基于李雅普诺夫函数的 BP 神经网络算法的收敛性分析

张茂元，卢正鼎

（华中科技大学 计算机科学与技术学院，湖北 武汉 430074）

摘　要：针对前馈神经网络应时变输入的自学习机制，采用李雅普诺夫函数来分析权值的收敛性，从而揭示 BP 神经网络算法朝最小误差方向调整权值的内在因素，并在分析单参数 BP 算法收敛性基础上，提出单参数变调整法则的离散型 BP 神经网络算法.

关 键 词：李雅普诺夫函数；前馈神经网络；BP 算法；单参数；变调整法则.

中图分类号：TP183　　**文献标识码**：A　　**文 章 编 号**：1000-1220(2004)01-0093-03

Lyapunov-based Analyse of Weights' Convergence on Backpropagation Neural Networks Algorithm

ZHANG Mao-yuan，LU Zheng-ding

(*Department of Computer Science and Technology*，*HuaZhong University of Science and Technology*，*Wuhan* 430074，*China*)

Abstract：In this letter，with the use of a backpropagation algorithm，Lyapunov function method is adopted to analyze the convergence of weights for feedforward Neural Networks learning with time varying inputs. So the factor of convergence can be proposed to minima of the error function. With this factor，a backpropagation algorithm based on single parameter is sound and quick. Then a backpropagation algorithm named with single parameter and multi-regulations is proposed.

Key words：lyapunov；feedforward neural networks；backpropagation algorithm；single parameter；multi-regulations

1 引　言

李雅普诺夫 V 函数方法的特点是不用求解系统的动态方程或分析参数特征根的实数部分正负特性，直接用确定并分析所研究状态点处的广义能量函数来判别系统的稳定性. 从物理观念上，如果系统的某个平衡状态是渐进稳定的，那随着系统的状态变化，其储存的能量应当是随着时间的推移而逐渐衰减的，直至趋于平衡状态而使能量趋于极小值，就如水泼在高低不平的地面上逐渐流动到低凹处并稳定下来.

揭示并模拟大脑神经系统的学习机理是研制智能信息处理系统的关键之一，这里，将深入探讨基于时变输入前馈神经网络的 BP 学习算法收敛性，进而分析单参数 BP 神经网络算法的收敛性. 李雅普诺夫函数方法应用于自组织神经网络（如 Kohonen 和 Hopfield 网络）的稳定分析，可用来推导条件来保证权值的收敛，将可以看到基于 BP 的学习算法具有收敛局部最小值的内在特性. 假如权值的初始值处于全局最小值或权值的分布有利于权值收敛于全局最小值，那 BP 学习机制将使权值收敛到全局最小值.

2 分　析

逆传播 BP 算法把网络输出出现的误差归结为各权值的"过错"，通过把输出网络的误差逐层向输入层逆向传播以"分摊"给各层单元，从而获得各层单元的参考误差以便调整相应的权值.

首先，如下定义前馈神经网络的输入、权值、期望输出和实际输出：

$$x(t)=(x1,x2,\cdots,xn)^T\in R;$$

$$w(t)=(w1,w2,\cdots,wl)^T\in R;$$

$$yd(t)=(yd1,yd2,\cdots,ydm)^T\in R;$$

$$y(t)=(y1,y2,\cdots,ym)T\in R.$$

其中，$X(t)$ 表示 FNN 的输入矢量，$w(t)$ 表示 FNN 感知器的权值矢量，$yd(t)$ 表示 FNN 的期望输出矢量，$y(t)$ 表示 FNN 的实际输出矢量. 符号 T 表示矢量的转置，并且这里的 $x(t)$ 是通用的类型，既可以是离散类型，也可以是连续类型. 于是平方误差可表示为

$$\begin{aligned} e(t) &= (y(t)-yd(t))^T(y(t)-yd(t))/2 \\ &= \| y(t)-yd(t) \| /2. \end{aligned} \tag{1}$$

如果用 τ 表示某时间段的区间长度，那该时间段的平均平方误差可以如下表示：

$$J=\frac{1}{\tau}\int_{t-\tau}^{t} e(\theta)d\theta \tag{2}$$

收稿日期：2002-07-26　作者简介：张茂元，博士生，主要从事人工智能和网络信息系统等方面研究；卢正鼎，教授，博士生导师，主要从事数据库，软件工程和信息系统等方面研究.

$$J_k = \lim_{\tau \to 0} \frac{1}{\tau} \int_{t-\tau}^{t} e(\theta) d\theta = e(k) \tag{3}$$

公式(2)适合于连续时间类型的 FNN 训练,公式(3)则用于离散类型的 FNN 训练.

