http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 11 (7): 794-798, 2012
ISSN 1812-5638 / DOI: 10.3923/itj.2012.794.798

An Intrusion Detection Model Based on GS-SVM Classifier

Xiangyu Lei and Ping Zhou
Guilin University of Electronic Technology, Guilin, 541004, Guangxi, China

Abstract: The Coarse-to-Refined Grid Search Support Vector Machine (GS-SVM) is an improvement on One-class Support Vector Machine (SVM). This study provides a solution for Intrusion Detection System (IDS) based on support vector machine. In practice, it is inefficient for SVM to identify massive intrusive behaviors which will exhaust memory resources. In addition, the accuracy of classification is subject to data preprocessing and parameter selection. In order to obtain precise detection rate, it is crucial to optimize the related parameters for proper kernel function. For this reason, an optimization algorithm based on grid search is proposed. Experiments over networks connection records from KDD 99 data set are implemented for 1-vs-N SVM to evaluate the proposed method. This approach reduces the training time, accelerates the speed of Cross Validation and improves the adaptability of the IDS.

Key words: Grid search, intrusive behavior, support vector machine, kernel function, cross validation

INTRODUCTION

The accuracy of intrusion detection system is subject to unstable and time-consuming detection algorithms. In recent years, intrusion detection technology can be divided into two categories: intellectualization (Macia-Perez *et al*., 2011) and distribution (Rehak *et al*., 2009). With the development of electronic computer technology, researches of the anomaly detection methods (Mabu *et al*., 2009) play a much more significant role in intrusion detection systems. Based on statistical learning theory, the intrusion detection system is capable of continuous learning which enhances recognition rate and reliability.

In reality, when the network scale is large, the training time becomes too long to detect the incoming attacks. Besides of this main issue, the complexities of matrix operation will cause memory-hungry. In order to monitor traffic on a high-speed link continuously, a quick-response SVM classifier is needed. However, it is not practical for SVM to handle unlabeled data stream. For this reason, this study provides convenience and assistant of preprocessing and analyzing audit data for security auditor.

The difficulty in SVM-based IDS is that of optimizing the parameters from the training phase of the classifier. Because of this, an optimization algorithm based on grid search is proposed. By our use of the KDD CUP 99 data preprocessed as a source in these experiments, the related parameters were selected rapidly and appropriately. The test result showed that the classification accuracy ratio has been improved significantly.

PRINCIPLE OF SVM FOR MULTI-CLASSIFICATION

Support Vector Machine (SVM), a kind of machine learning algorithm, can efficiently solve the classification issue. Since it can obtain better performance under less sample training conditions, SVM possesses stronger generalized capability (Keerthi *et al*., 2000). SVM had been widely used in the small data sets and high dimension feature spaces due to its good applicability and high efficiency, especially in abnormal intrusion detection.

SVM is a linear classification machine which was based on the premise that the problem is linearly separable. For two-class samples in training set, the linear separable function works perfectly if samples can be apart by a hyper-plane. But there are special circumstances which appears discrete samples.

Figure 1 indicates that it is impossible to separate the white samples from the black ones which surround them by a straight line. The classifier will fall into an infinite loop apparently. In order to obtain an available optimal hyper-plane, the data should be logically mapped into a high-dimensional feature space (Liejun *et al*., 2008). As Fig. 1 depicts, a dotted parabola is utilized to separate the two classes according to the function value. Suppose that $<x, z>$ is the inner product of original features. While mapping to $<\Phi(x), \Phi(z)>$, the kernel function can be defined as follows:

$$K(x, z) = \phi(x)^T \phi(z) \quad (1)$$

According to the Mercer Principle, different kernel function generates different classification machines.

