

研究论文

一种基于增量式 SVR 学习的在线 自适应建模方法

王 平^{1,2}, 田华阁¹, 田学民¹, 黄德先²

(¹ 中国石油大学(华东)信息与控制工程学院, 山东 东营 257061; ² 清华大学自动化系, 北京 100084)

摘要: 训练样本的数量与质量对于过程建模至关重要, 在很大程度上影响所建模型的质量。基于增量式支持向量回归(SVR)学习算法, 提出一种在线自适应建模方法以实现有选择地添加和删除训练样本。该方法利用 SVR 模型的 KKT 条件选择出那些包含足够多新信息的样本进行增量学习, 能够在保证模型泛化能力的同时降低模型更新频率。另外, 为快速准确地跟踪过程特性的变化, 将通过评价当前模型对新增训练样本的学习能力来决定是否需要删除旧样本。当需要删除样本时, 基于样本间的相似度, 选择淘汰与当前过程特性差别最大的旧样本。将该方法用于建立工业聚丙烯熔融指数预报模型, 结果表明, 与其他方法相比, 获得的预测模型具有更好的泛化性能, 且模型更新频率明显降低, 能有效地适应工况的变化。

关键词: 支持向量回归; 在线建模; 样本相似度; 熔融指数

中图分类号: TP 301.6

文献标识码: A

文章编号: 0438-1157(2010)08-2040-06

A new approach for online adaptive modeling using incremental support vector regression

WANG Ping^{1,2}, TIAN Huage¹, TIAN Xuemin¹, HUANG Dexian²

(¹ College of Information and Control Engineering, China University of Petroleum, Dongying 257061, Shandong, China; ² Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract: Considering the performance of a predictive model is heavily depended on its training samples, a new on-line adaptive modeling approach based on incremental support vector regression (SVR) is presented. When a new sample arrives, it is firstly checked by the Karush-Kuhn-Tucker (KKT) condition of established model, only those which contain sufficient new information can be introduced into the training sample set. In this way, the model generalization ability will be maintained while the update frequency can be reduced. If the new sample cannot be described by the established model and, therefore, has a large prediction error, the model must be updated and the useless sample should be deleted from the model, to adapt the process characteristics. In this case, the useless sample, while not the oldest one, is selectively deleted from the model based on the similarity between samples. The proposed method is illustrated through the application to an industrial polypropylene unit to predict its melt index. The results show that, compared with other methods, the proposed method exhibits good generalization ability while the update frequency significantly lower, and therefore trace the process characteristics effectively.

Key words: support vector regression; on-line modeling; sample similarity; melt index

2010-05-05 收到初稿, 2010-05-12 收到修改稿。

联系人: 田学民。第一作者: 王平(1983—), 男, 博士研究生。

基金项目: 国家高技术研究发展计划项目(2007AA04Z193);
中国石油大学(华东)研究生创新基金项目(Z10-09)。

Received date: 2010-05-05.

Corresponding author: TIAN Xuemin, tianxm@hdpu.edu.cn

Foundation item: supported by the High-tech Research and Development Program of China (2007AA04Z193).

引言

工业过程中存在着许多无法或难以直接测量的变量,这些变量通常与产品质量密切相关,是必须进行监控的参数。软测量技术是解决该问题的一种有效方法,目前已在实际工业中得到广泛应用^[1]。

近年来,基于数据驱动的软测量建模方法得到了人们的普遍关注^[2]。其中,支持向量回归(support vector regression, SVR)基于结构风险最小化原则,较好地解决了小样本、非线性和局部极小等问题,已被广泛应用于多个研究领域^[1]。目前的 SVR 学习算法大多是离线、批处理式的,而工业过程一般具有动态时变的特点,过程特性和工作点都可能发生变化。这样离线训练得到的模型不能保证长期适用,因此对 SVR 在线学习方法的研究具有重要的理论意义和使用价值。文献[3]研究了增加和删减训练样本对 SVR 模型的影响,在此基础上,文献[4]提出一种精确 SVR 在线学习算法,它以 Karush-Kuhn-Tucker (KKT) 条件为基础,当有样本加入或删除时,通过修改拉格朗日乘子和偏置项来更新 SVR 模型,避免了重复求解二次规划问题。为进一步加快训练速度,文献[5]提出一种基于核函数缓存的 SVR 在线训练方法,以牺牲计算内存为代价,加快 SVR 在线学习速度。文献[6]则提出用分块矩阵的方法来解决 SVR 在线更新算法复杂的问题。

