

带有稳定学习的递归神经网络动态偏最小二乘建模

王 魏¹, 柴天佑^{1,2}, 赵立杰^{1,3}

(1. 东北大学 流程工业综合自动化国家重点实验室, 辽宁 沈阳 110819;

2. 东北大学 自动化研究中心, 辽宁 沈阳 110819; 3. 沈阳化工学院 信息工程学院, 辽宁 沈阳 110142)

摘要: 针对目前非线性动态偏最小二乘(PLS)建模方法在拟合较强非线性化工过程时存在的问题, 提出一种基于稳定学习的递归神经网络动态PLS建模方法. 该算法将递归神经网络与Hammerstein模型相结合, 对外部PLS提取的特征向量进行内部建模, 具有逼近较强非线性化工过程的能力, 改善了模型的适用范围. 此外, 采用带有稳定学习的参数更新算法对模型参数进行在线修正, 改善了模型的预测精度和自适应能力. 将此方法应用于氧化铝生产过程中铝酸钠溶液组分浓度建模实验, 仿真结果表明, 本方法是可行有效的.

关键词: 偏最小二乘; 递归神经网络; Hammerstein模型; 软测量

中图分类号: TP183 **文献标识码:** A

Dynamic partial least squares modeling with recurrent neural networks of stable learning

WANG Wei¹, CHAI Tian-you^{1,2}, ZHAO Li-jie^{1,3}

(1. State Key Laboratory of Integrated Automation for Process Industries, Northeastern University, Shenyang Liaoning 110819, China;

2. Research Center of Automation, Northeastern University, Shenyang Liaoning 110819, China;

3. Information Engineering School, Shenyang Institute of Chemical Technology, Shenyang Liaoning 110142, China)

Abstract: A dynamic modeling algorithm is proposed for a strongly nonlinear chemical process, it is based on partial least squares (PLS) and recurrent neural networks with a stable learning rate. The outer PLS algorithm reduces the dimensionality of data and extracts score vector, and the inner model which combines recurrent neural networks with Hammerstein model captures the nonlinear characters to extend the model application scope. Besides, the stable learning algorithm updates the model parameters to improve the prediction precision and adaptation ability. This method is implemented in the process of alumina production to measure the component concentrations of sodium aluminate solution. Simulation results show that the modeling method is effective.

Key words: partial least squares; recurrent neural networks; Hammerstein model; soft sensing

1 引言(Introduction)

偏最小二乘(PLS)作为一种线性系统辨识方法, 已广泛应用于过程建模和监控领域. 它的主要局限性在于只能提取数据中的线性信息. 许多实际数据本质上是非线性的, 需要能够建立非线性关系的模型. 因此出现了一系列非线性PLS方法, 主要分为内部拟合方法, 如二次函数(QPLS)、样条函数(SPL-PLS)、神经网络(NNPLS)^[1]、模糊推理(FPLS)^[2]等和外部拟合方法, 如核函数(KPLS)^[3]、小波函数(Wavelet-PLS)^[4]等. 在非线性的动态PLS建模方面, 文献[5]中提出一种基于动态偏最小二乘(DFPLS)的多变量非线性动态建模方法, 通过对PLS提取的特征向量进行ARX建模来实现动态, 内部利用模糊聚类或多个TSK线性模型加权求和的方式拟合系统非线

性, 适用于具有多工况和非线性不强的工业过程. 在化工过程方面, S. Lakshminarayanan等在文献[6]中提出了将PLS与Hammerstein模型相结合的非线性动态建模方法, 其中Hammerstein模型由一个非线性增益串接一个线性系统构成, 用线性子系统描述对象的动态特性, 用非线性增益来校正线性模型, 是化工过程中最常用的模型之一. 然而, 文献[6]用多项式描述Hammerstein模型的非线性部分, 只能代表一类轻微的非线性过程, 对于具有中等或严重非线性的化工过程, 精度和适用性将会有所下降.

本文针对文献[5]和文献[6]中模型适用于多工况或轻微非线性过程的局限性, 结合大部分化工过程存在较强非线性的特点, 提出一种基于稳定学习的递归神经网络动态PLS建模方法, 外部采用PLS方

收稿日期: 2010-10-08; 收修稿日期: 2011-04-27.

