计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.201200239

基于遗传优化 PNN 神经网络的网络安全态势评估

王金恒[1]　单志龙[2]　谭汉松[3]　王煜林[1]
1 广州理工学院计算机科学与工程学院　广州 510540
2 华南师范大学计算机学院　广州 510631
3 中南大学计算机学院　长沙 410006
(11403404@qq.com)

摘　要　为提高网络安全态势评估的准确率,提出了一种基于遗传优化概率神经网络的网络安全态势评估。首先,在网络安全态势评估建模过程中,根据网络安全态势特点和常见评估等级建立了概率神经网络的网络安全态势评估模型,以便充分挖掘概率神经网络在网络安全态势评估细粒度方面的优势。然后,为了防止因网络安全态势参数细粒度评估而造成收敛速度过慢的情况发生,对概率神经网络的修正因子进行遗传算法优化,并再次进行概率神经网络训练,从而得到稳定的概率网络安全态势评估模型。经过实验证明,相比传统的概率神经网络算法,基于遗传算法优化概率神经网络的网络安全态势评估准确度更高,平均准确率达到 90%以上,且训练速度更快。

关键词:概率神经网络;网络安全态势;网络攻击;遗传算法
中图法分类号　TP393.08

Network Security Situation Assessment Based on Genetic Optimized PNN Neural Network

WANG Jin-heng[1],SHAN Zhi-long[2],TAN Han-song[3] and WANG Yu-lin[1]
1 School of Computer Science & Engineering,Guangzhou Institute of Science and Technology,Guangzhou 510540,China
2 School of Computer Science,South China Normal University,Guangzhou 510631,China
3 School of Computer Science,Central South University,Changsha 410006,China

Abstract　In order to improve the performance of network security situation assessment,this paper presents a network security situation assessment method based on genetic optimization probabilistic neural network. Firstly,In the process of network security situation assessment modeling,according to the characteristics of network security situation and common evaluation levels,the network security situation assessment model of PNN neural network is established,and the advantages of PNN neural network in fine-grained network security situation assessment are fully exploited. Then,in order to prevent the slow convergence caused by the fine-grained evaluation of network security situation parameters,the correction factors of PNN are left,and then the stable PNN network security situation assessment model is obtained by iterative training of PNN neural network. Experiments show that compared with the traditional PNN neural network algorithm,by using genetic algorithm to optimize the PNN network security situation assessment classification,evaluation accuracy is higher,average accuracy rate is more than 90%,and training speed is faster.

Keywords　Probabilistic neural network,Network security situation,Network attack,Genetic algorithm

互联网应用及服务日渐成熟,用户数量快速增长,随着网络服务量及访问用户数的增加,网络结构越来越复杂,而用户对网络服务的质量需求却在提升,这些都影响着互联网的发展与更新,特别是互联网安全管理及控制,关乎着互联网的发展。互联网安全问题一直是用户最关注的问题之一[1]。

对接入网络中的所有主机进行有效的网络安全态势评估,能够有效发现网络中存在的不稳定因素[2],从而为网络安全决策提供强有力的数据支持。网络安全态势评估既可以从网络全局考虑,评估整个网络的安全程度,又可以对接入网络中的单个主机进行安全评估,以检测网络中安全性较差的主

到稿日期:2020-12-31　返修日期:2021-03-06　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金项目(61671213);广东省质量工程建设项目(2020SZL02);广东省教学改革建设项目(2018SJG04)
This work was supported by the National Natural Science Foundation of China(61671213),Quality Engineering Construction Project of Guangdong Province,China(2020SZL02) and Teaching Reform and Construction Project of Guangdong Province, China(2018SJG04).
通信作者:王煜林(43498000@qq.com)

机，对重点主机采取严密监控，不断应对接入网络主机的安全考验。通过网络态势评估值确定网络安全等级，采取不同程度的安全控制策略和方法，这样既能保证整个网络的防御性，又能够为访问网络的所有用户提供持续型的服务。