对 J 和权值向量 w 建立李雅普诺夫 V 函数:

$$V(J,w) = \alpha J + \frac{1}{2}\beta \left| \frac{\partial J}{\partial w} \right|^2 \tag{4}$$

其中,符号 ‖ ‖表示矢量的二次范数,参数 α 和 β 都大于零,用来决定两项的比重程度. 且

$$\left| \frac{\partial J}{\partial w_l} \right|^2 = \left(\frac{\partial J}{\partial w_l} \right) \left(\frac{\partial J}{\partial w} \right)^T = \left(\frac{\partial J}{\partial w} \right) \left(\frac{\partial J}{\partial w^T} \right),$$

$$\left(\frac{\partial J}{\partial w} \right) = \left(\frac{\partial J}{\partial w_l}, \cdots, \frac{\partial J}{\partial w_1} \right),$$

$\left(\frac{\partial J}{\partial w^T} \right) = \left(\frac{\partial J}{\partial w_l}, \cdots, \frac{\partial J}{\partial w_1} \right)^T$. 显然,$J$ 和 $\| \partial J / \partial W \|^2$ 在极小值附近都是正值,加上参数 α,β>0,所以函数 V(J,w)是正定的.

采用如下 FNN 权值 w 调整法则,其中参数 u,v 大于零,I_1 为 $l * l$ 阶单位矩阵.

$$\frac{dw}{dt} = \begin{cases} = (-\alpha I_1 + \beta \frac{\partial^2 J}{\partial w \partial w^T})^{-1} (\frac{\frac{\partial J}{\partial w^T}}{\left| \frac{\partial J}{\partial w} \right|^2}) \\ (\alpha \frac{\partial J}{\partial t} + \beta \frac{\partial J}{\partial w} \frac{\partial^2 J}{\partial t \partial w^T} + u \left| \frac{\partial J}{\partial w} \right|^2 + vJ^2), if \left| \frac{\partial J}{\partial w} \right| \neq 0 \\ = 0, if \left| \frac{\partial J}{\partial w} \right| = 0 \end{cases} \tag{5}$$

对离散数据输入的自学习算法,$\partial J / \partial t = 0$,$\partial^2 J (\partial t \partial w^T) = 0$,可用(w(k+1)-w(k))/Δt 来表示 w'. 此函数 V 对现存的 BP 算法有通用性,例如:取 β=0,v=0 时,调整法则变为 $w' = -\alpha^{-1} u (\partial J_k / \partial w^T) = -\lambda (\partial J_k / \partial w^T)$,其中 $\lambda = \alpha^{-1} u$,这就是传统的梯度下降算法.

以下推导BP学习算法具有误差收敛到极小值的内在性.

定理:设 FNN 结构中的输入输出关系为 y(t)=f(w(t),x(t)),其中 x(t)是隐式表示的. 如果满足式(5),那么 w 渐进稳定收敛到极小值,$\partial J / \partial w$ 能渐进稳定趋于零.

证明:对 w 按式(4)构建 V(J,w)函数,显然,函数 V 在局部最小值附近是正定的,如果 w 达到局部最小值平衡点 w^*,V(J,w)满足 $\partial J / \partial w |_{w=w^*} = 0$. 函数 V 对 t 求导,得:

$$\begin{aligned} V' &= \alpha \left(\frac{\partial J}{\partial w} w' + \frac{\partial J}{\partial x} x' + \frac{\partial J}{\partial t} \right) \\ &+ \beta \frac{\partial J}{\partial w} \left(\frac{\partial^2 J}{\partial t \partial w^T} + \frac{\partial^2 J}{\partial w \partial w^T} w' + \frac{\partial^2 J}{\partial x \partial w^T} x' \right) \\ &= \frac{\partial J}{\partial w} \left(\alpha I_l + \beta \frac{\partial^2 J}{\partial w \partial w^T} \right) w' + \left(\alpha \frac{\partial J}{\partial x} + \beta \frac{\partial J}{\partial w} \frac{\partial^2 J}{\partial x \partial w^T} \right) x' \\ &+ \left(\alpha \frac{\partial J}{\partial t} + \beta \frac{\partial J}{\partial w} \frac{\partial^2 J}{\partial t \partial w^T} \right) \end{aligned}$$

由于函数 V 中的 x(t)是隐式形式存在的,所以 $\partial J / \partial x = 0$,$\partial^2 J / (\partial x \partial w^T) = 0$,从而

$$V' = \frac{\partial J}{\partial w} \left(\alpha I_l + \beta \frac{\partial^2 J}{\partial w \partial w^T} \right) w' + \alpha \frac{\partial J}{\partial t} + \beta \frac{\partial J}{\partial w} \frac{\partial^2 J}{\partial t \partial w^T}$$

