

控制与决策

Control and Decision

基于多块信息提取的AUV资源勘查系统故障检测

郭大权, 杨宗圣, 周晓锋, 李帅, 徐春晖

引用本文:

郭大权, 杨宗圣, 周晓锋, 等. 基于多块信息提取的AUV资源勘查系统故障检测[J]. 控制与决策, 2021, 36(4): 790–800.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.0732>

您可能感兴趣的其他文章

Articles you may be interested in

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>

基于交叉熵的改进NPE间歇过程故障检测算法

Improved NPE batch process fault detection algorithm based on cross entropy

控制与决策. 2021, 36(2): 411–417 <https://doi.org/10.13195/j.kzyjc.2019.0725>

事件触发机制下分布时滞网络化控制系统 H_∞ 故障检测

Event-triggered H_∞ fault detection for networked control systems with distributed delays

控制与决策. 2020, 35(12): 3059–3065 <https://doi.org/10.13195/j.kzyjc.2019.0456>

基于不变网络模型和故障注入的分布式信息系统故障溯源方法

Fault source location algorithm for distributed information system based on invariant network and fault injection

控制与决策. 2020, 35(11): 2723–2732 <https://doi.org/10.13195/j.kzyjc.2019.0214>

基于多块信息提取的AUV资源勘查系统故障检测

郭大权^{1,2,4}, 杨宗圣^{1,2,3,4}, 周晓锋^{1,2,4†}, 李 帅^{1,2,3,4}, 徐春晖^{1,2}

(1. 中国科学院沈阳自动化研究所, 沈阳 110016; 2. 中国科学院机器人与智能制造创新研究院, 沈阳 110169; 3. 中国科学院大学, 北京 100049; 4. 中国科学院网络化控制系统重点实验室, 沈阳 110016)

摘要: 针对“潜龙二号”AUV在实际航行过程中,资源勘查系统传感器数据具有多重变量相关性、故障类型多样、受运行状态和环境变化影响数值变化大以及噪声强等问题,提出一种新的基于多块信息提取的主元分析(PCA)故障检测方法. 首先,针对变量之间的多重相关性,通过滑窗和相关系数的方法提取变量间相关性信息;然后,根据变化率在不同运行状态和环境下基本稳定的特点,对于不同类型故障,分别提取变化率信息和变化率信息的各阶统计量累积误差信息;最后,基于提取的特征信息建立3个子块,对每个子块分别建立PCA模型并进行检测,将检测的结果通过中值滤波去噪后,用贝叶斯推断进行融合. 通过对“潜龙二号”实际运行数据进行检测,验证所提方法的有效性.

关键词: AUV; 主元分析; 故障检测; 信息提取; 多块建模

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.0732

开放科学(资源服务)标识码(OSID):



引用格式: 郭大权, 杨宗圣, 周晓锋, 等. 基于多块信息提取的AUV资源勘查系统故障检测[J]. 控制与决策, 2021, 36(4): 790-800.

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

GUO Da-quan^{1,2,4}, YANG Zong-sheng^{1,2,3,4}, ZHOU Xiao-feng^{1,2,4†}, LI Shuai^{1,2,3,4}, XU Chun-hui^{1,2}

(1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; 2. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China; 3. University of Chinese Academy of Sciences, Beijing 100049, China; 4. Key Laboratory of Network Control System, Chinese Academy of Sciences, Shenyang 110016, China)

Abstract: Aiming at the problems of AUV of Qianlong 2 in the actual navigation process, such as multi-variable correlation, multiple fault types, large numerical variation influenced by operation status and environmental changes, and strong noise, a new principal component analysis (PCA) fault detection method based on block information extraction is proposed. Firstly, according to the multiple correlations among variables, the correlation information between variables is extracted by sliding window and correlation coefficient method. Secondly, according to the basic stable characteristics of change rate in different operating states and environments, for different types of faults, the cumulative error information of each order statistics of change rate information and change rate information is extracted separately. Finally, three sub-blocks are built based on the extracted feature information, and the PCA model is built and tested for each sub-block respectively. After de-noising by median filtering, the detected results are fused by Bayesian inference. The effectiveness of the proposed method is verified by testing the actual operation data of Qianlong 2.

Keywords: AUV; principal component analysis; fault detection; information extraction; multi-block modelling

0 引言

工作在复杂深海环境中,自主水下机器人(autonomous underwater vehicle, AUV)发生故障后,不仅无法完成勘测任务,严重时会造成潜水器损坏甚至丢失^[1]。“潜龙二号”AUV资源勘查系统受到严苛环

境影响,是潜水器故障发生率最高的子系统之一,并且因其肩负着采集海洋环境数据的重任,勘测任务失败将带来巨大的损失,因而亟需研究对资源勘查系统的故障检测方法.

当前潜水器故障检测方法主要分为3种^[2-3]:需

收稿日期: 2019-05-27; 修回日期: 2019-10-09.

基金项目: 国家重点研发计划项目(2017YFC0306800).

责任编辑: 胡庆雷.

†通讯作者. E-mail: zhouxf@sia.cn.

要建立数学机理模型的解析模型方法,依赖于逻辑规则的定性模型方法和基于传感器历史数据的数据驱动方法。但是,潜水器系统结构越来越复杂并且专家经验和过程知识难以获取,因此定量模型和定性模型都难以建立^[4-5]。而基于数据驱动的多元统计方法,如主元分析(principal component analysis, PCA)、独立成分分析(independent component analysis, ICA),通过提取原始数据中的隐含信息建立检测模型,已经广泛应用到故障检测领域^[6]。PCA作为一种典型的多元统计方法,在降维同时保持了数据中的重要信息和特征^[7],并且它本身的降维和抗干扰特性非常适合潜水器的故障检测。但是传统PCA存在缺陷,如主元选择的随机性以及无法处理非线性数据等,因此一些学者从各个方面进行了改进:在文献[8]中,改进了传统PCA主元选取的不确定性,提出了基于平均特征值的累积百分比方差的新方法来选取主元;在文献[9]中,为了提高传统PCA处理非线性数据的能力,使用核主元(KPCA)进行潜水器故障检测;在文献[10]中,通过加权互信息选出包含质量变量信息量大的一组过程变量,然后通过PCA进行检测,降低了冗余数据的影响;在文献[1]中,为了提高传统PCA模型的实时性,将基于信度分配的模糊小脑神经网络(FCA-CMAC)与PCA相结合,建立了电流的预测模型。

这些方法都是对过程数据建立单一的PCA模型,面对潜水器资源勘查系统产生的复杂过程变量,无法取得准确且全面的检测效果。针对复杂过程数据,通常采用多块建模的方法全面提取信息,降低复杂性。传统的多块建模方法通常按照过程知识进行分块^[11-12],但是实际研究中,过程知识难以获取,因此基于过程数据分块的方法成为研究热点。在文献[13-14]中考虑变量间的关系,利用相关性信息划分子块,在工业过程监控领域取得了较好的检测成果,但是忽略了数据自身的分布和变化情况;在文献[15]中考虑数据自身分布特点,采用改进的Jarque-Bera检测方法将数据划分为高斯和非高斯信息块,然后对高斯子块和非高斯子块分别用PCA和ICA进行监测;在文献[16]中,不仅从原始数据中提取特征信息,而且考虑数据自身的变化,提取了变化率信息和累积误差信息,可以更好地监控各种类型故障。

