ATLANTIS PRESS

Advances in Engineering Research (AER), volume 148

3rd Workshop on Advanced Research and Technology in Industry Applications (WARTIA 2017)

RF-SVM Based Awareness Algorithm in Intelligent Network Security Situation Awareness System

Chen Gang[1,a,*], Zhao Yu-qian[1,b]

[1]Academy of Information Communication, National University of Defense Technology, WuHan, China

[a] besteel@aliyun.com, [b] brad@aliyun.com

Keywords: Intelligent System, Network Security Situation Awareness, Regression Forecast, Support Vector Machine

Abstract: With the increasing outstanding of network security situation, the intelligent network security situation awareness system has been an important weapon in cyberspace battle field, its research and development are also being a matter of great urgency. How to improve accurate rate of situation awareness has been a primary key problem which must be dealt with by intelligent network security situation awareness system. A network security situation awareness algorithm based on regression forecast support vector machine (RF-SVM) was put forward. With adopting regression idea of regression, this algorithm can forecast potential threat in future network data flow referring to historical network attack data thoroughly in process of network awareness. Experiment indicates it can improve accurate rate of situation awareness effectively and reduce forecasting error.

1. Introduction

With network going deep into social life, malicious activities aiming at network are growing and attacks such as Dos/DDos are becoming a growing threat. The traditional network security management is usually dependent on discrete deploying network security devices involving firewall, anti-virus and intrusion detection. So many problems including single function, larger warning information and too much irrelevant warning information will appear. Network security manager is hard to control being confronted with security situation of total network due to facing too much warning information. It would lead effective responses can't be taken in time. So research and development of intelligent network security situation awareness system are becoming a matter of great urgency. Network security situation awareness has been a hot field developing in recent years and how to improve accurate rate of situation awareness has been a primary solving problem. It can provide the foundation for network security managers with decision by analysing current network security situation and forecasting attack behaviour which will happen in next step. It is of great importance to improving monitoring capability of network, emergency response capability and forecasting development tendency of network security.

Recently, many network security situation awareness algorithms have been presented by researchers such as log audit and performance correction algorithm, basing on D-S evidence theory, basing on hybrid system model, basing on artificial neural network, multi-dimension data streams mining algorithm, Markov game model, and so on. These algorithms have been proven in process of network security situation awareness and achieved good result. Meanwhile, many network security situation forecast models have been built such as quantitative hierarchical threat evaluation model, information fusion model, complex network model, and so on.

Considering successful application in forecast and comprehensive evaluation, support vector machine was adopted to network security situation awareness. The low accurate rate will be got due to forecast network security situation only by current data. To resolve this problem, regression forecast is introduced to improve accurate rate of network security situation awareness. It can forecast potential threat in future network data flow referring to historical network attack data thoroughly in process of network awareness. Experiment indicates it can improve accurate rate of situation awareness effectively and reduce forecasting error.

2. Algorithm Design of RF-SVM

2.1. Theory of RF-SVM

SVM is a kind of data mining technique, which can find hidden modes from magnanimity data and mine information hidden in data. It is easy to recognize time series and trend of data by intelligent system with providing related information. SVM can look for the perfect class plane in solution space in linearly separable standard. It can also map sample in low dimension input space to high dimension space effectively by adopting nonlinear mapping method after introducing partial variable. So the optimization solution will be found in high dimension solution space with changing solution space to linearly separable. At the same time, the optimization solution will be found in solution space of vector using the principle of structural risk minimization by SVM.

For forecasting network security situation more effectively, thought of regression forecast was introduced and RF-SVM algorithm was put forward. Its principle can be described with following.

Supposing the given sample set (x, y) followed probability distribution $P(x, y)$ and the regression function is listed as Eq. 1.

$$F = \{f \mid f(x) = w^{T}\Phi(x) + b, w \in R^{n}\} \tag{1}$$

The structural risk function is also introduced as Eq. 2.

$$R_{reg} = \frac{1}{2}\|w\|^{2} + C \bullet R_{emp}^{\varepsilon}[f] \tag{2}$$

In Eq. 2, $\|w\|^{2}$ expresses describing function and C is constant. $f()$ expresses item of complexity. Eq. 2 can gain compromised balance by balancing structural risk and model complexity. By introduced ε item of insensitive loss function in Eq. 2, ε can be defined as Eq. 3.

$$|y - f(x)|_{\varepsilon} = \begin{cases} 0 & |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon & other \end{cases} \tag{3}$$

Eq. 3 expresses it will not punish item whose deviation less than ε. It will improve robust of regression function remarkably.

$$R_{emp}^{\varepsilon} = \frac{1}{1}\sum_{i=1}^{1}|y_{i} - f(x_{i})|_{\varepsilon} \tag{4}$$

Eq. 4 showed core idea of introducing regression forecast to SVM. Both training error and model complexity can be controlled to take a small expectation risk. Its minimum cost generalization function can be expressed as Eq. 5.

$$\min \frac{1}{2}w^{t}w + C\sum_{i=1}^{1}(1 + 1^{*}) \tag{5}$$

In Eq. 5, $1, 1^{*}$ expresses introduced relax variable. By using Lagrangian function and duality principle, Eq. 6 can be deduced.

