第 27 卷第 3 期
2010 年 3 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 27 No. 3
Mar. 2010

文章编号: 1000−8152(2010)03−0395−05

独立序列均值与方差变点的累积和估计及应用

袁　芳[1], 田　铮[1,2], 苏晓丽[1], 陈占寿[1]

(1. 西北工业大学 应用数学系, 陕西 西安 710072; 2. 中国科学院 遥感应用研究所, 北京 100101)

摘要: 产品生产加工过程中, 反映产品质量参数的均值和方差都有可能出现异常波动, 及时发现质量异常波动, 对于控制产品质量十分重要. 在工业生产、通信工程等领域中独立序列有广泛应用背景. 本文研究了独立序列中均值和方差都存在变点且变点时刻不相同时的变点估计问题, 给出变点时刻的累积和(CUSUM)型估计, 并得到变点估计的收敛速度. 最后将该方法应用于航空发动机管路生产质量控制中和放射医学研究的扫描仪质量控制问题中, 模拟结果和实例分析都说明方法的有效性.

关键词: 独立序列; 均值与方差变点; CUSUM估计

中图分类号: O213.1　　**文献标识码:** A

Cumulative sum estimation for change-points in the mean and variance of independent observations and applications

YUAN Fang[1], TIAN Zheng[1,2], SU Xiao-li[1], CHEN Zhan-shou[1]

(1. Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an Shaanxi 710072, China;
2. State Key Laboratory of Remote Sensing Science,Institute of Remote Sensing Applications,
Chinese Academy of Sciences, Beijing 100101, China)

Abstract: In the process of product manufacturing, abnormal fluctuations of the mean and variance, which reflect product quality parameters, are always possible. It is extremely important to discover the abnormal variation of the quality in time for controlling the quality of products. Besides, independent observations have been applied extensively in industrial production, communication engineering and other fields. The estimation problem of change-points of independent observations whose mean and variance change respectively is studied. The cumulative sum(CUSUM) estimators of change-points are presented and the rate of convergence is established. Finally, the proposed method is applied to the production quality control of aero engine pipelines and the quality control of the scanners in the radiation medicine research. Both the simulation results and instance analysis support the validity of our method.

Key words: independent observations; change-point in the mean and variance; CUSUM estimation

1　引言(Introduction)

变点问题最早是从质量控制问题中提出的. 人们从生产线抽检产品以检测产品质量是否发生显著性波动, 特别是产品质量是否超过其质量检测范围, 当产品质量超过质量控制警戒线时, 希望能及时预警, 以免出现更多的次品, 这个质量发生质变的时刻就称为变点. 变点不仅在工业自动控制领域有重要应用, 而且大量应用于经济、金融、气象、地质等其它方面. 近20年来, 关于变点问题的研究无论在理论方面还是在应用方面皆有快速发展, 出现了一系列研究成果[1,2]. 关于独立序列均值变点或方差变点的文献很多, 如Bhattacharya[3], Yao[4]研究了正态情形独立序列的均值变点的估计, Kokoszka和Leipus[5]用CUSUM方法对方差有穷的相依随机变量序列的均值变点进行估计, 但在实际观测数据中往往会出现均值和方差同时存在变点, 且两种变点出现的时刻有可能不相同的情况. 如生产加工过程中会产生大量数据, 产品质量参数的均值与反映精度的方差都有可能出现异常波动, 及时发现质量的异常变异, 对于控制产品质量十分重要. 所以关于序列中均值与方差变点估计的研究是很有意义的. 而已有文献却很少讨论此类问题, Potarakis[6]用最小二乘法对回归模型中的均值和方差变点进行估计, 但最小二乘法需首先估计未知参数, 计算量比较大,未能给出变点估计的收敛速度, 且对方差变点进行估计时未考虑均值变点对其影响. 由于在工业生产、通信工程等

收稿日期: 2008−12−19; 收修改稿日期: 2009−04−22.

