第43卷 第6A期
2016年6月

计 算 机 科 学
Computer Science

Vol. 43 No. 6A
June 2016

分段多向核主元分析的啤酒发酵过程故障检测

吕 宁 颜鲁齐 白光远

(哈尔滨理工大学自动化学院 哈尔滨 150080)

摘 要 基于主元分析的故障诊断模型应用在非线性时变过程中具有局限性。基于间歇过程具有周期性这一特点,在非线性空间的数据提取中,将核变换理论引入其中,提出了一种改进的多向核主元分析故障诊断模型,该方法对于过程数据的非线性问题的解决和非线性信息的充分提取表现出很好的性能,使得非线性主元能够在高维特征空间中被快速提取。对比实验结果表明,该方法对于缓慢时变的间歇过程具有很好的准确性与实时性。

关键词 间歇过程,故障检测,多向核主元分析,分段建模

中图法分类号 TS26 文献标识码 A

Fault Detection for Beer Fermentation Process Based on Segmentation Multiway Kernel Principal Component Analysis

LV Ning YAN Lu-qi BAI Guang-yuan

(School of Automation, Harbin University of Science and Technology, Harbin 150080, China)

Abstract The fault diagnosis model based on principal component analysis has limitation in nonlinear time varying process. Based on the characteristics of the batch process, we introduced the theory of kernel transformation into the data extraction of nonlinear space, and proposed an improved fault diagnosis model based on multiple kernel principal component analysis. This method shows good performance for the nonlinear problem of process data and the full extraction of nonlinear information, where the nonlinear principal element can be rapidly extracted in the high dimensional feature space. The method was tested by comparison. The results show that the method has good accuracy and real-time performance in the process of slow time varying batch process.

Keywords Batch process, Fault detection, Multiway kernel principal component analysis, Piecewise modeling

多操作阶段是间歇过程的一个固有特性,过程中的非线性与过程的阶段往往是密不可分的[1-3]。这就导致了一个问题,通过对过程整体的运行状况进行建模分析,不能够准确地判断运行情况。过程分为多个阶段,在每一个阶段它的主导变量与过程特征都不同,由于操作过程的变化,过程之间一些变量的相互关系也无法判断是否随时间时刻变化,它呈现一定的分阶段性。所以,对于它的统计建模与在线监控,我们不仅要对整体进行建模分析,还要对每一阶段进行细致分析。

多向主元分析法(Multi-way Principal Component Analysis, MPCA)以及多向部分最小二乘法(Multi-way Partial Least Squares, MPLS)是目前针对间歇过程故障诊断的常用方法[5]。这两种方法已经在间歇过程中取得了成功的应用,但是忽视了间歇过程中的个别数据特征,间歇过程运行复杂且具有多阶段性,用单一模型去分析这个过程显得捉襟见肘,这会导致很大的误差,无法得到想要的结果。为了解决这个问题,一个可行的方法就是将整个的过程分成多个子时段,将一个复杂的阶段,分阶段进行处理,对每一个子时段进行MPCA处理、过程监测与故障诊断。

由此,本文提出一种间歇过程时段划分的方法。为了对间歇过程的时段划分准确,在这里首先对其进行一个粗划分,然后再进行细化分,粗划分是根据时间片矩阵核主元分析后的主成分个数的不同来进行划分,然后在此基础上再针对每一个时段根据负载矩阵相似度的大小进行细化分。在此基础上,本文提出一种基于子时段划分的多向核主元分析(MKPCA)模型的间歇过程监测方法,并通过对该方法的实际实验,验证了它的可行性。

1 核主元分析故障诊断模型

主元分析方法是一种线性方法,所以把其强行地应用到具有非线性特点的生产过程中显得不够准确,核函数可对不可分的非线性数据进行处理,通过映射关系将不可分数据映射到高维特征空间,通过这种映射,数据的可分性得到加强,PCA对高维空间的数据进行一个处理就能够得到非线性主元。KPCA计算过程引入了核矩阵,与PCA的求解过程相比,计算量得到了简化[3-5]。

