

文章编号:1671-6833(2020)01-0044-05

基于自适应主成分分析的化工过程在线监测

吕照民<sup>1</sup>, 周 革<sup>2</sup>, 苗 晨<sup>3</sup>

(1.上海工程技术大学 城市轨道交通学院,上海 201620; 2.上海机电工程研究所,上海 201109; 3.上海工程技术大学 电子电气工程学院,上海 201620)

**摘 要:**当主成分分析(principal component analysis, PCA)应用于过程监测时,不适当的成分选择方法会导致变异特征被分散或被淹没从而影响监测性能。针对这个问题提出了成分的自适应选择方法并用于过程监测,即自适应主成分分析(adaptive principal component analysis, APCA)。自适应主成分应用于过程监测时主要包括 3 个步骤:首先,在离线建模时基于载荷矩阵通过欧氏距离计算各个成分的相似性,并基于每个成分选出与其相似性较高的成分构成多个成分子空间;其次,在线监测时基于在线样本的各成分通过核密度估计计算各个成分的变异概率,选择出变异概率最高的成分作为特征成分;最后,挑选出与特征成分对应的成分子空间,并构造  $T^2$  统计量。通过数值仿真案例和田纳西伊斯曼(tennessee eastman, TE)过程证明了提出方法 APCA 的有效性。

**关键词:**过程监测;主成分分析;子空间;自适应

**中图分类号:** TP277      **文献标志码:** A      **doi:**10.13705/j.issn.1671-6833.2019.04.006

0 引言

在现代工业中,过程监测是一种被广泛采用的提高过程安全性和产品质量的技术。由于分布式控制系统(distributed control system, DCS)的广泛应用,大量的数据被采集,多元统计过程控制(multivariate statistical process control, MSPC)方法发展迅速<sup>[1]</sup>。

PCA 作为最基本的 MSPC 方法,已有很多研究成果。然而,如何选择主成分并利用主成分信息进行监测仍然是一个开放的问题。主成分选择中使用最广泛的方法是累计方差贡献率法(cumulative percentage variance, CPV),选择具有较大方差的成分来构造主成分空间,剩余具有较小方差的成分构成残差空间<sup>[2]</sup>。不过这种方法不是以过程监测为目的设计出来的,故障发生时表现出的变异性不一定体现在方差信息较大的主成分,在主成分空间和残差空间都有可能体现,这会引变异特征信息分散,而且每个子空间内的大量正常信息可能淹没变异特征信息。

已经有许多研究人员专门为过程监测提出了

选择主成分的方法。Jiang 等<sup>[3]</sup>提出了一种敏感 PCA ( sensitive principal component analysis, SPCA)方法,该方法直接检测每个主成分上  $T^2$  统计量的变化率,并在线选择  $T^2$  统计量变化率较大的主成分作为敏感主成分进行过程监测。Jiang 等<sup>[4]</sup>提出了一种实时重组的 PCA 模型,通过核密度估计在线评估主成分的变异性,将具有故障检测的主成分集中到一个子空间进行监测。这两种方法客观地将大部分变异特征信息集中到一个子空间,不过只集中了变异性较大的主成分,忽略了变异性小的主成分,而部分变异性小的主成分也是有利于故障监测的。Jiang 等<sup>[5]</sup>提出了一种数据驱动的多块 PCA 方法,该方法利用互信息(mutual information, MI)自动划分块,每个变量块通过支持向量数据描述(support vector data description, SVDD)监测,并将全部块中的结果集成在一起。Tong 等<sup>[6]</sup>根据变量与主成分子空间和残差子空间的相关性或不相关性,将变量分成 4 个子空间,提出了一种 4 子空间贝叶斯(four-subspace construction and Bayesian inference, FSCB)监测方法。江伟等<sup>[7]</sup>提出了一种基于混合分块 DMICA-

收稿日期:2018-12-30;修订日期:2019-03-11  
基金项目:国家自然科学基金青年基金资助项目(61703275);上海市青年科技英才杨帆计划(18YF1409200);上海工程技术大学人才计划项目——展翅计划。  
作者简介:吕照民(1990—),男,内蒙古兴安盟人,上海工程技术大学讲师,博士,主要从事机器学习研究,E-mail: zhaomin\_lv@sues.edu.cn。

