to avoid the "dimension disaster". At present, the main forms of kernel functions are polynomial kernel, multilayer perception (MLP) kernel, Gaussian kernel and so on.

(1) Polynomial kernel function: $K(x, x_i) = [(x \square x_i) + 1]^d$;

(2) Double-level perception kernel function: $K(x, x_i) = S[v(x \square x_i) + c]$;

(3) Gauss kernel function: $K(x, x_i) = \exp\left(\frac{\|x-x_i\|^2}{2\sigma^2}\right)$.

2.2.3. Multi-Class Support Vector Machine

Multi-class support vector machine (MSSVM) is proposed for two-class classification. At present, there are mainly several methods to realize multi-class classification of SVM.

(1) One-to-many method: N support vector machine sub-classifiers are established for N-ary classification problems, and the i SVM sub-classifier is to separate class i data from other data.

(2) One-to-one method: N (N-1)/2 SVMs are established for n-ary classification problems, and one SVM is trained between each two classes to separate the two classes. The multivariate classifier constructed by "one-to-one" method has less training scale, balanced training data and easy to expand.

(3) To improve the objective function directly and establish k classification support vector machine. Because the number of variables is excessive, this method can only be used in solving small problems.

Support Vector Machine (SVM) has been successfully applied in many fields, such as face recognition, handwritten numeral recognition, automatic text classification, multi-dimensional function prediction, etc., and has produced a lot of deformation algorithms.

2.2.4. Multiple Classifier Design for Support Vector Machines

In this paper, the key feature recognition model of network intrusion signal based on multi-class support vector machine algorithm is constructed, as shown in Figure 4.

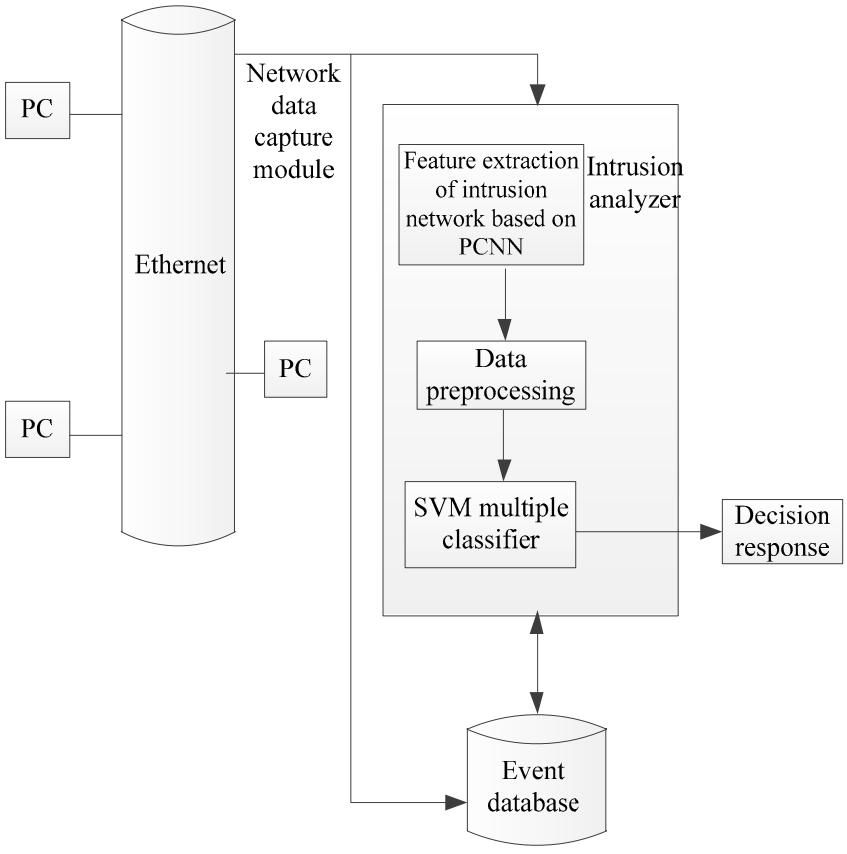

Figure 4. Key feature recognition model of network intrusion signal based on multi-class support vector machine algorithm.

The key feature recognition model of network intrusion signals described in Figure 4 is mainly composed of network data acquisition, principal component neural network extraction of

intrusion signal features, data pre-processing, support vector machine classifier, event database and decision-making response modules, in which SVM classifier is the core part of the whole model.

