一种基于改进BPNN的网络安全态势预测方法

申请号：201610871327.5
申请日：2016-09-30

申请(专利权)人　重庆邮电大学
地址　400065 重庆市南岸区南山街道崇文路2号
发明(设计)人　朱江　明月　王森

主分类号　H04L29/06(2006.01)I
分类号　H04L29/06(2006.01)I
公开(公告)号　106453293A
公开(公告)日　2017-02-22
专利代理机构　北京一格知识产权代理事务所(普通合伙) 11316
代理人　滑春生

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106453293 A
(43)申请公布日 2017.02.22

(21)申请号 201610871327.5

(22)申请日 2016.09.30

(71)申请人 重庆邮电大学
地址 400065 重庆市南岸区南山街道崇文路2号

(72)发明人 朱江 明月 王森

(74)专利代理机构 北京一格知识产权代理事务所(普通合伙) 11316
代理人 滑春生

(51)Int.Cl.
H04L 29/06(2006.01)

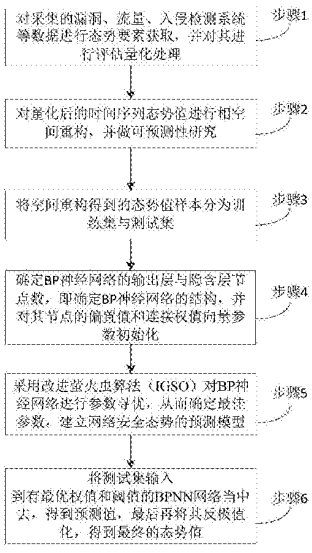
权利要求书4页 说明书9页 附图3页

(54)发明名称

一种基于改进BPNN的网络安全态势预测方法

(57)摘要

本发明涉及网络安全评估技术领域，特别涉及一种基于混沌理论与神经网络相结合的网络安全态势预测方法，包括：采用互信息法和cao氏法对归一化后的网络安全态势值序列集合进行处理得到网络安全态势样本值的最佳嵌入维数并进行相空间重构，分析重构后样本的最大李雅普诺夫指数来得到评估出来的样本是否具有混沌预测性；根据非线性时间序列的特点与经验确定反向传播神经网络的输出层与隐含层的节点数；利用改进的萤火虫算法进行参数寻优，从而确定网络权值和偏置值，建立网络安全态势的预测模型；测试集样本输入到BP神经网络中进行预测，并将得到的预测值反归一化；本发明能够较精确地对网络安全态势进行预测，同时能够提高网络安全态势预测收敛速度。



CN 106453293 A

1.一种基于改进反向传播神经网络BPNN的网络安全态势预测方法，其特征在于，包括以下步骤：

步骤1，对采集的漏洞、流量、入侵检测系统数据进行态势要素获取，并通过层次化网络安全态势评估量化方法对收集到的态势要素信息进行评估量化处理；

步骤2，运用极值化公式对量化后产生的非线性时间序列态势值进行预处理，再寻找最适合的嵌入维数与延迟时间进行相空间重构，并通过计算该非线性的时间序列的李雅普诺夫指数来确定是否有可预测性；

步骤3，将空间重构得到的态势值样本分为训练集与测试集；

步骤4，根据非线性时间序列的特点与经验BPNN的输出层与隐含层的节点数，设定输入层节点数为嵌入维数，从而确定神经网络的结构，并初始化BPNN的向量参数Θ；

步骤5，采用改进萤火虫算法IGSO对BPNN进行参数寻优，从而确定网络权值和偏置值，建立网络安全态势的预测模型；

步骤6，将测试集输入至有最优权值和阈值的BPNN中，得到预测值，最后再将其反极值化，得到最终的态势值。

2.根据权利要求1所述的基于改进BPNN的网络安全态势预测方法，其特征在于，所述步骤2进一步包括以下步骤：

步骤21，建模极值标准化公式：

$$x'(i)=\frac{x(i)-x(i)_{\min}}{x(i)_{\max}-x(i)_{\min}} \qquad i=1,2,\cdots,n$$

其中，x(i)与x'(i)分别为处理前后的网络安全态势值，$x(i)_{\min}$与$x(i)_{\max}$分别表示处理前所有网络安全态势值中的最小值与最大值，且通过处理后得到的网络安全态势数据x'(i)，i＝1,2,…n.是一组一维时间序列，其中n为一段时间内的网络安全态势样本数；

步骤22，采用最小互信息法计算最佳时间延时τ，并将τ和cao氏法相结合确定嵌入维数，从而得出BPNN的输入节点数m；

步骤23，根据cao氏法与互信息法得到的m与τ，引入最大Lyapunov指数来验证数据具有可预测性。

3.根据权利要求2所述的基于改进BPNN的网络安全态势预测方法，其特征在于，所述步骤22中的最佳时间延时τ的计算公式为：

$$I(\tau)=\sum_{i,j}P_{ab}(x'(t_i),x'(t_i+\tau))\log_2\left[\frac{P_{ab}(x'(t_i),x'(t_i+\tau))}{P_a(x'(t_i))P_b(x'(t_i+\tau))}\right]$$

其中，定义事件a表示网络安全态势样本序列x'(t_i)，事件b表示进行时间延迟的网络安全态势样本序列x'(t_i+τ)，p_a(x'(t_i))与p_b(x'(t_i+τ))分别表示a、b两事件中x'(t_i)与x'(t_i+τ)会发生的概率，P_{ab}(x'(t_i)，x'(t_i+τ))为x'(t_i)和x'(t_i+τ)两事件联合分布概率；如果最佳时间延时I(τ)等于0，则代表x'(t_i)与x'(t_i+τ)无相关，即x'(t_i+τ)是不可以预测的；若I(τ)取得极小值，表示x'(t_i)与x'(t_i+τ)具有最大可能的不相关，取I(τ)的第一个极小值为最佳时间延迟τ。

4.根据权利要求2所述的基于改进BPNN的网络安全态势预测方法，其特征在于，所述步骤22中将τ和cao氏法相结合确定嵌入维数，从而得出BPNN的输入节点数m包括：

$$a(i,m)=\frac{\left\|X_i(m+1)-X_{n(i,m)}(m+1)\right\|}{\left\|X_i(m)-X_{n(i,m)}(m)\right\|}$$

$$E(m)=\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}a(i,m)$$

$E_1(m)=E(m+1)/E(m)$

其中$X_i(m)$和$X_i(m+1)$分别表示嵌入维为m和m+1时重构相空间的第i个向量，$X_{n(i,m)}(m)$和$X_{n(i,m)}(m+1)$分别表示与$X_i(m)$和$X_i(m+1)$最近的向量，||·||为欧几里得距离，a(i,m)用于判断$X_{n(i,m)}(m)$是否为$X_i(m)$的真实临近点，若在m维相空间临近的两个点在m+1维相空间依然临近，则为“真实临近点”，否则为“虚假临近点”；E(m)和E(m+1)分别表示在m维和m+1维下非线性时间序列上点与其相邻点之间的平均统计距离，N表示态势值时间序列；如果网络安全态势的非线性时间序列当中包含确切的规律，那么能够找到一个合适m，当m大于某固定值m_0时，$E_1(m)$若停止较大变化，则可将m_0+1当作最小嵌入维数，其中判断是否停止较大变化包括：设置一个在0到1范围内波动的$E_2(m)$，来对比$E_1(m)$是否大幅增加还是已经停止较大变化，$E_2(m)$设置准则如下：

