基于分块核主成分分析和支持向量机的故障检测

李锦冰[1], 韩　冰[2], 冯守渤[1], 张佳冬[1], 李　宇[1], 钟　凯[1], 韩　敏[1†]

(1. 大连理工大学 电子信息与电气工程学部, 辽宁 大连 116024;

2. 上海船舶运输科学研究所 航运技术与安全国家重点实验室, 上海 200000)

摘要: 针对工业系统监测数据为非线性, 且难以辨识复杂工作过程中故障位置的问题, 提出一种基于分块核主成分分析(BKPCA)和最小二乘支持向量机(LS–SVM)的集成故障检测方法. 首先对系统监测变量进行分块, 使用KPCA对每个分块在特征空间中建立T^2和平方预测误差(SPE) 统计量来实时监测系统健康状态, 并使用LS–SVM对上述过程检测出来的故障数据进行再次判断. 随后计算出现故障后计算每一分块的故障贡献率, 进而确定发生故障的分块. 由于采用了并行分块算法, 可以较简单的确定故障发生位置, 提高计算效率, 同时LS–SVM方法的应用也可以提升故障检测的精度. 使用田纳西–伊斯曼化工(TE)过程数据对本文所提方法进行仿真验证, 试验结果表明所提方法取得了较好效果.

关键词: 故障检测; 分块核主成分分析; 最小二乘支持向量机; 特征提取

引用格式: 李锦冰, 韩冰, 冯守渤, 等. 基于分块核主成分分析和支持向量机的故障检测. 控制理论与应用, 2020, 37(4): 847 – 854

DOI: 10.7641/CTA.2019.80923

Fault detection based on block kernel principal component analysis and support vector machine

LI Jin-bing[1], HAN Bing[2], FENG Shou-bo[1], ZHANG Jia-dong[1],
LI Yu[1], ZHONG Kai[1], HAN Min[1†]

(1. Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian Liaoning 116024, China;
2. State Key Laboratory of Navigation and Safety Technology, Shanghai Ship and Shipping Research Institute, Shanghai 200000, China)

Abstract: The measurement data of industrial system is nonlinear and difficult to extract the characteristic information. In the complex large-scale industrial process, an integrated fault detection method based on the block kernel principal component analysis (BKPCA) and least squares support vector machine (LS–SVM) is proposed. Firstly, the measurement variables are partitioned. And KPCA is used to establish the T^2 and squared prediction error (SPE) monitoring statistics in the feature space for each block to monitor the health status in real time. The LS–SVM is used to rejudge the faulty data detected by above process. Calculating the contribution rate of each block after the fault occurs, and then the faulty block can be determined. Due to the parallel block algorithm, the location of the fault can be simply found, and the computational efficiency is improved. What is more, the application of LS–SVM can also improve the accuracy of fault detection. The Tennessee-Eastman (TE) process data is used to verify the method proposed in this paper. The results show the effectiveness of proposed method.

Key words: fault detection; block kernel principal component analysis; least square support vector machine; feature extraction

Citation: LI Jinbing, HAN Bing, FENG Shoubo, et al. Fault detection based on block kernel principal component analysis and support vector machine. *Control Theory & Applications*, 2020, 37(4): 847 – 854

1　引言

随着人们对工业过程稳定性要求的提升, 故障检测方法受到越来越广泛地关注. 经过近30年的不断发展, 故障检测逐渐形成了3类方法: 基于解析模型方法、基于定性经验知识方法、基于数据驱动方法[1]. 基于数据驱动的方法是近年来热门研究方法之一, 当过程的精确数学模型难以建立时, 对于历史数据的处理研究成为了处理故障检测问题的重要渠道[2–4].

收稿日期: 2018−11−24; 录用日期: 2019−08−23.
†通信作者. E-mail: minhan@dlut.edu.cn; Tel.: +86 411-84707847.
本文责任编委: 陈增强.
国家自然科学基金项目(61773087), 中央高校基本科研业务费专项资金(DUT18RC(6)005)资助.
Supported by the National Natural Science Foundation of China (61773087) and the Fundamental Research Funds for the Central Universities (DUT 18RC(6)005).

为应对工业系统监测过程复杂的问题, 大量多元统计分析方法被用于故障检测领域, 如主成分分析(principal component analysis, PCA)[5]、独立成分分析(independent component analysis, ICA)[6]及费舍尔判别分析(fisher discriminant analysis, FDA)[7]等. 其中PCA方法用较少的互不相关的综合指标代替原始维度较高的指标, 同时尽可能多的保持原始数据的有效信息[8]. 通过构建T^2和平方预测误差(squared prediction error, SPE)统计量分别对主元空间和残差空间中的投影变化进行度量, 除监测变量均值的增长和减少外还可以发现监测变量之间相关性的变化. 由于其良好的特性, PCA方法已经广泛应用于复杂工业流程的过程监控. 然而PCA方法在处理非线性、动态、多尺度过程中显得力不从心, 一些基于PCA的改进和扩展的方法被逐渐提出, 包括: 基于核的方法[9–10], 基于神经网络的方法[11–12]等. 核主成分分析(kernel PCA, KPCA)是对PCA算法的非线性扩展, 能够挖掘到数据样本中所隐含的非线性信息, 具有很好的性能. Ren, L等人[13]使用小波变换和KPCA方法相结合处理初始故障检测问题, 同时还分析了可能引起结果不确定的干扰因素. 韩敏等[14]提出一种结合核独立成分分析和流形学习保持局部结构的思想, 解决了之前容易忽略各独立成分分量对系统故障贡献度差异的问题. 赵孝礼等[15]基于KPCA与流形学习方法结合的算法, 可以实现转子故障数据的降维操作.

