

一种改进的 Elman 神经网络模型¹

王常虹 高晓智 徐立新 庄显义 Gao Xiaoming*

(哈尔滨工业大学 329 信箱自动控制理论及应用教研室 哈尔滨 150001)

*(Helsinki University of Technology, Faculty of Electrical Engineering Laboratory of Signal Processing and Computer Technology(STI) Otakaari 5A, FIN-02150, Espoo, Finland)

摘要 本文首先详细地阐述了 Elman 神经网络的结构、原理和学习算法。为了进一步提高 Elman 神经网络的逼近能力和动态特性,我们提出了一种改进的 Elman 神经网络模型。这种新的 Elman 神经网络在关联节点与输出节点之间又增加了一组可调权值,利用误差回馈原理推导出了其相应的学习算法。仿真实验结果表明,改进的 Elman 神经网络比原来的网络具有更好的动态性能,对于贯序输入输出数据的逼近收敛速度更快。

关键词 Elman 神经网络, 逼近, 学习算法

中图分类号 TN-052

1 前言

近些年来,人工神经网络越来越引起控制理论工作者的极大兴趣^[1,2]。由于神经网络具有并行处理、联想记忆、分布式知识存储、鲁棒性强等特点,尤其是它的自组织、自适应、自学习功能,从而在复杂非线性对象的辨识和控制中得到了广泛应用。目前,系统辨识中普遍采用的神经网络模型是多层前馈式反传网络(BP网络),使用的权值调整学习算法为有教师信号的误差逆向传播方法。虽然,从理论上讲,三层以上的BP网络能够逼近任何一个属于 L_2 的非线性函数,从而可以用于辨识被控对象的动态模型和逆动态模型。但是,由于BP网络从本质上来说是静态网络,它所能实现的只是一一对应的静态非线性映射关系,不适合动态系统的实时辨识,因此,非常有必要探讨其它适合于辨识和控制的动态神经网络模型和学习算法^[3]。

Elman神经网络是Elman于1990年首先针对语音处理问题而提出来的^[4]。它是一种典型的局部递归网络(Global Feedforward Local Recurrent)。由于Elman神经网络对于处理贯序输入输出数据的优越性而在系统建模、时序预报、自适应控制等方面获得了广泛的应用。本文针对Elman神经网络的特点,提出了一种在关联层与输出层之间加入可调权值的改进的Elman神经网络模型。仿真实验表明,这种改进的Elman神经网络在贯序输入输出训练数据的逼近速度上具有很大的优越性,取得了令人满意的效果。可以预见,将这种改进的Elman网络用于动态系统的实时辨识和控制是具有很大意义的。

2 Elman神经网络的基本原理

Elman神经网络可以看作一个具有局部记忆单元和局部反馈连接的前向BP神经网络,其基本结构如图1所示。从图中我们可以看到Elman神经网络主要由输入层、中间层、关联层、输出层组成。每一个中间层节点都有一个与之对应的关联层节点(context node)连接。输入层

¹ 1995-12-11 收到, 1997-03-06 定稿
航天基础研究基金资助项目

节点与中间层节点, 中间层节点与输出层节点以及关联层节点与中间层节点之间都有可调权值相连接。

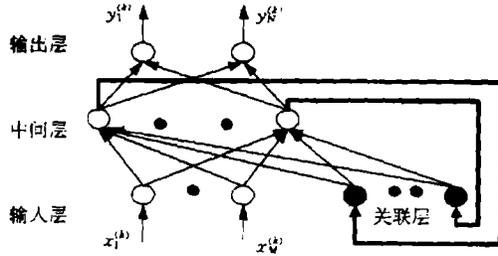


图 1 Elman 神经网络的基本结构

为了推导 Elman 神经网络的学习算法, 我们首先定义下列符号: 连接输入层节点 i 到中间层节点 j 的权值记作 $w_{1i,j}$, 连接中间层节点 i 到输出层节点 j 的权值记作 $w_{2i,j}$, 连接关联层节点 i 到中间层节点 j 的权值记作 $w_{3i,j}$, 网络中间层节点 i 的输入记作 $\text{net}-h_i$, 中间层节点 i 的输出记作 $o-h_i$. 关联层节点 i 的输入记作 $\text{net}-c_i$, 关联层节点 i 的输出记作 $o-c_i$, 网络中间层的节点函数为 $f(x)$, $f(x)$ 通常取 Sigmoid 函数, 例如 $f(x) = 1/(1 + e^{-ax})$. Elman 神经网络的输入记为 $x_i^{(k)}$, $i = 1, 2, \dots, M$. 网络输出为 $y_j^{(k)}$, $i = 1, 2, \dots, N$. $x_i^{(k)}, y_j^{(k)}$ 均为贯序输入输出数据. 我们设 Elman 神经网络的运行迭代次数为 $k, k = 1, 2, \dots, T$. M, N, L 分别为输入层, 输出层和中间层的节点数. 关联层节点数与中间层节点数相等, 均为 L . 显然可以得到