上述方法只注重研究变量间的相关性或者数据自身的分布和变化,无法全面提取原始数据中的信息。基于以上分析,结合当前潜水器数据特点和故障特点,提出一种基于多块信息提取的PCA故障检测方法(MIE-PCA)。首先,提取反映变量之间关系的相

关性信息、反映数据自身变化的变化率信息和变化率各阶统计量累积误差信息;然后,对每个子块分别用PCA进行监测;最后,为了降低误报率,在线建模时将每个子块的统计量先进行滤波,再构建BIC统计量进行检测。通过在“潜龙二号”潜水器实际航行数据(30~40共10个潜次)上进行实验来验证该方法不仅对各种故障类型都有较低的漏报率,并且在降低噪声影响后有较低的误报率,因此相对于其他算法,有非常高的准确率,可满足实际应用要求。

1 “潜龙二号” AUV

“潜龙二号”AUV资源勘查系统主要包括深/高度计、温盐深仪、多参数水质仪、甲烷传感器和三分量磁力仪。因此,在研究时选择17个变量,如表1所示。

表1 变量名称

深/高度计深度值	温盐仪声速	甲烷值
载体深度	磁力仪X轴分量	甲烷传感器温度
温盐仪电导率	磁力仪Y轴分量	多参数仪温度
温盐仪压力	磁力仪Z轴分量	多参数仪浊度
温盐仪温度	磁力仪总场量	甲烷电压
温盐仪盐度	多参数仪氧化还原电位值	

由于特殊的工作环境,采集到的数据有以下特点:

- 1) 数据量大,运行状态复杂。每个航行潜次都能采集到数万甚至数十万个样本,并且能够反映潜水器运行的各种状态。
- 2) 变量间有着多重相关性,一些变量之间的相关性十分明显。
- 3) 不同运行状态、不同潜次、变量的数值变化范围不同。
- 4) 传感器容易受到干扰,因此采集到的数据噪声大。

为了方便检测,对已经存在或者可能存在的故障进行分类汇总,如表2所示。

表2 故障汇总

故障类型	故障描述
卡死型故障	传感器的输出值某时刻不变
阶跃型故障	输出值在某时刻发生跳变
缓变型故障	传感器的输出值缓慢上升或下降
震荡型故障	传感器的输出值在某个时刻发生震荡
混合型故障	多种类型故障混合

根据以上分析,本文提出分块建模方法(图1)来解决上述问题。首先,充分利用变量间已存在的相关性,横向提取相关性信息;然后,对于相关性无法反映的故障,纵向提取变量自身的变化率信息;其次,对于上面提取的信息无法反映的其他故障,进一步提取变

化率的各阶统计量累积误差信息;接着,为了反映出潜水器的各个运行状态,选择尽可能多的样本来训练PCA模型;最后,在线检测时原始数据中存在噪声,因此对每个子块的统计量分别进行滤波,再用贝叶斯方法,将每个滤波后的统计量进行融合,形成一个BIC统计量,方便检测。

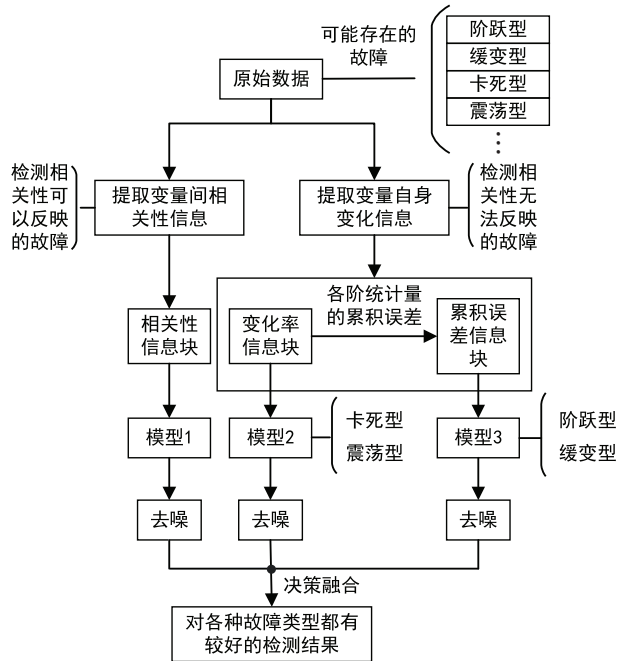


图1 本文多块建模思想

2 主元分析

主元分析(PCA)是常用统计分析方法,该方法可以选择能最大化表征数据特征的方差信息^[17].方法的核心思想是降维,并且通过构造一组正交向量,尽可能保持了原始数据中的变化^[18].

假设给定经过标准化后的数据矩阵 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T \in \mathbf{R}^{n \times m}$, n 为数据样本个数, m 为变量个数. 首先通过下式计算数据矩阵的协方差矩阵^[14]:

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X}. \quad (1)$$

然后对得到的协方差矩阵 \mathbf{S} 通过下式^[19]进行分解:

$$\mathbf{S} = \mathbf{V} \mathbf{A} \mathbf{V}^T. \quad (2)$$

其中: $\mathbf{V} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, λ_1 为协方差矩阵的特征值, 并且 $\lambda_1 > \lambda_2 > \dots > \lambda_n$. \mathbf{V} 中的列向量为特征值对应的特征向量, 两两正交. 选取 \mathbf{V} 中的前 d 个向量组成负载矩阵 \mathbf{P} , 其中 d 表示保留主元的个数, 通常通过累积百分比方差获得:

$$\sum_{i=1}^d \lambda_i / \sum_{i=1}^m \lambda_i \times 100\% \geq 85\%. \quad (3)$$

使用负载矩阵对原始数据进行降维:

$$\mathbf{T} = \mathbf{X} \mathbf{P}, \quad (4)$$

$$\mathbf{E} = \mathbf{X} - \mathbf{T} \mathbf{P}^T = \mathbf{X} (\mathbf{I} - \mathbf{P} \mathbf{P}^T). \quad (5)$$

其中: $\mathbf{T} \in \mathbf{R}^{n \times k}$ 为得分矩阵, 即降维后的矩阵; $\mathbf{E} \in \mathbf{R}^{n \times m}$ 为残差矩阵. 经过主元提取后, 原始数据空间被划分为 $\mathbf{T} \in \mathbf{R}^{n \times k}$ 所代表的主元子空间和 $\mathbf{E} \in \mathbf{R}^{n \times m}$ 所代表的残差子空间. 对于最新获取的样本 \mathbf{x}_i , 在主元子空间和残差子空间分别建立 Hotelling T^2 和 SPE 检测统计量:

$$T_i^2 = \mathbf{x}_i^T \mathbf{P} \mathbf{L} \mathbf{P}^T \mathbf{x}_i, \quad (6)$$

$$\text{SPE}_i = \mathbf{e}_i^T \mathbf{e}_i, \quad (7)$$

其中 $\mathbf{e}_i = \mathbf{x}_i^T (\mathbf{I} - \mathbf{P} \mathbf{P}^T)$. 将所获取的 T^2 和 SPE 统计量分别与各自的控制限进行比较(计算方法见文献[20]), 进而判断是否发生故障。

