Acta Technica 62 (2017), No. 6B, 469–478

Low energy consumption intrusion detection algorithm based on differential evolution strategy

Xingzhu Wang[1]

Abstract. Aiming at the problem of the mismatching for feature selection and classifier parameter in network intrusion, a network intrusion detection model (FSCO-SVM) of the feature selection and classifier optimization coupling is proposed. Firstly, the evaluation criteria of network features are mapped to high dimensional space by radial basis function kernel to carry out the calculation; then the relationship between the network feature evaluation and the subsequent network intrusion classifier is established, which solves the problem of parameter design in the feature selection stage; finally, the network intrusion detection model is established, and the performance of FSCO-SVM is tested by KDD 99 data set. The results show that FSCO-SVM can not only eliminate the useless and redundant features, reduce the dimensionality of the network features, but can obtain the optimal parameters of the network intrusion classifier, so that the accuracy and detection efficiency of network intrusion detection are improved.

Key words. Feature selection, Intrusion classification, Network intrusion, Differential evolution, Feature classification.

1. Introduction

With the increase of Internet users, the network scale is becoming larger and larger, and the network intrusion events are increasing; as the last line of defense, it is concerned whether the intrusion detection system can detect all forms of intrusion behaviors, and the network intrusion detection has been a hot topic in network security research [1,2].

Network intrusion detection is essentially a classification problem of pattern recognition; firstly, the network state data is collected, and then the network behavior is analyzed, finally the network behaviors are divided into the abnormal and the normal, and the corresponding safety measures are taken according to the test results[3]. The network intrusion detection mainly includes two key steps: ① network

[1] Furong College Hunan, University of Arts and Science, Hunan Changde, 415000, China

feature extraction and selection; ② classifier design and parameter optimization [4]. The network feature has high dimension, containing redundant and useless features, and only some of the features are related to the test results; current feature selection mainly includes: principal component analysis, rough set, swarm intelligence algorithm, etc [5-7]; where, the dimension reduction effect of principal component analysis is good, but the result interpretability is poor; swarm intelligence algorithms mainly contain genetic algorithm, particle swarm optimization, ant colony algorithm, artificial bee colony algorithm, etc, which simulate the biological behavior of nature, with the advantages of fast speed, intelligence and so on, able to find a better subset of network features in a quick manner [8]. However, these algorithms do not take into account the design of the subsequent network intrusion classifier in feature selection, and the intrinsic relation between feature selection and classifier is completely separated, so the optimal feature subset selected in this way is not the optimal for subsequent classifier, and the network intrusion detection model with the best overall performance is difficult to obtain [9]. At present, the network intrusion classifier mainly includes the machine algorithm based on neural network and supporting vector machine (SVM); the neural network needs a large number of samples, while SVM is a kind of machine learning algorithm for small samples and high dimension classification problem, with excellent generalization ability; a large number of studies have shown that generally the performance of SVM classification performance is better than that of neural network [10,11]. Therefore, this study chooses SVM to build a network intrusion detection classifier. In the process of current network intrusion modeling, single objective optimization is usually used to solve the problem of feature subset and classifier optimization; in fact, the network intrusion feature subset and classifier are interrelated and interacted with each other, jointly acting on network intrusion detection results, therefore, the mismatching problem of feature selection and classifier occurs, influencing the effect of network intrusion detection [12].

In order to solve the mismatching problem of feature selection and classifier in network intrusion, a network intrusion detection model (FSCO-SVM) of the feature selection and classifier optimization coupling is proposed, and the KDD Cup 99 data set is used to carry out the simulation testing to the FSCO-SVM line, to verify the validity and superiority of the model.

2. Network intrusion classifier

Given network state data set: $\{x_i, y_i\}$, where, x_i refers to the network characteristics, and y_i refers to the output; according to the principle of structural risk minimization, the SVM optimal classification plane is:

$$y = w^T \varphi(x) + b\,. \tag{1}$$

In the formula(1), w is the normal vector and b is the offset vector.

By introducing Lagrange multiplier and kernel function, the classification function of SVM becomes:

$$f(x) = \operatorname{sign}\left(\sum_i \alpha_i y_i \cdot k(x_i \cdot x) + b\right) . \tag{2}$$

In the formula(2), is the Lagrange multiplier and $k(x_i \cdot x)$ is the kernel function [13].

