

# 神经网络并联辨识算法的收敛性研究

卢 进 徐文立 韩曾晋

(清华大学自动化系·北京, 100084)

**摘要:** 神经网络可用来建立非线性动态系统的模型, 其辨识模型可分为串并联辨识模型和并联辨识模型两种. 后者的思想源于基于参考模型自适应方案的输出误差辨识模型, 对观测扰动有较强的抑制能力. 本文对这种神经网络并联辨识结构的收敛性进行了研究, 指出在网络参数满足一定条件时并联预测过程收敛, 且并联辨识算法具有局部收敛性. 仿真实验验证了上述结论.

**关键词:** 神经网络; 非线性动态系统; 并联辨识算法

## 1 概 述

前馈型神经网络的一类重要应用是对非线性系统进行逼近. 由于在处理大规模非线性问题时的简单一致性和潜在的并行处理机制, 神经网络技术在非线性动态系统的建模、控制中有其独特的优点, 在某些应用中实现了传统控制理论难以完成的控制, 展示出良好的应用前景. Narendra<sup>[1]</sup>提出过一种动态 BP 网络模型, 通过在输入端引入输出的延迟反馈来构成动态结构, 学术界对其评价颇高. 然而, 神经网络建模技术是基于数据的, 无论采用哪种结构或学习算法, 辨识模型的建立都完全依赖于数据, 这就要求我们认真地对待数据噪声问题. Narendra 等人在[1]中提出两种辨识结构(图 1(a)(b)). 图 1(a)称为串并联辨识方式, 这是我们目前在绝大多数实际应用中采用的辨识方式, 但由于所有延迟的输出数据都取自系统的实际观测值, 因而对数据要求高. 图 1(b)称为并联辨识方式, 辨识模型的输出只和自己过去的值有关, 和对象的观测过程无关, 因而相对串并联辨识方式来讲对扰动有较强的抑制能力. 但正如 Narendra 等人在[1]中指出的, 目前还没有理论结果保证这种并联辨识方式的参数收敛或者保证此并联辨识模型与实际系统的输出偏差收敛. 本文针对单输出三层前馈型神经网络的并联辨识结构进行了研究, 指出在网络参数满足一定条件情况下, 并联预测方式的输出偏差收敛, 并且指出, 通过对经典 BP 算法进行适当的修改, 可以保证并联辨识方式的学习过程具有局部收敛性.

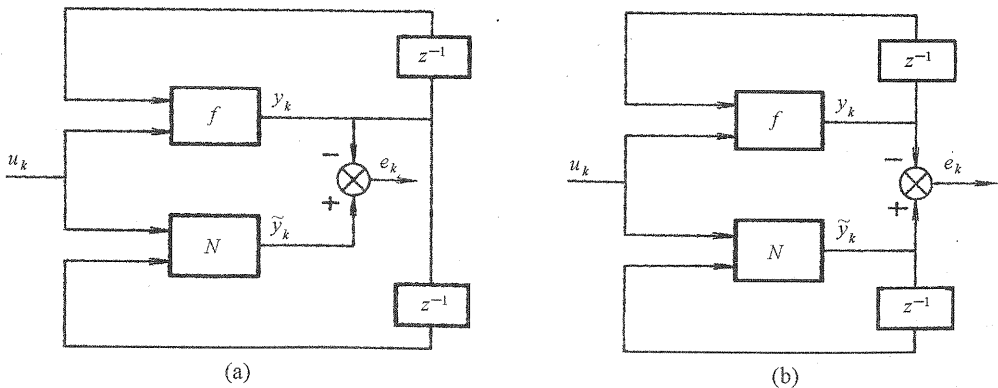


图 1 神经网络辨识的两种结构

本文的其它部分是这样安排的:第二部分是预备知识,介绍本文要用到的一种改进型 BP 算法——死区 BP 算法;第三部分讨论并联预测方式的输出偏差收敛性问题,以及死区 BP 算法对并联辨识结构学习的局部收敛性问题,并做出必要的证明;第四部分给出仿真实验来说明前面的结果;最后在第五部分给出结论.

