第 28 卷 第 5 期
2002 年 9 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 28, No. 5
Sept., 2002

神经网络 BP 学习算法动力学分析[1)]

梁久祯[1] 何新贵[2] 周家庆[1]

[1](浙江师范大学计算机科学与工程学院 金华 321004)

[2](北京系统工程研究所 北京 100101)

(E-mail: jiuzhen@yahoo.com)

摘 要 研究神经网络 BP 学习算法与微分动力系统的关系. 指出 BP 学习算法的迭代式与相应的微分动力系统数值解 Euler 方法在一定条件下等价，且二者在解的渐近性方面是一致的. 给出了神经网络 BP 学习算法与相应的微分动力系统解的存在性、唯一性定理和微分动力系统的零解稳定性定理. 从理论上证明了神经网络的学习在一定条件下与微分动力系统的数值方法所得的数值解在渐近意义下是等价的，从而借助于微分动力系统的数值方法可以解决神经网络的学习问题. 最后给出了用改进 Euler 方法训练 BP 网的例子.

关键词 神经网络，学习算法，动力系统，存在性，唯一性，稳定性

中图分类号 TP183

DYNAMIC ANALYSIS OF BP ALGORITHM FOR NEURAL NETWORKS

LIANG Jiu-Zhen[1] HE Xin-Gui[2] ZHOU Jia-Qing[1]

[1](*College of Computer Science & Engineering, Zhejiang Normal University, Jinhua* 321004)

[2](*Beijing Institute of System Engineering, Beijing* 100101)

(E-mail: jiuzhen@yahoo.com)

Abstract This paper deals with the relationship between BP algorithm for neural networks and differential dynamic systems. It is proposed that the iteration formula of BP algorithm is equivalent to Euler method of differential dynamic system under certain conditions, and the asymptotic solutions of the two formulas are consistent. For neural networks' BP algorithm and their corresponding differential dynamics, the solution existence theorem, exclusiveness theorem and zero solution stability theorem of dynamic system are presented. It is also theoretically proved that asymptotic solutions given by neural networks' BP algorithm are equivalent to that computed by any numerical method for differential dynamic systems under certain conditions. Therefore, training a neural network can be converted to computing numerical solution of differential dynamic systems. Also, an example to train the BP network by modified Euler method is presented.

Key words Neural networks, learning algorithm, dynamic system, existence, exclu-

1) 浙江省自然科学基金重点项目(ZD0108)资助

收稿日期 2000-09-01 收修改稿日期 2001-08-20

siveness, stability

1 引言

自从多层网络 BP 学习算法发明以来[1]，神经网络得到了广泛的应用和重视. 由于 BP 学习算法存在学习速度慢和易陷入局部极小等特点，许多人致力于 BP 算法的改进，如发展了改进 BP 学习算法[2]、自适应 BP 学习算法[3]、超线性 BP 学习算法[4]等. 但由于 BP 类学习算法都属于梯度下降搜索算法，故目标函数在搜索点的梯度信息是关键的因素. 另外，就 BP 类算法的迭代式而言，在权值修改的过程中要形成一个权值序列. 该序列的收敛性、收敛速度和极限点构成了算法的基本特征. 对于线性 BP 网络的学习问题，这些基本特征在 LMS 算法已有了经典的结论[5]. 对于非线性 BP 网，分析这些特征还没有好的方法. 另一方面，由于 BP 算法迭代式本身可以看作是一个离散的动力系统，它与微分动力系统有着许多的相似性，因此借助于微分动力系统研究 BP 学习算法迭代式的特征是一种有价值的尝试. 目前，虽然神经网络学习算法仍然是一个研究热点，但关于这一方面报道的文章除文献[6,7]中提及外为数少见.

本文考察神经网络 BP 学习算法与关于权值向量的微分动力系统的关系. 指出 BP 学习算法的迭代式是微分动力系统数值解 Euler 方法的一种特殊形式. 在一定条件下，BP 学习算法的数值解与微分动力系统的解在渐近性方面是一致的. 通过研究微分动力系统解的存在性、唯一性和稳定性来分析 BP 算法解序列的特征. 最后指出，神经网络的学习在一定条件下可以借助微分动力系统常用的数值解法，从而开拓出更广泛的神经网络学习算法.

