

# 控制与决策

Control and Decision

基于核典型相关性-熵成分分析的工业过程质量监测方法

彭开香, 张丽敏

引用本文:

彭开香, 张丽敏. 基于核典型相关性-熵成分分析的工业过程质量监测方法[J]. *控制与决策*, 2021, 36(12): 2999–3006.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.0548>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

[基于局部信息增量与MPLS的质量相关故障检测方法](#)

Quality-related fault detection method based on local information increment and MPLS

*控制与决策*. 2021, 36(7): 1647–1654 <https://doi.org/10.13195/j.kzyjc.2019.1402>

[基于批次图像化的卷积自编码故障监测方法](#)

Fault detection of batch image-based convolutional autoencoder

*控制与决策*. 2021, 36(6): 1361–1367 <https://doi.org/10.13195/j.kzyjc.2019.1342>

[基于自适应混合核典型变量分析的工业过程质量相关故障检测](#)

Quality-related fault detection for industrial processes based on adaptive mixed kernel canonical variable analysis

*控制与决策*. 2021, 36(4): 801–807 <https://doi.org/10.13195/j.kzyjc.2020.0690>

[基于广义主成分分析的重构故障子空间建模方法](#)

Reconstructed fault subspace modelling method based on generalized principal component analysis

*控制与决策*. 2021, 36(4): 808–814 <https://doi.org/10.13195/j.kzyjc.2019.0818>

[基于多块信息提取的AUV资源勘查系统故障检测](#)

Fault detection of AUV resource exploration system based on multi-block information extraction

*控制与决策*. 2021, 36(4): 790–800 <https://doi.org/10.13195/j.kzyjc.2019.0732>

# 基于核典型相关性-熵成分分析的工业过程质量监测方法

彭开香<sup>†</sup>, 张丽敏

(1. 北京科技大学 自动化学院, 北京 100083; 2. 工业过程知识自动化教育部重点实验室, 北京 100083)

**摘要:** 工业过程多变量、数据高维度和非线性特点使得对其质量监测及质量相关的故障诊断变得复杂。融合核熵成分分析(KECA)及典型相关分析(CCA)方法的思想,进行特征提取降维的同时确保所提取特征与质量变量的最大相关性,提出一种新的质量相关的工业过程故障检测方法。首先,采用 KECA 对输入数据进行核空间的映射及特征提取,同时融合 CCA 算法思想使得所提取特征与质量变量间关联最大化;然后,构建监测统计量并用 Parzen 窗估计其控制限,用于过程的故障检测;最后,运用所提方法对带钢热连轧工业过程实际生产数据进行分析,并与其他 4 种传统非线性算法对比分析,实验结果验证了所提方法的准确性、有效性及先进性。

**关键词:** 质量监测; 故障诊断; 带钢热连轧; 核熵成分分析; 典型相关分析; Parzen 窗

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.0548

开放科学(资源服务)标识码(OSID):



**引用格式:** 彭开香,张丽敏.基于核典型相关性-熵成分分析的工业过程质量监测方法[J].控制与决策,2021,36(12):2999-3006.

## A quality monitoring method for industrial process based on kernel canonical correlation-entropy component analysis

PENG Kai-xiang<sup>†</sup>, ZHANG Li-min

(1. School of Automation, University of Science and Technology Beijing, Beijing 100083, China; 2. Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, Beijing 100083, China)

**Abstract:** The characteristics of industrial process such as multivariate, high-dimensional data and nonlinearity complicate the quality monitoring and quality-related fault diagnosis. In this paper, we present a novel quality-related fault detection method for industrial process by combining the kernel entropy composition analysis(KECA) and the canonical correlation analysis(CCA) algorithms for feature extraction, which reduces the number of input space dimension and ensures the maximum correlation between the extracted features and quality variables simultaneously. Firstly, the KECA algorithm is used to extract the features of the standardized data, the idea of the CCA algorithm is used to maximize the correlation between the extracted features and quality variables. Then, the monitoring statistics are constructed for process failure detection and the control limits are estimated via invoking a Parzen window density estimator. The proposed method is applied to the actual data of hot strip mill process (HSMP). Comparing with the performance of other four classical algorithms which are also suitable for nonlinear data, the results verify the accuracy, efficiency and advance of the method proposed.

**Keywords:** quality monitoring; fault diagnosis; HSMP; KECA; CCA; Parzen window

## 0 引言

现代工业日益朝着规模化、复杂化、集成化方向发展,以适应市场对多品种、多规格、高质量产品的需求。但随着生产规模的扩大及复杂性增加,过程发生故障的可能性及危害程度也随之上升。因此,为保证复杂工业过程的安全性、产品质量的稳定性、生产效率的高效性,对生产过程进行有效地在线监测、准确

地故障诊断是当前过程监控与故障诊断领域的重中之重。采用合理的故障诊断方法保障复杂工业过程的安全稳定运行及持续稳定的产品质量已经逐渐成为过程控制领域的重要研究方向<sup>[1]</sup>。

就复杂工业过程而言,想要对系统建立准确的机理模型变得极为不易,原始的基于机理模型的故障诊断技术不能很好地适应复杂工业过程的故障

收稿日期: 2020-05-10; 修回日期: 2020-10-26.

基金项目: 国家自然科学基金项目(61873024).

责任编委: 方华京.

<sup>†</sup>通讯作者. E-mail: kaixiang@ustb.edu.cn.

