ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

Intrusion Detection in Power Information Network Based on Deep Learning and Cloud Computing

Jianxun Wang[1,a], Jia Li[1,b], Chi Zhang[2,c*]

[1]*Beijing GD Power New Energy Technology Col, Ltd, Chaoyang 100101, Beijing, China*
[2]*Beijing ZTXY Technology Co., Ltd, Chaoyang 100101, Beijing, China*
[a]*63199630@qq.com*, [b]*15040088825@139.com*, [c]*zhangc@cnztxy.com*
**Corresponding author*

Abstract: *With the electric power information network system has been popularized to all aspects of life, and its accompanying problems also come one after another, that is, its internal and external invasion ways are diversified, then the information security problems will affect the national economy, infrastructure and many other livelihood issues. On the one hand, the information security of power network needs the general characteristics of computer information security, while considering the characteristics of high security. According to the structure of power information network, this paper proposes an intrusion detection model to detect power information network, which uses both deep learning theory and cloud computing platform. This model not only uses parallel anomaly detection and misuse detection methods, but also can deal with the problem that a single misuse detection model can't detect new attack variants. At the same time, it can capture a large number of data flows of power information network, learn and extract the essential characteristics of data flows through the deep feature learning and extraction ability of deep learning, and quickly and accurately detect through the platform Invasion. Based on the analysis of all kinds of intrusion behaviors in power information network, this paper proposes a data extraction and analysis method based on Hadoop, which improves the accuracy of intrusion detection by analyzing massive captured data flow packets and generating characterization data. At the same time, considering the important factors affecting the detection of power information network intrusion feature selection, we study the automatic encoder algorithm using spark platform. According to the experimental results, it is found that this method can improve the feature selection of power intrusion detection system.*

Keywords: *Power Information Network, Intrusion Detection, Cloud Computing, Deep Learning*

1. Introduction

Sound power information network is very important for the long-term development of the society. According to the characteristics of power production in China, power information network can be divided into power information management system and power operation control system. From the current situation, the power information management system has developed its own customized security management system. When the power information network works, at the same time, the network firewall and anti-virus software set up on the computer will also work at the same time. The relevant personnel of the power grid can detect the work of the power information network in real time. At the same time, the electronic computer will back up the important power information data, in order to prevent the loss of power due to the lack of data. In addition, technicians use the power operation control system to operate all the power facilities through the special software of the relevant enterprises. The staff can check the operation of the equipment according to the operation data obtained from the operation of the equipment, which can not only reduce the manual operation, but also reduce the probability of safety accidents for the staff by the unified operation of the computer, greatly increased the efficiency of power enterprise operation. However, the electric power information network is not equipped in many small electric power enterprises. Due to the lack of funds, the electric power information network of many enterprises has not installed the basic firewall and anti-virus software, not to mention the backup of data, which greatly increases the probability of accidents in the electric power information network, and there will be certain security risks [1-4].

At present, the security risks in the field of electric power information network mainly include the following points: first, electronic computers require technical personnel to have very professional tech-

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

nical ability, and relevant technical personnel need to have excellent professional technical ability to operate the electric power operation control system [5]. In today's society, electric power information network operation technicians lack of professional knowledge of electric power information network operation, so it is difficult to operate the electric power operation control system efficiently and accurately, which to a large extent causes the instability of electric power control system operation and affects the electric power operation control work. Secondly, many managers of electric power enterprises do not attach great importance to information security management, and the leakage of this information data will cause great economic losses. Because of the lack of information security awareness and prevention concept, the staff of electric power enterprises ignore the information security system for their own convenience or careless mistakes, so they are used as crime tools by network crime. Therefore, the lack of awareness of employees and its consequences will inevitably lead to the company's information and security into an irreparable dangerous state. In addition, the data in the power information network system is also relatively weak management; the power information network will leak out the data in its system and affect the normal operation of the power enterprise. Third, the leakage of data management, most enterprises use the database management system to save the database data, this data storage method is plaintext storage, which means that the data is easy to leak in the storage process, it must get the storage medium of the power enterprise when extracting the data, which will make all the information data in the medium be read, which affects the safety of the data Sex. At the same time, many power information network security defense software is not equipped, so hackers can directly bypass the firewall and read system information, and some power information network software and hardware providers can enter the power information network system background to easily read data. Therefore, in the electric power information network, the careless data preservation will directly lead to the leakage and damage of the electric power information network data, thus affecting the economic loss of the electric power enterprises [6-14].

The breakthrough of technology in the field of deep learning is mainly traced back to neural network model, and the theory of deep learning is mainly formed by imitating the reflection of human brain on every layer of objective things. Different from neural network model, deep learning model can solve the over fitting problem of multilayer neural network. The depth model is mainly constructed by multi-layer neural network. In the neural network model, there is no connecting line between nodes of the same layer, but the layers will be connected. The learning of the network is trained by greedy algorithm. When the training of the network layer reaches the specified accuracy requirements, the next layer will be trained. At the same time, through research and investigation, it is found that Hadoop is a distributed system model framework that allows users to develop distributed programs in a simple programming mode without knowing the underlying details.

This paper uses spark platform, which uses HDFS storage layer of Hadoop to store data permanently. Spark can directly reuse the workload of working data set in cluster computing, put the workload into memory cluster computing and optimize it. At the same time, memory cache data set, which shortens the data anti-counterfeiting delay. Also, through such a theoretical overview, the elastic distributed data set can be separated from the spark- based data processing system. In this paper, the spark cloud computing platform and the self-encoder network intrusion detection power system are used, and the parallel model design is adopted. In each self-coding network, the assigned network connection record sample files are trained iteratively. The initial weights of each self-coding network are randomly generated by the central node, and the updated weights are all processed by the central processing section point to update [6].

