

文章编号: 1001-0920(2012)12-1800-05

基于灰关联分析的KFCM-LSSVM蒸发过程软测量

钱晓山^{1,2}, 阳春华¹

(1. 中南大学 信息科学与工程学院, 长沙 410083; 2. 宜春学院 物理科学与工程技术学院, 江西 宜春 336000)

摘要: 针对氧化铝蒸发过程出口浓度在线检测问题, 提出一种基于灰色关联分析和核模糊聚类相结合的多支持向量机蒸发过程建模方法. 该方法采用灰色关联分析提取影响出口浓度较大的变量, 再用核模糊聚类算法将样本划分成不同的子类, 并对各子类构建的支持向量机模型的输出综合加权得到最终输出. 对蒸发过程生产数据进行实验, 并与核模糊聚类最小二乘支持向量机(KFCM-LSSVM)相比较, 结果表明新模型鲁棒性较强、精度较高、泛化性更好, 可以用于蒸发过程的优化控制.

关键词: 灰关联分析; 蒸发过程; 核模糊聚类; 最小二乘支持向量机; 软测量

中图分类号: TP273

文献标志码: A

Soft sensor of based on grey correlation analysis and KFCM-LSSVM in evaporation process

QIAN Xiao-shan^{1,2}, YANG Chun-hua¹

(1. School of Information Science and Engineering, Central South University, Changsha 410083, China; 2. Physical Science and Technology College, Yichun University, Yichun 336000, China. Correspondent: QIAN Xiao-shan, E-mail: qianxiaoshan@126.com)

Abstract: Aiming at the online testing problem of outlet concentration in the evaporation process of alumina production, a multi-support vector machine modeling method based on the gray relational analysis and kernel fuzzy *c*-means clustering in the evaporation process is proposed. Variable factors of greater correlation with outlet concentration are selected by using grey relational analysis, these training samples constituted by these factors are divided into subsets with different cluster center by using kernel fuzzy *c*-means clustering, and output is obtained finally by summing each support vector machines model's output comprehensive weighting. The experimental results of industrial production data of evaporation process show that, compared with kernel fuzzy cluster method and least squares support vector machine(KFCM-LSSVM) model, the proposed model can get good robustness, high accuracy, and better generalization, which can be used for the optimal control of the evaporation process.

Key words: grey relational analysis; evaporation process; kernel fuzzy *c*-means clustering; least squares support vector machine; soft sensor

1 引言

氧化铝是电解铝生产的主要原料, 蒸发过程是氧化铝生产的关键工艺过程, 主要是对蒸发母液进行浓缩, 使其苛性碱的浓度达到规定的指标的同时能耗最低. 蒸发母液分别进入VI效及IV效, VI效的出料进入V效, V效的出料经过料泵与III效出料混合后送到调配槽. 产生蒸汽的通入量主要根据出料浓度的化验结果和人工经验设定. 作为该过程控制的重要参数(出口铝酸钠溶液浓度)的检测采用人工化验的方

式, 由于生产流程长, 化验分析的结果滞后严重, 导致蒸汽通入量不准确, 不能满足实时在线控制的要求, 也导致了出口浓度难以达标; 与此同时, 蒸汽的消耗量也较大. 研究一种出口浓度在线检测的方法, 对提高氧化铝生产质量和节能降耗具有重要意义.

蒸发过程具有高度的非线性、流程长、高维、变量多等特点, 难以建立精确的数学模型. 软测量为解决这一问题提供了可行的解决途径^[1]. 传统的软测量方法如多元回归以及时间序列分析等方法很难取

收稿日期: 2011-04-04; 修回日期: 2011-11-11.

基金项目: 国家自然科学基金项目(60874069); 国家863计划项目(2009AA04Z124, 2009AA04Z137); 国家杰出青年科学基金项目(61025015).

作者简介: 钱晓山(1980-), 男, 讲师, 博士生, 从事复杂工业过程建模、优化与控制的研究; 阳春华(1965-), 女, 教授, 博士生导师, 从事复杂工业过程建模、优化与控制等研究.

得理想的预测效果. 而神经网络虽具有强大的处理非线性数据的能力, 在复杂工业过程的建模与优化方面获得了广泛的应用, 但其神经网络结构的确定须依赖人为经验, 并且容易出现过学习问题. 文献[2]利用灰关联分析结合神经网络构建的粮食产量预测模型能够全面地筛选输入变量, 而不是凭主观经验判断, 增强了 BP 网络的适应能力且有较好的预测精度和稳定性. 文献[3]根据几个模型相加可提高模型的预测精度及鲁棒性的思想, 提出了一种基于模糊聚类法 (FCM) 的多模型结构方法. 文献[4]将多最小二乘支持向量机模型应用到草酸钴粒度软测量中, 得到了比单一最小二乘支持向量机模型更高的预测精度.

