

多子波神经网络的预滤波*

杨新星 焦李成 潘进

(西安电子科技大学雷达信号处理重点实验室, 西安 710071)

摘要 函数的采样值进行预滤波处理后作为多子波神经网络的初始权值, 明显地改善了多子波神经网络的逼近性能, 提高了训练速度.

关键词 多子波 神经网络 逼近性能 初始权值

函数逼近是神经网络所研究的一个重要课题. 网络节点函数的选取, 对于网络的逼近性能和收敛速度起着决定性的作用. 由于以 Sigmoid 函数作为节点函数的网络的初始权值和隐节点个数的选取缺乏有效的理论指导, 对此网络的训练不但速度慢而且经常陷入意想不到的局部极小值. 结构简单的径向基函数网络^[1]虽然使其逼近性能有所改善, 但径向基函数的非正交性和冗余性使它并不能有效地表示所给定的函数. 而正交子波的时频局域性和正交性^[2]使得用正交子波代替网络中 Sigmoid 函数具有以下几个优点^[3,4]: (1) 提供了给定函数的唯一和有效表示, (2) 渐进性能的分析较为方便, (3) 具有全局的逼近性和 L^2 逼近性, (4) 结构简单.

多子波克服了单子波不能同时具有正交性、紧支撑、对称性和高的逼近阶的缺点. 人们在此基础上提出了多子波神经网络^[5,6], 其收敛速度明显优于单子波神经网络^[1]. 然而, 初始化问题是多子波神经网络的基本问题, 初始权值直接影响神经网络的逼近性能和收敛速度. 对此, 我们提出了较为完整的多子波神经网络初始权值优化的预滤波器方法, 分析和实验结果都表明, 这一方法显著地提高了网络性能和加快了网络的逼近速度, 较好地解决了多子波神经网络的初始权值的优化问题.

1 多子波网络模型

对于 N 个紧支撑尺度函数 $\phi^l(t)$ ($l=1, 2, \dots, N$), 已知 $\phi^l(t-k)$, ($k \in \mathbb{Z}$) 相互正交, 并且设

$$\Phi(t) \triangleq (\phi^1(t), \phi^2(t), \dots, \phi^N(t))^T,$$

且 $\Phi(t)$ 满足双尺度方程

$$\Phi(t) = \sqrt{2} \sum_k H_k \Phi(2t-k), \quad (1)$$

其中 H_k 为 $N \times N$ 矩阵.

1999-03-09 收稿, 1999-09-29 收修稿

* 国家自然科学基金(批准号: 69772029)和国家“八六三”高技术资助项目

(1) 潘进, 焦李成. 多子波神经网络及其逼近性. 电子学报, 待发表

$\phi^l(t)$ 的伸缩平移 $\phi_{J,k}^l(t) \triangleq 2^{J/2} \phi^l(2^J t - k)$, $l=1, \dots, N, J, k \in \mathbb{Z}$, 张成尺度空间 V_J . 由多分辨分析可知 $V_J \subseteq V_{J+1}, \bigcup_{J \in \mathbb{Z}} V_J = L^2(\mathbb{R})$. 多子波 $\Psi(t) \triangleq (\varphi^1(t), \varphi^2(t), \dots, \varphi^N(t))^T$ 使得 $\varphi_{J,k}^l \triangleq 2^{J/2} \varphi^l(2^J t - k)$ 张成 V_J 在 V_{J+1} 中的直交补空间 W_J , 并且 $\{\varphi_{J,k}^l, J, k \in \mathbb{Z}\}$ 构成 W_J 的标准正交基. 所以, 对任意的 $f \in L^2(\mathbb{R})$, 有正整数 J_0 , 使得当 $J > J_0$ 时, 总有

$$\|f - f_J\|_2 < \varepsilon, \tag{2}$$

其中 $\|\cdot\|_2$ 表示 L^2 空间范数, ε 是事先给定的任意小的正数, 而

$$f_J = \sum_{l=1}^N \sum_{k \in \mathbb{Z}} \langle f, \phi_{J,k}^l \rangle \phi_{J,k}^l(t). \tag{3}$$