这些改进方法的应用有效改善了故障检测的效果, 然而由于传统KPCA在建立模型过程中, 需要考虑到过程所涉及的所有变量特征信息. 主成分模型一旦建立就基本不会发生更改, 但这也会忽视监测变量内部固有联系, 所以根据实际情况建立起更为细致划分的主成分模型就显得十分重要. 此外, 由于工业系统监测数据维度较多且随时间而大量生成, 传统方法降低了过程监控和故障检测的效率. 本文提出一种分块核主成分分析(block KPCA, BKPCA)算法, 将复杂过程按照一定规则进行分块处理, 确保各个模块的检测能力. 对各个分块进行全局的特征提取, 从而提高检测正确率, 降低计算复杂度. 同时通过计算每一分块的贡献率确定故障位置, 虽然无法明确确定发生故障的变量, 但随着现代工业系统集成化和自动化水平的提升, 本文所提方法简化了故障位置确定过程, 仍然具有十分重要的意义.

此外T^2和SPE统计量的控制限是根据概率方法确定的, 为保证发现较小的故障数据波动就一定会存在较高的误判率. 本文通过最小二乘支持向量机(least square support vector machine, LS–SVM)方法对系统的健康状态进行再一次判断, 可以进一步减小误报率, 提高判断精确度. LS–SVM 可以在保证识别精度的同时, 较好地解决小样本、非线性等实际问题, 且具有更快的求解速度[16–17], 因此可以获得更好的故障检测结果. 最后使用田纳西–伊斯曼化工(Tennessee-Eastman, TE)过程故障数据进行仿真实验, 证明了本文所提算法的有效性, 可以为工业系统实际故障检测和定位提供辅助决策建议.

2 分块核主成分分析算法

KPCA算法可以通过核映射将监测数据映射到高维核空间, 通过确定一个坐标系统的正交变换, 将监测数据变换为一组新的各维度线性无关的表示. 在新的坐标系下, 变换后的数据方差沿新的坐标轴最大化[18]. 本文所提BKPCA是对KPCA的改进, 对于复杂的工业系统建立更加精细的主元模型, 可以有效提升计算效率, 减小故障检测的计算复杂度.

2.1 核主成分分析算法

设$X=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\}$为样本集, 其中$N$为样本总数, 通过非线性映射$\phi$将原始数据映射到维数为无穷大的高维空间$\Theta$中, 记为$\phi(\boldsymbol{x}_k)$. 核函数通过映射实现原始空间点到高维空间的映射, 由此映射而得的特征空间中的数据是去均值的, 则特征空间中的协方差矩阵为

$$\boldsymbol{C}^{\Theta}=\frac{1}{N}\sum_{j=1}^{N}\phi(\boldsymbol{x}_j)\phi^{\mathrm{T}(\boldsymbol{x}_j)}, \tag{1}$$

$\boldsymbol{C}^{\Theta}$的特征值λ, 特征向量$\boldsymbol{v}$满足

$$\boldsymbol{C}^{\Theta}\boldsymbol{v}=\lambda\boldsymbol{v}. \tag{2}$$

一般而言, λ会存在N个结果, 可以表述为$\lambda_1,\lambda_2,\cdots,\lambda_N$且$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_N$. 存在系数$\boldsymbol{\alpha}_i(i=1,2,\cdots,N)$, 使得

$$\boldsymbol{v}=\sum_{i=1}^{N}\boldsymbol{\alpha}_i\phi(\boldsymbol{x}_i). \tag{3}$$

代入式(2), 并对式(2)两边同乘$\phi(\boldsymbol{x}_k)$可化为

$$\lambda(\phi(\boldsymbol{x}_k)(\sum_{i=1}^{N}\alpha_i\phi(\boldsymbol{x}_i)))=\phi(\boldsymbol{x}_k)\boldsymbol{C}^{\Theta}(\sum_{i=1}^{N}\alpha_i\phi(\boldsymbol{x}_i)). \tag{4}$$

式(4)可简化为

$$N\lambda\boldsymbol{\alpha}=K\boldsymbol{\alpha}. \tag{5}$$

核矩阵$K_{jk}=K(\boldsymbol{x}_j,\boldsymbol{x}_k)=\phi(\boldsymbol{x}_j)\phi^{\mathrm{T}}(\boldsymbol{x}_k)$, $\boldsymbol{\alpha}=[\alpha_1\ \alpha_2\ \cdots\ \alpha_n]^{\mathrm{T}}$为核矩阵$K$的特征向量.

上式推导需要假设$\phi(\cdot)$为去均值的, 实际情况并不一定成立, 而$\phi(\cdot)$与核函数密切相关, 因此在进行主成分分析之前, 需要对K进行中心化处理, 即将式(5)中的K用$\bar{K}$代替:

$$\bar{K}_{ij}=(\phi(\boldsymbol{x}_i)-\frac{\sum_{m=1}^{N}\phi(\boldsymbol{x}_m)}{N})(\phi(\boldsymbol{x}_j)-\frac{\sum_{n=1}^{N}\phi(\boldsymbol{x}_n)}{N})^{\mathrm{T}}=$$

$$(\boldsymbol{K} - 1_N\boldsymbol{K} - \boldsymbol{K}1_N + 1_N\boldsymbol{K}1_N)_{ij}, \tag{6}$$

其中: 1_N为每一个元素都为$1/N$的$N \times N$矩阵, N为K的维数. 结合式(3) 和式(5), 由矩阵K的特征向量$\boldsymbol{\alpha}$可求出协方差矩阵$\boldsymbol{C}^{\Theta}$ 的特征向量$\boldsymbol{v}$.

依据式(7)累积方差贡献率准则选取前p个特征值确定主成分个数:

$$\frac{\sum\limits_{k=1}^{p}\lambda_k}{\sum\limits_{k=1}^{N}\lambda_k} > 85\%. \tag{7}$$

对于新观测样本$\boldsymbol{x}_{\rm new}$, 其主元空间得分向量可以表示为

$$\boldsymbol{t}_{\rm new} = [t_{{\rm new},1}\ \ t_{{\rm new},2}\ \cdots\ t_{{\rm new},p}]^{\rm T}, \tag{8}$$

$$t_{{\rm new},l} = \upsilon_l\bar{\phi}(\boldsymbol{x}_{\rm new}) = \sum_{j=1}^{N}\alpha_j^l(\bar{\phi}(\boldsymbol{x}_j)\bar{\phi}(\boldsymbol{x}_{\rm new})), \tag{9}$$

其中: $l = 1, 2, \cdots, p$, $\bar{\phi}(\boldsymbol{x}_j)$为中心化的特征向量.