Corresponding Author: Xiangyu Lei, Guilin University of Electronic Technology, Guilin, 541004, Guangxi, China

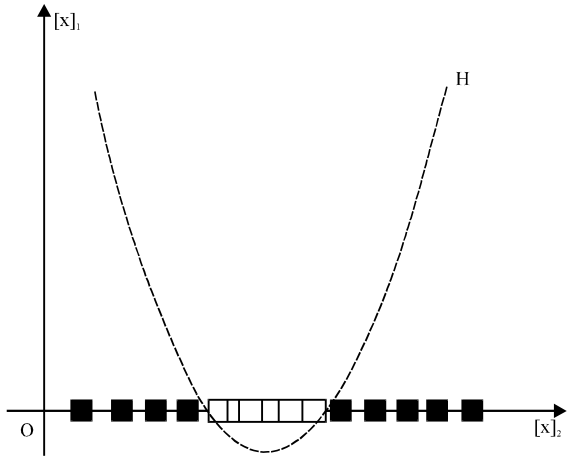

Fig. 1: The impartible data sample

RBF (Radial Basis Function) is one of the kernel functions and it can be defined as follows:

$$K(x,z)=\exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right) \quad (2)$$

In related researches (Lijia *et al*., 2011; El-Kouatly and Salman, 2008; Shuchun *et al*., 2011), the RBF possess advantages of other kernel function for three reasons: Keerthi proved that linear kernel, with one penalty parameter and RBF, with a penalty parameter and a kernel parameter, can provide the same performance. Moreover, it is hardly to distinguish the performance between Sigmoid kernel and RBF. RBF is able to deal with samples in which relationship between Class Label and feature is nonlinear.

Classification model might be severely affected by noise. An outlier will result in hyper plane dislocation. According to the largest interval law, the interval distance of hyper planes reduced by the deflected support vectors. To avoid this, the nonnegative Slack Variable ξ_i is introduced to largest interval and noise samples are partially allowed to maximize the interval distance. Here is the expression so-called soft margin separating hyper plane:

$$\min_{\xi,w,b} \quad \frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}\xi_i \quad (3)$$

$$\text{s.t. } y_i((w.X_i)+b)\geq 1-\xi_i;\ \xi_i\geq 0,\ i=1,\ldots,l \quad (4)$$

In expression 3, the target function value can be largely affected by parameter C which is the weight of outliers. The larger the C value is, the bigger the function value will be. Since the amount of outliers is controlled by accumulated polynomial, most of the sample points obey the constraint. The corresponding Lagrange formula can be defined as follows:

$$\mathcal{L}(w,b,\xi,\alpha,r)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}\xi_i-\sum_{i=1}^{l}\alpha_i\left[y_i(w\cdot x_i+b)-1+\xi_i\right]-\sum_{i=1}^{l}r_i\xi_i \quad (5)$$

In expression 3, both α_i and r_i are Lagrange multipliers. The expressions of w and b can be derived using partial derivatives to parameters of this formula:

$$\max_{\alpha} \quad F(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\alpha_i\alpha_j\langle x_i,x_j\rangle \quad (6)$$

$$\text{s.t. } 0\leq\alpha_i\leq C,i=1,\cdots,l;\ \sum_{i=1}^{l}\alpha_iy_i=0 \quad (7)$$

According to expression 6 and constraint expression 7, the anti-noise ability can be improved and the affect of outliers can be reduced by looking for the biggest interval. Aimed at the problem of multi-class classification, Hsu and Lin (2002) approaches compared theoretical results with experimental data. Some of the typical methods such as 1-vs 1, 1-vs-r and DAG-SVM are highly effective. Centered on the thinking of two-class classification (Yang *et al*., 2011), to obtain effective multi-class classifier, it is necessary to pretreat the input data before training the classifier. After the data preprocessing phase, parameter optimization will be the most important part in this study.

DATA PREPROCESSING FOR INTRUSION DICTION MODEL

Security audit dataset, KDD 99, processed by Lee (Stolfo *et al*., 2001) was considered as a noted dataset for intrusion detection systems. Aiming at making evaluation (Muda *et al*., 2011) on intrusion detection models, this audit dataset, provides TCP-dump-format data recorded from a simulative local network, serve as training input to regulate multi-class classifier based on SVM. There are 38 kinds of attacks in this text-format file. Basic, content-based, 2-second-flow and host stream characteristics are included in this dataset. Because of more than million records are included in KDD dataset, 10% data are selected in this study.

