

一种小波—神经网络多变量混合过程模型 及其应用

杨尔辅 徐用懋

(清华大学自动化系 北京 100084)

摘要 提出了一种小波—神经网络多变量混合过程模型, 将小波分析对信息分解和重构的能力与神经网络处理多变量非线性映射的能力很好地结合了起来, 发挥了各自的优势. 在多分辨率分析和 Mallat 算法研究的基础上, 提出了样本数据分解和空间划分的新方法, 使得模型输出信息熵最小, 提高了整个模型的可靠性、鲁棒性和适应性. 最后, 将该模型应用于常压塔顶粗汽油干点估计, 结果验证了它的合理性和有效性.

关键词 小波分析, 神经网络, 多分辨率分析, Mallat 算法, 过程建模, 常压塔, 软测量

1 引言

人工神经网络目前已成功地应用于许多复杂对象的建模、优化等方面^[1-3]. 为了改善神经网络模型的性能, 近年来多神经网络建模方法的研究与应用引起了人们的极大关注^[1-2]. 在多神经网络建模与应用过程中, 为了给每个子神经网络分配输入输出, 需要人为地对样本空间进行分解, 如通常采用的聚类方法等, 由于样本空间的复杂性以及所选择聚类算法的影响, 很难保证所建模型的通用性和有效性, 多网络模型的规模也往往事先无法确定; 同时, 由于误差、干扰甚至过失数据的存在会极大地影响聚类的结果, 使得多神经网络模型的推广性能进一步变差. 因此, 研究样本数据空间的适应性、确定性分解方法便成为多神经网络建模研究的一个重要内容.

小波分析是目前强有力的一种信号时频处理手段, 可以对指定频带和时段内的信号成分进行任意尺度的分析^[4-6]. 本文对过程数据样本空间根据其在小波变换下的“频带”特征进行自适应的划分, 形成每个子神经网络的子数据样本空间, 然后用多个神经网络来描述和解释整个数据空间, 模型的综合应用了小波反变换, 实现了样本空间中输出数据的重构. 这一信息分解和综合的过程便是小波—神经网络多变量混合过程模型的主要思想, 模型中的每一个子神经网络建立了样本数据空间中从“低频”到“高频”各个信号分量的映射关系. 同时, 由于在离散小波变换中一旦尺度确定, 子神经网络的数目便是确定的, 所以模型的建立不受样本空间划分结果的影响.

2 基本理论

2.1 多分辨率分析

多分辨率分析 (Multiresolution Analysis, MRA) 是在 $L^2(\mathbf{R})$ 函数空间内, 将函数 $f(t)$ 描述为一系列近似函数的极限, 这些近似都是在不同尺度上得到的. 令 \mathbf{R} 表示实数域, \mathbf{Z} 表示整数域, $f(t) \in L^2(\mathbf{R})$, 现考虑用分辨率 2^j 逼近函数 $f(t)$ 的问题^[5].

定义 1: 空间 $L^2(\mathbf{R})$ 的闭子空间列 $\{V_j; j \in \mathbf{Z}\}$ 称为 $L^2(\mathbf{R})$ 的一个多分辨率分析, 如果:

① 单调性: $V_j \subset V_{j+1}, \forall j \in \mathbf{Z}$, 即: $\dots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \dots$;

② 逼近性: $\bigcap_{j \in \mathbf{Z}} V_j = \{0\}, \bigcup_{j \in \mathbf{Z}} V_j = L^2(\mathbf{R})$ ③ 伸缩性: $f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}$;

④ 平移不变性: $f(t) \in V_j \Rightarrow f(t-k) \in V_j, \forall k \in \mathbf{Z}$;

⑤ Riesz 基: 存在 $g(t) \in V_j$, 使得 $\{g(2^j t - k) | k \in \mathbf{Z}\}$ 构成 V_j 的 Riesz 基, 即对任何 $f(t) \in V_j$, 存在唯一序列 $\{c_k \in l^2\}$ 使得 $f(t) = \sum_{k \in \mathbf{Z}} c_k g(2^j t - k)$; 反之, 任意序列 $\{c_k \in l^2\}$ 确定唯一函数 $f(t) \in V_j$, 且存在 $0 < A \leq B < \infty$, 使得 $A \|f\|^2 \leq \sum_{k \in \mathbf{Z}} |c_k|^2 \leq B \|f\|^2$.

