Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9510858, 11 pages
https://doi.org/10.1155/2021/9510858

Research Article

Network Intrusion Detection Based on Sparse Autoencoder and IGA-BP Network

Hongli Deng and Tao Yang

Education and Information Technology Center, China West Normal University, Nanchong City, Sichuan Province, China

Correspondence should be addressed to Tao Yang; yangt@cwnu.edu.cn

Received 13 April 2021; Revised 20 May 2021; Accepted 21 June 2021; Published 6 July 2021

Academic Editor: Chien-Ming Chen

Network intrusion detection system provides a better network security solution than other traditional network defense technologies. Aiming at the increasingly serious problem of Internet security in the big data environment, a network intrusion detection model based on autoencoder network model and improved genetic algorithm BP (IGA-BP) network is constructed. In order to reduce the data dimension and eliminate redundant information, the autoencoder network model is firstly used to denoise and dedimension. A new population was formed by selecting some of the best parent individuals for cross mutation and replacing the worst parent individuals. The improved genetic algorithm and new population generation model will provide more reasonable initial parameters for BP network, namely, IGA-BP network model. Based on IGA-BP network model, the problems of slow detection rate and easy to get into local optimality in BP network are solved. The experiments were performed on KDD CUP99 dataset, which simulated different types of user organizations and different types of network intrusion. Compared with the existing intrusion detection methods, the experimental results show that the proposed method has a great effect on classification accuracy, false positives, and detection rate.

1. Introduction

With the rapid development of network and its wide application in various fields, the situation of network security is becoming more and more serious. The firewall of early security measures can no longer meet the current network security needs, so how to find out the intrusion behavior by network has become the primary target to prevent network intrusion [1]. Network intrusion detection system is usually set between secure network and insecure network. Abnormal behavior is detected by obtaining and analyzing data information flowing through users or user organizations. When the abnormal behavior is found, the security module is called for effective defense [2, 3]. Network intrusion detection has the advantages of high detection efficiency, flexible use, not occupying normal service resources, etc., making it another effective security measure behind the firewall.

Accuracy and real-time are the necessary requirements of current intrusion detection systems. Only correct identification of normal data and abnormal data will not cause false positives and false positives. Similarly, only by timely processing the information in the network can measures be taken to avoid losses [4]. The network data processed by intrusion detection system usually contains a lot of redundancy and noise [5]. The existence of redundancy and noise features seriously consumes the resources of the computer system, making the detection time of intrusion detection longer, less real-time, and less accurate. Feature dimension reduction method can reduce data dimension and eliminate redundant features. Therefore, in order to carry out intrusion detection accurately and in real time, it is necessary to reduce network characteristics.

Relevant scholars have proposed different data dimension reduction strategies. Paper [6] proposed a new hybrid algorithm, namely, PCA-ANN algorithm. Principal component analysis (PCA) was used to reduce the dimension of input features, and artificial neural network (ANN) was used as the classification model. Experiments show that this method can effectively reduce the training time and test time.

Paper [7] proposed a hybrid dimension reduction intrusion detection method based on information gain (IG) and principal component analysis (PCA). The experimental results show that the hybrid dimension reduction method based on the base class learner integration has more key characteristics, is significantly better than the single method, and achieves higher accuracy and lower false positives. Paper [8] combined PCA dimensionality reduction method with PSO global optimization capability to optimize the weight and threshold of BP neural network. Then, the PCA-PSO-BP intrusion detection model is proposed, which effectively improved the detection accuracy and convergence rate. Paper [9] designed an algorithm combining KPCA and SVM. Kernel principal component analysis (KPCA) was used to reduce the data dimension of the network. Paper [10] proposed an intelligent system based on information gain and correlation, which used the level of information gain and correlation to identify useful and useless features and achieve feature reduction. At present, the network data feature presents complex nonlinear relation, and the above method has good effect on the linear correlation feature. In the face of nonlinear data, it is difficult for high-dimensional data to map effectively to low-dimensional space. These methods cannot eliminate redundancy and noise in network data.

In paper [11], a network security detection architecture based on rough set and back propagation algorithm is proposed. In its structure, rough set preprocesses the intrusion information, which improves the efficiency of detecting the intrusion behavior in the big data network. However, the problems of local minima and long optimization time in these structures have not been solved, resulting in a low detection rate of this model. Paper [12] considered the advantages of rough sets and artificial neural networks and established an improved rough set theoretical algorithm. Combining with artificial neural network, an intelligent fault diagnosis method of wireless sensor network node was constructed. Paper [13] proposed a new map reduce method for data mining and pattern recognition in the big data environment. This algorithm can determine the minimum reduction rough set and realize the parallel genetic algorithm. Its experimental results show that the proposed model can effectively reduce attributes in large decision-making systems. Paper [14] proposed to solve the shortage of data analysis in BP network through optimization. However, due to the lack of effective reduction of redundant attributes, the detection rate and even the correctness of detection will be reduced when the big data of network intrusion is analyzed. Paper [15, 16] proposed intrusion detection system optimized based on particle algorithm and k-nearest neighbor algorithm, which is suitable for dealing with datasets with crossover or overlap in samples. However, it is difficult to achieve accurate classification when the number of different samples varies greatly, and the classification errors of KNN often occur when the number of samples is small. Paper [17] showed the network security detection structure established by support vector machine. In paper [18], the kernel principal component analysis (KPCA) is used to assist support vector machine (SVM) for reducing the dimension of feature vectors. In order to remove the redundancy and noise in the data, an IDS combining deep belief network (DBN) is combined with the feature weighted support vector machine (WSVM) to detect the intrusion [19]. Since the number of SVM is directly proportional to the complexity of SVM calculation, the problem of too many computers due to the dimension is almost nonexistent. However, in the process of matrix storage and calculation, SVM has the problem that the detection efficiency is reduced due to the increased resource consumption when the sample size is large.

