基于自联想核回归的带离群值化工过程故障检测

沈飞凤, 杨慧中†

(江南大学 教育部轻工过程先进控制重点实验室, 江苏 无锡 214122)

摘要: 基于数据驱动的故障检测模型通常要求训练数据必须是正常操作条件下的测量值. 然而在实际工业生产过程中, 即使在正常工况下, 数据集中也难以避免存在离群值. 此时若仍采用传统的基于多元统计分析的方法, 其监测模型的控制限会受到严重影响, 造成故障漏报. 因此, 为了确保当训练数据包含离群值时, 监测模型仍然呈现较好的故障检测效果, 本文提出了一种基于自联想核回归的故障检测方法. 首先基于最小化β散度的鲁棒预白化算法对训练集进行白化计算, 消除变量之间相关性对样本相似度度量的影响. 然后通过自联想核回归算法重构正常工况下的验证数据, 根据重构误差建立模型监测指标. 为了消除离群值对故障样本重构的影响, 构造截断函数来避免离群样本参与相似故障数据的重构, 并对所有参与构建Q统计量的残差变量基于指数加权滑动平均方法自适应加权, 得到新的监测统计量. 将该方法运用于田纳西–伊斯曼过程并与其他方法进行比较, 验证了本文所提故障检测算法的有效性.

关键词: 离群值; 鲁棒白化; 自联想核回归; 指数加权滑动平均; 故障检测

引用格式: 沈飞凤, 杨慧中. 基于自联想核回归的带离群值化工过程故障检测. 控制理论与应用, 2023, 40(3): 583 – 592

DOI: 10.7641/CTA.2022.11013

Fault detection for chemical processes with outliers based on auto-associative kernel regression

SHEN Fei-feng, YANG Hui-zhong†

(Key Laboratory of Advanced Process Control for Light Industry of Ministry of Education, Jiangnan University, Wuxi Jiangsu 214122, China)

Abstract: The data driven fault detection models usually require that the training data must be measured under normal operating conditions. However, in the actual industrial processes, it is possible that the collected data set contains outliers even under normal working conditions. In this case, the control limits of the traditional method based on multivariate statistical analysis are often heavily influenced by the outliers, which results in a large number of missed failures. Therefore, in order to ensure that the monitoring model still has good performance even when the training data contains outliers, this paper proposed a fault detection method based on auto-associative kernel regression (AAKR). First, the training set is whitened on the basis of a robust whitening algorithm that minimizes β divergence to eliminate the influence of correlation between variables on sample similarity measurement. Then, AAKR reconstructs the validation data under normal working conditions to obtain the residuals and establish the correct detection index. In order to avoid the influence of outliers on the reconstructions of faulty test data, a truncation function is constructed to avoid the involvement of outliers similar to the faulty samples in reconstruction. All residual variables involved in Q statistic construction were weighted based on the exponentially weighted moving average (EWMA) to obtain the new monitoring statistic. The proposed method is applied to the Tennessee Eastman (TE) process to verify the effectiveness of the proposed fault detection algorithm.

Key words: outliers; robust whitening; auto-associative kernel regression; exponentially weighted moving average; fault detection

Citation: SHEN Feifeng, YANG Huizhong. Fault detection for chemical processes with outliers based on auto-associative kernel regression. *Control Theory & Applications*, 2023, 40(3): 583 – 592

收稿日期: 2021−10−23; 录用日期: 2022−05−17.

†通信作者. E-mail: yhz@jiangnan.edu.cn.

本文责任编委: 王大轶.

国家自然科学基金项目(61773181), 中央高校基本科研业务费专项资金项目(JUSRP51733B)资助.

Supported by the National Natural Science Foundation of China (61773181) and the Fundamental Research Funds for the Central Universities (JUSRP51733B).

1 引言

流程工业的迅速发展使得现代工业系统的规模越来越大且越来越复杂. 为了确保过程安全且高效地平稳运行, 对过程进行实时监控是十分必要的. 目前运用最为广泛的是基于数据驱动的故障检测方法, 常见的有主元分析 (principal component analysis, PCA)、偏最小二乘(partial least square, PLS)、神经网络以及以这些为基础进行改进的方法[1–3]. 然而大部分算法在建立监测模型和确定故障检测指标时, 都假设训练数据是不包含离群值的正常工况采样值[4–6]. 但是实际的大规模工业系统会因为一些特殊情况例如传感器测量扰动、操作失误或者数据传输存储不稳定等而导致所采集的正常数据中存在离群值[7]. 这些离群样本往往会造成故障检测指标的计算产生严重偏差, 致使监测模型变得不可靠. 为了应对这一情况, 对数据进行预处理或研究鲁棒的数据驱动方法是十分必要的.

由于被广泛应用于故障检测的数据驱动方法–PCA对离群值非常敏感, 因此在PCA的基础上提出了主成分追踪(principal component pursuit, PCP)的鲁棒主元分析(robust PCA, RPCA)方法, 并将其应用于故障检测中[8–9]. 通过对PCP分解得到的低秩矩阵和稀疏矩阵分别建立T^2统计量和基于相关系数的统计量, 证明了基于PCP的故障检测效果要优于PCA[10]. Yvon等人[11]基于尺度M 估计器, 提出了基于RPCA模型的故障检测和隔离方法. 虽然使用RPCA可以建立稳健的监测模型, 降低训练数据在建立监测模型时受到离群值的影响, 但是模型的监测指标仍然会因为离群值的存在而产生偏差. 除了鲁棒多元统计分析的方法, 最传统的离群值处理方法是基于一定的判定法则, 将高于设定阈值的数据作为离群值进行剔除. K均值聚类(K-means clustering, KMC)算法常常被用来剔除多元数据集中的离群值[12]. Cai等人[13]使用最小协方差行列式 (minimum covariance determinant, MCD) 直接寻找不含离群值的鲁棒样本数据. Luo等人[14]在同时考虑稀疏性和鲁棒性的基础上, 提出了稀疏鲁棒主元分析 (sparse robust PCA, SRPCA), 借助MCD估计器来检测训练集中的离群值并提出了一种新的鲁棒T^2统计量. 然而这类处理方法容易存在离群值剔除过度或者不足的问题, 也会破坏样本间原有的结构信息[15]. Yu等人[16]通过结合降噪自编码器 (denoising auto-encoder, DAE)和弹性网络(elastic net, EN), 成功挖掘了带噪声的过程数据的非线性结构, 正确识别了带故障的变量.

本文提出一种基于自联想核回归(auto-associative kernel regression, AAKR)的故障检测方法, 针对包含离群值的训练样本集, 直接采用最小化β散度的鲁棒预白化方法, 通过对远离正常工况的离群值给定较小的权重以降低其对均值和协方差计算的影响. 对数据进行鲁棒预白化也可以消除变量之间的线性相关性, 降低其对样本相似度计算带来的偏差. 鉴于AAKR是按照样本之间的相似度大小对训练数据进行加权以重构当前数据的, 离群值的存在并不会影响正常样本的重构, 但是当故障样本与训练集中的离群值非常接近时, 则可能会因为较小的重构误差而无法正确将其识别, 所以需要尽可能选择正常样本对其进行重构. 本文基于训练样本的鲁棒均值和协方差, 利用每个训练样本的比例因子构造截断函数来挑选重构当前故障数据的正常训练样本. 在AAKR算法中, 所有残差变量在参与Q统计量计算时权重都是一致的, 在没有任何先验知识的情况下, 某些只影响部分变量的微小故障容易因此而被淹没, 造成系统漏报. 为了提高模型的监测性能, 需要对可能携带故障信息的残差进行放大, 提高其在统计量计算中的比重, 从而进一步提高模型对于故障的检出率. 本文提出了基于指数加权滑动平均 (exponentially weighted moving average, EWMA)的自适应加权Q统计量计算方法. 通过在田纳西–伊斯曼(tennessee eastman, TE)过程上的仿真试验, 验证了本文所提自适应自联想核回归(adaptive auto-associative kernel regression, Ada-AAKR)方法在故障检测方面的有效性.

