第9卷第4期
1998年4月

软件学报
JOURNAL OF SOFTWARE

Vol. 9, No. 4
Apr. 1998

区间小波神经网络(II)——性质与模拟*

高协平[1,3] 张钹[2,3]

[1](湘潭大学数学系 湘潭 411105)

[2](清华大学计算机科学与技术系 北京 100084)

[3](清华大学智能技术与系统国家重点实验室 北京 100084)

摘要 证明了区间小波神经网络具有一致及 L^2 逼近性质，且为相容的函数估计子，其学习收敛速度在 d 维情形不随 d 增大而减慢，本质上克服了神经网络高维学习的"维数灾难"问题，模拟实例验证了理论的正确性.

关键词 神经网络，小波，多尺度分析，收敛.

中图法分类号 TP18

在文献[1]中，我们通过重新定义 $L^2[0,1]$ 上的正交多尺度分析，提出了区间小波神经网络模型. 对于实际应用中通常囿于有界区域(或空间)上的信号处理，该模型与已有其它神经网络[2~4]相比能使用更少的神经元，完全避免了训练过程陷入不期望局部极小的问题. 进一步，我们给出了该模型结构设计即隐层神经元数目的较好的上下界估计. 本文考虑区间小波神经网络与学习能力相关的几个理论结果，证明了该模型的学习收敛速度在 d 维情形下不随 d 增大而减慢，本质上解决了神经网络高维学习的"维数灾难"问题，模拟实例与理论结果吻合.

1 逼近性质

定义 1. 设 F_j 是 R^d 中函数集合子列，$F=\bigcup_j F_j$，U 是 R^d 中紧子集，称 F 在 U 中具有 $L^p(1\leqslant p\leqslant\infty)$ 逼近性质，如果对任意 $f\in L^p(U)$，存在 $f_n\in F_n$，在 L^p 意义下有 $f_n\to f$. 特别地，$p=\infty$ 指一致(Universal)逼近性.

从文献[1]中区间小波神经网络的构造(式(1))，容易知道它有 L^2 逼近性质. 实际上，关于其学习能力，我们有下列进一步的定理.

定理 1. 区间小波神经网络具有一致和 L^2 逼近性质. 设被分析信号 f 在 $[0,1]^d$ 上有所有 N_0 阶连续偏导数，则存在区间小波神经网络序列 $f_M\in V_M$，使 $\|f-f_M\|_\infty=O(h^{N_0})$，$\|f-f_M\|_{L^2}=O(h^{N_0})$，这里 $h=\frac{1}{2^M}$.

证明：记 $M_{p,k}=\langle x^p,\varphi_{0k}\rangle$，$N_{p,k}=\langle x^p,\psi_{0k}\rangle$，$m_k=\langle|\varphi_{0k}|,1\rangle$，$n_k=\langle|\psi_{0k}|,1\rangle$，

$$\|f^{(N_0+1)}\|_\infty^*=\max\left\{\left|\frac{\partial^{N_0+1}}{\partial x_1^{p_1}\ldots\partial x_d^{p_d}}f(x_1,\ldots,x_d)\right|;p_1+\ldots+p_d=N_0+1,p_i\geqslant 0\right\},$$

$$D^m=\{g(x)\,|\,|g(x)|\leqslant c\cdot(1+|x|)^{-(m+1+\varepsilon)},\varepsilon>0\},$$

其余记号沿用文献[1].

设 $\varphi,\psi\in D^{N_0}$，由 $L^2[0,1]$ 上的正交多尺度分析构造知，V_j 包含所有 $\leqslant N_0-1$ 次代数多项式全体，因此有

$$N_{p,k}=0,\text{对 }0\leqslant p\leqslant N_0-1 \tag{1}$$

这样，我们甚至可以把结论证明得更多一点，即给出误差渐近展式.

设 $f(x_1,\ldots,x_d)$ 在 V_n，W_n 空间的投影算子分别为 P_n，Q_n，则

$$P_nf(x_1,\ldots,x_d)=\sum_{k_1\ldots k_d\in J_n}\langle\langle f(y_1,\ldots,y_d),\prod_{l=1}^d\varphi_{nk_l}(y_l)\rangle\rangle\prod_{l=1}^d\varphi_{nk_l}(x_l) \tag{2}$$

* 本文研究得到国家自然科学基金，国家863高科技项目基金和国家攀登计划基金资助. 作者高协平，1965年生，副教授，主要研究领域为计算智能，计算数学及应用软件. 张钹，1935年生，教授，博士导师，中国科学院院士，主要研究领域为人工智能，计算机应用技术.