In order to eliminate the redundancy and noise in the network data, this paper introduces a sparse autoencoder to reduce the dimensionality of the nonlinear network data and introduces a denoising autoencoder network to improve the robustness of the data after dimensionality reduction. The IGA-BP network is obtained by using the improved genetic algorithm (IGA) to optimize the BP neural network. Finally, the IGA-BP network is used for intrusion detection. This model is intended to solve the problems of information dimension redundancy and BP network local area minimization in network intrusion. Finally, the proposed model of IGA-BP network is compared with other algorithms.

The rest of the paper is organized as follows. The related work is introduced in Section 2. In Section 3, the sparse autoencoder network, BP neural network, and genetic algorithm are introduced. In Section 4, the method of this paper is introduced. Section 5 gives the experimental results. At last, Section 6 draws the conclusion of this paper.

2. Related Work

As early as 40 years ago, machine learning has been applied in the field of network security, such as support vector machine, Bayesian, logistic regression, and other machine learning methods, and has made great achievements. With the development of the information age, large-scale network attacks are complex and diverse. With the development of computer hardware, deep learning algorithm has made great achievements in the field of multimedia. Researchers in the field of network security try to apply deep learning to network intrusion detection. Compared with the traditional machine learning method, deep learning improves the detection accuracy and reduces the false alarm rate. Deep learning can automatically and intelligently identify attack features and help find potential threats.

Literature [20] applied neural network to intrusion detection for the first time; then, deep learning is applied to intrusion detection. The accepted dataset used in intrusion detection is KDD99. This dataset contains 4898431 traffic data; each data contains 41 characteristics such as protocol type and service type and contains 22 kinds of attacks. These attacks can be divided into four categories: denial of service attacks (DoS), remote to local attack (R2L), users to the remote detection attack (U2R), and probing attack (probing). In order to solve the problems existing in KDD99 dataset, literature [21] proposed NSL-KDD based on KDD99 dataset, which deleted some redundant data in KDD99, and its characteristic dimensions and attack types were the same as KDD99 dataset.

There are many feature extraction and classification algorithms for intrusion detection based on deep learning. In literature [22], AE was first used for dimensionality reduction, and then, classification was conducted. Literature [23] used LSTM for network intrusion detection for the first time. The input feature is the original 41 features of the KDD dataset, and the output vector length is 5, including 4 attacks and normal requests. LSTM performs network intrusion detection and parameter selection on KDD99 dataset and obtains a high detection rate. However, the false positive rate of LSTM is also high, reaching 10.04%. Improper selection of LSTM initial weight value may be one of the main factors leading to high false positive rate. Literature [24] proposed applying GRU to intrusion detection in the field of Internet of Things. However, experiments were only carried out on KDD99 dataset, and the accuracy rate was higher than 99%. DBN is applied to intrusion detection as a classification model, which verifies that DBN can be applied to the classification of intrusion detection.

On the private dataset, literature [25] converts each byte of enterprise private traffic into pixels. Thus, the traffic is converted into pictures, and then, the pictures are used as the input of CNN for training and classification, and good results are obtained. Although good results were obtained on private datasets, they were not comparable to results on recognized datasets. Therefore, the convolutional neural network algorithm is applied to the identification dataset in the field of network intrusion detection, and a multiscale convolutional neural network is constructed according to the characteristics of network data.

3. The Basic Theory

In this section, we will introduce the sparse autoencoder network firstly. Secondly, we present BP neural network. The genetic algorithm (GA) will be introduced at last.

3.1. Sparse Autoencoder Network. The autoencoder network (AN) is an unsupervised learning algorithm that does not need to use tag information to obtain data [26]. And it consists of encoder and decoder. The encoder reduces the dimension of the original data, and the decoder reconstructs the reduced data. The learning process of the AN is to reduce the reconstruction error between the reconstructed data and the input data through training and to learn the internal feature representation of the data.

The traditional autoencoder network structure is shown in Figure 1. The original spatial data is R^{m*n}, m is the number of data instances in the original space, and n is the dimension of each instance data. $x_i \in R^n$, $(i = 1, 2 \cdots, m)$, the expressions for encoding and decoding are shown as equation (1) and equation (2).

$$h = S(f(x)) = S(W_x + P), \tag{1}$$

where W is the weight matrix between the input layer and the hidden layer and P is the offset of the hidden layer neurons.

$$\hat{x} = S(g(h)) = S\left(W^T h + q\right), \tag{2}$$

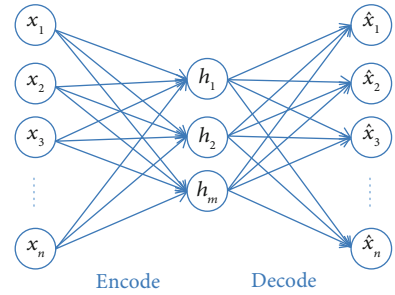

Figure 1: The structure of autoencoder network.

where W^T is the weight matrix between the hidden layer and the output layer and q is the offset of the output layer neurons.

$$G(S) = \frac{1}{1 + e^{-S}}. \tag{3}$$

Equation (3) is the sigmoid activation function. The goal of learning autoencoder network is to minimize the value of reconstruction error L, that is, to make the input value and the output value as close as possible. And the error function L is to select the mean square error loss function shown as

$$L(x, \hat{x}) = \frac{1}{m} \sum_{m}^{1} (x\wedge - x)^2. \tag{4}$$

If there are no constraints, the autoencoder network is easy to output direct copy input, which means that otherwise useless information is added to the dimensionality reduction feature. The purpose of the autoencoder network training is to reduce the reconstruction error, so reducing the dimension of the target feature is meaningless if the output is a direct copy of the input information. The sparse constraint can automatically remove the unnecessary information in the dimension reduction process. Therefore, in order to prevent the replication of input information, regularization correction can be added after the error function to obtain the regular autoencoder network. Then, the sparse autoencoder network can be obtained as follows.

$$J_s(x, \hat{x}) = L(x, \hat{x}) + \beta \sum_{j}^{m} KL(\rho || \widehat{\rho}), \tag{5}$$

$$KL\left(\rho || \widehat{\rho}_j\right) = \rho \log \frac{\rho}{\widehat{\rho}_j} + (1 - p) \log \frac{1 - \rho}{1 - \widehat{\rho}_j}, \tag{6}$$

where $KL(\rho || \widehat{\rho}_j)$ is the sparse penalty term, β is the weight of the sparse penalty term, ρ is the sparsity parameter, and $\widehat{\rho}_j$ is the average activation value of the jth hidden layer.

