Available online at www.sciencedirect.com
ScienceDirect
June 2017, 24(3): 33–43
www.sciencedirect.com/science/journal/10058885
The Journal of China Universities of Posts and Telecommunications
http://jcupt.bupt.edu.cn

Network security situation automatic prediction model based on accumulative CMA-ES optimization

Wang Jian[1] (✉), Li Ke[1], Zhao Guosheng[2]

1. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China
2. School of Computer Science and Information Engineering, Harbin Normal University, Harbin 150025, China

Abstract

To improve the accuracy of the network security situation, a security situation automatic prediction model based on accumulative data preprocess and support vector machine (SVM) optimized by covariance matrix adaptive evolutionary strategy (CMA-ES) is proposed. The proposed model adopts SVM which has strong nonlinear ability. Also, the hyper parameters for SVM are optimized through the CMA-ES which owns good performance in finding optimization automatically. Considering the irregularity of network security situation values, we accumulate the original sequence, so that the internal rules of discrete data can be revealed and it is easy to model. Simulation experiments show that the proposed model has faster convergence-speed and higher prediction accuracy than other extant prediction models.

Keywords security situation, automatic prediction, covariance matrix adaptive evolution strategy, support vector machine

1 Introduction

Network security situation can reflect the network status, and it is the important information of the active defense in the network. In order to determine the network status and make the accurate decision, it is necessary to perceive the network security situation in real time. Meanwhile, perceiving the network situation allows the network administrators to take the initiative to grasp the changes that will occur in the network. Furthermore, it can provide the basis for the network administrators to make the initiative decision. The network security situation awareness (NSSA), a comprehensive technology which obtains and processes the security information, has received extensive attention.

NSSA can be viewed as a three-phase process [1]: network security situation recognition, network security situation comprehension, and network security situation prediction (NSSP). The NSSA, a comprehensive technology which obtains and processes the security information, has received extensive attention. The research on NSSA mainly focuses on these three aspects, too. The NSSP, as the final step of the whole situation awareness, evaluates the current network security situation and forecasts the future network security situation. Harmer et al. [2] proposed metrics for how to analyze wireless network traffic in order to detect attacks. Zakrzewska et al. [3] employed Petri Nets to model real time cyber conflicts. Vu et al. [4] collected the vulnerability information of the network to evaluate the network security level. However, these scholars assessed the network security situation by extracting a certain point of situation factors and the situation factor extraction was too single. In Refs. [5–6], the authors introduced the alert correlation technology based on attack graph into NSSP. Although the method could predict the network security situation through the data association technology, the fusion data source was still a single source. Qi et al. [7] used fuzzy logic and rough set to deal with multiple sources of information, so as to assess the network security situation. A prediction method based on gray theory was proposed in Ref. [8]. In Ref. [9], Man et al. proposed a combined forecasting

Received date: 28-12-2016
Corresponding author: Wang Jian, E-mail: wangjianlydia@163.com
DOI: 10.1016/S1005-8885(17)60209-7

method. Klein et al. [10] had also done a lot of research on data fusion model and situation visualization. Tang et al. [11–12] used dynamic Bayesian networks and hidden Markov model to model insider threats, aiming to foresee future behavior. Although the method could fuse uncertain information, it was easy to cause dimension explosion when the amount of data was large, which made modeling more difficult. Qu et al. [13] fused alarm information, node vulnerability information and threat information by D-S evidence fusion technology. Chen et al. [14] proposed a hierarchical quantitative analysis method for information fusion. The two methods proposed in Refs. [13–14] had some deficiencies in multi-index relevance, and the experience of experts had a great impact on the results and lacked objective evidence. Based on the nonlinear of situation prediction, a neural network was introduced to solve the problem of situation prediction in Ref. [15]. Although the proposed method had advantages in nonlinear prediction, the neural network mainly relied on the principle of empirical risk minimization, which made the generalization ability of the method constrained. The SVM model was used to make the prediction in Ref. [16], which could make up for the shortcomings of neural network method. However, the blind selection of SVM parameters could not be neglected. Therefore, scholars began taking some in-depth exploration into parameter optimization. SVM could be optimized by particle swarm algorithm [17], artificial fish swarm algorithm [18], genetic algorithm (GA) [19] and so on. To a certain extent, these methods improved the blindness of parameter selection in SVM, but the models based on SVM optimized by these algorithms were easy to fall into premature. As a result, the prediction results cannot properly reflect the network security situation.

In order to solve the above problems, a network security situation automatic prediction model is proposed based on accumulative data preprocess method and SVM optimized by covariance matrix adaptive evolutionary strategy (CMA-ES-SVM). Firstly, SVM with better nonlinear capability is suitable for NSSP. Secondly, the proposed model uses the CMA-ES algorithm to optimize the hyper parameters of SVM, which has strong global searching ability and can update parameters automatically. Besides, the step size update often adapts nearly optimal step sizes usually leading to considerably larger step lengths, which improves convergence speed. Thirdly, it accumulates the original sequence so that the impact of disturbing factors in the original sequence is weakened, and the regularity of sequence is strengthened, and a new monotonically increasing sequence that is easy for CMA-ES-SVM to learn is obtained, which assists the prediction model. Ultimately, the better prediction model is achieved.

2 Modeling of NSSP system

NSSP is essentially a time series forecasting problem, and the time series prediction is to use the time series of the hidden relationship between the data to predict the future data. This paper presents the framework of the NSSP system consisting of four modules: data acquisition and evaluation module; data preprocessing module; network security situation value prediction module; warning strategy module. The framework of the system is shown in Fig. 1.

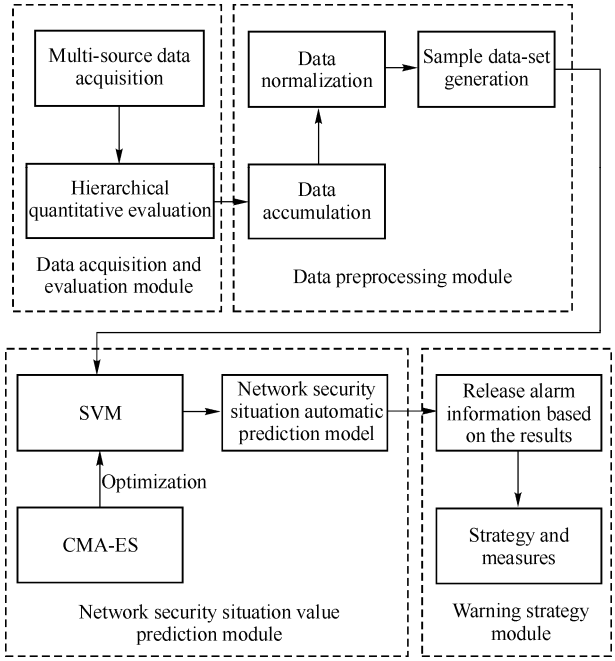

Fig. 1 Framework of NSSP system

The design of NSSP technology framework is based on the security situation forecasting process. It focuses on four key points: assessment, preprocessing, forecasting and alarming. Through the coordination of the four modules, the network security situation can be predicted accurately.