对于一个回归模型而言,其性能在很大程度上取决于所使用的训练样本,用于训练的样本越具有代表性,模型质量越高。鉴于此,结合增量式 SVR 学习算法的特点,提出一种在线自适应建模方法。该方法基于 SVR 模型的 KKT 条件,实现对新增样本的区分学习,降低了模型更新频率;另外,基于样本相似度的选择性淘汰算法,删除与新增训练样本相似度最小的旧样本,获得能够反映过程当前工况的训练样本集,由此提高模型跟踪过程特性变化的能力。最后以工业聚丙烯熔融指数预报模型为例验证了其有效性。

1 增量式 SVR 学习算法

给定样本集 $T = \{\mathbf{x}_i, y_i\}, i=1, 2, \dots, l, \mathbf{x}_i \in R^N$ 和 $y_i \in R$, SVR 利用非线性映射 $\phi(\cdot)$ 将数据 $\mathbf{x}_i$ 映射到高维特征空间 F , 然后构造函数 $f(\cdot)$ 拟合样本集

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b, \phi: R^N \rightarrow F, \mathbf{w} \in F \quad (1)$$

其中, $\mathbf{w}$ 和 b 可以通过求解下述二次规划问题得到

$$\begin{aligned} \min_{\mathbf{w}, b} & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s. t.} & \begin{cases} y_i - (\mathbf{w}^T \cdot \phi(\mathbf{x}_i) + b) \leq \epsilon + \xi_i \\ (\mathbf{w}^T \cdot \phi(\mathbf{x}_i) + b) - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, l \end{cases} \end{aligned} \quad (2)$$

式中 C 为平衡模型复杂度和训练误差的权系数, $\epsilon > 0$ 为容许误差。由拉格朗日乘子法得到式(2)的对偶优化问题

$$\begin{aligned} \min_{a, a^*} D = & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l Q_{ij} (a_i - a_i^*) (a_j - a_j^*) + \\ & \sum_{i=1}^l \epsilon (a_i + a_i^*) - \sum_{i=1}^l y_i (a_i - a_i^*) \end{aligned} \quad (3)$$

$$\text{s. t. } 0 \leq a_i, a_i^* \leq C, \sum_{i=1}^l (a_i - a_i^*) = 0, i = 1, \dots, l$$

式中 $Q_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$ 为核函数, a_i, a_i^* 为对偶参数。

定义样本系数偏差 $\theta_i = a_i - a_i^*$ 和边界函数 $h(\mathbf{x}_i) = f(\mathbf{x}_i) - y_i$ 。则根据系数偏差值和边界函数值,基于优化问题(3)的 KKT 条件可将样本集 T 分为如图 1 所示的 3 个子集^[4,7]

$$\begin{cases} E = \{i \mid |\theta_i| = C, |h(\mathbf{x}_i)| > \epsilon\} \\ S = \{i \mid 0 < |\theta_i| < C, |h(\mathbf{x}_i)| = \epsilon\} \\ R = \{i \mid \theta_i = 0, |h(\mathbf{x}_i)| < \epsilon\} \end{cases} \quad (4)$$

其中, E 为误差支持向量集, S 为边界支持向量集, R 为剩余样本集。增量式 SVR 学习算法包括递增和递减两种算法。递增算法的基本思想是:在有限时间步内,逐步改变新样本系数偏差,直到其满足 KKT 条件,同时保证在各更新时间步内,样本集中其他样本能够继续满足 KKT 条件。随着样本个数增多,可能会面临“维数灾难”使增量学习无法进行下去。因此需要利用递减算法丢弃一些训练数据,并调整剩余训练样本继续满足 KKT 条件^[7]。增量式学习算法对支持向量和非支持向量进行了区别对待和处理,且用简单的迭代修正代替复杂的优化求解,使得算法具有较高的计算效率^[7]。

2 在线自适应建模算法

实际生产过程中,离线建立的模型往往不能准确描述系统当前状况,有必要利用在线新数据不断更新模型^[8-9]。增量式 SVR 学习算法可实现对训练样本的在线实时添加与删除。然而,为进一步提高建模精度和降低学习时间,还有一系列问题需要解决。下面分别对这些问题进行研究分析,并提出相

应的解决方案。

2.1 基于 KKT 条件检验的新增样本区分学习

若不加区分地对每一个新增样本都进行增量学习,势必造成模型频繁更新。这样,一方面浪费了大量计算资源;另一方面,在正常生产时,各变量一般都保持在设定点周围,新样本反映的都是在工作点附近的信息,将扩大过程干扰、测量噪声对建模的影响,导致模型泛化能力降低。

