Research on Information Security Risk Evaluation of Cloud Storage System Based on BP-GA

Yanfeng Kong [1,2], LianXing jia [1], Jiang Zhang [1]

[1] National Defense Information Institute, Hubei Wuhan , China

[2] 92815 army branch of weapon , Xiang Shan, China

[a]13407160738@163.com, [b]jialianxing@163.com, [c]liondangcing@qq.com

Keywords：cloud storage, information security risk evaluation, BP neural network, genetic algorithm

Abstracts: At present, it is difficult to make a good security evaluation of cloud storage system because of qualitative, quantitative, linear and nonlinear factors, so in this paper, the artificial neural network technology and genetic algorithm(BP-GA) is introduced to evaluate the information risk degree of cloud storage. The mathematical evaluation model of cloud storage risk degree is set up, with illustration of its application process by samples. The results show that the BP-GA method put forward in this paper not only has enough engineering accuracy, but also has the advantage of good convenience and applicability.

Introduction

Cloud storage is a new conception after cloud computing, and is a new kind of network storage technology. On the basis of cloud application, network technology and distributed documentary function, cloud storage can provide data storage and service access function by cooperating numerous different kinds of storage devices to work together through application software[1]. But in recent years, many individuals and companies are unwilling to put their crucial or private data in the cloud server, for information security events are frequently occurred. So it is urgent to make information security risk evaluation of cloud storage system. However, as information security involves many factors, and there are many difficulties such as qualitative, quantitative, linear and nonlinear situations, which make it very complicated to have security evaluation of cloud storage.

Soft computing is a computing method developed currently, which is corresponding with mankind brain, and is of exceptional performance to learn and deduce in an uncertain and accurate environment[2]. Soft computing includes many computing methods, such as fuzzy algorithm, neuron network algorithm, fish-swarm algorithm and genetic algorithm, etc. among them, BP neural network combined with genetic algorithm(BP-GA) has the good effect to make an evaluation of the system contained qualitative, quantitative, linear and nonlinear factors[3].

Basic conception of BP-GA

BP-GA is a new optimization algorithm raised recently, it has the characteristics of high efficiency, strong global searching capability, etc[4]. GA is independent of gradient information, and has no requirement of the object function's continuity, even its definite expression. BP-GA takes the advantage of artificial neural network and genetic algorithm, not only overcomes the low efficiency and long convergence time of GA in structure optimization, but also improves global solving capability of GA, so it is an effective and applicable solution. Its working process is shown in Fig.1.

Construction of BP neural network risk evaluation

Multi-layer feed forward neural networks divides the network into several layers, which are arrayed in order, the neuron in layer i only receives signals from neuron in layer (i-1), and the neurons in all layers have no feedback. For a forward network, when vector x is inputted, vector y is

outputted through the network, so the forward neural network can be seen as a transformer of x to y mapping[5-6].

Three layers BP neural network is employed in this paper. The input vector in the first layer can be adjusted according to practice. The first layer is a normalized layer, where three nodes are taken. The input vectors are software threat, hardware threat and management vulnerability. Experts are organized to form an evaluation group to make an evaluation of each factor. The values in $(0,1)$ are supposed as the risk evaluation results, i.e. the better the results are, the lower the risk of information subject's success is. The second layer is the input layer of BP network, five nodes are taken. The third layer is the output layer, one node is taken, the output characteristic function is s-style function, the output value is information construction risk degree, which are the continual numbers in $(0,1)$.

$\omega 1_{ij}(i=1,2,3; j=1,2,3)$ is the weight of input layer, $o1_i(i=1,2,3)$ is the node of input layer, $\omega 2_j(j=1,2,3)$ is the weight of middle layer, $o2_i(i=1,2,3)$ is the node of middle layer, $\theta 2_j(j=1,2,3)$ is the weight of the second middle layer, θ_3 is the weight of output layer, ω_0 is risk degree. The mathematical expression of the node are:

$$o2_j = \frac{2}{1+e^{-u2_j}} - 1 , \quad u2_j = \sum_{i=1}^{3} \omega 1_{ij} o1_i - \theta 2_j \quad \text{（j=1，2，3）} \tag{1}$$

