

多子波神经网络及其逼近性

潘 进^{1,2}, 焦李成¹

(1. 西安电子科技大学雷达信号处理国家重点实验室, 西安 710071; 2. 西安通信学院, 西安 710106)

摘 要: 本文提出了多子波神经网络的概念和模型, 证明了其泛逼近性、 L^2 逼近性和相容性, 并给出了相应的收敛速度估计. 理论分析和实验结果表明, 多子波网络的逼近性能明显优于单子波网络. 类似的结果在国内外还未见报导.

关键词: 多子波神经网络; 神经网络; 子波; 多子波

中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112 (2000) 10-0060-05

Multiwavelet Neural Networks and Their Approximation Properties

PAN Jin^{1,2}, JIAO Li-cheng¹

(1. Key Lab for Radar Signal Processing, Xidian University, Xi'an 710071, China;

2. Xi'an Institute of Communications, Xi'an 710106, China)

Abstract: The concept and the model of the multiwavelet neural network are proposed. Its universal and L^2 approximation properties as well as its consistency are proved, and the convergence rates associated with these properties are given. Theoretical analyses and the experimental results show that multiwavelet networks can have obvious advantages in approximation performance over simple wavelet networks, and similar works have not been seen at home and abroad.

Key words: multiwavelet neural network; neural network; wavelet; multiwavelet

1 引言

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. 近年来, 随着子波理论研究的不断深入, 尺度函数的诸多特性以及用不同层级的尺度空间去匹配被逼近信号的思想, 已经在神经网络的研究中有了些成功的应用. 自从 92 年提出子波神经网络以来, 已经有一系列的文章对这一模型进行了研究^[1~4]. 网络权值和隐层节点数目的初始值选取是进行网络训练时的难点, 而子波网络由于有子波理论的指导, 因此可以在训练前对这两个值给出较好的估计. 按尺度空间的增大方式来进行网络训练, 使子波网络有更快的收敛速度. 作为子波概念的推广, 多子波已成为国际上研究的前沿和热点. 多子波可以同时具有正交性、正则性、对称性和紧支撑, 并且其光滑程度可以通过两尺度变换根据问题的要求而提高^[5~8]. 多尺度函数的这些特性在信号处理中是非常有用的. 因此, 用多尺度函数作为节点函数而构成的网络具有许多值得研究的优良性质. 但对于这方面的实质性工作, 目前在国内尚未见报道. 基于以上考虑, 本文提出了多子波神经网络的概念和模型, 研究了其泛逼近性和 L^2 逼近性. 理论分析和实验结果表明, 多子波网络的逼近性能明显优于单子波网络.

2 网络模型

设有多尺度函数 $\phi(t) = [\phi^1(t), \phi^2(t), \dots, \phi^r(t)]^T$ 满足双尺度方程

$$\phi(t) = \sum_{k \in \mathbb{Z}} p_k \phi(2t - k) \quad (1)$$

$\phi(t)$ 的伸缩平移 $\phi_{j,k}^l(t) = 2^{j/2} \phi_{j,k}^l(2^j t - k)$, $l = 1, \dots, r$, $k \in \mathbb{Z}$ 张成了尺度空间 V_j . 考虑正交多尺度函数的情况. 这时, $\{\phi_{j,k}^l | l = 1, \dots, r, k \in \mathbb{Z}\}$ 构成了 V_j 的标准正交基, 并且存在多子波 $\psi(t) = [\psi^1(t), \psi^2(t), \dots, \psi^r(t)]^T$ 使得 $\psi_{j,k}^l(t) = 2^{j/2} \psi^l(2^j t - k)$, $l = 1, \dots, r$, $k \in \mathbb{Z}$ 张成了 V_j 在 V_{j+1} 中的直交补空间 W_j , 并且 $\{\psi_{j,k}^l | l = 1, \dots, r, k \in \mathbb{Z}\}$ 构成了 W_j 的标准正交基. 由多分辨理论可知 $V_j \subseteq V_{j+1}$, 并且 $\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$. 所以, 对于任意的 $f \in L^2(\mathbb{R})$, 有正整数 J_0 , 使得当 $J > J_0$ 时, 总有

$$\|f - f_J\|_2 < \epsilon \quad (2)$$

ϵ 表示 L^2 空间范数, 是事先给定的任意小的正数,

$$f_J = \sum_{l=1}^r \sum_{k \in \mathbb{Z}} f_{j,k} \phi_{j,k}^l(t) \quad (3)$$

从神经网络的角度看, 式 (3) 中的 $f_{j,k}$ 可以用图 1 所示的网络来学习 (图为 $r=2$ 的情况). 由于这一网络来源于多子波理论, 故将这一网络称之为多子波神经网络. 该网络的输入输出

