

基于角结构统计量的 MKECA 间歇过程故障监测*

王 普^{1,2,3,4}, 李春蕾^{1,2,3,4}, 高学金^{1,2,3,4}, 常 鹏^{1,2,3,4}, 齐咏生⁵

- (1. 北京工业大学信息学部 北京 100124; 2. 数字社区教育部工程研究中心 北京 100124;
3. 城市轨道交通北京实验室 北京 100124; 4. 计算智能与智能系统北京市重点实验室 北京 100124;
5. 内蒙古工业大学电力学院 呼和浩特 010051)

摘 要:针对间歇过程复杂非线性的特点,提出一种基于角结构统计量的多向核熵成分分析(MKECA)间歇过程监测方法。该方法首先将间歇过程数据进行标准化预处理,然后采用 KECA 提取间歇过程数据的主成分矩阵。研究表明,经过 KECA 投影后的主成分数据具有良好的角结构,因此利用主成分矩阵构造基于角结构的统计量,并且采用核密度估计算法计算其控制限。与传统的统计量相比,无需假设过程变量服从高斯分布。最后通过青霉素发酵的仿真平台和大肠杆菌实际生产过程验证,实验结果表明,相比于传统 MKPCA 方法,能够有效利用主成分的结构信息,明显降低了故障的误报率、漏报率。

关键词:核熵成分分析;角结构统计量;核密度估计;故障监测

中图分类号: TH165⁺3 文献标识码: A 国家标准学科分类代码: 510.8010

Research on batch process monitoring method based on multi-way kernel entropy component analysis and angular structure statistic

Wang Pu^{1,2,3,4}, Li Chunlei^{1,2,3,4}, Gao Xuejin^{1,2,3,4}, Chang Peng^{1,2,3,4}, Qi Yongsheng⁵

- (1. The Information Department of Beijing University of Technology, Beijing, 100124, China; 2. Engineering Research Center of Digital Community, Ministry of Education, Beijing, 100124, China; 3. Beijing Laboratory of Urban Rail Transit, Beijing, 100124, China; 4. Beijing Laboratory of Computational Intelligence System, Beijing, 100124, China; 5. School of Electric Power, Inner Mongolia University of Technology, Huhhot 010051, China)

Abstract: Aiming at monitoring the batch process with complex nonlinear characteristic, a multi-way kernel entropy component analysis (MKECA) method based on the angle structure statistic is proposed. In this method, the process data is firstly preprocessed, and then the principal component matrices of the batch process data are extracted by KECA. Research shows that KECA reveals angular structure relating to the Renyi entropy of the input space data set, and angular structure statistic is constructed using the principal component matrix structure. And then the control limits are calculated by the kernel density estimation algorithm. Finally, through the simulation of the penicillin fermentation and the actual production process of recombinant, the experiment results show that the proposed method effectively uses the structural information of the principal components compared to the traditional method of process monitoring. So error rate and false alarming rate are significantly lowered.

Keywords: kernel entropy component analysis (KZCA); angular structure statistics; kernel density estimation; fault detection

1 引 言

在工业生产过程中,间歇生产过程占据着重要地位,其应用价值在生物制药、聚合物反应等领域都有体现。

在间歇过程故障监测中,主成分分析(principal component analysis, PCA)的理论有着广泛应用^[1-3]。但是,PCA 算法需要假设生产过程是线性的,这在包含复杂非线性的间歇过程中并不满足,因此过程监测的结果并不理想。针对该问题, Scholkopf B 等人^[4]提出核主元分析(kernel

principal component analysis, KPCA), 通过把原始数据投影到高维特征空间来解决非线性问题;之后 Lee J. M. 等人^[5]将 KPCA 应用到间歇过程监测领域,与 PCA 算法相比,该方法表现出更好的监测效果^[6-8]。然而 KPCA 是以数据方差信息最大化来进行数据降维,未考虑特征向量的变化对监测结果的影响。常鹏等人^[9]提出基于核熵成分分析(kernel entropy component analysis, KECA)的间歇过程监测算法,与现存 KPCA 算法相比,KECA 方法以信息熵的大小作为选取主元的指标,而不是以特征数据方差贡献作为指标,考虑特征值和特征向量的变化对监测结果的影响。但是,采用传统的 Hotelling- T^2 和 SPE 统计量进行间歇过程监测,要求过程变量服从高斯分布的假设,而这在复杂的间歇生产过程中并不满足。

