

小波神经网络的构造及其算法的鲁棒性分析

Construction of Wavelet Neural Networks and its Robustness Analysis of Learning Algorithm

张邦礼

Zhang Bangli

(重庆大学信息工程学院, 重庆, 630044)

TP. 8

李银国

Li Yingguo

(重庆建筑高等专科学校)

曹长修

Cao Changxiu

(重庆大学信息工程学院)

摘要 研究了用于非线性函数逼近的小波神经网络的结构设计方法;证明了在存在噪声干扰及网络设计误差的情况下,网络训练过程具有指数收敛性和鲁棒平稳性。

关键词 神经网络;小波变换;收敛性;鲁棒性

中国图书资料分类法分类号

ABSTRACT In this paper, a method of constructing the wavelet neural networks for nonlinear functional approximation is discussed. The exponential convergence of the training process and its robust stability to the noise perturbances and the network design errors are also proved.

KEYWORDS neural networks; wavelet transform; convergence; robustness

0 引言

由于前馈型人工神经网络(包括BP网络、RPFN网络等)具有极强的非线性逼近功能,所以它在非线性系统辨识、控制、信号处理等领域有着重要的应用^[1],同时也因其在网络结构设计、学习速度、局部最优、收敛性证明及抗噪能力等方面遇到的困难而受到众多的批评^[2]. Y. H. Pao曾提出了一种性能较优的函数基网络(简记FBPN)^[3],其神经元激活函数(activation function),即基函数可分别选择相当广泛的函数类,如sigmoid函数、Gauss函数、样条函数等。如图1所示的单隐层结构,它十分有利于进行学习算法设计和分析,而且能证明:当隐层神经元数充分大时,FBPN网络能以任意精度一致逼近于紧集 $A \subset R^N$ 中的 Lebesgue 可测函数^[3]。

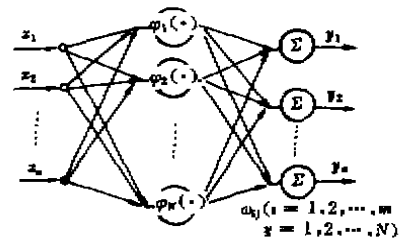


图1 FBPN网络基本结构

* 收文日期 1995-04-24
· 国家教委博士点基金项目

但是 FBPN 网络仍然解决的只是函数最优逼近的存在性问题。在实际中,如何根据学习样本数据确定隐层神经元数 N ? 如何确定基函数 $q_i(\cdot)$? 这些仍缺乏充分的依据。

小波函数具有良好的时(空)-频域的定位特性。选取适当的母小波(mother wavelet) $g(x)$, 并通过离散小波变换而获得一系列体现了原函数(信号)谱局部特性的小波基 $\{g_{m,n}(x)\}$ 。本文中研究了如何在一定精度要求下,从 $\{g_{m,n}(x)\}$ 中选取 N 个小波函数作为网络的基函数,从而得到拓扑结构同 FBPN 相似的小波神经网络;并且证明了该小波网络按梯度法修正权值,学习训练过程具有指数收敛性和抗噪声干扰的鲁棒性。

1 小波神经网络的构造

设待辨识的函数 $f: A \subset R^s \rightarrow R^n, f \in L^2(R^s), A$ 为有界紧子集。为了讨论方便,不妨设 $m = n = 1$ (即 SISO 系统)。

1.1 Hilbert 空间的标架 $\{h_n\}$ 及函数展开

定义 1 Hilbert 空间 H 中的向量序列 $\{h_n\}_{n=1}^{\infty}$, 称为标架(Frame) 如果存在常数 $A > 0, B < +\infty$, 使得

$$A \|f\|^2 \leq \sum_{n=1}^{\infty} |\langle f, h_n \rangle|^2 \leq B \|f\|^2 \quad (1)$$

