时变过程在线辨识的即时递推核学习方法研究

刘 毅 [1] 金福江 [2] 高增梁 [1]

摘 要 为了及时跟踪非线性化工过程的时变特性, 提出即时递推核学习 (Kernel learning, KL) 的在线辨识方法. 针对待预测的新样本点, 采用即时学习 (Just-in-time kernel learning, JITL) 策略, 通过构造累积相似度因子, 选择与其相似的样本集建立核学习辨识模型. 为避免传统即时学习对每个待预测点都重新建模的繁琐, 利用两个临近时刻相似样本集的异同点, 采用递推方法有效添加新样本, 并删减旧模型的样本, 以快速建立新即时模型. 通过一时变连续搅拌釜式反应过程的在线辨识, 表明了所提出方法在保证计算效率的同时, 较传统递推核学习方法提高了辨识的准确程度, 能更好地辨识时变过程.

关键词 过程辨识, 即时学习, 核学习, 最小二乘支持向量回归, 递推辨识

引用格式 刘毅, 金福江, 高增梁. 时变过程在线辨识的即时递推核学习方法研究. 自动化学报, 2013, **39**(5): 602−609

DOI 10.3724/SP.J.1004.2013.00602

Online Identification of Time-varying Processes Using Just-in-time Recursive Kernel Learning Approach

LIU Yi[1] JIN Fu-Jiang[2] GAO Zeng-Liang[1]

Abstract An online identification method using just-in-time recursive kernel learning (KL) is proposed to trace the time-varying characteristics of nonlinear chemical processes. For each query sample, a just-in-time kernel learning (JITL) model is established using the similar set constructed by a presented cumulative similarity factor. Different from traditional just-in-time learning approaches discarding their models at each time, an efficient modeling strategy is proposed to reduce the computational load by utilizing the similarity between two neighborhood models. Consequently, a new just-in-time kernel learning model can be quickly constructed using the recursive updating algorithm, by introducing new samples and deleting different samples. The superiority of the proposed online identification method is demonstrated by a continuous stirred tank reactor process with time-varying parameters, showing better prediction performance compared with conventional recursive kernel learning approaches.

Key words Process identification, just-in-time learning (JITL), kernel learning (KL), least squares support vector regression (LSSVR), recursive identification

Citation Liu Yi, Jin Fu-Jiang, Gao Zeng-Liang. Online identification of time-varying processes using just-in-time recursive kernel learning approach. *Acta Automatica Sinica*, 2013, **39**(5): 602−609

非线性过程辨识始终是自动化控制领域的研究热点和难点[1−3]. 针对某些化工过程, 其较强的非线性和时变性等复杂特性制约着辨识模型的准确程度. 因此, 有效地在线辨识过程的状态信息对于监测过程的运行、保障其产品质量及进一步实施先进控制和过程优化均具有重要意义. 神经网络和模糊系统等方法在非线性过程辨识领域已有广泛研究和应用[1−3]. 但前者在网络结构的选取和训练上仍存在困难; 后者则对模糊规则的提取缺乏普遍规律. 此外, 两类辨识方法都需要大量建模样本. 针对有限样本的非线性辨识问题, 支持向量回归 (Support vector regression, SVR) 及最小二乘支持向量回归 (Least squares SVR, LSSVR) 等核学习 (Kernel learning, KL) 建模方法表现出了更好的性能, 最近受到了更多的关注[4−12].

为了及时跟踪过程的非线性和时变性, 常采用递推方法来更新辨识模型. 大部分递推辨识方法引入滑动时间窗, 依靠窗口的滑动来更新辨识模型[11−12]. 此类方法针对辨识模型的修正主要有 3 类: 1) 随窗口滑动增/删个别数据, 但只能适应过程的缓慢变化; 2) 增/删块状的数据窗, 但在辨识模型修正前也难以适应过程的变化特性; 3) 引入遗忘因子以突出新的数据信息, 但针对不同过程, 遗忘因子

收稿日期 2012-05-15 录用日期 2012-12-19
Manuscript received May 15, 2012; accepted December 19, 2012
国家自然科学基金 (61004136, 61143005, 61273069) 资助
Supported by National Natural Science Foundation of China (61004136, 61143005, 61273069)
本文客座编委 王宏
Recommended by Guest Editor WANG Hong
1. 浙江工业大学化工机械设计研究所 杭州 310032 2. 华侨大学信息科学与工程学院 厦门 361021
1. Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Hangzhou 310032 2. School of Information Science and Engineering, Huaqiao University, Xiamen 361021

并不好选取. 此外, 数据窗口的长度也不好选取[11−12].

为了更好地在线辨识时变过程, 本文提出采用即时学习 (Just-in-time learning, JITL) 的方式[13−18], 针对待预报新样本, 根据 "相似输入产生相似输出" 的原则, 选择与其相似的样本建立辨识模型. 为了克服传统 JITL 对每个待预测点都重新建模的繁琐, 利用两个临近时刻相似样本集的异同点, 采用递推方法有效增/删模型的新/旧样本, 以快速获得新即时模型的参数. 在目前常用的 KL 辨识方法中, LSSVR 较 SVR 更容易获得递推形式, 且计算量较小, 更适合在线辨识[10−12, 15]. 因此, 为便于比较研究, 本文采用 LSSVR 及其递推形式 (Recursive LSSVR, R-LSSVR)[11, 15] 作为所提出即时递推 KL 辨识方法的基础. 主要的创新点体现在: 1) 提出 JITL 方法辨识时变非线性过程, 在辨识的方法上区别于基于滑动时间窗等传统的递推方法, 通过选择更相似的样本集建立的即时模型能更好地辨识当前的过程. 2) 提出快速建立 JITL 模型, 利用前后临近时刻模型样本间的异同, 有效地更新模型以提高计算效率.