1) Firstly, collect multi-source security data, and then integrate the security factors into macro value according to hierarchical quantitative evaluation method [14].

2) The network security situation values are accumulated and normalized to generate a sample data set,

which is one of the innovative points of this paper.

3) Network security situation value prediction module is mainly composed of SVM regression algorithm and CMA-ES algorithm. Firstly, the training data set is input into SVM for training, and meanwhile the hyper parameters of SVM are optimized by CMA-ES algorithm. The mean square error (MSE) between the predicted value and the real value is used as the fitness function to determine whether to terminate the optimization process. If the stopping condition is satisfied, the current prediction model can be used as the final prediction model.

Otherwise, the optimization process continues to iterate until the final parameters are determined, so as to establish the optimal NSSP model. Using CMA-ES algorithm to optimize the SVM prediction model is the other innovation point of this paper.

4) Warning strategy module. To issue alarm information according to the forecast results, and make the corresponding protection measures.

3 Network security situation automatic prediction model based on accumulative CMA-ES-SVM

3.1 Accumulated and normalized data preprocessing

Due to the irregular and nonlinear of network security situation, it is difficult to establish an ideal prediction model. The accumulative process [20] does not affect the law hidden in the original sequence, but also makes the integral characteristics contained in scattered original data obvious. Therefore, all the original situation values $\{x^{(0)}(1), x^{(0)}(2),... x^{(0)}(i),...x^{(0)}(n)\}$ are accumulated one time and a new sequence $\left\{x^{(1)}(1),x^{(1)}(2),...,x^{(1)}(k),...,x^{(1)}(n)\right\}$ is obtained, where,

$$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i);\quad k=1,2,...,n,\ i=1,2,...,n \tag{1}$$

Because SVM is more sensitive to the data of 0~1, the training speed is faster, so the accumulated data should be normalized, the specific formula is as follows:

$$x^{(1)}(k)'=\frac{x^{(1)}(k)-x^{(1)}_{\min}(k)}{x^{(1)}_{\max}(k)-x^{(1)}_{\min}(k)};\quad k=1,2,...,n \tag{2}$$

where $x^{(1)}(k)$ denotes accumulated values. Among $x^{(1)}(k)$, $x^{(1)}_{\max}(k)$ and $x^{(1)}_{\min}(k)$ are the maximum and minimum values respectively. $x^{(1)}(k)'$ denotes accumulated and normalized values. One dimensional network security situation data preprocessed by the first two steps is converted into a multi-dimensional situation data sample by determining the embedding dimension a $(1<a<n)$, and the embedding dimension is determined by the trial and error method, and the multi-dimensional situation data are shown in Table 1.

Table 1 Multi-dimensional situation data

Training sample	Situation forecast value
$\boldsymbol{x}_1=\left[x^{(1)}(1)',x^{(1)}(2)',...,x^{(1)}(a)'\right]^{\mathrm{T}}$	$x^{(1)}(a+1)'$
$\boldsymbol{x}_2=\left[x^{(1)}(2)',x^{(1)}(3)',...,x^{(1)}(a+1)'\right]^{\mathrm{T}}$	$x^{(1)}(a+2)'$
...	...
$\boldsymbol{x}_{n-a+1}=\left[x^{(1)}(n-a+1)',x^{(1)}(n-a+2)',...,x^{(1)}(n)'\right]^{\mathrm{T}}$	$x^{(1)}(n+1)'$

3.2 The description of SVM

The basic idea [21] of SVM to solve the problem of regression prediction is to give an input sample $\boldsymbol{x}$, and it will be mapped to the high-dimensional feature space from the low-dimensional feature space through an appropriate non-linear map $\boldsymbol{\varphi}(\boldsymbol{x})$, then SVM can carry out linear prediction in high-dimensional feature space, so it has good nonlinear fitting capacity.

A given data sample $\left\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),...,(\boldsymbol{x}_{n-a+1},y_{n-a+1})\right\}$ is provided, where $\boldsymbol{x}_k\in\mathbb{R}^a$ $(k=1,...,n-a+1)$ denotes an input value, $y_k\in\mathbb{R}$ denotes an output value, n denotes the number of samples. The regression prediction function is expressed as follows:

$$f(\boldsymbol{x})=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\varphi}(\boldsymbol{x})+b \tag{3}$$

where $\boldsymbol{w}$ denotes the weight vector of the SVM hyper plane, b denotes bias.

Through analysis, we can know that the SVM regression prediction is, at bottom, to solve the following minimization problem:

$$\left.\begin{aligned}&\min_{\boldsymbol{w},b,\xi_k,\xi_k^*}\left\{\frac{1}{2}\|\boldsymbol{w}\|^2+c\sum_{k=1}^{n}(\xi_k+\xi_k^*)\right\}\\&\text{s.t.}\\&f(\boldsymbol{x}_k)-y_k\leqslant\varepsilon+\xi_k;\quad \xi_k,\xi_k^*\geqslant 0,k=1,2,...,n-a+1\\&y_k-f(\boldsymbol{x}_k)\leqslant\varepsilon+\xi_k^*;\quad \xi_k,\xi_k^*\geqslant 0,k=1,2,...,n-a+1\end{aligned}\right\} \tag{4}$$

where c denotes the penalty factor, ξ_k, ξ_k^* denote the

relaxation factor, ε denotes the insensitive loss function, which is used to control the regression error, and is defined as:

$$\psi_{\varepsilon}\left(f\left(\boldsymbol{x}_k\right),y_k\right)=\begin{cases}\left|f\left(\boldsymbol{x}_k\right)-y_k\right|-\varepsilon; & \left|f\left(\boldsymbol{x}_k\right)-y_k\right|\geqslant\varepsilon\\ 0; & \text{otherwise}\end{cases} \tag{5}$$

By introducing the Lagrange multiplier $\alpha_k\geqslant 0$, $\alpha_k^*\geqslant 0$, $\gamma_k\geqslant 0$, $\gamma_k^*\geqslant 0$, we can obtain:

$$\begin{aligned}\mathrm{L}\left(\boldsymbol{w},b,\alpha,\alpha^*,\xi,\xi^*,\gamma,\gamma^*\right)=&\frac{1}{2}\|\boldsymbol{w}\|^2+c\sum_{k=1}^{n-a+1}\left(\xi_k+\xi_k^*\right)-\\ &\sum_{k=1}^{n-a+1}\gamma_k\xi_k-\sum_{k=1}^{n-a+1}\gamma_k^*\xi_k^*+\sum_{k=1}^{n-a+1}\alpha_k\left(f\left(\boldsymbol{x}_k\right)-y_k-\varepsilon-\xi_k\right)+\\ &\sum_{k=1}^{n-a+1}\alpha_k^*\left(y_k-f\left(\boldsymbol{x}_k\right)-\varepsilon-\xi_k^*\right)\end{aligned} \tag{6}$$

