基于免疫算法和神经网络的新型抗体网络

陈 科，许家珀，程永新

(电子科技大学应用数学学院 成都 610054)

【摘要】构建了基于免疫算法和神经网络的新型抗体网络入侵检测系统，系统与网络入侵检测功能相结合，应用于大型网络的入侵检测任务，具有良好的可扩充性；重点讨论了新型抗体网络原理，引进BP神经网络自学习能力，对已有的抗体网络模型进行改进；通过对网络数据集的测试表明，该算法相对于传统抗体网络，其检测效率得到了明显的改善。

关 键 词 入侵检测算法； 免疫算法； 神经网络； 抗体网络
中图分类号 TP389.1 文献标识码 A

A Novel Antibody Network Base on Immune Algorithm and Neural Network

CHEN Ke，XU Jia-yi，CHENG Yong-xin

(School of Applied Mathematics, Univ. of Electron. Sci. & Tech. of China Chengdu 610054)

Abstract This paper designs an antibody network (ABNET) based on immune algorithm and neural network used for Intrusion Detection System. The ABNET system combines network-based intrusion detection functions. It can be used to protect large area network and has relatively good expansibility. This paper also discusses the theory of the ABNET. The algorithm has been tested on a network data set. The result shows that it had much better performance than traditional ABNET.

Key words intrusion detection algorithm; immune algorithm; neural network; antibody network

入侵检测系统(Intrusion Detection System，IDS)是对计算机网络系统中的入侵行为进行自动检测的系统，可检测未授权用户，以及误用和滥用系统的内部和外部用户。传统的入侵检测方法是从定义入侵模式库开始，把采样并标准化的数据与模式库里的数据进行匹配检测，缺乏多样性和灵活性，尤其是对未知入侵行为显得无能为力。入侵检测大体可以分为基于主机的入侵检测和基于网络的入侵检测。基于主机的入侵检测主要利用系统日志跟踪入侵，而基于网络的入侵检测主要利用网络数据包分析入侵。本文利用免疫算法[1]神经网络生物技术，构造一种新型的、具有自适应能力的抗体网络(Antibody Network，ABNET)。

1 免疫算法和神经网络技术

近年来，受生物系统启发而设计的智能算法越来越受到人们的重视。免疫算法、神经网络、遗传算法并称为三大仿生算法。将免疫算法与神经网络技术结合，既弥补了神经网络收敛速度慢，容易陷入局部最优解的弊端，又增强了免疫算法的分布性和自适应性[2]。

1.1 免疫算法

免疫算法中有否定选择、高频变异和克隆选择三大关键技术。否定选择是区别自我/非我的方式，通过自体耐受，删掉与自身细胞产生应答的免疫细胞；高频变异使系统保持多样性，可对未知的抗原产生一定的应答；克隆选择根据抗原和抗体的亲和力，促进和抑制抗体的繁殖。

1.2 人工神经网络

人工神经网络(Artificial Neural Networks，ANN)根据外界的输入或刺激来调整神经网络的参数，以达到某种期望的自适应过程。本文拟采用BP神经网络作为入侵检测的基本分析工具。

收稿日期：2005 - 11 - 21
基金项目：四川省科技厅基础项目(04JY029-017-1)
作者简介：陈 科(1981 -)，男，硕士，主要从事信息技术与计算智能方面的研究.

BP神经网络为多层前馈网络，其权值学习常采用误差逆传递学习算法(Error Back Propagation，BP)。将采用这一学习算法进行训练的多层前向网络简称为BP网络。其训练模式可描述为：训练样本输入网络，计算出实际输出和理想输出之间的误差E，通过误差来调整权重，记忆训练，最终收敛到一个局部或是全局的最优解。