1　LSSVR 辨识方法及其递推算法

1.1　非线性过程辨识的 LSSVR 方法

非线性过程建模可统一表示为用函数 f: $\mathcal{X} \to \mathcal{Y}$ 来拟合样本集 $\mathcal{S} = \{(\boldsymbol{x}_1,\ y_1),\ \cdots,\ (\boldsymbol{x}_k,\ y_k)\} \subset \mathcal{X} \times \mathcal{Y}$, 其中 $\boldsymbol{x}$ 为自变量, $\boldsymbol{y}$ 为因变量. KL 框架下的建模问题用非线性映射 $\phi(\cdot)$ 将数据集 $\mathcal{X}$ 映射到高维特征空间 $\mathcal{H}$, 即 $\phi(\boldsymbol{x}_i)$: $\mathcal{X} \to \mathcal{H}$, 并在其中构建回归函数[12, 15]. 一大类非线性过程可表示为[1−3]

$$y_k = f(y_{k-1}, \cdots, y_{k-n_y}, u_{k-1}, \cdots, u_{k-n_u}) + e_k \quad (1)$$

其中, u_k 和 y_k 分别是 k 时刻过程的调节变量和被控变量, n_u 和 n_y 为相应阶次, $f(\cdot)$ 为待求的非线性函数. e_k 为相应的噪声项. 记模型的广义输入向量为 $\boldsymbol{x}_k = [y_{k-1}, \cdots, y_{k-n_y}, u_{k-1}, \cdots, u_{k-n_u}]^{\mathrm{T}}$, 则 KL 建模框架下的非线性模型可表示为[12,15]

$$y_k = f(\boldsymbol{w}_k, b_k, \boldsymbol{x}_k) + e_k = \boldsymbol{w}_k^{\mathrm{T}} \phi(\boldsymbol{x}_k) + b_k + e_k \quad (2)$$

其中, $\boldsymbol{w}_k$ 和 b_k 分别表示 k 时刻过程模型的权值向量和偏置, 采用 LSSVR 求解模型 f[5]:

$$\min J(\boldsymbol{w}_k) = \frac{1}{2}\|\boldsymbol{w}_k\|^2 + \frac{\gamma}{2}\|\boldsymbol{e}_k\|^2$$

$$\text{s.t. } y_i - \boldsymbol{w}_k^{\mathrm{T}} \phi(\boldsymbol{x}_i) - b_k - e_i = 0, i = 1, \cdots, k \quad (3)$$

其中, $\boldsymbol{e}_k = [e_1, e_2, \cdots, e_k]^{\mathrm{T}}$; $\frac{\|\boldsymbol{w}_k\|^2}{2}$ 为正则化项. 惩罚系数 $\gamma > 0$ 用于控制模型解的光滑度, 其值越大代表对误差的惩罚力度越强. 引入拉格朗日乘子法求解上述等式约束优化问题, 最终可化简为[5]

$$\begin{bmatrix} K_k + \frac{I_k}{\gamma} & \mathbf{1}_k \\ \mathbf{1}_k^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_k \\ b_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_k \\ 0 \end{bmatrix} \quad (4)$$

其中, $\boldsymbol{a}_k = [a_{k,1}, \cdots, a_{k,k}]^{\mathrm{T}}$ 为 k 时刻的拉格朗日乘子, $\boldsymbol{y}_k = [y_1, \cdots, y_k]^{\mathrm{T}}$, $I_k \in \mathbf{R}^{k\times k}$ 为单位阵, $\mathbf{1}_k = [1 \cdots 1]^{\mathrm{T}} \in \mathbf{R}^k$ 为全 1 向量; K_k 为核矩阵, 由 Mercer 定理[4−5], 其元素可表示为 $K_k(i,j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle \quad \forall i,j = 1, \cdots, k$. 记 $H_k = K_k + \frac{I_k}{\gamma}$, $P_k = H_k^{-1}$, 求解可得模型参数:

$$\begin{cases} \boldsymbol{a}_k = P_k \left[\boldsymbol{y}_k - \frac{\mathbf{1}_k \mathbf{1}_k^{\mathrm{T}} P_k \boldsymbol{y}_k}{\mathbf{1}_k^{\mathrm{T}} \boldsymbol{y}_k \mathbf{1}_k} \right] \\ b_k = \frac{\mathbf{1}_k^{\mathrm{T}} P_k \boldsymbol{y}_k}{\mathbf{1}_k^{\mathrm{T}} \boldsymbol{y}_k \mathbf{1}_k} \end{cases} \quad (5)$$

进而对 k+1 时刻进行预报:

$$f(\boldsymbol{w}_k, b_k,\ \boldsymbol{x}_{k+1}) = \boldsymbol{a}_k^{\mathrm{T}} \boldsymbol{k}_{k+1} + b_k \quad (6)$$

其中, $\boldsymbol{w}_k = \sum_{i=1}^{k} a_{k,i} \phi(\boldsymbol{x}_i)$, $\boldsymbol{k}_{k+1}(i) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_{k+1}) \rangle, \forall i = 1, \cdots, k$.

