

doi: 10.3969/j.issn.1674-8425(z).2014.01.019

# 运用主成分分析法的过程控制和诊断

肖枝洪,冉小华

(重庆理工大学 数学与统计学院,重庆 400054)

**摘 要:**针对控制过程中多变量的共线性,提出一种主成分质量控制方法,对化工中甲烷气体生产过程进行质量监控与诊断。首先运用在正常状态下所观测的13组样本数据建立过程控制模型,然后根据所建模型对在异常状态下观测的12组样本数据进行监控与诊断。结果表明:该方法能够及时地监测出故障并能准确查找出引起故障的过程变量。

**关键词:**主成分分析;故障监控与诊断; $T^2$ 控制图

中图分类号:TP273

文献标识码:A

文章编号:1674-8425(2014)01-0096-06

## Monitoring and Fault Diagnosis Based on PCA Method

XIAO Zhi-hong, RAN Xiao-hua

(School of Mathematics and Statistics,

Chongqing University of Technology, Chongqing 400054, China)

**Abstract:** To the deficiencies of commonly used multivariate statistical process control, the principal component quality control method is put forward to proceed quality monitoring and diagnostics on the production process of practical industrial methane gas. First, the process control model was established using the first 13 groups of sample data we observed in the normal state. Then, we used the established model to proceed quality monitoring and diagnostics on the known 12 groups of sample data we observed in the abnormal state. The results show that the method can monitor the fault timelier and find out the process variables that cause malfunction accurately.

**Key words:** principal component analysis (PCA); fault detection and diagnosis; controlling figure of  $T^2$

在工业生产过程中,经常会发生各种故障,影响产品质量,甚至会引起重大人员伤亡和财产损失。因此在工业生产过程中进行故障检测与诊断已成为当前控制理论研究的热点之一。工业生产系统

一般都具有过程精确、建模困难、过程变量众多且相互间强耦合,并且在实际中存在各种随机因素的影响等特点,这就使得基于机理模型的诊断方法的应用极为不便<sup>[1]</sup>。因此,在实际工业生产过

收稿日期:2013-09-07

基金项目:重庆市科委资助项目(cstc2012jjA00018);重庆市教委项目(KJ130810)

作者简介:肖枝洪(1965—),男,博士,教授,主要从事工程概率、可靠性理论等方面研究。

引用格式:肖枝洪,冉小华.运用主成分分析法的过程控制和诊断[J].重庆理工大学学报:自然科学版,2014(1):96-101.

Citation format: XIAO Zhi-hong, RAN Xiao-hua. Monitoring and Fault Diagnosis Based on PCA Method [J]. Journal of Chongqing University of Technology: Natural Science, 2014(1): 96-101.

程中统计建模运用的较多,多元统计过程控制是比较常用的多元质量控制方法。该方法能够检测生产过程中的多个相关变量,通过诊断、寻找、消除异常因素,达到提高产品质量的目的。然而当前运用的多元统计过程控制技术存在的主要问题是:随着过程变量维数的增加,监控的复杂性急剧增加,而变量间严重相关时会给过程的诊断带来极大的困难,且监控效率极低<sup>[2]</sup>。为了克服多元统计过程控制的不足,就需要将降维的主成分分析方法与质量控制过程有机地结合起来,对此已有大量的研究。例如文献[3-4]将此方法分别运用于食品加工厂设备监控系统和大坝安全多测点数据监控得到很好的效果。在文献[5]中提出运用主成分原理的多元质量控制图来监测荫罩生产过程,根据得分综合值绘制出多元控制图来检测和控制生产过程,但不足的是:利用得分综合做出的多元控制图重叠了较多的原始数据信息,不能反映真实的生产过程情况。在此基础上,本文将运用主成分分析法进行多变量质量控制与诊断,但不同于文献[5]的是:建立了故障检测与诊断模型,通过对 $T^2$ 和 $Q$ (或 $SPE$ )统计量与其阈值的判断,进行系统的故障检测与诊断。

## 1 主成分分析理论

### 1.1 基本思想

主成分分析是将原始变量进行综合,找出影响某一质量问题的少数几个综合特性。这些综合过程特性是将原来多个过程进行线性组合,这样不仅保留了原始数据中的主要变量,并且消除了变量间的相关性,与原来的过程特性相比更优越,是将原来多个过程特性降维为少数几个具有代表性的过程特性的质量控制方法<sup>[2]</sup>。

