文章编号:1001-506X(2009)11-2685-04

样条神经网络的权值直接确定法

张雨浓[1],杨逸文[1,2],肖秀春[1],邹阿金[1],李 巍[1]
(1. 中山大学信息科学与技术学院,广东 广州 510275;
2. 中山大学软件学院,广东 广州 510275)

摘 要:根据样条逼近理论和神经网络原理构造了一种样条神经网络模型,以一组样条基函数作为隐神经元的激励函数。依据误差回传(BP)思想推导出该网络模型的权值修正迭代公式,利用该公式迭代训练可得到该网络的最优权值。而对于构造的具有特定网络结构的样条神经网络,依据伪逆思想提出了一种直接计算权值的方法,从而避免冗长的迭代训练过程。仿真结果表明该权值直接确定法不仅能一步确定权值从而获得更快的运算速度,而且能达到更高的计算精度。

关键词:样条函数;神经网络;权值直接确定;伪逆
中图分类号:TP 183 **文献标志码**:A

Weights direct determination of a spline neural network

ZHANG Yu-nong[1], YANG Yi-wen[1,2], XIAO Xiu-chun[1], ZOU A-jin[1], LI Wei[1]
(1. School of Information Science and Technology, Sun Yat-Sen Univ., Guangzhou 510275, China;
2. School of Software, Sun Yat-Sen Univ., Guangzhou 510275, China)

Abstract: Based on spline approximation theory and neural networks, a spline neural network is constructed, where the hidden-layer neurons are activated by a group of spline functions. Based on the standard error back-propagation (BP) method, the neural-weights updating formula is derived. More importantly, a pseudo-inverse based method is then proposed, which could directly determine the network weights without iterative training. Computer simulation results show that the one-step weights-determination method could be more effective than the standard BP iterative-training method.

Keywords: spline function; neural network; one-step weigths-determination; pseudo-inverse

0 引 言

人工神经网络因其高度并行、分布式存储、自适应自学习等显著特点,在信号处理、模式识别、机器人以及非线性控制等领域有着广泛的应用[1-4]。

BP(back propagation)神经网络模型及其改进模型是当前应用最广泛的神经网络模型之一。但其自身存在一些缺陷,如收敛速度慢和易陷入局部极小点等,因此人们对BP算法做了很多改进。这些工作大多是基于对其权值(或阈值)修正算法的改进[5],即期望通过改进网络训练的迭代训练规则来克服BP网络的缺陷。它们主要分为两类[6-7]:针对标准梯度下降法的改进和基于数值优化方法的改进。

本文不再局限于对BP学习算法的改进,而是通过改进网络结构和激励函数类型来解决收敛速度慢和局部极小点的问题。由此,构造了一种样条神经网络(本文主要研究三次样条神经网络,以下所言样条函数都是指三次样条函数),以样条基函数作为隐神经元的激励函数。网络的隐神经元权值可以在学习过程中通过BP算法获得,但这种迭代训练方法过于复杂与漫长;本文利用伪逆思想来一步确定网络的权值,这种方法在计算速度和精度上都体现出更好的性能。

1 网络模型及理论基础

标准BP神经网络是一种多层前向神经网络,其核心是误差反向传播算法。其做法是依据负梯度下降方向,迭代调整网络的权值和阀值以实现误差函数值的减小。其权值w修正迭代公式可描述为

$$w(k+1) = w(k) + \Delta w = w(k) - \eta(\partial E/\partial w)\big|_{w=w(k)}$$

式中,$k=0,1,2,\cdots$,为迭代次数;η为迭代步长;$\Delta w = -\eta(\partial E/\partial w)\big|_{w=w(k)}$为第$k$次迭代时的权值修正量;$\partial E/\partial w$代表误差函数$E$对权值$w$的偏导。但是,按照这种标准的BP算法调整权值,其调整过程可能存在着收敛速度慢和易

收稿日期:2008-03-12; **修回日期**:2009-04-19。
基金项目:国家自然科学基金(60643004,60775050)资助课题
作者简介:张雨浓(1973-),男,教授,博导,博士,主要研究方向为神经网络理论及应用。E-mail:zhynong@mail.sysu.edu.cn

陷入局部极小点等缺点。为了改善网络的性能,本文从网络结构和激励函数的角度出发,基于样条插值和函数逼近理论[8-9],构造了一种三次样条神经网络(如图 1 所示)。在分析该样条神经网络的逼近能力之前,我们可以首先给出样条函数的基本定义。