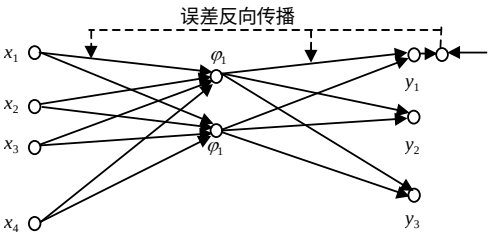

图1　BP网络的基本模型

BP算法是一种比较成熟的有指导的训练方法，它包含输入层X、隐含层φ和输出层Y。同层节点之间不连接，输入信号通过输入层节点，依次传递给隐含层节点，与权值W以及偏移值b作用，传递到输出层节点，每一层节点的输出作为下一层的输入。网络训练的目标是使误差函数E最小。E的定义如下：

$$E=\frac{1}{2}\sum_{p}\sum_{i}(t_{pi}-y_{pi})^2 \tag{1}$$

式中　E为网络输出误差；p代表p个训练样本；用实际输出y_{pi}和期望输出t_{pi}的误差来修改其连接权值和阈值，利用最快梯度下降法，对每个权值进行修正，有：

$$W(k+1)=W(k)+\eta D(k) \tag{2}$$

$$D(k)=\frac{-\partial E}{\partial W(k)} \tag{3}$$

式中　$W(k)$为k时刻的权值；η 为学习率。η 过小会造成迭代次数过多，降低神经网络的学习效率；过大会造成网络的区域波动，有可能徘徊在几个局部极小值点上。通过对式(3)进行推导，可以得出：

$$\Delta W_{ji}=-\eta\frac{\partial E}{\partial \sigma_i}x_{ji} \tag{4}$$

式中　σ_i 代表第i层上神经元接收的输入总和。至此，可以根据误差和前一层的输入计算出新的权值。

用三层的传统BP网络可以任意逼近任何连续函数，但由于采用了非线性的梯度优化方法，存在局部极小值，不能收敛到全局最优解上。如果η 的选取不当，会使收敛速度变得异常缓慢，甚至发散。采用神经网络来检测入侵，准确率低，不能满足实时检测的需要。

2　基于免疫算法和神经网络的新型抗体网络

入侵检测网络模型是入侵检测系统的核心部分。神经网络检测技术具有很强的非线性映射能力和学习能力，成为异常检测技术的研究热点。神经网络结合免疫算法的检测技术，提高了入侵检测的性能和速度。

2.1　基于免疫算法和神经网络的新型抗体网络结构

2.1.1　新型抗体网络结构模型

如图2所示，抗体(A_{Ab})表示神经网络中输入神经元到输出神经元的权向量w_k，而抗原(A_{Ag})表示神经网络中的输入样本,代表某个网络数据包。在基于网络的入侵检测中，均以二进制串表示。S代表自体库集合。在本文中，权值以及输入样本和输出单元都用二进制表示，即仅用0和1表示。

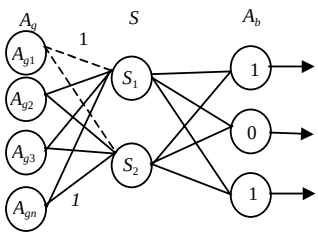

图2　Abnet网络模型

沿用传统的抗体网络[3]特征，规定ξ_j 表示网络中抗体j的抗原浓度，即抗体j所能识别的抗原个数。一个抗体细胞k与某种抗原细胞的亲和力，由权向量与该抗体的汉明距离(Hamming)决定，可由式(5)得到与抗原A_{Ag}可能性的最大状态序列：

$$\delta=\arg\max\left\|A_{\mathrm{Ag}}-A_{\mathrm{Ab}_k}\right\| \tag{5}$$

某种抗体的亲和力越大，说明这种抗体能够对抗原进行较好的应答、保留；亲和力小的抗体，可通过剪枝的方式，从网络中删掉。

2.1.2　新型抗体网络学习步骤

整个网络的竞争学习步骤分为两部分：

(1) 从抗体库里根据概率密度P_{Ab}选择出一个抗体，并和自体库S进行运算。设定阈值ε，当

$$\arg\max\left\|A_{\mathrm{Ab}}-S_i\right\|\le\varepsilon \qquad i=1,2,\cdots,n \tag{6}$$