2.3. Identification of Key Features of Network Intrusion Signals Based on PCNN-SVM Algorithm

In this paper, principal component neural network is used to extract the key features of network intrusion signals, and multi-class support vector machine multi-classifier (SVM) is used to identify the key features of network intrusion signals.

(1) Data pre-processing: Invasive data sets include continuous attributes and discrete attributes; discrete attributes are converted by decimal number, and data sets are then standardized.

(2) Given the training data set, the neural network is trained by GHA learning rules, and its weight is obtained to extract the principal component.

(3) Using principal component neural network projection to validate samples and test samples.

(4) With the transformed training samples and validation samples as input, multi-class support vector machine classifier is used to optimize the parameters, and test samples are used to evaluate the key features of network intrusion signals.

3. Results

3.1. Experimental Data Preparation

To verify the validity of the proposed algorithm, the Defense Advanced Research Projects Agency (DARPA) experimental data is used to test the effectiveness of the proposed algorithm. The DARPA experimental data is the basic data established by the 1999 KDD competition. The data set provides network connection data collected from a simulated typical U.S. Air Force LAN, which is used to evaluate the performance of the algorithm for key feature recognition of network intrusion signals.

According to relevant data, 38 kinds of intrusion signal data in KDD data are classified into four categories: Probing, DOS (Denial of Service), U2R (is a type of network intrusion that is difficult to detect) and R2L. Researchers extracted 41 features from each TCP/IP connection to determine normal and abnormal, including the following three aspects: the basic attributes of TCP/P connections, data domain information and traffic characteristics in a specific time.

This paper selects some data from Training Data and Test Data of the KDD dataset as the experimental dataset, and the data types and distribution are shown in Table 1.

Table 1. Type and number of experimental data.

Category		**Number of Training Sets**		**Number of Test Sets**	
Normal		**2976**		**4000**	
DOS	Back	404	1396	1098	3100
	Neptune	696		1001	
	Smurf	296		1001	
PROBE	Ipsweep	400	1296	306	2293
	Portsweep	484		354	
	Satan	412		1633	
U2R	buffer_ overflow	30	40	52	65
	rootkit	10		13	
R2L	guess_ passwd	53	500	1480	2500
	warezclient	447		1020	

Pre-processing experimental data before classification includes converting symbolic fields into numerical fields and normalizing the data.

3.2. Validity Analysis of SVM Multiple Classifiers

3.2.1. SVM Classifier Settings for DOS and Probe Intrusion

This paper divides the data set into two parts and carries on the experiment separately. Normal, DOS and Probing are one part; Normal, U2R and R2L are another part.

Firstly, we design the classifier of Normal, DOS and Probe mixed data. A 3-class SVM classifier is designed, which adopts a "one-to-one" multi-classifier design strategy and PCNN-SVM algorithm.

Gaussian kernel is selected as kernel function, and kernel parameters C and σ are selected by parameter space search method. Some parameter combinations are selected for training and testing. Figures 5 and 6 show that the actual generalization ability of SVM varies with kernel parameters G when C or σ is fixed.

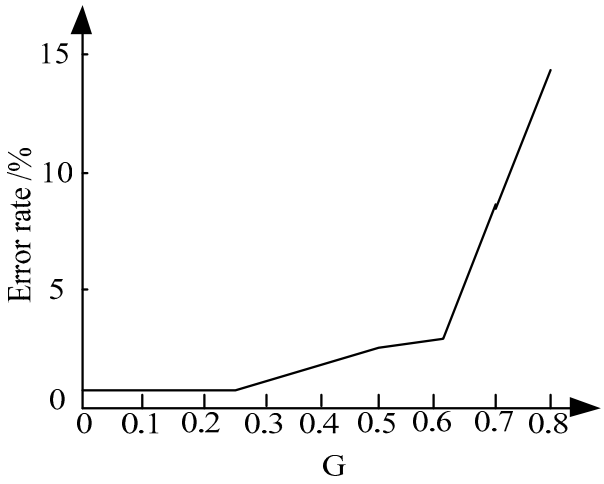

Figure 5. The error rate of SVM multiple classifiers varies with the parameter $G = \frac{1}{\sigma^2}$.

