Research on Feature Selection Method of Intrusion Detection Based on Deep Belief Network

BaoyiWang[1, a],Shan Sun[2, b], Shaomin Zhang[3, c]

[1,2,3]Schoolof Control and Computer Engineering, North China Electric Power University,Baoding,071003, China

[a]email:wangbaoyiqj@126.com,[b]email:sunshan514@126.com,

[c]email:zhangshaomin@126.com

Keywords: Feature Selection;Deep Belief Network ;Intrusion Detection

Abstract. Feature selection is one of the important factors that affect the intrusion detection system. Aiming at the problems due to selecting the high feature dimension and the redundancy causelow detection accuracy and high missing rate in the traditional intrusion detection system.In this paper, the deep belief network algorithm is given to select featureslayer by layer to reduce the feature dimension.As the deep belief network algorithm is an unsupervised learning algorithm, it is more suitable for selecting features from a large number of unlabeled data.Compared with other feature selection algorithm,the experiment shows the deep belief network algorithm is more effective than other algorithm in intrusion detection network.

Introduction

With the improvement of network bandwidth, one of the main problems of the traditional intrusion detection system faces is a high missed rate,slow detection speed and too late to deal with vast amounts of network data. One of the main reasons is because of the large number of feature extraction and selection and the redundancy between features[1].

Therefore, many feature selection algorithms are applied to intrusion detection system. Literature [2] the kernel principal component analysis is used for the feature selection of network data. The improved kernel function N-RBF is used to eliminate the noise caused by the difference between attributes. This method can effectively reduce the dimension of data, but it cannot detect the intrusion behavior quickly when the data set is large. Literature [3] uses the optimized particle swarm optimization algorithm for feature extraction, the method is simple, so the characteristics of the selection of a wide range of applications.However,the particle swarm algorithm will converge prematurely and will be trapped in Local minimum value.

Intrusion Detection Based on Deep Belief Network

In the case of limited samples and computing units, the representation of complex functions with shallow learning algorithm is limited. Deep learning has a strong ability to learn in the feature， it can be abstracted from the high dimensional features of the low dimensional features of the representation of data. In 2006 Hinton et put forward that the deep belief networkis a kind of non-supervised greedy layer by layer training algorithm, which brings the hope to solve the problem of the deep structure. In the field of image recognition, speech recognition, human behavior recognition, multi-modal learning, spam filtering, deep learning's application is very extensive, and achieved excellent results [5].

Deep Belief Network

Deep belief network is a common method of deep learning. By selecting the feature layer by layer, the feature space of the original sample is transformed into another new feature space, which

makes it easier to classify or predict.DBN is usually composed of a plurality of restricted Boltzmann machines, as shown in Figure 2 is a model of RBM [6].

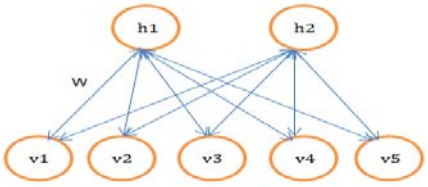

Fig.1. RBMmodel

RBM is a kind of energy model,assuming that a RBM has a n visible units and m hidden layer units,with vector v and h respectively state of visible and hidden layer units.$v_i$represents the state of the i unit, and$h_j$represents the state of the junit. For a given state (V, H), the energy is defined:

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

From formula 1, θ equals$\{W_{ij}, a_i, b_j\}$,W_{ij}, a_i and b_j are all RBM parameters. W_{ij} represents connection weights between i visible unit and j hidden unit.$a_i$is the bias of visible unit,, b_j means a bias of hidden unit.The joint probability of the visual layer and the hidden layer is

$$\mathrm{P(v,h|\theta)} = \frac{1}{Z(\theta)} exp(-E(v,h|\theta)) \quad (2)$$

The $Z(\theta)$ as the normalization factor, $Z(\theta) = \sum_{v,\mathrm{h}} exp(-E(v,h|\theta))$ .The distribution of observational data defined for P(v|θ)by the RBM, the joint probability P(v, h|θ)of the marginal distribution,

$$\mathrm{P(h|\theta)} = \frac{1}{\mathrm{Z(\theta)}} \sum_{\mathrm{v}} \exp(-\mathrm{E(v,h|\theta)}) \quad (3)$$

$$\mathrm{P(v|\theta)} = \frac{1}{Z(\theta)} \sum_{h} exp(-E(v,h|\theta)) \quad (4)$$

