

动态非线性连续时间系统的小波神经网络辨识*

张兆宁¹, 喻文焕², 郁惟镛¹

(1. 上海交通大学 电力工程系, 上海 200030; 2. 天津大学 数学系, 天津 300072)

摘要: 将小波神经网络应用于动态非线性连续时间系统的辨识, 同时为了使神经网络的训练达到全局最优和加速小波神经网络训练的收敛速度, 提出了信赖域算法, 并研究了信赖域算法的收敛性. 随后进行了算例仿真, 证明了所提辨识方法的有效性.

关键词: 小波; 神经网络; 非线性连续时间系统; 辨识

中图分类号: O174.2

文献标识码: A

Continuous time nonlinear system identification with wavelet neural networks

ZHANG Zhao-ning¹, YU Wen-huan², YU Wei-yong¹

(1. Department of Electric Power Engineering, Shanghai Jiaotong University, Shanghai 200030, China;

2. Department of Mathematics, Tianjin University, Tianjin 300072, China)

Abstract: The wavelet neural networks are used for identification of continuous time nonlinear systems. For the improvement of convergent speed and the global optimization in the training of wavelet neural networks, the trust region algorithm is proposed for the training. Then we study the convergence properties about the trust region algorithm. Finally, the simulation results are given to illustrate the efficiency of the method proposed.

Key words: wavelets; neural networks; nonlinear continuous time system; identification

1 引言(Introduction)

20 世纪 80 年代后期以来, 系统的非参数辨识方法的研究颇受人们重视, 其主要原因是这类方法要求的假设条件低于以前的参数辨识方法, 通常它们只要求依照测量空间的某种范数量测噪声有界, 而且它们能够适用于一般的时不变线性系统. 详细情况可参见文献[1~5], 尤其是综述性文献[4, 5]介绍了有关的大量参考文献. 上述文献主要研究处理线性系统的辨识, 而文献[6]则用正交基函数来研究非线性系统的辨识.

众所周知, 人工神经网络具有很强的非线性逼近能力, 因此被用来作为非线性系统的非参数辨识工具. 文献[7]简单介绍了这方面的应用情况.

近年来深受人们青睐的小波分析有很多优点, 由小波构造的级数不仅具有 Fourier 级数的正交规范性质(该性质使得级数的系数易于计算), 而且还具有良好的局部性质(即常用的母波函数、尺度函数具有紧支集), 这使计算进一步简化. 故已有人用母

波函数或尺度函数作为人工神经网络的基函数, 从而构成一类新的非参数辨识研究方法. 文献[8, 9]提出了用小波神经网络来辨识一些动态连续时间非线性系统类型.

本文研究了应用小波神经网络来辨识动态连续时间非线性系统的问题, 给出了一类比较复杂的动态连续时间非线性系统辨识方法. 考虑到神经网络通常使用的最速下降算法的收敛较慢且常常陷于局部最优, 本文提出了信赖域算法来改善网络的收敛速度与大范围性质, 接着对其收敛性给出了证明. 最后的算例仿真说明了所提辨识方法的可行性.

2 小波神经网络(Wavelet neural network)

设 $\varphi(t)$ 是具有紧支集 $\text{supp } \varphi = [a, b]$ 的尺度函数, 选取与 φ 相伴随并由它按一定方式产生的母波函数 $\Psi(t)$ 也是具有紧支集 $\text{supp } \Psi = [c, d]$. 经过二进制的平移和伸缩后, 生成函数系 $\{\varphi_{n,k}(t)\}_{n,k \in \mathbb{Z}}$ 与函数系 $\{\Psi_{n,k}(t)\}_{n,k \in \mathbb{Z}}$, 其中

$$\varphi_{n,k}(t) = 2^{n/2} \varphi(2^n t - k), \quad n, k \in \mathbb{Z}, \quad (2.1)$$

* 基金项目: 国家自然科学基金(69774012)资助项目.

