

小波变换、神经网络和小波网络的函数逼近能力分析与比较

刘志刚, 王晓茹, 何正友, 钱清泉

(西南交通大学电气化自动化研究所, 四川省成都市 610031)

摘要: 基于对小波变换和神经网络之间内在联系的分析, 利用神经网络不同激励函数的线性组合构造出了相应的小波函数, 得出小波函数作为神经网络的激励函数与普通神经网络的激励函数在本质上是一致的结论, 并引入了小波网络。通过对小波变换、神经网络和小波网络函数逼近能力的理论分析与比较, 认为小波网络在函数逼近方面具有明显的优势, 并且分别利用这 3 种方式对一典型函数进行了仿真逼近的验证。

关键词: 小波变换; 神经网络; 小波网络; 函数逼近

中图分类号: TN911; TP182

0 引言

小波理论是在最近 10 年不断完善和发展起来的, 其独特的优点使其在信号处理、图像处理、通信、地球物理等许多领域里得到青睐。与 Fourier 级数在 Fourier 分析中的函数逼近能力类似, 小波级数在小波分析中的函数逼近能力是小波理论中最重要的特征之一。神经网络则为人们提供了一种具有学习和记忆能力的、并行处理的结构和算法。近几年来最吸引人们研究的是它的泛逼近能力 (universal approximants), 泛逼近能力是指可以逼近任意函数的能力。

神经网络对模式识别、函数逼近等有着独特的优点, 而小波分析对非平稳随机信号具有逐级精细描述和特征提取的能力, 这两个工具结合起来就是小波神经网络, 简称小波网络。小波网络是 Zhang Qinghua 等人在 1992 年作为对前馈神经网络逼近任意函数变换的概念提出的^[1]。近几年来, 这种小波和神经网络结合的思想被不断地提出和改进^[2~7]。Zhang Jun 等人对小波网络用于函数学习, 即小波网络的函数逼近能力做了研究, 并与其他方法进行了比较^[4]; B. Delyon 等人从理论上对小波逼近的精度进行了分析^[7]; Y. C. Pati 等人利用离散仿射小波变换对前馈神经网络进行了分析和综合^[2]; 最近, Jiao Licheng 等人对多小波网络及其特性进行了研究^[5]。

本文首先介绍小波变换和前馈神经网络的函数

逼近能力, 从两个不同的角度介绍了小波变换和前馈神经网络之间的关系。在此基础上讨论了小波网络的函数逼近能力以及三者之间的对比, 并利用不同的方法对一典型函数进行了仿真。

1 小波变换的函数逼近能力

设 $V_n \subset L^2(0, 2\pi)$, $V_n = \text{span}\{e^{-in\pi}, \dots, e^{-i\pi}, 1, e^{i\pi}, \dots, e^{in\pi}\}$, 其中: $L^2(0, 2\pi)$ 是指在 $(0, 2\pi)$ 内的平方可积空间; $n = N + M + 1$; N, M 为正整数。若函数 $f(x) \in L^2(0, 2\pi)$, 则 $f(x)$ 在 V_n 中的最佳平方逼近元为 $\sum_{k=-M}^N c_k e^{ikx}$, 其中 c_k 为函数 $f(x)$ 在各个正交基上的投影系数, $c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx$ 。取最佳平方逼近元的极限, 则 $f(x)$ 表示为 $f(x) = \sum_{-\infty}^{\infty} c_k e^{ikx}$ 。也就是说, 每个 2π 周期平方可积函数都可以用 $e^{i\pi}$ 整数膨胀的叠加生成。

注意到 $e^{i\pi} \notin L^2(\mathbf{R})$, $\mathbf{R} = (-\infty, \infty)$, 如果把空间 $L^2(0, 2\pi)$ 扩展到空间 $L^2(\mathbf{R})$, 那么很显然, $f(x) \in L^2(\mathbf{R})$ 不可能用 $e^{i\pi}$ 整数膨胀的叠加生成。所以, 必须用其他类型的基来生成空间 $L^2(\mathbf{R})$, 这种基就是小波。小波与空间 $L^2(0, 2\pi)$ 的正交基 $e^{i\pi}$ 非常类似, 都是生成相应空间并对相应空间上函数进行逼近的基, 由于小波必须是衰减很快的函数, 这就决定了小波的多样性和复杂性。