2 包含离群值的自联想核回归故障检测方法

2.1 自联想核回归算法

根据正常工况下的历史数据, AAKR算法通过最小化样本重构误差来估计模型的参数并依据误差值构造Q 统计量, 建立模型的监测指标[17–18]. 对于每一个待测样本, 其采样值与通过模型估计的正常工况下的重构值之间的差值被用来判断过程是否发生故障. 假设正常工况下的训练数据$\boldsymbol{X}=[\boldsymbol{x}_1\ \ \boldsymbol{x}_2\ \ \cdots\ \ \boldsymbol{x}_n]^{\mathrm{T}}\in\mathbb{R}^{n\times m}$, 其中$n$是样本个数, 每个样本包含$m$个过程变量. AAKR首先基于测试样本$\boldsymbol{x}_*$与每一个训练样本之间的相似度对训练样本给定不同的权重. 其中欧氏距离和马氏距离作为使用最广泛的距离度量函数, 常被用来计算样本之间的相似度. 马氏距离的计算不受变量之间相关性的干扰, 即如下所示:

$$D_{\mathrm{Ma}}^2=(\boldsymbol{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\mu)=(\boldsymbol{x}-\mu)^{\mathrm{T}}(\boldsymbol{U}\Lambda^{-1/2})(\boldsymbol{U}\Lambda^{-1/2})^{\mathrm{T}}(\boldsymbol{x}-\mu).\quad(1)$$

数据集在经过旋转以及压缩或者拉伸之后, 变量之间的线性相关性被剔除. 若给定样本$\boldsymbol{x}_*$和$\boldsymbol{x}_i$, $i=1,\ 2,\ \cdots,\ n$, 则两个样本之间的马氏距离如下:

$$d_i(\boldsymbol{x}_i,\boldsymbol{x}_*)=\sqrt{(\boldsymbol{x}_i-\boldsymbol{x}_*)^{\mathrm{T}}\Sigma_{\boldsymbol{x}}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_*)},\quad(2)$$

其中$\Sigma_{\boldsymbol{x}}$是训练数据的协方差. 本文采用高斯权重函

数来估计每个训练数据在重构$\boldsymbol{x}_*$时的权重

$$K_h(d_i)=\frac{1}{\sqrt{2\pi}h}\exp[-\frac{(d_i)^2}{2h^2}], \tag{3}$$

其中h是待估计的带宽参数. 在分配权重时, 与样本$\boldsymbol{x}_*$的距离越小, 样本之间相似度就越高, $\boldsymbol{x}_i$在重构$\boldsymbol{x}_*$时所分配的权重就越大, 反之亦然. 加权重构函数如下所示:

$$\hat{\boldsymbol{x}}_*=\frac{\sum\limits_{i=1}^{n}K_h(d_i)\cdot\boldsymbol{x}_i}{\sum\limits_{i=1}^{n}K_h(d_i)}. \tag{4}$$

根据$\boldsymbol{x}_*$的实际测量值与重构值之差, 构建平方预测误差(squared prediction error, SPE)或Q统计量,

$$Q_*=\parallel\mathbf{r}_*\parallel^2=(\boldsymbol{x}_*-\hat{\boldsymbol{x}}_*)^{\mathrm{T}}(\boldsymbol{x}_*-\hat{\boldsymbol{x}}_*). \tag{5}$$

AAKR基于训练数据通过k折交叉验证来估计带宽参数, 并采用留一交叉验证的方法计算每个训练样本的统计量. 由于该统计量并不服从某一特定分布, 所以采用核密度估计(kernel density estimation, KDE)方法对模型的控制限进行估计并根据待测样本的统计量是否超过控制限进一步判断过程是否发生故障.

2.2　基于最小化β散度的鲁棒预白化算法

数据白化作为一种剔除变量之间线性相关性的数据预处理方法, 在降低独立成分分析(independent component analysis, ICA)算法复杂度中扮演着重要的角色. 因此在本文中, 对过程变量进行预白化, 降低残差各变量之间的相关性, 提高Q统计量计算的可靠性. 在标准的白化方法中, 样本$\boldsymbol{x}$经过白化后$\boldsymbol{y}=\Sigma^{-\frac{1}{2}}(\boldsymbol{x}-\mu)$, 各变量之间线性无关, 且$E(\boldsymbol{y})=0, E(\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}})=I$, 其中$\mu$和$\Sigma$分别是$\boldsymbol{x}$的均值和协方差. 然而在实际工业过程中, 离群样本的存在会严重影响均值和协方差的计算, 进而导致标准白化方法在剔除变量相关性时出现偏差. 因此, 基于最小化β散度的鲁棒预白化方法被作为ICA中数据去相关预处理步骤而被提出[19].

两个不同概率密度$p(\boldsymbol{x})$和$q(\boldsymbol{x})$之间的β散度为

$$D_\beta(p(\boldsymbol{x}),q(\boldsymbol{x}))=\int[\frac{1}{\beta}(p^\beta(\boldsymbol{x})-q^\beta(\boldsymbol{x}))p(\boldsymbol{x})-\frac{1}{\beta+1}(p^{\beta+1}(\boldsymbol{x})-q^{\beta+1}(\boldsymbol{x}))]\mathrm{d}\boldsymbol{x},\ \beta>0, \tag{6}$$

为了对带离群值的样本进行预白化, 最小化β散度的目标是找到一个与当前带噪声以及离群值的数据集最相似的未标准化高斯密度函数.

$$(\hat{\kappa},\hat{\mu},\hat{\boldsymbol{S}})=\arg\min_{\kappa,\mu,\boldsymbol{S}}D_\beta(p(\boldsymbol{x}),\kappa\psi_{\mu,\boldsymbol{S}}(\boldsymbol{x})), \tag{7}$$

其中: κ是一个正常数, μ和$\boldsymbol{S}$分别是高斯密度函数$\psi_{\mu,\boldsymbol{S}}(\boldsymbol{x})$的均值和协方差. 通过分别求$D_\beta(p(\boldsymbol{x}),\kappa\psi_{\mu,\boldsymbol{S}}(\boldsymbol{x}))$对$\kappa,\mu,\boldsymbol{S}$的偏导进一步求参数的表达式, 最终得到如下所示的迭代公式:

$$\kappa_{t+1}=\frac{1}{n}\sum_{i=1}^{n}\{\psi_{\mu_t,\boldsymbol{S}_t}(\boldsymbol{x}_i)\}^\beta\{\det(2\pi\boldsymbol{S}_t)\}^{\frac{\beta}{2}}\times(\beta+1)^{\frac{m}{2}}, \tag{8}$$

$$\mu_{t+1}=\frac{\sum\limits_{i=1}^{n}\{\psi_{\mu_t,\boldsymbol{S}_t}(\boldsymbol{x}_i)\}^\beta\boldsymbol{x}_i}{\sum\limits_{i=1}^{n}\{\psi_{\mu_t,\boldsymbol{S}_t}(\boldsymbol{x}_i)\}^\beta}, \tag{9}$$

$$\boldsymbol{S}_{t+1}=(\beta+1)\times\frac{\sum\limits_{i=1}^{n}\{\psi_{\mu_t,\boldsymbol{S}_t}(\boldsymbol{x}_i)\}^\beta(\boldsymbol{x}_i-\mu_t)(\boldsymbol{x}_i-\mu_t)^{\mathrm{T}}}{\sum\limits_{i=1}^{n}\{\psi_{\mu_t,\boldsymbol{S}_t}(\boldsymbol{x}_i)\}^\beta}, \tag{10}$$

