

一种鲁棒 BP 算法及其在非线形动态系统 辨识中的应用

郭创新 景雷 梁年生 叶鲁卿 曾杰

(华中理工大学电力工程系 武汉 430074)

摘要 利用多层前馈神经网络的非线性建模特性, 基于动态 BP 网络的串并联和并联模型, 提出一种高鲁棒性 BP 算法. 与传统的 BP 算法相比, 鲁棒 BP 算法有 5 个优点: (1) 适合于非线性动态系统辨识; (2) 辨识精度高; (3) 不必内插所有训练样本; (4) 具有高鲁棒性, 能抵制过失误差和量测误差; (5) 收敛速度得到了改进, 因为错误样本的影响得到了适度的抑制. 将该算法用于非线性动态系统辨识, 仿真结果表明此方法是有效的。

关键词 非线性动态系统辨识, 鲁棒 BP 算法, 最大似然法, 动态神经网络

1 引言

BP 网络能逼近任意的非线性映射^[1], 已成功地应用于非线性系统的辨识^[2~3]. Narendra 等人提出一种推广的动态神经网络系统及相应的算法^[2]; 文献[3]也提出了一种基于动态 BP 神经网络系统的系统辨识方法, 并用于离线非线性系统的辨识. 他们都是基于传统的 BP 算法, 被训练的模型要内插所有的训练点, 当噪声存在时, 学习的模型在受噪声影响的数据点振荡得很厉害, 达不到必要的精度, 也影响辨识速度, 这显然是系统辨识所不期望的.

提高辨识算法的抗干扰能力可谓是系统辨识发展的一条主线, 而现有的大多数辨识算法都要求噪音过程具有平稳的特性, 这在实际中较难得到解决, 这样就对辨识算法的鲁棒性提出了要求.

本文所提出的鲁棒 BP 算法, 能够进行非线性系统的动态辨识, 在小噪声扰动时处于稳定, 对过失误差有很强的鲁棒性. 与传统 BP 算法相比, 鲁棒 BP 算法特别适合于非线性动态系统辨识, 它具有逼近精度高、收敛速度快、不易陷入局部极小、鲁棒性好等优点.

2 基于多层前馈网络的非线性动态系统辨识

前馈神经网络一般由输入层、隐层、输出层组成. 基于神经网络的建模与辨识问题实际上即是: 在给定输入、输出样本 (x_p, y_p) 下, 通过某种训练算法对神经网络的数值进行调节、学习的过程, 以使预报输出与实际过程输出之差极小. 亦即, 对系统 $y_p = f(x_p) + e_p$ (e_p 是输出的测量噪声), 神经网络的一个逼近函数 \hat{f} , 使 \hat{f} 和 f 之间的误差极小.

本文所讨论的非线性动态系统为如下单输入单输出 n 阶系统

$$y^{(n)} = f\left[y, y, \dots, y^{(n-1)}, u\right] \quad (1)$$

(1) 式中 u 和 y 分别是系统的输入和输出, $f(\cdot): R^n \times R \rightarrow R$ 是非线性函数. 对非线性动态系统, 可采用如下两种辨识模型

$$\hat{y}(k) = \hat{f}[y(k-1), \dots, y(k-n), u(k-1), \dots, u(k-m)] \quad (2)$$

$$\hat{y}(k) = \hat{f}[\hat{y}(k-1), \dots, \hat{y}(k-n), u(k-1), \dots, u(k-m)] \quad (3)$$

(2) 式和(3) 式分别表示神经网络辨识的串并联模型(SPM) 和并联模型(PM). 如图 1 和图 2. 在这里, 我们进行神经网络的非线性动态辨识模型主要采用含有神经网络的串并联模型(SPM). 原因在另文讨论. 下两节将剖析旧的 BP 算法, 导出新的鲁棒 BP 辨识算法.