基于KPCA方法离线建模的步骤如下:

1)采集正常操作条件下的数据,根据式(1)对其进行标准化处理:

本文受黑龙江省自然科学基金(F201222)资助。

吕 宁(1970—),男,博士,教授,主要研究方向为智能控制、故障诊断、数字图像处理等;颜鲁齐(1988—),男,硕士,主要研究方向为控制工程,E-mail:yanluqi123@126.com(通信作者);白光远(1990—),男,硕士,主要研究方向为控制工程。

$$\tilde{x}_{ij}=\frac{x_{ij}-\bar{x}(j)}{s(j)},i=1,2,\cdots,n,j=1,2,\cdots,m \tag{1}$$

其中，$\bar{x}(j)$为 $x(j)$的样本均值，$S(j)$为 $x(j)$的样本标准差。

2)考虑 m 维的正常操作条件数据 $x_k\in R^m,k=1,\cdots,N$，根据式(2)计算核矩阵 $K\in R^{N\times N}$。

$$[K]_{ij}=K_{ij}=\langle\Phi(x_i),\Phi(x_j)\rangle=[k(x_i,x_j)] \tag{2}$$

3)在特征空间中，由式(3)执行中心化处理，使得 $\sum_{k=1}^{N}\tilde{\Phi}(x_k)=0$。

$$\tilde{K}=K-1_NK-K1_N+1_NK1_N \tag{3}$$

其中，$1_N=1/N[1,\cdots 1]\in R^{N\times N}$。

4)求解 $N\lambda\alpha=\tilde{K}\alpha$，依据式(4)对 α_k 进行标准化处理，然后求取主元个数，采用的是累计方差贡献率百分比法。

$$\langle\alpha_k,\alpha_k\rangle=1/\lambda_k \tag{4}$$

5)对于正常运行状况下得到的数据，由式(5)对其进行非线性主元提取：

$$t_k=\langle v_k,\tilde{\Phi}(x)\rangle=\sum_{i=1}^{N}\alpha_i^k\langle\tilde{\Phi}(x_i),\tilde{\Phi}(x)\rangle=\sum_{i=1}^{N}\alpha_i^k\tilde{k}(x_i,x) \tag{5}$$

6)由式(6)计算正常运行情况下的 T^2 检测统计量，由式(7)计算正常运行情况下的 SPE 检测统计量。

$$T^2=[t_1,\cdots,t_p]\Lambda^{-1}[t_1,\cdots,t_p]^{\mathrm{T}} \tag{6}$$

$$SPE=\|\phi(x)-\phi_p(x)\|^2=\sum_{i=1}^{N}t_i^2-\sum_{j=1}^{p}t_j^2 \tag{7}$$

7)通过式(8)、式(9)分别确定 T^2 和 SPE 控制限。

$$T^2_{m,n,\alpha}=\frac{m(n-1)}{n-m}F_{m,n-1,\alpha} \tag{8}$$

其中，m 为主元个数 n 为样本个数，$F_{m,n-1,\alpha}$ 是检验水平为 α、自由度为 $m,n-1$ 时的 F 分布临界值。

$$\begin{cases}SPE_\alpha=\theta_1\left[\dfrac{C_\alpha\sqrt{2\theta_2h_0^2}}{\theta_1}+1+\dfrac{\theta_2h_0(h_0-1)}{\theta_1^2}\right]^{\frac{1}{h_0}}\\ \theta_i=\sum\limits_{j=m+1}^{M}\lambda_j^i(i=1,2,3)\\ h_0=1-\dfrac{2\theta_1\theta_3}{3\theta_2^2}\end{cases} \tag{9}$$