Sparse autoencoder network is prone to overfitting. In model fitting, the coefficient of fitting function is usually large, which leads to the sharp jitter of fitting curve and makes the absolute value of derivative larger in some intervals. Regularization can reduce the coefficient value of the

fitting function by constraining the coefficients in the model, thus making the fitting curve more stable and alleviating the overfitting problem. The norm penalty L regularization method can be added to the error function to prevent overfitting, which is shown as follows.

$$J(x, \widehat{x}) = L(x, \widehat{x}) + \beta \sum_{j}^{m} KL\left(\rho || \widehat{\rho}_j\right) + \frac{\lambda}{n} \sum_{\omega} |\omega|, \quad (7)$$

where λ is the penalty factor, ω is the weight, and n is the number of training set samples.

The basic design idea of the sparse autoencoder network is as follows. (1) For a given tableless data, unsupervised learning is used to learn characteristics. For data without class labels, the input data is encoded through an encoder, and then, an output information is obtained using a decoder. If the output is approximately equal to the input data, the reconstruction error can be minimized by adjusting the parameters in the encoding and decoding stages. (2) The characteristic generated by the encoder is used as input; the network of the lower layer is trained layer by layer. Training with the first layer of code can be seen as a repetition of input data, so the following training process similar to the first layer. Supervised to fine-tune the neural network. After training all layers of network, study obtained since the encoder can better represent the characteristics of the input, and these characteristics can be optimally said the original input signal.

3.2. BP Neural Network. The BP network is a neural network algorithm based on multilayer backward learning [27]. The basic principle is the steepest descent method based on the optimization theory. BP neural network repeats the search until the algorithm finds the minimum error function value and its position in a certain region. The purpose of the BP network algorithm is to use the output possible error and back propagation multiple adjustments to obtain the optimal weights and thresholds during the training process and finally obtain the best derivation results. BP neural network has the characteristics of simple structure and adapting to various training algorithms and easy to implement. For a long time, BP network algorithm has not only applied in intrusion detection but also image processing. The x and y in the three-layer BP network structure with one hidden layer are input and output, respectively. And the values between the layers are, respectively, the weight and threshold of the adjustment error. When a dataset contains N samples, the error function L is shown as

$$L = \sum_{n=1}^{N} (t_n - y_n)^2, \quad (8)$$

where t_n is the category vector and y_n is the output value obtained when the BP network input is x_n.

BP uses the iterative learning method to find the optimal weight and threshold. But in the initial training of the BP network normalized between [0,1], the training function will generate a random value between 0 and 1. The random value will become the weight and threshold of the first training of the BP network. The first use of random values will lead to the instability of BP network, and the results of operation will vary greatly. Another problem is that the convergence rate is too slow to guarantee the global minimum value of convergence.

3.3. Genetic Algorithm. The genetic algorithm (GA) is an evolutionary algorithm whose principle is to mimic the survival law of the survival of the fittest in the process of evolution. Its essence is an efficient, parallel, and global search method. In the search process, the hidden knowledge of search space is automatically acquired and accumulated, and then, the search process is adaptively controlled to obtain the global optimal solution [28].

Genetic algorithm is a random global search method developed by imitating the biological evolution mechanism in nature. It borrows from Darwin's theory of evolution and Mendel's theory of heredity. Its essence is an efficient, parallel, and global search method. It can automatically acquire and accumulate the tacit knowledge of search space in the process of searching. And the global optimal solution can be obtained by controlling the search process adaptively. Genetic algorithm starts from the initial population and carries out selective evolutionary operation, crossover, and mutation genetic operation according to the fitness value of each individual. This process leads to the evolution of the individuals in the original population, giving rise to new populations. In this way, generation after generation, until it converges to a group of individuals with the best fitness value and finds the optimal solution.

The GA algorithm does not have strict requirements for initial conditions during the operation. It encodes the data and evaluates it using the fitness function. It exchanges the information of chromosomes through iterative methods of multiple selection, crossover, and mutation and finally chooses to produce the optimal new population. The genetic algorithm not only overcomes the shortcomings of traditional evolutionary algorithms that can only deal with a single body but also has the advantages of global optimization that are not available in traditional evolutionary algorithms. The BP network uses the initial value provided by the GA algorithm to improve the new population generation method to better solve the BP network local minimum problem.

4. Method

In this section, the algorithm framework will be introduced firstly. Then, we present preprocessing data and the improved sparse autoencoder to reduce the dimension. Finally, an improved BP network based on the improved new population generation algorithm, namely, IGA-BP network, is proposed.

4.1. Algorithm Framework. The architecture of network intrusion detection based on sparse autoencoder and BP network is shown in Figure 2. The specific process of the intrusion detection framework is as follows.