监测统计量$T^2_{\rm new}$和$\rm SPE_{new}$可以定义为

$$T^2_{\rm new} = \boldsymbol{t}^{\rm T}_{\rm new}\Lambda^{-1}\boldsymbol{t}_{\rm new}, \tag{10}$$

$$\rm SPE_{new} = k(\boldsymbol{x}_{\rm new}, \boldsymbol{x}_{\rm new}) - \boldsymbol{t}^{\rm T}_{\rm new}\boldsymbol{t}_{\rm new}, \tag{11}$$

其中Λ为p个主成分对应特征值构成的对角矩阵. 通过分析可知T^2统计量变化符合F 分布:

$$T^2 \sim \frac{p(l-1)}{n-p}F_{p;l-p;a}, \tag{12}$$

其中: n为样本个数; a为模型数据检验水平, 可以根据a的取值确定T^2的控制限.

SPE控制限计算公式如下:

$$\rm SPE_{lim} = \theta_1[1 + \frac{N_a\sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0(h_0-1)}{{\theta_1}^2}]^{\frac{1}{h_0}}, \tag{13}$$

其中:

$$\theta_l = \sum_{j=p+1}^{n}\lambda_j^i (i = 1, 2, 3),\ h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2},$$

λ_i为特征值, N_a是正态分布a临界值.

2.2　分块原理介绍

实际工业过程运行时, 内部结构错综复杂, 变量之间也存在相互的关联, 建立单一模型很难对变量内部的关系做出准确正确的解释. 本文提出一种有效建立分块模型的思想, 将整个工业系统运行过程分成若干等级或者过程, 每一过程都建立相应模型, 这就达到分块或多层次的思想, 具体原理如图1所示. 其中X^*可以看作是原始监测数据, 按照工业过程将其分类, 如图 1 中X_1^1, X_2^1等, 当第 1 次分类之后仍然具有多个不同单元时仍可以对其进行第2次分类. 因此图1只是一个示意图, 并不是必须分为3层.

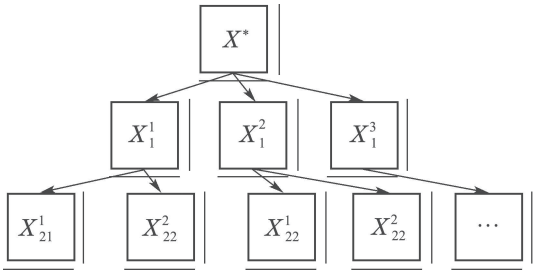


图 1　数据分块原理介绍

Fig. 1　The introduction for data blocking principle

传统KPCA算法可以得到了主成分的载荷矩阵$\boldsymbol{v}_k(m \times k)$, 即其主元个数为$m$, 采样个数为$k$. 针对不同环境下, 设备运行的先验知识, 将原始数据划分为B个分块, 其中, 第b个分块包含了m_b个变量, 载荷矩阵可以表述如下:

$$\boldsymbol{v}_k = [\boldsymbol{v}_{k1}\ \cdots\ \boldsymbol{v}_{kb}\ \cdots\ \boldsymbol{v}_{kB}]^{\rm T}, \tag{14}$$

其中$\boldsymbol{v}_{kb}(mb \times k)$表示第$b$个分块中独立的主成分载荷矩阵. 当输入新的数据信息时, 将数据依据分块标准分成B个分块.

为方便计算T^2统计量, 将协方差矩阵根据分块原则也进行了分块处理:

$$\boldsymbol{C}^{\Theta} = \begin{bmatrix} \boldsymbol{C}_1^{\Theta} & & & & \\ & \cdots & & & \\ & & \boldsymbol{C}_b^{\Theta} & & \\ & & & \cdots & \\ & & & & \boldsymbol{C}_B^{\Theta} \end{bmatrix}, \tag{15}$$

其中$\boldsymbol{C}_b^{\Theta}(mb \times mb)$表示数据块$\boldsymbol{X}_b(n \times mb)$的协方差矩阵.

因此, T^2统计量与式(10)类似.

$$T_b^2 = \boldsymbol{t}_b^{\rm T}\Lambda^{-1}\boldsymbol{t}_b. \tag{16}$$

SPE统计量与式(11)类似, 第b块分块的SPE统计量计算如下:

$$\rm SPE_b = k(\boldsymbol{x}_b, \boldsymbol{x}_b) - \boldsymbol{t}_b^{\rm T}\boldsymbol{t}_b. \tag{17}$$

分别使用每一分块中的T^2和SPE统计量对其对应分块的健康状态进行估计, 减少了其他分块变量的影响, 可以提高故障检测精度.

2.3　贡献率分析

系统故障表现在数据上主要是出现不正常波动, 同时这些波动的存在也可以帮助确定故障发生的位置. 本文使用BKPCA的方法对每一个数据分块引起的T^2和SPE统计量变化进行计算, 考虑到T^2统计量主要用于衡量现有监测数据距离主元子空间原点的距离, 使用各分块的T^2统计量的变化衡量对应分块的故障贡献率, 可以有效简化故障位置发生位置确定过程.

由于每一分块的T^2统计量控制限不同, 简单的使用统计量衡量故障贡献率效果较差. 因此, 本文构建

的第b块数据故障贡献率如下:

$$C_b = T_b^2 - T_{b,\lim}^2, \tag{18}$$

$T_{b,\lim}^2$是指第b块数据的T^2统计量控制限, 当监测数据T^2统计量不大于对应控制限时则认为贡献率为0.

使用分块的贡献率方法虽然无法直接确定故障变量, 但随着现代工业系统集成度和自动化水平的提升, 隔离故障所属分块同样具有非常重要的意义. 同时由于采用分块方法对数据进行处理, 可以减少特定故障的监测变量个数, 降低KPCA的计算复杂度, 提高故障检测效率. 同时, BKPCA方法考虑了监测变量内部固有联系, 根据实际情况建立起更为细致划分的主成分模型, 减少了其他分块中变量的影响, 可以在一定程度上提升故障检测精度.