其中，C_α 是当检验水平为 α 时的正态分布临界值，λ_j 为协方差矩阵的特征值，为建模时用到的数据，m 为主元个数，M 是全部主元的个数。

基于 KPCA 方法在线检测的步骤如下：

1)采集生产过程故障数据，将其作为测试数据，根据式(1)对其进行标准化处理。

2)考虑 m 维的测试数据 $x_t\in R^m$，通过 $[k_t]_j=[k_t(x_t,x_j)]$ 计算核向量 $k_t\in R^{1\times N}$，x_j 为正常操作条件下的数据，$x_j\in R^m,j=1,\cdots N$。

3)根据式(10)中心化核向量 k_t。

$$\tilde{k}_t=k-1_tK-k_t1_N+1_tK1_N \tag{10}$$

其中，K 和 1_N 在训练时得到，$1_t=1/N[1,\cdots 1]\in R^{1\times N}$。

4)对于测试 x_t，根据式(11)对非线性主元进行提取。

$$t_k=\langle v_k,\tilde{\Phi}(x_t)\rangle=\sum_{i=1}^{N}\alpha_i^k\langle\tilde{\Phi}(x_i),\tilde{\Phi}(x_t)\rangle=\sum_{i=1}^{N}\alpha_i^k\tilde{k}_t(x_i,x_t) \tag{11}$$

5)根据式(6)、式(7)分别计算测试数据的检测统计量 T^2 和 SPE。

6)对测试数据监测统计量控制限进行判断，看是否超过。

2 分段多向核主元分析故障诊断模型

分段 MKPCA 法的建模思想：首先，将采集到的三维数据按批次方式展开，三维数据为 $X(I\times J\times K)$，其中 J 和 I 分别为变量数目和批次数目，K 为采样点数[6]。三维数据展开后的数据矩阵为 $X(I\times KJ)$，将其按照时间片分别用 KPCA 进行处理，看主成分个数是否相同，按其连续时间片进行第一步的粗划分。若相邻两片负载矩阵的相似度越大，则这两片应该属于同一阶段；反之，则这两片不属于同一阶段。定义如式(12)所示的相似度，时段粗划分以后，时段 D_s 内的两个时间片负载矩阵之间的相似性可以据此来判断。

$$d_{i,j}^{D_s}=\sum_{l=1}^{a}\gamma_l\frac{|P_{il}^{\mathrm{T}}P_{jl}|}{\|P_{il}\|\cdot\|P_{jl}\|} \tag{12}$$

其中，γ_l 为加权系数，目的是用来突出在不同的投影方向，其重要性不同，γ_l 按下式取值：

$$\gamma_l=\frac{\frac{1}{l}}{\sum_{h=1}^{a}\frac{1}{h}},l=1,2,\cdots,a \tag{13}$$

两个时间片负载矩阵间的相似度高低是根据式(12)的值是否接近 1 来进行判断的。对每片数据矩阵分别进行 KPCA 建模，依据负载矩阵的相似度大小对同一时段数据进行细化分。