The relative merits of data preprocessing are highly related to the precision of classification (Gu *et al*., 2008). For some unrecognized features are included in KDD dataset, such as the character data, all features should be numeralized before the classification phase. For example,

the protocol type, of which the values are TCP, ICMP and UDP, are replaced by 3 integer values from 1 to 3. Similarly, numeralization replacement is implemented for the third, fourth and forty-second features. The 40 sec feature can individually serve for classification label. According to the rest of the dataset which includes 41 features, the metrics of features are different in standard. To avoid the unbalanced data impact, a phenomenon that some of the smaller values are covered by the larger values, it is essential to normalize the value of each attribute from metrical to non-unit-related by the following formula:

$$y = \frac{(y_{max} - y_{min})(x - x_{min})}{(x_{max} - x_{min})} + y_{min} \quad (8)$$

In the expression 8 for normalization, y_{max} is range of the upper limit whereas y_{min} is range of the lower limit. x_{max} and x_{min} are maximum and minimum value of the original feature. With the input of x, the value from current record of the corresponding attribute, the output *y* is recorded as normalized value through calculation.

However, it is complicated for computer to calculate enormous matrix generated from 10% KDD dataset after numeralization and normalization. To reduce memory consumption in large scale computation, feature extraction (Al-Tameemi *et al*., 2011) is utilized to accelerate convergence. According to related research, rough set theory is proposed to extract KDD dataset which can be indicated using the methods provided by ZAINA A (Zaina *et al*., 2006). Six attributes are included in the extracted dataset: 3, 4, 5, 24, 32 and 33. To train the SVM multi-classifier, this experiment selected 2251 samples, including 1000 normal behaviors, 817 denial-of-service attacks, 23 probe behaviors, 37 user-to-root behaviors and 374 remote-to-local behaviors.

TRAINING SVM-BASED MULTI-CLASSIFIER WITH GS

Cross validation provides reliable basis for the evaluation and stability to the trained multi-classifier. In this phase, SVM classification accuracy can be greatly affected by the nuclear parameter σ in formula 2 and the error penalty parameter C in expression 3. In order to obtain the best combination of these parameters, an improved GA (genetic algorithm) was applied to parameter optimization (Chen and Wang, 2008). To large-scale and unbalanced datasets, such as KDD dataset, GA has provided a global optimal solution depending on the fitness function. However, it is easy to get into the local optimum combination because of the randomness of genetic variation algorithm.

To avoid the shortcoming of weak local search ability of traditional genetic algorithm, a GS algorithm (Coarse-to-refined grid search) is proposed. The steps are as follow:

Step 1: First, search at large range. Parameter C is selected from a geometric series of which the range is 2^{-4} to 2^{8} and the common ratio is $2^{0.5}$. Parameter σ is selected from a geometric series of which the range is 2^{-4} to 2^{4} and the common ratio is $2^{0.5}$. According to the best accuracy in cross validation, the optimal combination of *C* and σ are recorded as C* and σ*. The distribution map of parameters and accuracy and three-dimensional one are shown in Fig. 2

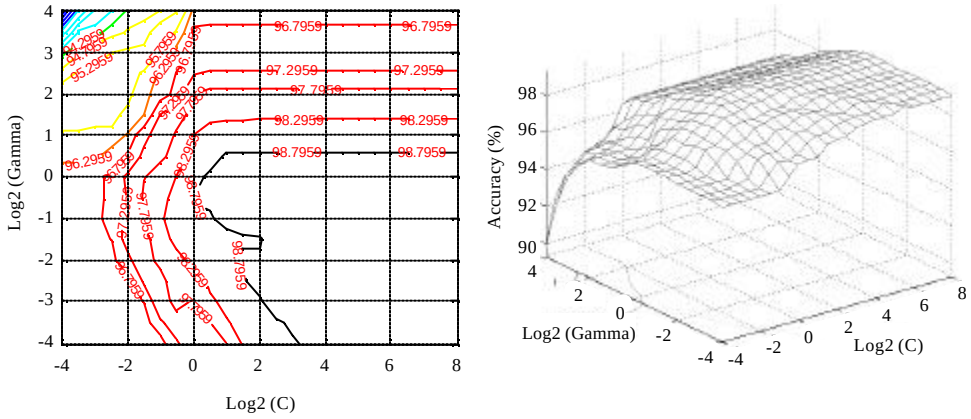

Fig. 2: Distribution map of parameters and accuracy and that in three-dimensional coordinate using rough GS; (a) Distribution map of parameters and accuracy and (b) Distribution map of parameters and accuracy in three-dimensional coordinate