收稿日期: 2000-09-26; 收修改稿日期: 2001-06-04.

$$\Psi_{n,k}(t) = 2^{n/2} \varphi(2^n t - k), \quad n, k \in \mathbb{Z}, \quad (2.2)$$

当自变量取自高维空间, 即 $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, 则可以由 φ 及 Ψ 按照张量积方式形成 d 元函数

$$\begin{aligned} \phi(x) &= \varphi(x_1) \cdots \varphi(x_d), \\ \Psi(x) &= \Psi(x_1) \cdots \Psi(x_d). \end{aligned} \quad (2.3)$$

因而得到

$$\phi_{n,k}(x) = 2^{dn/2} \varphi(2^n x_1 - k_1) \cdots \varphi(2^n x_d - k_d),$$

$$\Psi_{n,k}(x) = 2^{dn/2} \Psi(2^n x_1 - k_1) \cdots \Psi(2^n x_d - k_d).$$

其中 $k = (k_1, \dots, k_d) \in \mathbb{Z}^d$. 根据小波理论, 对于 n , $V_n = \text{span}\{\varphi_{n,k}\}_{k=-\infty}^{\infty}$ 张成 $L^2(\mathbb{R}^d)$ 的一个线性子空间, 这些线性子空间序列构成 $L^2(\mathbb{R}^d)$ 的一个多分辨率分析; 而 $\{\Psi_{n,k}(x)\}_{n \in \mathbb{Z}, k \in \mathbb{Z}^d}$ 则成为 $L^2(\mathbb{R}^d)$ 的一个规范正交基.

于是, 根据小波理论, 对任意 $f \in L^2(\mathbb{R}^d)$, 当取 M 较大时, 我们可以用下列近似式来表示 $f(x)$:

$$f(x) = \sum_{k \in \mathbb{Z}^d} \alpha_k \phi_{M,k}(x). \quad (2.4)$$

从形式上看, 级数(2.4)有无穷多项, 通过简单的证明可知, 实际上只有有限项, 这是由于信号存在的时间区间有限之故.

因此式(2.4)成为

$$f(x) = \sum_{k \in B} \alpha_k \phi_{M,k}(x). \quad (2.5)$$

其中 B 是 $\mathbb{R}^d$ 中的一个超长方体.

所以式(2.4)可以由一个小波神经网络来逼近, 小波神经网络由输入层、隐层和输出层决定; 网络的输入是 x , 输出是 $f(x)$, 权值是 W , 由 α_k 决定.

3 一类非线性连续时间动态系统的辨识 (The identification of a class of nonlinear continuous time system)

3.1 系统的小波级数表示(Representation of systems by wavelet series)

我们考虑下面动态非线性连续时间动态系统:

$$\begin{aligned} \dot{x} &= f(x, u), \quad 0 < t < T, \\ y &= g(x, u), \quad 0 \leq t \leq T. \end{aligned} \quad (3.1)$$

其中 $x \in \mathbb{R}^n$ 是系统的状态向量, $u \in \mathbb{R}^m$ 是系统的输入(控制)向量, 而 $y \in \mathbb{R}^d$ 是系统的输出(观测)向量.

为了辨识此系统, 我们的任务是根据输入向量 $u(t)$ 及输出向量 $y(t)$ 来确定未知的非线性函数

$$f: \mathbb{R}^{n+m} \rightarrow \mathbb{R}^n \text{ 与 } g: \mathbb{R}^{n+m} \rightarrow \mathbb{R}^d.$$

如果 f 与 g 已知, 则可以将它们展开成小波级数:

$$\begin{aligned} f(x, u) &= \sum_{k \in B} \alpha_k \phi_{M,k}(x, u), \\ g(x, u) &= \sum_{k \in B} \beta_k \phi_{M,k}(x, u). \end{aligned} \quad (3.2)$$