基金项目: 国家自然科学基金资助项目(60375003); 国家航空基金基础项目资助项目(03I53059); 西北工业大学科技创新基金资助项目(2007KJ01033).

领域中独立序列有广泛应用背景, 因此, 本文研究独立序列中均值与方差变点的估计问题, 用CUSUM方法进行估计, 直接构造估计量, 无需讨论与比较各种条件, 与文献[6]的方法相比更为实用.

时间序列模型的变点分析是重要的统计问题. 本文研究独立序列$\{X_t\},\ t = 1,2,\cdots,T$满足$\mathrm{E}X_t = \mu(t),\ \mathrm{Var}X_t = \sigma(t)$中的均值和方差均存在变点的情况下, 这两种参数变点的估计问题. 假设$\mu(t),\sigma(t)$都只取两个值, 分别为μ_1,μ_2和σ_1,σ_2, 即

$$\mu(t)=\begin{cases}\mu_1,\ t\leqslant k_0,\\ \mu_2,\ t>k_0,\end{cases}\quad \sigma(t)=\begin{cases}\sigma_1,\ t\leqslant k_1,\\ \sigma_2,\ t>k_2.\end{cases}\tag{1}$$

对变点k_0和k_1进行估计, 为方便起见, 在这里将X_t记为$X_t = \mu(t) + \sigma(t)\epsilon_t$, 其中$\epsilon_t$为独立同分布随机变量序列, 且$\mathrm{E}\epsilon_t=0, \mathrm{Var}\epsilon_t=1, \exists\, r>2$, 有$\mathrm{E}|\epsilon_t|^r<\infty$. 本文用CUSUM方法估计均值和方差的变点时刻, 得到了变点的一致估计和收敛速度, 并把它应用于航空发动机管路生产质量控制中及放射医学研究的扫描仪质量控制中, 有效估计产品质量参数的变点.

2 均值变点估计(Estimation for change-point in the mean)

首先应用CUSUM型估计量对独立序列$\{X_t\}$的均值变点进行估计, 变点k_0的估计定义为

$$\hat{k}_0 = \arg\max_k |R_k|,$$

其中

$$R_k=\frac{k(T-k)}{T^2}\Big\{\frac{1}{k}\sum_{t=1}^{k}X_t-\frac{1}{T-k}\sum_{t=k+1}^{T}X_t\Big\}.\tag{2}$$

在进行均值变点估计时, 无需对模型中的未知参数事先估计, 而是直接构造CUSUM统计量.

记$\tau = k/T,\ \tau_0 = k_0/T,\ \Delta = \mu_1 - \mu_2$.

定理 1 τ_0的估计$\hat{\tau}_0$满足$|\hat{\tau}_0 - \tau_0| = O_{\mathrm{p}}(\frac{1}{T^{\delta}})$, 其中$\delta < 1/2$.

$|\hat{\tau}_0 - \tau_0|$收敛速度还可以进一步改进, 下面是更强的结果.

定理 2 τ_0的估计$\hat{\tau}_0$满足$|\hat{\tau}_0 - \tau_0| = O_{\mathrm{p}}(\frac{1}{T\Delta^2})$.

3 方差变点估计(Estimation for change-point in the variance)

本节分两种情形讨论方差变点估计, 首先考虑均值及均值变点已知的情形下方差变点估计问题, 为此取方差变点k_1的估计$\hat{k}_1$为

$$\hat{k}_1 = \arg\max_k |V_k|,$$

其中

$$V_k = \frac{k(T-k)}{T^2}\Big\{\frac{1}{k}\sum_{t=1}^{k}(X_t-\mu(t))^2 - \frac{1}{T-k}\sum_{t=k+1}^{T}(X_t-\mu(t))^2\Big\}.\tag{3}$$

这里

$$\mu(t)=\mu_1, t\leqslant k_0; \mu(t)=\mu_2, t>k_0,$$

μ_1,μ_2,k_0均已知.