本文基于以上研究基础, 结合该氧化铝生产蒸发过程的特点, 提出一种灰色关联分析和核模糊聚类的多最小二乘支持向量机的软测量建模方法. 该方法首先采用灰色关联分析, 全面筛选出与出口浓度强相关的输入变量, 减少建模数据的维数. 另外, 考虑到支持向量机的泛化能力受到样本数据分布特别是孤立点或者不完整数据的影响较大这一问题, 利用模糊核聚类对新的样本进行聚类, 增强其鲁棒性, 提高支持向量机的泛化能力. 对于每个子类构建最小二乘支持向量机模型, 利用综合加权得到最终的输出, 建立出口浓度预测模型. 为了克服参数误差对支持向量机 (SVM) 建模准确性的影响, 本文利用混沌粒子群算法来优化 SVM 参数, 最终得到了出口浓度预测值.

2 灰色关联分析

灰关联分析 (GRA)^[5-6]是一种相对性的排序分析, 目的是定量地度量某个系统内部各因素之间的关联程度, 从而为确定影响铝酸钠溶液浓度各因素的重要性提供了理论基础. GRA 的基本思想是根据序列在空间中的几何形状的相似程度来判断关联程度^[7-8], 具体计算步骤如下:

1) 确定反映系统行为特征的输出序列和影响系统行为的输入序列并进行标准化处理.

2) 计算关联系数. 对于 $t = n$ 时刻的输入序列 $\{X_i(n)\}$ 和输出序列 $\{X_o(n)\}$ 的关联系数为

$$\xi_{io}(n) = \frac{\Delta_{\max} + \Delta_{\min}}{\Delta_{io}(n) + \lambda \Delta_{\max}}. \quad (1)$$

其中: λ 为分辨系数, $0 < \lambda < 1$; Δ_{\min} 为第 2 级最小差, Δ_{\max} 为两级最大差; $\Delta_{io}(n)$ 为各输入数列 $\{X_i(n)\}$ 曲线上的每一个点与输出数列 $\{X_o(n)\}$ 曲线上的每一个点的绝对差值.

3) 求关联度 r_{io} . 输入序列和输出序列的关联程度数量表示如下:

$$r_{io} = \frac{1}{L} \sum_{n=1}^L \xi_{io}(n). \quad (2)$$

4) 关联度排序. 按照计算出的关联度从大到小进行排序.

3 模糊核聚类原理

聚类是将一组给定的未知类标号的样本分成内在的多个类别, 使得同一类中的样本具有较高的相似度, 而不同类中的样本差别较大. 聚类分析的目的是揭示和刻画数据的内在结构. 模糊 c -均值聚类算法将模糊数学引入聚类分析中, 更具合理性, 是使用最广泛的聚类技术之一^[9]. 引入核函数至 FCM 中, 构成核模糊聚类分析, 使之在原输入空间中诱导出类核依赖的新的距离度量, 也是将 FCM 在欧氏距离下的执行推广到同一空间中不同距离度量的新的聚类, 且对不完整或缺失数据、包含噪声和野值数据的聚类更有鲁棒性.

定义模糊核聚类算法的目标函数为^[10]

$$J_m(u, v) = \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m \|\phi(x_j) - \phi(v_i)\|^2. \quad (3)$$

其中: x_j 为样本集, v_i 为聚类中心, u_{ij} 应该满足以下 3 个条件:

$$\begin{cases} \sum_{i=1}^c u_{ij} = 1, 1 \leq j \leq N; \\ 0 \leq u_{ij} \leq 1, 1 \leq i \leq c, 1 \leq j \leq N; \\ 0 < \sum_{j=1}^N u_{ij} < N, 1 \leq i \leq c. \end{cases} \quad (4)$$

定义核函数 $K(x, y)$, 满足 $K(x, y) = \phi(x)^T \phi(y)$, 将式 (3) 展开并代入核函数, 在式 (4) 的约束下优化得到

$$u_{ij} = \frac{\left(\frac{1}{K(x_k, x_k) + K(v_i, v_i) - 2K(x_k, v_i)} \right)^{1/(m-1)}}{\sum_{j=1}^N \left(\frac{1}{K(x_k, x_k) + K(v_j, v_j) - 2K(x_k, v_j)} \right)^{1/(m-1)}}, \quad (5)$$

$$v_i = \frac{\sum_{j=1}^N (u_{ij})^m \tilde{K}(x_k, v_i) x_k}{\sum_{j=1}^N (u_{ij})^m \tilde{K}(x_k, v_j)}. \quad (6)$$

4 最小二乘支持向量机

支持向量机回归方法的根本思想是, 先通过某种非线性变换将输入向量映射到高维特征空间; 然后在高维特征空间中构造线性最优决策函数, 使模型结构风险最小.