诊断<sup>[2]</sup>. 而利用过程历史数据的数据驱动方法逐渐得到工程界研究人员的关注, 并且随着传感器技术、存储技术的迅猛发展, 海量的过程数据得到了保存, 为数据驱动的故障诊断方法提供了坚实的基础, 使其逐渐成为故障诊断技术的中流砥柱. 其中于20世纪90年代提出的多元统计过程监控(multivariate statistical process monitoring, MSPM)方法发展最为成熟, 应用最为广泛. 基于多元统计的故障诊断方法强调变量间的潜结构<sup>[3]</sup>, 有别于传统的输入输出模型, MSPM方法利用过程正常运行数据建立潜结构模型, 在此基础上构建故障检测指标及其控制限进行故障检测和诊断. 常见的MSPM方法包括主元分析(principal component analysis, PCA)、偏最小二乘(partial least squares, PLS)、典型相关分析(canonical correlation analysis, CCA)等. 其中以PCA算法应用最为普遍, 它以保留原始空间数据的最大方差为准则, 有效地完成由高维空间到低维空间的投影. 但是, PCA作为一种线性降维方式无法对非线性数据进行有效处理. 文献[4]提出通过非线性映射将原始输入空间的非线性不可分问题转化为高维特征空间的线性可分问题, 在高维特征空间再进行主元分析, 成功将PCA的应用领域拓展到非线性数据, 也就是之后被广泛使用的KPCA算法; 文献[5]在核主元分析(KPCA)的理论基础上引入了信息熵, 提出了核熵成分分析(kernel entropy component analysis, KECA), 成功地将信息论的思想融入到特征提取及维数约简的过程中. 与KPCA算法依据最大方差选取主元不同, KECA算法进行特征提取时着重于特征值及对应的特征向量对原始特征数据集的Renyi熵的最大贡献率, 从而实现高维空间到低维空间的转换. 之后国内外学者对KECA及其改进算法的研究表明, 其在遥感数据的聚类、人脸图像识别、音频情感识别、医学影像数据识别等方面都具有独特的优势<sup>[6-9]</sup>. 在故障诊断方面, 有研究人员将其用于设备的故障诊断, 例如: 文献[10]将KECA与支持向量机(SVM)算法结合用于电力系统设备老化程度的在线监测; 文献[11]将多尺度主元分析提取的故障特征作为KECA分类器的输入, 最终实现冷水机组的故障实时监测和自动诊断. 针对过程的故障诊断主要集中在对数据的前期处理方面, 即采用合适的方法先对数据进行处理, 再运用KECA算法, 例如: 文献[12]提出将批次、变量、时间3个维度的数据阵沿时间片展开转化为二维数据阵, 然后在独立成分分析算法前使用KECA实现对数据的白化处理, 以此来降低数据的复杂度, 提高故障诊断的检测效果; 文献[13]将小波变换做预处理实现

数据去噪, 再结合KECA算法并将其用于工业过程的故障检测, 增强了原本算法的抗干扰能力.

总体上, KECA用于工业过程的故障诊断的研究相对较少, 而且同KPCA等算法类似, KECA并没有考虑过程变量主元对于质量变量变化的解释作用, 然而比起单纯地监测过程变量的波动和异常, 企业管理人员和工程师们可能更加关心的是过程变量引起的故障是否导致最终产品质量的波动<sup>[14]</sup>. 然而, 在实际生产过程中, 过程变量和产品质量总是无法同时在线测量, 因此建立过程变量与质量变量数据间的相关关系, 然后使用在线可用的过程度量监视产品质量并检测与质量相关的故障变得尤为重要<sup>[15]</sup>. 基于此, 本文提出核典型相关性-熵成分分析(kernel canonical correlation-entropy component analysis, KCC-ECA)算法提取数据特征, 在尽可能多地保留原始输入空间数据的信息熵的同时兼顾过程变量与质量指标间的相关性, 构建质量相关的故障检测模型, 并利用带钢热轧工业过程实际生产数据进行验证.

本文按照如下结构展开: 首先, 介绍涉及到算法的思想及理论基础; 其次, 重点对本文提出的算法进行介绍; 然后, 介绍新算法的实验仿真结果, 并与其他非线性工业过程质量监测算法进行对比分析; 最后, 根据仿真结果得出结论.

## 1 基于KECA的特征提取

### 1.1 KECA算法

KECA是一种基于信息熵的特征提取算法. 信息熵值可以直接展现信息的不确定性, 若作为量化指标表征工业过程运行状态中存在的确定, 则会对运行工况有很强的刻画能力. 本文主要使用的是Renyi熵的二阶形式, 对于一个给定的数据集 $D = [x_1, x_2, \dots, x_N]$ , 其中 $N$ 为样本数量, 设 $p(x)$ 是 $D$ 的概率密度函数, 则 $D$ 的二阶Renyi熵具体形式为

$$H(p) = -\lg\left(\int p^2(x)dx\right). \quad (1)$$

由于对数函数具有单调性, 只需要关注式(1)中的 $V(p) = \int p^2(x)dx$ , 通过估计 $V(p)$ 的值得到 $H(p)$ . 引入Parzen窗得到密度函数的估计值

$$\hat{p}(x) = \frac{1}{N} \sum_{x_i \in D} k_\sigma(x, x_i), \quad (2)$$

其中 $k_\sigma(x, x_i)$ 即为Parzen窗, 它是以 $x_i$ 为中心、 $\sigma$ 为宽度参数的核函数. 选取一个Mercer核, 例如高斯核函数, 则可得到如下形式的 $V(p)$ 的估计值:

$$\begin{aligned} \hat{V}(p) &= \int \hat{p}^2(x)dx = \\ &= \frac{1}{N^2} \sum_{x_i \in D} \sum_{x_j \in D} \int k_\sigma(x, x_i)k_\sigma(x, x_j)dx = \end{aligned}$$

$$\frac{1}{N^2} \sum_{x_i \in D} \sum_{x_j \in D} k_{\sqrt{2\sigma}}(x_i, x_j) = \frac{1}{N^2} \mathbf{1} K \mathbf{1}^T. \quad (3)$$

其中:  $k_{\sqrt{2\sigma}}(x_i, x_j)$  为核矩阵  $K \in R^{N \times N}$  中的元素,  $\mathbf{1}$  为  $1 \times N$  维的元素全为1的行向量. 因此, 基于样本集得到的 Renyi 熵估计值完全保留在相应核矩阵元素中. 对  $K$  进行特征值分解  $K = EDE^T$ . 其中:  $D = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_i]$  为特征值构成的对角矩阵,  $E = [e_1, e_2, \dots, e_i]$  是以特征向量为列向量的矩阵. 则 Renyi 熵估计值可由核矩阵特征值和特征向量的形式表示为

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

式(4)中, 每一个  $(\sqrt{\lambda_i} e_i^T \mathbf{1}^T)^2$  都对 Renyi 熵的估计值有相应的贡献.