In a large number of high-speed power information network data, the two main factors that affect the ability of power information network intrusion detection system are the number of connection records and the performance of selection and classification. Considering these two important factors, this paper studies the feature selection method and BP classifier optimization algorithm, puts forward the deep learning method based on spark to detect the intrusion of power information network, and puts forward the corresponding optimization parallel algorithm, designs and implements a multilayer automatic encoder algorithm based on BP algorithm and DBN parallel optimization.

With the rise of internet technology, a large number of enterprises transfer their core business to the Internet, so network security has become an inevitable important issue for the people [7]. Generally speaking, ordinary enterprises usually set the first defense line of security as firewall. However, with the gradual maturity of hacker technology, intrusion means and technology have become increasingly complex. Simple firewall strategy has been unable to protect the security of many departments, so the defense network has become increasingly difficult, and more stable means should be designed to defend. At the same time, in the face of the increasingly complex network environment and atmosphere,

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

various devices need to constantly upgrade the system, which makes the company's network management personnel more and more responsible, and the carelessness of the administrator will cause major security risks. In the current network environment, intrusion detection system has become a new focus of security issues, not only people pay more and more attention to its security awareness, but also it has played a very important role in many enterprise network environments.

2. Intrusion Detection Analysis of Power Information Network

2.1. Principle of Intrusion Detection

The detection part is a very important part of P2DR intrusion detection security management model, which extends and does not fill a security function of firewall. Compared with other border security measures, intrusion detection model is the core of P2DR model. Detection can help network managers understand and analyze dynamic network data flow in real time.

Network intrusion is a very broad concept, which not only includes the hacker who has obtained the illegal control of the system, but also includes many behaviors such as collecting vulnerability information and refusing to access DOS, which do harm to the computer hardware and software system. Intrusion detection DOS is an intrusion detection behavior. It collects and analyzes the information of many key points in the computer network system, so as to judge whether there is any sign of behavior attack by non-security policy in the network system. With the rapid change of computer network and the complexity of its structure, people put forward higher requirements for security risk control. At present, detection and corresponding infrastructure based on intrusion detection behavior is the most effective way for dynamic control and management of security and persuasion protection, and corresponding intrusion detection has also become the infrastructure of network security.

Intrusion detection technology is a technology to detect the abnormal behavior of computer system, which is designed to ensure the security of computer network system. Intrusion detection system detects intrusion behavior by checking network traffic and various system events, such as system call, CPU utilization and file operation.

2.2. Analysis of Power Information Network Based on Cloud Computing

At present, the development of cloud computing has been expanding to different application areas, followed by security and privacy issues in the cloud environment are gradually highlighted [8]. Traditional security protection methods have been unable to meet the current complex cloud environment detection needs. It is limited in response speed, detection range and system scale. So how to build an efficient intrusion detection system in the cloud environment is a very important research direction in the field of intrusion detection [9].

At present, the security technology applied in the cloud needs to solve various challenges such as high concurrent access, scale data, 24-hour service availability, software compatibility, etc. the future development direction of intrusion detection also becomes how to ensure the efficiency and quality of cloud services and improve the security of the whole cloud system. At present, most of the traditional security detection methods have been unable to meet the dynamic complexity in the cloud environment. The traditional security protection technology has many limitations in detection accuracy response speed and system scale. The problem of a large number of data processing computational intrusion detection is that at present, the traditional intrusion detection algorithms are only suitable for processing small-scale data. When the amount of data increases, the time consumption of traditional algorithms will increase because of the speed of calculation, resulting in the computer cannot run, but cloud computing can just avoid this minefield, which can improve the super high-speed computing capacity and storage capacity, which can be the same as At the same time, cloud computing can build a large-scale platform for intrusion detection analysis to improve the security situation of the whole cloud environment.

In the cloud environment, large-scale nodes can collect and process cloud security events in a distributed way, and analyze cloud security events through the cloud intrusion detection and Analysis Center, so as to improve the ability of collecting and processing security events in real time. At the same time, if the traditional intrusion detection algorithm can be parallelized to improve the running speed and applied to the cloud environment, many challenges faced by the intrusion detection will be resolved. Cloud computing is a very important research field which represents an important trend of the

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

future development of information industry, so we consider to combine many algorithms of intrusion detection with cloud computing platform, which can expand the application scope of cloud computing and improve the performance of intrusion detection technology to achieve a qualitative leap [10-11].