Let partial derivatives are 0, we can obtain the following dual problem:

$$\left.\begin{aligned}&\max_{\alpha,\alpha^*}\left\{-\frac{1}{2}\sum_{k,j=1}^{n-a+1}\left(\alpha_k^*-\alpha_k\right)\left(\alpha_j^*-\alpha_j\right)\boldsymbol{\varphi}\left(\boldsymbol{x}_k\right)^{\mathrm{T}}\boldsymbol{\varphi}\left(\boldsymbol{x}_j\right)+\right.\\ &\left.\sum_{k=1}^{n-a+1}\left(\alpha_k-\alpha_k^*\right)y_k-\sum_{k=1}^{n-a+1}\left(\alpha_k+\alpha_k^*\right)\varepsilon\right\}\\ &\text{s.t.}\\ &\sum_{k=1}^{n-a+1}\left(\alpha_k^*-\alpha_k\right)=0;\quad \alpha_k,\alpha_k^*\in[0,c]\end{aligned}\right\} \tag{7}$$

According to Karush-Kuhn-Tucker (KKT) conditions, we obtain:

$$\boldsymbol{w}=\sum_{k=1}^{n-a+1}\left(\alpha_k^*-\alpha_k\right)\boldsymbol{\varphi}\left(\boldsymbol{x}_k\right) \tag{8}$$

$$b=y_k+\varepsilon-\sum_{j=1}^{n-a+1}\left(\alpha_j^*-\alpha_j\right)\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{x}_k;\quad \alpha_k,\alpha_k^*\in[0,c] \tag{9}$$

Putting Eqs. (8) and (9) into Eq. (3), we can obtain:

$$f\left(\boldsymbol{x}\right)=\sum_{k=1}^{n-a+1}\left(\alpha_k^*-\alpha_k\right)\boldsymbol{\varphi}\left(\boldsymbol{x}_k\right)\boldsymbol{\varphi}\left(\boldsymbol{x}\right)+b \tag{10}$$

In Eq. (7), $\boldsymbol{\varphi}\left(\boldsymbol{x}_k\right)\boldsymbol{\varphi}\left(\boldsymbol{x}_j\right)$ is inner product operation for higher dimensional space and the amount of computation is very large, and the nonlinear mapping $\boldsymbol{\varphi}\left(\boldsymbol{x}\right)$ is not easy to determine. To solve this problem, the kernel function is introduced.

$$K\left(\boldsymbol{x},\boldsymbol{x}_k\right)=\boldsymbol{\varphi}\left(\boldsymbol{x}_k\right)^{\mathrm{T}}\boldsymbol{\varphi}\left(\boldsymbol{x}\right) \tag{11}$$

The final expression of the regression function is:

$$f\left(\boldsymbol{x}\right)=\sum_{k=1}^{n-a+1}\left(\alpha_k^*-\alpha_k\right)K\left(\boldsymbol{x},\boldsymbol{x}_k\right)+b \tag{12}$$

Algorithm are as follows:

Step 1 Look for a kernel function $K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)$.

Step 2 Find the solutions α_i and α_i^* of the optimization problem in Eq. (7).

Step 3 Calculation $\boldsymbol{w}$, b according to Eqs. (8) and (9) and the obtained α_i, α_i^*.

Step 4 Construct a regression function $f\left(\boldsymbol{x}\right)$, as shown in Eq. (12).

3.3 Optimization and selection of SVM hyper parameters

Because of the excellent performance of CMA-ES in solving the global optimization problem, we use CMA-ES to optimize the hyper parameters of SVM in real time.

3.3.1 Problem formulation

After a large number of studies, Vapnik et al. found the kernel function was also an aspect of affecting SVM performance. The radial basis kernel function (RBF) $K\left(\boldsymbol{x}_k,\boldsymbol{x}\right)=\exp\left(-\left\|\boldsymbol{x}_k-\boldsymbol{x}\right\|^2/\left(2\sigma^2\right)\right)$ is the most widely used kernel function. RBF kernel function is applicable, whether in low-dimensional or high-dimensional space, for small sample or large samples and it has a wide convergence domain, so it is the ideal kernel function. Therefore, the RBF kernel function is selected as the SVM kernel function. At this point, the parameters which affect the performance of SVM are mainly the penalty parameter c, the width of RBF kernel function δ and the insensitive loss function ε. The so-called SVM parameter optimization selection is to find the optimal combination of parameters $\boldsymbol{u}=\left[c,\delta,\varepsilon\right]$, so that SVM has better predictive performance and improve the learning and generalization ability of SVM. The individual in the algorithm can be expressed as a combination of three-dimensional real numbers. The algorithm uses the MSE between the predicted value and the real value as the individual evaluation index (fitness function).

$$\tau_{\mathrm{MSE}}=\frac{1}{n-a+1}\sum_{k=1}^{n-a+1}\left(y_k-y_k'\right)^2 \tag{13}$$

The smaller the value of this index is, the higher the accuracy of the prediction is.

3.3.2 SVM hyper parameters optimization selection based on CMA-ES algorithm

The CMA-ES [22] is an improved global optimization algorithm based on the evolutionary strategy. It inherits the

advantages of the standard evolutionary algorithm and introduces the high guidance into evolutionary algorithm so that it avoids the dependence on population size and premature in the traditional evolutionary algorithm. In the CMA-ES, the adaptation of the covariance matrix is complemented with step size control. The adjustment of the step size is based on the relationship between the mean value of contemporary optimal subgroup and the mean value of the previous generation, so that the population converges to the global optimal solution adaptively. Cumulative path length control often adapts nearly optimal step sizes usually leading to considerably larger step lengths. This improves convergence speed and global search capability at the same time. Therefore, it is highly adaptable to solving complex optimization problems, and has been widely used in the field of optimization.

CMA-ES uses the Gaussian distribution $N\left(\boldsymbol{m},\sigma^2\boldsymbol{C}\right)$ to sample in the solution space of the optimization problem, and the population is generated by multivariate normal distribution. Where $\boldsymbol{m}$ denotes the center of the population distribution, $\boldsymbol{C}$ denotes the covariance matrix of the population, which reflects the shape of population distribution. On this basis, an additional parameter, global step size σ, is introduced as a global degree factor of $\boldsymbol{C}$. According to a certain sample selection mechanism, the first μ samples with the best fitness are selected from the samples to dynamically update the Gaussian distribution parameters $\boldsymbol{m}$, $\boldsymbol{C}$ and σ in real time. The sampling and updating process continue iteratively until a satisfactory solution is found or the maximum permission iterative number of times is reached.