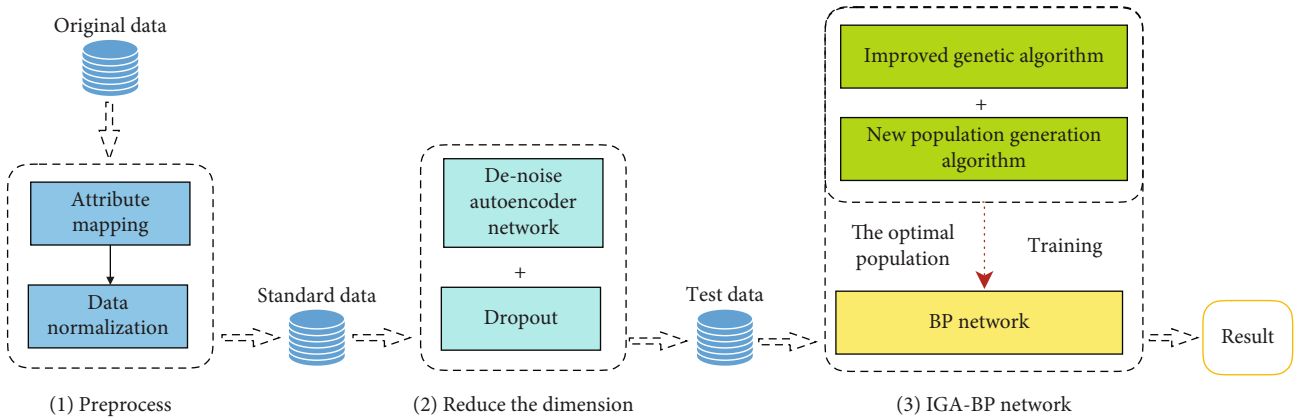

Figure 2: The structure of intrusion detection.

(1) *Preprocessing Data*. Attribute mapping, which converts character network data features into numeric data. Due to the great difference in the data of attributes of the same type, the training effect is affected, so the data is normalized into an interval [0,1]

(2) *Reduce the Dimension*. In order for the encoder to learn better features, noise is added during the training phase to train the network. At the same time, in order to reduce the phenomenon of data overfitting, the dropout method is used for training during the training process

(3) *IGA-BP Network*. Firstly, the genetic algorithm is improved, and a new population generation algorithm is obtained. Then, the improved new generated population is applied to the training of BP network; then, the IGA-BP network is obtained. Input the prediction data into the trained IGA-BP network to obtain the prediction result of each prediction data

4.2. Preprocessing Data. Network intrusion data have the characteristics of large data volume and many attributes, and the value size of the same attribute often has 10^7 or more differences. If these data are directly analyzed and trained during the test, it will result in unclear data attributes, abnormal convergence, and even failure to obtain correct training results. After data normalization, the attribute value of each attribute can be controlled in a fixed region, which is convenient for further analysis and processing of data. Therefore, data normalization speeds up the convergence of the analysis program. The normalized value range of the data is the interval [0,1], and the specific normalization method is shown in the following equation.

$$v'_i = \frac{v_i - \min_a}{\max_a - \min_a}, \tag{9}$$

where $\max_a$ and $\min_a$ are the maximum and minimum values in attribute a, respectively. v'_i is the eigenvalue of normalized attribute v_i.

4.3. Reduce the Dimension. In order to make the encoder learn the features better, noise is added in the training phase to train the network. At the same time, in order to reduce the phenomenon of data overfitting, dropout method was used in the training process. Then, the improved sparse autoencoder algorithm is used to reduce the dimension of the data.

4.3.1. Denoise Autoencoder Network. Characteristic of the current network data is complex nonlinear relations. In this paper, the dimensionality reduction of the nonlinear network data is carried out through the autoencoder network, and the denoising autoencoder network is introduced to improve the robustness of the data after dimensionality reduction. Assuming that the characteristics learned from the coding network are highly representative, the original data can be effectively reconstructed even if the input data is corrupted. On the basis of this assumption, a noise reduction decoder network can be proposed; that is, the network carries on certain damage to the input data for training. Since the input data contains noise data, the autoencoder network with explicit denoising can make the learned feature data more robust, so this advantage is used to train the conventional autoencoder network.

The structure of denoise autoencoder network is shown in Figure 3. x is the original data instance, and qD is the random mapping function. x is converted to $\tilde{x}$ using a random mapping function that randomly adds noise data to the original data x. Then, the encoded data y and decode data $\hat{x}$ can be obtained as follows.

$$y = S(W\tilde{x} + p), \tag{10}$$

$$\hat{x} = S\left(W^T h + q\right). \tag{11}$$

Finally, the reconstruction error J is obtained by x and $\hat{x}$. The gradient descent algorithm is used to train and reduce the reconstruction error to restore the original data x as much as possible.

4.3.2. Dropout. Dropout is a method to reduce overfitting of data. In this paper, dropout is applied to train deep neural

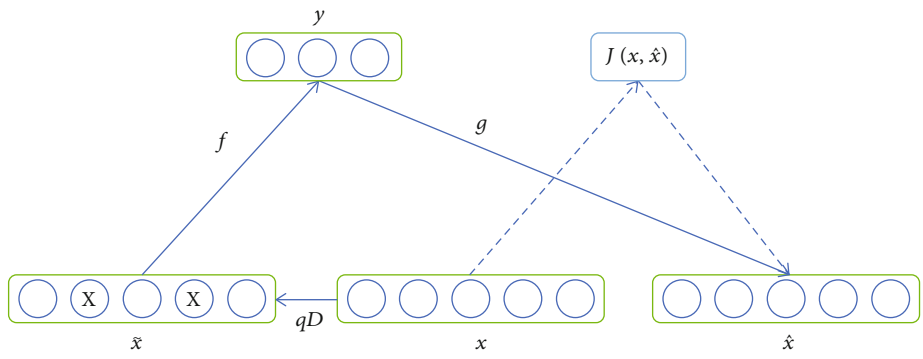

Figure 3: Denoise autoencoder network.

networks to avoid the repetition of extracted features due to the mutual adaptation of hidden layer neurons.

In the implementation of dropout, the output of hidden layer neurons is randomly set to zero in a certain proportion, so that some neurons do not participate in the training process of forward propagation. However, dropout is not a simple zeroing operation. In the training process and the test process, the forward propagation algorithm is different. During the training phase, the zeroed neurons do not participate in the forward propagation and do not contribute to the back propagation algorithm. However, their weights are preserved. In the test phase, the idea of mean network was used. Although all neurons (including the zeros) were involved in the forward propagation, the output of the neurons was attenuated in proportion to the dropout to maintain the equilibrium of the whole test network. In the BP network training process, this paper will use dropout method to improve the ability of feature extraction and classification.