从上述均值和方差变点估计量可知均值变点估计量形式简单且不受方差变点的影响, 而方差变点的估计由于会受到均值及均值变点的影响, 使得其一致性与收敛速度的证明更为复杂. 定理3和定理4分别给出方差变点估计的一致性和收敛速度.

记$\tau_1 = k_1/T,\ \lambda = \sigma_1 - \sigma_2$.

定理 3 τ_1的估计$\hat{\tau}_1$满足$|\hat{\tau}_1 - \tau_1| = O_{\mathrm{p}}(\frac{1}{T^{\delta}})$, 其中$\delta < 1/2$.

定理 4 τ_1的估计$\hat{\tau}_1$满足$|\hat{\tau}_1 - \tau_1| = O_{\mathrm{p}}(\frac{1}{T\lambda^2})$.

在上述讨论中, 假设μ_1,μ_2,k_0已知. 而实际数据这些值一般是未知的. 下面讨论μ_1,μ_2,k_0未知时的情况, 为此, 首先对均值及均值变点作出估计, 再对方差变点进行估计. 变点k_1的估计定义为

$$\tilde{k}_1 = \arg\max_k |U_k|,$$

$$U_k = \frac{k(T-k)}{T^2}\Big\{\frac{1}{k}\sum_{t=1}^{k}(X_t-\bar{X})^2 - \frac{1}{T-k}\sum_{t=k+1}^{T}(X_t-\bar{X})^2\Big\}.\tag{4}$$

其中:

$$\bar{X} = \bar{X}_1 = \frac{1}{\hat{k}_0}\sum_{t=1}^{\hat{k}_0}X_t,\ t\leqslant\hat{k}_0;$$

$$\bar{X} = \bar{X}_2 = \frac{1}{T-\hat{k}_0}\sum_{t=\hat{k}_0+1}^{T}X_t,\ t>\hat{k}_0.$$

定理 5 τ_1的估计$\tilde{\tau}_1$满足$|\tilde{\tau}_1 - \tau_1| = O_{\mathrm{p}}(\frac{1}{T\lambda^2})$

证 选择$\delta > 0$, 满足$\tau_1 \in (\delta, 1-\delta)$, 可证得$\tilde{k}_1/T$是$\tau_1$的相合估计. 故对任意的$\epsilon > 0$, 当$T\to\infty$时, $\mathrm{P}\{\tilde{k}_1/T \notin (\delta,1-\delta)\} < \epsilon$. 需要证明: 对于一个任意大的实数$M > 0$, 当$T\to\infty$时, $\mathrm{P}\big\{|\tilde{\tau}_1 - \tau_1| > M(T\lambda^2)^{-1}\big\}\to 0$. 对上面的$M$, 定义

$$D_{T,M} = \{k : T\delta\leqslant k\leqslant T(1-\delta), |k-k_1| > M\lambda^{-2}\},$$

则

$$\mathrm{P}\{|\tilde{\tau}_1 - \tau_1| > M(T\lambda^2)^{-1}\}\leqslant$$

$$\mathrm{P}\{\tilde{\tau}_1 \notin (\delta, 1-\delta)\} + \mathrm{P}\{|\tilde{\tau}_1 - \tau_1| > M(T\lambda^2)^{-1}, \tilde{\tau}_1 \in (\delta, 1-\delta)\} \leqslant \epsilon + \mathrm{P}\{\sup_{k\in D_{T,M}} |U_k| \geqslant |U_{k_1}|\}. \quad (5)$$

可证

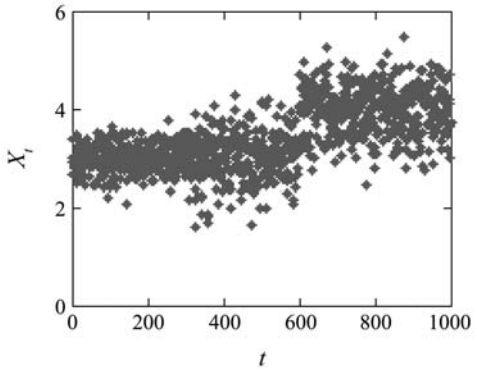
$$\mathrm{P}\{\sup_{k\in D_{T,M}} |U_k| \geqslant |U_{k_1}|\} \longrightarrow 0.$$

证毕.