其中 ϕ 是由尺度函数 φ 按张量积方式生成, 这样

$$\begin{aligned} \alpha_k &= \langle f, \phi_{M,k} \rangle = \int_{\mathbb{R}^{n+m}} f(x, u) \phi_{M,k}(x, u) dx du, \\ \beta_k &= \langle g, \phi_{M,k} \rangle = \int_{\mathbb{R}^{n+m}} g(x, u) \phi_{M,k}(x, u) dx du. \end{aligned} \quad (3.3)$$

其中 $\alpha_k \in \mathbb{R}^n, \beta_k \in \mathbb{R}^d$. 但是 f 与 g 未知, 无法用(3.3)来求出 α_k 和 β_k , 因此我们考虑下列模型:

$$\begin{aligned} \dot{\hat{x}} &= \sum_{k \in B} \alpha_k \phi_{M,k}(\hat{x}, u), \\ \hat{y} &= \sum_{k \in B} \beta_k \phi_{M,k}(\hat{x}, u). \end{aligned} \quad (3.4)$$

其中 $\hat{x} \in \mathbb{R}^n, \hat{y} \in \mathbb{R}^d$ 分别是模型的状态与输出.

假定指标集合 B 的长度为 K , 则未知的系数向量是 $\gamma = (\alpha_1^T, \dots, \alpha_K^T, \beta_1^T, \dots, \beta_K^T)^T \in \mathbb{R}^{(n+d)K}$. 显然模型的输出量 $\hat{y}$ 依赖于系数向量 γ , 记为 $\hat{y}(t) = \hat{y}(t; \gamma)$. 为了使得模型(3.4)追踪实际系统(3.1), 应当选择系数向量 γ 使得下述目标泛函取得极小:

$$J(\gamma) = \int_0^T \|y(t) - \hat{y}(t; \gamma)\|^2 dt \rightarrow \min J(\gamma). \quad (3.5)$$

很容易证明: $\hat{y}$ 是 γ 的连续函数, 因此存在 $\hat{\gamma}$ 使得式(3.5)达到极小值:

$$J(\hat{\gamma}) = \min_{\gamma \in \mathbb{R}^{(n+d)K}} J(\gamma). \quad (3.6)$$

3.2 最速下降法(Steepest descent method)

为了计算 $\hat{\gamma}$, 使用最速下降法, 故必须计算 J 对向量 γ 的梯度:

$$G = \text{grad} J = \left(\left(\frac{\partial J}{\partial \alpha_1} \right)^T, \dots, \left(\frac{\partial J}{\partial \alpha_K} \right)^T, \left(\frac{\partial J}{\partial \beta_1} \right)^T, \dots, \left(\frac{\partial J}{\partial \beta_K} \right)^T \right)^T. \quad (3.7)$$

其中各分量(也是向量)计算如下:

$$\begin{cases} \frac{\partial J}{\partial \alpha_k} = 2 \int_0^T \left(\frac{\partial \hat{y}}{\partial \alpha_k} \right)^T (\hat{y} - y) dt, \\ \frac{\partial J}{\partial \beta_k} = 2 \int_0^T \left(\frac{\partial \hat{y}}{\partial \beta_k} \right)^T (\hat{y} - y) dt, \quad k = 1, \dots, K. \end{cases} \quad (3.8)$$

通过模型(3.4), 容易计算(3.8)中的 $d \times n$ 矩阵 $\frac{\partial \hat{y}}{\partial \alpha_k}$

及 $d \times d$ 的方阵 $\frac{\partial \hat{y}}{\partial \beta_k}$:

$$\begin{cases} \frac{\partial \hat{y}}{\partial \alpha_k} = \sum_{i \in B} \beta_i \left[\frac{\partial \phi_{M,i}(\hat{x}, u)}{\partial \hat{x}} \right]^T \frac{\partial \hat{x}}{\partial \alpha_k}, \\ \frac{\partial \hat{y}}{\partial \beta_k} = I_d \phi_{M,k}(\hat{x}, u). \end{cases} \quad (3.9)$$