将式(5)代入,并化简得

$$V' = \begin{cases} -u \left| \frac{\partial J}{\partial w} \right|^2 - VJ^2, if = \left| \frac{\partial J}{\partial w} \right| \neq 0 \\ \alpha \frac{\partial J}{\partial t}, if = \left| \frac{\partial J}{\partial w} \right| = 0 \end{cases}$$

当 $\| \partial J / \partial w \| \neq 0$ 时,$V' = -u \| \partial J / \partial w \|^2 - vJ^2 < 0$,即 V 在局部附近是单调递减,V 将在 w 的调整下单调减小,直至 $\| \partial J / \partial w \| = 0$,即 $\partial J / \partial w = 0$. 可见$\lim_{t \to \infty} (\partial J / \partial w) = 0$,$\partial J / \partial w$ 趋于零,w 的极小值是满足式(5)w(t)的解.

因函数 V 是定正函数,且 V′是定负函数,由应用于非自治系统的李雅普诺夫定理(对于局部范围正定的 V(J,w),且 w 在 w^* 的附近,如果导数 V′为常负函数,则平衡点 w * 是李雅普诺夫稳定点;如果导数 V′是定负函数的,则平衡点 w^* 是李雅普诺夫渐进稳定点.)可知满足式(5)的 w=w(t)解是渐进稳定收敛的,即 w 是渐进稳定收敛到极小值,那么 $\partial J / \partial w$ 渐进稳定收敛到零,命题得证.

综上所述,V 在 w 的调整下单调减小,直至 $\| \partial J / \partial w \| = 0$,此时 w 渐进稳定收敛到局部最小值,并不再调整,自学习停止,不再有减小 J 的机制. 可见式(5)可以使 ∂J/∂w 渐进趋于零,但不能使 J 渐进趋于零. 因此,函数 V′与 J′及 $\partial J / \partial w$ 内在相关性,只能保证在全局最小值 w^* 的稳定性,并非渐进稳定性. 对时变输入,如果$(\partial J / \partial t) \leqslant 0$ 满足,当自学习停止时,即 $\| \partial J / \partial w \| = 0$,仍可保持 J 的收敛趋势,且约束了权值 w 的收敛.

3 应 用

单参数 BP 神经网络算法,是多层前向神经网络在反向传播误差时,每次仅调整一个权值,其它参数不变. 其计算步骤为:

1) 给各权值参数赋予初值;

2) 如果权值都已处理过,跳至步骤 5,否则从未处理的权值中取一权值,设为 w_i,;

3) 计算权值 w_i 的调整值,并对权值 w_i 调整;

4) 如果 $|\partial J / \partial w_i| > 0.01$,转至步骤 3,否则转至步骤 2;

5) 算法计算结束.

3.1 用李雅普诺夫函数分析单参数 BP 神经网络算法的收敛性;

对步骤 3)、4)分析权值 w_i 收敛性,因仅调整权值 w_i,所以 $\partial J / \partial w_j = 0 (j \neq i)$. 调整法则式(5)化简为:

$$\frac{dw}{dt} = \begin{cases} = -(\alpha I_1 + \beta \frac{\partial^2 J}{\partial w \partial w^T})^{-1} (\frac{\frac{\partial J}{\partial w^T}}{(\frac{\partial J}{\partial w_i})^2}) \\ (\alpha \frac{\partial J}{\partial t} + \beta \frac{\partial J}{\partial w_i} \frac{\partial^2 J}{\partial t \partial w_i} + u (\frac{\partial J}{\partial w_i})^2 + vJ^2), if \frac{\partial J}{\partial w_i} \neq 0 \\ = 0, if \frac{\partial J}{\partial w_i} = 0 \end{cases}$$

$$\frac{dw_j}{dt}=\begin{cases}=-(\alpha+\beta\frac{\partial^2 J}{\partial w_i^2})^{-1}(\frac{\partial J}{\partial w_i})^{-1}(\alpha\frac{\partial J}{\partial t}+\beta\frac{\partial J}{\partial t_i}\cdot\frac{\partial^2 J}{\partial t\partial W_i}+\\ u(\frac{\partial J}{\partial w_i})^2+vJ^2),if(j=i,\text{且}\frac{\partial J}{\partial w_i}\neq 0)\\ =0,if(j\neq 1,\text{或}\frac{\partial J}{\partial w_i}=0)\end{cases}$$