抗体经过了自体耐受，变为成熟的检测器，且不与自体发生免疫应答，是合格的检测器[4]；如果超过阈值ε，则从网络中删除此抗体节点。如图2所示，抗体A_{Ab_1} [101]和自体集S中的S_1[010]产生了免疫应答。

(2) 经过自体耐受的抗体和抗原A_{Ag_k}的作用，在抗体内部设定一个ξ_i,一旦抗原的权向量和抗体的汉明距离超过阈值ε，ξ_i增加1。如果ξ_i长时间等于0，抗体被从网络中删去；当ξ_i增加到一个常值M的时候，不再继续增加。启动一个计时器，按一定的间隔时间t递减ξ_i，可避免长时间未产生应答的抗体继续残留在抗体库中。

2.1.3 自体库的建立

随机获取一输入向量抗原X，按照式(5)产生与X最匹配的中心序列k，把此序列k添加到抗体群中，如现有抗原X=[01010110]，则根据具有最大Hamming距离法则，生成抗体k=[10101001]，把k加入新的抗体群中。

2.1.4 抗体的克隆变异

为了保持抗体的多样性和自适应性，必须对抗体进行变异，以适应新的抗原变化，传统的克隆选择算法利用单点或多点变异，破坏了抗体的整体性。利用BP神经网络的自学习功能，可以使抗体的变异收敛于一个与抗原匹配的最优解，步骤如下：

步骤 1 从抗体库A_{Ab}找出一个和输入抗原最匹配的抗体A_{Ab_k}，如果超过阈值ε，则输入下一个抗原；如果未超过，则进行步骤2。

步骤 2 利用式(1)的变形，计算出A_{Ab_k}和A_{Ag}的误差函数，定义为：

$$E=\frac{1}{2}\sum_i\sum_j\left|A_{\mathrm{Ab}_{ki}}-A_{\mathrm{Ag}_j}\right|^2 \tag{7}$$

步骤 3 对于和抗原Ag有最大E的抗体权A_{bk}进行更新，由于采用二进制编码，重新设定学习率η以及一个门限函数F，有：

$$\beta_i=\Delta A_{\mathrm{Ab}_k}=-\eta\frac{\partial E}{\partial A_{\mathrm{Ab}_k}}+\alpha A_{\mathrm{Ab}_k} \tag{8}$$

式中 β为由最快梯度下降法得出的抗体权向量改变值[5]；α为影响矩阵因子。

$$F_i=\begin{cases}1 & \left\|\beta_i\right\|\ge T\\ 0 & \left\|\beta_i\right\|\le T\end{cases} \tag{9}$$

T为预设的，最终抗体单位权向量改变由F函数决定，

$$A_{\mathrm{Ab}_k}(t+1)=A_{\mathrm{Ab}_k}(t)\oplus F \tag{10}$$

式中 $\oplus$为二进制的加法运算，即1+1=0，1+0=1，0+0=0。

步骤 4 更新抗体群。

2.2 对抗体网络的测试

为测试改进后的网络在入侵检测应用中的效果，采用具有30万条数据记录的测试数据集，每条数据包括网络数据包的包头信息、网络连接信息和数据信息等。数据包的96位二进制代码中，前32位为源IP地址；第32～64位为目标IP地址；第64～96位为一些数据信息。每条数据记录被标记为异常或正常。首先用前1万条数据记录进行网络的训练，训练结果的参数见表1。

表**1** 抗体网络的训练

自体库	单次变异率/(%)	单抗体扰动率/(%)	用时/s
100	23.45	29.34%	632
200	21.93	19.33%	1 965
500	15.25	13.92%	12 200

表**2** 三种算法网络模型的对比

网络名称	误报率/(%)	漏报率/(%)	用时/s
免疫算法模型	1.83	0.36	6 003
传统抗体模型	1.15	0.55	7 264
新型抗体模型	0.98	0.23	7 940

