* 1

基于反向传播神经网络的入侵检测算法

韩仲祥，段　弢，董淑福，张　锐，陶晓燕
(空军工程大学　电讯工程学院,陕西　西安　710077)

摘　要:讨论了多层神经网络算法缺陷,提出了一种基于改进反向传播(Back Propagation,BP)的快速入侵检测算法——IBP 算法:在 BP 算法中的梯度下降算式中,加入一个动量项 $\alpha[\omega(t)-\omega(t-1)]$,改善计算神经元 j 到神经元 i 的级联权值;采用学习速率可变的策略;算法训练网络时采用批处理的样本输入方式。改进后的算法选取较大的学习速率 $\eta=0.5$ 和 $\eta=0.65$,并采用 3 层神经网络的结构,输入、输出样本是 16 维和 15 维,各进行 100 次独立仿真实验,结果证明可加快算法收敛速度,另外,仿真实验还证明:改进后的算法对初始权值的敏感性、网络所表现出的稳定性等都比传统算法性能优越。

关键词: BP 算法;收敛速度;入侵检测
DOI:10.3969/j.issn.1009-3516.2009.04.012
中图分类号: TN915　　**文献标识码**:A　　**文章编号**:1009-3516(2009)04-0053-05

目前已经有很多方法用于入侵检测,例如数据挖掘的方法[1]、支持向量机的方法[]、计算免疫[3]和遗传算法[4]方法等,有采用神经网络方法的诸如 SOM(自组织映射)[5]网络和 RBF[6]网络等。BP 神经网络技术应用于入侵检测系统既有优势,也有缺陷和不足,如执行速度比较慢的问题等。基本的 BP 传播算法(也称为最速下降反传算法 Steepest Descent Back Propagation,SDBP)。收敛速度慢,网络易陷于局部极小,为了克服这些不足,出现了许多改进算法,比较成功的算法有:共轭梯度算法和 Levenberg Marquardt 算法(牛顿法的变形)[7]等。传统的入侵检测产品在技术上难以满足入侵检测所需要的实时性、适用性、可用性、可靠性和准确性等方面的需求。而神经网络在概念和处理方法上都很适合入侵检测系统的要求[8-10]。

1　多层前馈神经网络

1.1　网络结构

神经网络通常包含许多层,如图 1 所示的 3 层网络,其中 $\boldsymbol{u}$、$\boldsymbol{y}$ 是网络的输入、输出向量,每一神经元用一节点来表示。这种网络特点是只有前后相邻 2 层之间神经元相互联接,各种神经元之间没有反馈。每个神经元可以从前一层接受多个输入,并用同一个输出送给下一层的各神经元。

3 层神经神经网络分为输入层、隐含层和输出层。在前向网络中有计算功能的节点被称为计算单元,而输入节点无计算功能。由于用 BP 学习算法,所以称之为 BP 神经网络。

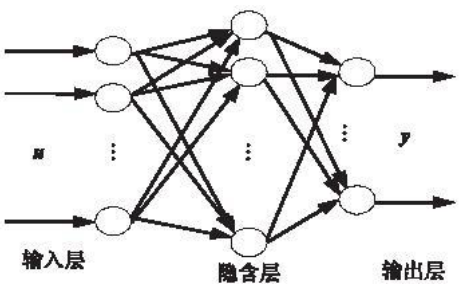

图 1　3 层神经网络
Fig.1　Three layer of neural network

1.2　梯度下降算法

* **收稿日期**:2009-02-19
基金项目:陕西省自然科学基础研究计划资助项目(DG070302);空军工程大学电讯工程学院博士启动基金资助项目
作者简介:韩仲祥(1971-),男,山东莒南人,讲师,博士,主要从事计算机网络安全与管理研究.
E-mail:zhongxianghan@sina.com

梯度下降算法是一种对某个准则函数的迭代寻优算法。设 $J(\boldsymbol{a})$是准则函数,$\boldsymbol{a}$为一向量。$\nabla J(\boldsymbol{a})$是 $J(\boldsymbol{a})$在点 $\boldsymbol{a}_k$ 的梯度,为一向量,其方向为 $J(\boldsymbol{a})$增长最快的方向,负梯度方向,则是 $J(\boldsymbol{a})$减小最快的方向。因此。若求某函数的极大值,沿梯度方向走,可以最快的速度到达极大点;反之,沿负梯度方向走,可最快地到达极小点。梯度下降法是求函数极值的迭代算法,但是梯度下降法的迭代在向极小值靠近的过程中,走的是曲折的路径,即所谓的"锯齿现象",这是该算法收敛速度慢的原因。

