CSCanada
Advances in Natural Science
Vol. 9, No. 2, 2016, pp. 39-43
DOI:10.3968/8209

ISSN 1715-7862 [PRINT]
ISSN 1715-7870 [ONLINE]
www.cscanada.net
www.cscanada.org

Intrusion Detection System Based on Integration of Artificial Neural Network and Support Vector Machine

XU Xiaolong[a],* GAO Zhonghe[b]; HAN Lijuan[a]

[a]Experiment Teaching Center, Qufu Normal University, Rizhao, China.
[b]Institute of Software, Qufu Normal University, Rizhao, China.
*Corresponding author.

Received 30 April 2016; accepted 14 June 2016
Published online 26 June 2016

Abstract

Soft computing techniques are more and more widely used to solve a variety of practical problems. This paper applied the integration of different soft computing techniques in intrusion detection system (IDS). Due to the increasing incidents of network attacks, building effective intrusion detection system is necessary, but it faces great challenges. Two sorts of soft computing techniques are studied:Artificial Neural Network (ANN) and Support Vector Machines (SVM). Experimental results show that integration of ANN and SVM is superior to individual approaches for intrusion detection in terms of classification accuracy.

Key words: Intrusion detection; Neural network; Support vector machines; Soft computing techniques

Xu, X. L., Gao, Z. H., & Han, L. J. (2016). Intrusion Detection System Based on Integration of Artificial Neural Network and Support Vector Machine. *Advances in Natural Science*, *9*(2), 39-43. Available from: http://www.cscanada.net/index.php/ans/article/view/8209
DOI: http://dx.doi.org/10.3968/8209

INTRODUCTION

Because of the important role of the computer network in a country, intrusion detection system is of great importance for information protection. This paper briefly summarizes the function of Artificial Neural Network (An & Liang, 2012) Support Vector Machine (Chow & Chu, 1997) and integration of different artificial intelligence technologies in the construction of intrusion detection system. A good detection technology is the key factor for the intrusion detection system to achieve the best performance, So we study different soft computing technologies and their integration to establish intrusion detection system model based on the experimental data.

Most of the intrusions can be found by examining the behavior patterns of users, so many intrusion detection systems using the identified attack mode for misuse detection (Ding, Li, Su, Yu, & Jin, 2013). Study found that SVM is superior to ANN in many important aspects of intrusion detection system (Huang, Lan, Hoffmann, & Lacey, 2011). This paper will focus on the integration of SVM and ANN to obtain higher detection accuracy.

1. INTRUSION DATA SET

In a DARPA intrusion detection assessment project, a LAN was built to simulate the US air force network, it collects raw TCP/IP data for analysis, like a real environment of multiple attacks. For each TCP/IP connection, it can extract more than 41 kinds of quantitative or qualitative attack features. In the database, a subset of 494,020 data were used, 20% of them represent normal mode.

There are four different types of attack patterns:

a) Denial of Service(DoS): DoS is making some calculation or storage resources too busy to handle normal user requests. Examples of denial of service attacks are Mail bomb, SYN flood, Ping of Death, Smurf, etc..

b) User to Root(U2R): Attackers login system with a regular user identity, then use some of the weaknesses of the system to obtain super user's authority. Examples are buffer overflow attack, Xterm, Perl, etc..

c) Remote to Local(R2L): A remote user without a local account gets local user's authority by sending packets. Examples are dictionary attack, IMAP attack, Xlock, etc..

d) Probe: Attackers can gather information or find loopholes by scanning computers on the network. They can master the distribution of available services or computers and then make use of them. Examples are IP sweep attack, Nmap scan, etc..

2. CONNECTIONIST MODELS

The connection model study by adjusting the connection between layers. When the network is fully trained, it can generate the corresponding output of the input data.