其中: 样本均值可以作为μ的迭代初始值, 而协方差$\boldsymbol{S}$的初始值可以设定为单位阵, 通过交替迭代式(9)–(10), 最终得到样本的鲁棒均值和协方差. 其中比例因子$\{\psi_{\mu,\boldsymbol{S}}(\boldsymbol{x}_i)\}^\beta$是每个样本$\boldsymbol{x}_i$在计算$\hat{\mu},\hat{\boldsymbol{S}}$时的权重

$$\phi(\boldsymbol{x}_i|\mu,\boldsymbol{S})=\{\psi_{\mu,\boldsymbol{S}}(\boldsymbol{x}_i)\}^\beta. \tag{11}$$

明显地, 当样本$\boldsymbol{x}_i$越偏离正常数据点, 其在计算均值和协方差时的权重就越小. 这种迭代加权方法确保了当数据集中存在离群值时对于均值和协方差计算的鲁棒性. 算法步骤如下所示:

步骤 1　给定输入: $\boldsymbol{X}$, 参数β_0.

步骤 2　将n个训练样本近似均匀地随机分为K个子集: $\mathcal{T}(1),\mathcal{T}(2),\cdots,\mathcal{T}(K)$.

步骤 3　给定$\beta=0,\ \zeta,\ 2\zeta,\ 3\zeta,\ \cdots,\ \beta_0$, 其中$\zeta$是步长.

- 将$\mathcal{T}(j)$作为验证集, 剩余子集构成训练样本$\mathcal{T}^{-j}=\{\boldsymbol{x}|\boldsymbol{x}\notin\mathcal{T}(j)\}$;
- 基于训练集$\mathcal{T}^{-j}$, 通过最小化$D_\beta(p(\boldsymbol{x}),\kappa\psi_{\mu,\boldsymbol{S}}(\boldsymbol{x}))$估计参数$\hat{\kappa},\hat{\mu},\hat{\boldsymbol{S}}$;
- 基于估计所得$\hat{\kappa}_\beta,\hat{\mu}_\beta,\hat{\boldsymbol{S}}_\beta$, 利用$\mathcal{T}(j)$计算交叉验证目标函数$CV_j=-\mathcal{L}_{\beta_0}(\boldsymbol{x}\in\mathcal{T}(j);\hat{\kappa}_\beta,\hat{\mu}_\beta,\hat{\boldsymbol{S}}_\beta)$;
- $\hat{\mathcal{O}}_{\beta_0}(\beta)=\dfrac{1}{n}\sum\limits_{j=1}^{K}CV_j$.

步骤 4　输出$\hat{\beta}=\arg\min_\beta\hat{\mathcal{O}}_{\beta_0}(\beta)$.

目标函数$\mathcal{L}_{\beta_0}$为

$$\mathcal{L}_{\beta_0}(\boldsymbol{x};\hat{\kappa}_\beta,\hat{\mu}_\beta,\hat{\boldsymbol{S}}_\beta)=\frac{1}{n\beta_0}\sum_{\boldsymbol{x}\in\mathcal{T}(j)}\{\hat{\kappa}_\beta\psi_{\hat{\mu}_\beta,\hat{\boldsymbol{S}}_\beta}(\boldsymbol{x})\}^{\beta_0}-\frac{(\hat{\kappa}_\beta)^{\beta_0+1}}{(\beta_0+1)^{\frac{m+2}{2}}[\det(2\pi\hat{\boldsymbol{S}}_\beta)]^{\frac{\beta_0}{2}}}, \tag{12}$$

其中: β_0的取值并不影响β的估计, 只需确保$\beta_0>\beta$即可. 当数据集不存在噪声及离群值时, $\hat{\beta}_{\mathrm{opt}}=0$; 如果样本存在离群值, $\hat{\mathcal{O}}_{\beta_0}(\beta)$的估计曲线存在一个拐点, 该点横坐标即为最优值$\hat{\beta}_{\mathrm{opt}}$. 基于最小化$\beta$散度的

预白化方法在一定程度上避免了离群值对于样本均值以及协方差计算的影响, 使得最终的估计结果更接近正常工况下样本的均值和协方差. 经过鲁棒预白化之后, 训练样本$\boldsymbol{x}$被转换为$\boldsymbol{y}=\hat{\boldsymbol{S}}^{-\frac{1}{2}}(\boldsymbol{x}-\hat{\mu})$.

2.3 包含离群值的自联想核回归故障检测算法

通常, AAKR都是基于训练数据交叉验证的方法估计带宽参数h以及模型的监测指标的. 但是当数据集中包含离群值时, 不仅样本之间的相似度计算存在偏差, 影响重构值的估计, 也会为了最小化离群值的重构误差造成带宽参数的估计出现严重错误. 因此为了提高故障的检出率, 减少因离群值而造成故障漏报, 本文基于式(11)的比例因子构造截断函数. 在经过迭代计算得到$\hat{\mu}, \hat{\boldsymbol{S}}$的值之后, 各训练样本的比例因子为$\hat{\phi}(\boldsymbol{x}_i|\hat{\mu},\hat{\boldsymbol{S}})=\{\psi_{\hat{\mu},\hat{\boldsymbol{S}}}(\boldsymbol{x}_i)\}^{\hat{\beta}}$, 为了便于表达, 比例因子简写为$\hat{\phi}_{\boldsymbol{x}_i}$, 则截断函数如下所示:

$$\rho(\boldsymbol{x}_i)=\begin{cases}1,\ \hat{\phi}_{\boldsymbol{x}_i}\geqslant\phi_{\rm c},\\ 0,\ \hat{\phi}_{\boldsymbol{x}_i}<\phi_{\rm c},\end{cases}\tag{13}$$

$\phi_{\rm c}$是一个介于(0,1)之间的参数, 其值是根据训练数据的$\hat{\phi}_{\boldsymbol{x}_i}=\{\psi_{\hat{\mu},\hat{\boldsymbol{S}}}(\boldsymbol{x}_i)\}^{\hat{\beta}}$给定的. 将训练集中所有样本的$\hat{\phi}_{\boldsymbol{x}_i}, i=1,2,\cdots,n$, 按从大到小排列, 可以发现, 正常数据的比例因子之间都非常接近, 而离群值的比例因子明显小于正常样本. 因此从经过排序后的比例因子曲线图1中, 可以直观地发现该曲线初始部分几乎呈水平状, 有轻微的下降趋势, 然而出现离群样本时, 该曲线骤然下降, 因此该拐点可以被设定为截断函数的参数$\phi_{\rm c}$. 除了按照比例因子曲线图中的拐点来设定参数$\phi_{\rm c}$, 也可以依据曲线图中每个点的梯度来设定. 对于正常样本的比例因子, 其梯度接近于0, 然而$\phi_{\rm c}$这一点的梯度远远偏离正常值, 因此选择第一个梯度较大的数据点的比例因子作为参数$\phi_{\rm c}$的值.