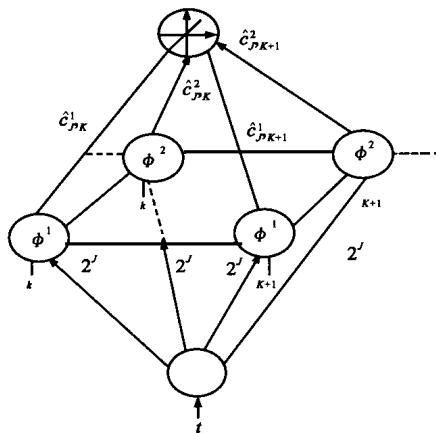


图 1 多子波神经网络

表达式为

$$f_j = \sum_{l=1}^r \sum_{k=z}^z \hat{c}_{j,k}^l \phi_{j,k}^l(t) \quad (4)$$

隐层节点的权值均为 2^j , 前排节点的节点函数均为 $\phi^l(t)$, 阈值分别为 $k, k+1, \dots$, 后排节点的节点函数均为 $\phi^2(t)$, 阈值分别为 $k, k+1, \dots$ 输出层相应的权值分别为 $\hat{c}_{j,k}^1$ 和 $\hat{c}_{j,k+1}^1$, \dots 和 $\hat{c}_{j,k}^2, \hat{c}_{j,k+1}^2$. 要用这样的网络来进行函数逼近, 首先要考虑如下的两个问题.

(1) 是否存在一组权值使网络能够逼近到所需的精度. 如果存在, 其收敛速度如何.

(2) 对于存在的理论权值, 网络是否可以通过样本点集上的训练来逼近这组权值. 如果可以, 怎样估计其收敛速度.

本文的定理 1 和定理 2 分别回答了这两个问题.

3 主要结果

定义 1 设 F_j 是 R 上的函数集合序列, $F = \bigcup_j F_j$. 称 F 在 R 上具有 L^p ($1 \leq p \leq \infty$) 逼近性, 如果对于任意的 $f \in L^p(R)$, 存在 $f_j \in F_j$ ($j=1, 2, \dots$), 使得 L^p 在意义下有 $f_j \rightarrow f$. 特别地, 当 $p = \infty$ 时, 称 F 具有泛逼近性. 函数集合 F 通常称为函数估计子.

对于多子波网络而言, F_j 为对于固定的 j 网络选取不同权值所得到的全部函数. 由式(2)、式(3)和式(4)可知, 多子波神经网络具有 L^2 逼近性. 而对于泛逼近性, 则要求多尺度函数 $\phi(t)$ 具有一定的精确阶并且 $f(t)$ 具有一定的光滑性.

定义 2 多尺度函数 $\phi(t)$ 具有精确阶 m , 如果所有不超过 $m-1$ 阶的多项式均属于 V_0 , 即任何一个不超过 $m-1$ 阶的多项式均可以表示为 $\sum_{l=1}^r \sum_{k=z}^z \phi(t-k) |l=1, \dots, r, k=z|$ 的线性组合.

在实际问题中, 总是在一定的时间段上来分析和处理信号的. 并且目前常使用的子波, 不论是单子波还是多子波, 其支撑都具有 $[0, u]$ 的形式, 并且其尺度函数的支撑也同样具有 $[0, u]$ 的形式. 因此本文考虑 $\text{supp}(\phi) = \text{supp}(\phi(\cdot - k)) \subseteq [0, u]$ 的多子波神经网络对紧支撑连续函数的逼近问题.

定理 1 设有 r 维多尺度函数 $\phi(t)$, 其支撑为 $[0, u]$, 并具有精确阶 m . 设 $f \in L^p(R)$, 若 $f \in C_0^m(R)$, 其中 $C_0^m(R)$ 表

示具有紧支撑的 m 次连续可微函数空间, 则存在函数序列 $\{f_j(t)\}, f_j \in V_j$, 使得

$$f - f_j \leq c_1 r(1+u)^m u^{1/2} 2^{-jm} \quad (5)$$

$$f - f_j \leq c_2 r(1+u)^m u^{1/2} 2^{-jm} \quad (6)$$

证明 设 f_j 为 f 在 V_j 空间上的投影, 其表达式由式(3)给出, 则由式(2)可知 $f - f_j \rightarrow 0$. 因此, 为证明 $f - f_j \rightarrow 0$ ($j \rightarrow \infty$), 只需证明函数序列一致收敛. 设

$$(Qf)(t) = \sum_{l=1}^r \sum_{k=z}^z f_{j,k}^l \phi_{j,k}^l(t)$$

则由序列 f_j 的定义可知, 对于任意的正整数 q ,

$$f_{j+q}(t) - f_j(t) = \sum_{j=J}^{J+q-1} (Qf)(t) \quad (7)$$

由于 $f_{j,k}^l = \int_{j,k}^l f_{j,k}^l(t)$

$$\begin{aligned} &= 2^j \int_{j,k}^l f(x) \phi_{j,k}^l(2^j x - k) dx \\ &= 2^j \int_{j,k}^l \frac{f^{(m)}(x)}{m!} [x - (2^{-j}k - 2^{-j}t_{j,k})]^m \phi_{j,k}^l(2^j x - k) dx \end{aligned}$$