1.2　辨识模型的递推更新

为了跟踪过程的非线性和时变特性, 辨识模型需要在线递推更新. 假设 k 时刻的 R-LSSVR 模型有 N_k 个节点 (可理解为滑动时间窗模型包含的样本). 与式 (4) 类似, 需求解如下方程组[11, 15]:

$$\begin{bmatrix} K_{N_k} + \frac{I_{N_k}}{\gamma} & \mathbf{1}_{N_k} \\ \mathbf{1}_{N_k}^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{N_k} \\ b_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_{N_k} \\ 0 \end{bmatrix} \quad (7)$$

其中, 各项含义参照式 (4), 区别在于此处的 R-LSSVR 模型阶次为 N_k. 记 $H_{N_k} = K_{N_k} + I_{N_k}$, $P_{N_k} = H_{N_k}^{-1}$, 则此时辨识模型的参数为

$$\begin{cases} \boldsymbol{a}_{N_k} = P_{N_k} \left[\boldsymbol{y}_{N_k} - \frac{\mathbf{1}_{N_k} \mathbf{1}_{N_k}^{\mathrm{T}} P_{N_k} \boldsymbol{y}_{N_k}}{\mathbf{1}_{N_k}^{\mathrm{T}} \boldsymbol{y}_{N_k} \mathbf{1}_{N_k}} \right] \\ b_{N_k} = \frac{\mathbf{1}_{N_k}^{\mathrm{T}} P_{N_k} \boldsymbol{y}_{N_k}}{\mathbf{1}_{N_k}^{\mathrm{T}} \boldsymbol{y}_{N_k} \mathbf{1}_{N_k}} \end{cases} \quad (8)$$

当 $[\boldsymbol{x}_{k+1}, y_{k+1}]$ 加入 R-LSSVR 辨识模型后, 记相应的节点为 $[\boldsymbol{x}_{N_{k+1}}, y_{N_{k+1}}]$, 并更新模型参数. 注意到求解 $\boldsymbol{a}_{N_k}$ 和 b_{N_k} 的关键在于求矩阵 H_{N_k} 的逆 P_{N_k}. 根据矩阵和的求逆公式, 可得 $P_{N_{k+1}}$ 的递推式为[19]

$$P_{N_{k+1}} = \begin{bmatrix} P_{N_k} & 0 \\ 0 & 0 \end{bmatrix} + \boldsymbol{r}_{N_{k+1}} \boldsymbol{r}_{N_{k+1}}^{\mathrm{T}} z_{N_{k+1}} \quad (9)$$

其中, $\boldsymbol{r}_{N_{k+1}} = \left[\boldsymbol{V}_{N_{k+1}}^{\mathrm{T}} P_{N_k}, -1\right]^{\mathrm{T}}$; $\boldsymbol{V}_{N_{k+1}} = \left[K\left(\boldsymbol{x}_1, \boldsymbol{x}_{N_{k+1}}\right), \cdots, K\left(\boldsymbol{x}_{N_k}, \boldsymbol{x}_{N_{k+1}}\right)\right]^{\mathrm{T}}$; $z_{N_{k+1}} = 1\Big/\left(v_{N_{k+1}} - \boldsymbol{V}_{N_{k+1}}^{\mathrm{T}} P_{N_k} \boldsymbol{V}_{N_{k+1}}\right)$ 和 $v_{N_{k+1}} = \boldsymbol{K}\left(\boldsymbol{x}_{N_{k+1}}, \boldsymbol{x}_{N_{k+1}}\right) + 1/\gamma$ 为标量. 因此, 模型每增加一个节点, $P_{N_{k+1}}$ 可由 P_{N_k} 递推求得, 无需求逆, 运算复杂度由原来的 $\mathrm{O}(N_k^3)$ 减至 $\mathrm{O}(N_k^2)$.

当过程时变特性较为严重时, 有必要修剪模型使其适应过程; 或者模型节点过多时, 也需要修剪模型以控制其复杂度[11, 15]. 当某个节点被删除时, 相应的矩阵 H_{N_k} 要删掉一行一列 (记新矩阵为 H_N), 为避免直接求其逆矩阵 P_N, 采用文献 [15] 的递推方法. 每减少一个节点, 运算复杂度也由 $\mathrm{O}(N_k^3)$ 减至 $\mathrm{O}(N_k^2)$. 当 R-LSSVR 辨识模型的节点数目保持不变, 且每次删减的是最旧的样本时, 可视为固定滑动时间窗的在线辨识方法. 带有遗忘因子的 R-LSSVR, 本质上也是一种加权 R-LSSVR, 即突出新数据信息的权重, 但遗忘因子不好选取[11], 限于篇幅本文对带遗忘因子的 R-LSSVR 辨识方法不作探讨.

2 即时递推核学习在线辨识方法

2.1 在线辨识的 JITL 方法

按照建模的主动/被动方式, SVR, LSSVR 及其相应的递推方法均可归为主动型, 即先建立辨识模型, 认为所建模型包含了当前数据的信息; 而后可进一步用新数据更新和修正辨识模型. JITL 则是一种被动型的学习方式, 它不对当前已有的样本建立模型, 而是等到待预报新样本 $\boldsymbol{x}_q$ 到来之时, 采取 "分而治之" 的策略, 选择与其相似的样本, 建立辨识模型并预测, 预测结束后则抛弃当前模型[13−18].

由于 JITL 是在 $\boldsymbol{x}_q$ 到来时才启动, 故又称为惰性学习 (Lazy learning) 或局部学习[13−18]. 经典的 JITL 方法如局部加权回归, 它可提高普通多元回归的预测性能. JITL 优点在于: 不需要像主动学习方式建立全局模型, 而是基于 "相似输入产生相似输出", 针对每个 $\boldsymbol{x}_q$ 能使用更丰富的假设空间, 因此提高了当前模型的推广能力, 预测更准确. 现有的 JITL 建模方法包括即时最小二乘法、即时主元回归、即时偏最小二乘回归、即时 SVR、即时 LSSVR 等, 它们侧重于提高已有全局建模方法的非线性描述能力, 主要用于软测量建模、在线监测和控制[13−18], 但针对时变过程的在线辨识的研究还很少.