从神经网络的角度看, (3) 式中的 $f_J(t)$ 可以用图 1 所示的网络来学习 (图为 $N=2$ 的情况). 网络的输入层到隐层节点的权值均为 2^J , 输入层是直通的, 前排节点的节点函数均为尺度函数 $\phi^1(t)$, 阈值分别为 $k, k+1, \dots$, 后排节点的节点函数均为尺度函数 $\phi^2(t)$, 阈值分别为 $k, k+1, \dots$. 输出层相应的权值分别为 $\hat{c}_{J,k}^1, \hat{c}_{J,k+1}^1, \dots$ 和 $\hat{c}_{J,k}^2, \hat{c}_{J,k+1}^2, \dots$. 该网络的输入输出表达式为

$$\hat{f}_J = \sum_{l=1}^N \sum_{k \in \mathbb{Z}} \hat{c}_{J,k}^l \phi_{J,k}^l(t), \tag{4}$$

对此网络训练的目的就是使 $\hat{c}_{J,k} \triangleq (\hat{c}_{J,k}^1, \hat{c}_{J,k}^2, J, k \in \mathbb{Z})^T$ 逼近最优值

$$c_{J,k} = \langle f, \Phi_{J,k} \rangle = \int f(t) \Phi_{J,k}(t) dt.$$

从而使函数 $\hat{f}_J(t)$ 逼近目标函数 $f_J(t)$.

对于单子波网络, 由于单尺度函数具有低通特性, 故当 J 足够大时,

$$c_{J,k} \approx 2^{-J/2} f\left(\frac{k}{2^J}\right), \tag{5}$$

因此在单子波神经网络中, 可以把网络的初始权值取为函数采样点的值, 这样就可以把初始权值选到最优值的附近. 但是多尺度函数不具有低通特性, 所以无论 J 取多大, 多子波系数 $c_{J,k}$ 都不可能具有 (5) 式中的形式. 因此, 函数的采样值与理论权值的偏差较大. 本文通过对函数的采样值进行预滤波得到 $c_{J,k}$ 的近似值, 而且这个近似值随着尺度系数 J 的增大逼近 $c_{J,k}$ 的速度显著加快, 因此把这个近似值作为网络的初始权值来进行训练, 会使网络初始误差随着尺度系数 J 的增大而按 2^{-NJ} 衰减. 因此 J 越大, 即网络节点数 N 越多, 越能显示它的优越性和潜力.

2 预滤波器的设计方法

为了设计多子波网络的预滤波器, 我们构造函数

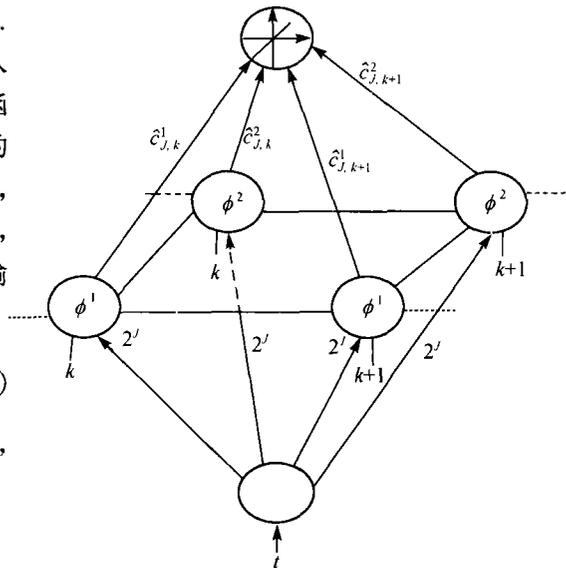


图 1 多子波神经网络

$$g(t) = \sum_n \mathbf{a}^T(n) \Phi(t - n), \quad (6)$$

其中 $\mathbf{a}(n) \triangleq (a_1(n), a_2(n), \dots, a_N(n))^T$ 是实常数矢量. 其频率形式为

$$\hat{g}(\omega) = \mathbf{A}^T(\omega) \hat{\Phi}(\omega), \quad (7)$$

此处 $\mathbf{A}(\omega) = \sum_n \mathbf{a}(n) e^{-jn\omega}$.