如果 $\Psi \in L^2(\mathbf{R})$, 并且满足允许条件^[8]:

$$C_{\Psi} = \int_{-\infty}^{\infty} \frac{|\hat{\Psi}(w)|^2}{|w|} dw < \infty \quad (1)$$

式中: $\hat{\Psi}(w)$ 为 $\Psi(t)$ 的 Fourier 变换, $\Psi(t)$ 称为“母小波”。

为简单起见,首先考虑由 $\Psi(t)$ 构成正交基的情况。 $\Psi_{j,k}(t) = 2^{j/2}\Psi(2^j t - k)$ ($j, k \in (-\infty, \infty)$), 生成空间 $L^2(\mathbf{R})$ 的一组正交基, 即:

$$W_j = \text{span}\{2^{j/2}\Psi(2^j t - k), \dots, 2^{j/2}\Psi(2^j t - k)\}$$

$$L^2(\mathbf{R}) = \bigoplus_j W_j$$

利用对空间 $L^2(\mathbf{R})$ 的多分辨率分析(MRA), 有以下的一系列闭子空间:

$$\dots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \dots \quad (2)$$

式中: $\bigcap_j V_j = \{0\}$; $\text{close}\{\bigcup_j V_j\} = L^2(\mathbf{R})$; $V_j = \text{span}\{2^{j/2}\varphi(2^j t - k)\}$; $V_{j+1} = V_j \oplus W_j$; $\varphi(t)$ 是相应小波的尺度函数。

由上面的讨论, 可以把空间 $L^2(\mathbf{R})$ 中的函数 $f(t)$ 进行分解^[8]:

$$f(t) = \sum_{j,k} \langle f, \Psi_{j,k} \rangle \Psi_{j,k}(t) \quad (3)$$

也就是说,

$$\lim_{M_1, N_1, M_2, N_2 \rightarrow \infty} \left\| f - \sum_{j=-M_1}^{N_2} \sum_{k=-M_2}^{N_1} c_{j,k} \Psi_{j,k} \right\|_2 = 0$$

式中: $c_{j,k} = \langle f, \Psi_{j,k} \rangle$ 。

则必然存在某个整数 J_0 , 使得对于任何 $\varepsilon > 0$, 满足下列不等式:

$$\left\| f(t) - \sum_k c_{J_0,k} \Psi_{J_0,k} \right\|_2 < \varepsilon \quad (4)$$

这种形式与函数用 Fourier 级数来逼近是完全一样的, 只是除了基的整数膨胀之外, 还有基在实直线上的平移。

下面介绍非正交小波的函数逼近。

在空间 $L^2(\mathbf{R})$ 中, 函数 $f(t)$ 的离散小波变换被定义为: $d_{j,k} = \langle f, \Psi_{j,k} \rangle = |a|^{-1/2} \int f(t) \bar{\Psi}_{j,k} dt$ 。在介绍函数 $f(t)$ 的重构之前, 有必要引入小波理论中的框架理论^[8]。假定一个 $L^2(\mathbf{R})$ 空间和一个向量序列 $\{\Psi_{j,k}\}$, $\{\Psi_{j,k}\} \subset L^2(\mathbf{R})$, 则 $\{\Psi_{j,k}\}$ 称为一个框架; 如果存在一个常数 $A > 0$, 且 $B < \infty$, $B \geq A$, 并满足:

$$A \|f\|_2^2 \leq \sum_n |\langle f, \Psi_{j,k} \rangle|^2 \leq B \|f\|_2^2 \quad (5)$$

对于任何函数 $f \in L^2(\mathbf{R})$, A 和 B 称为框架界。

任何函数 $f \in L^2(\mathbf{R})$ 都可以被写为^[9]:

$$f = \sum_{j,k} \langle f, S^{-1}\Psi_{j,k}(t) \rangle \Psi_{j,k} \quad (6)$$

式中: S 称为框架算子。

在小波变换的实际计算中, 式(6)中多项式的有限项之和就可以对函数 $f \in L^2(\mathbf{R})$ 进行有效的逼近, 即:

$$f = \sum_{j,k} w_{j,k} \Psi_{j,k}(t) = \sum_{j,k} w_{j,k} \Psi\left(\frac{t-b_j}{a_j}\right) \quad (7)$$