3. Network feature selection method

In the process of network intrusion feature selection, two problems need to be solved: ① network feature selection criterion; ② network feature selec-tion method [14]. Current feature selection includes two types of Filter and Wrapper. The Filter method is independent with the subsequent network intru-sion classifier design, and the selected feature can hardly obtain the ideal effect in the subsequent network intrusion classifier design; while, Wrapper method needs to evaluate the performance of the network intrusion classifier in adding or deleting the features; the method combines the feature selection with the subsequent classifier, therefore this study uses wrapper method to select the network intrusion features.

3.1. Network feature selection criteria

Supposed that there are N types of network intrusion, the number of samples for each type of network intrusion is L, and the network feature dimension is K, then the network intrusion feature set can be expressed as:F=$\{f1$,$f2$,. . . ,fk$\}$, where, fi refers to the *ith* feature. Using "1-v-r" to design network intrusion multi classifier, it need N nos. of two-stage classifier. Each two-stage classifier needs two classes of samples, the first class is of fi("+1" mark) and the second class is of si("-1" mark); if a feature enlarges the distance between the two classes of samples, and reduces the variance within the class, it indicates that the feature is better, which is conductive to network intrusion modeling; therefore, the network feature selection criterion is defined as:

$$J(i) = \frac{|m_1^i - m_2^i|}{\sigma_1^i + \sigma_2^i} . \tag{3}$$

In the formula(3), m_1^i and m_2^i are respectively in the mean value center of f^i and s^i in the *ith*feature; σ_1^i and σ_2^i are respectively the standard variance of f^i and s^i in the *ith*feature;

Definitions of m_1^i, m_2^i, σ_1^i, and σ_2^i are as follows:

$$\begin{cases} m_1^i = \dfrac{1}{L}\displaystyle\sum_{j=1}^{L} f_j^i\,, \\ m_2^i = \dfrac{1}{M}\displaystyle\sum_{j=1}^{M} s_j^i\,, \\ \sigma_1^i = \sqrt{\dfrac{1}{L-1}\displaystyle\sum_{j=1}^{L}(f_j^i - m_1^i)^2}\,, \\ \sigma_2^i = \sqrt{\dfrac{1}{M-1}\displaystyle\sum_{j=1}^{M}(s_j^i - m_2^i)^2}\,. \end{cases} \tag{4}$$

Since the SVM classification is to transform the input space into a high dimensional space by mapping function of $\phi(\cdot)$, if the distance between two classes of samples is enlarged by a sample feature to a greater extent in a high dimensional space, it is more conducive to the subsequent classification of network intrusion classifier based on SVM, therefore we can use $\phi(\cdot)$ to map the feature selection criteria to high dimensional space, and the network feature selection criterion in the high dimensional feature space is:

$$\phi J(i) = \frac{\left|\frac{1}{L}\sum\limits_{j=1}^{L}\Phi(f_j^i) - \frac{1}{M}\sum\limits_{j=1}^{M}\Phi(s_j^i)\right|}{\{\sqrt{[\frac{1}{L-1}\sum\limits_{j=1}^{L}[\Phi(f_j^i) - \frac{1}{L}\sum\limits_{j=1}^{M}\Phi(f_k^i)]^2} + \sqrt{[\frac{1}{M-1}\sum\limits_{j=1}^{L}[\Phi(s_j^i) - \frac{1}{M}\sum\limits_{j=1}^{M}\Phi(s_k^i)]^2}\}}. \tag{5}$$

Since the radial basis kernel function $k(x_i, x_j) = \exp(-|x_i - x_j|^2/2\sigma^2)$ only needs to determine one parameter (σ core width) and is used in practice most widely, the radial basis kernel function is chosen as the kernel function of SVM, and then the feature selection criterion based on the radial basis kernel function in high dimensional space becomes:

$$\phi J(i)) = \frac{\sqrt{\frac{1}{L^2}\sum\limits_{j=1}^{L}\sum\limits_{i=1}^{L}K(f_j^i, f_i^i) - \frac{2}{LM}\sum\limits_{j=1}^{L}\sum\limits_{i=1}^{L}K(f_j^i, s_i^i) + \frac{1}{M^2}\sum\limits_{j=1}^{M}\sum^{M}K(s_j^i, s_i^i)}}{\{\sqrt{\frac{1}{L-1} - \frac{1}{L(L-1)}\sum\limits_{j=1}^{L}\sum\limits_{i=1}^{L}K(f_j^i, f_i^i)} + \sqrt{\frac{1}{M-1} - \frac{1}{M(M-1)}\sum\limits_{j=1}^{M}\sum\limits_{i=1}^{M}K(d_j^i, d_i^i)}\}}. \tag{6}$$

According to the Formula (6), it is found that the radial basis kernel function parameter σ is included in the network feature selection criteria of high dimensional space, so the relationship between network intrusion feature selection and network intrusion classifier parameters is established through σ. The value of parameter σ is not only related to the result of network intrusion feature selection, but also related to the classifier of SVM; this paper first sets the range of the σ value, and then uses the grid search algorithm to select the optimal parameters.