2 问题的提出

为确定起见，我们只考虑多层 BP 神经网络. 设 $\boldsymbol{w}$ 为任意多层 BP 网络的连接权值(包括阈值)，即 $\boldsymbol{w}=(w_1,w_2,\cdots,w_n)$，$g(x)=1/(1+e^{-x})$ 为非线性传递函数. 样本空间为 (X,Y)，X 为样本输入，Y 为样本输出(期望输出). 误差能量函数定义为

$$E = \| Y_r - Y \|^2 \tag{1}$$

其中 $\|\cdot\|$ 为某一范数，Y_r 为神经网络的实际输出. 由于当样本空间确定以后，E 只是权值向量 $\boldsymbol{w}$ 的函数，故上式可改写为

$$E = E(\boldsymbol{w}) \tag{2}$$

通常，训练多层神经网络的 BP 算法标准形式为

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k - h\nabla E(\boldsymbol{w}^k), \quad \boldsymbol{w}^0 = \boldsymbol{w}_0 \qquad (3a),(3b)$$

其中 $h>0$ 为学习步长，而 $\nabla E(\boldsymbol{w}^k)=\left(\frac{\partial E}{\partial w_1},\frac{\partial E}{\partial w_2},\cdots,\frac{\partial E}{\partial w_n}\right)^{\mathrm{T}}$ 为 $E(\boldsymbol{w}^k)$ 在点 $\boldsymbol{w}^k$ 的梯度.

考虑如下微分动力系统

$$\frac{\mathrm{d}\boldsymbol{w}(t)}{\mathrm{d}t} = -\nabla E(\boldsymbol{w}(t)), \quad \boldsymbol{w}(t_0) = \boldsymbol{w}_0 \qquad (4a),(4b)$$

其中 $\boldsymbol{w}(t)$，$E(\boldsymbol{w}(t))$ 为关于 t 的连续可微函数，且 $\boldsymbol{w}(t)\in C_n^1[0,\infty)$，$E(\boldsymbol{w}(t))\in C_n^2[0,\infty)$，这里 $C_n^1[0,\infty)$ 为 n 维 1 阶连续可微函数类. 众所周知，在数值分析中求解微分动力系统(4)的一阶 Euler 方法为

$$w^{k+1} = w^k - (t_{k+1} - t_k)\nabla E(w^k) \tag{5}$$

其中 $w^0=w(t_0)$ 为已知的初值. 若取 $t_{k+1}-t_k=h$，对 $k=0,1,2,\cdots$，则式(5)可化为式(3)的形式. 由此可见，标准 BP 算法(3)是求解微分动力系统(4)的 Euler 方法步长为学习速度 h 的特例. 另一方面，考虑是否可以通过考察动力系统(4)的特征来研究 BP 算法(3)的解序列的性质. 与此相关的问题有

1) 式(4)和(5)的解存在性和唯一性如何；

2) 若式(4)和(5)的解都存在且唯一，那么二者在渐近意义下是一致的吗，即是否有

$$\lim_{k\to\infty} w^k = \lim_{t\to\infty} w(t) \tag{6}$$

其中 w^k 为(3)的解，而 $w(t)$ 为(4)的解；

3) 动力系统(4)的初值稳定条件是什么；

4) 若在上述问题都得到肯定的回答之后，用微分方程其它数值解法(如 Runge-Kutta 方法)求解(4)，所得数值解是否与 BP 算法(3)的解在渐近意义下是一致的.

3　解的存在性和唯一性

首先考察 BP 学习算法(3). 若式(3)的渐近解存在，即

$$\lim_{k\to\infty} w^k = w^* \tag{7}$$

其中 w^* 为 w^k 的极限，或聚点. 由式(3)知，式(7)等价于

$$\lim_{k\to\infty} \nabla E(w^k) = \lim_{t\to\infty} \nabla E(w(t)) = 0 \tag{8}$$

事实上，只要式(3)的解存在，设 w^*，w^{**} 都是式(3)的任意两个渐近解，且具有相同的初值 w_0，显然有 $w^*=w^{**}$.

引理 1. BP 学习算法(3)的解存在且唯一的充分必要条件为式(8)成立 .

再考察式(4)的渐近解. 对于动力系统(4)引入如下定义.