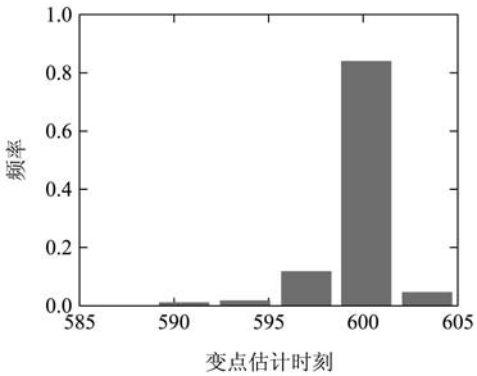
从以上收敛速度可以看出, 变点的估计与变点前后跳跃度有关, 跳跃度越大, 变点估计就越精确.

4　模拟结果(Simulation)

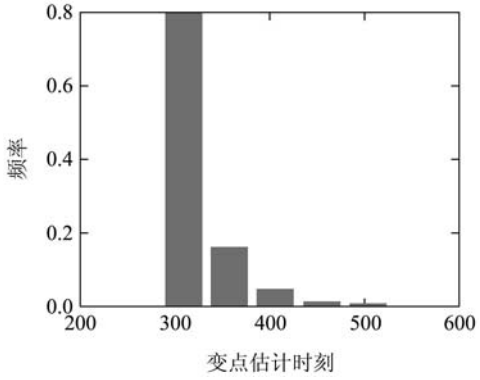
对随机序列$\{X_t\}, X_t = \mu + \sigma\epsilon_t$, 设参数向量$\theta = (\mu, \sigma)$, 样本容量$n = 1000$. 考虑以下数据生成过程$t \leqslant 300, \theta = (3, 0.3); 300 < t \leqslant 600, \theta = (3, 0.5); t > 600, \theta = (4, 0.5)$. 模拟1000 次, 变点估计的直方图见图1. 其中图1(a)~(d)分别表示随机生成一组数据的散点图, 均值变点估计直方图, 均值和均值变点已知时方差变点估计直方图及均值和均值变点未知时方差变点估计直方图. 均值变点的跳跃度较大, 从图1(b)可看出均值变点估计较为精确. 而方差变点跳跃度相对较小, 估计效果相对差些. 均值及均值变点已知时方差变点的估计比未知时估计效果好, 图1(c)和图1(d)说明了这一点.



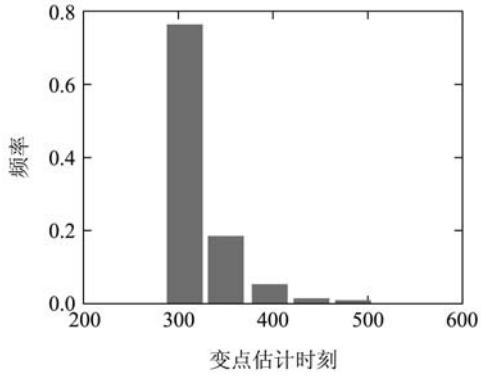
(a)



(b)



(c)



(d)

图 1　变点估计直方图

Fig. 1　Histograms of change-points

5　实例分析(Empirical illustration)

5.1　例1(Example 1)

航空发动机及其管路是飞机的“心脏”和“心血管”, 管路质量将直接影响发动机的可靠性和寿命. 由于管路布局空间限制严格, 管道层叠交错、形状复杂, 管路质量检测在管路生产过程中具有举足轻重的地位. 弯管工艺是航空发动机管路生产中的一项关键技术, 弯曲角度准确与否直接影响管路的装配质量. 弯曲回弹、模具的调节、机械振动、材料属性差异等都会影响弯管的弯曲角度, 弯曲角的检测是管路质量检测中的重要环节[7,8].