定理 1 设 $f(\mathbf{x})$ 为 SVR 回归函数, $\{\mathbf{x}_{\text{new}}, y_{\text{new}}\}$ 为新增样本。则满足 KKT 条件的新增样本不会改变支持向量集,违背 KKT 条件的新增样本将使支持向量集变化,且新增训练样本违反 KKT 条件的充要条件为新增样本位于 $|h(\mathbf{x}_{\text{new}})| > \varepsilon$ 区域中。

定理 1 的证明可参考文献 [4, 10]。由该定理知,若新样本 $\{\mathbf{x}_{\text{new}}, y_{\text{new}}\}$ 位于图 1 $|h(\mathbf{x}_{\text{new}})| \leq \varepsilon$ 区域内,则其满足 KKT 条件,说明此时的 SVR 模型能很好地描述新增样本,为避免样本重复训练,新样本将不被纳入到训练集中。若新样本 $\{\mathbf{x}_{\text{new}}, y_{\text{new}}\}$ 位于图 1 中 $|h(\mathbf{x}_{\text{new}})| > \varepsilon$ 区域内,则其违反了现有模型的 KKT 条件,这些样本包含新的信息,需要对其学习以反映系统当前状态。最终,通过检验样本是否满足现有模型的 KKT 条件,实现对样本区分学习。这就减少模型更新频率,且保留已有模型包含的历史信息,维持了模型的泛化能力。

2.2 基于相似度的样本选择性淘汰方法

在线建模时,为提高在线训练速度及更好地反映过程当前特性,冗余样本必须被删除。最常用的是滑动时间窗法^[1-2],其一般删除最古老的样本,然而,这种做法不尽合理,最古老的样本未必不能反映系统当前特性。另一种方法是,对样本进行不同程度的加权^[11-12]以体现各样本对模型的不同贡

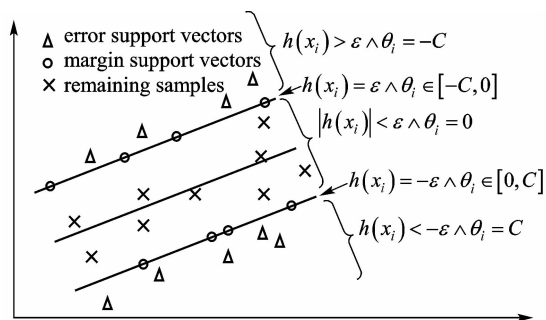


图 1 训练样本集的划分

Fig. 1 Three subsets gotten from training samples set

献。但给各样本分配一个合理的权值是很困难的。

针对该问题,提出一种基于相似度的样本选择性淘汰方法。通常是以欧氏距离计算相似度的,但仅以距离作为衡量尺度并不能充分挖掘样本的信息。为更全面地反映出样本间的差异性,将采用基于距离和角度信息^[13-14]作为样本删除准则。定义新增样本与训练样本集中样本间的相似度 (s_i)

$$s_i = \gamma \exp(-d(\mathbf{x}_i, \mathbf{x}_{\text{new}})) + (1 - \gamma) \cos \beta(\mathbf{x}_i, \mathbf{x}_{\text{new}}) \quad (5)$$

$$d(\mathbf{x}_i, \mathbf{x}_{\text{new}}) = \sqrt{\|\mathbf{x}_i - \mathbf{x}_{\text{new}}\|_2} = \sqrt{\sum_{j=1}^N (\mathbf{x}_{i,j} - \mathbf{x}_{\text{new},j})^2} \quad (6)$$

$$\beta(\mathbf{x}_i, \mathbf{x}_{\text{new}}) = \arccos\left(\frac{\mathbf{x}_i^T \mathbf{x}_{\text{new}}}{\|\mathbf{x}_i\|_2 \cdot \|\mathbf{x}_{\text{new}}\|_2}\right) \quad (7)$$

式(6)是距离计算式,式(7)是角度计算式, $0 \leq \gamma \leq 1$ 为加权系数。

以二维空间中的 3 个向量为例,说明计算样本相似度时考虑角度信息的重要性。如图 2 所示,此时有 $d_i = d(\mathbf{x}_i, \mathbf{x}_{\text{new}}) = d(\mathbf{x}_k, \mathbf{x}_{\text{new}}) = d_k$, 若只以距离来定义相似度,则 $\mathbf{x}_i$ 、 $\mathbf{x}_k$ 和新样本具有相同的相似度,无法判断哪个样本需要删除。若考虑角度信息,则由 $\beta_1 < \beta_2$ 知,与 $\mathbf{x}_k$ 相比, $\mathbf{x}_i$ 更靠近新样本,更能反映新样本的特性,此时只可能选择删除样本 $\mathbf{x}_k$ 。因此,综合考虑样本间的距离和角度信息可以得到更准确全面的删除准则。样本删减操作时,首先将训练样本集中的样本按照相似度进行排序,然后删除与当前操作状态相似度最小的旧样本。