KECA 在寻找最优投影方向时的原则是利用最少的特征保证输入样本集的二阶 Renyi 熵的损失量最小. 存在一个非线性变换  $\phi: R^d \rightarrow F$  实现输入空间到特征空间的映射  $x_i \rightarrow \phi(x_i)$ , 映射后的数据为  $\Phi = [\phi(x_1), \phi(x_2), \dots, \phi(x_N)]$ , 其在希尔伯特空间  $F$  的内积可以通过 Mercer 核函数求得:  $k(x_i, x_j) \langle \phi(x_i), \phi(x_j) \rangle$ , 则  $K = \Phi^T \Phi = EDE^T$  是  $F$  空间的 Gram 矩阵,  $\Phi$  在空间内某个主轴  $w_i$  上的投影可以通过  $w_i^T \Phi = \sqrt{\lambda_i} e_i^T$  计算得到. 因此, 式(4)中 Renyi 熵值的估计是基于在  $F$  空间中所有主轴的投影实现的, 即  $W^T \Phi = D^{0.5} E^T$ , 其中  $W = [w_1, w_2, \dots, w_N]$  为投影矩阵. 综上, KECA 特征提取降维过程可以转化为优化问题求解, 即

$$\min_{\lambda_i, e_i | i \in [1, N]} \hat{V}(p) - \hat{V}_p(p). \quad (5)$$

其中:  $\hat{V}_p(p) = \frac{1}{N^2} \mathbf{1} E_p D_p E_p^T \mathbf{1}^T = \frac{1}{N^2} \mathbf{1} K_{\text{ECA}} \mathbf{1}^T$  是提取的  $p$  个特征所携带的那部分 Renyi 熵, 问题转化为求取

$$\min_{\lambda_i, e_i | i \in [1, N]} \frac{1}{N^2} \mathbf{1} (K - K_{\text{ECA}}) \mathbf{1}^T = \frac{1}{N^2} \sum_{j=p+1}^N \psi_j, \quad (6)$$

其中  $\psi_j$  对应式(4)中第  $j$  大的项. 因此, KECA 算法通过将  $\Phi$  投影到由对 Renyi 熵估计值贡献量最大的前  $p$  个特征空间主轴张成的特征子空间  $W_p$ , 得到提取到的 KECA 特征

$$\Phi_{\text{ECA}} = W_p^T \Phi = D_p^{0.5} E_p^T. \quad (7)$$

## 1.2 CS 散度统计量

CS 散度统计量<sup>[16]</sup>由 Cauchy-Schwarz 不等式的积分形式  $\left[ \int f(x)g(x)dx \right]^2 \leq \int f^2(x)dx \int g^2(x)dx$  演化得到

$$D_{\text{CS}}(p_i, p_j) = -\lg \left[ \frac{\int p_i(x)p_j(x)dx}{\sqrt{\int p_i^2(x)dx \int p_j^2(x)dx}} \right], \quad (8)$$

$D_{\text{CS}}(p_i, p_j)$  可衡量两种概率密度函数  $p_i(x)$  与  $p_j(x)$  的相似度, 由不等式的取等条件可知, 二者相等时,  $D_{\text{CS}}(p_i, p_j)$  取最小值为0. 同样用 Parzen 窗对  $p(x)$  进行估计, 结合  $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$  及式(2)可得

$$D_{\text{CS}}(p_i, p_j) = -\lg \cos \angle(\gamma_i, \gamma_j). \quad (9)$$

其中:  $\gamma_i = \frac{1}{N_i} \sum_{x_n \in D_i} \phi(x_n)$ ,  $\gamma_j = \frac{1}{N_j} \sum_{x_t \in D_j} \phi(x_t)$ . 令  $V_{\text{CS}} = \cos \angle(\gamma_i, \gamma_j)$ , 则 CS 散度转化为核特征空间均值向量的角度余弦值, 其值落在区间  $[0, 1]$  内. 实验验证 KECA 变换后的数据具有一定的角度结构, 即不同类型的数据聚集在不同的方向, 故选取  $D_{\text{CS}}$  作为统计量更有助于识别故障数据, 其控制限可通过核密度估计的方法求得.

## 2 KCC-ECA 的质量监测模型与故障检测

### 2.1 CCA 算法基本思想

典型相关分析(CCA)是研究两组变量相关性的一种统计方法. 为研究两组变量  $U^T = [u_1, u_2, \dots, u_p]$  和  $V^T = [v_1, v_2, \dots, v_q]$  的相关性, 需找到投影方向  $A^T = [a_1, a_2, \dots, a_p]$  和  $B^T = [b_1, b_2, \dots, b_q]$ , 通过研究投影  $A^T U$  与  $B^T V$  的相关关系确定两组变量间的相关关系.

Pearson 相关系数  $\rho$  是用来度量变量的相关性的指标, 取值范围为  $[-1, 1]$ , 负值和正值分别代表负相关和正相关, 值的大小代表相关性的强弱. 值越大, 相关性越强, 若值为0, 则代表二者无线性相关关系, 具体形式如下:

$$\rho(U, V) = \frac{E[(U - \mu_U)(V - \mu_V)]}{\sigma_U \sigma_V}, \quad (10)$$

其中  $\mu_U$ 、 $\mu_V$ 、 $\sigma_U$ 、 $\sigma_V$  分别为  $U$  和  $V$  的均值及标准差. CCA 方法的核心思想就是寻找最优的投影方向  $A$  和  $B$  使得变量集在此方向的投影的相关性达到最强, 其准则函数为

$$\rho_{uv}(A, B) = \frac{A^T C_{uv} B}{\sqrt{A^T C_{uu} A} \sqrt{B^T C_{vv} B}}; \quad \text{s.t. } A^T C_{uu} A = B^T C_{vv} B = 1. \quad (11)$$

其中:  $C_{uu} = E[UU^T]$  和  $C_{vv} = E[VV^T]$  分别为  $U$  和  $V$  的自协方差矩阵,  $C_{uv} = E[UV^T]$  为二者的互协方差矩阵, 且有  $C_{vu} = E[VU^T] = E[UV^T] = C_{uv}^T$ .