给定训练样本 $(x_i, y_i) \in R^m \times R, i = 1, 2, \dots, N$, N 为样本总数, m 为样本空间的维数. 通过非线性映射 $z_i = \phi(x_i)$, 将 m 维输入空间映射到 $d(d \geq m)$ 维

特征空间,在这个高维特征空间中采用线性函数 $f(x) = \omega\phi(x) + b$ 对其进行拟合,并容许出现拟合误差.目标是使回归模型在模型推广能力和经验风险之间找到最佳平衡点,即结构风险最小.

在 LSSVM 中确定最优超平面需要求解如下二次规划问题:

$$\begin{cases} \min J(\omega, \varepsilon) = \frac{1}{2}\omega^T\omega + \frac{\zeta}{2} \sum_{i=1}^N \varepsilon_i^2; \\ \text{s.t. } y_i = \omega\phi(x) + b + \varepsilon_i, i = 1, 2, \dots, N. \end{cases} \quad (7)$$

其中: $J(\omega, \varepsilon)$ 为结构风险, ζ 为经验风险调整因子, ε_i 为容许误差, $\omega^T\omega$ 控制着模型的推广能力.

利用 Lagrange 法求解式 (7) 的优化问题,根据 KKT 优化条件整理得到如下的线性方程组:

$$\begin{bmatrix} 0 & 1_N^T \\ 1_N & \theta + \frac{1}{\zeta} I_N \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}. \quad (8)$$

其中: $y = [y_1, \dots, y_N]^T$, $1_N = [1, \dots, 1]$, I_N 为单位阵, $\alpha = [\alpha_1, \dots, \alpha_N]^T$, $\theta_{i,j} = \phi(x_i)^T \phi(x_j)$, $i, j = 1, 2, \dots, N$.

根据 Mercer 条件定义如下的核函数:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j), \quad (9)$$

可以求解得到最小二乘支持向量机回归模型为

$$f(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b. \quad (10)$$

其中: N 为输入样本的个数, x_i 为第 i 个输入样本, x 为某一个输入变量.选择不同的核函数能够建立不同的 LSSVM 模型,此时模型的性能也不同,本文选用径向基核函数

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right), \quad (11)$$

其中 σ 为径向基宽度.

5 过程建模与验证分析

5.1 输入样本的灰关联分析

蒸发过程是一个能量传递的过程,在蒸发器内加热蒸汽释放潜热,转移到料液中,使溶剂发生相变,溶液浓度得以提高.大多数多效蒸发工艺的操作是以最终产品的浓度作为主要控制目标.以某氧化铝厂采用带闪蒸和强制循环的七效逆流降膜蒸发的过程为例.由于最终产品浓度软测量模型的影响因素较多,仅凭现场操作工的经验定性分析和机理分析的结果来确定主要影响因素,可信度不高,并且缺乏理论依据.因此,在机理分析和现场操作人员的经验定性分析的基础上,采用灰色关联分析方法^[11],从定量分析的角度确定各因素对软测量模型的影响程度.分析得到的选择变量有:进料温度 A_1 , 进料流量 A_2 , 进料浓度 A_3 , 新蒸汽温度 A_4 , 新蒸汽流量 A_5 ; I 效

至 VI 效的液温分别为 $A_6, A_7, A_8, A_9, A_{10}, A_{11}$; I 效至 VI 效的气压分别为 $A_{12}, A_{13}, A_{14}, A_{15}, A_{16}, A_{17}$; 出料浓度为 A_0 . 以 2009-01-04 的生产数据为训练样本和测试样本.为消除量纲的影响,采用均值法对样本进行处理,然后进行灰关联分析,得到结果为

$$\begin{aligned} G(0, 1) &= 0.9276, G(0, 2) = 0.8534, \\ G(0, 3) &= 0.9448, G(0, 4) = 0.9102, \\ G(0, 5) &= 0.8176, G(0, 6) = 0.7814, \\ G(0, 7) &= 0.9123, G(0, 8) = 0.7618, \\ G(0, 9) &= 0.9178, G(0, 10) = 0.6906, \\ G(0, 11) &= 0.9222, G(0, 12) = 0.3053, \\ G(0, 13) &= 0.9336, G(0, 14) = 0.8939, \\ G(0, 15) &= 0.9352, G(0, 16) = 0.9432, \\ G(0, 17) &= 0.9380; \end{aligned}$$