4.4. IGA-BP Network. A new population generation algorithm can be obtained by improving genetic algorithm. In this way, better weights and thresholds can be trained and become parameters of the first simulation experiment of BP network, namely, IGA-BP network. Compared with BP neural network, IGA-BP network can better solve the problems of slow optimization speed.

4.4.1. Improved Genetic Algorithm. Genetic algorithm is an evolutionary algorithm whose principle is to mimic the survival law of the survival of the fittest in the process of evolution. The GA algorithm does not have strict requirements for initial conditions during the operation. It encodes the data and evaluates it using the fitness function. It exchanges the information of the chromosomes through iterative methods of multiple selection, crossover, and mutation and finally chooses to produce the optimal new population. The genetic algorithm not only overcomes the shortcomings of traditional evolutionary algorithms that can only deal with a single body but also has the advantages of global optimization that are not available in traditional evolutionary algorithms. The optimal population can be obtained by using GA algorithm which improves the generation method of new population. The optimal weights and thresholds can be obtained by using the optimal population training BP network, which can solve the local minimum problem of BP network.

(1) *Improved Selection Algorithm.* The fitness is used to judge the quality of the individual in the GA algorithm. Scroll selection method is selected; that is, the smaller the individual fitness, the greater the probability of selection. The greater the probability of being selected, the more its gene will expand in the population, and vice versa, it may be eliminated. The fitness is related to the error of the test results. In this paper, the absolute value of the final test result is used by the BP network to determine the fitness of the individual. The individual fitness function and the selection function are as shown in equations (12) and (13), respectively

$$F = \sum_{t=1}^{m} |y_t - o_t|, \tag{12}$$

where m is the number of output nodes of the BP neural network. y_t is the expected result of the first node of the BP neural network. o_t is the possible result of the tth node.

$$P_k = \frac{1/F_k}{\sum_{K=1}^{M} 1/F_k}, \tag{13}$$

where F_k is the fitness of the individual k of the population and M is the total number of all individuals in the population.

When selecting individuals by fitness in the improved algorithm, partial optimal parent individuals are selected according to fitness. In order to conduct crossover operation, the number of selected individuals must be even. Through the test, it is concluded that the parent with a ratio of 0.95 is the best choice for the next step.

(2) *Crossover Operation.* The crossover algorithm is to let the offspring produce a new individual with both parental cross-individual features. If the new individual obtains the optimal characteristics of the father when crossing, the new individual generated will be

better than the individual before the intersection, and the cross will help the new population to evolve. There are single point, two points, and multiple points in the cross mode, and the single point crossover algorithm is used in this paper. The principle is to randomly select two individuals as cross objects and then randomly generate intersections. The two individuals exchange some genes at the intersection, thus producing two different individuals than before the intersection. The crossover operation usually adopts the same algorithm as the individual coding method. The crossover operation result of the two chromosomes R and R' at the k position is equal to the uncrossed value plus the intersection value of the opposite party, and the crossover operation is shown as following equation

$$R_k = R_k(1 - S) + R'_k S, \tag{14}$$

$$R'_k = R_k(1 - S) + R_k S, \tag{15}$$

where R_k and R'_k are the crossover values of the two chromosomes R and R' at the k position, respectively. S is a random number, and $S \in [0, 1]$.

(3) *Mutation Operation*. The principle of mutation operation is to generate a new individual through gene mutation. If an inferior individual is produced, the individual will be eliminated after the selection operation. However, if a better individual is produced, it will produce more progeny individuals after the selection operation, so that the individual will occupy the dominant address in the population. In order to avoid premature convergence, the commonly used methods are basic bit variation or uniform variation. In this paper, the variation of the basic position is adopted. The individual population consisting of binary genes is flipped with a small probability of the gene; that is, 0 and 1 are mutually variable. The randomly selected mutation position in this paper is greater than 0.5, and the variation of gene R is selected as

$$R = R + \left(R - R'\right) * r\left(1 - \frac{g}{g_1}\right)^2, \tag{16}$$

where R' is the upper bound of the gene. In order to prevent the degradation of the genetic algorithm, a small probability variation is usually adopted. The mutation probability $r \in [0.001,0.1]$ and $r = 0.01$ is set in this paper. g is the current number of iterations, $g \in [1,200]$. g_1 is the maximum number of iterations, and $g_1 = 50$ is set in this paper.

4.4.2. Improved Population Generation Algorithm. Although the classical genetic algorithm solves the shortcomings of its local optimum by optimizing the initial value of the BP network, there is still a problem of poor improvement. The problem is that the population generated by each iteration cannot be guaranteed to be better than the population of the original parent. Aiming at the shortcomings of classical genetic algorithm, an improved genetic algorithm is proposed, which is aimed at the improvement of new population generation process. The description process of the improved new population generation algorithm is as follows.

(1) The initial population, objective function, and fitness were calculated, and the parent generation was generated

(2) The optimal individuals with a ratio of 0.95 in the parent generation were selected and treated with crossover and mutation

(3) On the premise of retaining the best individuals in the parent population in step 1, the reinsertion function is used to replace the individuals with the worst fitness in the original parent population, so as to form the current optimal new population

(4) Perform multiple iterations to arrive at the optimal new population

In the improved new population generation algorithm, optimization is performed by replacing the individuals with the least fitness in the parent. In the improved new population generation algorithm, it is possible to generate a population superior to the original parent by iteration each time. After the algorithm iteration, a better population can be generated than the classical genetic algorithm without the improved new population generation method. Thus, better weights and thresholds can be trained and become parameters of the first simulation test of BP network. These initial parameters are used by BP network for back propagation learning adjustment. In this way, the problems of slow

Table 1: The definition of parameters.

Definition	The definition of parameters
T_P	True positive, anomaly instances correctly classified as an anomaly
F_P	False positive, normal instances wrongly classified as an anomaly
T_N	True negative, normal instances correctly classified as normal
F_N	False negative, anomaly instances wrongly classified as normal

Table 2: Results of independent intrusion experiment.