定义 1. 设 $S\subset R^n$ 为一开集，函数 $F(t,w)$ 为 $T\times S\to R^n$ 上的连续映射，称 F 对 w 具有局部 Lipschitz 条件，指对任何有界闭集 $M\subset T\times S$ 有常数 $K(M)>0$，使

$$\| F(t,w_1) - F(t, w_2) \| \leqslant K(M) \| w_1 - w_2 \|, \quad \forall\ (t,w_1),(t,w_2) \in M \tag{9}$$

其中 $T\subset R$. 记 $F(t,w)\in L_M(T\times S)$ 表示 F 在 $T\times S$ 上对 w 具局部 Lipschitz 条件，若 $K(w)$ 可以选择与 M 无关，则称 $F(t,w)$ 在 $T\times S$ 上具 Lipschitz 条件，并记 $F\in L(t,w)$.

由文献[8]知，对于动力系统(4)成立如下结论.

引理 2. 若 $\nabla E(w):T\times S\to R^n$，$S\subset R^n$，又 $\nabla E(w(t))\in C_n(T\times S)\cap L_M(T\times S)$，则微分动力系统(4)对初值问题 $w(t_0)=w_0$，$F(t_0, w_0)\in L_M(T\times S)$ 存在唯一解，记为 $w(t;t_0, w_0)$，即它是式(4)的解，且有 $w(t_0;\ t_0,w_0)=w_0$.

引理 3. 动力系统(4)存在唯一解，若 M 为有界闭集，S 为开集，且 $E(w(t))\in C_n^2(T\times S)$.

证明. 由于 $E(w(t))$ 关于 $w(t)$ 一阶连续可微，即 $\nabla E(w(t))\in C_n^1(T\times S)$，显然成立 $\nabla E(w(t))\in C_n(T\times S)$. 另一方面，由定义 1，$\forall\ (t,w_1),(t,w_2)\in M\subset T\times S$，由于 $\nabla E(w)$ 关于 $w(t)$ 可微，且 $\nabla^2 E(w(t))$ 在有界闭集 M 上有上确界，记

$$K(M) = \sup_{w(t)} \| \nabla^2 E(w(t)) \| \tag{10}$$

由微分中值定理，有

$$\nabla E(\boldsymbol{w}_1)-\nabla E(\boldsymbol{w}_2)=\nabla^2 E(\xi)(\boldsymbol{w}_1-\boldsymbol{w}_2) \tag{11}$$

其中 $\xi\in M$. 于是

$$\|\nabla E(\boldsymbol{w}_1)-\nabla E(\boldsymbol{w}_2)\| \leqslant \|\nabla^2 E(\xi)\| \cdot \|\boldsymbol{w}_1-\boldsymbol{w}_2\| \leqslant K(M)\|\boldsymbol{w}_1-\boldsymbol{w}_2\|,$$

即 $\nabla E(\boldsymbol{w})$ 对 $\boldsymbol{w}$ 具 Lipschitz 条件,或 $\nabla E(\boldsymbol{w}(t))\in C_n(T\times S)\cap L_M(T\times S)$. 由引理 2,动力系统(4)存在唯一解. 证毕.

定理 1. 若 M 为有界闭集, S 为开集,且 $E(\boldsymbol{w}(t))\in C_n^2(T\times S)$, $\lim\limits_{k\to\infty}\nabla E(\boldsymbol{w}^k)=\lim\limits_{t\to\infty}\nabla E(\boldsymbol{w}(t))=0$. 则 BP 学习算法(3)和动力系统(4)的解存在且唯一.

4 BP 学习算法与微分动力系统

记 $e(t)=\boldsymbol{w}^t-\boldsymbol{w}(t)$,即 $e(t)$ 为 BP 算法(3)的解与动力系统(4)的解的偏差. 现考察极限 $\lim\limits_{t\to\infty}e(t)$ 的性质. 由微分方程数值解法 Euler 方法的截断误差公式知

$$e(t)=\boldsymbol{w}^t-\boldsymbol{w}(t)=-\sum_{n=2}^{\infty}\frac{h^n}{n!}\frac{\mathrm{d}^n\boldsymbol{w}(t)}{\mathrm{d}t^n}=O(h^2) \tag{12}$$

下面逐项考察如下极限的性质.

$$\lim_{t\to\infty}\frac{\mathrm{d}^n\boldsymbol{w}(t)}{\mathrm{d}t^n},\quad n=2,3,\cdots \tag{13}$$

当 $n=2$ 时,

$$\frac{\mathrm{d}^2\boldsymbol{w}(t)}{\mathrm{d}t^2}=\nabla^2 E(\boldsymbol{w}(t))\nabla E(\boldsymbol{w}(t)) \tag{14}$$