关联序为

$$\begin{aligned} A_3 &> A_{16} > A_{17} > A_{15} > A_{13} > A_1 > \\ A_{11} &k > A_9 > A_7 > A_4 > A_{14} > A_2 > \\ A_5 &> A_6 > A_8 > A_{10} > A_{12}. \end{aligned}$$

从结果中可以看出,如果选择关联度大于 0.85 或 0.9,则舍弃了 A_2 或 A_5 ,而进料流量 A_2 和新蒸汽流量 A_5 是两个重要的控制变量,因此结合该蒸发过程的实际操作经验和机理建模结果,选择关联度大于 0.8 的 13 个变量为影响因子.另外,通过对关联度的分析发现, I 效至 VI 效的液温及 VI 效的气压对出料浓度的关联度较高,而文献 [12-13] 构建出料浓度软测量模型时仅凭经验选取了 A_1, A_2, A_3, A_4, A_5 为影响因子作为模型的输入,势必会造成预测效果不佳.

5.2 软测量模型的建立

来自工业现场的样本数据常常含有噪声,同时,由于生产过程中根据车间调度而进行的一些生产调整(如蒸发过程的水洗和酸洗等)常常会造成一些孤立点的存在和数据缺失,这些特点影响了支持向量机模型的训练,而支持向量机对于训练集中数据噪音或者孤立点又过分敏感^[14].为解决这一问题,利用适合对不完整或缺失数据、包含噪声和野值数据聚类的模糊核聚类算法对灰关联后的输入样本进行聚类,尽量减少或消除异常点对支持向量机模型的影响,最终确定为 5 类;然后对每个子类构建最小二乘支持向量机模型,根据样本在时间和空间上的分布特征,对每类不同的样本赋予不同的模糊隶属度 μ_i ;最后利用综合加权得到最终的输出

$$Y_o = \sum_{j=1}^N \mu_{ij} y_j / \left(j \sum_{i=1}^N \mu_{ij} \right). \quad (12)$$

软测量建模过程如图 1 所示。

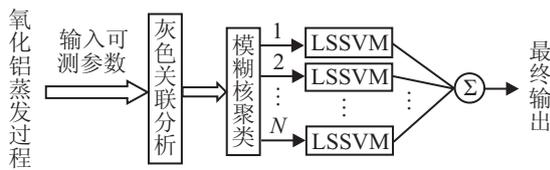


图 1 基于 GRA 的 KFCM-LSSVM 软测量建模过程

为得到较好的建模效果, 需要选择合适的惩罚系数 C 以及核函数宽度 σ , 本文采取混沌粒子群参数优选方法以参数对 (C, σ) 对目标函数寻优, 得到最优的参数集。

5.3 模型验证

为检验该过程建模方法的有效性, 基于实验平台配置为 windows XP 系统, 2.6G 主频, 1G 内存, Matlab7.0。本文以某氧化铝厂蒸发过程的生产历史数据为基础, 随机选取经过数据插值和数据滤波等处理后的 4 个月数据中的 300 组作为训练集, 60 组作为测试集, 用来比较分析算法的性能。根据蒸发过程数据的分布情况, 聚类个数初始值设定为 4, 参数的优化区间设为 $1 \leq C \leq 1000, 0.001 \leq \sigma \leq 1$; 采用的粒子群算法的参数设置为: 种群规模为 40, $V_{max} = 0.9, V_{min} = 0.4, c_1 = c_2 = 2$, 迭代次数为 1000, 每个子集的样本不少于 30 个。每一个样本集采用最小二乘支持向量机进行训练, 采用相对均方根误差 (RMSE) 作为评价指标, 惩罚系数 C 以及核函数宽度 σ 的初始值相应选取为 30 和 0.5。采用混沌粒子群算法最终得到最优参数对 $(C, \sigma) = (42, 0.7)$ 。灰关联分析前后的仿真结果如图 2 和图 3 所示。误差分析如表 1 所示。