## 2 预备知识

我们在这一部分首先简单介绍 Chen 提出的死区 BP 算法<sup>[2]</sup>. 设多输入单输出系统  $f(X)$  和神经网络  $N(X;W)$  的输出分别为  $y$  和  $\tilde{y}$ ,  $W$  为网络参数列向量. 定义误差度量  $E = \frac{1}{2}(y - \tilde{y})^2$ , 则网络参数  $W$  应按下式寻优:

$$W_{k+1} = W_k - \mu \frac{\partial E}{\partial W} \Big|_{W_k}.$$

$\mu$  为一正常数,称为学习步长. BP 算法为我们提供了一个计算  $\frac{\partial E}{\partial W}$  的方法,即

$$\frac{\partial E}{\partial W} \Big|_{W_k} = (\tilde{y}_k - y_k) \frac{\partial \tilde{y}_k}{\partial W} \Big|_{W_k}. \quad (1)$$

BP 算法的训练过程是一个迭代的过程,该过程至  $E$  充分小后停止. 而死区 BP 算法<sup>[2]</sup>则将(1)式修改为

$$\frac{\partial E}{\partial W} \Big|_{W_k} = D(\tilde{y}_k - y_k) \frac{\partial \tilde{y}_k}{\partial W} \Big|_{W_k}.$$

其中

$$D(e) = \begin{cases} 0, & \text{if } |e| \leq d_0, \\ e - d_0, & \text{if } e > d_0, \\ e + d_0, & \text{if } e < -d_0. \end{cases}$$

$d_0$  称为此算法的死区,取充分小的正数. 这样,权重的修改公式变为:

$$W_{k+1} = W_k - \mu D(\tilde{y}_k - y_k) \frac{\partial \tilde{y}_k}{\partial W} \Big|_{W_k}. \quad (2)$$

对于  $e_k = W_k^T \frac{\partial \tilde{y}}{\partial W} \Big|_{W_k} + \eta_k$ ,文[2]中证明当  $|\eta_k| < d_0$  时,存在  $0 \leq \alpha_k \leq 1$ ,使

$$D(e_k) = \alpha_k W_k^T \frac{\partial \tilde{y}}{\partial W} \Big|_{W_k}. \quad (3)$$

式(3)在我们第三部分的讨论中将会用到. 以下我们称并联辨识结构下的算法(2)为并联辨识算法.

本文研究神经网络对动态系统的辨识方法. 不失一般性,本文假设动态系统具有如下的形式:

$$y_k = f(y_{k-1}, u_k). \quad (4)$$

$f$  为非线性函数,且动态系统(4)满足下列条件:

**条件 2.1** 若动态系统(4)的控制域为  $U_D \subset \mathbb{R}^m$ , 则  $\forall u \in U_D$ , 都有  $\|y\| < \infty$ , 即动态系统是稳定的.

为下面的讨论方便,现给出动态系统估计的概念.

**定义 2.1** 动态系统的估计:神经网络  $N(y_{k-1}, u_k; W^*)$  称为动态系统(4)的估计当且仅当

$$\|f(y_{k-1}, u_k) - N(y_{k-1}, u_k; W^*)\| \leq \varepsilon, \quad y_{k-1} \in Y_D, \quad u_k \in U_D. \quad (5)$$

其中  $\varepsilon$  为预先给定的任意小正数,  $Y_D, U_D$  为我们关心的输出域和控制域.

神经网络映射定理告诉我们,用图 2 所示的三层神经网络做为辨识模型,则满足条件 2.1 的动态系统(4)的估计总是存在的.