近年来，关于网络安全态势评估的研究较多，Xie 等[3]采用 BP 神经网络进行网络安全态势评估，为提高 BP 神经网络在评估过程中的准确率，其结合了布谷鸟算法进行优化，并取得了较高的安全态势评估准确率，但是存在评估稳定性不高的问题。Liu 等[4]采用马尔可夫链对云计算环境的网络安全态势进行评估，解决了云计算复杂网络环境（大规模网络）的网络安全态势评估，但是马尔可夫链算法的局部优化效果不好，因此其最佳评估准确率不理想。Ye 等[5]采用深度学习算法进行网络安全态势评估，该方法的评估准确率较高，但牺牲了算法执行效率，导致评估时间较长。

综上，有的算法获得了较高的网络安全态势评估准确率，但出现了评估效率不高的问题；部分算法可以用于大规模网络的安全态势评估，但评估准确率不高。因此，要获得更稳定的网络安全态势评估准确率，还需要不断优化评估算法。本文将 PNN(Probabilistic Neural Network)算法应用于网络安全态势评估，并结合遗传算法优化 PNN，以便获得较好的网络安全评估性能，提高评估效率。

1 网络安全态势评估

网络安全态势分析需要从海量网络数据中将网络威胁和攻击按照网络安全标准进行感知，从网络流量中提取网络攻击类型，并且根据 CIC-IDS2017 标准对不同的攻击类型进行态势评估[6]。

表 1 按照从低到高的顺序列出了网络攻击类别对网络安全的评估分值。根据表 1 对网络中所有主机受到的攻击类型及攻击个数分别计算态势值。设网络中包含 N 台主机，攻击类别为 $i(1\leqslant i\leqslant 8)$，每种攻击对应的分值为 s_i，设第 j 台主机受到的第 i 类攻击的个数为 m_i，那么第 j 台主机的态势值为：

$$M_j=\frac{1}{m_{ij}}\sum_{i=1}^{8}s_i\cdot m_{ij} \tag{1}$$

表 1 网络攻击分级

Table 1 Network attack classification

标准	攻击类别	分值
CIC-IDS2017	Scan	0.1
	暴力破解	0.2
	Dos	0.3
	bot	0.4
	Web 攻击	0.5
	DDos	0.6
	安全漏洞	0.7
	bleed 攻击	0.8

分别计算 N 台主机的态势值，然后加权求和归一化得到网络的安全态势值：

$$M=\sum_{j=1}^{N}q_jE_j \tag{2}$$

其中，q_j 表示第 j 台主机在网络中的安全权重 $\sum_{j=1}^{N}q_j=1$。

式(2)求解出的态势值可用来评估网络安全等级，如表 2 所列[7]。

表 2 基于网络安全态势值的安全等级分类

Table 2 Security classification based on network security situation value

网络安全态势值	网络安全等级
0	安全
(0,0.2]	轻度
(0.2,0.5]	一般
(0.5,0.8]	较重
(0.8,1.0]	严重

根据网络安全等级可以采取相应策略来维护网络安全。但在实际情况中，网络安全态势的计算受到多方面的影响，导致获取的态势值精确度不高。一方面是因为网络中主机的网络态势感知算法不够灵敏，造成攻击的漏检和误检；另一方面是网络结构复杂，特别是在异构网络中，无法准确设定各主机占整个网络安全态势的权重。基于这些影响，网络安全态势需要智能算法根据网络流量样本来评估其安全等级。

2 概率神经网络(PNN)