(下转第840页)

3 结 束 语

通过理论分析和实验证实，利用本文提出的基于傅立叶技术的快速预测方法对基因组序列的编码区进行预测可取得良好的效果。该方法的显著优点是运算速度比利用FFT的方法快，容易应用，不需要基因组序列的任何先验知识；并且可同时实现基因的预测和定位。预测出编码区的大概位置，为进一步用实验方法精确定位编码区打下基础。正如文献[7]所指出的，通常难以用一种方法将各种生物DNA序列的编码区预测问题全部解决，需要多种方法融合，才能达到准确预测和定位编码区的目的。面对世界范围内急剧增长的生物序列信息，相信对简便、快速、准确和适应性强的编码区预测方法的需求将会越来越大。

参 考 文 献

[1] Dodin G, Vandergheynst P, Levoir P, et al. Fourier wavelet transform analysis a tool for visualizing regular patterns in DNA sequences[J]. J Theo. Biol., 2000, 206: 323-326.
[2] Berger J A, Mitra S K, Carli M, et al. Visualization and analysis of DNA sequences using DNA walks[J]. Journal of the Franklin Institute, 2004, 341: 37-53.
[3] Buldyrev S V, Goldberger A L, Havlin S, et al. Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis[J]. Physical Review, 1995, 51(5): 5084-5091.
[4] Tiwari S, Ramachandran S, Bhattacharya A, et al. Prediction of probable genes by Fourier analysis of genomic sequences[J]. CABIOS, 1997, 13(3): 263-270.
[5] Shepherd S J, Kay N D, Van Eetvelt P. An efficient algorithm for computing genetic spectra[C]// Oxford Bioinformatics Forum, Oxford, UK, 2003: 11-15.
[6] Chechetkin V R, Turygin A Y J. Study of correlation in DNA sequences[J]. Theo. Biol., 1996, 178: 205-217.
[7] Fickett J W. The gene identification problem: An overview for developers[J]. Comp. Chem., 1996, 20: 103-118.

编 辑 孙晓丹

(上接第806页)

由表1可以得出，自体库选过小，会造成单抗体的高扰动率，频繁更新抗体群，缺乏抗体的多样性，覆盖范围减小；而自体库选过大，会造成训练网络的时间急剧增多。本文采用200条为自体库大小，然后通过新型网络模型对这30万条数据记录进行检测，并与单免疫算法模型和传统的抗体网络模型进行对比。虽然此网络模型在时间上略逊于其他两种已知算法模型，但在准确率上却有明显的提高，如表2所示。

3 结 束 语

本文构建了基于免疫算法和神经网络的新型抗体网络，针对传统BP神经网络在入侵检测应用中学习性能的不足，引入免疫算法原理，对已有的抗体网络进行改造。通过对网络模拟数据集的测试，相对于单免疫网络和传统的抗体网络，检测效率和学习性能有明显的提高。

参 考 文 献

[1] 赵俊忠, 黄厚宽. 免疫机制在计算机网络入侵检测中的应用[J]. 计算机研究与发展, 2003, 40(9): 1293-1299.
[2] 吴 知, 许家玷. 免疫原理在多Agent入侵检测系统中的应用[J]. 电子科技大学学报, 2005, 6(3): 381-384.
[3] D Castro L N, Von Zuben F J, de Deus J G A. The construction of a boolean competitive neural network using Ideas from Immunology[J]. Neurocomputing, 2003, (50c): 51-85.
[4] Kim J, Peter J B. Towards an artificial immune system for network intrusion detection:An investigation of dynamic clone selection [J/OL]. IEEE2002, 0-7803-7282-4/02, 2005-10-21.
[5] D`haeseleer P, Forrest S. An immunological approach to change detection: algorithms, analysis and implications[C]//IEEE Symposium on Research in Security and Privacy, Oakland, 1996.

编 辑 熊思亮