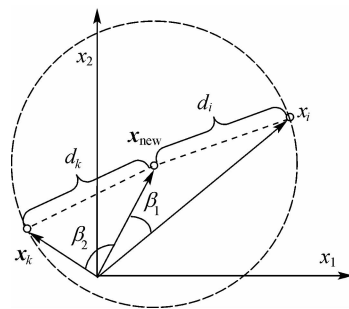


图 2 基于距离和角度信息的样本相似度计算

Fig. 2 Illustration of importance of angle measure

2.3 基于模型性能评价的自适应建模策略

若训练样本过多,模型可能过度拟合,精度下降,且影响在线学习速度;若样本数过少,则不能确切地表达出系统动态特性。通常是利用留一法交叉验证^[10, 15-16]、正交最小二乘算法^[17]等方法来确定训练样本数。但这些方法或者计算量大,不适合在线使用;或者依赖特殊模型结构形式,难以推广。

近年来，基于模型性能评价的样本选择方法得到人们的普遍关注^[9]，它具有在线计算量小、不依赖于具体模型形式等优点。基于增量式 SVR 学习算法的特点，提出一种简单有效的样本个数选择方法：增量学习后，检验新增样本是否进入训练样本集中的误差支持向量集中，即检验 $|\bar{h}(\mathbf{x}_{\text{new}})| > \varepsilon$ 是否成立，若成立，表明当前模型并不能很好地描述新样本，修正后的模型对新样本仍有较大的预测误差。此时应删除旧样本以降低对新样本的预测误差。与 2.1 节中的边界函数 $h(\cdot)$ 不同，此时 $\bar{h}(\cdot)$ 是增量学习新样本后的边界函数。

最终得到一种基于增量式 SVR 学习算法的在线自适应建模方法，其流程如图 3 所示。在线获得新的训练样本后，首先，检测新样本是否满足现有模型的 KKT 条件，若不满足，则进行递增学习新样本信息；然后，进行模型预测性能评价和最大样本数检验以确定模型中是否包含冗余样本，若存在冗余样本，则计算样本的相似度并删除与新增样本相似度最小的旧样本。

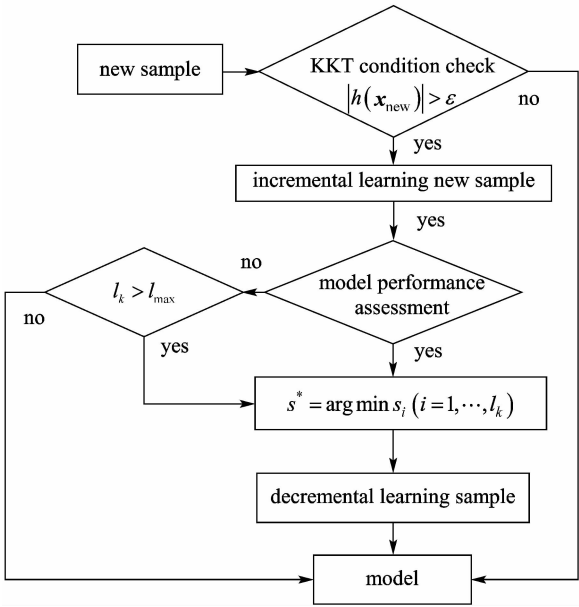


图 3 在线自适应建模流程图
Fig. 3 Online adaptive modeling flowsheet

3 应用分析

熔融指数 (melt index, MI) 是衡量聚丙烯产品质量的关键指标之一。目前 MI 是通过离线化验分析获得，难以满足在线监控的要求。从某公司聚丙烯装置 DCS 系统采集过程数据，作为模型的输入变量，同时从化验车间取 MI 化验值作为输出变