式中 $f \in H, A, B$ 称为标架的界。由此得到的 H 空间上的算子 $T: H \rightarrow H$

$$Tf = \sum_{n=1}^{\infty} \langle f, h_n \rangle h_n \quad (2)$$

称为标架算子。

不难看出, T 为 H 中的有界线性可逆算子, $\forall f \in H, f$ 可展开为^[4]:

$$f = \sum_{n=1}^{\infty} \langle f, T^{-1}h_n \rangle h_n \quad (3)$$

式中 T^{-1} 为 T 的逆算子。特别地, 当 $A = B = 1$ 时, $\{h_n\}$ 为 H 中的标准正交基, 此时

$$f = \sum_{n=1}^{\infty} \langle f, h_n \rangle h_n \quad (4)$$

1.2 离散仿射小波变换及 $L^2(R)$ 中标架的构造

由于一般的标架并不要求 $\{h_n\}$ 正交, 所以使我们可以付里叶展开, Hear 基展开之外, 寻找更广泛、更方便的展开方法。小波基展开(wavelet expansion) 可以很好地解决 $L^2(R)$ 中函数表示问题。

在 Hilbert 空间 $L^2(R)$ 中选取 $g(x)$ 作为母波, $g(x)$ 应满足相容性条件:

$$\int_{-\infty}^{\infty} \frac{|G(w)|}{|w|} dw < +\infty \quad (5)$$

式中 $G(w)$ 为 $g(x)$ 的付里叶变换。对 $g(x)$ 作仿射离散小波变换, 得函数列 $\{g_{m,n}(x)\}$

$$g_{m,n}(x) = a_0^{-\frac{m}{2}} g(a_0^{-m} \cdot x - b_0 n) \quad (6)$$

这样, 适当选择伸缩尺度 $a_0 > 0$ 和平移尺度 $b_0 > 0$, $\{g_{m,n}(x)\}$ 便构成 $L^2(R)$ 的仿射标架(affine

frame)^[5]. f 的小波基展开为

$$f(x) = \sum_{m,n} c_{m,n}(f) \cdot g_{m,n}(x) \quad (7)$$

式中 $c_{m,n}(f) = \langle f, T^{-1}g_m \rangle$ 为待定的展开系数。

1.3 母波函数 $g(x)$ 的确定

满足相容条件(5)的函数很多,如 Gaussian 函数,墨西哥帽(Mexican hat)函数等。由于 sigmoid 函数模拟较方便,故可采用如下的由三个 sigmoid 函数组合而成的母波函数

$$g(x) = s(x+2) - 2s(x) + s(x-2) \quad (8)$$

$s(x) = 1/(1 + \exp(-\lambda x))$, $(\lambda > 0)$ 。可以验证, $g(x)$ 符合相容性条件,而且选择 $a_0 = 2, b_0 = 1$ 时, $\{g_{m,n}(x)\}$ 便是一个 $L^2(\mathbb{R})$ 中的仿射标架^[6]。

对 $g(x)$ 作付氏变换可得

$$G(w) = \int_{\mathbb{R}} g(x) \cdot e^{-jwx} dx = -j \frac{4\pi \sin^2 w}{\lambda \cdot \operatorname{sh}(\pi w/\lambda)} \quad (9)$$

若取坡度参数 $\lambda = 2$, 经过推导、计算,可得到下列各量:

$|g(x)|^2$ 在时(空)域的中心(能量分布中心):

$$x_0 = \frac{1}{\|g\|^2} \int_{\mathbb{R}} x |g(x)|^2 dx = 0$$

$|G(w)|^2$ 在 $w > 0$ 上的分布中心:

$$w_0 = \frac{1}{\int_{(0,+\infty)} w |G(w)|^2 dw} \int_{(0,+\infty)} w |G(w)|^2 dw = 0.94$$

$|g(x)|^2$ 关于中心 $x_0 = 0$ 的紧支撑为:

$$\begin{aligned} \operatorname{supp}(g, \varepsilon) &= \{(-T, T) \mid \int_{x \in \mathbb{R} \setminus (-T, T)} |g(x)|^2 dx < \varepsilon \|g\|^2\} \\ &= [-2.15, 2.15] \quad (\varepsilon = 0.1) \end{aligned}$$

$|G(w)|^2$ 在 $w > 0$ 上关于中心 w_0 的紧支撑为:

$$\begin{aligned} \operatorname{supp}(G, \varepsilon) &= \{(-\Omega_0, \Omega_1) \mid \int_{\substack{x \in \mathbb{R} \setminus [-\Omega_0, \Omega_1] \\ |\Omega_1 - \Omega_0| = |\Omega_0 - w_0|}} |G(x)|^2 dx < \varepsilon \|g\|^2\} \\ &= [0.29, 1.59] \quad (\varepsilon = 0.1) \end{aligned}$$