2.3. Parallel Design of Automatic Encoder Network Based on Spark Cloud Computing Platform

Sparse self-encoder is an unsupervised network learning algorithm, which uses back propagation algorithm to adjust the input and output values to make them equal. The hidden layer in the network is a general expression of input data information [12-14]. The internal structure of the self-encoder hidden layer network is shown in Figure 1 below:

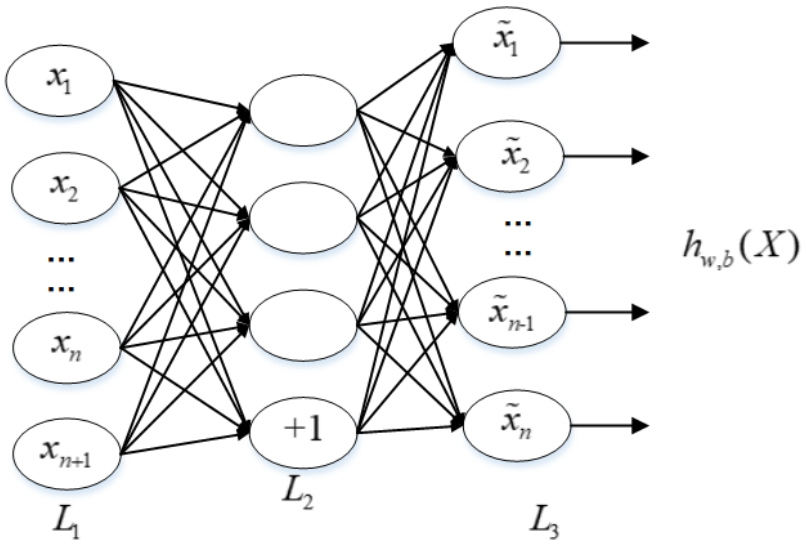

Figure 1: Structure of automatic encoder

As shown in Figure 1 above, L_1 layer mapping to L_2 layer represents the feature mapping from n dimension to m dimension, in which layer L_1 is the input vector of n dimension, and $x \in [0,1]^n$ its mapping to layer is $y \in [0,1]^m$, what's more $m < n$, when the activation function selects sigmoid function:

$$y = f_\theta(x) = s(Wx + b) \tag{1}$$

In the above equation $\theta = \{W, b\}$ is the expression of network parameters, W is a dimension weight matrix $m \times n$, where b is the offset vector. L_1 map to L_2 layer and then get reconstruction vector $z \in [0,1]^n$ according to inverse mapping. The expression of inverse mapping is:

$$z = g_{\theta'}(\mathrm{y}) = \mathrm{s}(\mathrm{W}^T\ y + b') \tag{2}$$

It is worth mentioning that $\theta' = \{W^T, b'\}$, W^T is the transposed weight matrix, b' is expressed as offset vector. The goal of network training is to minimize the error between the reconstructed value and the original sample. Then the error between each sample $x^{(i)}$ and the reconstructed value $z^{(i)}$ is: $\left|x^{(i)} - z^{(i)}\right|$, we define the error function as $L(x^{(i)}, z^{(i)})$, then the square error can be expressed as follows:

$$L(x,z) = \frac{1}{2}\sum_{i=1}^{i=n}\left|x^i - z^i\right|^2 \tag{3}$$

By substituting equation (1) (2) into equation (3), we can get the cost function of the weight matrix W and offset b in the automatic encoder, assuming that the cost function $J(W,b)$ is defined as:

$$J(W,b) = \frac{1}{2}\sum_{i=1}^{i=n}\left|x^i - s\left(W^T s\left(Wx^i + b\right) + b'\right)\right|^2 \tag{4}$$

Adding other constraints to the cost function of sparse automatic encoder, keep most L_2 layer nodes

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

suppressed, then the cost function of SAE can be expressed as:

$$J_{sparse}(W,b) = J(W,b) + \beta\sum_{i=1}^{m} KL(\rho\|\rho_j) \tag{5}$$

It is worth mentioning that ρ_j is the output value of the hidden layer, ρ is the average output value of the layer L_2 and the penalty factor is β.

$KL(\rho\|\rho_j)$ represents the information entropy of ρ and ρ_j the output values sum.

$$\mathrm{KL}(\rho\mathrm{P}\rho_j) = \rho\log\frac{\rho}{\rho_j} + (1-\rho)\log\frac{1-\rho}{1-\rho_j} \tag{6}$$

When $\rho = \rho_j, KL(\rho\|\rho_j) = 0$. Using the gradient descent method, the weight and offset changes are obtained, and the following results are obtained:

$$W_{ij}^{(\mathrm{l})} = W_{ij}^{(\mathrm{l})} - \frac{\partial J_{sparse}(\mathrm{W},\mathrm{b})}{\partial W_{ij}^{(\mathrm{l})}} \tag{7}$$

$$b_i^{(\mathrm{l})} = b_i^{(\mathrm{l})} - \frac{\partial J_{sparse}(\mathrm{W},\mathrm{b})}{\partial b_i^{(\mathrm{l})}} \tag{8}$$

After the training, error back propagation algorithm is needed to fine tune the self-coding network.

2.3.1. Encoder Parameter Design

For the connection records in the pre-processed power information network packets, we can reduce the data dimension by designing encoder parameters and selecting effective features. The task of encoder is mainly composed of training and fine tuning. The connection record training sample can be expressed as $(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\cdots,(x^{(n)},y^{(n)})$, the sample dimension is d, then the input layer has d nodes, The binary heuristic method can determine the number of nodes in the hidden layer. The specific process is as follows:

a. The initial number of hidden layer nodes sets the $\frac{1}{2}$ number of input layer nodes

b. From the initial value, the number of convergence iterations and classification accuracy are classified.