3 基于MIE-PCA的AUV资源勘查系统故障检测

3.1 分块信息提取

根据第2节的分析, 针对潜水器数据特点和已知的故障类型, 分别提取原始数据的相关性信息、变化率信息和变化率统计量累积误差信息。

3.1.1 基于斯皮尔曼的相关性信息提取

相关系数表示两个变量之间的相关关系, 反映数据的发展模式^[21]. 如果相关性高, 则意味着当一个变量发生变化时, 另一个变量也会很大可能发生变化, 可以根据这种关系监测到有着强相关关系的变量中发生的故障。

斯皮尔曼(Spearman)相关系数是对取值的等级进行的计算, 而不是对取值的本身进行计算, 因此该方法与取值的尺度无关, 对于数据错误和极端值的反应不敏感, 相对于其他的相关性度量方法更加适合当前数据的特点. 假设 $\mathbf{X}, \mathbf{Y} \in \mathbf{R}^{n \times 1}$ 表示待计算的变量, x_i, y_i 表示对应第 i 个值的等级, 具体的计算过程如下^[22]:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}. \quad (8)$$

其中: $d_i = x_i - y_i$ 为观察等级的差值, $-1 \leq \rho \leq 1$ 为相关系数. 为了充分利用原始数据变量之间的相关性信息, 通过滑窗计算窗口内的斯皮尔曼相关系数。

假设原始数据集为 $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n]^T \in \mathbf{R}^{n \times m}$, n 为样本数量, m 为变量个数, 滑窗的窗口大小为 len , 步长为 step . 因此, 第 i 个窗口内的数据为 $\mathbf{W}_i = [\mathbf{X}_i, \mathbf{X}_{i+1}, \dots, \mathbf{X}_{i+\text{len}-1}]^T \in \mathbf{R}^{\text{len} \times m}$.

对窗口内的变量通过式(8)两两计算斯皮尔曼相关系数,提取的相关性信息为

$$\mathbf{X}_i^c = [p_{1,2}, p_{1,3}, \dots, p_{m-1,m}] \in \mathbf{R}^{1 \times \frac{m \times (m-1)}{2}}$$

其中: \mathbf{X}_i^c 为相关性信息中的第 i 个样本, $p_{1,2}$ 为第 i 个滑窗内变量 1 和变量 2 的斯皮尔曼相关系数,通过滑窗构造新的相关性信息数据集为

$$\mathbf{X}^c = [\mathbf{X}_1^c, \mathbf{X}_2^c, \dots, \mathbf{X}_{n-1-\text{len}}^c]^T \in \mathbf{R}^{(n-1-\text{len}) \times \frac{m \times (m-1)}{2}}$$

相关性信息只能反映出一部分故障信息,在某些情况下会造成漏报,例如: 1) 故障造成几个相关性强的变量发生趋势相同变化的情况; 2) 故障造成几个相关性关系较弱的变量发生变化的情况。

3.1.2 变化率信息提取

当利用相关性信息无法检测出故障时,按照已知的故障类型,从每个变量的自身数值分析. 变化率信息反映样本中每个变量的变化速率,对变量数值的突然变化敏感,虽然不同潜次变量的数值尺度会发生变化,但是变化率与尺度无关,并且在潜水器正常运行状况下基本稳定. 因此可用变化率信息检测变化率持续不变的卡死型故障或者持续变化的震荡型故障.

已知原始数据集为 $\mathbf{X} = \mathbf{R}^{n \times m}$, 原始数据中第 i 个样本 ($i \geq 1$) 的变化率为

$$\mathbf{X}^L(i) = \mathbf{X}(i+1) - \mathbf{X}(i). \quad (9)$$

其中: $\mathbf{X}^L(i)$ 为变化率信息中的第 i 个样本, $\mathbf{X}(i)$ 为原始数据中的第 i 个样本. 构造的新的变化率信息矩阵为 $\mathbf{X}^L = [\mathbf{X}_1^L, \mathbf{X}_2^L, \dots, \mathbf{X}_{n-1}^L]^T \in \mathbf{R}^{(n-1) \times m}$.

3.1.3 变化率的统计量累积误差信息提取

变化率与尺度无关,并且能够反映出变量在时间上的变化规律,但是对于缓变型故障和阶跃型故障无法取得很好的效果. 而统计量是指计算原数据的一阶、二阶或者高阶统计量,可以发现原数据中更加详细的信息,因此可以通过滑窗提取变化率各阶统计量的信息,能够反映出更加细微的变化.

上节获取的变化率信息矩阵为 $\mathbf{X}^L \in \mathbf{R}^{(n-1) \times m}$, 为了和相关性信息提取时一致,设定滑窗的长度为 $\text{len} - 1$, 步长为 step . 第 i 个窗口内的数据为

$$\mathbf{W}_i = \begin{bmatrix} X_{i,1}^L & X_{i,2}^L & \dots & X_{i,m}^L \\ X_{i+1,1}^L & X_{i+1,2}^L & \dots & X_{i+1,m}^L \\ \vdots & \vdots & \ddots & \vdots \\ X_{i+\text{len}-2,1}^L & X_{i+\text{len}-2,2}^L & \dots & X_{i+\text{len}-2,m}^L \end{bmatrix} \in \mathbf{R}^{(\text{len}-1) \times m}. \quad (10)$$

对该窗口内的每一列变量分别计算一阶统计量均值 ($\mu_{i,j}$), 二阶统计量方差 ($v_{i,j}$), 以及高阶统计量偏斜度 ($s_{i,j}$) 和峰度 ($k_{i,j}$), 具体计算公式^[23] 如下:

$$\mu_{i,j} = \frac{1}{\text{len} - 1} \sum_{q=0}^{\text{len}-2} \mathbf{X}_{i+q,j}^L, \quad (11)$$

$$v_{i,j} = \frac{1}{\text{len} - 1} \sum_{q=0}^{\text{len}-2} [\mathbf{X}_{i+q,j}^L - \mu_{i,j}]^2, \quad (12)$$

$$s_{i,j} = \frac{\frac{1}{\text{len} - 1} \sum_{q=0}^{\text{len}-2} [\mathbf{X}_{i+q,j}^L - \mu_{i,j}]^3}{(v_{i,j})^{3/2}}, \quad (13)$$

$$k_{i,j} = \frac{\frac{1}{\text{len} - 1} \sum_{q=0}^{\text{len}-2} [\mathbf{X}_{i+q,j}^L - \mu_{i,j}]^4}{(v_{i,j})^2} - 3. \quad (14)$$

将同一滑窗内提取的统计量,按照行向量进行排序,构成一个统计量样本 $\tilde{\mathbf{X}}^S$. 然后通过滑窗更新窗口内的样本,提取变化率中的统计量信息,得到信息矩阵为

$$\tilde{\mathbf{X}}^S = [\tilde{\mathbf{X}}_1^S, \tilde{\mathbf{X}}_2^S, \dots, \tilde{\mathbf{X}}_{n-\text{len}-1}^S]^T \in \mathbf{R}^{(n-\text{len}-1) \times (m \times 4)}$$

虽然统计量信息可以反映原始数据中的细微变化,但是建立的统计模型仍然会对这些细小的变化不敏感,对于阶跃型故障也仅能反映出故障发生时刻附近的变化,无法持续反映故障,从而造成很高的漏报率. 而累积误差是累积观测值信息与设定的标准值的差而得到的信息,可以将统计量中的细微变化放大,并且当阶跃故障发生后可以持续反映故障^[16].