本文通讯联系人：高协平，湘潭 411105，湘潭大学数学系

本文 1997-01-21 收到原稿，1997-09-01 收到修改稿

$$Q_n f(x_1,\ldots,x_d)=\sum_{i=0}^{d-1}\sum_{u=D}\sum_{\substack{k_1,\ldots,k_i\in J_n\\ q_{i+1}\cdots q_d\in L_n}}\langle\langle f(y_1,\ldots,y_d),\prod_{l=1}^{i}\varphi_{nk_l}(y_{u_l})\cdot\prod_{l=i+1}^{d}\psi_{nu_l}(y_{u_l})\rangle\rangle\cdot\prod_{l=1}^{i}\varphi_{nk_l}(x_{u_l})\cdot\prod_{l=i+1}^{d}\psi_{oq_l}(x_{u_l})\quad(3)$$

注意到 $L^2([0,1]^d)=V_M\bigoplus_{j\geq M}W_M$，所以

$$\varepsilon_M f(x_1,\ldots,x_d)\triangleq f(x_1,\ldots,x_d)-P_M f(x_1,\ldots,x_d)=\sum_{j\geq M}Q_j f(x_1,\ldots,x_d)\quad(4)$$

设在 $[0,1]^d$ 上 $f(x_1,\ldots,x_d)$ 有所有 N_0+1 阶连续偏导数，将式(3)中 $f(y_1,\ldots,y_d)$ 在 $(x_1,\ldots,x_d)$ 处作 Taylor 展开，我们有

$$\begin{aligned}Q_j f(x_1,\ldots,x_d)=&\sum_{i=0}^{d-1}\sum_{u=D}\sum_{\substack{k_1\cdots k_i\in J_j\\ q_{i+1}\cdots q_d\in L_j}}\sum_{0\leq p_1+\cdots+p_d=p\leq N_0}\frac{\partial^p f(x_{u_1},\ldots,x_{u_d})}{\partial x_{u_1}^{p_1}\cdots\partial x_{u_d}^{p_d}}\cdot\frac{1}{(p_1)!\cdots(p_d)!}\times\\&\langle\langle\prod_{l=1}^{d}(y_{u_l}-x_{u_l})^{p_l},\prod_{l=1}^{i}\varphi_{jk_l}(y_{u_l})\cdot\prod_{l=i+1}^{d}\psi_{jq_l}(y_{u_l})\rangle\rangle\prod_{l=1}^{i}\varphi_{jk_l}(x_{u_l})\cdot\prod_{l=i+1}^{d}\psi_{jq_l}(x_{u_l})+\\&\sum_{i=0}^{d-1}\sum_{u=D}\sum_{\substack{k_1\cdots k_i\in J_j\\ q_{i+1}\cdots q_d\in L_j}}\sum_{\substack{p_1+\cdots+p_d=p=N_0+1\\ p_l\geq 0}}\frac{1}{(p_1)!\cdots(p_d)!}\prod_{l=1}^{i}\varphi_{jk_l}(x_{u_l})\cdot\prod_{l=i+1}^{d}\psi_{jq_l}(x_{u_l})\times\\&\langle\langle\frac{\partial^{N_0+1}f(\zeta^1,\ldots,\zeta^d)}{\partial x_{u_1}^{p_1}\cdots\partial x_{u_d}^{p_d}}\cdot\prod_{l=1}^{d}(y_{u_l}-x_{u_l})^{p_l},\prod_{l=1}^{i}\varphi_{jk_l}(y_{u_l})\cdot\prod_{l=i+1}^{d}\psi_{jq_l}(y_{u_l})\rangle\rangle\\=&\Lambda_j+\rho_j\end{aligned}\quad(5)$$

这里 $\zeta^l=\zeta^l(x_{u_l},y_{u_l}),l=1(1)d$.

