

【电子与信息科学 / Electronics and Information Science】

# 基于KPCA-SVM的相关和独立变量故障检测方法

郭金玉, 于欢, 李元

沈阳化工大学信息工程学院, 辽宁沈阳 110142

**摘要:** 实际工业过程中有一些独立于其他变量的过程变量, 为能够分别检测独立变量与相关变量, 提出一种基于核主成分分析(kernel principal component analysis, KPCA)和支持向量机(support vector machine, SVM)相结合的相关和独立变量故障检测方法, 记为KPCA-SVM. 首先, 应用基于互信息的变量划分策略, 通过计算变量之间的互信息将过程变量划分为相关变量和独立变量. 然后, 在相关变量空间和独立变量空间分别建立KPCA和SVM模型对测试数据进行监测. 与传统KPCA和SVM算法相比, KPCA-SVM方法结合了KPCA在检测相关变量和SVM方法在检测独立变量上的优势, 提高了KPCA和SVM方法的故障检测性能. 最后, 将KPCA-SVM方法应用于田纳西-伊斯曼(Tennessee-Eastman, TE)工业过程进行故障检测, 并与KPCA、核熵成分分析(kernel entropy component analysis, KECA)和SVM方法进行比较. 仿真结果表明, KPCA-SVM方法具有较好的检测效果, 对于多种故障的检测效果有所提升, 其中对于微小故障5的检测效果有明显提升, 进一步验证KPCA-SVM方法的有效性.

**关键词:** 自动控制技术; 核主成分分析; 支持向量机; 故障检测; 相关变量; 独立变量

**中图分类号:** TP277

**文献标志码:** A

**doi:** 10. 3724/SP. J. 1249. 2023. 01014

## Related and independent variable fault detection method based on KPCA-SVM

GUO Jinyu, YU Huan, and LI Yuan

College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, Liaoning Province, P. R. China

**Abstract:** In the real industrial process, some process variables are independent of other variables, a fault detection method of related and independent variable based on kernel principal component analysis and support vector machine (KPCA-SVM) is proposed to detect these independent variables separately from related variables. Firstly, a variable division strategy based on mutual information is applied to divide the process variables into related variables and independent variables by calculating the mutual information between variables. Then, KPCA and SVM models are established in the related variable space and the independent variable space to monitor the test data. Compared with the traditional KPCA and SVM methods, the KPCA-SVM method combines the advantages of KPCA in detecting related variables and SVM methods in detecting independent variables, and improves the fault detection performance of KPCA and SVM methods. Finally, the KPCA-SVM method is applied to the Tennessee-Eastman (TE) industrial process for fault detection, and compared with KPCA, kernel entropy component analysis (KECA) and SVM methods. The results show that the proposed KPCA-SVM method has a good detection effect and improves the detection effect of multiple faults, among which the detection effect of minor fault 5 is significantly improved, which

**Received:** 2022-04-24; **Accepted:** 2022-10-09; **Online (CNKI):** 2022-12-16

**Foundation:** National Natural Science Foundation of China (62273242)

**Corresponding author:** Associate professor GUO Jinyu. E-mail: shandong401@sina.com

**Citation:** GUO Jinyu, YU Huan, LI Yuan. Related and independent variable fault detection method based on KPCA-SVM [J]. Journal of Shenzhen University Science and Engineering, 2023, 40(1): 14-21. (in Chinese)



further verifies the effectiveness of the KPCA-SVM method.

**Key words:** automatic control technology; kernel principal component analysis; support vector machine; fault detection; related variables; independent variables

随着现代工业规模不断发展扩大, 对工业生产过程进行故障监测和诊断, 在保证生产安全和提高产品质量等方面发挥了重要作用. 过程监控日益受到学术界和工业领域研究人员的广泛关注<sup>[1-2]</sup>. 在各种监测方法中, 数据驱动方法, 尤其是多元统计过程监控(multivariate statistical process monitoring, MSPM)近年来得到了迅速发展<sup>[3-5]</sup>. 一些具有代表性的方法, 如主成分分析(principal component analysis, PCA)、典型变量分析(canonical variate analysis, CVA)和偏最小二乘(partial least squares, PLS)方法等, 已被广泛用于过程监控和故障诊断<sup>[6]</sup>.

作为最普遍的MSPM方法, PCA已经得到了深入的研究. PCA方法将线性数据矩阵投影到含有所有相关信息的低维子空间上实现降维, 同时保留了重要信息. PCA方法使用平方预测误差(squared prediction error, SPE)和Hotellings  $T^2$ 统计量对样本进行监控<sup>[7]</sup>. 然而, 工业过程中数据变量之间的非线性和动态性会导致传统PCA方法的检测性能较差. 针对此类非线性问题, SCHÖLKOPF等<sup>[8]</sup>提出核主成分分析(kernel principal component analysis, KPCA)方法, 通过使用核函数避免了复杂的非线性优化过程; LEE等<sup>[9]</sup>首先将KPCA应用于故障检测; CHOI等<sup>[10]</sup>开发了用于故障识别的KPCA贡献图; ZHANG等<sup>[11]</sup>将小波分解技术与滑动中值滤波相结合, 提出改进的KPCA方法; FAN等<sup>[12]</sup>提出的核独立成分分析-主成分分析(independent component analysis-principal component analysis, KICA-PCA)方法利用了局部数据结构分析的同时考虑了非线性和非高斯特性; JENSSEN<sup>[13]</sup>提出核熵成分分析(kernel entropy component analysis, KECA)方法, 并用于无高斯假设的非线性工业过程故障检测. 然而, 传统的PCA、KPCA和PLS方法都是基于全局模型, 忽略了局部结构信息. DENG等<sup>[14]</sup>提出一种2步局部KPCA(two-step localized KPCA, TSLKPCA)方法, 通过设计一种2步局部信息挖掘框架, 在提取核成分的同时, 保留了全局和局部数据结构信息, 有效地提高了早期故障诊断的性能. 邓晓刚等<sup>[15]</sup>提出一种基于双层局部KPCA的非线性检测方法, 将过程变量划分为多个变量块, 通过构建函数挖掘样本局