将统计量信息矩阵进行标准化后,各变量服从均值为 0, 方差为 1 的标准正态分布,将均值作为设定的标准值,可以将标准化后的样本通过下式相加:

$$\mathbf{X}_i^S = \sum_{i=1}^{n-\text{len}-1} \tilde{\mathbf{X}}_i^S, \quad (15)$$

得到统计量的累积误差信息矩阵

$$\tilde{\mathbf{X}}^S \in \mathbf{R}^{(n-\text{len}-1) \times (m \times 4)}$$

3.2 基于中值滤波的去噪方法

在故障检测领域,大部分学者的研究目标是降低漏报率,但是,在实际应用中,误报率也是非常重要的指标. 中值滤波是一种非线性数字滤波技术,常用来消除图像噪声,尤其对于椒盐噪声有着比其他滤波方法更好的效果^[24].

本文在进行离线建模时,输入所有训练数据,分块提取信息后使用 PCA 建立各个子模型,因为信息提取 PCA 本身具有抗干扰特性,所以计算出的各统

计量的控制限受噪声影响较小. 但是在线监测时, 每次输入单个样本, 模型无法去噪, 因此计算出来的统计量仍然含有大量噪声. 如果直接对各个子块的统计量用贝叶斯推断进行融合, 则每个子块中的噪声都会影响最终融合结果, 获得的BIC统计量会含有大量噪声, 不仅影响研究人员对检测结果的直接观察, 而且会增高误报率(图2). 传统的去噪方法是对原始数据中的单变量进行去噪, 没有考虑变量之间的关系和变量前后变化的信息, 可能将正常数据错判为噪声数据, 因此会导致信息丢失, 而且统计量包含了输入样本的大部分信息, 因此本文从理论分析和实际应用出发, 通过中值滤波对模型计算出的统计量进行去噪, 这样不仅避免信息丢失, 而且最后融合的BIC统计量噪声较小, 误报率也明显降低, 方便研究人员对检测结果分析(图3).

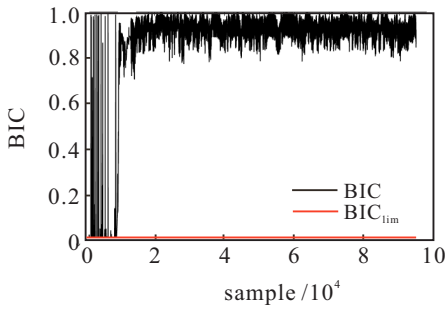


图2 混合型故障未去噪的BIC

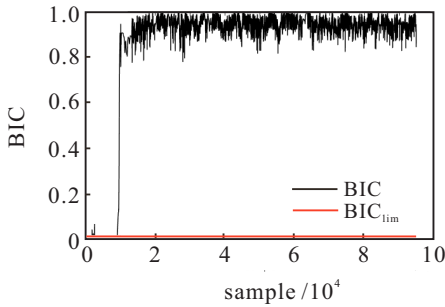


图3 混合型故障去噪后的BIC

假设原始数据集 $\mathbf{X} \in \mathbf{R}^{n \times 1}$, 滤波后构造的数据集为 $\tilde{\mathbf{X}} \in \mathbf{R}^{n \times 1}$. $\tilde{\mathbf{X}}(i)$ 计算公式如下:

$$\begin{cases} \text{Median}[\tilde{\mathbf{X}}(i-2), \tilde{\mathbf{X}}(i-1), \mathbf{X}(i), \\ \mathbf{X}(i+1), \mathbf{X}(i+2)], & i \geq 3; \\ \mathbf{X}(i), & 1 \leq i \leq 2. \end{cases} \quad (16)$$

其中: $\tilde{\mathbf{X}}(i)$ 为滤波后的第 i 个值, $\mathbf{X}(i)$ 为滤波前的第 i 个值. 对原始数据第 i 个值 ($i \geq 3$) 进行去噪时, 构造一个窗口, 窗口内包含当前未去噪值 $\mathbf{X}(i)$, 当前样本的后两个未去噪值 $\mathbf{X}(i+1)$ 和 $\mathbf{X}(i+2)$, 以及前两个去噪后的值 $\tilde{\mathbf{X}}(i-2)$ 和 $\tilde{\mathbf{X}}(i-1)$, 对窗口内的数据进

行排序, 计算中值为当前窗口的输出.

3.3 基于贝叶斯推断的故障检测

对每个子块进行监测时, 每个子块都有两个统计量指标 T^2 和 SPE, 通过贝叶斯推断计算统一的统计量进行检测, 会大大提高检测效率. 以提取的相关性信息矩阵 \mathbf{X}^c 为例, T^2 统计量的故障概率可以通过如下公式^[25] 计算:

$$P_{T^2(F|\mathbf{X}^c)} = \frac{P_{T^2(\mathbf{X}^c|F)}P_{T^2F}}{P_{T^2\mathbf{X}^c}}, \quad (17)$$

$$P_{T^2\mathbf{X}^c} = P_{T^2(\mathbf{X}^c|N)}P_{T^2N} + P_{T^2(\mathbf{X}^c|F)}P_{T^2F}, \quad (18)$$

$$P_{T^2(\mathbf{X}^c|N)} = \exp\left(-\frac{T_{\text{lim}}^2}{T_{\text{lim}}^2}\right), \quad (19)$$

$$P_{T^2(\mathbf{X}^c|F)} = \exp\left(-\frac{T_{\text{lim}}^2}{T_{\text{new}}^2}\right). \quad (20)$$

其中: $P_{T^2(F|\mathbf{X}^c)}$ 为相关性信息矩阵 \mathbf{X}^c 中 T^2 统计量的故障概率; P_{T^2N} 为正常样本的先验概率, 置信度为 α (本文 α 取值 0.01); P_{T^2F} 为故障样本的先验概率, 置信度为 $1 - \alpha$; T_{lim}^2 为相关性信息矩阵的 T^2 控制限; T_{new}^2 为新样本的 T^2 统计量.