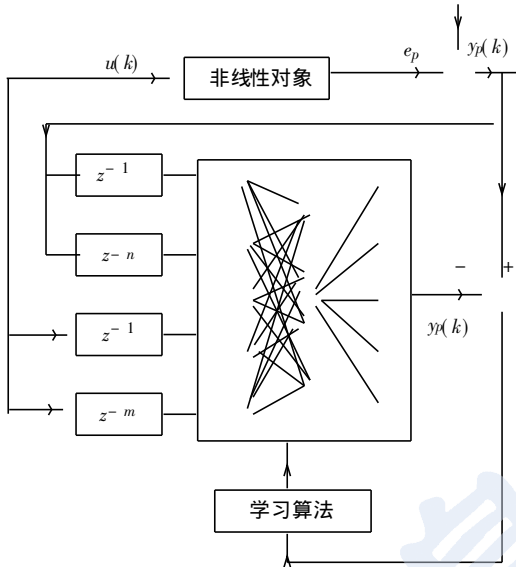


图1 非线性动态系统SPM辨识框图

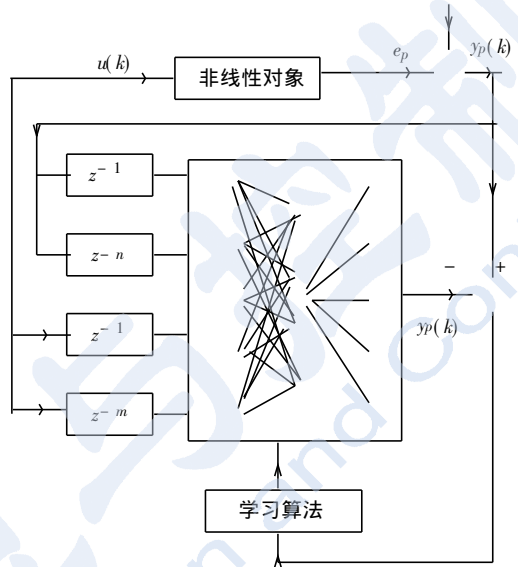


图2 非线性动态系统PM辨识框图

3 BP 算法应用于辨识存在的问题

我们知道成功地运用 BP 算法进行辨识依赖于训练样本的质量, 但如果过误差、量测误差存在于训练样本中, 则 BP 算法效率较低, 系统辨识达不到满意的结果. 主要表现如下.

(1) 非平滑内插. 我们知道, BP 算法是通过迭代调整权值使误差平方函数达到最小的,

即使得 $E_{LS} = \sum_{p=1}^N \sum_{j=1}^{n_j} (y_{pj} - \hat{y}_{pj})^2$ 达到最小, 其中 \hat{y}_{pj} 是输出层第 j 个神经元的估计输出, y_{pj} 则是相应的期望输出(即系统的实际输出). 由此易看出, 使 E_{LS} 达到 0, BP 算法则不顾样本的质量而必须使辨识模型内插所有的训练样本. 若在训练样本中存在过误差或量测噪声并且神经网络的层数及神经元的个数选择不当时, 则内插会使辨识达不到必要的精度. 而训练前就能选好正确的网络结构至今仍是研究神经网络用于辨识的难题.

(2) 噪声的影响. 辨识中的干扰通常有: 环境噪声、原器件及负荷扰动、观测随机误差、计算误差等. 这时辨识模型一般为: $y_p = f(x_p) + e_p$. 这样, 样本 (x_p, y_p) 就偏离了正确的位置. 如果误差很大, 则 BP 算法在这一错误样本位置产生反复振荡, 给收敛速度造成很大的影响.

(3) 当所辨识的系统高度非线性时, 样本的学习逼近精度不高. 表现为在一些样本点学习效果较好, 而在另一些样本点学习效果较差. 以下例来说明这一现象, 例如有两个样本点, 其期望输出分别为 $y_1 = 0.9, y_2 = 0.001$, 学习后样本点的实际输出分别为 $y_1 = 0.81, y_2 = 0.091$. 两个样本点的实际输出和期望输出间绝对误差 $E_1 = E_2$, 但第一个样本点的实际输出和

期望输出间误差为 10%，而第二个样本误差为期望输出的 90 倍。究其原因，BP 算法中的误差函数 E_{LS} 是误差的绝对量，不能有效地表征样本学习的误差程度。因为对不同的样本点，即使其绝对误差量相同，但样本点的实际输出和期望输出间误差程度也不同。

