https://doi.org/10.31449/inf.v44i2.3195 Informatica **44** (2020) 269–273 **269**

Research on the Detection of Network Intrusion Prevention with SVM Based Optimization Algorithm

Debing Wang
Anhui Vocational & Technical College of Industry & Trade, Huainan, Anhui 232007, China
E-mail: bdaie2@163.com

Guangyu Xu
Anhui University of Science & Technology, Huainan, Anhui 232001, China

Keywords: support vector machine, intrusion prevention, intrusion detection, whale optimization algorithm

Received: June 6, 2020

Support vector machine (SVM) has a good application in intrusion detection, but its performance needs to be further improved. This study mainly analyzed the SVM optimization algorithm. The principle of SVM was introduced firstly, then SVM was improved using the improved whale optimization algorithm (WOA), the improved WOA (IWOA)-SVM based intrusion detection method was analyzed, and finally experiments were carried out on KDD CUP99 to verify the effectiveness of the algorithm. The results showed that the IWAO-SVM algorithm was more accurate in attack detection; compared with SVM, PSO-SVM and ant colony optimization (ACO)-SVM algorithms, the performance of the IWAO-SVM algorithm was better, the detection rate was 99.89%, the precision ratio was 99.92%, the accuracy rate was 99.86%, and the detection time was 192 s, showing that it had high precision in intrusion detection. The experimental results verify the reliability of the IWAO-SVM algorithm, and it can be promoted and applied in the detection of network intrusion prevention.

Povzetek: Algoritem SVM je bil prilagojen za iskanje napadov v omrežjih.

1 Introduction

With the development of technology and the further popularization of computer, the use of network has become more extensive [1], which not only changes the way people study and work, but also creates great values for economic development. However, the network security problem is becoming more and more prominent [2], means of intrusion attack is becoming more complex and diverse [3], which means greater and stronger harms, and the difficulty of intrusion prevention is becoming higher. In order to deal with all kinds of network intrusion, more and more methods have been applied in intrusion detection. Li et al. [4] studied relevance vector machine (RVM), determined the parameters of RVM using the cloud particle swarm optimization algorithm (CPSO), and verified its high accuracy through experiments. Sangeetha et al. [5] designed a method based on application layer signature. If the signature did not match the rule base, the system would generate an alarm. The method could effectively reduce the false alarm rate and improve the accuracy. Kannan et al. [6] designed an enhanced C4.5 for intrusion detection in hybrid virtual cloud environment and verified the effectiveness of the method through the data set and feeding. Geng et al. [7] designed an intrusion detection algorithm based on rough set and Bayes and combining with weighted average and found through experiments that the resource consumption of the method was low and it was easy to realize and had higher efficiency. This study optimized support vector machine (SVM), applied it to the detection of network intrusion, carried out an experiment on the data set, and compared the performance of different SVM optimization algorithms to verify the effectiveness of the designed optimization algorithm, which provides some theoretical bases for its further application in the actual network and offers more ideas for the design of intrusion detection methods.

2 Network intrusion prevention detection

Network intrusion refers to the behavior of trying to access or destroy a system without authorization to make it unavailable [8]. Detection of network intrusion is to analyze the key information collected from the inside and outside of the computer, such as security log, etc. [9], find out the characteristics that may generate attacks [10], and give responses such as alarm and network outage [11], and its flow is shown in Figure 1.

Firstly, multiple monitoring points are set in the network to collect data such as system log, firewall log, software information and intrusion information as much as possible and comprehensively to ensure the detection effect. Then, the collected data are normalized to reduce the detection error, and the processed data are analyzed

Figure 1: The detection process.

using detection methods to obtain the detection results. Finally, the system makes response to defend according to the detection results.

3 Detection method combined with SVM optimization algorithm

3.1 SVM algorithm

SVM is a machine learning algorithm [12], which has advantages of strong generalization ability, learning ability and applicability. Its classification idea is that two separate categories are on both sides of the hyperplane and have as large an interval as possible (Figure 2).

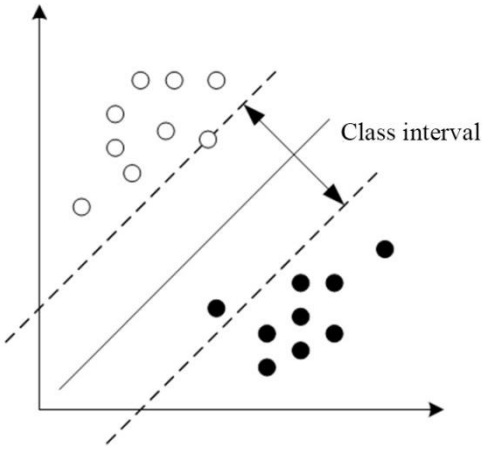

Figure 2: The principle of SVM.