$E_2(m)=E^*(m+1)/E^*(m)$

$$E^*(m)=\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}\left|X(i+m\tau)-X_{n(i,m)}(i+m\tau)\right|$$

对于随机事件序列，数据内部无关联，因此是不可预测的，$E_2(m)$将始终为1，而对于确定性时间序列，相邻点之间的关系会随着嵌入维数m的值变化，因此总有一些m使得$E_2(m)$不等于1，因此，$E_2(m)$的波动程度能够用来度量时间序列中的确定性元素。

5.根据权利要求1所述的基于改进BPNN的网络安全态势预测方法，其特征在于，所述步骤2所述相空间重构方法为：

$$\begin{cases}\boldsymbol{X}_i(m)=\left\{x'(i),x'(i+\tau),\cdots,x'(i+(m-1)\tau)\right\},i=1,2,\cdots M\\ M=N-(m-1)\tau\end{cases}$$

其中x’(i)为极值化后的一维时间序列，M表示重构相点的数量，m为嵌入维数，即输入层节点数，τ为延迟时间。

6.根据权利要求1所述的基于改进BPNN的网络安全态势预测方法，其特征在于，所述步骤5进一步包括以下步骤：

步骤51，将萤火虫群的个体位置映射为BPNN的向量参数Θ，并指定种群内萤火虫个体的数目，对所有的个体进行随机实数编码，使得萤火虫种群均匀分布在D维的搜索空间里；

步骤52，初始化IGSO算法的参数，其中包括：最大迭代次数t_{max}、最小移动步长s_{min}、最大移动步长s_{max}、萤火素更新参数ρ、适应度函数参数γ、萤火素初始值l_0、萤火虫感知范围r_s；

步骤53，按照IGSO算法进行迭代寻优，得到萤火虫种群在搜索空间中的全局最优解，即得到BPNN对网络安全态势训练样本预测精度最高的一组向量参数Θ，并基于该组向量参数Θ来构建BP网络中各层之间的连接权值与各节点之间的阈值，进而得到网络安全态势值泛化能力最强的BPNN网络模型。

7.根据权利要求6所述的基于改进BPNN的网络安全态势预测方法，其特征在于，所述步骤53中IGSO算法进一步包括以下步骤：

步骤531，设定种群个体数目并在解空间中随机初始化个体位置，计算初始化种群每个个体的适应度函数值，同时生成公告板；

步骤532，对种群中的所有萤火虫个体按$l_i(t)=(1-\rho)l_i(t-1)+\gamma J(x_i(t))$更新萤火素值，其中，$l_i(t)$表示第t次迭代中第i个萤火虫所携带的萤火素，$\rho\in(0,1)$为萤火素更新参数，$\gamma$为适应度函数参数，$J(x_i(t))$为适应度函数，$x_i(t)$为萤火虫i在第t次迭代的位置；

步骤533，进入迭代阶段，求解种群中个体的邻居萤火虫的集合，如果邻居集合存在则转到步骤535，否则转到步骤536；

步骤534，根据轮盘赌的方法计算萤火虫i在其决策域内的移动方向，同时为了摆脱陷入局部最优，引入变步长来代替固定步长进行移动步长的更新，并设定变步长公式为：$s(t)=s_{max}e^{c\cdot t}$，$c=\frac{1}{t_{max}}\ln(\frac{s_{min}}{s_{max}})$，其中，$t_{max}$为最大迭代次数，$s_{min}$为最小移动步长，$s_{max}$为最大移动步长；

步骤535，根据534的步长$s(t)$进行位置更新，则萤火虫在t+1次迭代的位置$x_i(t+1)$的更新公式为：

$$x_i(t+1)=x_i(t)+s\left(\frac{x_j(t)-x_i(t)}{\left\|x_j(t)-x_i(t)\right\|}\right)$$

其中$x_i(t)$表示萤火虫i在第t次迭代的位置，$x_j(t)$表示萤火虫i在第t次迭代时决策域内的第j只萤火虫的位置，同时更新萤火虫个体的决策域，设定第i只萤火虫在t+1次迭代时刻的动态决策范围$r_d^i(t+1)$为：

$$r_d^i(t+1)=\min\left\{r_s,\max\left\{0,r_d^i(t)+\beta(n_t-|N_i(t)|)\right\}\right\}$$

其中$N_i(t)=\left\{j:\left\|x_j(t)-x_i(t)\right\|<r_d^i(t);l_i(t)<l_j(t)\right\}$表示第i只萤火虫在第t次迭代时，它的决策域内所包含的萤火虫的集合，$l_i(t)$表示第t次迭代中第i个萤火虫所携带的萤火素，$l_j(t)$表示第t次迭代中第j个萤火虫所携带的萤火素，其中，$j\in N_i(t)$，$||x||$表示x的范数；r_s为萤火虫感知范围，$r_d^i(t)$为第i只萤火虫t次迭代时刻的动态决策范围，β为比例常数，n_t为邻居阈值；

步骤536，计算当前种群所有个体对应的适应度函数值，取其中最好的适应度函数值与公告板中的值进行对比，若优于公告板信息，则选择更新公告板；

步骤537，根据条件判断，如果发生变异即当迭代次数大于2且公告板中连续3代的最优适应度函数值变化都小于u，则执行步骤538，若不发生变异执行步骤539；

步骤538，执行自适应t分布变异，具体为：在萤火虫算法中引入自适应t分布变异操作，利用目前为止所有迭代次数中最优适应度函数值所属萤火虫个体的状态替换当前种群中最差萤火虫个体的状态，然后对本次迭代种群中的最优个体进行高斯变异，对其他的个体按式$\boldsymbol{X}_i^{New}=\boldsymbol{X}_i(1+k\cdot t(t_{max}))$进行t分布变异，其中，$\boldsymbol{X}_i^{New}$是变异后个体的位置，k是1到0之间递减的变量，$t(t_{max})$是以$t_{max}$为参数自由度的学生分布，$t_{max}$为最大迭代次数，进而计算所有个体变异后的适应度函数值，若优于公告板信息，则更新公告板；

步骤539，完成一次迭代，判断迭代次数是否达到t_{max}，若满足则退出迭代，输出公告板上最优的适应度函数值；若不满足执行步骤533，进行下一次迭代。

8.根据权利要求7所述的基于改进BPNN的网络安全态势预测方法，其特征在于，所述步骤532中的适应度函数为：

$$J(\Theta)=\sqrt{\frac{1}{N}\sum_{t=1}^{N}[\varepsilon(t,X)]^2}$$

$\varepsilon(t,X)=y(t)-y_N(t,\Theta)$

其中$y(t)$为期望输出，$y_N(t,\Theta)$为实际输出，N为训练集的样本数。

一种基于改进BPNN的网络安全态势预测方法

技术领域

[0001] 本发明涉及网络安全评估技术领域，尤其涉及一种基于改进反向传播神经网络(Back propagation neural network,BPNN)的安全态势预测方法。