由于 $\nabla^2 E(\boldsymbol{w}(t))$ 在有界闭集 M 上按范数有上确界,而 $\lim\limits_{t\to\infty}\nabla E(\boldsymbol{w}(t))=0$,故

$$\lim_{t\to\infty}\frac{\mathrm{d}^2\boldsymbol{w}(t)}{\mathrm{d}t^2}=0 \tag{15}$$

当 $n>2$ 时, 由于 $\frac{h^n}{n!}\frac{\mathrm{d}^n\boldsymbol{w}(t)}{\mathrm{d}t^n}=O\left(\frac{h^2}{2!}\frac{\mathrm{d}^2\boldsymbol{w}(t)}{\mathrm{d}t^2}\right)$,这里 $h<1$. 故当 $h<1$ 时,有

$$\lim_{t\to\infty}e(t)=0 \tag{16}$$

于是,可将上述结果总结为如下定理.

定理 2. 若式(4)由 Euler 方法求解,当步长 $h<1$,且 $\lim\limits_{t\to\infty}\nabla E(\boldsymbol{w}(t))=0$ 时,有

$$\lim_{t\to\infty}\boldsymbol{w}^t=\lim_{t\to\infty}\boldsymbol{w}(t) \tag{17}$$

即 BP 算法(3)的解与动力系统(4)的解具有渐近等价性.

注 1. 定理 2 的前提条件说明多层神经网络的误差曲面(1)只要存在平衡点,且 BP 学习算法(3)收敛到该平衡点,则相应的微分动力系统的解也收敛到该平衡点.

5 动力系统的初值稳定性

由于式(4)为一微分动力系统,在给定初值后系统的运行状态是必须要考虑的. 若微分动力系统不稳定,则即使得到了收敛的解,该解只能是一个不稳定的平衡点,而不是一个吸引子. 我们所期望的是得到系统的稳定吸引子,即要考察系统(4)稳定的条件. 设 $\boldsymbol{w}^*(t)$ 为动力系统(4)的一个特解,考虑 $U(t)=\boldsymbol{w}(t)-\boldsymbol{w}^*(t)$,则

$$\frac{\mathrm{d}U(t)}{\mathrm{d}t} = \nabla E(\boldsymbol{w}(t) - U(t)) - \nabla E(\boldsymbol{w}(t)) \stackrel{\text{记}}{=} f(t,U(t)) \tag{18}$$

显然有 $f(t,0)=0$. 于是系统(4)的稳定性等价于系统(18)的零点稳定性. 引入如下定义.

定义 2. 系统(18)的零解 $U(t)=0$ 是稳定的,是指$\forall\ \varepsilon>0$, $\forall\ t_0\in T$, $\exists\ \delta>0$,对$\forall\ U_0\in B_\delta$,$\forall\ t\geqslant t_0$, 有 $U(t;t_0,U_0)\in B_\varepsilon$. 其中 $B_\varepsilon=\{U|U\in R^n, \|U\|<\delta\}$, $B_\varepsilon=\{U|U\in R^n, \|U\|<\varepsilon\}$.

引理 4[8]. 若系统(18)存在 Lyapnov 函数 $V(t,U(t))$,满足

1) $V(t,0)=0$;

2) $V(t,U(t))$正定;

3) $\dot V(t,U(t))\leqslant 0$.

则称系统(18)是零点稳定的，其中 $\dot V(t,U(t))=\dfrac{\partial V}{\partial t}+\nabla V\cdot\dfrac{\partial U}{\partial t}$.

下面构造系统(18)的 Lyapnov 函数. 考察

$$V(t,U(t)) = \frac{1}{2}U(t)^{\mathrm{T}}U(t) \tag{19}$$

显然引理 4 的 1)和 2)满足,而

$$\frac{\partial V}{\partial t} = U(t)^{\mathrm{T}}\frac{\partial U(t)}{\partial t},\quad \nabla V = \frac{\partial V(t,U(t))}{\partial U} = U(t),$$

$$\dot V(t,U(t)) = \frac{\partial V}{\partial t} + \nabla V\cdot\frac{\partial U}{\partial t} = 2U(t)^{\mathrm{T}}\frac{\partial U(t)}{\partial t},$$

又

$$\frac{\partial U(t)}{\partial t} = \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{w}(t) - \boldsymbol{w}^*(t)) = \nabla E(\boldsymbol{w}^*(t)) - \nabla E(\boldsymbol{w}(t)) =$$

$$-\nabla^2E(\xi(t))(\boldsymbol{w}(t) - \boldsymbol{w}^*(t)) = -\nabla^2E(\xi(t))U(t).$$

这里 $\xi(t)\in[\min(\boldsymbol{w}(t), \boldsymbol{w}^*(t)), \max(\boldsymbol{w}(t), \boldsymbol{w}^*(t))]$,于是

$$\dot V(t,U(t)) = -2U(t)^{\mathrm{T}}\nabla^2E(\xi(t))U(t) \tag{20}$$

故,当$\nabla^2E(\xi(t))$半正定时,必有 $\dot V(t,U(t))\leqslant 0$.