2.1 PNN 算法

PNN 是神经网络的一种，通过引入概率估计来解决多类别分类问题，其主要结构如图 1 所示[8]。

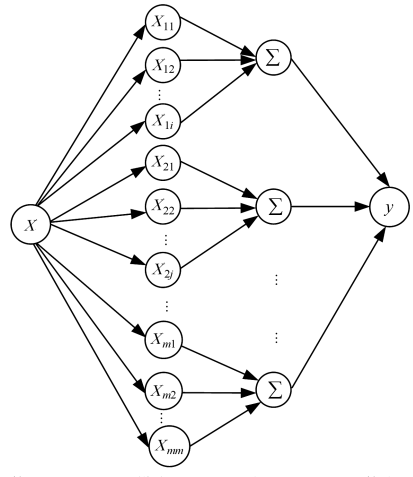


图 1 PNN 的结构

Fig. 1 PNN structure

在图 1 中，模式层的模式种类一般与输入样本的属性个数相同，PNN 模式层的求解方法[9]为：

$$\phi_{ij}(X)=\frac{1}{(2\pi)^{\frac{1}{2}}\sigma^d}\exp\left[-\frac{(X-X_{ij})(X-X_{ij})^{\mathrm{T}}}{\sigma^2}\right] \tag{3}$$

其中，X 为输入样本，σ 为修正因子，d 为样本属性个数。

根据式(3)进行求和并取均值，则第 i 种模式的求解方法[10]为：

$$g_i(x)=\frac{1}{L}\sum_{j=1}^{L}\phi_{ij}(x) \tag{4}$$

其中，L 为 i 模式下的样本数。

对所有模式求和，并根据求和结果进行判别，具体计算方式为：

$$y=\arg\max(g_i) \tag{5}$$

设测试个数为 N，y_i 为 PNN 输出值，$\hat{y}_i$ 为实际值，那么：

$$E(y)=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 \tag{6}$$

不断优化 $E(y)$，调整优化因子等参数，获得所需的稳定网络模型。

2.2 遗传优化

为了提高 PNN 在网络安全态势评估方面的效率及性能，利用遗传算法来优化 PNN 修正因子。

首先根据式(6)求解适应度函数 f：

$$f(y)=\frac{1}{1+E(y)} \tag{7}$$

在 GA 种群构建时，个体 i 被选择进化的概率 P_i 为[11]：

$$P_i=f_i/\sum_{i=1}^{N}f_i, i=1,2,\cdots,N \tag{8}$$

其中，f_i 为个体适应度值。

设 t 时刻两个体 x_A^t 和 x_B^t 进行交叉处理后得到 x_A^{t+1} 和 x_B^{t+1}[12]：

$$\begin{cases}x_A^{t+1}=\alpha x_B^t+(1-\alpha)x_A^t\\x_B^{t+1}=\alpha x_A^t+(1-\alpha)x_B^t\end{cases} \tag{9}$$

个体 x_K 变异得到：

$$x_k'=W_{\min}^k+\lambda(W_{\max}^k-W_{\min}^k) \tag{10}$$

其中，λ 为[0,1]随机数。

设交叉和变异概率分别为 P_c 和 P_m，限制范围为[$P_{c\min}$，$P_{c\max}$]和[$P_{m\min}$，$P_{m\max}$]，其中 $P_{c\min}=0$，$P_{c\max}=0.9$，$P_{m\min}=0.01$，$P_{m\max}=0.1$。设全部个体适应度均值为 f_{avg}，交叉与变异的适应度分别为 f' 和 f，则有[13]：

$$P_c=\begin{cases}P_{c\max}-\dfrac{(P_{c\max}-P_{c\min})(f_{\max}-f')}{f_{\max}-f_{avg}}, & f'\geqslant f_{avg}\\P_{c\max}, & f'<f_{avg}\end{cases} \tag{11}$$

$$P_m=\begin{cases}P_{m\max}-\dfrac{(P_{m\max}-P_{m\min})(f_{\max}-f)}{f_{\max}-f_{avg}}, & f\geqslant f_{avg}\\P_{m\max}, & f<f_{avg}\end{cases} \tag{12}$$