3.2. Network feature selection method

The basic idea of network intrusion feature selection: firstly, the influence of each feature on the network intrusion classifier is evaluated, and then the network features are sorted in descending order according to the above network intrusion selection criteria, finally, the No. $d(d \leq K)$ of former feature is chosen as the optimal feature. Network feature selection steps are as follows:

step1: The *kth* feature value of the *ith* class $(i = 1, 2, \dots, n)$ ofnetwork state sample is copied to set class_I, and the *kth* feature value of the other network state samples are copied to set class_II

step2: Calculate the *jth* kernel parameter and the ϕJ_{ijk} of the *kth* feature value, and save them.

step3: If $i = i + 1$ and $i \leq N$, skip to step2; otherwise, continue down.

step4: When all the feature values of the kernel function are calculated, calculate the following evaluation criteria:

$$\overline{\phi J_{ik}} = \frac{1}{N} \sum_{i=1}^{N} \phi J_{ijk} \times [\min(\phi J_{ijk})] \,. \tag{7}$$

Formula (7) describes the overall contribution of a feature to network intrusion classifier, the greater the value is, the greater the contribution to the result of network intrusion detection is, and namely the stronger the ability to distinguish network intrusion classes is; $\min(\phi J_{ijk})$ is the worst reflection to the network intrusion classifier ability; in this way, the formula (7) establishes the connection between the network feature evaluation and the subsequent network intrusion classifier, and the parameter design problem of classifier is considered in the feature selection phase.

step5: Under the *jth* kernel function, sort $\overline{\phi J_{ik}}$ and get the maximum value, namely the criterion coefficient is:

$$JO_j = \arg(\max_k \overline{\phi J_{ik}}) \,.$$

Select the network state feature under the $h = \arg(\max_k \overline{\phi J_{ik}})$th kernel function as the final optimization feature.

step6: Under the condition of choosing the optimal SVM radial basis kernel function parameters, the former d features are selected, and meeting $\phi J_h(1) > \phi J_h(2) > \cdots > \phi J_h(d)$.

4. Simulation experiment

4.1. Data sources

Experimental data comes from KDD99 data set, and the dataset consists of 4 types of intrusions: Probe (scanning and detecting), DoS (denial of service attack), U2R (illegal access to local super users) and R2L (unauthorized remote access); there are 41 attributes for each connection record in the data set: 9 discrete attributes and 32 continuous attributes [15]. In simulation environment of P4 dual core 2.8G CPU, 1G memory, and Windows XP operating system, use Matlab2012b for programming. 4000 samples are randomly selected to form the training set and the other 1000 samples for testing set, and the features are conducted with normalized processing, to be within the [0,1] range. Some of the codes for the normalized processing are:

```
[MaxV,I]=max(Data);
[MinV,I]=min(Data);
[R,C]==size(Data);
Scaled=(Data-ones(R,1)*MinV).*(ones(R,l)*(Upper-
            Lower)*ones(1,C))./(MaxV-MinV)))+Lower;
```

4.2. Feature selection

Range of kernel parameter σ is: {0.001, 0.01, 0.1, l, 2, 4, 8, 16, 32, 64, 128, 256}, and the training set is used to perform the above feature selection steps; criteria coefficient JOj for feature selection sees 1. From Table 1 we can see, when $\sigma = 8$, largest feature criterion coefficient can be obtained. When σ=8, the sorting sequence of the network feature numbers from large to small is: 4, 5, 8, 10, 39, 13, 15, 18, 19, 22, 25, 26, 28, 36, 3, 30, 16, 18, 19, 20, 21, 31, 23, 33, 36, 27, 29, 37, 24, 14, 41, 11, 6, 1, 7, 9, 12, 17, 32, 34, 35, 40.

