基于变量贡献率的 MSPC 异常识别方法

杜福洲 唐晓青

(北京航空航天大学 机械工程及自动化学院, 北京 100191)

摘 要: 异常识别是多元统计过程控制(MSPC, Multivariate Statistical Process Control)方法有效应用的关键. 针对现有研究对历史异常信息利用的不足,综合考虑了主成分变量贡献率与重构误差变量贡献率对异常识别的影响,将两种变量贡献率进行归一化处理并求和得到综合变量贡献率;提出了一种基于综合变量贡献率的 MSPC 异常识别方法,并基于 matlab 计算平台实现了该算法. 通过田纳西过程故障模式仿真及异常识别,对该方法的应用及算法有效性进行了实例验证.

关 键 词: 质量控制; 统计过程控制; 多变量控制系统; 异常识别

中图分类号: TH 165$^+$.4

文献标识码: A **文 章 编 号**: 1001-5965(2012)10-1295-05

Method of MSPC fault detection and diagnosis based on variable contributions

Du Fuzhou Tang Xiaoqing

(School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Fault detection and diagnosis is one of the key technologies on the effective application of multivariate statistical process control(MSPC). In order to overcome the historical fault information using shortage, considering the influence of principal components variable contributions and the reconstructive errors, the synthetical variable contributions were calculated by normalizing and summing these two different variable contributions. A novel MSPC fault detection and diagnosis method was proposed based on the integrated variable contributions, and the relevant algorithm and program were presented and implemented. A case study was illustrated through the Tennessee Eastman challenge process simulation platform. The experimental results demonstrate that the proposed method is feasible and valid.

Key words: quality control; statistical process control; multivariable control systems; fault detection and diagnosis

多元质量控制作为一种通过监控多个关键质量特性的变化来达到控制产品质量目的的方法,已经引起了学术界和工业界的广泛关注,并开始在生产过程中得到应用[1-2]. 在实际生产过程运行中,难免会出现过程的异常波动和各种故障,其在控制图上通常表现为超出控制线的异常点. 根据采样数据和具体生产过程的专业知识来识别过程的异常是多元质量控制图能够有效应用的关键.

基于主成分分析(PCA, Principal Component Analysis)的多元质量控制图通过对采样数据构建 PCA 模型,将过程变量数据由高维采样空间投影到少数隐变量定义的低维特征空间,同时保留过程主要的特征信息,从而利用特征空间中的样本 T^2 统计量和平方预测误差(SPE, Squared Prediction Error)统计量实现过程监测.

收稿日期: 2011-11-06; **网络出版时间**: 2012-06-15 15:44

网络出版地址: www.cnki.net/kcms/detail/11.2625.V.20120615.1544.029.html

基金项目: 国家自然科学基金资助项目(50905010)

作者简介: 杜福洲(1976-),男,山东潍坊人,讲师,du_fuzhou@163.com.

对于多元质量控制图的异常识别常用的方法有基于 PCA 的贡献图法、基于独立成分分析的异常识别方法等,文献[3]将多变量系统的异常信息输入到人工神经网络,利用人工神经网络实现过程异常的识别;文献[4]针对瞬变过程,提出了一种改进的基于 PCA 异常识别过程,该方法可实现对过程稳态与瞬态异常识别;文献[5]针对部分半导体制造过程非正态性分布的特点,对其过程的异常检测进行了研究;文献[6]则采用动态独立成分分析的方法,对过程的异常识别与诊断进行了研究. 以上几种方法主要采用当前异常点的信息进行异常的识别与故障诊断,而实际的过程中当过程出现异常时,往往会出现多个过程异常点,如果仅采用当前异常点的信息,容易受到随机误差或噪声等因素的影响而出现误判或者漏判. 针对基于 PCA 的多元质量控制图的特点,将过程异常时过程变量对于监测统计量 T^2 和 SPE 的贡献率进行归一化后综合,提出了一种基于综合累积变量贡献率的 MSPC 异常识别方法,并给出了该方法的程序实现. 最后,采用田纳西过程仿真平台对该方法的应用进行了实例研究.

1 基于 PCA 的多元统计过程控制

利用 PCA 对生产过程进行监测,首先要获得过程正常运行状态下的采样数据矩阵 $\boldsymbol{X}_{n\times p}$(其中,$n$ 为样本数,p 为监控变量的维数),并使得 $\boldsymbol{X}$ 包含生产过程中可视为正常运行状态尽可能多的信息,然后根据 $\boldsymbol{X}$ 建立过程的 PCA 模型,其步骤如下:

1)对采样数据进行标准化处理,处理后其采样数据矩阵仍记为 $\boldsymbol{X}$.

