基于 HHGA-RBF 神经网络的网络安全态势预测模型

孟 锦 马 驰 何加浪 张 宏
(南京理工大学计算机科学与技术学院 623 教研室 南京 210094)

摘 要 针对网络安全态势感知中的预测问题,提出了采用径向基函数(RBF)神经网络对态势值进行预测的方法。为了提高 RBF 神经网络的预测精度,使用混合递阶遗传算法(HHGA)对 RBF 神经网络进行训练,获得了神经网络结构参数。实验结果说明了此预测方法的有效性,并通过与已有的预测方法进行对比实验,验证了所提算法在精度方面的优越性。

关键词 混合递阶遗传算法,网络安全态势,态势预测,RBF 神经网络
中图法分类号 TP393.08,TP18 **文献标识码** A

Network Security Situation Prediction Model Based on HHGA-RBF Neural Network

MENG Jin MA Chi HE Jia-lang ZHANG Hong
(Dept. of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract To address the prediction problem in the Network Security Situational Awareness, the RBF neural network method was proposed to predict the situation value. In order to improve the prediction accuracy of RBF neural network, the HHGA was used for training to acquire the neural network parameters. Experimental results show the validity of this prediction method, and the accuracy of the proposed algorithm was verified by comparative experiment with the existing forecasting methods.

Keywords HHGA, Network security situation, Situation prediction, RBF neural network

1 引言

网络安全态势感知(Network Security Situational Awareness)是目前网络安全领域的一个研究热点,已经引起了相关科研机构和研究人员的足够重视。网络安全态势评估技术能够综合各方面的安全因素,融合防火墙、防病毒软件、入侵检测系统(IDS)、安全审计系统等安全措施的数据信息,从整体上反映网络安全状况,对当前的网络安全状态进行定量和定性的评估,为增强网络安全提供可靠的参照依据。

1999 年 Bass 等人首次提出了网络态势感知概念[1],态势感知源自于空中交通监管(ATC)态势感知概念。Bass 将两者进行类比,旨在把 ATC 态势感知的成熟理论和技术运用到网络态势感知中去,随后 Bass 提出应用多传感器数据融合建立网络空间态势感知的框架,利用入侵检测系统的分布式传感器进行数据融合的方法,对网络安全态势进行评估[2]。此后,国内外的研究人员依据不同的技术思路,设计并实现了大量针对网络安全的态势评估方法[3],但这些方法只能提供给管理员过去和当前的网络安全态势。研究人员逐渐意识到态势感知并不仅仅是态势评估,还需要对网络安全态势进行预测,以辅助网络管理者进行决策和安全管理,使得网络安全管理从被动变为主动。网络安全态势预测通过统计过去到现在的态势值,对未来的发展趋势进行预测。

本文根据网络安全态势值具有非线性时间序列的特点,利用神经网络处理非线性数据的优势,提出基于 RBF 神经网络进行态势预测的方法。为了得到更精确的预测数据,采用混合递阶遗传算法(HHGA, Hybrid Hierarchy Genetic Algorithm)对 RBF 神经网络进行优化。最后进行了仿真实验,运用提出的态势预测模型得到了预测态势值。为了进一步说明本文模型的有效性,将它与已有的态势预测模型进行了对比实验,通过误差分析说明了本文模型的效果。

2 RBF 神经网络

RBF 神经网络是以函数逼近理论为基础而构造的一种典型的 3 层前向型网络。它的每个隐层神经元都构成了拟合平面的一个基函数。其结构由输入层、隐层和输出层构成:输入层负责将输入信号传递到隐层;隐层包含一组径向基函数,负责从输入层接收信号后进行最主要的数学处理,最后送至输出层;输出层负责将隐层的输出信号加权聚合为神经网络的输出信号。

$X=(x_1,x_2,\cdots,x_n)^T\in R^n$ 为网络输入矢量,$y\in R^m$ 为网络输出矢量,W 为隐层与输出层之间的 $m\times 1$ 阶权值矩阵,$\Phi=(\phi_1,\phi_2,\cdots,\phi_m)^T$ 为隐层输出,Φ 的各个分量均定义为具