为此,本文提出一种基于角结构统计量的多向 KECA(multi-way KECA, MKECA)间歇过程监测算法。该方法首先将过程数据通过核函数映射到高维空间,然后把数据信息熵的值作为投影坐标轴来进行特征提取。因为经过 KECA 投影后的数据具有良好的角结构,构造基于角结构的统计量。采用核密度估计方法确定控制限,不需假设过程变量满足高斯分布。基于角结构的统计量能够充分利用过程数据的信息,提高了间歇过程监测性能。

2 基本理论

2.1 KECA

KECA 是 Jenssen R 在 2010 年提出的^[10]:

$$H(p) = -\log \int p^2(x) dx \quad (1)$$

式中: $p(x)$ 是样本 $D = x_1, \dots, x_N$ 服从的概率密度函数。

令 $V(p) = \int p^2(x) dx$, 需要对 $V(p)$ 进行估计。利用 Parzen 窗概率密度估计来计算 $V(p)$, 如式(2)所示。

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^N K_{\sigma}(x, x_i) \quad (2)$$

$$\text{令 } K_{\sigma}(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right), \text{ 经过公式的代换,}$$

得:

$$\hat{V}(p) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N K_{\sigma}(x_i, x_j) = \frac{1}{N^2} \mathbf{1}^T \mathbf{K} \mathbf{1} \quad (3)$$

式中: $\mathbf{1}$ 是 1 的 $N \times 1$ 的向量, 矩阵 \mathbf{K} 为核矩阵。将核矩阵进行分解, $\mathbf{K} = \mathbf{E} \mathbf{D}_{\lambda} \mathbf{E}^T$, 其中, \mathbf{D}_{λ} 为特征值 $[\lambda_1, \dots, \lambda_N]$ 组成的矩阵, \mathbf{E} 为对应的特征向量 $[\mathbf{e}_1, \dots, \mathbf{e}_N]$ 组成的矩阵^[11-12]。即:

$$\hat{V}(p) = \frac{1}{N^2} \sum_{i=1}^N (\sqrt{\lambda_i} \mathbf{e}_i^T \mathbf{1})^2 \quad (4)$$

式中: 每一个特征值和特征向量对 Renyi 熵产生不同程

度的贡献。因此在 KECA 算法中选择对 Renyi 熵贡献最大的前 l 个特征值和特征向量^[13]。得到经过投影后的数据矩阵, 即主成分矩阵如下:

$$\mathbf{M}_k = \mathbf{D}_l^{1/2} \mathbf{E}_l^T \quad (5)$$

2.2 基于角结构的统计量

CV(cosine value) 统计量的定义如下:

$$CV_k = \cos \angle(\mathbf{M}, \mathbf{M}_k) = \frac{\sum_{j=1}^l \frac{\mathbf{m}_j^T \mathbf{m}_{kj}}{\|\mathbf{m}_j\| \|\mathbf{m}_{kj}\|}}{\quad} \quad (6)$$

式中: $\mathbf{M}_k = [\mathbf{m}_{k1}, \mathbf{m}_{k2}, \dots, \mathbf{m}_{kl}]$ 是第 k 采样时刻数据的主成分矩阵, 而 $\mathbf{M} = \frac{1}{N} \sum_{k=1}^N \mathbf{M}_k$ 是 N 个时刻主成分矩阵的均值, $\mathbf{M} = [m_1, m_2, \dots, m_l]$, l 是主成分的个数。

统计量 CV 实际上是计算两个矩阵的相似度。正常操作条件下在线模型和历史模型差异不大, 相似程度高, 统计量值也大^[14]; 一旦发生故障, 两个模型相差较大, 相似度迅速减小, 表现为统计量值下降, 低于控制限, 从而检测到故障。采用核密度估计计算 CV 统计量的控制限(CK), 具体计算过程参考相关文献[13]。