量。经一系列数据预处理后获得 444 组数据，前 244 组用于离线训练使用，剩余数据用于模型验证。

分别用偏最小二乘法 (PLS)、SVR (SVR-I)、带滑动窗口的增量式 SVR (SVR-II, $l_{\text{max}} = 50$) 及本文方法 (SVR-III, $l_{\text{max}} = 50$, $\gamma = 0.95$) 建立 MI 预报模型。SVR 模型均使用 RBF 核函数，且参数设置相同 ($\sigma = 0.6$, $C = 20$, $\varepsilon = 0.05$)。为比较各方法的预报精度和速度，以均方根误差 (RMSE)、相对均方差误差 (RRMSE) 和最大绝对误差 (MAE) 作为模型精度指标；以 CPU 时间 (time/s, Matlab 7.3, 2.1GHz, 1G RAM)、样本学习次数 (No.1)、样本删除次数 (No.2) 作为建模速度指标，结果列于表 1。

表 1 四种建模方法的性能比较
Table 1 Performance of four modeling methods

Model	RMSE	RRMSE	MAE	Time/s	No. 1	No. 2
PLS	0.4786	0.1039	1.6822	×	×	×
SVR-I	0.425	0.0951	1.9088	×	×	×
SVR-II	0.3728	0.085	1.2899	47.06	200	195
SVR-III	0.3389	0.0749	1.1307	28.17	32	29

由表 1 可知，与 PLS、SVR-I 相比，SVR-II 和 SVR-III 的预报精度更高。这是因为聚丙烯过程是一个复杂的非线性过程，生产不同牌号产品时，变量间的对应关系可能会发生变化，使用固定的模型不能保证准确地预报 MI 的变化，而利用在线数据不断更新模型可以有效地跟踪这些变化，进而提高模型预报精度。另外，与 SVR-II 相比，SVR-III 在预报精度和模型更新速度上都具有优势。SVR-III 对训练样本有选择性地增删，在保证模型预报精度的同时，降低了模型更新次数 (由 200 次降为 32 次)，其对 MI 的预报结果如图 4 所示。

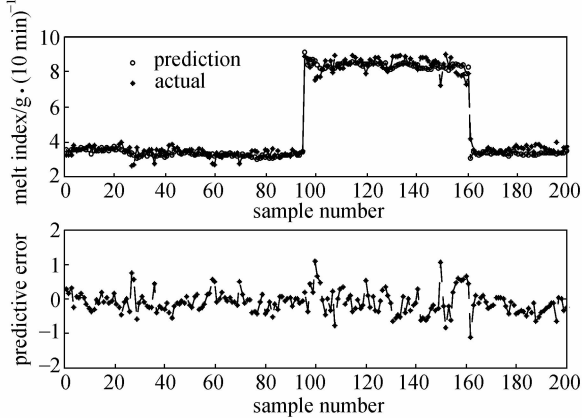


图 4 基于 SVR-III 的 MI 在线自适应预报
Fig. 4 Online adaptive prediction of MI by SVR-III

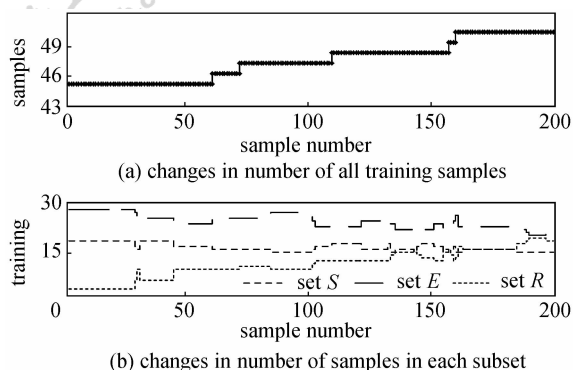


图 5 训练样本数量的变化

Fig. 5 Changes in the number of training samples

为充分利用历史训练结果及提高在线学习速度, 离线训练 244 组数据得到 SVR 模型 (即 SVR-I) 后, 删除剩余样本集并以此初始化 SVR-III 模型, 该初始模型包含 45 个训练样本。选择最大建模样本数 $l_{\max} = 50$, 则使用 SVR-III 在线建模时, 模型的训练样本个数及 3 个子集中样本个数的变化情况如图 5 所示。由图 5(a) 知, 与固定的滑动窗口建模方法不同, SVR-III 的样本数不是固定不变的, 而是自适应变化的。此外, 当最大建模样本数限制在 $l_{\max} = 60$ 时, 总训练样本数增至 52 个后就不再变化。由图 5(b) 知, 3 个子集中样本个数的变化情况也反映出 SVR-III 模型是随着工况变化而变化的。