2.1 Artificial Neural Network

Artificial neural network technology enables us to design a nonlinear system which can accept a large amount of input. This design is only based on instances of input and output relations.

a) Resilient Back Propagation (RP)

The purpose of RP is to eliminate the negative effects of partial derivative magnitudes. Only the sign of the derivative is used to determine the direction of the weight update. The magnitude of the derivative has no effect on the update of the weight. The change of weight is determined by a separate update value. When the derivative of the performance function has the same symbol by successive iteration, each update value of weight and bias will increase. When the sign of the derivative changed after the previous iteration, the update values will be reduced. If the derivative is 0, the update value will not change. If the weight has been changed in the same direction after multiple iterations, the magnitude of weight will be increased.

b) Scaled Conjugate Gradient Algorithm (SCG)

In order to avoid the complex linear search procedure of traditional conjugate gradient algorithm, Moller introduced scaled conjugate gradient algorithm(Majumder, 2015). According to SCG, the value of Hessian matrix can be estimated as

$$E^{''}(\omega_k)p_k = \frac{E^{'}(\omega_k+\sigma_k p_k)-E^{'}(\omega_k)}{\sigma_k}+\lambda_k p_k .$$

Where E' and E'' is the first and second order derivative of global error function $E(\omega_K)$. Other symbol p_k,σ_k and λ_k represent weights, search direction, which is used to control two order derivative approximation weight change and the uncertainty of Hessian matrix. In order to get better quadratic approximation of function E, when the Hessian matrix is positive, a mechanism is needed to increase or decrease λ_k.

c) One-Step-Secant Algorithm (OSS)

Quasi-Newton method produces a series of matrix G^k which can approximate inverse of a Hessian matrix more accurately. It only uses first derivative information of function E. The updated expression is:

$$G^{(k+1)} = G^{(k)} + \frac{pp^T}{p^T v} - \frac{(G^{(k)}v)v^T G^{(k)}}{v^T G^{(k)} v} + (v^T G^{(k)} v)uu^T ,$$

where

$$p = w^{(k+1)} - w^{(k)}, v = g^{(k+1)} - g^{(k)}, u = \frac{p}{p^T v} - \frac{G^{(k)}v}{v^T G^{(k)} v} .$$

T represents transposed matrix. The problem of this approach is the calculation and storage of approximate Hessian matrix in each iteration.

OSS is a bridge between conjugate gradient algorithm and Quasi-Newton algorithm. OSS doesn't storage the whole Hessian matrix, it assumes that the previous Hessian matrix is unit matrix in each iteration. Another advantage is that the inverse matrix does not need to be computed when the new search direction is calculated (Mukkamala, Janoski, & Sung, 2002).

2.2 Support Vector Machine (SVM)

Support vector machine will transform data into a multi-dimensional feature space F. It is notable that the generalization ability of SVM depends on the geometric features of the training data, not the dimension of the input space. Training SVM will produce a quadratic optimization problem with bounded constraints and linear equality constraints. Vapnik shows training a SVM for pattern recognition how to raise the following quadratic optimization problem(Olabelurin, Kallos, Veluru, & Rajarajan, 2015).

$$\text{Minimize: } W(\alpha) = -\sum_{i=1}^{l}\alpha_i + \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j k(x_i, x_j) , \quad (4)$$

$$\text{subjugate: } \sum_{i=1}^{l} y_i \alpha_i, \forall i : 0 \le \alpha_i \le C . \quad (5)$$

Where l is the number of training samples, α is a vector representing variable l, each α_i corresponds to a training sample (x_i, y_i). The solution of formula (4) is the vector which meet condition (5) and make (4) the minimum value.

2.3 Integration of Soft Computing Techniques

Study found that the optimal linear combination of the neural network algorithm have a good effect (Sutskever, Vinyals, & Le, 2014). When the mean square error is minimized, the ordinary least squares regression coefficient determines the optimal weight. When we have to optimize other error measurements, the problem becomes more complicated. In terms of intrusion detection, our goal is to design a classifier which can give the highest accuracy for each attack pattern. The first step is to construct a different connection model to obtain the optimal generalization performance of the classifier. Test data goes through these models, the corresponding output is recorded. Suppose the classification performance generated by SVM, RP, SCG and OSS are a_n, b_n, c_n and d_n, the corresponding expected value is x_n. Our goal is to integrate a_n, b_n, c_n and d_n, so as to get the optimal output value which can maximum the distinguishing accuracy. We use the following integrated method:

Calculate the absolute value of a single deviation (e.g. $|x_n\text{-}a_n|$). The output value we use is corresponding to $\min(|a_n - x_n|,|b_n - x_n|,|c_n - x_n|,|d_n - x_n|)$. This approach can be represented in Figure 1.