CCA 关注的是变量集的整体相关程度, 丢弃了两个变量组内各个维度上的信息, 而 KECA 算法只侧重保留变量组内的信息, 因此, 希望提取的特征能融合尽可能多的原始数据的信息, 同时兼顾自变量主

元与因变量间的相关性,本文提出基于KCC-ECA的方法。

## 2.2 基于KCC-ECA的质量监测模型

### 2.2.1 KCC-ECA算法

设  $X = [x_1, x_2, \dots, x_N] \in R^{L \times N}$  和  $Y = [y_1, y_2, \dots, y_N] \in R^{M \times N}$  分别代表过程变量集和质量变量集,  $x_i$  和  $y_i$  分别为  $L$  维和  $M$  维的列向量, 其中  $i \in (1, 2, \dots, N)$ . 对原始数据做中心化处理以及无量纲化处理得到  $X_0$  和  $Y_0$ , 实现简化数据模型结构, 消除不同测量量纲对模型的影响。

采用高斯核函数进行密度估计, 即

$$k_\sigma(q_i, q_j) = \exp\left(-\frac{\|q_i - q_j\|^2}{2\sigma^2}\right). \quad (12)$$

分别得到过程变量集及质量变量集的核矩阵  $K_X = \Phi_X^T \Phi_X$  和  $K_Y = \Phi_Y^T \Phi_Y$ , 其中  $\Phi_X$  和  $\Phi_Y$  分别为  $X_0$  和  $Y_0$  经过非线性映射后的数据. 分别对核矩阵进行特征值分解  $K_X = E_X D_X E_X^T$ ,  $K_Y = \Delta \Gamma \Delta^T$ ,  $\lambda_i$ 、 $e_i$  和  $d_i$ 、 $t_i$  分别为  $K_X$  和  $K_Y$  的特征值和特征向量, 而每一个特征值和特征向量对于 Renyi 熵的估计值都有一定的贡献. 根据式(4), 以  $K_X$  为例, 构建  $S_X = (\sqrt{\lambda_i} e_i^T \mathbf{1}^T)^2$  作为所提取的特征对 Renyi 熵估计值贡献度的指标。

记  $u_i$  为过程变量集中的一个特征, 表示  $\Phi_X$  在高维特征空间内与之相应的主轴  $\alpha_i$  上的投影  $u_i = \Phi_X^T \alpha_i$ ,  $u_i$  可用如下形式计算得到:  $u_i = \Phi_X^T \alpha_i = \sqrt{\lambda_i} e_i$ . 同理记  $v_i$  为质量变量集中的一个特征,  $\beta_i$  是高维特征空间内与之相应的主轴, 则  $v_i = \Phi_Y^T \beta_i = \sqrt{d_i} t_i$ . 为使  $u_i$ 、 $v_i$  更好地保留原数据集中的信息, 根据熵成分分析原理, 应有  $S_X \rightarrow \max$ ,  $S_Y \rightarrow \max$ . 另一方面, 希望过程变量对质量变量有最大的解释能力, 结合CCA的思想,  $u_i$ 、 $v_i$  的相关度应达到最大, 综上, KCC-ECA算法的思想是在保留原数据信息熵的同时考虑过程变量与质量变量间的相关性. 则KCC-ECA算法的准则函数定义为

$$J = \max_{i,j \in [1,N]} S_X S_Y r(u_i, v_j)^2. \quad (13)$$

其中:  $S_X$  和  $S_Y$  分别为所提取过程变量特征和质量指标特征对 Renyi 的熵值贡献度,  $r(u_i, v_j)$  用来衡量所提取过程变量特征与质量指标特征间的相关程度。

$r(u_i, v_j)$  参照式(10), 结合KECA算法未经中心化的特点, 可以定义为 Pearson 相关系数的未中心化的形式, 即

$$r(u_i, v_j) = \frac{\text{cov}(u_i, v_j)}{\sqrt{\sum_{u_{im} \in u_i} u_{im}^2} \sqrt{\sum_{v_{jm} \in v_j} v_{jm}^2}} = \frac{\langle u_i, v_j \rangle}{\|u_i\| \|v_j\|}. \quad (14)$$

可以看出其实质为向量间的余弦值, 即用余弦相似度可以衡量二者的相关度, 代入  $u_i$ 、 $v_j$  可推导出如下形式:

$$r(u_i, v_j) = \frac{\langle \sqrt{\lambda_i} e_i, \sqrt{d_j} t_j \rangle}{\sqrt{\lambda_i} e_i^T e_i \sqrt{d_j} t_j^T t_j} = e_i^T t_j, \quad (15)$$

其中  $\|e_i\| = \sqrt{e_i^T e_i} = \|t_j\| = \sqrt{t_j^T t_j} = 1$ . 则新算法的准则函数可以改写为如下形式:

$$\max_{i,j \in [1,N]} [(\sqrt{\lambda_i} e_i^T \mathbf{1}^T) (\sqrt{d_j} t_j^T \mathbf{1}^T) (e_i^T t_j)]^2. \quad (16)$$

此外, 根据式(4), 每一个特征值和特征向量对于 Renyi 熵估计值都有一定的贡献, 按照下式定义  $m$  个特征对 Renyi 熵估计值的累计贡献率

$$\text{CR}_m = \sum_{j=1}^m \psi_j / \sum_{j=1}^N \psi_j \times 100\%. \quad (17)$$

其中:  $\psi_j$  为式(6)中的项, 提取保留原始数据集95%以上 Renyi 熵的特征. 最终提取到的过程变量集的  $m$  个特征为  $\Omega_X = D_m^{1/2} E_m^T$ .