部信息, 提高故障检测能力. 为改善这些全局方法在故障检测中的不足, 几种局部方法, 如等距特征映射、局部线性嵌入和局部保持投影(locality preserving projections, LPP)<sup>[16]</sup>在故障检测中被广泛使用. 这些方法的主要思想是寻找数据集中存在的局部结构. LPP是一种线性投影方法, 可以保留数据的局部特征, 提高故障检测的性能. ZHANG等<sup>[17]</sup>提出一种全局-局部结构(global-local structure analysis, GLSA)方法, 利用底层局部几何流形的同时保留了全局信息. 该方法结合了LPP和PCA的优点.

在实际应用中, 正常条件下的数据较多, 而关于故障状态的信息较少. 在故障样本有限的情况下, 提高学习能力具有重要意义. 支持向量机(support vector machine, SVM)<sup>[18]</sup>是一种经典的模式分类方法, 对小样本、非线性数据进行故障检测时效果更显著, 在解决高维模式识别和函数拟合等问题时具有许多优势, 是统计学习理论中结构风险最小化原则的一种实现<sup>[19]</sup>. SVM将采集到的数据通过函数映射到高维特征空间, 并在高维特征空间中建立最优分离超平面, 有效地解决了非线性数据分类问题. XUE等<sup>[20]</sup>构建了一个先进的故障检测与诊断框架, 通过引入高斯核支持向量机递归特征消除方法, 提取非线性特征并选择出最优参数, 该方法用于解决故障检测中的非线性问题. 然而, 这种方法没有考虑过程变量间的相关性, 这会影响SVM的故障检测性能.

HUANG等<sup>[21]</sup>提出一种相关和独立变量划分策略并运用不同方法检测这2类变量. 本研究将相关和独立变量的特征作为变量划分的基础, 提出一种基于KPCA-SVM的相关和独立变量故障检测方法. 传统的KPCA方法假设变量之间的关系是非线性的, 但在工业过程中, 有些变量是独立于其他变量的, 这类变量应被单独分离和监控. 首先, 采用互信息来判断变量是否独立. 互信息值接近0表示两个变量之间不相关. 互信息向量是由变量与其他变量之间的互信息生成的, 如果互信息向量中的每个值都接近于0, 则认为该变量是独立的. 对每个变量进行检验后, 可完成独立变量与相关变量的划分. 然后, 对独立变量和相关变量分别采用SVM和KPCA进行故障检测.

# 1 KPCA 和 SVM 方法原理

## 1.1 KPCA 原理

PCA 是一种线性降维方法, 在处理非线性相关的数据时不能很好地发挥作用. KPCA 的思想是将原始输入数据通过非线性核函数映射到高维特征空间, 在高维特征空间中计算核矩阵的特征值和特征向量. KPCA 方法可以对非线性数据进行有效监控.

设输入数据矩阵  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T \in \mathbf{R}^{n \times m}$ , 其中,  $\mathbf{x}_i (i = 1, 2, \dots, n) \in \mathbf{R}^m$ ,  $n$  为样本数,  $m$  为变量数. 利用非线性映射函数  $\Phi$  将原始空间的数据映射到高维特征空间  $F$ , 则特征空间中样本的协方差矩阵为

$$C = \frac{1}{n} \sum_{i=1}^n \Phi(\mathbf{x}_i) [\Phi(\mathbf{x}_i)]^T \quad (1)$$

设  $C$  的特征值为  $\lambda$ , 特征向量为  $V$ , 则有

$$\lambda V = CV = \left\{ \frac{1}{n} \sum_{i=1}^n \Phi(\mathbf{x}_i) [\Phi(\mathbf{x}_i)]^T \right\} V = \frac{1}{n} \sum_{i=1}^n \langle \Phi(\mathbf{x}_i), V \rangle \Phi(\mathbf{x}_i) \quad (2)$$

其中, 符号  $\langle \rangle$  表示内积;  $V$  在特征空间中表示为

$$V = \sum_{i=1}^n \alpha_i \Phi(\mathbf{x}_i) \quad (3)$$

其中,  $\alpha_i$  为线性相关系数.

在式(2)等号两边同时用  $\Phi(\mathbf{x}_k) (k = 1, 2, \dots, n)$  做内积, 结合式(3)可得

$$\lambda \sum_{i=1}^n \alpha_i \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_k) \rangle = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle \langle \Phi(\mathbf{x}_k), \Phi(\mathbf{x}_j) \rangle \quad (4)$$

为避免高维运算, 引入  $n \times n$  维的核矩阵

$$K_{ij} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle \quad (5)$$

内积可通过引入核函数计算. 本研究采用高斯核函数

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp \left( \frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2} \right) \quad (6)$$

其中,  $\sigma$  为窗口宽度. 则式(4)可简化为

$$\lambda \sum_{i=1}^n \alpha_i K_{ik} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \alpha_i K_{ij} K_{kj} \quad (7)$$

对核矩阵  $K = (K_{ij}) \in \mathbf{R}^{n \times n}$  进行中心化处理, 可得

$$\bar{K} = K - IK - KI + IKI \quad (8)$$

其中,  $I$  为系数为  $1/n$  的  $n$  阶单位矩阵.

