



基于互信息的多块 k 近邻故障监测及诊断

郑静, 熊伟丽

引用本文:

郑静, 熊伟丽. 基于互信息的多块 k 近邻故障监测及诊断[J]. 智能系统学报, 2021, 16(4): 717–728.

ZHENG Jing, XIONG Weili. Multiblock k -nearest neighbor fault monitoring and diagnosis based on mutual information[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(4): 717–728.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202007035>

您可能感兴趣的其他文章

移动通信网络的中性集故障诊断方法研究

Research on neutral set fault diagnosis method for mobile communication networks

智能系统学报. 2020, 15(5): 864–869 <https://dx.doi.org/10.11992/tis.201906031>

反馈式 K 近邻语义迁移学习的领域命名实体识别

Domain-named entity recognition based on feedback K -nearest semantic transfer learning

智能系统学报. 2019, 14(4): 820–830 <https://dx.doi.org/10.11992/tis.201804013>

多层信息网络故障定位综述

Survey of fault localization in multilayer information networks

智能系统学报. 2019, 14(1): 44–56 <https://dx.doi.org/10.11992/tis.201804062>

多特征融合的lncRNA识别与其功能预测

LncRNA recognition by fusing multiple features and its function prediction

智能系统学报. 2018, 13(6): 928–934 <https://dx.doi.org/10.11992/tis.201806008>

基于高维 k -近邻互信息的特征选择方法

Feature selection method based on high dimensional k -nearest neighbors mutual information

智能系统学报. 2017, 12(5): 595–600 <https://dx.doi.org/10.11992/tis.201609020>

改进D-S证据理论在电动汽车锂电池故障诊断中的应用

Application of improved D-S evidence theory in fault diagnosis of lithium batteries in electric vehicles

智能系统学报. 2017, 12(4): 526–537 <https://dx.doi.org/10.11992/tis.201605001>



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202007035

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20210412.1514.013.html>

基于互信息的多块 k 近邻故障监测及诊断

郑静^{1,2}, 熊伟丽^{1,2}

(1. 江南大学 轻工过程先进控制教育部重点实验室, 江苏 无锡 214122; 2. 江南大学 物联网工程学院, 江苏 无锡 214122)

摘要: 由于传统的 k 近邻故障监测不考虑过程的局部信息, 只建立一个全局模型, 因此提出一种基于互信息的多块 k 近邻故障监测方法。首先, 考虑建模数据的非线性和非高斯等特性, 基于变量间的互信息进行子块构建; 然后, 利用 k 近邻方法对每个子块进行建模与监测, 子块中的 k 近邻模型反映了更多的过程局部特征; 最后, 将所有子块的监测结果通过贝叶斯推断方法进行融合, 并采用基于马氏距离的故障诊断方法辨识故障源。通过对田纳西-伊斯曼过程和高炉炼铁过程中的应用仿真, 监测结果表明所提方法的可行性和有效性。

关键词: 互信息; 多块建模; k 近邻; 过程监控; 故障检测; 贝叶斯推断; 故障诊断; 马氏距离

中图分类号: TP277 **文献标志码:** A **文章编号:** 1673-4785(2021)04-0717-12

中文引用格式: 郑静, 熊伟丽. 基于互信息的多块 k 近邻故障监测及诊断 [J]. 智能系统学报, 2021, 16(4): 717-728.

英文引用格式: ZHENG Jing, XIONG Weili. Multiblock k -nearest neighbor fault monitoring and diagnosis based on mutual information[J]. CAAI transactions on intelligent systems, 2021, 16(4): 717-728.