式中: $w_{j,k} = \langle f, S^{-1}\Psi_{j,k}(t) \rangle$; $N < \infty$ 。

小波的一个显著特点是, 它不仅能成为 $L^2(\mathbf{R})$ 空间的标准正交基, 而且还为许多其他空间提供无条件基, 例如 $L^p(\mathbf{R})$ 空间、Sobolev 空间、Hölder 空间等。也就是说, 利用小波可以在许多空间上进行函数逼近。

2 神经网络的函数逼近能力

为简单起见, 这里以人工神经网络中最常用的前馈神经网络来说明神经网络的逼近能力。从本质上看, 前馈神经网络的学习过程就是通过各层连接权的调整和组合, 达到一种满意的拓扑结构, 这种拓扑结构能将学习样本的给定输入矢量空间, 映射到给定的输出矢量空间^[10]。

对于一个给定的连续函数 $f(x)$, $x \in [0, 1]$, 可

以用一个 Fourier 级数来逼近, $f_F(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx}$, c_k

为相应的 Fourier 系数。若 x 为一个 n 维空间矢量, 即用函数 $f(x)$ 完成 $[0, 1]^n \subset \mathbf{R}^n \rightarrow \mathbf{R}^m$ 的映射, 当 $f(x) \in L^2(\mathbf{R})$ 时, 则根据 Fourier 级数理论, 仍存在一个可以逼近 $f(x)$ 的级数:

$$f_F(x, N, f) = \sum_{k_1=-N}^N \sum_{k_2=-N}^N \dots \sum_{k_n=-N}^N c_{k_1, k_2, \dots, k_n} e^{i \sum_{i=1}^n k_i x_i} = \sum_k c_k e^{ikx} \quad (8)$$

当 $N \rightarrow \infty$ 时, 满足:

$$\lim_{N \rightarrow \infty} \int_{[0, 1]^n} |f(x) - f_F(x, N, f)| dx = 0 \quad (9)$$

对于任意多维函数的映射, 也可以得到类似的结果。给定一个函数 $h(x)$, $x \in \mathbf{R}^n$, $[0, 1]^n \subset \mathbf{R}^n \rightarrow \mathbf{R}^m$, 其 $h(x) = [h_1(x), h_2(x), \dots, h_m(x)]^T$, 则 h 中的每一分量也可以用相应的 Fourier 级数来近似, 这就是神经网络的映射定理。

映射定理^[11]: 给定一个 $\varepsilon > 0$, 一个连续矢量 h , 矢量中的各个分量满足 $\int_{[0, 1]^n} |h_i(x)| dx$ 存在, 且 h 完成映射 $h: [0, 1]^n \subset \mathbf{R}^n \rightarrow \mathbf{R}^m$, 则必定存在一个 3 层 BP 神经网络来逼近 H , H 为所要逼近函数组成的空间, 且使逼近误差保持在 ε 之内。

从上面的讨论可以看出, 神经网络的映射定理实际上是先用一个 Fourier 级数来逼近 h 中的各个分量 $h_i(x)$; 然后用一个输入层、隐含层和输出层单元数分别为 $n, n_1, 1$ 的 BP 子网络, 来逼近 Fourier 级数的任一三角函数; 最后将这些子网络组合起来, 形成一个能够逼近给定映射函数 h 的 3 层 BP 网络。它实际上是 Fourier 级数逼近函数的推广。虽然利用神经网络来逼近函数时, 无法给出具体的公式解释其