到稿日期:2010-11-11 返修日期:2011-02-23 本文受国家自然科学基金项目(90718021,60903027),自主科研先期投入计划(2010XQTR04)资助。

孟 锦(1984—),女,博士生,主要研究领域为信息安全、风险评估、态势感知,E-mail:mengjin1107@yahoo.com.cn;马 驰(1984—),男,博士生,主要研究领域为信息安全、无线自主网抗毁;何加浪(1984—),男,博士生,主要研究领域为可信软件;张 宏(1956—),男,博士生导师,主要研究领域为信息安全、数据挖掘。

有径向对称的基函数，通常取如下形式的高斯函数：

$$\phi_i=\exp(-\frac{\|X-c_i\|^2}{2\sigma_i^2}),i=1,2,\cdots,L \quad (1)$$

式中，c_i 为高斯函数的中心，σ_i 为高斯函数的宽度，L 为隐层节点个数。

RBF 网络的输出层对隐层基函数的输出进行线性加权组合，网络映射输出为：

$$y_k=\sum_{j=1}^{L}w_{kj}\varphi_j,j=1,2,\cdots,L \quad (2)$$

式中，w_{kj} 为隐层第 j 个输出与输出层第 k 个神经元之间的连接权值。

RBF 神经网络的设计包括结构和参数两方面，其中结构设计方面主要是指隐节点数目的选择，参数设计方面是指选择合适的隐节点对应径向基函数的中心和宽度以及隐层与输出层之间的连接权值。

3 HHGA-RBFNN 预测模型

遗传算法(GA，Genetic Algorithm)是基于自然选择在计算机上模拟生物进化机制的寻优搜索算法。遗传算法具有全局搜索、收敛速度快的特点，将其与神经网络结合起来，不仅能发挥神经网络的泛化映射能力，而且能使神经网络克服收敛速度慢和易陷入局部最优的缺点。GA 改变了过去运用试探法来确定隐层节点数目的方法，采用全局搜索的方法，以输出误差最小为目标在隐层节点定义域中寻找网络的隐层节点数目。

递阶遗传算法(HGA，Hierarchy Genetic Algorithm)是近年来提出的一种新型遗传算法[4]，采用二进制编码和实数编码相结合的混合编码方法，可以在对神经网络参数优化求解的同时对神经网络的结构进行优化，具有较高的学习效率。

基于 HGA 的 RBF 神经网络算法能够根据样本数据确定 RBF 神经网络的结构和参数，但在学习过程中算法的收敛速度较慢。它将输出层神经元的连接权重放到染色体中用遗传算法进行搜索。分析 RBF 神经网络的结构可知，RBF 神经网络输出层为线性神经元，因此在确定了中心和扩展宽度后，输出层权值可以采用最小二乘法进行设计。为此，将 HGA 与最小二乘法相结合，采用基于混合递阶遗传算法(HHGA)的 RBF 神经网络学习算法，其中递阶染色体中只包含隐层参数，输出层的设计在 GA 的评价函数中完成。

混合递阶遗传算法优化神经网络的算法流程如图 1 所示。

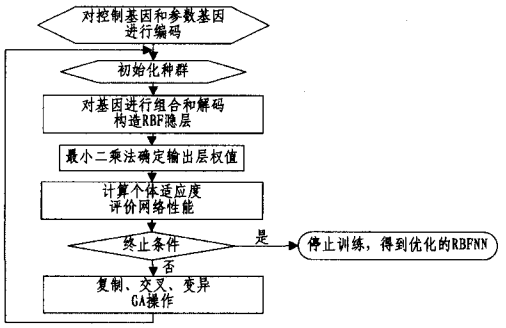

图 1 HHGA-RBFNN 流程图

3.1 编码：采用 HGA 的编码方法

HGA 中染色体由控制基因和参数基因两部分组成。控制基因采用二进制编码，编码长度为最大隐层节点个数，每一位对应一个隐层节点，“1”表示隐层节点存在，对应的参数基因处于有效状态；“0”表示隐层节点不存在，对应的参数基因处于无效状态。控制基因中“1”的个数即为隐层节点的个数。为了加强遗传算法在解空间的搜索能力，参数基因采用实数编码，表示隐层节点中心和宽度。