将式(7)进一步简化, 可得

$$\lambda \alpha = \frac{1}{n} K \alpha \quad (9)$$

为了保证  $V_i^T V_i = 1, i = 1, 2, \dots, n$ . 由式(3)和式(9)得到  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$  归一化的条件为

$$\alpha_i^T \alpha_i = \frac{1}{\lambda_i}, \quad i = 1, 2, \dots, n \quad (10)$$

核主成分的选取取决于  $\lambda$  的值, 本研究仅选取前  $k (k < n)$  个较大的特征值对应的特征向量, 实现对数据的降维,  $k$  个特征向量组成的矩阵记为  $V_l = [V_1, V_2, \dots, V_k]$ .

提取测试样本  $\mathbf{x}_{\text{new}}$  的核主成分, 计算其在特征向量上的投影为

$$t_l = \langle V_l, \Phi(\mathbf{x}_{\text{new}}) \rangle = \sum_{i=1}^n \alpha_i \langle \Phi(\mathbf{x}_{\text{new}}), \Phi(\mathbf{x}_i) \rangle \quad l = 1, 2, \dots, k \quad (11)$$

测试样本的  $T^2$  和 SPE 统计量分别为

$$T^2 = [t_1, t_2, \dots, t_k] \Lambda^{-1} [t_1, t_2, \dots, t_k]^T \quad (12)$$

$$\text{SPE} = \sum_{i=1}^n t_i^2 - \sum_{i=1}^k t_i^2 \quad (13)$$

其中,  $\Lambda$  为对角矩阵, 其元素是核矩阵  $\bar{K}$  的前  $k$  个较大的特征值.

## 1.2 SVM 原理

SVM 是机器学习中经典的分类方法之一, 通过构造一个决策超平面, 对正负样本进行有效分类, 以最大超平面分离得到最好的泛化能力. 在处理线性数据时, SVM 能够构建最大间隔超平面并对数据进行分类. 给定一组训练数据样本  $\{(\mathbf{x}_i, y_i), i = 1, 2, \dots, n\}$ ,  $\mathbf{x}_i \in \mathbf{R}^m, y_i \in \{-1, +1\}$ , 则最优超平面将由少数样本(支持向量)的最近点决定. 超平面  $(\omega, b)$  的基本形式为

$$\omega \mathbf{x} + b = 0 \quad (14)$$

其中,  $\omega$  为特征空间中分离超平面的法向量;  $b$  为位移项. 若  $(\omega, b)$  能够正确分类样本, 则有

$$\begin{cases} \omega \mathbf{x}_i + b \geq 1, & y_i = 1 \\ \omega \mathbf{x}_i + b \leq -1, & y_i = -1 \end{cases} \quad (15)$$

其中,  $\omega$  为特征空间中分离超平面的法向量;  $b$  为位移项.

2 个类别的支持向量到超平面的间隔距离之和为  $\gamma$ . 为找到能使分类间隔最大的超平面, 对分离超平面 SVM 的优化问题进行归一化. SVM 的基本形式为

$$\begin{cases} \min \left( \frac{1}{2} \|\omega\|^2 \right) \\ \text{s. t. } y_i(\omega x_i + b) \geq 1, \quad i = 1, 2, \dots, n \end{cases} \quad (16)$$

为降低SVM模型对噪声点的敏感性, 提高模型的泛化能力, 引入惩罚因子  $C > 0$  和松弛变量  $\xi_i \geq 0$ , 则SVM基本形可改写为

$$\begin{cases} \min \left( \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \right) \\ \text{s. t. } y_i(\omega x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, n \end{cases} \quad (17)$$

为求解式(17)问题, 将该原始问题转化为对偶问题, 构造拉格朗日函数

$$L(\omega, \alpha, b, \mu, \xi) = \frac{1}{2} \|\omega\|^2 + \sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i [1 - \xi_i - y_i(\omega x_i + b)] - \sum_{i=1}^n \mu_i \xi_i \quad (18)$$

其中, 拉格朗日乘子  $\alpha_i \geq 0$ ,  $\mu_i \geq 0$ .  $L(\omega, \alpha, b, \mu, \xi)$  对  $\omega$ 、 $b$  和  $\xi_i$  的偏导, 令其结果等于0, 求解式(18), 可得

$$f(x') = \omega x' + b = \sum_{i=1}^n \alpha_i y_i x_i^T x' + b \quad (19)$$

处理非线性样本问题时, 通过非线性映射函数  $\Phi$  将输入数据映射到高维空间并建立最大超平面. 通过定义合适的内积核函数可避免高维运算.

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle \quad (20)$$

引入核函数后的SVM模型表示为

$$f(x') = \text{sgn} \left[ \sum_{i,j=1}^n \alpha_i y_i K(x_i, x_j) + b \right] \quad (21)$$

其中,  $\text{sgn}$  为符号函数.