然后对 3 个信息块中计算出的 T^2 和 SPE 统计量进行中值滤波, 最后进行融合形成 BIC 统计量^[26]:

$$\text{BIC} = \frac{\sum (P_{T^2(F|\mathbf{X}^c)}^2 + P_{\text{SPE}(F|\mathbf{X}^c)}^2 + P_{T^2(F|\mathbf{X}^L)}^2)}{\sum (P_{T^2(F|\mathbf{X}^c)} + P_{\text{SPE}(F|\mathbf{X}^c)} + P_{T^2(F|\mathbf{X}^L)}^2)} + \frac{\sum (P_{\text{SPE}(F|\mathbf{X}^L)}^2 + P_{T^2(F|\mathbf{X}^S)}^2 + P_{\text{SPE}(F|\mathbf{X}^S)}^2)}{\sum (P_{\text{SPE}(F|\mathbf{X}^L)}^2 + P_{T^2(F|\mathbf{X}^S)}^2 + P_{\text{SPE}(F|\mathbf{X}^S)}^2)}. \quad (21)$$

3.4 MIE-PCA 算法流程

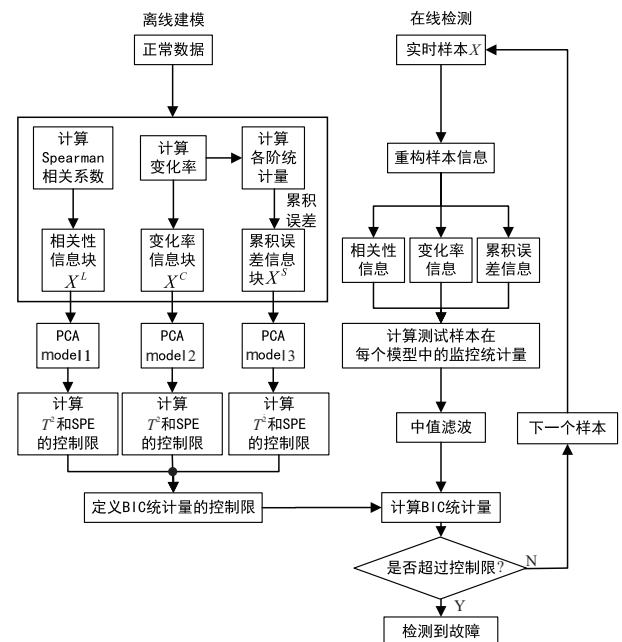


图4 MIE-PCA 算法流程

离线建模:

step 1: 对采集到的数据进行预处理,选取实际运行过程中的正常数据进行建模;

step 2: 对正常数据进行分块信息提取,分别获得相关性信息矩阵、变化率信息矩阵和变化率的累积误差信息矩阵;

step 3: 在每个信息矩阵中分别建立PCA模型;

step 4: 计算每个子模型的监控统计量控制限并定义BIC控制限.

在线监测阶段:

step 1: 对于新来的样本,延时 $len - 1$ 个样本后开始检测;

step 2: 利用分块信息提取的方法,重构原始样本的样本信息,获得新的测试样本;

step 3: 对于每个子块用建立的PCA模型进行检测,获得每个子块检测的监控统计量;

step 4: 延时两秒后获得每个子块中监控统计量的滤波结果;

step 5: 利用贝叶斯推断得到3个子块的联合统计量BIC;

step 6: 与定义的BIC控制限比较,获得检测结果.

离线建模时采用滑窗方法,因此会损失前 $len - 1$ 个样本. 虽然会造成样本损失,但是原始数据中包含的信息并没有丢失,并且数据集中的前几千个样本包含信息较少(潜水器还未进入航行状态),丢失前 len (通常选择1000左右)个样本不会造成影响. 在线监测时,为了提取多块信息,需要对统计量滤波,不仅用到了当前时刻的统计量,而且用到了当前时刻后两秒的样本,因此在线监测时,对每一个样本延迟两秒后进行检测. 但是,由于传感器噪声大,当前时刻的数

据可能是无效数据或者噪声数据,因此延迟两秒给出监测结果反而会降低误报率,提高监测性能.

4 检测结果分析

本文选取“潜龙二号”AUV在33~39潜次采集到的共50多万条数据进行实验,其中正常运行数据30多万条,故障数据20多万条. 已知的故障是漏电故障(即混合型故障),为了验证算法的性能,选取30000个正常样本模拟4种类型故障,并和现存的漏电故障一起进行验证,各类故障如下:

1) 模拟故障1:卡死型故障,变量“温盐仪温度值”在第21000个样本开始不变;

2) 模拟故障2:阶跃型故障,变量“多参数仪氧化还原电位值”在21000点开始加上一个大小为0.7的阶跃;

3) 模拟故障3:缓变型故障,变量“多参数仪氧化还原电位值”在21000点开始加上一个斜率为0.0001的缓慢上升故障;

4) 模拟故障4:震荡型故障,多参数仪氧化还原电位值在21000个点开始增加一个服从 $N(0, 1.5^2)$ 的随机扰动;

5) 现存故障5:混合型故障,由于电机漏电,多个传感器的输出值发生异常.

训练数据选取正常运行状态下的30万条数据,漏电故障的测试数据为95679条,从第11000个样本开始发生故障. 实验中,通过累计方差贡献率(CPV)选取主元个数,其中设定累积贡献率为90%,各个统计量的置信度为99%,滑窗长度 $len = 1000$,步长 $step = 1$. 为了验证算法的性能,表3给出3个子块分别对5种故障的漏报率.

表3 故障汇总

故障类型	子块						BIC
	子块1		子块2		子块3		
	T^2	SPE	T^2	SPE	T^2	SPE	
1	66.2	11.6	99.3	94.7	62.1	20.4	6.9
2	98.2	20.4	98.8	95.6	63.9	0	0
3	98.2	74.6	98.8	96.6	39.6	0	7.63
4	100	95.1	0.38	6.45	1.5	2.0	0.2
5	57.8	3.2	10.8	24.9	0.5	0.3	0.2

图5是各子块对模拟故障1的监测结果. 因为有许多变量与故障发生变量有着强相关关系,可以发现子块1的SPE统计量很好地检测到故障的发生;而由于该变量存在很强的噪声,子块2无法检测到故障,

而子块3的SPE统计量虽然也可以检测到故障,但是根据表3可以发现漏报率较高. 最后的BIC统计量结合了各个子块的优势,漏报率达到最低,这说明提取相关性信息是有必要的.

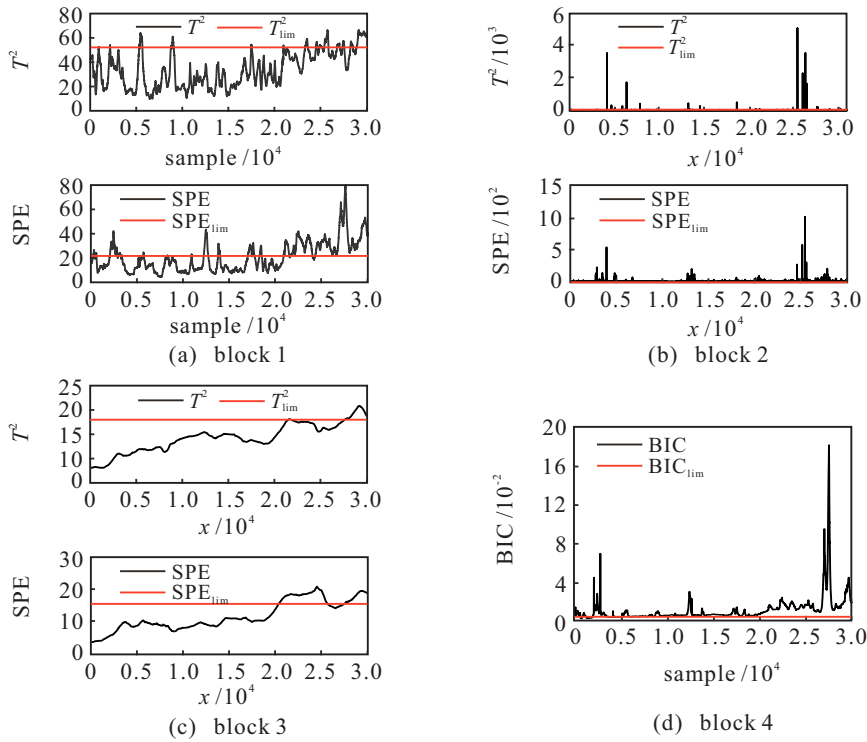


图5 各子块对模拟故障1的监测结果

图6是各子块对于模拟故障2的监测结果. 因为数据集中, 没有与故障发生的变量有很强相关性的变量, 子块1无法检测出来; 虽然子块2在故障刚发生时检测得到, 但是变化率只在故障发生时产生变化,

因此不能持续地反映故障; 而变化率的累积误差信息可以将变化率发生的变化进行累加, 能持续反映故障, 因此只有子块3检测出来该故障. 而最后的BIC统计量继承了子块3的优势, 漏报率也为0.