不断进化迭代，直到网络安全态势评估精度达到要求或者达到最大迭代次数时算法停止，从而获得经过优化的 PNN 最佳权重和阈值。

2.3 评估流程

获得稳定的 PNN 网络安全态势评估模型的核心是确定合适的 PNN 修正因子，得到修正因子最优解就能够确定 PNN 网络安全态势评估模型。在 GA 优化过程中，根据式(7)求解适应度，不断调整修正因子，以满足网络安全态势评估准确率的要求。结合常用网络安全态势评估流程及 GA-CPP 优化方法，本文提出的网络安全态势评估的主要流程如图 2 所示。

在网络安全态势评估过程中，以安全态势评估准确率阈值和最大迭代次数为算法停止条件，满足任一条件即停止运行，得到 GA-PNN 网络安全态势评估模型，然后输入待评估的网络数据样本，最后输出网络安全态势评估值。

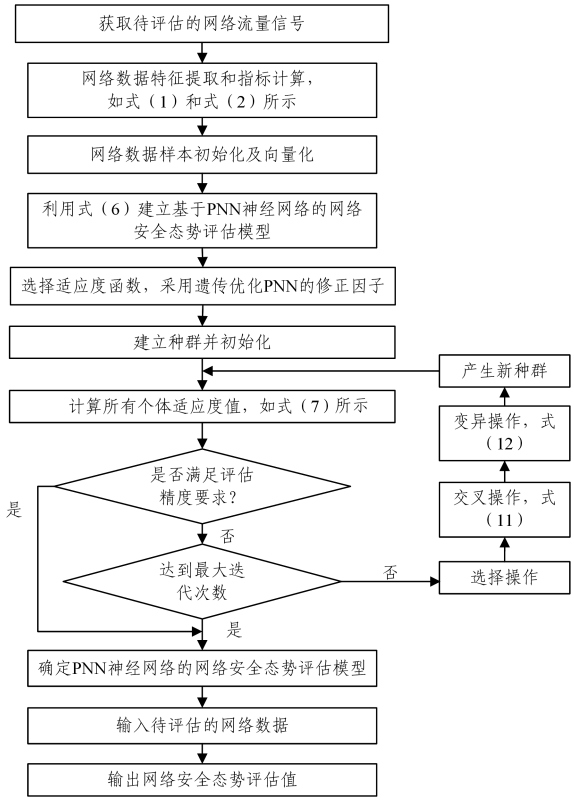


图 2 基于遗传优化的 PNN 网络安全态势评估流程

Fig. 2 PNN network security situation assessment process based on genetic optimization

3 实例仿真

为了验证遗传优化的 PNN 算法在网络安全态势评估中的性能，首先对某网络安全公司 12 个月的网络监测数据进行仿真，分别采用 PNN 算法和 GA-PNN 算法对数据样本的网络安全态势评估的准确率和收敛性进行对比；其次选择 KD CUP99 数据集，差异化选择样本容量，比较不同样本容量的网络态势评估准确率和标准差；最后使用常用的网络安全态势评估算法和本文算法分别对 KD CUP99 数据集进行仿真，对比不同算法在网络安全态势评估中的性能。

将网络安全公司 12 个月的样本按照时间顺序分为 6 组，构成 6 组网络安全数据集，其中每组数据集按照 3∶1 分为训练样本和测试样本；仿真选取的 KD CUP99 网络样本数量为 494010，其中正常样本和攻击样本的数量分别为 97 276 和 396 734，训练样本和测试样本数按照 3∶1 分配。

3.1 遗传优化的性能对比

3.1.1 PNN 和 GA-PNN 安全态势评估的准确率

为了分析 GA 引入后对 PNN 网络安全态势评估的性能影响，分别对 6 组数据集的训练样本进行传统 PNN 和 GA-PNN 网络安全态势评估计算，同时根据 6 组数据集的网络攻击类型及个数，对照表 1 的分级值计算实际态势值，将实际态势值与网络训练得到的态势值进行对比获得稳定的训练模型，之后输入 6 组测试集，具体训练结果如表 3 所列。