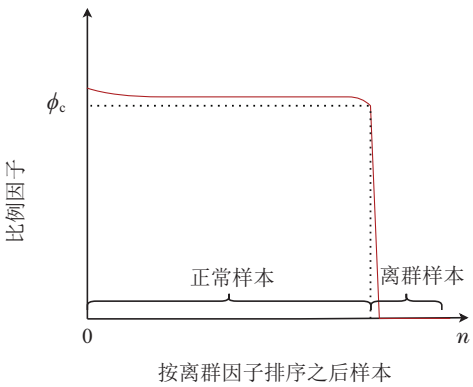


图 1 样本比例因子降序图

Fig. 1 Descending diagram of sample scaling factors

按照式(13)给定的截断函数

$$\tau_i=\rho(\boldsymbol{x}_i)||\rho(\boldsymbol{x}_*),\ i=1,2,\cdots,n,\tag{14}$$

对式(3)中在计算$\boldsymbol{x}_*$重构值时的权重$K_h(d_i)$作进一步加权处理, 得到新的样本权重计算

$$K_{\rm new}(d_i)=\tau_iK_h(d_i)=\tau_i\frac{1}{\sqrt{2\pi}h}\exp(-\frac{(d_i)^2}{2h^2}).\tag{15}$$

当测试集中的样本$\boldsymbol{x}_*$的截断函数值为0时, 用来重构当前样本的训练数据不包含离群值, 只选择正常数据点对当前样本进行重构计算. 但是对于正常的待测样本, 其截断函数为1, 无论是训练集中的正常数据还是离群值, 都能经过距离度量在重构待测样本时合理分配权重.

2.4 基于残差加权的统计量计算方法

基于AAKR的故障检测方法在估计数据重构误差时, 没有考虑对样本的降维处理, 即在式(5)的$\|\boldsymbol{r}_*\|^2=r_{*,1}^2+r_{*,2}^2+\cdots+r_{*,m}^2$计算中, 各残差变量的权重均为1. 事实上, 当只有少部分过程变量因为故障而偏离正常工况时, 最理想的方法是只基于这一部分残差变量计算统计量, 但是这往往需要大量的先验知识, 并且在实际系统中并不适用. 为了确保故障信息不被淹没, 考虑对部分可能携带故障的残差变量进行加权处理, 即在计算统计量时, 对正常变量给定较小的权重, 缩小其在统计量计算中所占的比重, 而对于可能携带故障信息的变量, 给定较大的权重, 进一步放大该故障信息[20–21].

在正常工况下, 基于KDE方法, 分别计算各个残差变量基于99%置信度的控制限$\mathrm{cl}_k, k=1,2,\cdots,m$, 然后估计待测样本残差各变量与控制限的比值

$$R_{*,k}=\frac{r_{*,k}^2}{\mathrm{cl}_k},\tag{16}$$

将其作为判断各残差变量携带故障信息量的依据, 如果$R_{*,k}>1$, 则认为该变量可能携带故障, 需要被进一步放大; 而当$R_{*,k}\leqslant 1$时, 则认为该变量运行在正常工况下, 可以缩小其在Q统计量中所占的比重. 因此, 经过变量加权之后的统计量变为

$$Q_*=[r_{*,1}\ \ r_{*,2}\ \ \cdots\ \ r_{*,m}]\cdot\begin{bmatrix}R_{*,1}&&&\\&R_{*,2}&&\\&&\ddots&\\&&&R_{*,m}\end{bmatrix}\times[r_{*,1}\ \ r_{*,2}\ \ \cdots\ \ r_{*,m}]^{\rm T}.\tag{17}$$

通过对残差各变量赋予不同的权重, 最终使得Q统计量中包含更多的故障信息, 降低微小故障被淹没的可能. 在新统计量的基础上, 基于KDE方法, 重新估计模型控制限. 然而, 在系统实际运行过程中, 样本会受到各种噪声因素的影响, 例如测量噪声以及过程噪声, 进而导致部分变量可能会在正常工况的临界点附近上下波动, 出现式(16)中部分变量的$R_{*,k}$略大于1的情况, 经过残差加权之后会进一步放大这种误差, 使得

正常样本的统计量超过控制限, 产生误报. 因此, 为了减小模型的误报率, 可以认为这种由随机噪声引起的差异是瞬时扰动, 能够通过滤波的方法予以消除. 本文基于EWMA的方法构建一种新的统计量, 在时间轴上对最近的数据赋予更大的权重, 对距离当前采样点较远的历史数据给定一个较小的权重, 根据当前a时刻的统计量Q_a以及前$w-1$个样本的统计量, 构造新的指数加权统计量

$$Q_{a_{wi}} = \frac{cQ_{a-w+1} + c^2Q_{a-w+2} + \cdots + c^wQ_a}{\sum\limits_{l=1}^{w} c^l}, \tag{18}$$

如果$Q_{a_{wi}} \geqslant \mathrm{Cl}_\delta$, 则认为系统出现故障, 其中$\mathrm{Cl}_\delta$是基于训练数据的统计量$Q_i, i = 1, 2, \cdots, n$计算的控制限. 将式(18)进行变换, 当前待测样本的统计量Q_a满足以下条件时, 样本被判定为故障数据[22]:

$$Q_a \geqslant \mathrm{Cl}_\delta + \sum_{l=1}^{w-1} c^{l-w}[\mathrm{Cl}_\delta - Q_{a-w+l}] = \mathrm{Cl}_{ad,\delta}, \tag{19}$$

其中: c为常数, 且$c \geqslant 1$. 当c值等于1时, 基于指数加权滑动平均的自适应统计量转变为基于滑动窗的平均值自适应统计量. 根据式(19), 当前时刻待测样本自适应控制限为原本固定的控制限加上控制限与前$w-1$个时刻样本的统计量之间偏差的指数加权滑动平均. 如果系统运行在正常工况下, 样本统计量位于控制限以下, 此时自适应控制限大于原始固定控制限, 即$\mathrm{Cl}_{\mathrm{ad},\delta} > \mathrm{Cl}_\delta$. 但是当系统发生故障时, 样本的统计量超过控制限, $\mathrm{Cl}_\delta - Q_{a-w+l} < 0$, 随着故障幅值进一步增大, 整个自适应控制限的值小于0, 即$\mathrm{Cl}_{\mathrm{ad},\delta} < 0$. 因此在判断样本是否发生故障时, 需要在式(19) 的基础上给定一个正常数$0 < \lambda < \mathrm{Cl}_\delta$. 即当如下公式满足时, 样本$a$ 就认为是携带故障信息的:

$$Q_a \geqslant \max\{\mathrm{Cl}_{\mathrm{ad},\delta}, \lambda\}. \tag{20}$$

结合式(16)–(17), (20), 针对每个变量的控制限cl_k, 结合EWMA的思想, 重新构造各变量自适应加权控制限

$$\mathrm{cl}_{ad,k} = \max\{\mathrm{cl}_k + \sum_{l=1}^{w-1} c^{l-w}[\mathrm{cl}_k - r^2_{a-w+l,k}], \lambda_k\}. \tag{21}$$

此时, 残差各变量的自适应权重就变为

$$R^{\mathrm{new}}_{a,k} = \frac{r^2_{a,k}}{\mathrm{cl}_{\mathrm{ad},k}}. \tag{22}$$

将其代入式(17), 可以构造样本新的自适应加权统计量Q^{new}_*.