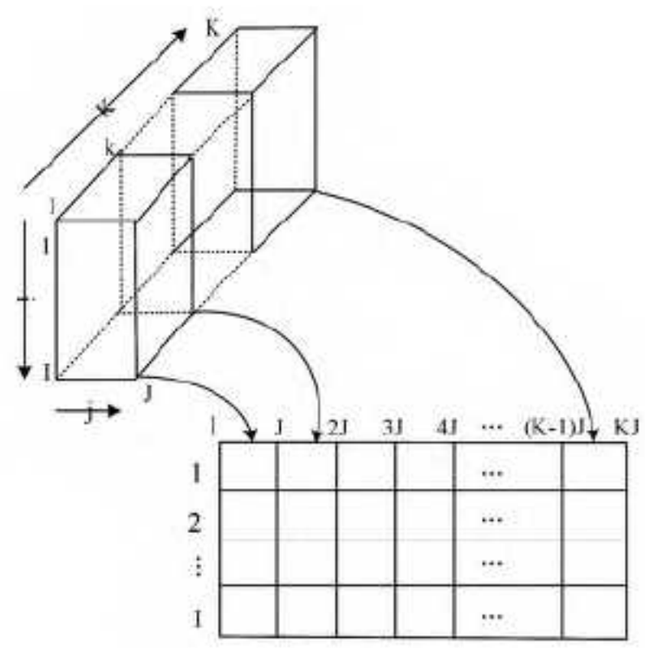

图 1 MKPCA 建模

基于分段多向核主元分析的故障监测，其离线建模的步骤如下：

1)选取 I 个历史批次数据，各批次数据构成矩阵 $X(K\times J)$ 的形式，将其横向排列成 $X(1\times KJ)$，对各批次矩阵 $X(1\times KJ)$ 中少于 KJ 列的补零，使各批次矩阵 $X(1\times KJ)$ 均有 KJ 列，多于 KJ 列的直接截取，再将矩阵 $X(1\times KJ)$ 按批次纵向形成矩阵 $X(I\times KJ)$。

2)对矩阵 $X(I\times KJ)$ 按下式执行标准化处理：

$$\begin{cases}\hat{x}_{ijk}=\dfrac{x_{ijk}-\bar{x}_{jk}}{s_{jk}}\\ \bar{x}_{jk}=\dfrac{1}{I}\sum\limits_{i=1}^{I}x_{ijk}\\ s_{jk}=\sqrt{\dfrac{1}{I-1}\sum\limits_{i=1}^{I}(x_{ijk}-\bar{x}_{jk})^2}\end{cases} \tag{14}$$

3)$\tilde{X}(I\times KJ)$ 为标准化以后的矩阵，对其建立 KPCA 模型，将每片的主元个数和负载矩阵 P_k 分别求出。建模的对象是 K 个时间片矩阵 $\tilde{X}_k(I\times J),k=1,2,\cdots,K$。

$$\begin{cases}\tilde{X}_k = T_k P_k^T + E_k \\ \hat{X}_k = T_k P_k^T \\ E_k = \tilde{X}_k - \hat{X}_k\end{cases} \tag{15}$$

各时间片主成分的个数按照 $\delta\%$ 的累计方差贡献率来确定：

$$\frac{\sum_{i=1}^{i=a}\lambda_i}{\sum_{i=1}^{i=J}\lambda_i} \geqslant \delta\% \tag{16}$$

其中，λ_i 为各时间片数据的主元，a 为主元个数。当各相邻时间片具有相同数目的主元满足式(16)时，就把这些连续的时间片划分为同一阶段。由此完成操作时段的粗划分。

4)选择时段 D_1 内的所有时间片矩阵。按式(12)和式(13)计算第 1 个时间片负载矩阵 $P_1^{D_1}(I\times a)=[P_{1,1}^{D_1}, P_{1,2}^{D_1}, P_{1,3}^{D_1},\cdots,P_{1,a}^{D_1}]$ 与第 2 个时间片负载矩阵 $P_2^{D_1}(I\times a)=[P_{2,1}^{D_1}, P_{2,2}^{D_1}, P_{2,3}^{D_1},\cdots,P_{2,a}^{D_1}]$ 之间的相似度 $d_{1,2}^{D_1}$。若满足 $d_{1,2}^{D_1}>\zeta$（ζ 为给定的阈值），则将时间片 1 和 2 归为时段 $D_{1,1}$，并计算负载矩阵 $P_1^{D_1}(I\times a)$ 和 $P_2^{D_1}(I\times a)$ 的均值负载矩阵 $\bar{P}^{D_{1,1}}=\frac{1}{2}\sum_{i=1}^{2}P_i^{D_1}$。再计算时段 $D_{1,1}$ 的均值负载矩阵 $\bar{P}^{D_{1,1}}$ 与第 3 个时间片负载矩阵 $P_3^{D_1}$ 之间的相似度 $d_{1,3}^{D_1}$。若满足 $d_{1,3}^{D_1}\geqslant\zeta$，则将时间片 3 归为时段 $D_{1,1}$，并重新计算时段 $D_{1,1}$ 的均值负载矩阵 $\bar{P}^{D_{1,1}}=\frac{1}{3}\sum_{i=1}^{3}P_i^{D_1}$，再依次计算均值负载矩阵与其它时间片的相似度，直至第 k 个时间片，当 $d_{1,k}^{D_1}\leqslant\zeta$ 时，把第 k 个时间片纳入新的子时段 $D_{1,2}$。