表 3 PNN 和 GA-PNN 的安全态势评估准确率

Table 3 Accuracy rate of PNN and GA-PNN security situation assessment

数据集编号	PNN		GA-PNN	
	平均准确率	标准差	平均准确率	标准差
1	0.8195	0.4711	0.9117	0.2627
2	0.8625	0.5323	0.9367	0.3811
3	0.7963	0.7646	0.9001	0.4325
4	0.9041	0.3827	0.9846	0.1463
5	0.8333	0.5224	0.9249	0.2152
6	0.8247	0.4619	0.9212	0.2719

从表 3 可以看出，相比 PNN 算法，GA-PNN 算法的网络安全态势评估准确率更高，GA-PNN 对 6 组数据集的安全态势评估准确率均达到 0.9 以上，特别是对于数据集 4，平均准确率达到了 0.9846，在所有数据集中态势评估性能最优。在标准差方面，GA-PNN 的性能也优于 PNN，对所有数据集的评估标准差均小于 0.5，表明 GA-PNN 的稳定性更好。

下面将以天为单位，将 GA-PNN 计算的网络安全态势值与实际态势值进行对比，本文选取了第二季度该公司的网络安全态势值可视化结果，如图 3 所示。

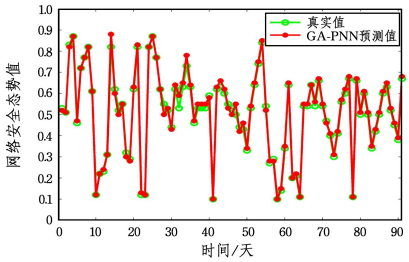


图 3 第二季度网络安全态势评估值的对比

Fig. 3 Comparison of network security situation assessment values in the second quarter

由图 3 可知，GA-PNN 的安全态势评估值与真实值的拟合较好，91 天中部分天数有较小的评估偏差，大部分天数的预测性能较好。当网络安全态势值超过 0.5 时，就需要开启网络安全的控制预警，超过 0.8 时就需要开启网络安全管控措施。图 3 中安全态势值超过 0.5 的天数较多，表明第二季度该公司监测的网络受到的网络攻击较多。

3.1.2 遗传优化的收敛分析

从 6 组数据集中各抽取 100 个样本进行 PNN 和 GA-PNN 网络安全态势评估模型的训练，稳定之后，从 6 组样本中各抽取 30 个测试样本输入到模型中，记录两种算法收敛时的标准差。安全态势评估标准差随迭代次数的变化如图 4 所示。

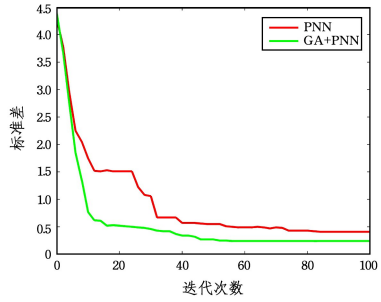


图 4 PNN 和 GA-PNN 收敛性比较

Fig. 4 Comparison of convergence between PNN and GA-PNN

从图 4 可以看出，随着迭代次数的增加，PNN 和 GA-PNN 算法的误差均不断减少，最后标准差不再随着迭代次数的增加而变化，从而得到稳定的网络安全态势评估模型。通过对比可知，PNN 算法迭代 82 次左右趋于稳定；而 GA-PNN 在迭代 56 次便达到了稳定。在收敛性方面，GA-PNN 的优势明显，并且提高了安全态势评估的效率。从图 4 也可发现，PNN 在训练过程中有两个时间段的稳定假象，分别是[20，24]，[31，38]，虽然在这两个时间段标准差未发生变化，但是从后续迭代中发现，PNN 算法远未收敛，这表明 PNN 的网络安全态势训练过程效率不高。而 GA-PNN 训练时每次迭代标准差均发生变化，GA 对 PNN 的网络安全态势评估收敛性能优化明显，这主要是因为 GA 算法对修正因子的优化起到了关键作用，使得 PNN 训练模型能够快速获得适应度值高的修正因子。