满足 Strang-Fix 条件和正交性条件^[7], 因此对于任意函数 $f(t)$, 存在 $J > 0$, 使得 $f(t)$ 能够用 $g_{J,k} = 2^{J/2} g(2^J t - k)$ 线性组合来很好的逼近, 即

$$f(t) \approx \sum_k d_{J,k} g_{J,k}(t),$$

式中

$$d_{J,k} = \int f(t) g_{J,k}(t) dt \propto f\left(\frac{k}{2^J}\right).$$

对于 $f(t) \in V_J$, 有

$$f(t) = \sum_k d_{J,k} g_{J,k}(t),$$

把(6)式代入上式, 得

$$f(t) = 2^{J/2} \sum_m \left(\sum_k d_{J,k} \mathbf{a}^T(m - k) \right) \Phi(2^J t - m).$$

由上式得预滤波表达式

$$\tilde{\mathbf{c}}_{J,m}^T = \sum_k d_{J,k} \mathbf{a}^T(m - k), \quad (8)$$

由(8)式可以看出, 函数的采样值经过预滤波器 $\{\mathbf{a}(n)\}_{n \in \mathbb{Z}}$ 后得到 $\mathbf{c}_{J,m}$ 的近似值 $\tilde{\mathbf{c}}_{J,m}$, 如果把 $\tilde{\mathbf{c}}_{J,m}$ 作为多子波网络的初始权值, 就可以使网络的初始权值直接选取在目标权值的附近, 因而可使网络的收敛速度加快. 下面我们给出收敛速度的表达式.

由于 $\sum_k g(t - k)$ 是周期为 1 的周期函数, 且 $g(t)$ 满足 Strang-Fix 条件, 因此将此信号展开成 Fourier 级数可得

$$\sum_k g(t - k) = 1.$$

所以

$$f(t) - \sum_k f\left(\frac{k}{2^J}\right) g(2^J t - k) = \sum_k \left(f(t) - f\left(\frac{k}{2^J}\right) \right) g(2^J t - k). \quad (9)$$

将函数 $f\left(\frac{k}{2^J}\right)$ 在 t 附近作 Taylor 展开, 则有

$$f\left(\frac{k}{2^J}\right) = f(t) + \sum_{m=1}^{N-1} \frac{f^{(m)}(t)}{m!} \left(\frac{k}{2^J} - t\right)^m + \frac{f^{(N)}(\xi)}{N!} \left(\frac{k}{2^J} - t\right)^N. \quad (10)$$

这里 ξ 在 t 与 $2^{-J}k$ 之间. 将(9)式代入(10)式, 得

$$f(t) - \sum_k f\left(\frac{k}{2^J}\right) g(2^J t - k) = \frac{(-1)^N}{2^{NJ} N!} \sum_k f^{(N)}(\xi) (2^J t - k)^N g(2^J t - k) = O(2^{-NJ}). \quad (11)$$

由(11)式可知, 误差随着尺度系数 J 的增大而按 2^{-NJ} 衰减. 这就说明了多子波预滤波的可行性和有效性. 下面我们给出确定预滤波器 $\{\mathbf{a}(n)\}_{n \in \mathbb{Z}}$ 的一般步骤.

设 $\Phi(t)$ 是具有逼近阶 m 的紧支撑正交尺度函数. 则存在一个满足

$$y_0^r = \sum_{n \in \mathbb{Z}} n^r a(n), \quad (r = 0, 1, \dots, m-1) \quad (12)$$

的有限支撑 $\{a(n)\}_{n \in \mathbb{Z}}$, 使得(6)式中的构造函数 $g(t)$ 满足 Strang-Fix 条件:

$$\hat{g}^{(r)}(0) = 1; \hat{g}^{(r)}(2k\pi) = 0, \quad k \in \mathbb{Z}, k \neq 0, r = 0, 1, \dots, m-1. \quad (13)$$

y_0^0 满足 $(y_0^0)^T \hat{\Phi}(0) = 1$.