5)以时间片 k 为起始位置，以阈值公式为条件，当满足条件时，逐一重新计算时段 $D_{1,2}$ 的均值负载矩阵 $\bar{P}^{D_{1,2}}=P_k^{D_1}$，将他与后面的负载矩阵进行对比，计算相似度，否则进入新时段 $D_{1,3}$，依此类推，将时段 D_1 内的所有时间片的时段细化分结束作为结束，得到子时段 $D_{1,1}, D_{1,2},\cdots$，以及各子时段的均值负载矩阵 $\bar{P}^{D_{1,1}}, \bar{P}^{D_{1,2}},\cdots$。

6)对粗划分获得的时段 $D_2, D_3,\ldots$ 的所有时间片矩阵，依照步骤 4)和步骤 5)分别对其进行时段细化分。

7)对所有的时段进行细化分，然后最终的结果是得到 1，2，…，M 的操作子时段。针对不同的子时段搭建各自的 MK-PCA 模型。根据式(8)和式(9)分别确定各子模型的 T^2 和 SPE 控制限。

对于监测过程的测试数据，需要的应该是完整的批次数据，而现在得到的监测数据为自批次开始到监测时刻的采用数据，因此需要对数据进行评估。针对这个问题，已经提出了多种解决方法，本文采用各变量的均值来代替其估计值。当三维数据矩阵展开成二维数据矩阵之后，基于分段多向核主元分析在线检测的步骤与基于 KPCA 的步骤完全一样。区别仅在于基于分段多向核主元分析在线检测是对离线建模时得到的 1，2，…，M 的操作子时段分别在线监测，然后分别求出各个时段的 T^2 和 SPE 统计量，并检测其是否超限。

3 实验研究

本实验采用微型啤酒生产设备对算法进行验证，将发酵过程的监控数据作为测试数据，同时根据生产过程中各变量对生产状态的影响考虑生产工艺的特点，在这里选择了压力、PH 值、温度、糖度、酒精度和液位这 6 个过程检测变量，这些变量反应了酵母菌菌体生长和发酵产物的合成状况。发酵过程共 15 天，1 小时对数据进行 1 次采样，每一批次共采样 360 次。本实验选取正常批次的历史数据共 12 个。考虑到每一批次数据的反应时间不一样，在这里将 $X(K\times 6)$ 转换成 $X(1\times K6)$ 之后，将多于 2160 列的数据进行截取，不足 2160 列的批次补零，使得每批次都为 2160 列，然后将矩阵 $X(1\times 2160)$ 排列成 $X(12\times 2160)$ 形式，对其标准化处理，然后分别对每一个时间片段进行 KPCA 处理，按照累计贡献率为 93% 来提取主元，根据主成分个数的不同，对过程数据进行时段粗划分，其划分结果如图 2 所示。