3.2 不同样本容量的故障识别准确率

为了验证 GA 优化的 PNN 安全态势评估对不同类型样本的适应度，本文采用常用的网络安全分析样本 KD CUP99，通过差异化设置训练样本和测试样本数量，计算网络安全态势评估的准确率。为了对比准确率，需要对 KD CUP99 样本的攻击类型及数量进行预计算，并结合表 1 各类攻击评分值计算所有样本的安全态势评估值。根据 KD CUP99 的样本总量，实验分别选取不同数量的训练样本和测试样本进行 GA-PNN 训练。

从表 4 可得，当参与 GA-PNN 训练的样本数量变化时，网络安全态势评估的准确率缓慢上升，评估的标准差也逐渐减小。这是因为 KD CUP99 的样本量较大，训练样本量过小则不足以反映系统的样本特点及攻击类型和数量的全貌。通过对比发现，当训练样本量达到 10000 时，安全态势评估的平均准确率有明显的提升，而训练样本从 30000 变化至50000，平均准确率的提升不明显。但总体来说，合理选择参与 GA-PNN 训练的样本数量，对于 KD CUP99 数据集能够获得较好的安全态势评估准确率。

表 4 不同样本数量的安全态势评估准确率

Table 4 Accuracy rate of security situation assessment with different sample sizes

训练样本数量	测试样本数量	平均准确率	标准差
1000	300	0.8947	2.5216
3000	900	0.9062	1.7153
5000	1500	0.9145	1.3409
10000	3000	0.9427	1.0726
30000	9000	0.9541	0.8395
50000	15000	0.9628	0.6527

3.3 不同算法的网络安全态势评估性能

为了进一步验证 GA-PNN 算法在网络安全态势评估中的性能，采用层次分析法[14]、马尔可夫链(HMM)[15]、支持向量机(SVM)[16-21]和 GA-PNN 算法分别对 KD CUP99 数据集的 10000 个样本进行安全态势评估性能仿真，仿真结果如图 5 所示。从图 5 可以看出，在网络安全态势评估准确率方面，GA-PNN 最优，HMM 算法次之，层次分析法最差；从评估收敛性来看，SVM 和层次分析法最优，本文算法和 HMM 算法次之，但 4 种算法的收敛速度差距较小。综合而言，本文算

法的安全态势评估性能更好。

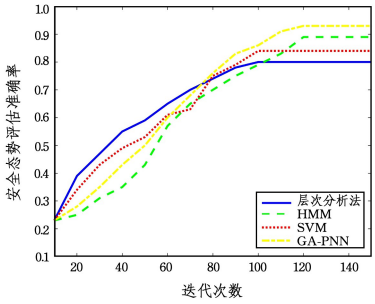


图 5 不同算法的网络安全态势评估性能

Fig. 5 Network security situation assessment performance of different algorithms

结束语 本文采用 PNN 进行网络安全态势评估，并采用 GA 优化 PNN 的修正因子。实验证明，GA 优化提高了 PNN 算法在网络安全态势评估上的准确率和稳定性，相比常用的网络安全态势评估算法，本文算法的评估准确率优势明显。但在训练样本过小时，GA-PNN 算法存在评估时间较长的问题。下一步将进一步优化 PNN 算法，提高其安全态势评估效率，降低评估时间，以满足大规模网络实时进行网络态势评估的要求，不断提高 PNN 算法在网络安全态势评估方面的适应度。

参 考 文 献

[1] LIU J, SU P R. Review of software and network security [J]. Acta Sinica Sinica, 2018, 29(1): 42-68.