其中 $t_{j,k} = 2^j t - k$, 介于 $2^{-j}k - 2^{-j}t_{j,k}$ 与 x 之间. 令

$$c_3 = \frac{1}{m!} \max_{t \in R} |f^{(m)}(t)| \quad (8)$$

则 $f_{j,k}^l = \int_{j,k}^l f_{j,k}^l(t)$

$$\begin{aligned} &2^j c_3 \int_{j,k}^l |x - (2^{-j}k - 2^{-j}t_{j,k})|^m \phi_{j,k}^l(2^j x - k) dx \\ &= c_3 \int_{j,k}^l |x - t_{j,k} + 2^{-j}t_{j,k} - 2^{-j}k|^m \phi_{j,k}^l(x - k) dx \\ &= c_3 \int_{j,k}^l |x - t_{j,k} + 2^{-j}t_{j,k} - 2^{-j}k|^m \phi_{j,k}^l(x - k) dx \\ &\leq 2^{-jm} c_3 \int_{j,k}^l |x - t_{j,k} + 2^{-j}t_{j,k} - 2^{-j}k|^m \phi_{j,k}^l(x - k) dx \\ &\leq 2^{-jm} c_3 u^{1/2} \int_{j,k}^l |x - t_{j,k} + 2^{-j}t_{j,k} - 2^{-j}k|^m \phi_{j,k}^l(x - k) dx \end{aligned}$$

从而 $|(Qf)(t)| \leq \sum_{l=1}^r \sum_{k=z}^z |f_{j,k}^l| \phi_{j,k}^l(t)$

$$\leq 2^{-jm} c_3 u^{1/2} \sum_{l=1}^r \sum_{k=z}^z \int_{j,k}^l |x - t_{j,k} + 2^{-j}t_{j,k} - 2^{-j}k|^m \phi_{j,k}^l(x - k) dx$$

由于对于任意的 $n, 0 \leq n \leq m$, $|x^n|$ 是紧支撑的连续函数, 故 $|(x - k)^n \phi_{j,k}^l(x - k)|$ 有界, 设

$$c^l = \max_{k \in \mathbb{Z}} \left\{ |(x - k)^n \phi_{j,k}^l(x - k)| \mid n = 0, \dots, m \right\}, \text{ 并设}$$

$$c_4 = \max\{c^l \mid l = 1, \dots, r\}, c_5 = c_3 c_4 \quad (9)$$

则 $|(Qf)(t)| \leq c_5 r u^{1/2} (1+u)^m 2^{-jm}$. 由式(7)可知

$$|f_{j+q}(t) - f_j(t)| \leq \sum_{j=J}^{J+q-1} |(Qf)(t)| \leq c_1 r u^{1/2} (1+u)^m 2^{-Jm}$$

其中 $c_1 = \frac{c_5}{1 - 2^{-m}}$. 这样就证明了函数序列 $\{f_j(t)\}$ 的一致收敛性. 由 q 的任意性可知式(5)正确. 为证式(6), 只需注意到对于任意的紧支撑函数 $g(x)$, 均有

$$g(x) \leq U \max_{x \in R} |g(x)|$$

其中 U 表示 $\text{supp}(g)$ 的长度.

定理 1 不仅给出了网络逼近的速度为 $O(2^{-Jm})$, 而且表明了多子波相对于单子波的优越性. 如果在上面的讨论中令 $r=1$, 则得出了单子波时的相应结论, 因此该定理包含了已有

的结果. 由式(8)和式(9)还可知, 结论(5)和(6)中的常数与尺度函数的个数无关. 虽然在多子波的情况下 r 大于 1, 但与 r 相比, 支撑长度 u 对于逼近的速度起着更大的作用. 定理中的 u 是多子波 (t) 的支撑长度. 事实上, 可以证明, 对于很大的一类多尺度函数 $\phi(t)$, 可以选取 (t) , 使得 (t) 中有较多分量的支撑长度为 $u/2^{[8]}$. 因此, 从定理 1 的证明中可以看出, 实际的逼近效果比定理的结论更好. 本文实验中所使用的 GHM 多子波就是这种情况. 下面讨论问题(2).

定义 3 称一个函数估计子是相容的, 如果当训练样本点的数目趋于无穷大时函数估计子收敛于被逼近函数.