李雅普诺夫 V 函数:$V(J,w_i)=\alpha J+\frac{1}{2}\beta(\partial J/\partial w_i)^2$,对 t 求导:

$$\begin{aligned}V'&=\alpha(\frac{\partial J}{\partial w_i}w_i'+\frac{\partial J}{\partial x}x'+\frac{\partial J}{\partial t})+\beta\frac{\partial J}{\partial w_i}\\&(\frac{\partial^2 J}{\partial t\partial w_i}+\frac{\partial^2 J}{\partial w_i^2}w_i'+\frac{\partial^2 J}{\partial x\partial w_i}x')\\&=\frac{\partial J}{\partial w_i}(\alpha+\beta\frac{\partial^2 J}{\partial w_i^2})w_i'+(\alpha\frac{\partial J}{\partial x}+\beta\frac{\partial J}{\partial w_i}\frac{\partial^2 J}{\partial x\partial w_i})x'\\&+(\alpha\frac{\partial J}{\partial t}+\beta\frac{\partial J}{\partial w_i}\frac{\partial^2 J}{\partial t\partial w_i})\\&=\frac{\partial J}{\partial w_i}(\alpha+\beta\frac{\partial^2 J}{\partial w_i^2})w_i'+(\alpha\frac{\partial J}{\partial t}+\beta\frac{\partial J}{\partial w_i}\frac{\partial^2 J}{\partial t\partial w_i})\end{aligned}$$

(因为 x 是隐式存在)

$$\begin{aligned}&=-(\alpha\frac{\partial J}{\partial t}+\beta\frac{\partial J}{\partial w_i}\frac{\partial^2 J}{\partial t\partial w_i}+u(\frac{\partial J}{\partial w_i})^2+vJ^2)+\\&(\alpha\frac{\partial J}{\partial t}+\beta\frac{\partial J}{\partial w_i}\frac{\partial^2 J}{\partial t\partial w_i})\\&=-u(\frac{\partial J}{\partial w_i})^2-vJ^2\end{aligned}$$

$$\text{所以 }V'=\begin{cases}-u(\frac{\partial^2 J}{\partial w_i})^2-vJ^2<0,if\ \frac{\partial J}{\partial w_i}\neq 0\\ \alpha\frac{\partial J}{\partial t},if\ \frac{\partial J}{\partial w_i}=0\end{cases}\tag{7}$$

当 $\|\partial J/\partial w_i\|\neq 0$ 时,$V'=-u\|\partial J/\partial w_i\|^2-vJ^2<0$,即 V 在局部附近是单调递减,V 将在 w_i 的调整下单调减小,直至 $\|\partial J/\partial w_i\|=0$. 可见 $\lim\limits_{t\to+\infty}\partial J/\partial w_i=0$,$\partial J/\partial w_i=0$,$\partial J/\partial w_i$ 趋于零,w_i 的极小值是满足式(6)$w_i(t)$的解. 因函数 V′是定负函数,V 是定正函数,由应用于非自治系统的李雅普诺夫定理,可知 w_i 是渐进稳定收敛到极小值.

从上面的权值 w_i 的收敛性分析可知,单参数 BP 神经网络算法仅调整一个权值,可以使该权值渐进稳定收敛到极小值;依次类似调整其它权值,逐步把它们收敛到各自的极小值. 这就像在多维空间求极小值,可以依次在各维上求得极小值,这体现了 BP 神经网络算法收敛的内在性.

3.2　分析单参数 BP 神经网络算法的收敛速度

比较采用式(5)的非单参数 BP 神经网络算法 A 和采用式(6)的单参数 BP 神经网络算法 B 的收敛速度. 假定算法 B 的对一个权值调整的迭代次数为 K_i,平均迭代次数为 K_c,最大迭代次数为 K_m,则算法 B 总的迭代次数为 l^*k_c 次;算法 A 中,由于各个权值互相制约,使得对向量 w 的调整次数大于 K_c 次,甚至大于 K_m 次. 又由于在每次对向量 w 的调整中,要对 L 个权值调整,因此算法 A 的总的迭代次数大于 l^*k_c 次,也就是算法 A 的迭代次数大于算法 B 的迭代次数. 算法 B 每次迭代的计算复杂度为 O(1);算法 A 中由于每次对向量 w 的调整,要计算矩阵的逆,计算复杂度为 $O(l^*\ln l)$,因此对每个权值的每次调整算法 A 的计算复杂度为 $O(\ln l)$. 经比较,算法 B 不仅在迭代次数上,而且在迭代的计算复杂度上优于算法 A,可见算法 B 的收敛速度大于算法 A.