其中 I_d 是 $d \times d$ 的单位方阵, 而 $n \times n$ 的方阵 $\frac{\partial \hat{x}}{\partial \alpha_k}$ 由下列微分方程组确定:

$$\begin{cases} \frac{d}{dt} \left(\frac{\partial \hat{x}}{\partial \alpha_k} \right) = \sum_{i \in B} \alpha_i \left[\frac{\partial \phi_{M,i}(\hat{x}, u)}{\partial \hat{x}} \right]^T \frac{\partial \hat{x}}{\partial \alpha_k} + I_n \phi_{M,k}(\hat{x}, u), \\ \left. \frac{\partial \hat{x}}{\partial \alpha_k} \right|_{t=0} = 0. \end{cases} \quad (3.10)$$

其中 $\hat{x}$ 由(3.4)确定.

3.3 信赖域算法(Trust region algorithm)

众所周知, 最速下降法往往只是开始下降快, 后来收敛很慢, 为此我们采用信赖域法来改善收敛速度. 为此引进一些定义.

定义 信赖域的子问题的目标泛函为

$$\theta_l(h) = \min_{h \in \mathbb{R}^{(n+d)K}, \|h\| \leq \Delta_l} G_l^T h. \quad (3.11)$$

其中

$$h = (a_1^T, \dots, a_K^T, b_1^T, \dots, b_K^T)^T \in \mathbb{R}^{(n+d)K}, \quad a_i \in \mathbb{R}^n, \quad b_i \in \mathbb{R}^d.$$

显然子问题(3.11)的解为

$$h_l = - \frac{\Delta_l}{\|G_l\|} G_l, \quad (3.12)$$

相应的最小值为 $\theta_l(h_l) = -\Delta_l \|G_l\|$. 再定义预估下降量为

$$\text{Pre}h_l = \theta(0) - \theta(h_l) = \Delta_l \|G_l\|, \quad (3.13)$$

以及实际下降量

$$\text{Are}h_l = J(\gamma^{(l)}) - J(\gamma^{(l)} + h_l). \quad (3.14)$$

最后定义比值

$$r_l = \frac{\text{Are}h_l}{\text{Pre}h_l}. \quad (3.15)$$

于是有下列信赖域算法:

第 1 步 给出 $\gamma_1 \in \mathbb{R}^{(n+d)K}$, $\Delta_1 > 0$, $\varepsilon > 0$, $\delta_1 > 1 > \delta_2 > 0$, $0 < \delta_3 < \delta_4$, 置 $l := 1$.

第 2 步 通过式(3.7) ~ (3.10) 计算 G_l , 如果 $\|G_l\| \leq \varepsilon$, 则停止计算; 否则由式(3.12)计算出 h_l .

第 3 步 根据式(3.13) ~ 式(3.15) 计算 r_l . 令

$$\gamma^{(l+1)} = \begin{cases} \gamma^{(l)} + h_l, & r_l > 0, \\ \gamma^{(l)}, & r_l \leq 0. \end{cases} \quad (3.16)$$

选取 Δ_{l+1} 使其满足:

$$\Delta_{l+1} \in \begin{cases} [1, \delta_1] \|h_l\|, & r_l \geq \delta_2, \\ [\delta_3, \delta_4] \|h_l\|, & \text{否则}. \end{cases}$$

第 4 步 置 $l := l + 1$, 转向第 2 步.

3.4 信赖域算法的收敛性(Convergence of trust region algorithm)