2.5　模型参数

本文所提出的故障检测模型涉及到几个需要人为设定的参数:

1) 鲁棒白化算法中的参数β_0: 只要保证$\beta_0 > \hat{\beta}$, 其大小对白化算法影响不大.

2) 截断函数中的参数ϕ_c: 该参数值是根据训练样本比例因子降序排序之后人为给定的, 也可以同时参照比例因子的曲线图及其梯度图设定. 由于该参数的大小决定了某些训练样本是否参与待测样本的重构, 因此会出现以下几种情况: 1) 如果故障样本与训练样本中的离群值相似度很小, 那么ϕ_c的取值是否精确对最终过程监测的效果影响很小; 2) 如果带故障的测试样本与离群值非常接近, 此时ϕ_c设置过小会导致训练数据中部分离群值参与故障样本的重构, 使得故障数据的重构误差较小, 最终由于Q统计量低于控制限而产生漏报; 3) 如果ϕ_c值过大, 测试集中的正常样本会因为ρ值为0无法依靠训练集中某些正常且ρ值也为0的相似样本进行重构计算, 最终导致重构误差变大, 系统产生误报. 在实际应用中, 由于离群值明显偏离正常数据, 正常训练样本与离群样本的比例因子会存在明显的差异, 因此能准确设定ϕ_c的值.

3) 权重因子c: 权重因子的大小决定了历史残差信息在构建当前残差变量的权重中所占据的比重. 若c值等于1, 基于EWMA的加权方法就退化成基于滑动窗的均值加权方法, 虽然能在一定程度上降低误报率, 但是也会导致故障检测延迟; c值越大, 式(21)中c^{l-w}就越小, 历史残差信息对于权重计算的影响也越小.

4) 滑动窗口宽度w: 窗口宽度的大小决定了参与当前残差变量权重计算的历史数据数量, w值过大不仅会影响计算速度, 也会减弱较远时刻数据点的指数加权值对式(21)的影响.

5) 自适应加权中的参数λ_k: λ_k值的取值不能超过cl_k, 否则会因为$R^{\mathrm{new}}_{a,k}$过小故障信息并没有被放大而产生漏报. 合适的取值范围为$(0, \mathrm{cl}_k)$, 在该范围内, λ_k值取较小值, 带故障信息的变量残差会被进一步放大, 故障的检出率会有所提升, 但是对正常数据的尖刺, 也会因为残差变量的权重过大而可能导致误报率升高.

2.6　故障检测步骤以及流程图

离线建模:

1) 将带离群值的训练集$\boldsymbol{X}$按照鲁棒预白化的方法, 经过5折交叉验证, 得到最优参数$\hat{\beta}$;

2) 计算基于最优$\hat{\beta}$的鲁棒均值和协方差$\hat{\mu}$, $\hat{\boldsymbol{S}}$, 并对原始样本进行白化计算, 得到白化后的数据$\boldsymbol{Y}$;

3) 根据鲁棒均值和协方差, 按照式(11), 计算每个训练样本的比例因子$\hat{\phi}_{\boldsymbol{x}_i}$;

4) 根据比例因子排序后的曲线图或者该曲线图上每个点的梯度值, 设置截断函数参数ϕ_c;

5) 正常的验证集$\boldsymbol{V}_A$经过鲁棒预白化之后, 通过最小化重构误差来估计带宽参数h并计算每个变量的Q统计量以及基于99%置信度的控制限cl_k;

6) 正常的验证集$\boldsymbol{V}_B$经过鲁棒预白化之后, 基于式(22)EWMA自适应加权, 按照式(17)得到每个样本新的统计量Q_*^{new};

7) 根据新的统计量, 通过KDE方法确定基于99%置信度的样本最终控制限CL;

在线检测:

1) 对采集的测试样本鲁棒预白化, 得到白化后数据$\mathbf{y}_{\text{test}}$以及对应的比例因子$\hat{\phi}_{\text{test}}$;

2) 根据式(13)–(15)以及式(4)计算待测数据的重构误差;

3) 对残差各变量基于EWMA自适应加权, 得到最终的样本统计量Q_{test};

4) 根据统计量是否超过控制限CL, 判断该样本是否携带故障信息. 本文所提算法流程图如图2所示.