从图 2、图 3 和表 1 中看出, GRA-KFCM-LSSVM

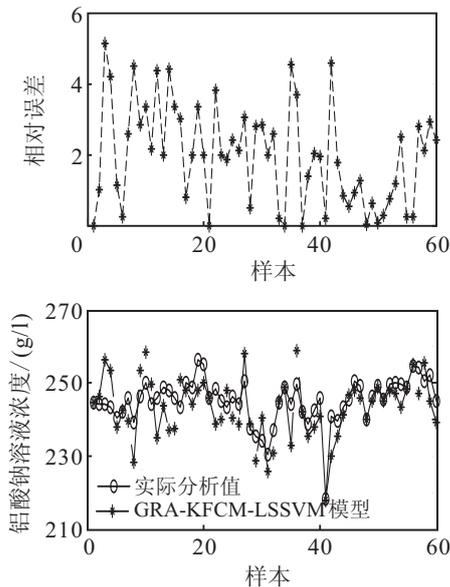


图 2 GRA-KFCM-LSSVM 仿真结果

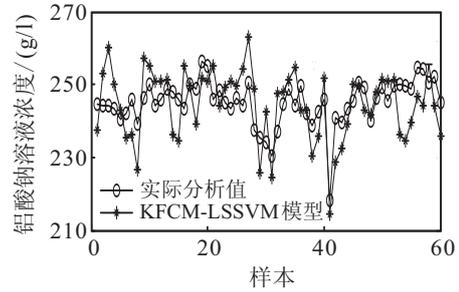
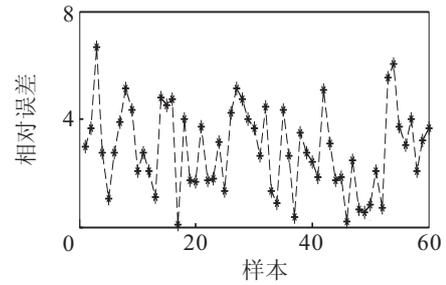


图 3 KFCM-LSSVM 的仿真结果

表 1 误差结果分析

算法	$E_{max}/\%$	RMSE	RRMSE/%
GRA-KFCM-LSSVM	4.91	0.00566	1.45
KFCM-LSSVM	6.28	0.00801	3.35

模型相比于 KFCM-LSSVM 软测量效果更好, 且最大相对误差仅为 4.91%, 表明模型软测量精度高; 均方根差为 0.00566, 表明模型可行度高; 均方根相对误差为 1.45%, 表明模型的跟踪性能强。另外, 将本文所建模型与文献 [12-13] 凭经验选取的输入变量构建的模型比较, 结果如表 2 所示。

表 2 两种算法的测试结果比较

训练样本数	方法	训练时间/s	RRMSE/%
300	文献[12-13]	6.20	9.8
300	GRA-KFCM-LSSVM	10.60	1.45

通过分析, 文献 [12-13] 的模型仿真结果最大相对误差为 12%, 且大于 10% 的样本占 12%, 而 GRA-KFCM-LSSVM 模型的最大相对误差仅为 6.28%, 说明模型具有较高的精度且跟踪样本数据变化趋势的能力强, 具有较强的适应能力。文献 [12-13] 由于凭经验选择输入变量造成漏选影响了出料浓度较大的因素变量, 模型的适应能力较差, 虽然训练时间相对于 GRA-KFCM-LSSVM 模型较短, 但是 10 min 足够满足工厂现场的工艺要求, 完全可以用来指导过程的操作优化。

6 结 论

本文针对蒸发过程机理复杂、流程长和影响因素多等特点以及铝酸钠溶液浓度在线检测困难, 提出了基于灰色关联分析的模糊核聚类和支撑向量机相结合的蒸发过程软测量模型。通过对经过灰关联筛选

出的输入样本空间的模糊核聚类,减少或消除了缺失数据和异常点对支持向量机模型训练的影响.采用现场生产数据的验证分析结果表明,所提出的方法在尽可能全面地选择样本的前提下,建立的蒸发过程软测量模型精度高、适应能力强,完全可以用于蒸发过程的实时操作优化.