背景技术

[0002] 近年来，随着移动互联网和智能终端时代的到来与普及，人们的线上行为越来越频繁，营销规模越来越大，各种社交网络组成了复杂、异构的大规模网络。然而，由于通信网络存在可移动性、可扩展性、大规模性、泛在性等特性，在网络深入人们社会生活的同时，也成为黑客攻击的首要目标，导致网络安全漏洞数量持续快速增长。因此，安全问题必将成为未来大规模网络首要解决的问题。在传统技术无法满足人们对大规模网络安全需求的情况下，各国专家学者继而将研究重点转向了网络安全态势感知研究。

[0003] 网络安全态势预测就是借助过去和现在的黑客攻击行为要素信息，得到对网络未来状态的预测，其本质就是一种根据现在的黑客行为特征推测未来网络安全发展态势的技术方法。一个完整的网络安全态势感知体系包括：在对真实网络的安全要素信息进行提取、理解的前提下，通过对历史和当前数据的观测和分析，进而对网络的未来安全趋势做出推测，辅助网络管理员及时了解网络系统即将发生的攻击行为，并做出及时的防御措施。网络安全态势预测作为态势感知过程的最高层，是网络安全态势感知研究的最终目的。

[0004] 目前，各国对于网络安全态势感知的研究还处于起步阶段，虽然相关理论和技术都还不太成熟，但研究人员已尝试从不同角度出发研究和提出相关的网络态势预测方法。

[0005] Endsley最早给出了态势感知的概念，即从空间和时间两个维度感知环境中 的要素，综合理解感知信息并预测未来的状况。

[0006] 卓颖等人提出了基于广义回归神经网络的态势预测方法，首先对历史数据进行分类，针对各个类别的数据建立广义神经网络模型，进行态势预测，具有较好的预测精度。

[0007] Zhang Guiling等人借助模糊神经网络在处理模糊性、非线性等问题上的优势，提出了基于模糊神经网络的入侵攻击评估模型，用于预测入侵行为。

[0008] Liu Z等人从不同角度对网络态势感知开展了研究，提出了采用数据挖掘的方法进行态势感知和预测，但是上述研究存在态势要素提取不全面，计算复杂度过大导致维数爆炸等问题。

[0009] 谢丽霞等人提出基于神经网络的网络安全态势感知方法，采用遗传算法优化径向基函数(Radical Basis Function,RBF)神经网络，有效提高了预测精度，但在对历史数据集进行相空间重构时，人为指定输入维数缺乏一定的理论依据，具有一定的局限性。

[0010] 针对上述提出的各种网络安全态势预测方法存在的不足与缺陷，需要寻找一种高效准确地网络安全态势预测方法。

发明内容

[0011] 本发明的目的是提供一种基于改进BPNN的网络安全态势预测方法，用以解决现有

的人为指定输入维数导致网络不可预测以及网络容易陷入局部最优导致网络安全态势预测精度低的问题。

[0012] 本发明为解决上述技术问题，提供一种基于改进BPNN的网络安全态势预测方法，该方法包括以下步骤：

[0013] 步骤1，对采集的漏洞、流量、入侵检测系统等数据进行态势要素获取，并通过层次化网络安全态势评估量化方法对收集到的态势要素信息进行评估量化处理；

[0014] 步骤2，运用极值化公式对量化后产生的非线性时间序列态势值进行预处理，再寻找最适合的嵌入维数与延迟时间进行相空间重构，并通过计算该非线 性的时间序列的李雅普诺夫(Lyapunov)指数来确定是否有可预测性；

[0015] 步骤3，将空间重构得到的态势值样本分为训练集与测试集；

[0016] 步骤4，根据非线性时间序列的特点与经验确定BP神经网络的输出层与隐含层的节点数，设定输入层节点数为嵌入维数，从而确定神经网络的结构，并初始化BP神经网络的向量参数Θ；

[0017] 步骤5，采用改进萤火虫算法(Improved glowworm swarm optimization，IGSO)对BP神经网络进行参数寻优，从而确定网络权值和偏置值，建立网络安全态势的预测模型；

[0018] 步骤6，将测试集输入至有最优权值和阈值的BPNN中，得到预测值，最后再将其反极值化，得到最终的态势值。

[0019] 优选地，所述步骤2进一步包括以下步骤：

[0020] 步骤21，建模极值标准化公式如下所示：

[0021] $$x'(i)=\frac{x(i)-x(i)_{\min}}{x(i)_{\max}-x(i)_{\min}} \qquad i=1,2,\cdots,n$$

[0022] 其中，x(i)与x'(i)分别为处理前后的网络安全态势值，$x(i)_{\min}$与$x(i)_{\max}$分别表示处理前所有网络安全态势值中的最小值与最大值，且通过处理后得到的网络安全态势数据x'(i)，i＝1，2，…n.是一组一维时间序列，其中n为一段时间内的网络安全态势样本数；

[0023] 步骤22，采用最小互信息法计算最佳时间延时τ，并将τ和cao氏法相结合确定嵌入维数，从而得出BP网络的输入节点数m；

[0024] 步骤23，根据cao氏法与互信息法得到的m与τ，引入最大Lyapunov指数来验证数据具有可预测性。

[0025] 优选地，所述步骤22中的最佳时间延时τ的计算公式为：

[0026] $$I(\tau)=\sum_{i,j}P_{ab}(x'(t_i),x'(t_i+\tau))\log_2\left[\frac{P_{ab}(x'(t_i),x'(t_i+\tau))}{P_a(x'(t_i))P_b(x'(t_i+\tau))}\right]$$

[0027] 其中，定义事件a表示网络安全态势样本序列$x'(t_i)$，事件b表示进行时间延迟 的网络安全态势样本序列$x'(t_i+\tau)$，$p_a(x'(t_i))$与$p_b(x'(t_i+\tau))$分别表示a、b两事件中$x'(t_i)$与$x'(t_i+\tau)$会发生的概率，$P_{ab}(x'(t_i),x'(t_i+\tau))$为$x'(t_i)$和$x'(t_i+\tau)$两事件联合分布概率；通过对该公式分析可知，如果I(τ)等于0，则代表$x'(t_i)$与$x'(t_i+\tau)$无相关，即$x'(t_i+\tau)$是不可以预测的；若I(τ)取得极小值，表示$x'(t_i)$与$x'(t_i+\tau)$具有最大可能的不相关，因此取I(τ)的第一个极小值为最佳时间延迟τ。

[0028] 优选地，所述步骤22中的根据cao氏法确定输入神经元数m的计算公式为：

[0029] $$a(i,m)=\frac{\left\|X_i(m+1)-X_{n(i,m)}(m+1)\right\|}{\left\|X_i(m)-X_{n(i,m)}(m)\right\|}$$