引理 1 设 $G(\gamma) = \nabla J(\gamma)$ 由式(3.8) 确定,

若由信赖域算法产生的点列 $\{G_l\}$ 满足

$$\|G_l\| \geq \delta > 0. \quad (3.17)$$

其中 δ 是正常数, 则存在正常数 δ_5 使得

$$\|h_l\| \geq \delta_5, \quad l = 1, 2, \dots.$$

证 根据公式(3.12): $\|h_l\| = \Delta_l$. 故只需证明存在 δ_5 使得

$$\Delta_l \geq \delta_5, \quad l = 1, 2, \dots. \quad (3.18)$$

假设式(3.18)不成立, 于是存在子序列 $\{l_i\}$ 使得

$$\lim_{i \rightarrow \infty} \Delta_{l_i} = 0.$$

不妨设 $\Delta_{l_i} < \Delta_{l_{i-1}}$ (即是严格单调减少序列). 由于 $\text{Pre}h_{l_i} = \Delta_{l_i} \|G_{l_i}\| \geq \delta \Delta_{l_i}$, 以及 $G(\gamma)$ 的一致连续性, 便有

$$\begin{aligned} \text{Are}h_{l_i} &= J(\gamma_{l_i}) - J(\gamma_{l_i} + h_{l_i}) = \\ &= G_{l_i}^T h_{l_i} + o(\|h_{l_i}\|) = \text{Pre}h_{l_i} + o(\|h_{l_i}\|), \end{aligned}$$

从而得到

$$\lim_{i \rightarrow \infty} r_{l_i} = \lim_{i \rightarrow \infty} \frac{\text{Are}h_{l_i}}{\text{Pre}h_{l_i}} = 1. \quad (3.19)$$

于是当 i 充分大时, 便有 $r_{l_i} > \delta_2$, 由信赖域算法可得知:

$$\Delta_{n_i} \geq \|h_{l_{i-1}}\| \Delta_{l_{i-1}}.$$

这与 $\Delta_{l_i} < \Delta_{l_{i-1}}$ 的假设相矛盾. 故引理得证.

定理 1 由信赖域算法产生的梯度序列 $\{G_l\}$ 必定满足

$$\liminf_{l \rightarrow \infty} \|G_l\| = 0. \quad (3.20)$$

证 假设式(3.20)不成立, 则必存在常数 δ 使得式(3.17)成立. 定义集合

$$I = \{k; r_k \geq \delta_2\}. \quad (3.21)$$

根据引理 1 的证明知: 集合 I 的基数为可列无穷.

由于 $J(\gamma_k) \geq 0$, 故而有

$$\begin{aligned} +\infty &> \sum_{k=1}^{\infty} [J(\gamma_k) - J(\gamma_{k+1})] \geq \sum_{k \in I} [J(\gamma_k) - J(\gamma_{k+1})] \geq \\ &\sum_{k \in I} \delta_2 \text{Pre}h_k = \delta_2 \sum_{k \in I} \Delta_k \|G_k\| \geq \delta_2 \sum_{k \in I} \delta_5 \delta = +\infty. \end{aligned}$$

显然是矛盾的, 从而式(3.20)成立.

注 由于算法对初始条件没有要求, 固信赖域法是一种大范围收敛法.

4 算例仿真(Simulation of an example)

为了验证提出方法的可行性, 我们考虑下列受控 Van der Pol 方程

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = \mu(1 - x_1^2)x_2 - x_1 + u(t), \\ y = x_1. \end{cases} \quad (4.1)$$

当用式(4.1)描述电路中的 $R-L-C$ 回路的电流运动特性时,参数 $\mu = \frac{\sigma M - RC}{\sqrt{LC}}$, 其中 σ 是跨导, M 是互感, R 是电阻, C 是电容, L 为电感. 可将其设定为 $\mu = 1$. $u(t)$ 是输入函数, $y(t)$ 是输出函数, 它们都是能测量的量, 而 (x_1, x_2) 则表示系统的状态.

在建立模型测试时, 我们使用下列输入量:

$$u(t) = 0.5 \sin \frac{2\pi}{15}t + 0.2 \sin \frac{2\pi}{40}t.$$

而在训练时, 使用均匀分布于 $[-0.5, 0.5]$ 的伪随机变量作为系统的输入.