$$\begin{aligned} o-c_i^{(k)} &= \text{net}-c_i^{(k)}, \quad \text{net}-h_i^{(k)} = \sum_{i=1}^M w_{1i,j} X_i^{(k)} + \sum_{i=1}^L w_{3i,j} o-c_i^{(k)}, \\ o-h_j^{(k)} &= f(\text{net}-h_j^{(k)}), \quad y_j^{(k)} = \sum_{i=1}^L w_{2i,j} o-h_j^{(k)}, \end{aligned}$$

其中在 k 时刻关联层节点的输入 $\text{net}-c_i^{(k)}$ 为 $k-1$ 时刻其所对应的中间层节点的输出 $o-h_i^{(k-1)}$, 如 (1) 式所示

$$\text{net}-c_i^{(k)} = o-h_i^{(k-1)}. \quad (1)$$

Elman 神经网络的学习算法仍然使用 BP 神经网络的梯度下降误差回馈原理, 所不同的是其指标函数定义为在某一时间片 $[0, T]$ 内的整体逼近误差

$$E = \frac{1}{2} \sum_{k=1}^T \sum_{i=1}^N (Y_i^{(k)} - y_i^{(k)})^2,$$

$Y_i^{(k)}$ 为 k 时刻网络的期望输出. 权值 $w_{1i,j}$ 、 $w_{2i,j}$ 的修正等同于静态 BP 神经网络中的权值学习算法. 特别地, 对于权值 $w_{3i,j}$, 应用链式导数规则, 我们有

$$\Delta w_{3i,j} = -(\partial E / \partial o-h) (\partial o-h_j / \partial w_{3i,j}),$$

其中在 k 时刻有

$$\begin{aligned} \frac{\partial E}{\partial o-h_j^{(k)}} &= \sum_{i=1}^N \frac{\partial E}{\partial y_i^{(k)}} \frac{\partial y_i^{(k)}}{\partial o-h_j^{(k)}} = - \sum_{i=1}^N [(Y_i^{(k)} - y_i^{(k)}) w_{2i,j}^{(k)}], \\ \frac{\partial o-h_j^{(k)}}{\partial w_{3i,j}^{(k)}} &= f'(\text{net}-h_j^{(k)}) \frac{\partial \text{net}-h_j^{(k)}}{\partial w_{3i,j}^{(k)}} = f'(\text{net}-h_j^{(k)}) o-c_i^{(k)} = f'(\text{net}-h_j^{(k)}) o-h_i^{(k-1)}, \end{aligned}$$

因此

$$\Delta w_{3i,j} = \sum_{k=1}^T \sum_{i=1}^N [(Y_i^{(k)} - y_i^{(k)}) w_{2j,i}^{(k)}] f'(\text{net-}h_j^{(k)}) o-h_i^{(k-1)}. \quad (2)$$

当 $k = 1$ 时, $o-h_i^{(0)}$ 为 Elman 神经网络运行状态的初始值, 此值可由设计者确定或者直接令 $o-h_i^{(0)} = 0$.

3 改进的 Elman 神经网络模型及其学习算法

从上面的学习算法推导中可以看到 Elman 神经网络关联层节点的输出实际上起到了存储网络的“内部状态”的作用^[5], 关联层节点与中间层节点的连接类似于系统内部的状态反馈。事实上, Elman 神经网络区别于其它静态前馈网络(如 BP 网络)的特点就在于其关联层节点的存在。通过关联层节点的输出, Elman 神经网络不仅可以存储当前的贯序输入数据, 而且可以存储贯序输入数据中的过去的某些信息。这样, 在训练过程中就能够动态地回溯到与当前 k 时刻期望输出有关的 $k-1, k-2, \dots, k-n, n$ 个过去时刻的网络输入。所以, 为了进一步提高 Elman 神经网络的逼近能力和动态性能, 就必须充分利用其关联层节点的输出。因此, 我们提出了一种新的改进的 Elman 神经网络模型, 其基本结构如图 2 所示。从图中可以看到, 改进的 Elman 神经网络在网络的关联层与输出层之间又增加了一组新的连接权值 $w_{4i,j}$ 。对于输出 $y_j^{(k)}$, 显然有