Table 1. Criteria coefficient in selection

σ	JOj	σ	JOj	σ	JOj
0.001	0.1210	2	0.4642	32	0.4762
0.001	0.1971	4	0.4839	64	0.4449
0.1	0.2185	8	0.5130	128	0.4405
l	0.4096	16	0.4984	256	0.4359

Kernel parameter of the subsequent SVM$\sigma = 8$, penalty factor C=100 (C value is not related with the feature selection process). The value of feature dimension: $d = 5 \sim 41$, train the network intrusion two-stage classifier of "1-v-r", and the testing set is used to test the classifier performance; the obtained average detection accuracy of the network intrusion detection is shown in Fig. 1. According to the results of Fig. 1, when the feature number of the network gradually increases from 5 to 11, the maximum value of the average detection accuracy is 93.98%; then, with the increase of the number of features, the average detection accuracy gradually decreases, and the change curve of average detection accuracy has some fluctuations, which indicates that the simple increase of the network feature without selection will adversely affect the performance of the classifier, resulting the reducing of network intrusion detection accuracy. Therefore, in the selection of 4th,5th, 8th, 39th, 13th, 15th, 18th, 19th, 22nd and 25th features as the input of the classifier, the kernel function parameters of SVM are optimized, so that the overall performance of the network intrusion detection model is optimized.

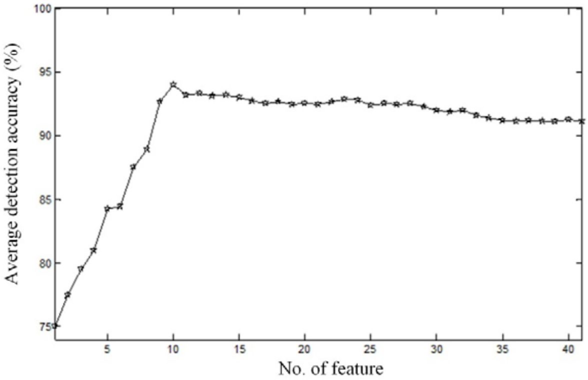

Fig. 1. Change curve between intrusion detection accuracy and feature dimension

4.3. Compared with the original feature detection performance

In order to make the comparison between the FSCO-SVM algorithm and the original SVM algorithm convictive, 5 groups of experiments are carried out, with the average value of the detection results obtained; average detection accuracy and detection are shown in Table 2 and 3 respectively. As can be seen from Table 2, compared to the network intrusion detection model without feature selection, FSCO-SVM network intrusion detection accuracy increases by 2.79%; the comparison results show that the feature selection can effectively eliminate redundant and useless features; select the important features closely related to the detection results, so that the network detection model can more accurately describe the change of network state information, so as to improve the effect of network intrusion detection.

Table 2. Detection accuracy before and after feature selection

Intrusion type	Original feature	Post selection feature	Increase value
Probe	90.63	93.82	3.20
DoS	92.86	97.07	4.21
U2R	92.55	94.10	1.55
R2L	88.72	90.91	2.18
Average value	91.19	93.98	2.79

As can be seen from Table 3, after the feature selection, the average detection time of FSCO-SVM is greatly reduced, which indicates that through the network feature selection, the redundant and useless features are eliminated, the input dimension of SVM is reduced, the number of support vectors is reduced and the computation time is reduced, so that the network intrusion detection speed increases; the network intrusion detection model can satisfy the real-time requirement of network intrusion detection to a greater extent by the selection of network features.

Table 3. Comparison of average detection time (second, s) before and after feature selection

Intrusion type	Original feature	Post selection feature	Decrease value
Probe	3.65	2.47	1.18
DoS	3.28	1.88	1.41
U2R	0.82	0.53	0.29
R2L	0.97	0.48	0.49
Average value	2.18	1.34	0.84

4.4. *Performance comparison with other feature selection methods*

In order to make the detection results of the FSCO-SVM more convictive, SVM model (PSO-SVM) of particle swarm optimization algorithm selection feature is used to be the contrast model. The optimal feature dimension of PSO-SVM is 15, which is larger than FSCO-SVM's dimension of 11; the average network intrusion detection accuracy (%) of the 5 results of FSCO-SVM and PSO-SVM model is as shown in Fig. 2. As can be seen from Fig. 2, the detection rate of FSCO-SVM is higher than that of PSO-SVM model; the comparison results show that FSCO-SVM uses the feature selection and classifier coupling to build the network intrusion model, to explore the relationship between feature selection and classifier, to describe the network state change information more accurately, to effectively eliminate redundant and useless features, to overcome the mismatching problem of feature selection and classifier in network intrusion, so as to improve the detection accuracy of network intrusion.

5. Conclusion

In order to solve the mismatching problem of feature selection and classifier in network intrusion, a network intrusion detection model of the feature selection and classifier optimization coupling is proposed. The simulation results show that FSCO-SVM not only reduces the feature dimension and achieves the feature optimization combination, but also considers the optimization of classifier parameters and establishes a network intrusion detection model with the best overall performance; the optimized feature subset contains only 11 features, and the number of features is reduced by 73.17% compared with the original 41 features; the average accuracy of network intrusion detection can reach more than 93%; at the same time, compared with other network intrusion models, FSCO-SVM also has a better effect of intrusion detection, which verifies the effectiveness and superiority of FSCO-SVM.