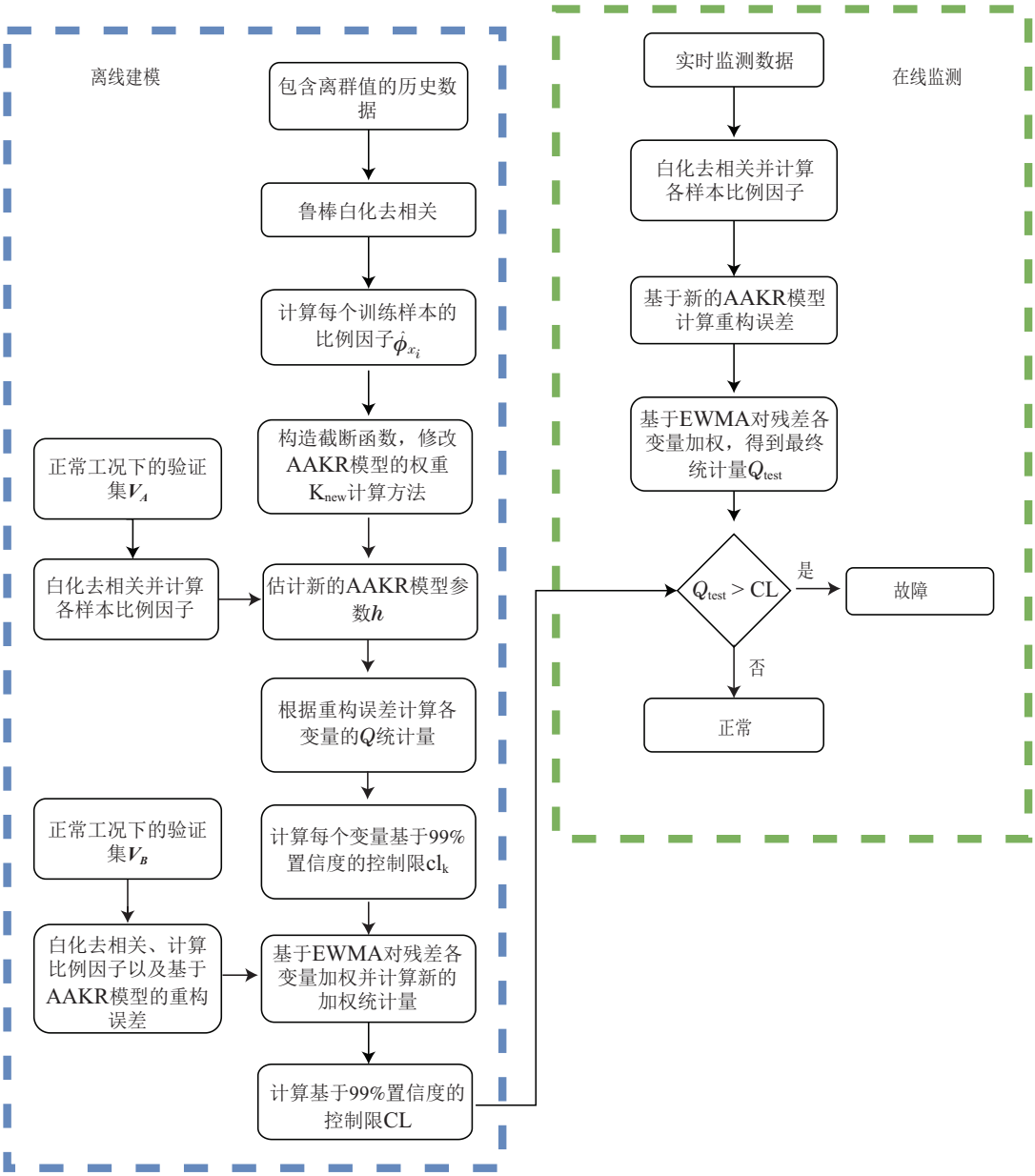


图 2 Ada-AAKR故障检测流程图

Fig. 2 The flowchart of Ada-AAKR for fault detection

3 仿真验证

本文选择TE过程数据验证带离群值的AAKR算法的故障检测效果. TE过程可以通过模拟实际工业过程得到较为理想条件下的样本集, 因此被广泛用来验证模型故障检测的效果[23–24]. TE过程包含22个过程测量变量、19个成分测量变量以及11个控制变量共52个变量. 选择其中的过程变量和控制变量共33个变量的采样数据建立监测模型. 首先利用500个正常工况下的样本组成训练数据集$\boldsymbol{X}$并随机挑选5%的样本加入离群信息. 针对这25个样本, 在每一个样本中随机挑选$l(l \in [1, 10])$个变量, 添加$5\sigma_i, i = 1, 2, \cdots, m$的离群信息, 其中$\sigma_i$是变量$i$的标准差. 基于最小化$\beta$散度

鲁棒预白化的方法, 对带离群值的训练样本进行白化计算, 其中参数β_0设置为1. 将鲁棒预白化后训练样本的比例因子降序排序, 图3是当前带离群值的各训练样本的比例因子.

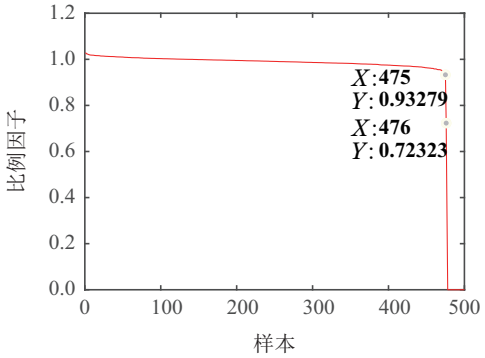


图 3　训练样本的比例因子

Fig. 3　The scaling factors of the training data

图4是各样本比例因子梯度的绝对值. 结合这两张图, 可以发现在第475个点之后, 样本的比例因子发生严重变化, 因此该点的比例因子作为截断函数的参数ϕ_c. 在另一组正常工况下采集的960 个样本中, 挑选前400个样本作为验证集$\boldsymbol{V}_A$, 经过最小化重构误差估计带宽参数, 计算此时验证集每个残差变量的Q统计量并基于统计量计算99%置信度的控制限cl_k; 再将401～700 个样本作为验证集$\boldsymbol{V}_B$, 基于EWMA计算残差变量加权的自适应权重$\mathrm{cl}_{ad,k}$以及经过加权之后样本新的统计量, 得到模型最终的控制限CL.

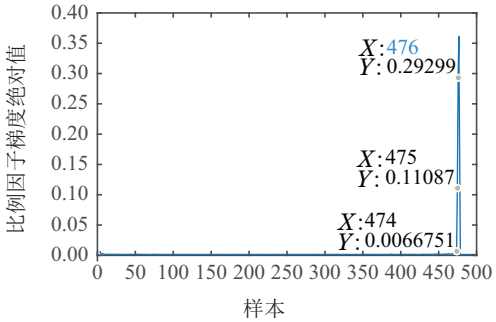


图 4　训练样本的比例因子梯度绝对值

Fig. 4　The absolute gradient of the scaling factors of the training data

本章设置3个仿真对比实验, 分别验证: 1) 基于马氏距离的AAKR故障检测方法对比基于欧式距离的AAKR故障检测方法的优越性; 2) 数据预白化在本文所提的Ada-AAKR故障检测方法中的重要性以及变量自适应加权是否有助于提高故障检测性能; 3) 截断函数的加入是否能进一步提高故障的检出率.

3.1　基于欧氏距离/马氏距离的AAKR故障检测方法

本节主要对比在AAKR算法的样本相似度计算时, 选择欧式距离还是马氏距离对于模型最终的故障检测结果的影响. 首先对训练集进行预处理, 将离群值通过MATLAB中的“rmoutliers”函数进行剔除, 之后分别基于欧式距离与马氏距离计算变量之间的相似度, 按照式(3)计算样本权重, 基于原始的AAKR算法重构观测数据, 得到重构误差. 表1是两种方法以及基于PCA的故障检测结果.

表 1　3种方法故障检测效果

Table 1　Fault detection performance of three methods

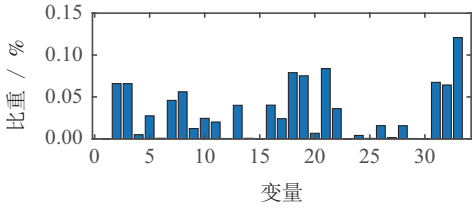
故障	误报率/%			漏报率/%		
	PCA	AAKR[1] Euc	AAKR[2] Mah	PCA	AAKR Euc	AAKR Mah
1	4.375	**1.25**	2.5	**0**	0.125	0.25
2	5	1.875	**1.25**	**0.75**	1.25	1.625
3	6.25	**5.625**	9.375	87.125	**80**	85.625
4	5.625	3.125	**1.25**	**0**	**0**	0.625
5	5.625	3.125	**1.25**	64.75	59.125	**28**
6	4.375	**0.625**	**0.625**	**0**	**0**	**0**
7	5	5.625	**1.875**	**0**	**0**	**0**
8	**3.125**	5.625	5.625	1.625	**0.875**	1.25
9	**15.625**	26.25	23.125	88.75	**84**	86.625
10	**3.125**	5	4.375	32.75	31.625	**20.375**
11	7.5	6.875	**3.75**	**16**	18.125	21.625
12	**4.375**	11.875	11.875	0.875	**0.25**	**0.25**
13	2.5	1.25	**0.625**	4.75	**4.5**	**4.5**
14	5	5	**2.5**	**0**	**0**	**0**
15	3.125	**1.25**	**1.25**	82.125	**76.625**	78.625
16	**20**	38.125	28.75	37.625	31.875	**14.625**
17	6.25	3.75	**2.5**	**3**	3.75	5.75
18	7.5	5.625	**1.875**	**8.5**	9	**8.5**
19	3.125	3.125	**2.5**	54.25	62.375	**50.25**
20	4.375	0.625	**0**	31.75	28.625	**19.375**
21	11.25	10	**5.625**	**39.75**	47.125	50

[1] 基于欧式距离的AAKR故障检测方法.

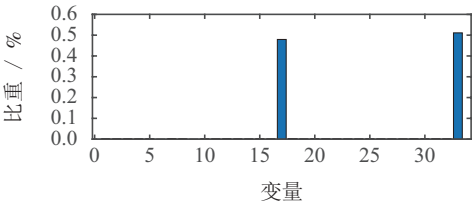
[2] 基于马氏距离的AAKR故障检测方法.