3.2 初始化

设种群大小为 Q。合适的群体规模对遗传算法的收敛具有重要的意义。群体太小难以求得满意的结果，群体太大则计算复杂。依据经验，群体规模一般取 20～160。

将控制基因和参数基因分别初始化，前者设置最大值 M(即最大隐层节点数为 M)，最小值为 1。后者初始化为[0，1]区间的随机数。

3.3 适应度函数

训练 RBF 神经网络的目标是使其在满足一定精度的要求下具有最简单的网络结构，也就是使网络的精确度和网络的复杂度的综合指标达到最小。

网络的精度目标函数：

$$F_1=SSE=\sum_{k=1}^{N}(x_k-y_k)^2 \quad (3)$$

式中，SSE 为网络输出与期望输出之间的误差平方和。

网络复杂度由隐层节点数决定，目标函数：

$$F_2=L \quad (4)$$

为使递阶遗传算法有效地训练 RBF 网络，需要建立能同时反映这两个目标的适应度函数。本文采用了最小信息量准则(AIC)适应度函数：

$$f=\left[N\log[\frac{1}{N}\sum_{i=1}^{N}(y_i-y_i')2]+4L\right]+b \quad (5)$$

式中，N 为样本数，L 为隐层节点数，y_i 为期望输出值，y_i' 为训练 RBF 网络输出值，b 为一足够大的值。SSE 越小，L 越小，f 将越大。

3.4 遗传操作

3.4.1 选择与复制

选择若干适应度值最大的染色体作为父本，直接遗传给下一代。HHGA 与 GA 的选择操作一样，适应度越大的个体被选择的概率也越大。

采用期待值法来求个体的期望值：

$$v_i=\frac{f_i}{\bar{f}}=\frac{f_i}{f_{sum}/N} \quad (6)$$

式中，f 为个体 i 的适应度，f_{sum} 为种群的总适应度，N 为种群规模。

个体期待值确定种群中的个体是否进入下一代进行优化，个体 i 被复制的个数为 v_i。初始化种群经过选择与复制由 P_1 成为 P_2。

3.4.2 交叉与变异

交叉的目的在于产生新的基因组合，交叉后形成种群 P_3。

控制基因和参数基因使用不同的编码方式，因此分别进行交叉处理。控制基因的交叉遵循二进制编码的交叉规则：一点交叉操作，即在个体串中随机设定一个交叉点，实行交叉时该点前后的两个个体的部分结构进行交换，并产生两个新个体。

参数基因采用的是实值编码，因此需要采用模拟二进制交叉操作。模拟二进制交叉从父代群体中随机选取两个个体 x_1 和 x_2，按下式定义的线性组合交叉方式，将 x_1，x_2 对应交

叉位的值相组合产生新后代：

$$\begin{cases} y_1 = \alpha x_2 + (1-\alpha) x_1 \\ y_2 = \alpha x_1 + (1-\alpha) x_2 \end{cases} \tag{7}$$

式中，α 是一个随机数，$\alpha \in [0,1]$。

变异运算用来模拟生物在自然的遗传环境中的基因突变，通过变异操作，可确保种群中遗传基因类型的多样性，以使搜索能在尽可能大的空间中进行。变异操作是按一定的概率从种群 P_3 中每次随机选取一个个体，随机变化选定个体的某一个或某些基因位，形成种群 P_4。