先估计第 2 项 ρ_j（记 $H=2^{-j},h=2^{-M}$）：

$$\begin{aligned}&①\overset{\text{def}}{=\!=}\langle|(y_{u_l}-x_{u_l})^{p_l}|,|\varphi_{jk_l}(x_{u_l})|\rangle\cdot|\varphi_{jk_l}(x_{u_l})|\overset{t_{u_l}=2^jy_{u_l}}{=\!=\!=\!=}H^{p_l}\cdot\langle|(t_{u_l}-2^jx_{u_l})^{p_l}|,|\varphi_{0k_l}(t_{u_l})|\rangle\cdot|\varphi_{0k_l}(2^jx_{u_l})|\\&\quad\leq H^{p_l}\cdot\sum_{v_l=0}^{p_l}\left[\binom{p_l}{v_l}\cdot|2^jx_{u_l}|^{v_l}\cdot\langle|t_{u_l}|^{p_l-v_l},|\varphi_{0k_l}(t_{u_l})|\rangle\right]\cdot|\varphi_{0k_l}(2^jx_{u_l})|\\&\quad\leq H^{p_l}\cdot(1+|2^jx_{u_l}|)^{p_l}\cdot|\varphi_{0k_l}(2^jx_{u_l})|\end{aligned}$$

$$\left[\because\varphi\in\mathrm{D}^{N_0},\therefore\langle|t_{u_l}|^{p_l-v_l},|\varphi_{0k_l}(t_{u_l})|\rangle\leq\int_0^1\frac{|t_{u_l}|^{p_l-v_l}}{(1+|t_{u_l}|)^{N_0+1+\varepsilon}}dt_{u_l}\leq 1\right]$$

$$②\overset{\text{def}}{=\!=}\langle|(y_{u_l}-x_{u_l})^{p_l}|,|\psi_{jq_l}(y_{u_l})|\rangle\cdot|\psi_{jq_l}(x_{u_l})|\leq H^{p_l}\cdot(1+|2^jx_{u_l}|)^{p_l}\cdot|\psi_{oq_l}(2^jx_{u_l})|$$

$$\begin{aligned}\therefore|\rho_j|\leq&\sum_{i=0}^{d-1}\sum_{u=D}\sum_{\substack{k_1\cdots k_i\in J_j\\ q_{i+1}\cdots q_d\in L_j}}\sum_{\substack{p_1+\cdots+p_d=N_0+1\\ p_l\geq 0}}\frac{\|f^{(N_0+1)}\|_\infty^*}{(p_1)!\cdots(p_d)!}\cdot\prod_{l=1}^{i}①\cdot\prod_{l=i+1}^{d}②\\\leq&H^{N_0+1}\cdot\sum_{\substack{p_1+\cdots+p_d=N_0+1\\ p_l\geq 0}}\frac{\|f^{(N_0+1)}\|_\infty^*}{(p_1)!\cdots(p_d)!}\sum_{i=0}^{d-1}\sum_{u=D}\sum_{\substack{k_1\cdots k_i\in J_j\\ q_{i+1}\cdots q_l\in L_j}}\\&\prod_{l=1}^{i}\{(1+|2^jx_{u_l}|)^{p_l}\cdot|\varphi_{0k_l}(2^jx_{u_l})|\}\cdot\prod_{l=i+1}^{d}\{1+|2^jx_{u_l}|)^{p_l}\cdot|\psi_{oq_l}(2^jx_{u_l})|\}\\\leq&\mathrm{C}\cdot H^{N_0+1}(\because\varphi,\psi\in\mathrm{D}^{N_0}).\end{aligned}$$

这里 C 是与 h,j 无关的有界常数. 所以

$$\sum_{j\geq M}|\rho_j|\leq\mathrm{c}\cdot h^{N_0+1}\quad(6)$$

再看第 2 项 Λ_j. 令

$$\begin{aligned}\lambda_{j,i,k_i,q_{i+1},u}=&\langle\langle\prod_{l=1}^{d}(y_{u_l}-x_{u_l})^{p_l},\prod_{l=1}^{i}\varphi_{jk_l}(y_{u_l})\cdot\prod_{l=i+1}^{d}\psi_{jq_l}(y_{u_l})\rangle\rangle\\=&\prod_{l=1}^{i}\langle(y_{u_l}-x_{u_l})^{p_l},\varphi_{jk_l}(y_{u_l})\rangle\cdot\prod_{l=i+1}^{d}\langle(y_{u_l}-x_{u_l})^{p_l},\psi_{jq_l}(y_{u_l})\rangle\end{aligned}$$

注意 $N_{p,k}=0$，对 $0\leq p\leq N_0-1$，

$\therefore$　$\langle(y_{u_l}-x_{u_l})^{p_l},\psi_{jq_l}(y_{u_l})\rangle=0$，对 $0\leq p_l\leq N_0-1$.