4 用于辨识的鲁棒 BP 算法

4.1 能量函数修正

由于在 BP 算法的学习过程中，学习效果提高即样本实际输出和期望输出间误差的减少是通过反向传播过程进行的，因此不能有效地表征样本的学习误差就不能有效地提高学习效果。

BP 算法使用最小平方方法获得能量函数 E_{LS} 最小的权值。最小平方方法仅当偏差是服从独立的且恒等的高斯分布时才能产生最佳值。由于误差的概率分布都不知道，并且多层前馈网络是高度的非线性，所以最小平方方法的最优性不能保证。结合上一节分析 BP 存在的问题，如下我们讨论几个最小平方方法的修改，它们是推导鲁棒 BP 算法的基础。

- (1) 首先用相对量形式的误差 $r_{pj} = 1 - y_{pj} / \hat{y}_{pj}$ 代替原先绝对量形式的误差函数；
- (2) 把最小平方方法修改为最大似然方法。我们首先定义一个新的能量函数

$$E_R = \sum_{p=1}^N \sum_{j=1}^{n_l} \mathcal{Q}(r_{pj})$$

$\mathcal{Q}(\cdot)$ 是一个时变函数。 $r_{pj} (j = 1, \dots, n_l)$ 是服从独立和恒等的分布，在概率密度函数 $P_e(r)$ 下，最大似然方法等价于使目标函数 E_R 最小，这里 $\mathcal{Q}(r) = -\ln P_e(r)$ 。

显然当 $P_e(r) = e^{-r^2/2}$ 时， $\mathcal{Q}(r) = r^2$ 。亦即当误差服从独立和恒等的高斯分布时，最小平方方法的解也是最大似然方法的解。这样，利用最大似然方法和误差概率知识，可以得到一个不同的目标函数。这就需要知道误差分布的概率密度函数 $P_e(r)$ ，我们知道

$$\hat{e}_p = (r_{p1}, \dots, r_{pj}, \dots, r_{pm}^T) = y_p - \hat{f}(x_p)$$

求 e_p 的估计值 \hat{e}_p 需要 $f(x_p)$ 的估计值 $\hat{f}(x_p)$ ，同样反过来求 $f(x_p)$ 的估计值 $\hat{f}(x_p)$ 也需要 e_p 的估计值 \hat{e}_p 。为了克服这一矛盾，并且避免求解 $P_e(r)$ ，我们利用文献[4]设计的 M 估计器，它一般设计成较小的噪声扰动下是稳定的和对数据有过失误差、量测误差时是鲁棒的，主要用来处理当准确的概率模型未知的情况。

取 $\Psi_i(r) = d\mathcal{Q}(r) / dr$ ，并把它设计成 Hample 双曲正切估计器的形状^[4]，但与文献[4]不同的是它的形状是随着学习时间的变化而变化的。这里 $\Psi_i(r)$ 定义为

$$\Psi_i(r) = \begin{cases} r, & r < a(t) \\ C_1 \tanh(C_2(b(t) - r)) \operatorname{sign}(r), & a(t) < r < b(t) \\ 0, & r > b(t) \end{cases} \quad (4)$$

其中 $a(t)$ 和 $b(t)$ 是随时间变化的截点， C_1 和 C_2 是常数。

在 Hample 的双曲正切估计器中的常数截点是基于噪声知识可统计出来的，而在函数 $\Psi_i(r)$ 中，时变的截点 $a(t)$ 和 $b(t)$ 则需要处理统计知识并不能直接可用的情况。下面，我们将介绍如何基于噪声统计知识迭代计算 $a(t)$ 和 $b(t)$ 。

设 q 是线外点百分比的上界，则 $(1 - q)$ 是正确数据点百分比的下界。对 $(1 - q)N$ 个正确数据点的误差排序

$$r(t)_{(1)} \quad r(t)_{(2)} \quad \dots \quad r(t)_{((1-q)N)}$$

利用文献[5]自举方法, 我们可以估计出第 $(1-q)N$ 个数据绝对误差的置信区间. 设 $l_{(1-q)}(t)$ 和 $u_{(1-q)}(t)$ 分别是其下界和上界, 即

$$\text{Prob}(\{l_{(1-q)}(t) < r(t) < u_{(1-q)}(t)\}) = 0.95 \quad (5)$$

这样 $a(t)$ 和 $b(t)$ 分别设为 $l_{(1-q)}(t)$ 和 $u_{(1-q)}(t)$, 因为(4)式成立的条件是仅当 $l_{(1-q)}(t)$ 和 $u_{(1-q)}(t)$ 的估计值正确, 所以它们是在学习过程中不断地迭代重新置值的.