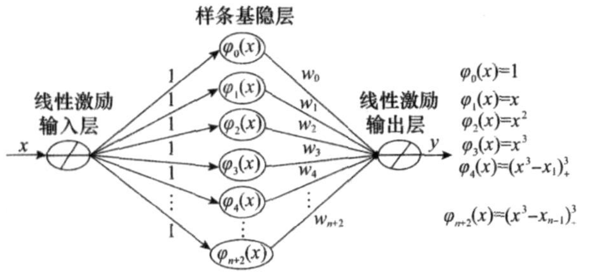

图 1　构造的三次样条神经网络模型

定义 1[10]　设 $\Phi(x)$ 是定义在区间 $[a,b]$ 上的函数,在 $[a,b]$ 上有一个划分

$$\Delta: a = x_0 < x_1 < \cdots < x_n = b \tag{1}$$

$\Phi(x)$ 满足如下条件:

(1) $\Phi(x)$ 在每个子区间 $\Delta_i = [x_{i-1}, x_i]\ (i = 1,2,\cdots,n)$ 上都是不超过 r 次的多项式,至少在一个子区间上为 r 次的多项式;

(2) $\Phi(x)$ 是 $r-1$ 次连续,即 $\Phi(x) \in C^{r-1}[a,b]$ 则称 $\Phi(x)$ 是关于划分 Δ 的一个 r 次样条函数。

在此,本文主要讨论 $r = 3$ 的情形。设该三次样条函数 $\Phi(x)$ 在节点 x_j 上还满足如下关于未知函数 $\phi(x)$ 的插值条件

$$\Phi(x_i) = \phi(x_i),\ i = 0,1,\cdots,n$$

则其截断幂函数为

$$x_+^r = \begin{cases} 0, & x < 0 \\ x^r, & x \geq 0 \end{cases}$$

式中,r 是正整数,并可以验证 x_+^r 具有直至 $r-1$ 阶的连续导数。

引理 1[10]　函数系 $1, x, x^2, x^3, (x-x_1)_+^3, \cdots, (x-x_{n-1})_+^3$ 在 $[a,b]$ 上线性无关。

定理 1[10]　设划分 Δ 由式(1)给出,则任一关于划分 Δ 的三次样条函数 $\Phi(x)$ 可以唯一地表示

$$\Phi(x) = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + \sum_{i}^{n-1} w_i (x - x_i)_+^3,\ a \leq x \leq b \tag{2}$$

图 1 给出了三次样条神经网络的模型,它由输入层、隐层和输出层三层构成。输入层和输出层采用线性激励函数 $f(x) = x$,且可以将输入层与隐层之间的权值固定设为 1,并固定所有神经元的阈值为 0。该神经网络隐层有 $n+3$ 个神经元并使用如下 $\varphi(x), i = 0,1,2,\cdots,n+2$ 作为第 $i+1$ 个神经元的激励函数

$$\begin{cases} \varphi_0(x) = 1 \\ \varphi_1(x) = x \\ \varphi_2(x) = x^2 \\ \varphi_3(x) = x^3 \\ \varphi_4(x) = (x - x_1)_+^3 \\ \cdots \\ \varphi_{i+2}(x) = (x - x_{i-1})_+^3 \end{cases}$$
$$i = 2,3,\cdots,n$$

值得指出的是上述三次样条神经网络仍可以看成是一种 BP 前向神经网络,并能通过误差回传算法来迭代修正权值 $\mathrm{w} = [w_0, w_1, w_2, \cdots, w_{n+2}]^{\mathrm{T}}$。也值得注意的是,隐神经元的个数与输入的样本个数是显性相关的;即,当输入的样本个数是 $n+1$ 时,可以调整 $n+3$ 个权值系数 $w_i\ (i = 0,1,\cdots,n+2)$ 从而达到函数逼近要求[10-11]。另外,将输入层与隐层之间的连接权值固定设为 1,三层的网络结构某种意义上也可理解为两层(考虑到只有后两层之间的连接权值需要学习调整);但就网络的结构形式和信息传递方式而言,该样条网络仍然呈现出输入层、隐层和输出层的三层网络结构,且输入信息在正向传递的过程中,依然经过了隐层神经元样条激励函数的变换,而后在输出层神经元处累加求和。

另外,当该三次样条神经网络应用于辨识实际问题中的未知函数关系式 $y = \phi(x)$ 时,可以在某区间 $[a,b] \subseteq R$ 上得到一系列输入数据点 x_i 及其对应的函数值 $y_i = \phi(x_i)$,从而组成训练样本集 $\{(x_i, y_i), i = 0,1,2,\cdots,n\}$,如表 1 所示。