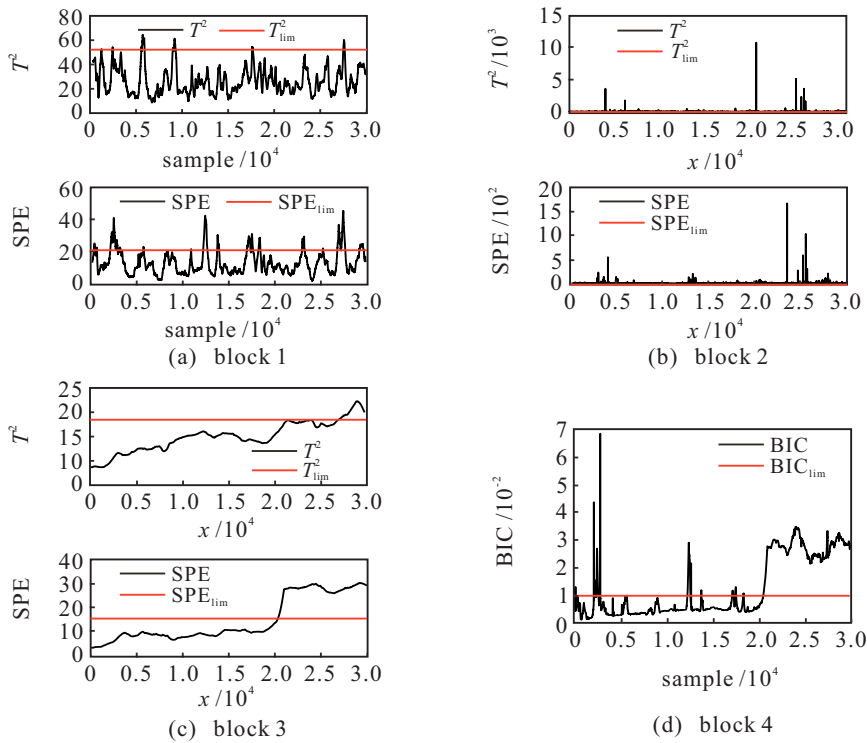


图6 各子块对模拟故障2的监测结果

图7是各子块对模拟故障3的检测结果. 因为噪声太强, 缓变的故障特征被噪声覆盖, 相关性信息和

变化率信息均无法在强噪声中提取出有用的故障特征, 因此子块1和子块2均无法检测出故障; 而变化率

的各级统计量信息能够反映出数据中更加细微的变化, 累积误差又将这些变化进行放大, 所以子块3可

以很好地检测出故障. 最后的BIC虽然受到了子块1和子块2的影响, 但相对而言漏报率也较低.

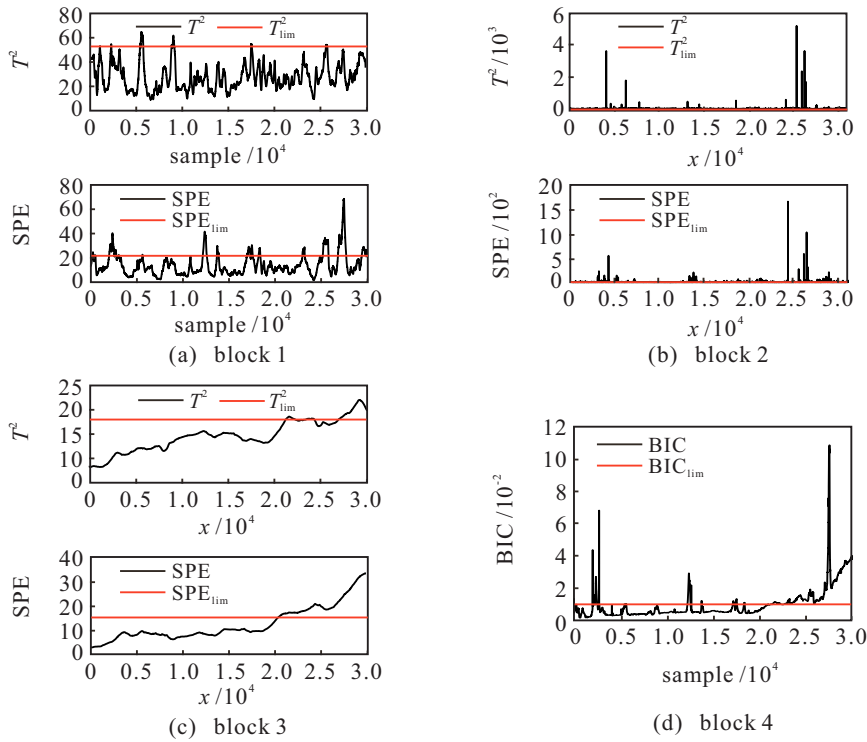


图7 各子块对模拟故障3的监测结果

图8是各子块对模拟故障4的检测结果. 由于噪声的存在, 相关性信息仍然无法反映出故障特征, 子块1无法检测出故障; 变化率持续发生变化, 因此子块2可以很好地检测出故障, T^2 的漏报率仅为

0.34%; 虽然变化率的统计量累积误差可以检测出故障, 但是累积误差是逐渐将微小变化放大, 因此有一定的漏报率. 而最终的BIC统计量在计算时, 会综合各个子块统计量的优势, 因此使得误报率最低.