令 A_j 是用分辨率 2^j 逼近函数 $f(t)$ 的算子, 则有以下关系成立:

$$\|u(t) - f(t)\| \geq \|A_j f(t) - f(t)\|, \forall u(t) \in V_j \quad (1)$$

又令 D_j 是用分辨率 2^j 逼近函数 $f(t)$ 的细节算子, $D_j f(t) = f(t) - A_j f(t)$, 则 $A_j f(t)$ 是在子空间 V_j 中最近似函数 $f(t)$ 的分量, 故称近似分量. 由于逼近算子 A_j 是在向量空间 V_j 上的正交投影, 为了数值计算的方便, 需要求出 V_j 的正交基.

定理 1: 设 $\{V_j; j \in \mathbf{Z}\}$ 为 $L^2(\mathbf{R})$ 的一个多分辨率逼近, 则存在一个唯一的尺度函数 $\phi(t)$, 令 $\phi_j(t) = 2^j \phi(2^j t)$, $j \in \mathbf{Z}$, 则 $\{\sqrt{2^{-j}} \phi_j(t - 2^{-j} k), k \in \mathbf{Z}\}$ 是 V_j 的一个正交基.

定理 2: (小波构造定理) 设 $\{V_j, j \in \mathbf{Z}; \phi(t)\}$ 是一个正交 MRA, 对 $\forall j \in \mathbf{Z}$, W_j 为 V_{j+1} 在 V_j 中的正交补空间, 即 $V_j \perp W_j$, 并且 $V_{j+1} = V_j \oplus W_j$, 如果: $\phi(t) = \sum_{k \in \mathbf{Z}} c_k \phi(2t - k)$ (2)

那么, 函数 $\psi(t) = \sum_{k \in \mathbf{Z}} (-1)^k \bar{c}_{1-k} \phi(2t - k)$ (3)

是 $L^2(\mathbf{R})$ 上的正交小波, 并且当 $W_j = \text{close}_{L^2(\mathbf{R})} \{2^{\frac{j}{2}} \psi(2^j t - k), k \in \mathbf{Z}\}, j \in \mathbf{Z}$ 时还有:

$$W_j \perp W_{j'}, W_j \perp V_{j'}, W_j \oplus V_{j'} = V_{j+1}, j, j' \in \mathbf{Z} \quad (4)$$

式(2)为双尺度方程, 式(3)为构造方程, $\phi(t)$ 为尺度函数, $\int_{-\infty}^{\infty} \phi(t) dt = 1$, $\psi(t)$ 为小波函数, $\int_{-\infty}^{\infty} \psi(t) dt = 0$. 定理 1 指出了尺度函数和正交基的存在性, 定理 2 则给出了具体的构造过程, 从定理 2 可知, $A_j f(t) \in V_j, D_j f(t) \in W_j, \langle A_j f(t), D_j f(t) \rangle = 0$.

2.2 Mallat 算法^[5-6]

设 $f(t) \in V_j, \{V_j; j \in \mathbf{Z}\}$ 是一给定分辨率为 2^j 的多分辨率分析, $\phi(t)$ 和 $\psi(t)$ 分别是尺度函数和小波函数, 由于 $f(t) \in V_j$, 故有: $f(t) = A_j f(t) = \sum_{k \in \mathbf{Z}} c_{j,k} \phi_{j,k}(t)$ (5)

记 $h(k) = \frac{1}{2} \int_{-\infty}^{\infty} \phi(\frac{t}{2}) \bar{\phi}(t+k) dt, g(k) = (-1)^{1-k} h(1-k)$, 从而 $A_j f(t) = A_{j+1} f(t) + D_{j+1} f(t)$ (6)

$A_{j+1} f(t) = \sum_{m=-\infty}^{\infty} c_{j+1,m} \phi_{j+1,m}(t), D_{j+1} f(t) = \sum_{m=-\infty}^{\infty} d_{j+1,m} \psi_{j+1,m}(t)$ (7)

