ACO and SVM Selection of Network Intrusion Detection Method

Jianjun WU
Hunan City University, Yiyang, Hunan, 413000, China

Abstract: Feature selection and classifier design is the key to network intrusion detection. In order to improve network intrusion detection rate for feature selection problem, this paper proposed a network intrusion detection method (ACO-FS -SVM) combining ant colony algorithm to select the features with a feature weighting SVM. First, the use of support vector machine classification accuracy and feature subset dimension construct a comprehensive fitness weighting index. Then use the ant colony algorithm for global optimization and multiple search capabilities to achieve optimal solutions feature subset search feature. And then selected the key feature of network data and calculated information gain access to various features weights and heavy weights to build support vector machine classifier based on the characteristics of network attacks right. At last, refine the final design of the local search methods to make the feature selection results without redundant features while improve the convergence resistance, and verify the data set by KDD1999 effectiveness of the algorithm. The results show that ACO-FS-SVM can effectively reduce the dimension of features, and have improved network intrusion detection accuracy and detection speed.

Keywords: Feature Selection; Feature Weighting; Ant Colony Optimization Algorithm; Support Vector Machines; Network Intrusion Detection

1. Introduction

In recent years, the Internet scale growing, coupled with its open, non-executives and undefended, the complex network intrusions, the number of intrusions and the growing degree of harm, network intrusion detection has been the focus of network security defense research[1]. With the rapid development of information technology, government departments, research institutions, enterprises, business organizations' dependence of information system is growing, and threats facing information security is also increasing. Traditional information security technology has been unable to meet the requirements of modern information security. Information security situation assessment (information Security Situation Awareness, ISSA) came into being. On the basis of the integration of information security elements in macroscopic real-time assess the security situation of information, and it can estimate the development trend [2].

Network intrusion detection is a pattern recognition classification problems, including feature selection, classifier selection and optimization modules. Network data is very complex, with a high dimensional feature. The feature set contains some redundant features and useless features, which will increase the model training time and computational complexity, and have a negative impact intrusion detection results[3]. To this end, before network intrusion detection modeling, it often used the feature selection algorithm to select beneficial feature subset of test results, in order to reduce the feature dimension. There are main sequence search algorithm, based on principal component analysis, genetic algorithms, particle swarm optimization and other feature selection method [3,4]. In addition to the feature subset selection, network intrusion detection results are also closely related to the classifier and parameter. The current network intrusion detection model is mainly Bayesian networks, neural networks and support vector machines and other non-linear classification algorithm[5]. ACO-FS-SVM network intrusion detection flow ACO-FS-SVM algorithm use wrapper feature to select model. It uses the automatic optimization capability ACO to global search in the feature space, gets a different combination of features. According to the results of SVM classification to determine the features of the combination of classification performance, and constantly update the selected feature sets, until the search results to obtain the best classification feature combination. First extract the characteristics of network status information, and then send into the ACO-FS-SVM classification feature selection module to select the best feature set for network intrusion detection. ACO-FS-SVM intrusion detection network flow model is shown in Figure 1.

2. ACO-FS-SVM Network Intrusion Detection Method

Ant Colony Optimization (ACO) is a collective intelligence algorithms, simulated ants foraging information exchange and mutual cooperation, with positive feedback, global search ability and distributed computing, etc. The

time of the proposed algorithm is although late, but developed rapidly. It has been widely used in solving the TSP, job-shop scheduling, network route(QOS), knapsack and other aspects. And the simulation results show that ant colony algorithm has good results[18,19]. In network intrusion detection feature selection, the need for network intrusion detection features to be accessed as a place of ants, which will be converted feature optimization problem into a path search problem.