3 基于角结构统计量的 MKECA 间歇过程监测

过程数据为三维矩阵 $\mathbf{X}(I \times J \times K)$ 。其中 I 表示批次数, J 表示变量数, K 表示采样点数。对三维数据矩阵 $\mathbf{X}(I \times J \times K)$ 沿时间片方向展开^[15], 如图 1 所示, 进而得到 k 个时间片矩阵 $\mathbf{X}_k(I \times J), k = 1, 2, \dots, K$ 。

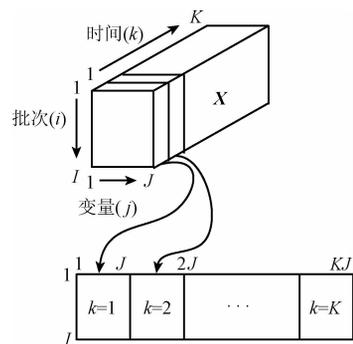


图 1 沿时间片方向展开示意图

Fig. 1 Unfolding method along time slices

3.1 基于角结构统计量的 MKECA 离线模型

先对正常运行批次的的数据作离线建模, 建立在正常操作条件下的模型, 计算出统计量以及相应的控制限, 之后对生产过程进行在线监测。

1) 将建模的三维数据沿时间片方向展开, 对数据进行标准化处理。

2)对每个时间片矩阵进行 KECA 建模,得到 N 个主成分矩阵, N 为采样点数。设定核函数和核参数,对每个预处理后的时间片矩阵计算核矩阵 \mathbf{K} ,对核矩阵进行特征值和特征向量分解,根据式(4)计算每个特征值对应的瑞利熵,根据对瑞利熵贡献的大小,选择前 l 个主成分向量,构成主成分矩阵,共有 N 个主成分矩阵。

3)确定统计量的控制限。对每个时刻 k 按式(5)计算统计量 CV_k ,采用核密度估计计算 CV 统计量的控制限 CK 。

3.2 基于角结构统计量的 MKECA 在线监测

1)在时刻 t ,采集到的时间片矩阵数 $\mathbf{X}_t(I \times J)$, $t = 1, 2, \dots, N$ 。首先对数据进行标准化,然后计算核向量 \mathbf{K}_t ;

2)按式(5)计算主成分矩阵;

3)按式(6)计算新时刻的统计量 CV_k ,并与控制限 CK 比较,如果高于控制限,表明生产过程正常,否则即为故障。

4 算法仿真应用

Pensim2.0 是青霉素发酵生产过程的仿真平台^[16],可以模拟青霉素发酵过程,通过该平台产生用于发酵过程故障监测的实验数据。

选择对过程起重要作用的 10 个变量,用于过程的建模和离线分析,选择的变量如表 1 所示。设定青霉素发酵过程需要 400 h,采样时间间隔为 1 h。总共产生 30 批次的过程数据,因此得到三维数据 $\mathbf{X}(30 \times 10 \times 400)$ 。

表 1 青霉素发酵过程变量

Table 1 Process variables of Penicillin fermentation

符号	变量名称	符号	变量名称
x_1	反应温度/K	x_6	溶解氧浓度(%)
x_2	PH 值	x_7	通风速率/($L \cdot h^{-1}$)
x_3	底物流加速率($L \cdot h^{-1}$)	x_8	搅拌速率/($r \cdot \min^{-1}$)
x_4	补料温度/K	x_9	冷水流加速度($L \cdot h^{-1}$)
x_5	菌体浓度/($g \cdot L^{-1}$)	x_{10}	产生热(cal)

将三维数据矩阵进行标准化预处理,并沿时间片方向展开为二维数组。之后对每个时间片数据矩阵进行 KECA 建模,提取主成分矩阵。最后计算 CV 统计量,并且采用核密度估计确定控制限。

对多批正常数据分别采用 MKPCA 和本文方法进行在线监测。仿真结果表明,MKPCA 的误报率为 8.3%,高于控制限设定值 5%,而本文提出的方法误报率为 1%。分析可知,对于包含复杂非线性的发酵过程而言,基于 CV 统计量的 MKECA 方法考虑了特征值和特征向