在一般情况下, 训练点是通过随机采样而获得的, 因此收敛概念应该是某种统计意义下的. 在这里, 本文考虑 L^2 范数的均方误差. 设有训练点集

$$T_N = \{ (t_i, f(t_i)) \mid i = 1, \dots, N \} \quad (10)$$

为了强调式(4)中估计量 $f_{J,N}(t)$ 对训练点数目 N 的依赖关系, 将式(4)重写为

$$f_{J,N}(t) = \sum_{l=1}^r \sum_{k=0}^{2^J-1} c_{J,k}^{l,N} \phi_{J,k}^l(t) \quad (11)$$

定理 2 设在训练样本式(10)中, $t_i (i = 1, \dots, N)$ 是独立同分布的, 且均为一致分布, 多子波神经网络式(11)的学习规则由 BP 算法给出. 并设网络输出 $f_{J,N}(t)$ 的全体所构成的函数估计子为 F . 则对于任意的紧支撑函数 $f \in L^2(R)$, 存在 J , 使得 $E[|f - f_{J,N}|^2] = O(1/N)$, 即多子波神经网络式(4)是均方意义下 L^2 相容的, 并且其收敛速度为 $O(1/N)$.

证明 不妨设 $\text{supp}(f) \subseteq [0, 1]$. 分两种情况讨论. (1) 设存在某个 J , 使得 $f \in V_J$. 这时,

$$f(t) = f_J(t) = \sum_{l=1}^r \sum_{k=0}^{2^J-1} f_{J,k}^l \phi_{J,k}^l(t) \quad (12)$$

由于 $\text{supp}(\phi) \subseteq [0, u]$, 故上式中只有有限项非零. 当 $k \geq 2^J$ 时, $f_{J,k}^l = 0$. 当 $k < 0$ 时, 设 $s = \text{supp}(\phi_{J,k}^l) \subseteq [0, 1]$, 则 $|s| < 2^{-J}u$, 其中 $|s|$ 表示 s 的测度, 则

$$|f_{J,k}^l|^2 = \left| \int_s f(t) \phi_{J,k}^l(t) dt \right|^2 \leq |s| \max_{x \in s} |f(x)|^2$$

$$x \in [0, |s|] \Rightarrow |f(x)| \leq u |2^{-J} \max_{x \in [0, 2^{-J}u]} |f(x)|^2 \Rightarrow x \in [0, 2^{-J}u] \}.$$

由于 $J \rightarrow \infty$ 时, $\max_{x \in [0, 2^{-J}u]} |f(x)|^2 \rightarrow 0$, 因此当 $2^{-J} = O(1/N)$ 时

时有

$$\sum_{k=0}^{2^J-1} |f_{J,k}^l|^2 = O(1/N) \quad (14)$$

其中 $O(1/N)$ 表示 $1/N$ 的高阶无穷小. 所以, 这时可以只考虑网络估计子

$$f_{J,N}(t) = \sum_{l=1}^r \sum_{k=0}^{2^J-1} c_{J,k}^{l,N} \phi_{J,k}^l(t) \quad (15)$$

对函数

$$f(t) = \sum_{l=1}^r \sum_{k=0}^{2^J-1} f_{J,k}^l \phi_{J,k}^l(t) \quad (16)$$

的逼近. 这时和式中的基函数 $\phi_{J,k}^l(t)$ 的支撑均含于 $[0, 1]$ 之内. 设

$$c_{J,k}^l = f_{J,k}^l, \quad \bar{c}_{J,k}^{l,N} = \frac{1}{N} \sum_{i=1}^N f(t_i) \phi_{J,k}^l(t_i), \quad l = 1, \dots, r, k = 1, \dots, 2^J - 1$$

$$\text{设 } \bar{f}_{J,N}(t) = \sum_{l=1}^r \sum_{k=0}^{2^J-1} \bar{c}_{J,k}^{l,N} \phi_{J,k}^l(t), \text{ 则}$$

$$E[|f - f_{J,N}|^2] = E[|f - \bar{f}_{J,N}|^2] = E[\sum_{l=1}^r \sum_{k=0}^{2^J-1} (c_{J,k}^l - \bar{c}_{J,k}^{l,N})^2] = \sum_{l=1}^r \sum_{k=0}^{2^J-1} E[(c_{J,k}^l - \bar{c}_{J,k}^{l,N})^2] \quad (17)$$

$$E[|f - \bar{f}_{J,N}|^2] = E[\sum_{l=1}^r \sum_{k=0}^{2^J-1} (c_{J,k}^l - \bar{c}_{J,k}^{l,N})^2] = \sum_{l=1}^r \sum_{k=0}^{2^J-1} E[(c_{J,k}^l - \bar{c}_{J,k}^{l,N})^2] \quad (18)$$

$$\text{由于 } E[\bar{c}_{J,k}^{l,N}] = \frac{1}{N} \sum_{i=1}^N f(t_i) \phi_{J,k}^l(t_i) = \int_0^1 f(t) \phi_{J,k}^l(t) dt = f_{J,k}^l,$$

