文章编号: 1001-0920(2002)03-0346-03

# 一种新的对角回归神经网络快速学习算法

王振雷, 王建辉, 顾树生
(东北大学 信息科学与工程学院, 辽宁 沈阳 110004)

**摘　要**: 提出一种新的动态对角回归神经网络学习算法——局部动态误差反传算法(LDBP), 该算法定义了一种新的局部均方差函数, 并为回归单元建立一种新的学习结构。如果估计出各层的期望输出值, 多层回归网络便可分解成一组自适应单元(Adaline), 而每个单元可通过二次优化方法进行训练。采用可在有限步内找出全局最优解的共轭梯度法(CG)进行寻优。由于学习过程采用超线性搜索, 大大减少了循环步数和计算时间。

**关键词**: 动态误差反传算法(DBP); 共轭梯度法(CG); 对角回归神经网络(DRNN); 动态非线性系统; 系统辨识

**中图分类号**: TP 18　　　　**文献标识码**: A

# New method for diagonal recurrent neural networks with local error-backpropagation

*WANG Zhen-lei, WANG Jian-hui, GU Shu-sheng*
(School of Information Science and Engineering, Northeastern University, Shenyang 110004, China)

**Abstract**: A fast new local dynamic error-backpropagation algorithm (LDBP) is presented for the training of diagonal recurrent neural networks(DRNN). This algorithm is based on the definition of a new local mean-squared error function. The approximation to a recurrent node has the similar construction as the Adaline (Adaptive linear element). When the local desired outputs of the elements have been estimated, the DRNN can be decomposed into a set of node elements that can be trained by quadratic optimization methods. The conjugate gradient (CG) method is used. The simulation result shows the advantages of the new algorithm.

**Key words**: dynamic error-backpropagation(DBP); conjugate gradient(CG); diagonal recurrent neural networks(DRNN); dynamic non-linear system; system identification

## 1 引　　言

对角回归神经网络是一种特殊的动态网络, 是由 Chao-Chee Ku 等[1]首次提出并对其收敛性进行了证明。它既具有一般动态网络易于处理动态非线性问题的特点, 又具有结构简单、容易构造训练算法的优点。所以, 对角回归神经网络在系统辨识与控制器的设计等方面均得到广泛应用[1,2]。

动态误差反传算法(DBP)[1]是在对角回归网络训练中应用最多的一种学习方法, 它是对BP算法

**收稿日期**: 2000-05-19; **修回日期**: 2000-07-12
**基金项目**: 辽宁省自然科学基金项目(26237)
**作者简介**: 王振雷(1975—), 男, 山东德州人, 博士生, 从事智能控制和复杂系统分析与控制等研究; 顾树生(1939—), 男, 黑龙江绥化人, 教授, 博士生导师, 从事智能控制、复杂系统控制及电力拖动等研究。

的改进, 充分考虑了回归单元对算法的影响。网络参数通常用梯度下降法进行调整。

DBP 算法的主要缺点是收敛速度慢和容易陷入局部极小值, 对此人们提出了许多改进方法[3], 但这些方法多是基于线性搜索的优化方法, 其搜索效率不高[4]。

本文提出的局部动态误差反传算法(LDBP)将简单的梯度下降法与全局高效的二次优化搜索技术相结合。该算法定义了一种新的均方差函数, 既可保证全局收敛, 又可有效地应用于各自适应单元用 CG 法独立寻优。仿真结果表明, 该算法是一种高效的全局方法。

## 2　回归单元和输出线性单元的二次优化方法

把回归层和输出层拆分成多个自适应单元, 则回归层和输出层得到的单元结构分别如图 1 和图 2 所示。其中 $P$ 是样本对的个数, $p$ ($p = 1, 2, \cdots, P$) 表示样本序列数, $u_p$ 和 $\hat{u}_p$ 分别为线性输出和线性估计输出。采用教师强制学习法[5,6] 对回归单元进行训练: 用非线性估计输出 $\hat{y}_{p-1}$ 代替回归输入 $y_{p-1}$。