Multiblock k -nearest neighbor fault monitoring and diagnosis based on mutual information

ZHENG Jing^{1,2}, XIONG Weili^{1,2}

(1. China Key Laboratory of Advanced Process Control for Light Industry Ministry of Education, Jiangnan University, Wuxi 214122, China; 2. School of the Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

Abstract: The traditional k -nearest neighbor (kNN) fault monitoring does not take into account the process of local information and only builds a global model. Thus, a multi-block kNN fault monitoring algorithm based on mutual information is proposed. First, with the nonlinear and non-Gaussian characteristics of the modeled data taken into consideration, subblocks are constructed based on mutual information between variables. Then, the kNN algorithm is used to model and monitor each subblock, in which the kNN model reflects more local characteristics of the process. Lastly, the monitoring results of all subblocks are fused by the Bayesian inference method, and a fault diagnosis method based on Mahalanobis distance is used to identify the source of faults. Through the application simulation in the Tennessee Eastman process and the blast furnace ironmaking process, the monitoring results show the feasibility and effectiveness of the proposed method.

Keywords: mutual information; multi-block modeling; k -nearest neighbor; process monitoring; fault detection; Bayesian inference; fault diagnosis; Mahalanobis distance

随着新型传感器、数据采集设备和系统的迅速发展, 一些先进工业过程积累了丰富的过程数据, 使得多元统计过程监控 (multivariate statistical process monitoring, MSPM) 技术不断进步^[1-3]。其中的主成分分析 (principal component analysis,

PCA) 和 k 近邻方法 (k -nearest neighbor, kNN) 是较为为基础的方法, 得到了大量的研究和应用^[4-7]。

作为一种降维技术, PCA 通过将数据投影到低维空间以有效地处理高维和线性相关的数据, 通过建立主元子空间和残差子空间的统计量进行过程监控。但是, 对于具有非线性和非高斯特性的过程数据, PCA 方法可能无法进行有效的监测。He 等^[8] 提出基于 k 近邻规则的故障监测算

收稿日期: 2020-07-22. 网络出版日期: 2021-04-12.

基金项目: 国家自然科学基金项目 (61773182); 国家重点研发计划子课题 (2018YFC1603705-03).

通信作者: 熊伟丽. E-mail: greenpre@163.com.

法,该算法不局限于线性和高斯数据,使用局部近邻距离度量样本相似度,根据故障样本和正常样本的相似度不同实现故障监测。但是由于每一个样本都需要计算与其他样本的距离,计算量明显增大。为此,学者们提出了许多改进的 kNN 故障监测算法。例如:文献[9]利用改进 K-means 聚类将原始建模数据分成多个类,对每个类分别建立 kNN 监测模型,大大缩短故障检测时间;文献[10]提出将动态 PCA 和 kNN 相结合的故障诊断方法,先建立主元模型,再利用 kNN 获取样本的 k 个近邻,明显提高了故障的报警率;文献[11]针对 kNN 模型不能及时更新的问题,提出了一种特征空间自适应 k 近邻故障检测方法,有效提高模型实时监测的能力;文献[12]考虑到多模态过程数据具有多中心、方差差异大等特点,通过构造标准距离,实现了 kNN 方法对多模态数据的有效监测。

由于现代工业过程具有多个操作单元、变量关系复杂等特点,全局建模策略无法更加准确地对过程建模,多块建模策略成为有效的解决方案。Macgrego 等^[13]首次提出了多块投影方法,为每个子块以及整个过程建立监测模型。文献[14]采用 Jarque-Bera(J-B)检测方法并利用变量间的 Hellinger 距离获得高斯和非高斯子块,然后分别采用不同的方法进行建模,并对每个子块的统计量进行加权得到总的联合指标实现在线监控;文献[15]将整个过程划分成多个子块单元,然后在每个子块单元内分别进行相对变换独立主元分析处理,实现故障排查和识别;Ge 等^[16]提出分布式 PCA 的全流程过程监控方法,利用过程变量在主元方向上的贡献度划分子块,有效地提高了监控效果。