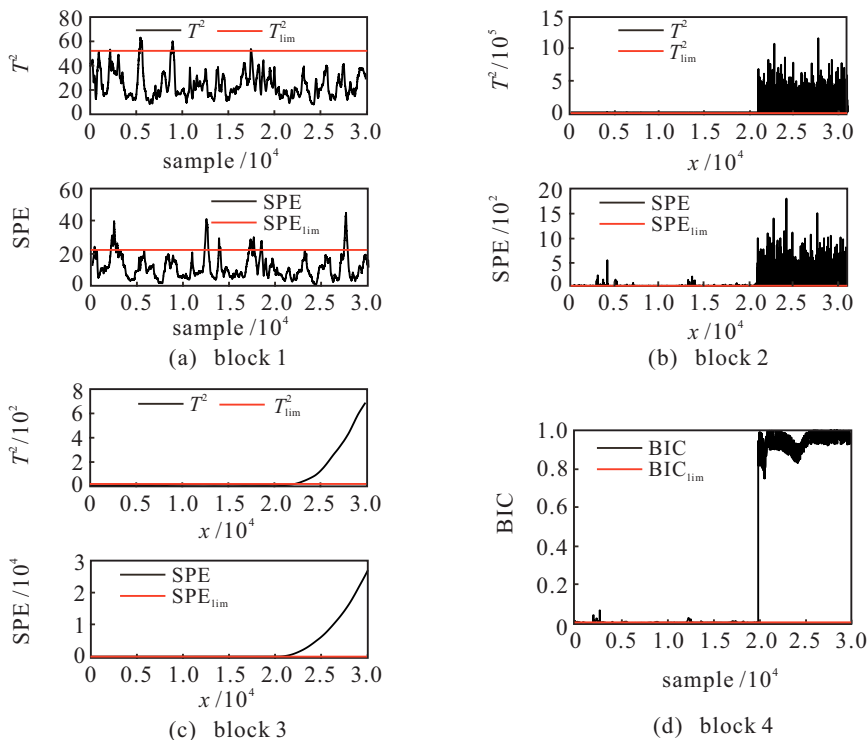


图8 各子块对模拟故障4的监测结果

图9是各子块对现存故障5的检测结果,可以看出各子块均能检测出故障,但本文的MIE-PCA提取

的信息最全面,因此最终的BIC统计量误报率仍然最低.

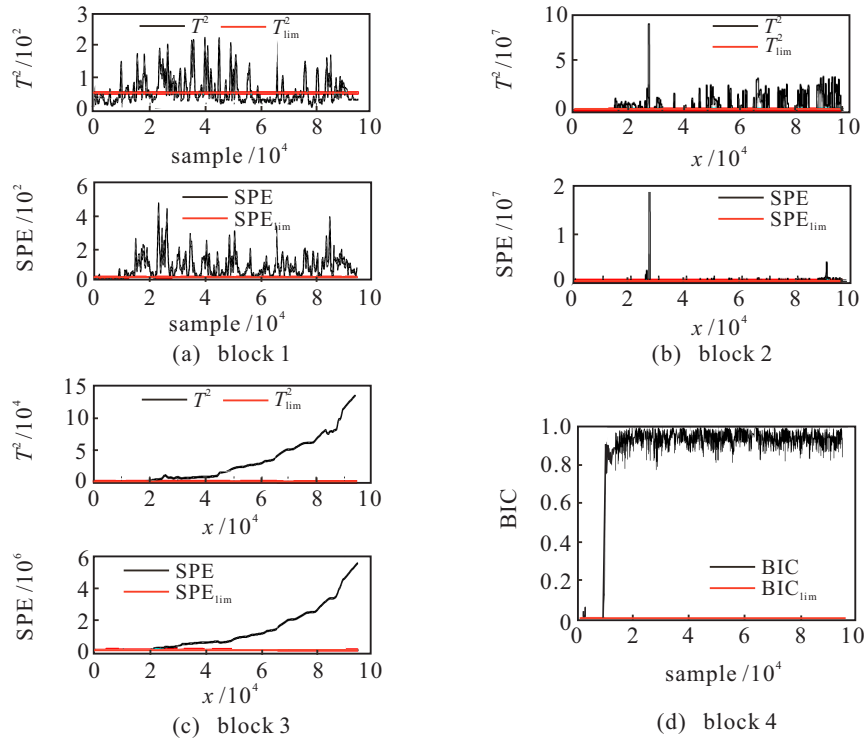


图9 各子块对现存故障5的监测结果

表4是本文提出的MIE-PCA算法与基础算法PCA、文献[27]的DPCA(dynamic principal component analysis)、文献[28]的MIMBPCA(mutual information-spectral clustering multiblock principal component analysis)、文献[15]的IJB-PCA-ICA(improved Jarque-Bera-principal component analysis-independent component analysis)和文献[16]的MBI-PCA(multi-block

information-principal component analysis)准确率对比结果,对比算法中训练集和测试集的选择以及PCA参数与先前实验一致,其余参数按照实际情况进行了调优.可以发现,本文提出的BIC算法对5种故障检测的准确率都达到90%以上,优于其他算法.总体而言,所提出的算法对于检测潜水器资源勘查系统中的各种故障类型都有非常好的效果.

表4 几种故障检测方法的准确率

fault	PCA		DPCA		MIMBPCA		MBI-PCA		IJB-PCA-ICA	MIE-PCA
	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE	BIC	BIC
1	69.8	29.7	32.3	32.9	32.5	67.8	27.4	30.1	84.1	93.3
2	99.7	30	32.2	33.5	33.4	69.1	60.1	34.2	89.3	92.9
3	69.8	30.1	31.4	33.4	32.7	67.3	63.7	31.4	90.1	93.2
4	68.8	24.3	35.2	32.6	22.5	76	53.4	33.5	80.9	92.5
5	88.1	84.1	89.6	83.9	83.1	92.6	82.7	68.9	92.2	99.3
avg	79.2	39.6	44.1	43.2	40.8	72.8	57.4	39.6	87.3	94.2

5 结论

为了提高对AUV资源勘查系统中多种故障类型的检测准确率并降低漏报率,本文提出了一种基于多块信息提取的PCA故障检测方法(MIE-PCA).首先,分析了资源勘查系统传感器在实际运行中得到的数

据特点,并对可能存在的故障进行了总结分类;然后,为了能够对各种类型故障都有较好的检测结果,针对当前数据特点,将分块建模的思想应用到潜水器故障检测领域;最后,对实验结果进行分析,发现分块提取的信息非常完备.一方面,变化率信息和变化率的

统计量累积误差信息能够反映出存在的5种故障;另一方面,相关性信息可以进一步对提取的信息进行补充,在对故障类型1的实验中可以发现,提取的相关性信息大大降低了漏报率.最后的BIC统计量将各子块的优势进行结合,因此对数据中存在的各种故障类型均有非常低的漏报率和非常高的准确率.将本文提出的方法与PCA、DPCA、MIMBPCA、MBI-PCA和IJB-PCA-ICA进行了比较,可以发现,所提出的方法对各类故障的检测准确率都远远高于对比算法,说明该方法对潜水器资源勘查系统有着更好的故障检测性能.但是,本文提出的算法模型无法随着时间更新,当用来训练的数据无法表征所有的运行状态时,检测效果不理想.因此,如何做到模型动态更新是下一步的工作重点.另外,本文并没有研究故障的定位问题,未来还需要进一步研究.