通常考虑样本间的距离和角度信息来描述相似度, 即待预报新样本 $\boldsymbol{x}_q$ 与样本集 $\mathcal{X}$ 间 $\boldsymbol{x}_i$ 的相似性程度 s_{qi} 可表示为[13−15]:

$$s_{qi} = \rho\|\boldsymbol{x}_i - \boldsymbol{x}_q\| + (1-\rho)\cos(\theta_{qi}) \tag{10}$$

其中, ρ 为权系数, 特殊的当 $\rho = 1$(或 $\rho = 0$) 时即只考虑距离 (或角度) 信息[13−15]. 获得相似度后, 对 s_{qi} 进行排序以选择最相似的 l 个样本构造相似样本集 $\mathcal{S}_{\mathrm{sim}}$, 通常设定在 $[l_{\min}, l_{\max}]$ 搜索合适的 l[13−15]. 然而, 由于过程在不同阶段通常具有不同的特性, 搜索范围太小难以选择到合适的 l, 范围太大则计算量太大. 此外, 仅通过 l 的大小并无法直接判断 $\boldsymbol{x}_q$ 与所选择样本的相似度, 所选择出来的 l 带有一定的盲目性. 因此, 本文提出累积相似度因子 s_l:

$$s_l = \frac{\sum_{i=1}^{l} s_{qi}}{\sum_{i=1}^{l_{\max}} s_{qi}},\ l_{\min} \le l \le l_{\max} \tag{11}$$

其中, 分母表示当前时刻所有与 $\boldsymbol{x}_q$ 相似样本的相似度总和, 分子表示所选择的具有较大相似度 l 个样本所占的贡献. 与经典的主元分析类似, 可以直接按照贡献度来选择 l (如 $s_l = 0.9$), 或者设定在贡献度的某一范围内 (如 $0.75 \le s_l \le 0.95$) 选择 l, 这样就能获得大部分相似样本. 由于 s_l 通过相似度将 $\boldsymbol{x}_q$ 与相似样本集联系起来, 比直接选择 l 意义更明确, 可使得在建立每个即时模型时所选择的样本数更合理和简便.

2.2 即时辨识模型的递推更新算法

虽然 JITL 较传统全局建模方法预测更好, 但计算量较大是其最显著缺点, 因为 JITL 针对每个待预测样本都重新建模[13−18]. 为此, 针对软测量建模, Ge 等提出在工况较稳定时不建立即时模型, 而是沿用已保存的即时模型, 以降低计算量[18]. 但是, 该方法针对那些没有建立即时模型的时刻, 由于沿用了旧模型, 降低了预测的准确程度. 因此通常只能用于稳态, 并无法用于时变过程.

为了提高 JITL 建模方法的计算效率, 提出利用前后时刻即时模型节点间的异同, 快速递推, 从而在显著降低计算量的同时, 保证当前即时模型的预报性能. 所提出的更新策略如图 1 所示, 左图为传统的采用滑动时间窗的递推辨识方法[11−12], 右图为即时递推在线辨识方法. 采用滑动时间窗的递推辨识通常在增加新样本时, 删减旧样本, 默认最旧的样本对当前辨识没用. 然而, 这种基于时间的假设并不总是成立的, 某些 "旧" 样本其实也包含了过程的一些重要信息[12, 15]. JITL 方法通过相似的样本建立即时辨识模型来描述过程的非线性, 并及时跟踪时变特

性.

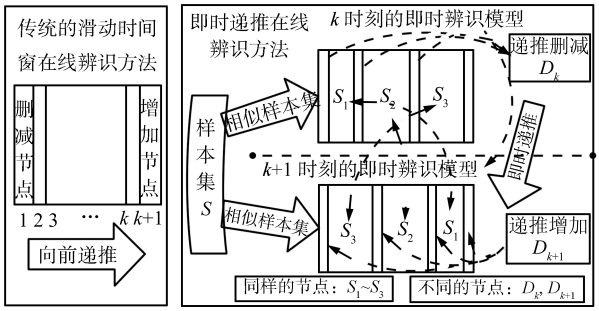


图 1　即时递推在线辨识方法示意图

Fig. 1　Flowchart of just-in-time recursive learning method for online identification

如图 1 所示, 在 k 和 $k+1$ 两个相邻时刻所建立的即时辨识模型, 其模型节点可分成相同和不同部分. 简单起见, 相同节点集部分为 S_1, S_2, S_3, 这些节点集保持不变; 针对 k 时刻的即时辨识模型的旧节点为 D_k, 采用递推算法删减这些旧节点; 针对 $k+1$ 时刻的即时辨识模型的新节点为 D_{k+1}, 采用递推算法增加这些新节点. 在算法程序的实现上, 所有的样本按照时间序列均有各自的编号, 以此作为标记进行即时模型节点的递推增删. 由于两个时刻是近邻的, 样本空间特性不会相差太大, 即使过程始终表现出时变性 (只要不总处于剧烈的跳变阶段), 通常情况下还是有大部分节点是相同的. 因此, $k+1$ 时刻的即时辨识模型便可在 k 时刻的即时辨识模型基础上递推得到, 通过加入新的相似样本, 删除旧的相似样本, 从而不用在每一时刻都重新建立即时辨识模型. 假设 $k+1$ 时刻模型的节点数为 N_k, 和 k 时刻不同的节点数为 m, 重新建立模型的计算复杂度为 $\mathrm{O}({N_k}^3)$, 采用递推算法的计算复杂度为约为 $\mathrm{O}(m{N_k}^2)$, 由于通常 $m \ll N_k$, 因此计算量大大减少.