The optimization algorithm of SVM parameters based on CMA-ES is as follows:

Step 1 Parameter setting and initialization.

We mainly set penalty parameters and the upper and lower bounds of kernel function parameters δ. We also need set the maximum number of iterations G and problem dimensions a, initial step size σ, search range, initialization strategy parameters and convergence conditions.

Step 2 Set the SVM corresponding data set, including training data set and test set.

Step 3 Population sampling.

The population which is composed of λ individuals $\boldsymbol{u}=[c,\delta,\varepsilon]$ is generated by multivariate normal distribution:

$$\boldsymbol{u}_t^{(g+1)}=\boldsymbol{m}^{(g)}+\sigma^{(g)}N_t\left(0,\boldsymbol{C}^{(g)}\right);\quad t=1,2,...,\lambda \tag{14}$$

where $\boldsymbol{u}_t^{(g+1)}$, $t=1,2,...,\lambda$ denotes the tth candidate of the $(g+1)$th generation population. $\boldsymbol{m}^{(g)}$ denotes the mean of the gth generation population distribution (it also denotes the center position of the population). $\sigma^{(g)}$ denotes the step size of the gth generation population distribution. $\boldsymbol{C}^{(g)}$ denotes the covariance matrix of the gth generation population distribution, which satisfies the following equation.

$$\boldsymbol{C}^{(g)}=\boldsymbol{B}^{(g)}\left(\boldsymbol{D}^{(g)}\right)^2\left(\boldsymbol{B}^{(g)}\right)^{\mathrm{T}} \tag{15}$$

where $\boldsymbol{B}^{(g)}$ is a orthogonal matrix, whose column vector is orthonormal basis of the eigenvector of $\boldsymbol{C}^{(g)}$, which is used to realize the rotation of hyper spherical sphere. $\boldsymbol{D}^{(g)}$ is a diagonal matrix, whose diagonal elements are the square root of the characteristic values of $\boldsymbol{C}^{(g)}$ and corresponding to each column vector of $\boldsymbol{B}^{(g)}$, which is used to realize the scale expansion of the hyper ellipsoidal surface of population distribution.

Step 4 Calculate the individual fitness of the population.

Using $\boldsymbol{u}=[c,\delta,\varepsilon]$ as the training parameter, SVM is trained by the training set, and the model is used to forecast the situation of the test set. The MSE is used as the fitness value of the individual. Train one by one and select the μ individuals with the smallest fitness function value to form the current optimal subgroup.

Step 5 Parameter update.

Substep 1 Update the mean $\boldsymbol{m}$.

$$\left.\begin{aligned}&\boldsymbol{m}^{(g+1)}=\sum_{t=1}^{\mu}w_t'\,\boldsymbol{u}_{t:\lambda}^{(g+1)}\\&\sum_{t=1}^{\mu}w_t'=1;\quad w_1'>w_2'>\cdots>w_\mu'>0\end{aligned}\right\} \tag{16}$$

where μ denotes the offspring population size. λ denotes the parent population size, w_t' denotes the weight coefficients, and $w_1'>w_2'>\cdots>w_\mu'>0$. $\boldsymbol{u}_{t:\lambda}$ denotes the tth solution in the λ solutions which are arranged in descending order according to the fitness values.

After the sampling and mean updating, the center position of the population will be moved to the nearby position of the offspring population, whose solutions are more excellent than parent population.

Substep 2 Update covariance matrix $\boldsymbol{C}$.

$$\left.\begin{aligned}\boldsymbol{C}^{(g+1)}&=\left(1-c_1-c_\mu\right)\boldsymbol{C}^{(g)}+c_1\left(\boldsymbol{p}_{\mathrm{c}}^{(g+1)}\boldsymbol{p}_{\mathrm{c}}^{(g+1)^{\mathrm{T}}}+\delta\left(h_\sigma^{(g+1)}\right)\boldsymbol{C}^{(g)}\right)+\\&\quad c_\mu\sum_{t=1}^{\mu}w_t'\,\boldsymbol{v}_{t:\lambda}^{(g+1)}\boldsymbol{v}_{t:\lambda}^{(g+1)^{\mathrm{T}}}\\\boldsymbol{p}_{\mathrm{c}}^{(g+1)}&=\left(1-c_{\mathrm{c}}\right)\boldsymbol{p}_{\mathrm{c}}^{(g)}+h_\sigma^{(g+1)}\sqrt{c_{\mathrm{c}}\left(2-c_{\mathrm{c}}\right)}\sqrt{\mu_{\mathrm{eff}}}\,\frac{\boldsymbol{m}^{(g+1)}-\boldsymbol{m}^{(g)}}{\sigma^{(g)}}\\\boldsymbol{v}_{t:\lambda}^{(g+1)}&=\frac{\boldsymbol{u}_{t:\lambda}^{(g+1)}-\boldsymbol{m}^{(g)}}{\sigma^{(g)}}\end{aligned}\right\}\tag{17}$$

where c_1 and c_μ are the learning rate for updating the covariance matrix. $\boldsymbol{p}_{\mathrm{c}}$ denotes the evolution path, the initial $\boldsymbol{p}_{\mathrm{c}}=\boldsymbol{0}$. $c_{\mathrm{c}}\leqslant 1$ denotes the backward time horizon of the evolution path.

Substep 3 Update step size.

$$\left.\begin{aligned}\sigma^{(g+1)}&=\sigma^{(g)}\mathrm{e}^{\frac{c_\sigma}{d_\sigma}\left(\frac{\left\|\boldsymbol{p}_\sigma^{(g+1)}\right\|}{E\left(\left\|N(0,\boldsymbol{I})\right\|\right)}-1\right)}\\\boldsymbol{p}_\sigma^{(g+1)}&=\left(1-c_\sigma\right)\boldsymbol{p}_\sigma^{(g)}+\sqrt{c_\sigma\left(2-c_\sigma\right)}\sqrt{\mu_{\mathrm{eff}}}\,\boldsymbol{C}^{-\frac{1}{2}}\cdot\\&\quad\frac{\boldsymbol{m}^{(g+1)}-\boldsymbol{m}^{(g)}}{\sigma^{(g)}}\end{aligned}\right\}\tag{18}$$

where, $\exp\left(\left(c_\sigma/d_\sigma\right)\left(\left\|\boldsymbol{p}_\sigma^{(g+1)}\right\|/E\left(\left\|N(0,\boldsymbol{I})\right\|\right)-1\right)\right)$ can be regarded as the expansion factor of the step size. d_σ denotes a damping parameter. $E\left(\left\|N(0,\boldsymbol{I})\right\|\right)$ denotes the expectation of the Euclidean norm of a $N(0,\boldsymbol{I})$ distributed random vector. c_σ denotes the backward time horizon. $\boldsymbol{p}_\sigma$ denotes the conjugate evolution, the initial $\boldsymbol{p}_\sigma=\boldsymbol{0}$.