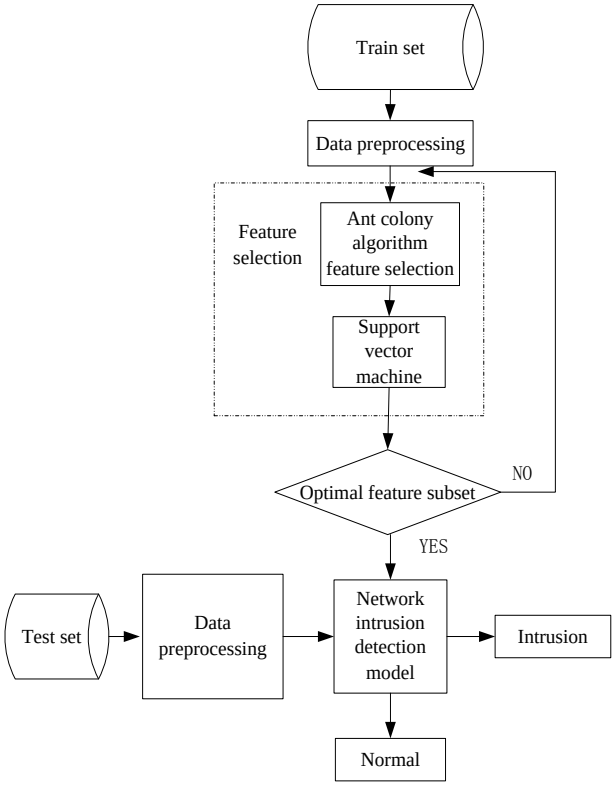


Figure 1. The Working Flow of ACO-FS-SVM Network Intrusion Detection Model

2.1. The Fitness Function Establishment

Network intrusion feature selection include two aspects: ①select features subset to make the network attack detection accuracy rate higher. ②the feature dimension as far as possible smallest. But in fact a contradiction between the two. In order to make the balance, this research of fitness function is defined as:

$$f(s) = \lambda P_{error} + (1-\lambda)\frac{d}{D} \tag{1}$$

Where, d is the dimension of feature selection subset s. D is the dimension of candidate feature set. P_{error} is classification error rate. λ is the weighting coefficient of classification error rate.

The computational formula of weighting coefficient λ is:

$$\lambda = \frac{100}{100 + D \times x} \tag{2}$$

Where, x presents the percentage of network intrusion detection reduced error rate when features increase one dimension(x%).

2.2. Feature Weighting SVM

Use feature weighting kernel function constructed SVM, we call feature weighting SVM. Weighting function is weighted for one feature. The kernel function K_p is defined as:

$$k_p(x_i, x_j) = k(x_i^T P, x_j^T P) \tag{3}$$

P is called feature weighting matrix, where P_{ii} = ω_i (1< i <n) means the weighting of i_{th} feature. Generally speaking, ω_i is not all equal. If an ω_i =0, represents the k_{th} feature has nothing to do with the output of the classifier.

$$\begin{aligned} k_p(x_i, x_j) &= \exp(-\frac{\left\| x_i^T P - x_j^T P \right\|^2}{\sigma^2}) \\ &= \exp(-\frac{((x_i - x_j)^T P P^T (x_i - x_j))}{\sigma^2}) \end{aligned} \tag{4}$$

Feature weighting support vector machine(FS-SVM) algorithm is described as follows.

$$\begin{aligned} &\min_{\omega,b,\xi} \frac{1}{2}\omega T\omega + C\sum_{i=1}^{l}\xi_i \\ &s.t. \\ &y_i(\omega^T\varphi(x_i)+b) \geq 1-\xi_i \\ &\xi_i \geq 0, i = 1,2,\cdots,l \end{aligned} \tag{5}$$

Where, C>0 is a penalty parameter.

FS-SVM can be described as a quadratic programming problem. In the case of only the minimum required points, we can use the Lagrange multiplier method for solving minimum ω_i :

$$\begin{aligned} L(\omega,\xi_i,b,\alpha_i) = \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{l}\xi_i - \\ \sum_{i=1}^{l}\alpha_i(y_i(\omega^T\varphi(x_i)+b)-1+\xi_i) \end{aligned} \tag{6}$$

Where, α_i is lagrange multipliers. Partial derivative ω, b, ξ respectively, and make them equal 0.

$$\begin{cases} \frac{\partial L}{\partial \omega} = \omega - \sum_{i=0}^{l}\alpha_i y_i \varphi(x_i) = 0 \\ \frac{\partial L}{\partial b} = -\sum_{i=0}^{l}\alpha_i y_i = 0 \\ \frac{\partial L}{\partial \xi_i} = Cs_i - \alpha_i = 0 \end{cases} \tag{7}$$