其中, $c_{j+1,m} = \sum_{k=-\infty}^{\infty} \bar{h}(k-2m) c_{j,k}, d_{j+1,m} = \sum_{k=-\infty}^{\infty} \bar{g}(k-2m) d_{j,k}$ (8)

记矩阵 $H = (H_{m,k}), G = (G_{m,k})$, 其中 $H_{m,k} = \bar{h}(k-2m), G_{m,k} = \bar{g}(k-2m)$ 则有:

$$c_{j+1} = H c_j, d_{j+1} = G c_j \quad (j=0,1,\dots,J) \quad (9) \quad c_j = H^* c_{j+1} + G^* d_{j+1} \quad (j=J,\dots,1,0) \quad (10)$$

其中, H^*, G^* 分别为 H, G 的共轭转置矩阵. (9) 式便是 Mallat 的塔式分解算法,

(10)式则为重构算法. 以上过程如同对信号 $f(t) \in V_j$ 共轭镜像滤波, 故 H 、 G 分别称为分解低通、高通滤波器, H^* 、 G^* 则分别称为重构低通、高通滤波器. 因此, $A_j f(t)$ 可以理解为函数 $f(t)$ 频率不超过 2^j 的分量, 而 $D_j f(t)$ 是函数 $f(t)$ 的频率介于 2^j 和 2^{j+1} 的分量.

3 样本数据分解及空间划分

令 $A_j f(t), 0 \leq j \leq J$ 为尺度级数为 j 时的函数 $f(t)$ 的近似分量, $D_j f(t), 0 \leq j \leq J$ 为尺度级数为 j 时的细节分量, 根据 MRA 和 Mallat 算法, 对正交小波有如下关系式成立:

$$D_j f(t) = \sum_{k \in Z} c_{j,k} \psi_{j,k}(t), A_j f(t) = \sum_{j > J} D_j f(t), A_{j-1} f(t) = A_j f(t) + D_j f(t), f(t) = A_j f(t) + \sum_{k \leq j} D_j f(t) \quad (11)$$

因此, $f(t) \in L^2(\mathbf{R})$ 可由 $(A_j f(t), (D_j f(t))_{1 \leq j \leq J})$ 完全刻画, 在所有的尺度上信息都无损失, 且 $f(t)$ 可由其分量精确重构; 同时 Mallat 算法表明, 通过小波分析, 可以将函数 $f(t)$ 分解成不同频率通道的成分, 并将每一频率通道的成分又按相位进行了分解, 频率越高者相位划分越细, 反之则越疏, 这一现象为本文样本数据聚类及空间划分研究、混合多网络模型拓扑结构的确定与每一子网络功能的分配提供了直接的理论依据.

设有样本数据集 $S = \{(X, Y)_k; k=1, \dots, N\}$, N 为样本数, $X = (x_1, x_2, \dots, x_n)$, $Y = (y_1, y_2, \dots, y_m)$, n, m 分别为样本中输入和输出的变量数目. 在 S 中数据通常可以根据其某些特点进行分解, 并进一步分配到相应的子神经网络模型中去. 现对 S 中各变量进行小波分解, 令:

$$A_j X = (A_j x_1, A_j x_2, \dots, A_j x_n), A_j Y = (A_j y_1, A_j y_2, \dots, A_j y_m), D_j X = (D_j x_1, D_j x_2, \dots, D_j x_n), D_j Y = (D_j y_1, D_j y_2, \dots, D_j y_m), 1 \leq j \leq J. \text{ 记:}$$

$$SV_j = \{(A_j X, A_j Y); k=1, \dots, N\}, SW_j = \{(D_j X, D_j Y); k=1, \dots, N\},$$

根据 MRA 和定理 2, 一定有:

$$SW_j \perp SW_{j'}, SV_j \perp SV_{j'}, SW_j \oplus SV_{j'} = SV_{j+1}, j, j' \in Z \quad (12)$$

根据 (11) 式, 样本数据空间 S 可以分解为一系列正交子空间的直和, 即:

$$S = SW_J \oplus SV_J \oplus SW_{J-1} \oplus SV_{J-2} \oplus SV_J \oplus SV_1 \quad (13)$$