3 LS–SVM故障检测

LS–SVM是SVM的一种扩展改进, 通过二次规划将非线性问题转换为线性问题, 不仅可以降低计算的成本还大大提高了模型的泛化能力[19]. 基于BKPCA与LS–SVM的集成故障检测方法可以有效降低误报率, 提高检测效率.

3.1 最小二乘支持向量机

设训练样本集$D=\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}$, $i=1,2,\cdots,m$, $\boldsymbol{x}_i\in\mathbb{R}^n$, $\boldsymbol{y}_i\in\{-1,1\}$为期望输出. 通过非线性映射$\phi(\boldsymbol{x}_i)$将样本数据从原始空间映射到高维特征空间, 其中特征空间中的两个类间的距离为$2/\|\omega\|$, 为了增大分类边缘且减小训练的误差ε_i, 得到函数

$$\begin{aligned}&\min_{\omega,b,\varepsilon} L = \min_{\omega,b,\varepsilon}\{\frac{1}{2}\|\omega\|^2 + c\sum_{i=1}^{N} s_i\varepsilon_i\},\\ &\text{s.t. } \boldsymbol{y}_i[\sum_{i=1}^{N}\omega\phi(\boldsymbol{x}_i)+b]-1+\varepsilon_i\geqslant 0,\end{aligned} \tag{19}$$

c为超出控制误差的样本惩罚程度, s_i为加权系数.

LS–SVM通过将等式约束替代SVM中的不等式约束, 引入Lagrange系数α_i, 定义如下的Lagrange函数:

$$\begin{aligned}L = &\frac{1}{2}\|\omega\|^2 + \frac{1}{2}\gamma\sum_{i=1}^{N}{\varepsilon_i}^2 -\\ &\sum_{i=1}^{N}\alpha_i\{\boldsymbol{y}_i[\sum_{i=1}^{N}\omega\phi(\boldsymbol{x}_i)+b]-1+\varepsilon_i\},\end{aligned} \tag{20}$$

其中: γ为惩罚因子; b是偏差量; α_i为Lagrange乘子系数. 利用KKT(Karush Kuhn Tucker)条件[20]求出最优的α_i和b, 引入$K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\phi^{\mathrm{T}}(\boldsymbol{x}_i)\phi(\boldsymbol{x}_j)$核函数, 求$L$对$\omega,b,\varepsilon_i,\alpha_i$的偏导为0, 并代入公式(20), 得到如下矩阵方程:

$$\begin{bmatrix}0 & 1_T^{\mathrm{T}}\\ 1_V & \Omega+\dfrac{1}{\gamma}I\end{bmatrix}\begin{bmatrix}b\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\ \boldsymbol{y}(\boldsymbol{x})\end{bmatrix}, \tag{21}$$

其中: $1_V=[1\ \ 1\ \cdots\ 1]$; $\alpha=[\alpha_1\ \ \alpha_2\ \cdots\ \alpha_n]$; Ω中的元素$\Omega_{ij}=K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\phi^{\mathrm{T}}(\boldsymbol{x}_i)\phi(\boldsymbol{x}_j)$, 通过求解矩阵, 可得到最后的最小二乘支持向量机的分类函数表达式为

$$\boldsymbol{y}(\boldsymbol{x}) = \sum_{k=1}^{l}\alpha_k K(\boldsymbol{x}_i,\boldsymbol{x}_j)+b. \tag{22}$$

另外本研究选用高斯核函数对数据处理:

$$K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp(\frac{-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2}{2\sigma^2}). \tag{23}$$

3.2 故障检测过程

本文所提方法主要是使用BKPCA与LS–SVM相结合实现故障检测, BKPCA方法主要用于分块的监测系统的健康状态, 为提高其故障检测的精度, 在发现故障之后使用LS–SVM方法对监测数据的健康状态进行二次判断, 并通过贡献率的方法对故障发生分块进行确定. 这种方法可以实现在线的系统健康状态监测, 并提高故障检测效率和精度. 具体的工作过程如下所示:

步骤1 收集系统历史运行状态数据, 根据实际设备运行过程进行实际可操作的分块处理;

步骤2 对数据进行预处理并分别构建每一分块KPCA监测模型, 计算每一分块的T^2和SPE控制限, 同时训练建立每一分块的LS–SVM故障分类器;

步骤3 监测系统运行状态数据, 并对收集的监测数据进行预处理, 分别计算每一分块T^2和SPE统计量, 确定是否超出对应控制限;

步骤4 当分块的T^2或SPE统计量未超出控制限则重复进行步骤3, 认为此时没有发生故障并持续对系统运行状态进行监测, 反之则将超出控制限的监测数据导入对应分块的LS–SVM故障分类器检测故障因素, 当LS–SVM检测为故障时认为故障已经发生;

步骤5 根据公式(18)分别计算并比较每一分块的故障贡献率, 确定发生故障的分块.

4 仿真实验

为验证本文所提方法的有效性, 本文使用TE过程数据进行仿真验证. TE过程数据是由Downs等根据Eastman化学公司的实际工艺流程作少许修改后于1993年提出的[21], 已经被广泛用于连续工业过程的状态监测和诊断的优化等领域. 整个过程由5个主要的操作单元组成, 即反应器、冷凝器、气液分离塔、循环压缩机和汽提塔, 具体流程图如图2所示. 该数据集共包含41个可供监控的测量变量和12个操作变量(如表1所示), 其中41个测量变量又包含22个连续过程测量变量(如表2所示)和19个成分测量值. 本文选用22个连续测量变量和11个操作变量结合对系统的健康状态进行监测. 为保证故障检测问题的应用, TE过程数据集一共被设计了21组故障数据和1组正常数据, 每组数据包含训练数据共480个采样点, 测试数据共960个采样点, 其中测试数据在第161个采样点处开始加入故障因素.