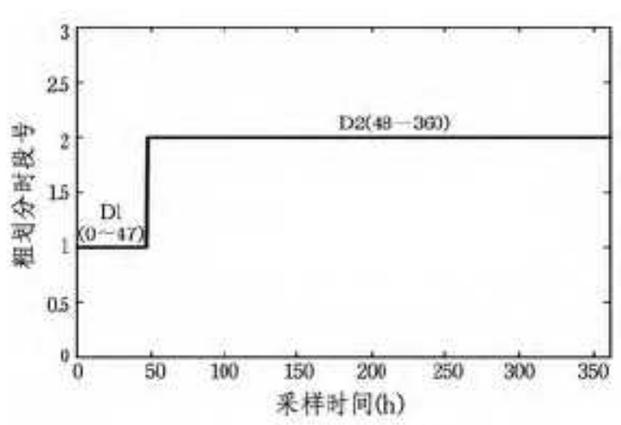

图 2 时段粗划分结果

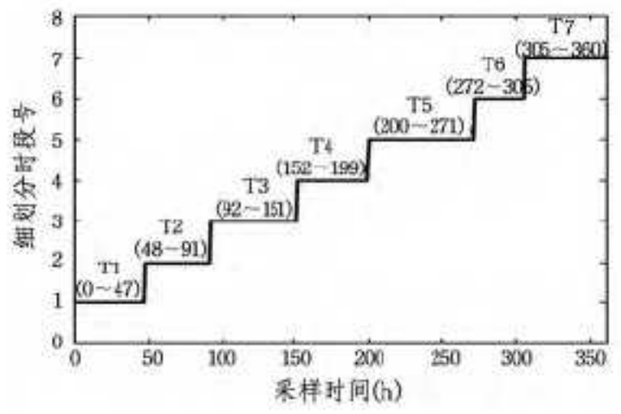

图 3 时段细划分结果

由图 2 所示，整个生产过程被粗划分为两个阶段，第一阶段为 $D1$，时间为 0～47h，与实际生产过程的时间进行对比分析可知：这一阶段为啤酒发酵的自然升温过程；$D2$ 为 48～360h，对应啤酒的发酵过程。由此完成了整个过程的粗划分。

图 3 所示为时段细化分的结果，$D1$、$D2$ 为粗划分，在这个基础上，依据负载矩阵相似度的大小对其进行进一步细化分。$D1$ 被细化为一段，即 $T1$，它对应的是啤酒发酵中的自然升温阶段；将 $D2$ 分成 6 段，其中 $T2$ 对应主发酵阶段，该阶段糖被分解释放热量；$T3$ 对应后酵阶段，这一阶段发酵罐温度下降，要求降温速度为 0.3℃/H；$T4$、$T5$ 对应还原阶段，这一阶段继续分解残糖，降低氧含量，沉淀蛋白质，啤酒发酵进一步稳定；$T6$ 阶段对应的是贮酒阶段，这一阶段要求温度下降，降温的速度为 0.15℃/H；$T7$ 对应贮酒阶段，根据工艺要求，要将发酵罐的温度控制在 0～1℃。

阶段划分后，各个阶段的 T^2 及 SPE 控制限也相应地确定，图 4、图 5 所示为 T^2 统计量和 SPE 统计量的控制限，它由 12 个正常批次数据所得到。

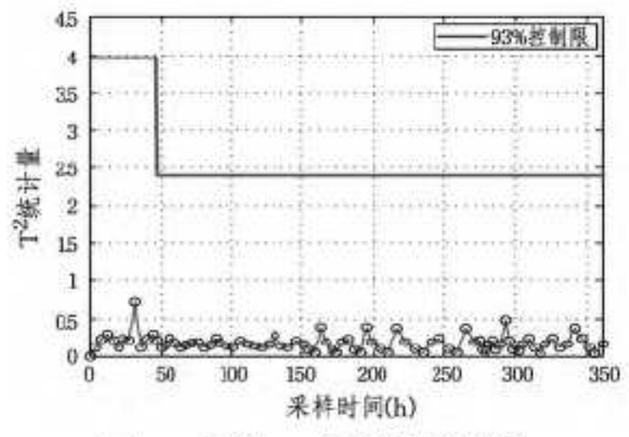

图 4 离线 T^2 统计量控制限

(下转第 33 页)

[4] Robert S. Integrated tools for managing the total pipeline [J]. Annual Conference Proceedings (Chicago: Council of Logistics Management), 2012, 15(7): 93-108

[5] 中国民用航空局. 中国航空运输发展报告(2012/2013)[OL]. http://www.caac.gov.cn/H1/H2/201308/t20130828_18587.html