$$\min\left\{\frac{1}{2}\left[a, (a^{*})^{T}\right]\begin{bmatrix} Q & -Q \\ -Q & Q \end{bmatrix}\begin{bmatrix} a \\ a^{*} \end{bmatrix} + \left[\varepsilon I^{T} + y^{T}\varepsilon I^{T} - y^{T}\right]\begin{bmatrix} a \\ a^{*} \end{bmatrix} w^{t}w + C\sum_{i=1}^{l}(1 + 1^{*})\right\} \tag{6}$$

In Eq. 6, a, a^{*} expresses Lagrangian operator and it can be solved as Eq. 7.

$$w^{*} = \sum_{i=1}^{n} a_{i}^{*}y_{i}x_{i} \tag{7}$$

In Eq. 7, α_i^* is support vector and it is non-zero sample. The weight coefficient vector of optimization class plane is linear combination of support vector. So the obtained optimization class function can be expressed as Eq. 8.

$$f(x) = \mathrm{sgn}((w^*)^T x + b^*) = \mathrm{sgn}(\sum_{i=1}^{n} a_i^* y_i x_i^* x + b^*) \quad (8)$$

In Eq. 8, $\mathrm{sgn}(\ )$ is sign function and b can be computed with constraint condition $\alpha_i[y_i(w^T x_i + b) - 1] = 0$.

2.2. Algorithm Design and Parameter Setting

The control methods of RF-SVM regression mode consists of capability control factor C, loss function and kernel function. These methods can realize efficient control and regression of RF-SVM mode. For testing validity of RF-SVM algorithm, it is controlled by adopting ε insensitive loss function of Vapnik and using kernel function with Gauss radial basis function. $\varepsilon = 0.008$ is set and value of control factor C is unlimited. Gauss radial basis function $\sigma = 0.2$ is set to finish training in process of model training. For forecasting network security situation index better, realization of RF-SVM algorithm involves two modules including network security situation forecast training module and forecast module. Function of each module is described as follow.

2.3. RF-SVM Training Module

In process of algorithm execution, RF-SVM training module includes four parts such as overall control, database read, situation forecast and training. Its specific execution step can be described with following.

Step One: Setting input training data time serialize condition, confirming time serial and its value range.

Step Two: Calling database read control function and network security situation forecast function, forecasting situation and classifying by time serial, storing forecast value to variable LIST.

Step Three: Building a forecast model by introducing LIST to forecast mode training modules.

2.4. RF-SVM Forecast Module

RF-SVM forecast module will forecast network security situation according to recent historical data adopting RF-SVM algorithm. Its specific execution step can be described with following.

Step one: Confirming time serial of training data based on input condition, setting value range of time serial.

Step two: Finishing work of time serial.

Step three: Calling network security situation forecast module, forecasting network security situation based on model built by time serial.

3. Experiment

3.1. Experimental Environment and Experimental Data

In order to demonstrate validity of RF-SVM algorithm, a powerful IDS was constructed with two FTP servers, one Web server, twenty computer terminals and six simulation attack computer. During January 1 and March 10 in 2014, 300000 selected intrusion warning information bits including 70000 log warning information bits, 100000 network warning information bits, 80000 device warning information bits and 50000 agent warning information bits are collected. According to classification standard of network security situation, these information bits can be classified as 40000 level one warning information bits, 80000 level two warning information bits, 80000 level three warning information bits and 100000 level four warning information bits.

Due to enormous data collection of network security situation in different time period, the computation situation value must be normalized for avoiding larger error generated by RF-SVM algorithm. The normalization expression can be expressed as Eq. 9.

$$\hat{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{9}$$

In Eq. 9, x is current value of network security situation, $\hat{x}$ is value of network security situation after normalization, x_{max} and x_{min} are maximum and minimum of network security situation.

By setting data during January 1 and February 28 in 2014 as training data collection, learning and training of RF-SVM algorithm are finished. So accuracy and reliability of RF-SVM algorithm will be demonstrated by setting network security situation data during March 1 and March 10 in 2014 as test data. The running result of RF-SVM algorithm is shown in Table 1.

Table 1 The running result of RF-SVM algorithm

Time	Actual Situation	Forecast of RF-SVM	Error of RF-SVM	Forecast of SVM	Error of SVM
March 1	0.4864	0.3463	0.1401	0.3363	0.1501
March 2	0.2589	0.2785	-0.0197	0.2943	-0.0355
March 3	0.4733	0.3835	0.0898	0.3527	0.1207
March 4	0.1361	0.1967	-0.0606	0.2163	-0.0803
March 5	0.4539	0.4031	0.0508	0.3723	0.0816
March 6	0.3898	0.1900	0.1998	0.1472	0.2426
March 7	0.6078	0.6879	-0.0801	0.7287	-0.1209
March 8	0.4443	0.5844	-0.1401	0.6146	-0.1703
March 9	0.3388	0.4583	-0.1195	0.4932	-0.1544
March 10	0.6106	0.6465	-0.0360	0.6808	-0.0703

3.2. Analysis of Experimental Results

To demonstrate validity of RF-SVM algorithm, the forecast results of RF-SVM algorithm are compared with forecast results of SVM algorithm. It is proved that forecast accurate rate of network security situation can be improved better by RF-SVM algorithm, as shown in Figure 1.