对控制基因，即染色体以二进制编码的系统中，以一定的概率对其进行求反运算，随机地将染色体的某一个基因由 1 变成 0，或由 0 变成 l。

对于参数基因的实值编码，用偏置变异，以一定概率给该位加上一个随机偏置值变。

4 实验分析

本文实验数据来自于 HoneyNet[6]，采用文献[7]提出的态势评估方法来计算 2000 年 10 月 1 日到 2001 年 1 月 31 日的网络安全态势值，得到 123 个实验数据。为避免原始数据跨度大对 RBF 网络训练产生负面影响，对所得态势值均作极值标准化处理。

对于原始数据 $X=(x_1,x_2,\cdots,x_n)$，$x \in R$，运用极值标准化公式：

$$x_i' = (x_i - \min(x))/(\max(X) - \min(X)) \tag{8}$$

得到的标准化数据如图 2 所示。

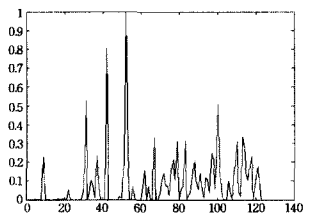

图 2 网络安全态势标准化值

前 110 个数据作为网络训练样本，余下的 13 个数据为测试样本。根据对 HoneyNet 数据集的统计分析，得出一次完整的网络攻击所用的时间为 $t \leqslant 3$ 天。即攻击周期大约为 3 天，因此，输入向量维数应大于等于 3，本文选用一周的天数作为输入向量维数，设定 RBF 神经网络预测模型的输入向量维数为 7，输出向量维数为 1，即根据神经网络输入向量预测未来一天的网络安全态势值。应用 HHGA 优化 RBF 神经网络，得到隐层节点数为 15。网络安全态势预测模型为 7-15-1 的 RBF 神经网络。

利用训练好的 RBF 神经网络对数据集中最后的 13 天的态势值进行预测。同时为了验证本文算法预测精度的先进性，采用已有的网络安全态势预测算法 RBF 神经网络[8]、GABPNN[9]进行相同的实验，得到如图 3 所示的结果。

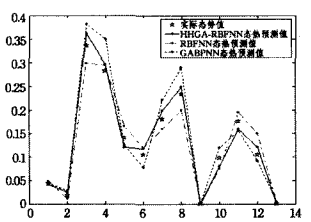

图 3 HHGA-RBFNN、RBFNN 和 GABPNN 态势预测值

从图 3 中可以看出，HHGA-RBFNN 模型预测结果优于其它两种方法，下面进一步进行定量的对比。

首先采用绝对误差指标 $e=|y_i'-y_i|$ 对 3 种方法的 13 个预测值分别进行度量，所得结果如图 4 所示。

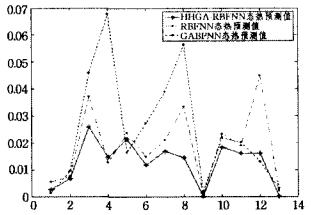

图 4 绝对误差指标

从图 4 中可以看到，HHGA-RBFNN 有个别预测值的绝对误差略大于对应时间点的 RBFNN 或者 GABPNN 预测值，但从误差值大小和整体个数上来看，HHGA-RBFNN 态势预测模型的预测精度仍明显地优于另两个预测模型。

接着采用平均相对误差(MAPE)和均方根误差(RMSE)两项性能指标来定量评价预测模型的精确度。显然，MAPE 和 RMSE 值越小，则预测模型的精度越高。

平均相对误差：

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i'-y_i|}{y_i} \tag{9}$$

均方根误差：

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i'-y_i)^2} \tag{10}$$

式中，y_i' 为预测值，y_i 为实际值，n 为样本数。

表 1 HHGA-RBFNN，RBFNN 和 GABPNN 预测误差对比

	HHGA-RBFNN	RBFNN	GABPNN
MAPE	0.1125	0.5825	1.1796
RMSE	0.0149	0.0230	0.1164

从表 1 中可以看出，HHGA-RBFNN 模型对于网络安全态势的预测，所得预测值的平均相对误差(MAPE)和均方根误差(RMSE)均小于 RBFNN 和 GABPNN。