A maximum likelihood function is used to maximize the marginal distribution and get all parameters of RBM. Assuming the N sample, the maximum likelihood function is:

$$\mathrm{L(\theta)} = \frac{1}{N} \sum_{n=1}^{N} \log P(v|\theta) \quad (5)$$

We use stochastic gradient descent to maximum L(θ), L(θ)forθ the partial derivative is:

$$\frac{\partial L}{\partial \theta} = \sum_{n=1}^{N} \left(< \frac{\partial(-E(v^{(n)},h|\theta)}{\partial \theta} >_{P(h|v^{(n)},\theta)} - < \frac{\partial(-E(v,h|\theta)}{\partial \theta} >_{P(v,h|\theta)} \right) \quad (6)$$

$<\cdot>_P$is the mathematical expectation of P.$P(h|v^{(n)},\theta)$indicates the probability distribution of the hidden layer when the visible element is known as $v^{(n)}$.$P(v,h|\theta)$represents the Joint probability distribution between hidden unit and visible unit.

The following diagram is a structural model based on a deep belief network.

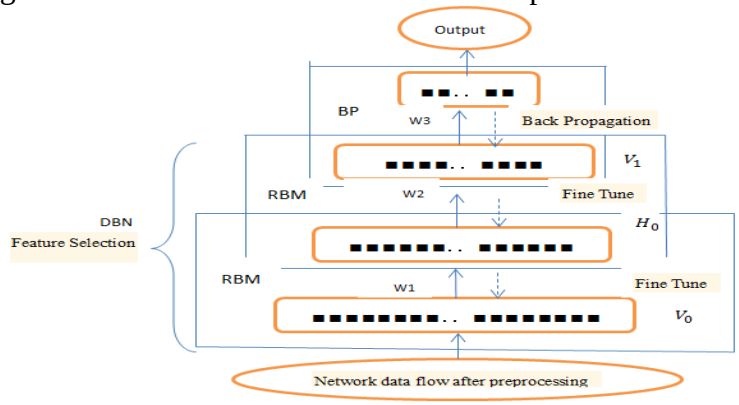

Fig.2. DBNstructural model

As figure 3 shows, the DBN is composed of two layers of RBM.First of all, training the first layer of RBM, the first RBM's output will treat as the input of second RMB. After the end of the RBM training, we can use BP network to fine tune to more easily classified [6].

Network Intrusion Detection Based on DBN

The intrusion detection system is used to obtain the normal behavior of user, and to gather information from the key nodes of the network and computer to detect if there are abnormal behaviors.Among them, describing the user's normal behavior habits is the key technology of system design.Although the network connection records have the different values of the different fields, it is possible to determine the entire connection record as a category.Network behavior can be divided into normal and abnormal behavior, and abnormal network behavior can be subdivided into denial of service attack, port scanning, unauthorized remote connections, etc. DBN is used to select the essential features of connection records, and training the BP neural network to establish the network behavior of normal profile, to identify various types of network attack [8].The following figure is the DBN network intrusion detection model.

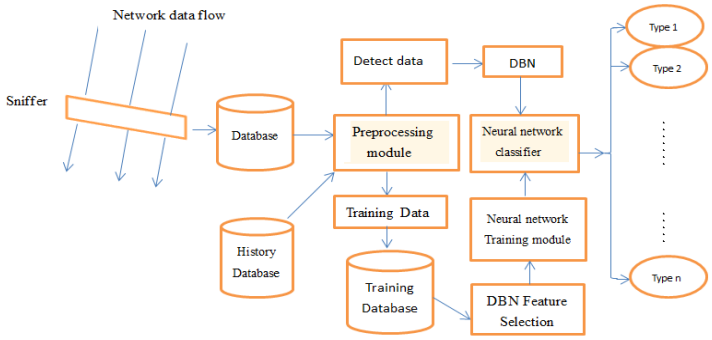

Fig.4.Network intrusion detection model based on DBN

The method of this paper includes three parts, data collection and preprocessing, off-line training and on-line detection. The workflow of the model is as follows:

1) Data collection and preprocessing

①Sniffer deployed in the core node on the network .

②The preprocessing module takes the network connection records from the database, analyzes the acquisition time and the connection description, and obtains the characteristic field values of the whole connection.Convert the non-numeric data into numerical data.

③The generated training data format is the target value of the connection record number, the characteristic field value (as DBN input), and the test result field (as the BP neural network classifier).

④Normalized training data.