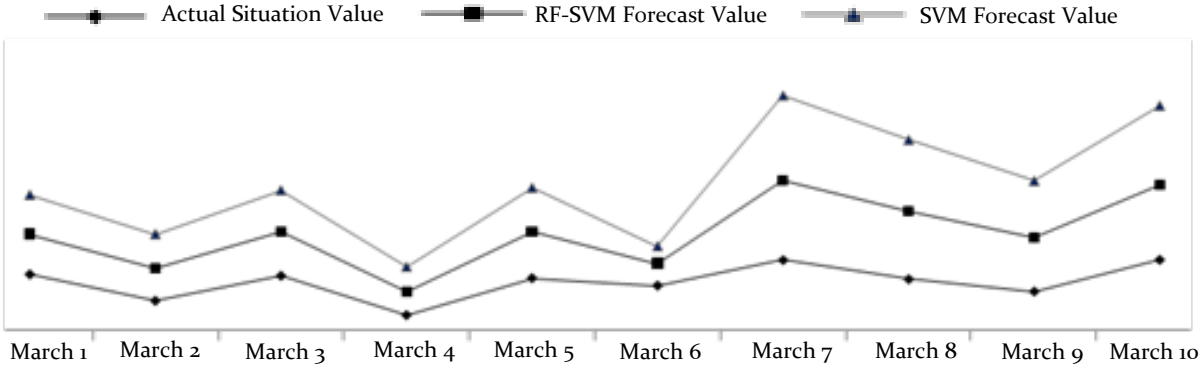

Figure 1 The forecast results of RF-SVM algorithm are compared with forecast results of SVM algorithm

4. Conclusion and Future Works

In view of shortcoming of SVM algorithm in network security situation awareness process, including unable to store historical data flow and low network security situation awareness accurate rate, RF-SVM algorithm was put forward by introducing idea of regression forecast. This algorithm can forecast potential threat in future network data flow referring to historical network attack data

thoroughly in process of network awareness. Experiment indicates it can improve accurate rate of situation awareness effectively and reduce forecasting error. The emphasis of future works for RF-SVM algorithm involve improving kernel function, parameter setup optimization and realizing intelligent automatic forecast of network security situation.

References

[1] GONG Zheng-Hu, ZHUO Ying. “Research on Cyberspace Situational Awareness”. Journal of Software. 2010, 21(7):1605-1619.

[2] WEI Yong, LIAN Yi-Feng. “A Network Security Situational Awareness Model Based on Log Audit and Performance Correction” . Chinese Journal of Computers. 2009, 32(4):763-772.

[3] SHI Bo, XIE Xiao-quan. “Research on Network Security Situation Forecast Method Based on D-S Evidence Theory”. Computer Engineering and Design. 2013, 34(3):821-825.

[4] LI Wen, DAI Ying-Xia, LIAN Yi-Feng and FENG Ping-Hui. “Context Sensitive Host- based IDS Using Hybrid Automaton”. Journal of Software. 2009, 20(1):138-152.

[5] ZHONG Zhao-man, LI Cun-hua and GUAN Yan. “Instant Intrusion Detection System Based on Neural Network”. Computer Engineering and Applications. Computer Engineering and Applications. 2007, 43(30):120-123.

[6] Mao Guojun, Zong Dongjun. “An Intrusion Detection Model Based on Mining Multi-Dimension Data Streams”. Journal of Computer Research and Development. 2009, 46(4):602-609.

[7] ZHANG Yong, TAN Xiao-Bin, CUI Xiao-Lin and XI Hong-Sheng. “Network Security Situation Awareness Approach Based on Markov Game Model”. Journal of Software. 2011, 22(3):495-508.

[8] CHEN Xiu-Zhen, ZHENG Qing-Hua, GUAN Xiao-Hong and LIN Chen-Guang. “Quantitative Hierarchical Threat Evaluation Model for Network Security”. Journal of Software. 2006, 17(4):885-897.

[9] Wei Yong, Lian Yifeng and Feng Dengguo. “A Network Security Situational Awareness Model Based on Information Fusion”. Journal of Computer Research and Development. 2009, 46(3):353-362.

[10] Yang Jia-xing. “Simulation Analysis of Complex Network Security Situation Assessment Model”. Computer Simulation. 2013, 30(8):289-292.

[11] RAO Xian, DONG Chun-Xi and YANG Shao-Quan. “An Intrusion Detection System Based on Support Vector Machine”. Journal of Software. 2003, 14(4):798-803.

[12] LIU Ji-ke, ZHAO Wei. “Response Surface Method for Reliability Computation Based on Support Vector Regression”. ACTA SCIENTIARUM NATURALIUM UNIVERSITATIS SUNYATSENI. 2008, 47(1):1-4.