量的变化对监测结果的影响,充分利用间歇过程数据的信息,因此监测灵敏度比 MKPCA 方法要高。

图 2、3 分别是传统 MKPCA 方法和本文方法对一个批次正常过程数据的监测图。其中虚线为 99% 控制限,SPE 和 T^2 控制限的具体计算过程参考相关文献[8]。从图 2 可以看出,传统 MKPCA 算法在发酵开始阶段出现误报警现象。从图 3 可以看出,在线模型和历史模型相似度高,统计量 CV 较大,高于最低控制限,基本上没有出现误报的现象。对比发现,本文算法有更好的监测性能。

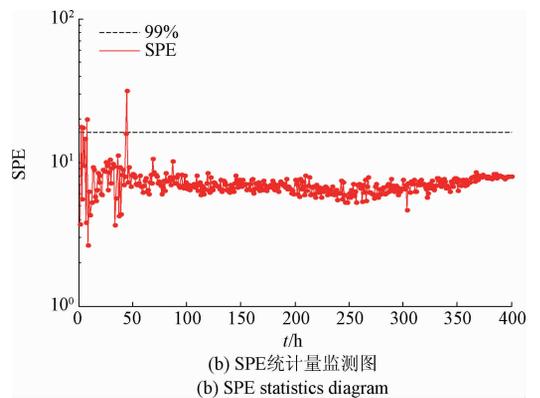
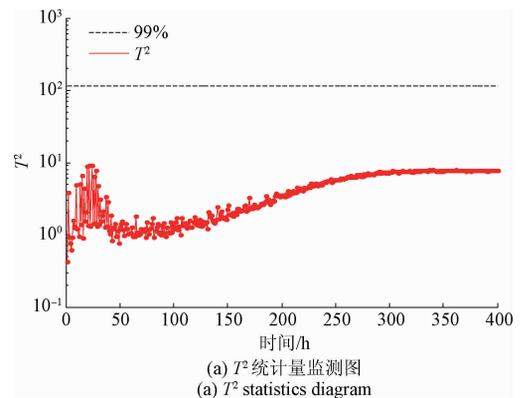


图 2 在正常操作条件下的 MKPCA 实验结果
Fig. 2 Experimental results of MKPCA in normal operating conditions

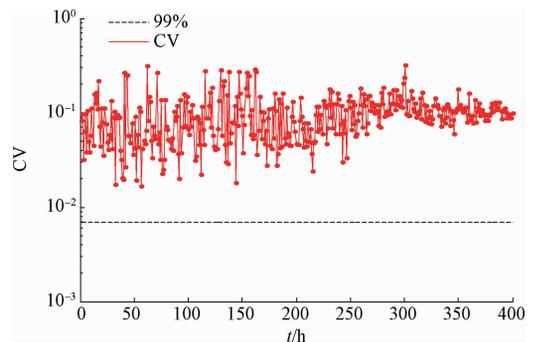


图 3 在正常操作条件下的 KECA-CV 实验结果
Fig. 3 KECA-CV results in normal operation conditions

故障 1:对通风速率引入幅值为 10% 的阶跃故障。在 200 h 施加一阶跃信号,直到反应结束。

故障 2:对底物流加速率引入幅值为 0.15 的斜坡故障。在 200 h 施加一斜坡信号,直到反应结束。

图 4 和 5 所示是两种方法在故障 1 操作条件下的监测图。从图 4 可以了解,MKPCA 方法较快地检测到故障,但在发酵过程开始阶段出现误报警现象。由图 5 可以知道,发生故障后,在线监测模型和历史模型的相似度迅速减小,CV 统计量的值下降,低于控制限,从而判断故障的发生。本文方法也能迅速地检测到故障,并且没有出现误报警的现象。分析可知,KECA 在数据特征提取上与 KPCA 算法相比更有优势,CV 统计量将 KECA 投影后数据的角度信息很好地表示出来,有利于故障监测。与 MKPCA 方法相比,本文方法监测性能更好。