PCA 的过程监控方法。该方法先利用已知的部分过程知识对变量初步分块,再利用各分块变量之间改进的广义 Dice's 系数(modified general dice's coefficient, MGDC)进行进一步的分块。然后采用 DMICA-PCA 方法对每个子块进行建模,通过加权方法求得总的联合指标进行监测。这些多块方法通常分解过程变量空间或成分空间,以简化监测过程变量间的复杂度。

PCA 模型中载荷矩阵的向量体现了变量对主成分的影响,将其定义为对应主成分的特征向量。如果对主成分有重要影响的过程变量发生变异,相应的主成分也会具有变异特征,因此,具有相似特征向量的主成分具有相似的变异特征,将它们集中在同一子空间中监测有利于减少变异特征的分散。而在离线状态下选择的主成分不能保证具有变异特征,因此,根据在线信息选择具有变异特征的主成分作为特征成分并将与其具有相似变异特征的主成分集中监测可以提高监测效果<sup>[8]</sup>。笔者提出一种在线自适应挑选主成分的方法即自适应主成分分析方法(APCA)。首先利用载荷矩阵计算各成分与其余成分间特征向量的欧氏距离,为每个成分选择出与其欧氏距离较小的成分构成该成分的成分子空间;其次通过核密度估计计算出在线成分的变异概率,挑选变异概率最高的主成分作为 CC;最后利用 CC 对应的子空间进行监测。

1 基于 PCA 的过程监测

PCA 是应用最广泛的多变量过程监测算法。过程的数据矩阵表示为  $X \in \mathbf{R}^{n \times m}$  ( $n$  为样本个数,  $m$  为测量变量个数),先对该数据矩阵进行归一化处理,然后利用 PCA 方法建立模型,可以表示为<sup>[9]</sup>:

$$X = \tilde{T}\tilde{P}^T + \hat{T}\hat{P}^T = \tilde{T}\tilde{P}^T + E, \tag{1}$$

式中:  $\tilde{T}$ 、 $\tilde{P}$  为主成分得分和其载荷矩阵;  $\hat{T}$ 、 $\hat{P}$  为残差得分和其载荷矩阵;  $E$  为 PCA 残差矩阵;  $\tilde{T}(n \times k)$  中包含  $k$  个主成分个数,选取主成分可采用累积方差贡献率法。通过对  $X$  的协方差矩阵进行特征值或者奇异值分解可以得到主成分得分、残差得分和相应的载荷矩阵,即

$$\sum = X^T X / (n - 1); \tag{2}$$

$$\sum = [\tilde{P}\hat{P}]\Lambda[\tilde{P}\hat{P}]^T; \tag{3}$$

$$\tilde{T} = X\tilde{P}; \tag{4}$$

$$\hat{T} = X\hat{P}; \tag{5}$$

$$\Lambda = \text{diag } \lambda_i, \tag{6}$$

式中:  $\lambda_i(i = 1, 2, \dots, m)$  为协方差矩阵的  $m$  个特征值,也就是主成分得分矩阵  $\tilde{T}$  和残差得分矩阵  $\hat{T}$  每列的方差。经过 PCA 分解之后,原来的  $m$  维变量空间就被  $k$  维主成分空间和残差空间所代替。对于数据样本  $x$ ,主成分空间的  $T^2$  统计量可以构造为:

$$T^2 = x^T \tilde{P}(\Lambda)^{-1} \tilde{P}^T x. \tag{7}$$

残差空间的  $SPE$  统计量被构造为:

$$e = x - x\tilde{P}\tilde{P}^T; \tag{8}$$

$$SPE = e^T e, \tag{9}$$

式中:  $e$  是残差向量。

$T^2$  和  $SPE$  统计量的控制限为:

$$T^2_{\text{lim}} = \frac{k(n-1)}{n-k} F_{k, (n-k), \alpha}; \tag{10}$$

$$SPE_{\text{lim}} = g\chi^2_{h, \alpha}; \tag{11}$$

$$g = \frac{v}{2a}; \tag{12}$$

$$h = \frac{2a^2}{v}, \tag{13}$$

式中:  $\alpha$  表示显著水平;  $a$  和  $v$  分别表示样本  $SPE$  的均值和方差<sup>[10]</sup>。

2 基于 APCA 的过程监测

具有相似特征向量的主成分有相似的变异特征,应将其集中到一个子空间。基于 APCA 的过程监测以此建立多个成分子空间模型,根据在线信息挑选变异概率最高的主成分作为 CC,以 CC 对应的成分子空间作为 ACS(adaptive component space, ACS)构造  $T^2$  统计量进行监测。本节对所提方法进行了详细描述。

2.1 APCA 方法

2.1.1 子空间模型

假设有历史正常数据  $X \in \mathbf{R}^{n \times m}$ ,其中  $n$  是每个变量的样本数量,  $m$  是过程变量的数量。对过程数据  $X$  进行归一化处理后,利用奇异值分解得到主成分:

$$T = [t_1, t_2, \dots, t_m]. \tag{14}$$

$P \in \mathbf{R}^{n \times n}$  为载荷矩阵,载荷矩阵  $P$  每一列都是载荷向量,可以表示为:

$$P = [p_1, p_2, \dots, p_i, \dots, p_m]. \tag{15}$$

采用主成分载荷向量之间的欧氏距离  $d$  来度量相似程度;

$$d_{ij} = (p_i - p_j)^T (p_i - p_j); \tag{16}$$

$$\bar{d}_i = \frac{\sum_{j=1}^n d_{ij}}{m}, \tag{17}$$

式中:  $i = 1, 2, \dots, m; j = 1, 2, \dots, m; d_{ij}$  表示第  $i$  和第  $j$  个主成分的欧氏距离;  $\bar{d}_i$  为第  $i$  个主成分与其他主成分的欧氏距离的平均值。子空间设计结果为:

如果  $d_{ij} \leq \bar{d}_i$ ,  
那么  $t_j \in A_i, p_j \in \bar{P}_i$ , \tag{18}

式中:  $t_j$  为第  $j$  个成分;  $A_i$  为第  $i$  个子空间;  $\bar{P}_i$  为第  $i$  个子空间的载荷矩阵。

2.1.2 特征成分的选择

为选出具有变异特征的主成分,利用核密度估计每个主成分的变异概率,挑选变异概率最高的成分作为特征成分。具体如下:

对于在线数据样本  $x \in \mathbf{R}^{m \times 1}$ , 进行归一化处理后,映射到主成分空间得到在线主成分得分  $\bar{T} \in \mathbf{R}^{m \times 1}$ , 利用主成分  $T$  通过核密度估计可求出在线样本各个主成分概率密度  $\hat{f}(\bar{T}_i)$ :

$$\hat{f}(\bar{T}_i) = \frac{1}{n} \frac{1}{h\sqrt{2\pi}} \sum_{j=1}^n \exp\left(-\frac{T_{ij} - \bar{T}_i}{2h^2}\right). \tag{19}$$

根据当前样本每一主成分对应的概率密度,计算第  $i$  个成分对应的变异概率  $f_i$  和主成分的变异概率  $f$ :

$$f_i = [1 - \hat{f}(\bar{T}_i)] \times 100\%; \tag{20}$$

$$f = [f_1, f_2, \dots, f_m]. \tag{21}$$

选取变异概率最大的主成分作为特征成分  $C_c$ :

如果  $f_\alpha = \max(f_1, f_2, \dots, f_m)$ ,  
那么  $C_c = t_\alpha$ . \tag{22}

2.1.3 构造  $T^2$  统计量

在线信息的变异特征主要集中在 ACS 中,构造  $T^2$  统计量用于监测工业过程。特征成分对应的自适应子空间为  $A_\alpha, P_\alpha$  为自适应子空间  $A_\alpha$  的载荷矩阵。 $T^2$  统计量构造如下:

$$T^2 = A_\alpha^T P_\alpha (\Lambda)^{-1} P_\alpha^T A_\alpha. \tag{23}$$

2.2 方法实现

离线建模:

(1) 得到正常运行观测数据  $X \in \mathbf{R}^{n \times m}$ , 通过各变量的均值和方差对其进行归一化。

(2) 对归一化后的数据进行 PCA 分解。

(3) 利用载荷矩阵计算各成分间载荷向量的欧氏距离来表示各主成分间的相似程度。

(4) 为每个主成分挑选相似特征向量的主成

分,并建立子空间模型。

在线监测:

(1) 使用训练数据的平均值和方差来归一化当前时间点数据,并将归一化后的当前数据映射到主成分空间。

(2) 通过核密度估计法计算每个主成分对应的概率密度,选概率密度小的成分为 CC。

(3) 挑选特征成分对应的子空间作为 ACS。

(4) 计算 ACS 的  $T^2$  统计量和  $T^2$  控制限。如果  $T^2$  统计量超出控制限,则表明当前数据发生故障,发出警报;否则当前数据正常,继续运行。

基于 APCA 的化工过程监测方法流程图如图 1 所示。

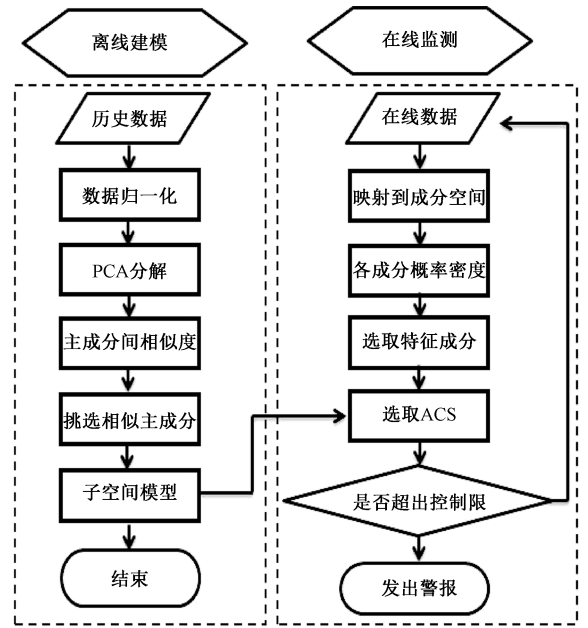


图 1 APCA 过程监测流程图

Figure 1 APCA process monitoring flow chart

3 TE 仿真实验研究

TE 过程有 22 个过程变量和 12 个操作变量,还可以仿真模拟 21 种不同的故障类型<sup>[11-12]</sup>。笔者选择其中的 33 个变量作为监测变量。为了性能评估,生成 21 个故障场景。每个故障场景有 960 个样本组成,故障场景中所有故障在 161 样本点引入<sup>[13]</sup>。

分别利用 PCA 和 APCA 方法对 TE 过程的 21 种故障进行监测,计算漏报率,并与 SPCA 和 FSCB 进行比较,结果如表 1 所示,其中最小的漏报率已用黑体标出。其中,故障 1、2、4、6、7、8、12、13、14 中每种监控方法的检测率几乎相同,因为故障幅度非常大,以至于所有这些方法都可以检测到几乎 100% 的故障。而故障 3、9 和 15 监测

难度太大,因此,这些故障在本研究中没有被考虑。对于故障 5、10、16、19、20,基于 APCA 的方法监测效果相对于其他方法有显著的提升。具体地,故障 5、16、19 的详细监测效果分别在图 2、3、4 中给出。从图中可以看出,APCA 对这些故障有更高的灵敏度。

表 1 TE 过程 APCA、PCA、SPCA 和 FSCB 漏报率比较  
Table 1 Comparison of missing alarm rates of APCA, PCA, SPCA and FSCB in TE process