## 2 基于KPCA-SVM的故障检测方法

### 2.1 相关变量和独立变量的划分

传统的故障检测方法通常假设变量之间是相关的. 然而, 在实际工业过程中, 各变量之间关系复杂, 有些是线性相关, 有些是非线性相关, 还有些是相互独立的或弱相关. 因此, 应该对独立变量和相关变量进行分离和独立监测. 变量划分可以突出相关变量之间的强相关性, 消除不相关变量对检测结果的影响.

给定数据矩阵  $X = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^{n \times m}$ , 采用互信息探讨任意两个变量之间的相关关系. 互信息对线性和非线性依赖关系敏感, 可估计二阶或高阶关系. 互信息作为衡量随机变量间独立性的指

标, 由其理论背景信息定义<sup>[21]</sup>. 变量之间的相关程度可由互信息的值反映. 当且仅当2个变量相互独立时, 互信息等于0.

变量  $x_i$  的香农熵估计为

$$H(x_i) = - \int p(x_i) \lg p(x_i) dx_i \quad (22)$$

其中,  $p(x_i)$  是变量  $x_i$  的概率密度. 因此, 两变量  $x_1$  和  $x_2$  互信息为

$$I(x_1, x_2) = \iint P(x_1, x_2) \lg \left[ \frac{P(x_1, x_2)}{P(x_1)P(x_2)} \right] dx_1 dx_2 \quad (23)$$

用  $x_1$  和  $x_2$  的熵表示式(23), 则有

$$I(x_1, x_2) = H(x_1) + H(x_2) - H(x_1, x_2) \quad (24)$$

其中,  $H(x_1, x_2)$  是  $x_1$  和  $x_2$  的联合熵.

对变量  $x_i \in \mathbb{R}^m$  建立互信息向量  $p_i \in \mathbb{R}^{m-1}$ , 其中元素是变量  $x_i$  与其他  $m-1$  变量之间的互信息值. 计算  $p_i$  的范数为  $D_i = \|p_i\|$ .

生成随机矩阵  $R = [r_1, r_2, \dots, r_{m-1}]^T \in \mathbb{R}^{n \times (m-1)}$ , 其中,  $r_i (i = 1, 2, \dots, m-1)$  是零均值单位方差的高斯分布数据. 计算  $x_i$  与  $R$  中其余  $m-1$  个变量的互信息得到互信息向量  $q_i \in \mathbb{R}^{m-1}$ .  $R$  中任意两个变量都是不相关的, 且任意  $r_i$  都独立于原始数据矩阵  $X$  中的变量  $x_i$ . 这意味着  $q_i$  中每个元素的值都接近0. 计算  $q_i$  的范数  $N_i = \|q_i\|$ ,  $D_i < N_i$  表示变量  $x_i$  独立于其他变量, 此时  $x_i$  要分到独立变量矩阵  $X_I$  中, 其余变量则分到相关变量矩阵  $X_R$  中. 通过独立变量的划分将原始数据分成  $X_I$  和  $X_R$  两部分, 即  $X = [X_I, X_R]$ .

### 2.2 KPCA-SVM的故障检测步骤

将原始数据矩阵  $X$  划分为相关变量矩阵  $X_I$  和独立变量矩阵  $X_R$ ,  $X_R$  的变量之间包括线性和非线性相关关系, KPCA 对非线性相关问题有很好的处理效果. 利用 KPCA 对相关变量进行检测, 分别得到  $T^2$  和 SPE 的控制限  $T_{lim}^2$  和  $SPE_{lim}$ . 对于独立变量, 采用 SVM 进行分类. 最终, KPCA-SVM 的统计量 KS 定义为

$$KS = \begin{cases} \text{正常,} & \text{SVM 标签为0,} \\ & T^2 \leq T_{lim}^2 \text{ 和 } SPE \leq SPE_{lim} \\ \text{故障,} & \text{其他} \end{cases} \quad (25)$$

离线建模流程为

1) 在正常条件下, 采集训练数据矩阵  $X$  并进行标准化;

2) 对于  $X$  中的每个变量  $x_i$ , 计算对应的互信息向量  $p_i$  和  $q_i$ ;

- 3) 计算  $D_i$  和  $N_i$ , 并划分相关变量和独立变量;
- 4) 将  $X$  划分为相关变量矩阵  $X_R$  和独立变量矩阵  $X_I$ ;
- 5) 建立相关变量矩阵  $X_R$  的 KPCA 模型, 运用核密度估计得到控制限  $T_{lim}^2$  和  $SPE_{lim}$ ;
- 6) 使用训练数据和部分故障数据的独立变量矩阵生成 SVM 模型的训练数据, 并对训练数据贴标签, 其中正常数据标签为 0, 故障数据标签为 1, 用其训练 SVM 模型.

在线故障检测流程为

- 1) 收集测试数据  $Y$ , 使用离线建模数据的均值和标准差进行标准化, 再将测试数据划分为相关变量矩阵  $Y_R$  和独立变量矩阵  $Y_I$ ;
- 2) 将  $Y_R$  投影到 KPCA 模型上, 计算统计量  $T^2$  和  $SPE$  进行故障检测;
- 3) 运用 SVM 模型对  $Y_I$  进行故障分类, 分类标签为 0 的数据为正常数据, 而分类标签为 1 的数据为故障数据;
- 4) 根据式(25)得到最终统计量, 计算故障检测率.