从而得到时(空)-频域上的两个矩形域:

$$Q_{0,0} = [-T, T] \times [-\Omega_1, -\Omega_0] \cup [-T, T] \times [\Omega_0, \Omega_1] \quad (10)$$

根据付里叶变换性质,可得到小波基函数 $g_{m,n}(x)$ 的时-频率集中区域(Time-frequency concentration):

$$\begin{aligned} Q_{m,n} &= [a_0^m[-T + nb_0], a_0^m(T - nb_0)] \times \{-a_0^m \Omega_1, -a_0^m \Omega_0\} \\ &\quad \cup [a_0^m \Omega_0, a_0^m \Omega_1] \end{aligned} \quad (11)$$

1.4 小波网络基函数的选择

设被辨识函数(信号) $f(x)$ 的时(空)-频域的支撑为: $Q_c(f) = [x_{\min}, x_{\max}] \times \{[w_{\min}, w_{\max}] \cup [-w_{\max}, -w_{\min}]\}$,即 Q_c 表示它在误差范围 ε 之内覆盖了原函数(信号)的时-频域中最集中的部分.在实际应用中, $Q_c(f)$ 是通过将样本数据作频率带宽(bandwidth)估计而得到的.

从标架 $\{g_{m,n}(x)\}$ 中选取有限个函数,作为小波网络基函数,得

$$\hat{f}(x) = \sum_{m,n \in B_i} c_{m,n}(f) \cdot g_{m,n}(x) \tag{12}$$

其中,下标集 B_i 的选取原则:使选中的 $g_{m,n}(x)$ 所对应的 $Q_{m,n}$ 覆盖或部分覆盖 $Q_c(f)$,即

$$B_i = \{(m,n) | Q_{m,n} \cap Q_c(f) \neq \emptyset, m,n \in Z\} \tag{13}$$

Daubechies^[6]已证明:存在 B_i 以及 $c_{m,n}(f)$ 使(12)式以任意精度逼近于 $f(x)$.到此,我们便完成了小波网络结构的设计,其中 $g_{m,n}(x)$ 便是隐层神经元激活函数.

设 B_i 中共有 N 个元素,对 $g_{m,n}(x) ((m,n) \in B_i)$ 重新编号,记为 $\varphi_1(x), \varphi_2(x), \dots, \varphi_N(x)$.这相当于取小波函数作为基的FBPN网络.上述讨论,主要针对SISO系统,实际上,小波分析过程也适合于MIMO系统^[7].一般地,我们得到如下结构的小波网络:

$$\begin{aligned} \hat{y}_i &= \hat{f}_i(x) = \sum_{j=1}^N w_{ij} \varphi_j(x) \\ \hat{\mathbf{y}} &= \hat{\mathbf{f}}(x) = W\Phi(x) \end{aligned} \tag{14}$$

式中, $x \in A \subset R^n, \hat{\mathbf{y}} \in R^m$

$$\Phi(x) = (\varphi_1(x), \dots, \varphi_N(x))^T$$

$$W = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} & \dots & w_{2N} \\ \dots & \dots & \dots & \dots \\ w_{m1} & w_{m2} & \dots & w_{mN} \end{pmatrix}$$

2 小波网络学习算法的收敛性和鲁棒性

设真实函数(信号)为

$$\bar{y} = \bar{f}(x) \quad (x \in A \subset R^n, \bar{y} \in R^m) \tag{15}$$

$\bar{f} \in L^2(R^n)$,在噪声扰动下,实际得到的样本为:

$$y_i = \bar{f}(x_i) + v_i \quad (i = 1, 2, \dots) \tag{16}$$

由此得样本集 $J = \{x_i, y_i | i = 1, 2, \dots\}$.按上节的构造方法,可得到小波基网络

$$\hat{\mathbf{y}} = \hat{\mathbf{f}}(x) = W\Phi(x)$$

及输出

$$\hat{y}_i = \hat{f}_i(x_i) = W\Phi(x_i) \tag{17}$$

这样便出现下述偏差:1) 系统噪声偏差 $\bar{e}_i = y_i - \bar{y}_i$; 2) 样本拟合偏差 $e_i = y_i - \hat{y}_i$; 3) 真实拟合偏差 $\bar{e}_i = \bar{y}_i - \hat{y}_i$.学习样本系列,分批地输入网络进行学习,设每批取 P 个样本,记第 k 批第 p 个样本为 $(x_i^{(k)}, y_i^{(k)}) (p = 1, 2, \dots, P, k = 1, 2, \dots)$ 相应的偏差便是 $\bar{e}_i^{(k)}, e_i^{(k)}$ 和 $\bar{e}_i^{(k)}$,现在的问