表 1　未知函数关系 $y = \phi(x)$ 的给定样本数据表

x	x_0	x_1	x_2	$\cdots$	x_{n-1}	x_n
y	y_0	y_1	y_2	$\cdots$	y_{n-1}	y_n

下面给出该三次样条神经网络模型逼近能力的更多理论基础。依据样条/插值逼近理论[8,10],可以构造一个多项式函数 $\Phi(x)$ 来逼近未知函数 $y = \phi(x)$,其相关定义和定理可以整理如下。

定义 2[10,12]　假设 $\phi(x), \varphi_j(x) \in C[a,b], j = 0,1,\cdots,n+2$,且 $\{\varphi_j(x)\}_{j=0}^{n+2}$ 是区间 $[a,b]$ 上的一个线性无关函数系。确定样条多项式 $\Phi(x) = \sum_{j=0}^{n+2} w_j \varphi_j(x)$ 的权值系数 $w_0, w_1, \cdots, w_{n+2}$ 使得 $\int (\phi(x) - \Phi(x))^2 \mathrm{d}x$ 最小。这样得到的样条函数 $\Phi(x)$ 被称为 $\phi(x)$ 在 $[a,b]$ 上的最佳平方逼近(样条函数)或最小二乘逼近(样条函数)。

定理 2[10,12]　设 $\phi(x) \in C[a,b]$,则定义 2 中所述及的 $\phi(x)$ 的最佳平方逼近样条函数 $\Phi(x)$ 存在而且唯一,且其 $w_0, w_1, \cdots, w_{n+2}$ 系数可以求解得到。

显然,把输入层与隐层之间的权值固定为 1,图 1 所示的三次样条神经网络的输入输出关系可以直接描述为如下数学公式

$$y = w_0 \varphi_0(x) + w_1 \varphi_1(x) + \cdots + w_{n+2} \varphi_{n+2}(x) \tag{3}$$

也即上述定理 1 和定义 2 所给出的三次样条多项式 $\Phi(x)$,其对未知函数 $\phi(x)$ 的逼近能力因此可由定理 2 所保证。更为重要的是,后文可以在此基础上推导得出本文重点研究的权值直接确定方法,且也已为计算机仿真数据所验证。另外,这种输入层与隐层之间的连接权值固定为 1 的做法也利于简化网络的结构和连接,对未来可能的硬件实现的简化有一定的益处。也值得指出的是,类似的对权值固定为 1 的处理做法实际上在其他经典神经网络模型中也曾出现过(如 RBF 径向基函数神经网络)。

2　权值迭代与直接确定

对于图 1 所给出的三次样条神经网络模型,其隐层神经元个数为 $n+3$,隐神经元与输出层神经元之间的权值记为 $w_j\ (j=0,1,2,\cdots,n+2)$。该神经网络输入 x 与输出 y 的关系因此可以描述成式(3)的形式。如果采用样本集 $\{(x_i,y_i),i=0,1,\cdots,n\}$ 对神经网络进行训练,并定义其总体误差函数 E 如下

$$E=\frac{1}{2}\sum_{i=0}^{n}\left(y_i-\sum_{p=0}^{n+2}w_p\varphi_p(x_i)\right)^2 \tag{4}$$

则图 1 所示神经网络基于标准 BP 算法的权值 w_j 修正公式可如定理 3 所述。

定理 3[13]　(元素迭代法) 图 1 所示的三次样条神经网络的权值迭代公式可设计为

$$w_j(k+1)=w_j(k)-\eta\frac{\partial E}{\partial w_j}\bigg|_{w=w(k)}=$$

$$w_j(k)-\eta\sum_{i=0}^{n}\left(\varphi_j(x_i)\left(\sum_{p=0}^{n+2}w_p(k)\varphi_p(x_i)-y_i\right)\right),$$

$$j=0,1,\cdots,n+2 \tag{5}$$

式中,迭代次数 $k=0,1,2,\cdots$;学习步长 $\eta>0$。

若将上述定理 3 以矩阵和向量形式描述,则有如下定理 4,即更简洁的权值迭代形式。

定理 4[13]　(矩阵迭代法) 图 1 所示的三次样条神经网络的权值迭代式(5)可以简化为如下矩阵向量迭代形式

$$\mathrm{w}(k+1)=\mathrm{w}(k)-\eta\mathrm{X}^{\mathrm{T}}(\mathrm{X}\mathrm{w}(k)-\gamma) \tag{6}$$