### 3 神经网络并联辨识方式的局部收敛性

一个如图 2 所示的神经网络  $N(X;W)$ ,  $W = [w_i, v_{ij}, t_i, h]^T, i = 1, \dots, n, j = 1, \dots, m$ , 取 sigmoid 激励函数  $g(x) = \frac{1}{1 + e^{-x}}$ , 其输入输出关系可以表示为:

$$z_i = g\left(\sum_{j=1}^m v_{ij}x_j - t_i\right), \tag{6}$$

$$N(X;W) = g\left(\sum_{i=1}^n z_i w_i - h\right).$$

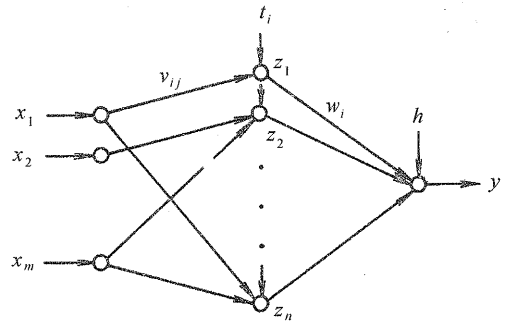


图 2 三层前馈型神经网络的结构

如不特别说明,以下提到的神经网络均指结构如图 2 的神经网络,而动态系统(4)均满足条件 2.1.

**引理 3.1** 具有  $n$  个隐节点的神经网络  $N(X;W)$ ,  $X = [x_1, x_2, \dots, x_m]^T$ , 若  $w_0 v_0 < \frac{16}{n}$ , 则有  $\left\| \frac{\partial N}{\partial x_j} \right\| \triangleq c < 1, j = 1, 2, \dots, m$ . 其中,  $w_0 = \max_{1 \leq i \leq n} \{ |w_i| \}, v_0 = \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} \{ |v_{ij}| \}$ .

证 由  $g(x) = \frac{1}{1 + e^{-x}}$ , 不难得出

$$\|g'(x)\| \leq \frac{1}{4}.$$

再由式(6)出发,有

$$\begin{aligned} \left\| \frac{\partial N}{\partial x_j} \right\| &\triangleq c = \left\| g' \left( \sum_{i=1}^n z_i w_i - h \right) \sum_{i=1}^n w_i g' \left( \sum_{j=1}^m v_{ij} x_j - t_i \right) v_{ij} \right\| \\ &\leq \sum_{i=1}^n \frac{\|w_i v_{ij}\|}{16} < 1. \end{aligned}$$

证毕.

**定理 3.1** 若神经网络  $N(y_{k-1}, u_k; W^*)$  为动态系统(4)的估计,且  $w_0^* u_0^* < \frac{16}{n}$ , 则并联预测过程  $\tilde{y}_k = N(\tilde{y}_{k-1}, u_k; W^*)$  收敛. 其中,  $w_0 = \max_{1 \leq i \leq n} \{ |w_i| \}, v_0 = \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} \{ |v_{ij}| \}$ .

证 由条件(5),有

$$\begin{aligned} \|y_k - \tilde{y}_k\| &= \|f(y_{k-1}, u_k) - N(\tilde{y}_{k-1}, u_k; W^*)\| \\ &\leq \|f(y_{k-1}, u_k) - N(y_{k-1}, u_k; W^*)\| \\ &\quad + \|N(y_{k-1}, u_k; W^*) - N(\tilde{y}_{k-1}, u_k; W^*)\| \\ &\leq \varepsilon + \left\| \frac{\partial N}{\partial y_{k-1}} \right\|_{\xi} \|y_{k-1} - \tilde{y}_{k-1}\|, \end{aligned}$$

其中  $\xi \in [y_{k-1}, \tilde{y}_{k-1}]$ .

由引理 3.1,  $\left\| \frac{\partial N}{\partial y_{k-1}} \right\|_{\xi} \leq c < 1$ , 于是有

$$\begin{aligned} \|y_k - \tilde{y}_k\| &\leq \varepsilon + c\varepsilon + \dots + c^{k-1}\varepsilon + c^k \|y_0 - \tilde{y}_0\| \\ &= \frac{1}{1-c}\varepsilon + c^k \left( \|y_0 - \tilde{y}_0\| - \frac{1}{1-c}\varepsilon \right) \rightarrow \frac{1}{1-c}\varepsilon. \end{aligned}$$

证毕.

并联预测过程在动态系统的仿真和控制中经常用到,定理 3.1 说明并联预测过程在一定条件下是收敛的.

