

一种基于相对主元分析的故障检测方法

欧阳高强 曹柳林*

(北京化工大学 信息科学与技术学院, 北京 100029)

摘要: 通过主元分析方法进行主元提取时协方差矩阵特征值有时会出现变化“均匀”缺陷,并且将该方法运用于故障检测时,通常使用单一的性能指标 T^2 指标或 SPE 指标作为检测判据,容易遗漏一些重要信息,降低故障的检测效果。针对这一状况,本文研究了相对主元分析方法,并且将 T^2 指标和 SPE 指标有机融合成综合指标,结合 TE 过程进行故障检测。通过单一指标 SPE-时间图和综合指标-时间图的对比,发现综合指标图比 SPE 图报警时间早、误报少,说明了将综合指标运用于相对主元分析方法进行故障检测的优越性及有效性。

关键词: 相对主元; 故障检测; 综合指标

中图分类号: TP206+.3

引言

主元分析方法(PCA)被广泛运用于工业系统的故障检测中,并已取得很大成效^[1]。但这种方法仍存在一定缺陷,使得对故障的检测有时不太准确。由于传统的主元分析方法通过计算数据矩阵协方差的特征值确定过程的主元变量,而其协方差的数值与各变量的量纲有密不可分的关系,因此各变量的数量级而非重要性会对主元的提取造成很大影响^[2]。虽然在主元分析建模中对变量的数据矩阵进行了标准化处理,消除了各变量量纲间的差异,但这种标准化处理存在着使新的协方差阵特征值被平均化的趋势,因而难以提取有代表性的主元^[3]。为此,文成林等^[4]提出了相对主元分析方法(RPCA),对在标准化处理后的变量前增加相应的权值,既保持了变量量纲的归一,又使各变量的重要性得到体现,因此可以改善故障检测过程中难以提取有代表性主元的问题。

本文以相对主元分析方法为基础,进一步对其进行改进,将主元分析方法中通常用于检测系统故障的 T^2 指标和 SPE 指标进行有机融合,组成综合指

标对系统进行故障检测,并结合 TE 过程进行仿真。

1 相对主元分析方法

1.1 数据预处理

假设 X 为数据矩阵,其中 m 为样本个数, n 为变量个数。首先,对数据矩阵进行去噪处理,本文使用小波去噪对数据进行去噪,以减小噪声对检测结果的影响。然后,对数据矩阵进行标准化处理,即对每个变量减去均值后再除以其标准差,公式如式(1)所示。

$$X_i^* = \frac{X_i - E(X_i)}{\sqrt{\text{Var}(X_i)}} \quad (1)$$

1.2 相对化变换

相对化变换是指根据工业系统的先验知识,在每一个过程变量前面乘以一个相应的权值,其算法如式(2)所示。

$$X^R = X \cdot W \quad (2)$$

其中, X^R 为相对化变换矩阵, $W = \text{diag}(w_1, w_2, \dots, w_n)$ 为相对化变换算子。

$$w_i = \mu_i m_i \quad (3)$$

式(3)中 μ_i 是比重因子, m_i 是标准化因子。标准化因子的选取有多种,本文用于计算标准化因子的公式(1)是最常用的一种。

1.3 相对主元模型

仿照主元分析方法,对相对化变换矩阵进行主元分析,将相对化变换矩阵分为 n 个向量外积之和。

$$X^R = t_1 p_1^T + t_2 p_2^T + \dots + t_n p_n^T \quad (4)$$

其中 t_i 为得分向量,也称为主元向量, p_i 为负荷向

收稿日期: 2013-10-15

基金项目: 国家自然科学基金(61174128);北京市自然科学基金(4132044)