Intrusion type	Normal	Intrusion	AC%	FA%
Back	2800	200	95.59	1.13
guess_passwd	2800	00	95.83	1.15
ipsweep	2800	200	98.49	1.14
Neptune	2800	200	99.53	1.04
Smurf	2800	200	95.68	1.09
portsweep	2800	200	99.72	1.01

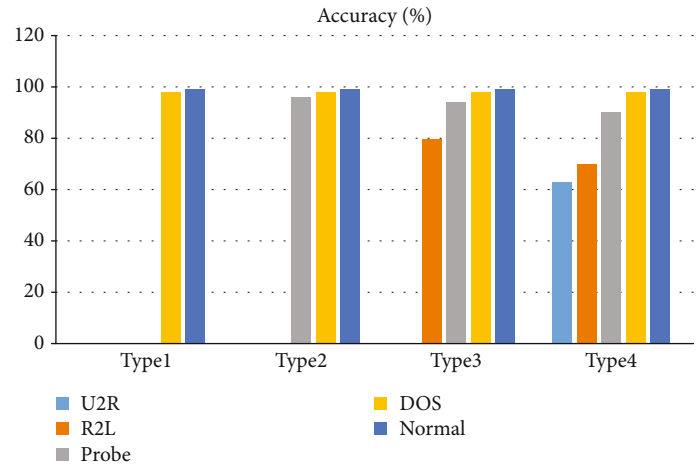

Figure 4: Accuracy of combined type intrusions.

optimization speed and local optimization are better solved, and the efficiency of data analysis is improved.

In order to get the best parameters, IGA-BP network needs to be tested many times. After many tests, the optimal parameters are as follows: the crossover probability of genetic algorithm is 0.7, the mutation probability is 0.01, the training target of BP neural network is 0.01, and the learning rate is 0.1.

5. Experiment

The experiment was carried on the laboratory sever which is equipped with Intel Core i7 CPU (8 GHz), 16 G RAM memory, and Windows 10 operating system. The experiment source code is developed using Python 3.5.

5.1. Dataset. The KDD CUP99 dataset is the general standard dataset for current intrusion detection experiments [29]. The KDD CUP99 dataset is derived from an intrusion detection assessment project at Lincoln Laboratories. It simulates a network environment in the Air Force LAN that simulates a variety of different types of users and types of cyber intrusions, which makes it like a real network environment. It is a collection of 9 weeks of simulated raw TCPdump (∗) data on a LAN. Training data is obtained from 7 weeks of network traffic, with approximately 5 million connection records and approximately 2 million connection records for the last two weeks. There are 4 types of intrusions in the data, which are divided into 39 subcategories. There are 22 types of training data. The new 17 intrusions are additional intrusions in the test dataset but not in the training dataset. Each instance data in the dataset contains 41 feature attributes and a tag attribute. The tag attributes are divided into 5 categories, namely, normal, DoS, probe, R2L, and U2R.

5.2. Evaluation Index. In the comparison experiment, Accuracy (AC), False Rate (FA), and Recall (RE) [30] were used as the evaluation criteria for the merits and demerits of this experiment.

$$\mathrm{AC} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}, \tag{17}$$

$$\mathrm{FA} = \frac{F_N}{T_P + F_P}, \tag{18}$$

$$\mathrm{RE} = \frac{T_P}{T_P + F_N}. \tag{19}$$

The specific parameters are shown in Table 1.

5.3. Intrusion Analysis

5.3.1. Independent Intrusion Experiment. Six of the most common types of intrusion were selected from the dataset as experimental subjects for individual intrusion detection. Each type of data selected 3000 pieces of data for testing (2800 for normal types and 200 for intrusion types). The experimental results are shown in Table 2.

It can be seen from Table 2 that the proposed method has a good recognition effect for common intrusion types. Therefore, the proposed method can effectively identify common independent intrusions.

5.3.2. Combined Intrusion Experiment. Because the types of intrusions in network data are usually complex, different intrusion combinations are set to test the effectiveness of the proposed algorithm for complex network intrusions. The intrusion detection experiments were carried out on the four types of intrusion type data and the normal type data in the dataset, respectively. The detection accuracy of different intrusion combination types is shown in Figure 4. The horizontal axis in the figure represents the data of different intrusion combinations, and the vertical axis represents the accuracy. The types of intrusions included in the intrusion combination are shown in Table 3.

Table 3 gives information on the types of intrusions included in the intrusion combination in the experiment. The "1" indicates that the intrusion combination contains the intrusion type of the class, and the "0" indicates that the intrusion combination does not contain the intrusion.

It can be seen from the experimental results that the detection rate of normal and abnormal is higher when only normal and abnormal data are included. When it contains 3 types of data, it has a good detection effect on the normal

Table 3: The combination of various intrusion.

Type	DoS	Probe	R2L	U2R
Type 1	0	0	0	1
Type 2	0	0	1	1
Type 3	0	1	1	1
Type 4	1	1	1	1

Table 4: Details of five groups of data.

Group	Normal	Intrusion	Total
G1	10380	9719	20099
G2	12334	13521	25855
G3	9413	11819	21232
G4	11752	12653	24405
G5	14742	10413	25155

Table 5: The comparison result of AC (%).

Methods	G1	G2	G3	G4	G5
BPNN [31]	94.5	92.7	95.12	95.57	93.35
PCA-PSO-BP [8]	90.6	89.5	90.12	91.27	90.44
PCA-ANN [6]	94.35	92.7	95.12	95.57	93.35
KPCA-SVM [9]	95.4	94.1	95.8	96.45	95.32
RNN [32]	97.1	95.6	97.21	97.1	96.51
DBN [33]	97.31	95.01	97.12	97.57	96.65
CNN-WDLSTM [34]	97.52	96.04	97.45	97.71	97.10
Proposed	97.68	97.26	97.72	98.27	97.25

type and the DoS intrusion type. When it contains 4 types of data, it has a good detection effect on the data of normal type, DoS type, and probe type. Due to the small amount of training data for U2R and R2L types of intrusions, the training is insufficient, so the accuracy of the obtained test results is slightly lower among the five data types.