基金项目: 国家重点基础研究发展计划资助项目(2009CB320601); 国家自然科学基金资助项目(61020106003, 60904079, 61004009); 国家创新研究群体科学基金资助项目(60821063); 高等学校学科创新引智计划资助项目(B08015); 中国博士后自然科学基金资助项目(20100471464).

法将高维数据空间投影到低维特征空间,有效地克服了普通最小二乘回归的共线性问题,内部将递归神经网络与Hammerstein模型相结合,更具有通用性,能够拟合非线性较强的工业过程,适用范围更广泛.此外,由于采用固定学习速率的梯度下降法时系统性能较差,相比较时变学习速率具有更快的收敛性,可以保证误差的稳定性^[7-8].因此,本文还提出一种带有稳定学习的参数辨识算法.将其应用于氧化铝生产过程中铝酸钠溶液组分浓度仿真实验,结果表

明本方法精度较高,泛化能力较强,适用于更广泛的化工过程建模.

2 带有稳定学习的动态PLS建模(Dynamic PLS modeling with stable learning rate)

本文提出的基于稳定学习的递归神经网络动态PLS建模方法,如图1所示.外部采用PLS算法从样本数据中成对地提取最优成分,使多变量系统降维,并消除其间的耦合关系,内部采用Hammerstein递归神经网络模型拟合系统的非线性及动态特性.

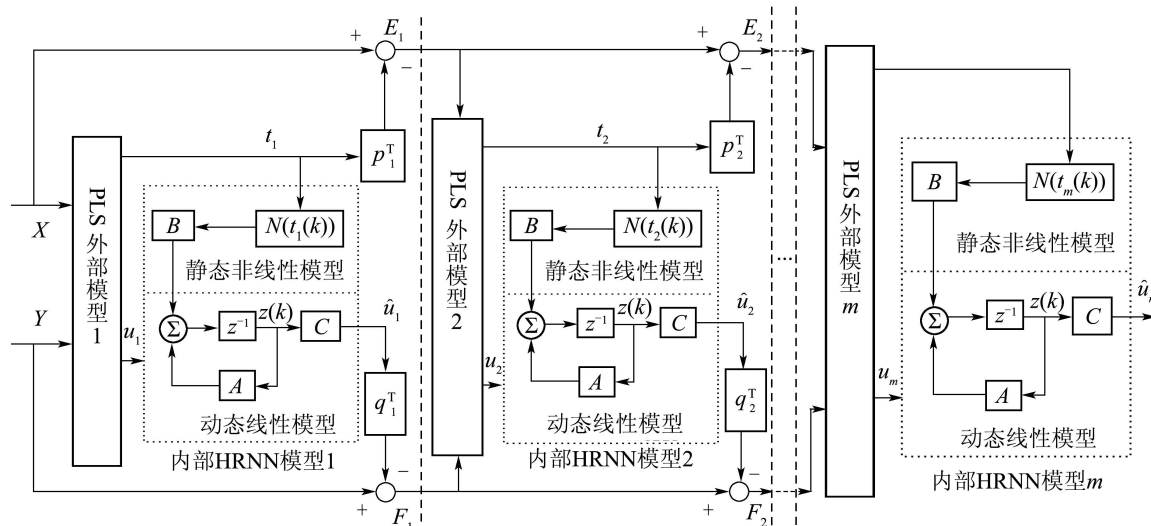


图1 Hammerstein递归神经网络动态PLS建模方法

Fig. 1 Dynamic PLS modeling with Hammerstein recurrent neural networks

2.1 Hammerstein递归神经网络(Hammerstein recurrent neural networks(HRNN))

Hammerstein模型结构如图1内部所示,分成静态非线性和动态线性两部分.静态非线性模型通过一个非线性变换将输入空间映射到状态空间,然后状态空间通过一个线性动态映射到输出空间.状态空间方程表示如下:

$$z(k+1) = Az(k) + BN(t(k)), \quad (1)$$

$$u(k) = Cz(k), \quad (2)$$

其中: $t = [t_h]$, $u = [u_h]$ 表示第 $h(h = 1, 2, \dots, m)$ 个内部模型的输入和输出变量,系数矩阵 $A \in \mathbb{R}^{J \times J}$, $B \in \mathbb{R}^{J \times J}$, $C \in \mathbb{R}^{1 \times J}$, $N \in \mathbb{R}^J$.