第一作者: 男,1988年生,硕士生

* 通讯联系人

E-mail: caoll@mail.buct.edu.cn

量。得分向量和负荷向量都是正交的,且负荷向量的长度为 1,在式(4)两边同时乘以 p_i ,得

$$t_i = X^R p_i \tag{5}$$

实际过程中,负荷向量可以通过计算相对化变换矩阵协方差阵的特征向量来得到,继而通过式(5)获得主元向量。在主元模型的建立过程中,可以使用累计贡献分数来确定需要选取主元的个数。假设相对化变换矩阵协方差阵的特征值分别为 $\lambda_1, \lambda_2, \dots, \lambda_n$,且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$,则累计方差分数(CPV) C 的计算公式为

$$C(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \tag{6}$$

其中, k 为选取的主元的个数, n 为变量个数, CPV 的值一般选为 85%。此时, $t_1, t_2, \dots, t_k, p_1, p_2, \dots, p_k$ 构成主元空间, $t_{k+1}, t_{k+2}, \dots, t_n, p_{k+1}, p_{k+2}, \dots, p_n$ 构成残差空间。

这时,相对变换阵可近似的表示为

$$X^R \approx t_1 p_1^T + t_2 p_2^T + \dots + t_k p_k^T \tag{7}$$

1.4 比重因子的选取

相对主元分析方法通常是在假定对工业某过程具有先验知识的基础上进行的,在对过程各变量进行工艺分析后,会对过程各变量的重要性作出估计,据此可以量化变量的比重因子。然而,在无任何系统的先验知识时,也可以训练数据,获取比重因子,提高相对主元在过程监控中的代表能力。胡静^[3]在分别考虑主元空间和残差空间中各变量的敏感度之后,定义了敏感系数,并参考其求得比重因子引入到相对化变换中,进而定量分析故障在 RPCA 框架下的被检测能力。

第 i 个变量的 T^2 敏感系数 $S_T(i)$ 定义如式(8)所示。

$$S_T(i) = b_i^2 \sum_{j=1}^k \frac{p_{ji}}{\lambda_j} \tag{8}$$

第 i 个变量的 SPE 敏感系数 $S_{SPE}(i)$ 为

$$S_{SPE}(i) = b_i^2 \sum_{j=k+1}^n p_{ji}^2 \tag{9}$$

其中, b_i 为一常数,可以通过对 TE 过程进行多次仿真实验辨识出来^[3],在此不再赘述。

2 SPE 与 T^2 综合性能指标

许多应用场合大多使用单一的性能指标监控过

程^[5]。 T^2 指标在第 i 个监控时刻的计算公式如式(10)所示。

$$T_i^2 = X_i P_k \lambda_z^{-1} P_k^T X_i^T = [t_{i,1}, t_{i,2}, \dots, t_{i,k}] \lambda_z^{-1} [t_{i,1}, t_{i,2}, \dots, t_{i,k}]^T \tag{10}$$

其中, X_i 表示第 i 时刻的采样数据, P_k 表示由前 k 个负荷向量构成的矩阵, λ_z 表示由前 k 个特征值构成的对角矩阵, $t_{i,k}$ 表示第 i 时刻主元向量的选值。

T^2 指标的控制限为

$$UCL = \frac{k(m^2 - k)}{m(m - k)} F_\alpha(k, m - k) \tag{11}$$

其中, k 为选取的主元个数, m 为样本个数, $F(k, m - k)$ 表示自由度为 k 和 $m - k$ 的 F 分布, α 为分位数。

SPE 指标 S 在第 i 个监控时刻的计算公式如式(12)所示。

$$S_i = X_i (I - P_k P_k^T) X_i^T = [t_{i,k+1}, \dots, t_{i,n}] [t_{i,k+1}, \dots, t_{i,n}]^T \tag{12}$$

SPE 指标的控制限为

$$Q_\alpha = \theta_1 [A + 1 + B]^{\frac{1}{h_0}} \tag{13}$$

其中, $A = C_\alpha (2\theta_2 h_0^2)^{1/2} / \theta_1$; $B = \theta_2 h_0 (h_0 - 1) / \theta_1^2$, $\theta_1 = \sum \lambda_i, \theta_2 = \sum \lambda_i^2, \theta_3 = \sum \lambda_i^3, i = k + 1, k + 2, \dots, n, h_0 = 1 - 2\theta_1 \theta_3 / 3\theta_2^2$; C_α 为置信极限; I 为单位矩阵。