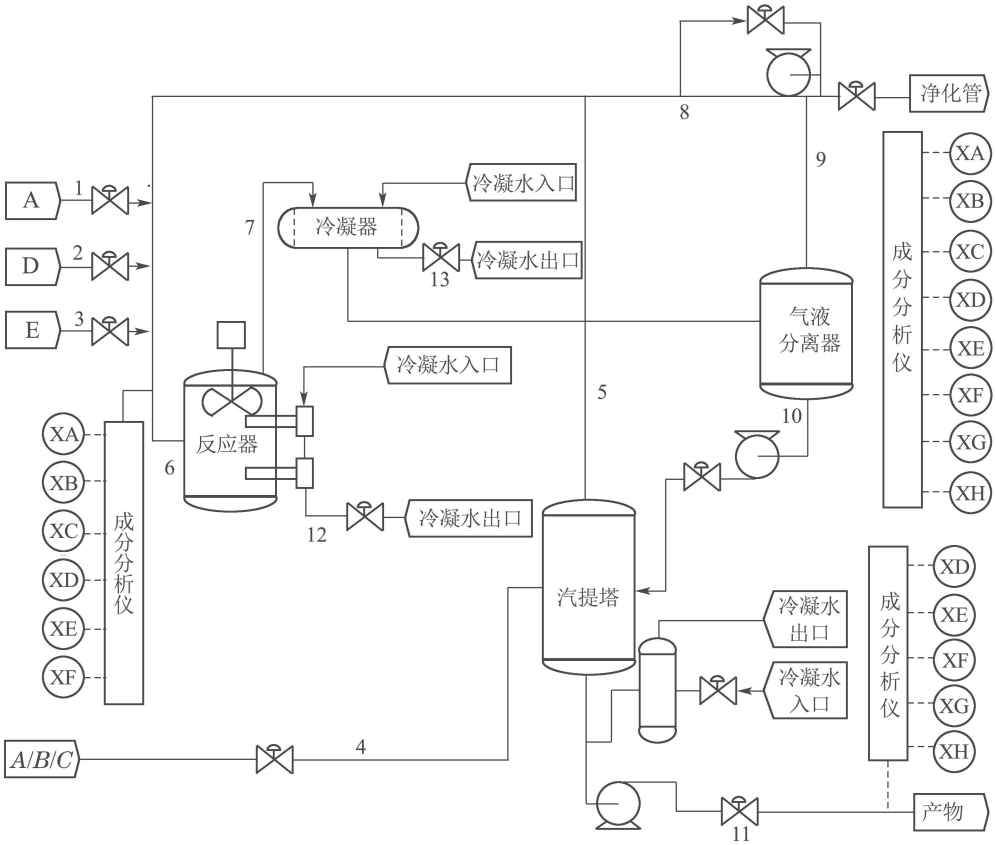


图 2 TE过程流程图

Fig. 2 The flowchart of the TE process

表 1 TE过程操纵变量

Table 1 The operating variables of TE process

编号	变量名	单位
X_1	物料D的流量	kg/h
X_2	物料E的流量	kg/h
X_3	物料A的流量	1000 m^3/h
X_4	物料A和C的流量	1000 m^3/h
X_5	压缩机循环阀	%
X_6	放空阀	%
X_7	分离器液位流量	m^3/h
X_8	液体产品流量	m^3/h
X_9	汽提塔水流阀	%
X_{10}	反应器冷却水流量	m^3/h
X_{11}	冷凝器冷却水流量	m^3/h
X_{12}	搅拌速度(无)	r/min

首先按照对TE过程结构原理的理解将完整的故障数据分成5块, 包括:

1) 原料块 1: $F_1, F_2, F_3, F_4, X_1, X_2, X_3, X_4$;

2) 反应器块 2: $F_6, F_7, F_8, F_{21}, X_{10}, X_{11}$;

3) 分离器块 3: $F_{11}, F_{12}, F_{13}, F_{14}, X_7$;

4) 汽提塔块 4: $F_{15}, F_{16}, F_{17}, F_{18}, F_{19}, F_8, F_9$;

5) 压缩机块 5: $F_5, F_{10}, F_{20}, F_{22}, X_5, X_6$.

分别检测每一分块中故障因素可以完成故障位置的确定, 同时减小计算复杂度.

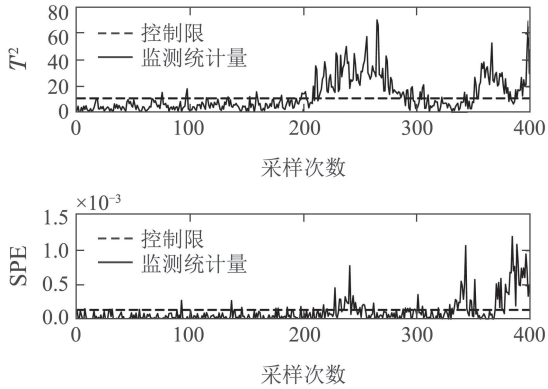
表 2 TE过程的22个测量变量

Table 2 The 22 measurement variables of TE process

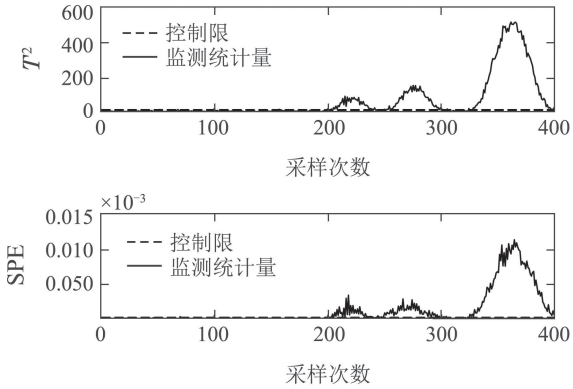
编号	变量名	单位
F_1	物料A流量	1000 m^3/h
F_2	物料D流量	kg/h
F_3	物料E流量	kg/h
F_4	物料A和C流量	1000 m^3/h
F_5	循环流量	1000 m^3/h
F_6	反应器进料流量	1000 m^3/h
F_7	反应器压力	kPa(表值)
F_8	反应器液位	%
F_9	反应器温度	°C
F_{10}	放空速率	1000 m^3/h
F_{11}	分离器温度	°C
F_{12}	分离器液位	%
F_{13}	分离器压力	kPa(表值)
F_{14}	分离器底部流量	m^3/h
F_{15}	汽提器液位	%
F_{16}	汽提器压力	kPa(表值)
F_{17}	汽提塔底部流量	m^3/h
F_{18}	汽提塔温度	°C
F_{19}	汽提塔流量	kg/h
F_{20}	压缩机功率	kW
F_{21}	反应器冷却水出口温度	°C
F_{22}	分离器冷却水出口温度	°C