[6] 刘浩, 钱小燕, 汪荣. 随机需求 VRP 的一个算法[J]. 南京工业大学学报(自然科学版), 2013, 5(3): 9-11

[7] 陆琳, 谭清美. 一类随机需求 VRP 的混合粒子群算法研究[J]. 系统工程与电子技术, 2013, 28(2): 244-247

[8] Dethloff J. Relation between vehicle routing problems: An insertion heuristic for the vehicle routing problem with simultaneous delivery and pick-up applied to the vehicle routing problem with backhauls[J]. Journal of the Operational Research Society, 2012, 53(6): 115-118

[9] Tang F A, Galvao R D. Vehicle routing problems with simultaneous pick-up and delivery service[J]. Journal of the Operational Research Society, 2012, 39(8): 19-33

(上接第 27 页)

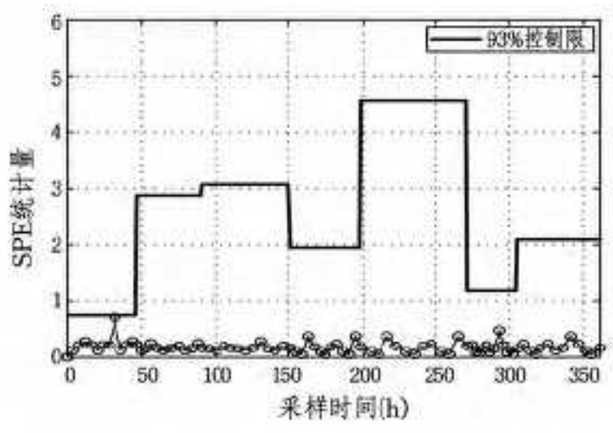

图 5 离线 SPE 统计量控制限

测试数据从第 317 个采样时刻到第 360 个采样时刻引入发酵温度故障，图 6—图 9 分别示出了采用 MPCA 算法和 MKPCA 算法的检测结果。

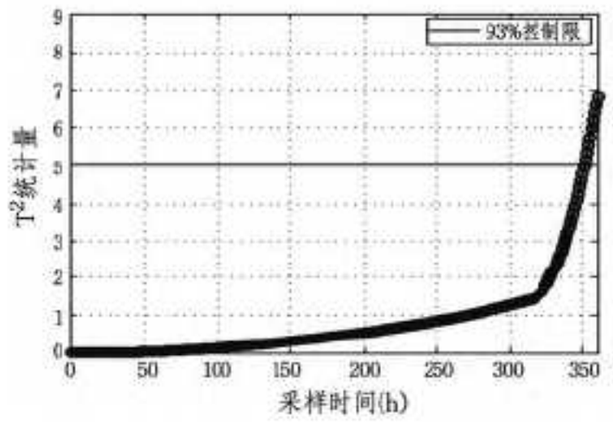

图 6 MPCA T^2 统计量监测图

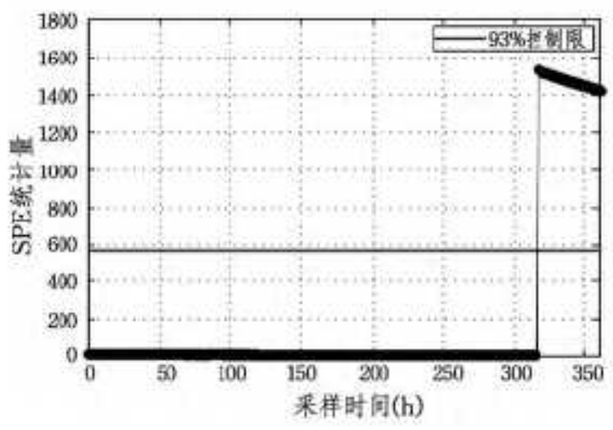

图 7 MPCA SPE 统计量监测图

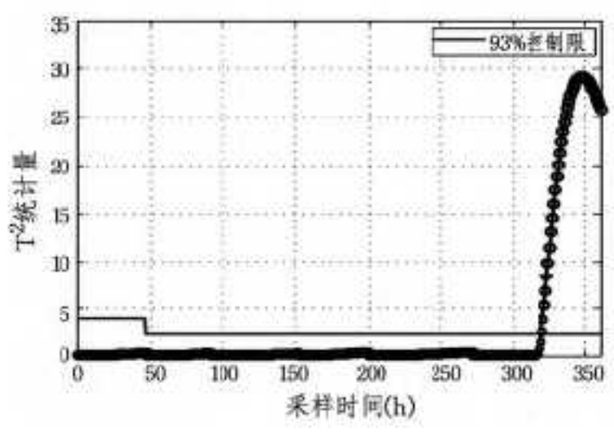

图 8 MKPCA T^2 统计量监测图

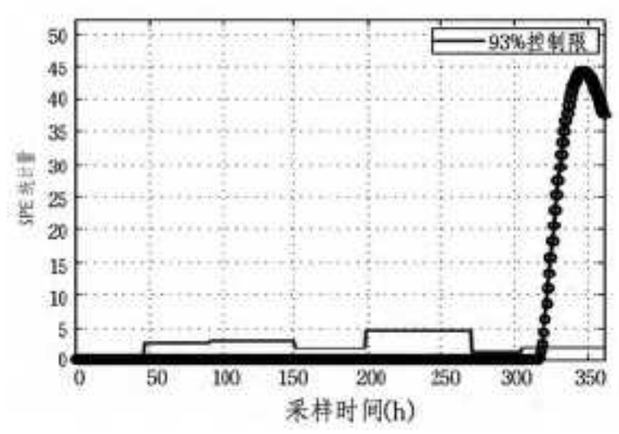

图 9 MKPCA SPE 统计量监测图