### 2.2.2 统计量的构建及控制限确定

本文提出的KCC-ECA方法在将矩阵核中心化后, 会导致 Renyi 熵无限大, 影响监测模型的构建, 但非中心化的核矩阵又会影响传统的  $T^2$  以及 SPE 统计量的使用, 故本文构建如下形式的 CS 统计量用于过程监测:

$$D_{\text{CS}} = -\lg \cos \angle(\gamma_i, \gamma). \quad (18)$$

对数函数为单调函数, 可通过对  $V_{\text{CS}} = \cos \angle(\gamma_i, \gamma)$  的检测反映 CS 统计量的变化,  $V_{\text{CS}}$  的值落在区间  $[0, 1]$  内. 最大的相似度值1意味当前的过程运行条件属于正常工况, 但是由于过程噪声的存在, 使得相似度不可能为1, 而是接近于1的数. 对  $V_{\text{CS}}$  进行核密度估计, 选取高斯核函数, 得到如下形式的概率密度函数估计:

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^N \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-x_i)^2}{2h^2}\right). \quad (19)$$

进而可以确定核密度显著程度为  $\alpha$  的控制限  $X_\alpha$  为  $P(X \leq X_\alpha) = \alpha$ , 可通过样本的分布情况计算得到. 从而可以确定检测统计量  $D_{\text{CS}}$  的控制限, 当统计量值超过控制限时, 认为有故障发生. 需要注意进行核密度估计时, 参数  $h$  的选取对于核密度估计的准确度至关重要, 影响控制限的确定, 故检测的效果依赖于控制限的确定. 本文选取了  $l_1$  范数约束下的交叉检验的方法确定  $h$ , 具体实现方式参见文献[17].

KCCA算法关注的是两组变量在某个方向上的投影的相关程度, 注重组间的相关性考虑, 忽略了组内各变量间的关联, 而KECA算法虽保留了变量组内

的信息, 却缺乏自变量主元对于因变量变化的解释作用的考量. 本文提出的KCC-ECA算法在进行特征提取时既完成了组内变量信息的融合, 又兼顾了过程变量与质量变量的相关性, 因此, 它在保留原始数据信息熵的同时, 实现了过程变量在这些方向上的投影与质量指标的相关性相互最大化, 从而提高了识别精度. 基于KCC-ECA的质量相关的故障检测流程框架如图1所示.

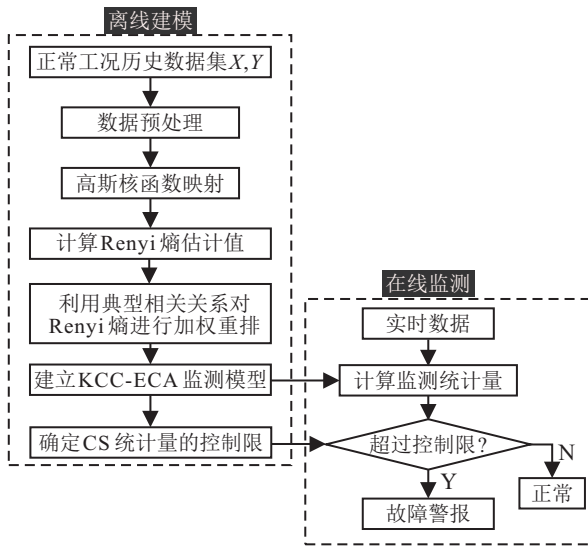


图1 基于KCC-ECA质量相关故障检测流程

### 3 带钢热连轧过程仿真分析

带钢热轧过程是钢铁生产流程的关键工序, 是一条以高速、高效、高质著称的全自动化生产作业线, 主要包含加热炉、粗轧机组、传送带和飞剪、精轧机组、层流冷却装置、卷取机等装置<sup>[18]</sup>. 此外, 还需在生

产线各个关键点布置不同种类的检测仪器, 例如测厚仪、测温仪、测宽仪和压头等, 详细可参考图2, 从而实时观察生产状态, 检测板带质量, 提高生产效率.

各类仪表及传感器记录并存储了大量的现场数据, 为基于数据驱动的故障检测提供了大量资源. 在热连轧的生产过程中, 精轧过程是质量控制至关重要的一环, 精轧机组是带钢热连轧的核心设备, 基于此, 本文以精轧机组为主要研究对象, 用以验证所提出方法的实用性和有效性.

精轧机组一般由6或7个机架串联组成, 每台均为四辊轧机, 主要由一对工作辊、一对支撑辊以及相应的液压压下装置等部分构成. 在轧机的下支撑辊下部一般装有轧制力检测传感器——压头, 用于测量带钢轧制力. 每台轧机的驱动一般由一台大型的交流电机完成. 上、下工作辊之间的辊缝主要由高精度的液压伺服控制系统完成, 以保证一定厚度的带钢经该机架轧制后得到相应的出口厚度. 在整个带钢热连轧过程, 本文关心的质量变量为厚度、宽度及出口温度等, 尤其是出口厚度, 是直接影响产品质量的关键因素. 但是, 受到传感器安装位置的限制, 轧制系统存在明显的时滞现象, 对质量控制产生消极影响. 过程中一旦发生故障, 仅靠延迟滞后的反馈控制策略很难保证产品质量<sup>[19]</sup>, 故本文重点研究出口厚度这一关键因素, 利用某钢铁公司1700 mm带钢热连轧生产线现场采集的大量反映实际生产过程的数据, 考虑生产过程的“厚度质量变量”及影响质量的“关键过程变量”, 实现过程厚度质量监测. 相关变量的详细情况如表1所示.

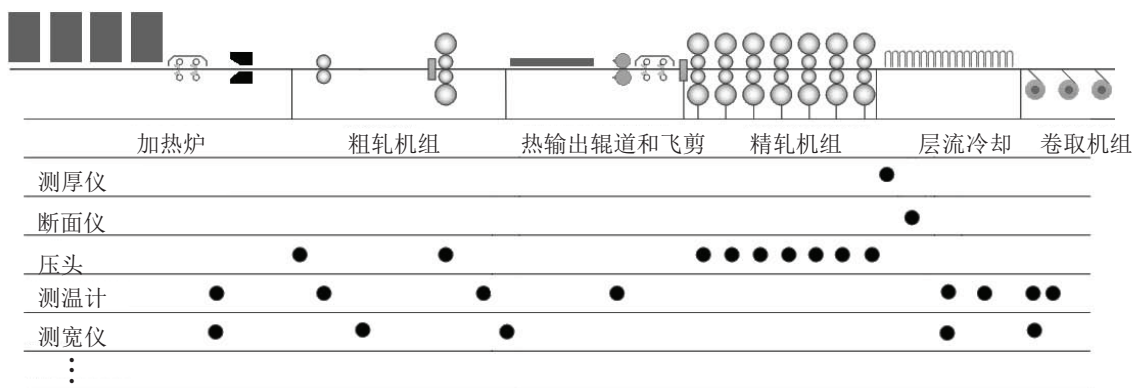


图2 带钢热连轧生产流程及主要检测仪器分布

表1 精轧过程各变量分配表

变量号	类型	物理意义	单位
1 ~ 7	过程变量	第 <i>i</i> 机架辊缝值, $i = 1, 2, \dots, 7$	mm
8 ~ 14	过程变量	第 <i>i</i> 机架轧制力, $i = 8, 9, \dots, 14$	MN
15 ~ 20	过程变量	第 <i>i</i> 机架弯辊力, $i = 15, 16, \dots, 20$	MN
21	质量变量	精轧末机架出口厚度	mm