定理 3. 若$t\in T,\boldsymbol{w}(t)\in S$ 为开集,$\nabla^2E(\boldsymbol{w}(t))$在 $M=(T,S)$上存在且半正定,则微分动力系统(4)关于初值$(t_0, \boldsymbol{w}_0)$稳定.

注 2. 定理 3 说明当微分动力系统(4)所对应的多层神经网络的误差曲面(1)在 S 上二阶连续光滑,且具有单极值时,系统必最终处于稳定状态.

6 BP 学习算法与动力系统数值解的关系

设由微分动力系统(4)的某一数值解法得到的解序列为$\{\overline{\boldsymbol{w}}^k\}$. 问题是这一解序列与由 BP 学习算法(3)得到的解序列$\{\boldsymbol{w}^k\}$关系如何? 是否有如下关系

$$\lim_{k\to\infty}\overline{\boldsymbol{w}}^k = \lim_{k\to\infty}\boldsymbol{w}^k \tag{21}$$

成立? 由定理 3 知,这只要考察是否成立

$$\lim_{k\to\infty}\overline{\boldsymbol{w}}^k = \lim_{t\to\infty}\boldsymbol{w}(t) \tag{22}$$

其中 $\boldsymbol{w}(t)$为动力系统(4)的解. 由此,我们得到 BP 学习算法与微分动力系统数值解之间的关系表述定理.

定理 4. 设$\{\bar{w}^k\}$为由微分动力系统(4)的某一数值解法得到的解序列，$\{w^k\}$为由 BP 学习算法(3)得到的解序列，其中 BP 学习算法的学习速度 $0<h<1$. 若$\lim\limits_{t\to\infty}\nabla E(w(t))=0$ 且式(22)满足，则式(21)成立，即该数值解与 BP 学习算法的解具有渐近一致性.

定理 4 表明，在一定条件下神经网络的 BP 学习问题与求一个相应的微分动力系统是等价的，微分动力系统的解同时也是 BP 学习解序列. 因此，可以借助于求解微分动力系统的较成熟的数值算法解决神经网络的学习问题.

7 训练神经网络的改进 Euler 方法

下面，我们以改进 Euler 方法为例讨论如何将求解微分动力系统(4)的数值解法用于多层神经网络的学习. 考虑改进 Euler 方法公式如下：

$$w^{k+1} = w^k + \frac{1}{2}(K^1 + K^2), \quad K^1 = hf(t_k, w^k) \tag{23a),(23b}$$

$$K^2 = hf(t_k + h, w^k + K^1) \tag{23c}$$

其中 $f(t,w)=-\nabla E(w(t))$.

现对上述算法(23)与 BP 算法(3)的计算复杂度进行简单的分析. 对改进 Euler 方法，由于需要计算 2 次 $f(t,w)$，因此它在一次迭代中的计算量为 BP 算法的 2 倍. 另一方面，改进 Euler 方法的截断误差为$O(h^3)$，而与 BP 算法(3)对应的 Euler 方法的截断误差为$O(h^2)$，由此可见，较大的计算量换取的是较高的精度，这在要求高精度的计算时是值得考虑的. 另外由数值分析知，Euler 方法为显式法，而改进 Euler 方法为隐式法，因此改进 Euler 方法较 BP 算法具有较好的关于初值的稳定性.

用改进 Euler 方法训练多层神经网络的算法描述如下：

第 1 步. 给定误差精度 ε，最大学习次数 M，学习累记数 $s=0$，初始化权重和阈值 w；

第 2 步. 按式(1)计算 E，若 $E<\varepsilon$ 或 $s>M$，则转第 4 步，否则转第 3 步；

第 3 步. 按改进 Euler 方法的式(23)修正权重和阈值，$s+1\to s$，转第 2 步；

第 4 步. 输出学习结果，结束.