有时为了简化计算,统计量 SPE 的控制限也用式(14)表示^[6]。

$$\delta_\alpha^2 = g \chi_{h,\alpha}^2 \tag{14}$$

其中, $g = \theta_2 / \theta_1, h = \theta_1^2 / \theta_2, \chi_{h,\alpha}^2$ 表示自由度为 h 的卡方分布的上侧分位数。

检测故障时,会出现 4 种可能的结果:(a) T^2 和 SPE 的值都没有超过控制限;(b) T^2 的值超过控制限, SPE 的值未超过控制限;(c) T^2 没有超过控制限, SPE 的值超过控制限;(d) T^2 和 SPE 的值都超过了控制限。一般来说, c 和 d 两种情况被认为是由系统故障造成,而 b 情况可能是由于工况引起的,不能被认为是有故障发生,因此,许多场合使用单一的性能指标如 SPE 指标来检测故障。虽然这样的检测方法在大多数情况下能够检测出系统故障,但是也可能遗漏一些信息。

因此,本文将 T^2 和 SPE 指标融合,组成综合性性能指标 ϕ ,其表达式如式(15)所示。

$$\phi = \frac{S(x)}{\delta_\alpha^2} + \frac{T^2(x)}{\chi_{k,\alpha}^2} = x^T \varphi x \tag{15}$$

其中, $\varphi = P_k \lambda_z^{-1} P_k^T / \chi_{k,\alpha}^2 + (I - P_k P_k^T) / \delta_\alpha^2$, χ^2 为卡方分布。

T^2 指标是前 k 个得分向量共同构成的一个多变量指标,通过描述主元模型内部主元向量模的波动反映过程变量变化。而 SPE 指标反映的是当前样本点对主元模型的偏离程度,如果当前样本点与主元模型的偏离程度较大时,SPE 的值就会显得比较大,这时可以认为是系统发生了故障,但是 SPE 指标仅仅覆盖了残差子空间,而忽略了主元子空间的变化情况。因此,将描述主元子空间变化情况的 T^2 和描述残差空间变化情况的 SPE 指标融合,应该可以更加全面的检测过程故障。

3 TE 过程仿真实例

3.1 TE 过程

TE 过程^[7] 包括 5 个主要操作单元,即反应器、冷凝器、汽液分离器、循环压缩机和产品解吸塔,共 4 个反应,生成 2 种产物,包括 12 个操纵变量和 41 个测量变量(包括 22 个连续测量变量和 19 个成分测量值)。TE 过程流程图如图 1 所示,基本过程主要是 A、C、D、E 4 种反应物进行反应,得到产品 G 和 H,同时生成副产物 F,反应方程式见式(16)~(19)。

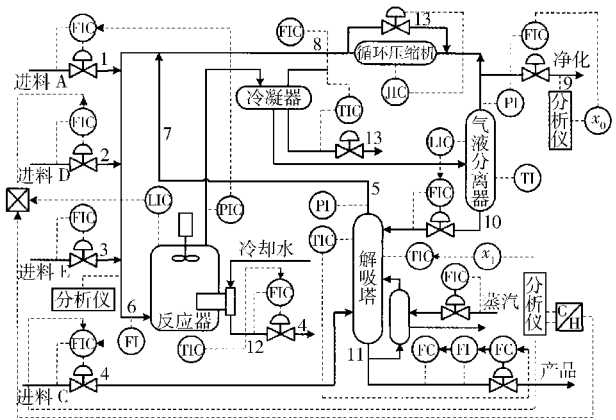
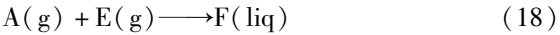
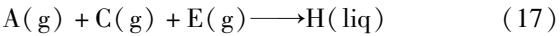
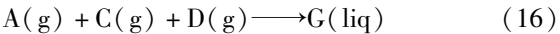