在信息论领域里,互信息 (mutual information, MI) 是一种相对成熟的统计分析技术,可以通过信息熵度量两个随机变量之间的依赖性,并且这种度量不局限于数据线性关系的假设条件^[17-19],已经在数据分析与建模领域得到了比较多的应用。文献[20]利用变量间的互信息定义数据的相关性矩阵,为过程数据建立更为精确的描述模型。文献[21]利用互信息矩阵之和替代传统主成分分析中的协方差矩阵,计算其特征向量与特征值,得到较主成分分析更好的降维效果。文献[22]利用高维 k 近邻互信息方法,有效解决建模过程中的特征选择问题。

综上所述,为了更加充分地对复杂过程变量之间的关系进行描述,并提取过程的局部特征,利用多块建模策略,提出一种基于互信息的 k 近邻故障监测算法。该算法在计算训练集样本间的互信息基础上,根据互信息值的大小将变量分成

多个子块,对每个子块建立相应的 kNN 模型,并利用核密度估计方法求出控制限,最后利用贝叶斯推断将各子块的监测结果融合,使得整体的监测效果更为直观。本文进一步采用基于马氏距离的故障诊断方法,通过计算样本中各变量与其均值的马氏距离,找出引发故障的源变量并对其隔离。利用田纳西-斯曼 (Tennessee Eastmann, TE) 和实际高炉炼铁过程数据,对所提方法进行了仿真,并与几种传统监测方法进行了对比,验证了本文方法的性能。

1 相关算法介绍

1.1 kNN 算法

k 近邻算法是数据挖掘和数据分类中最常用的方法之一,传统 kNN 算法通过寻找 k 个近邻样本,采用投票的方法确定待测样本的类别。基于 kNN 的故障监测,其基本思想是通过计算近邻距离度量样本间的相似度,若样本点与训练集中前 k 个近邻样本距离的平方和大于正常样本的相应距离平方和,则该样本点被定义为故障点。监测过程包括模型建立和故障检测两步,具体描述如下:

1) 建立模型

首先在训练集中,寻找每个样本 x_i 的前 k 个近邻样本,记做 $T(x_i, k) = \{x_i^1, x_i^2, \dots, x_i^j, \dots, x_i^k\}$, 其中, x_i^j 表示样本 x_i 的第 j 个近邻样本。然后,计算每个样本 x_i 与其 k 个近邻样本的欧式距离平方和作为统计量,如式 (1) 所示,其中, d_{ij}^2 表示样本 x_i 与它的第 j 个近邻样本的欧氏距离平方。接着,根据置信度 α 确定训练模型的控制限 D_α^2 。

$$D_i^2 = \sum_{j=1}^k d_{ij}^2, d_{ij}^2 = \|x_i - x_i^j\|^2 \quad (1)$$

式中: D_i^2 表示样本 x_i 的统计量。

2) 在线故障检测

首先,在训练集中寻找待测样本 x 的前 k 个近邻。然后,计算 x 与其 k 个近邻样本的欧式距离平方和,记做 D_x^2 。最后,比较 D_x^2 与 D_α^2 的大小,若 $D_x^2 \geq D_\alpha^2$, 则判定为故障点,反之为正常点。

1.2 互信息

在概率论和信息论领域,互信息是一种非常实用的信息度量方法。它可以度量两个随机变量相互依赖的程度,表示出两个变量共享的信息,反映两个变量的相关性^[21],这种度量同样适用于非线性相关的变量。对于密切相关的变量,它们拥有较大的互信息。令随机变量 X 和 Y 的联合概率分布及边缘概率分布分别为 $p(x, y)$ 、 $p(x)$ 和 $p(y)$, 其中 $x \in X, y \in Y$, X 的熵定义如式 (2) 所示。