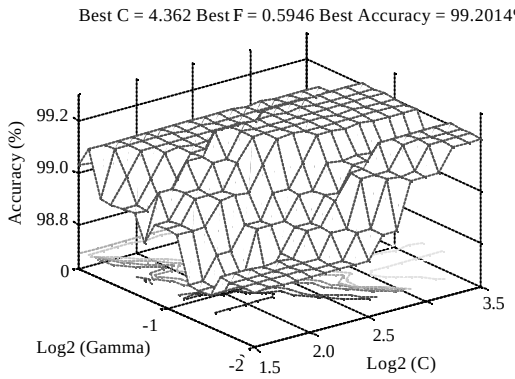

Fig. 3: Three-dimensional distribution map of parameters and accuracy using refined GS

As Fig. 2a shows, the cross validation accuracies are distributed in the curves. The combination of C and σ is gradually converging to smaller areas.

Step 2: Then, search at small range according to the selected C* and σ*. New parameter C is selected from a geometric series of which the range is $C^*\times2^{-1}$ to $C^*\times2^{1}$ and the common ratio is $2^{0.125}$. Similarly, new parameter σ is selected from a geometric series of which the range is $\sigma^*\times2^{-1}$ to $\sigma^*\times2^{-1}$ and the common ratio is $2^{0.125}$. According to the best accuracy in cross validation, new optimal combination of C and σ are recorded as C* and σ*. The distribution map of parameters and accuracy and three-dimensional one are shown in Fig. 3

As Fig. 2a shows, the distribution of cross validation accuracies are in close proximity to plain areas. In the process selecting *C* and σ, over-learning may arise (The multi-classifier runs with high accuracy in training dataset but poor accuracy in test dataset.). In order to avoid this phenomenon in highest plain area, minimum C is selected as best parameter C*, maximum σ is selected as best parameter σ*.

SIMULATION RESULTS

By using the 5% KDD dataset, selected and pretreated as test data, is utilized for classification. The Table 1 compares the SVM multi-classifier based on Grid Search Algorithm with that based on Genetic Algorithm.

Experimental result shows that the testing accuracy optimized by GS is better then that optimized by GA and the best C and best σ selected by GS are smaller than that

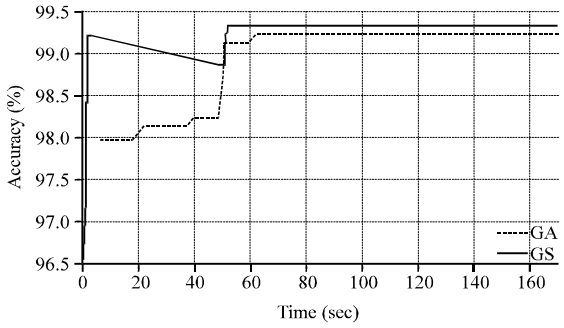

Fig. 4: The training time and cross-validation accuracy of GS and GA during the parameter optimization phase

Table 1: Grid search algorithm vs. Genetic algorithm

	Parameter optimizing algorithm	
	GS	GA
The best C	4.362	9.2603
The best σ	0.5946	0.6347
training time (second)	80.812	233.734
Cross-validation accuracy	99.2014%	99.2014%
testing time (second)	0.063	0.094
Testing accuracy	98.1317%	97.8648%

selected by GA. Although cross-validation accuracies are the same, the GS is advantageous in both training and testing time. The training result is shown in Fig. 4.

From the Table 1 and Fig. 4, we can see that the training time of GA-based multi-classifier is three times the training time of GS-based. So the intrusion detection model based on GS-SVM is an effective system, with faster reaction velocity, higher accuracy and stronger generalization.