Put formula(13) into formula(12), get the dual problem of formula(11) :

$$\min_{a} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j K(Px_i, Px_j) - \sum_{j=1}^{l}\alpha_j$$
$$s.t. \qquad (8)$$
$$\sum_{i=1}^{l} y_i \alpha_i = 0; 0 \le \alpha_i \le C; i = 1, 2, \cdots, l$$

Solve the dual problem to get optimal decision function

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{l} \alpha_i y_i K(Px_i, Px_j) + b) \qquad (9)$$

Network intrusion classification FW-SVM algorithm is as follows:

Step1: Collect network data training sample set {(x_1 , y_1),…,(x_i , y_i)}, x_i =(x_{1i} , x_{2i} ,…, x_{di}) is a d-dimensional vector, $y_i \in$(+1,-1), i =(1,…, l) .

Step2: Select high-impact features by ACO.

Step3: Calculate weight value w of each feature method based on the information gain method, and construct its feature vector β = diag (β_1 , β_2 ,…, β_n)T .

Step4: Select the appropriate penalty parameter C > 0, according to formula (14) constructed and solved quadratic programming optimization problem, obtain the optimal solution for α=(α_1 ,…, α_l)T.

Step5: α_j (0< α_j <C) is a component of the α. (x_i , x_j) is its corresponding sample points. Calculated b, construct optimal classification Hyper plane ($\omega \cdot x$)+b=0, get network intrusion classification decision function: f(x)=sgn(($\omega \cdot x$)+b).

2.3. Determination of Ants State Transition Probability

Feature is the nodes that each ant must go through. Each complete a cycle, an ant traverse all features. Each feature has a probability of selection, each of the ants through a feature node according to the selected characteristics of the probability to determine whether features are selected. Ants use feature selection probability, the greater the probability of selection feature , the greater the likelihood of being selected. The probability of ants from feature i to j is:

$$p_{ij}^{k}(t) = \begin{cases} \dfrac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum \tau_{is}^{\alpha}(t)\eta_{is}^{\beta}(t)} & j, s \notin tabu_k \\ 0 & otherwise \end{cases} \qquad (10)$$

where, η_{ij} is inspiring factor determined by the intrusion detection accuracy. The larger η_{ij} is, the greater the ants move to feature j. $\tau_{ij}(t)$ is the pheromone from feature i to feature $N_{\max}$ at t time. Tabuk is the tabu list of ant k .

In the state transition probability, α represents the weighting of pheromone, and β represents the weighting of inspiration factor. According to reference [5], α in this study is a constant, β is determined by the formula (11).

$$\beta = \beta_0 \left(1 - \frac{n}{N_{\max}}\right) \qquad (11)$$

Where, n is iterations. β_0 is the initial value of inspiration factor weighting. $N_{\max}$ is the maximum iteration.

2.4. Local Refine the Search Process

After k sub-set of ants important feature search, has got k important feature. In order to prevent some features of the network intrusion detection irrelevant or redundant features retained in the feature subset, search for the optimal feature u_j in k feature to meet:

$$F(S_j) = \min_i (S_i, \forall i) \qquad (12)$$

where U is for any feature subset u_i , denote $S_i = S_m \cup u_i \cup \{ f_n \}$.

2.5. The Ants Aearch Termination Condition

Under normal circumstances it is difficult to determine the dimensions of optimal feature subset. The study of ants search termination conditions is 3 consecutive increasing features, F(s) does not happen too much change, said the current round of search termination.

3. Simulation Test

3.1. Data Sources

The experimental data selected KDD CUP 99 data sets, and the data sets of data contained 41-dimensional features, 34 numeric fields and seven symbolic field. 41-dimensional features can be divided into four parts: the basic features of TCP connections (1 ~ 9 No. features), content features of TC connections(10 to 22 No. features) , network traffic statistics feature based on time(23 to 31 No. features), network traffic statistics feature based on host (32 to 41 No. features) . The data set divided into four types of intrusions: Probe (scanning and detection), DOS (denial of service attack), U2R (unauthorized access to local super user) and R2L (unauthorized remote access), see Table 1. In P4 dual-core 2.8G CPU, 1G RAM, Windows XP operating system for the simulation environment, using VC ++ 6.0 algorithm.