编程仿真的实际输出 $\tilde{y}(t)$ 与理想输出 $y(t)$ 结果为:

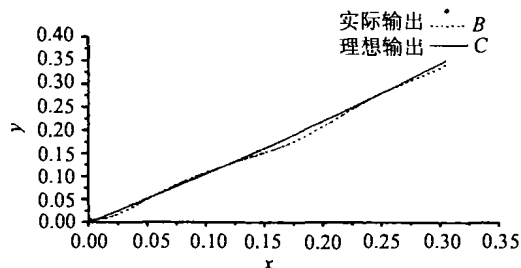


图1 实际输出曲线与理想输出曲线

Fig. 1 Actual output and ideal output curves

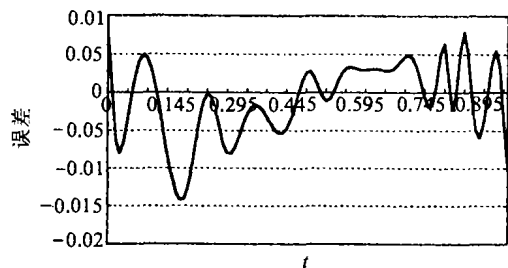


图2 误差曲线

Fig. 2 Error curve

5 结论(Conclusions)

本文将小波神经网络应用于动态非线性连续时间系统的辨识问题, 同时考虑到神经网络通常应用的最速下降算法的局限性, 为了使训练达到全局最优和加速小波神经网络训练的收敛速度, 提出了信

赖域算法, 并给出了信赖域算法的收敛性研究结果. 随后的算例仿真证明了所提辨识方法的有效性.

参考文献(References)

- [1] Helmicki A J, Jacobson C A, Nett C N. Control oriented system identification: a worst-case/deterministic approach in H_∞ [J]. IEEE Trans. on Automatic Control, 1991, 36(10): 1163 - 1176
- [2] Gu G, Khargonekar P P. A class of algorithms for identification in H_∞ [J]. Automatica, 1992, 28(2): 299 - 312
- [3] Wei C, Guo L. H_∞ -identification of infinite-dimensional linear stochastic systems [J]. Systems Science and Mathematical Sciences, 1995, 8(2): 88 - 96
- [4] Milance M, Vicino A. Information-based complexity and nonparametric worst-case system identification [J]. Journal of Complexity, 1993, 9(3): 427 - 446
- [5] Milance M, Vicino A. Optimal estimation theory for dynamic systems with set membership uncertainty: an overview [J]. Automatica, 1997, 27(6): 997 - 1009
- [6] Nelles O. Orthonormal basis functions for nonlinear system identification with local linear model trees (LOLIMOT) [A]. In Y. Sawaragi and S. Sagara. Preprints of 11th IFAC Symp. on System Identification [C]. Kitakyushu, Japan, 1997, 2: 667 - 672
- [7] Narendra K S, Mukhopadhyay S. Neural networks for system identification [A]. In Y. Sawaragi and S. Sagara. Preprints of 11th IFAC Symp. on System Identification [C]. Kitakyushu, Japan, 1997, 2: 763 - 269
- [8] Ho D W C, Xu J. Wavelet neural networks for continuous time nonlinear system identification [A]. In H. F. Chen and B. Whalberg. Preprints of 14th World Congress of IFAC [C]. Beijing, China, 1999, 115 - 120
- [9] Liu G P, Billings S A, Kadir kamanathan V. Nonlinear identification via variable wavelet networks [A]. In H. F. Chen and B. Whalberg. Preprints of 14th World Congress of IFAC [C]. Beijing, China, 1999, 109 - 114

本文作者简介

张兆宁 1964年生, 副教授, 1999年于天津大学博士毕业. 研究方向为小波分析, 神经网络及其应用.

喻文焕 1940年生, 教授, 1989年毕业于美国康涅狄克大学, 获得博士学位. 主要从事系统辨识, 最优控制, 偏微分方程反问题的理论与应用研究.

郁惟镛 1940年生, 教授, 博士生导师, 1964年于上海交通大学获得学士学位. 长期从事人工智能在继电保护中的应用研究.