递归神经网络学习能力强、结构灵活,是一种建模和控制复杂动态系统的有效工具,具有收敛速度快、收敛精度高的特点.因此,结合递归神经网络与Hammerstein模型的优点,构成HRNN模型,网络结构如图2所示.

整个模型结构可以看成是一个具有两个隐含层的神经网络,第1个隐层用来描述Hammerstein模型的静态非线性部分,第2个隐层描述动态线性部分.以第 h 个内部模型为例,输出量 $u_h(k)$ 和状态变量 $z_h(k)$ 通过每层节点计算获得,步骤如下:

$$u_h(k) = c_h z_h(k) = \sum_{j=1}^J c_{hj} z_{hj}(k), \quad (3)$$

$$z_{hj}(k) = \sum_{i=1}^J (a_{hji} z_{hi}(k-1)) + b_{hjj} n_{hj}(k-1), \quad (4)$$

$$n_{hj}(k) = f(g_{hj}(k)) = \frac{e^{g_{hj}(k)} - e^{-g_{hj}(k)}}{e^{g_{hj}(k)} + e^{-g_{hj}(k)}}, \quad (5)$$

$$g_{hj}(k) = w_{hj} t_h(k) + d_{hj}, \quad (6)$$

其中: w_{hj} 是输入层与隐含层第 j 个神经元之间的权值, d_{hj} 是第 j 个隐含层节点的阈值, a_{hji} , b_{hjj} , c_{hj} 是式(1)-(2)中的系数矩阵 A , B , C 的元素,其中 h 表示第 h 个模型, c_h 表示第 h 个模型的 C 系数矩阵.

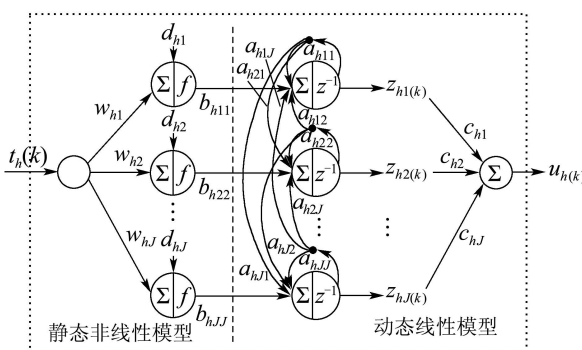


图2 HRNN模型拓扑结构图

Fig. 2 The topology of HRNN model

2.2 稳定学习(Stable learning)

根据第 h 个内部模型的输出公式, 双隐层单输出网络模型可表示为

$$\begin{aligned} \hat{y}(k) &= W_C(k)\{W_A(k)Z(k-1) + \\ &W_B(k)\varphi[W(k)X(k) + W_D(k)]\}, \quad (7) \\ Z(k-1) &= W_A(k-1)Z(k-2) + W_B(k-1) \times \\ &\varphi[W(k-1)X(k-1) + W_D(k-1)], \quad (8) \end{aligned}$$

其中: $X(k) = t_h(k)$, $y(k) = u_h(k)$ 表示网络的输入输出向量. $W = [w_j] \in \mathbb{R}^{1 \times J}$ 表示第1个隐含层权值矩阵, $W_D = [d_j] \in \mathbb{R}^{1 \times J}$ 表示第1个隐含层阈值矩阵, $W_B = [b_{jj}] \in \mathbb{R}^{J \times J}$ 表示第2个隐含层权值矩阵, $W_A = [a_{ij}] \in \mathbb{R}^{J \times J}$ 表示第2个隐含层另一权值矩阵, $W_C = [c_j] \in \mathbb{R}^{J \times 1}$ 表示输出层权值向量. 网络的输入层节点个数为1, J 表示神经网络的隐含层节点个数, φ 表示第1个隐含层节点的基函数.

