

蝙蝠算法优化神经网络的网络入侵检测

刘 羿

(河南农业大学信管学院,河南 郑州 450002)

摘要:研究网络入侵安全问题,网络入侵具有隐蔽性、随机性和突发性等复杂变化特点,传统方法无法描述变化规律,导致入侵检测正确率低。为了提高网络入侵检测效果,针对BP神经网络参数优化问题,提出一种蝙蝠算法优化BP神经网络的神经网络入侵检测模型(BA-BPNN)。首先将BP神经网络参数编码为蝙蝠个体,并以网络入侵检测正确率作为个体适应度函数,然后通过模拟蝙蝠飞行过程找到BP神经网络最优参数,最后根据最优参数建立网络入侵检测模型。在Matlab 2012平台采用KDD CUP 99数据集仿真测试,结果表明,BA-BPNN解决了传统神经网络模型存在的难题,提高网络入侵检测正确率。

关键词:网络入侵;蝙蝠算法;神经网络;参数优化

中图分类号:TP391 **文献标识码:**B

Network Intrusion Detection Model Based on Neural Network Optimized by Bat Algorithm

LIU Yi

(Henan Agricultural University, College of Information Management Science, Zhengzhou Henan 450002, China)

ABSTRACT: In order to improve the detection effect of the network intrusion, this paper proposes an intrusion detection model (BA-BPNN) based on BP neural network in which parameters are optimized by bat algorithm. Firstly, parameters of BP neural network are coded as bats individual, and network intrusion detection rate is taken as the fitness function, and then the optimal parameters of the BP neural network are found by simulating the bat flying, finally, network intrusion detection model is established according to the optimal parameters. The simulation test is carried out on the Matlab 2012 platform by using KDD CUP 99 data sets, and the results show that compared with other network intrusion detection model, BA-BPNN improves the learning ability and generalization ability of BP neural network, and can obtain satisfactory detection effect for network intrusion.

KEYWORDS: Network intrusion; Bat algorithm; Neural network; Parameters optimization

1 引言

随着Internet规模增大,网络入侵事件日益增多,一旦设网络受到恶意攻击,网络安全将遭受极大的破坏,导致用户私人数据泄露、用户权限被恶意篡改、网络瘫痪等,然而入侵检测系统作为安全防御的最后一道防线,能够检测各种形式的入侵行为,因此网络入侵检测一直是网络安全研究中的一个热点^[1]。

网络入侵检测首先收集网络状态数据,然后对网络行为进行分析,最后将网络行为分为异常和正常两种,并根据检

测结果采取相应安全措施^[2]。网络状态受到多种因素的综合影响,具有随机性和时变性等变化特点,因此网络入侵检测实质是一个非线性分类问题,传统线性分类算法难以建立最优分类器,而机器学习算法具有非线性、智能学习能力,因此出现了基于神经网络、支持向量机的网络入侵检测分类器^[3-6]。尤其是前向传播(back propagation, BP)神经网络是一种非线性分类能力优异的神经网络,具有简单、易实现、自学习、自组织和适应能力强等优点,成为应用最为广泛的网络入侵检测方法^[7,8]。当采用BP神经网络建立网络入侵模型时,如果BP神经网络参数如果选择不当,就难以建立准确、全面反映网络入侵动态、随机性变化趋势,导致网络入侵正确低,网络入侵误报、漏报率相当的高。为了解决该难题,有学者利用提出采用群智能学习算法对BP神经网络参数进

基金项目:国家自然科学基金项目(61202285)

收稿日期:2014-05-04

行优化,如利用粒子群算法、遗传算法、蚁群算法对 BP 神经网络参数进行优化,以获到较优的 BP 神经网络参数,建立入侵检测正确率高的检测模型,并取得了不错的应用效果,为保证网络安全起着重要的作用。然而这些智能算法存在各自的不足,易出获得局部最优解以及收敛速度慢等难题,从而对网络入侵检测结果产生不利影响^[9,10]。蝙蝠算法(Bat Algorithm, BA)是一种模拟蝙蝠回声定位的新型群智能算法,相比较于粒子群算法、遗传算法等传统智能算法,蝙蝠算法具有发挥更大作用的潜能,可以实现动态控制局部搜索和全局搜索间的相互转换过程,避免陷入局部最优缺陷,具有更好的收敛性,为神经网络参数优化提供了一种新的研究方法^[11-13]。