2.3.2. Training Phase

The training iteration is set to γ, the error condition is represented by e, the penalty factor is represented as β, and the value between 0-1 is selected randomly as the initial weight and offset value W,b. The first input sample is $(x^{11},x^{12},x^{13},\cdots x^{1d})^T$, to calculate the output value $y^{(1)} = (y^{11},y^{12},y^{13},\cdots y^{1\frac{d}{2}})^T$ of each hidden layer node according to equation (1), the reconstruction vector $(z^{11},z^{12},z^{13},\cdots z^{1d})^T$ from equation (2), and the updated weight matrix and offset vector from equation (7) (8). Next, other training samples update the weight matrix and bias according to the above steps, so as to determine whether the actual total error E meets the set error threshold requirements e. If the number of iterations has reached the threshold, the training is stopped, and the sparse encoder is fine-tuned again [16-20].

2.3.3. Refine Phase

If the error back propagation algorithm is used to fine tune the automatic encoder, the cost function expression becomes:

$$\mathrm{J}(\mathrm{W},\mathrm{b}:\mathrm{x},\mathrm{y}) = \frac{1}{n}\sum_{i=1}^{i=n}\left(\frac{1}{2}Ph_{W,b}\left(x^{(i)}\right) - y^{(i)}P^2\right) \tag{9}$$

$J(W,b:x,y)$ express as mean square error in supervised network training, x is sample input, y is solved target value, $h_{W,b}(x^{(i)})$ is network output. For the W and b partial derivatives of the sum $J(W,b:x,y)$, the updated formula of the sum of weights is obtained:

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \alpha \frac{\partial J(\mathrm{W},\mathrm{b})}{\partial w_{ij}^{(l)}} \tag{10}$$

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \alpha \frac{\partial J(\mathrm{W},\mathrm{b})}{\partial w_{ij}^{(l)}} \tag{11}$$

α represents the learning rate, $W_{ij}^{(l)}$ represents the connection weight from the i node at the l layer to the j node at the $l+1$ layer. b_i^j represents the offset of the l layer i node. In the training stage, the weight of the encoder will be adjusted by the above two equations, so the adjustment process of the encoder can be regarded as the training process of BP neural network.

In this paper, the automatic encoder identification intrusion detection technology can be divided into two parts: the first part is to calculate the output of hidden nodes according to the input, and set the number of input samples as N, m as the sample dimension, L is the hidden layer node, so the time complexity of time calculation is O (NML); the second part is to update the encoder weight through the previously calculated complexity, which is also o (NML). The computing time will increase greatly with the increase of the number of samples, the dimension of samples and the number of hidden nodes. Therefore, the parallel implementation of the automatic encoder algorithm can improve the operation speed by allocating the calculation scale, thus reducing the requirements of the algorithm on the computer memory and processing capacity. The parallel encoder algorithm based on spark proposed in this paper can be divided into two parts: the first part is to calculate the hidden layer output of information data set and calculate the output of all hidden layer nodes in parallel; the second part is to reduce the weight of the same key value and update the weight in parallel [20-24].

3. Optimization of BP Neural Network Based on DBN

BP network uses its feedforward network model to learn and store a large number of mapping relationships between input and output patterns, which can be applied to classification and prediction. BP network is generally composed of three layers. Each layer is connected by weight. The process of learning and training mainly includes forward weight calculation and reverse error propagation. Because BP network is very dependent on the initial set of weights, it is easy to fall into the local minimum [25-28]. At the same time, its convergence speed is very slow, resulting in a long training time. In view of these shortcomings, many improved BP algorithms have been proposed. In order to solve the problem that BP network is very sensitive and dependent on the initial weight, researchers propose a global search strategy based on the global search algorithm in the value space. At present, the typical algorithms are mainly based on genetic algorithm, ion swarm algorithm, ant colony algorithm and other intelligent algorithms. In consideration of the slow convergence speed, the researchers put forward an improved algorithm based on BP algorithm to drive the variable term to change the step length adaptively. In this way, the disadvantages of BP network can be effectively solved, and the global minimum can be solved efficiently to avoid too many iterations [29-30].

Boltzmann machine is a kind of energy model including the visible layer and the hidden layer with no connection between the nodes of the layers. n and m are the number of neurons in the visible layer $P(v|\theta) = \frac{1}{Z(\theta)} \sum_h \exp(-E(v,h|\theta))$ and the hidden layer h, respectively. v and h represent the state vectors of the visible layer and the hidden layer. For a given RMB state vector, the energy of RMB is defined as:

$$E(\mathrm{v},\mathrm{h}|\theta) = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n}\sum_{j=1}^{m} v_i W_{ij} h_j \tag{12}$$

The parameters in the equation are RBM, which W_{ij} are the connection weights of the i visible unit and the j hidden layer unit. a_i represent the offset of the i visible unit and the j is hidden layer unit b_j. The joint probability between visible layer and hidden layer is:

$$P(v,h|\theta) = \frac{1}{Z(\theta)} \exp(-E(v,h|\theta)) \tag{13}$$

Where $Z(\theta)$ is the normalization factor, $Z(\theta) = \sum_v \exp(-E(v,h|\theta))$. The distribution of observation data

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

defined by RMB $P(h|\theta)$ and $P(h|\theta)$:

$$P(h|\theta)=\frac{1}{Z(\theta)}\sum_{v}\exp(-E(v,h|\theta)) \tag{14}$$

$$P(v|\theta)=\frac{1}{Z(\theta)}\sum_{h}\exp(-E(v,h|\theta)) \tag{15}$$

To maximize the marginal distribution of RBM, the parameters of RBM can be obtained. Generally speaking, we use the method of maximum likelihood function to solve the problem. If there is a sample N , the maximum likelihood function can be expressed as:

$$L(\theta)=\frac{1}{N}\sum_{n=1}^{N}\log P(v|\theta) \tag{16}$$

In this paper, the random gradient descent method is used to solve the maximization $L(\theta)$, in which the partial derivation $L(\theta)$ about θ can be obtained as follows:

$$\frac{\partial L}{\partial\theta}=\sum_{n=1}^{N}(<\frac{\partial(-E(v^{(n)},h|\theta))}{\partial\theta}>_{P(h|v^{(n)},\theta)}-<\frac{\partial(-E(v^{(n)},h|\theta))}{\partial\theta}>_{P(v,h|\theta)}) \tag{17}$$