结束语 对网络安全态势预测的研究有利于网络管理者获得网络安全态势，更快地采取相应的安全措施。本文提出了基于 HHGA-RBF 神经网络的网络安全态势预测方法，其利用 HHGA 的全局搜索能力来更好地优化 RBF 神经网络的结构和参数，然后使用训练好的 RBF 网络构建一个预测模型来预测未来网络安全态势。实验结果表明此方法能够更好地预测未来网络安全态势。

参 考 文 献

[1] Bass T. Multi-sensor Data Fusion for Next Generation Distributed Intrusion Detection Systems[C]// 1999 IRIS National Symposium on Sensor and Data Fusion. May 1999:24-27

[2] Bass T. Intrusion detection systems & multisensor data fusion: Creating Cyberspace Situational Awareness[J]// Communications of the ACM, 2000, 43(4):99-105

[3] 韦勇，连一峰，冯登国. 基于信息融合的网络安全态势评估模型[J]. 计算机研究与发展，2009，46(3)：353-362

[4] Ferreira P M, RUuano A E, Fonseca C M. Genetic assisted selection of RBF model structures for greenhouse inside air temperature prediction[C]// IEEE Conference on Control Applications Proceedings. 2003:576-581

(下转第 75 页)

2)簇内密钥更新的具体方法

第1步　$\forall V_i, i\in[1,N]$按照密钥建立协议中的方法根据其新私钥计算 P_i'；

第2步　$\forall H_j\rightarrow V_i, V_i\in C(i,N_i,H_j)$：$E(P_H', P_H)$和标识 ID；

第3步　$\forall V_i\rightarrow H_j, V_i\in C(i,N_i,H_j)$：执行 $D(P_H', P_H)$后，计算并传递 $E(E(P_i', P_i), P_H')$和标识 ID；

第4步　$\forall H_j$ 分别用 P_H'，P_i 解密获得 P_i'，并更新 V_i 对应密钥列表项的值 P_i 为 P_i'，且 $V_i\in C(i,N_i,H_j)$。

新周期起始时刻各簇首节点通过新密钥参数计算相互之间通信的对等密钥 K_{ij}'，完成簇首密钥更新。

(2) 簇首节点更新

当有新的簇首节点加入时，需先按照式(5)和式(6)计算新簇头节点的密钥份额和簇间密钥，然后按照集中式密钥建立方法建立该簇首节点所在簇的簇内密钥。由于簇间相互独立，因此其余簇内节点不受影响。簇首节点退出时，按照簇管理方法必然会生成新的簇首节点，新簇首节点生成后按照簇首节点加入的方法更新密钥。

(3) 簇内节点更新

小卫星节点作为簇内节点加入和退出网络而引起的密钥更新过程比较简单，因为其不会影响簇首节点的密钥份额，也不会影响其他节点的密钥计算。这种情况的密钥更新分别对应于其所在簇簇首节点密钥列表中一条记录的增加和删除。

4　协议分析

4.1　安全性分析

多种文献表明，基于 ECC 密钥管理协议建立在椭圆曲线离散对数难题之上具有较高的安全性。本文在文献[10]的基础上改进了周期性密钥更新方式，采用在上一周期结束前通过该周期的会话密钥传递下一周期密钥信息的方式，进一步提高了安全性。