针对 BP 神经网络在网络入侵检测中存在的难题,为了提高网络入侵检测率,将蝙蝠算法引入到 BP 神经网络参数优化中,提出一种蝙蝠算法优化 BP 神经网络的网络检测模型(BA-BPNN),并通过仿真对 BA-BPNN 性能进行测试。

2 网络入侵的检测原理

网络入侵是对计算机网络进行有计划性、隐蔽性的访问,其会降低网络系统的安全性和稳定性,而网络入侵检测是指依据计算机网络中的核心信息收集网络目标信息,再分析目标信息判断相应的网络系统是否存在不符合安全体系的行为以及网络被破坏的特征的一种计算机网络安全技术。图 1 为一个网络系统一周每小时的网络入侵数量变化曲线。

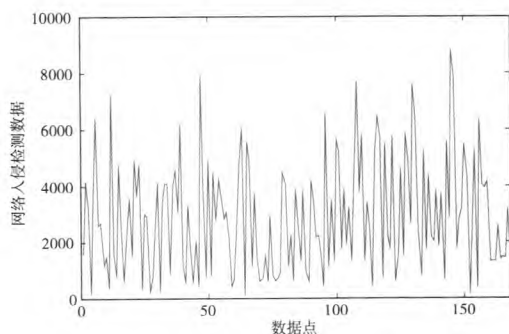


图 1 网络入侵数量的变化情况

对图 1 的网络入侵化特性的分析可知,网络入侵变化有如下特点:

- 1) 影响网络状态因素很多,影响程度又随入侵行为而异,导致网络入侵变化的随机性、突变性。
- 2) 网络入侵数据历史数据相当多,数据之间的关联性为非线性关系,传统建模算法难以实时反映网络状态的变化趋势。
- 3) 网络入侵行为同时也具有周期性,而且入侵行为检测实时要求相当高,建立的入侵检测模型必须满足在线性检测要求。

从网络入侵行为变化特点以及要求可以知道,网络入侵

不仅具有隐蔽性,还具有多样性、随机性,是一种典型非线性变化系统,而且检测速度相当快,传统方法难以准确跟踪网络入侵非线性变化特征,从而导致检测正确率,而这些问题恰好是当前网络入侵建模过程中面临的困难。

对于网络入侵建模过程来说,选择良好的检测模型是提高网络入侵检测正确的关键,而且模型参数优化是保证高正确率网络入侵检测结果的核,本文采用 BP 神经网络建立网络入侵检测模型,并结合蝙蝠优化算法全面挖掘网络入侵的变化规律,以更好准确、全面刻画网络入侵变化趋势,从而获得更高网络入侵检测正确率。网络入侵检测原理如图 2 所示。

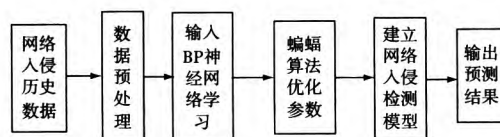


图 2 BA-BPNN 的网络入侵检测原理

3 BA-BPNN 的网络入侵检测模型

3.1 BP 神经网络

一个三层的前向网络能够以任意精度逼近任何一个非线性函数,因此 BP 神经网络只需输入层、隐含层和输出层。其网络参数(权值和阈值)的调整公式为

$$w_{kj}(t+1) = w_{kj} + \alpha \delta_k H_j \quad (1)$$

$$w_{ji}(t+1) = w_{ji} + \alpha \delta_j I_i \quad (2)$$

$$\theta_k(t+1) = \theta_k(t) + \beta \delta_k \quad (3)$$

$$\theta_j(t+1) = \theta_j(t) + \beta \sigma_j \quad (4)$$

式中, $w_{kj}(t+1)$ 和 $w_{kj}(t)$ 分别为前后两次训练时隐含节点 j 与输出层节点 k 的连接权值; $w_{ji}(t+1)$ 和 $w_{ji}(t)$ 分别为前后两次训练时隐含节点 i 与输出层节点 j 的连接权值; H_j 为隐含层节点的输出; I_i 为从输入节点 i 输入的信号; α, β 分别为学习参数; θ_k 和 θ_j 分别为输出节点和隐含节点 j 处的阈值; δ_k 和 σ_j 分别为输出层节点 k 和隐含层节点 j 的误差信号。