为了测试本文提出的KCC-ECA联合驱动的质量监测算法, 利用Matlab对带钢热连轧过程的历史数据建立质量监测模型, 数据为30 s内采集, 采样周期为10 ms. 再分别利用包含3000个采样点的不同类型故障数据进行故障检测, 实现在线监控的目的. 表2列举了本文选取的案例中的3种典型故障.

表2 精轧机组典型故障

故障	故障描述	持续时长/s
故障1	第5机架的弯辊力测量传感器故障	10
故障2	第4机架液压压下辊缝控制系统失效	10
故障3	第2、3机架间冷却水控制阀的执行器故障	15

选择同样用于处理非线性数据核典型相关分析(KCCA)算法<sup>[20]</sup>、传统KECA算法、传统核偏最小二乘(KPLS)算法以及改进的KPLS(MKPLS)算法<sup>[21]</sup>与本文提出的算法进行对比,用来验证KCC-ECA算法的有效性和优越性。

首先,考虑的故障1为第5个机架测量弯辊力的传感器出现故障,测量值发生阶跃跳变.由于系统自动厚度控制(automatic gauge control, AGC)对后面两

个机架的反馈调节,后面机架的弯辊力的值也会发生相应变化,但是这种故障只会引起带钢板形的变化,对钢板出口厚度的影响甚小,因此这种故障是质量不相关的. KCCA算法以及KPLS算法选用SPE统计量检测质量无关的故障, MKPLS算法选用 $T_u^2$ 统计量,而KECA和KCC-ECA算法均选用CS统计量检测故障,故障发生在开始采样第10s,持续10s.图3分别展示了几种算法对于故障1的检测结果。