⑤The training sample is divided into several small training sets and test sets.

2) Off-LineTraining

①DBN network gets the training data set from the training database, and set the number of the initial neural units as the number of training data. Until the RBM is trained to meet the conditionsof iteration or the deviation condition.

②BP neural network adopts three layer structure: the number of hidden units is the same as that of the input unit of the last layer RBM, and the output layer unit is the number of the intrusion attack types.

③Backpropagation algorithm for correcting deviations on BP networks.Wake-Sleep algorithm for tuning DBN from the bottom to the top.

④Hypothesis testing sample set contains a total of m normal network behavior and n attack behavior data. After detection is completed, a total of m'normal network behavior and n' attack behavior record is correctly identified,then the overall accuracyR_C ,the false positive rate R_W,false negative rate R_L of the test sample data set:

$$R_C = \frac{m\prime+n\prime}{m+n} \quad (7)$$

$$R_W = \frac{m-m\prime}{m} \quad (8)$$

$$R_L = \frac{n-n\prime}{n} \quad (9)$$

⑤Return to step 1, and adjust the RBM layer, and then compare the correct rate ofR_C, the rate of false positive samples rate$R_W$and negative rate R_L.

3) On-Line Detection

①DBN selects the data feature, BP network classifies feature dataaccording to different attack types to take different response method.

②Update the history database.

Algorithm Analysis and Explanation

It is set up by the state of the visible cell into a training sample v, using the formula (3) to calculate thevalue of all the hidden units.

Contrastive DivergenceAlgorithm

The main steps of the contrastive divergence algorithm are:

Input:training samplesx_0, visible layer unit number$V_n^{(i)}$, hidden layer unit number $H_n^{(i)}$, gamma iterationsγ, deviation epsilonε, learning rate ofμ.

Output: the connection weight matrix W, the visible layer unit bias a′, the hidden layer unit biasb′.

①The random number generated by the Gauss distribution to initialize the connection weight matrix, a andb. Input sample values.

②Sampling input data based on the Gibbs sampling method, which makes the reconstructed data more close to the real sample data.Using formula (3) computes the probability p of hidden nodes, while producing a 0-1 random number p', when p' >=p set 1 as the valueof the node, otherwise to 0.

③After get the probability value of node in hidden layer, using formula (4) calculates the reconstructed value of the visible unit.

④Update the parameter of RBM.$W \leftarrow W + \mu(P(h_1 = 1|v_1)v_1^T - P(h_2 = 1|v_2)v_2^T)$, $a' \leftarrow a' + \mu(v_1 - v_2)$, $b' \leftarrow b' + \mu(P(h_1 = 1|v_1) - P(h_2 = 1|v_2))$.

⑤Take the next sample data. Calculate the step ①-④ repeatedly.

⑥Repeat the steps ②-⑤ forγ times.

⑦The formula for the calculation of the reconstruction deviation is $E = \sum_1^N |v - v'|$.N is the sample number.If the E<ε, stop training andtraining next layer of RBM,or continue the iteration.

BP Algorithm

BP neural network is easy to appear the local minimum and the slow convergence speed in the case of the random initial weights. In this paper, the initial weights of the DBN weights are assigned to the BP network. The training process of DBN is also the process of searching the sample space. Here are the main steps of the BP neural network algorithm:

Input: features DBN selected x_0', the initial weight of W,learning rate μ', the number of iterationsγ'.

Output:W', bias a of hidden layer, bias b of output.

①BP network Initialization. The number of features of DBN selection is used as a number of input layer neurons.$H_n^{(l)}$, l represents the output of last RBM. The number of output layer is Q, the value of each unit is 0 or 1, then the Q unit can be represented by the Q times of the 2. The number of hidden unitsis determined by the formula $h_b = (H_n^{(l)} + q)^{1/2} + c$,where C is constant. The activation function of the hidden layer unit and the output layer unit is the sigmoid function, The actual output value of the output layer unit is greater than 0.5, then set to 1, otherwise it is 0.

②Signalforward propagation. The Formula $h_o^{(j)} = \sigma(\sum_{i=1}^{H_n^{(l)}} x_0^{(i)} w_{ij} + a_j)$can calculate the output

of j unit on hidden layer. And $O_o^{(k)} = \sigma(\sum_{j=1}^{h_b} h_o^{(j)} w_{jk} + b_k)$ can calculate the k unit output of the output layer. The actual output and target output deviation of the K unit is $e = \frac{1}{2}(O_o^{(k)} - O_{target}^{(k)})^2$.