$$E[(\bar{c}_{J,k}^{l,N})^2] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E[f(t_i) \phi_{J,k}^l(t_i) f(t_j) \phi_{J,k}^l(t_j)]$$

$$= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E[f(t_i) \phi_{J,k}^l(t_i) f(t_j) \phi_{J,k}^l(t_j)] + \frac{1}{N^2} \sum_{i=1}^N E[f^2(t_i) (\phi_{J,k}^l(t_i))^2] = S_1 + S_2,$$

$$S_1 = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E[f(t_i) \phi_{J,k}^l(t_i) E[f(t_j) \phi_{J,k}^l(t_j)]]$$

$$= \frac{N-1}{N} |f_{J,k}^l|^2 = (1 - \frac{1}{N}) (f_{J,k}^l)^2,$$

$$S_2 = \frac{1}{N} f^2, (\phi_{J,k}^l)^2,$$

$$E[(c_{J,k}^l - \bar{c}_{J,k}^{l,N})^2] = \text{Var}[\bar{c}_{J,k}^{l,N}] = E[(\bar{c}_{J,k}^{l,N})^2] - (f_{J,k}^l)^2 = S_1 + S_2 - (f_{J,k}^l)^2$$

$$= \frac{1}{N} f^2, (\phi_{J,k}^l)^2 - \frac{1}{N} (f_{J,k}^l)^2 \leq \frac{1}{N} f^2 + \frac{1}{N} f^2.$$

故由式(18)可知 $E[|f - \bar{f}_{J,N}|^2] = O(1/N)$. 因此, 由式(17)可知, 为证明定理 2, 只需证明当 N 充分大以后, 总有

$$\sum_{l=1}^r \sum_{k=0}^{2^J-1} E[(c_{J,k}^l - \bar{c}_{J,k}^{l,N})^2] \leq \sum_{l=1}^r \sum_{k=0}^{2^J-1} E[(c_{J,k}^l - \bar{c}_{J,k}^{l,N})^2] \quad (19)$$

由于网络是按使均方误差最小来训练权值的, 故

$$\frac{1}{N} \sum_{i=1}^N (f(t_i) - f_{J,N}(t_i))^2 \leq \frac{1}{N} \sum_{i=1}^N (f(t_i) - \bar{f}_{J,N}(t_i))^2 \quad (20)$$

两边取数学期望得 $E[(f(t) - f_{J,N}(t))^2] \leq E[(f(t) - \bar{f}_{J,N}(t))^2]$.

因为

$$E[(f(t) - f_{J,N}(t))^2] = E[\sum_{l=1}^r \sum_{k=0}^{2^J-1} (c_{J,k}^l - \bar{c}_{J,k}^{l,N})^2],$$

$$E[(f(t) - \bar{f}_{J,N}(t))^2] = E[\sum_{l=1}^r \sum_{k=0}^{2^J-1} (c_{J,k}^l - \bar{c}_{J,k}^{l,N})^2]$$

所以式(19)成立.

式(2)不存在 J , 使得 $f \in V_J$ 的情况. 由定理 1 知, 存在 J

使得 $\|f - f_J\|_2^2 \leq c2^{-2Jm}$. 与前面一样设 $2^{-J} = O(1/N)$. 这时 $\|f - f_J\|_2^2 = O(1/N^{2m})$, 从而由第(1)种情况的讨论可知 $E[\|f - f_{J,N}\|_2^2] = E[\|f - f_J + f_J - f_{J,N}\|_2^2] \leq E[2\|f - f_J\|_2^2 + 2\|f_J - f_{J,N}\|_2^2] \leq 2E[\|f - f_J\|_2^2] + 2E[\|f_J - f_{J,N}\|_2^2] = O(\frac{1}{N^{2m}}) + O(\frac{1}{N}) = O(\frac{1}{N})$

定理 2 证毕.

4 学习算法

不失一般性, 考虑 f 的支撑 $U = [0, 1]$ 时的情况. 并采用式(15)所给出的模型. 这类似于区间子波, 但由于多子波可以在保留其它特性的同时具有紧支撑, 并且可以比单子波的支撑更短. 所以它比区间子波具有更好的时间局部性, 从而收敛速度更快.

多子波网络学习算法

- (1) 独立、一致分布地选取训练点集 $T_N = \{t_i, f(t_i) \mid i = 1, \dots, N\}$, 并根据信号估计一个较小的 J 值.
- (2) 适当选取初始权值 $\hat{c}_{J,k}^l, l = 1, \dots, r, k = 0, \dots, 2^J - 1$, 产生一个初始的函数估计

$$\hat{f}(t) = \sum_{l=1}^r \sum_{k=0}^{2^J-1} \hat{c}_{J,k}^l \phi_{J,k}^l(t)$$

- (3) 令 $e_N(t_i) = f(t_i) - \hat{f}(t_i) (i = 1, \dots, N)$, $E_N = \frac{1}{N} \sum_{i=1}^N [e_N(t_i)]^2$, 并用梯度下降法近似求出使 E_N 达到最小的组合系数 $\hat{c}_{J,k}^l$.