通过对比发现几种算法都能较好地检测出故障,但是在故障刚结束后都存在一定程度的误报警,其主要与热连轧过程是连续的轧制过程且反馈控制存在一定程度的时滞有关。

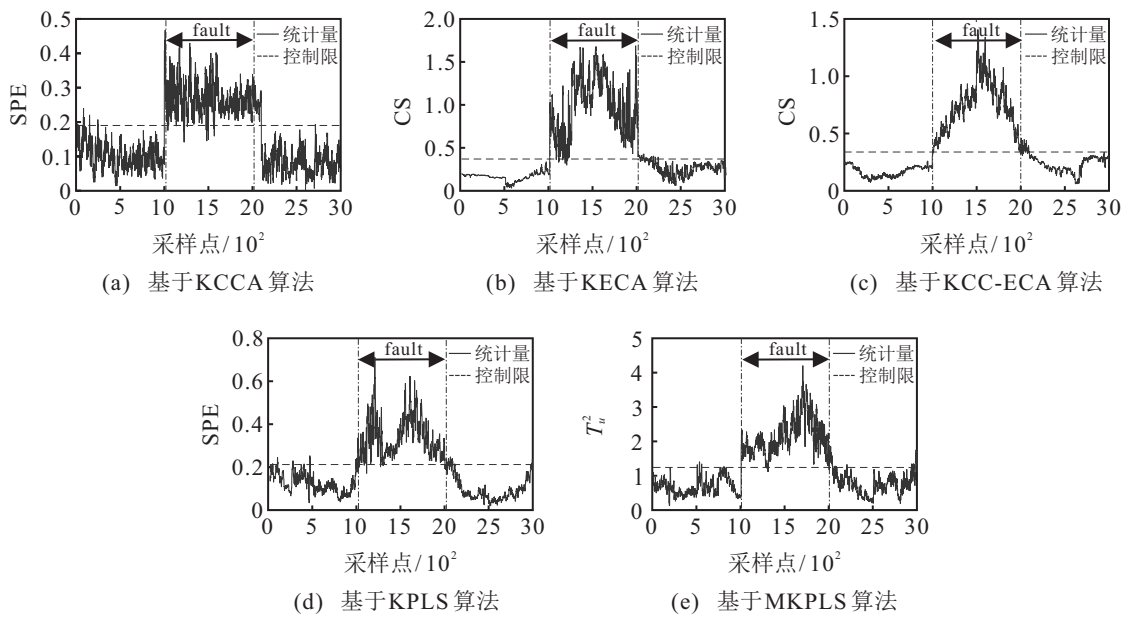


图3 带钢热连轧过程故障1的各算法的检测结果

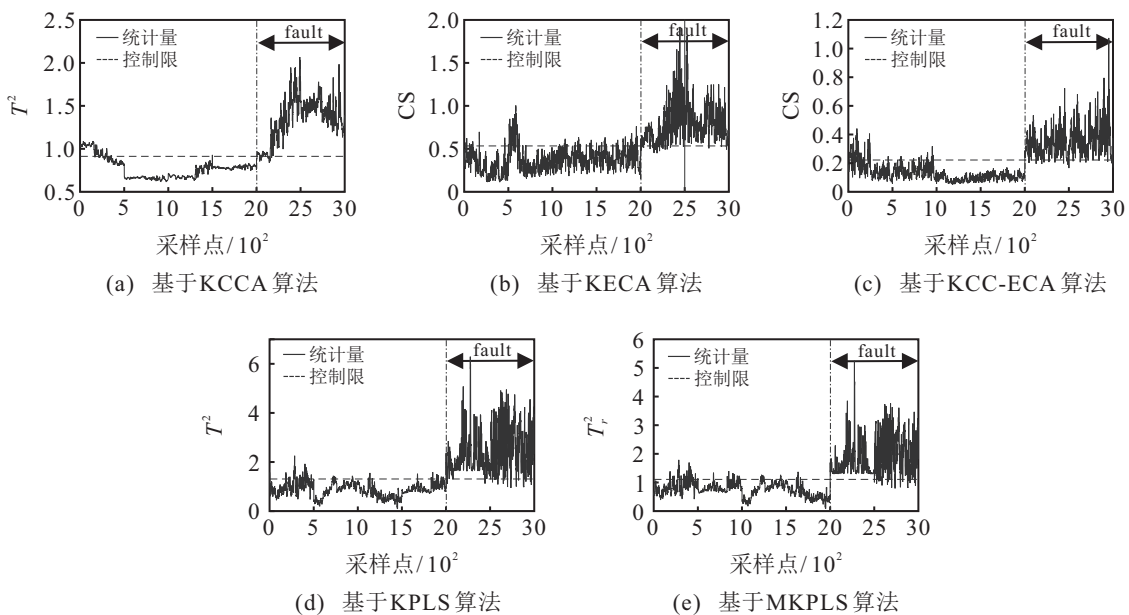


图4 带钢热连轧过程故障2的各算法的检测结果

其次,考虑的故障2是第4机架液压压下辊缝控制系统失效,这种故障发生时,直接影响第4机架辊缝采样值,第4机架的轧制力采样值也受到影响,此故障还会波及后面机架的辊缝、轧制力的采样值,进而影响最后的出口厚度,因此该故障为质量相关的故障.故障在采样开始后的第20s时发生,一直持续到采样结束,KCCA以及KPLS算法选用 $T^2$ 统计量,MKPLS选用 $T_r^2$ 统计量检测质量相关故障,KECA算法及KCC-ECA算法均采用CS统计量进行检测.对比图4展示的故障检测结果即可发现,5种算法中KPLS误报率最低,但是其检测率并非最佳,总体而言在采样开始阶段及故障发生阶段较容易产生误报、漏报,其中KCCA在故障发生的初始阶段存在较多的

漏报,说明其对于故障2的检测存在延迟,相比较而言,KCC-ECA算法建立的模型能及时检测出故障并且漏报、误报较少.

最后,考虑的故障3为精轧机轧制过程中时常发生的第2机架与第3机架之间冷却水控制阀的执行器故障,其不能按照设定模式关闭,对第3及之后机架的轧制力造成影响,并且在AGC的调节作用下,各机架辊缝也相应地发生变化,从而导致出口带钢厚度产生正向偏差,影响最终的产品质量,同样为质量相关的故障,在采样开始后的第10s时发生故障,第25s时故障结束.同样用几种算法分别建立的模型对该采样过程进行监测,其中,各算法相应统计量的选取与故障2相同,几种算法模型的检测结果如图5所示.

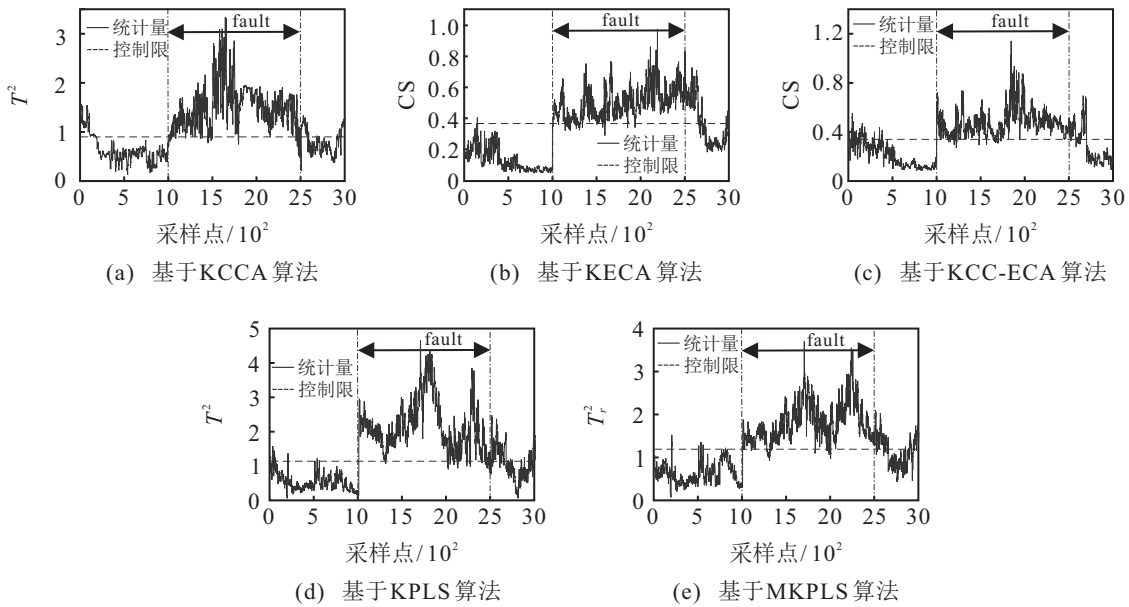


图5 带钢热连轧过程故障3的各种算法的检测结果

对比发现,现有的几种检测方法也可以实现故障3的检测,但KCCA算法在采样起始阶段存在较大的误报,而且故障发生阶段的漏报也较多,而KECA、KPLS以及MKPLS在故障结束后以及采样结束前均存在较多的误报,综合对比后,应用KCC-ECA算法建模对故障3的检测效果更优.

表3 故障检测率和故障误报率对比结果

算法种类	故障1		故障2		故障3	
	检测率	误报率	检测率	误报率	检测率	误报率
KCCA	0.980	0.062	0.937	0.135	0.941	0.215
KECA	0.985	0.049	0.943	0.110	0.947	<b>0.143</b>
KCC-ECA	0.987	<b>0.039</b>	<b>0.959</b>	0.072	<b>0.965</b>	0.196
KPLS	0.985	0.042	0.952	<b>0.068</b>	0.956	0.167
MKPLS	<b>0.989</b>	0.049	0.955	0.091	0.961	0.161

从表3中各算法检测结果的对比情况来看,针对带钢热连轧生产过程发生的质量相关故障,KCC-ECA算法建立的模型具有较高的故障检测率和较低的误报率,而对于生产过程中的与质量无关的故障,各种算法建立的模型的差别较小,主要与本文选择的分析案例中的质量变量较少有关.除检测效果的对比外,还记录了几种算法离线建模及实时检测的耗时,对几种算法的计算复杂度进行量化对比,表4展示了量化对比结果.