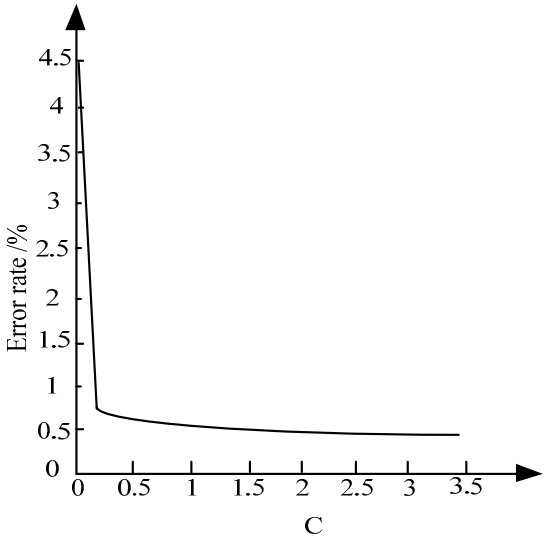

Figure 6. The error rate of SVM multiple classifiers varies with the parameter C.

From Figure 5, we can see that with the increase of parameter G, the error rate of SVM multiple classifiers is on the rise.

From Figure 6, we can see that the generalization ability of SVM increases gradually (the error rate decreases) with the increase of C when σ remains unchanged.

The result of using SVM multiple classifiers to recognize DOS and Probe class data is shown in Table 2.

Table 2. Recognition results of SVM multiple classifiers for Denial of Service (DOS) and Probe data.

Actual Category	Forecast Category			
	Normal	**DOS**	**Probe**	**Correct Rate/%**
Normal	3966	0	36	99.10
DOS	2	3085	13	99.52
Probe	199	3	2091	91.19
Training time			0.625 s	

In Table 2, rows represent actual categories, columns represent predicted categories, such as the number of data that actually belongs to the Normal class in the first row, and the first column represents the number of data recognized by the classifier as Normal class. Therefore, from the first row of the table, 3966 actual Normal data are correctly identified by SVM, 0 Normal data are identified as DOS attacks, 32 Normal data are misidentified as Probing attacks, and the last column of the first row is 99.10%. The correct recognition rate of displaying normal data and the false alarm rate of normal data are 0.90%. From this we can see that the designed SVM has a high recognition rate for DOS and Probing intrusion signals.

3.2.2. SVM Classifier for U2R and R2L Attacks

The same classifier is designed for the mixed data of U2R, R2L and Normal. The method is the same as above. Using the Gauss kernel as the kernel function, the control factor $G = \frac{1}{\delta^2}$ is 0.3 and the penalty factor C is 10000. Multi-class support vector machine for U2R, R2L intrusion signal recognition results are shown in Table 3.

Table 3. Multi-class support vector machine for identification of U2R and R2L intrusion signals.

Actual Category	Forecast Category			
	Normal	**DOS**	**Probe**	**Correct Rate/%**
Normal	3986	12	2	99.65
DOS	18	43	4	66.15
Probe	1438	52	1010	42.40
Training time			0.485 s	

It can be seen from Tables 2 and 3 that the recognition rate of SVM for DOS and Probe attacks is very high, but for U2R, R2L attacks is not ideal.

3.3. Comparative Analysis of Support Vector Machine and BP Neural Network

BP neural network is used to test the same data to study the difference between similar algorithms. The results are shown in Tables 4 and 5, Figures 7 and 8.

Table 4. Recognition results of BP network against DOS and Probe attacks.

Actual Category	Forecast Category			
	Normal	**DOS**	**Probe**	**Correct Rate/%**
Normal	3962	1	37	99.05
DOS	146	2935	19	94.68
Probe	270	3	2020	88.09
Training time			107.69 s	

For the identification of DOS and Robe attacks, after many experiments, BP network adopts the structure of 41-45-35-3 double hidden layer, the hidden layer uses the S-type transfer function with tangent characteristics, the output layer uses the S-type transfer function with logarithmic characteristics, and the conjugate gradient descent algorithm. According to Table 4, the training time of

BP network is 107.69 s, the training error is 9.962×10^{-4}, and the training frequency is 101 times. From Figure 7, we can see that with the increase of training times, the convergence speed of BP network has decreased.

Table 5. Recognition results of BP network against U2R and R2L attacks.