### 1.2 主成分分析原理

设有 $n$ 个样品,设每个样品的主成分模型<sup>[7]</sup>为:

$$\begin{cases} F_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \\ F_2 = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p \\ \vdots \\ F_p = a_{1p}X_1 + a_{2p}X_2 + \dots + a_{pp}X_p \end{cases} \quad (1)$$

满足以下条件:

- 1) 每个主成分系数平方和为1,即 $a_{1i}^2 + a_{2i}^2 + \dots + a_{pi}^2 = 1 (i = 1, 2, \dots, m)$

- 2) 主成分之间互不相关,即

$$\text{cov}(F_i, F_j) = 0$$

- 3) 主成分方差依次递减,即

$$\text{Var}(F_1) \geq \text{Var}(F_2) \geq \dots \geq \text{Var}(F_p)$$

从代数观点看,主成分就是 $p$ 个变量 $X_1, \dots, X_p$ 的一些特殊的线性组合,在几何上这些线性组合将 $X_1, \dots, X_p$ 构成的坐标系旋转产生新坐标系,新坐标轴使之通过变差最大的方向(即说具有最大的样品方差)。

## 2 故障监控与诊断模型

基于主成分模型的多变量质量控制图<sup>[2]</sup>常有主成分 $SPE$ 控制图、主成分 $T^2$ 控制图以及主成分贡献控制图。

假设多变量过程 $X = (X_1, X_2, \dots, X_p)' \sim N(\mu, \Sigma)$ 的目标值为 $\mu_0$ ,从中选取 $n$ 个样本,记第 $i$ 个样品的观测向量为 $X_{(i)} = (X_{i1}, X_{i2}, \dots, X_{ip})'$ , $n$ 个样品组成的矩阵为 $X_{n \times p} = (X_{(1)}, X_{(2)}, \dots, X_{(n)})'$ 。在进行主成分分析过程之前,应该先将原始数据标准化,即作如下标准化变换:

$$X_s = [X_{n \times p} - (1, 1, \dots, 1)\bar{X}'_{n \times p}] \text{diag}\left(\frac{1}{s_1}, \frac{1}{s_2}, \dots, \frac{1}{s_p}\right) \quad (2)$$

其中: $\bar{X} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p)'$ 为过程 $X$ 的样本均值; $s_i$ 为 $X_i$ 的样本标准差。

对标准化后的数据 $X_s$ 进行主成分分析,可以得到 $X_s = t_1u'_1 + t_2u'_2 + \dots + t_pu'_p$ 。假如用前 $m$ 个( $m < p$ )主成分来代表原始数据中的主要变化,即可以得到如下主成分模型:

$$X_s = t_1u'_1 + t_2u'_2 + \dots + t_pu'_p + E = \hat{X} + E \quad (3)$$

### 2.1 SPE 控制图

对于主成分 $SPE$ 控制图上的打点值实际上为主成分模型在第 $i$ 个子组样本的平方预测误差,其统计量为 $SPE = \sum_{j=1}^p (X_{ij} - \hat{X}_{ij})^2$ ,其中: $X_{ij}$ 为第 $i$ 个子组样本的第 $j$ 个变量的观测值; $\hat{X}_{ij}$ 为第 $i$ 个子

组样本的第  $j$  个变量的模型预测值。对于第  $i$  个子组样本,其打点值为

$$Q_i = e_i e_i' = X_i (I - U_m U_m') X_i' \quad (4)$$

这里的  $e_i$  是误差阵  $E$  的第  $i$  行,且  $U_m = (u_1, u_2, \dots, u_m)$ 。

其实统计量  $SPE$  在第  $i$  时刻的值只是一个标量,主要刻画的是第  $i$  时刻的过程观测值  $X_i$  与模型的偏离程度。因通过多个变量综合形成  $e_i$ ,所以  $SPE$  控制图可以对多变量过程  $X$  中的多个相关特性进行监控,这有别于其他一般的多变量质量控制。如取置信水平  $\alpha$ ,可以得到  $SPE$  控制图的控制线为:

$$Q_\alpha = \theta_1 \left[ \frac{u_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{h_0} \quad (5)$$