在 BP 神经网络开始训练前,需要选择最合适的参数,蝙蝠算法(BA)是一种新型群智能算法,其通过模拟自然界中蝙蝠通过超声波搜索、捕食猎物的生物学特性找到问题最优解,具有模型简单、潜在并行性和分布式等特点,为此,本文采用 BA 算法优化 BP 神经网络参数以提高网络入侵检测的正确率。

3.2 蝙蝠算法

BA 是模拟自然界中蝙蝠利用一种声呐来探测猎物、避免障碍物的随机搜索算法,其工作原理为:将种群数量为 m 的蝙蝠个体映射为 D 维问题空间中的 m 个可行解,将优化过程和搜索模拟成种群蝙蝠个体移动过程和搜寻猎物,利用求解问题的适应度函数值来衡量蝙蝠所处位置的优劣,将个体的优胜劣汰过程类比为优化和搜索过程中用好的可行解替代较差可行解的迭代过程。BA 的实施过程

1) 种群初始化,即蝙蝠以随机方式在 D 维空间中扩散分布一组初始解。具体包括:初始种群个体数(蝙蝠数目) m ,最大脉冲音量 A_0 ,最大脉冲率 R_0 ,搜索脉冲频率范围 $[f_{\min}, f_{\max}]$,音量的衰减系数 α ,搜索频率的增强系数 γ ,最大迭代次数 $iter_max$ 。

2) 随机初始化蝙蝠的位置 x_i ,并根据适应度值的优劣寻找当前最优解 x^* 。

3) 蝙蝠的搜索脉冲频率、速度和位置更新。种群在进化过程中,每一代个体的搜索脉冲频率、速度和位置按如下公式进行变化

$$f_i = f_{\min} + (f_{\max} - f_{\min}) \times \beta \quad (5)$$

$$v_i^t = v_i^{t-1} + (x_i^t - x^*) \times f_i \quad (6)$$

$$x_i^t = x_i^{t-1} + v_i^t, \quad (7)$$

式中, $\beta \in [0, 1]$ 是均匀分布的随机数; f_i 是蝙蝠 i 的搜索脉冲频率 $f_i \in [f_{\min}, f_{\max}]$; v_i^t, v_i^{t-1} 分别表示蝙蝠 i 在 t 和 $t-1$ 时刻的速度; x_i^t, x_i^{t-1} 分别表示蝙蝠 i 在 t 和 $t-1$ 时刻的位置; x^* 表示当前所有蝙蝠的最优解。

4) 生成均匀分布随机数 $rand$, 如果 $rand > r_i$, 则对当前最优解进行随机扰动, 产生一个新的解, 并对新的解进行越界处理。

5) 生成均匀分布随机数 $rand$, 如果 $rand < A_i$ 且 $f(x_i) < f(x^*)$, 则接受步骤 4) 产生的新解, 然后按如下公式对 r 和 A_i 进行更新

$$A_i^{t+1} = \alpha A_i^t \quad (8)$$

$$r_i^{t+1} = R_0 [1 - \exp(-\gamma t)] \quad (9)$$

6) 对所有蝙蝠的适应度值进行排序, 找出当前的最优解和最优值。

7) 重复步骤 2) ~ 6), 直至满足设定的最优解条件为止。

8) 输出全局最优值和最优解。

为了验证 BA 的性能, 选用 2 个常用基准函数对 BA、遗传算法(GA)和粒子群算法(PSO)性能进行对比, 测试函数如下

$$f_1 = 100 \times (x_1^2 - x_2)^2 + (1 - x_1)^2 \quad (10)$$

$$f_2 = \frac{1}{4000}(x_1^2 + x_2^2) - \cos x_1 \cos(\frac{x_2}{\sqrt{2}}) + 1 \quad (11)$$