### 3 仿真实验

#### 3.1 TE 过程

目前, 田纳西-伊斯曼(Tennessee-Eastman, TE)过程<sup>[22]</sup>已被广泛用于检验故障检测方法的效果, 包括 5 个主要单元, 在化学反应中存在 4 种反应物  $A$ 、 $C$ 、 $D$ 、 $E$  和惰性气体  $B$ , 产生产物  $G$ 、 $H$  和副产物  $F$ . 在 TE 过程仿真数据集中有 22 个连续测量和 12 个操纵变量. 基准数据集包括 33 个变量和 960 个样本. 其中, 第 161 ~ 960 个样本产生了 21 种不同类型的故障, 本研究选用其中 12 种过程故障(表 1)来验证 KPCA-SVM 方法的有效性.

#### 3.2 数据采集

以 TE 工业过程为背景, 选取 500 个正常样本作为 KPCA 的训练数据集, 选取后 300 个故障样本作为测试数据集. 基于互信息变量划分策略的变量划分结果如图 1. 由图 1 可见, 变量 1、7、9、10、12、13、15、16、17、18、19、20、25、27、28、29、30、31 和 33 为相关变量, 其余变量为独立变量. KPCA、KECA 以及 KPCA-SVM 方法的置信度设为 0.99, 通过寻优测试设置 KPCA 和 KECA 中核函数参数  $\sigma = 2.500$ . SVM 和 KPCA-SVM 方法采用

表 1 12 种 TE 过程故障  
Table 1 12 faults of TE process

| 编号 | 描述                             | 类型   |
|----|--------------------------------|------|
| 1  | $A/C$ 进料比率变化, $B$ 成分不变(流 4)    | 阶跃   |
| 2  | $B$ 成分发生变化, $A/C$ 进料比率不变(流 4)  | 阶跃   |
| 3  | $D$ 的进料温度发生变化(流 2)             | 阶跃   |
| 5  | 冷凝器冷却水的入口温度发生变化                | 阶跃   |
| 8  | $A$ 、 $B$ 和 $C$ 的进料成分发生变化(流 4) | 随机变化 |
| 10 | $C$ 的进料温度发生变化(流 2)             | 随机变化 |
| 12 | 冷凝器冷却水的入口温度发生变化                | 随机变化 |
| 13 | 反应动态特性发生变化                     | 慢漂移  |
| 15 | 冷凝器冷却水阀门粘滞                     | 粘滞   |
| 18 | 未知                             | 未知   |
| 19 | 未知                             | 未知   |
| 20 | 未知                             | 未知   |

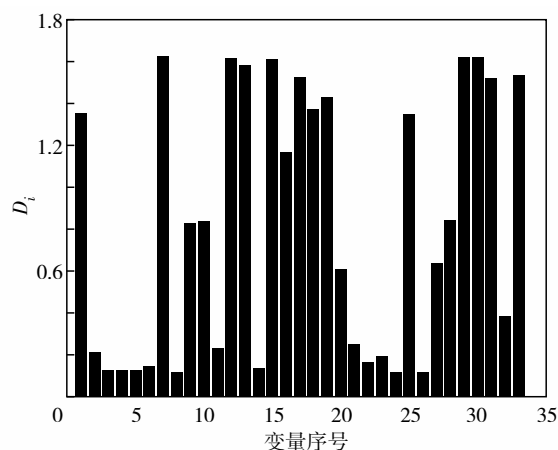


图 1 TE 过程变量划分柱状图

Fig. 1 Bar graph of variable division in TE process.

高斯核函数处理非线性数据, 通过寻优测试设置惩罚因子  $C = 0.25$ , 窗宽  $= 0.05$ .

#### 3.3 仿真结果

故障 3 是反应物  $D$  的进料温度出现阶跃变化, KPCA、KECA、SVM 和 KPCA-SVM 四种方法对故障 3 的后 300 个故障样本的检测效果如图 2.

KECA 方法中柯西-施瓦茨(Cauchy-Schwarz, CS)统计量表示了角度信息, 比 KPCA 方法的  $SPE$  统计量检测效果有所提升. 传统的 SVM 方法在训练时加入了故障数据, 能够有效学习到故障的变化特征, 因此检测效果优于 KPCA 和 KECA. KPCA-SVM 方法在独立变量空间中的故障检测效果比传统 SVM 有明显提升, 说明剔除相关变量后可改进传统 SVM 的检测效果.