如式 (10) 所示, 选择样本间的距离和角度信息来描述相似度, 不同的 ρ 会导致 s_l 以及后续所选择的相似样本集 $\mathcal{S}_{\mathrm{sim}}$ 的差异. 为提高参数选择的效率, 针对 ρ 和 s_l 各设定一范围, 采用快速留一交叉验证法 (Fast leave-one-out cross-validation, FLOO-CV) 来优化这两个参数[15]. 当选择的 $\mathcal{S}_{\mathrm{sim}}$ 样本数为 l 时, 其 FLOO-CV 计算的误差可表示为[15]:

$$E_l^{\mathrm{FLOO-CV}} = \sum_{i=1}^{l} \|\boldsymbol{e}_i^{\mathrm{FLOO-CV}}\| \tag{12}$$

每次采用 FLOO-CV 方法可以将模型参数选择的计算量从传统的 $\mathrm{O}(l^4)$ 减小为 $\mathrm{O}(l^3)$, 仅相当于求解一次 LSSVR, 而无需额外的计算量. 如果备选的参数对有 t 组, 则参数选择的总计算量将从传统的 $\mathrm{O}(tl^4)$ 减小为 $\mathrm{O}(tl^3)$[15].

因此, 所提出的即时递推 KL 在线辨识方法的主要步骤可归纳为:

步骤 1. 即时辨识模型的初始化: 在 k 时刻, 当首次建立即时辨识模型时, 设定 $\mathcal{S}_{\mathrm{sim}}$ 的参数 $[\rho, s_l]$ 的范围, 以及 LSSVR 模型的参数范围, 采用 FLOO-CV 方法建立初始的 LSSVR 模型, 进行预测, 并保存模型.

步骤 2. 即时辨识模型的递推更新: 在 $k+1$ 时刻, 通过每一组待选参数 $[\rho, s_l]$, 选择 $\mathcal{S}_{\mathrm{sim}}$. 首先, 判断和上一时刻即时辨识模型的异同节点, 进行递推更新, 快速计算出此参数相应的 $E_l^{\mathrm{FLOO-CV}}$; 其次, 在此基础上, 通过判断其他参数 $[\rho, s_l]$ 相应的 $\mathcal{S}_{\mathrm{sim}}$ 和计算得到的 $\mathcal{S}_{\mathrm{sim}}$ 的异同, 递推更新以快速计算出此参数相应的 $E_l^{\mathrm{FLOO-CV}}$; 最终, 选择具有最小 $E_l^{\mathrm{FLOO-CV}}$ 的参数 $[\rho, s_l]$ 及其对应的即时辨识模型, 作为该时刻的模型, 进行预报, 并只保存当前模型. 在新时刻时, 重复此递推更新步骤以获得新即时模型.

注意到针对参数 $[\rho, s_l]$ 的选择, 理论上有无数种组合, 为了获得较好的相似样本用于建立模型, 参数 $[\rho, s_l]$ 需要寻优. 在算法的实现上, 为了提高效率, 本文简单的采用选取有限组具有典型代表的参数进行组合. 通常情况下, 距离比角度有更重要的相似作用, 因此可取较大权重[13]. 此外, 如前所述, s_l 主要用于确定 l 的范围, 其上下限都较容易确定, 通常下限不小于 0.7, 这样就能确定大部分相似的样本, 本文取 $0.75 \le s_l \le 0.95$ 用于确定相似样本数 l.

综上所述, 针对时变过程的在线辨识, 所提出方法和传统方法的主要区别有:

1) 在辨识方法上: 针对当前待预测点选择相似的样本集, 即时建立辨识模型并预测, 以 FLOO-CV 选择模型参数, 以保证辨识模型的性能, 这点和传统基于滑动时间窗的辨识方法无论是在模型样本集的选择还是模型参数的选择上都有较大区别.

2) 在 JITL 模型的建立策略上: 传统的 JITL 通常在预测结束后则抛弃当前模型, 到下一时刻再重新建立[13−18], 这种 "推倒重建" 方式没有考虑模型间的相似性, 导致计算量较大, 值得商榷. 本文将递推算法引入 JITL, 利用辨识模型节点的相同和不同部分, 快速递推更新相邻的模型, 只需额外保存当前的即时模型, 便能在此基础上有效提高 JITL 的效率.

3　时变过程在线辨识: 仿真和讨论

以一非线性连续搅拌釜式反应过程为对

象[12, 20], 来验证所提出即时递推在线辨识方法的有效性和优点. 它是一典型的非线性辨识测试案例, SVR, LSSVR, 多种神经网络和模糊系统等方法都曾辨识过此过程[12, 20], 但对其时变性却研究较少.

C_a 为反应器组分流出时的浓度, T 是反应器温度, q_c 是冷却剂流量, 控制目标是通过调节 q_c 来控制 C_a, 其余操作条件及参数详见文献 [20]. 过程的非线性表现在设定点附近阻尼比变化, 高浓度区振荡较厉害[20]. 过程的辨识目标为 $C_{a,k+1}$, 广义输入向量取为 $\boldsymbol{x}_k = [C_{a,k}, C_{a,k-1}, q_{c,k}, q_{c,k-1}]^{\mathrm{T}}$[12, 20], 产生辨识数据时所有的过程变量均加入 1% 的白噪声. 图 2 给出了主要过程变量和时变参数 k_p, 其中 k_p 为频率因子[20]. $1 \sim 50$ 为训练样本集, 主要用于训练 LSSVR 模型以及作为基础的样本库: $51 \sim 100$ 为测试样本集时, 过程参数在此期间表现出常见的时变特性[20]. 注意到, 由于测试集中的过程参数 k_p 发生了时变, 相应的, 过程变量也逐渐表现出训练集没有的特性, 以此来测试所提出方法和传统辨识方法针对此类具有时变特性过程的有效性和区别.