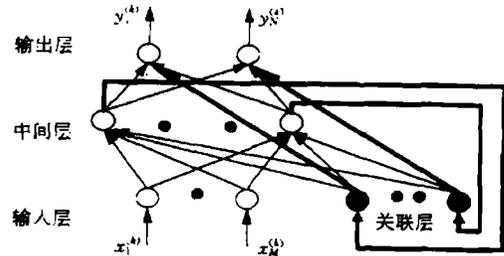


图 2 改进的 Elman 神经网络结构

$$y_j^{(k)} = \sum_{i=1}^L (w_{2i,j} o-h_i^{(k)} + w_{4i,j} o-c_i^{(k)}).$$

这时的关联层既具有原来 Elman 神经网络的关联层的作用, 又充当了改进后网络的另外一个“中间层”。权值 $w_{4i,j}$ 的调整方法与原有权值 $w_{1i,j}$, $w_{2i,j}$ 相同, 见 (3) 式:

$$\Delta w_{4i,j} = \sum_{k=1}^T \sum_{i=1}^N [(Y_i^{(k)} - y_i^{(k)}) o-h_i^{(k)}]. \quad (3)$$

由于 $w_{4i,j}$ 是将延迟一步的中间层节点输出与输出层节点相连, 因此, 在计算网络权值的修正误差时应加上由这一部分权值从输出层回馈回来的逼近误差。注意到关联层的输出是延迟一步的中间层的输出, 所以它在 $k+1$ 步回馈的误差也必须延迟一步后再用于 k 步的权值 $w_{3i,j}$ 修正上。改进的 $w_{3i,j}$ 学习公式为

$$\begin{aligned} \Delta w_{3i,j} &= -\frac{\partial E}{\partial o-h_j} \frac{\partial o-h_j}{\partial w_{3i,j}} - \frac{\partial E}{\partial o-c_j} \frac{\partial o-c_j}{\partial w_{3i,j}} \\ &= \sum_{k=1}^T \left[\frac{\partial E}{\partial o-h_j^{(k)}} \frac{\partial o-h_j^{(k)}}{\partial w_{3i,j}^{(k)}} + \frac{\partial E}{\partial o-c_j^{(k+1)}} \frac{\partial o-c_j^{(k+1)}}{\partial w_{3i,j}^{(k)}} \right], \end{aligned}$$

其中令上式右端中 $k = T$ 时的 $\frac{\partial E}{\partial o-c_j^{(T+1)}} \frac{\partial o-c_j^{(T+1)}}{\partial w_{3,i,j}^{(T)}} = 0$. 已知 $o-c_j^{(k+1)} = o-h_j^{(k)}$. 类似上述推导过程, 有

$$\Delta w_{3,i,j} = \sum_{k=1}^T \left\{ \sum_{i=1}^N [(Y_i^{(k)} - y_i^{(k)})w_{2,j,i}^{(k)}] f'(\text{net}_j^{(k)})o-h_i^{(k-1)} + \sum_{i=1}^N [(Y_i^{(k+1)} - y_i^{(k+1)})w_{4,j,i}^{(k+1)}] f'(\text{net}_j^{(k)})o-h_i^{(k-1)} \right\}$$

改进后的 Elman 神经网络实际上可以看作是一个具有局部反馈的 4 层前向网络, 即前向通路由输入层、中间层、关联层、输出层组成, 反馈回路由关联层、中间层组成, 见图 3 . 图中 z^{-1} 为延迟一步因子. 由于权值 $w_{4,j,i}$ 的引入, 将 Elman 网络的内部状态通过 $w_{4,j,i}$ 与输出节点相连接, 这类似于系统中的状态前馈. 改进的 Elman 神经网络的状态方程可表达如下:

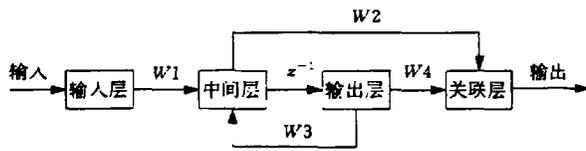


图 3 改进的 Elman 神经网络原理图

$$X(k+1) = W3f(X(k)) + W1U(k+1),$$

$$Y(k+1) = W4X(k) + W2X(k+1).$$

$U(k)$ 、 $Y(k)$ 、 $X(k)$ 分别为 Elman 神经网络的输入、输出和关联层节点的输出. 从上式中可以发现, 对 Elman 网络本身的运行而言, 权值 $w_{4,j,i}$ 增强了其动态性能, 提高了网络每次迭代的学习效率, 从而加速了网络的收敛.