因此, 得确定预滤波器 $\{a(n)\}_{n \in \mathbb{Z}}$ 的一般步骤:

- (1) 由文献[6]中的定理 3.2 求解序列 $y_0^r, r = 0, 1, \dots, m-1$;
- (2) 根据正交性条件参数化 $A(\omega)$;
- (3) 根据 $A^{(k)}(0) = \sum_n (-jn)^k a(n) = (-j)^k, y_0^k, (k = 0, 1, \dots, m-1)$ 得方程组;
- (4) 求解方程组, 确定 $A(\omega)$ 的参数.

3 算例结果

下面以 GHM 两子波作为本文的多子波来给出一个算例.

已知矢量 $H(\omega)$ 的冲击响应为

$$\begin{aligned} H_0 &= \begin{bmatrix} 3/10 & 2\sqrt{2}/5 \\ -\sqrt{2}/40 & -3/20 \end{bmatrix}, \\ H_1 &= \begin{bmatrix} 3/10 & 0 \\ 9\sqrt{2}/40 & 1/2 \end{bmatrix}, \\ H_2 &= \begin{bmatrix} 0 & 0 \\ 9\sqrt{2}/40 & -3/20 \end{bmatrix}, \\ H_3 &= \begin{bmatrix} 0 & 0 \\ -\sqrt{2}/40 & 0 \end{bmatrix}. \end{aligned}$$

由此可得

$$H(0) = \begin{bmatrix} 3/5 & 2\sqrt{2}/5 \\ 2\sqrt{2}/5 & 1/5 \end{bmatrix}.$$

由于多子波的正交性, 我们有

$$|\hat{\phi}(0)|^2 + |\hat{\psi}(0)|^2 = 1. \quad (14)$$

由(1)式得

$$\hat{\Phi}(0) = H(0)\hat{\Phi}(0), \quad (15)$$

因为 $H(0)$ 有单特征值 1, 从(14)式和(15)式可求得

$$\hat{\Phi}(0) = y_0^0 = \begin{bmatrix} \sqrt{6} \\ 3 \\ \sqrt{3} \\ 3 \end{bmatrix}.$$

又由于(6)式中构造函数的正交性, 因此, FIR 预滤波器 $A(\omega)$ 可参数化为^[7]

$$\mathbf{A}(\omega) = \mathbf{U}_\rho(\omega) \cdots \mathbf{U}_1(\omega) \mathbf{A}(0), \tag{16}$$

这里

$$\|\mathbf{A}(0)\|^2 = 1, \tag{17}$$

$$\mathbf{U}_r(\omega) = \mathbf{I}_N + (e^{-j\omega} - 1) \mathbf{u}_r \mathbf{u}_r^T, \tag{18}$$

$$\mathbf{u}_r = (u_{r1}, u_{r2}, \dots, u_{rN}) \text{ 且 } \|\mathbf{u}_r\|_2^2 = 1.$$

因此

$$\mathbf{A}(\omega) = \left(\mathbf{I}_2 + (e^{-j\omega} - 1) \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \end{bmatrix} \right) \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix}.$$

这里 θ 是一个角度. 进一步

$$\mathbf{A}(0) = \mathbf{y}_0^0 = \hat{\Phi}(0);$$

$$\mathbf{A}^{(1)}(0) = (-j) \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \end{bmatrix} \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix} = (-j) \mathbf{y}_0^1 = (-j) \begin{bmatrix} \frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} \end{bmatrix}. \tag{19}$$

求解(19)式可得

$$\theta = \arctan(\sqrt{2}).$$

这样我们就得到了预滤波器

$$\mathbf{A}(\omega) = \begin{bmatrix} \frac{\sqrt{6}}{9} \\ -\frac{\sqrt{3}}{9} \end{bmatrix} + \begin{bmatrix} \frac{2\sqrt{6}}{9} \\ \frac{4\sqrt{3}}{9} \end{bmatrix} e^{-j\omega}.$$