参考文献(References)

- [1] 程学龙,朱大奇,孙兵,等.深海载人潜水器推进器系统故障诊断的新型主元分析算法[J].控制理论与应用,2018,35(12):1796-1804.
(Cheng X L, Zhu D Q, Sun B, et al. Principal component analysis algorithm for fault diagnosis of thruster system in deep-sea human occupied vehicle[J]. Control Theory & Applications, 2018, 35(12): 1796-1804.)
- [2] 吴娟.自主式水下机器人推进器与传感器多故障诊断技术研究[D].哈尔滨:哈尔滨工程大学机电工程学院,2011.
(Wu J. Research on multi-fault diagnosis technology of thruster and sensor for AUV[D]. Harbin: College of Mechanical and Electrical Engineering, Harbin Engineering University, 2011.)
- [3] Raanan B Y, Bellingham J, Zhang Y W, et al. Detection of unanticipated faults for autonomous underwater vehicles using online topic models[J]. Journal of Field Robotics, 2018, 35(5): 705-716.
- [4] 张铭钧,吴娟,王玉甲.水下机器人多传感器并发故障检测方法[J].机器人,2010,32(3):298-305.
(Zhang M J, Wu J, Wang Y J. A method of multi-sensor simultaneous faults detection for autonomous underwater vehicle[J]. Robot, 2010, 32(3): 298-305.)
- [5] 张铭钧,宋炜胥,褚振忠.自主式水下机器人模糊定性建模方法研究[J].哈尔滨工程大学学报,2013,34(1):116-122.
(Zhang M J, Song W X, Chu Z Z. Research on the method of fuzzy qualitative modeling for AUV[J]. Journal of Harbin Engineering University, 2013, 34(1): 116-122.)
- [6] 王振雷,江伟,王昕.基于多块MICA-PCA的全流程过程监控方法[J].控制与决策,2018,33(2):269-274.
(Wang Z L, Jiang W, Wang X. Plant-wide process monitoring based on multiblock MICA-PCA[J]. Control and Decision, 2018, 33(2): 269-274.)
- [7] 王健,冯健,韩志艳.基于流形学习的局部保持PCA算法在故障检测中的应用[J].控制与决策,2013,28(5):683-687.
(Wang J, Feng J, Han Z Y. Locally preserving PCA method based on manifold learning and its application in fault detection[J]. Control and Decision, 2013, 28(5): 683-687.)
- [8] 王玉甲,张铭钧,郭勇.基于PCA的水下机器人故障诊断与数据重构[J].华中科技大学学报:自然科学版,2009,37(S1):135-139.
(Wang Y J, Zhang M J, Guo Y. Fault diagnosis and data restoration based on PCA for underwater vehicle[J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2009, 37(S1): 135-139.)
- [9] Navi M, Davoodi M, Meskin N. Sensor fault detection and isolation of an autonomous underwater vehicle using partial kernel PCA[C]. 2015 IEEE Conference on Prognostics and Health Management. Austin: IEEE, 2015: 1-9.
- [10] 赵帅,宋冰,侍洪波.基于加权互信息主元分析算法的质量相关故障检测[J].化工学报,2018,69(3):962-973.
(Zhao S, Song B, Shi H B. Quality-related fault detection based on weighted mutual information principal component analysis[J]. CIESC Journal, 2018, 69(3): 962-973.)
- [11] Zhang Y W, Zhou H, Qin S J, et al. Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares[J]. IEEE Transactions on Industrial Informatics, 2010, 6(1): 3-10.
- [12] Zhang Y W, Ma C. Decentralized fault diagnosis using multiblock kernel independent component analysis[J]. Chemical Engineering Research and Design, 2012, 90(5): 667-676.
- [13] Jiang Q, Yan X. Plant-wide process monitoring based on mutual information-multiblock principal component analysis[J]. ISA Transactions, 2014, 53(5): 1516-1527.
- [14] Jiang Q C, Yan X F. Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA[J]. Journal of Process Control, 2015, 32: 38-50.
- [15] 刘舒锐,彭慧,李帅,等.基于IJB-PCA-ICA算法的故障检测[J].化工学报,2018,69(12):5146-5154.
(Liu S R, Peng H, Li S, et al. Fault detection based on IJB-PCA-ICA[J]. CIESC Journal, 2018, 69(12): 5146-5154.)

- [16] 顾炳斌, 熊伟丽. 基于多块信息提取的PCA故障诊断方法[J]. 化工学报, 2019, 70(2): 736-749.
(Gu B B, Xiong W L. Fault diagnosis based on PCA method with multi-block information extraction[J]. CIESC Journal, 2019, 70(2): 736-749.)
- [17] Jiang Q C, Wang B, Yan X F. Multiblock independent component analysis integrated with hellinger distance and bayesian inference for non-gaussian plant-wide process monitoring[J]. Industrial & Engineering Chemistry Research, 2015, 54(9): 2497-2508.
- [18] Jing C, Hou J. SVM and PCA based fault classification approaches for complicated industrial process[J]. Neurocomputing, 2015, 167(1): 636-642.
- [19] Li S, Zhou X F, Shi H B, et al. Monitoring of multimode processes based on subspace decomposition[J]. Industrial & Engineering Chemistry Research, 2015, 54(15): 3855-3864.
- [20] Joe Qin S J. Statistical process monitoring: Basics and beyond[J]. Journal of Chemometrics, 2003, 17(8/9): 480-502.
- [21] Li S, Zhou X F, Pan F C, et al. Correlated and weakly correlated fault detection based on variable division and ICA[J]. Computers & Industrial Engineering, 2017, 112: 320-335.
- [22] Divya Pandove, Shivani Goel, Rinkle Rani. An intuitive general rank-based correlation coefficient[J]. Frontiers of Information Technology & Electronic Engineering, 2018, 19(6): 699-711.
- [23] 马贺贺. 基于数据驱动的复杂工业过程故障检测方法研究[D]. 上海: 华东理工大学信息科学与工程学院, 2013.
(Ma H H. Fault detection of complex industrial processes based on data-driven methods[D]. Shanghai: School of Information Science and Engineering, East China University of Science and Technology, 2013.)
- [24] Capovilla C E, Casella I R S, Costa F F, et al. DFIG-based wind turbine predictive control employing a median filter to mitigate impulsive interferences on transmitted wireless references[J]. IEEE Transactions on Industry Applications, 2019, 55(4): 4091-4099.
- [25] Zhang Y W, Li S. Modeling and monitoring between-mode transition of multimode processes[J]. IEEE Transactions on Industrial Informatics, 2013, 9(4): 2248-2255.
- [26] Tong C D, Palazoglu A, Yan X F. Improved ICA for process monitoring based on ensemble learning and bayesian inference[J]. Chemometrics and Intelligent Laboratory Systems, 2014, 135: 141-149.
- [27] Ku W F, Storer R H, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1995, 30(1): 179-196.
- [28] Jiang Q C, Yan X F. Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA[J]. Journal of Process Control, 2015, 32: 38-50.

作者简介

郭大权(1983—), 男, 副研究员, 从事MES系统的研发与企业应用, 工业大数据对企业生产可视化、数字化、智能化的研究, E-mail: guodaquan@sia.cn;

杨宗圣(1990—), 男, 硕士生, 从事数据挖掘和故障诊断与自愈控制的研究, E-mail: 1014837124@qq.com;

周晓锋(1978—), 女, 副研究员, 博士, 从事过程监测与故障诊断机器学习与大数据分析处理、智能信息处理及平台技术等研究, E-mail: zhouxf@sia.cn;

李帅(1988—), 男, 助理研究员, 博士生, 从事过程监测与故障诊断、数据挖掘与机器学习相关方法和智能信息处理及平台技术的研究, E-mail: lishuai@sia.cn;

徐春晖(1982—), 男, 副研究员, 硕士, 从事水下机器人软件控制的研究, E-mail: xchh@sia.cn.

(责任编辑: 闫妍)