2)计算处理后的采样数据矩阵 $\boldsymbol{X}$ 的协方差矩阵 $\boldsymbol{\Sigma}$.

3)计算协方差矩阵 $\boldsymbol{\Sigma}$ 的前 m 个特征值 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m$,及其对应的特征向量 $\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_m$,其中 m 为选取的主成分的个数,本文采用主成分贡献率法来选取 m[7].

对于生产过程中新采集到的一组数据 $\boldsymbol{X}_i = (x_{i1}, x_{i2}, \cdots, x_{ip})$,可由过程的 PCA 模型得到其主成分得分 $\boldsymbol{T}_i = (t_{i1}, t_{i2}, \cdots, t_{im})$、重构数据 $\hat{\boldsymbol{X}}_i = (\hat{x}_{i1}, \hat{x}_{i2}, \cdots, \hat{x}_{ip})$ 及重构误差 $\boldsymbol{X}_{ei}$[7]:

$$\boldsymbol{T}_i = \boldsymbol{X}_i \boldsymbol{P}_m \tag{1}$$

$$\hat{\boldsymbol{X}}_i = \boldsymbol{T}_i \boldsymbol{P}_m^{\mathrm{T}} = \boldsymbol{X}_i \boldsymbol{P}_m \boldsymbol{P}_m^{\mathrm{T}} \tag{2}$$

$$\boldsymbol{X}_{ei} = \boldsymbol{X}_i - \hat{\boldsymbol{X}}_i \tag{3}$$

其中,$\boldsymbol{P}_m = (\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_m)$ 为过程 PCA 模型的载荷矩阵;$\boldsymbol{T}_i$ 为第 i 个采样数据的得分向量. 定义预测误差平方和 SPE(也称为 Q 统计量)为

$$Q = \sum_{j=1}^{p} (x_{i,j} - \hat{x}_{i,j})^2 = \boldsymbol{X}_{ei} \boldsymbol{X}_{ei}^{\mathrm{T}} \tag{4}$$

经过 PCA 变换将过程的信息投影到 2 个子空间:主成分子空间和残差子空间. 过程正常运行时,主成分子空间包含了过程正常运行的大部分信息,而残差子空间主要包含了过程的冗余信息、测量或环境噪声等信息. 对于过程或产品的质量控制也就转化为对这 2 个子空间中的统计量 T^2 和 Q 的监控. 如果过程正常运行下的测量数据服从多元正态分布,则

$$T^2 = \boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} \boldsymbol{t} \sim \frac{m(n^2-1)}{n(n-m)} F(m, n-m) \tag{5}$$

式中,$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{bmatrix}$ 为过程主成分的协方差矩阵,为对角阵;$\boldsymbol{t}$ 为采样的主成分得分向量;m 为过程主成分个数;n 为用于建模的样本数;$F(m, n-m)$ 为第一自由度为 m,第二自由度为 $n-m$ 的 F 分布. T^2 控制图的上控制限为

$$L_{T^2} = \frac{m(n^2-1)}{n(n-m)} F_\alpha(m, n-m) \tag{6}$$

式中 α 为控制图的置信度,其取值根据具体的应用环境而定. Q 统计量的上控制限为[8]

$$L_Q = \theta_1 \left[1 - \frac{\theta_2 h_0 (1-h_0)}{\theta_1^2} + \frac{z_\alpha (2\theta_2 h_p^2)^{\frac{1}{2}}}{\theta_1} \right]^{\frac{1}{h_0}} \tag{7}$$

其中 $\theta_j = \sum_{i=m+1}^{p} \lambda_i^j \quad j = 1, 2, 3$

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2}$$

z_α 为选定置信度 α 时标准正态分布所对应的上分位点.

如果过程出现异常,则会导致 T^2 值或 Q 值出现异常,而超出其控制线. 如果仅仅是 T^2 值异常,则表明过程模型仍然有效,但过程测量变量的变化超出了正常的工作范围. 如果 Q 值出现了异常,则表明此时过程的 PCA 模型已经失效,该模型已经无法正确的解释过程中包含的信息.