其中输入整合矩阵

$$\mathrm{X}=\begin{bmatrix}\varphi_0(x_0) & \varphi_1(x_0) & \cdots & \varphi_{n+2}(x_0)\\ \varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_{n+2}(x_1)\\ \cdots & \cdots & \cdots & \cdots\\ \varphi_0(x_n) & \varphi_1(x_n) & \cdots & \varphi_{n+2}(x_n)\end{bmatrix}\in\mathbf{R}^{(n+1)\times(n+3)}$$

式中,输出向量 $\gamma=[y_0\ \ y_1\ \ \cdots\ \ y_n]^{\mathrm{T}}\in\mathbf{R}^{n+1}$;$k$ 代表迭代次数($k=0,1,2,\cdots$);学习步长 $\eta>0$。

定理 5[13]　(权值直接确定法) 定义 X 和 γ 等参量如同定理 4。图 1 所示的三次样条神经网络的最优权值可以直接给定为

$$\omega=\mathrm{X}^{+}\gamma \tag{7}$$

式中,$\mathrm{X}^{+}=\lim\limits_{\lambda\to 0}(\mathrm{X}^{\mathrm{T}}\mathrm{X}+\lambda\mathrm{I})^{-1}\mathrm{X}^{\mathrm{T}}$,表示输入整合矩阵 X 的伪逆,I 表示一个 $n+3$ 维单位矩阵。式(7)也可以简写为 $\omega=\mathrm{pinv}(\mathrm{X})\gamma$(并可调用 MATLAB 程序 pinv[6,8,13] 一步计算)。基于给出的最优权值式(7),该三次样条神经网络能够实现对输入输出关系 $\Phi(\cdot)$ 的最小平方逼近。

证明　根据定理 4 及权值迭代式(6),即 $\mathrm{w}(k+1)=\mathrm{w}(k)-\eta\mathrm{X}^{\mathrm{T}}(\mathrm{X}\mathrm{w}(k)-\gamma)$,考虑到步长 $\eta>0$,则当网络训练达到稳态后($k\to+\infty$ 时,$\mathrm{w}(k+1)=\mathrm{w}(k)$)有

$$\mathrm{X}^{\mathrm{T}}(\mathrm{X}\omega-\gamma)=0 \tag{8}$$

因此,满足式(8)的 ω 是该三次样条神经网络的稳态最优权值,且能够使网络学习误差 E 最小。如果 $\mathrm{X}^{\mathrm{T}}\mathrm{X}$ 非奇异,稳态权值 $\omega=(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\gamma$;但在本项研究中,鉴于 $\mathrm{X}\in\mathbf{R}^{(n+1)\times(n+3)}$,可以知道 $\mathrm{X}^{\mathrm{T}}\mathrm{X}$ 通常奇异而不直接可逆,实践中可以利用 $\omega=(\mathrm{X}^{\mathrm{T}}\mathrm{X}+\lambda\mathrm{I})^{-1}\mathrm{X}^{\mathrm{T}}\gamma$ 来进行计算[14],其中 $\lambda=10^{-7}$。利用矩阵伪逆思想,由上述分析可得 $\omega=\mathrm{X}^{+}\gamma$,简言之,该三次样条神经网络的最优权值向量可一步直接确定。定理 5 因此得证。

3　仿真分析

以 $\Phi(x)=2\sin(x)/(3x^2+x+1)$ 作为目标函数,对图 1 所示的三次样条神经网络进行计算机仿真验证。在区间 $[-1,0.99]$ 内以 0.01 的间隔进行采样,得到 200 个样本点。以样本数据点 $x_i\ (i=0,1,2,\cdots,199)$ 作为仿真输入,而相应的样本数据值 $y_i=\Phi(x_i)$ 作为目标输出。本文分别用元素迭代法(对应式(5))、矩阵迭代法(对应式(6))和权值直接确定法(对应式(7))对上述目标函数进行样条神经网络逼近验证,计算机仿真所得到的结果如表 2 和图 2 所示。

表 2　采用元素迭代法、矩阵迭代法与权值直接确定法的仿真结果对比

方法	计算过程/次	运行时间/s	总误差 E	均差 E/m	校验均差
元素迭代法	10^5	4.16×10^4	7.63×10^{-2}	3.81×10^{-4}	3.77×10^{-4}
矩阵迭代法	10^5	24.4	7.63×10^{-2}	3.81×10^{-4}	3.77×10^{-4}
权值直接确定	一步计算	0.38	5.70×10^{-10}	2.85×10^{-12}	2.25×10^{-12}