故

$$\lambda_{j,i,k_i,q_{i+1},u}\begin{cases}=0, & \text{当 } p<N_0 \text{ 或 } p=N_0 \text{ 但 } i<d-1 \\ & \quad\text{或 } p=N_0 \text{ 且 } i=d-1 \text{ 但 } p_{u_d}<N_0 \\ \neq 0, & \text{当 } p=N_0 \text{ 且 } i=d-1 \text{ 且 } p_{u_d}=N_0 \\ & \quad(\text{此时 } p_{u_l}=0, \text{对 } 1\leqslant l\leqslant d-1)\end{cases}$$

∴

$$\lambda_{j,i,k_i,q_{i+1},u}=H^{N_0}\cdot\prod_{l=1}^{d}M_{ok_l}\cdot N_{N_0,q_d}\cdot 2^{-jd/2}$$

定义函数 $\sigma_{j,u}(x_1,\ldots,x_d)\in L^2([0,1]^d)$如下：

$$\sigma_{j,u}(x_1,\ldots,x_d)=\sum_{\substack{k_1\cdots k_{d-1}\in J_j\\ q_d\in L_j}}2^{-(j-M)N_0}\cdot N_{N_0q_d}\varphi(2^jx_{u_l})\prod_{l=1}^{d-1}\left(M_{ok_l}\psi(2^jx_{u_l})\right)$$

则对 $j\geqslant M$，有

$$\Lambda_j=\frac{h^{N_0}}{N_0!}\sum_{u\in D}\left(\frac{\partial^{N_0}f(x_1,\ldots,x_d)}{\partial x_{u_d}^{N_0}}\cdot\sigma_{j,u}(x_1,\ldots,x_d)\right)\tag{7}$$

定义 $\tau_{M,u}(x_1,\ldots,x_d)$为一个一致收敛级数的有界函数：

$$\tau_{M,u}(x_1,\ldots,x_d)=\sum_{l=0}^{\infty}\sigma_{M+l,u}(x_1,\ldots,x_d)\tag{8}$$

于是由式(4)、(6)～(8)得到

$$\varepsilon_Mf(x_1,\ldots,x_d)=\frac{h^{N_0}}{N_0!}\sum_{u\in D}\left(\frac{\partial^{N_0}f(x_1,\ldots,x_d)}{\partial x_{u_d}^{N_0}}\cdot\tau_{M,u}(x_1,\ldots,x_d)\right)+O(h^{N_0+1}).$$

这样，定理 1 按 $\|\cdot\|_\infty$收敛结论显然成立．对 $[0,1]^d$ 上的 L^2 收敛性，只需注意

$$\int_{[0,1]^d}|f(x_1,\ldots,x_d)|^2\mathrm{d}x_1\ldots\mathrm{d}x_d\leqslant\operatorname*{Sup}_{x_i\in[0,1]}|f(x_1,\ldots,x_d)|^2$$