Step 6 Is it up to the stop condition? If so, stop and output of the optimal individual $\boldsymbol{u}^*=\left[c^*,\sigma^*,\varepsilon^*\right]$ and the optimal fitness value, otherwise return to Step 3.

It can be seen that the updating of parameters of CMA-ES takes advantage of the so-called 'evolutionary path', especially the updating of the covariance matrix, which can automatically adjust the parameters adaptively to the change of the population size with 'rank 1' and 'rank A'. The algorithm equipped with high guidance of covariance matrix and effective global step length, which makes the evolution process highly efficient.

3.4 NSSP based on accumulative CMA-ES-SVM

In this paper, the NSSP model is divided into the following steps:

1) Collect the information of network security situation, such as attack type, attack strength, network topology and other configuration information and so on. Analyze network situation through situation understanding and situation assessment. Quantitative calculation is carried out according to the hierarchical quantitative evaluation method, and the original trend value sequence is generated.

2) A series of new monotone increasing trend is obtained by accumulating the network security situation value. Moreover, SVM is more sensitive to the data of 0~1, so the accumulated situation value sequence is normalized by Eq. (2).

3) The data after preprocessing is divided into two parts: training data set and test data set. The training data set is input to the SVM to train the model, and its hyper parameters are optimized by CMA-ES algorithm. The optimal parameters are obtained and the optimal situation prediction model is built with the optimal parameters.

4) Test the generalization performance of the model with the test set data. According to Eq. (19), the predicted values produced by the model are anti-normalizing. Then, execute the inverse process of accumulation according to Eq. (20).

$$x^{(1)}(k)=x^{(1)}(k)'\left(x_{\max}^{(1)}(k)-x_{\min}^{(1)}(k)\right)+x_{\min}^{(1)}(k)\tag{19}$$

$$x^{(0)}(i+1)=x^{(1)}(i+1)-x^{(1)}(i);\quad i=n,n+1,n+2,...\tag{20}$$

The workflow of the NSSP model in this paper is summarized as shown in Fig. 2.

The model has wide application prospect and can be used for real-time monitoring of internal network security situation and can also be used for real-time monitoring of information critical system. The macroscopic quantitative prediction of the network security situation can make the network administer macroscopically control the overall security situation of the network, understand the changing trend of the network security situation, and discover the insecurity factors of the network in time, and make reasonable response to the threat of strategy, then make active defense, which can ensure that the network provide efficient services consistently.

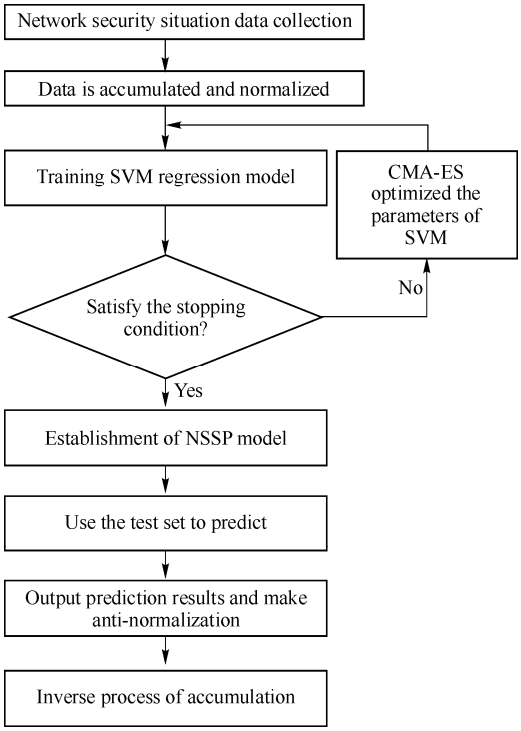

Fig. 2 Workflow chart of NSSP model

4 Experimental analysis

In this paper, we adopt the real safety data provided by the network security protection log of a network center in 2015. Here three important attacks that include deny of service (DoS) attack (it makes some services stop), Injection attack (insert malicious code in to the servers) and Port scanning attack (it is a common way to search target computer by hackers) are chosen. These attacks are sampled once every three days, a total of 120 samples. We effectively obtain the macro network security situation values by using the hierarchical quantitative evaluation method proposed in Ref. [14]. In order to ensure the results more realistic and credible, 10-fold cross-validations are carried out, where the 120 original samples are divided into 10 subsamples, and each subsample contains 12 samples. The situation values of each group were preprocessed by the first two steps mentioned in Sect. 3.1, and the processing results of the subsample set 1 are shown in Fig. 3. Due to limited space, the other 9 subsamples are not listed. Subsequently, the normalized sample sequence is reconstructed, and the trial and error method is used to determine the reconstructed dimension and the reconstructed dimension is 3. That is to say, the input variable is 3-dimensional (3D), so there are 9 sample pairs in each data set. In these subsamples, one subsample is randomly selected as the testing data, and the remaining nine subsamples are treated as training data. The cross-validation is repeated 10 times, and each subsample has been selected as the testing data, making full use of the experimental data.

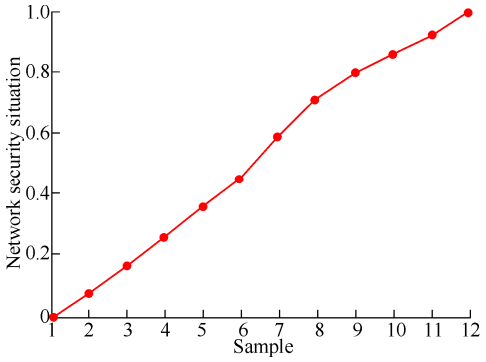

Fig. 3 Network security situation value of sub dataset 1

The SVM prediction model is established by the optimization method proposed in this paper, then initialize parameters, let the initial mean (population center) $\boldsymbol{m}^{(0)} = \boldsymbol{u}^{(0)} = [c, \delta, \varepsilon]^{(0)}$, the max-epoch for the training is 100, the SVM penalty parameter c and the RBF kernel function width δ are all set in the $[0.01, 100]$, the termination threshold is 1×10^{-5}. At the same time, we use the CMA-ES algorithm to optimize the SVM hyper parameters $[c, \delta, \varepsilon]$, and then we use the optimal solution $[c^*, \delta^*, \varepsilon^*]$ to establish prediction model. After repeated experiments, the optimal solution is obtained: $\varepsilon^* = 0.001\,6$, $c^* = 83$, $\delta^* = 2.21$. The predicted values generated by the CMA-ES-SVM prediction model trained by 10-fold cross-validations are shown in Fig. 4, and the MSEs of the predicted values in 10-fold cross-validations are shown in Table 2.