2 基于变量贡献率的异常识别

2.1 基于综合变量贡献率的异常识别

贡献率图是一种基于 PCA 的多元质量控制图异常识别方法,它考虑了变量之间的相关性,较单变量统计技术更为合理. 这种方法的思想是:当

控制图出现异常时,量化每个过程变量相对于异常点的贡献率,然后根据每个变量的贡献率来对该异常进行识别. 基于变量的贡献率图主要分为 2 种:主成分贡献率图和重构误差贡献率图. 本文首先研究了 2 种变量贡献率的计算方法,然后对变量贡献率进行归一化处理并加和,求得总的变量贡献率.

当 T^2 图出现异常时,量化每个过程变量相对于得分向量的贡献率,然后对那些造成失控状态的得分的贡献率进行求和,其算法如下:

1) 检查异常点 $\boldsymbol{X}_i$ 的标准化得分 $(t_{i,j}/\lambda_j)^2$,并确定造成失控状态的 $r \leqslant m$ 个得分,其标准为

$$(t_{i,j}/\sigma_{i,j})^2 > \frac{1}{m} UCL_{T^2}$$

其中 $t_{i,j}$ 为异常点 $\boldsymbol{X}_i$ 映射到第 j 个载荷向量上的得分,$j = 1,2,\cdots,m$.

2) 计算每个监控变量 $x_{i,j}$ 相对于失控得分 $t_{i,l}$ 的贡献率:

$$C_{i,j,l} = \left| \frac{t_{i,l}}{\lambda_l^2} p_{l,j} (x_{i,j} - \mu_j) \right| \tag{8}$$

其中,$p_{l,j}$ 为过程 PCA 模型载荷矩阵 $\boldsymbol{P}$ 的第 (l,j) 个元素,$l = 1,2,\cdots,r$;$j = 1,2,\cdots,p$.

3) 计算过程变量 $x_{i,j}$ 的总贡献率:

$$\Omega_{i,j} = \sum_{l=1}^{r} C_{i,j,l} \tag{9}$$

其中,$\Omega_{i,j}$ 为过程异常点 $\boldsymbol{X}_i$ 的第 j 个变量对于过程异常的贡献率.

4) 对贡献率进行标准化,计算方法为

$$\Omega'_{i,j} = \frac{\Omega_{i,j}}{\sum_{j=1}^{p} \Omega_{i,j}} \tag{10}$$

当 Q 图出现异常时,重构误差变量贡献率的计算方法如下:

① 计算重构误差,其计算公式为

$$\boldsymbol{X}_{ei} = \boldsymbol{X}_i - \hat{\boldsymbol{X}}_i = \boldsymbol{X}_i - \boldsymbol{X}_i \boldsymbol{P}_m \boldsymbol{P}_m^{\mathrm{T}} \tag{11}$$

② 对重构误差变量贡献率进行标准化,其计算公式为

$$x'_{ei,j} = \frac{x_{ei,j}}{\sum_{j=1}^{p} |x_{ei,j}|} \tag{12}$$

当分别计算出主成分变量贡献率和重构误差变量贡献率后,需要对它们进行综合,将重构误差变量贡献率取绝对值然后加和,计算方法为

$$C'_{i,j} = k_1 \Omega_{i,j} + k_2 |x_{ei,j}| \tag{13}$$

式中,k_1 为主成分变量贡献率系数,T^2 图出现异常时 $k_1 = 1$,正常时 $k_1 = 0$;k_2 为重构误差变量贡献率系数,Q 图出现异常时 $k_2 = 1$,正常时 $k_2 = 0$. 然后对综合的变量贡献率进行标准化,绘制主成分变量贡献率、重构误差变量贡献率以及综合变量贡献率曲线,根据这 3 个变量贡献率曲线图以及过程的专业知识对控制图的异常进行识别和诊断.

2.2　基于累积变量贡献率的异常识别

仅根据一个异常点来进行控制图的异常识别,很容易受到环境噪声的影响. 为此,本文提出了一种基于累积变量贡献率的异常识别方法,其主要思想是充分利用控制图异常的历史信息,对每个异常点的变量贡献率进行累积,其主要算法如下:

1) 采用 2.1 节所提出的变量贡献率的计算方法计算异常点的变量贡献率;

2) 如果控制图再次出现异常点,先计算该异常点的变量贡献率,然后与以前所有异常点的变量贡献率进行累积;

3) 对变量的累积贡献率进行标准化处理;

4) 绘制主成分变量累积贡献率、重构误差变量累积贡献率和综合变量累积贡献率曲线图,根据这些累积贡献率曲线进行异常的识别.