4.1 分块故障检测结果

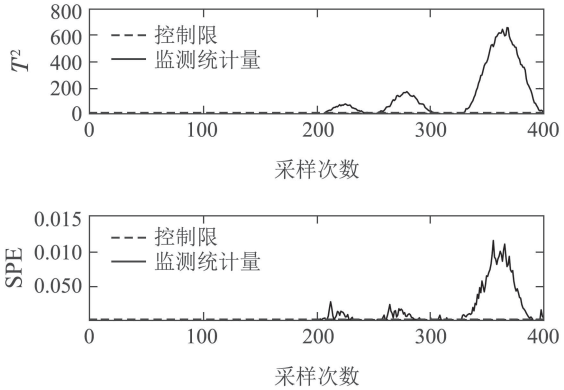
缓变故障可以更好的展现BKPCA方法的故障检测性能, 本文选用TE过程中故障13即反应动态偏移故障发生过程为例对所提方法进行仿真验证. 为了更加清楚的展现T^2和SPE统计量对故障的检测效果, 仿真实验只选用前400个采样点对仿真结果进行说明. 图3所示的是每个分块的T^2和SPE统计量故障检测结果, 其中虚线为对应的控制限. 可以发现在监测数据中加入故障因素之后每个分块的T^2和SPE统计量均开始出现变化趋势, 且经过几个采样点之后均超出控制限, 可以检测出故障因素, 说明这一故障发生会同时引起多个参数出现变化. 同时可以发现, 只有分块2即反应器块的T^2和SPE统计量均较早的超出控制限, 可以认为分块2参数变化对故障的影响最大, 符合故障13的产生原理.



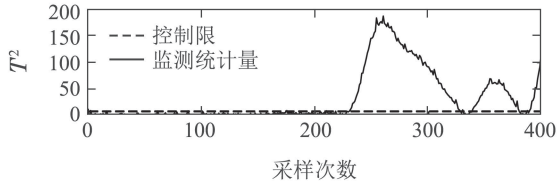
(a) 分块1

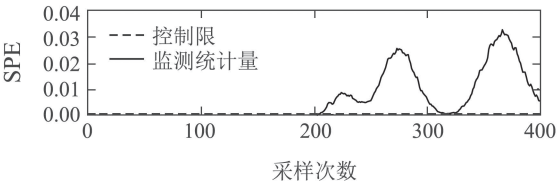


(b) 分块2

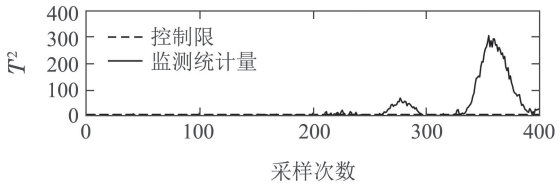


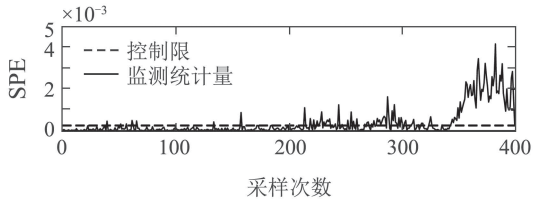
(c) 分块3





(d) 分块4





(e) 分块5

图 3 故障13的分块故障检测结果

Fig. 3 The result of block fault detection for Fault 13

为了更加清晰的定位故障发生的位置, 本文使用贡献率的方法对每一分块故障的贡献率进行了说明. 如图4展示了故障13中各个分块对故障的贡献率, 经过比较可以确定, 在故障发生初期, 只有分块2反应器块和分块3分离器块对故障具有较高的贡献率. 综合考虑TE过程反应原理, 出现这种现象主要是由于反应器产物进入分离器对分离器中各监测数据产生了影响, 比较两者的贡献率也可以发现分块2具有较高的故障贡献率, 即分块2对故障具有最大的影响.

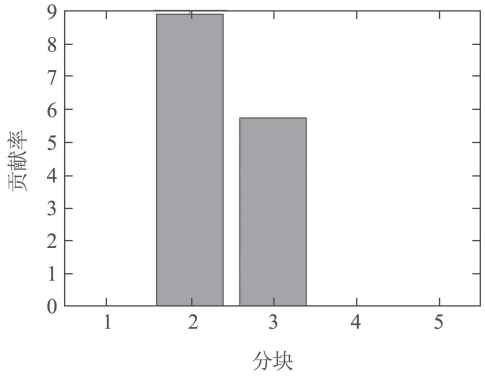


图 4 故障13下各分块的贡献率

Fig. 4 The block contribution rate for Fualt 13

为证明本文所提方法的客观性, 本文同样选用了一个阶跃故障, 即TE过程的故障1为例再次对分块贡

献率确定故障发生位置的方法进行验证. 分析每一分块中T^2和SPE统计量的变化过程可以发现随着A/C进料的阶跃变化不同块中的各监测变量也随之出现了变化, 然而经过一段时间之后各变量开始趋于平稳, 甚至回落到正常值, 因此出现了如图5所示的贡献率情况, 可以发现分块1贡献率最高, 可以认为故障1主要与原料块相关, 符合故障实际, 可以说明本文方法的正确性.

4.2　故障检测效率和精度

本文提出一种BKPCA与LS–SVM方法相结合以实现高效率高精度的故障检测的方法, 由于使用了分块的方法对数据进行了处理, 可以最大程度上的减少KPCA的监测模型和LS–SVM分类模型建立的复杂程度, 同时也可以减小在线故障检测过程的计算复杂度. 使用LS–SVM对BKPCA检测的分块监测数据的健康状态进行二次判断, 提高故障检测的精度. 如表3所示是使用几种方法比较TE数据集21种故障测试数据的检测结果和运行时间, 以误判率、漏判率和精确度3种参数共同对几种方法的故障检测结果进行对比.

