Research on Vehicle Network Security Situation Prediction Based on Improved CLPSO-RBF

Ruxiang Li, Fei Li*, Chunwang Wu, Jiaqi Song
*School of Cyberspace Security, Chengdu University of Information Technology, Chengdu 610225, China

*lifei@cuit.edu.cn

Abstract. In order to improve the accuracy and efficiency of the vehicle network security situation prediction, a prediction algorithm based on improved CLPSO-RBF is proposed. Firstly, for speeding up the optimization efficiency of CLPSO, a reasonable speed monitoring variable has been introduced. Secondly, we use the improved CLPSO algorithm to optimize the clustering radius of the RBF neural network, so that the optimal RBF network structure can be determined. Finally, we use the optimal RBF network to predict the security situation of the vehicle network. Simulation experiments have proved that the improved algorithm has higher accuracy and faster convergence rate in situation prediction, and has better prediction effects.

Keywords: Vehicle Network; Security Situation Prediction; CLPSO; RBF

1. Introduction

With the development of the Internet and industrial intelligence, China’s automobile industry is rapidly shifting to digitalization and intelligence. Besides, the market scale of the vehicle network is also growing rapidly [1]. At the same time, network security incidents related to the vehicle network are also constantly occurring.

In order to deal with various types of vehicle network security threats, more and more car companies have introduced security defense equipment such as intrusion detection system and firewall. However, these devices are independent of each other, and in many cases, they cannot cooperate with each other, nor can they effectively defend against security threats to the vehicle network. Therefore, it has become a trend to use network security situational awareness technology to build a network security defense system. Network security situation awareness technology can evaluate the current network security situation based on the collected security information, and predict the security situation for a period of time based on the current situation level, so as to help managers make the next decision.

Researchers have conducted a lot of research on situation prediction. Sun Weixi combines PSO and support vector machine to predict the network security situation, which improves the prediction accuracy, but the algorithm time complexity is relatively large [2]. Jiang Yang et al. combined PSO with RBF for situation prediction and used inertial weighting factors to improve PSO, but the algorithm was more complicated and the learning time was longer [3]. Jin Xin et al. introduced the Spark framework and combined with PSO to improve the neural network, but this method is also used...
in the forecast of the power communication network [4]. Li Fangwei proposed a prediction method based on the hybrid kernel function PSO-SVR, which improves the prediction ability, but iterative optimization times are large [5]. Chen Shanxue uses PSO-SVM to predict the security situation, which effectively improves the prediction accuracy, but the data preprocessing part is more complicated, so this method is not used in the vehicle network environment [6]. It can be seen that it has become a trend to combine the PSO algorithm in the situation prediction because the PSO algorithm can improve the optimization accuracy and is easy to construct. But PSO still has shortcomings such as slow convergence speed and easy to fall into local optimum. Therefore, some scholars have improved the procedure of PSO, namely CLPSO. CLPSO can expand the scope of optimization and avoid premature convergence, so it is more suitable for combining with RBF to predict the situation of the vehicle network. Based on this, we propose a prediction method based on improved CLPSO-RBF to predict the network security situation of the vehicle network.

2. RBF neural network
RBF neural network is a three-layer forward neural network composed of an input layer, hidden layer, and output layer. Among them, the input layer contains signal source nodes, and the hidden layer contains several radial basis functions. These radial basis functions can map the input linearly inseparable data to the linearly separable space, and the output layer is composed of the training results of the network. RBF neural network has the characteristics of strong learning ability and fast convergence speed, and its nonlinear fitting ability is also prominent. Figure 1 is a structure diagram of an RBF neural network, which has \( p \) input nodes, \( n \) hidden nodes, and \( j \) output nodes.

![RBF neural network structure diagram](image)

The core part of RBF neural network is the hidden layer. The hidden layer uses the radial basis function as the activation function, and the commonly used activation function is the Gaussian function, which can be expressed as equation (1).

\[
\phi_i(x_p) = \exp\left(-\frac{1}{2\sigma^2}\|x_p - c_i\|^2\right)
\]  

(1)
In equation (1), $x_p$ is the input sample $p$, $c_i$ is the center of the Gaussian function, $\sigma$ is the variance of the Gaussian function, and $\|x_p - c_i\|$ is the Euclidean distance. The output obtained by the RBF neural network can be expressed as equation (2).

$$y_j = \sum_{i=1}^{n} w_{ij} \phi_i = \sum_{i=1}^{n} w_{ij} \exp\left(-\frac{1}{2\sigma^2}\|x_p - c_i\|^2\right)$$

(2)

In equation (2), $y_j$ is the output corresponding to the input sample $j$, $c_i$ is the node center of the hidden layer, $w_{ij}$ is the connection weight of the hidden layer to the output layer, and $c_i$ is the center of the Gaussian function.