于是，定理 1 得证．

注记：在神经网络的逼近性质研究中，收敛速度的估计是十分重要的内容．设 q 表示被逼近函数 f 的可微指标，对径向基函数(RBF)神经网络，文献[4]提到：Girosi 和 Anzellotti 得到了其一致与 L^2 逼近敛速 $O(n^{-\frac{1}{2}})$；Corradi，White 和 Xu 等得到其 L^2敛速 $O(n^{-1/(1+d/2q)})$(详见文献[4]的参考文献[6～8])．这些收敛速度估计，后者显然依赖于维数 d，除非 f 无穷光滑；而前者虽然表面上独立于维数 d，但被逼近函数类将随维数 d 增加而缩小，因此，实质上是依赖于维数的；Zhang Q，Benveniste 在文献[3]中提出了一类(标架)小波神经网络，Delyon，Benveniste 等在文献[5]中得到了其 L_2 敛速 $O(n^{-q/d})$，显然随着维数 d 的增大，收敛速度减慢；Zhang J 等在文献[4]中提出了另一类(正交)小波神经网络，并给出其一致和 L^2 敛速度估计 $O(h^\alpha)$，同样不幸的是，α 依赖于维数 d 及函数的光滑度 q．虽然类似于 Girosi，Anzellotti 关于 RBF 的结果，随着 d 增大以缩小函数类为代价可以避开"维数灾难"，但这绝非根本性的．因此，"维数灾难"问题并没有真正解决．本文定理 1 指出，区间小波神经网络的一致及 L^2 收敛速度完全与维数 d 无关，不随 d 的增大而减慢，同时，对 f 的光滑性要求仅与所取多尺度分析相关而不依赖于维数 d，因此，从本质上彻底解决了高维学习的"维数灾难"问题．从我们的证明过程容易看出，文献[4]的"维数灾难"缘于其证明方法，而采用完全平行于本文方法，其"维数灾难"是可以本质解决的．

2　相容性质

定义 2.　一个函数估计子(如区间小波神经网络)称为是 L^p 相容的，如果随着训练样本数趋于无穷大，估计子在 L^p 意义下收敛于被逼近函数．

记

$$K=\{k\ k=(k_1,\ldots,k_d),k_i\in J_M,i=1(1)d\},x=(x_1,\ldots,x_d)$$

设对某足够大的 M，$f\in V_M$，则

$$f(x)=\sum_{k\in K}c_k^{(0)}\varphi_{M,k}(x)\tag{9}$$

这里

$$c_k^{(0)}=\langle\langle f,\varphi_{M,k}\rangle\rangle$$

设 $\widetilde{f}_{M,N}(x)$ 是关于 $f(x)$ 由训练数据集 T_N 的一个区间小波神经网络实现，则

$$\widetilde{f}_{M,N}(x)=\sum_{k\in K}\widetilde{C}_{N,k}\varphi_{M,k}(x) \tag{10}$$

其中系数 $\widetilde{C}_{N,k}$ 由文献[1]提供的算法(4)(5)获取.

由多尺度分析的规范正交性，有

$$\begin{aligned}E[\|f-\widetilde{f}_{M,N}\|^2]&=E[\int_{[0,1]^d}(f(x)-\widetilde{f}_{M,N}(x))^2\mathrm{d}x]\\&=E[\int_{[0,1]^d}(\sum_{k\in K}(c_k^{(0)}-\widetilde{C}_{N,k})\varphi_{M,k}(x))^2\mathrm{d}x]\\&=E[\sum_{u\in K}\sum_{v\in K}(c_u^{(0)}-\widetilde{C}_{N,u})(c_v^{(0)}-\widetilde{C}_{N,v})\langle\langle\varphi_{M,u},\varphi_{M,v}\rangle\rangle]\\&=E[\sum_{k\in K}(c_k^{(0)}-\widetilde{C}_{N,k})^2]=\sum^{k\in K}E[(c_k^{(0)}-\widetilde{C}_{N,k})^2]\end{aligned} \tag{11}$$

定理 2. 设训练样本集 $T_N=\{(x^{(i)},y^{(i)}),i=1(1)N\}$ 是独立恒等分布和一致分布的，$\widetilde{f}_{M,N}(x)$ 是 $f\in V_M$ 的区间小波神经网络实现，则

$$E[(c_k^{(0)}-\widetilde{C}_{N,k})^2]=O(\frac{1}{N}),\quad k\in K \tag{12}$$

从而区间小波神经网络在均方根期望意义下是 L^2 相容的，且收敛速度为 $O(\frac{1}{N})$.

证明：注意到式(11)，故只需证式(12)为真即可，过程完全类似于文献[4]，从略.

推论. 在定理 2 的条件下，设对任何 M，$f\in L^2([0,1]^d)$ 但 $f\notin V_M$，$\widetilde{f}_{M,N}$ 是 f 的区间小波神经网络实现，则它在均方根期望意义下是 L^2 相容的.