2　BP 学习算法

2.1　BP 学习算法原理

已知网络的输入/输出样本,即导师信号。正向传播是输入信号从输入层经隐含层传向输出层,若输出层得到了期望的输出,则学习算法结束;否则,转至反向传播。反向传播就是将误差信号(样本期望输出与网络实际输出之差)按原连接通路反向计算,由梯度下降法调整各层神经元的权值和阈值,使误差减小。

算法步骤:

1)设置初始权值 $W(0)$为较小的随机非零值。

2)给定输入/输出样本对,计算网络的输出;设 p 组样本输入、期望输出分别为 $\boldsymbol{u}_p=(u_{1p},u_{2p},\cdots,u_{np})$,$\boldsymbol{d}_p=(d_{1p},d_{2p},\cdots,d_{kp})$,$p=1,2,\cdots,L$,$p$ 为样本对数,n 为第 n 个输入结点,k 为第 k 个输出结点。

输出结点 i 在第 p 组样本输入时,实际输出为:

$$y_{ip}=f[x_{ip}(t)]=f[\sum_j w_{ij}(t)I_{jp}] \tag{1}$$

式中:I_{jp} 为在第 p 组样本输入时,结点 i 的第 j 个输入;t 为权值调整次数。$f(\cdot)$取可微分的S型激活函数式,如:$f(x)=\dfrac{1}{1+\mathrm{e}^{-x}}$,可由输入层经隐含层至输出层,求得网络输出层结点的输出。

3)计算网络的总目标函数 J。

设 E_p 为第 p 组样本输入时网络的目标函数,取 L_2 范数,则:

$$E_p(t)=\frac{1}{2}\|\boldsymbol{d}_p-y_p(t)\|_2^2=\frac{1}{2}\sum_k[d_{kp}-y_{kp}(t)]^2=\frac{1}{2}\sum_k \mathrm{e}_{kp}^2(t) \tag{2}$$

式中:$y_{kp}(t)$为在第 p 组样本输入时,经 t 次权值调整网络的输出;k 是输出层第 k 个结点。

网络的总目标函数为:

$$J(t)=\sum_p E_p(t) \tag{3}$$

作为对网络学习状况的评价。

4)判别。若 $J(t)\leqslant\varepsilon,\varepsilon>0$,则算法结束;否则,至步骤5)。

5)反向传播算法。由输出层,依据 J,按"梯度下降法"反向计算,逐层调整权值。神经元 j 到神经元 i 的级联权 $t+1$ 次调整算式:

$$w_{ij}(t+1)=w_{ij}(t)-\eta\frac{\partial J(t)}{\partial w_{ij}(t)}=w_{ij}(t)-\eta\sum_p\frac{\partial E_p(t)}{\partial w_{ij}(t)}=w_{ij}(t)+\Delta w_{ij}(t) \tag{4}$$

式中 η为步长,在此称学习算子或学习速率。

2.2　BP 学习算法缺陷

BP 算法有以下不足:①由于是非线性优化,不可避免地会存在局部极小问题;②学习算法的收敛速度慢,且收敛速度与初始权值有关;③网络的结构设计,即隐含层及节点数的选择,尚无理论指导;④新加入的样本会影响已学好的样本。

3　BP 学习算法改进(IBP)

输出层神经元输出值表示可以连接的安全状态,隐含层数由实验确定,采用的学习算法为在 BP 算法基础上改进后得到的,称之为 IBP 算法。

3.1　IBP 算法

为了克服传统 BP 学习算法收敛速率慢等不足,下面从几个方面对 BP 算法进行改进。

1)学习速率 η的选择很重要,η大则收敛快,但过大则可能引起不稳定(振荡);η小可避免不稳定,但收

敛速率就慢了。解决这一矛盾的最简单办法就是在式(4)中加入一个动量项,即:

$$w_{ij}(t+1)=w_{ij}(t)-\eta\sum_{p}\frac{\partial E_P(t)}{\partial w_{ij}(t)}+\alpha[\omega(t)-\omega(t-1)] \tag{5}$$