The mathematical expectation $<>_P$ is expressed in inequality of distribution. When $P(h|v^{(n)},\theta)$ the visible element is known as $v^{(n)}$, the probability distribution of the hidden layer and the joint probability distribution of the visible element and the hidden layer element are represented as $P(h|v^{(n)},\theta)$. When calculating this term P , it is difficult to get the unbiased estimation of the sample due to the existence of the normalization factor, so the contrast divergence algorithm is used to reconstruct the sample data for approximate sampling. As shown in Figure 2 below, the structure model of BP network based on DBN is composed of two RBMs. After the first RBM is trained, the output of the first RBM is taken as the input of the second RBM. After the second RBM is trained, the BP algorithm is used to fine tune to achieve the purpose of optimizing the BP network classifier.

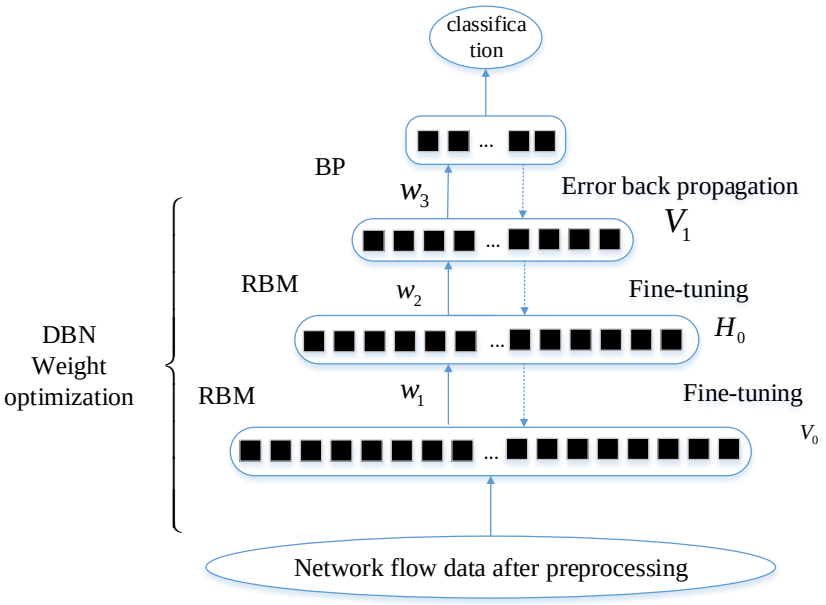

Figure 2: Structure of BP network based on DBN optimization

DBN-BP algorithm is mainly composed of two parts, the first part is to train one by one using contrast divergence algorithm; the second is to train BP neural network. The algorithm flow of contrast divergence is as follows:

Input: training sample x, number of explicit layer units expressed as $V_n^{(i)}$, number of hidden layer unit is $H_n^{(i)}$ s, number of iterations is γ , error as ε , learning rate as μ

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

Output: connection weight matrix W, display layer unit offset a', hidden layer unit offset b'.

1) Through Gaussian distribution, the connection weights a and offset b between 0-1 are generated, $a' = a$ and $b' = b$, the sample values are input.

2) According to the Gibbs sampling method, the input data is sampled to make the reconstructed data close to the real sample data. In other words, the probability p of hidden nodes is calculated by the equation, and a random number p' between 0-1 is generated at the same time. When $p' >= p$, the value of the hidden node is set to 1, otherwise it is 0.

3) After getting the probability value of the hidden layer node, the reconstruction value of the explicit layer node is calculated. The calculation method depends on 2) 60%, and the cyclic sampling updating calculation twice approaches the real value.

4) Update parameters for RBM. Update formula to $W \leftarrow W + \mu(P(h_1 = 1|v_1)v_1^T - P(h_2 = 1|v_2)v_2^T)$, $a' \leftarrow a' + \mu(v_1 - v_2)$, $b' \leftarrow b' + \mu(P(h_1 = 1|v_1) - P(h_2 = 1|v_2))$.

5) Take down one sample data, repeat 1) - 4).

6) Step 2) - 5) repeat iteration γ times.

7) Reconstruction error calculation $E = \sum_1^N |v - v'|$, N is the number of samples. If $E < \varepsilon$, stop training the RBM of this layer and train the next RBM, otherwise continue the iteration.

The fine-tuning process of BP algorithm is as follows：

Input: weight sample x_0' after DBN structure optimization, initial weight W, learning rate μ', iterations number γ'

Output: update the weight W' after iteration through network, hidden layer offset a and output layer offset b.

1) Initialization of BP neural network. The number of input layers in BP network is the number of output layers of the second RBM, that is $H_n^{(l)}$, the second RBM l. The number of output layers is q. If each cell is 0 or 1, then the q number of cells can be expressed as 2^q category. The number h_b of hidden layer elements is determined by formula $h_b = (H_n^{(l)} + q)^{1/2} + c$, where c is constant. Sigmoid function is the activation function of hidden layer unit and output layer unit. If the actual output of output layer unit is greater than 0.5, it is set to 1, otherwise it is 0.