When the RBF neural network is learning, it first needs to cluster the input sample data, calculate the center $c_i$ and $\sigma$ of each node in the hidden layer, and then train the network to obtain the connection weight $w_{ij}$ of the hidden layer to the output layer.

The nearest neighbor clustering algorithm is commonly used in clustering, and the key of the algorithm is to determine the clustering radius, so the selection of $\lambda$ is crucial to the selection of the hidden layer node center $c_i$ of the RBF network. If the $\lambda$ is too large, the number of $c_i$ will be too small, resulting in a large network convergence error. If the $\lambda$ is too small, the number of $c_i$ will be too large, which will cause overfitting, resulting in poor network generalization ability, and training time will be longer. Therefore, we will combine the improved CLPSO algorithm to select the appropriate clustering radius $\lambda$, therefore, the optimal RBF network parameters can be determined.

3. Prediction algorithm based on improved CLPSO-RBF

3.1. Improved algorithm of CLPSO

3.1.1. CLPSO

The standard particle swarm optimization algorithm (PSO) is an intelligent optimization algorithm that simulates bird swarm foraging cooperation. In the algorithm, each particle is regarded as an alternative solution to the problem, and the particle updates its speed and position through individual extreme values and group extreme values in each iteration [7]. For example, in a $D$ dimensional search space, the position and velocity update formulas of the generation $t$ of the particle $i$ are as follows:

$$V_i(t+1) = V_i(t) + c_1 \times rand_1(t) \times (pbest_i(t) - X_i(t)) + c_2 \times rand_2(t) \times (gbest(t) - X_i(t))$$

(3)

$$X_i(t+1) = X_i(t) + V_i(t+1)$$

(4)

In equation (3), $rand_1$ and $rand_2$ are random numbers between 0 and 1, $pbest_i$ is the historical extreme value of the particle $i$ in $D$ dimensional space, and $gbest_i$ is the current group extreme value of all particles in the entire particle swarm in $D$ dimensional space. $c_1$ and $c_2$ are the weights of each particle flying to $pbest$ and $gbest$. However, in the case of the multi-peak problem, when the global extremum $gbest$ is in a local optimal position, every particle will be attracted by it, which will cause the particle swarm to fall into the local optimal position, resulting in poor fitness. Therefore, J.J.Liang et al. proposed an improved algorithm of PSO, namely CLPSO algorithm. The learning strategy adopted by this algorithm is: in a $D$ dimensional search space, the generation $t$ of the particle $i$ will use the historical extreme values of other particles to update its speed. The speed update formula can be expressed as equation (5).
\[ V_i(t + 1) = w \times V_i(t) + c_1 \times \text{rand}_1(t) \times (pbest_i(t) - X_i(t)) \]  

(5)

In equation (5), \( pbest_i \) represents the historical extremum of a certain particle selected with a certain probability from all the historical extremums of the particles, which is used as the learning object to expand the search range and avoid falling into the local optimum. Among them, the learning probability \( P_c \) of particle \( i \) is:

\[ P_{ci} = 0.05 + 0.45 \times \frac{\exp \left( \frac{5(i-1)}{S-1} \right) - 1}{\exp(5) - 1} \]  

(6)

In equation (6), \( S \) is the overall scale, and \( i \) is the particle number. However, the new learning strategy ignores the sociology of the particle swarm in the original PSO algorithm. While guiding the particles to fly to the current swarm extremum \( pbest \), it also slows down the convergence speed and reduces the optimization efficiency of the algorithm.

3.1.2. CLPSO

Aiming at the problem of CLPSO algorithm convergence speed being too slow, He Sheng et al. proposed to restore the group extremum \( gbest \) to speed up the convergence speed. The speed update formula is:

\[ V_i(t + 1) = w \times V_i(t) + c_1 \times \text{rand}_1(t) \times (pbest_i(t) - X_i(t)) + c_2 \times \text{rand}_2(t) \times (gbest(t) - X_i(t)) \]  

(7)

Besides, they introduce the speed monitoring variable \( v_{monitor} \) to monitor the algorithm process, and reinitialize the algorithm process when the \( v_{monitor} \) is less than a certain threshold \( v_{threshold} \). At the same time, set \( renumber \) to monitor the number of reinitialization, and set \( c_2 \) to 0 when \( renumber > 3 \), and return to the CLPSO algorithm. However, He Sheng et al. did not provide a basis for determining the speed monitoring variable \( v_{monitor} \). So we will continue to propose algorithm improvements on its basis.