其中:  $\theta_i = \sum_{j=m+1}^p \lambda_j^i$ ,  $h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^3}$ ;  $\lambda_j (j = m+1, \dots, p)$  为过程  $X$  的协方差阵的  $(p-m)$  个特征值;  $p$  是全部的主成分个数;  $m$  是选取主成分的个数;  $\mu_\alpha$  是正态分布置信区间下的一个临界值。

## 2.2 $T^2$ 控制图

对于  $T^2$  控制图,首先建立多变量过程  $X$  的主成分模型,然后通过计算  $T^2$  统计量来对多个主成分进行监控。对于第  $i$  个子组样本  $X_{(i)}$ ,有统计量  $T_i^2 = t_i \Lambda_m^{-1} t_i' = X_i U \Lambda_m^{-1} U' X_i'$ ,其中:  $t_i$  由主成分模型的第  $i$  个主成分的得分向量组成,  $\Lambda_m$  是由与前  $m$  个主成分所对应的特征值而组成的对角矩阵。主成分  $T^2$  控制图主要通过模型内部的主成分向量模的波动来反映多变量过程  $X$  的变化情况。对于主成分模型的  $T^2$  控制图的控制线可以表示为:

$$T_{m,n,\alpha}^2 = \frac{m(n-1)}{n-m} F_{m,n-1,\alpha} \quad (6)$$

## 2.3 贡献控制图

当前面所计算的  $SPE$  统计量和  $T^2$  统计量超出其控制线时,可以判定多变量  $X$  出现了不正常情况,但从上述的 2 种控制图中无法得知到底哪里出了问题。对此,贡献控制图的研究和发展可以解决这个问题。主成分贡献控制图可以分析每个过程分量对  $SPE$  统计量的贡献大小,并确定哪些过程分量引起了多维过程变量  $X$  的不正常变化情况。

通常,过程  $X$  的第  $i$  个子组样本的第  $j$  个分量对  $SPE$  统计量的贡献可以表示为  $Q_{ij} = e_{ij}^2 = (X_{ij} - \hat{X}_{ij})^2$ ,而第  $i$  个子组样本的第  $j$  个分量对第  $m$  个主成分的贡献为  $X_{ij} u_{mj}$ ,其中,  $u_{mj}$  是  $u_m$  的第  $j$  个元素。最后通过对  $X$  过程的分量的主成分贡献控制进行分析,得知过程  $X$  的哪些分量的变化引起  $SPE$  统计量超出了控制线。

## 3 实例研究

本文针对中石化中甲烷液化气体的局部蒸馏过程  $X$  的输出建立多变量过程监控与诊断。蒸馏过程的好坏直接影响到甲烷气体的质量,然而对蒸馏过程效果不好评定,一般使用仪表板上的数据来检测整个过程的效果。本文选取塔顶的温度、入口的流速、塔顶回流的速度、再蒸馏的温度这 4 个过程变量作为主要检测数据指标。蒸馏过程  $X$  由下述变量构成:  $X_1$  为塔顶的温度 ( $^{\circ}\text{C}$ );  $X_2$  为入口的流速 ( $\text{L/h}$ );  $X_3$  为塔顶回流的流速 ( $\text{L/h}$ );  $X_4$  为再蒸馏的温度 ( $^{\circ}\text{C}$ )。

根据经验,塔顶的温度、入口的流速、塔顶回流的流速以及再蒸馏的温度的变动最能反映蒸馏过程的效果,从而影响气体质量;另外,设备或传感器的失效或故障所引起的偶然事件也需要及时报警和排除。因此,需要应用主成分质量控制方法对蒸馏过程实施统计监控,以便及时发现过程中的各种异常变化,并诊断出引起异常失控的原因。

对上述的 4 个变量过程进行监控,让生产过程正常运行一段时间。取正常运行阶段及故障阶段 25 个时刻(总 100 个数据)建立 PCA 模型,其中前 13 个时刻的样本点数据为正常运行阶段观测采集的,而后面 12 个时刻的样本点数据为故障阶段观测采集的。

