



# 多层前向网络的随机学习新算法 及其工业应用

戴连奎 马龙华 李晓东

(浙江大学工业控制技术研究所 杭州 310027)

**关键词** 多层前向网络, 学习算法, 催化裂化装置.

## 1 引言

许多复杂的工业过程建模问题, 如系统辨识、参数估计与函数逼近等, 都可转化为下面的非线性最优拟合问题.

给定样本集合  $\{x(k), y(k) | k=1, \dots, M\}$ , 求最优拟合函数

$$\hat{y}(k) = f(x(k)), \quad k = 1, M, \quad (1)$$

使

$$J = \frac{1}{M} \sum_{k=1}^M [y(k) - \hat{y}(k)]^2 \quad (2)$$

极小化, 其中  $x(k) = (x_1(k), \dots, x_n(k))^T$  为样本输入向量;  $M$  为样本数;  $y(k) \in R^1$  为样本实际输出;  $f(\cdot)$  为最优拟合函数;  $\hat{y}(k)$  为对应的输出估计值.

神经网络作为一种新型的非线性系统描述方法, 它通过对简单非线性函数的复合以实现复杂的函数, 具有模型结构简单、自学习功能强等特点, 目前已广泛应用于各种工业过程. 本文针对多层前向网络反向传播学习算法存在的不足, 提出了一种新的多层前向网络随机学习算法, 并将其应用于某催化裂化装置粗汽油干点的实时估计与控制. 工业应用结果表明, 该随机学习算法具有较快的学习速度和较高的拟合精度, 能够避免学习过程陷入局部极小点, 可广泛用于工业过程的建模与控制.

## 2 多层前向网络的随机学习算法

反向传播学习算法<sup>[1]</sup>解决了多层前向网络的学习问题, 使前向网络成为目前应用最广泛的神经网络之一, 但它存在以下主要问题:

1) 学习算法的收敛速度过慢;

2) 由于网络的权系数较多, 采用普通的梯度下降优化算法, 极易陷入局部极小点, 最终使估计误差无法进一步减少.

为此, 文[2]提出了利用动态规划思想训练前向网络的方法; Witts 等人基于随机优

化思想提出了前向网络的随机学习算法<sup>[3]</sup>. 本文结合网络的结构特点, 提出了相应的随机学习算法.

多层前向网络的存在定理<sup>[4]</sup>证明了一个仅含有隐层非线性连续变换函数的多输入单输出三层网络可以用来逼近有界闭区域上的任一连续函数  $f(x)$ , 因此以上最优拟合问题等价于以下非线性优化问题.

$$\min_{V,W} J = \frac{1}{M} \sum_{k=1}^M [y(k) - \hat{y}(k)]^2,$$

$$\hat{y}(k) = \sum_{i=1}^m v_i z_i(k) + v_0, \quad z_i(k) = g(h_i(k)), \quad h_i(k) = \sum_{j=1}^n w_{ij} x_j(k) + w_{i0},$$

$$i = 1, \dots, m; \quad k = 1, \dots, M. \quad (3)$$

其中  $V = (v_0, v_1, \dots, v_m)^T$  为输出层权系数阵;

$W = \{w_{ij} | i = 1, \dots, m; j = 0, 1, \dots, n\}$  为隐层权系数阵;

$m$  为隐节点数;  $n$  为输入节点数;  $\hat{y}(k)$  为该网络的输出估计值;

$h_i(k), z_i(k)$  分别为第  $i$  个隐节点的输入与输出,  $i = 1, \dots, m$ ;

$g(x)$  为 Sigmoid 函数  $g(x) = \frac{1}{1 + \exp(-x)}$ .

针对此类三层网络的具体特点: 若  $\{w_{ij}\}$  一定, 则  $v_i (i = 0, 1, \dots, m)$  的估计问题属典型的线性最小二乘问题, 可采用递推线性最小二乘算法予以求解; 而隐层权系数可采用随机寻优方法获得, 由此提出了以下随机学习算法.

多层前向网络的随机学习算法:

**第一步. 样本归一化**

$$x_i(k) \leftarrow \frac{x_i(k) - x_{i,\min}}{x_{i,\max} - x_{i,\min}}, \quad y(k) \leftarrow \frac{y(k) - y_{\min}}{y_{\max} - y_{\min}}, \quad i = 1, \dots, n; \quad k = 1, \dots, M. \quad (4)$$

其中  $x_{i,\max}, x_{i,\min}$  为  $\{x_i(k) | k = 1, \dots, M\}$  的最大最小值,  $i = 1, \dots, n$ ;  $y_{\max}, y_{\min}$  为  $\{y(k) | k = 1, \dots, M\}$  的最大最小值.