参考文献(References)

- [1] Adilson J, Filho R M. Soft sensor development for on line bioreactor state estimation[J]. *Computer and Chemical Engineering*, 2000, 24(2): 1099-1103.
- [2] 苏博, 刘鲁, 杨方廷. 基于灰色关联分析的神经网络模型[J]. *系统工程理论与实践*, 2008, 28(9): 98-104.
(Su B, Liu L, Yang F T. Research of artificial neural network forecasting model based on grey relational analysis[J]. *Systems Engineering-Theory and Practice*, 2008, 28(9): 98-104.)
- [3] 仲蔚, 俞金寿. 基于模糊 c -均值聚类的多模型软测量建模[J]. *华东理工大学学报: 自然科学版*, 2000, 26(1): 83-87.
(Zhong W, Yu J S. Study on soft sensing modeling via FCM based multiple models[J]. *J of East China University of Technology: Natural Science*, 2000, 26(1): 83-87.)
- [4] 张淑宁, 王福利, 何大阔, 等. 基于多最小二乘支持向量机的草酸钴粒度的软测量[J]. *仪器仪表学报*, 2010, 31(9): 2081-2087.
(Zhang S N, Wang F L, He D K, et al. Soft sensing procbalt oxalate particle size based on multiple LSSVM regression[J]. *Chinese J of Scientific Instruction*, 2010, 31(9): 2081-2087.)
- [5] Sallehuddin R, Shamsuddin S M H, Hashim S Z M. Application of grey relational analysis for multivariate time series[C]. *Proc of the 8th Int Conf on Intelligent Systems Design and Applications*. Piscataway, 2008: 432-437.
- [6] 邓聚龙. 灰色系统理论教程[M]. 武汉: 华中理工大学出版社, 1990.
(Deng J L. Grey system theory tutorial[M]. Wuhan: Huazhong University of Technology Press, 1990.)
- [7] 于润桥, 党蕾. 基于灰色关联分析的焊接超声检测缺陷类型识别[J]. *仪器仪表学报*, 2006, 27(10): 1358-1360.
(Yu R Q, Dang L. Welding defect recognition based on gray correlation analysis in ultrasonic testing[J]. *Chinese J of Scientific Instruction*, 2006, 27(10): 1358-1360.)
- [8] 黄亮, 侯建军. 基于灰关联熵的模拟电路测试点优选算法[J]. *电子测量与仪器学报*, 2009, 23(9): 27-32.
(Huang L, Hou J J. Algorithm of selecting optimal test point for analog circuit based on gray relational entropy[J]. *J of Electronic Measurement and Instrument*, 2009, 23(9): 27-32.)
- [9] Chen W C, Wang M S. A fuzzy c -means clustering-based fragile watermarking scheme for image authentication[J]. *Expert Systems with Applications*, 2009, 36(2): 1300-1307.
- [10] Chen T. Incorporation fuzzy c -means and backpropagation network ensemble to job completion time prediction in a semiconductor fabrication factory[J]. *Fuzzy Sets and Systems*, 2007, 158(19): 2153-2168.
- [11] Kuo Y, Yang T, Huang G W. The use of grey relational analysis in solving multiple attribute decision-making problems[J]. *Computers and Industrial Engineering*, 2008, 55(1): 80-93.
- [12] 桂卫华, 阳春华, 柴琴琴, 等. 一种氧化铝蒸发装备出口碱液浓度的在线预测方法[P]. 中国: 2009.
(Gui W H, Yang C H, Chai Q Q, et al. An online prediction method of alkali out concentration in alumina evaporation equipment[P]. China: 2009.)
- [13] 阳春华, 钱晓山. 基于变异 CPSO 算法的 LSSVM 蒸发过程软测量[J]. *化工进展*, 2010, 29(3): 440-448.
(Yang C H, Qian X S. Soft sensor study of evaporation process in alumina production based on mutant CPSO and LSSVM[J]. *Chemical Industry and Engineering Progress*, 2010, 29(3): 440-448.)
- [13] Xu Q, Zhou H, Wang Y J, et al. Fuzzy support vector machine for classification of EEG signals using wavelet-based features[J]. *Medical Engineering and Physics*, 2009, 31(7): 858-865.

《控制与决策》第四次荣获中国高校精品科技期刊奖

2012年11月22日,由教育部科技司举办的第四届中国高校精品·优秀·特色科技期刊表彰大会在济南隆重举行.中宣部、教育部等有关领导出席会议,并作了重要讲话,来自全国高校科技期刊编辑出版单位的500多名代表与会。

经权威专家评审,共评选出精品科技期刊60种,优秀科技期刊120种,特色科技期刊30种。《控制与决策》荣获教育部科学技术司颁发的第四届中国高校精品科技期刊奖,这是本刊第四次获此殊荣。