[0030] $$E(m)=\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}a(i,m)$$

[0031] $E_1(m)=E(m+1)/E(m)$

[0032] m代表嵌入维数，也即神经网络的输入节点数，就是通过这几个公式来确定，m从1开始取，一直到$E_1(m)$停止变化；

[0033] 其中，$X_i(m)$和$X_i(m+1)$分别表示嵌入维为m和m+1时重构相空间的第i个向量，$X_{n(i,m)}(m)$和$X_{n(i,m)}(m+1)$分别表示与$X_i(m)$和$X_i(m+1)$最近的向量，$||\cdot||$为欧几里得距离，则$a(i,m)$用于判断$X_{n(i,m)}(m)$是否为$X_i(m)$的真实临近点，若在m维相空间临近的两个点在m+1维相空间依然临近，则为“真实临近点”，否则为“虚假临近点”；$E(m)$和$E(m+1)$分别表示在m维和m+1维下非线性时间序列上点与其相邻点之间的平均统计距离，N表示态势值时间序列；进一步，通过对上述公式分析可知，如果网络安全态势的非线性时间序列当中包含确切的规律，那么就一定能够找到一个合适m，当m大于某固定值m_0时，$E_1(m)$开始停止较大变化则可将m_0+1当作最小嵌入维数，其中判断是否停止较大变化，可以设置一个在0到1范围内波动的$E_2(m)$，来对比$E_1(m)$是否大幅增加还是已经停止较大变化，$E_2(m)$设置准则如下：

[0034] $E_2(m)=E^*(m+1)/E^*(m)$

[0035] $$E^*(m)=\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}\left|X(i+m\tau)-X_{n(i,m)}(i+m\tau)\right|$$

[0036] 对于随机事件序列，数据内部无关联，因此是不可预测的，$E_2(m)$将始终为1，而对于确定性时间序列，相邻点之间的关系会随着嵌入维数m的值变化，因此总有一些m使得$E_2(m)$不等于1，因此，$E_2(m)$的波动程度能够用来度量时间序列中的确定性元素。

[0037] 优选地，所述步骤2的相空间重构方法为：

[0038] $$\begin{cases}X_i(m)=\{x'(i),x'(i+\tau),\cdots,x'(i+(m-1)\tau)\},i=1,2,\cdots M\\M=N-(m-1)\tau\end{cases}$$

[0039] 其中，m和τ根据步骤22得出，x’(i)为极值化后的一维时间序列，M表示重构相点的数量，m为嵌入维数，即输入层节点数，τ为延迟时间。

[0040] 优选地，所述步骤5进一步包括以下步骤：

[0041] 步骤51，将萤火虫群的个体位置映射为BP神经网络的向量参数Θ，并指定种群内萤火虫个体的数目，对所有的个体进行随机实数编码，使得萤火虫种群均匀分布在D维的搜索空间里；

[0042] 步骤52，初始化IGSO算法的参数，其中包括：最大迭代次数t_{max}、最小移动步长s_{min}、最大移动步长s_{max}、萤火素更新参数ρ、适应度函数参数γ、萤火素初始值l_0、萤火虫感知范围r_s；

[0043] 步骤53，按照IGSO算法进行迭代寻优，得到萤火虫种群在搜索空间中的全局最优解，即得到BPNN对网络安全态势训练样本预测精度最高的一组向量参数Θ，并基于该组向量参数Θ来构建BP网络中各层之间的连接权值与各节点之间的阈值，进而得到网络安全态势值泛化能力最强的BPNN网络模型。

[0044] 进一步，所述步骤53中IGSO算法具体步骤为：

[0045] 步骤531，参数及种群初始化，即设定种群个体数目并在解空间中随机初始化个体位置，计算初始化种群每个个体的适应度函数值，同时生成公告板；

[0046] 步骤532，对种群中的所有萤火虫个体按$l_i(t)=(1-\rho)l_i(t-1)+\gamma J(x_i(t))$更新萤火素值，其中，$l_i(t)$表示第t次迭代中第i个萤火虫所携带的萤火素，$\rho\in(0,1)$为萤火素更新参数，$\gamma$为适应度函数参数，$J(x)$为适应度函数；

[0047] 步骤533，进入迭代阶段，求解种群中个体的邻居萤火虫的集合，如果邻居集合存在则转到步骤534，否则转到步骤536；

[0048] 步骤534，根据轮盘赌的方法计算萤火虫i在其决策域内的移动方向，同时为了摆脱陷入局部最优，引入变步长来代替固定步长进行移动步长的更新，并设定变步长公式为：$s(t)=s_{\max}e^{c\cdot t}$，$c=\frac{1}{t_{\max}}\ln(\frac{s_{\min}}{s_{\max}})$；其中，$t_{\max}$为最大迭代次数，$s_{\min}$为最小移动步长，$s_{\max}$为最大移动步长；

[0049] 步骤535，根据534的步长$s(t)$进行位置更新，则萤火虫在t+1次迭代的位置$x_i(t+1)$的更新公式为：

[0050] $$x_i(t+1)=x_i(t)+s\left(\frac{x_j(t)-x_i(t)}{\left\|x_j(t)-x_i(t)\right\|}\right)$$

[0051] 其中，$x_i(t)$表示萤火虫i在第t次迭代的位置，$x_j(t)$表示萤火虫i在第t次迭代时决策域内的第j只萤火虫的位置，同时更新萤火虫个体的决策域，设定第i只萤火虫在t+1次迭代时刻的动态决策范围$r_d^i(t+1)$为：

[0052] $$r_d^i(t+1)=\min\left\{r_s,\max\left\{0,r_d^i(t)+\beta(n_t-|N_i(t)|)\right\}\right\}$$

[0053] 其中，r_s为萤火虫感知范围，$r_d^i(t)$为第i只萤火虫t次迭代时刻的动态决策范围，β为比例常数，n_t为邻居阈值；$N_i(t)=\left\{j:\left\|x_j(t)-x_i(t)\right\|<r_d^i;l_i(t)<l_j(t)\right\}$表示第i只萤火虫在第t次迭代时，它的决策域内所包含的萤火虫的集合，$l_i(t)$表示第t次迭代中第i个萤火虫所携带的萤火素，$l_j(t)$表示第t次迭代中第j个萤火虫所携带的萤火素，其中，$j\in N_i(t)$，$||x||$表示x的范数；

[0054] 步骤536，计算当前种群所有个体对应的适应度函数值，取其中最好的适应度函数值与公告板中的值进行对比，若优于公告板信息，则选择更新公告板；

[0055] 步骤537，根据条件判断，如果发生变异即当迭代次数大于2且公告板中连续3代的最优适应度函数值变化都小于u，则执行步骤538，若不发生变异执行步骤539；

[0056] 步骤538，执行自适应t分布变异，具体为：在萤火虫算法中引入自适应t分布变异操作，利用目前为止所有迭代次数中最优适应度函数值所属萤火虫个体的状态替换当前种群中最差萤火虫个体的状态，然后对本次迭代种群中的最优个体进行高斯变异，对其他的个体按式$\boldsymbol{X}_i^{New}=\boldsymbol{X}_i(1+k\cdot t(t_{\max}))$进行t分布变异，其中，$\boldsymbol{X}_i^{New}$是变异后个体的位置，k是1到0之间递减的变量，$t(t_{\max})$是以$t_{\max}$为参数自由度的学生分布，$t_{\max}$为最大迭代次数，进而计算所有个体变异后的适应度函数值，若优于公告板信息，则更新公告板；