For the velocity monitoring variable \( v_{monitor} \), its value will be calculated by selecting the velocity of the latest three generations of particle \( i \) as equation (6).

\[ v_{monitor} = \frac{\left| (V_i(t) - V_i(t-1)) + (V_i(t-1) - V_i(t-2)) + (V_i(t-2) - V_i(t-3)) \right|}{3} \]  

(8)

Taking account of the change of the speed, the threshold is: \( v_{threshold} = 10 \times e^{-3} \).

In terms of fitness function, the Griewank function is selected, which is a commonly used function to test the efficiency of optimization programs and has great advantages in global optimization.

The process of the improved CLPSO algorithm is as follows:

1. Initialize the variables: initialize \( v_{threshold} = 10 \times e^{-3} \), \( renumber = 0 \), initialize the position and velocity of the particle, \( pbest \) is the current value of the particle, and \( gbest \) is the historical extreme value of the particle.
(2) Update the particle velocity and position according to formulas (7) and (4). Calculate the fitness of each particle, compare the objective function value corresponding to the current fitness and the objective function value corresponding to gbest, and update the smaller of the two.

(3) Determine whether the vmonitor is less than the threshold, if so, save the optimization result, otherwise reinitialize the position and velocity of all particles, and set renumber = renumber+1.

(4) Judge whether renumber is greater than 3, if so, set $c_2$ to 0 and return to CLPSO. Otherwise, terminate the operation and select the best result from the saved optimization results to save.

3.2. Vehicle network security situation prediction method based on improved CLPSO

Literature [3] and literature [9] both proposed methods to optimize RBF neural networks based on improved PSO. Among them, in the training process of the RBF network, the literature [3] uses the method of encoding the neural network parameters first and then using the PSO algorithm to optimize the network parameters, but the literature does not give the detailed encoding and decoding steps. The method adopted in literature [9] is to first train RBF networks with the same number of particle swarms, compare the fitness of these networks, and then use PSO to find the optimal RBF network. However, this method only uses the mean square error to evaluate fitness, which lacks a theoretical basis. We improve on the above-mentioned literature methods and use the improved CLPSO algorithm in Section 3.1 to optimize the RBF neural network and predict the security situation of the vehicle network. The specific algorithm flow is as follows:

(1) Normalize the training samples of the vehicle network security situation prediction so that the input sample data is between 0 and 1. Equation (9) is the calculation formula. In the formula, $x_{\text{max}}$ and $x_{\text{min}}$ are the maximum and minimum values of the sample data.

\[
 x = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} 
\]  

(2) Initialize the particle swarm, the number of particles is n, the initial algebra $t=1$, and the initial velocity and position of each particle are randomly generated. The position of the particle swarm is the cluster radius $t$ during the clustering of the RBF network, and the pbest of each particle is also the initial value of the cluster radius.

(3) Train RBF neural network parameters and structure separately for each particle, that is, train $n$ RBF networks. The specific method is: first use the nearest neighbor clustering algorithm to cluster the input data according to different $\lambda$ values to obtain the number of clusters of the network and the hidden layer center $c_i$. Then use the least squares method to train the RBF network to obtain the prediction output value.

(4) Calculate the current fitness value $f(\lambda)$ of each particle, the historical best position fitness value $f(p\text{best})$, and the group historical best position $f(g\text{best})$. The fitness calculation formula is as formula (10), where $k$ represents the number of clusters. If $f(\lambda) < f(p\text{best})$, then $f(p\text{best}) = f(\lambda)$, if $f(\lambda) < f(g\text{best})$, then $g\text{best} = p\text{best}$.

\[
f(\lambda) = \frac{1}{4000} \sum_{i=1}^{t} y_i^2 - \prod_{i=1}^{t} \cos \left( \frac{y_i}{\sqrt{k}} \right) + 1
\]

(5) Use formula (7) and formula (4) in the improved CLPSO algorithm to update the particle velocity and position.

(6) If the optimization result is obtained, return the current particle swarm group extreme value pbest and the value of $\lambda$, and calculate the optimal RBF neural network parameters.
(7) Use the optimal parameters to construct an RBF network, analyze the test samples, and output the results.