图 1 TE 过程流程图

Fig. 1 Flow chart of the TE process

TE 过程是一个大样本的复杂非线性化工系统,它包括 21 种预先设定好的故障,分别代表阶跃、随机变化、慢漂移、粘滞和恒定位置等故障类型,如表

1 所示。

表 1 过程故障
Table 1 Process faults

故障	描述	类型
IDV(1)	A/C 进料比率, B 成分不变	阶跃
IDV(2)	B 成分, A/C 进料比率不变	阶跃
IDV(3)	D 的进料温度	阶跃
IDV(4)	反应器冷却水的入口温度	阶跃
IDV(5)	冷凝器冷却水的入口温度	阶跃
IDV(6)	A 进料损失(流 1)	阶跃
IDV(7)	C 存在压力损失, 可用性降低	阶跃
IDV(8)	A、B、C 进料成分	随机变量
...		

本文在进行仿真研究时,选取了 16 个关键变量^[2]进行仿真分析,各变量表示见表 2。

表 2 进行故障检测的过程变量
Table 2 Process variables for fault detection

序号	变量	基本工况值
1	物料 A 的流量, km ³ /h	0.25052
2	物料 D 的流量, kg/h	3664.0
3	物料 E 的流量, kg/h	4909.3
4	物料 A 和 C 的流量, km ³ /h	9.3477
5	循环流量, km ³ /h	26.902
6	反应器进料流量, km ³ /h	42.339
7	反应器温度, °C	120.4
8	放空速率, km ³ /h	0.33712
9	分离器温度, °C	80.109
10	分离器压力, kPa	2633.7
11	分离器底部流量, m ³ /h	25.16
12	解吸塔压力, kPa	3102.2
13	解吸塔温度, °C	65.731
14	解吸塔上部蒸汽流, kg/h	230.31
15	反应器冷却水出口温度, °C	94.599
16	分离器冷却水出口温度, °C	77.297

3.2 仿真结果

通过 TE 过程采集数据,将未加入故障的输出数据当作历史数据,加入故障后的数据作为在线实时数据,对过程进行仿真研究,仿真时间为 20 h,每 h 采样数据为 100 个,加入故障的时间点为 10 h。加入的故障为有代表性的 IDV(1)(A/C 进料比率, B 成分不变)、IDV(2)(B 成分, A/C 进料比率不变)和 IDV(8)(A、B、C 进料成分)。

为了降低噪声对报警的影响,首先定义有效报警,即有连续 3 个以上采样周期性能指标的值超过

控制限成为有效报警。

加入故障 $IDV(1)$, 使用 SPE 指标和综合指标两种评价指标的仿真效果见图 2。

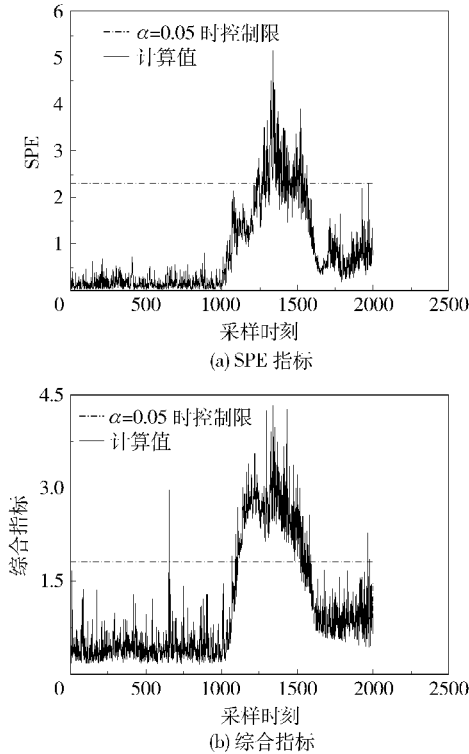


图 2 故障 $IDV(1)$ 的 SPE 指标及综合指标图
Fig.2 Fault $IDV(1)$ using the SPE index and the combined index