4 仿真实验

为了比较改进的 Elman 网络与原来的 Elman 网络的学习收敛速度, 我们选取一组两个具有不同幅值的正弦时序信号做为网络的输入, 网络的理想输出定义为各个不同正弦信号所对应的最大幅值, 见 (5) 式:

$$\left. \begin{aligned} x(k) &= 1.0 \sin(k), & Y(k) &= 1.0, & k &= 1, 2, \dots, 20; \\ x(k) &= 2.0 \sin(k), & Y(k) &= 2.0, & k &= 1, 2, \dots, 20. \end{aligned} \right\} \quad (5)$$

选取的 Elman 神经网络结构为输入层和输出层分别只有一个节点, 网络中间层节点均取 10 个. 设预期的逼近误差方差和 (Sum Squared Error) SSE=0.01 . 改进的 Elman 网络与原来的 Elman 网络的训练过程分别如图 4(a)、4(b) 所示. 从图 4 可以看到, 改进的 Elman 神经网络只用了 121 次迭代就使逼近误差达到了 0.01 , 而原来的 Elman 网络却需要 603 次. 在 486/DX66 计算机 Matlab 4.0 for Windows 仿真环境下原有的 Elman 网络训练时间为 5min 43s , 而改进的 Elman 网络只需要 2min 27s . 说明改进的 Elman 神经网络具有更好的动态性能, 逼近速度比原有的网络得以大大提高. 这为 Elman 神经网络在实时辨识和控制等方面的应用提供了有利条件. 另外, 在训练结束后, 我们对改进的 Elman 网络进行检验, 选择的测试数据分别为以前的训练样本和新的两组正弦输入, 见 (6) 式:

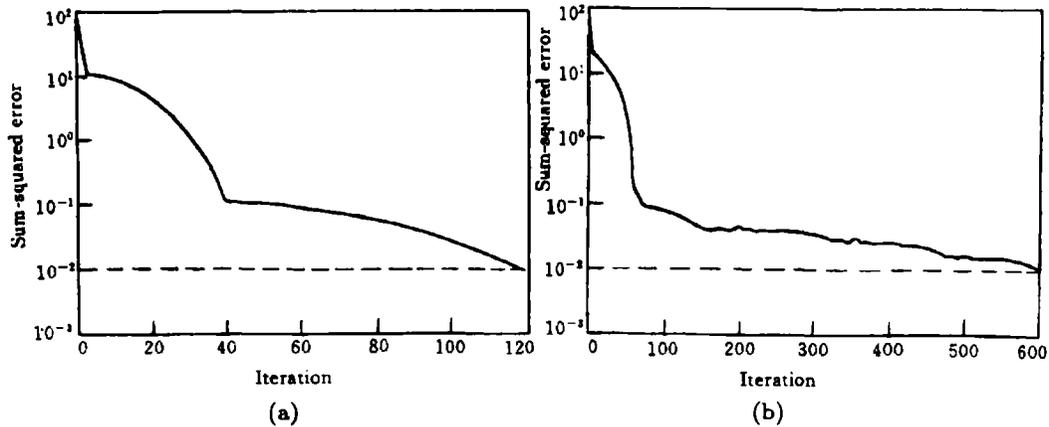


图 4 不同的 Elman 神经网络逼近速度的比较

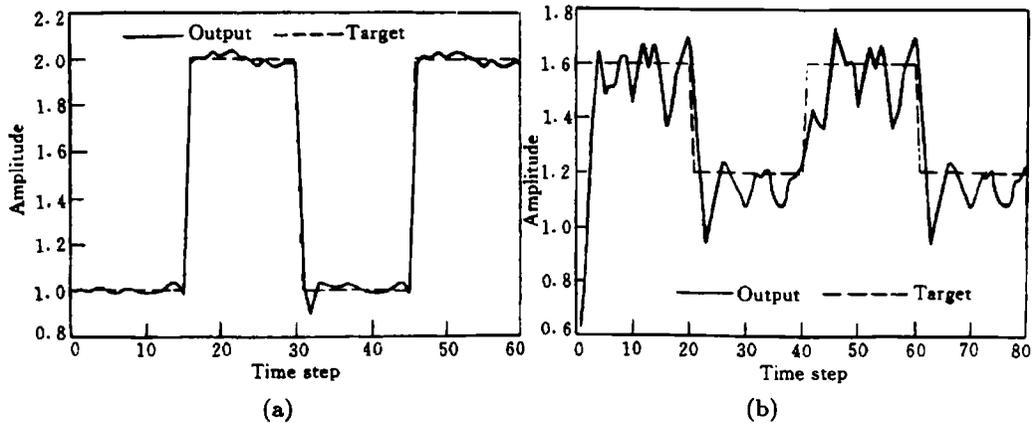


图 5 改进的 Elman 神经网络的逼近精度与泛化能力

$$\left. \begin{aligned} x(k) &= 1.2 \sin(k), & Y(k) &= 1.2, & k &= 1, 2, \dots, 20; \\ x(k) &= 1.6 \sin(k), & Y(k) &= 1.6, & k &= 1, 2, \dots, 20. \end{aligned} \right\} \quad (6)$$

改进的 Elman 网络的输出分别如图 5(a)、5(b) 所示。从图中可以看到, 网络不仅对已训练的样本逼近精度很高, 而且对未训练的测试输入也有一定的泛化能力。显然, 更多的训练样本和隐含层节点可以使其泛化能力得以进一步提高。可见 Elman 网络对于时序输入可以给出较好的逼近和预报。

5 结 论

本文提出了一种改进的 Elman 神经网络模型及其学习算法。仿真实验证明了这种新的 Elman 神经网络具有动态特性好, 逼近速度快等特点。进一步的工作应包括对 Elman 神经网络的快速学习算法, 训练过程中的鲁棒性、收敛性的理论分析等。有关基于 Elman 神经网络的非线性动态系统辨识和自适应控制问题研究将另文阐述。

参 考 文 献

- [1] Hunt K J, Sbarbaro D, Zbikowski R, *et al.* Neural networks for control system—A survey. *Automatica*, 1992, 28(6):1083–1112.
- [2] Narendra K S, Parthasarathy K. Identification and control of dynamical systems using neural networks. *IEEE Trans. on Neural Networks*, 1990, 1(1):4–27.
- [3] 高晓智, 王常虹, 徐立新, 等. CMAC 神经网络再励学习控制. CIAC'95 中国智能自动化学术会议暨智能自动化专业委员会成立大会论文集, 天津: 1995, 638–643.
- [4] Elman J. Finding structure in time. *Cognitive Science*, 1990, 14(2):179–211.
- [5] Sastry P S, Santharam G, Unnikrishnan K P. Memory neuron networks for identification and control of dynamical systems. *IEEE Trans. on Neural Networks*, 1994, 5(2):306–319.

A NEW MODIFIED ELMAN NEURAL NETWORK MODEL

Wang Changhong Gao Xiaozhi Xu Lixin Zhuang Xianyi Gao Xiaoming*

(*Department of Control Engineering, P. O. Box 329, Harbin Institute of Technology, Harbin 150001*)

**(Helsinki University of Technology, Faculty of Electrical Engineering, Laboratory of Signal Processing and Computer Technology(STI), Otakaari 5A, FIN-02150, Espoo, Finland)*

Abstract This paper first discusses the structure, principle and learning algorithm of Elman neural network model. A modified Elman neural network model is then proposed by adding new adjustable weights between the context nodes and the output nodes to enhance its dynamical character. The corresponding learning algorithm is also derived by using steepest descent principle. Theoretical analysis and simulation results show that this kind of modified Elman neural network learns much faster than the original model.

Key words Elman neural network, Approximation, Learning algorithm

王常虹: 男, 1961年生, 副教授, 主要研究方向为智能控制、模糊控制。

高晓智: 男, 1972年生, 博士生, 目前主要从事神经网络模型和智能控制方面的研究。

徐立新: 男, 1969年生, 博士生, 研究方向为神经网络自适应控制。

庄显义: 男, 1936年生, 教授, 研究方向为精密伺服系统、过程控制。