- (4) 计算均方误差 E_N .

- (5) 如果 E_N 小于所设置的阈值, 停机. 否则, $J = J + 1$, 转向第二步.

注意, 如果要用较粗的尺度空间对函数进行估计, 则应采用式(11)给出的网络模型, 即在网络模型式(15)中适当添加 $\phi_{J,k}^l (k < 0)$ 的项.

5 网络参数与初值的选择策略

在使用第 4 节中算法进行函数学习时, 样本点数目 N 是首先要选定的. 它代表着对目标信号的先验了解, 是受客观因素所制约的. 充分大的 N 是网络准确估计目标函数的先决条件. 算法中的 J 既反映了层级空间的尺度及空间基底函数的支撑, 也反映了网络隐层神经元的个数, 因此它受到两个方面的约束. 当 N 固定后, 首先, 式(13)对于初始 J 值的选取起到了指导作用. 其次, 由优化理论可知, N 个点的训练样本集最多只能有效地训练出 N 个权值. 故应有

$$r2^J \leq N \quad (21)$$

同时考虑式(13)和式(21)可知

$$J = \text{fix}(\log_2(N/r)) \quad (22)$$

是网络训练时 J 的上限. 其中 fix 是取整函数. 对于本文实验中的 GHM 子波, 当 $N = 64$ 时, J 的上限为 5. 在对尺度空间相对于信号的逼近误差有所了解的情况下, 式(21)给出了训练网络权值所需的样本点数的下限估计.

在进行网络训练时, 初始权值的选取直接影响着网络的

收敛速度. 当网络的初始权值取为目标函数在采样点处的函数值时, 将会得到一个较好的初始估计. 当然, 由于其固有的低通特性, 单子波的这一初始估计将比多子波更好. 特别当精度要求不高时, 甚至可以将这一初始估计直接作为训练的结果, 这是单子波网络的一个优越性. 在实验 1 中就是这样选取初始权值的. 从图 4(d) 中可以看到, 这时单子波初始逼近的均方误差在 10^{-3} 左右. 在 40 次迭代之内单子波网络一直保持着初始估计的优势. 但是, 如果要求均方误差小于 10^{-4} , 则多子波快速收敛的优势就明显地体现出来了. 表 2 给出了有关的数据.

6 实验结果

分别用 Daubechies2 单子波和 GHM 多子波对图 2 和图 3 所示的函数进行了逼近. 之所以使用 Daubechies2 是因为它在单子波中具有较短的支撑. 实验中, 均匀采样 64 个点. 对于每一个给定的 J , 当梯度的所有分量函数的最大模均小于 10^{-4} 时, 可以认为网络达到了空间 V_J 对目标函数的逼近要求. 当这时的结果不能达到所要求的逼近精度时, $J = J + 1$. 图 4(a) 是 Daubechies2 单子波网络的逼近结果. 迭代次数为 1440, 均方误差为 $1.4e - 8$. 图 4(b) 是 GHM 多子波网络的逼近结果. 迭代次数为 527, 训练点上的均方误差为 $3.5e - 10$. 图 4(c) 给出了 $\|f - \hat{f}\|_2^2$ 随 J 的增大而减小的曲线. 虚线对应于单子波网络, 实线对应于多子波网络. 这里及以下用 1024 点均方误差作为 $\|f - \hat{f}\|_2^2$ 的近似. 从图中可以看出, 多子波随层级空间的增大而逼近目标函数的速度明显快于单子波. 图 4(d) 给出了对于固定的 J , 训练点上的均方误差随迭代次数的增加而减小的曲线. 因为对于固定的 J , GHM 多子波网络的节点个数为 2^{J+1} , 所以, 为了使计算具有相同的复杂度, 同时也为了使层级空间具有相同的维数, 图 4(d) 分别以虚线和实线给出了 J 等于 6 时的单子波网络和 J 等于 5 时的多子波网络的均方误差收敛曲线. 由图可见, 多子波网络的收敛速度明显快于单子波网络. 为了更精确地给出实验结果, 表 1 给出了 J 从 1 到 5 的每一层级空间上网络的迭代次数、 $\|f - \hat{f}\|_2^2$ 和 $\|f - \hat{f}\|_1$. 表 2 给出了为使训练点均方误差达到所给精度, 网络所进行的迭代次数. 对于图 3 中的函数, 本文进行了同样的实验, 结果与前一实验基本相同. 其变化趋势及具体数值由图 5(a) ~ (d) 及表 3 和表 4 给出.