小波-神经网络混合模型的目的是完成非线性映射关系: $F: X \in S \rightarrow Y \in S$ (14)

每一个子神经网络的目的是完成相应子空间内的非线性映射关系:

$$F_j: D_j X \in SW_j \rightarrow D_j Y \in SW_j, F_j: A_j X \in SV_j \rightarrow A_j Y \in SV_j, 1 \leq j \leq J \quad (15)$$

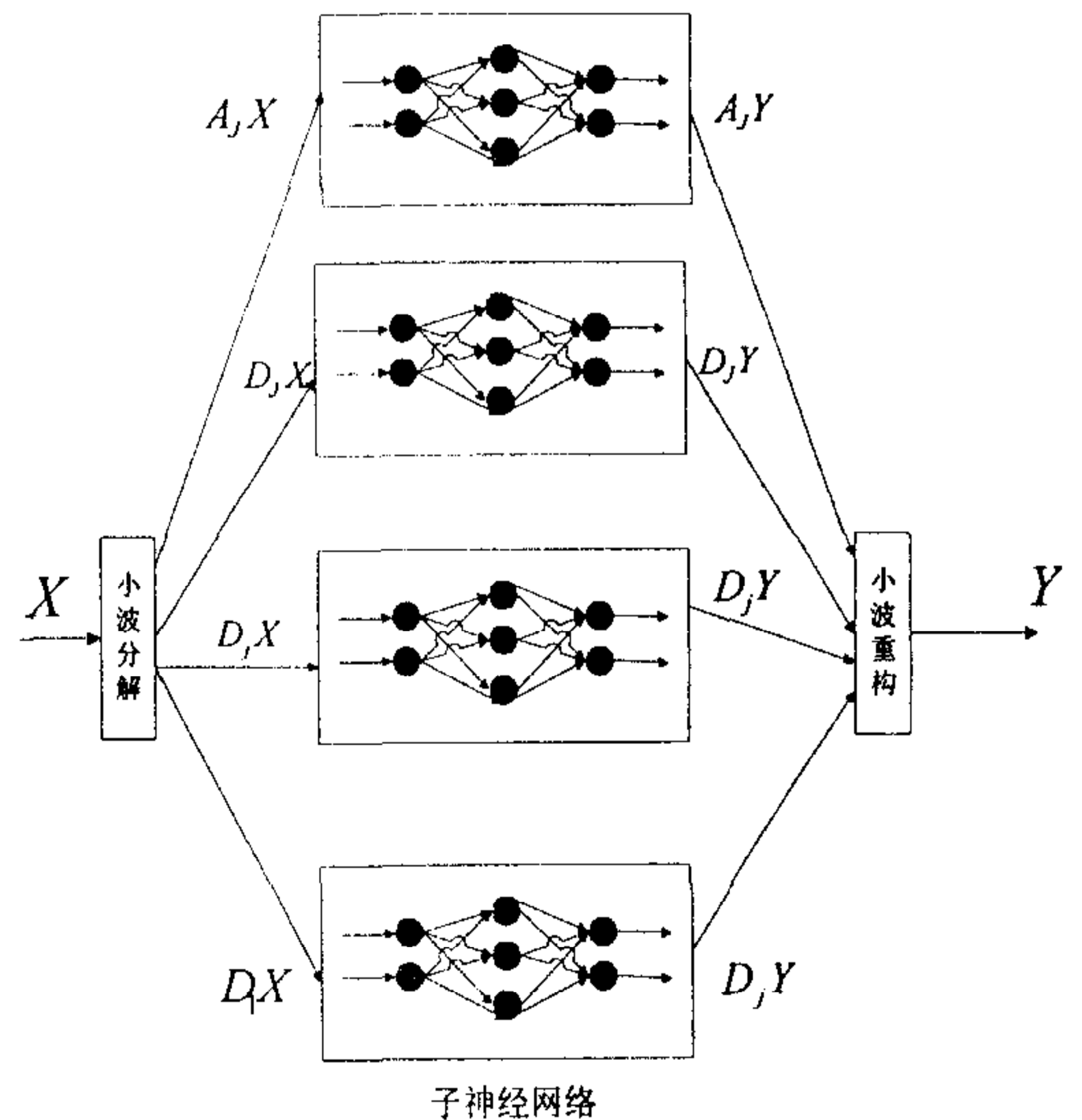


图1 小波-神经网络混合模型

4 小波-神经网络混合模型

根据 (12) ~ (15) 式, 可形成如图 1 所示的小波-神经网络多变量混合模型.