从表 2 可以看出,对于本文所构造的三次样条神经网络而言,采用权值直接确定法,无论在计算速度还是计算精度上,都要远优于上述两种迭代方法。

(1) 计算时间方面:两种迭代法(包括元素迭代法和矩阵迭代法)的网络训练时间分别为 4.16×10^4 s 和 24.4 s,而权值直接确定法仅需要 0.38 s。

(2) 计算精度方面:两种迭代法训练 100 000 次后的均方误差为 3.81×10^{-4},而权值直接确定法直接计算得到的均方误差仅为 2.85×10^{-12}。图 2 中,实线代表目标函数 $\Phi(x)=2\sin(x)/(3x^2+x+1)$ 的实际输出曲线,虚线代表神经网络的仿真输出结果。图 2 的三个子图分别对应于基于三种方法(元素迭代法、矩阵迭代法和权值直接确定法)所得到的网络逼近效果。由图 2 可知,基于迭代法的网络输出在规定区间内学习训练 100 000 次所得的网络输出与目标输出的误差仍较大。相比较而言,基于伪逆的权值直接确定法的网络仿真结果就显然更加理想:所得到的网络输出与目标输出基本上是完全重合的。因此,该三次样条神经网络模型基于伪逆思想的权值直接确定法的情况下表现出更好的函数逼近性能。另外,网络训练完毕后,我们采用一组未经训练的校验样本对该神经网络进行测试。如表 2 的最后一列所显示的,迭代法得到的校验均差为 3.77×10^{-4},而直接确定法为 2.25×10^{-12},可见该三次样条神经网络具有良好的泛化能力[15]。

作为更多的计算机仿真示例的展示,图 3 和图 4 分别给出了针对文献[11]所提到的两个不同的目标函数 $\Phi(x)=(x+2)/(x^2+4x+3)$ 和 $\Phi(x)=(2+x)\cdot\exp(-x^2)+\sin\log(5+x)$ 进行学习与逼近而得出的计算机仿真结果,其同样显示出了样条神经网络及其权值直接确定法的可行性和有效性。

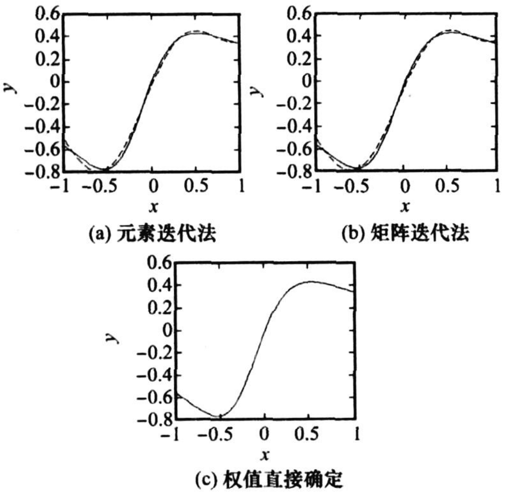

图2 基于三种学习方法的网络逼近效果(对应于目标函数 $\phi(x)=2\sin(x)/(3x^2+x+1)$)

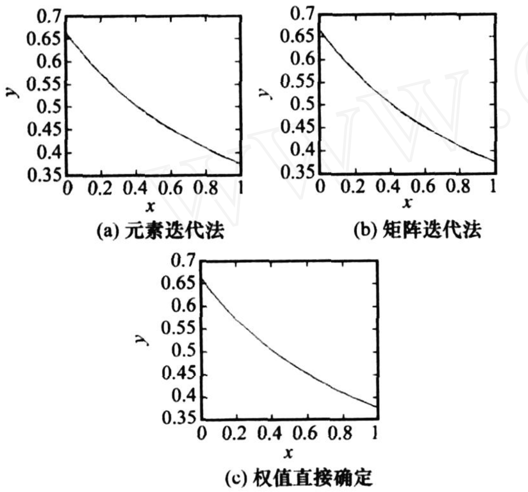

图3 基于三种学习方法的网络逼近效果(对应于目标函数 $\phi(x)=(x+2)/(x^2+4x+3)$)