7 结论

理论分析和实验结果都表明, 在同样的计算复杂度下, 多子波神经网络学习的收敛速度明显快于单子波网络. 其主要原因在于, 多子波不仅保持了单子波的全部优点, 而且可以比单子波具有更短的支撑. 这使得在局部调整系数时, 不会改变

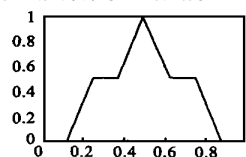


图 2

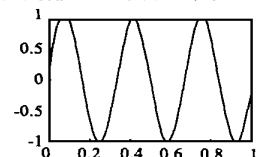


图 3

已经学习好了的其它系数. 因而本文提出的多子波网络和学习算法适用于广泛的逼近、学习与识别问题.

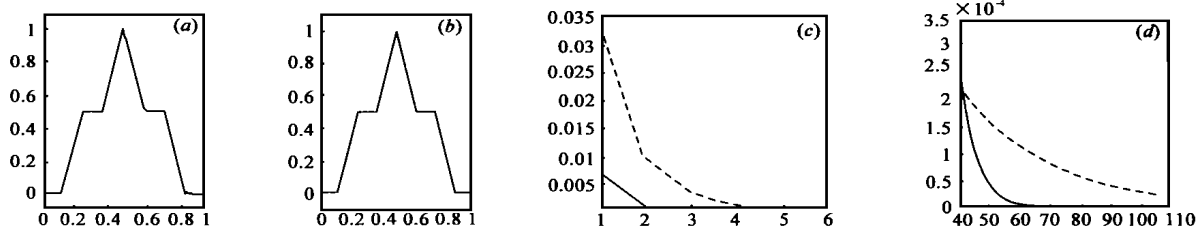


图 4

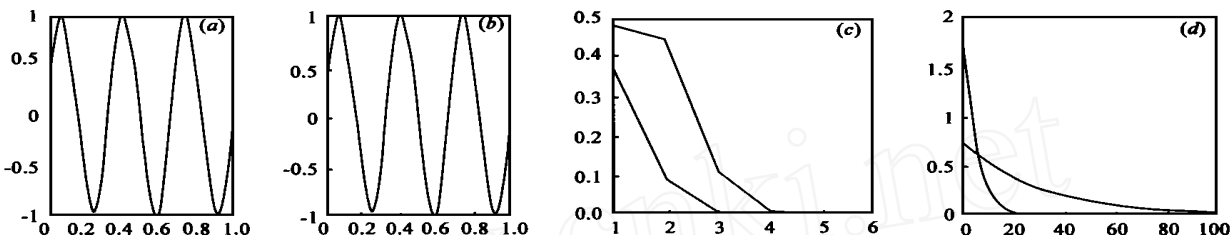


图 5

表 1

空间层级 J		1	2	3	4	5
迭代次数	GHM	22	33	50	71	112
	Daubechies2	27	53	79	134	230
$f - f_{\frac{1}{2}}$	GHM	0.0068	0.0002	0.0000	0.0000	0.0000
	Daubechies2	0.0340	0.0089	0.0031	0.0004	0.0001
$f - f$	GHM	0.1953	0.0408	0.0000	0.0001	0.0001
	Daubechies2	0.3650	0.2356	0.1503	0.0763	0.0420

表 2

迭代次数	均方误差	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
节点函数										
GHM ($J=5$)		10	22	34	46	59	71	83	95	107
Dau. ($J=6$)		1	1	2	64	133	204	274	345	416

表 3

空间层级 J		1	2	3	4	5
迭代次数	GHM	23	34	66	147	120
	Daubechies2	27	57	115	234	546
$f - f_{\frac{1}{2}}$	GHM	0.3824	0.0903	0.0055	0.0003	0.0001
	Daubechies2	0.4783	0.4440	0.1147	0.0120	0.0011
$f - f$	GHM	1.1175	0.7126	0.1737	0.0467	0.0951
	Daubechies2	1.1940	1.1568	0.9889	0.3013	0.964

表 4

迭代次数	均方误差	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
节点函数										
GHM ($J=5$)		17	29	41	53	65	77	90	102	114
Dau. 2 ($J=6$)		63	134	205	276	348	419	494		

作者简介:



潘 进 1990 年在陕西师范大学获得硕士学位, 现为西安电子科技大学雷达信号处理重点实验室博士研究生, 主要从事智能信号处理及子波方面的研究.



焦李成 1984 年和 1990 年在西安交通大学分别获得硕士和博士学位. 现为西安电子科技大学教授, 博士生导师, 主要从事非线性科学和智能信号处理以及神经网络与大规模并行处理等领域的研究.