根据Stone-Weierstrass定理^[9], 存在理想的权值矩阵使得

$$\begin{aligned} \hat{y}(k) &= W_C^*(k)\{W_A^*(k)Z(k-1) + \\ &W_B^*(k)\varphi[W^*(k)X(k) + W_D^*(k)]\} - \mu(k), \quad (9) \end{aligned}$$

其中: W^* , W_A^* , W_B^* , W_C^* 和 W_D^* 分别为建模误差最小时的神经网络隐含层和输出层权值矩阵, 辨识误差可表示为

$$\begin{aligned} e(k) &= \\ &W_C(k)\{W_A(k)Z(k-1) + W_B(k)\varphi[W(k)X(k) + \\ &W_D(k)]\} - W_C^*(k)\{W_A^*(k)Z(k-1) + \\ &W_B^*(k)\varphi[W^*(k)X(k) + W_D^*(k)]\} + \mu(k). \quad (10) \end{aligned}$$

Taylor级数展开:

$$\begin{aligned} e(k) &= \sum_{j=1}^J \frac{\partial \hat{y}}{\partial W_{Cj}} (W_{Cj}(k) - W_{Cj}^*(k)) + \dots + \\ &\sum_{j=1}^J \frac{\partial \hat{y}}{\partial W_{Dj}} [W_{Dj}(k) - W_{Dj}^*(k)] + \mu(k) + \varepsilon(k), \quad (11) \end{aligned}$$

其中 $\varepsilon(k)$ 为Taylor公式的高阶项, 对式中各项求导后, 误差可以表示成如下矩阵形式:

$$\begin{aligned} e(k) &= \tilde{W}_C(k)Z(k) + E^T \tilde{W}_A(k)A(k)E + \\ &E^T \tilde{W}_B(k)B(k)E + \tilde{W}(k)W_W(k) + \\ &\tilde{W}_D(k)D(k) + \delta(k), \quad (12) \end{aligned}$$

其中: $\tilde{W}_C(k) = W_C(k) - W_C^*$ 为 $1 \times J$ 矩阵, 第 $1 \times J$ 个

元素为 $c_{hJ} - c_{hJ}^*$; $\tilde{W}_A(k) = W_A(k) - W_A^*$ 为 $J \times J$ 矩阵, 第 $J \times J$ 个元素为 $a_{hJJ} - a_{hJJ}^*$; $\tilde{W}_B(k) = W_B(k) - W_B^*$ 为 $J \times J$ 矩阵, 第 $J \times J$ 个元素为 $b_{hJJ} - b_{hJJ}^*$; $\tilde{W}(k) = W(k) - W^*$ 为 $1 \times J$ 矩阵, 第 $1 \times J$ 个元素为 $w_{hJ} - w_{hJ}^*$; $\tilde{W}_D(k) = W_D(k) - W_D^*$ 为 $1 \times J$ 矩阵, 第 $1 \times J$ 个元素为 $d_{hJ} - d_{hJ}^*$; $Z(k)$ 为 $J \times 1$ 矩阵, 第 $J \times 1$ 个元素为 z_{hJ} ; E 为单位阵, $A(k)$ 为 $J \times J$ 矩阵, 第 $J \times J$ 个元素为

$$c_{hJ}(k)[z_{hJ}(k) + a_{hJJ}(k-1)z_{hJ}(k-2)];$$

$B(k)$ 为 $J \times J$ 三角阵, 对角线元素为

$$c_{hJ}(k)[n_{hJ}(k) + a_{hJJ}(k-1)n_{hJ}(k-2)];$$

$W_W(k)$ 为 $J \times 1$ 矩阵, 第 $J \times 1$ 个元素为

$$\begin{aligned} c_{hJ}(k)[t_h(k)b_{hJJ}(k) \times \frac{4}{(e^{g_{hJ}(k)} + e^{-g_{hJ}(k)})^2} + \\ a_{hJJ}(k-1)t_h(k-2)b_{hJJ}(k-2) \times \\ \frac{4}{(e^{g_{hJ}(k-2)} + e^{-g_{hJ}(k-2)})^2}], \end{aligned}$$