2) The signal is going forward. The output of the j-th unit of the hidden layer and the k-th unit of the output layer are calculated by the equation $h_0^j = \sigma(\sum_{i=1}^{H_n^{(l)}} x_0^{(i)} w_{ij} + a_j)$ and $O_0^{(k)} = \sigma(\sum_{j=1}^{h_l} h_0^{(i)} w_{jk} + b_k)$. Then the error between the actual output of the k-th unit and the target output is $e = \frac{1}{2}(O_0^{(k)} + O_{target}^{(k)})^2$.

3) Error back propagation. The negative gradient algorithm is used to calculate the sample W_w、W_a、W_b to minimize the total error of the sample, and the weights and offsets are modified according to the equation $a' \leftarrow a' + u'W_a, b' \leftarrow b' + u'W_b$ and $W' \leftarrow W' + u'W_W$.

4) When the total error can meet the given threshold error or reach the number of iterations, the training of BP network will be stopped immediately, otherwise repeat step 2) - 3)

4. Experimental Results and Analysis

4.1. Description of Experimental Environment

Experiment preparation Description: build a spark platform composed of four nodes in the laboratory. Each node machine is configured with Intel (R) core (TM) i5-2400 4-core CPU @ 2.60GHz, 2gbram, Ubuntu 14.04.1 LTS, Hadoop version 2.5.1, spark version 1.3.1.

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

4.2. Test Data Set Description

All the data sets used in this paper are from the network traffic audit log data of a power enterprise. At the same time, in order to increase the data of all kinds of intrusion behaviors, a part of security audit data set is added to the data set after analysis and processing on Hadoop platform to form the intrusion detection data set in this paper. This data set uses 2 million network traffic as training Data, while the other 1 million data sets are test data sets. There are four types of intrusion: port scanning attack, DoS attack, local user's unauthorized access and remote host's unauthorized access. Among the 39 types of intrusion attacks found, there are 22 kinds of training data set provided this time. It is worth mentioning that each record has 53 dimensional attributes, and the last attribute is its category. Data is generally composed of the following four aspects: first, fully consider all the basic characteristics of network connection: such as destination IP address, source IP address, source port, destination port and other attribute fields. Second, consider the content characteristics of network connection: the data part of the data package contains the user's remote access and operating system sensitive file instructions and login system password and other information. Third, consider the time characteristics of traffic: Based on the time correlation of network attacks, some connections with the connection within 2S before the current connection are counted, assuming that the percentage of the same host and service type with the current connection within 2S, etc. Fourth, fully consider the traffic statistical characteristics of the specified host: the actual network attack behavior will be longer than the time span of 2S. In order to find out the attack, count the relationship between the 100 connections before the current connection and the link, for example, count the percentage of the same host and service type between the first 100 connections and the current connection. The format of a normal and an abnormal network connection data is shown below.

192,112,211,25,202,206,187,45,4532,80,31,2,tcp,smtp,0,1684,363,0,0,0,0,01,0,0,0,0,0,0,0,0,0,0,1,1, 0.00,0.00, 0.00,1.00,0.00,0.00,104,66,0.63,0.03,0.01,0.00,0.00,0.00,0.00,0.00,normal.

192,119,131,65,202,206,225,130,7642,25,24,0,tcp,private,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,38, 1,0.00,0.00, 1.00,1.00,0.03,0.55,0.00,208,1,0.00,0.11,0.18,0.00,0.01,0.00,0.42,1.00,portsweep.

Because the data contains discrete and continuous data. So, we need to standardize and normalize the data in order to fit the input of neurons and avoid the situation of large numbers eating decimals.

4.2.1. Data Standardization

The input of neural network must be numerical data, and the attribute field is the unified coding of the character type. In general, dictionary sorting is used to assign ordinal numbers to character fields. At present, there are three types of protocols: TCP, UDP and ICMP. The sorted results are ICMP, TCP and UDP. As shown in Table 1, other character fields are standardized in the same way.

Table 1: Protocol type coding method

Character attributes	Coded number
TCP	3
UDP	2
ICMP	1

Data normalization: The size range of the data is normalized to the interval [0,1], and the data $a = \frac{a - \min}{\max - \min}$ is normalized using the equation. It is worth mentioning that a represents the value of the property field, $\max$ and $\min$ represents the maximum and minimum values of the property field respectively.

Description of evaluation index: If there are M pieces of normal network behavior and N pieces of attack behavior data in the test sample set, then the trained intrusion detection model can correctly identify m' the normal network behavior records and the n' attack behavior records, then the overall accuracy of the test sample data set can be expressed as $R_e = \frac{m' + n'}{m + n}$, the false alarm rate as $R_w = \frac{m - m'}{m}$ and the missed alarm rate $R_l = \frac{n - n'}{n}$.

4.2.2. Performance Test of MR_DBN_BP Algorithm

Data sample set construction:

Because there are relatively few R2L and U2Rin the whole KD data set, the training set and test set

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

can be divided into sample data sets according to the ratio of 3:1, as shown in Table 2: We randomly select 6590 network connection data as sample data sets.

Table 2: Sample data

Connection record / Sample data	Normal record	Attack record			
		Dos	Probe	U2R	R2L
Training sample data set	2500	1500	600	40	300
Testing sample data set	835	500	200	15	100

The sample data contains continuous discrete values, so in order to standardize and normalize the data, it is necessary to code the category of the records so as to make the BP classification correct.