3　实例分析

3.1　田纳西过程介绍

田纳西过程(TECP, Tennessee Eastman Challenge Process)是由 Downs 和 Vogel 于 1993 年提出的[9]. 该过程以实际的复杂工业化工过程为基础,已为不同研究者广泛采用并作为过程建模、优化、多元统计过程监控与异常诊断等不同研究目的的标准测试平台. 田纳西过程主要流程如图 1[10]所示.

田纳西过程包含 5 个主要操作单元:反应器、冷凝器、循环压缩机、汽液分离器和解吸塔. 它以 4 种反应物 A,C,D,E 为原料,通过其中同时发生的汽液两相放热反应生成 2 种主要产品 G 和 H:

$$A(\text{汽}) + C(\text{汽}) + D(\text{汽}) \rightarrow G(\text{液})$$

$$A(\text{汽}) + C(\text{汽}) + E(\text{汽}) \rightarrow H(\text{液})$$

同时进行的 2 个副反应则生成惰性物质 B 及副产品 F. 该过程共有 12 个操纵变量和 41 个测量变量,供检测和控制用. 其中有 22 个测量变量为连续,另外 19 个为气相色谱仪对各种浓度的测量值,其采样周期为 0.1 h 或 0.25 h. 此外,田纳西过程仿真平台还提供了 20 种不同的扰动,分别对应着 20 种不同的过程异常.

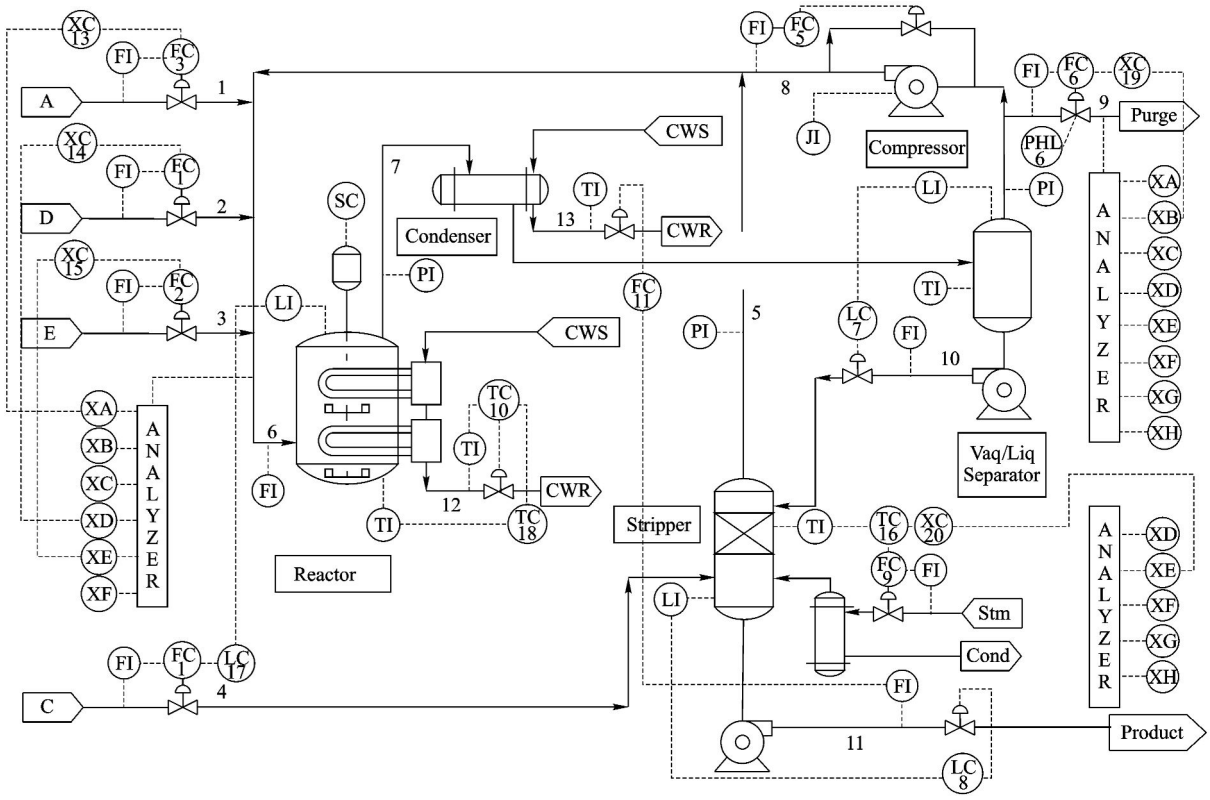

图 1 田纳西过程流程图[10]

3.2 仿真及结果分析

选用田纳西过程仿真平台进行基于变量贡献率的 MSPC 过程异常识别算法的实例研究,监测变量选取 22 个连续测量变量,采样时间间隔为 3 min. 首先,在过程平稳运行时对过程进行采样并建立过程的 PCA 模型,然后,采用基于 PCA 的多元质量控制图对过程进行监测.