加入故障 $IDV(2)$, 使用两种评价指标的仿真效果见图 3。

加入故障 $IDV(8)$, 使用两种评价指标的仿真效果见图 4。

从仿真结果图可以看到, 加入故障 $IDV(1)$ 时, 综合指标的值在故障出现后迅速上升, 立刻报警, 明显优于 SPE 指标图; 加入 $IDV(2)$ 时, 两种指标对故障的检测效果都不是很理想, 但是综合指标检测故障的误报点显然要少一些; 加入 $IDV(8)$ 时, 在故障发生后 SPE 指标未能检测到故障, 而综合指标却能清晰的检测到故障。因此, 相对主元分析方法用于故障检测时, 综合指标性能要优于 SPE 指标。为了更清楚的分析 and 对比仿真结果, 统计了各个故障报警时间和误报数, 见表 3。

由于本文加入故障的时间是在 10 h, 也就是在 1000 数据点左右报警。从表 3 可以看出, 相对于使用单一 SPE 指标进行故障检测, 使用综合指标在报警时间上更快, 误报个数更少。

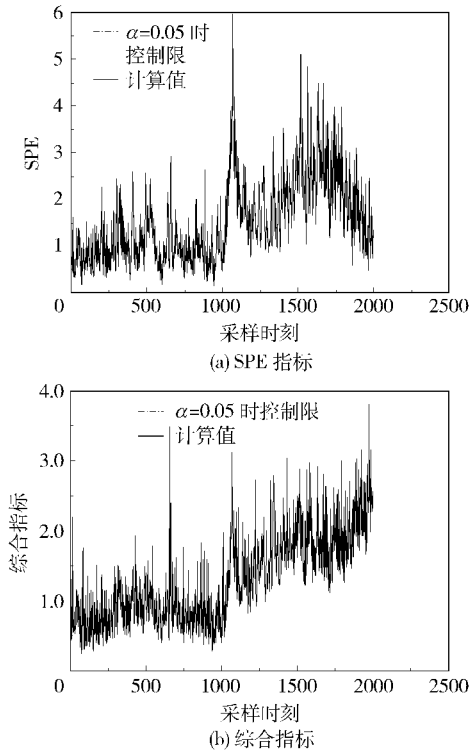


图 3 故障 $IDV(2)$ 的 SPE 指标及综合指标图
Fig.3 Fault $IDV(2)$ using the SPE index and the combined index

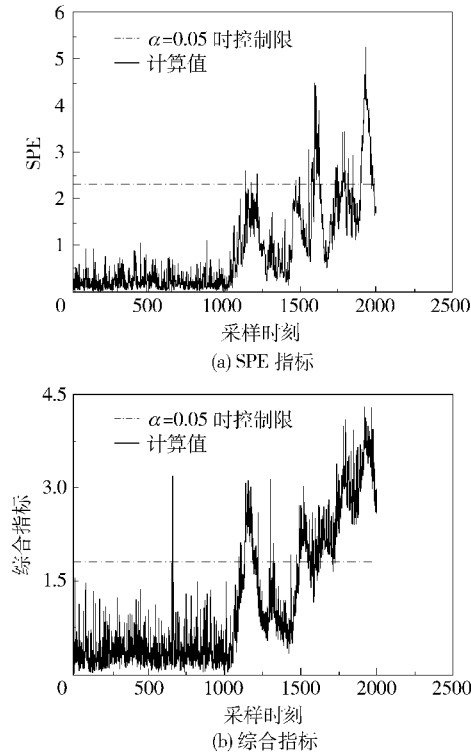


图 4 故障 $IDV(8)$ 的 SPE 指标及综合指标图
Fig.4 Fault $IDV(8)$ using the SPE index and the combined index