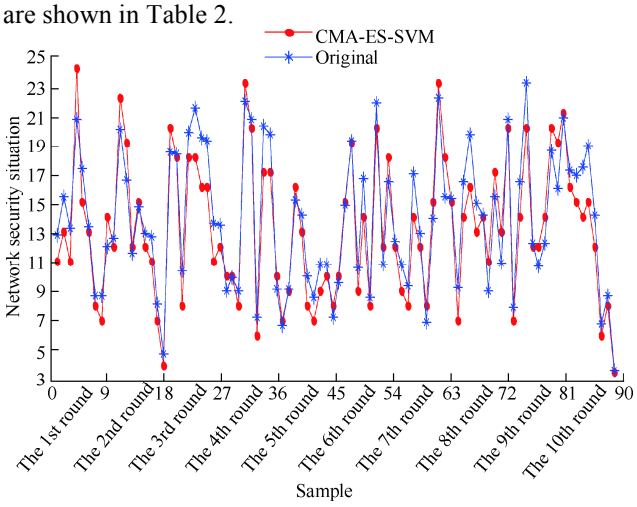

Fig. 4 The forecasting values by the CMA-ES-SVM model with 10 round cross-validations

Table 2 The MSEs of the forecasting results by the CMA-ES-SVM model with 10-fold cross-validations

Round	MSE
The 1 st	$3.491\,0\times 10^{-4}$
The 2 nd	$7.638\,1\times 10^{-4}$
The 3 rd	$2.512\,1\times 10^{-3}$
The 4 th	$8.131\,2\times 10^{-4}$
The 5 th	$2.103\,0\times 10^{-3}$
The 6 th	$7.910\,1\times 10^{-4}$
The 7 th	$1.031\,2\times 10^{-3}$
The 8 th	$2.910\,3\times 10^{-4}$
The 9 th	$1.921\,6\times 10^{-4}$
The 10 th	$6.919\,2\times 10^{-4}$

From Fig. 4, it can be seen that the forecasting values fit the original security situation well in each cross-validation. This proves that the proposed model can predict the network security situation effectively.

In order to further demonstrate the superiority of the proposed model, the following three comparative studies are carried out.

1) The comparative study between CMA-ES-SVM model and SVM model.

The same dataset is used in both two models. 10-fold cross-validations are carried to compare the SVM prediction model with the CMA-ES-SVM model. Fig. 5 shows the predicted values generated by these two models. Table 3 lists the average MSE and the average mean radial error (MRE) of the two models for prediction.

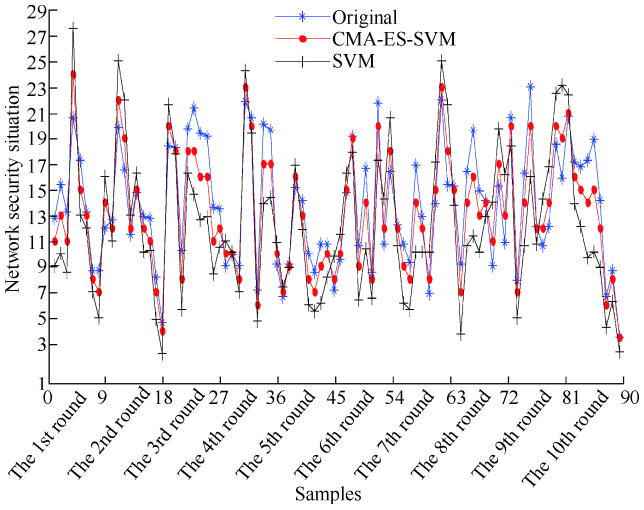

Fig. 5 The comparative results between CMA-ES-SVM model and SVM model with 10 round cross-validations

Table 3 The average MSEs and the average MREs generated by two model

Model	Average MSE	Average MRE
SVM	$5.370\,3\times 10^{-2}$	$9.013\,1\times 10^{-1}$
CMA-ES-SVM	$9.538\,5\times 10^{-4}$	$3.529\,1\times 10^{-1}$

As can be seen from Fig. 5, the red curve is closer to the blue curve than the black one, which means that the predicted values generated by the CMA-ES-SVM model are closer to the true value. As can be seen from Table 3, the average MSE of the predicted value generated by the CMA-ES-SVM prediction model is two orders of magnitude smaller than SVM model's, which means that the prediction error rate of the proposed model is smaller. This also reflects that CMA-ES plays an important role in SVM parameter selection. Therefore, it can be concluded that the CMA-ES-SVM model is more accurate than SVM prediction model. Meanwhile, this experiment can prove the validity of the model.

2) The comparative study between the CMA-ES-SVM prediction model and the SVM prediction model with other optimization algorithm.

There are a number of other optimization algorithms that can be used to optimize the parameters of the SVM prediction model. In this paper, three state-of-the-art optimization algorithms, particle swarm optimization (PSO) algorithm, GA and simulated annealing (SA) algorithm [23] are selected for comparative analysis. Meanwhile, convergence speed and prediction accuracy are selected as the performance evaluation index.

a) Convergence speed

Referring to Fig. 6, the abscissa represents the generation, and the ordinate represents the root mean square error (RMSE).

$$\tau_{\mathrm{RMSE}}=\sqrt{\tau_{\mathrm{MSE}}} \tag{21}$$

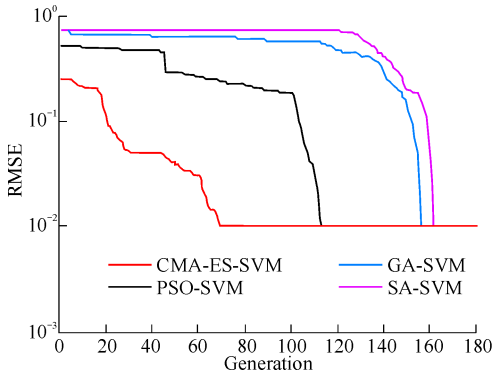

Fig. 6 Convergence speed of prediction models with 10 round cross-validations

We see that the different algorithms have different convergence speeds. The convergence speed of CMA-ES-SVM is fastest. It converges at the 69th generation. The second fastest algorithm is PSO-SVM which ends at the 113th generation. GA-SVM and

SA-SVM stop running to reach convergence precision at the 156th and 163th generations respectively. So, we draw a conclusion that the model combining CMA-ES and SVM reduces the convergence time effectively and improves the prediction efficiency.

b) Prediction accuracy

In this paper, we choose absolute error (AE), the MSE, mean relative error (MRE) as the judgment criteria to measure the accuracy of the algorithm.

$$\tau_{\mathrm{AE}} = y_k - y_k' \tag{22}$$

$$\tau_{\mathrm{MRE}} = \frac{1}{n-a+1}\sum_{k=1}^{n-a+1}\left|\frac{y_k - y_k'}{y_k'}\right| \tag{23}$$

Fig. 7 shows the AE generated by the four optimization algorithms (as space is limited, 10 sets of data will not be all listed here, only one set is listed). Fig. 8 shows the NSSP values generated by the four optimization algorithms.