[0057] 步骤539，完成一次迭代，判断迭代次数是否达到t_{max}，若满足则退出迭代，输出公告板上最优的适应度函数值；若不满足执行步骤533，进行下一次迭代。

[0058] 优选地，所述步骤532中的适应度函数为：

[0059] $$J(\Theta)=\sqrt{\frac{1}{N}\sum_{t=1}^{N}[\varepsilon(t,X)]^2}$$

[0060] $\varepsilon(t,X)=y(t)-y_N(t,\Theta)$

[0061] 其中，$y(t)$为期望输出，$y_N(t,\Theta)$为实际输出，N代表训练集的样本数。

[0062] 与现有技术相比，本发明达到的有益效果是：

[0063] 本发明提供了一种改进BPNN的网络安全态势预测方法，通过采集网络和主机的异常信息，过滤网络安全威胁报警事件，从而建立预测模型的训练样本集；使用混沌理论和BP神经网络相结合的方法建立网络安全态势预测模型，通过对样本数据进行相空间重构，避免了人为设定神经网络输入层节点数的问题，同时分析重构后样本的最大李雅普诺夫指数来得到评估出来的样本是具有混沌预测性；考虑到神经网络易陷入局部最优，因此用改进的萤火虫算法对其进行优化；进而本发明能够较为精确的对网络安全进行预测，同时能够提高网络安全态势预测收敛速度。

附图说明

[0064] 图1是本发明提供的基于改进BPNN的网络安全态势预测方法的流程图；

[0065] 图2是本发明中网络安全态势要素评估量化模型简化图；

[0066] 图3是本发明中神经网络输入维数m的仿真图；

[0067] 图4是本发明与BPNN、GSO-BPNN的仿真比较图；

[0068] 图5是本发明与其他智能优化算法的仿真比较图。

具体实施方式

[0069] 为了使本发明的目的、技术方案及优点更加清楚明白，下面结合附图对本发明的具体实施方式作进一步说明。

[0070] 本发明所提的基于改进BPNN的网络安全态势预测方法，通过对历史时刻的网络安全态势值进行相空间重构，得出训练集和测试集，同时用改进萤火虫算法优化反向传播神经网络，最后使用训练好的反向传播神经网络进行下一时刻的网络安全态势值预测，图1为本发明提供的基于改进BPNN的网络安全态势预测方法的流程图，该方法包括以下步骤：

[0071] 步骤1，对采集的漏洞、流量、入侵检测系统等数据进行态势要素获取，并通过层次化网络安全态势评估量化方法对收集到的态势要素信息进行评估量化处理；

[0072] 步骤2，运用极值化公式对量化后产生的非线性时间序列态势值进行预处理，再寻找最适合的嵌入维数与延迟时间进行相空间重构，并通过计算该非线性的时间序列的Lyapunov指数来确定是否有可预测性；

[0073] 步骤3，将空间重构得到的态势值样本分为训练集与测试集；

[0074] 步骤4，根据非线性时间序列的特点与经验确定BP神经网络的输出层与隐含层的节点数，设定输入层节点数为嵌入维数，从而确定神经网络的结构，并初始化BP神经网络的向量参数Θ；

[0075] 步骤5，采用改进萤火虫算法IGSO对BP神经网络进行参数寻优，从而确定网络权值和偏置值，建立网络安全态势的预测模型；

[0076] 步骤6，将测试集输入至有最优权值和阈值的BPNN中，得到预测值，最后再将其反极值化，得到最终的态势值。

[0077] 根据本发明，所述步骤2进一步包括以下步骤：

[0078] 步骤21，建模极值标准化公式如下所示：

[0079] $$x'(i)=\frac{x(i)-x(i)_{\min}}{x(i)_{\max}-x(i)_{\min}} \qquad i=1,2,\cdots,n$$

[0080] 其中，x(i)与x'(i)分别为处理前后的网络安全态势值，$x(i)_{min}$与$x(i)_{max}$分别表示处理前所有网络安全态势值中的最小值与最大值，通过处理后得到的网络安全态势数据x'(i)，i＝1，2，…n.是一组一维时间序列，n为一段时间内的网络安全态势样本数；

[0081] 步骤22，采用最小互信息法计算最佳时间延时τ，结合τ和cao氏法确定嵌入维数，从而得出BP网络的输入节点数m；

[0082] 其中，建模最佳时间延时τ的计算公式为：

[0083] $$I(\tau)=\sum_{i,j}P_{ab}(x'(t_i),x'(t_i+\tau))\log_2\left[\frac{P_{ab}(x'(t_i),x'(t_i+\tau))}{P_a(x'(t_i))P_b(x'(t_i+\tau))}\right]$$

[0084] 其中，定义事件a表示网络安全态势样本序列$x'(t_i)$，事件b表示进行时间延迟的网络安全态势样本序列$x'(t_i+\tau)$，$p_a(x'(t_i))$与$p_b(x'(t_i+\tau))$分别表示a、b两事件中$x'(t_i)$与$x'(t_i+\tau)$会发生的概率，$P_{ab}(x'(t_i),x'(t_i+\tau))$为$x'(t_i)$和$x'(t_i+\tau)$两事件联合分布概率；通过对该公式分析可知，如果I(τ)等于0，则代表$x'(t_i)$与$x'(t_i+\tau)$无相关，即$x'(t_i+\tau)$是不可以预测的；若I(τ)取得极小值，表示$x'(t_i)$与$x'(t_i+\tau)$具有最大可能的不相关，因此取I(τ)的第一个极小值为最佳时间延迟τ；

[0085] 所述cao氏法可以参见许小可等人论文《基于非线性分析的海杂波处理与目标检测》，大连海事大学，2008，不再详述。

[0086] 进一步，所述步骤22中利用cao氏法确定输入神经元数m的计算公式为：

[0087] $$a(i,m)=\frac{\left\|X_i(m+1)-X_{n(i,m)}(m+1)\right\|}{\left\|X_i(m)-X_{n(i,m)}(m)\right\|}$$

[0088] $$E(m)=\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}a(i,m)$$

[0089] $E_1(m)=E(m+1)/E(m)$

[0090] 其中，$X_i(m)$和$X_i(m+1)$分别表示嵌入维为m和m+1时重构相空间的第i个向量，$X_{n(i,m)}(m)$和$X_{n(i,m)}(m+1)$分别表示与$X_i(m)$和$X_i(m+1)$最近的向量，‖·‖为欧几里得距离，则a(i，m)用于判断$X_{n(i,m)}(m)$是否为$X_i(m)$的真实临近点，若在m维相空间临近的两个点在m+1维相空间依然临近，则为“真实临近点”，否则为“虚假临近点”；E(m)和E(m+1)分别表示在m维和m+1维下非线性时间序列上点与其相邻点之间的平均统计距离，N表示态势值时间序列；进一步，通过对上述公式分析可知，如果网络安全态势的非线性时间序列当中包含确切的规律，那么就一定能够找到一个合适m，当m大于某固定值m_0时，$E_1(m)$开始停止较大变化则可将m_0+1当作最小嵌入维数，其中判断是否停止较大变化，可以设置一个在0到1范围内波动的E_2