[2] SONG J, TANG G L. Research and application of network security situation awareness technology [J]. Communication Technology, 2018(6): 187-192.

[3] XIE L X, WANG Z H. Network security situation assessment method based on cuckoo search Optimized BP neural network [J]. Computer Applications, 2017, 37(7): 1926-1930.

[4] LIU H, LIU J H, HUI X Y. Network security situation assessment based on cloud model and Markov chain [J]. Computer and Digital Engineering, 2019(6): 155-159.

[5] YE L, TAN Z J. A network security situation assessment method based on deep learning [J]. Intelligent Computer and Application, 2019, 9(6): 73-75.

[6] HAN X L, LIU Y, ZHANG Z J, et al. Overview of network security situation awareness theory and technology and Research on difficult problems [J]. Information Security and Communication Security, 2019(7): 61-71.

[7] OLIMID R F. SecRet: How to Apply the 5E Model for a Master's Level Network Security Course[J]. IEEE Communications Magazine, 2019, 57(11): 54-59.

[8] FENG S, XIONG Z, NIYATO D, et al. Joint Traffic Routing and Virtualized Security Function Activation in Wireless Multihop Networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(99): 9205-9219.

[9] LU Z, QU G, LIU Z. A Survey on Recent Advances in Vehicular Network Security, Trust, and Privacy[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(2): 760-776.

[10] MAZURCZYK W, BISSON P, JOVER R P, et al. Challenges and Novel Solutions for 5G Network Security, Privacy and Trust [J]. IEEE Wireless Communications, 2020, 27(4): 6-7.

[11] SEVINÇ E, COŞAR A. An Evolutionary Genetic Algorithm for Optimization of Distributed Database Queries[J]. 2018, 54(5): 717-725.

[12] KANTOUR N, BOUROUBI S. Cryptanalysis of Merkle-Hellman cipher using parallel genetic algorithm [J]. Mobile Networks and Applications, 2019(8): 1-12.

[13] JIA W, ZHAO D, ZHENG Y, et al. A novel optimized GA-Elman neural network algorithm[J]. Neural Computing and Applications, 2019, 31(6): 1-11.

[14] QU X H, SHI X M. Research on network security situation assessment technology based on analytic hierarchy process [J]. Automation Technology and Application, 2018, 37(11): 43-45.

[15] WU J T, QIAO Y F, ZHU S F, et al. Network security situation assessment and prediction method based on HMM [J]. Navigation and Control, 2018(2): 13-20, 34.

[16] HUANG Z H. Heuristic network security situation prediction based on simplex vector machine [J]. Journal of Hunan Institute of Engineering (Natural Science Edition), 2020, 30(1): 53-56.

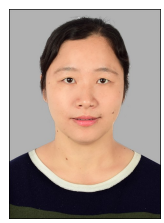
[17] CAO J M. Research on Network Security Framework of Hyper-heuristic SVM for Big Data [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2020, 32(1): 23-29.

[18] ZARCA A M, BERNABE J B, SKARMETA A, et al. Virtual IoT HoneyNets to mitigate cyberattacks in SDN/NFV-enabled IoT networks[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(99): 1262-1277.

[19] UBARHANDE S D, DOYE D D, NALWADE P S. A time stamp-based algorithm to improve security and performance of mobile ad hoc network[J]. Wireless Networks, 2019, 25(4): 1867-1874.

[20] ZHU N, YAO S H, ZHENG X L. Analysis and Model Establishment for Mobile Communication Network Risk Evaluation Index System[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2020, 37(3): 100-106.

[21] LI X. A Real-time Prediction Method of Network Security Risk Based on Predictive Model [J]. Journal of Chongqing University of Technology (Natural Science), 2019, 33(2): 132-137.



WANG Jin-heng, born in 1982, master, lecturer. Her main research interests include computer network technology, artificial intelligence and cloud computing.

WANG Yu-lin, born in 1982, master, associate professor. His main research interests include network security and cloud computing.