$D(k)$ 为 $J \times 1$ 矩阵, 第 $J \times 1$ 个元素为

$$\begin{aligned} c_{hJ}(k)[b_{hJJ}(k) \times \frac{4}{(e^{g_{hJ}(k)} + e^{-g_{hJ}(k)})^2} + \\ a_{hJJ}(k-1)b_{hJJ}(k-2) \times \\ \frac{4}{(e^{g_{hJ}(k-2)} + e^{-g_{hJ}(k-2)})^2}], \end{aligned}$$

$\delta(k) = \varepsilon(k) + \mu(k)$, $\tilde{W}_C(k)$, $\tilde{W}_A(k)$, $\tilde{W}_B(k)$, $\tilde{W}(k)$, $\tilde{W}_D(k)$ 分别表示权值估计误差, $\delta(k)$ 表示Taylor级数的高阶项与未建模动态之和. 根据误差反传原理, 第 h 个内部模型的参数更新算法如下:

$$c_{hj}(k+1) = c_{hj}(k) + \eta_{hk} e_h(k) z_{hj}(k), \quad (13)$$

$$\begin{aligned} a_{hji}(k+1) &= a_{hji}(k) + \eta_{hk} e_h(k) c_{hj}(k) \times \\ &[z_{hi}(k) + a_{hjj}(k-1) \frac{\partial z_{hj}(k-1)}{\partial a_{hji}}], \quad (14) \end{aligned}$$

$$\begin{aligned} b_{hjj}(k+1) &= b_{hjj}(k) + \eta_{hk} e_h(k) c_{hj}(k) [n_{hj}(k) + \\ &a_{hjj}(k-1) \frac{\partial z_{hj}(k-1)}{\partial b_{hjj}}], \quad (15) \end{aligned}$$

$$\begin{aligned} w_{hj}(k+1) &= w_{hj}(k) + \eta_{hk} e_h(k) c_{hj}(k) \times \\ &[t_h(k)b_{hjj}(k) \times \frac{4}{(e^{g_{hj}(k)} + e^{-g_{hj}(k)})^2} + \\ &a_{hjj}(k-1) \frac{\partial z_{hj}(k-1)}{\partial w_{hj}}], \quad (16) \end{aligned}$$

$$\begin{aligned} d_{hj}(k+1) &= d_{hj}(k) + \eta_{hk} e_h(k) c_{hj}(k) \times \\ &[b_{hjj}(k) \times \frac{4}{(e^{g_{hj}(k)} + e^{-g_{hj}(k)})^2} + \\ &a_{hjj}(k-1) \frac{\partial z_{hj}(k-1)}{\partial d_{hj}}], \quad (17) \end{aligned}$$

$$z_{hj}(k+1) = \sum_{i=1}^J a_{hji} z_{hi}(k) + b_{hjj} n_{hj}(t_h(k)), \quad (18)$$

其中:

$$\frac{\partial z_{hj}(k-1)}{\partial a_{hji}} = z_{hi}(k-2),$$

$$\frac{\partial z_{hj}(k-1)}{\partial b_{hjj}} = n_{hj}(k-2),$$

$$\frac{\partial z_{hj}(k-1)}{\partial w_{hj}} = t_h(k-2) b_{hjj}(k-2) \times$$

$$\frac{4}{(e^{g_{hj}(k-2)} + e^{-g_{hj}(k-2)})^2},$$

$$\frac{\partial z_{hj}(k)}{\partial d_{hj}} = b_{hjj}(k-2) \times \frac{4}{(e^{g_{hj}(k-2)} + e^{-g_{hj}(k-2)})^2}.$$

选择

$$\eta_{hk} = \frac{\eta_h}{1 + \Phi_k}, \quad 0 < \eta_h \leq 1,$$

$$\Phi_k = \|Z(k)\|^2 + \|A(k)\|^2 + \|B(k)\|^2 + \|W_W(k)\|^2 + \|D(k)\|^2.$$

那么, 平均辨识误差

$$\bar{J} = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T \|e_h(k)\|^2 \leq \frac{\eta_h \bar{\delta}}{\pi} \quad (19)$$

有界, 其中: $\pi = \frac{\eta_h}{[1 + \max(\Phi_k)]^2}$, $\bar{\delta} = \max[\|\delta(k)\|^2]$.