某管路设计弯曲角为120°, 公差范围为0.3°, 弯管机的误差范围±0.1°, 在生产线上每小时测量4个数据, 连续测量40个数据(见表1), 在理想的情况下, 测量值应随机地分散在参考值周围. 通过对40个数据进行估计, 发现均值在$k = 28$处存在变点. 弯制角度为120°±0.3°, 其均值应在120°附近, 而序号28后的实际角度均值却向大于120°的范围内偏离, 说明存在异常. 再根据文中关于均值及均值变点未知时方差变点估计的讨论, 估计得方差在$k = 35$处存在变点, 弯管精度发生偏差变化. 经质控人员检测与分

析, 发现均值偏离是因为弯管机与模具产生了间隙, 方差变化是因为弯管机电源电压不稳. 尽管生产中出现异常, 但已经生产的管路件弯曲角度仍在允许公差范围内, 仅仅是有发生质量问题的趋势与隐患. 通过及时查找原因, 进行整改, 使生产恢复到正常状态.

5.2 例2(Example 2)

在放射医学研究中, 总是期待所有仪器和设备在同一个体、同一条件下具有相同的精度误差. 当厂方将精度误差设置在适当的基准线内后, 监控设备确保精密就很重要了. 放射设备的可再生、设备改变、软件升级、仪器重校准、硬件老化或故障等都会使测量发生变化. 在理想的情况下,维护的好的设备测量值应随机地分散在参考值周围[9].

现对文献[9]中实例的两个月的DXA扫描仪质量控制数据进行分析(数据散点图如图2).

表 1 发动机管路弯曲角质量控制数据(单位: °)

Table 1 Data of aero engine pipeline(unit: °)

序号	测量值	序号	测量值	序号	测量值	序号	测量值	序号	测量值
1	120.055	9	119.967	17	119.978	25	119.931	33	120.062
2	119.912	10	119.958	18	120.037	26	119.789	34	120.165
3	120.231	11	119.890	19	119.998	27	119.994	35	120.497
4	120.021	12	119.990	20	120.261	28	119.979	36	120.115
5	119.721	13	119.783	21	120.243	29	120.313	37	120.223
6	119.932	14	120.092	22	120.094	30	120.163	38	120.186
7	119.902	15	119.802	23	120.014	31	120.060	39	120.161
8	120.015	16	119.901	24	119.879	32	120.006	40	120.250

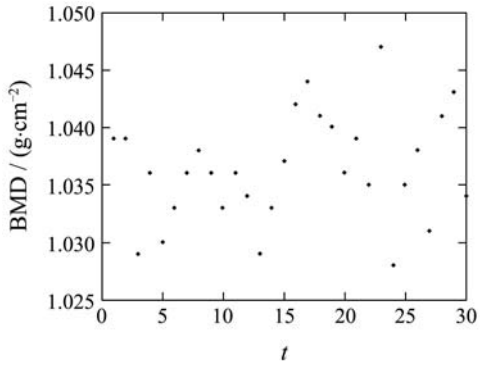


图 2 DXA扫描仪质量控制数据散点图

Fig. 2 Dot graph of quality control data by DXA scanner

在这个过程中, 一周扫描Hologic脊骨体膜3次, 目的是监控DXA的稳定性. 由于条件的不同均值会发生偏差, 而方差反映精度变化也会发生变化. 文献中应用CUSUM图方法, 为了检测标准差均值的变化, 需确定较多参考值进行比较. 现直接应用文中估计量对从03/13/89到05/15/89的30个数据进行估计, 发现均值在$k = 14$处存在变点, 而方差在$k = 21$处存在变点. 因此我们可从第20个数据对应的日期4月10日附近集中注意调查, 对方差变点前后的观测值进行分析, 发现方差变点后的观测值方差偏大, 这说明测量精度发生偏差, 扫描仪失控了. 此时需重新校准仪器. 这与文献中用CUSUM图的方法分析一致, 但应用本文中的估计量直接进行分析更为便捷, 从而可有效地指导精度偏差的调整.