(m)，来对比$E_1(m)$是否大幅增加还是已经停止较大变化，$E_2(m)$设置准则如下：

[0091] $E_2(m)=E^*(m+1)/E^*(m)$

[0092] $$E^*(m)=\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}\left|X(i+m\tau)-X_{n(i,m)}(i+m\tau)\right|$$

[0093] 对于随机事件序列，数据内部无关联，因此是不可预测的，$E_2(m)$将始终为1，而对于确定性时间序列，相邻点之间的关系会随着嵌入维数m的值变化，因此总有一些m使得$E_2(m)$不等于1，因此，$E_2(m)$的波动程度能够用来度量时间序列中的确定性元素；

[0094] 步骤23，根据cao氏法与互信息法得到的m与τ，引入最大Lyapunov指数来验证数据具有可预测性。

[0095] 根据本发明，所述步骤5具体包括以下步骤：

[0096] 步骤51，将萤火虫群的个体位置映射为BP神经网络的向量参数Θ，并指定种群内萤火虫个体的数目，对所有的个体进行随机实数编码，使得萤火虫种 群均匀分布在D维的搜索空间里；

[0097] 步骤52，初始化IGSO算法的参数，其中包括：最大迭代次数t_{max}、最小移动步长s_{min}、最大移动步长s_{max}、萤火素更新参数ρ、适应度函数参数γ、萤火素初始值l_0、萤火虫感知范围r_s；

[0098] 步骤53，按照IGSO算法进行迭代寻优，得到萤火虫种群在搜索空间中的全局最优解，即得到BPNN对网络安全态势训练样本预测精度最高的一组向量参数Θ，并基于该组向量参数Θ来构建BP网络中各层之间的连接权值与各节点之间的阈值，进而得到网络安全态势值泛化能力最强的BPNN网络模型。

[0099] 根据本发明，所述步骤51进一步包括以下步骤：

[0100] 步骤511，在解空间中，将具体的萤火虫个体编码为：

[0101] $\Theta=[w,v,\theta,\alpha]$

[0102] 其中，w为隐含层各节点与输入层各节点之间的连接权值，v为隐含层各节点与输出层各节点之间的连接权值，θ为隐含层节点的偏置值，α输出层节点的偏置值；

[0103] 步骤512，搜索空间维数的确定：设输入层节点的个数为m，隐含层节点的个数为p，输出层节点的个数为1，那么，输入层与隐含层的连接权值维数为m×p；隐含层与输出层之间的连接权值维数为p；隐含层节点对应的阈值维数为p；输出层节点对应的阈值维数为1；则算法中萤火虫个体的搜索空间维数为：

[0104] $D=(m\times p+p)+(p+1)$

[0105] 由上式可知，每个萤火虫个体在空间当中都具有D个维度，则萤火虫个体编码可以表示为：$\Theta=[x_1,x_2,\cdots,x_D]$，当搜索到最优的Θ时，该位置的目标函数适应度最大。

[0106] 进一步，所述步骤53中IGSO算法具体包括以下步骤：

[0107] 步骤531，设定种群个体数目并在解空间中随机初始化个体位置，计算初始化种群每个个体的适应度函数值，同时生成公告板；

[0108] 步骤532，对种群中的所有萤火虫个体按$l_i(t)=(1-\rho)l_i(t-1)+\gamma J(x_i(t))$更新萤火素值，其中，$l_i(t)$和$l_i(t-1)$分别表示第t次和第t-1次迭代中第i个萤火虫所携带的萤火素，$\rho\in(0,1)$为萤火素更新参数，γ为适应度函数参数，J(x)为适应度函数，其具体计算公式为：

[0109] $$J(\boldsymbol{\Theta})=\sqrt{\frac{1}{N}\sum_{t=1}^{N}[\varepsilon(t,X)]^2}$$

[0110] $\varepsilon(t,X)=y(t)-y_N(t,\Theta)$

[0111] 其中，$y(t)$为神经网络期望输出，$y_N(t,\Theta)$为神经网络实际输出，N为训练集的样本数；

[0112] 步骤533，进入迭代阶段，求解种群中个体的邻居萤火虫的集合，如果邻居集合存在则转到步骤535，邻居集合不存在就转到步骤536；

[0113] 步骤534，根据轮盘赌的方法计算萤火虫i在其决策域内的移动方向，同时为了摆脱陷入局部最优，引入变步长来代替固定步长进行移动步长的更新，并设定变步长公式为：$s(t)=s_{\max}e^{ct}$，$c=\frac{1}{t_{\max}}\ln(\frac{s_{\min}}{s_{\max}})$；

[0114] 步骤535，根据534的步长s进行位置更新，则萤火虫在t+1次代的位置$x_i(t+1)$的更新公式为：

[0115] $$x_i(t+1)=x_i(t)+s\left(\frac{x_j(t)-x_i(t)}{\left\|x_j(t)-x_i(t)\right\|}\right)$$

[0116] 其中，$x_i(t)$表示萤火虫i在第t次迭代的位置，$x_j(t)$表示萤火虫i在第t次迭代时决策域内的第j只萤火虫的位置，同时更新萤火虫个体的决策域，设定第i只萤火虫在t+1次迭代时刻的动态决策范围$r_d^i(t+1)$为：

[0117] $$r_d^i(t+1)=\min\left\{r_s,\max\left\{0,r_d^i(t)+\beta(n_t-|N_i(t)|)\right\}\right\}$$

[0118] 其中，r_s为萤火虫感知范围，$r_d^i(t)$为第i只萤火虫t次迭代时刻的动态决策范围，β为比例常数，n_t为邻居阈值；$N_i(t)=\left\{j:\left\|x_j(t)-x_i(t)\right\|<r_d^i;l_i(t)<l_j(t)\right\}$表示第i只萤火虫在第t次迭代时，它的决策域内所包含的萤火虫的集合，其中，$j\in N_i(t)$，$||x||$表示x的范数；

[0119] 步骤536，计算当前种群所有个体对应的适应度函数值，取其中最好的适应度函数值与公告板中的值进行对比，若优于公告板信息，则选择更新公告板；

[0120] 步骤537，根据条件判断，如果发生变异即当迭代次数大于2且公告板中连续3代的最优适应度函数值变化都小于u，则执行步骤538，若不发生变异执行步骤539；

[0121] 步骤538，执行自适应t分布变异，具体为：在萤火虫算法中引入自适应t分布变异操作，利用目前为止所有迭代次数中最优适应度函数值所属萤火虫个体的状态替换当前种群中最差萤火虫个体的状态，然后对本次迭代种群中的最优个体进行高斯变异，对其他的个体按式$\boldsymbol{X}_i^{New}=\boldsymbol{X}_i(1+k\cdot t(t_{\max}))$进行t分布变异，其中，$\boldsymbol{X}_i^{New}$是变异后个体的位置，k是1到0之间递减的变量，$t(t_{\max})$是以$t_{\max}$为参数自由度的学生分布，进而计算所有个体变异后的适应度函数值，若优于公告板信息，则更新公告板；

[0122] 步骤539，完成一次迭代，判断迭代次数是否达到$t_{\max}$，若满足则退出迭代，输出公告板上最优的适应度函数值；若不满足执行步骤533，进行下一次迭代。