从表中可以看出, 基于马氏距离的AAKR检测方法在故障误报率方面, 整体而言效果更好; 对于故障的检出率, 3种方法在不同类型的故障上取得的检测效果略有差异. 对于故障5, 10, 16和20, 基于马氏距离的AAKR算法的漏报率更低, 检出的故障更多; 而PCA对于故障11和21的检测效果更好. 对比基于欧氏距离和马氏距离的两种AAKR故障检测方法, 在大部分故障类型中, 其故障检出率差别不大, 但是在考虑变量之间相关性之后, 某些故障的检出率有了大幅度的提升. 以故障5为例, 冷凝水进口温度的阶跃变化引发了系统故障, 只有第11个操纵变量——冷凝器冷却水流量在系统发生故障后, 故障信息一直存在, 其他变量在第161个样本开始, 有较大幅值的波动, 但是由于TE过程中闭环控制回路的存在, 在故障发生一段时间后, 大部分变量的波动幅度恢复正常, 此时不考虑

变量之间相关性的欧式距离无法准确度量故障样本与训练样本之间的距离, 导致大量故障漏报. 而剥离变量之间相关性之后, 原本变化并不明显的其他变量的故障信息变得非常清晰, 因为这些变量与冷凝器冷却水流量的相关性最强, 因此在计算样本相似度时, 携带故障信息的变量能帮助准确度量该故障样本与正常样本之间的距离. 图5给出了故障5中第500个样本与训练集第200个正常样本之间在分别计算欧式距离与马氏距离时, 各变量所占的比重. 可以非常直观地发现, 马氏距离更能凸显故障信息对于距离计算的影响; 而在欧式距离中, 第500个测试样本的第18, 19, 31, 32个变量值与正常样本差别很小, 因此在度量样本相似度时, 容易导致该样本与正常样本之间相似度变高, 故障信息被淹没, 最终造成系统漏报.



(a)各变量在欧式距离中所占的比重



(b)各变量在马氏距离中所占的比重

图 5 各变量在距离计算中所占的比重

Fig. 5 The ratios of the variables in the Euclidean/Mahalanobis distance.

3.2 数据预白化以及变量自适应加权对于Ada-AAKR故障检测方法的影响

在其他条件不变的情况下, 本节分别比较了: 1) 省略预白化步骤的Ada-AAKR算法; 2) 未对残差变量自适应加权, 直接利用Q统计量的鲁棒预白化的AAKR算法; 3) 本文所提的经过鲁棒预白化之后自适应加权的Ada-AAKR算法的故障检测性能. 表2分别列出了21种故障下3种方法的误报率和漏报率. 对比这3种方法的误报率和漏报率, 本文提出的算法整体效果更好. 尤其是从第10个故障到第20个故障, 其故障的检出率更高. 但是对于故障3, 9, 16, 自适应加权的方法会使得系统误报增多, 其原因在于这种加权方法只对引起误报的中等和个别尖刺具有鲁棒性, 当引起误报的尖刺较大或者正常工况下连续几个样本都超过正常范围, 即式(22) 中因为当前待测样本是幅值较大的尖刺, 那么即使$\mathrm{cl}_{\mathrm{ad},k} > \mathrm{cl}_k$, 因为$r_{a,k}^2$较大, $R_{a,k}^{\mathrm{new}} \gg 1$; 或者由于连续几个样本的统计量都超过了控制限$\mathrm{cl}_k$, 使得$\mathrm{cl}_{\mathrm{ad},k} < \mathrm{cl}_k$, 可能会出现$r_{a,k}^2 > \mathrm{cl}_{\mathrm{ad},k}$的情况, 残差变量经过进一步放大之后, 会导致最终统计量变大, 产生误报. 以故障10为例, 该故障是物料C温度的随机扰动. 由于阀门的调节作用, 部分变量的采样值以较大的幅度来回振荡, 造成阶段性的部分样本包含的故障信息变少. 如图6所示, 3种方法在第490~560个样本的故障检出率存在明显的不同, 经过变量预白化之后, 更多故障样本能被检测出, 而经过进一步对残差各变量加权之后, 又有少量样本的统计量超过控制限, 故障的检出率得到了进一步的提高.

表 2 3种方法故障检测效果

Table 2 Fault detection performance of three methods

故障	误报率/%			漏报率/%		
	A[1]	B[2]	C[3]	A	B	C
1	1.25	0.625	**0**	0.625	**0**	0.125
2	1.875	**0**	0.625	3	1.375	**1.125**
3	**0.625**	5	13.75	97.75	93.75	**88.375**
4	1.875	1.875	**1.25**	90.125	**0**	**0**
5	1.875	1.875	**1.25**	60.125	**0**	**0**
6	0.625	**0**	**0**	**0**	**0**	**0**
7	**0.625**	**0.625**	3.125	1.25	**0**	**0**
8	**0**	**0**	0.625	4.75	**1.625**	**1.625**
9	**2.5**	13.125	24.375	96.5	95.75	**92.75**
10	**0.625**	1.875	**0.625**	45	9.75	**6.875**
11	**0.625**	1.875	1.25	83.75	20.625	**12**
12	**1.25**	**1.25**	**1.25**	4.625	**0.125**	**0.125**
13	**0**	**0**	**0**	5.625	**4.625**	**4.625**
14	2.5	2.5	**1.25**	41.875	**0**	**0**
15	1.875	**0**	**0**	90	82.25	**76.75**
16	**5.625**	10	15.625	40.375	6.625	**3.5**
17	**0.625**	**0.625**	1.25	38.75	3.375	**2.75**
18	1.875	1.25	**0.625**	11.375	9.75	**8.625**
19	0.625	**0**	**0**	92	6.875	**3.5**
20	0.625	**0**	**0**	49	8.5	**8.25**
21	**2.5**	4.375	8.75	61.375	36.875	**30.75**

[1] A: 缺少数据预白化的Ada-AAKR算法.

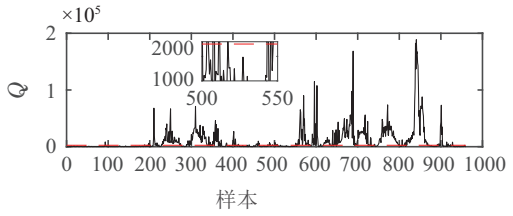
[2] B: 缺少残差变量自适应加权的Ada-AAKR算法.

[3] C: Ada-AAKR算法.

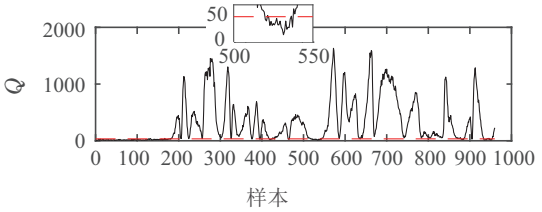
3.3 截断函数对于提高故障检测率的影响

本节以故障11为例, 通过比较包含截断函数与不包含截断函数的Ada-AAKR故障检测方法在部分故障样本检出率上的差异, 来证明加入截断函数能进一步提高模型的故障检测性能. 由于截断函数针对的是当故障样本与训练集中的离群点非常相似的情况, 因此在本小节中, 随机挑选故障11中第501~550个样本, 人为添加$e \sim \mathcal{N}(0, 0.001^2)$的高斯白噪声, 替换训练集中第31~80个正常样本来模拟这种情况. 图7是包含截断函数与不包含截断函数对这50个故障样本的检测情况.