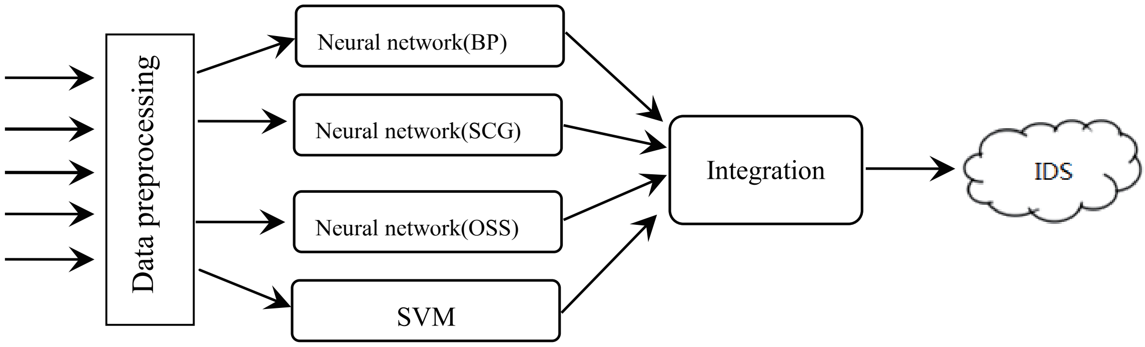

Figure 1
IDS Using Integration of Intelligent Methods

3 EXPERIMENTS

The data used in the experiment came from the Lincoln Laboratory of Massachusetts Institute of Technology. These data are used to evaluate intrusion detection systems by DARPA (Zhang, Lei, Zhou, & Wang, 2015). It is widely regarded as a benchmark for evaluation of intrusion detection systems.

Our experiments were divided into 5 classes, the training and test data set contains 11,980 randomly generated data from these 5 classes. The data volume of each category is commensurate with the size of each class. The normal data belongs to class 1, the probe data belongs to class 2, DoS data belongs to class 3, U2R data belongs to class 4, R2L data belongs to class 5. A random selection of 6,880 of the data sets (11,980) are used o test the different soft computing techniques.

3.1 Experiments Using Neural Networks

Data set consisting of 5,090 training data is divided into five classes: normal, probe, Dos, U2R and R2L. The attack is composed of 22 kinds of specific attack forms in four categories, the rest is normal data. We use two implicit levels in the study, each layer has 20 to 30 neurons, and the network is trained using RP, SCG and OSS algorithms, until the mean square error is reached 0.001.

In the training process, SCG algorithm achieves the goal through 303 iterations, RP algorithm used 66 times and OSS algorithm with 637 times. We use the same test data (6,880), the same network structure and the same activation function to identify the training function which plays a key role in the identification of intrusion. Table 1 shows the results of three different types of networks: network using SCG algorithm has an accuracy of 95.25%; the accuracy of the network using RP algorithm can reach 97.03%; network using OSS algorithm has an accuracy of 93.59%.

Table 1
Performance of Different Neural Network Training Algorithm

Training algorithm	Test 1 iterations	Test 2 iterations	Test 1 accuracy	Test 2 accuracy
RP	68	66	97.03	95.45
SCG	350	303	90.87	95.25
OSS	637	637	93.59	93.59

Table 2
Performance of the Optimal Neural Network Training Algorithm RP

	Normal	Probe	DoS	U2R	R2L	%
Normal	1393	6	1	0	0	99.5
Probe	49	648	3	0	0	92.6
DoS	4	102	4094	2	0	97.4
U2R	1	1	8	13	4	52.0
R2L	0	2	7	21	533	95.0
%	96.3	85.4	99.5	36.1	99.3	

The entry in the upper left corner of the table indicates that the 1,393 "normal" test data is detected as normal; the last column indicates that 99.5% of the "normal" data is correctly detected. In the same way, 648 "probe" data were correctly detected; the last column indicates that 92.6% of the "probe" data is correctly detected. The last line indicates 96.3% of the test data that are considered "normal" are real normal data. The correct rate of the whole classification is 97.03%, the false positive rate is 2.76% and the false negative rate is 0.2%.