新的鲁棒 BP 算法中, 权值的调整与 $Q(r)$ 的变化率成比例, 所以在每一个瞬时 Δt 大约有 $(1-q)N$ 个训练数据(它们的残差小于等于 $r(t)$ ($(1-q)N$)) 用于学习, 而另外 qN 个训练样本因被怀疑为线外点, 所以在学习过程中它们对学习过程的影响得到了抑制. 需要说明的是, q 值不宜过大, 也不能太小. 太大则学习过程中会排除很多有用的样本, q 太小则不能把线外点找出来.

4.2 鲁棒 BP 算法

4.2.1 多层前馈网的动力学特性

设一个 L 层多层前馈网, 各层神经元的个数为 $n_l, l=1, 2, \dots, L$, 设样本 (x_p, y_p) , 其中 $x_p = (x_{p1}, \dots, x_{pn_1})^T$ 为输入向量, $y_p = (y_{p1}, \dots, y_{pn_L})^T$ 为输出向量. 第 p 个样本第 l 层的第 i 个神经元的输出为 o_{pi}^l , 节点 $l+1$ 层第 j 个神经元的输入为 net_{pj}^{l+1} , 第 l 层的第 i 个神经元到第 $l+1$ 层的第 j 个神经元的权值为 w_{ij}^l , 则网络的估计输出 \hat{y}_p 由下列 BP 算法前向传播递归方程决定, 它表征了网络的动力学特性.

$$\text{net}_{pj}^{l+1} = \sum_{i=1}^{n_l} w_{ij}^l o_{pi}^l \quad (6)$$

$$o_{pj}^{l+1} = \hat{f}(\text{net}_{pj}^{l+1}) \quad (7)$$

$$\hat{y}_p = o_{pj}^L \quad (8)$$

式中 $l=1, 2, \dots, L-1, i=1, \dots, n_{l+1}, j=1, \dots, n_{l+1}, p=1, \dots, N$ (以下 l, i, j, p 均同此参数), $f(\cdot)$ 为 Sigmoid 函数.

4.2.2 鲁棒 BP 算法的推导

为了确立权值, 使目标函数 E_R 受网络的动力学特性非线性制约, 我们在这里利用一个基于拉格朗日乘子方法的数学结构^[6]用来推导 BP 算法.

目标是找出 w^* 使 E_R 最小, 且服从(6), (7), (8)式给出的网络的动力学特性非线性制约.

定义拉格朗日函数

$$\text{Lag} = \sum_{p=1}^N \sum_{j=1}^{n_L} Q(r_{pj}) + \sum_{p=1}^N \sum_{l=1}^{L-1} \sum_{j=1}^{n_{l+1}} \lambda_{pj}^l (o_{pj}^L - f(\sum_{i=1}^{n_{l+1}} w_{ij}^{l+1} o_{pi}^l)) \quad (9)$$

(9)式的第一项是一个时变的目标函数, 前面已经论述过; 第二项代表约束条件. 由拉格朗日乘子方法的原理求解 w^* 等同于使 Lag 最小的拉格朗日函数的一个临界点. 即满足条件

$$\begin{cases} \frac{\partial \text{Lag}}{\partial \lambda_{pj}^l} = 0 \\ \frac{\partial \text{Lag}}{\partial o_{pj}^l} = 0 \\ \frac{\partial \text{Lag}}{\partial w_{ij}^l} = 0 \end{cases} \quad (10)$$

(10)式中第一个方程的计算比较容易, 因为格朗日函数(9)式的第一项中不含 λ_{pj}^l , 第二项是

λ_{pj}^l 的线性函数. 这样计算第一个方程的结果即是得到前向的传播方程(6), (7), (8).