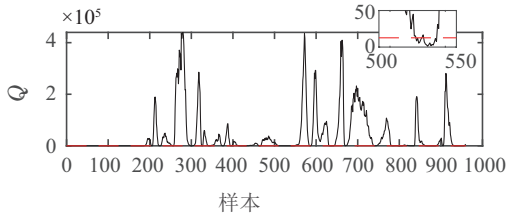
从图7中可以清晰地发现, 当故障样本与训练集中某些样本存在极高的相似度时, 该故障样本的重构值会接近于该训练样本, 进而导致重构误差变小, 系统发生漏报. 因此在这种情况下, 截断函数能通过设定的ϕ_c参数, 将该训练样本排除于故障样本的重构值计算过程.



(a) Ada-AAKR (无数据预白化)



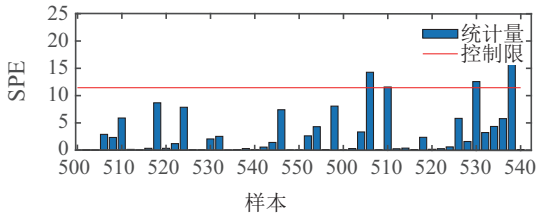
(b) Ada-AAKR (无残差变量自适应加权)



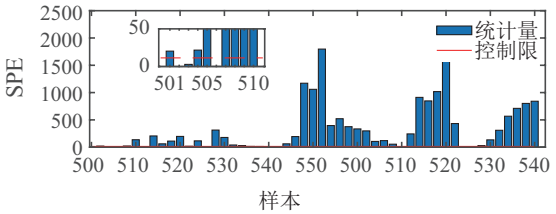
(c) Ada-AAKR

图 6　故障10的检测结果图

Fig. 6　The monitoring performance of fault 10



(a) Ada-AAKR (无截断函数)



(b) Ada-AAKR

图 7　故障11第501～550个样本的检测结果

Fig. 7　The monitoring performance of samples from 501 to 550 of fault 11

4　总结

本文介绍了一种训练样本带离群值的自适应加权AAKR故障检测方法, 将样本集经过最小化β散度的鲁棒预白化之后建立AAKR故障检测模型, 针对故障样本可能与样本集中的离群样本高度相似而被误判为正常样本的情况, 通过截断函数选择较为正常的训练样本对其进行重构计算. 为了进一步提高故障的检出率, 避免微小故障被噪声淹没, 同时为了降低故障的误报率, 提出了残差变量自适应加权的AAKR故障检测方法. 通过TE过程的仿真, 验证了本文提出的方法在过程监控方面具有一定的效果.

参考文献:

[1] GE Z. Review on data-driven modeling and monitoring for plantwide industrial processes. *Chemometrics & Intelligent Laboratory Systems*, 2017, 171: 16 – 25.

[2] MACGREGOR J, CINAR A. Monitoring, fault diagnosis, fault tolerant control and optimization: Data driven methods. *Computers & Chemical Engineering*, 2012, 47: 111 – 120.

[3] QIN S. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 2012, 36(2): 220 – 234.

[4] CHEN H, JIANG B, DING S, et al. Probability-relevant incipient fault detection and diagnosis methodology with applications to electric drive systems. *IEEE Transactions on Control Systems Technology*, 2019, 27(6): 2766 – 2773.

[5] ZHOU B, GU X. Multi-block statistics local kernel principal component analysis algorithm and its application in nonlinear process fault detection. *Neurocomputing*, 2020, 376: 222 – 231.

[6] LUO L, WANG J, TONG C, et al. Multivariate fault detection and diagnosis based on variable grouping. *Industrial & Engineering Chemistry Research*, 2020, 59(16): 7693 – 7705

[7] ZHOU B, YE H, ZHANG H, et al. Process monitoring of iron-making process in a blast furnace with PCA-based methods. *Control Engineering Practice*, 2016, 47: 1 – 14.

[8] CANDES E, LI X, MA Y, et al. Robust principal component analysis. *Journal of the ACM*, 2021, 58(3): 1 – 37.

[9] ISOM J, LABARRE R. Process fault detection, isolation, and reconstruction by principal component pursuit. *American Control Conference*. San Francisco: IEEE, 2011: 238 – 243.

[10] PAN Y, YANG C, SUN Y, et al. Fault detection with principal component pursuit method. *Journal of Physics Conference Series*, 2015, 659(1): 012035.

[11] THARRAULT Y, MOUROT G, RAGOT J, et al. Fault detection and isolation with robust principal component analysis. *International Journal of Applied Mathematics & Computer Science*, 2008, 18(4): 429 – 442.

[12] GADHOK N, KINSNER W. A study of outliers for robust independent component analysis. *Canadian Conference on Electrical and Computer Engineering*. Niagara Falls: IEEE, 2004, 3: 1421 – 1425.

[13] CAI L, TIAN X. A new fault detection method for non-gaussian process based on robust independent component analysis. *Process Safety and Environmental Protection*, 2014, 92(6): 645 – 658.

[14] LUO L, BAO S, TONG C. Sparse robust principal component analysis with applications to fault detection and diagnosis. *Industrial & Engineering Chemistry Research*, 2019, 58(3): 1300 – 1309.

[15] WANG D, ROMAGNOLI J. Robust multi-scale principal components analysis with applications to process monitoring. *Journal of Process Control*, 2005, 15(8): 869 – 882.

[16] YU W, ZHAO C. Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net. *IEEE Transactions on Control Systems Technology*, 2020, 28(3): 1083 – 1091.

[17] YU J, YOO J, JANG J, et al. A novel hybrid of autoassociative kernel regression and dynamic independent component analysis for fault detection in nonlinear multimode processes. *Journal of Process Control*, 2018, 68: 129 – 144.

[18] BARALDI P, MAIO F, TURATI P, et al. Robust signal reconstruction for condition monitoring of industrial components via a modified auto associative kernel regression method. *Mechanical Systems & Signal Processing*, 2015, 60/61: 29 – 44.

[19] MOLLAH M, EGUCHI S, MINAMI M. Robust prewhitening for ICA by minimizing β-divergence and its application to fastICA. *Neural Processing Letters*, 2007, 25: 91 – 110.

[20] JIANG Q, YAN X. Probabilistic monitoring of chemical processes using adaptively weighted factor analysis and its application. *Chemical Engineering Research and Design*, 2014, 92(1): 127 – 138.

[21] JIANG Q, YAN X. Non-Gaussian chemical process monitoring with adaptively weighted independent component analysis and its applications. *Journal of Process Control*, 2013, 23(9): 1320 – 1331.

[22] BAKDI A, KOUADRI A. A new adaptive PCA based thresholding scheme for fault detection in complex systems. *Chemometrics and Intelligent Laboratory Systems*, 2017, 162: 83 – 93.

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

[24] RICKER N. Optimal steady-state operation of the tennessee eastman challenge process. *Computers & Chemical Engineering*, 1995, 19(9): 949 – 959.

作者简介:

沈飞凤 博士研究生, 目前研究方向为软测量建模和工业过程故障检测, E-mail: 7161905004@vip.jiangnan.edu.cn;

杨慧中 教授, 博士, 目前研究方向为复杂工业过程建模与优化控制, E-mail: yhz@jiangnan.edu.cn.