即
$$\Delta w_{ij}(t)=-\eta\sum_{p}\frac{\partial E_P(t)}{\partial w_{ij}(t)}+\alpha\Delta w_{ij}(t-1)$$

式中的第 2 项即为动量项,$0<\alpha<1$ 其作用简单分析如下:

当顺序加入训练样本时,上式可写成以 n 为变量的时间序列,n 由 0 到 t,因此上式可看作是 Δw_q 的一阶差分方程,对 $\Delta w_{ij}(t)$ 求解可得:

$$\Delta w_{ij}(t)=\eta\sum_{n=0}^{t}\alpha^{t-1}\delta(n)y_j(n)=-\eta\sum_{n=0}^{t}\alpha^{t-1}\sum_{p}\frac{\partial E_p(n)}{\partial w_{ij}(n)} \tag{6}$$

当本次的 $\sum_{p}\frac{\partial E_p(n)}{\partial w_{ij}(n)}$ 与前一次同号时,其加权求和值加大,使 $\Delta w_{ij}(t)$ 较大,结果在稳定调节时加速了 w 的调节速率;当与前次符号相反时说明有一定振荡,此时加权求和结果使 $\Delta w_{ij}(t)$ 减小,起到稳定作用。

由此可见,通过加入动量项,滤掉高频变量,使权值空间的误差表面平滑,可以先取较大的 η 值来加快收敛速率而不致振荡。

2)采用学习速率可变策略,根据收敛性要求对 η 加以调整,以提高收敛速率。若本次迭代总误差 $J(t)>J(t-1)$,则这次迭代无效,并恢复以前的步长,减小步长重新迭代,此时 $\eta(t+l)=\eta(t-1)-\eta(t-l)/t$;反之,本次迭代有效,增大学习步长进行下一次迭代,此时 $\eta(t+1)=\eta(t)+\eta(t)/t$。

3)用 BP 算法训练网络时有 2 种方式,一种是每输入一个样本修改一次权值;另一种是批处理方式,即等待组成一个训练周期的全部样本都依次输入后,计算总的误差,修改权值。当训练样本数目不是很大时,后一种方式能加快收敛速率。因而,在提出的改进算法中,文中选择后一种方式。

改进后的BP算法步骤:① 设置初始权值 $W(0)$ 为较小的随机非零值,并输入其它参数与网络结构。② 给定输入/输出样本对,计算网络的输出。见式(1)。③ 计算网络的总目标函数 $J(t)$。$J(t)$ 的计算见式(2)、式(3)。若计算所得的 $J(t)>J(t-1)$,则令 $\eta(t+1)=\eta(t-1)-\eta(t-1)/t$,至步骤 ⑤;否则,令 $\eta(t+1)=\eta(t)+\eta(t)/t$。④ 判别。若 $J(t)\leqslant\varepsilon$(ε是预先确定的,$\varepsilon>0$)则算法结束;否则,至步骤 ⑤。⑤ 反向传播计算。由输出层,依据 J,按"梯度下降法"反向计算,依据下式逐层调整权值。

$$w_{ij}(t+1)=w_{ij}(t)-\eta(t+1)\sum_{p}\frac{\partial E_p(t)}{\partial w_{ij}(t)}+\alpha[w(t)-w(t-1)] \tag{7}$$

式中 $\eta(t+1)$ 为可变学习速率。具体计算步骤同前述 BP 算法。

4 仿真及分析

4.1 入侵检测样本设定

本文选择了一些基于 Linux 的典型入侵事件,它们来自以下 3 种情况:①用户登陆时;②访问网络资源;③访问跨域资源。这里选择了 16 种典型事件(分别记为 1 号事件,2 号事件,……),而对于这些事件,共有 15 种可能发生的入侵(1 号入侵,2 号入侵,……),因此用 16 维的向量来表示输入,如果某事件出现,则相应的输入神经元的输入被置为 1,同时用 15 维的向量表示输出,如某入侵出现,则相应的输出神经元输出 1,如某个输入向量为"0011110000000000",则表示发生了事件 3、4、5、6;若此时对应的输出量为"001010000000000",则表示对应以上事件的入侵为 3、5。实验中选择了 50 个输入输出向量样本对。