Table 1. THE 41 Data Features of Network Connection

Basic features	Connections content features	Network traffic statistics feature based on time	Network traffic statistics feature based on host
1. duration C	10. hot C	23. count C	32. dst_host_count C
2. protocol_type D	11. num_failed_logins C	24. serror_rate C	33. dst_host_srv_count C
3.service D	12. logged_in D	25. rerror_rate C	34. dst_host_same_srv_rate C
4. src_bytes	13. num_compromised C	26. same_srv_rate C	35. dst_host_diff_srv_rate C
5. dst_bytes C	14. root_shell D	27. diff_srv_rate C	36. dst_host_same_src_port_rate C
6. flagD	15. su_attempted D	28. srv_count C	37. dst_host_srv_diff_host_rate C
7. land D	16. num_root C	29. srv_serror_rate C	38. dst_host_serror_rate C
8. wrong_fragment C	17. num_file_creations C	30. srv_rerror_rate C	39. dst_host_srv_serror_rate C
9. urgent C	18. num_shells C	31. srv_diff_host_rate C	40. dst_host_rerror_rate C
	19. num_access_files C		41. dst_host_srv_rerror_rate C
	20. num_outbound_cmds C		
	21. is_hot_login D		
	22. is_guest_login D		

3.2. Results and Analysis

3.2.1 Performance Comparison before and after Feature Selection

(1) Because the original data set is too large, a small portion of this was chosen as the experimental data of the data set. The training set of randomly selected 5000 and 1000 test set, and they are normalized characteristics, reduced into [0, 1] range.
(2) The training set is input to the SVM for training and test sets for testing. Feature selection before the intrusion detection results are obtained.
(3) The training set is input to the SVM. Use ACO algorithm combining SVM feature to select. Optimal features as shown in Table 2.
(4) According to the results of the step (3) feature selection for screening training set and testing set.
(5) The step (4) obtained the training set put into the SVM for training and test sets for testing, intrusion detection result obtained after the feature selection.
(6) Comparison the detection result and the running time of step (2) and step (5) .

Table 2 . ACO Feature Selection

intrusion type	feature subset
probe	2，4，9，21，29，32，33，34，35
Dos	2，3，7，9，16，20，27，32，37，40
U2R	2，4，9，20，31，21，29，32，33，34，35
R2L	1，2，3，4，6，7，9，11，16，20，21，23，27
Normal	2，3，4，7，8，9，10，15，16，21，22，23，25

Using 5 experiments to take the average of the test results. It can obtain intrusion detection rates before and after feature selection are shown in Table 3, the running time shown in Table 4. As apparent from Table 3, the average rate of intrusion detection after feature selection network increased by 3.10% . The results show, feature selection can be more accurately depicts the network status change information, eliminated redundant and useless information. Intrusion detection performance of feature selection improved significantly.

Table 3 . Comparative the Average Rate of A Intrusion Detection before and after Feature Selection (%)

Intrusion type	Original features	After selection features	Increasing value
probe	89.76	93.27	3.51
Dos	92.71	95.49	2.78
U2R	87.8	92.53	4.73
R2L	87.48	90.85	3.37
Normal	92.12	93.25	1.13
Average value	89.97	93.07	3.10

From Table 4, the feature selection greatly reduced runtime network intrusion detection model. It indicates that the network through ACO feature selection can be achieved by a number of key features, which eliminate unwanted features, reduce the number of SVM input dimension and computing time, speed up the detection speed. Network feature selection to meet real-time requirements can be more network intrusion detection.