模型输出的直接重构信息为:
$$Y = A_j Y + \sum_{1 \leq j \leq J} D_j Y \quad (16)$$

模型输出的加权重构信息为:
$$Y = W_0 A_j Y + \sum_{1 \leq j \leq J} W_j D_j Y \quad (17)$$

其中, $W = [W_0, W_1, \dots, W_J]^T$ 为加权系数矢量, 根据各级分量的能量大小进行确定. 从定理 2 及 (12) ~ (15) 可知, 由于模型输出 $A_j Y, D_j Y, \dots, D_1 Y$ 相互独立, 根据信息论可知

此时由(16)~(17)式得到的信息不确定性最小,即信息熵最小,因此整个模型的可靠性、鲁棒性和适应性能得到提高,而且模型的结构严格满足文献[1]提出的准则。

在此模型中,充分利用了小波分析和神经网络各自的特点,将小波分析对信息很强的分解和重构能力与神经网络处理多变量非线性映射的能力很好地结合了起来,发挥了各自的优势,而且混合模型的拓扑结构是唯一确定的。设小波分析的最大尺度级数为 J ,则模型中神经网络的数目 $n_i = J + 1$ 。

模型中所使用的神经网络可以是任何一种具有非线性映射能力的前向网络,如BP(Back Propagation)网络、RBF(Radial Basis Function)网络等,但是BP网络存在局部极小、训练时间过长等问题,建议采用在此方面性能均优于BP网络的RBF网络^[5]。

5 模型分析

图 1 所示的多变量混合过程模型可根据不同的应用目的进行分析:(1) 信噪分离器。此时, $D_j Y, 0 < j \leq J$ 可大致理解为位于不同频带中的信号噪声,而 $A_j Y$ 为相应的真实信号,通过混合模型达到了多变量信噪分离的目的。(2) 噪声估计器。通过信号在不同频带的分解,可以在多个尺度对过程噪声的性质及程度进行估计。(3) 信号逼近器。根据(11)式, $A_{j-1} Y = A_j Y + D_j Y, 1 \leq j \leq J$,所以可对过程变量在任意尺度上逼近。(4) 信号滤波器。在信号重构时,可以对某一频带信号成分进行取舍达到滤波的目的。