各算法的收敛精度如图 3 和图 4 所示。从图 3 和图 4 可知, 对所有函数, BA 的收敛速度明显优于 GA, 而且避免了 GA、PSO 陷入局部最优的缺点, 这表明 BA 搜索能力、收敛精度和收敛速度均优于 GA、PSO。

3.3 BP-BPNN 入侵检测模型的工作流程

将 BA 中的蝙蝠个体的位置分量看作 BP 神经网络的权值和阈值, 每只蝙蝠唯一确定一个网络, 那么在网络训练过程中, 蝙蝠个体的前后两次位置的变化即为对应着的神经网络的权值和阈值的更新, 从则通过蝙蝠的位置更新可以搜索到网络的最优权值和阈值, 以达到网络训练的目的。

把 BP 神经网络的权值和阈值蝙蝠个体化后, 蝙蝠的适

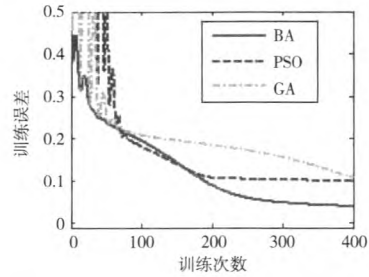


图 3 不同算法在函数 f_1 上的收敛结果

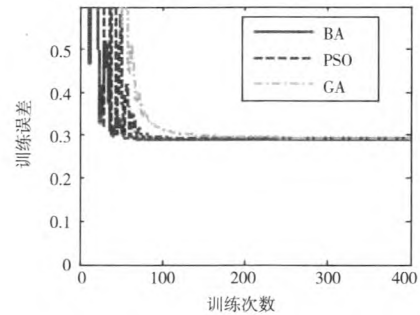


图 4 不同算法在函数 f_2 上的收敛结果

应度计算如下

$$f(x_i) = \frac{1}{n_i} \sum_{m=1}^{n_i} (O_{im} - T_{im})^2 \quad (12)$$

式中, n_i 为训练样本的个数; O_{im} 和 T_{im} 分别表示 m 个训练样本在第 i 只蝙蝠所确定的网络权值和阈值下的网络实际输出和期望输出。

BA-BPNN 网络入侵检测模型的工作步骤具体如下:

1) 收集网络入侵历史数据, 并对其进行预处理。

2) 初始化网络结构: 输入层、隐含层、输出层的节点个数; 导入训练样本和检验样本。

3) 初始化蝙蝠群体: 蝙蝠个体数 n ; 每只蝙蝠的音量 A 和脉冲频率 r 及其位置向量 x 和速度向量 v ; 蝙蝠的频率 f 范围; 蝙蝠的位置 x_i ; 迭代次数和误差精度。

4) 利用 BP 神经网络权值和阈值调整式(9)~(11)对每个蝙蝠的个体最佳位置和全局最佳位置进行更新。

5) 利用式(5)~(7)对蝙蝠的搜索脉冲频率、速度和位置进行更新。得到个体最佳位置和全局最佳位置。

6) 生成均匀分布随机数 $rand$, 如果 $rand < A_i$ 且 $f(x_i) < f(x^*)$, 则接受步骤 3) 产生的新解, 然后按式(8)~(9)对 r_i 和 A_i 进行更新。

7) 利用式(12)计算所有蝙蝠的适应度值 f 并对其进行排序, 得到全局最佳位置的适应度值 f_g 。若 f_g 达到网络的训练精度 ($f_g < \varepsilon$) 或当前迭代次达到最大迭代次数, 则迭代结束转步骤 7); 否则计算各蝙蝠的个体极值 q_i 和全局极值 q_g 的位置, 转步骤 3) 继续更新蝙蝠的速度和位置。