If there is a dataset, $(x_1, y_1), (x_2, y_2), \cdots, (x_i, y_i), i = 1,2,\cdots,n, x \in R$, $y \in \{+1,-1\}$, and the hyperplane of its classification can be written as: $y = wx_i + b$, where w stands for weight and b stands for the threshold value. To find the optimal classification plane, the constraints can be written as:

$$\min \frac{\|w\|^2}{2}$$
$$s.t. y_i(wx_i + b) \geq 1 \text{ ,(1)}$$

In order to improve the modeling speed, slack variable λ is introduced, then:

$$\min \frac{\|w\|^2}{2} + C\sum_{i=1}^{n} \lambda_i$$
$$s.t. y_i(wx_i + b) \geq 1 - \lambda_i, \lambda_i \geq 0 \text{ ,(2)}$$

where C refers to the penalty factor, and then the Lagrange method is introduced to transform it into a dual problem:

$$\max \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i,j=1}^{n} a_i a_j y_i y_j k(x_i, x_j) \text{ ,(3)}$$

where a_i is a Lagrange multiplier and $k(x_i, x_j)$ is a kernel function. The constraint is $\sum_{i=1}^{n} a_i y_i = 0, 0 \leq a_i \leq C$. The final classification function can be written as:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n} a_i y_i k(x_i, x_j) + b\right) \text{ .(4)}$$

The kernel function used in this study is RBF kernel function, and the formula is as follows:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{r^2}\right) \text{ ,(5)}$$

where r is a nuclear parameter.

3.2 SVM optimization algorithm

In SVM, penalty factor C and nuclear parameter r has a great impact on the final allocation performance. In order to be able to get optimal values of C and r , the whale optimization algorithm (WOA) [13] was used to obtain the optimal value of parameters in this study, and SVM was optimized. WOA is an optimization algorithm based on the simulation of whale hunting behavior. It is easy to operate and implement, but it also has the problem of slow convergence speed. Therefore, inertia weight σ was introduced to obtain an improved WOA (IWOA).

Suppose that the population size of whales is N , the position of the i -th whale in the d -th space is $X_i = (x_i^1, x_i^2, \cdots, x_i^d)$, $i = 1,2,\cdots,N$, and the position of the prey of whale is the optimal solution of problem. In the process of surrounding prey, the formula of the position updating of whale can be written as:

$$X(t+1) = \sigma X_i(t) - A|CX_i(t) - X(t)| \text{ , (6)}$$

where t stands for the times of iterations, σ is an inertia weight, $\sigma = \sigma_{\max} - (\sigma_{\max} - \sigma_{\min})\left(\frac{t}{t_{\max}}\right)^{1/t}$, A and C are coefficient vectors, $A = 2ar_1 - a$, $C = 2r_2$, $a = 2 - \frac{2t}{t_{\max}}$, and r_1 and r_1 are random quantity in $[0,1]$.

The hunting strategy of whales is called bubble-net [14], which means generating bubbles through the spiral path to surround the prey. This process can be expressed as follows:

$$X(t+1) = D' e^{bz} \cos(2\pi z) + \sigma X_i(t) \text{ , (7)}$$

where $D' = |X_i(t) - X(t)|$, b stands for the constant defining the spiral shape, and z is a random number in $[-1,1]$.

In addition to bubble-net, whales also conduct random search, which can be expressed as:

$$X(t+1) = \sigma X_{rand}(t) - A|CX_{rand}(t) - X(t)| \text{ , (8)}$$

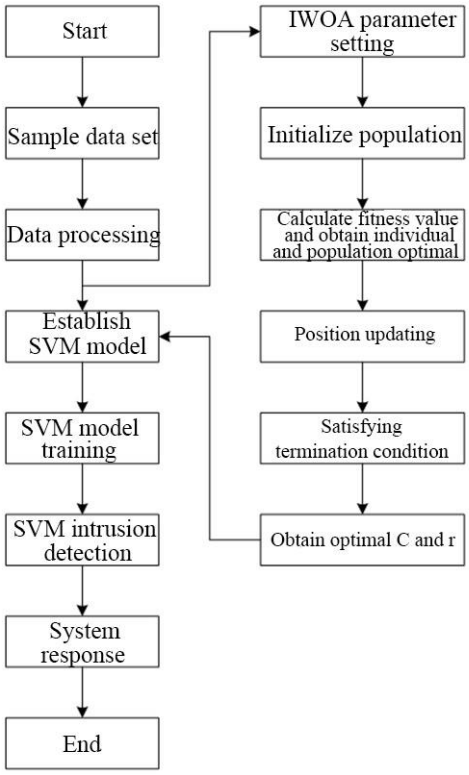

Figure 3: The flow of the intrusion detection algorithm.

where X_{rand} refers to a randomly selected whale position.

3.3 IWOA-SVM intrusion detection algorithm

After optimization with IWOA, the flow of the IWOA-SVM algorithm is shown in Figure 3.

The specific steps of the algorithm are as follows. For the collected sample data set, after preprocessing, the parameters of IWOA are set, and parameter C and r which need optimization in SVM are taken as whale individuals. The population is initialized, and then the fitness value of the individual is calculated to obtain the optimal value of the individual and population. Then, the location is updated by IWOA to obtain new solutions until the termination conditions are met, and optimal value C and r are obtained and regarded as the parameters of SVM. The SVM model is established. After training, the model is tested using the testing samples. Finally, the system responds according to the test results.

4 Experimental results

4.1 Experimental environment and data set

The experiment was carried out on Linux operating system, with Intel Core i7 CPU@2.40GHz, 8 GB memory, and Python language. The size of the IWOA population was 20, $t_{\max}$ was 50, $\sigma_{\min}$ was 0.3, and $\sigma_{\max}$ was 0.9. The experimental data set was KDD CUP99, including probe, DOS, U2R and R2L in addition to Normal. As KDD CUP99 is too large, only a part of data was randomly selected in this study. There were 3500 normal data, 8260 attack data in the training set; there were 1500 normal data and 3540 attack data in the testing set, as shown in Table 1.

Category	Training set	Testing set
Normal	3500	1500
Probe	2100	900
DOS	4900	2100
U2R	700	300
R2L	560	240

Table 1: Experimental data set.

		Detection result	
		Attack data	Normal data
Actual condition	Attack data	A	B
	Normal data	C	D

Table 2: Confusion matrix.

	Normal	Probe	DOS	U2R	R2L
Normal	1497	1	2	0	0
Probe	0	899	1	0	0
DOS	0	1	2198	1	0
U2R	0	0	1	299	0
R2L	0	0	0	0	240

Table 3: Confusion matrix results of the IWAO-SVM algorithm.

4.2 Evaluation index

The detection algorithm was evaluated using the confusion matrix, as shown in Table 2.

In Table 2, A represents that attack data is correctly judged as attack data; B represents that normal data is misjudged as attack data, C represents that attack data is misjudged as normal data, and D represents that normal data is correctly judged as normal data.

(1) Detection rate = A/(A+B)

(2) Precision ratio = A/(A+C)

(3) Accuracy = (A+D) / (A+B+C+D)

4.3 Experimental results

In order to verify the detection effect of the IWOA-SVM algorithm, it was compared with SVM, particle swarm optimization-SVM (PSO-SVM) [15] and ant colony optimization-SVM (ACO-SVM) algorithms [16]. The confusion matrix result of the IWAO-SVM algorithm is shown in Table 3, and the result comparison between different algorithms is shown in Table 4.

The four numbers separated by slashes in Table 4 represent the results of SVM, PSO-SVM, ACO-SVM and IWOA-SVM algorithms respectively. According to the

			Detection result	
			Attack data	Normal data
Actual condition	Attack data	3540	3320/3478/3486/3536	220/62/54/4
	Normal data	1500	134/38/12/3	1366/1462/1488/1497

Table 4: Comparison results of different algorithms.

	SVM	PSO-SVM	ACO-SVM	IWAO-SVM
Detection time/s	189	197	198	192

Table 5: Comparison of testing time.

data in Table 4, the detection rate of the algorithms was calculated, and the results are shown in Figure 4.