5.4. Comparative Analysis with Existing Methods. In order to verify the effectiveness of the proposed algorithm, five groups of data were randomly selected from the dataset for experimental verification. The experimental results were compared with the traditional dimensionality reduction algorithm and the existing methods. Five groups of data are shown in Table 4.

5.4.1. Accuracy, False Positive Rate, and Time Analysis. The effectiveness of the proposed algorithm was tested by comparing the accuracy of the five sets of data (AC/%), false positive rate (FA/%), and test time (Te/s).

The proposed algorithm in this paper is compared with other seven algorithms. The AC, FA, and Te of these algorithms on five sets of data are compared, as shown in Tables 5–7. As shown in Table 5, AC of the proposed algorithm is the largest, which indicates that the accuracy of the proposed algorithm is higher. As shown in Table 6, the FA of the proposed algorithm is the minimum, which means that the false rate of the proposed algorithm is the lowest. As shown in Table 7, the Te of the proposed algorithm is the minimum, which means that test time of the proposed algorithm is the lowest. As a result, the proposed algorithm is superior to other used intrusion detection algorithms in terms of AC, FA, and Te (test time).

Table 6: The comparison result of FA (%).

Methods	G1	G2	G3	G4	G5
BPNN [31]	6.2	5.7	5.21	4.87	5.05
PCA-PSO-BP [8]	5.6	5.5	5.12	5.27	5.2
PCA-ANN [6]	9.35	8.7	8.9	8.57	8.35
KPCA-SVM [9]	8.1	7.9	8.2	7.45	6.8
RNN [32]	2.4	1.6	1.71	1.65	1.53
DBN [33]	3.31	3.1	3.22	2.9	2.85
CNN-WDLSTM [34]	2.7	1.82	1.53	1.44	1.19
Proposed	1.03	0.98	1.01	0.87	0.75

Table 7: The comparison result of Te (s).

Methods	G1	G2	G3	G4	G5
BPNN [31]	27.5	28.71	27.12	28.7	26.98
PCA-PSO-BP [8]	12.61	13.52	13.12	12.57	13.44
PCA-ANN [6]	16.52	17.07	16.13	17.53	18.32
KPCA-SVM [9]	11.24	12.1	11.48	12.46	11.37
RNN [32]	12.63	13.62	12.23	13.21	12.51
DBN [33]	11.34	12.01	11.42	11.56	11.67
CNN-WDLSTM [34]	9.32	7.65	6.73	4.22	5.72
Proposed	3.31	3.26	3.78	2.21	2.97

Table 8: The accuracy result of various intrusion types (%).

Methods	Normal	DoS	Probe	R2L	U2R
BPNN [31]	96.24	97.51	91.18	23.22	64.28
PCA-PSO-BP [8]	97.21	98.52	92.14	26.52	65.44
PCA-ANN [6]	96.52	97.07	91.13	27.63	68.63
KPCA-SVM [9]	98.56	98.51	93.11	26.72	71.28
RNN [32]	98.73	98.69	93.26	28.71	73.51
DBN [33]	99.03	98.71	94.22	32.56	75.97
CNN-WDLSTM [34]	99.12	98.97	99.10	54.25	78.64
Proposed	99.51	99.27	99.38	61.21	81.94

5.4.2. Comparison of Accuracy of Various Types of Data. To further verify the effectiveness of the proposed algorithm, five different types of data, normal, DoS, probe, U2R and R2L, were compared and analyzed. The experimental results are shown in Table 8. The proposed algorithm has a high detection rate for normal and DoS type data. Because the training data of U2R and R2L types of intrusion data is less and there are more unknown intrusions, the detection accuracy of proposed algorithm is slightly lower, but its detection effect is better than other classifier algorithms.

6. Conclusion and Future Work

Compared with the past, the current Internet data is expanding every day, and the data is created rapidly from ZB to PB. So the data is bigger, more complex, and more dimensional than ever before. In this case, the traditional network intrusion detection methods can not meet the requirements of real-time and accuracy. To solve this problem, this paper proposed a network intrusion detection algorithm based on autoencoder network model and IGA-BP model. Firstly, the autoencoder network model is used to denoise the network data and reduce the data dimension. Then, the population generation algorithm of GA model is improved, and the improved genetic algorithm which improves the generation of new population will provide more reasonable initial parameters for BP network. Finally, IGA-BP network model is used for intrusion detection of network data. Experiments were performed on KDD CUP99 dataset which simulated different types of user organizations and different types of network intrusion. The experimental result shows that the false positives and false positives of the proposed method are better than other intrusion detection methods. And the proposed method is applicable to the current high dimensional and complex network data and provides a new idea for the current network intrusion detection research.

The study of obtaining the optimal parameters by automatic learning is one of the goals of future work. And the detection effect of proposed algorithm on other network intrusion needs further testing.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

This work is supported by the Sichuan Science and Technology Program (No. 2020JDRC0075).

References

[1] B. Selvakumar and K. Muneeswaran, "Firefly algorithm based feature selection for network intrusion detection," *Computers & Security*, vol. 81, pp. 148–155, 2019.

[2] J. Fontaine, C. Kappler, A. Shahid, and E. De Poorter, "Log-Based Intrusion Detection for Cloud Web Applications Using Machine Learning," in *International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, pp. 197–210, Springer, Cham, 2019.

[3] R. Pokhrel, P. Pokharel, and A. Kumar Timalsina, "Anomaly-based intrusion detection system using user profile generated from system logs," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, no. 2, 2019.

[4] Y. Chen, X. Q. Cheng, Y. Li, and L. Dai, "Lightweight intrusion detection system based on feature selection," *Ruan Jian Xue Bao (Journal of Software)*, vol. 18, no. 7, pp. 1639–1651, 2007.