题是:网络(17)在最小均方差型目标函数

$$\begin{aligned} E_{L\delta}(k) &= \frac{1}{P} \sum_{j=1}^P \|e_j^{(k)}\|^2 \\ &= \frac{1}{P} \sum_{j=1}^P \|y_j^{(k)} - W(k) \cdot \Phi(x_j^{(k)})\|^2 \end{aligned} \quad (18)$$

下学习训练,权矩阵 $W(k)$ 是否收敛?真实拟合偏差 $\tilde{e}_j^{(k)}$ 是否能收敛于零或零点较小的邻域?

在下面的讨论中,取向量范数为 $\|x\| = \left\{ \sum_{i=1}^n x_i^2 \right\}^{1/2}$,矩阵 A 的特征值降序排列为 $\lambda_1[A] \geq \lambda_2[A] \geq \dots \geq \lambda_n[A]$,取和 $\|x\|$ 相容的矩阵范数为 $\|A\| = (\lambda_1[A^T A])^{1/2}$,不难证明,它和 Frobenius 范数 $\|A\|_F = \left(\sum_{i,j} a_{ij}^2 \right)^{1/2}$ 等价。

设 $\sup_j \|\tilde{e}_j\| = \sup_j \|v_j\| < \varepsilon_f$ (19)

(ε_f 由对噪声作估计而得到),又根据小波逼近原理^[5]知,存在最优权矩阵 W^* ,使得

$$\hat{y}_j^* = W^* \Phi(x_j) \quad (20)$$

且

$$\sup_j \|y_j - \hat{y}_j^*\| < \varepsilon_i \quad (21)$$

(ε_i 为网络设计偏差,其估计见^[5]),对网络采用梯度法进行迭代训练,不难得到如下的权矩阵修正的 δ -规则

$$\begin{aligned} W(k+1) &= W(k) + \frac{a_k}{P} \sum_{j=1}^P e_j^{(k)} \cdot \Phi^T(x_j^{(k)}) \\ &= W(k) - \frac{a_k}{P} \sum_{j=1}^P W(k) \cdot \Phi(x_j^{(k)}) \cdot \Phi^T(x_j^{(k)}) + \frac{a_k}{P} \sum_{j=1}^P y_j^{(k)} \cdot \Phi^T(x_j^{(k)}) \end{aligned} \quad (22)$$

记矩阵

$$\begin{aligned} \Phi^{(k)} &= \frac{1}{P} \sum_{j=1}^P \Phi(x_j^{(k)}) \cdot \Phi^T(x_j^{(k)}) \\ A(k) &= I - a_k \Phi^{(k)} \\ B(k) &= \frac{a_k}{P} \sum_{j=1}^P \Phi(x_j^{(k)}) (y_j^{(k)} - \hat{y}_j^{*(k)})^T \\ \bar{W}(k) &= W(k) - W^* \end{aligned}$$

由(20)、(22)式可得

$$\bar{W}^T(k+1) = A(k) \cdot \bar{W}(k) + B(k) \quad (23)$$

定义1 对于 R^n 中向量序列 $\{\Phi(x_i)\}$,若存在正整数 $S \geq 1$ 及 $\rho > 0$,使得

$$\sum_{j=1}^S \Phi(x_{i-j}) \cdot \Phi^T(x_{i-j}) > \rho I \quad (24)$$

对于所有的 $k = 0, 1, 2, \dots$ 均成立,则称序列 $\{\Phi(x_i)\}$ 具有持续激励性(persistently exciting,简称 PE 性),其中最小正整数 $S = S_0$ 称为激励度(degree for excitation)。