单参数算法由于每次仅调整一个参数,因此每个参数能较快地收敛到极小值,大大减少了各误差函数的计算量,从而加快了 BP 神经网络算法自学习的收敛速度.

3.3　单参数变调整法则的离散型 BP 神经网络算法

式(5)是通用的 BP 神经网络调整法则,而式(6)是通用的单参数 BP 神经网络调整法则. 对离散数据输入,取 β=0,v=0时,式(6)调整法则变为式(8):$w_i'=-\alpha^{-1}u(\partial J_K/\partial w_i)=-\lambda(\partial J_K/\partial w_i)$,其中 $\lambda=\alpha^{-1}u$,这就是单参数梯度下降算法;取 v=0 时,调整法则变为式(9):$w_i'=-(\alpha+\beta(\partial^2 J_K/\partial w_i^2))^{-1}u(\partial J_K/\partial w_i)$,这是单参数 Levenberg－Marquardt 算法.

由于小的学习率使网络需要较长的时间收敛,大的学习率又可能会产生学习过程的振荡. 在前面分析的基础上,提出一个单参数变调整法则的离散型 BP 神经网络算法. 其算法思想是:当权值离极小值较远时,采用调整幅度大的调整法则来加快学习速度;当权值离极小值较近时,采用调整幅度小的调整法则,来避免学习过程中的振荡. 采用多个调整法则后,运用前一个调整法则得到权值的最后结果就是运用后一个调整法则的权值初值,并由于都朝极小值方向收敛,因此权值最后渐进稳定收敛到极小值.

其算法步骤为:1)给各权值参数赋予初值;2)如果权值都已处理过,跳至步骤 7,否则从未处理的权值中取一权值,设为 w_i;3)计算 $\partial^2 J/\partial w_i^2$,如果小于零,转至步骤 4,否则跳至步骤 5;4)采用调整法则式(9),对权值 w_i 调整,转至步骤 3;5)如果 $|\partial J/\partial w_i|<0.01$,转至步骤 2,如果 $|\partial J/\partial w_i|>0.3$ 采用调整法则式(8),否则采用调整法则式(9);6)对权值 wi 调整,转至步骤 5;7)算法计算结束.

4　结束语

通过用李雅普诺夫函数方法对 BP 神经网络算法收敛的分析,BP 神经网络算法确实具有朝最小误差方向调整权值的内在因素. 从加快收敛速度上,本文提出的单参数变调整法则离散型 BP 神经网络算法是对 BP 神经网络算法的深入研究. 在这个基础上,BP 神经网络自学习的收敛速度、归纳性和容错性都可以继续深入研究.

References:

1 Xinhuo Yu, M. Onder Efe, and Okyay Kaynak. A general backpropagation algorithm for feedforward neural networks learning [J]. IEEE Trans. Neural Networks, Jan. 2002, 13(1): 251～254.

2 Chen Xiao-dong, Feng Ying-jun. A Research of the backpropagation algorithm based on single parameter[J]. Journal of Applied Sciences. 2000, 18(1): 24～26.

3 Le Fan, Fan Jun-bo, Tan Yong-dong, Neural networks and neural computer[M]. Chengduo: Southwest JiaoTong University, Press, 1991.

附中文参考文献:

2 陈晓东,冯英浚. 单参数快速搜索 BP 算法的研究与应用[J]. 应用科学学报,2000,18(1):24～26.

3 勒蕃,范俊波,谭永东. 神经网络与神经计算机[M]. 成都. 西南交通大学出版社,1991.

基于李雅普诺夫函数的BP神经网络算法的收敛性分析

作者： 张茂元， 卢正鼎
作者单位： 华中科技大学,计算机科学与技术学院,湖北,武汉,430074
刊名： 小型微型计算机系统ISTIC PKU
英文刊名： MINI-MICRO SYSTEMS
年，卷(期)： 2004,25(1)

参考文献(5条)

1. 勒蕃;范俊波;谭永东 神经网络与神经计算机 1991

2. 陈晓东;冯英浚 单参数快速搜索BP算法的研究与应用[期刊论文]-应用科学学报 2000(01)

3. Le Fan;Fan Jun-bo;Tan Yong-dong Neural networks and neural computer 1991

4. CHEN Xiao-dong;Feng Ying-jun A Research of the backpropagation algorithm based on single parameter 2000(01)

5. Xinhuo Yu;M.Onder Efe;Okyay Kaynak A general backpropagation algorithm for feedforward neural networks learning 2002(01)

引证文献(3条)