8) 输出蝙蝠全局最优位置对应的权值和阈值,即为 BP 神经网络的最优初始权值和阈值。

9) 将蝙蝠全局最优位置对应的权值和阈值作为 BP 神经网络参数,建立最优的网络入侵检测模型。

综合上述可知,复杂多变的网入侵检测模型的工作流程如图 5 所示。

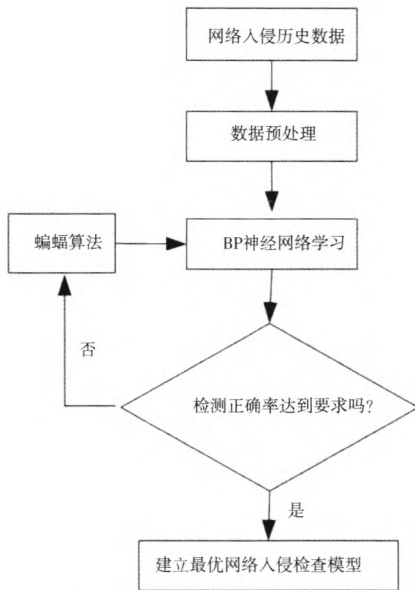


图 5 网络入侵检测模型的工作流程

4 仿真研究

4.1 仿真环境

数据来自网络入侵标准测试集 KDD CUP 99 数据集,其包括 4 种入侵类型:DoS、Probe、U2R 和 R2L,同时包括正常样本,每一个样本共有 41 个特征,7 个符号型字段和 34 个数值型字段。由于 KDD Cup 99 数据集样本多,从中随机选择部分数量的数据进行测试,数据具体分布见表 1。

表 1 样本集分布情况

入侵类型	含义	训练样本	测试样本
DoS	拒绝服务攻击	1000	500
Probe	监视和其它探测活动	2000	1000
R2L	远程机器的非法访问	200	50
U2R	普通用户对本地超级用户特权的非法访问	400	100

4.2 特征归一化处理

由于网络入侵特征值采用不同度量单位,在 BP 神经网络训练过程,如果某些特征占了主导地位,就会掩盖了其它一些特征对检测结果的贡献,为了消除该现象的出现,对特征值进行归一化处理。数据归一化的 Matlab2012 代码为:

$[MaxV, I] = \max(Data);$

$[MinV, I] = \min(Data);$

$[R, C] = \text{size}(Data);$

$Scaled = (Data - \text{ones}(R, 1) * \text{Min}(V) * (\text{ones}(R, 1) * ((Upper - Lower) * \text{ones}(1, C) / (\text{Max}(V) - \text{Min}(V)) + Lower;$

4.3 对比模型及评价标准

为了使 BA - BPNN 的结果具有可比性,采用 PSO 算法优化 BP 神经网络(PSO - BPNN),遗传算法优化 BP 神经网络(GA - BPNN)作为对比模型。模型性能的评价标准为:检测率、误报率和运行速度。其中,检测率和误报率分别定义如下

$$rate = \frac{\text{检测出的入侵样本数}}{\text{入侵样本总数}} \times 100 \quad (13)$$

$$error = \frac{\text{被误报为入侵的正常样本数}}{\text{正常样本总数}} \times 100 \quad (14)$$

4.4 结果与分析

1) 检测结果对比

PSO - BPNN、GA - BPNN 和 BA - BPNN 的检测率和误报率如图 6 和 7 所示。从图 6 和 7 可知,相对于对比模型,BA - BPNN 的检测结果最佳,这表明采用 BA 可以获得比 PSO 算法更优的 BP 神经网络参数,从而建立了更优的网络入侵检测模型,有效降低了网络入侵检测的误报率,提高了网络入侵的检测率得。