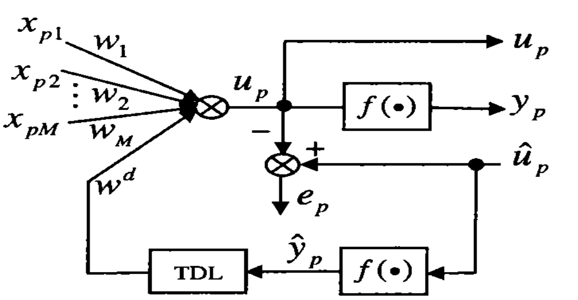


图 1　回归单元结构

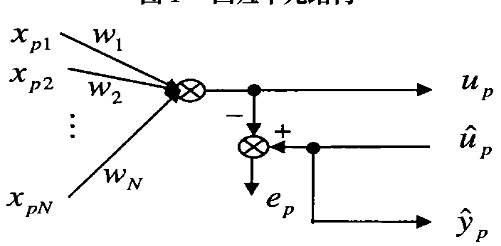


图 2　输出单元结构

定义回归单元局部均方差函数为

$$E_{\mathrm{H}}(W, w^d) = \frac{1}{2P}\sum_{p=1}^{P}[\hat{u}_p - (x_p^{\mathrm{T}} W + \hat{y}_{p-1} w^d)]^2 = \frac{1}{2P}\sum_{p=1}^{P}[\hat{u}_p - x_p'^{\mathrm{T}} W']^2 \tag{1}$$

其中

$$x_p = [x_{p1}, x_{p2}, \cdots, x_{pN}], \quad x_p' = [x_p \quad \hat{y}_{p-1}]^{\mathrm{T}}$$

$$W = [w_1, w_2, \cdots, w_M]^{\mathrm{T}}, \quad W' = [W^{\mathrm{T}} \quad w^d]^{\mathrm{T}}$$

定义输出线性单元的局部均方差函数为

$$E_{\mathrm{O}}(W) = \frac{1}{2P}\sum_{p=1}^{P}[\hat{u}_p - x_p^{\mathrm{T}} W]^2 \tag{2}$$

其中

$$x_p = [x_{p1}, x_{p2}, \cdots, x_{pN}], \quad W = [w_1, w_2, \cdots, w_N]^{\mathrm{T}}$$

可以看出式(1) 和式(2) 在结构上是完全相同的, 因此在下面论述的用共轭梯度法进行优化的算法中, 具有完全相同的表示形式。对节点单元的学习采用文献[6] 提出的改进的共轭梯度法。

## 3　局部动态误差反传算法(LDBP)

设第 $i$ 个回归神经元的线性估计输出为 $\hat{u}_{p,i}$, 非线性估计输出为 $y_{p,i}$; 第 $j$ 个输出神经元的线性输出为 $o_{p,j}$, 线性估计输出为 $o_{p,j}$, 该估计值输出直接由参考输出 $Y_{p,j}$ 代替。

### 3.1　局部均方差函数

为对角回归神经网络定义新的均方差函数

$$\tilde{E}(W, \hat{u}) = \sum_{j=1}^{N} E_{\mathrm{H}}(w^{\mathrm{I}}_{i,j}, w^{\mathrm{D}}_{j}) + \sum_{j=1}^{L} E_{\mathrm{O}}(w^{\mathrm{O}}_{i,j}) \tag{3}$$

其中, $E_{\mathrm{H}}(\bullet)$ 和 $E_{\mathrm{O}}(\bullet)$ 分别是式(1) 和式(2) 定义的回归单元和输出单元的线性估计误差。