Actual Category	Forecast Category			
	Normal	DOS	Probe	Correct Rate/%
Normal	3990	3	7	99.75
DOS	13	40	12	61.54
Probe	1500	0	1000	40.0
Training time			1022.2 s	

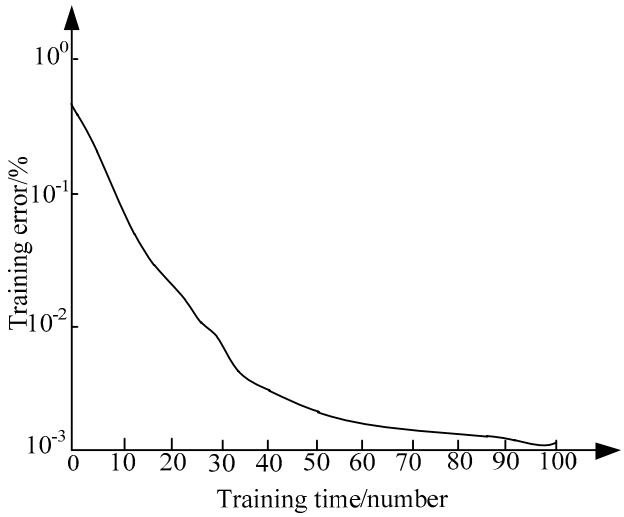

Figure 7. Training error convergence curve of BP network for DOS and Probe attacks.

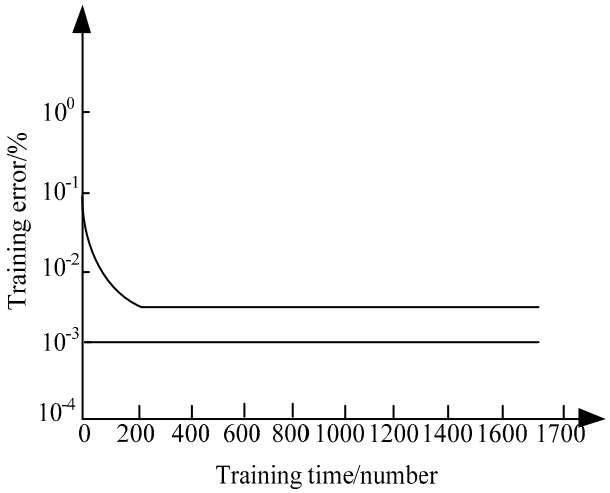

Figure 8. Training convergence curve of BP network for U2R and R2L attacks.

For the identification of U2R and R2L attacks, BP network uses 41-50-40-3 network structure, the hidden layer uses S-type transfer function with tangent property, the output layer uses S-type transfer function with logarithmic property and the conjugate gradient descent algorithm. According to Table 5, the training time of BP network is 1022.2 s, the training error is 2.258×10^{-3}, and the training frequency is 1701 times. It can be seen from Figure 8 that after 200 times of training, the rate of error convergence has slowed down.

3.4. Comparison and Analysis of Key Characteristics of Different Algorithms for Identifying Network Intrusion Signals

In order to highlight the advantages of the algorithm in this paper, the algorithm is used to carry out the experiment again, and the experiment adopts the KDD99 data set. Each sample in the dataset has 41 input variables and one response variable. A total of 5 million samples, the response variables are divided into Normal, DOS, Probing, U2R and R2L five types, it also provides a reduced 10% subset, containing 494021 samples. For the convenience of calculation, this paper only distinguishes each connection as "Normal" or "Attack", which classifies various intrusion methods as "Attack". In this way, the original multi-value classification problem can be changed into a binary classification problem, and the effect of feature extraction can be paid more attention to.

A total of 2000 samples were randomly selected as training sets, 1900 samples as validation sets, and 1900 samples as test sets. The normal samples were 1000, 1000 and 1000 respectively, and the attack samples were 1000, 900 and 900 respectively. The training set is used to train PCNN, the verification set is used to determine the number m of output layers of PCNN and to optimize the parameters of SVM algorithm, and the test set is used to evaluate the performance of intrusion detection algorithm. In order to evaluate the performance of the above methods more accurately, repetitive samples, such as those samples with the same 42 attribute values, are removed during the sampling process. In the process of selecting attack samples, the proportion of probing attacks which are easy to identify in the original samples is reduced, while the proportion of U2R and R2L attacks which are difficult to identify is increased, so, it is more challenging to recognize. Firstly, the feature fields whose attribute values are almost invariant in the sample are removed, and then the attribute values of the remaining fields are pre-processed with standardized data. The C-SVM algorithm is selected to carry out the comparative experiment. The experimental results are shown in Tables 6 and 7. The curve of SVM accuracy varying with PCNN is shown in Figure 9.