6 结束语(Conclusion)

本文主要讨论了随机序列均值与方差变点的估计, 数值模拟结果和实例分析结果说明方法的可行性. 对于本文的估计问题可以作进一步推广, 将随机序列模型推广到线性过程或是非参数函数模型, 进行等距观测, 用文中方法对函数变点进行估计. 但是本文只考虑了独立序列中均值和方差分别存在一个变点的估计问题, 对于多变点的问题有待今后进一步研究.

参考文献(References):

[1] 韩四儿. 两类厚尾相依序列的变点统计分析[D]. 西安: 西北工业大学, 2006.
(HAN Si'er. *Analysis of change-point for two kinds of the heavy-tailed dependent observations*[D]. Xi'an: Northwestern Polytechnical University, 2006.)

[2] 齐培艳, 田铮. 噪声为单位根过程的非参数函数变点的小波检测[J]. 控制理论与应用, 2009, 26(1): 57 – 61.
(QI Peiyan, TIAN Zheng. Wavelet detection of jumping points in a nonparametric function with the unit-root noise[J]. *Control Theory & Applications*, 2009, 26(1): 57 – 61.)

[3] BHATTACHARYA P K. Maximum likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case[J]. *Multivariate Analysis*, 1987, 23(2): 183 – 208.

[4] YAO Y C. Approximating the distribution of the ML estimate of the

change-point in a sequence of independent random variables[J]. *Annals of Statistics*, 1987, 3(6): 1321 – 1328.

[5] KOKOSZKA P, LEIPUS R. Change-point in the mean of dependent observations[J]. *Statistics and Probability Letters*, 1998, 40(4): 385 – 393.

[6] PITARAKIS J Y. Least squares estimation and tests of breaks in mean and variance under misspecification[J]. *Econometrics*, 2004, 7(1): 32 – 54.

[7] 刘元朋. 航空发动机管路测量数据分割方法[J]. 航空学报, 2008, 29(2): 285 – 290.
(LIU Yuanpeng. Segmentation method for point cloud of aeroengine pipelines[J]. *Acta Aeronautica et Astronautica Sinica*, 2008, 29(2): 285 – 290.)

[8] 张艳丽. 用SPC统计技术控制机车弯管质量[J]. 铁道技术监督, 2008, 36(1): 12 – 14.
(ZHANG Yanli. Quality control of locomotive air pipe in SPC statistics technology[J]. *Raiway Quality Control*, 2008, 36(1): 12 – 14.

[9] 方积乾, 陆盈. 现代医学统计学[M]. 北京: 人民卫生出版社, 2002.
(FANG Jiqian, LU Ying. *Modern Medical Statistics*[M]. Beijing: People's Medical Press, 2002.)

[10] 葛春蕾, 史晓平. 正态分布参数变点估计的强相合性[J]. 合肥工业大学学报, 2008, 31(2): 280 – 283.
(GE Chunlei, SHI Xiaoping. Consistency of the estimators for the change point of the parameters of a norm al distribution[J]. *Journal of Hefei University of Technology*, 2008, 31(2): 280 – 283.)