### 3.1 主成分模型的建立

本文利用前 13 个时刻的样本点数据进行主成分分析<sup>[8]</sup>,首先对原始数据进行标准化处理,即可得到表 1。

针对表 1 中的数据建立主成分模型,在此利用 SAS 软件进行主成分分析。经过运行程序<sup>[9]</sup>得

到表2。

表1 实例中标准化后的数据

序号	$X_1$	$X_2$	$X_3$	$X_4$
1	-0.078	-1.423	-0.900	1.792 3
2	-1.098	-1.230	0.504 4	1.314 4
3	0.601 5	0.504 2	-0.588	-0.597
4	0.601 5	-1.102	-0.588	1.015 6
5	-0.078	0.247 2	-0.900	0.179 2
6	0.601 5	0.440 0	-0.432	-0.477
7	-0.758	1.468 2	0.816 7	-1.433
8	-1.098	-1.102	1.597 3	0.836 4
9	-0.928	0.375 7	0.972 8	-0.477
10	2.301 5	-0.074	-1.213	-0.239
11	-1.098	-0.524	1.753 4	0.239 0
12	0.601 5	1.146 9	-0.432	-1.075
13	0.431 5	1.275 4	-0.588	-1.075

表2 相关阵的特征和特征向量

	特征值	差值	比值	累积值
1	2.235 708 51	0.659 644 79	0.558 9	0.558 9
2	1.576 063 72	1.389 459 35	0.394 0	0.352 9
3	0.186 604 37	0.184 980 97	0.046 7	0.999 6
4	0.001 623 40		0.000 4	1.000 0
	特征向量			
$x_1$	0.475 959	-0.508 97	0.675 489	0.241 081
$x_2$	0.563 863	0.413 943	-0.314 440	0.641 745
$x_3$	-0.394 069	0.604 968	0.637 687	0.268 475
$x_4$	-0.547 934	-0.451 229	-0.195 440	0.676 730

从表2可以看出:前2个主分量的累计贡献率达95.29%,远高于85%。因此,这里只需用前2个主分量就能更好地概括大部分信息。

根据表2可以得到负荷向量矩阵为:

$$U = \begin{pmatrix} 0.475 959 & -0.508 98 & 0.675 489 & 0.241 081 \\ 0.563 863 & 0.413 948 & -0.314 44 & 0.641 745 \\ -0.394 07 & 0.604 968 & 0.637 687 & 0.268 475 \\ -0.547 93 & -0.451 23 & -0.195 44 & 0.676 73 \end{pmatrix}$$

从而有

$$U_2 = \begin{pmatrix} 0.475 959 & 0.563 863 & -0.394 07 & -0.547 93 \\ -0.508 98 & 0.413 948 & 0.604 968 & -0.451 23 \end{pmatrix}$$

根据负荷向量矩阵得出前2个主成分为:

$$PC_1 = X \times u_1 = 0.475 9X_1 + 0.563 8X_2 - 0.3940X_3 - 0.547 9X_4$$

$$PC_2 = X \times u_2 = -0.508 9X_1 + 0.413 9X_2 + 0.604 9X_3 - 0.451 2X_4$$

在此,根据式(4)利用Matlab软件计算出主成分SPE控制图的 $Q_i$ 打点值见表3。

表3 主成分SPE控制图的打点值

样本序号	$Q_i$	样本序号	$Q_i$
1	2.61	8	1.21
2	1.15	9	1.06
3	5.62	10	4.31
4	0.97	11	2.97
5	3.50	12	1.93
6	0.12	13	1.34
7	0.21		

从表2中可得4个特征值分别为 $\lambda_1 = 2.235 7, \lambda_2 = 1.576 1, \lambda_3 = 0.186 6, \lambda_4 = 0.001 6$ 。

由此得出

$$\theta_1 = \sum_{j=3}^4 \lambda_j = 0.188 2, \theta_2 = \sum_{j=3}^4 \lambda_j^2 = 0.034 8$$

$$\theta_3 = \sum_{j=3}^4 \lambda_j^3 = 0.006 5, \theta_4 = \sum_{j=3}^4 \lambda_j^4 = 0.326 5$$

然后根据式(5),给定置信水平 $\alpha = 0.01$ ,有 $\mu_{0.01} = 2.325$ ,由此可得

$$Q_{0.01} = \theta_1 \left[ \frac{u_{0.01} \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} = 5.99$$