编号	APCA	PCA	PCA	SPCA	FSCB
	$T^2$	$T^2$	$SPE$	$T^2$	$BIC$
1	0.001	0.007	<b>0</b>	<b>0</b>	0.002 5
2	0.008	0.017	0.01	<b>0.006</b>	0.018 7
3	0.806	0.937	0.966	<b>0.734</b>	—
4	<b>0</b>	0.688	<b>0</b>	<b>0</b>	<b>0</b>
5	<b>0</b>	0.722	0.755	0.578	<b>0</b>
6	<b>0</b>	0.006	<b>0</b>	0.004	<b>0</b>
7	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
8	0.012	0.026	0.058	<b>0.011</b>	0.021 2
9	0.811	0.947	0.973	<b>0.748</b>	—
10	<b>0.078</b>	0.543	0.585	0.343	0.186 2
11	0.133	0.518	0.233	<b>0.095</b>	0.280 0
12	<b>0.001</b>	0.015	0.065	0.003	0.002 5
13	<b>0.033</b>	0.057	0.047	0.040	0.052 5
14	<b>0</b>	0.005	<b>0</b>	<b>0</b>	0.001 2
15	0.717	0.915	0.940	<b>0.671</b>	—
16	<b>0.040</b>	0.700	0.583	0.406	0.135
17	0.025	0.200	0.043	<b>0.023</b>	0.056 2
18	0.087	0.101	0.097	<b>0.074</b>	0.102 5
19	<b>0.07</b>	0.855	0.765	0.495	0.168 7
20	<b>0.087</b>	0.575	0.435	0.193	0.196 2
21	0.353	0.593	0.466	<b>0.336</b>	0.527 5

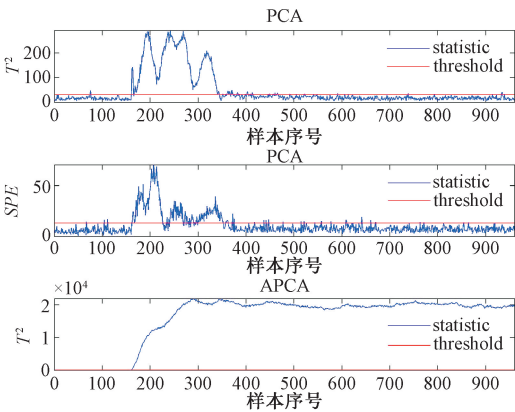


图 2 TE 过程中 PCA 和 APCA 对故障 5 的监测效果  
Figure 2 Monitoring results of PCA and APCA for case 5 in TE process

4 结论

笔者提出一种在线自适应挑选主成分的过程

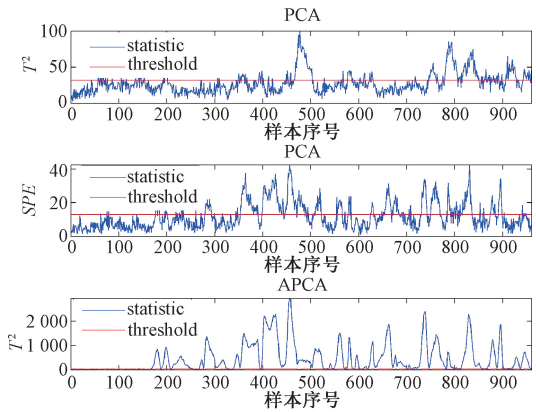


图 3 TE 过程 PCA 和 APCA 对故障 16 的监测效果  
Figure 3 Monitoring results of PCA and APCA for case 16 in TE process

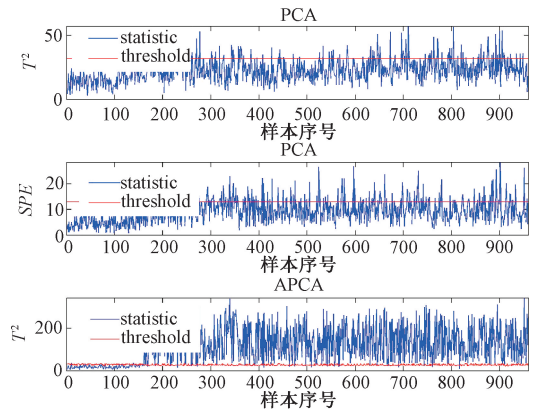


图 4 TE 过程 PCA 和 APCA 对故障 19 的监测效果  
Figure 4 Monitoring effect of PCA and APCA on fault 19 in TE process