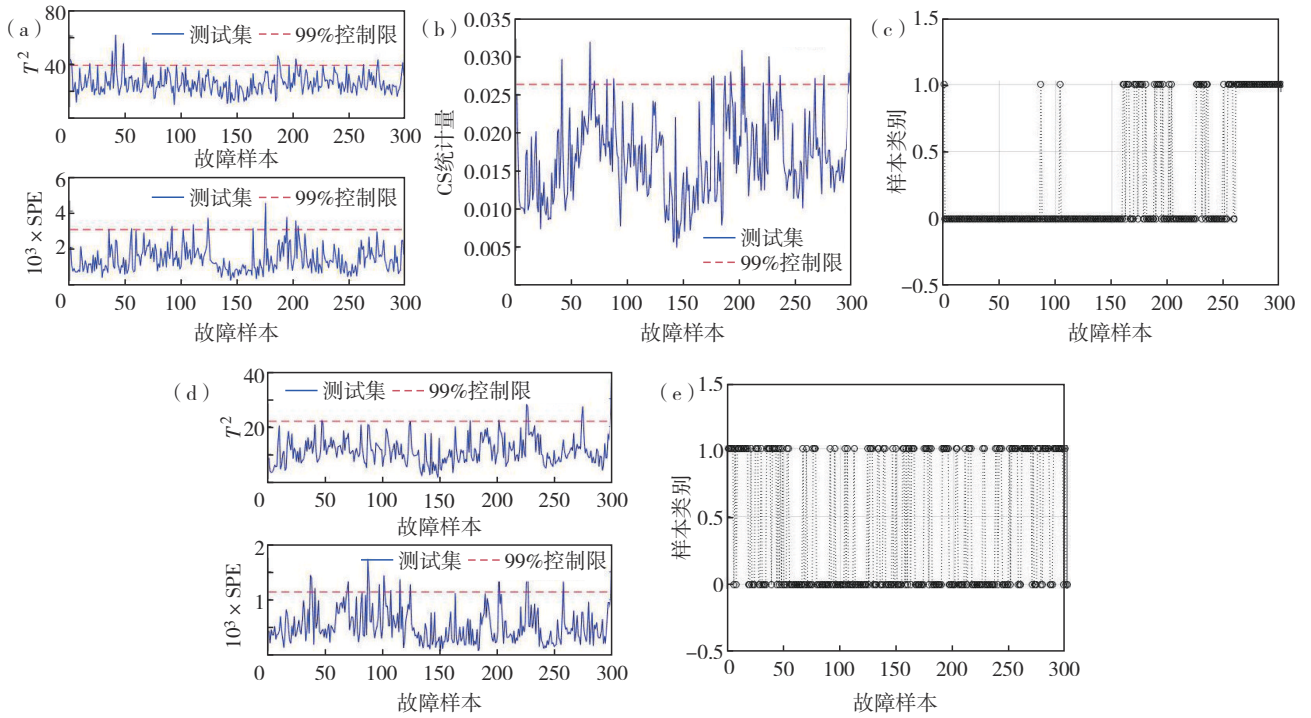


图2 (a) KPCA、(b) KECA、(c) SVM、(d) KPCA-SVM检测相关变量和(e) KPCA-SVM检测独立变量方法对故障3的检测结果  
Fig. 2 (Color online) Detection results of (a) KPCA, (b) KECA, (c) SVM, (d) KPCA-SVM detecting related variables and (e) KPCA-SVM detecting independent variables methods for fault 3. In (a), (b) and (d), the solid line is for test set and the dash line is for 99% control limit.

故障5为冷凝器的冷却水入口温度出现阶跃变化, 冷却水的流量也随之发生变化。但是, 由于控制系统的存在, 阶跃变化会在一段时间后得到补

偿。本研究选择后300个故障样本进行检测, 此时故障已经非常微小。KPCA、KECA、SVM和KPCA-SVM四种方法对故障5的检测效果如图3。

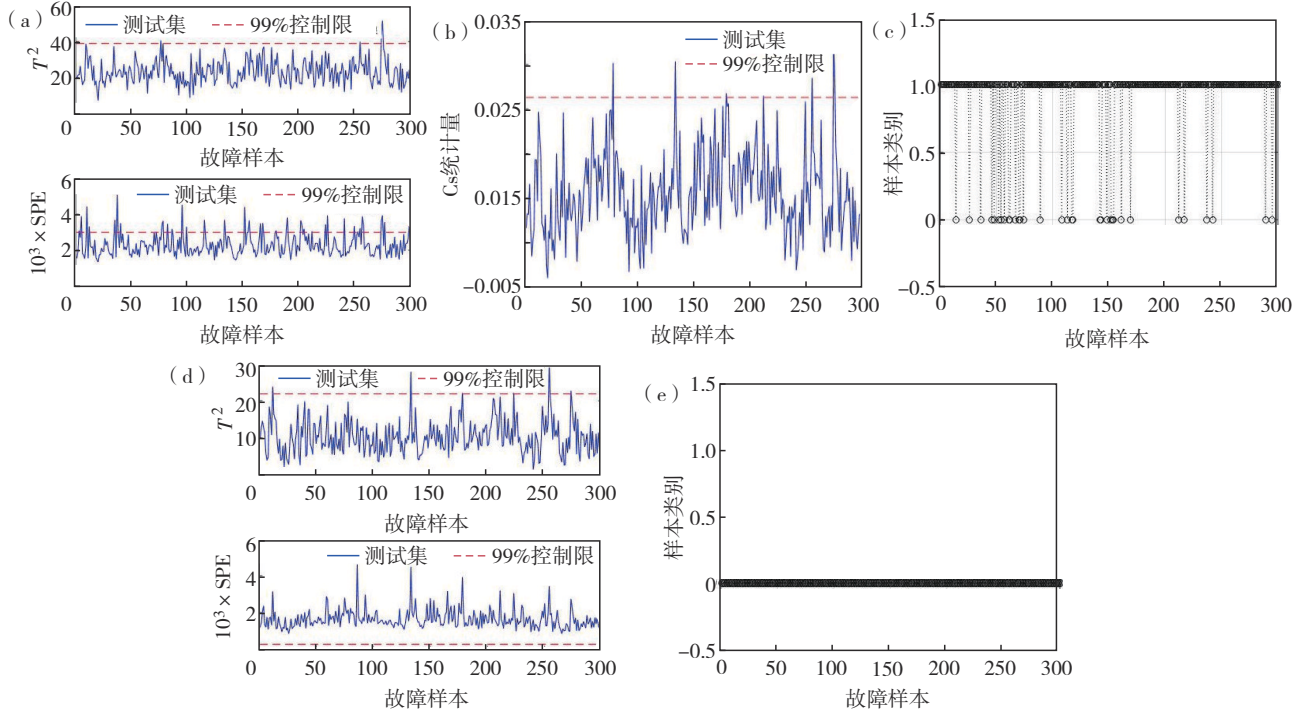


图3 (a) KPCA、(b) KECA、(c) SVM、(d) KPCA-SVM检测相关变量和(e) KPCA-SVM检测独立变量方法对故障5的检测结果  
Fig. 3 (Color online) Detection results of (a) KPCA, (b) KECA, (c) SVM, (d) KPCA-SVM detecting related variables and (e) KPCA-SVM detecting independent variables methods for fault 5. In (a), (b) and (d), the solid line is for test set and the dash line is for 99% control limit.