对于主成分 $T^2$ 控制图的控制线根据式(6)可得

$$T_{2,13,0.01}^2 = \frac{12 \times 2}{11} F_{2,12,0.01} = 0.130 9$$

在此利用Matlab软件计算出主成分 $T^2$ 控制图中的 $T_i^2$ 打点值见表4。

相应地,运用主成分控制研究的方法利用R软件绘制图1~3。从图1中可明显看出:第9、第7个样本点离第一主成分较近;第2、4、6、11样本点离第二主成分较近;第10个样本点偏离第一、

第二主成分较远。

表4 主成分  $T^2$  控制图的打点值

样本序号	$T_i$	样本序号	$T_i$
1	0.052 11	8	0.089 33
2	0.028 84	9	0.050 93
3	0.085 12	10	0.044 72
4	0.089 20	11	0.065 79
5	0.065 62	12	0.064 76
6	0.013 44	13	0.11
7	0.011 99		

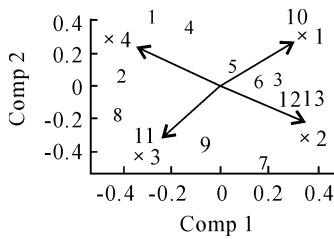


图1 原始数据分别在第一、第二主成分下的散点图

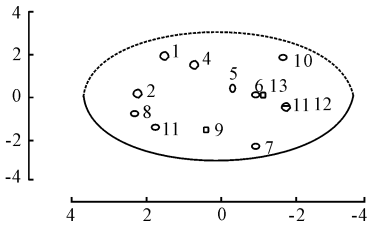


图2 原始数据主成分椭圆质量控制图

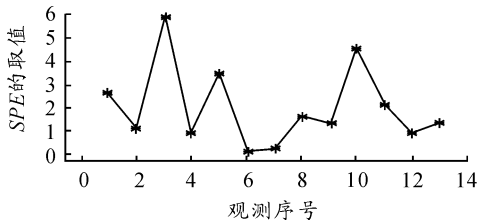


图3 原始数据的  $SPE$  控制图

在此绘制出主成分椭圆控制图和  $SPE$  控制图来检测原始数据的观测值是否都处于受控状态。从图2可明显看出:原始数据值都在控制限内。由图3可知:13组  $Q_i$  点都落在控制线内,所以可认为多变量蒸馏过程  $X$  处于统计受控状态,所建立的主成分模型有效,于是可以用  $Q_\alpha$  作为控制用控制图的控制线,用于故障监控与诊断。

### 3.2 故障监控与诊断

现在不仅已知13组原始数据都处于受控状

态,还可认为所建立的主成分模型成立可用于监控。为了找出12组失控的新观测值,可通过本文建立的主成分模型来检测,从而确定引起的失控的变量。

通过 Matlab 软件绘制出控制图,见图4、5。

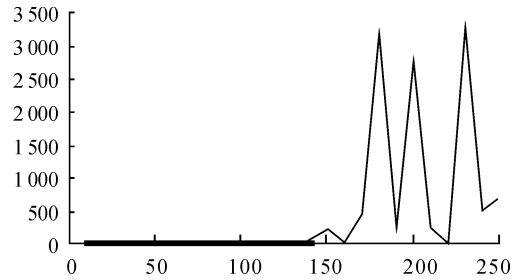


图4 对应于监测的  $SPE$  控制图

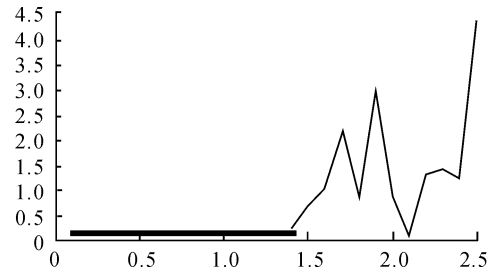


图5 对应于监测的  $T^2$  控制图

图4(粗线代表控制线,细线代表新观测值的  $SPE$ )和图5(粗线代表控制线,细线代表新观测值的  $T^2$ )分别是新采集的12组观测数据的  $SPE$  控制图和  $T^2$  控制图。对于图4、5的控制线分别为  $UCL = Q_{0.01} = 5.99$  和  $UCL = T^2 = 0.130$ 。很容易看出:在后面提供的12组样本点中,从第14个采样时刻点开始, $Q$  统计量以及  $T^2$  统计量都严重超出了其控制线。在图4、5中,因图形尺度的关系, $Q$  统计量以及  $T^2$  统计量的控制限均与横坐标重合,所以可判定过程出现异常,表现在第20个采样时刻点远远偏离了由前面数据所建立的主成分模型,因此,在该点上绘制变量的主成分贡献图见图6。