引理1 取输入样本序列 $x_i \in A$,得网络隐层输出回归向量(regression vector) $\Phi(x_i) \in$

R^N , 若 A 为有界紧集, $\{\Phi(x_i)\}$ 满足 PE 性条件, 激励度为 S_0 , 每批学习样本数 $P \geq S_0$. 则

- 1) 存在 $B > 0$, 使 $\|\Phi(x_i)\| \leq B \quad (i = 1, 2, \dots)$
- 2) $\Phi^{(k)}$ 的最大特征值 $\lambda_{\max}[\Phi^{(k)}] = \lambda_{\max}^{(k)} \leq B^2$
- 3) $\Phi^{(k)}$ 的最小特征值 $\lambda_{\min}[\Phi^{(k)}] = \lambda_{\min}^{(k)} > \frac{\rho}{P} > 0$.

证明 1) 由 A 的紧性及小波函数基 $\Phi(x)$ 的连续性直接推出(略)。

2) 记 $\Phi_p^{(k)} = \Phi(x_p^{(k)}) \cdot \Phi^T(x_p^{(k)})$, 由于 $\Phi_p^{(k)}$ 非负定, 且 $\text{rank}(\Phi_p^{(k)}) = 1$, 所以, 只有唯一正惯性指数 $\lambda_1[\Phi_p^{(k)}] > 0$, 而 $\lambda_2[\Phi_p^{(k)}] = \dots = \lambda_v[\Phi_p^{(k)}] = 0$.

由此得

$$\lambda_1[\Phi_p^{(k)}] = \text{tr}[\Phi_p^{(k)}] = \|\Phi(x_p^{(k)})\|^2$$

根据特征值不等式(weyl)得:

$$\begin{aligned} \lambda_{\max}^{(k)} &= \lambda_1[\Phi^{(k)}] = \lambda_1\left[\frac{1}{P} \sum_{p=1}^P \Phi(x_p^{(k)}) \cdot \Phi^T(x_p^{(k)})\right] \\ &= \lambda_1\left[\frac{1}{P} \sum_{p=1}^P \Phi_p^{(k)}\right] \leq \frac{1}{P} \sum_{p=1}^P \lambda_1[\Phi_p^{(k)}] \\ &= \frac{1}{P} \sum_{p=1}^P \|\Phi(x_p^{(k)})\|^2 \leq B^2 \end{aligned}$$

3) 因为 $P \geq S_0$, 所以

$$\begin{aligned} \lambda_{\min}^{(k)} &= \lambda_{\min}[\Phi^{(k)}] = \lambda_{\min}\left[\frac{1}{P} \sum_{p=1}^P \Phi(x_p^{(k)}) \cdot \Phi^T(x_p^{(k)})\right] \\ &\geq \frac{1}{P} \lambda_{\min}\left[\sum_{p=1}^{S_0} \Phi(x_p^{(k)}) \cdot \Phi^T(x_p^{(k)})\right] \\ &= \frac{1}{P} \lambda_{\min}\left[\sum_{p=1}^{S_0} \Phi(x_{kp+s}) \cdot \Phi^T(x_{kp+s})\right] \end{aligned}$$

根据 PE 性条件(24) 式得

$$\lambda_{\min}^{(k)} > \frac{1}{P} \lambda_{\min}[\rho I] = \frac{\rho}{P} > 0 \quad \text{证毕}$$

定理 1 在引理 1 条件之下, 再取学习率 $\alpha_k = \eta_k / B^2$, 其中 $0 < \eta_{\min} \leq \eta_k \leq \eta_{\max} < 2$, 则按规则(22) 式训练小波网络, 能保证:

1) 网络权矩阵 $W(k)$ 满足

$$\|W(k) - W^*\| < r^k \|W(0) - W^*\| + \frac{1-r^k}{1-r} \cdot \frac{\eta_{\max}}{B} \cdot \varepsilon_j \quad (25)$$

2) 真实拟合偏差 $\bar{e}_i^{(k)} = \bar{y}_i^{(k)} - \hat{y}_i^{(k)}$ 满足

$$\|\bar{e}_i^{(k)}\| < a \cdot r^k + b\varepsilon_i + \varepsilon_j \quad (26)$$