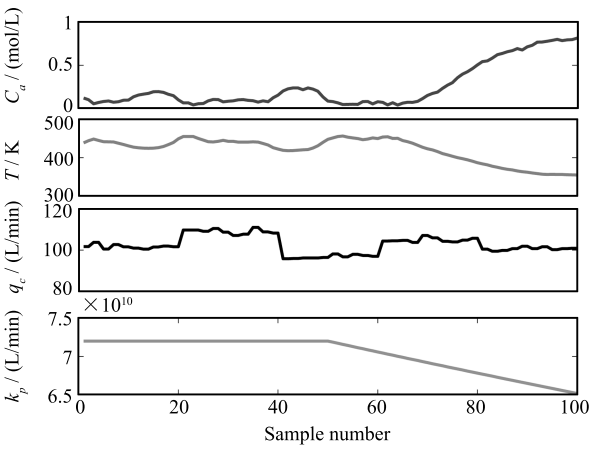


图 2 过程主要输入输出变量和时变参数

Fig. 2 Main input and output variables and the time-varying parameter of the process

所提出的在线辨识方法采用 JITL 和递推算法, 并且以 LSSVR 为基础, 因此简记为 JR-LSSVR (Just-in-time recursive LSSVR, JR-LSSVR); 普通的 R-LSSVR 作为主要的比较方法, 并对其采用不同长度的滑动时间窗进行测试[11, 15]; 没有递推的 LSSVR 作为第 3 种方法[5]. 采用预测均方误差 (Root mean squares error, RMSE), 预测相对均方误差 (Relative RMSE, RE) 和预测最大绝对误差 (Maximal absolute error, MAE), 以全面衡量这 3 种辨识方法的预测性能[12]. 之前的研究表明, 针对此过程多项式核函数 (其阶次为 d) 辨识效果较好[12], 因此以下辨识均采用多项式核.

首先通过训练集, 采用 FLOO-CV 方法训练得到 LSSVR 辨识模型, 其参数 $[\gamma, d]$ 供后续 R-LSSVR 以及 JR-LSSVR 使用. 对采用不同长度滑动时间窗 N 的 R-LSSVR 方法进行测试, 由于最优的 N 事先并无法得到, 简单起见, N 取为 10 的整数倍. 在建立每一时刻的 JR-LSSVR 辨识模型时, 相似样本集的参数 $[\rho, s_l]$ 通过 FLOO-CV 方法得到, 其中 $\rho \in [0.8, 0.85, 0.9, 0.95]$, 通过 $0.75 \leq s_l \leq 0.95$ 确定 l 的范围. 获得的新样本, 均将其加入样本库, 供下一时刻的 JR-LSSVR 和 R-LSSVR 模型使用.

各种辨识方法的预测性能指标如表 1 第 $2 \sim 4$ 列所示. 综合各种误差性能指标可知, 针对过程的时变特性, JR-LSSVR 和 R-LSSVR 两种在线辨识模型都能够获得比仅采用离线辨识的 LSSVR 更高的精度. 针对 R-LSSVR, 虽然存在合适的滑动时间窗 N 能提高辨识效果, 但 N 固定的模型缺乏对过程的自适应调整能力, 如 N 太大模型更新缓慢, N 太小则模型过于敏感. 而且, 针对过程的不同特性 N 也不好选择, 如果测试集表现出来的特性和训练集不一样, 针对训练集得到的 N 往往不适合测试集; 此外, 随着过程特性的变化, N 的范围也会发生变化. 图 3 和图 4 分别给出了 JR-LSSVR, R-LSSVR ($N = 40$ 时, 预测性能大致为平均水平) 和 LSSVR 这 3 种方法针对测试集的辨识效果和绝对误差比较. 从中可知, LSSVR 方法大约从第 75 个样本开始, 辨识精度开始显著下降, 对照训练集和测试集可知, 测试集的后半部分 (对应图 2 的第 $75 \sim 100$ 个样本) 和训练集差异较大, 因此仅靠离线辨识难以跟踪过程的时变特性. 图 3 和图 4 中 R-LSSVR 辨识效果略逊于 JR-LSSVR, 个别点辨识误差较大. 综合对比表 1 以及图 3 和图 4 的结果可知, JR-LSSVR 在线辨识方法能够较好的跟踪过程的时变特性, 辨识精度始终高于 R-LSSVR, 且不用通过反复调试来指定滑动时间窗的大小.

图 5 给出了针对测试集, JR-LSSVR 在线辨识方法每次通过 FLOO-CV 自动获得相似度参数 ρ 和样本数 l, 以选择 $\mathcal{S}_{\mathrm{sim}}$ 的示意图. 从图中可知, 选择的相似样本数一开始是增加的 (从第 $51 \sim 75$ 样本), 这是因为随着过程的时变性变强, 样本 $\mathcal{X}$ 间 $\boldsymbol{x}_i$ 与待预测样本 $\boldsymbol{x}_q$ 的相似度变小, 需要更多的样本才能使累积相似度因子 s_l 达到所要求的范围, 这样 $\mathcal{S}_{\mathrm{sim}}$ 也才能较好的描述 $\boldsymbol{x}_q$. 随着具有时变特性的新样本逐渐加入样本集 $\mathcal{X}$, 只需要较少的样本便可以获得较高相似度的 $\mathcal{S}_{\mathrm{sim}}$. 这直接反映了传统的 JITL 通常

设定在 $[l_{\min}, l_{\max}]$ 搜索 l[13−15,18] 并不合理, 因为不同时刻的最优 l 差别很大. 采用 s_l 来选择 l 可通过总相似度有效地缩减搜索范围, 提高了效率和可解释性. 同时, 这也间接说明了针对递推学习[11−12,15], 选择固定长度的滑动时间窗, 在过程的不同阶段也不合适.