Training parameter setting of experimental method: Compared introducing steepness factor λ with BP method, PSO-BP method and GA-BP method, the network parameters are set as follows: the number of RBM in DBN network is set as 2, and the number of nodes is 53-22-121000 as the number of iterations.

4.3. Experimental Results and Analysis

As shown in Table 3, this method and the other three traditional methods are compared in three aspects: false alarm rate, detection rate and false alarm rate.

From the experimental data in the table, we can know that: in the case of a single machine, the detection time of BP network based on DBN optimization is slightly longer than the other three existing methods, but because DBN method can select more essential features in the data more accurately during training, so the detection accuracy is higher than other methods, and at the same time, it is also very effective to reduce the detection error rate and missing detection rate.

Table 3: Comparison of BP network detection rate with other optimization methods

Connection Record / Sample data	Normal record	Attack record				Detection rate (%)			Detection time(s)
		Dos	Probe	U2R	R2L	R_c	R_w	R_L	
	835	500	200	15	100				
λ BP	789	462	169	8	76	91.15	5.51	12.23	19.84
PSO-BP	817	491	178	10	87	95.94	2.23	6.01	21.09
GA-BP	823	487	185	11	89	96.67	1.47	5.33	20.18
DBN-BP	826	496	189	10	93	97.82	1.13	3.36	21.25

As shown in Figure 3, we show the relationship between the number of iterations and the mean square error of the DBN based BP network algorithm in single machine mode and the training on the spark platform.

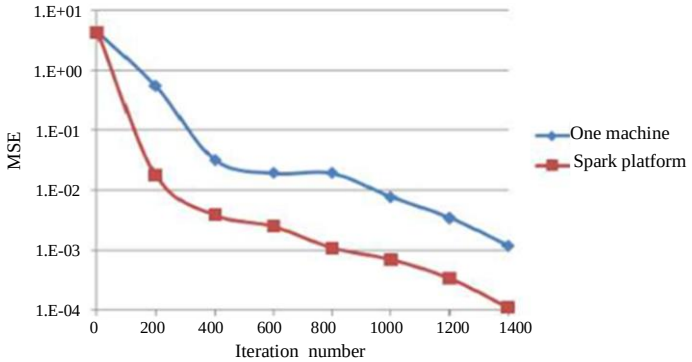

Figure 3: The training contrast relationship of DBN_BP

According to Figure 3, it takes longer to train DBN_BP under a single machine, and many iterations are needed to reach the established goal requirements. However, in the cluster experiment, three nodes in the experiment can train DBN_BP at one time, so that the goal requirements can be achieved with very few iterations.

In a very ideal situation, the acceleration ratio and expansion rate of parallel cloud computing system are set to 1, which cannot reach the ideal state in practical application due to the time signal transmission delay between nodes. At the same time, because of the increase of data set, the expansion rate

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

of cluster will decrease with the increase of data set. With the increase of the number of cluster nodes, the communication overhead between nodes increases gradually, and the acceleration ratio of the system also decreases. As shown in Figure 4, the growth rate of the acceleration ratio of MR_DBN_BP algorithm decreases with the increase of the data volume of the experimental data set. As shown in Figure 5, the decline slope of the algorithm's expansion rate also decreases. From the above analysis, the proposed MR_DBN_BP algorithm performs well in all expansion indexes.