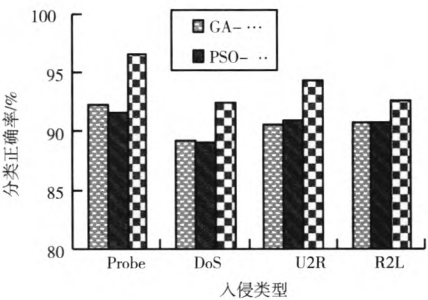


图 6 几种模型的检测率比较

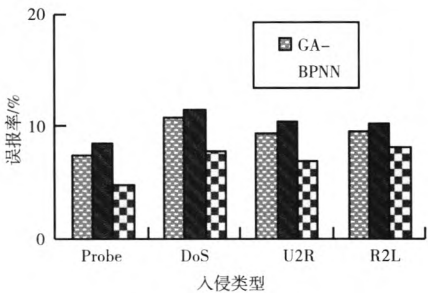


图 7 几种模型的误报率比较

2) 检测速度对比

为了检测模型的运行速度,采用模型对验证集的检测时间作为衡量指标,各模型的检测时间(秒,s)

(下转第 445 页)

出版社, 2001.

- [7] Woo Zhiwei, Chung Hungyuan, Lin Jinjye. A PID type fuzzy controller with self-tuning scaling factors[J]. Fuzzy Sets and Systems, 2000, 115(2): 321-326.
- [8] 王正林, 等. matlab/simulink 与控制系统仿真[M]. 北京: 电子工业出版社, 2012.
- [9] 关新民, 于斌, 李新勤. 电弧炉电极调节系统 Fuzzy-PD 控制器的仿真研究[J]. 计算机仿真, 2002, 19(5): 100-103.

(上接第 314 页)

见表 2。从表 2 可知, 相对于 PSO-BPNN、GA-BPNN、BA-BPNN 检测速度得到大幅度提高, 主要由于 BA-BPNN 采用了分层技术, 减少了计算时间, 同时对 PSO 算法进行改进, 加快了收敛速度, 因此, BA-BPNN 能够更加满足现代网络入侵检测系统的实时性、在线要求。

表 2 不同模型的检测时间(s)对比

入侵类型	BA-BPNN	GA-BPNN	PSO-BPNN
DOS	0.66	0.95	1.62
U2L	0.68	1.33	1.41
U2R	0.71	0.98	1.34
Probe	1.03	1.43	1.73

5 结束语

针对 BP 神经算法参数选择不当导致收敛速度慢、易陷入局部极值等难题, 提出一种蝙蝠算法优化 BP 神经网络的网络入侵检测模型。在神经网络参数优化过程中, 不仅考虑了 BP 神经网络的训练误差反传, 同时过跟踪个体的最佳权值和群体的最佳权值来更新个体和群体的权值, 这样, 既充分利用了蝙蝠算法的全局搜索能力, 又较好地利用了 BP 神经网络的误差反向传播特点, 提高了 BP 神经网络的整体性能。仿真结果表明, 相对于对比算法, BA-BPNN 不仅可以加快网络入侵检测速度, 提高了网络入侵的检测率, 同时误报率明显降低, 在网络入侵检测中有着广泛的应用前景。

参考文献:

- [1] 唐正军, 李建华. 入侵检测技术[M]. 北京: 清华大学出版社, 2004.
- [2] 颜谦和, 颜珍平. 遗传算法优化的神经网络入侵检测系统[J]. 计算机仿真, 2011, 28(4): 141-144.
- [3] 黄小龙, 梁碧珍. 基于粗糙集和 CP 神经网络的入侵检测模型[J]. 计算机仿真, 2011-10: 115-117.

[作者简介]



贾 华(1961-), 男(汉族), 内蒙古包头市人, 硕士, 副教授, 硕士研究生导师, 主要研究方向为电力电子技术应用及传动。

郭向超(1986-), 男(汉族), 河北石家庄市人, 硕士研究生, 主要研究方向为控制理论及应用。

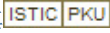
刘慧博(1972-), 女(汉族), 内蒙古包头市人, 博士, 副教授, 硕士研究生导师, 主要研究方向为导航制导与控制。