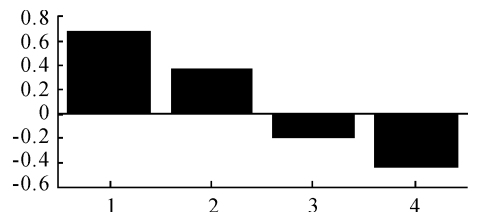


图6 第20时刻样本点的主成分贡献图

图6表示了各个过程变量在第20个采样时刻对前2个主元的贡献图。由于变量的变化程度与相对应的主元负荷一起决定每个变量对主元的贡献图,则从图6可以看出:第1个变量即塔顶的温度和第4个变量即再蒸馏的温度对前2个主元的贡献最大。所以有理由判定在蒸馏过程中出现异常与塔顶的温度有很大关系。结合实际工业生产过程知:在蒸馏过程中塔顶温度异常是由塔顶回流引起的,所以要降低回流温度,以防引起塔顶温度失控,影响整个生产过程。

表5 主元和变量间的相关系数

	Prin1	Prin2	Prin3	Prin4
$X_1$	0.648	0.572	0.808	0.267
$X_2$	0.367	0.381	0.193	-0.56
$X_3$	-0.23	-0.02	0.183	0.652
$X_4$	-0.57	-0.47	-0.52	0.432

表5为输出主元与各个变量间的相关系数。以主元Prin1为例,由表5可以看出:该主元与第1和第4个变量(即塔顶的温度和再蒸馏的温度)密切正相关,因此可以用温度来解释该主元。

## 4 结束语

本文将主成分质量控制方法用于实际甲烷气体生产过程中的监控诊断,能够及时判断生产过程中的异常情况,并精确查找出引起异常情况的过程变量<sup>[10]</sup>。本文利用主成分质量控制对工业生产过程进行质量监控,取得了较好的效果。一方面为该方法基于统计降维思维的过程控制和诊断提供了新的思路<sup>[11]</sup>;另一方面也为解决工业生产质量监控提供了一种新的方法和方向。

与常用多变质量控制方法相比,主成分分析法在质量控制监测过程中更加有效、精确。其不

足之处在于:当多个过程变量出现异常时,该方法则不理想,甚至出现错误结论,对此还需进一步研究。

## 参考文献:

- [1] 李伯奎,刘远伟.表面粗糙度理论发展研究[J].工具技术,2004,38(1):63-67.
- [2] 蒋家东,冯允成.统计过程控制[M].北京:中国质检出版社,2011.
- [3] 杜福洲,唐晓表.基于PCA的多元质量控制与诊断方法研究[J].制造业自动化,2006,28(8):10-18.
- [4] 陈瑞兴,程琳.基于PCA和多元控制图的大坝安全多测点数据监控方法[J].水利与建筑工程学报,2012,10(5):68-73.
- [5] 耿修林.基于主成分原理的多元质量控制的构造[J].数理统计与管理,2007,26(1):106-111.
- [6] 陈勇,梁军,陆浩.基于PCA的多变量控制系统的故障检测与诊断[J].工程设计学报,2002,9(5):257-260.
- [7] 肖枝洪,余家林.多元统计及SAS应用[M].武汉:武汉大学出版社,2013.
- [8] 苏理云,邓燕,冉雪竹,等.低合金高强度钢力学性能与化学成分之间的统计建模与分析[J].重庆理工大学学报:自然科学版,2009,23(2):36-40.
- [9] 杜福洲,唐晓青.基于PCA的多元质量控制与诊断方法研究[J].制造业自动化,2006,28(8):10-13.
- [10] 骆顺志,程越,李鑫.机械电子控制金属带式变速器速比控制及仿真[J].重庆理工大学学报:自然科学版,2011,25(8):120-126.
- [11] Evan L R, Leo H. Fault detection in industrial Processes using canonical variate Analysis and dynamic principal component analysis[J]. Chemometrics and Intelligent-Laboratory Systems, 2000, 51(8): 81-93.

(责任编辑 刘 舸)