$$\omega = \frac{1}{1+e^{-u3}} , \quad u3 = \sum_{j=1}^{3} \omega 2_j o2_j - \theta 3 \quad \text{（j=1，2，3）} \tag{2}$$
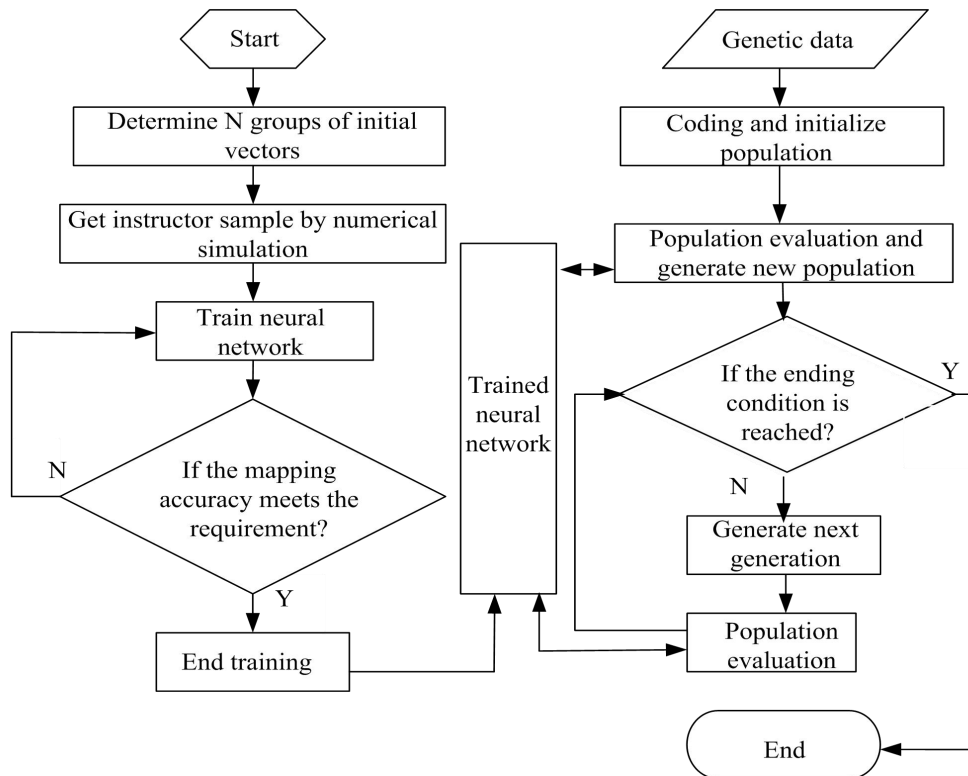

Fig.1 Flow chart of BP-GA

Design of genetic algorithm

Genetic algorithm contains five basic parts: coding, initial population, fitness function, genetic manipulation and parameter control and end rules[7]. Coding is the bridge of problem and algorithm. In order to carry out genetic searching in a big space and improve algorithm accuracy, float-encoding is employed. Each individual of the initial population is composed of 40-bit binary strings, the value is determined as follows: generate a random number in $(0,1)$, if it is bigger than 0.5, the value of this

bit is 1, or it is 0. Now the most used method is based on the penalty function idea. As the problem talked about in this paper is the least error optimization problem, then the fitness function can be expressed as:

$$Min \quad f = \frac{1}{\sum\limits_{i=1}^{m} e_i^2}, e_i = \omega_i - \overline{\omega_i}(i=1,2,\cdots,m) \tag{3}$$

Genetic manipulation contains four parts, they are selection, crossover, mutation and population update.

Selection: individual is selected by roulette wheel method.

Crossover: in an even crossover way, the mask sample is generated at random.

Mutation: mutation bit is selected randomly, then get the bit value reversed.

Population update: using pareto solution reserve strategy, the parent pareto solution is taken directly into child generation.

This paper employs two end rules, for one rule, when the difference of the maximum and minimum of the object function is smaller than given precision $1e-4$, the algorithm is considered convergent, and the program is terminated. For the other rule, suppose the largest genetic generation value is 100, when the iteration time of algorithm reaches 100, the program is terminated.