其中, $0 < r < 1$, 且 $r = \max\left\{\left|1 - \frac{\eta_{\min}}{B^2 P} \rho\right|, |1 - \eta_{\max}|\right\}$

$$a = B \cdot \|W(0) - W^*\| - \frac{1}{1-r} \eta_{\max} \cdot \varepsilon_i$$

$$b = 1 + \frac{1}{1-r} \eta_{\max}$$

证明 1) 因为 $A^{(k)} = I - \alpha_k \cdot \Phi^{(k)}$, 所以 $\lambda[A(k)] = 1 - \alpha_k \cdot \lambda_1[\Phi^{(k)}] = 1 - \frac{\eta_k}{B^2} \lambda_1[\Phi^{(k)}]$.

由引理 1 得:

$$\begin{aligned}\lambda_n[A(k)] &\leq 1 - \frac{\eta_{\min}}{B^2} \cdot \lambda_n[\Phi^{(k)}] < 1 - \frac{\eta_{\min}}{B^2} \cdot \frac{\rho}{P} < 1 \\ \lambda_n[A(k)] &\geq 1 - \frac{\eta_{\min}}{B^2} \cdot \lambda_n[\Phi^{(k)}] \geq 1 - \frac{\eta_{\min}}{B^2} \cdot \lambda_{\max}^{(k)} \\ &\geq 1 - \frac{\eta_{\min}}{B^2} B^2 = 1 - \eta_{\max} > -1\end{aligned}$$

所以

$$|\lambda_n[A(k)]| \leq \max\left\{\left|1 - \frac{\eta_{\min}}{B^2} \cdot \frac{\rho}{P}\right|, |1 - \eta_{\max}|\right\} = r < 1$$

另一方面, 利用 $P \geq S_0 \geq N \geq 1$ 及 PE 条件, 可以验证 $r > 0$.

$$\|A(k)\| = (\lambda_1[A^T(k) \cdot A(k)])^{1/2} = |\lambda_1[A(k)]| \leq r$$

$$\begin{aligned}\|B(k)\| &= \left\| \frac{\alpha_k}{P} \sum_{r=1}^P \Phi(x_r^{(k)}) (y_r^{(k)} - \hat{y}_r^{(k)}) T \right\| \\ &\leq |\alpha_k| \cdot \frac{1}{P} \sum_{r=1}^P \|y_r^{(k)} - \hat{y}_r^{(k)}\| \cdot \|\Phi(x_r^{(k)})\| \\ &< \frac{\alpha_k}{B^2} \cdot \varepsilon_k \cdot \frac{1}{P} \sum_{r=1}^P B \leq \frac{\eta_{\max}}{B} \cdot \varepsilon_k\end{aligned}$$

解差分方程(23)得:

$$\bar{W}^T(k) = T(k, 0) \bar{W}^T(0) + \sum_{j=0}^{k-1} T(k, j+1) \cdot B(j)$$

其中状态转移矩阵

$$\begin{aligned}T(k, j) &= A(j) \cdot A(j-1) \cdots A(k-1) \\ \|T(k, j)\| &\leq \prod_{i=j}^{k-1} \|A(i)\| \leq r^{k-j}\end{aligned}$$

所以

$$\begin{aligned}\|\bar{W}^T(k)\| &\leq \|T(k, 0)\| \cdot \|\bar{W}^T(0)\| + \sum_{j=0}^{k-1} \|T(k, j+1)\| \cdot \|B(j)\| \\ &< r^k \|\bar{W}^T(0)\| + \sum_{j=0}^{k-1} r^{k-j-1} \cdot \frac{\eta_{\max}}{B} \cdot \varepsilon_k \\ &= r^k \|\bar{W}^T(0)\| + \frac{1-r^k}{1-r} \cdot \frac{\eta_{\max}}{B} \cdot \varepsilon_k\end{aligned}$$

2) 由于 $\tilde{e}_r^{(k)} = \hat{y}_r^{(k)} - \hat{y}_r^{*(k)} = (\tilde{y}_r^{(k)} - y_r^{(k)}) + (y_r^{(k)} - \hat{y}_r^{*(k)}) + (\hat{y}_r^{*(k+1)} - \hat{y}_r^{*(k)}) = (y_r^{(k)} - \hat{y}_r^{*(k)}) - \tilde{e}_r^{(k)} - (W(k) - W^*) \Phi(x_r^{(k)})$.