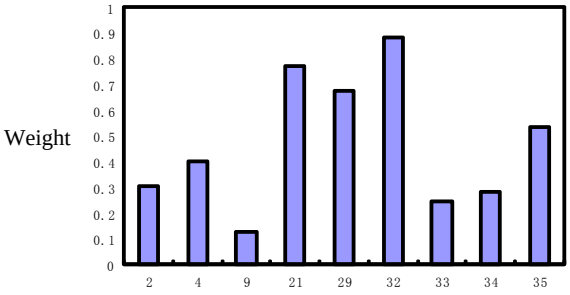
Table 4 . Comparison Running Time (ms) before and after Feature Selection

Intrusion type	Original features	After selection feature	Reduction values

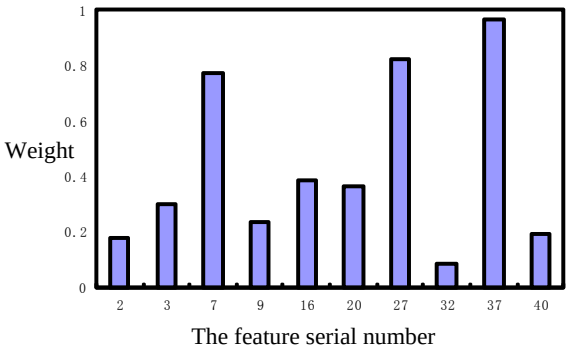
Probe	32.00	23.88	8.12
Dos	23.51	17.99	5.52
U2R	7.78	4.99	2.79
R2L	5.27	3.84	1.43
Normal	10.55	6.89	3.66

3.2.2 Network Intrusion Performance Comparison before and after Feature Weighting

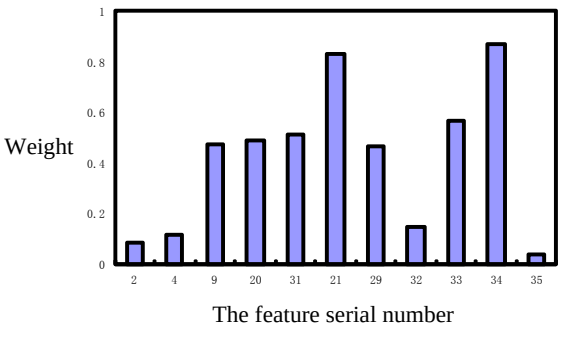
Firstly, use information gain method to calculate the weight of each feature, the results shown in Figure 3. And then deal with the weight of the feature and build weighted support vector machine classifier network intrusion, get test results are shown in Table 5.



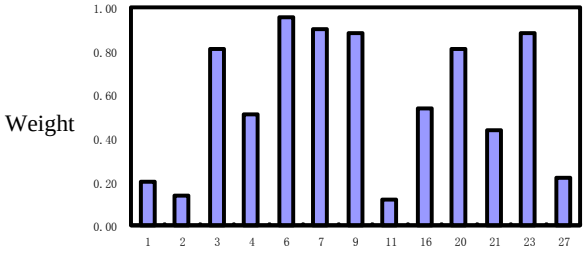
(a) The Probe Feature Subset Weight Distribution



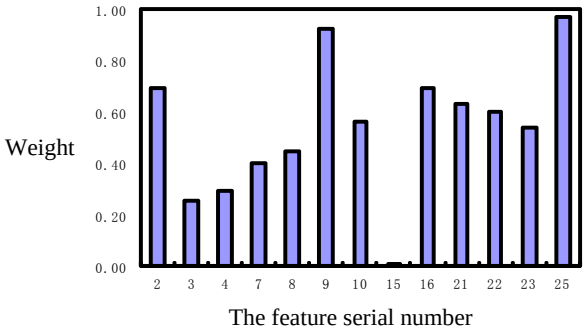
(b) The DOS Feature Subset Weight Distribution



(c) The U2R Feature Subset Weight Distribution



(d) The R2L Feature Subset Weight Distribution3



(e) The Normal Feature Subset Weight Distribution

Figure 2. All Kinds of Invasion of the Feature Subset of Weight Distribution

From Table 5, under the same experimental conditions, the network intrusion detection methods ACO-FS-SVM, both in terms of time efficiency or network intrusion detection rates are higher than the original SVM method. Intrusion detection reached for Normal of 99.13%, which was mainly due to the increased value of the classification affect more significant characteristic quantities. The original in the network data samples near the classification surface is misclassified been corrected, and the original classified correct network data sample basically did not happen change. Therefore, by the weighting network intrusion of all kinds of intrusion detection accuracy have varying degrees increase.

