Research on Network Security Situation Awareness System Based on Machine Learning

YANG Ye-ling

Collage of Electronic Information, Chongqing Institute Of Engineering, Chongqing, China

Yelingyang0087@163.com

Keywords: Network Security, Security Situational Awareness, Machine Learning

Abstract: With the popularity of computers, the Internet has entered the production and all aspects of social life, but the attendant problem of network security has become the focus of widespread concern. Network security situation awareness to effectively respond to network security issues provide a viable solution: for complex network environments and malicious attacks, a comprehensive analysis of attacks against various parts of the network system, from a macro point of view of network security situation be assess and predict the future of network security situation based on this information. For the predictive accuracy of prediction system for network security situation has improved significantly, and network security situation prediction method based on machine learning for the network security situation prediction have a high degree feasible, in the real network security situation awareness applications have certain research and practical value.

Introduction of Network Security Situational Awareness

Network security situation is the current status and trends of the entire information from a variety of factors operating conditions of various network devices, network behavior and user behavior constituted. Network security situational awareness can acquire, understand and display the network environment of security elements. Through a series of technical means in time and space, are fully aware of network security and access and associated elements as possible in a pluralistic, and the establishment of a network based on complex behaviors modeling and simulation situational analysis and pre-side system, and then integrate and analyze vast amounts of security associated with the network-related data.

Network security situation awareness is a scientific and effective network security situation assessment and use of relevant technologies to make reasonable predictions about trends over time network security, network management personnel in advance to remind the network system for network equipment, network peer node hosts and data resources to make reasonable adjustments, upgrades and backup, network environment to address the risk of possible future harm to the network system, losses may result down to an acceptable range [1].

Extract information network security situation is carried out on the basis of situational awareness that only a comprehensive collection of data and the use of sophisticated index system, to ensure the correctness of the results of the assessment. Therefore, in the design process model or system must pay attention to select a metric system. Network security situation is a source of diverse, different collection methods, different devices collect data formats, network security situation letter from mainly contains configuration, operating status, traffic, user behavior and other information.

Machine Learning

Machine learning object from given training data set of known samples corresponding to the input output mapping to try to solve the input and output between the training samples, the training to obtain this mapping relationship to the entire system for the next the output of the unknown input, an output corresponding to a prediction as accurate as possible. Wherein the system is studied, when the input X will be able to output Y, machine learning is the mapping function to be solved

according to the object system input and output, the purpose is to learn to get the machine input X is output Y. Machine learning can be simply expressed as: Suppose the sample space has N numbers independent and identically distributed sample spaces, where x represents the input, Y represents the corresponding output, and x to certain dependencies between Y, that is, x and Y follow a unknown joint distribution rate [2]. Fig.1 shows the trusted mobile platform structural system.

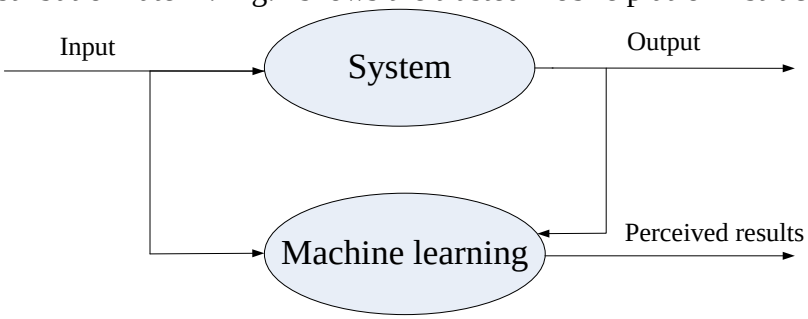

Fig. 1.The trusted mobile platform structural system

Machine learning is based on the purpose of the training sample set in a set of prediction functions set by adjusting the form and function parameters, solving an optimal mapping as a dependency between the input and output of Y is estimated to be an effective unknown input or predict, so that the expected risk function R gamma) is taken to the theoretical minimum. VC dimension is to study the concept of uniform convergence machine learning speed and generalization, a core concept of the statistical theory defined set of functions related to learning and an important indicator of performance. If the sample for any number of in the function set has a function which can break apart, then VC dimensional function set is infinite.