③Backpropagation.The total deviation of the sample is minimized by the negative gradient method, and get the Δw，Δa，Δb, Modified weight and bias based on $W' \leftarrow W' + \mu' \Delta w$, $a \leftarrow a + \mu' \Delta a$, $b \leftarrow b + \mu' \Delta b$.

④When the total deviation meets the requirement or the number of iterations, the training of BP network hasbeen finished. Or repeat step ②-③.

Experiment

In order to verify the validity of intrusion detection method based on DBN algorithm, and compare the current intrusion detection method, the following experiments are carried out.

1）Constructed data sample set

In this paper, the data set is derived from the security audit data set named CUP99 KDD, which is formed by the IDS Laboratory of Columbia University. The data set includes four major categories: Dos attack, Probe, R2L, and U2R, and each record has 42 dimension attributes, and the last one belongs to category. The sample data set is divided into training set and test set according to the ratio of 3:1. The sample data set is selected from the 10% data provided by CUP99 KDD, a total of 6590 network connection data. As shown in table 1.CR means connection records.

Table1data sample set

CR / Data Sets	Normal Record	Attack Record			
		Dos	Probe	U2R	R2L
Training data sets	2500	1500	600	40	300
Test data sets	835	500	200	15	100

The sample data contains continuous values and discrete values, so the data is need to be standardized and normalized.

2) The training parameters setting of experiment method

In order to compare this method with PCA-BP, PSO-BP and GA-BP[11][12], the network parameters are set as follows.

①The DBN network has two RBM while the number of nodes is 41-22-12 and the iteration is 1000 times.

②BP neural network is a the three layer structure that the node number is 12-6-3, the iteration is 500 times, the learning rate is 0.01 and the deviation rate is 0.01

3) Results and analysis of the experiment

Table 5 shows the comparison results in the three aspects of detection rate, false positive rate and false negative rate with the other three methods. The experimental environment ismatlab 2013a.

Table5 The comparison of this method with other methods

CR / Method	Normal Record	Attack Record				Detection rate (%)			Time (s)
		Dos	Probe	U2R	R2L				
	835	500	200	15	100	R_C	R_W	R_L	
PCA-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
My method	826	496	189	10	93	97.82	1.13	3.36	21.25

Conclusions

In this paper, the deep belief network is applied to network intrusion detection, and the combination of BP neural network model, by comparing the results with other methods. The results show that the method in the intrusion detection feature selection has certain advantages, it can

reduce the miss rate and the dimension of data, to accelerate the speed of data processing, provides a method for the quasi real time detection of intrusion detection system.

Acknowledgement

In this paper, the research was sponsored by the Scientific research project of Hebei Province (Pr oject No. Z2012077).

References

[1]WangFeng,LiuDongdong,NiuLei,GuoBo.Featureselection and classifier optimization coupled to the network intrusiondetection [J]. computer engineering and applications, 2013,20:87-90.

[2]Yu Wenli, Yu Jianjun, Fang Jianwen. A new based on KPCA and improved epsilon SVMintrusion detection model[J].Computer engineering and application, 2015,11:93-98.

[3]HuangHuiqun,Sun Hong. The characteristics of particle swarm optimization and information gain determination of feature weights for intrusion detection [J]. computer applications, 2014,06:1686-1688+1693.

[4]Lichunlin, Yue Jiang Huang, wanghong, bovine Changxi. A depth study of network based intrusion detection method [J]. Information and communication security. 2014,10:68-71.

[5]Zhang Chunxia, JiNannan, Wang Guanwei. Journal of Engineering Mathematics [J].restricted Boltzmann machine, 2015,02:159-173.

[6]AsjaFischer,ChristianIgel. An Introduction to Restricted Boltzmann Machines[J], Lecture Notes in Computer Science,2012:14-36

[7]Yang Kunpeng. Intrusion detection model based on deep belief network [J], modern computer (Professional Edition), 2015,02:10-14.

[8]Du Qijun,WangShu,YuGuixian,LiGuangping,XuYafei,XueYang,WangXiaowei, the optimal feature selection method for network intrusion detection system [J].

[9]Liu Chun, a network intrusion detection model based on combination algorithm selection features [J]. computer and modern, 2014,08:75-80.

[10]Li Xuefeng. The network intrusion detection of genetic algorithm and support vector machine parameters [J].computer application and software, 2014,03:301-303+333.