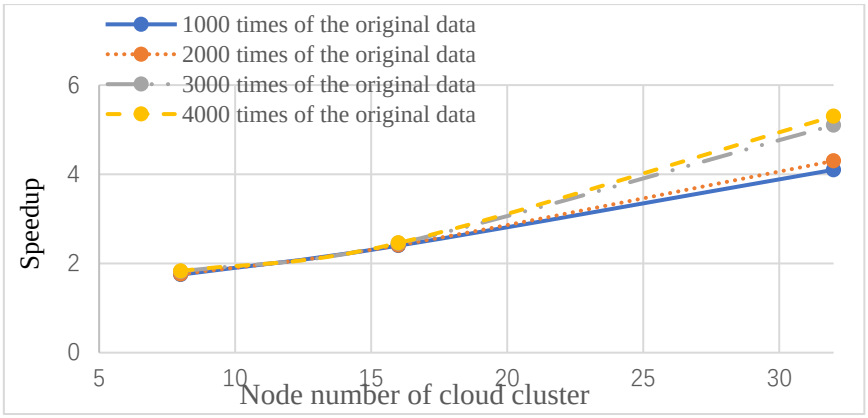

Figure 4. The accelerate of MR_DBN_BP Algorithm

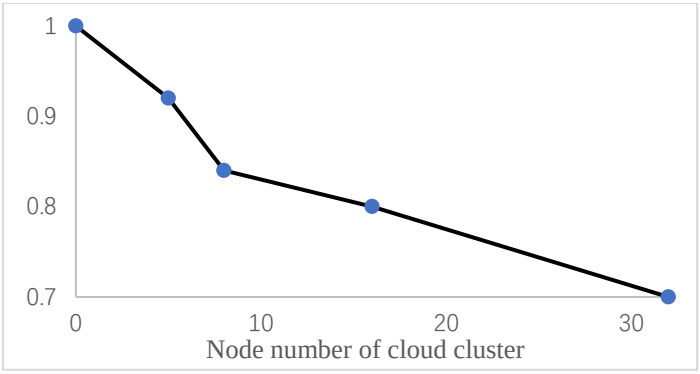

Figure 5. The expansion rate of MR_DBN_BP Algorithm

5. Conclusions

With the development of Internet technology and the high-speed development of power information system in China, the information security problems in power information system are gradually highlighted. Because the security threats generated by the Internet are more complex and diversified, the current security issues have greatly affected the production and normal operation of the power system. In the infrastructure of national economy, electric power information network is very important, which to some extent determines that information network should not only consider the characteristics of computer information security, but also consider the characteristics of high security requirements. In this paper, by analyzing the intrusion threats in the current power information network, and in-depth research and exploration of the power information network intrusion detection methods that have been studied, in the case of thorough study of the current power information network structure, fully combining the advantages of anomaly detection and misuse detection mode. This paper proposes an intrusion detection model based on deep learning and cloud computing. This paper not only analyzes all kinds of intrusion behaviors in the current power information network, but also puts forward a method of data packet extraction based on Hadoop platform, which is based on the factors of traffic, protocol and time. This method analyzes a large number of captured data packet characterization data and improves the accuracy of intrusion detection. The spark cloud computing platform with four nodes can provide a good technical service and performance test for the theory of power network information intrusion detection technology based on cloud computing. We process the data set through standardization and normalization, and input the processed data into the neural network unit. A large number of experimental data and tests show that this method can meet a large number of high-dimensional power network intrusion detection data and provide a large number of computer resources to make real-time

ISSN 2616-5767 Vol.4, Issue 8: 10-21, DOI: 10.25236/AJETS.2021.040802

detection.

References

[1] Valenzuela J, Wang J, Bissinger N. Real-time intrusion detection in power system operations. IEEE transactions on Power Systems, 2013, 28(2): 1052-1062.
[2] Anderson James P. Computer security threat monitoring and surveil lance [R] Fort Washington, PA: James P. Anderson Co., 1980.
[3] Denning, Dorothy E.An intrusion detection mode. IEEE Transactions on Software Engineering (SE-13), 1987, 4(2):222-232.
[4] Lunt T. IDES: An intellingent system for detecting intruders [A]. In Proceedings of the Symposium: Computer Security, Threat and Counterneasures, 1990.
[5] GE Liepins, H S Vaccaro. Anomaly detection: purpose and framework[C]. In Proceedings of the 12th National Computer Security Conference, 1989, 10:494-504.
[6] Forrest S, HofmeyrS A, Somayaji A. Computer immunology. Communications of the ACCM, 1997, 40(10):88-96.
[7] Y.Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35(2013) 1978-1828.
[8] Asja Fischer, Christian Igel. An introduction to restricted Boltzmann machines, Lecture Notes in Computer Science, 2012:14-36.
[9] Kui Jia, Lin Sun. Laplacian Auto-Encoders: An explicit learning of nonlinear data manifold. Neurocomputing, 2014.10.16:1-1.
[10] Armbrust M, Fox A, Griffith R. A view of cloud computing. Communication of the ACM, 2010, 53(4):50-58.
[11] D. Song, E. Shi, l. Fischer, U. Shankar. Cloud data protection for the masses. IEEE Computer.2012, 45(1):39-45.
[12] Shao-Zi Li, Bin Yu. Feature learning based on SAE-PCA network for human gesture recognition in RGBD images. Neurocomputing 151 (2015) 565-573.
[13] Sascha Lange, Martin Riedmiller. Deep Auto-Encoder Neural networks in reinforcement learning. IEEE Conference on Computer Vision and Pattern Recognition, 2010.
[14] Chelsea Finn, Xinyu Tan, Yan Duan, Trevor Darrell, Sergey Levine, Pieter Abbeel. Deep spatial autoencoders for visuomotor learning. arXiv preprint arXiv:1509.06113, 2015.
[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In OVPR, 2014.
[16] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, A. Oliva, Learning deep feature for scene recognition using places database. Advances in Neural Information Processing Systems, 2015.
[17] Karpathy. A, Fei-fei. L, Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[18] J. Donahue, L.A. Hendricks, S. Guadarrama, M. Rohrbach. Long-term recurrent convolutional networks for visual recognition and description. In CVPR, 2015.
[19] Alex Krizhevsky, Ilya Sulskever, Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, 2013.
[20] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In ECCV, vol.8689:818-833, 2015.
[21] Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature 521:436-444, 2015.
[22] C. Szegedy, et. al, Going deeper with convolutions. 2015.
[23] Kaiming. He, Xiangyu. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2016.
[24] Ronen Eldan and Ohad Shamir. The power of depth for feedforward neural networks. 2016.
[25] V. Mnih, Human-level control through deep reinforcement learning. Nature 518, 529-533, 2019.
[26] David Silver, Mastering the game of go with deep neural networks and tree search, Nature, 529:484-489, 2016.
[27] Sergey Levine, Chelsea Finn, Trevor Darrell and Pieter Abbeel. End-to-End Training of Deep Visuomotor Policies. 2016.
[28] S. Venugopalan, H. Xu, J. Donahue, M. Rohrbach, R. Mooney, K. Saenko, translating videos to natural language using deep recurrent neural networks. In NAACL, 2015.
[29] L. C. Chen, Y. Wang, J. Xu, Attention to scale: Scale-aware semantic image segmentation, 2015.
[30] X. Kelvin, L.B. Jimmy, K. Ryan, Attend and Tell: Neural image caption generation with visual attention, 2016.