4.2　性能分析

因为卫星网络信息传输过程中传播距离远，时延较长，所以总时延主要取决于消息传输时延，也即消息传播经历的跳数。各簇首节点到其簇内各个节点消息传播的距离，以跳数 H 来表示：

$$H=\sum_{i=1}^{N} hop(v_i) \tag{7}$$

式中，N 为节点总数，hop 函数用于求节点 V_i 到其所在簇簇首的跳数。通过簇首节点到其簇内节点的跳数之和可分析密钥更新速度。

以 Iridium 为例将 LEO 骨干卫星网络划分为 6 个簇，根据式(7)分别计算文献[9]的跳数 H_1 和本文方法所得的跳数 H_2：

$H_1=(1+2+3+4+5)*2*6=180$

$H_2=(1*4+2*5+3*2)*4+(1*4+2*4)*2=104$

二者都是将 Iridium 网络划分为 6 个簇，显然本文方法进一步降低了密钥更新的平均时延，更新速度更快。

结束语　卫星网络自身的特点使其比传统网络的安全问题更加复杂，而密钥管理是实现卫星网络安全通信的基本保障。本文分析了卫星网络密钥管理的安全性需求，对 LEO 卫星网络密钥管理协议进行了改进；提出了分布式和集中式相结合的密钥建立方法，既实现了密钥管理不依赖于地面 CA 的效果，也降低了广播密钥信息的额外通信量；引入了更合理的卫星网络分簇方法，提高了密钥更新速度；完善了密钥更新方法，增强了密钥管理协议的安全性。

参 考 文 献

[1] 张焕国，王张宜. 密码学引论(第二版)[M]. 武汉：武汉大学出版社，2009

[2] 蹇波，郭永辉，罗长远，等. 基于 ECC 的无线传感器网络密钥管理协议[J]. 计算机工程，2010，36(3)：142-144

[3] 宗家然，王阳阳. 椭圆曲线在 WSN 密钥管理中的应用[J]. 电脑知识与技术，2010，6(4)：828-830

[4] 李喆，李冬妮，王光兴. LEO/MEO 卫星网络中运用自组网思想的动态路由算法[J]. 通信学报，2005，26(5)：50-57

[5] 吴举，杜学绘，钱雁斌. 改进的空间网络密钥交换协议[J]. 计算机工程，2009，35(18)：113-115

[6] 蔡晓秋，张建中. 基于椭圆曲线的多银行电子现金系统[J]. 计算机应用研究，2007(5)：14-15

[7] Kumar K，Begum J N，Sumathy V. A Novel Approach towards Cost Effective Region-based Group Key Agreement Protocol for Ad-hoc Networks Using Elliptic Curve Cryptography[J]. Int. J. Communications，Network and System Sciences，2010，3：369-379

[8] 纪豫宣，马恒太，郑刚，等. 卫星网络密钥管理模型设计与仿真[J]. 系统仿真学报，2009，21(13)：4153-4158

[9] 张志强，张永健，王宇，等. 低轨卫星网络中基于轨道分簇的密钥更新算法[J]. 电子与信息学报，2010，32(3)：687-692

[10] 闫少阁，苏锦海. 基于椭圆曲线的分布式密钥建立协议[J]. 计算机工程与应用，2009，45(1)：113-115

[11] 李喆，刘军. 卫星网络安全路由研究[J]. 通信学报，2006，27(8)：113-119

[12] 李冬妮，张大坤. 卫星网络中多因素均衡的分簇算法[J]. 北京理工大学学报，2008，28(1)：62-65

(上接第 72 页)

[5] 郑丕谔，马艳华. RBF 神经网络的递阶遗传训练新方法[J]. 控制与决策，2000，15(2)：165-168

[6] Honeynet Project. Know Your Enemy：Statistics. 2001 [EB/OL]. http://old.honeynet.org/papers/stats/

[7] 陈秀真，郑庆华，管晓宏，等. 层次化网络安全威胁态势量化评估方法[J]. 软件学报，2006，17(4)：885-897

[8] 任伟，蒋兴浩，孙锬锋. 基于 RBF 神经网络的网络安全态势预测方法[J]. 计算机工程与应用，2006，31：136-138，144

[9] Wang Hui-qiang，Lai Ji-bao，Wu Xiao. A Quantitative Forecast Method of Network-Secuity-Situation-Based on the BP Neural-Network with Genetic Algorithm[C] // Second International Multisymposium on Computer and Computational Sciences. 2007，65：374-380