从(10)式第二个方程中, 我们得到两个对 λ_{pj}^l 递归的公式, 一个是对输出层单元, 另一个对于隐层单元

$$\lambda_{pj}^l = \Psi_l(r_{pj}) \quad (11)$$

$$\lambda_{pj}^l = \prod_{i=1}^{n_l} w_{ij}^l f \left(\prod_{i=1}^{n_l} w_{ij}^l o_{pi}^l \right) \lambda_{pj}^{l+1} \quad (12)$$

式中, $\Psi_l(\cdot)$ 是 $\mathcal{Q}_l(\cdot)$ 的导数, 令

$$\delta_{pj}^l = f \left(\prod_{i=1}^{n_{l-1}} w_{ij}^{l-1} o_{pi}^{l-1} \right) \lambda_{pj}^l \quad (13)$$

这样(11)式和(12)式转化为

$$\delta_{pj}^l = f' \left(\prod_{i=1}^{n_{l-1}} w_{ij}^{l-1} o_{pi}^{l-1} \right) \Psi_l(r_{pj}) \quad (14)$$

$$\delta_{pj}^l = f \left(\prod_{i=1}^{n_{l-1}} w_{ij}^{l-1} o_{pi}^{l-1} \right) \prod_{i=1}^{n_l} w_{ij}^l \delta_{pj}^{l+1} \quad (15)$$

这就是计算误差信号的反向传播方程. 注意, 当使用最小平方能量函数时, (14)式变成

$$\delta_{pj}^l = f \left(\prod_{i=1}^{n_{l-1}} w_{ij}^{l-1} o_{pi}^{l-1} \right) r_{pj} \quad (16)$$

求解(10)式第三个方程, 我们就得到权值的修正方程.

$$\begin{aligned} \frac{\partial \text{Lag}}{\partial w_{ij}^l} &= \sum_{p=1}^N \frac{\partial}{\partial w_{ij}^l} [\lambda_{pj}^{l+1} (o_{pj}^{l+1} - f(\prod_{i=1}^{n_l} w_{ij}^l o_{pi}^l))] = \\ &- \sum_{p=1}^N f \left(\prod_{i=1}^{n_l} w_{ij}^l o_{pi}^l \right) \lambda_{pj}^{l+1} o_{pi}^l = - \sum_{p=1}^N \delta_{pj}^{l+1} o_{pi}^l = 0 \end{aligned}$$

这样, 求使 E_R 最小的 w^* 解可以运用梯度下降法迭代求解. 迭代方程为

$$w_{ij}^l(t+1) = w_{ij}^l(t) - \eta \frac{\partial \text{Lag}}{\partial w_{ij}^l(t)} = w_{ij}^l(t) + \eta \sum_{p=1}^N \delta_{pj}^{l+1} o_{pi}^l \quad (17)$$

w_{ij}^l 是随机产生的. 这个方程类似于 BP 算法, 它的解是局部最优, 而非最优解.

鲁棒 BP 算法包括方程(14), (15), (17)以及计算 $\Psi_l(r)$ 的方法. 鲁棒 BP 算法与传统 BP 算法的数学区别是在输出层权值的变化不同, 前者是按 $\Psi_l(r)$ 比例变化, 而后者按残差 r 变化.