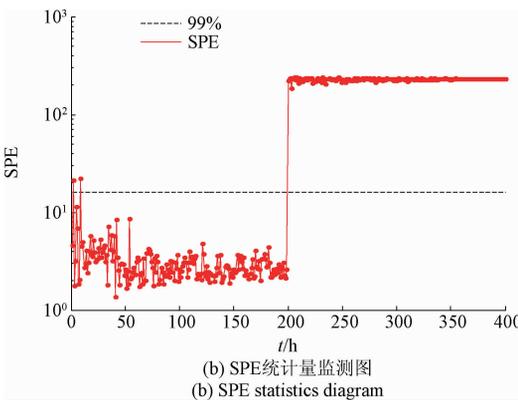
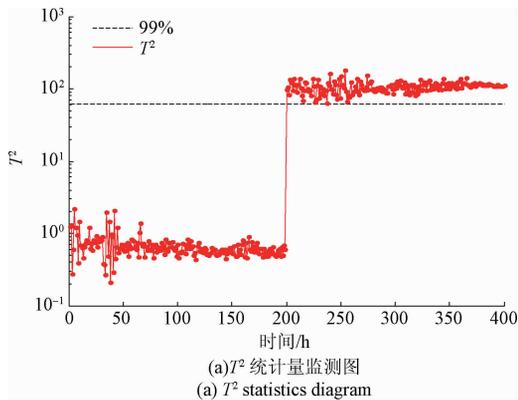


图 4 故障 1 批次 MKPCA 方法实验结果图

Fig. 4 Experimental results of fault 1 batch of MKPCA

图 6 和 7 分别为两种方法在故障 2 操作条件下的监测图。由于故障 2 为缓慢变化的斜坡故障,两种监测方法均有一定的延时性。从图 6 看出,MKPCA 的 T^2 统计量在 231 h 开始检测到故障,SPE 统计量在第 234 h 检测到故障。而本文方法在 218 h 开始检测到故障。相比之下,本文方法能更迅速地检测到故障。

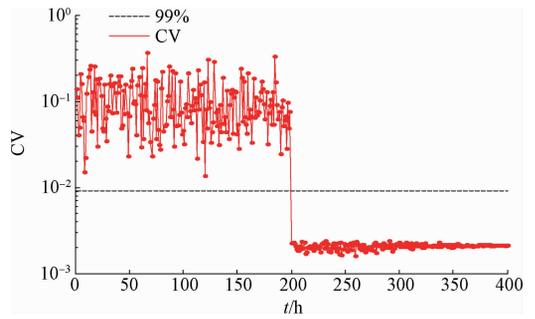


图 5 故障 1 批次的 KECA-CV 实验结果

Fig. 5 The experimental results of fault 1 batches of KECA-CV

发酵过程是一种典型的间歇过程,具有较强的非线性,并且过程变量在很大程度上不满足高斯分布。通过以上 3 组实验结果的对比可以看出,本文方法利用信息熵的大小来提取主元,考虑了特征值和特征向量的变化对监测结果的影响,这在包含复杂非线性的发酵过程是更合适的。此外,本文方法不需假设过程变量满足高斯分布,也是符合发酵过程特性,因此监测灵敏度比 MKPCA 方法要高。

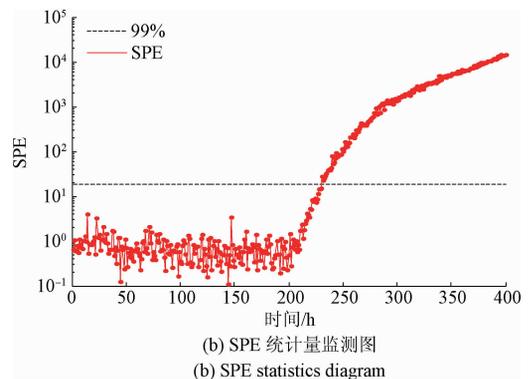
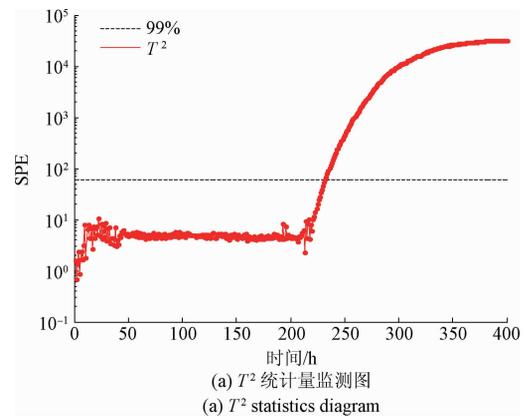