事实上，取 $N>n=2^M$，并让 $n\to\infty$(从而 $N\to\infty$)，由定理 1，存在一个区间小波神经序列 $f_n\in V_M$，使 $f_n\xrightarrow{L^2}f(n\to\infty)$；由定理 2，存在序列 $\widetilde{f}_{M,N}$，使 $\widetilde{f}_{M,N}\xrightarrow{L^2}f_n(N\to\infty)$. 这样，对充分大的 n 和 N，有

$$E[\|f-\widetilde{f}_{M,N}\|^2]\leqslant\|f-f_n\|^2+E[\|f_n-\widetilde{f}_{M,n}\|^2]\xrightarrow{L^2}0.$$

3 模拟实验

本节给出了区间小波神经网络及 $L^2(R)$ 小波神经网络[4]的模拟实例结果. 为了科学地评价逼近效果，对样本数据集 $T_N=\{(x^{(i)},y^{(i)})\}_{i=1}^N$ 和神经网络输出 $f_{M,N}(x)$，采用如下评判标准：

$$\varepsilon\triangleq\varepsilon(f_{M,N})=\sqrt{\frac{\sum_{j=1}^{N}[f_{M,N}(x^{(j)})-y^{(j)}]^2}{\sum_{j=1}^{N}(\bar{y}-y^{(j)})^2}}$$

其中 $\bar{y}=\frac{1}{N}\sum_{j=1}^{N}y^{(j)}$. 分别使用 $N_0=2$ 的 Daubechies 小波和相应的区间小波. 模拟结果见表 1,2，它们验证了理论的正确性.

例 1：

$$f(x)=\begin{cases}43.72x-8.996, & 0\leqslant x<0.4\\-84.92x+42.46, & 0.4\leqslant x\leqslant 0.5\\10e^{-x}\sin(12x^2+2x-4), & 0.5<x\leqslant 1\end{cases} \tag{13}$$

取一致样本数 $N=128$.

例 2：

$$g(x,y)=(x^2+y^2)\sin\frac{\pi}{2}x\cdot\sin\frac{\pi}{2}y,0\leqslant x,y\leqslant 1 \tag{14}$$

取一致样本数 $N=1\,024$.

表1 对 $f(x)$ 的逼近

方法	隐层单元数	迭代次数	误差控制 ϵ
区间小波网	32	5	0.049 449
$L^2(R)$ 小波网	34	62	0.048 261

表2 对 $g(x,y)$ 的逼近

方法	隐层单元数	迭代次数	误差控制 ϵ
区间小波网	462	4	0.046 756
$L^2(R)$ 小波网	510	17	0.042 194

参考文献

1 高协平,张钹.区间小波神经网络(I)——理论与实现.软件学报,1998,9(3):217~221
(Gao Xie-ping, Zhang Bo. Interval-wavelets neural networks(I)— theory and implements. Journal of Software, 1998,9(3): 217~221)

2 Pati Y C, Krishnaprasad P S. Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations. IEEE Transactions on Neural Networks, 1993,4(1):73~85

3 Zhang Q, Benveniste A. Wavelet networks. IEEE Transactions on Neural Networks, 1992,3(6):889~898

4 Zhang J, Walter G G, Miao Y *et al*. Wavelet neural networks for function learning. IEEE Transactions on Signal Processing, 1995,43(6):1485~1497

5 Delyen B, Juditsky A, Benveniste A. Accuracy analysis for wavelet approximations. IEEE Transactions on Neural Networks, 1995,6(2):332~348

Interval-wavelets Neural Networks (II)——Properties and Experiment

GAO Xie-ping[1,3] ZHANG Bo[2,3]

[1](*Department of Mathematics Xiangtan University Xiangtan* 411105)

[2](*Department of Computer Science and Technology Tsinghua University Beijing* 100084)

[3](*National Key Laboratory of Intelligent Technology and Systems Tsinghua University Beijing* 100084)

Abstract In the present paper, it is proved that the interval wavelets neural networks has universal and L^2 approximation properties and is a consistent function estimator. Convergence rates associated with these properties do not decrease as d increases in d-dimensional function learning, *i.e.*, the "curse of dimensionality" is eliminated substantially. In the experiments, the proposed interval wavelet neural networks, compared to traditional wavelet networks, has performed better.

Key words Neural network, wavelets, multiresolution analysis, convergence.