CONCLUSION

This study focused on the One-Class SVM which was applied to the intrusion detection module. According to the scale of KDD dataset, the accuracy was strongly influenced by parameter C and σ. After sample numeralization and normalization, feature extraction based on rough set theoretical model was implemented to reduce memory consumption result from high dimensionality. In order to improve the generalization capacity of SVM, a parameter selection algorithm was proposed. Computation on simulation examples and comparison with genetic algorithm demonstrated that the GS doubled the speed of convergence in the cross validation phase. Moreover, the classification accuracy was improved by 0.2669%. For this reason, this paper recommended GS-SVM into intrusion detection system for the purpose of reducing the training time and enhancing the stability of the anomaly detection model.

ACKNOWLEDGMENT

This study is supported in part by the National Natural Science Foundation of China, Grant No. 60961002.

REFERENCES

Al-Tameemi, A.M., L. Zheng and M. Khalifa, 2011. Off-line arabic words classification using multi-set features. Inf. Technol. J., 10: 1754-1760.

Chen, Z.G. and S. Wang, 2008. Minimax probability machine with genetic feature optimized for intrusion detection. Inform. Technol. J., 7: 185-189.

El-Kouatly, R. and G.A. Salman, 2008. A radial basic function with multiple input and multiple output neural network to control a non-linear plant of unknown dynamics. Inform. Technol. J., 7: 430-439.

Gu, Y., B. Zhou and J. Zhao, 2008. PCA-ICA ensembled intrusion detection system by pareto-optimal optimization. Inform. Technol. J., 7: 510-515.

Hsu, C.W. and C.J. Lin, 2002. A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Networks, 13: 415-425.

Keerthi, S.S., S.K. Shevede, C. Bhattacharyya and K.R.K. Murthy, 2000. A fast iterative nearest point algorithm for support vector machine classifier design. IEEE Trans. Neural Networks, 11: 124-136.

Liejun, W., J. Zhenhong and L. Zhaogan, 2008. Multi-resolution signal decomposition and approximation based on SVMS. Inform. Technol. J., 7: 320-325.

Lijia, C., Z. Shengxiu, L. Xiaofeng, L. Yinan and L. Ying, 2011. Nonlinear adaptive block backstepping control using command filter and neural networks approximation. Infor. Technol. J., 10: 2284-2291.

Mabu, Y.G., S.C. Chen, Y. Wang and K. Hirasawa, 2009. Intrusion detection system combining misuse detection and anomaly detection using Genetic Network Programming. Proceedings of the ICCAS-SICE, August 18-21, 2009, Waseda University, Kitakyushu, Japan, pp: 3463-3467.

Macia-Perez, F., F. Mora-Gimeno, D. Marcos-Jorquera, J.A. Gil-Martinez-Abarca, H. Ramos-Morillo and I. Lorenzo-Fonseca, 2011. Network intrusion detection system embedded on a smart sensor. IEEE Trans. Ind. Electron., 58: 722-732.

Muda, Z., W. Yassin, M.N. Sulaiman and N.I. Udzir, 2011. A K-means and naive bayes learning approach for better intrusion detection. Inform. Technol. J., 10: 648-655.

Rehak, M., M. Pechoucek, M. Grill, J. Stiborek, K. Bartos and P. Celeda, 2009. Adaptive multiagent system for network traffic monitoring. Intell. Sys., 24: 16-25.

Shuchun, Y., Y. Xiaoyang, S. lina, Z. Yuping and S. Yongbin *et al*., 2011. A reconstruction method for disparity image based on region segmentation and RBF neural network. Inform. Technol. J., 10: 1050-1055.

Stolfo, S.J., W. Lee, P.K. Chan, W. Fan and E. Eskin, 2001. Data mining-based intrusion detectors: An overview of the Columbia IDS project. ACM Sigmod Rec., 30: 5-14.

Yang, M.N., X.J. LI and X.H. Zhang, 2011. Robust description method of SIFT for features of license plate characters. Inf. Technol. J., 10: 2189-2195.

Zaina, A., M.A. Maarof and S.M. Shamsuddin, 2006. Feature selection using rough set in intrusion detection. Proceedings of the TENCON 2006 IEEE Region of 10 Conference, November 14-17, 2006, Teknologi Malaysia, Johor, pp: 1-4.