$$H(X) = - \sum_{x \in X} p(x) \log p(x) \quad (2)$$

联合熵为

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y) \quad (3)$$

则变量 X 和 Y 之间的互信息可以定义为

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (4)$$

若 X 和 Y 相互独立, 则 X 不对 Y 提供任何信息, 此时互信息值最小, 结果为 0。反之, 两个变量间的相关性越高, 互信息值越大。

2 基于互信息的多块建模 kNN 故障监测及诊断

2.1 基于互信息的分块策略

在实际的工业过程中, 变量之间大多是线

性、非线性共存, 高斯、非高斯混合分布, 传统的 PCA 与 kNN 方法往往从全局的角度出发, 系统的本质特征无法得到充分的展示。因此, 首先对变量进行 MI 计算, 将互信息大的多个变量放在一起组成子块, 使得子块内的变量拥有更多相同的信息, 最大化地反映变量的一个或者多个局部特征, 同时也大大降低了监控过程的复杂度, 从而有效地提高系统的监控效果。

对于训练集 $X \in R^{n \times m}$, $x_i \in X$, $x_j \in X$, 计算变量 x_i 与变量 x_j 之间的互信息 I_{ij} , 即

$$I_{ij} = I(x_i, x_j) (i = 1, 2, \dots, m; j = 1, 2, \dots, m) \quad (5)$$

若 $I_{ij} \geq I_{il}$, 则把变量 x_j 与变量 x_i 放到相同的子块中。 I_{il} 一般根据经验获得, 本文结合互信息针状图为了更好地划分变量, I_{il} 取 $1.3I_{im}$, 其中 I_{im} 是 I_{ij} 的中值。本文所采用的多块建模方法如图 1 所示。

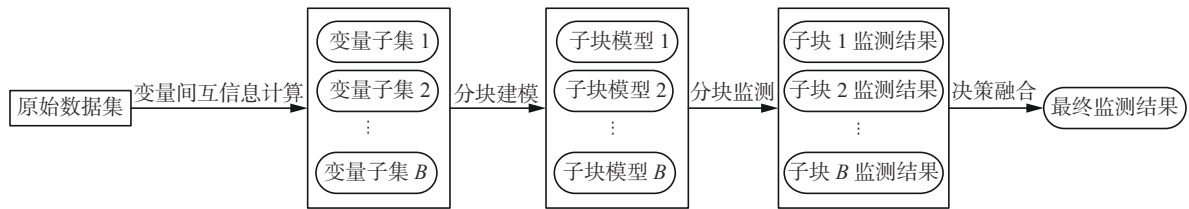


图 1 本文所采用的多块建模方法

Fig. 1 Multi-block modeling method in this paper

2.2 故障监测及诊断

针对划分好的子块, 建立 kNN 监测模型。寻找各子块中样本的 k 近邻样本集, 记做

$$T(x_{ib}, k) = \{x_{ib}^1, x_{ib}^2, \dots, x_{ib}^k\} (b = 1, 2, \dots, B) \quad (6)$$

其中 x_{ib} 表示第 b 个子块中的变量 x_i , x_{ib}^k 表示样本 x_{ib} 的第 k 个近邻样本。

然后计算每个子块中样本与其 k 近邻样本的欧式距离平方和作为子块统计量, 即 $D_{ib}^2 = \sum_{j=1}^k d_{ibj}^2$, 其中 D_{ib}^2 表示第 b 个子块的统计量。

通过核密度估计法 (kernel density estimation, KDE) 估计每个统计量的概率密度, 再绘制累加概率密度图, 根据置信度确定每个块中统计量的控制限。由于子块数目较多且产生多个监测结果, 无法得到一个直观的最终决策, 因此, 采用贝叶斯融合策略^[23], 将正常事件和故障事件与贝叶斯推断相结合, 从概率的角度将所有子块的统计量组合成一个新的统计量来得到最终的监测结果。