**第二步.** 设置允许误差  $\epsilon$ , 最大迭代次数  $q_{\max}$  与隐节点数  $m$ , 令迭代次数  $q = 0$ ; 初始化均方误差  $E_{\text{opt}}$  与隐层权系数  $\{w_{ij}\}$  及其最优值  $\{w_{ij}^{\text{opt}}\}$ ,

$$w_{ij} = \xi_i, w_{ij}^{\text{opt}} = w_{ij}, \quad i = 1, \dots, m, j = 0, 1, \dots, n,$$

其中  $\xi_i$  为  $(-1, 1)$  之间的均匀分布随机数.

**第三步.** 针对每组样本, 计算网络的隐单元输出

$$z_i(k) = g\left(\sum_{j=1}^n w_{ij} x_j(k) + w_{i0}\right), \quad i = 1, \dots, m, k = 1, \dots, M.$$

**第四步.** 基于  $\{z_1(k), \dots, z_m(k); y(k) | k = 1, \dots, M\}$  用递推最小二乘算法估计权系数阵  $\{v_i | i = 0, 1, \dots, m\}$ .

**第五步.** 计算整体均方误差

$$E = \sqrt{\frac{1}{M} \sum_{k=1}^M [y(k) - \hat{y}(k)]^2}, \quad q \leftarrow q + 1.$$

**第六步.** 若  $E < E_{\text{opt}}$ , 则令  $w_{ij}^{\text{opt}} \leftarrow w_{ij}, v_i^{\text{opt}} \leftarrow v_i$ ; 置  $q = 0$ . 若  $E < \epsilon$ , 则迭代结束; 否则, 令

$w_{ij} \leftarrow w_{ij} + \Delta w_{ij}$ , 并转第三步. 其中  $\Delta w_{ij} = \sigma \sum_{i=1}^{12} \xi_i$  为近似  $N(0, \sigma^2)$  的高斯分布随机数,  $\sigma$  为



平均方差.

**第七步.** 若  $E \geq E_{\text{opt}}$ , 则令  $w_{ij} \leftarrow w_{ij}^{\text{opt}}$ . 若  $q \geq q_{\text{max}}$ , 则迭代过程结束; 否则, 令  $w_{ij} \leftarrow w_{ij} + \Delta w_{ij}$  并转第三步. 其中  $\Delta w_{ij}$  为近似  $N(0, \sigma^2)$  的高斯分布随机数.

### 3 工业应用——催化裂化装置粗汽油干点的实时估计与控制<sup>[5]</sup>

基于某工业催化裂化装置的现场采集数据, 采用前面提出的神经网络随机学习算法, 建立了粗汽油干点的估计模型, 实现了该质量指标的预测控制. 长期工业现场运行数据表明, 估计模型具有较高的精度, 而且可适应处理量的变化与原料性质的波动. 所开发的闭环预测控制系统具有良好的调节性能与较强的鲁棒性, 该系统的引入提高了分馏塔的平稳性, 而使粗汽油干点的波动范围显著减少, 并使分馏塔的操作处于较佳的水准上, 产生了明显的经济效益.

工业应用结果表明, 由于该随机学习算法克服了反向传播学习算法收敛速度慢, 极易陷入局部极小点的不足, 使得多层前向网络用于实时估计与控制成为可能. 同时, 由于该网络的输出层为线性函数, 有时为了提高实时性, 可采用递推最小二乘算法仅更新输出层权系数, 这就给前向网络的工业应用带来了便利.

#### 参 考 文 献

- [1] Rimelhart D E, McClelland J L. Parallel distributed processing. MIT Press, 1986.
- [2] Saratchandram P. Dynamic programming approach to optimal weight selection in multilayer neural networks. *IEEE Trans. Neural Network*, 1991, **2**: 465—467.
- [3] Witts M J *et al.* Artificial neural networks in process engineering. *IEEE Processings-D*, 1991, **138**: 256—266.
- [4] Robert H N. Theory of the backpropagation neural network. In: Proc. of IJCNN, 1989, **1**, 593—598.
- [5] 李晓东, 戴连奎, 孙优贤. 粗汽油干点的实时估计与控制. 化工自动化及仪表, 1996, **23**增刊: 23—27.

## A NOVEL RANDOM LEARNING ALGORITHM FOR MULTILAYER NEURAL NETWORKS AND ITS APPLICATION

DAI LIANKUI MA LONGHUA LI XIAODONG

(Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027)

**Key words** Multilayer networks, learning algorithm, fluidized catalytic cracking units.