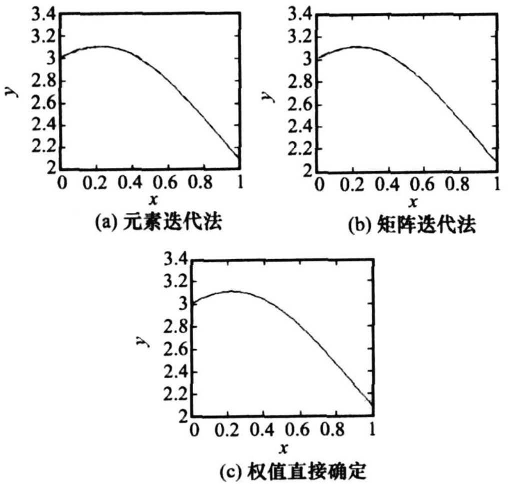

图4 基于三种学习方法的网络逼近效果(对应于目标函数 $\phi(x)=(2+x)\exp(-x^2)+\sin\log(5+x)$)

4 结束语

标准BP神经网络存在着诸如收敛速度慢和易陷入局部极小点等缺点,人们在对BP网络进行改进时大多是局限于神经网络的学习算法,而网络结构和激励函数的优化同样也可以提高神经网络的学习性能。本文利用样条插值逼近的思想,构造了一种样条神经网络模型,给出了该神经网络模型基于误差回传思想的权值迭代公式,同时针对该特定网络结构提出了一种基于伪逆的权值直接确定方法。仿真结果表明,本文构造的基于伪逆的权值直接确定法相比权值迭代调整的方法具有更快的运算速度和更高的计算精度,能够一步计算出正确的稳态最优权值,从而避免了冗长的权值迭代训练过程。

也值得指出的是,本文探讨的是一种单输入单输出前向样条神经网络模型(其中输入输出神经元个数均设置为1),且结合矩阵伪逆思想提出了一种可直接确定网络连接权值的方法;多输入多输出神经网络(如RBF神经网络)权值直接确定法的研究已经开展,因此输入输出神经元个数最终而言将不会有理论上的限制(当然这方面的研究工作还有待进一步的完善及更多计算机仿真验证)。

参考文献:

[1] 张雨浓,徐小文,毛宗源. Java语言与人工神经网络应用[J]. 暨南大学学报(自然科学版),1998,19(1):108-112.

[2] Zhang Y N, Wang J. Recurrent neural networks for nonlinear output regulation[J]. *Automatica*, 2001, 37(8): 1161-1173.

[3] Zhang Y N, Ge S S, LEE T H. A unified quadratic programming based dynamical system approach to joint torque optimization of physically constrained redundant manipulators[J]. *IEEE Trans. on Systems, Man, and Cybernetics*, 2004, 34(5): 2126-2132.

[4] Zhang Y N, Wang J. Obstacle avoidance for kinematically redundant manipulators using a dual neural network[J]. *IEEE Trans. on Systems, Man, and Cybernetics*, 2004, 34(1): 752-759.

[5] 苏小红,王亚东,马培军. 基于反馈调控参数的BP学习算法研究[J]. 哈尔滨工业大学学报,2005,37(10):1311-1314.

[6] 蒲春,孙政顺,赵世敏. Matlab神经网络工具箱BP算法比较[J]. 计算机仿真,2006,23(5):142-144.

[7] 高雪鹏,丛爽. BP网络改进算法的性能对比研究[J]. 控制与决策,2001,16(2):167-171.

[8] Mathews J H, Fink K D. Numerical methods using MATLAB[M]. *Beijing: Pearson Education Inc.*, 2004.

[9] 莫国端,刘开第. 函数逼近论方法[M]. 北京:科学出版社,2003.

[10] 吴勃英,王德明. 数值分析原理[M]. 北京:科学出版社,2003.

[11] 袁晓红. 一组样条神经网络建模[J]. 长沙电力学院学报(自然科学版),2002.

[12] 林成森. 数值分析[M]. 北京:科学出版社,2007.

[13] 张雨浓,李巍,刘巍,等. 幂激励前向神经网络的权值直接确定法[C]//2007全国模式识别学术会议论文集,2007:72-77.

[14] Wang J. Recurrent neural networks for computing pseudoinverses of rank-deficient matrices[J]. *SIAM Journal on Scientific Computing*, 1997, 18(5): 1479-1493.

[15] 盛守照,姜斌. 前向神经网络泛化问题研究[J]. 系统工程与电子技术,2006,28(9):1388-1390. (Sheng Shouzhao, Jiang Bin. Research on generalization capability feedforward neural networks[J]. *Systems Engineering and Electronics*, 2006, 28(9): 1388-1390.)