考虑到在多数情况下误差曲面(2)是较复杂的，即存在局部极值，因此改进 Euler 方法在训练神经网络时初值的选取对学习的收敛影响很大. 另一方面，BP 算法在学习的初始阶段使误差函数下降较快，而在后期较慢；改进 Euler 方法对初值较敏感，且具有较高的精度. 结合两种算法的优点，给出如下 BP-Euler 结合算法：

第 1 步. 给定误差精度 δ，ε，其中 $\delta>\varepsilon$，最大学习次数 M，学习累记数 $s=0$，初始化权重和阈值 w；

第 2 步. 按式(1)计算 E，若 $E<\delta$，则转第 4 步；否则转第 3 步；

第 3 步. 按 BP 算法(3)修正权重和阈值，$s+1\to s$，转第 2 步；

第 4 步. 按式(1)计算 E，若 $E<\varepsilon$ 或 $s>M$，则转第 6 步；否则转第 5 步；

第 5 步. 按改进 Euler 方法的式(23)修正权重和阈值，$s+1\to s$，转第 4 步；

第 6 步. 输出学习结果，结束.

8 实例及分析

在这一节中利用两个实例对上述三种算法(即 BP 算法，改进 Euler 方法，BP-Euler 结

合法)进行实验,并加以比较分析.

九点模式分类问题. 平面上 9 个点按图 1 排列,可分成◇和◆两类. 选取多层 BP 网络的拓扑结构为 2-5-1,即 2 个输入节点、5 个隐节点和 1 个输出节点.

分别用上述 BP 算法、改进 Euler 方法及 BP-Euler 结合法进行 20 次随机选取网络权值初值的重复学习,取平均值结果如表 1.

◇ ◆ ◇
◆ ◇ ◆
◇ ◆ ◇

图 1　九点模式

表 1　九点模式问题改进 Euler 方法与标准 BP 算法的学习结果对比

误差精度	标准 BP 算法迭代次数	改进 Euler 法迭代次数	BP-Euler 结合法迭代次数
0.1	5038	761	333
0.05	14011	2012	1943

由上例可知,改进 Euler 方法在求解多层神经网络的学习问题中是奏效的. 从迭代次数来看,改进 Euler 方法比 BP 算法有了明显的减少,但由于改进 Euler 方法的计算复杂度为 BP 算法的 2 倍,所以从这个意义上而言,改进 Euler 方法与 BP 算法的收敛速度应是同阶的. 对于 BP-Euler 结合法,计算结果基本上与上一节的分析相符,但如何确定两种算法的转换(即如何确定 δ 的大小)是一个较难把握的问题,实际中需由具体问题而定. 另外,由定理 2 知,微分动力系统(4)的解一定是 BP 网(3)的局部极小点,所以用改进 Euler 方法解多层神经网络的学习问题仍存在局部极值问题.

参 考 文 献

1 Rumelhart D E, Hinton G E, Williams R J. Learning Internal Representations by Error Propagation. In: Parallel Distributed Proceeding, Cambridge MA: MIT Press, 1986. 318～362

2 Reyneri L M, Filippi E. Modified backpropagation algorithm for fast learning in neural networks. *Electronics Letters*, 1990, **26**(19): 1564～1566

3 Qiu G, Varley M R, Terrell T J. Accelerated training of backpropagation neural networks by using adaptive momentum step. *Electronics Letters*, 1992, **28**(4): 377～379

4 梁久祯,何新贵,黄德双. 前馈神经网络的一种超线性收敛 BP 算法. 软件学报, 2000, **11**(8): 1094～1096

5 Haykin S. Introduction to Adaptive Filters, New York: Mcmillan, 1984

6 Pinaki Roychowdhury, Y P Singh, R A Chansarkar. Dynamic tunneling technique for efficient training of multilayer perceptron. *IEEE Transactions on Neural Networks*, 1999, **10**(1): 48～54

7 Forsythe G E, Motzkin T S. Asymptotic properties of the optimum gradient method. *American Mathematical Society Bulletin*, 1951, **57**: 183

8 黄　琳. 稳定性理论. 北京:北京大学出版社, 1992

梁久祯　博士,主要研究方向为神经网络学习理论及其应用.

何新贵　中国工程院院士. 主要研究方向为模糊理论与技术.

周家庆　1991 年和 2001 年分别从西南师范大学和西南交通大学获理学学士和工学硕士学位,现在浙江师范大学计算机科学与工程学院工作. 研究方向为计算机网络管理、信息安全技术以及神经网络等.