[0123] 为了说明本发明的有益效果，本发明将结合具体的态势值进行仿真分析。取某公司10-11月中60天里防火墙、入侵检测系统(Intrusion Detection Systems,IDS)等历史日

志信息作为原始数据源。对每天的日志信息进行5次采样，并将采样得到的日志信息按照图2所示方法进行网络安全评估量化，从而得到原始态势值。实验中IGSO算法的具体参数如表1所示。

[0124] 表1仿真参数

[0125]

s_{min}	s_{max}	ρ	γ	l_0	r_s	μ
0.005	0.1	0.4	0.6	5	0.8	10^{-4}

[0126] 图3描述了最小嵌入维数m的确定，对归一化后的网络安全态势值进行互信息法得到最佳时间延时τ＝1，再将τ与cao氏法相结合计算出m。从图中可知，从m＝5开始，$E_1(m)$和$E_2(m)$差值控制在一定范围内，即$E_1(m)$不再发生较大变化，所以确定利用cao氏法求出的最小嵌入维数为5。

[0127] 图4为本发明提出的IGSO-BPNN算法与通过单纯的BPNN算法和未经改进的萤火虫算法优化BPNN(GSO-BPNN)算法得到的态势预测精度对比图。在实验中，设定IGSO、GSO等算法的种群个体数目均取值为30，相当于同时在空间中30个点上一起并行的进行学习，选取拟合最好的点作为权值和阈值进行预测，则对于BPNN模型而言，进30次仿真，取预测精度最高的一组与其他算法进行对比。通过智能算法与神经网络相结合的组合模型比单纯的神经网络预测算法更加符合真实值的趋势。将IGSO-BPNN与GSO-BPNN预测模型进行对比，可以看出改进后的IGSO算法相比GSO算法在寻优过程中更具有优势，经过IGSO优化后的BPNN神经网络模型精度更高，且IGSO-BPNN模型预测的态势趋势更接近于真实趋势。

[0128] 图5给出了通过本发明提出的IGSO-BPNN算法、遗传算法以及粒子群算法优化BPNN算法得到的网络安全态势预测效果的对比图。在仿真中，设定网络的最大迭代次数为100次，种群最大数量为30，对20组数据进行预测。通过比较可以看出，以实际值曲线作为衡量准则，本发明所提IGSO-BPNN预测模型相比其它两种优化算法所预测得到的结果，其所预测出的趋势走向更贴近真实态势值的趋势。

[0129] 本发明所举实施方式或者实施例对本发明的目的、技术方案和优点进行了进一步的详细说明，所应理解的是，以上所举实施方式或者实施例仅为本发明的优选实施方式而已，并不用以限制本发明，凡在本发明的精神和原则之内对本发明所作的任何修改、等同替换、改进等均应包含在本发明的保护范围之内。

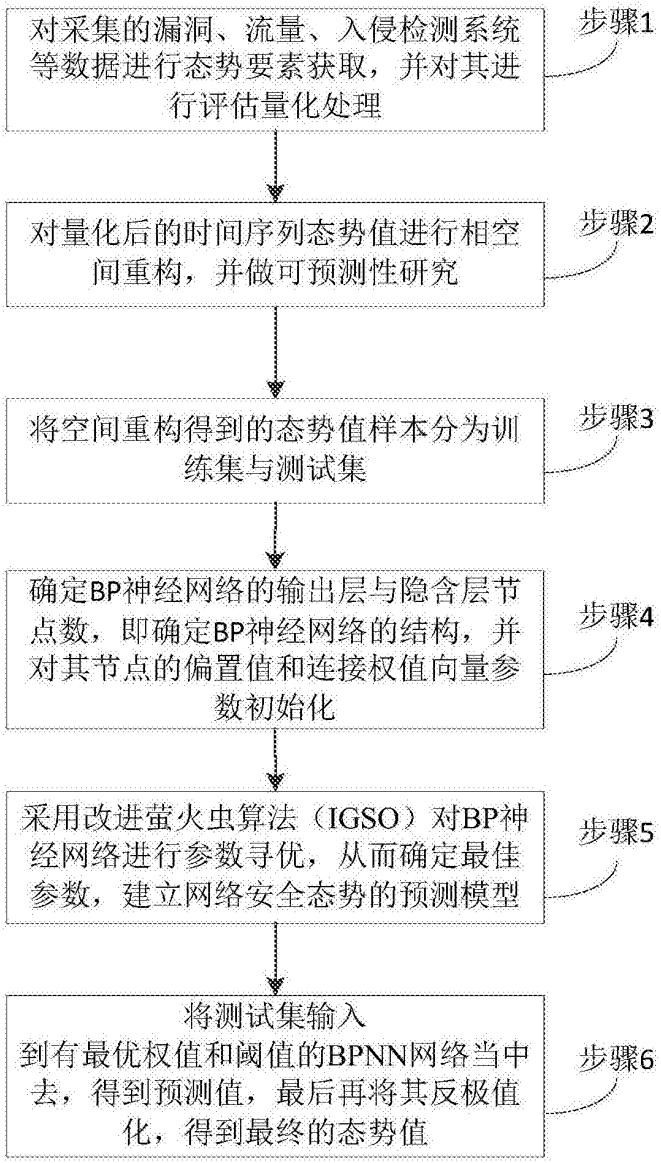

对采集的漏洞、流量、入侵检测系统等数据进行态势要素获取，并对其进行评估量化处理
步骤1
对量化后的时间序列态势值进行相空间重构，并做可预测性研究
步骤2
将空间重构得到的态势值样本分为训练集与测试集
步骤3
确定BP神经网络的输出层与隐含层节点数，即确定BP神经网络的结构，并对其节点的偏置值和连接权值向量参数初始化
步骤4
采用改进萤火虫算法（IGSO）对BP神经网络进行参数寻优，从而确定最佳参数，建立网络安全态势的预测模型
步骤5
将测试集输入到有最优权值和阈值的BPNN网络当中去，得到预测值，最后再将其反极值化，得到最终的态势值
步骤6