According to Figure 4, first of all, the detection rate of the PSO, ACO and IWAO optimized SVM algorithms was 6.21%, 8.71% and 14.88% higher than that of SVM, respectively. It was seen that the detection rate of the IWAO-SVM algorithm significantly improved; the precision ratio of the four algorithms were all over 90%, of which the IWAO-SVM algorithm was the highest, 99.92%; from the perspective of accuracy rate, the optimization by PSO and ACO improved the accuracy rate of the SVM algorithm, but not as significant as IWAO; the accuracy of the IWAO-SVM algorithm was 15.51% higher than that of the SVM algorithm.

The detection time of different algorithms was compared, and the results are shown in Table 5.

It was seen from Table 5 that the time complexity of the SVM optimization algorithms increased compared with the SVM algorithm, the detection time of the PSO-SVM algorithm increased by 4.23% compared with the SVM algorithm, the detection time of the ACO-SVM algorithm increased by 4.76%, and the detection time of the IWAO-SVM algorithm only increased by 1.59%, 2.54% lower than the PSO-SVM algorithm and 3.03% lower than the ACO-SVM algorithm, which showed that the optimization algorithm designed in this study not only had obvious advantages in the detection rate, but also had a good performance in the detection time, i.e., it could provide more excellent service for network intrusion detection.

5 Discussion

It is very important for network protection and control to detect intrusion attacks effectively [17]. In the network intrusion detection, clustering algorithm [18], Apriori algorithm, decision tree [19], Q-learning, neural network [20] and hidden Markov [21] have a wide range of applications. This study mainly analyzed SVM. As a common classification and prediction algorithm, SVM has a good application in many fields, such as face recognition [22], risk assessment [23], electricity price prediction [24] and image classification [25].

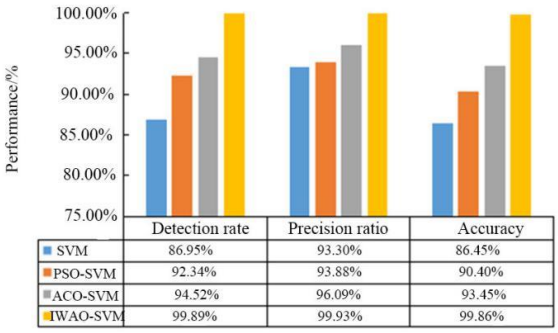

Figure 4: Comparison of the performance between different algorithms.

In order to improve the effectiveness of SVM in intrusion detection, it was optimized by the WAO algorithm in this study, and then it was verified by KDD CUP99 data set. It was seen from Table 3 that the IWAO-SVM algorithm had excellent accuracy in the classification of intrusion attacks, and only seven data were wrongly classified. Then, it was seen from Table 4 and Figure 4 that the IWAO-SVM algorithm had a better detection performance, with the detection rate reaching 99.89%, 14.88%, 8.18% and 5.68% higher than the other three algorithms respectively; the precision ratio improved by 7.10 %, 6.43% and 3.93% respectively; the accuracy increased by 13.41%, 10.64% and 6.86% respectively, which verified the effectiveness of IWAO in SVM optimization and the good precision of the IWAO-SVM algorithm in the intrusion detection. Finally, the comparison of the detection time showed that the method proposed in this study had a good advantage in time compared to the other optimization algorithms, only 1.59% longer than the SVM algorithm.

Although some achievements have been made in the research of network intrusion prevention and detection, there are still some shortcomings that need to be solved in the future work:

(1) the detection effect of the SVM algorithm should be compared when choosing different kernel functions;
(2) the performance of more optimization algorithms in SVM should be compared;
(3) the performance of the IWAO-SVM algorithm in practical application should be studied.

6 Conclusion

Aiming at the detection of network intrusion prevention, this study analyzed the optimization of SVM, designed an improved WAO algorithm, and compared it with other optimization algorithms on the data set. The results suggested that:

(1) the IWAO-SVM algorithm could detect intrusion attacks accurately;
(2) the detection rate of the IWAO-SVM algorithm was 99.89%, the precision ratio was 99.92%, and the accuracy rate was 99.86%, which were all higher than the other excellent algorithms;
(3) the detection time of the IWAO algorithm was 192s , only 1.59% longer than the SVM algorithm.

7 References

[1] Elekar KS (2015). Combination of data mining techniques for intrusion detection system. International Conference on Computer. IEEE.

[2] Shah AA, Khiyal MSH, Awan MD (2015). Analysis of Machine Learning Techniques for Intrusion Detection System: A Review. International Journal of Computer Applications, 119(3), pp. 19-29.