图 2 给出了构造函数同时满足 Strang-Fix 条件和正交性条件的频谱与只满足正交性条件的频谱的比较结果,符号 A 表示归一化的幅度. 图 3 表示所要逼近的原函数. 表 1 列出了尺度系数 $J=2 \sim 4$ 时分别用函数的采样值和采样值经过预滤波器的值作为初始权值所得到的初始误差. 表 2 列出了分别收敛到某一较小误差时的迭代次数. 表 3 则列出了分别迭代同样次数后

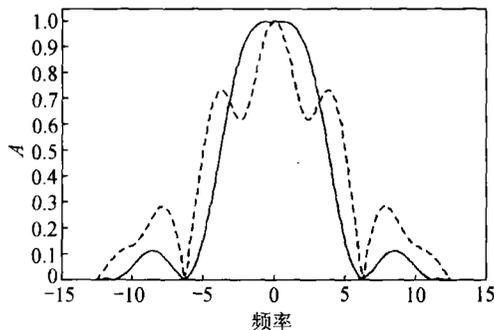


图 2 构造函数的频谱

——表示本文的方法, ---表示文献[7]的方法

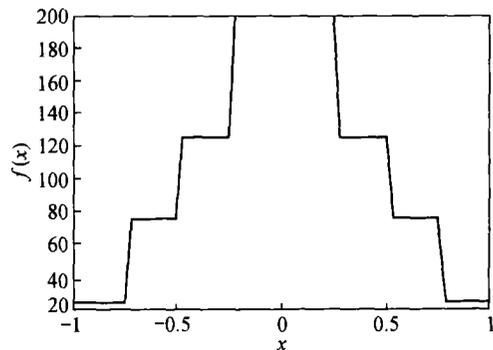


图 3 原函数

表1 两种初始误差的比较

尺度系数	初始误差	
	预滤波后的 $\ f - \hat{f}\ _2$	不预滤波的 $\ f - \hat{f}\ _2$
$J=2$	2.3394×10^2	1.9947×10^3
$J=3$	1.5696×10^2	3.1465×10^3
$J=4$	1.1951×10^2	4.6269×10^3

表2 $J=4$ 两种权值的迭代次数

方法	$\ f - \hat{f}\ _2$			
	$< 1 \times 10^{-3}$	$< 1 \times 10^{-4}$	$< 1 \times 10^{-5}$	$< 1 \times 10^{-10}$
预滤波后	17	21	24	42
不预滤波	29	34	41	90

表3 $J=4$ 两种权值的迭代误差

方法	迭代次数			
	17	21	30	45
预滤波后	7.6288×10^{-4}	5.1648×10^{-5}	1.4445×10^{-7}	9.2818×10^{-12}
不预滤波	1.6053×10^{-1}	1.2448×10^{-2}	3.8895×10^{-4}	2.6211×10^{-6}

的误差. 由表1,表2很明显地看到,采用预滤波后的值作为初始权值所得到的初始误差比直接采用函数采样值作为初始权值所得到的初始误差至少小一个数量级,而且前者的误差随着 J 的增大衰减速度显著加快,从而显著提高了网络的训练速度,也明显地改善了网络的逼近性能.

参 考 文 献

- 1 Poggio T, Girosi F. Networks for approximation and learning. Proc IEEE, 1990, 78: 1 481
- 2 赵松年,熊小芸. 子波变换与子波分析. 北京:电子工业出版社,1997
- 3 Zhang J, Walter G G, Miao Y, et al. Wavelet neural networks for function learning. IEEE Trans on Signal Processing, 1995, 43(6): 1 485
- 4 Pati Y C, Krishnaprasad P S. Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations. IEEE Trans Neural Net., 1993, 4(1): 73
- 5 Cohn, Daubechies I, Plonka G. Regularity of refinable function vectors. J Fourier Anal Appl, 1997, 3: 295
- 6 Plonka G. Approximation order provided by refinable function vectors. Constr Approx, 1997, 13: 221
- 7 Xia X G. A new prefilter design for discrete multiwavelet transforms. IEEE Trans Signal Process, 1998, 46: 1 558