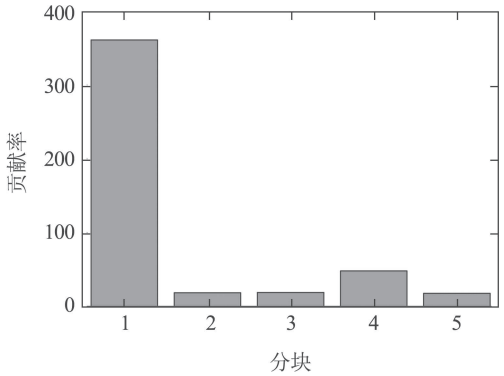


图 5　故障1下各分块的贡献率

Fig. 5　The block contribution rate for Fualt 1

表 3　TE数据集故障检测结果(%)和运行时间(s)对比

Table 3　The comparison of operation time (s) and results of fault detection (%) for TE data

故障序号	KPCA/T^2			KPCA/SPE			运行时间/s	LS–SVM			运行时间/s	本文			运行时间/s
	误判率/%	漏判率/%	精确度/%	误判率/%	漏判率/%	精确度/%		误判率/%	漏判率/%	精确度/%		误判率/%	漏判率/%	精确度/%	
1	3.75	0.75	98.75	16.88	0.13	97.08	2.684	0	0.75	99.38	24.581	1.25	0.25	**99.58**	4.197
2	2.5	1.38	98.44	14.37	1.13	96.67	2.608	0	2.26	97.81	25.971	0	0.13	**99.89**	5.547
3	1.87	96.13	19.58	17.5	80.13	30.31	2.621	79.37	25.12	65.83	25.995	3.13	0.75	**98.85**	3.439
4	2.5	48.25	59.37	13.13	0	97.81	2.822	8.13	2	96.98	25.86	2.5	0	**99.58**	3.457
5	2.5	70.63	40.73	13.13	61.38	46.67	2.656	1.87	0.13	**99.58**	28.053	2.5	1.13	98	3.687
6	0.63	1	99.06	10	0	98.33	2.723	1.25	0.63	99.27	25.033	0	0	**100**	5.87
7	0.63	0	**99.9**	10	0	98.33	2.664	0.63	0	**99.9**	25.346	0.63	2	98.23	4.116
8	1.25	2.62	97.6	12.5	2.38	95.94	2.568	1.87	0.13	**99.58**	25.273	1.25	1.25	98.75	5.651
9	7.5	93.87	20.52	15.63	83.75	28.6	2.589	56.87	30.13	65.42	25.455	9.38	1.87	**96.87**	3.331
10	1.87	63	47.19	9.38	35.38	68.96	2.664	27.37	22.38	77.29	25.934	0.63	0.37	**99.58**	3.704
11	2.5	43.38	63.44	13.75	19.38	81.56	2.631	46.25	22.5	73.54	25.433	4.37	0.13	**99.17**	3.427
12	3.75	1.13	98.44	16.88	2.38	95.21	2.936	1.25	3.25	97.08	26.941	3.75	0.25	**99.17**	6.499
13	3.75	4.88	95.31	8.13	3.88	95.42	2.882	0	12.63	89.48	25.498	1.87	0.5	**99.27**	5.448
14	4.37	0.13	99.17	16.25	0.13	97.19	2.695	44.37	9.12	85	25.643	1.25	0	**99.79**	2.857
15	1.87	95	20.52	12.5	81.63	29.9	2.668	31.25	31.13	68.85	25.346	1.87	0.37	**99.31**	3.349
16	10.63	78.13	33.13	18.12	38.87	64.58	2.736	70.63	7.88	81.67	25.501	15	0.13	**97.4**	3.714
17	1.25	19.25	83.75	11.87	2.75	95.73	2.567	12.5	16	84.58	25.346	3.75	1.25	**98.33**	3.798
18	3.13	10.25	90.14	16.25	8.25	90.42	2.83	20.62	8	89.9	24.497	1.25	0.13	**99.69**	5.499
19	1.87	81.25	31.98	13.75	51.38	54.9	2.607	86.88	9.5	77.6	24.780	2.5	0.25	**99.37**	2.976
20	2.5	61.38	48.44	8.13	30.88	72.92	2.968	10	22.38	79.69	25.145	1.25	0.13	**99.69**	4.587
21	6.88	54.13	53.75	25	38.25	63.96	2.554	5	88.38	25.52	25.146	0	3.13	**97.39**	4.283

可以发现本文所提方法对各种故障均具有较低的误判率和漏判率, 可以保证其故障检测结果的可信度. 同时在大部分情况下本文所提方法均获得最高的故障识别精确度, 未获得最高识别精度情况下, 其精度与最高精度相差不大, 这也同时证明了方法的可靠性. 由于新方法综合了分块核主成分分析和最小二乘支持向量机, 因此无法获得最小的运行时间, 但本文所提方法运行时间远远小于核主成分分析和最小二乘

支持向量机的运行时间之和, 可以说明本文所提方法具有较高的故障检测效率.

5 结论

随着人们对工业过程稳定性要求的提升, 智能化的故障检测方法逐渐引起了人们的广泛关注. 本文针对复杂工业过程所面对的数据规模大和故障位置难以确定等问题, 提出一种BKPCA与LS–SVM结合的故障检测方法. 该方法主要根据工业过程原理对大规模样本数据进行分块处理, 使用BKPCA方法并行处理样本数据, 分别计算每一分块的故障贡献率, 可以有效提高计算效率, 简化发现故障位置的过程. 通过LS–SVM方法对故障检测结果进行二次判断, 可以较小故障误判和漏判, 提高故障检测精度. 使用TE过程数据集对本文所提方法进行验证, 仿真结果证明本文所提算法能有效地辨识出故障发生位置, 提高了故障检测精度和效率, 可以为后期的故障类型确定和健康维护等过程提供辅助决策建议.

参考文献:

[1] NANDI S, TOLIYAT H, LI X. Condition monitoring and fault diagnosis of electrical motors–a review. *IEEE Transactions on Energy Conversion*, 2005, 20(4): 719 – 729.