**定理 3.2**(并联算法局部收敛性定理) 记动态系统(4)的估计为  $N(y_{k-1}, u_k; W^*)$ , 若  $\omega_0^* v_0^* < \frac{16}{n}$ , 则并联辨识算法(2)局部收敛, 此时:  $\|W_k - W^*\| \leq \|W_{k-1} - W^*\|$ , 且  $\|\tilde{y}_k - y_k\| \rightarrow c < d_0$ .

证 记  $\tilde{W}_k = W_k - W^*$ ,  $J_k = \frac{\partial N}{\partial W} |_{W_k}$ , 则

$$\begin{aligned} e_k &\triangleq \tilde{y}_k - y_k \\ &= N(\tilde{y}_{k-1}, u_k; W_k) - N(\tilde{y}_{k-1}, u_k; W^*) + N(\tilde{y}_{k-1}, u_k; W^*) - f(y_{k-1}, u_k) \\ &= \tilde{W}_k^T J_k + o(\|W_k - W^*\|) + N(\tilde{y}_{k-1}, u_k; W^*) - f(y_{k-1}, u_k). \end{aligned}$$

取

$$\eta_k = o(\|W_k - W^*\|) + N(\tilde{y}_{k-1}, u_k; W^*) - f(y_{k-1}, u_k),$$

由定理 3.1, 有

$$\begin{aligned} |\eta_k| &\leq o(\|W_k - W^*\|) + \frac{1}{1-c}\epsilon + c^k(\|y_0 - \tilde{y}_0\| - \frac{1}{1-c}\epsilon) \\ &\rightarrow o(\|W_k - W^*\|) + \frac{1}{1-c}\epsilon. \end{aligned}$$

由式(2)

$$\tilde{W}_{k+1} = \tilde{W}_k - \mu D(e_k) J_k.$$

若  $W_k$  与  $W^*$  足够近, 即  $\|W_k - W^*\|$  足够小, 使  $|\eta_k| < d_0$ , 则利用(3)式可得:

$$\begin{cases} D(e_k) \leq \tilde{W}_k^T J_k, & \text{当 } D(e_k) > 0 \text{ 时,} \\ D(e_k) \geq \tilde{W}_k^T J_k, & \text{当 } D(e_k) \leq 0 \text{ 时.} \end{cases}$$

于是有

$$\begin{aligned} \tilde{W}_k^T \tilde{W}_k - \tilde{W}_{k-1}^T \tilde{W}_{k-1} &= -2\mu D(e_{k-1}) \tilde{W}_{k-1}^T J_{k-1} + \mu^2 D^2(e_{k-1}) J_{k-1}^T J_{k-1} \\ &\leq -2\mu D^2(e_{k-1}) + \mu^2 D^2(e_{k-1}) J_{k-1}^T J_{k-1}, \end{aligned}$$

取  $\mu > 0$ , 且使  $\mu J_{k-1}^T J_{k-1} < 1$ , 则

$$\tilde{W}_k^T \tilde{W}_k - \tilde{W}_{k-1}^T \tilde{W}_{k-1} \leq -\mu D^2(e_{k-1}) \leq 0,$$

即  $\|\tilde{W}_k\| \leq \|\tilde{W}_{k-1}\|$ , 等号成立当且仅当  $D(e_{k-1}) = 0$ . 证毕.

在实际计算过程中, 为保证总有  $\eta_k < d_0$ , 不妨在算法开始阶段取稍大的  $d_0$ , 随  $k$  的加大逐渐将  $d_0$  减少至要求的值.

#### 4 仿真实验

这一部分我们用仿真实例来验证第三部分的收敛性结论.