3 仿真实验(Simulation experiment)

氧化铝生产过程铝酸钠溶液主要组分浓度(包括苛性碱 c_K , 氧化铝 c_A , 碳酸碱 c_C)机理模型难以建立, 采用本文提出的方法对其进行数据建模. 采集溶液在不同温度 T_1, T_2, T_3 下的电导率 d_1, d_2, d_3 , 利用其中的350组样本建模, 150组测试.

采用留一交叉验证法计算预测误差累积平方和最小时, 特征向量数目为1, 因此选择第一对主元建立内部HRNN模型. 网络结构选择[1 10 10 1], 将训练好的模型用于预测, 结果如图3-5所示. 从实验结果可看出, 通过PLS的外部变换与HRNN模型的内部拟合, 达到了良好的模型预测效果.

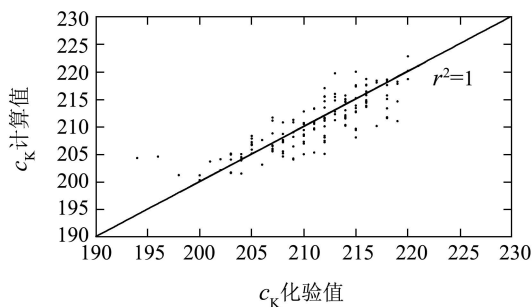


图 3 苛性碱浓度化验值与计算值相关系数
Fig. 3 The correlation coefficient of c_K

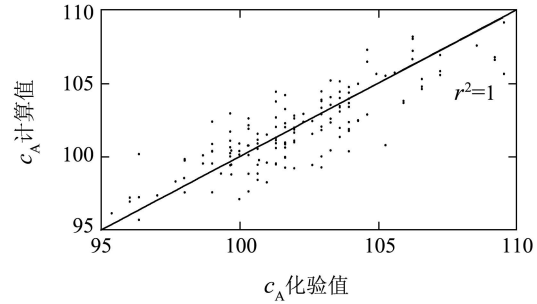


图 4 氧化铝浓度化验值与计算值相关系数
Fig. 4 The correlation coefficient of c_A

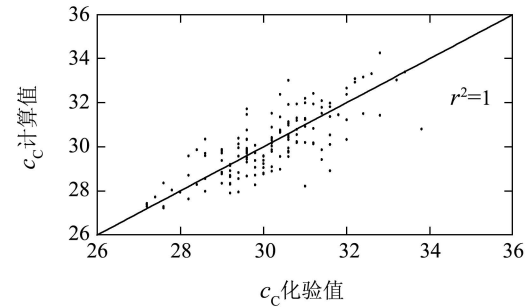


图 5 碳酸碱浓度化验值与计算值相关系数
Fig. 5 The correlation coefficient of c_C

为评价模型性能, 使用最大误差和均方相关系数对模型的精度进行统计, 公式如下:

$$\text{MAXE} = \max_{i=1}^n (\|\hat{y}_i - y_i\|), \quad (20)$$

$$r^2 = \frac{(n \sum_{i=1}^n \hat{y}_i y_i - \sum_{i=1}^n \hat{y}_i \sum_{i=1}^n y_i)}{(n \sum_{i=1}^n \hat{y}_i^2 - (\sum_{i=1}^n \hat{y}_i)^2)(n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2)}. \quad (21)$$

与已有建模方法DFPLS^[5]和PLS与Hammerstein^[6]进行比较, 不同方法的预测结果如表1所示.

表 1 不同方法精度比较

Table 1 Accuracy of different methods

性能指标	方法			
	DFPLS	PLS+Hammerstein	PLS+HRNN	
c_K	MAXE c_K	7.32	20.26	7.23
	r^2	0.22	0.38	0.70
c_A	MAXE c_A	14.31	10.88	8.48
	r^2	0.34	0.33	0.68
c_C	MAXE c_C	4.60	3.32	2.46
	r^2	0.24	0.37	0.61

可见本文提出的建模方法最大误差较低, 均方相关系数较大, 测试精度较高.