Table 5 . The Intrusion Detection Rate(%) Comparison before and after Weighting

invasion type	before weighting (ACO- SVM)	after weighting (ACO-FS-SVM)	increasing value
probe	93.27	98.46	5.19
Dos	95.49	97.09	1.6
U2R	92.53	98.68	6.15
R2L	90.85	98.56	7.71

Normal	93.25	99.13	5.88
average value	93.07	98.38	5.30

Because support vector (SV) can fully characterize the training dataset, the division of the set of SV equivalent to the division of the entire training data set. In order to further illustrate the effectiveness and advantages of ACO-FS-SVM, further comparative analysis of support vector set before and after the feature weighting, the results shown in Figure 3.

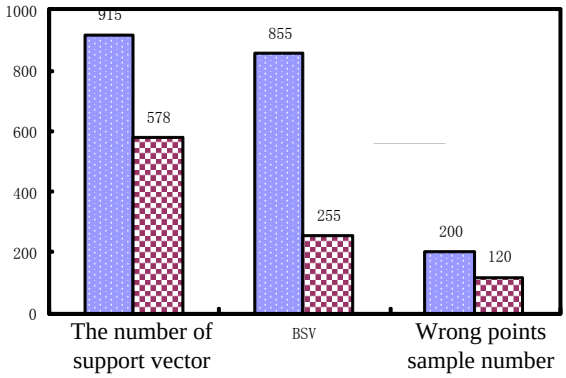


Figure 3. Support Vector Set Comparison before and after Feature Weighted

Figure 3 shows that after weighting, the total number of support vectors dropped from 915 to 578. It indicates that ACO-FS-SVM generalization is better, border support vector (bounded support vector, BSV) number dropped from 855 to 255, decreased by70.17%. Wrong points sample also declined dramatically. The result showed that before weighting, the data samples at the network boundary hyper plane are many. By weighting processing, significantly reducing the data samples in the classification of these networks polygon boundaries. It indicates that ACO -FS-SVM on feature weighted, can increase network intrusion detection.

4. Conclusion

AS for the current network intrusion detection problem, this paper put forward an ant colony algorithm to select and feature weighting fusion of support vector machine network intrusion detection methods. By optimizing the feature selection, weighting processing improve the detection affect of support vector machine classifier for network intrusion. Use the simulation experiments to verify its validity. Simulation results show that, ACO-FS-SVM can select and test results associated with a higher degree of feature subset, effectively reduce the dimension of features and improve the efficiency of network intrusion detection and detection accuracy.

For today's complex network robustness evaluation model only consider the robustness of the network topology and the defect of local effect of the failed node, so this paper proposes complex network function evaluation algorithm based on node efficiency. The algorithm overall consider the global influence of node failure, and use the efficiency of the node on network to define the load of each node, Limit load and failure model, with the rate of striking the ultimate failure nodes on network to measure the functionality of the network, the result of robustness experiment proofs: the algorithm is suitable for assessing the robustness of large-scale and small-world network function, the complexity of algorithm time is O (n2).

References

[1] Andrew R,Peters G P, Lennox J. Approximation and regional aggregation in multi-regional input-output analysis for national carbon footprint accounting . Economic Systems Research, 2009, 21 (3): 311-335.

[2] Boyd J P,Fitzgerald W J,Mahutga M C,Smith D A. Computing continuous core/periphery structures for social relations data with MINRES/SVD, Social Networks, 2010, 32 (2): 125-137.

[3] Dietzenbacher E. More on Multipliers. Journal of Regional Science, 2005, 45 (2): 421-426,

[4] Holub H W,Schnabl H. Qualitative input-output analysis and structural information. Economic Modelling, 2012, 2(1): 67-73.

[5] Holub H W? Tappeiner G. A general qualitative technique for the comparison of economic structures. Quality & Quantity, 2010, 22 (3): 293-310.