逼近原理,但是利用神经网络来逼近函数与利用多项式(包括 Fourier 级数和小波级数等)、样条函数逼近函数相比较,具有独特的优点。

下面对神经网络与小波级数进行比较:

a. 利用小波级数来逼近函数时,求解未知参数的过程和步骤随自变量的增加而变得非常复杂,限制了它们在解决高维问题中的应用;而前馈神经网络逼近函数时,其运算的复杂程度主要体现在运算时间的加长,而不是运算步骤的复杂。

b. 对于不同的函数,小波的选择是至关重要的,如果小波函数选得不合适,则必须重新计算;前馈神经网络的结构则非常灵活,很容易进行调整。

c. 需要逼近函数所在的空间,这是小波级数不能进行逼近的首要条件;前馈神经网络却没有这样的限制。

d. 小波级数对突变信号很敏感,而前馈神经网络却没有该优点。

在前馈神经网络中,目前所采用的 BP 算法收敛速度慢,容易出现局部极小化。因此,小波变换和前馈神经网络相结合将能够弥补两者在函数逼近方面的缺陷,综合两者的优点,达到最满意的函数逼近效果。

3 小波变换和前馈神经网络之间的关系

在介绍小波网络的函数逼近能力之前,分析小波变换和前馈神经网络之间的内在联系是非常必要的。下面将从各自的角度来分析它们之间的内在联系。

3.1 从小波变换的角度

由小波变换可得到式(7)所示的逼近函数。前馈神经网络的神经元输出 y 为:

$$y = f\left(\sum_{i=1}^n w_i x_i\right) \quad n \in \mathbf{R}, n < \infty \quad (10)$$

其中,偏差被包含在多项式中。

比较式(7)和式(10),可以发现它们之间是非常相似的。事实上,式(7)可以看成是一个 3 层神经网络结构,它以小波函数 Ψ 作为隐含层神经元的激励函数,在输出层上是一个线性神经元输出; $w_{j,k}$ 可以作为输入节点和输出节点之间的权重; $\Psi\left(\frac{x-b_j}{a_j}\right)$ 可以作为输入节点的输出值。由前馈神经网络和小波级数逼近的等式,可以构成如图 1 所示的小波网络^[3]。

从图 1 所示的小波网络可以得到:

$$\hat{f}(t) = \sum w_k \Psi\left(\frac{t-b_k}{a_k}\right) \quad (11)$$

式中; w_k 为权重; b_k 为该小波网络的位移参数; a_k 为尺度参数。

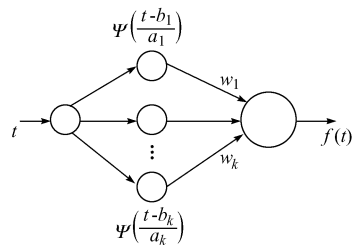


图 1 由式(7)构成的小波网络

Fig. 1 Wavelet network based on equation (7)

如果恰当地选择网络的权重和位移、尺度参数,那么,该小波网络就可以逼近函数 $f(t)$ 。当然,位移参数、尺度参数和权重的选取可以通过类似 BP 算法调整得到。

3.2 从前馈神经网络的角度

我们知道,前馈神经网络中激励函数的不同选择将决定该网络的输入/输出特性,常用的激励函数有 3 种,分别是:阶跃函数、sigmoid 函数和径向基函数,如图 2(a)、(b)和(c)所示。

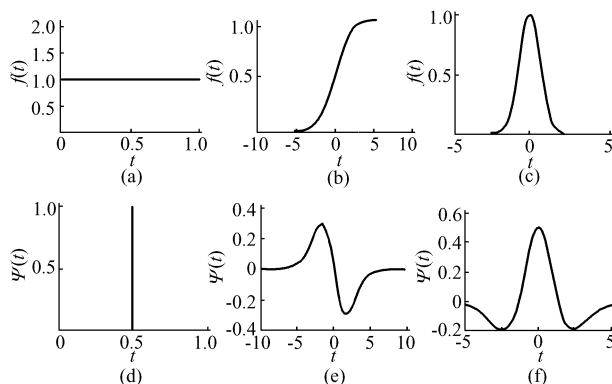


图 2 3 个激励函数和由它们产生的小波
Fig. 2 Three activation functions and three wavelets from activation functions

给定阶跃函数:

$$U = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases}$$

如果令 $\Psi(t) = U(2t) - U(2t-1)$, $t \in [0, 1]$, $\Psi(t)$ 如图 2(d)所示,事实上, $\Psi(t)$ 就是 Haar 小波的尺度函数。可以认为,具有激励函数为阶跃函数的前馈神经网络中的 2 个神经元可以构成一个 Haar 小波的尺度函数。这个过程如图 3 所示。