以本文提出的方法为基础, 结合氧化铝生产过程铝酸钠溶液组分浓度的特性分析, 设计并开发了组分浓度检测系统并成功应用于原矿浆制备工序, 取得了很好的效果^[10].

4 结论(Conclusion)

本文针对化工过程数据维数较高且非线性较强的特点, 提出一种新的非线性动态建模方法. 通过PLS方法从数据中成对提取主成分, 克服高维共线性问题; 内部HRNN模型对提取的每对成分建模, 拟合了系统的非线性和动态特性; 最后, 采用稳定学习算法更新模型参数, 保证了内部模型建模误差有界. 氧化铝生产过程仿真实验表明了方法的有效性.

参考文献(References):

- [1] BAFFI G, MARTIN E B, MORRIS A J. Non-linear projection to latent structures revisited (the neural network PLS algorithm)[J]. *Computers and Chemical Engineering*, 1999, 23(9): 1293 – 1307.
- [2] BANG Y H, YOO C K, LEE I B. Nonlinear PLS modeling with fuzzy inference system[J]. *Chemometrics and Intelligent Laboratory Systems*, 2003, 64(2): 137 – 155.
- [3] 姚林, 阳建宏, 何飞, 等. 基于核偏最小二乘的锌层重量预测模型[J]. *控制工程*, 2008, 2(15): 154 – 157, 167.
(YAO Lin, YANG Jianhong, HE Fei, et al. Forecasting model for zinc coating weights based on kernel partial least squares[J]. *Control Engineering of China*, 2008, 2(15): 154 – 157, 167.)
- [4] TEPPOLA P, MINKKINEN P. Wavelet-PLS regression models for both exploratory analysis and process monitoring[J]. *Journal of Chemometrics*, 2000, 14(5/6): 383 – 399.
- [5] 王魏, 赵立杰, 岳恒, 等. 多变量动态模糊偏最小二乘建模方法及其应用[J]. *系统仿真学报*, 2010, 22(5): 1309 – 1312, 1318.
(WANG Wei, ZHAO Lijie, YUE Heng, et al. Multivariable modeling based on dynamic fuzzy PLS and its application[J]. *Journal of System Simulation*, 2010, 22(5): 1309 – 1312, 1318.)
- [6] LAKSHMINARAYANAN S, SHAH S L, NANDAKUMAR K. Modeling and control of multivariable processes: dynamic PLS approach[J]. *AIChE Journal*, 1997, 43(9): 2307 – 2322.
- [7] YU W, MARCO A, MORENO A, et al. System identification using hierarchical fuzzy neural networks with stable learning algorithm[J]. *Journal of Intelligent & Fuzzy System*, 2007, 18(2): 171 – 183.
- [8] 丛秋梅, 柴天佑, 余文. 污水处理过程的递归神经网络建模[J]. *控制理论与应用*, 2009, 26(1): 8 – 14.
(CONG Qiumei, CHAI Tianyou, YU Wen. Modeling wastewater treatment plant via hierarchical neural networks[J]. *Control Theory & Applications*, 2009, 26(1): 8 – 14.)
- [9] CYBENKO G. Approximation by superposition of sigmoidal activation function[J]. *Mathematics of Control, Signals, and Systems*, 1989, 2(4): 303 – 314.
- [10] WANG W, CHAI T Y, YU W, et al. Modeling component concentrations of sodium aluminate solution via Hammerstein recurrent neural networks[J]. *IEEE Transactions on Control System Technology*, in press.

作者简介:

王 魏 (1982—), 女, 博士研究生, 目前研究方向为复杂工业过程的智能建模, E-mail: ww_wangwei@163.com;

柴天佑 (1947—), 男, 中国工程院院士, 教授, IEEE Fellow, IFAC Fellow, 欧亚科学院院士, 目前研究方向为自适应控制、多变量智能解耦控制、流程工业过程综合自动化理论、方法与技术, E-mail: tychai@mail.neu.edu.cn, 通讯作者;

赵立杰 (1972—), 女, 副教授, 目前研究方向为复杂工业过程的智能建模、软测量方法研究等, E-mail: zlj_lunlun@163.com.