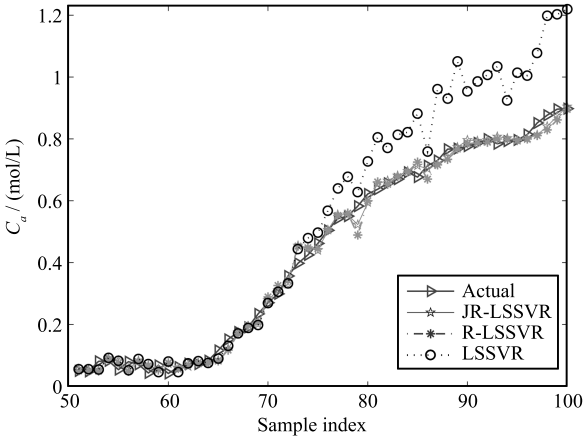


图 3　3 种辨识方法预测结果比较

Fig. 3　Comparison of prediction results with three identification methods

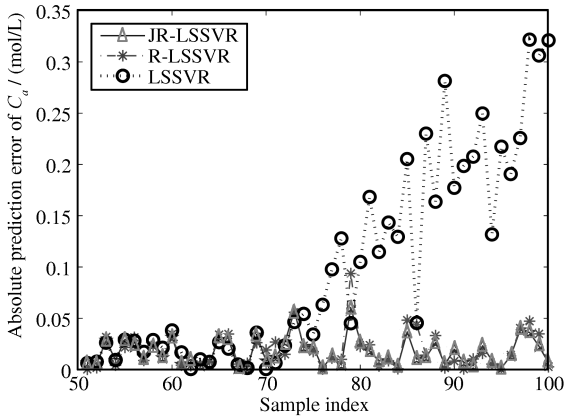


图 4　3 种辨识方法绝对误差比较

Fig. 4　Comparison of absolute prediction errors with three identification methods

此外, 如图 5 所示, 相似样本数在临近的两个测试点通常变化不大. 例如, 针对第 69 ∼ 77 个测试点, 此时可供选择的样本库分别对应着 68 ∼ 76 个样本, 而它们最终建立的即时模型节点数量大致分别为 55 ∼ 60. 仿真结果表明, 针对这几个待预测点, 所建立的即时辨识模型有大部分节点都是一样的, 相邻两个即时模型的不同的节点数只有不到 10 %, 这种情况下采用递推算法更新模型, 较重新建立即时模型能极大地降低计算量. 从图 5 可知, 有一半以上的测试点情况类似, 因此, 采用 JR-LSSVR 利用前后即时模型节点的相似性进行递推是可行和有效的.

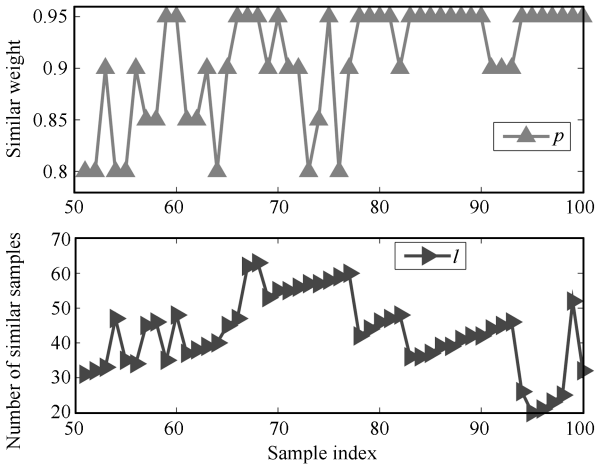


图 5　每个测试点相似度参数 ρ 和样本数 l 的选择过程

Fig. 5　Online selection of the similarity parameter ρ and the number of similar samples l for each query sample

最后, 为了进一步验证 JR-LSSVR 方法对量测噪声的鲁棒性, 对训练和测试集均加入 3 % ∼ 5 % 幅度的噪声, 进行 20 次仿真. 表 1 第 5 列和第 6 列给出了 JR-LSSVR 和具有不同滑动时间窗长度的 R-LSSVR 比较的平均结果, 从中可知 JR-LSSVR 在噪声幅度较大的情况下仍旧具有较好的辨识性能. 这是因为所提出方法针对当前待预测点选择相似的样本集 (而不是按照滑动时间窗口选择样本), 即时建立辨识模型并预测; 并以 FLOO-CV 作为模型参数的选择标准, 以保证辨识模型的性能. 因此, 在每个待辨识点都建立了一个较好的模型, 故能在过程有噪声的条件下获得比传统递推辨识较好的效果. R-LSSVR 在 $N = 30 \sim 40$ 能够取得较好的辨识效果, 但当 N 选择不合适时, 辨识结果并不理想. 相比之下, JR-LSSVR 较 R-LSSVR 对量测噪声具有较好的鲁棒性, 始终能够获得较好的辨识性能. 值得说明的是, 所提出的方法只是在有高斯噪声的条件下进行测试, 如何从理论上分析噪声幅度在多大程度上能够影响方法的准确性和鲁棒性还有待进一步研究, Gao 等提出的辨识方法是值得借鉴的[21].

4　结论

针对化工过程的非线性和时变特性, 研究简单通用的在线辨识方法具有重要意义. 本文提出将 JITL 和递推核学习辨识方法有效结合: 发挥 JITL

表 1 JR-LSSVR, R-LSSVR 和 LSSVR 辨识方法的结果比较

Table 1 Comparison of identification results with JR-LSSVR, R-LSSVR, and LSSVR methods

高斯噪声幅度	1 %	1 %	1 %	3 % ∼ 5 %	3 % ∼ 5 %
辨识方法	RMSE	RE (%)	MAE	RMSE	RE (%)
JR-LSSVR	**0.022**	**18.26**	**0.060**	**0.053**	**22.47**
R-LSSVR (增长的时间窗)	0.028	20.15	0.092	0.074	24.69
R-LSSVR ($N = 10$)	0.025	20.11	0.061	0.069	23.86
R-LSSVR ($N = 20$)	0.024	18.81	0.058	0.074	24.63
R-LSSVR ($N = 30$)	0.024	18.97	0.064	0.063	23.57
R-LSSVR ($N = 40$)	0.026	18.67	0.094	0.068	24.78
R-LSSVR ($N = 50$)	0.027	19.90	0.095	0.079	26.93
R-LSSVR ($N = 60$)	0.027	20.20	0.092	0.073	24.39
R-LSSVR ($N = 70$)	0.028	20.15	0.091	0.077	26.82
LSSVR	0.139	27.48	0.322	—	—

局部建模能力更好的优点; 且利用递推算法快速获得当前的模型, 以提高计算效率. 与传统 JITL 在每一时刻预测结束即抛弃当前模型不同, 提出保留当前的模型信息, 便于下一时刻更有效的递推. 通过一时变过程的仿真研究表明了其优越性, 可作为非线性时变过程在线辨识的一种新途径. 进一步需研究的包括从理论上分析噪声幅度在多大程度上能影响方法的准确性和鲁棒性.