Table 6. Confusion matrix based on PCNN-SVM algorithm.

Predicted Value / Actual Value	Normal	Intrusion	Total	Correct Rate/%
Predicted value	987	13	1000	98.7
Attack	36	864	900	96
%	-	-	-	97.42

Table 7. Performance comparison of intrusion recognition algorithms.

Algorithm (Number of Features)	Correct Rate/%	False Alarm Rate/%	Average Recognition Time (ms)
PCNN-SVM (6)	97.42	1.48	0.38
C-SVM (41)	95.16	2.69	0.92

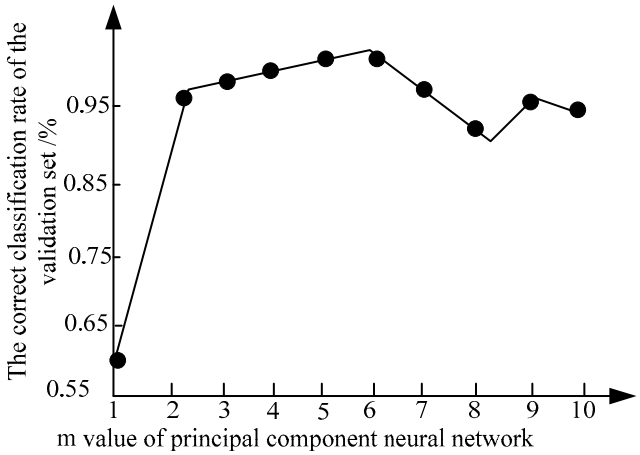

Figure 9. The SVM accuracy of the validation set varies with the m value of PCNN.

Figure 9 shows that the correct classification rate of the verification set is the highest when the number of principal components is six, so the m value of the recognition model is 6. The obfuscation matrix shown in Table 6 is a criterion for measuring the performance of binary classifiers, in which 987 is the number of normal samples correctly classified, 13 is the number of normal samples wrongly classified, 36 is the number of attack samples wrongly classified, and 864 is the number of attack samples correctly classified. Compared with C-SVM algorithm, PCNN feature extraction method can effectively reduce the dimension of input data and reduce the recognition time.

3.4.1. Performance Comparison of Different Feature Selection Algorithms

In order to highlight the advantages of this algorithm, we use RBF algorithm and GA-RBF algorithm to carry out comparative experiments. GA-RBF algorithm and this algorithm are used to optimize the feature selection and parameters of network intrusion signal data. Finally, the feature selection is shown in Table 8.

Table 8. Comparison of feature selection numbers for each algorithm.

Type of Invasion	All Characteristic Numbers	GA-RBF Algorithm Selects Characteristic Number	The Algorithm Selects Characteristic Numbers in This Algorithm.
Probe	41	35	30
DOS	41	31	27
U2R	41	22	18
R2L	41	34	20

From the comparison results in Table 8, we can see that after feature selection, the optimal number of features obtained by the two feature selection algorithms is less than the original number of features, which indicates that there are some redundant features in the network intrusion feature set and useless features in the recognition results, so feature selection must be carried out.

3.4.2. Comparison of Correct Rate, False Positive Rate and Missing Report Rate

The best network intrusion features selected in Table 8 are used to learn the training samples. Finally, the test sets are tested. The comparison of recognition accuracy, false alarm rate and false alarm rate of key feature recognition algorithms for network intrusion signals with different algorithms is shown in Figures 10–12, respectively.