参考文献:

- [1] Q. Zhang, A. Benveniste. Wavelet network [J]. IEEE Trans. on NN, Nov. 1992, 3:889 - 898.
- [2] J. Zhang et al. Wavelet neural networks for function learning [J]. IEEE Trans. on SP, 1995, 43(6):1485 - 1497.
- [3] 高协平, 张钊. 区间小波神经网络(I) — 性质与模拟 [J]. 软件学报, 1998, 9(3).
- [4] 高协平, 张钊. 区间小波神经网络(II) — 理论与实现 [J]. 软件学报, 1998, 9(4).
- [5] G. Strang and V. Strela. Short wavelets and matrix dilation equations [J]. IEEE Trans. on SP, 1995, 43:108 - 115.
- [6] H. S. Geronimo, D. P. Hardin, and P. R. Massopust. Fractal functions and wavelet expansions based on several scaling functions [J]. J. Approx. Theory, 1994, 78:373 - 401.
- [7] G. Strang and V. Strela. Orthogonal multiwavelets with vanishing moment [J]. J. Optical Eng., 1994, 33:2104 - 2107. (下转第 55 页)

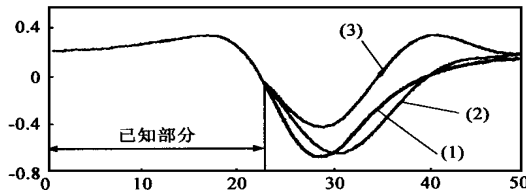


图4 待外推信号中直流分量对网络收敛速度的影响

根据片段信号的外推结果;(3)是将原信号加一直流分量(其片段信号也相应加上此分量)后,根据片段信号外推,再将外推部分减去预加直流分量之后的结果。外推效果截然不同,说明直流分量在其中起了作用。

上面四个例子说明了:(1)改进算法的有效性;(2)算法对噪声有一定的抑制能力;(3)加窗处理可改善因截断而使频谱严重泄漏的待外推信号的外推效果;(4)去直可加快收敛速度。从而验证了文中分析的合理性。各例中选用不同的信号都取得了满意结果说明该算法对不同类别带限信号都是有效的。

5 结论

本文引进了交替投影神经网络,将其应用范围从实数域拓广到复数域,并给出了在复数域仍然成立的若干结论。在研究频带受限数字信号的离散傅立叶变换特性之后,结合网络的这些结论,通过选取一组正交基作为网络的独立模式,来对网络进行训练。本文在对(1)网络对噪声的抑制能力;(2)网络收敛速度;(3)因截断而使待外推信号频谱严重外泄等问题的分析与讨论的基础上,提出了一种基于交替投影神经网络的带限信号外推算法。仿真实验表明该方法是行之有效的。文中的分析与设计是合理的。另外,该算法对频谱外推同样适用;由于它采用全互连神经网络结构,因而易于并行计算和 VLSI 实现,从而可满足军事上实时处理的需要。

参考文献:

[1] D. Slepian, H. O. Pollak and H. J. Landau, Prolate spheroidal wave functions, [J]. Bell Syst. J., Jan. 1966, 40:43 - 84.

- [2] R. W. Gerchberg, Superresolution through error energy reduction [J]. Opt. Acta, 1975, 21 (9): 709 - 720.
- [3] A. Papoulis, A new algorithm in spectral analysis and bandlimited extrapolation [J]. IEEE Trans. CAS-22, 1975: 735 - 742.
- [4] A. K. Jain and S. Ranganath, Extrapolation algorithms for discrete signal with application in spectral estimation [J]. IEEE, Trans. ASSP, 1981, 29(4): 830 - 845.
- [5] C. K. Rushforth et al., Least-squares reconstruction of objects with missing high-frequency components [J]. J. Opt. Soc. Am, Feb 1982, 72 (2)
- [6] 陈宗鹭等, 带宽有限信号外推的泛函算子逼近法 [J]. 电子学报, 1987, 15(2): 70 - 76.
- [7] 王思雄等, 基于傅立叶变换的带限信号外推算法 [J]. 电子学报, 1992, 20(12): 75 - 78.
- [8] R. J. Marks, S. Oh, and L. E. Atlas, Alternating projection neural networks [J]. IEEE Trans. CAS, 1989, 36(6): 846 - 857.

作者简介:



王金根 1972 年出生, 现为海军工程大学在读博士生。研究兴趣为: 弱信号检测与处理, 神经网络, 小波分析等。

林春生 1983 年毕业于海军工程学院 (现为海军工程大学), 1996 年获工学博士学位, 现为海军工程大学教授, 主要从事数字信号处理、人工智能、传感器技术等方面的教学与科研工作。



龚沈光 1939 年 7 月出生, 1962 年毕业于哈军工。现任海军工程大学教授, 博士生导师, 出版专著 2 部, 获国家科技进步奖、军队科技进步奖、全国科学大会奖等各项奖共 14 项。主要研究领域为: 舰船物理场, 非触发引信, 现代信号处理等。

(上接第 64 页)

[8] PAN Jin and JIAO Licheng, Construction of orthogonal multiwavelets with short sequence via genetic algorithm [A]. Neural Networks and

Brain Proceedings [C], ICNN&B '98, Beijing: 1998, 492 - 495.