图1

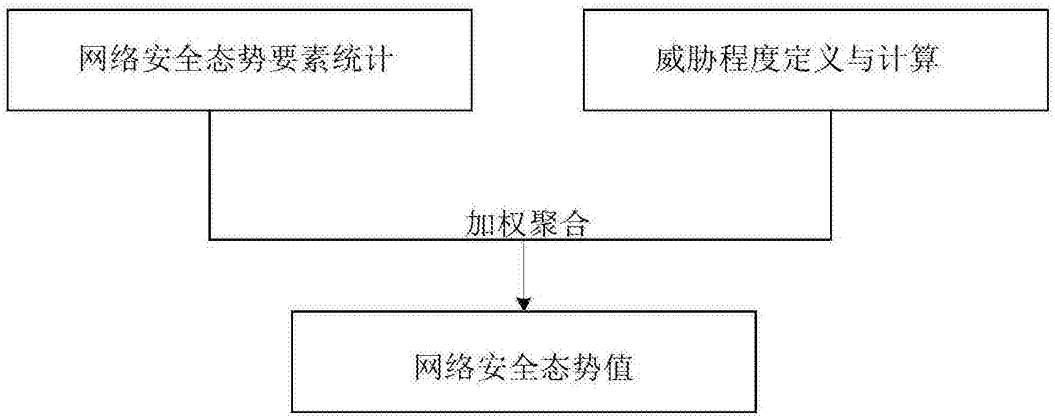

网络安全态势要素统计
威胁程度定义与计算
加权聚合
网络安全态势值

图2

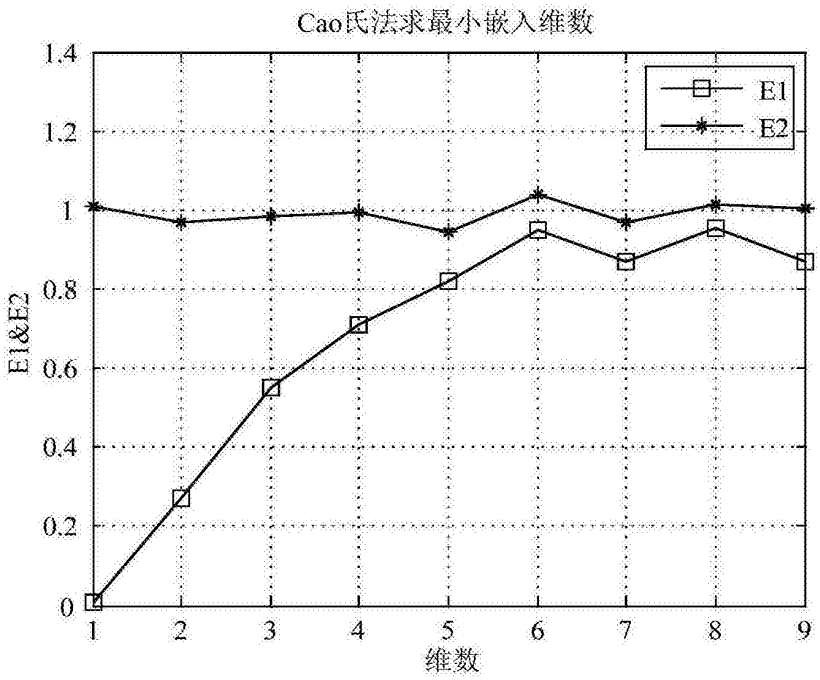

Cao氏法求最小嵌入维数
E1
E2
E1&E2
维数
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1
2
3
4
5
6
7
8
9

图3

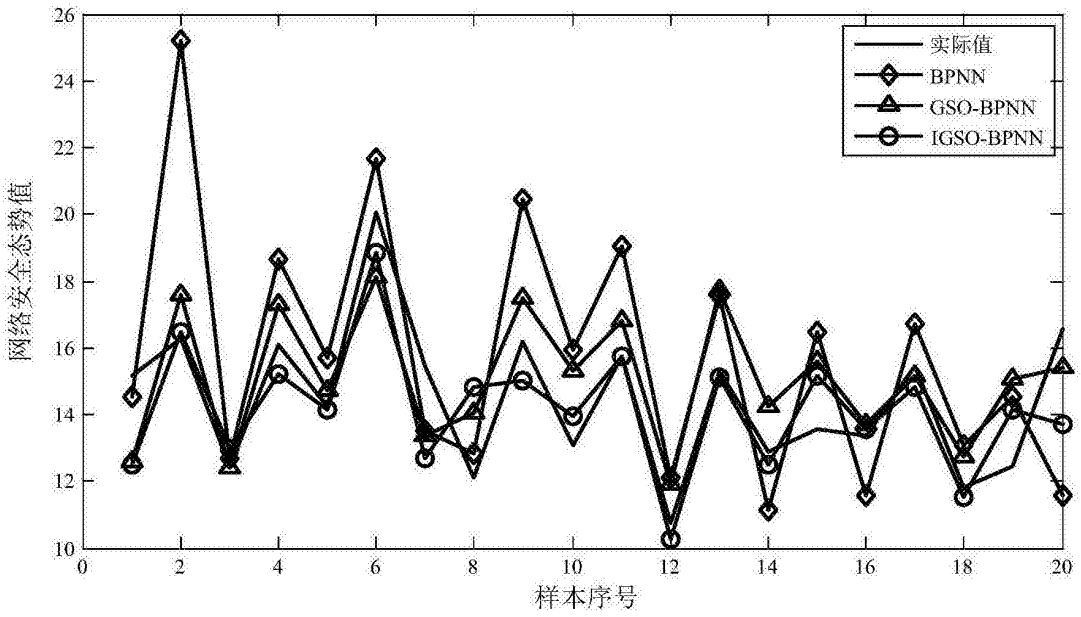

实际值
BPNN
GSO-BPNN
IGSO-BPNN
网络安全态势值
样本序号

图4

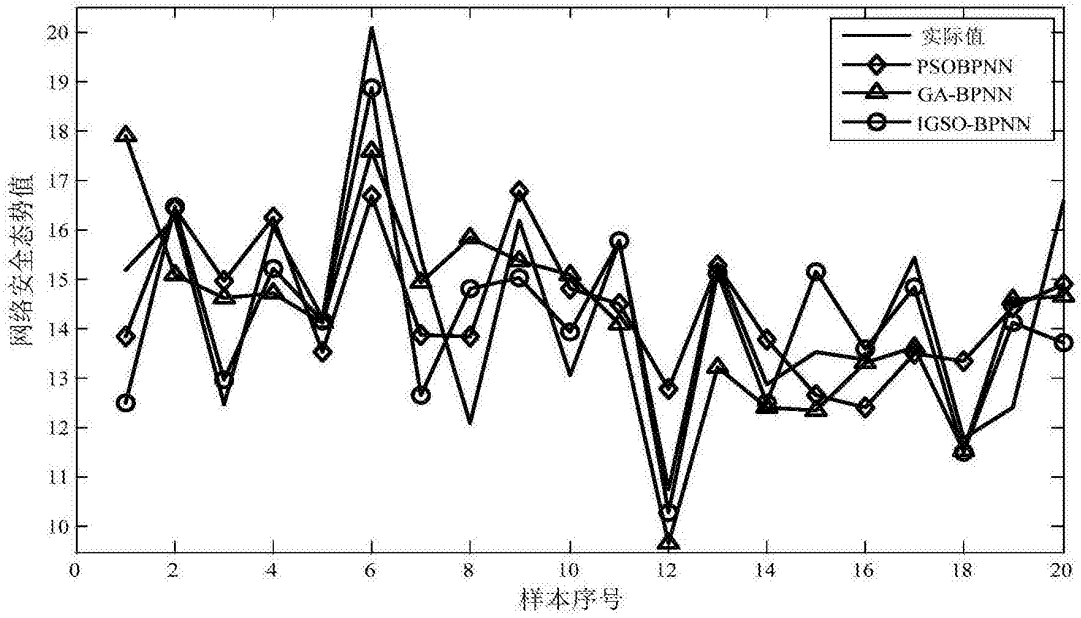

实际值
PSOBPNN
GA-BPNN
IGSO-BPNN
网络安全态势值
样本序号

图5