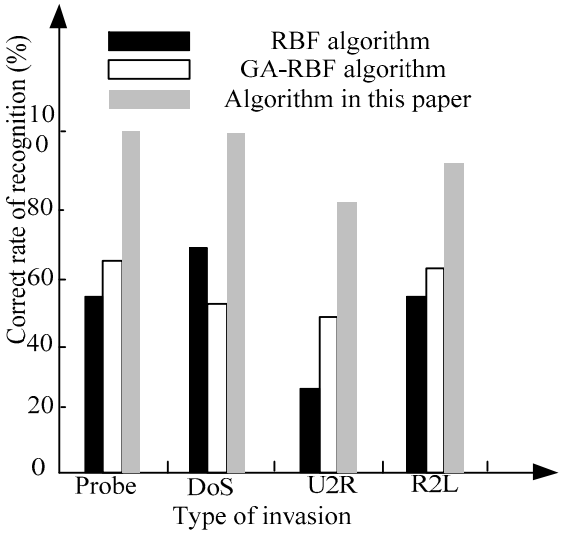

Figure 10. Comparison of recognition accuracy rates of different algorithms.

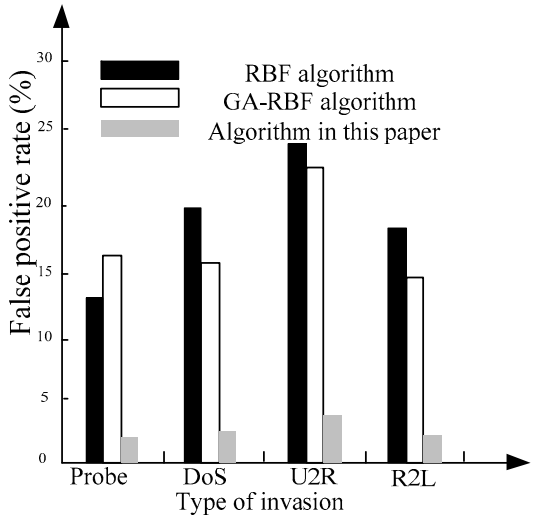

Figure 11. Comparison of false alarm rates of different algorithms.

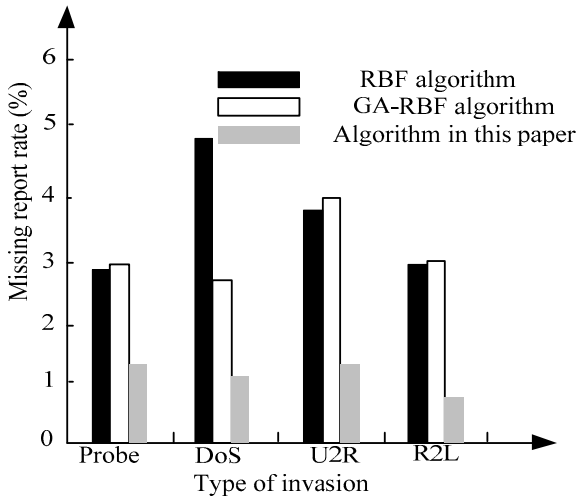

Figure 12. Comparison of missing rate of different algorithms.

In Figure 10, the recognition accuracy of the algorithm is 100% twice, and the other two times are 81% and 90%, respectively; the recognition accuracy of the RBF algorithm is between 28% and 70%; the recognition accuracy of GA-RBF algorithm is between 50% and 68%; the comparison of the three groups of data shows that the recognition accuracy of the key features of the network intrusion signal of this algorithm is higher than that of the same kind. Compared with the algorithm, it has advantages.

In Figure 11, the false alarm rate of this algorithm is less than 4%, which is 14%-16% lower than that of RBF and GA-RBF. The correctness of this algorithm is verified again.

The missing rate of the three algorithms can be seen from Figure 12. Generally, the missing rate of the three algorithms is low. The missing rate of the proposed algorithms is always about 1%, the maximum missing rate of the RBF algorithm is 5%, and the minimum missing rate of the GA-RBF algorithm is 2.8%. Therefore, the proposed algorithm has the advantage of low missing rate in identifying network intrusion. The key features of intrusion signals can be identified in the process of signal detection.

In summary, the recognition index of this algorithm is better than the contrast algorithm.

3.4.3. Comparison of Network Intrusion Recognition Speed

In modern large-scale network system, the efficiency and speed of network intrusion detection are very important. Therefore, based on the training time and recognition time of each algorithm, the recognition speed of each algorithm is calculated. The specific running time is shown in Table 9.

Table 9. Comparison of recognition time for different algorithms.