4. Simulation

4.1. Experimental environment and data
We collect vehicle-related data and construct the vehicle network security situation data set by carrying out real attacks on vehicles. Attacks on vehicles include DOS attacks, replay attacks, tampering with ECU data, and disrupting vehicle communications [8]. Vehicle security situation assessment attributes include attack propagation range, attack recurrence difficulty, attack discoverability, and attack can cause losses. The method in the literature [1] is used to calculate the vehicle network security situation value, and these situation values are composed of vehicle network security situation data set.

4.2. Analysis of results
We will adopt a method similar to that in the literature [3], and predict the situation value of the next time period through the situation data of the first 5 time periods. In terms of data selection, samples are selected from the security situation data set at 10-minute intervals. A total of 70 data samples are selected, 50 of which are used for training the network, and the remaining 20 samples are used for prediction testing. At the same time, the improved CLPSO-RBF method is compared with the RBF prediction algorithm and BP prediction algorithm[10]. The comparison of the training results and prediction results of the three algorithms is shown in Figure 2 and Figure 3.

![Fig.2 Comparison of training results](image-url)
Through comparison, it can be concluded that the method we proposed is superior to the traditional RBF algorithm and BP algorithm in terms of the accuracy of training data and prediction data. The number of iterations of the three algorithms is shown in Table 1. It can be seen that the improved algorithm we proposed is also better than the other two methods in terms of convergence speed.

| Algorithm                    | Number of iterations |
|------------------------------|----------------------|
| Improved CLPSO-RBF algorithm | 36                   |
| RBF algorithm                | 64                   |
| BP algorithm                 | 107                  |

Tab.1 Algorithm iteration number table

5. Conclusion
Applying cybersecurity situational awareness technology to the security of the vehicle network can better deal with the complex and diverse security threats of the vehicle network. We propose a prediction algorithm based on improved CLPSO-RBF. In this method, we improve the existing improved CLPSO algorithm and combine the RBF neural network to predict the value of the vehicle network security situation. According to the optimized algorithm, experiments are carried out using real vehicle safety data sets. The experimental results show that, compared with the RBF prediction method and BP prediction method, the method proposed in this paper is better in terms of prediction accuracy and convergence speed. However, the improved algorithm has higher complexity and longer running time. The next step is to combine CLPSO with other artificial intelligence optimization algorithms to ensure the accuracy of predictions and improve operational efficiency.

Acknowledgments
This work was supported by Natural Science Foundation of Sichuan Province under grant No.2019YFG0201, No.2018GZDZX0011, and Chengdu Science and Technology Project No.2018-YF05-00707-SN.
References

[1] Ruxiang L, Fei L and Jiayan Z 2020 Vehicle network security situation assessment method based on attack tree *IOP Conference Series Earth and Environmental Conference* 428 012021

[2] Weixi S 2019 PSO and SVM for network security situation prediction *J. Computer Applications and Software* 36(06) pp308-316

[3] Yang J, Chenghai L, Xiaohui W and Zhipeng L 2018 Research on network security situation prediction based on RBF optimized by improved PSO *J. Measurement & Control Technology* 37(05) pp56-60

[4] Xin J, Longwei L, Guohua S and Xiaolei L 2017 Prediction about network security situation of electric power telecommunication based on spark frame and PSO algorithm *J.Computer Science* 44(S1) pp366-371

[5] Fangwei L, Jia L, Jiang Z and Haibo Z 2015 A method of network security situation prediction based on hybrid kernels PSO-SVR *J. Microelectronics & Computer* 32(12) pp110-115

[6] Shanxue C, Zheng Y, Jiang Z and Fangwei L 2015 Network security situation prediction method based on PSO-SVM *J.Application Research of Computers* 32(06) pp1778-1781

[7] Qichang D, Min Z and Daxing W 2009 A novel radial basis function neural network method based on improved particles swarm optimization *J. Computer Simulation* 26(12) pp126-129

[8] Jian G, Xiao dong Z, Qi S, Xiaoyan H and Jie X 2017 Survey of Network Security Situation Awareness *J.Journal of Software* 28(04) pp1010-1026

[9] Liang C 2018 Security threats and research status of vehicle network *J.China Information Security* 6 pp103-105

[10] Okutan A , Yang S J and Mcconky K 2017 Predicting cyber attacks with bayesian networks using unconventional signals *Conference on Cyber & Information Security Research*