References

1 Ljung L, Hjalmarsson H, Ohlsson H. Four encounters with system identification. *European Journal of Control*, 2011, **17**(5): 449−471

2 Himmelblau D M. Accounts of experiences in the application of artificial neural networks in chemical engineering. *Industrial and Engineering Chemistry Research*, 2008, **47**(16): 5782−5796

3 Wang L X. *Adaptive Fuzzy Systems and Control: Design and Stability Analysis*. New Jersey: Prentice-Hall, 1994.

4 Scholkopf B, Smola A J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2002.

5 Suykens J A K, van Gestel T, de Brabanter J, De Moor B, Vandewalle J. *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.

6 Rojo-Alvarez J L, Martinez-Ramon M, de Prado-Cumplido M, Artes-Rodriguez A, Figueiras-Vidal A R. Support vector method for robust ARMA system identification. *IEEE Transactions on Signal Processing*, 2004, **52**(1): 155−164

7 Toivonen H T, Totterman S, Akesson B. Identification of state-dependent parameter models with support vector regression. *International Journal of Control*, 2007, **80**(9): 1454−1470

8 Totterman S, Toivonen H T. Support vector method for identification of Wiener models. *Journal of Process Control*, 2009, **19**(7): 1174−1181

9 Li C H, Zhu X J, Cao G Y, Sui S, Hu M R. Identification of the Hammerstein model of a PEMFC stack based on least squares support vector machines. *Journal of Power Sources*, 2008, **175**(1): 303−316

10 Wang H, Pi D Y, Sun Y X. Online SVM regression algorithm-based adaptive inverse control. *Neurocomputing*, 2007, **70**(3): 952−959

11 Tang H S, Xue S T, Chen R, Sato T. Online weighted LS-SVM for hysteretic structural system identification. *Engineering Structures*, 2006, **28**(12): 1728−1735

12 Liu Y, Wang H Q, Yu J, Li P. Selective recursive kernel learning for online identification of nonlinear systems with NARX form. *Journal of Process Control*, 2010, **20**(2): 181−194

13 Cheng C, Chiu M S. A new data-based methodology for nonlinear process modeling. *Chemical Engineering Science*, 2004, **59**(13): 2801−2810

14 Pan T H, Li S Y, Cai W J. Lazy learning-based online identification and adaptive PID control: a case study for CSTR process. *Industrial and Engineering Chemistry Research*, 2007, **46**(2): 472−480

15 Liu Yi. Research on Kernel Learning Adaptive Modeling and Control for Industrial Batch Processes [Ph. D. dissertation], Zhejiang University, China, 2009
(刘毅. 间歇过程的核学习自适应建模与控制研究及工业应用 [博士学位论文]. 浙江大学, 中国, 2009)

16 Fujiwara K, Kano M, Hasebe S, Takinami A. Soft-sensor development using correlation-based just-in-time modeling. *AIChE Journal*, 2009, **55**(7): 1754−1765

17 Liu Y Q, Huang D P, Li Y. Development of interval soft sensors using enhanced just-in-time learning and inductive confidence predictor. *Industrial and Engineering Chemistry Research*, 2012, **51**(8): 3356−3367

18 Ge Z Q, Song Z H. A comparative study of just-in-time-learning based methods for online soft sensor modeling. *Chemometrics and Intelligent Laboratory Systems*, 2010, **104**(2): 306−317

19 Golub G H, van Loan C F. *Matrix Computations*. Baltimore: The John Hopkins University Press, 1996.

20 Nikravesh M, Farell A E, Stanford T G. Control of non-isothermal CSTR with time varying parameters via dynamic neural network control (DNNC). *Chemical Engineering Journal*, 2000, **76**(1): 1−16

21 Gao Z W, Dai X W, Breikin T, Wang H. Novel parameter identification by using a high-gain observer with application to a gas turbine engine. *IEEE Transactions on Industrial Informatics*, 2008, **4**(4): 271−279



刘 毅 浙江工业大学副教授. 2009 年获浙江大学博士学位. 主要研究方向为过程系统工程, 包括统计学习方法及其在工业过程建模和控制的应用. 本文通信作者. E-mail: yliuzju@zjut.edu.cn

(**LIU Yi** Associate professor at Zhejiang University of Technology. He received his Ph. D. degree from Zhejiang University in 2009. His research interest covers process systems engineering, including statistical learning methods with applications to industrial process modeling and control. Corresponding author of this paper.)



金福江 华侨大学教授. 2002 年获得浙江大学博士学位. 主要研究方向为流程工业过程建模, 控制和优化.
E-mail: jinfujiang@163.com

(**JIN Fu-Jiang** Professor at Huaqiao University. He received his Ph. D. degree form Zhejiang University in 2002. His research interest covers industrial process modeling, control and optimization.)



高增梁 浙江工业大学教授. 1997 年获得华东理工大学博士学位. 主要研究方向为流程工业过程装备与控制.
E-mail: zlgao@zjut.edu.cn

(**GAO Zeng-Liang** Professor at Zhejiang University of Technology. He received his Ph. D. degree from East China University of Science and Technology in 1997. His research interest covers process equipment and control engineering in industrial processes.)