给定 sigmoid 函数为 $f(\cdot) = (1 + e^{-ax})^{-1}$, 其中 $a > 0$ 。如果考虑在空间 $L^2(\mathbf{R})$ 上令 $\Psi(x) = f(x+2) - f(x) + f(x-2)$, $x \in \mathbf{R}^{[2]}$, $\Psi(x)$ 则如图 2(e)所示。可以发现 $\Psi(x)$ 是一个奇函数,以指

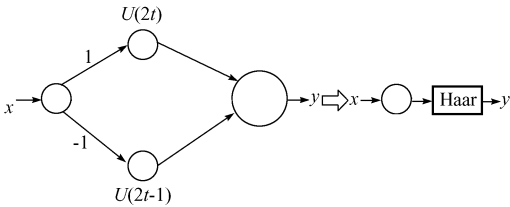


图 3 由 2 个神经元构成 Haar 小波的尺度函数
Fig. 3 Two neurons construct a dilation function of Haar wavelet

数形式衰减,而且很显然, $\int \Psi(x)dx = 0$, 也就是说 $\Psi(x)$ 满足小波的允许条件(式(1))。可以认为,具有激励函数为 sigmoid 函数的前馈神经网络中的 3 个神经元可以构成一个小波。这个过程如图 4 所示。

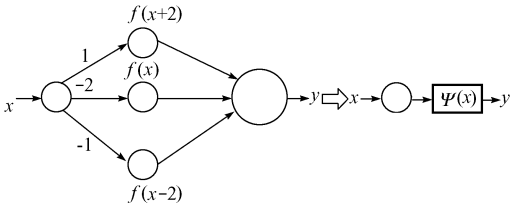


图 4 由 3 个神经元构成的小波
Fig. 4 Three neurons construct a wavelet

给定径向基函数 $f(x) = \frac{1}{2\sqrt{\pi a}} e^{-\frac{x^2}{4a^2}}$, 其中, $f \in L^2(\mathbf{R})$, $x \in \mathbf{R}$, $a > 0$ 。令 $a = \frac{1}{\sqrt{2}}$ 在一个 $f(x)$ 中, $a = \sqrt{2}$ 在另一个 $f(x)$ 中, 则

$$\Psi(x) = \sqrt{2} \sqrt{2\pi} f(x)_{a=\frac{1}{\sqrt{2}}} - \frac{2}{2} \sqrt{\sqrt{2\pi}} f(x)_{a=\sqrt{2}} = e^{-\frac{x^2}{2}} - \frac{1}{2} e^{-\frac{x^2}{8}}$$

$\Psi(x)$ 如图 2(f) 所示。可以发现, $\Psi(x)$ 满足小波的允许条件(式(1))。可以认为,具有激励函数为径向基函数的前馈神经网络中的 2 个神经元可以构成一个小波, 这个函数实际上是 DOG (difference of Gaussian) 小波。这个过程如图 5 所示。

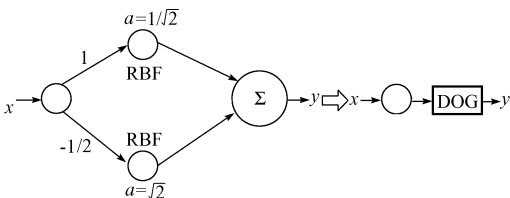


图 5 由 2 个神经元构成的一个 DOG 小波
Fig. 5 Two neurons construct a DOG wavelet

从上面的讨论可以看出,在具有激励函数分别为阶跃函数、sigmoid 函数和径向基函数的前馈神

经网络中,可以通过设置神经元的权值和激励函数的平移,分别构造相应的小波函数。也就是说,前馈神经网络中的神经元的某种线性组合与某一小波的作用在一定程度上是等效的。可以认为,小波函数作为前馈神经网络的激励函数与普通的前馈神经网络没有本质上的区别。