VC bounded real function of the dimension can be defined by certain thresholds to convert it to the indicator function. VC dimension directly reflects the learning function set, it indicates that the greater the VC dimension of learning more complex machine, that the greater the capacity, the greater the confidence interval, which led to the deviation between expected risk and risk the possibility of the greater experience which the main reason is an excessive learning. Thus, machine learning process, not only to make the empirical risk minimization, but also as far as possible to narrow the range of the confidence interval, so that both channels and the minimum value when, in order to obtain a smaller expected risk, ensure the promotion of learning machine [3].

Design of Network Security Situation Awareness System

Support vector machine is a machine learning statistical learning theory, the latest theoretical achievements for outstanding practicality, currently used as statistical learning theory in the field of research focus, and in constantly evolving. Linear support vector machine is classified based and developed. Linear classifier basic theory in two-dimensional space represents. Fig.2 shows theory of linear classifier.

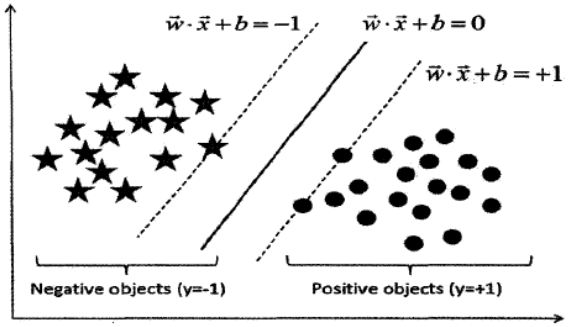

Fig. 2.The theory of linear classifier

Five-pointed star and the dots represent the two kinds of points, the hyperplane symmetry exists in the classification hyper plane around both sides, respectively, through two types of sample points from the hyperplane nearest sample points three straight line parallel to the middle, the

distance between them is called a class interval, two sample points is a straight line through the SVM (support vector). Solving linear classifier is to the n-dimensional data space for training set to find a hyperplane to the sample point be classified separately, and the classification of the maximum distance in n-dimensional space. Inherited proposed machine learning method based on binary tree of many small sub-classifiers training methods, the advantages of faster decision-making, and to overcome each sub-classifier training needs of the entire sample, the disadvantage caused by the slow training, while drawing on the advantages of fewer training samples and advantages of one algorithm directed acyclic graph method of decision-making speed etc.

In this paper, simulated annealing algorithm to optimize the parameters of SVM used to obtain the approximate optimal solution. Simulated annealing algorithm is introduced Monte Carlo iterative strategy for solving stochastic optimization algorithm. Since the physics of solids and the annealing process optimization process has many similarities. Physical principles of simulated annealing algorithm corresponding to: solid after being heated from a higher initial temperature of the starting temperature at a temperature falling to ambient temperature process, the binding characteristics of the probability of the sudden jump in the solution space in the form of random find the global objective function optimal solution, namely local optimal solution can be in the form of probabilistic jumping out of local, and find the most optimal solution to a final approximate approaches to the global optimal solution [4].

Achievement of Network Security Situation Awareness System

In order to improve accuracy and generalization ability of machine learning, select the appropriate kernel function is very important. Seen from the introduction, the most commonly used kernel functions include linear kernel, polynomial kernel and Gaussian kernel of three. Using their default parameters set using the training data set of nuclear tests respectively linear kernel, polynomial kernel function and Gaussian kernel. Using the optimal parameters of simulated annealing algorithm, the optimal parameters will be used to establish the appropriate network security situation prediction model, and extract data normalized concentration in the first seven weeks as a training data set, function prediction model training.