- [4] D E Denning. An Intrusion Detection Model[J]. IEEE Transaction on Software Engineering, 2010, 13(2): 222-232.
- [5] C L Hang, C J Wang. A GA-based feature selection and parameters optimization for support vector machines[J]. Expert Systems with Applications, August 2009, 31(2): 231-240.
- [6] 肖海军, 洪帆, 张昭理, 廖俊国. 基于融合分类和支持向量机的入侵检测研究[J]. 计算机仿真, 2008, 25(4): 130-132.
- [7] 陈仕涛, 等. 基于粒子群优化和邻域约简的入侵检测日志数据特征选择[J]. 计算机研究与发展, 2010, 47(7): 1261-1267
- [8] J Hong, et al. A novel intrusion detection system based on hierarchical clustering and support vector machines[J]. Expert Systems with Applications, 2011(38): 306-313.
- [9] L Khan, M Awad, B Thuraisingham. A new intrusion detection system using support vector machines and hierarchical clustering[J]. The VLDB Journal, 2007, (16): 507-521.
- [10] 杨富华, 彭刚. PCA-SVM 在网络入侵检测中的仿真研究[J]. 计算机仿真, 2011, 28(7): 146-149.
- [11] X S Wang. A new metaheuristic bat-inspired algorithm[C]. Nature Inspired Cooperative Strategies for Optimization, Studies in Computational Intelligence, Springer-Verlag, Berlin Heidelberg, 2010-10: 65-74.
- [12] X S Yang. Bat Algorithm for Multi-objective Optimization[J]. Int. J. Bio-Inspired Computation, 2011, 3(5): 267-274.
- [13] X S Yang, A H Gandomi. Bat Algorithm: A Novel Approach for Global Engineering Optimization[J]. Engineering Computation, 2012, 29(5): 267-289.

[作者简介]



刘 羿(1973-), 男(汉族), 河南唐河县人, 硕士, 讲师, 主要研究方向为: 数据挖掘、软件工程。

作者: [刘羿](#), [LIU Yi](#)
作者单位: [河南农业大学信管学院, 河南郑州, 450002](#)
刊名: [计算机仿真](#) 
英文刊名: [Computer Simulation](#)
年, 卷(期): 2015, 32(2)

参考文献(13条)

1. [唐正军;李建华](#) [入侵检测技术](#) 2004
2. [颜谦和;颜珍平](#) [遗传算法优化的神经网络入侵检测系统](#)[期刊论文]-[计算机仿真](#) 2011(04)
3. [黄小龙;梁碧珍](#) [基于粗糙集和CP神经网络的入侵检测模型](#)[期刊论文]-[计算机仿真](#) 2011(10)
4. [D E Denning](#) [An Intrusion Detection Model](#) 2010(02)
5. [C L Hang;C J Wang](#) [A GA-basod feature seleetion and parameters optimization for support vector machines](#) 2009(02)
6. [肖海军;洪帆;张昭理;廖俊国](#) [基于融合分类和支持向量机的入侵检测研究](#)[期刊论文]-[计算机仿真](#) 2008(04)
7. [陈仕涛](#) [基于粒子群优化和邻域约简的入侵检测日志数据特征选择](#)[期刊论文]-[计算机研究与发展](#) 2010(07)
8. [J Hong](#) [A novel intrusion detection system based on hierarchical clustering and support vector machines](#) 2011(38)
9. [L Khan;M Awad;B Thuraisingham](#) [A new intrusion detection system using support vector machines and hierarchical clustering](#) 2007(16)
10. [杨富华;彭刚](#) [PCA-SVM在网络入侵检测中的仿真研究](#)[期刊论文]-[计算机仿真](#) 2011(07)
11. [X S Wang](#) [A new metaheuristic bat-inspired algorithm](#) 2010
12. [X S Yang](#) [Bat Algorithm for Multi-objective Optimization](#) 2011(05)
13. [X S Yang;A H Gandomi](#) [Bat Algorithm:A Novel Approach for Global Engineering Optimization](#) 2012(05)

引用本文格式: [刘羿](#). [LIU Yi](#) [蝙蝠算法优化神经网络的网络入侵检测](#)[期刊论文]-[计算机仿真](#) 2015(2)