4 小波网络的函数逼近能力

径向基函数神经网络是典型的局部逼近神经网络,与激励函数为 sigmoid 函数的全局逼近神经网络相比,它具有明显的优点,但是径向基函数是非正交性的,其逼近函数的表达式并不惟一。小波网络与径向基函数神经网络相比,有着独特的优点:①小波函数可以是正交性的,保证逼近函数表达式的惟一性;②对于不同的逼近函数,小波函数可以有更广泛的选择,可以根据逼近函数的特性来选择不同的小波函数;③小波函数具有对突变函数逐步精细的描述特性,使得函数的逼近效果更好,这是径向基函数所不具备的。径向基函数神经网络在空间 $L^2(\mathbf{R})$ 的收敛率为 $O(n^{-1/2})$,而小波网络在空间 $L^2(\mathbf{R})$ 的收敛率为 $O(n^{-\alpha})$, $\alpha > 0$ [4]。两者相比,小波神经网络的收敛率比径向基函数神经网络的收敛率更好。

一般说来,前馈神经网络与小波变换的结合有两种形式,即小波变换作为神经元的激励函数和作为前馈神经网络的输入前置处理工具。在小波网络进行函数逼近之前,选择非正交基小波和正交基小波,以及小波基的个数,都将直接影响函数逼近的能力和计算的复杂程度。下面分别采用不同的方法对一阶梯函数(如图 6(a)所示,本文所采用的函数为一典型的阶梯突变函数,仿真结果不具有普遍性)进行逼近,所有方法均采用学习步数为 1 000、能量函数误差为 0.000 01 的标准。图 6(b)采用加动量的 BP 算法逼近,图 6(c)采用径向基网络逼近,图 6(d)采用 Daubechies 小波级数逼近,图 6(e)采用图 1 形式的小波网络逼近,并且分别绘制了它们的误差曲线。

从图 6 可以看出,小波级数与小波网络的逼近效果最好,BP 算法神经网络的逼近效果最差;小波级数的逼近误差最小,接着为小波网络、径向基网络、BP 网络。小波网络虽然在误差精度和逼近速度方面比小波变换差些,但是它具有神经网络的所有优点,其误差分布均匀(所有误差都在 0.02 以下),不像径向基网络和 BP 网络受到信号突变的影响,而且训练速度明显快于径向基网络和 BP 网络。所

以从某种意义上说,小波网络是逼近效果最好的神经网络。但是,小波网络却存在小波本身所具有的缺点,在仿真过程中,如果尺度 a_k 、位移 b_k 以及权重 w_k 的初始值选得不对,小波网络可能不收敛。

的函数逼近能力,分别从小波变换和神经网络两个方面介绍了它们之间的内在关系,并且利用不同的激励函数构造相应的小波函数。然后介绍了小波网络的函数逼近能力,并与小波变换和神经网络的函数逼近能力进行了比较。最后采用不同的方法对一阶梯函数进行了逼近,分析了逼近结果。根据分析认为,在某种程度上小波网络的函数逼近能力最好,但是存在小波变换本身的缺陷。

参考文献

- 1 Zhang Qinghua, Benveniste A. Wavelet Network. IEEE Trans on Neural Network, 1992, 3: 889~898
- 2 Pati Y C, Krishnaprasad P S. Analysis and Synthesis of Feedforward Neural Network Using Discrete Affine Wavelet Transformations. IEEE Trans on Neural Network, 1993, 4: 73~85
- 3 Szu H H, Telfer B, Kadambe B. Neural Networks Adaptive Wavelets for Signal Representation and Classification. Optical Engineering, 1992, 31: 1907~1916
- 4 Zhang Jun, Walter G G, Miao Yubo, et al. Wavelet Neural Networks for Function Learning. IEEE Trans on Signal Processing, 1995, 43: 1485~1497
- 5 Jiao Licheng, Pan Jin, Fang Yangwang, et al. Multiwavelet Neural Networks and Its Approximation Properties. IEEE Trans on Neural Network, 2001, 12: 1060~1066
- 6 Zhang Q H. Using Wavelet Network in Nonparametric Estimation. IEEE Trans on Neural Network, 1997, 8: 227~236
- 7 Delyon B, Juditsky A, Benveniste A, et al. Accuracy Analysis for Wavelet Approximations. IEEE Trans on Neural Network, 1995, 6: 332~348
- 8 Chui C K. Wavelets: A Tutorial in Theory and Application. New York: Academic, 1992
- 9 Daubechies. Ten Lectures on Wavelets. In: CBMS-Conference Lecture Notes. SIAM Philadelphia, 1992
- 10 靳 蕃(Jin Fan). 神经计算智能基础原理·方法(Theory and Method of Neural Computation Intelligence). 成都:西南交通大学出版社(Chengdu: Southwest Jiaotong University Press), 2000
- 11 张立明(Zhang Liming). 人工神经网络的模型及其应用(Model and Applications of ANN). 上海:复旦大学出版社(Shanghai: Fudan University Press), 1993