新的误差函数对权值 $w^{\mathrm{I}}_{i,j}, w^{\mathrm{D}}_{j}, w^{\mathrm{O}}_{i,j}$ 及估计值 $\hat{u}_{p,i}$ 的偏导数分别为

$$\frac{\partial \tilde{E}(W, \hat{u})}{\partial w^{\mathrm{I}}_{i,j}} = \frac{\partial E_{\mathrm{H}}(w^{\mathrm{I}}_{i,j}, w^{\mathrm{D}}_{j})}{\partial w^{\mathrm{I}}_{i,j}} = -\frac{1}{P}\sum_{p=1}^{P}[\hat{u}_{p,j} - u_{p,j}] I_{p,i} \tag{4}$$

$$\frac{\partial \tilde{E}(W, \hat{u})}{\partial w^{\mathrm{D}}_{j}} = \frac{\partial E_{\mathrm{H}}(w^{\mathrm{I}}_{i,j}, w^{\mathrm{D}}_{j})}{\partial w^{\mathrm{D}}_{j}} = -\frac{1}{P}\sum_{p=1}^{P}[\hat{u}_{p,j} - u_{p,j}] \hat{y}_{(p-1),j} \tag{5}$$

$$\frac{\partial \tilde{E}(W, \hat{u})}{\partial w^{\mathrm{O}}_{i,j}} = \frac{\partial E_{\mathrm{O}}(w^{\mathrm{O}}_{i,j})}{\partial w^{\mathrm{O}}_{i,j}} = -\frac{1}{P}\sum_{p=1}^{P}[\hat{o}_{p,j} - o_{p,j}] \hat{y}_{p,j} \tag{6}$$

$$\frac{\partial \tilde{E}(W, \hat{u})}{\partial u_{p,i}} = \frac{1}{P}\left\{[\hat{u}_{p,j} - u_{p,j}] - \sum_{j=1}^{L}\left[(\hat{o}_{p,j} - o_{p,j})\frac{\partial o_{p,j}}{\partial u_{p,i}}\right]\right\} = \frac{1}{P}\left\{[\hat{u}_{p,j} - u_{p,j}] - f'(\hat{u}_{p,i})\sum_{j=1}^{L}\left[w^{\mathrm{O}}_{i,j}(\hat{o}_{p,j} - o_{p,j})\right]\right\} \tag{7}$$

我们取网络的期望输出值 $Y_{p,i}$ 作为输出单元的估计值 $\hat{o}_{p,i}$, 无需进行调整。

### 3.2 局部误差反传算法步骤

Step 1: 用小的随机数初始化所有的权值和估计输出 $\hat{u}_p$, 并设 $k = 0$;

Step 2: 取 $p = 1, 2, \cdots, P$, 用 $\hat{u}_{p,k+1} = \hat{u}_{p,k} + \alpha_k s_k$ 调整估计值[5], 其中 $\alpha_k$ 是学习系数, 而

$$\hat{u}_{p,k} = [\hat{u}_{p,1k}, \hat{u}_{p,2k}, \cdots, \hat{u}_{p,Nk}]^{\mathrm{T}}$$

$$s_{ik}^{(1)} = -\frac{\partial \widetilde{E}(W_k, \hat{u}_k)}{\partial u_{p,ik}}, \quad i = 1, 2, \cdots, N$$

Step 3: 用CG 二次寻优方法分别调整输入层到回归层、回归层到输出层各单元的权值 $w_{k,j}^{\mathrm{I}}, w_j^{\mathrm{D}}, w_{j,l}^{\mathrm{O}}$。

Step 4: 设定停止条件: $E_A^{(i)} \leq \epsilon$ 或 $\widetilde{E} \leq \widetilde{\epsilon}_{local}$, 若不满足终止条件, 则设 $k = k + 1$, 并返回 Step 2。

## 4 仿真研究

以对非线性系统的辨识为例, 对DBP 算法和LDBP 算法进行仿真比较。

被辨识的非线性对象为

$$y(k+3) = (y(k+2) * y(k+1) * y(k) * u(k+) * (y(k) - 1) + u(k+2)) / (1 + y(k)^2 + y(k+1)^2)$$