Type of Invasion	RBF Algorithm/ms	GA-RBF Algorithm/ms	Algorithm in This Paper/ms
Probe	4.14	20.33	17.33
DOS	20.38	11.5	11.02
U2R	4.3	2.98	1.25
R2L	3.99	3.26	3.18

As can be seen from Table 9, the time to detect the key features of network intrusion signals using the proposed algorithm is the shortest.

4. Discussions

4.1. Discussion on Validity of SVM Classifier

When the penalty factor C remains unchanged, G ($G = \frac{1}{\sigma^2}$ as the control factor) is too large or too small, which leads to SVM classifier recognition rate decrease (error rate increases), because G is smaller (σ is larger), the dimension of the feature space is reduced, so that the learning ability (complexity) of SVM is too low and its generalization ability is reduced; when G is larger (σ is smaller). When the dimension of feature space of SVM classifier is too high, the VC dimension of the constructed SVM increases, and the confidence range is too large, which leads to the decrease of the ability to recognize the key features of network intrusion. When σ is unchanged, the generalization ability of SVM increases with the increase of C (error rate decreases). This is mainly due to the gradual reduction of empirical risk, resulting in the result that structural risk tends to minimize. When C is greater than a certain value, the complexity of SVM reaches the maximum allowed in the feature space, so the change of C has little impact on its generalization ability, and the error rate remains at a relatively stable level. We can see that different parameters have a certain effect on classifier performance. After comparing several groups of parameters, the final parameters are as follows: the control factor G is 0.25, the penalty factor C is 10 000. Support Vector Machine (SVM) has high recognition rate for DOS and Probe attacks, but not for U2R and R2L attacks. The main reason is that these two types of network intrusions use system vulnerabilities to gain access to the target host to achieve network intrusion, the main features are concentrated in the content of network packets, while the 41 features selected in the experimental data cannot fully reflect these information, according to the existing characteristics it is difficult to classify the two types of numbers. It is separated from normal data. Therefore, to identify these two types of attacks, it is necessary to provide more detailed content features, or to combine other identification results such as host-based intrusion detection.

4.2. Comparison between Multi-Support Vector Machine Classifier and BP Network

From the experimental results of Tables 4, 5, 7 and 8, it can be seen that the classification effect of the classifier and BP network classifier in this paper is not good because of the inseparability between the network intrusion signals of U2R and R2L datasets and Normal data; for DOS and Robe attacks, SVM is slightly better than BP network in recognition rate. Support vector machines have great advantages in training time.

4.3. Performance Comparison of Different Algorithms

In this paper, the key feature recognition algorithm of network intrusion signal based on neural network and support vector machine is proposed. In fact, it is a method of network intrusion recognition based on feature extraction, which uses principal component neural network to extract the features of network intrusion connection samples and acts as the input of SVM classifier. The algorithm is validated with KDD99 data set. Simulation results show that SVM based on PCNN can get better generalization ability than SVM without feature extraction. The recognition accuracy of RBF algorithm without feature selection is lower than that of GA-RBF and the algorithm presented in this paper, which shows that feature selection can improve the network intrusion detection effect.

The comparison results in Table 9 show that the proposed algorithm can effectively eliminate the adverse effects of useless and redundant features on the algorithm, reduce the input dimension of the classifier, and speed up the network training. At the same time, it shows that the algorithm can obtain better network state features, make network intrusion detection faster, and meet the real-time network identification.