刘志刚(1975—),男,博士研究生,研究方向为电力系统及其自动化。E-mail: zhigang-liu457@sohu.com

王晓茹(1965—),女,副教授,研究方向为电力系统及其自动化、小波分析及应用、神经网络等。

何正友(1970—),男,副教授,研究方向为电力系统及其自动化、小波分析及应用等。

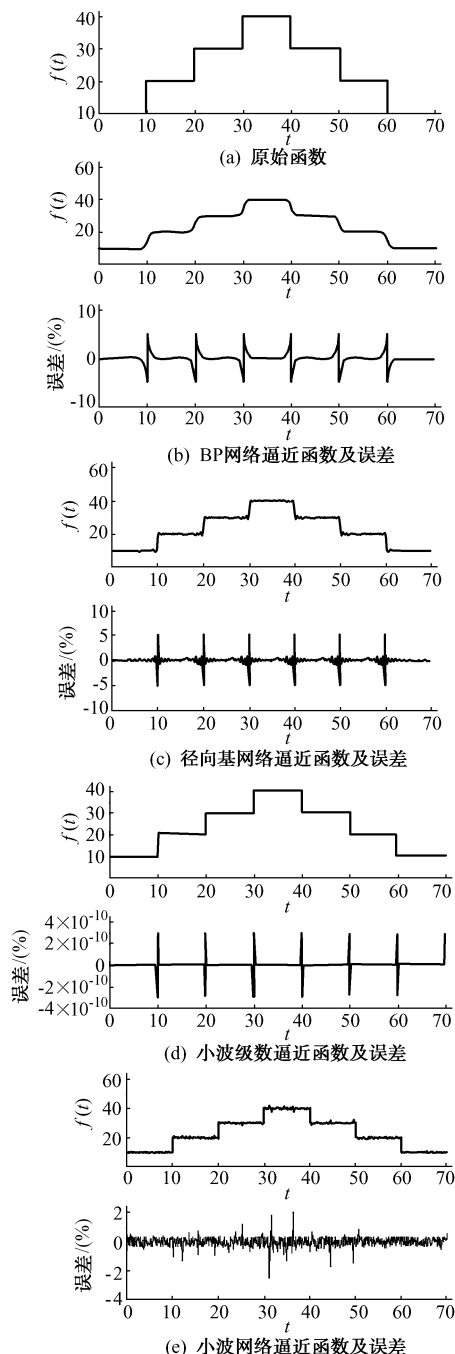


图6 不同方法逼近结果及误差

Fig. 6 Function approximation results and errors with different tools

5 结语

本文首先从理论上介绍了小波变换和神经网络

ANALYSIS AND COMPARISON OF FUNCTION APPROXIMATION ABILITY BASED ON WAVELET TRANSFORMATION, NEURAL NETWORK AND WAVELET NETWORK

Liu Zhigang, Wang Xiaoru, He Zhengyou, Qian Qingquan

(Southwest Jiaotong University, Chengdu 610031, China)

Abstract: Based on the analysis of inherent relations between wavelet transformation and neural network, the corresponding wavelet functions can be constructed by the linear combination of different neural network activation functions. As neural network activation functions, the wavelet functions are consistent with common neural network activation functions. This paper introduces wavelet network and gives theoretical analysis and comparisons of function approximation ability of wavelet transformation, neural network and wavelet network. The wavelet network is considered as having distinct advantages. The function approximation simulations for a typical function with these three means are done to support the conclusion.

This project is supported by National Natural Science Foundation of China (No. 59977019).

Key words: wavelet transformation; neural network; wavelet network; function approximation