传统 KPCA 方法的故障检测效果并不理想, 在控制限上方的故障样本数较少, 说明在故障补偿时 KPCA 方法不能很好地检测到故障. KECA 的检测效果好于 KPCA 的  $T^2$ , 但是低于 KPCA 的 SPE. 由于故障 5 是微小故障, 统计分析方法很难检测出来. 传统 SVM 方法由于在训练时加入了故障数据, 能够有效地学习到故障的变化特征, 检测效果优于 KPCA 和 KECA. 相比 KPCA、KECA 和 SVM 方法, KPCA-SVM 方法在相关变量空间中的 SPE 统计量总是超出控制限, 因为 KPCA 方法能够很好地去除变

量的相关性. 对比图 3(a)和(d), 可以明显看出, KPCA-SVM 方法对相关变量数据的检测效果更好, 并且能够检测到微小故障.

表 2 是 4 种故障检测方法对不同故障类型的故障检测率对比. 由表 2 可见, KPCA-SVM 算法的故障检测率部分有明显提升, 表明该方法整体检测效果较好. 其中, 对故障 1、2、5、8、12、13 和 18 的故障检测率达到 100%; 对故障 3、5、10、15、19 和 20 的故障检测率都有不同程度的提升.

表 2 KPCA、KECA、SVM 和 KPCA-SVM 方法对 TE 过程故障检测率对比<sup>1)</sup>

Table 2 Comparison of fault detection rates of KPCA, KECA, SVM, and KPCA-SVM for TE process

%

| 故障<br>类型 | KPCA   |        | KECA   | SVM    | KPCA-SVM |        |        |        |
|----------|--------|--------|--------|--------|----------|--------|--------|--------|
|          | $T^2$  | SPE    | CS     |        | $T^2$    | SPE    | SVM    | KS     |
| 1        | 100    | 100    | 100    | 100    | 100      | 100    | 100    | 100    |
| 2        | 100    | 100    | 100    | 100    | 100      | 100    | 100    | 100    |
| 3        | 6. 67  | 3. 33  | 6. 67  | 26. 67 | 3. 33    | 5. 67  | 41. 00 | 45. 33 |
| 5        | 1. 33  | 13. 00 | 2. 33  | 89. 00 | 2. 00    | 100    | 0      | 100    |
| 8        | 100    | 100    | 100    | 55. 67 | 100      | 100    | 18. 00 | 100    |
| 10       | 72. 33 | 82. 33 | 73. 33 | 71. 00 | 71. 67   | 87. 33 | 31. 67 | 89. 00 |
| 12       | 91. 00 | 100    | 100    | 100    | 95. 67   | 100    | 93. 67 | 100    |
| 13       | 84. 00 | 100    | 100    | 86. 33 | 98. 67   | 100    | 72. 67 | 100    |
| 15       | 36. 33 | 43. 67 | 53. 67 | 67. 33 | 37. 33   | 55. 33 | 35. 67 | 74. 33 |
| 18       | 0      | 100    | 100    | 100    | 0        | 100    | 100    | 100    |
| 19       | 29. 67 | 66. 00 | 18. 67 | 41. 67 | 10. 33   | 73. 30 | 29. 33 | 84. 00 |
| 20       | 62. 00 | 75. 33 | 60. 33 | 55. 33 | 64. 00   | 78. 33 | 47. 00 | 83. 33 |

<sup>1)</sup>灰底数据为该故障类型下所得故障检测率显著高于其他算法

## 结 语

提出一种基于 KPCA-SVM 的故障检测方法, 采用基于互信息的变量划分策略将原始数据变量划分为相关变量和独立变量, 运用 KPCA 和 SVM 方法分别检测相关变量空间和独立变量空间. 将 KPCA-SVM 方法应用到 TE 过程中, 仿真结果表明, 与 KPCA、KECA 和 SVM 方法相比, KPCA-SVM 方法有效提高了故障检测率. 本研究主要研究 KPCA-SVM 方法在工业过程故障检测的实现与仿真, 然而, 实际工业中的过程数据变量之间关系往往更加复杂, 对检测方法要求更高, 未来可对变量进行更细致地划分并分别采用不同的方法检测, 以提升方法的检测效果. 本研究尚缺少实际工业数据的检测结果, 这也是未来研究工作的方向.

基金项目: 国家自然科学基金资助项目 (62273242)

作者简介: 郭金玉 (1975—), 沈阳化工大学副教授、博士. 研究方向: 故障诊断、生物特征识别算法及应用.

E-mail: 969554959@qq.com

引 文: 郭金玉, 于欢, 李元. 基于 KPCA-SVM 的相关和独立变量故障检测方法[J]. 深圳大学学报理工版, 2023, 40 (1): 14-21.