表 2 比较了不同相似度加权系数 γ 对模型预报精度的影响。以 RMSE 指标为例, 可以看出, γ 与模型预报精度不是简单的线性关系, 随着 γ 逐步增大, RMSE 先降后升, 一般需要采用交叉验证的方法来确定合适的 γ 值。另外, 可以看出, 与只以距离来衡量样本相似度 (即 $\gamma = 0$) 相比, 综合考虑距离和角度信息会得到更高预测精度的模型。

表 2 不同相似度加权系数对预测精度的影响

Table 2 Prediction accuracy for various values of weight coefficient

Item	$\gamma = 0$	$\gamma = 0.9$	$\gamma = 0.95$	$\gamma = 0.98$
RMSE	0.3485	0.347	0.3389	0.3563
RRMSE	0.0798	0.0779	0.0749	0.0774
MAE	1.1691	1.42	1.1307	1.4631

4 结 论

过程工业在正常生产时, 各操作参数都被尽量保持在设定值周围, 样本所包含的信息都是在工作

点附近的; 同时, 为满足不同的生产要求, 需要改变操作参数的目标值以满足工况变迁要求, 这导致样本呈现按工作点聚类 and 迁移的特性。针对这些特性, 提出一种基于增量式 SVR 的在线自适应建模方法。通过对训练样本有选择地添加或删除, 实现模型在线更新, 以有效地跟踪过程工况的变化。工业应用实例验证了该方法的可行性和有效性。

References

- [1] Fortuna L, Graziani S, Rizzo A, Xibilia M G. Soft Sensors for Monitoring and Control of Industrial Processes. Berlin: Springer-Verlag, 2007: 1-10
- [2] Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry. *Computers & Chemical Engineering*, 2009, **33** (4): 795-814
- [3] Cauwenberghs G, Poggio T. Incremental and decremental support vector machine learning. *Advances in Neural Information Processing Systems*, 2001, **44** (13): 409-415
- [4] Ma J, Theiler J, Perkins S. Accurate online support vector regression. *Neural Computation*, 2003, **15** (11): 2683-2704
- [5] Wang Hui (汪辉), Pi Daoying (皮道映), Sun Youxian (孙优贤). On-line support vector machine training algorithm and its application. *Journal of Zhejiang University: Engineering Science* (浙江大学学报: 工学版), 2004, **38** (12): 1642-1645, 1649
- [6] Wang Lingyun (王凌云), Gui Weihua (桂卫华), Liu Meihua (刘梅花), Yang Chunhua (阳春华). Prediction model of ion concentration based on improved online support vector regression. *Control and Decision* (控制与决策), 2009, **24** (4): 537-541
- [7] Laskov P, Gehl C, Kruger S, Muller K. Incremental support vector learning: analysis, implementation and application. *Journal of Machine Learning Research*, 2006, **7** (9): 1909-1936
- [8] Alexandridis A, Sarimveis H, Bafas G. A new algorithm for online structure and parameter adaptation of RBF networks. *Neural Networks*, 2003, **16** (7): 1003-1017
- [9] Liu Y, Hu N, Wang H, Li P. Soft chemical analyzer development using adaptive least-squares support vector regression with selective pruning and variable moving window size. *Ind. Eng. Chem. Res.*, 2009, **48** (12): 5731-5741
- [10] Zhou Weida (周伟达), Zhang Li (张莉), Jiao Licheng (焦李成). An analysis of SVMs generalization performance. *Acta Electronica Sinica* (电子学报), 2001, **29** (5): 590-594
- [11] Suykens J, Brabanter J, Lukas L, Vandewalle J. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing*, 2002, **48** (1-4): 85-105

- [12] Cui W, Yan X. Adaptive weighted least square support vector machine regression integrated with outlier detection and its application in QSAR. *Chemometrics and Intelligent Laboratory Systems*, 2009, **98** (2): 130-135
- [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. Generalised predictive control for nonlinear process systems based on lazy learning. *International Journal of Modelling, Identification and Control*, 2006, **1** (3): 230-238
- [15] Cawley G, Talbot N. Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. *Neural Networks*, 2004, **17** (10): 1467-1475
- [16] Liu Yi (刘毅), Wang Haiqing (王海青), Li Ping (李平). Adaptive local learning based least squares support vector regression with application to online modeling for fermentation processes. *Journal of Chemical Industry and Engineering (China)* (化工学报), 2008, **59** (8): 2052-2057
- [17] Chen S, Hong X, Luk B, Harris C. Orthogonal-least-squares regression: a unified approach for data modeling. *Neurocomputing*, 2009, **72** (10-12): 2670-2681