监测方法——APCA。该方法离线建模时为每个成分选出与其相似性较高的成分构成多个成分空间,在线监测时实时选出变异概率最高的成分作为特征成分,以特征成分对应的子空间进行监测。所提方法实时进行成分的自适应挑选,降低了变异特征被分散和淹没的风险,同时可以有效地集中有利于故障检测的变异特征。通过数值仿真案例和 TE 过程案例证明了 APCA 的漏报率较低,具有较好的监测性能。未来的工作可以专注于开发一种确定主成分数量的策略,并使所提出的方法适用于非线性和非高斯过程。

参考文献:

[1] GE Z Q, SONG ZH, GAO F R. Review of recent research on data-based process monitoring[J]. Industrial & engineering chemistry research, 2013, 52 (10): 3543–3562.

[2] JACKSON J E. A user's guide to principal components [M]. New York: John Wiley & Sons, 2005.

[3] JIANG Q C, YAN X F, ZHAO W X. Fault detection



- and diagnosis in chemical processes using sensitive principal component analysis [J]. Industrial & engineering chemistry research, 2013, 52 (4): 1635–1644.
- [4] JIANG Q C, YAN X F. Just-in-time reorganized PCA integrated with SVDD for chemical process monitoring [J]. AIChE journal, 2014, 60(3): 949–965.
- [5] JIANG Q C, YAN X F. Plant-wide process monitoring based on mutual information-multiblock principal component analysis [J]. ISA transactions, 2014, 53(5): 1516–1527.
- [6] TONG C D, SONG Y, Yan YAN X F. Distributed statistical process monitoring based on four-subspace construction and bayesian inference [J]. Industrial & engineering chemistry research, 2013, 52 (29): 9897–9907.
- [7] 江伟, 王振雷, 王昕. 基于混合分块 DMICA-PCA 的全流程过程监控方法 [J]. 化工学报, 2017, 68 (2): 759–766.
- [8] 吕照民. 基于数据驱动子空间设计的间歇过程监测 [D]. 上海: 华东理工大学, 2017.
- [9] 葛志强. 复杂工况过程统计监测方法研究 [D]. 杭州: 浙江大学, 2009.
- [10] 葛志强, 宋热环. 基于 PICA 的过程监控方法 [J]. 化工学报, 2008, 59(7): 1665–1670.
- [11] 童楚东, 蓝艇, 史旭华. 基于互信息的分散式动态 PCA 故障检测方法 [J]. 化工学报, 2016, 67(10): 4317–4323.
- [12] 王振雷, 江伟, 王昕. 基于多块 MICA-PCA 的全流程过程监控方法 [J]. 控制与决策, 2018, 33(2): 269–274.
- [13] JIANG Q C, YAN X F. Plant-wide process monitoring based on mutual information-multiblock principal component analysis [J]. ISA transactions, 2014, 53(5): 1516–1527.

## Online Monitoring of Chemical Process Based on Adaptive Principal Component Analysis

LYU Zhaomin<sup>1</sup>, ZHOU Ge<sup>2</sup>, MIAO Chen<sup>3</sup>

(1.School of Urban Rail Transportation, Shanghai University of Engineering Science, Shanghai 201620, China; 2.Shanghai Electro-mechanical Engineering Institute, Shanghai 201109, China; 3.School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

**Abstract:** When Principal Component Analysis (PCA) was applied to process monitoring, improper component selection method would cause variation characteristics to be dispersed or submerged, thus affecting monitoring performance. In order to solve this problem, An adaptive selection method of components called Adaptive Principal Component Analysis (APCA) was proposed and applied it to process monitoring. The application of adaptive principal components to process monitoring mainly included three steps. Firstly, the similarity of each component was caculated based on the load matrix through Euclidean distance during offline modeling, and components with high similarity to each component was selected to form multiple molecular spaces. Secondly, during on-line monitoring, the variation probability of each component was calculated by kernel density estimation based on each component of the on-line sample, and the component with the highest variation probability was selected as the characteristic component. Finally, the molecular space corresponding to CC was selected and statistics were constructed. The result of numerical simulation and Tennessee Eastman (TE) process proved the effectiveness of the proposed APCA.

**Key words:** process monitoring; principal component analysis; subspace; adaptive