这是一个典型的高阶非线性对象, 用传统的辨识手段很难获得良好的效果。下面对DRNN 网络进行辨识, 网络结构为 $2 \times 7 \times 1$。输入信号为

$$u(k) = 0.2 * \sin(2k\pi/25) + 0.8 * \sin(2k\pi/250)$$
$$k = 0, 1, \cdots, P$$

其中$P$ 为训练样本个数, 输入训练对为$u(k)$ 和$y(k)$。

采用DBP 算法在Matlab 5.0 环境中训练 10 min (Inter Pentium 100, 16M DRAM) 的仿真结果如图 3 所示(其中实线为原系统输出, 虚线为网络输出), 网络输出值与实际系统输出值之间误差的范数为 0.446 4。图 4 是采用LDBP 算法在同样条件下学习 1.5 min 所得到的结果, 误差的范数为 0.097。

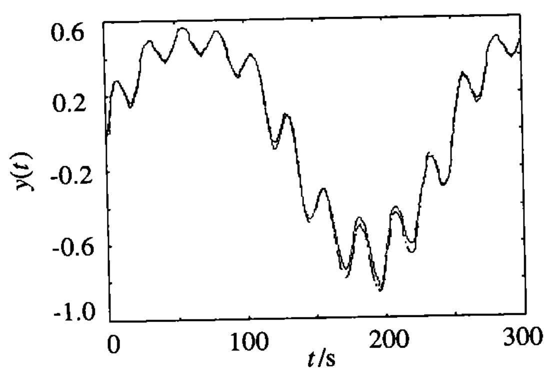


图 3 DBP 算法网络输出与原系统输出

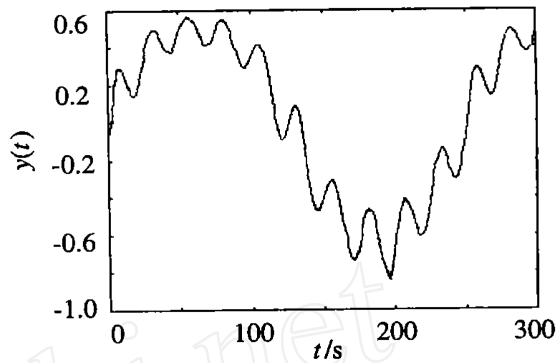


图 4 LDBP 算法网络输出与原系统输出

## 5 结 论

本文提出一种训练对角回归网络的新算法——LDBP 算法, 通过对局部误差的估计和对回归单元的重新构造, 对角回归网络可分解为一组独立的单元, 这些单元可通过二次型优化方法进行训练。本文采用改进的共轭梯度法(CG) 作为二次寻优方法, CG 法能在有限步内找到全局最优解。仿真结果表明,LDBP 法具有比DBP 法更快的收敛速度和更好的逼近性能。

参考文献(References):


[1] Chao-Chee Ku, Kwang Y Lee. Diagonal recurrent neural networks for dynamic systems control[J]. IEEE Trans on Neural Networks, 1995, 6(2): 144-155.

[2] 沈清. 神经网络应用技术[M]. 长沙: 国防科技大学出版社, 1998. 170-210.

[3] Charalambous C. Conjugate gradient algorithm for efficient training of artificial neural networks[J]. IEE Proc Part G, 1992, 139(2): 301-310.

[4] Johansson E M, Dowla F U, Goodman D M. Backpropagation learning for multilayer feed-forward neural networks using the conjugate gradient method [J]. Int J of Neural Systems, 1992, 2(3): 291-301.

[5] Williams R J, Zipser D. A learning algorithm for continually running fully recurrent neural networks[J]. Neural Computation, 1991, 3(4): 375-385.

[6] Chin-Sung Liu, Ching-Huan Tseng. Quadratic optimization method for multilayer neural networks with local error-backpropagation[J]. Int J of Systems Science, 1999, 30(7): 889-898.