4.2 仿真结果及分析

采用 3 层神经网络结构,由于输入输出样本分别是 16 维和 15 维,所以输入层和输出层分别有 16 个和 15 个神经元。而由于隐含层神经元数目 b 的选择无确定的法则,所以采用试值法测试隐含层节点数对网络性能的影响,另外,初始学习速率 η 的调整也只能采用试值法。

首先,使用 20 个隐含层神经元,取初始学习速率 $\eta=0.65$,并且在其它网路参数均相同的情况下,分别对 2 种不同的 BP 算法各进行了 100 次独立实验,传统 BP 算法有 16 次不收敛的情况,而改进的方法均收敛,其收敛的速度均比传统 BP 算法要快,而且对初始值不敏感。如图 3 所示,即为在基本初始条件均相同的情况下,两算法在其中一次运行后的情况。

图 3 和图 4 均采用类似的方法和过程各进行了 100 次独立实验。从中能够明显看出,改进方法收敛的速度均比传统 BP 算法要快,而且对初始权值不敏感。

在进行 12 000 次数的训练之后,用其中的一个样本对训练后的网络性能进行测试后发现,改进方法的实际输出结果也比传统 BP 算法要好。而在对一个样本加入比较小的噪声信号之后,传统 BP 算法有时甚至会出现错误的结果,但改进方法有比较强的抗噪声能力,在一定范围内的噪声信号的影响下,它仍然能输出误差允许范围内的结果,图 5 是加入方差为 0.1 的高斯白噪声后 2 种算法的运行结果。

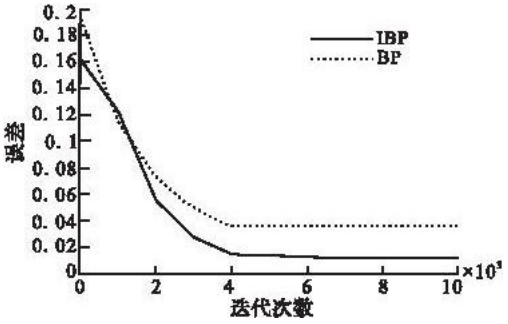

图 2　$b=20$, $\eta=0.65$ 的情况

Fig. 2　Simulation of $b=20$, $\eta=0.65$

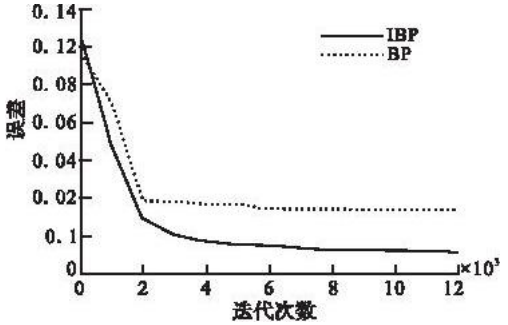

图 3　$b=20$, $\eta=0.5$ 的情况

Fig. 3　Simulation of $b=20$, $\eta=0.5$

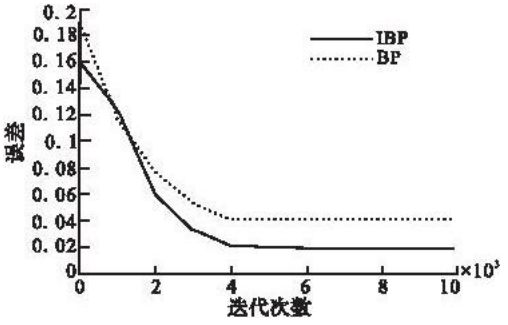

图 4　$b=15$, $\eta=0.5$ 的情况

Fig. 4　Simulation of $b=20$, $\eta=0.5$

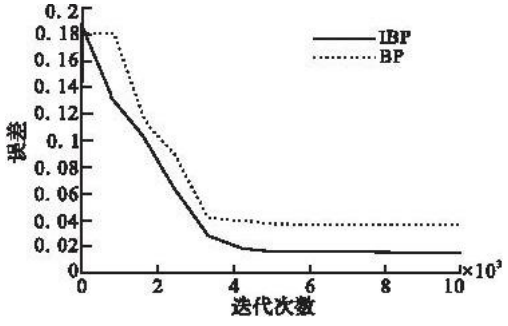

图 5　加入噪声后 2 种算法的迭代情况

. 5　Simulation of adding noises with two arithmetic

5　结束语

实验证明,改进方法效果比较理想。它表现出比传统算法收敛速度更快的优势,对初始权值的敏感性也优于传统 BP 算法,另外,它还表现出更好的稳定性和更强的抗噪声能力。这些优势使得改进方法可以更好地满足入侵检测技术所需要的实时性、可靠性和准确性等方面的要求。

参考文献:

[1] Stefanos M, Marvin C, Dan Z, et al. A Data Mining Analysis of RTID Alarms [J]. Computer Networks, 2000, 34(5): 571—577.