3.2 Experiments Using Support Vector Machines

The previously used data set is still used to test the performance of the SVM. 5,090 training data and 6,880 test data used by neural network algorithm are still used to verify the performance of this algorithm.

Because SVM can only carry out binary classification, we need five SVMs, corresponding to five kinds of situations in intrusion detection. We divide the data into two classes, "normal" and "rest". The "rest" represents four types of attack instances in the data set. We repeat this process for all classes. This training uses a radial basis function, a very important point of this function is that it defines the classified feature space of the training set. Table 3 summarizes the results of the experiment:

Table 3
Performance of SVM for 5 Kinds of Behavior

Category	Training time(sec)	Testing time(sec)	Accuracy(%)
Normal	7.67	1.25	99.56
Probe	49.14	2.11	99.71
DoS	22.86	1.93	99.24
U2R	3.37	1.06	99.86
R2L	11.53	1.03	99.77

3.3 Experiments Using Soft Computing Integration

Different soft computing paradigms are carefully constructed to obtain the optimal generalization performance of the classifier. Test data goes through these models, the corresponding output is recorded. Table 4 summarizes the results of the three types of neural networks, SVM and the integration of neural networks and SVM.

Table 4
Performance Comparison of Five Types of Algorithms

Category	SVM accuracy (%)	RP Accuracy (%)	SCG accuracy (%)	OSS accuracy (%)	Integration accuracy (%)
Normal	98.41	99.56	99.57	99.63	99.70
Probe	98.56	92.70	85.57	92.72	99.85
DoS	99.12	97.46	72.02	91.75	99.94
U2R	65	49	0	16	75
R2L	97.32	95.03	98.21	96.79	99.63

CONCLUSION

Our research clearly shows the importance of using the integrated methods to construct intrusion detection system. Integration technology can make the different learning modes cooperate and complement each other. Since many performance measurement methods can be optimized like this, this method can be extended to different fields. The following conclusions are drawn through the experimental results:

(a) This integration method is beyond the single use of SVM or ANN in accuracy. If a suitable soft algorithm is chosen, the integration of them can obtain higher accuracy.

(b) SVM is superior to ANN in the aspects of the expansibility, running time and accuracy predicting.

(c) In the neural network algorithm, the RP algorithm has the best performance with the accuracy rate and the number of iterations.

However, we note that they are not very different in terms of accuracy, the difference in statistics is not very significant either, certain conclusions can only be made after the analysis of more network traffic data.

REFERENCES

An, W., & Liang, M. (2012). A new intrusion detection method based on svm with minimum within-class scatter. *Security & Communication Networks, 6*(9), 1064-1074.

Chow, C. R., & Chu, C. H. (1997). A concurrent training algorithm for supervised learning in artificial neural networks.. *International Journal of Advanced Research in Engineering & Technology, 13*, 267-291.

Ding, S., Li, H., Su, C., Yu, J., & Jin, F. (2013). Evolutionary artificial neural networks: A review. *Artificial Intelligence Review, 39*(3), 251-260.

Huang, Y., Lan, Y., Hoffmann, W. C., & Lacey, R. E. (2011). A pixel-level method for multiple imaging sensor data fusion through artificial neural networks. *Advances in Natural Science, 4*(1), 1-13.

Majumder, M. (2015). *Artificial neural network. Impact of urbanization on water shortage in face of climatic aberrations*. Springer Singapore.

Mukkamala, S., Janoski, G., & Sung, A. (2002). Intrusion detection using neural networks and support vector machines. *Neural Networks, 2002. IJCNN &#039;02. Proceedings of the 2002 International Joint Conference on* (Vol.2, pp.1702-1707). IEEE.

Olabelurin, A., Kallos, G., Veluru, S., & Rajarajan, M. (2015). Multi-agent based framework for time-correlated alert detection of volume attacks. *Lecture Notes in Electrical Engineering, 339*, 499-507.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems, 4*, 3104-3112.

Zhang, L., Lei, J., Zhou, Q., & Wang, Y. (2015). Using genetic algorithm to optimize parameters of support vector machine and its application in material fatigue life prediction. *Advances in Natural Science, 8*(1), 21-26.