[2] ZHANG Ping, WANG Guizeng, ZHOU Donghua. Fault diagnosis methods for dynamic systems. *Control Theory & Applications*, 2000, 17(2): 153 – 158.
(张萍, 王桂增, 周东华. 动态系统的故障诊断方法. 控制理论与应用, 2000, 17(2): 153 – 158.)

[3] CAI B, ZHAN Y, LIU H, et al. A data-driven fault diagnosis methodology in three-phase inverters for PMSM drive systems. *IEEE Transactions on Power Electronics*, 2016, 32(7): 5590 – 5600.

[4] XIA Lisha, YANG Yuying, FANG Huajing. Fault diagnosis performance improvement for chemical process based on EasyEnsemble method. *Control Theory & Applications*, 2017, 34(1): 49 – 53.
(夏丽莎, 杨玉英, 方华京. 基于EasyEnsemble的化工过程故障诊断性能改进. 控制理论与应用, 2017, 34(1): 49 – 53.)

[5] FAN J, QIN S, WANG Y. Online monitoring of nonlinear multivariate industrial processes using filtering KICA–PCA. *Control Engineering Practice*, 2014, 22(1): 205 – 216.

[6] CHEN M, HSU C, MALHOTRA B, et al. An efficient ICA-DW-SVDD fault detection and diagnosis method for non-Gaussian processes. *International Journal of Production Research*, 2016, 54(17): 5208 – 5218.

[7] LU Q, JIANG B, GOPALUNI R, et al. Locality preserving discriminative canonical variate analysis for fault diagnosis. *Computers & Chemical Engineering*, 2018, 117(1): 309 – 319.

[8] GARCIA D, FUENTE M, SAINZ G. Fault detection and isolation in transient states using principal component analysis. *Journal of Process Control*, 2012, 22(3): 551 – 563.

[9] ZHANG Jialiang, CAO Jianfu, GAO Feng, et al. Fault diagnosis of complex system based on nonlinear spectrum and kernel principal component analysis. *Control Theory & Applications*, 2012, 29(12): 1558 – 1564.
(张家良, 曹建福, 高峰, 等. 结合非线性频谱与核主元分析的复杂系统故障诊断方法. 控制理论与应用, 2012, 29(12): 1558 – 1564.)

[10] DUFRENOIS F. A one-class kernel fisher criterion for outlier detection. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(5): 982 – 994.

[11] HAN M, JIANG L. Endpoint prediction model of basic oxygen furnace steelmaking based on PSO–ICA and RBF neural network. *Proceedings of International Conference on Intelligent Control and Information Processing*. Dalian: IEEE, 2010: 388 – 393.

[12] XI Jianhui, HAN Min. Prediction of multivariate time series based on principal component analysis and neural networks. *Control Theory & Applications*, 2007, 24(5): 719 – 724.
(席剑辉, 韩敏. 主成分分析与神经网络的结合在多变量序列预测中的应用. 控制理论与应用, 2007, 24(5): 719 – 724.)

[13] REN L, XU Z, YAN X. Single-sensor incipient fault detection. *IEEE Sensors Journal*, 2011, 11(9): 2102 – 2107.

[14] HAN Min, ZHANG Zhankui. Fault detection method based on weighted kernel independent component analysis. *Control and Decision*, 2016, 31(2): 242 – 248.
(韩敏, 张占奎. 基于加权核独立成分分析的故障检测方法. 控制与决策, 2016, 31(2): 242 – 248.)

[15] ZHAO Xiaoli, ZHAO Rongzhen. A method of dimension reduction of rotor faults data set based on fusion of global and local discriminant information. *Acta Automatica Sinica*, 2017, 43(4): 560 – 567.
(赵孝礼, 赵荣珍. 全局与局部判别信息融合的转子故障数据集降维方法研究. 自动化学报, 2017, 43(4): 560 – 567.)

[16] WANG Dingcheng, JIANG Bin. OS-LSSVMR internal model control for nonlinear uncertain systems. *Control Theory & Applications*, 2008, 25(5): 905 – 907.
(王定成, 姜斌. 非线性不确定系统的OS-LSSVMR内模控制. 控制理论与应用, 2008, 25(5): 905 – 907.)

[17] ZHANG J, LIU Y. Application of complete ensemble intrinsic timescale decomposition and least-square SVM optimized using hybrid DE and PSO to fault diagnosis of diesel engines. *Frontiers of Information Technology and Electronic Engineering*, 2017, 18(2): 272 – 286.

[18] JIANG Q, YAN X. Parallel PCA–KPCA for nonlinear process monitoring. *Control Engineering Practice*, 2018, 80(4): 17 – 25.

[19] LIU X, BO L, LUO H. Bearing faults diagnostics based on hybrid LS–SVM and EMD method. *Measurement*, 2015, 59(7): 145 – 166.

[20] DILMEN E, BEYHAN S. An enhanced online LS–SVM approach for classification problems. *Soft Computing*, 2018, 22(13): 4457 – 4475.

[21] DOWNS J, VOGEL E. A plant-wide industrial process control problem. *Computers & Chemical Engineering*, 1993, 17(3): 245 – 255.

作者简介:

李锦冰 硕士研究生, 研究方向为故障诊断和预测, E-mail: ljs@mail.dlut.edu.cn;

韩 冰 博士, 硕士生导师, 研究方向为船舶动力平台故障诊断和故障预测, E-mail: hanbing@sssri.com;

冯守渤 博士研究生, 研究方向为神经网络与数据挖掘, E-mail: fsb@mail.dlut.edu.cn;

张佳冬 本科生, 研究方向为神经网络与数据挖掘, E-mail: zhangjiadong@mail.dlut.edu.cn;

李 宇 硕士研究生, 研究方向为特征提取与故障诊断, E-mail: liyu0512@mail.dlut.edu.cn;

钟 凯 博士研究生, 研究方向为数据挖掘与过程监控, E-mail: zhongkai0402@mail.dlut.edu.cn;

韩 敏 博士, 博士生导师, 研究方向为复杂系统建模与预测等, E-mail: minhan@dlut.edu.cn.