[2] Mukkamala S, Janoski G, Sung A. Intrusion Detection Using Support Vector Machine [C]//Proceedings of High Performance Computing Symposium—HPC. Washington DC: IEEE Computer Society, 2002: 1702—1707.

[3] 罗文坚. 面向入侵检测的人工免疫模型和算法研究[D]. 合肥:中国科学技术大学, 2003.
LUO Wenjian. Research on Artificial Immune Model and Algorithms [D]. Hefei: University of Science and Technology of China, 2003. (in Chinese)

[4] 张凤斌. 基于免疫遗传算法的入侵检测技术研究[D]. 哈尔滨:哈尔滨工程大学, 2005.
ZHANG Fengbin. Research of Intrusion Detection Based on Immunogenetic Algorithm[D]. Harbin: Harbin Engineering University, 2005. (in Chinese)

[5] Teuvo. Self—Organizing Maps [J]. Springer Series in Information Sciences, 2001, 30(2): 155—163.

[6] Bishop. Neural Networks for Pattern Recognition [M].Oxford:Clarendon Press,1995.
[7] Hagan M T, Demuth H B, Mark Beale. Neural Network Design [M]. Beijing:Beijing Mechanical Industry Press, 2002.
[8] Kantzavelou I, Patel A. An Attack Detection System for Secure Computer System—design of the ADS Information Systems Security: Facing the Information Society of 21st Century[M]. London:Chapman & Hall, 1997.
[9] Debar H, Dorizzi B. An Application of A Recurrent Network to An Intrusion Detection System[C]//Proceedings of the International Joint Conference on Neural Networks. New York: IJCNN,1992:468—474.
[10] Teresa F Lunt. IDES: An Intelligent System for Detecting Intruders[C]//Proceedings of the Symposium: Computer Security, Threat and Countermeasures. Rome:CERIAS,1990:528—534.

(编辑:徐楠楠)

An Intrusion Detection Algorithm Based on BP Neural Network

HAN Zhong—xiang, DUAN Tao, DONG Shu—fu, ZHANG Rui, TAO Xiao—yan
(Telecommunication Engineering Institute, Air Force Engineering University, Xi′an 710077,China)

Abstract:This paper discusses the limitation of multi—ply feedback neural network algorithm, presents a fast intrusion detection algorithm, i.e. IBP algorithm, based on the modified BP algorithm: adding a momentum in decreasing gradient formula to calculate the cascade weight value of neuron j to neuron i,adopting alterable learning rate strategy, choosing batch processing sample input while training the neural network. Bigger learning rates $\eta=0.5$ and $\eta=0.65$ are selected in the improved algorithm and the structure of three—ply neural network is adopted. The simples of input and output are of fifteen dimension and sixteen dimension. The simulation with computer shows that the modified algorithm is superior to the traditional algorithm in constringency speed, susceptiveness to the initial weight value, stabilization in network.
Key words:BP algorithm; constringency speed; intrusion detection

(上接第 52 页)

A Phase Spectrum Reservation Receiving Method of DS Signal

WANG Yong—min, GUO Jian—xin
(Telecommunication Engineering Institute, Air Force Engineering University, Xi′an 710077,China)

Abstract:In order to mitigate narrow—band interference in DS system, a phase spectrum reservation processing is proposed in DS signal receiving and an iterative algorithm is employed in the method to reduce computational complexity. The suppressing narrow—band interference principle is discussed and the performance loss in AWGN channel is analyzed. A method of calculating the performance loss is presented. The numeric simulation results of receiving DS signal by phase reservation processing are presented for variety of interference conditions. These results demonstrate that the narrow band interference can be effectively suppressed and no obvious performance loss is caused after the phase spectrum reservation processing
Key words:DS spread spectrum; narrow—band interference suppression; phase spectrum; DFT; iterative algorithm