Samples of risk evaluation of cloud storage based on BP-GA

In this sample, ten training samples are taken, i.e. ten different information construction subject are studied(seen in Tab.1), with the risk degree values determined by experts from experience. The GA program is made up in MATLB 7.0 simulation environment, the initial population number is supposed 100, the crossover probability is taken 0.51,and the mutation probability is taken 0.032.

Table 1   BP neural network trained samples

subject	Software threat	Hardware threat	Management ulnerability	Expected risk
subject 1	0.3	0.2	0.9	0.11
subject 2	0.9	0.8	0.9	0.72
subject 3	0.4	0.2	0.7	0.11
subject 4	0.5	0.9	0.9	0.61
subject 5	0.4	0.3	0.7	0.16
subject 6	0.9	0.7	0.8	0.59
subject 7	0.2	0.5	0.6	0.17
subject 8	0.9	0.6	0.8	0.51
subject 9	0.7	0.3	0.5	0.18
subject 10	0.4	0.8	0.9	0.48

Having the sample values in Tab.1 self-trained in BP neural network model, then get the final weights and thresholds in Tab.2.

Table 2   Weights and thresholds in BP neural network model

j	$\omega 1_{1j}$	$\omega 1_{2j}$	$\omega 1_{3j}$	$\omega 2_j$	$\theta 2_j$	$\theta 3_0$
1	-0.21093	-0.09377	-0.050774	-0.15648	-0.44185	-2.1478
2	0.86752	0.90625	-0.29687	0.32587	-0.85469	-2.1478
3	-0.62541	0.21875	-0.63945	0.25846	0.698742	-2.1478

To test the model's validity, four test samples are selected(seen in Tab.3). Take the self-studied weights and thresholds in Tab.2 into BP neural network model, and make the computation of the risk degree of the test samples, then get the output risk degree(seen in Tab.3).

Table 3   Comparison of numerical simulation results and neural network mapping results

No.	Software threat	Hardware threat	Management vulnerability	Information evaluation risk value		Relative error
				By numerical simulation	Prediction by BP	
1	0.7	0.4	0.6	0.2598	0.26348	1.4%
2	0.3	0.6	0.9	0.3118	0.30646	-1.7%
3	0.1	0.4	0.7	0.1085	0.10653	-1.8%
4	0.8	0.3	0.3	0.1950	0.19227	-1.4%

Easy to see, the output risk degree values of the model are basically coincident with the expected risk degree values, and the errors are all less than 0.05, thereby the validity of the model has been tested.

Conclusions

On the basis of normal information risk evaluation models, the nonlinear global mapping of each risk factor and the final risk value is determined with BP neural network algorithm, and the optimization of weights and thresholds of BP neural network is made through GA. Introducing BP-GA in the security risk evaluation domain for cloud storage system is a new attemp, which brings a satisfactory result.

References

[1] An Shijie, Ouyang Guangyao. Analysis of the design feature of common rail fuel injection system for marine diesel engines[J]. Diesel Engine, 2002, 22(6): 13-16.

[2] Zen Wei, Gu Dongliang, Song Guoming, etc. New driving method of electronic control injector for high ressure common rail system [J]. Vehicle Engine, 2010, 22(8): 32.

[3] Wang Guihua, Lu Chen, Zhang Xichao, etc. Expriment study of high speed electromagnet feature in diesel electronic fuel injection system[J]. Chinese Internal Combusion Engine Engineering, 2003, 24(4): 13.

[4] Ren Liang. Software design of high pressure common rail diesel electric-control injection system[D]. Beijing: Tsinghua University, 2004.

[5] Zhang Qi, Zhang Kexun, Li Jianqiu, etc. Optimized design of electric control diesel solenoid valve driving circuit[J]. Chinese Internal Combusion Engine Engineering, 2005, 41(2):2-3.

[6] Chen Hailong. Theory and test study of high pressure common rail system in rising-voltage mode[D]. Wuhan: Naval University of Engineering, 2012. .

[7] Gao Wei, Shan Qinlin, Zhu Ying. Design of Boost transformer based on UC3842 control chip[J]. Journal of Daqing Petroleum Institute, 2008, 32(3): 71-73. .

[8] Ding Yu, Zeng Wei, Shen Yan, etc. New design of boost and closed-loop controlled on electric control unit for common rail system [J]. Modern Vehicle Power, 2012, 34(1): 46-50.