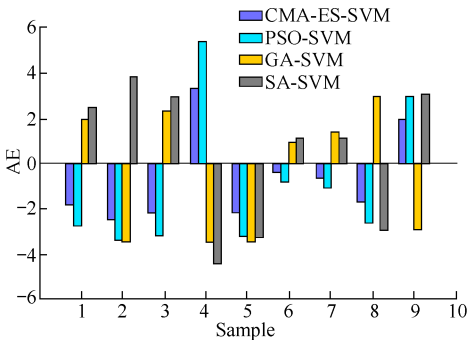

Fig. 7 AE bar graphs for prediction models

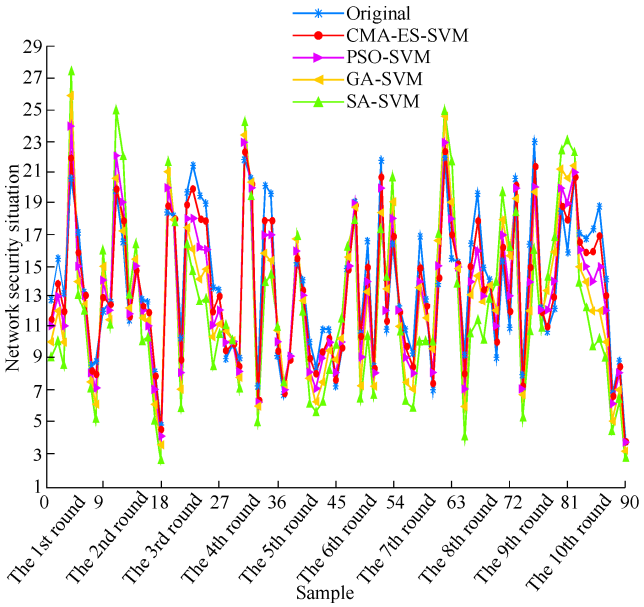

Fig. 8 The comparative results predicted by four models with 10 round cross-validations

Based on Figs. 7 and 8, we can find that four prediction models have achieved good prediction effect fairly. But CMA-ES-SVM is closer to the real network situational values and has smaller value in prediction error than PSO-SVM, GA-SVM, SA-SVM. Based on the analysis of the results the main reason is that CMA-ES-SVM adopts CMA-ES to optimize SVM so that it has higher fitting capacity in situation values. As a result, the premature converge is avoid effectively. Consequently, CMA-ES-SVM can converge a better solution.

Table 4 shows the average MSEs and the average MREs of the four algorithms.

Table 4 Prediction model accuracy inspection standard

Model	Average MSE	Average MRE
CMA-ES-SVM	$9.538\,5 \times 10^{-4}$	$3.529\,1 \times 10^{-1}$
PSO-SVM	$1.272\,4 \times 10^{-3}$	$4.256\,1 \times 10^{-1}$
GA-SVM	$1.452\,5 \times 10^{-3}$	$5.382\,5 \times 10^{-1}$
SA-SVM	$1.472\,1 \times 10^{-3}$	$5.529\,2 \times 10^{-1}$

As can be seen from Table 4, the average MSE of the predicted values generated by the CMA-ES-SVM prediction model is one order of magnitude smaller than PSO-SVM model's. Considering the average MSE and the average MRE, we can conclude that CMA-ES-SVM has the lowest prediction error value of the four models, which means that CMA-ES-SVM has the strongest capacity of prediction and CMA-ES is better than the other three methods for optimizing SVM hyper parameters.

3) The comparative study between accumulative CMA-ES-SVM prediction model and CMA-ES-SVM model

In order to prove that the performance of CMA-ES-SVM model is improved by calculation, the CMA-ES-SVM model which is not preprocessed by data accumulation is selected and compared with the proposed model in this paper. Fig. 9 shows the values produced by the two prediction models. Table 5 shows the average MSEs and the average MREs of the two models.

Table 5 The average MSEs and the average MREs generated by different model

Model	Average MSE	Average MRE
Accumulative CMA-ES-SVM	$5.420\,6 \times 10^{-4}$	$2.315\,3 \times 10^{-1}$
CMA-ES-SVM	$9.538\,5 \times 10^{-4}$	$3.529\,1 \times 10^{-1}$

As can be seen from Fig. 9, the green curve is closer to the blue curve than the red one, which means that the predicted values of the accumulative CMA-ES-SVM prediction model is closer to the true value. As can be seen from Table 5, the average MSE generated by the model which is not preprocessed by data accumulation is about 2

times than the average MSE generated by accumulative CMA-ES-SVM model. Considering the average MSE and the average MRE, we can conclude that accumulative CMA-ES-SVM has the lowest prediction error value of the four models. That is to say the accumulative method can effectively improve the prediction accuracy. This is because, to a certain extent, the accumulative method weakens the irregularity of the original sequence, highlights the regularity of the sequence, so that the model can better reflect the real situation.

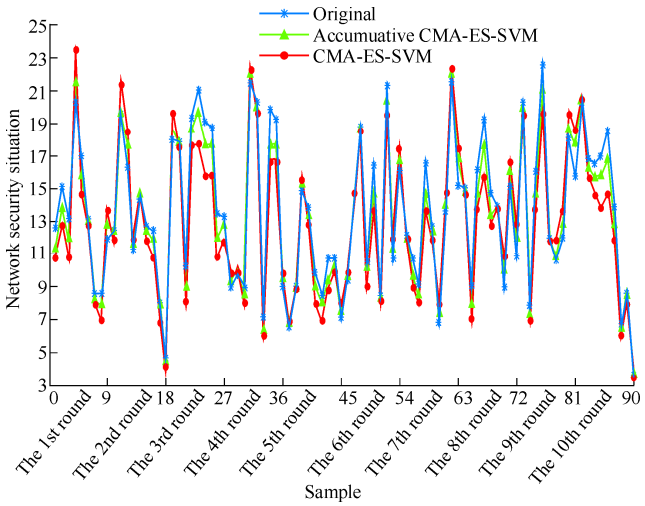

Fig. 9 The comparative results between accumulative CMA-ES-SVM model and PSO-SVM model with 10 round cross-validations

In contrast to Figs. 4 and 5, Figs. 8 and 9, we can see that the predicted value generated by accumulative CMA-ES-SVM model is closest to the real value, that is, the proposed accumulative CMA-ES-SVM prediction model has better prediction ability and performance.