[3] Keegan N, Ji S Y, Chaudhary A, Concolato C, Yu B, Jeong DH (2016). A survey of cloud-based network intrusion detection analysis. Human-centric Computing and Information Sciences, 6(1), pp. 19.

[4] Li GD, Hu JP, Xia KW (2015). Intrusion detection using relevance vector machine based on cloud particle swarm optimization. Control & Decision, 30(4), pp. 698-702.

[5] Sangeetha S, Devi BG, Ramya R, Dharani MK, Sathya P (2015). Signature Based Semantic Intrusion Detection System on Cloud. Advances in Intelligent Systems and Computing, 339, pp. 657-666.

[6] Kannan A, Venkatesan KG, Stagkopoulou A, Li S (2015). A Novel Cloud Intrusion Detection System Using Feature Selection and Classification. International Journal of Intelligent Information Technologies, 11(4), pp. 1-15.

[7] Geng X, Li Q, Ye D, Wu Z, Jiang Y (2017). Intrusion detection algorithm based on rough weightily averaged one-dependence estimators. Journal of Nanjing University of Science & Technology, 41(4), pp. 420-427.

[8] Milliken M, Bi Y, Galway L, Hawe GI (2015). Ensemble learning utilising feature pairings for intrusion detection. World Congress on Internet Security. IEEE.

[9] Ghosh P, Mandal AK, Kumar R (2015). An Efficient Cloud Network Intrusion Detection System. Advances in Intelligent Systems & Computing, 339, pp. 91-99.

[10] Jinny SV, Kumari JJ (2015). Encrusted CRF in Intrusion Detection System. Advances in Intelligent Systems & Computing, 325, pp. 605-613.

[11] Tedesco G, Aickelin U (2016). Adaptive Alert Throttling for Intrusion Detection Systems. Social Science Electronic Publishing, 730, pp. 194-201.

[12] Abdiansah A, Wardoyo R (2015). Time complexity analysis of support vector machines (SVM) in LibSVM. International Journal of Computer Applications, 128(3), pp. 975-8887.

[13] Aljarah I, Faris H, Mirjalili S (2016). Optimizing connection weights in neural networks using the whale optimization algorithm. Soft Computing, 22(1), pp. 1-15.

[14] Friedlaender A, Weinrich M, Bocconcelli A, et al (2011). Underwater components of humpback whale bubble-net feeding behaviour. Behaviour, 148(5), pp. 575-602.

[15] Wang L, Dong C, Hu J, Li G (2015). Network Intrusion Detection Using Support Vector Machine Based on Particle Swarm Optimization. Plant Biotechnology Reports, 4(3), pp. 237-242.

[16] Zan P, Ai YT, Zhao J, Shao Y (2014). A Prediction Model of Rectum's Perceptive Function Reconstruction Based on SVM Optimized by ACO. 461, pp. 121-128.

[17] Deng S, Zhou A, Yue D, Hu B, Zhu L (2017). Distributed intrusion detection based on hybrid gene expression programming and cloud computing in cyber physical power system. IET Control Theory and Applications, 11(11), pp. 1822-1829.

[18] Chahal JK, Kaur A (2016). A Hybrid Approach based on Classification and Clustering for Intrusion Detection System. International Journal of Mathematical Sciences & Computing, 2(4), pp. 34-40.

[19] Modinat M, Abimbola A, Abdullateef B, Opeyemi A (2015). Gain Ratio and Decision Tree Classifier for Intrusion Detection. International Journal of Computer Applications, 126(1), pp. 975-8887.

[20] Gautam SK, Om H (2016). Computational Neural Network Regression Model for Host based Intrusion Detection System. Perspectives in Science, 8(C), pp. 93-95.

[21] Sharma SK, Manoria M (2015). Intrusion Detection using Hidden Markov Model. International Journal of Computer Applications, 115(4), pp. 35-38.

[22] Prakash N, Singh Y (2015). Fuzzy Support Vector Machines for Face Recognition: A Review. Maropoulos P G, 131(3), pp. 24-26.

[23] Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016). Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides, 13(2), pp. 361-378.

[24] Shrivastava NA, Khosravi A, Panigrahi BK (2015). Prediction Interval Estimation of Electricity Prices Using PSO-Tuned Support Vector Machines. Industrial Informatics, IEEE Transactions on, 11(2), pp. 322-331.

[25] Tan K, Zhang J, Du Q, Wang X (2015). GPU Parallel Implementation of Support Vector Machines for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 8(10), pp. 1-10.