图 6 故障 2 批次的 MKPCA 方法实验结果

Fig. 6 The experimental results of fault 2 batches of MKPCA

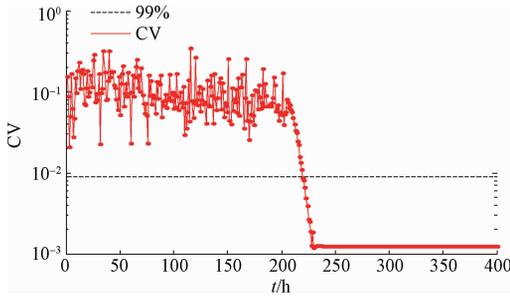


图7 故障2批次的 KECA-CV 实验结果
Fig.7 The experimental results of fault 2 batches of KECA-CV

5 算法实际验证

利用大肠杆菌制备白介素-2 的发酵过程是一个典型的间歇过程。该发酵生产过程变量之间也不是简单地服从建模要求的高斯正态分布。该生产过程的周期大约为 6~7 h,实验在北京某生物制药厂进行。在大肠杆菌实验时,选择 7 个变量,具体如表 2 所示。

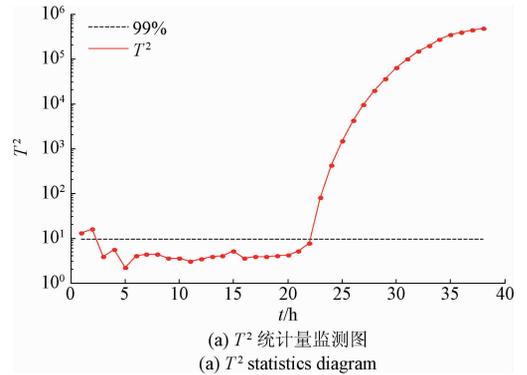
表 2 生产过程变量

Table 2 The production process variables

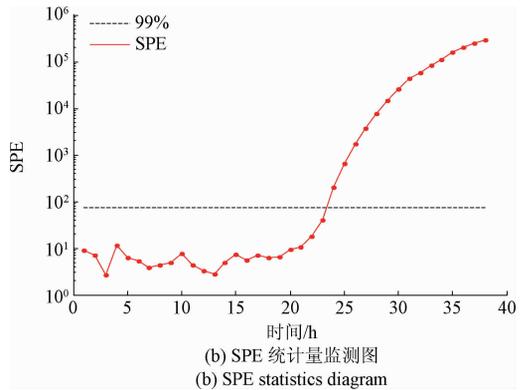
符号	变量名称
x_1	通气量/($L \cdot \min^{-1}$)
x_2	溶解氧浓度(%)
x_3	温度/ $^{\circ}C$
x_4	罐压/bar
x_5	搅拌转速/($r \cdot \min^{-1}$)
x_6	大肠杆菌浓度/($g \cdot L^{-1}$)
x_7	补葡萄糖速/($L \cdot h^{-1}$)

在实验过程中,发酵时间为 6.3 h,采样间隔为 10 min。采集 30 批正常数据用于建模,得到三维数据矩阵 $X(30 \times 7 \times 38)$ 。故障批次为搅拌速率在 3.2 h 引入幅值为 10% 的斜坡故障。

图 8 和 9 分别为在故障操作条件下的监测图。从图 8 可以看出, MKPCA 的 T^2 统计量和 SPE 统计量对故障数据的报警时刻存在较大的延迟,并且在发酵开始阶段出现误报警的现象。图 9 为本文方法的监测图,可以看出没有出现误报警,并且检测到故障的时间明显提前。通过实际发酵生产过程验证, KECA 在数据特征提取上与 KPCA 算法相比更有优势。此外, CV 统计量利用主成分数据的角度结构信息,其控制限的计算过程不需假设过程变量满足高斯分布,这在包含复杂非线性的间歇过程中是更合适的。两种方法监测结果的比较如表 2 所示。