表3 相对主元分析方法 SPE 指标和综合指标的报警情况

Table 3 Table of RPCA alarms using the SPE index and the combined index

故障	报警时间		误报个数	
	SPE 指标	综合指标	SPE 指标	综合指标
IDV(1)	1319	1113	3	1
IDV(2)	1064	1063	2	0
IDV(8)	1595	1135	1	0

4 结论

以 TE 过程为背景,通过对相对主元分析方法检测故障的研究,将 T^2 指标和 SPE 指标进行有机融合组成综合性能指标对过程进行故障检测,并比较了单一 SPE 指标和综合指标的检测效果。由仿真结果可知,使用相对主元分析方法的综合指标进行故障检测,无论是从报警时间上还是从误报个数上都要优于传统的 SPE 指标,因此,相对主元方法使用综合指标进行故障检测具有一定的优越性。

参考文献:

[1] 王海清,宋执环,王慧. PCA 过程监控方法的故障检测行为分析[J]. 化工学报, 2002, 53(3): 297-301.
Wang H Q, Song Z H, Wang H. Fault detection behavior analysis of PCA-based process monitoring approach[J]. Journal of Chemical Industry and Engineering (China), 2002, 53(3): 297-301. (in Chinese)

[2] 许洁. 基于统计理论的工业过程性能监控与故障诊断

研究[D]. 南京: 南京航空航天大学, 2010: 23-31.
Xu J. Researches on performance monitoring and fault diagnosis for process industry based on statistical theory [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2010: 23-31. (in Chinese)

[3] 胡静. 相对主元分析理论及其运用研究[D]. 河南 开封: 河南大学, 2008: 42-46.
Hu J. Research on relative PCA theory with its application [D]. Kaifeng, Henan: Henan University, 2008: 42-46. (in Chinese)

[4] 王天真, 汤天浩, 文成林, 等. 相对主元分析方法及其在故障检测中的应用[J]. 系统仿真学报, 2007, 19(13): 2889-2894.
Wang T Z, Tang T H, Wen C L, et al. Relative principal component analysis algorithm and its application in-fault detection[J]. Journal of System Simulation, 2007, 19(13): 2889-2894. (in Chinese)

[5] Qin S J. Statistical process monitoring: basics and beyond [J]. Journal of Chemometrics, 2003, 17: 480-502.

[6] 张光明, 李柠, 李少远. 一种数据驱动的预测控制器性能监控方法[J]. 上海交通大学学报, 2011, 45(8): 1113-1118.
Zhang G M, Li N, Li S Y. A data-driven approach for model predictive control performance monitoring [J]. Journal of Shanghai Jiaotong University, 2011, 45(8): 1113-1118. (in Chinese)

[7] Downs J J, Vogel E F. A plant-wide industrial process-control problem[J]. Computers Chem Engng, 1993, 17(3): 245-255.

A fault detection approach using relative principal component analysis

OUYANG GaoQiang CAO LiuLin

(College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: The eigenvalues of the covariance matrix are almost the same in principal component analysis (PCA), and the single index T^2 or squared prediction error (SPE) has often been utilized when using relative principal component analysis (RPCA) in fault detection. However, some important information in the fault-detection process is omitted when using this single index, and incorrect detection results can be obtained. To solve this problem, a comprehensive index has been introduced, by combining the index T^2 and SPE in an effective way. Then, the combined index can be used in fault detection by relative principal component analysis in a typical Tennessee Eastman (TE) process. By comparing the simulation results in terms of the single index SPE and the combined index, it can be seen that the warning time was earlier and number of false alarms was less when using the combined index in the process of fault detection. The outcomes of this study demonstrate the effectiveness and feasibility of the method proposed.

Key words: relative principal component; fault detection; combined index

<http://www.journal.buct.edu.cn>