**算例 1** 动态系统  $y_k = f(y_{k-1}) = 0.65 - 0.3e^{-0.01y_{k-1}} \sin(y_{k-1}\pi)$ , 取神经网络隐节点数  $n = 3$ , 给定误差限  $\epsilon = 0.005$ , 并认为满足此误差限的神经网络是该系统的估计. 在无观测噪声的情况下用 BP 算法训练, 迭代 137102 步后得到估计  $N(y_{k-1}; W^*)$ , 其网络参数  $w^* = [1.804225, 1.335228, 2.042832]$ ,  $v^* = [-1.425353, -0.891813, -1.674767]$ . 显然  $\omega_0^* v_0^* = 2.04 \times 1.67 = 3.4 < \frac{16}{3} = 5.3$ , 可以进行并联预测. 利用神经网络  $N(\tilde{y}_{k-1}; W^*)$  对  $f(y_{k-1})$  进行并联预测, 系统初值  $\tilde{y}_0 = 0.0, y_0 = 0.6$ , 输出误差示于图 3, 从图中可以看到, 预测过程是收敛的.

**算例 2** 由算例 1 知, 动态系统  $y_k = f(y_{k-1}) = 0.65 - 0.3e^{-0.01y_{k-1}} \sin(y_{k-1}\pi)$  的估计满足

$w_0^* v_0^* < \frac{16}{3}$ , 可以利用并联算法辨识. 取  $d_0 = 0.001$ , 并将算例 1 的  $W$  作为网络参数的初值, 以满足并联辨识算法的局部收敛性条件. 迭代后辨识过程很快收敛, 收敛后新的估计的网络参数为

$$w^* = [1.804140, 1.335176, 2.042734],$$

$$v^* = [-1.425793, -0.892386, -1.675168].$$

显然, 只要满足一定的条件, 并联辨识算法的收敛性是能够得到保证的.

## 5 结 论

本文讨论了三层神经网络并联辨识算法的收敛性问题,

从理论和仿真实验两个方面说明只要某动态系统估计的参数  $w_0^* v_0^* < \frac{16}{n}$ , 则有: 1) 并联预测过程收敛; 2) 并联辨识算法局部收敛. 本文的结果不难推广到多层神经网络结构.

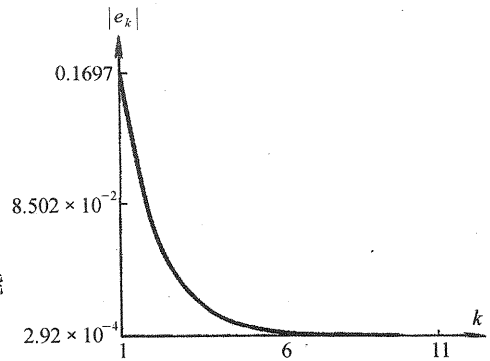


图 3 前馈型神经网络的并联预测过程

## 参 考 文 献

- 1 Narendra, K. S. and Parthasarathy, K.. Identification and control of dynamical systems using neural networks. IEEE Trans. Neural Network, March, 1990, 1(1): 4-27
- 2 Chen, F. C.. On the error and parameter convergence of back-propagation learning. Proc. of IJCNN'91, 1991, 1092-1097

# Research on Parallel Identification Algorithm of Neural Networks

LU Jin, XU Wenli and HAN Zengjin

(Department of Automation, Tsinghua University • Beijing, 100084, PRC)

**Abstract:** Neural networks can be used to set up models of nonlinear dynamic systems. Their identification models are classified as series-parallel model and parallel model. The latter one is developed from the output error identification model based on the reference model adaptive scheme, which has stronger capability to control observation noise. In this paper, we study the convergence of this parallel identification model and find that while the parameters of neural networks meet some prerequisites, the parallel prediction model converges and the parallel identification algorithm is locally convergent. Simulation results demonstrate the above conclusions.

**Key words:** neural network; nonlinear dynamic system; parallel identification algorithm

## 本文作者简介

卢 进 见本刊 1998 年第 1 期第 60 页.

徐文立 1947 年生. 清华大学自动化系教授. 博士生导师. 研究兴趣为智能控制, 计算机视觉, 计算机控制与管理系统等.

韩曾晋 1933 年生. 清华大学自动化系教授. 博士生导师. 研究兴趣为自适应控制, 智能控制, 离散事件动态系统的理论及应用, CIMS 体系结构等.