(a) T^2 统计量监测图
(a) T^2 statistics diagram



(b) SPE 统计量监测图
(b) SPE statistics diagram

图 8 故障批次 MKPCA 方法实验结果

Fig.8 Experimental results of fault batches of MKPCA

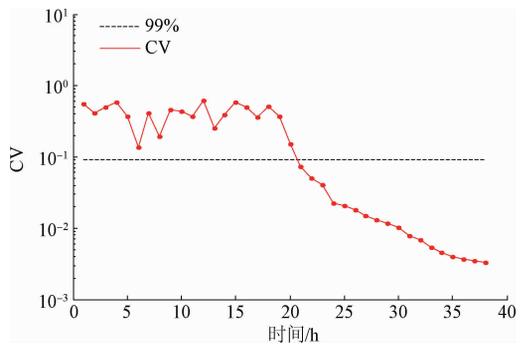


图 9 故障批次的 KECA-CV 实验结果

Fig.9 The Experimental results of fault batches of KECA-CV

表 3 两种方法监测结果对比

Table 3 Comparison of fault batch monitoring results (%)

方法	T^2/CV		SPE	
	误报率	漏报率	误报率	漏报率
传统 MKPCA	12	3.4	17	0.5
本文方法	1.5	2.8	-	-

6 结 论

在间歇生产过程中,过程数据包含非线性的特性,本文提出基于角结构统计量的 MKECA 间歇过程监测方法。该方法是基于信息熵的大小对数据信息进行提取,考虑特征向量的变化对监测结果的影响,提高了模型精度;其次,由于 KECA 投影后的数据具有良好的角结构,构造基于角结构的统计量,采用核密度估计计算控制限。与传统的 T^2 和 SPE 统计量相比,不需假设过程变量满足高斯分布,这在包含复杂非线性的间歇过程中更适合,且提高了间歇过程监测的性能。通过仿真实验和实际生产实验验证,与 MKPCA 方法相比,本文提出的方法能更迅速检测到故障的发生,并且误报率和漏报率明显降低。