表4 各算法计算复杂度量化对比结果

算法种类	离线建模时间/s	在线检测时间/s
KCCA	2.123	1.202
KECA	<b>1.769</b>	0.897
KCC-ECA	2.013	1.012
KPLS	2.324	1.034
MKPLS	1.996	<b>0.912</b>

由表4可以看出,各算法的计算复杂度相差不大,均可较快完成故障检测.综合仿真结果的对比情况可知,采用KCC-ECA算法建模能得到更优的故障检测效果.

## 4 结论

本文提出了一种基于KCC-ECA的工业过程质量检测方法,该方法将原始输入空间映射到高维核空间,以质量变量与过程变量间的典型相关性加权后的Renyi熵为指标进行特征提取,同时设计与之相适应的统计量及计算控制限,构建质量相关的工业过程故障检测模型.该算法在实现过程变量组内信息融合的同时兼顾其与质量变量组间的相关性,通过对带钢热轧实际生产过程数据的应用研究,表明本算法能及时准确地检测出故障,与其他两种算法的对比结果验证了本文算法的有效性.未来研究工作的主要内容是在KCC-ECA模型的监测架构下实现故障的诊断与识别.

## 参考文献(References)

- [1] Peng K X, Ma L, Zhang K. Review of quality-related fault detection and diagnosis techniques for complex industrial processes[J]. *Acta Automatica Sinica*, 2017, 43(3): 349-365.
- [2] Liu Q, Chai T Y, Qin S Z, et al. Progress of data-driven and knowledge-driven process monitoring and fault diagnosis for industry process[J]. *Control and Decision*, 2010, 25(6): 801-807.
- [3] Zhang K, Shardt Y A W, Chen Z W, et al. A KPI-based process monitoring and fault detection framework for large-scale processes[J]. *ISA Transactions*, 2017, 68: 276-286.
- [4] Schölkopf B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. *Neural Computation*, 1998, 10(5): 1299-1319.
- [5] Jenssen R. Kernel entropy component analysis[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(5): 847-860.
- [6] Gómez-Chova L, Jenssen R, Camps-Valls G. Kernel entropy component analysis for remote sensing image clustering[J]. *IEEE Geoscience and Remote Sensing Letters*, 2012, 9(2): 312-316.
- [7] Ruan X L, Wang S F. Face recognition based on weighted multi-resolution kernel entropy component analysis[C]. *The 29th Chinese Control and Decision Conference. Chongqing: IEEE*, 2017: 6118-6123.
- [8] Shi J, Jiang Q K, Zhang Q, et al. Sparse kernel entropy component analysis for dimensionality reduction of biomedical data[J]. *Neurocomputing*, 2015, 168(30): 930-940.
- [9] Gao L, Guan L. Information fusion via optimized KECA with application to audio emotion recognition[C]. *2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery. Zhengzhou: IEEE*, 2018: 255-264.
- [10] Yang D K, Chen W G, Zhou W R, et al. Recognition of aging stage of oil-paper insulation based on surface enhanced Raman scattering and kernel entropy component analysis[J]. *IEEE Access*, 2019, 7: 127862-127873.
- [11] 齐咏生, 张海利, 王林, 等. 基于MSPCA-KECA的冷水机组故障监测及诊断[J]. *化工学报*, 2017, 68(4): 1499-1508.  
(Qi Y S, Zhang H L, Wang L, et al. Fault detection and diagnosis for chillers using MSPCA-KECA[J]. *CIESC Journal*, 2017, 68(4): 1499-1508.)
- [12] 常鹏, 王普, 高学金. 基于高阶累计统计量的微生物发酵过程监测[J]. *控制与决策*, 2017, 32(12): 2273-2278.  
(Chang P, Wang P, Gao X J. Microbial fermentation process monitoring based on higher order cumulant statistics[J]. *Control and Decision*, 2017, 32(12): 2273-2278.)
- [13] Yang Y H, Li X L, Liu X Z, et al. Wavelet kernel entropy component analysis with application to industrial process monitoring[J]. *Neurocomputing*, 2015, 147(5): 395-402.
- [14] Qin S J, Zheng Y Y. Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures[J]. *AIChE Journal*, 2013, 59(2): 496-504.
- [15] Ma L, Dong J, Peng K X. A novel hierarchical detection and isolation framework for quality-related multiple faults in large-scale processes[J]. *IEEE Transactions on Industrial Electronics*, 2020, 67(2): 1316-1327.
- [16] Jenssen R, Eltoft T. A new information theoretic analysis of sum-of-squared-error kernel clustering[J]. *Neurocomputing*, 2008, 72(1/2/3): 23-31.
- [17] Suhre A, Arikian O, Cetin A E. Bandwidth selection for kernel density estimation using Fourier domain constraints[J]. *IET Signal Processing*, 2016, 10(3): 280-283.
- [18] Ma L, Dong J, Peng K X. A novel robust semisupervised classification framework for quality-related coupling faults in manufacturing industries[J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(5): 2946-2955.
- [19] Zhang K, Dong J, Peng K X. A novel dynamic non-Gaussian approach for quality-related fault diagnosis with application to the hot strip mill process[J]. *Journal of the Franklin Institute*, 2017, 354(2): 702-721.
- [20] Liu Q, Zhu Q Q, Qin S J, et al. Dynamic concurrent kernel CCA for strip-thickness relevant fault diagnosis of continuous annealing processes[J]. *Journal of Process Control*, 2018, 67: 12-22.
- [21] Jiao J F, Zhao N, Wang G, et al. A nonlinear quality-related fault detection approach based on modified kernel partial least squares[J]. *ISA Transactions*, 2017, 66: 275-283.

## 作者简介

张开香(1971—),男,教授,博士生导师,从事复杂工业系统故障诊断与容错控制等研究, E-mail: kaixiang@ustb.edu.cn;

张丽敏(1995—),女,硕士生,从事工业过程监控与故障诊断、机器学习及大数据分析的研究, E-mail: zlm41401008@126.com.

(责任编辑: 闫妍)