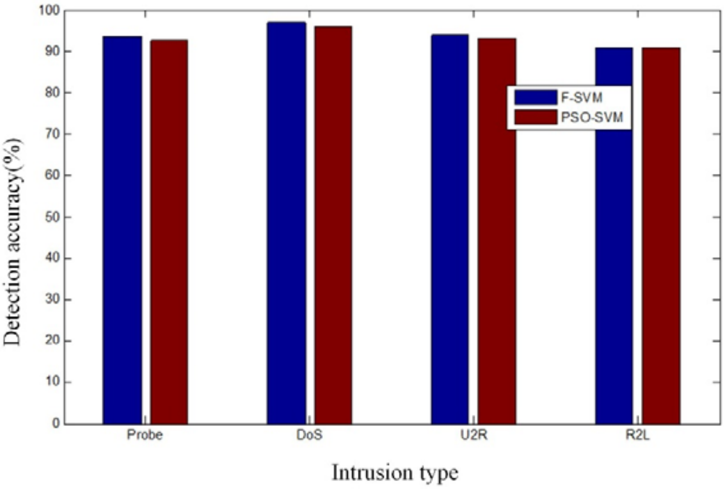

Fig. 2. Comparison between detection results for FSCO-SVM and PSO-SVM

Acknowledgement

This work was supported by the scientific research project of Education Department of Hunan Province No. 14C0792.

References

[1] Gude N, Koponen T, Pettit J, et al.: (2008) *NOX: towards an operating system for networks*[J]. AcmSigcomm Computer Communication Review, 38(3):105-110.

[2] Chandrakasan A P, Brodersen R W: (1995) *Minimizing power consumption in digital CMOScircuits*[J]. Proceedings of the IEEE, 83(4):498-523.

[3] Kansal A, Hsu J, Zahedi S, et al.: (2007) *Power management in energy harvesting sensor networks*[J]. Acm Transactions on Embedded Computing Systems, 6(4):32.

[4] Chen Y, Zhao Q: (2005) *On the lifetime of wireless sensor networks*[J]. IEEE Communications Letters, 9(11):976-978.

[5] Yang H, Meng X, Lu S : (2002) *Self-organized network-layer security in mobile ad hoc networks*[C]// ACM Workshop on Wireless Security, Atlanta, Ga, Usa, September. DBLP, 2002:11-20.

[6] Lee W, Stolfo S J, Mok K W: (2000) *Adaptive Intrusion Detection: A Data Mining Approach*[J]. Artificial Intelligence Review, 14(6):533-567.

[7] Inoue K, Ishihara T, Murakami K: (1999) *Way-predicting set-associative cache for high performance and low energy consumption*[C]// International Symposium on Low Power Electronics and Design, 1999. Proceedings. IEEE, 1999:273-275.

[8] Sebring M M, Shellhouse E, Hanna M F, et al.: (1988) *Expert systems in intrusion detection: a case study*[C]// World Conference on Photovoltaic Energy Conversion. 1988:32–38.

[9] Vasiliadis G, Antonatos S, Polychronakis M, et al.: (2008) *Gnort: High Performance Network Intrusion Detection Using Graphics Processors*[C]// International Symposium on Recent Advances in Intrusion Detection. Springer-Verlag, 2008:116-134.

[10] Zheng Y, Stute M, Geen A V, et al.: (2004) *Redox control of arsenic mobilization in Bangladesh groundwater*[J]. Applied Geochemistry, 19(2):201-214.

[11] Dunkels A, Finne N, Eriksson J, et al.: (2006) *Run-time dynamic linking for reprogramming wireless sensor networks*[C]// International Conference on Embedded Networked Sensor Systems, SENSYS 2006, Boulder, Colorado, Usa, October 31 - November. DBLP, 2006:15-28.

[12] Khan L, Awad M, Thuraisingham B: (2007) *A new intrusion detection system using support vector machines and hierarchical clustering*[J]. The VLDB Journal, 16(4):507-521.

[13] Manshaei M H, Zhu Q, Alpcan T, et al.: (2011) *Game theory meets network security and privacy*[J]. Acm Computing Surveys, 45(3):1-39.

[14] Modi C, Patel D, Borisaniya B, et al.: (2013) *Review: A survey of intrusion detection techniques in Cloud*[J]. Journal of Network & Computer Applications, 36(1):42–57.

Received May 7, 2017