参考文献

- [1] LEE J M, YOO CH K, LEE I B. Enhanced process monitoring of fed-batch penicillin cultivation using time-varying and multivariate statistical analysis [J]. *Journal of Bio-technology*, 2004, 110(2): 1119-1136.
- [2] 胡永兵, 高学金, 李亚芬, 等. 批次加权软化分的多阶段 AR-PCA 间歇过程监测 [J]. *仪器仪表学报*, 2015, 36(6): 1291-1230.
- HU Y B, GAO X J, LI Y F, et al. Multiphase AR-PCA monitoring for batch processes based on the batch weighted soft classifying [J]. *Chinese Journal of Scientific Instrument* 2015, 36(6): 1291-1230.
- [3] AGUADO D, FERRER A, FERRER J, et al. Multivariate SPC of a sequencing batch reactor for wastewater treatment [J]. *Chemometrics and Intelligent Laboratory Systems*, 2007, 85(1): 82-93.
- [4] SCHOLKOPF B, SMOLA A, MULLER K R. Nonlinear component analysis as a kernel eigenvalue problem [J]. *Neural Computation*, 1998, 10(5): 1299-1399.
- [5] LEE J M, YOO CH K, CHOI S W, et al. Nonlinear process monitoring using kernel principal component analysis [J]. *Chemical Engineering Science*, 2004, 59(1): 223-234.
- [6] VILLEZ K, YOO CH K, LEE I B. Multivariate nonlinear statistical process control of a sequencing batch reactor [J]. *Journal of Chemical Engineering of Japan*, 2006, 39(1): 43-51.
- [7] YOO CH K, LEE I B, VANROLLEGHEM P A. On-line adaptive and nonlinear process monitoring of a pilot-scale sequencing batch reactor [J]. *Environmental Monitoring and Assessment*, 2006, 119(1): 349-366.
- [8] 齐咏生, 王普, 高学金, 等. 改进 MKPCA 方法及其在发酵过程监控中的应用 [J]. *仪器仪表学报*, 2009, 30(12): 2530-2538.
- QI Y SH, WANG P, GAO X J, et al. Application of an improved multi-way kernel principal component analysis method in fermentation process monitoring [J]. *Chinese Journal of Scientific Instrument* 2009, 30(12): 2530-2538.
- [9] 常鹏, 王普, 高学金, 等. 基于核熵投影技术的多阶段间歇过程监测研究 [J]. *仪器仪表学报*, 2014, 35(7): 1655-1661.
- CHANG P, WANG P, GAO X J, et al. Research on batch process monitoring based on multistage kernel pattern entropy projection technology [J]. *Chinese Journal of Scientific Instrument*, 2014, 35(7): 1655-1661.
- [10] JENSEN R. Kernel entropy component analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(5): 847-860.
- [11] 张明咪, 吴立薪, 曾祥杰. 基于直达波和熵的一站固定式双基成像算法 [J]. *国外电子测量技术*, 2015, 34(12): 23-34.
- ZHANG M M, WU L X, ZENG X J. Algorithm based on direct signal and entropy optimization Spaceborne / fixed BISAR imaging [J]. *Foreign Electronic Measurement Technology*, 2015, 34(12): 23-34.
- [12] 崔文玲, 潘静, 何改云, 等. 基于类心和特征加权的特征选择算法 [J]. *电子测量技术*, 2015, 38(3): 25-29.
- CUI W L, PAN J, HE G Y, et al. Class center and feature weighting based feature selection algorithm [J]. *Electronic Measurement Technology*, 2015, 38(3): 25-29.
- [13] JROBERT J, ELTOFT T, MARK G, et al. Kernel maximum entropy data transformation and an enhanced spectral clustering algorithm [J]. *Advances in Neural Information Processing Systems*, 2007, 753(2): 633-640.
- [14] 王道明, 鲁昌华, 蒋薇薇, 等. 基于粒子群算法的决策树 SVM 多分类方法研究 [J]. *电子测量与仪器学报*, 2015, 29(4): 611-615.
- WANG D M, LU CH H, JIANG W W, et al. Study on PSO-based decision-tree SVM multi-class classification method. *Journal of Electronic Measurement and Instrumentation*, 2015, 29(4): 611-615.
- [15] LEE J M, YOO CH K, LEE I B. Statistical process

monitoring with independent component analysis [J].
Journal of Process Control, 2004, 14(5): 467-485.

- [16] BIROL G, UNDEY C, CINAR A. A modular simulation package for fed-batch fermentation: penicillin production [J]. Computer and Chemical Engineering, 2002, 26(11): 1553-1565.

作者简介



王普, 1988 年于中国矿业大学获得博士学位, 现为北京工业大学教授、博士生导师, 主要研究方向为工业过程控制与优化、复杂系统控制和计算机控制系统等。

E-mail: wangpu@bjut.edu.cn

Wang Pu received his Ph. D. degree from China University of Mining and Technology in 1988. He is currently a professor and Ph. D. supervisor in Beijing University of Technology. His main research fields include control and optimization of industry process, complex system control and computer control system.



李春蕾, 2014 年于洛阳理工学院获得学士学位, 现为北京工业大学硕士研究生, 主要研究方向为多元统计过程控制与故障诊断、优化控制技术。

E-mail: 15650752108@163.com

Li Chunlei received his B. Sc. degree from Luoyang Institute of Science and Technology in 2014. Now he is a post graduate in Beijing University of Technology now. His main research interests include multivariate statistical process monitoring, fault diagnosis and optimization control.



高学金 (通讯作者), 2006 年于北京工业大学获得博士学位, 现为北京工业大学副教授, 主要研究方向为复杂系统建模与控制、工业过程监控与故障诊断等。

E-mail: gaoxuejin@bjut.edu.cn

Gao Xuejin (Corresponding author) received his Ph. D. degree in 2006 from Beijing University of Technology, now he is an associate professor in Beijing University of Technology. His main research interests include modeling and control for complex system, industrial process monitoring and fault diagnosis.