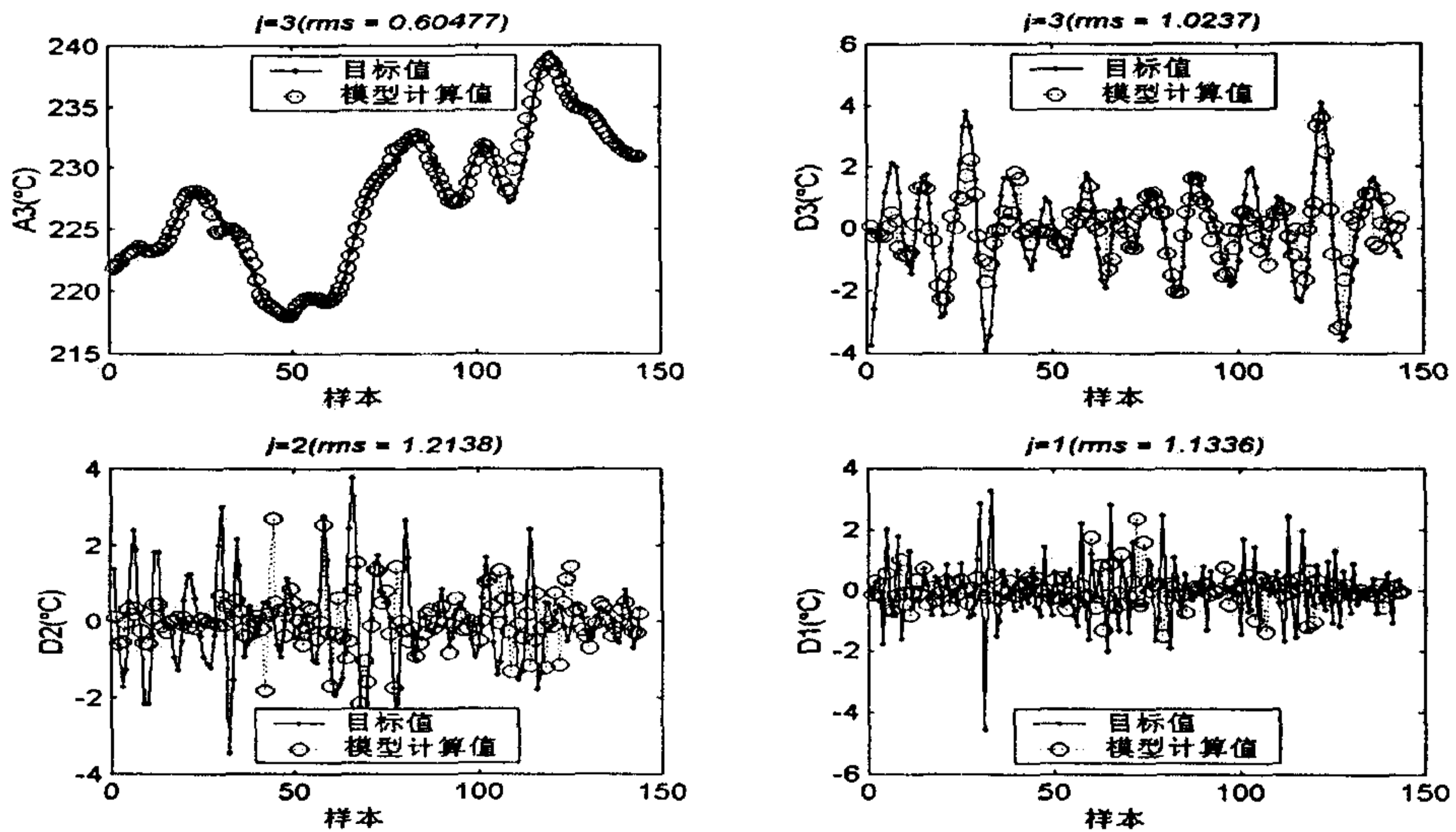


图 2 小波-神经网络混合过程模型输出各分量 ($J = 3$)