在 317 个采样时刻引入发酵温度故障以后，图 8、图 9 中 MKPCA 的 T^2 和 SPE 统计量均及时准确地检测出故障，而由图 6、图 7 可以看出，MPCA 的 SPE 统计量及时准确地检测出了故障，但是 T^2 统计量对于故障的响应表现出明显的滞后，没能将故障准确及时地监测出。

结束语 为了更好地对这一过程进行故障诊断，本文根据间歇生产过程的复杂性、非线性等问题，提出了一种过程分段的多向核主元分析方法(MKPCA)，通过对整个生产阶段的粗划分与细化分，建立了一个更加准确的统计模型，它克服了传统的 MPCA 对间歇过程分析的多种弊端。将这一方法应用到典型间歇过程——啤酒发酵过程中，通过仿真研究，不管是在监测的实时性还是准确性上都取得了满意的效果，验证了该方法对间歇过程故障诊断的优势。

参 考 文 献

[1] Qi Y S, Wang P, Fan S J, et al. Enhanced batch process monitoring using kalman filter and multiway kernel principal component analysis[C]// 2009 Chinese Control and Decision Conference(CCDC 2009). 2009: 5289-5294

[2] Zhang C, Li Y. Study on the fault-detection method in batch process based on statistical pattern analysis[J]. Chinese Journal of Scientific Instrument, 2013, 34(9): 2103-2110

[3] 陆宁云, 王福利, 高福荣, 等. 间歇过程的统计建模与在线监测[J]. 自动化学报, 2006, 32(3): 400-410

[4] 潘玉松. 基于主元分析的传感器故障检测与诊断[D]. 河北: 华北电力大学 2005

[5] 孔晓光, 郭金玉, 林爱军. 基于二维主元分析的间歇过程故障诊断[J]. 计算机应用, 2013, 33(2): 350-352

[6] 常玉清, 王姝, 谭帅, 等. 基于多时段 MPCA 模型的间歇过程监测方法研究[J]. 自动化学报, 2010, 36(9): 1312-1320