## 参考文献 / References:

- [1] TAO Yang, SHI Hongbo, SONG Bing, et al. A novel dynamic weight principal component analysis method and hierarchical monitoring strategy for process fault detection and diagnosis [J]. IEEE Transactions on Industrial Electronics, 2020, 67(9): 7994-8004.
- [2] LAN Ting, TONG Chudong, YU Haizhen, et al. Nonlinear process monitoring based on decentralized generalized regression neural networks [J]. Expert Systems with Applications, 2020, 150: 113273.

- [ 3 ] GE Zhiqiang. Review on data-driven modeling and monitoring for plant-wide industrial processes [J]. *Chemometrics and Intelligent Laboratory Systems*, 2017, 171: 16-25.
- [ 4 ] HE Q P, WANG Jin. Statistical process monitoring as a big data analytics tool for smart manufacturing [J]. *Journal of Process Control*, 2018, 67: 35-43.
- [ 5 ] WANG Youqing, SI Yabin, HUANG Biao, et al. Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008—2017 [J]. *The Canadian Journal of Chemical Engineering*, 2018, 96(10): 2073-2085.
- [ 6 ] LI Ning, GUO Shaojun, WANG Youqing. Weighted preliminary-summation-based principal component analysis for non-Gaussian processes [J]. *Control Engineering Practice*, 2019, 87: 122-132.
- [ 7 ] 张成, 郭青秀, 冯立伟, 等. 基于局部保持投影-加权 $k$ 近邻规则的多模态间歇过程故障检测策略[J]. *控制理论与应用*, 2019, 36(10): 1682-1689.  
ZHANG Cheng, GUO Qingxiu, FENG Liwei, et al. Fault detection strategy based on locality preserving projections-weighted  $k$  nearest neighbors in multimodal batch processes [J]. *Control Theory & Applications*, 2019, 36(10): 1682-1689. (in Chinese)
- [ 8 ] SCHÖLKOPF B, SMOLA A, MÜLLER K R. Nonlinear component analysis as a kernel eigenvalue problem [J]. *Neural Computation*, 1998, 10(5): 1299-1319.
- [ 9 ] LEE J M, YOO C K, CHOI S W, et al. Nonlinear process monitoring using kernel principal component analysis [J]. *Chemical Engineering Science*, 2004, 59(1): 223-234.
- [ 10 ] CHOI S W, LEE C, LEE J M, et al. Fault detection and identification of nonlinear processes based on kernel PCA [J]. *Chemometrics and Intelligent Laboratory Systems*, 2005, 75(1): 55-67.
- [ 11 ] ZHANG Yingwei, LI Shuai, HU Zhiyong. Improved multi-scale kernel principal component analysis and its application for fault detection [J]. *Chemical Engineering Research and Design*, 2012, 90(9): 1271-1280.
- [ 12 ] FAN Jicong, QIN S J, WANG Youqing. Online monitoring of nonlinear multivariate industrial processes using filtering KICA - PCA [J]. *Control Engineering Practice*, 2014, 22: 205-216.
- [ 13 ] JENSSEN R. Kernel entropy component analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(5): 847-860.
- [ 14 ] DENG Xiaogang, CAI Peipei, CAO Yuping, et al. Two-step localized kernel principal component analysis based incipient fault diagnosis for nonlinear industrial processes [J]. *Industrial & Engineering Chemistry Research*, 2020, 59(13): 5956-5968.
- [ 15 ] 邓晓刚, 邓佳伟, 曹玉苹, 等. 基于双层局部KPCA的非线性过程微小故障检测方法[J]. *化工学报*, 2018, 69(7): 3092-3100.  
DENG Xiaogang, DENG Jiawei, CAO Yuping, et al. Double-level local KPCA method for incipient fault detection in nonlinear process [J]. *CIESC Journal*, 2018, 69(7): 3092-3100. (in Chinese)
- [ 16 ] HU Kunlun, YUAN Jingqi. Multivariate statistical process control based on multiway locality preserving projections [J]. *Journal of Process Control*, 2008, 18(7/8): 797-807.
- [ 17 ] ZHANG Muguang, GE Zhiqiang, SONG Zhihuan, et al. Global - local structure analysis model and its application for fault detection and identification [J]. *Industrial & Engineering Chemistry Research*, 2011, 50(11): 6837-6848.
- [ 18 ] VAPNIK V N. An overview of statistical learning theory [J]. *IEEE Transactions on Neural Networks*, 1999, 10(5): 988-999.
- [ 19 ] 郭金玉, 李涛, 李元. 基于DW-ICA-SVM的工业过程故障检测算法[J]. *河北科技大学学报*, 2021, 42(4): 369-379.  
GUO Jinyu, LI Tao, LI Yuan. Fault detection algorithm of industrial process based on DW-ICA-SVM [J]. *Journal of Hebei University of Science and Technology*, 2021, 42(4): 369-379. (in Chinese)
- [ 20 ] XUE Yangtao, ZHANG Li, WANG Bangjun, et al. Nonlinear feature selection using Gaussian kernel SVM-RFE for fault diagnosis [J]. *Applied Intelligence*, 2018, 48(10): 3306-3331.
- [ 21 ] HUANG Jian, YAN Xuefeng. Related and independent variable fault detection based on KPCA and SVDD [J]. *Journal of Process Control*, 2016, 39: 88-99.
- [ 22 ] LYMAN P R, GEORGAKIS C. Plant-wide control of the Tennessee Eastman problem [J]. *Computers & Chemical Engineering*, 1995, 19(3): 321-331.

【中文责编: 英子; 英文责编: 木柯】