[5] Y. Xiao, C. Xing, T. Zhang, and Z. Zhao, "An intrusion detection model based on feature reduction and convolutional neural networks," *IEEE Access*, vol. 7, pp. 42210–42219, 2019.

[6] S. Lakhina, S. Joseph, and B. Verma, *Feature Reduction Using Principal Component Analysis for Effective Anomaly Based Intrusion Detection on NSL-KDD*, International Journal of Engineering Science and Technology, 2010.

[7] F. Salo, A. B. Nassif, and A. Essex, "Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection," *Computer Networks*, vol. 148, pp. 164–175, 2019.

[8] L. Shanshan, X. Xiaoyao, and J. Fengyi, "Research on network intrusion detection based on PCA PSO-BP," *Application Research of Computers*, vol. 33, no. 9, pp. 2795–2798, 2016.

[9] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," *Applied Soft Computing*, vol. 18, pp. 178–184, 2014.

[10] I. Manzoor and N. Kumar, "A feature reduced intrusion detection system using ANN classifier," *Expert Systems with Applications*, vol. 88, pp. 249–257, 2017.

[11] H. H. Inbarani, M. Bagyamathi, and A. T. Azar, "A novel hybrid feature selection method based on rough set and improved harmony search," *Neural Computing and Applications*, vol. 26, no. 8, pp. 1859–1880, 2015.

[12] X. ZHOU and S. XUE, "WSN fault diagnosis with improved rough set and neural network," *Computer Science*, vol. S2, 2016.

[13] E. S. M. El-Alfy and M. A. Alshammari, "Towards scalable rough set based attribute subset selection for intrusion detection using parallel genetic algorithm in MapReduce," *Simulation Modelling Practice and Theory*, vol. 64, pp. 18–29, 2016.

[14] L. Bowen, "Application of improved BP neural network in intrusion detection," *China Computer & Communication*, vol. 2017, no. 14, p. 28, 2017.

[15] A. A. Aburomman and M. B. I. Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," *Applied Soft Computing*, vol. 38, pp. 360–372, 2016.

[16] P. Kuttranont, K. Boonprakob, C. Phaudphut et al., "Parallel KNN and neighborhood classification implementations on GPU for network intrusion detection," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 9, no. 2-2, pp. 29–33, 2017.

[17] J. Peng, Y. Zhou, and C. L. P. Chen, "Region-kernel-based support vector machines for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 4810–4824, 2015.

[18] K. S. Sahoo, B. K. Tripathy, K. Naik et al., "An evolutionary SVM model for DDOS attack detection in software defined networks," *IEEE Access*, vol. 8, pp. 132502–132513, 2020.

[19] Y. Wu, W. W. Lee, Z. Xu, and M. Ni, "Large-scale and robust intrusion detection model combining improved deep belief network with feature-weighted SVM," *IEEE Access*, vol. 8, pp. 98600–98611, 2020.

[20] D. Debar HBecker MSiboni, "A neural network component for an intrusion detection system," in *IEEE Symposium on Security and Privacy*, p. 240, Oakland, CA, USA, 1992.

[21] Cnadian Institute for Cybersecurity, "NSL-KDD dataset [EB/OL]," 2017, https://www.unb.ca/cic/research/datasets/nsl.html.

[22] R. Li YMa RJiao, "A hybrid malicious code detection method based on deep learning," *International Journal of Software Engineering & Its Applications*, vol. 9, no. 5, pp. 205–216, 2015.

[23] R. C. Staudemeyer, "Applying long short-term memory recurrent neural networks to intrusion detection," *South African Computer Journal*, vol. 56, no. 1, 2015.

[24] M. K. Putchala, "Deep learning approach for intrusion detection system (IDS) in the Internet of Things (IoT) network using gated recurrent neural networks (GRU)," Wright State University, 2017.

[25] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, "Malware traffic classification using convolutional neural network for representation learning," in *2017 International Conference on Information Networking (ICOIN)*, pp. 712–717, Da Nang, Vietnam, 2017.

[26] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of ICML workshop on unsupervised and transfer learning*, pp. 37–49, Edinburgh, Scotland, UK, 2012.

[27] J. Li, J.-H. Cheng, J.-Y. Shi, and F. Huang, "Brief Introduction of Back Propagation (BP) Neural Network Algorithm and its Improvement," in *Advances in Computer Science and Information Engineering*, pp. 553–558, Springer, Berlin, Heidelberg, 2012.

[28] T. Xu, H. Wei, and G. Hu, "Study on continuous network design problem using simulated annealing and genetic algorithm," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1322–1328, 2009.

[29] A. M. Chandrashekhar and K. Raghuveer, "Performance evaluation of data clustering techniques using KDD cup-99 intrusion detection data set," *International Journal of Information and Network Security*, vol. 1, no. 4, p. 294, 2012.

[30] M. al-Qatf, Y. Lasheng, M. al-Habib, and K. al-Sabahi, "Deep learning approach combining sparse autoencoder with SVM for network intrusion detection," *IEEE Access*, vol. 6, pp. 52843–52856, 2018.

[31] R. Sen, M. Chattopadhyay, and N. Sen, "An efficient approach to develop an intrusion detection system based on multi layer backpropagation neural network algorithm: IDS using BPNN algorithm," in *Proceedings of the 2015 ACM SIGMIS Conference on Computers and People Research*, pp. 105–108, New York, NY, USA, 2015.

[32] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *IEEE Access*, vol. 5, pp. 21954–21961, 2017.

[33] N. Gao, L. Gao, Q. Gao, and H. Wang, "An intrusion detection model based on deep belief networks," in *2014 Second International Conference on Advanced Cloud and Big Data*, pp. 247–252, Huangshan, China, 2014.

[34] M. M. Hassan, A. Gumaei, A. Alsanad, M. Alrubaian, and G. Fortino, "A hybrid deep learning model for efficient intrusion detection in big data environment," *Information Sciences*, vol. 513, pp. 386–396, 2020.