5 Conclusions

NSSP plays a very important role on discovering potential threats, malicious attacks and reducing the corresponding damage. It is very important to formulate corresponding strategies of network security defense and to improve the speed of emergency response. In this paper, we put forward a NSSP model, which is based on combining accumulative data preprocess and SVM with the parameters optimized by the CMA-ES algorithm. Hence, the proposed model can effectively predict the network security situation and the prediction accuracy is higher and the convergence speed is faster. At the same time, we believe that the network security active defense should be combined with the visualization of the situation forecast; therefore, we will have the situation visualization research, seeking a good interactive visual display mode.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61403109, 61202458), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20112303120007), the Specialized Research Fund for Scientific and Technological Innovation Talents of Harbin (2016RAQXJ036).

References

1. Jajodia S, Liu P, Swarup V, et al. Cyber situational awareness. Berlin, Germany: Springer, 2010
2. Harmer P, Thomas R, Christel B, et al. Wireless security situation awareness with attack identification decision support. Proceedings of the 2011 IEEE Symposium on Computational Intelligence in Cyber Security (CICS'11), Apr 11–15, 2011, Paris, France. Piscataway, NJ, USA: IEEE, 2011: 144–151
3. Zakrzewska A N, Ferragut E M. Modeling cyber conflicts using an extended Petri net formalism. Proceedings of the Computational Intelligence in Cyber Security (CICS'11), Apr 11–15, 2011, Paris, France. Piscataway, NJ, USA: IEEE, 2011: 60–67
4. Vu H L, Khaw K K, Chen T Y. A new approach for network vulnerability analysis. Proceedings of the 33rd IEEE Conference on Local Computer Networks (LCN'08), Oct 14–17, 2008, Montreal, Canada. Piscataway, NJ, USA: IEEE, 2008: 200–206
5. Ahmadinejad S H, Jalili S, Abadi M. A hybrid model for correlating alerts of known and unknown attack scenarios and updating attack graphs. Computer Networks, 2011, 55(9): 2221–2240
6. Elshoush H T, Osman I M. Alert correlation in collaborative intelligent intrusion detection systems—a survey. Applied Soft Computing, 2011, 11(7): 4349–4365
7. Qi Y L, An H N. The evaluation model of network security based on fuzzy rough sets. Luo Q (ed). Advances in Wireless Networks and Information Systems. LNEE 72. Berlin, Germany: Springer, 2010: 517–525
8. Dong J F. The building of network security situation evaluation and prediction model based on grey theory. Proceedings of the 2010 International Conference on Challenges in Environmental Science and Computer Engineering (CESCE'10): Vol 2, Mar 6–7, 2010, Wuhan, China. Piscataway, NJ, USA: IEEE, 2010: 401–404
9. Man D P, Wang Y, Yang W, et al. A combined prediction method for network security situation. Proceedings of the 2010 International Conference on Computational Intelligence and Software Engineering (CiSE'10), Dec 10–12, 2010, Wuhan, China. Piscataway, NJ, USA: IEEE, 2010: 4p
10. Klein G, Günther H, Träber S. Modularizing cyber defense situational awareness-technical integration before human understanding. Aschenbruck N, Martini P, Meier M, et al (eds). Future Security. CCIS 318. Berlin, Germany: Springer, 2012: 307–310
11. Tang K, Zhou M T, Wang W Y. Insider cyber threat situational awareness framwork using dynamic Bayesian networks. Proceedings of the 4th International Conference on Computer Science and Education (ICCSE'09) , Jul 25–28, 2009, London, UK. Piscataway, NJ, USA: IEEE, 2009: 1146–1150
12. Liang Y, Wang H Q, Pang Y G. A kind of formal modelling for network

security situational awareness based on HMM. Proceedings of the 9th International Conference on Web-Age Information Management (WAIM'08), Jul 20–22, 2008, Zhangjiajie, China. Piscataway, NJ, USA: IEEE, 2008: 598–605
13. Qu Z Y, Li Y Y, Li P. A network security situation evaluation method based on DS evidence theory. Proceedings of the 2nd International Conference on Environmental Science and Information Application Technology (ESIAT'10): Vol 2, Jul 17–18, 2010, Wuhan, China. Piscataway, NJ, USA: IEEE, 2010: 496–499
14. Chen X Z, Zheng Q H, Guan X H, et al. Quantitative hierarchical threat evaluation model for network security. Journal of Software, 2006, 17(4): 885–897 (in Chinese)
15. Zhang H B, Huang Q, Li F W, et al. A network security situation prediction model based on wavelet neural network with optimized parameters. Digital Communications and Networks, 2016, 2(3): 139–144
16. Chen J, Tu X G. Network security risk assessment based on support vector machine. Proceedings of the IEEE 3rd International Conference on Communication Software and Networks (ICCSN'11), May 27–29, 2011, Xi'an, China. Piscataway, NJ, USA: IEEE, 2011: 184–187
17. Bamakan S M H, Wang H D, Tian Y J, et al. An effective intrusion detection framework based on MCLP/SVM optimized by time-varying chaos particle swarm optimization. Neurocomputing, 2016, 199: 90–102
18. Gao Y Y, Shen Y J, Zhang G D, et al. Information security risk assessment model based on optimized support vector machine with artificial fish swarm algorithm. Proceedings of the IEEE 6th International Conference on Software Engineering and Service Science (ICSESS'15), Sept 23–25, 2015, Beijing, China. Piscataway, NJ, USA: IEEE, 2015: 599–602
19. Zeng B, Zhong P. Simulation research on network security situation prediction method. Computer Simulation, 2012, 29 (5): 170–173 (in Chinese)
20. Deng J L. Grey system theory and course. Wuhan, China: Huazhong University of Science and Technology Press, 1990: 89–90 (in Chinese)
21. Wu C H, Ho J M, Lee D T. Travel-time prediction with support vector regression. IEEE Transactions on Intelligent Transportation Systems, 2004, 5(4): 276–281
22. Hu G Y, Zhou Z J, Zhang B C, et al. A method for predicting the network security situation based on hidden BRB model and revised CMA-ES algorithm. Applied Soft Computing, 2016, 48: 404–418
23. Lin S W, Lee Z J, Chen S C, et al. Parameter determination of support vector machine and feature selection using simulated annealing approach. Applied Soft Computing, 2008, 8(4): 1505–1512

(Editor: Wang Xuying)