5 鲁棒 BP 算法的实现

现在, 在鲁棒 BP 算法中有 5 个参数需要在学习之前确定. 它们是: 线外点所占百分数的上界 q , 它是基于用户对训练数据的可信度来选择的; 两次连续修正 $\mathcal{Q}_l(r)$ 的间隔时间 Δt ; 阈值 δ , 它是用作监测能量在初始估计量上有一个大的下降值; 最小能量水平 ϵ 和最小能量变化标准 ϵ_l , 它们是用来结束学习过程的; ϵ 和 ϵ_l 都根据近似精度的需要确定. 详细描述如下: ① 随机初始化权向量; ② 用前向传播方程(6), (7), (8)对每一个训练输入计算估计输出, 用残差平方和平均值(MSSR)计算能量函数的值; ③ 对输出层单元和隐层单元用 BP 算法计算误差信号; ④ 用权值修正方程(17)修正权向量; ⑤ 如果当前的 MSSR 和上一个的 MSSR 的差异大于 δ , 则转向⑥, 否则转向②; ⑥ 置 $k=0$, 它用于修正 $\mathcal{Q}_l(r)$; ⑦ 对每一个训练输入用前向传播方程计算估计输出, 计算 E_R , 如果 $E_R < \delta$ 或两次能量的差异小于 ϵ_l , 则结束学习过程; ⑧ 如果 k 是 Δt 的倍数, 即 $k \bmod \Delta t = 0$, 则修正 $a(t)$ 和 $b(t)$; ⑨ 用鲁棒 BP 算法计算输出层单元和隐层单

元的误差信号; 10 用权值修正方程修正网络权值, $k = k + 1$, 转到⑦。

6 仿真研究

考虑如下被辨识系统

$$y_p(k+1) = f[y_p(k), y_p(k-1), y_p(k-2), u(k), u(k-1)]$$

$$f[x_1, x_2, x_3, x_4, x_5] = \frac{x_1 x_2 x_3 x_4 \left(\frac{x_3 - 1}{1 + x_3^2 + x_2^2} \right) + x_5}{1 + x_3^2 + x_2^2} \quad (18)$$

可以验证对象 $[-1, 1]$ 间是 BIBO 稳定的, 这里取网络结构为 5-20-10-1, 将 $[-1, 1]$ 中随机分布的点作为输入分量, 取 $q = 0.0$; $\Delta t = 100$; $\delta = 0.10$; $\epsilon = 0.00005$; $\epsilon_l = 0.000005$; $\eta = 0.67$ 。用鲁棒 BP 算法, 运用图 1 的 SPM 模型, 对网络进行训练。

$$\hat{y}_p(k+1) = \hat{f}[\hat{y}_p(k), \hat{y}_p(k-1), \hat{y}_p(k-2); u(k), u(k-1)]$$

学习 68000 次后收敛, 将辨识模型改为 PM 模型, 并取

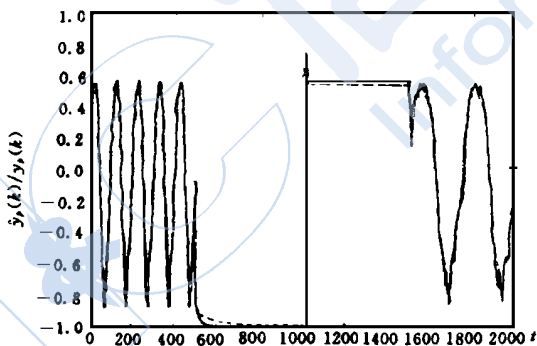
$$u(k) = \begin{cases} \sin(2\pi k/250), & k < 500 \\ 0.8\sin(2\pi k/250) + 0.2\sin(2\pi k/25), & 1500 < k < 2000 \\ -1.0, & 500 < k < 1000 \\ 1.0, & 1000 < k < 1500 \end{cases}$$

$y_p(k)$ 和 $\hat{y}_p(k)$ 如图 3 示。

为比较鲁棒 BP 算法与传统 BP 算法在防止辨识中的量测噪声和过失误差, 我们随机取 2 个量测噪声和过失误差点加入训练样本, 网络结构同上, 并取 $q = 0.05$; $\Delta t = 100$; $\delta = 0.10$; $\epsilon = 0.00005$; $\epsilon_l = 0.000005$; $\eta = 0.68$, 分别用两种辨识方法运用 SPM 模型进行辨识, 传统 BP 算法 90000 次收敛, 鲁棒 BP 算法 58000 次收敛, 之后都采用 PM 模型, 取

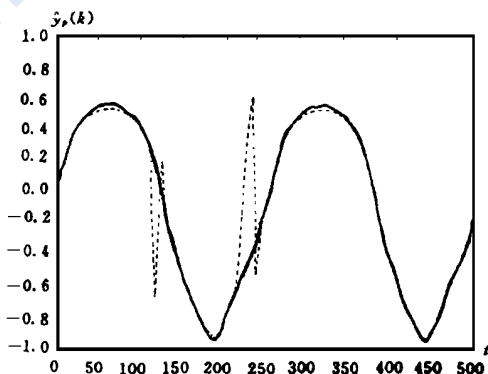
$$u(k) = \sin(2\pi k/250), \quad k < 500$$