5. Conclusions

This paper combines principal component neural network (PCNN) and multi-class support vector machine (MSSVM) algorithm to identify the key features of network intrusion signals. The feature of this algorithm is that PCNN is used to extract the features of network intrusion connection samples and is used as the input of SVM classifier. Principal component neural network is used to extract the characteristics of network intrusion signals, and multi-class support vector machines are used to build classifiers to accurately identify the key features of network intrusion signals. Experimental results show that the proposed algorithm can effectively identify the key features of network intrusion signals. Compared with similar algorithms, the proposed algorithm has the advantages of high accuracy and low false alarm rate. In addition, through the study of this paper, we can see that the selection of kernel function and parameters of SVM and the selection of neural network structure have a great impact on the generalization ability of the classifier. How to choose the appropriate parameters and network structure is a problem worthy of further study.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, B.; Hu, G.; Zhou, Z. Network intrusion detection based on directed acyclic graph and belief rule base. *ETRI J.* **2017**, *39*, 592–604. [CrossRef]
2. Jia, W.; Zhao, D.; Shen, T. An optimized classification algorithm by BP neural network based on PLS and HCA. *Appl. Intell.* **2015**, *43*, 1–16. [CrossRef]
3. Mansouri, M.; Golsefid, M.T.; Nematbakhsh, N. A hybrid intrusion detection system based on multilayer artificial neural network and intelligent feature selection. *Arch. Med. Res.* **2015**, *44*, 266–272.
4. Arthur, M.P.; Kannan, K. Cross-layer based multiclass intrusion detection system for secure multicast communication of MANET in military networks. *Wirel. Netw.* **2016**, *22*, 1–25. [CrossRef]
5. Xiao, X.; Li, T.; Zhang, R. An immune optimization based real-valued negative selection algorithm. *Appl. Intell.* **2015**, *42*, 289–302. [CrossRef]
6. Benmessahel, I.; Xie, K.; Chellal, M.X. A new evolutionary neural networks based on intrusion detection systems using multiverse optimization. *Appl. Intell.* **2015**, *48*, 2315–2327. [CrossRef]
7. Gao, W.; Baig, A.Q.; Ali, H.; Sajjad, W.; Farahani, M.R. Margin based ontology sparse vector learning algorithm and applied in biology science. *Saudi J. Biol. Sci.* **2017**, *24*, 132–138. [CrossRef]
8. Ge, S.; Chen, X.; Li, D.; Liu, Z.; Ouyang, H.; Peng, W.; Zhang, Z. Hemicellulose structural changes during steam pretreatment and biogradation of lentinus edodes. *Arab. J. Chem.* **2017**, *11*, 771–781. [CrossRef]

9. Yang, Y.; Zhong, M.; Yao, H.; Yu, F.; Fu, X.; Postolache, O. Internet of things for smart ports: technologies and challenges. *IEEE Instrum. Meas. Mag.* **2018**, *21*, 34–43. [CrossRef]
10. Zaidi, K.; Milojevic, M.B.; Rakocevic, V. Host-based intrusion detection for vanets: a statistical approach to rogue node detection. *Trans. Veh. Technol.* **2016**, *65*, 6703–6714. [CrossRef]
11. Hu, S.; Han, J.; Wei, X. A multi-hop heterogeneous cluster-based optimization algorithm for wireless sensor networks. *Wirel. Netw.* **2015**, *21*, 57–65. [CrossRef]
12. Liu, C.; Wang, W.; Wang, M. An efficient instance selection algorithm to reconstruct training set for support vector machine. *Knowl.-Based Syst.* **2017**, *116*, 58–73. [CrossRef]
13. Wieland, M.; Torres, Y.; Pittore, M. Object-based urban structure type pattern recognition from Landsat TM with a Support Vector Machine. *Int. J. Remote Sens.* **2016**, *37*, 4059–4083. [CrossRef]
14. Cheng, C.; Bao, L.; Bao, C. Network intrusion detection with Bat algorithm for synchronization of feature selection and support vector machines. *Comput. Sci.* **2016**, *46*, 69–76.
15. Kalteh, A.M. Enhanced Monthly Precipitation Forecasting Using Artificial Neural Network and Singular Spectrum Analysis Conjunction Models. *INAE Lett.* **2017**, *2*, 73–81. [CrossRef]
16. Supraja, P.; Gayathri, V.M.; Pitchai, R. Optimized neural network for spectrum prediction using genetic algorithm in cognitive radio networks. *Cluster Comput.* **2018**, *65*, 79–84. [CrossRef]
17. Ijjina, E.P.; Chalavadi, K.M. Human action recognition using genetic algorithms and convolutional neural networks. *Pattern Recognit.* **2016**, *59*, 199–212. [CrossRef]
18. Zhang, Z.; Wang, J.; Ren, X. Design of standard parts recycling system based on machine vision. *Autom. Instrum.* **2017**, *76*, 1–9.
19. Sun, Y.F.; Qi, G.-L.; Hu, Y.-L. Deep convolution neural network recognition algorithm based on improved fisher criterion. *J. Beijing Univ. Technol.* **2015**, *41*, 835–841.
20. Hodo, E.; Bellekens, X.; Hamilton, A. Threat analysis of IoT networks Using artificial neural network intrusion detection system. *Tetrahedron Lett.* **2017**, *42*, 6865–6867.

© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).