作者简介:

袁　芳　(1985—), 女, 硕士研究生, 主要研究方向为非线性时间序列分析与信息处理, E-mail: yuanfangyf2005@163.com;

田　铮　(1948—), 女, 教授, 博士生导师, 主要从事非线性时间序列分析、多尺度非线性随机模型、计算机视觉与图像处理等研究, E-mail: zhtian@nwpu.edu.cn;

苏晓丽　(1986—), 女, 硕士研究生, 主要研究方向为非线性时间序列分析与信息处理, E-mail: suxiaoli986@163.com;

陈占寿　(1982—), 男, 博士研究生, 主要研究方向为非线性时间序列分析与信息处理, E-mail: chenzhanshou@126.com.

(上接第394页)

[3] WALCZAK B, MASSART D L. The radial basis function-partial least squares approach as a flexible non-linear regression technique[J]. *Analytica Chimica Acta*, 1996, 331(3): 177 – 185.

[4] DURAND J F. Local polynomial additive regression through PLS and splines: PLSS[J]. *Chemometrics & Intelligent Laboratory Systems*, 2001, 58(2): 235 – 246.

[5] WOLD S, SJÖSTROM M, ERIKSSON L. PLS-regression: a basic tool of chemometrics[J]. *Chemometrics & Intelligent Laboratory Systems*, 2001, 58(2): 109 – 130.

[6] JONG S. SIMPLS: an alternative approach to partial least squares regression[J]. *Chemometrics & Intelligent Laboratory Systems*, 1993, 18(3): 251 – 263.

[7] SERNEELS S, CROUX C, FILZMOSER P, et al. Partial robust M-regression[J]. *Chemometrics & Intelligent Laboratory Systems*, 2005, 79(1/2): 55 – 64.

[8] WALCZAK B, MASSART D L. Local modelling with radial basis function networks[J]. *Chemometrics & Intelligent Laboratory Systems*, 2000, 50(2): 179 – 198.

[9] SERNEELS S, FILZMOSER P, CROUX C, et al. Robust continuum regression[J]. *Chemometrics & Intelligent Laboratory Systems*, 2005, 76(2): 197 – 204.

[10] CROUX C, FILZMOSER P, OLIVEIRA M R. Algorithms for projection-pursuit robust principal component analysis[J]. *Chemometrics & Intelligent Laboratory Systems*, 2007, 87(2): 218 – 225.

[11] KUTNER M H, NACHTSHEIM C J, NETER J. *Applied Linear Regression Models*[M]. Beijing: Higher Education Press, 2005: 439 – 441.

[12] DASZYKOWSKI M, KACZMAREK K, HEYDEN Y V, et al. Robust statistics in data analysis-a review: basic concepts[J]. *Chemometrics & Intelligent Laboratory Systems*, 2007, 85(2): 203 – 219.

[13] HUBERT M, BRANDEN K V. Robust methods for partial least squares regression[J]. *Journal of Chemometrics*, 2003, 17(10): 537 – 549.

[14] 王文忠, 郑平. Ni, Co萃取过程平衡pH值数学模型的建立[J]. 河北理工学院学报, 1999, 21(4): 19 – 22.
(WANG Wenzhong, ZHENG Ping. Establishing a mathematical model of equilibrium in the process of extraction Ni, Co[J]. *Journal of Hebei Institute of Technology*, 1999, 21(4): 19 – 22.)

[15] ANITHA M, SINGH H. Artificial neural network simulation of rare earths solvent extraction equilibrium data[J]. *Desalination*, 2008, 232(1/3): 59 – 70.

作者简介:

贾润达　(1981—), 男, 博士研究生, 主要研究方向为复杂工业系统建模与优化, E-mail: jiarunda@yahoo.com.cn;

毛志忠　(1961—), 男, 教授, 博士生导师, 主要研究方向为复杂工业系统建模、控制与优化, E-mail: maozhizhong@ise.neu.edu.cn;

常玉清　(1973—), 女, 副教授, 主要研究方向为软测量技术, E-mail: changyuqing@mail.neu.edu.cn;

周俊武　(1966—), 男, 博士, 研究员, 主要研究方向为复杂工业系统建模, E-mail: zhou_jw@bgrimm.com.