两种方法的 $\hat{y}_p(k)$ 输出如图 4 所示。



—系统实际输出; —神经网络模型输出

图 3 $\hat{y}_p(k)$ 与 $y_p(p)$



—鲁棒 BP 算法辨识输出; —传统 BP 算法辨识输出

图 4 两种方法的 $\hat{y}_p(k)$ 输出

仿真结果表明, 鲁棒 BP 算法能有效地进行非线性动态系统辨识, 并在逼近精度、收敛速度、防止量测噪声和过失误差对辨识影响优于传统 BP 算法。

7 结论

本文分析了传统 BP 算法在进行系统辨识时存在着非平滑内插、受噪声影响很大、样本的学习逼近精度不高等问题,并对传统 BP 算法能量函数和误差函数进行了修正,同时,进一步研究了把静态 BP 算法用于非线性动态系统辨识的可行性,在此基础上提出了高鲁棒性动态 BP 算法.把它用于非线性系统动态辨识的应用研究中,仿真结果表明,鲁棒 BP 算法能有效地进行非线性动态系统辨识,并在逼近精度、收敛速度、防止量测噪声和过失误差对辨识影响优于传统 BP 算法.从而为非线性动态系统辨识提供了一个新的鲁棒算法.

参 考 文 献

- 1 Funahashi K I. On the Approximate Realization of Continuous Mapping by Neural Networks. *Neural Networks*, 1989, (2): 183 ~ 192
- 2 Narendra K S, Parthasarathy K. Identification and Control of Dynamic Systems Using Neural Networks. *IEEE Trans Neural Network*, 1990, 1(1): 4 ~ 27
- 3 田 明等. 基于动态 BP 神经网络的系统辨识方法. *自动化学报*, 1993, 19(4): 450 ~ 453
- 4 Hampel F R, Rousseeuw P J, Rouchetti E, Stahel W A. *Robust Statistics— The Approach Based on Influence Function*. New York: Wiley, 1986
- 5 Efrom B, Tibshirani R. Bootstrap Methods for Standard Errors Confidence Intervals and Other Measures of Statistical Accuracy. *Statist Sci*, 1988, 1: 54 ~ 77
- 6 Cun Y Le. A Theoretical Framework for Backpropagation. In: D Touretzky, G Hinton, T Sejnowski Eds, *Proc 1988 Connectionist Models Summer School*, San Mateo, CA: Morgan Kaufmann, 1988: 21 ~ 28

A ROBUST BP ALGORITHM AND ITS APPLICATION ON THE IDENTIFICATION OF NONLINEAR DYNAMIC SYSTEM

GUO Chuangxin JING Lei LIANG Niansheng YE Luqing ZENG Jie

(Huazhong University of Science and Technology, Wuhan 430074)

Abstract The paper presents a high robust BP algorithm using the nonlinear of multilayer feedforward neural networks and based on the series-parallel model of the dynamic BP network. In contrast to the conventional BP algorithm, five advantages of the robust BP algorithm are: (1) fitting to the dynamic identification of nonlinear system; (2) the identifiability accuracy is very high; (3) not interpolating all the training points; (4) it is robust against gross errors and measuring errors; (5) its rate of convergence is improved since the influence of incorrect sample is gracefully suppressed. The algorithm is applied to the dynamic identification of nonlinear system and the simulation result shows the new method is efficient.

Key words nonlinear system, dynamic identification, robust BP algorithm, maximum likelihood method

作者简介

郭创新,男,27岁,博士生.研究领域为专家系统、神经网络、遗传算法及其在水电系统中的应用.

景 雷,男,26岁,博士生.研究领域为水轮机调速器的智能控制.

梁年生,男,58岁,教授,博士生导师.研究领域为水轮机调速器的自完善控制,自学习控制与智能控制.