当扰动发生时,对控制图异常点计算主成分变量贡献率、重构误差变量贡献率和综合变量贡献率以及与以前异常点的累积值,绘制相应的贡献率曲线. 过程运行时间为 48 h,在过程运行 5 h 时引入故障 1,故障 1 为 A/C 进料流量比发生变化,组分 B 含量保持不变,故障类型为阶跃,贡献率曲线及贡献率图(其中变量 1 为物料 A 的流量)如图 2 ~ 图 8 所示.

故障 1 发生时,在流 4 中的 A/C 进料比发生一个阶跃变化,从而导致流 4 中的 C 进料增加,而 A 进料较少,这又导致再循环流 5 中 A 进料的减少,并且控制回路作出反应使流 1 中的 A 进料增加,如图 9 所示.

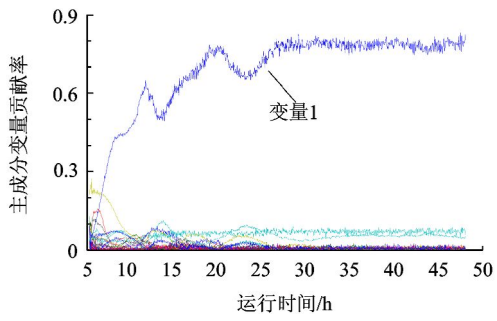

图 2 故障 1 下的主成分变量贡献率曲线

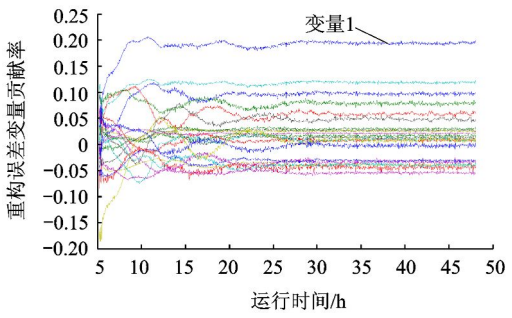

图 3 故障 1 下的重构误差变量贡献率曲线

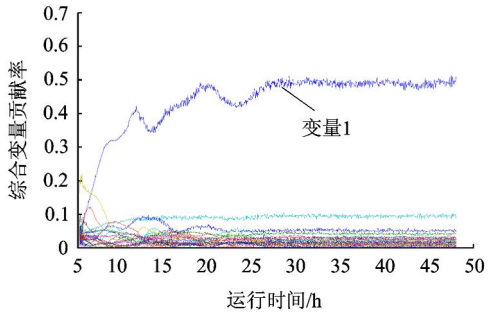

图 4 故障 1 下的综合变量贡献率曲线

随着时间的推移,这 2 种作用相互抵消,从而使得流 1 中 A 的进料在经过一段时间后(大约 24 h)成为常数,但较正常工况下其均值向上偏移

了大约 500 m^3/h. 从图 2 ~ 图 8 中可以证实此时过程的异常主要是由于变量 1 即流 1 中 A 的进料流量发生了异常引起的.