由(19)、(21)式得

$$\begin{aligned}\|\tilde{e}_r^{(k)}\| &\leq \|y_r^{(k)} - \hat{y}_r^{*(k)}\| + \|\tilde{e}_r^{(k)}\| + \|\bar{W}^T(k)\| \cdot \|\Phi(x_r^{(k)})\| \\ &< \varepsilon_k + \varepsilon_f + (r^k \cdot \|\bar{W}^T(0)\| + \frac{1-r^k}{1-r} \cdot \frac{\eta_{\max}}{B} \cdot \varepsilon_k) B \\ &= (B \|\bar{W}^T(0)\| - \frac{1}{1-r} \eta_{\min} \cdot \varepsilon_k) r^k + \left\{1 + \frac{1}{1-r} \eta_{\max}\right\} \varepsilon_k + \varepsilon_f \\ &= a \cdot r^k + b \cdot \varepsilon_k + \varepsilon_f\end{aligned}$$

证毕

从以上的证明过程可以看出, $\{\Phi(x_i)\}$ 的 PE 性条件对于网络学习过程的收敛性是至关

重要的。因为, 只有选取 $\{x_i\}$, 使得向量序列 $\{\Phi(x_i)\}$ 作为输入信号是“充分丰富”(sufficient rich) 的, 才能将系统的各模态都激励起来, 达到系统辨识的目的。

3 小 结

文中根据小波变换理论和方法, 讨论了确定 FBPN 网络基函数和隐层神经元数的一般方法, 从而构成具有良好逼近性能和自适应性的小波网络。定理 1 从理论上证明了适当选取学习样本, 便可使小波网络及其学习算法具有以下特性:

1) 在理想状态下, 即噪声为零 ($\varepsilon_f = 0$), 且 $\tilde{f}(x)$ 属于网络表示的函数空间 ($\varepsilon_s = 0$), 由 (25) 知

$$\|W(k) - W^*\| < r^k \|W(0) - W^*\| \rightarrow 0$$

即: 对于任何初始权利阵 $W(0)$, 均有 $W(k)$ 依范数指数收敛于 W^* , 其实偏差 $\|\tilde{e}_r^{(k)}\|$ 以指数率收敛于零。

2) 当系统存在噪声且网络结构设计存在偏差时, $\varepsilon_f \neq 0$, $\varepsilon_s \neq 0$ (这更符合应用实际), 则 $\|W(k)\|$ 和 $\|\tilde{e}_r^{(k)}\|$ 仍然能以指数率分别收敛于最优权值 W^* 和原点的一个较小范围, 即

$$\lim_{k \rightarrow +\infty} \|W(k) - W^*\| \leq \frac{1}{1-r} \cdot \frac{\eta_{\max}}{B} \cdot \varepsilon_s$$

$$\lim_{k \rightarrow +\infty} \|\tilde{e}_r^{(k)}\| \leq b \cdot \varepsilon_s + \varepsilon_f$$

这说明在存在噪声干扰和网络设计误差的情况下, 网络训练过程保持平稳, 学习算法具有鲁棒性。

参 考 文 献

- 1 Chu s R et al. Neural networks for system identification. Special section on neural networks for system and control. IEEE control systems magazine. 1990, 31~35
- 2 张铃, 张敏. 神经网络中 BP 算法的分析, 模式识别与人工智能, 1994, 7(3)
- 3 Sadegh N. Nonlinear identification and control via neural networks, in proc, 1991, ASME winter Annu. Meeting, Atlanta, GA, Nov. 1991
- 4 Heil C E, Walnut D F. Continuous and discrete wavelet transforms, SIAM Review. 1989, 31: 628~666
- 5 Daubechies I. The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. on Information theory, 1990, 30(5): 961~1005
- 6 Pati Y C et al. Analysis and synthesis of feedforward neural networks using discrete affine wavelet transforms, IEEE Trans, on neural networks. 1993, 4(1): 73~85
- 7 Mallat S G. A theory for multiresolution signal decomposition. The wavelet representation. IEEE Trans, on pattern analysis and machine intelligence, 1989, 11(7): 674~693