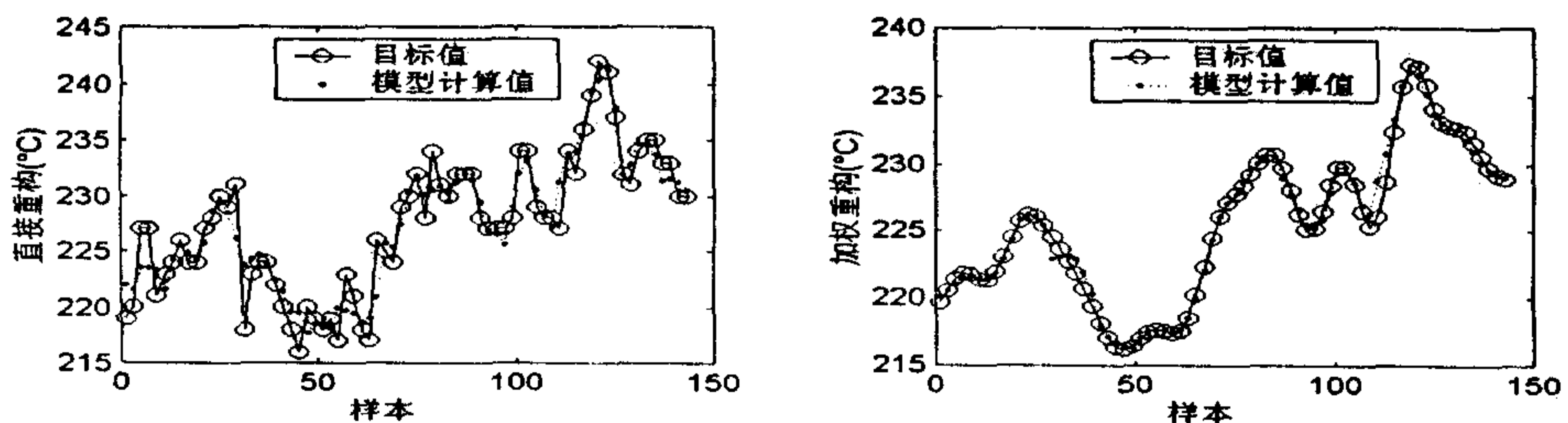


图 3 小波-神经网络混合模型输出重构结果 ($J = 3$)

6 应用实例

本文将以上混合过程模型应用于某常压塔顶粗汽油干点的软测量估计,建模时,输入变量取原油物性综合参数、进料流量、塔顶温度、塔顶压力、二线采出温度、三线采出温度等 6 个变量,输出为塔顶粗汽油干点 1 个变量.样本数据长度 $N = 288$,一半用于训练,一半用于验证.采用 Daubechies10 小波^[5],分解尺度 $J = 3$,使用 RBF 神经网络中的 GRNN(generalized regression neural networks)网络,模型训练精度 < 0.001 (规范后),采用小波-神经网络混合过程模型的应用结果如图 2~3 所示.

从图 2~3 可以看出,小波-神经网络混合模型能在不同尺度上得到信号的近似分量和细节分量,建模的同时还达到了信号中各分量分离的目的.本文采用均方根误差 rms (root of mean squared error)和输出序列中的最大误差 e_{max} 对单一模型和混合模型的应用结果进行比较研究,见表 1,从表中可以看出即使单一模型比混合模型训练要好的情况下,在验证阶段混合模型比单一模型有更大的优势,各项性能指标均优于单一模型.

表 1 不同模型的训练与验证结果比较

比较项目		单一模型	小波-神经网络混合模型(3级)				
			A_3	A_2	A_1	直接重构	加权重构
训练	rms	0.1025	0.1506	0.7424	0.9158	1.1095	0.1499
	e_{max}	0.5636	0.7185	2.6748	3.4413	4.4778	0.7175
验证	rms	2.9742	0.6048	1.2106	1.7243	2.0694	0.6003
	e_{max}	9.0022	2.7434	4.1954	5.8772	6.0427	2.7236

7 结论

为了提高神经网络模型的可靠性和鲁棒性,增强模型的抗干扰能力,本文提出了一种小波-神经网络多变量混合过程模型,将小波分析对信息分解和重构的能力与神经网络处理多变量非线性映射的能力很好地结合了起来,发挥了各自的优势.对过程数据样本空间根据其在小波变换下的“频带”特征进行自适应的划分,形成每个子神经网络的子数据样本空间,然后用多个神经网络来描述和解释整个数据空间.该混合过程模型在实际中可以应用于不同的目的,如信噪分离、噪声估计、信号逼近等.最后,将该模型应用于常压塔顶粗汽油干点估计,结果验证了它的合理性和有效性.

参 考 文 献

- 熊智华,王雄,徐用懋,基于广义信息熵融合的多神经网络建模方法,信息与控制,1999,28(增刊):346~349
- Yang Erfu, Zhang Zhenpeng, Liu Guoqiu, et al, A real-time system for condition monitoring of propulsion system using BP-ART hybrid neural networks, *Proceedings of the Second Asian-Pacific Conference on Aerospace Technology and Science*, Chinese Society of Aeronautics and Astronautics, October 6-10, 1997, Dun Huang, Gansu, China: 410—415
- 赵振宇,徐用懋,模糊理论和神经网络的基础与应用,北京:清华大学出版社,1996
- 杨尔辅,张振鹏,刘国球,小波分析及其在推进系统健康监控中的应用,推进技术,1997,18(1):73~78
- 刘贵忠,邸双亮,小波分析及其应用,西安:西安电子科技大学出版社,1992
- 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