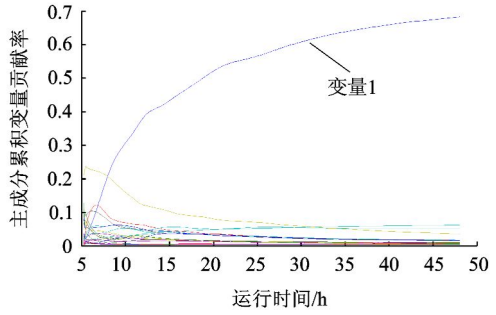

图 5　故障 1 下的主成分累积变量贡献率曲线

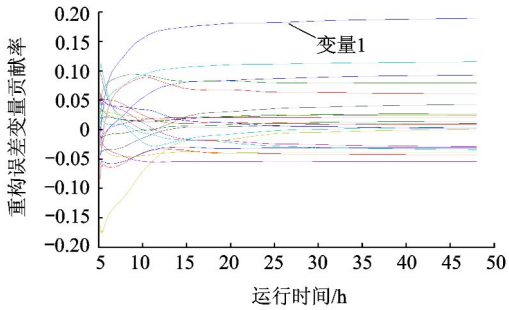

图 6　故障 1 下的重构误差累积变量贡献率曲线

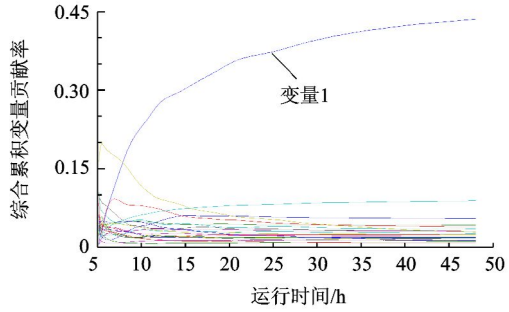

图 7　故障 1 下的综合累积变量贡献率曲线

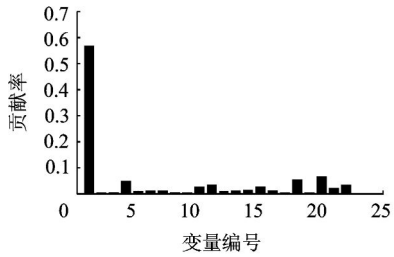

图 8　故障 1 下运行 10 h 后第 1 个异常点的综合累积变量贡献率图

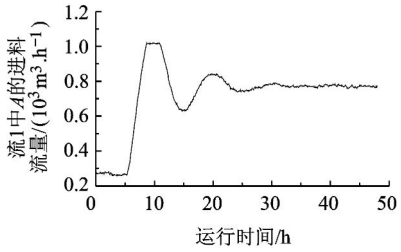

图 9　流 1 中物料 A 的进料流量曲线

4　结 束 语

MSPC 的异常识别是该方法有效应用的基础. 本文在现有研究的基础上,将变量的贡献率分为主成分变量贡献率和重构误差变量贡献率,并对两者进行了综合,提出了一种基于综合变量贡献率以及累积综合变量贡献率的多元统计过程异常识别方法,通过对田纳西过程的仿真研究表明,该方法是有效、可行的.

参考文献（References）

[1] Lieftucht D, Kruger U, Irwin G W. Improved reliability in diagnosing faults using multivariate statistics[J]. Computers and Chemical Engineering, 2006, 30: 901 – 912

[2] Bozorgtabar B, Noorian F, Rad G A. Comparison of different PCA based face recognition algorithms using genetic programming [C]//5th International Symposium on Telecommunications. Tehran: IEEE, 2010: 801 – 805

[3] El-Midany T T, El-Baz M A, Abd-Elwahed M S. A proposed framework for control chart pattern recognitions in multivariate process using artificial neural networks[J]. Expert Systems with Applications, 2010, 37(2): 1035 – 1042

[4] Alkaya A, Eker I. A new threshold algorithm based PCA method for fault detection in transient state processes[C]//7th International Conference on Electrical and Electronics Engineering. Piscataway, NJ: IEEE, 2011: 144 – 147

[5] Yu Jianbo. Fault detection using principal components-based Gaussian mixture model for semiconductor manufacturing processes[J]. IEEE Transactions on Semiconductor Manufactureing, 2011, 24(3): 432 – 444

[6] Stefatos G, Hamza A B. Fault detection and isolation of faults in a multivariate process with Bayesian network [J]. Journal of Process Control, 2010, 20(8): 902 – 911

[7] 杨英华. 多元统计过程监测和故障诊断方法及其应用研究[D]. 沈阳:东北大学信息科学与工程学院, 2002
Yang Yinghua. Multivariate statistics based process monitoring and fault diagnosis method and its application[D]. Shenyang: College of Information Science and Engineering, Northeastern University, 2002 (in Chinese)

[8] Yang Qingsong. Model-based and data driven fault diagnosis methods with applications to process monitoring[D]. Cleveland: Department of Electrical Engineering and Computer Sciences, Case Western Reserve University, 2004

[9] Downs J J, Vogel E F. A plant-wide industrial process control problem[J]. Computers and Chemical Engineering, 1993, 17(3): 245 – 255

[10] Chiang L H, Russell E L, Braatz R D. Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 2000, 50: 243 – 252

(编　辑:文丽芳)