In order to test the perceptual model based on machine learning network security situation prediction accuracy and generalization capability obtained from the normalized data after a process of centralized data extracted after two weeks, the security situation to get a time series that contains 336 hours, and sliding window of size 7 time series reconstructed. The test dataset input support vector machine prediction model, to predict the beginning of the test data, and get the security situation time series forecasting [5]. Fig.3 shows network security situation forecast result compare.

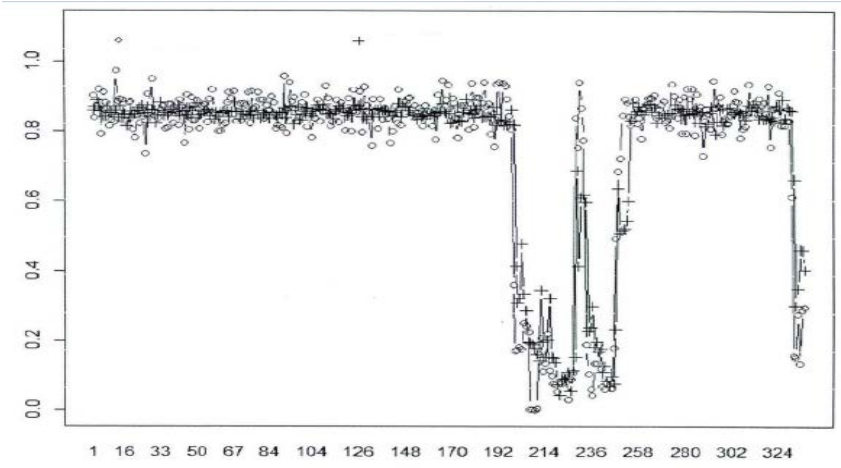

Fig. 3.The network security situation forecast result compare

The horizontal axis represents the time in hours, and the ordinate represents the normalized subsequent to the [0, 1] interval of trend value, hollow red dot indicates the actual security situation for each time interval, while the blue dots represent hollow corresponding to the time interval by the

value of the security situation prediction model results. As can be seen from the figure, the actual security situation and predicted trends broadly in line, when the value of the security situation mutation, predictive values and the real deviation is larger. Simulated annealing algorithm is a state transition factor contains a position distribution so that it can be seen as a continuous random parameter values within the range, set the proper cooling rule, fast convergence, to find global optimal solution. The grid size is based on cross-validation state transition step, easy to skip the global optimal solution, and the optimization process required for each point on the grid within the range of parameters to calculate the objective function value many times, when the training data set is too large, or fine-grid search optimization too long.

Conclusions

With the growing size of the network, the network structures are becoming increasingly complex, which gives the network viruses and other security risks in the intangibles to opportunity, and threat of loss of its network system consisting of growing. In this context, the proposed network security situation and build a framework for modeling based on machine learning, constitute the network security situation awareness system, an important part of network security assessment, emergency response network, network security early warning, through a series of analysis, it shows that the system can support network security situational awareness of evolving. Network security situation awareness can analyze the overall state of the network from a macro perspective, according to the contact between the various security events, integrated network security status given to the evaluation system, to make effective and accurate predictions.

References

[1] Yih-Lon Lin, Jer-Guang Hsieh, Hsu.-Kun Wu.et. Three-parameter sequential minimal optimization for support vector machines [J].Neuro computing. 2011, 72: 3467-3475.

[2] Stephane, Alexandre Lucas. Visual Intelligence for Crisis Management[D]. Florida: Florida Institute of Technology.2013.

[3] Haoliang Zhang, Jinqiao Shi, Xiaojun Chen. A Multi-Level Analysis Framework in Network Security Situation Awareness [J].Procedia Computer Science, 2013, 17: 530-536.

[4] Tiago P. F. Lima; Teresa B, Ludermir. An automatic method for construction of ensembles to time series prediction [J]. International Journal of Hybrid Intelligent Systems.2013, 10 (4):191-203.

[5] Song-song Lu, Xiao-feng Wang, Li Mao. Network security situation awareness based on network simulation [C]. Electronics, Computer and Applications, 2014 IEEE Workshop Ji-Nan, China. 2014.