在贝叶斯推断 (bayesian inference, BI) 中, 测试样本 x_{test} 在第 b 个子块中的 D^2 统计量的故障条件概率可以表示为

$$P_{D^2}(\text{F}|x_{\text{test}}, b) = \frac{P_{D^2}(x_{\text{test}}, b|\text{F})P_{D^2}(\text{F})}{P_{D^2}(x_{\text{test}}, b)} \quad (7)$$

$$P_{D^2}(x_{\text{test}}, b) = P_{D^2}(x_{\text{test}}, b|\text{N})P_{D^2}(\text{N}) + P_{D^2}(x_{\text{test}}, b|\text{F})P_{D^2}(\text{F}) \quad (8)$$

式中: $x_{\text{test}, b}$ 表示第 b 个子块中的测试样本。条件概率 $P_{D^2}(x_{\text{test}, b}|\text{N})$ 和 $P_{D^2}(x_{\text{test}, b}|\text{F})$ 可定义如式 (9) 所示。

$$P_{D^2}(x_{\text{test}, b}|\text{N}) = e^{-D_{b, \text{new}}^2 / D_{b, \text{lim}}^2} \quad (9)$$

$$P_{D^2}(x_{\text{test}, b}|\text{F}) = e^{-D_{b, \text{lim}}^2 / D_{b, \text{new}}^2}$$

其中, “N”和“F”分别代表“normal”和“fault”, 即正常样本和故障样本。 $P_{D^2}(\text{N})$ 为正常样本的先验概率, 由置信度 α 决定, 则 $P_{D^2}(\text{F})$ 为 $1-\alpha$; $D_{b, \text{new}}^2$ 为新样本在第 b 个子块中的统计量; $D_{b, \text{lim}}^2$ 是第 b 个子块中由核密度估计方法估计得出的统计量的控制限。最终, 融合的 BIC 统计量即为待测样本发生故障的概率, 如式 (10) 所示。

$$\text{BIC}_{D^2} = \sum_{b=1}^B \left\{ \frac{P_{D^2}(x_{\text{test}, b}|\text{F})P_{D^2}(\text{F}|x_{\text{test}, b})}{\sum_{j=1}^B P_{D^2}(x_{\text{test}, j}|\text{F})} \right\} \quad (10)$$

BIC 统计量的控制限由 $1-\alpha$ 确定 (本文 α 取 0.01), 当 BIC 的值大于 $1-\alpha$ 时, 判断发生故障; 否则, 过程正常。

当检测到故障后, 需要找出引发故障的源变量并对其进行分离。计算数据样本中各变量与其

均值的马氏距离^[17],即加权计算数据样本中各变量相较于其均值的偏移量,偏移量越大,说明该变量对于故障贡献越大。该方法可以有效辨识引发故障的源变量,即发生故障的根本原因。

2.3 基于 MI-MBkNN 故障监测算法流程

基于互信息的多块 kNN 故障监测算法流程如图2所示,具体步骤描述如下。

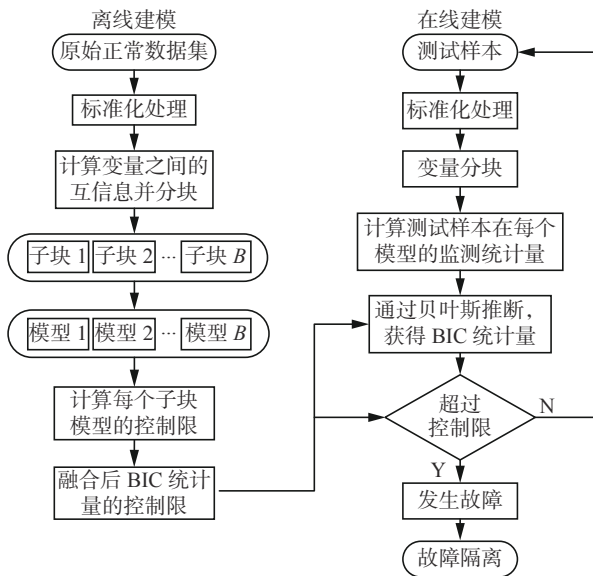


图2 基于 MI-MBkNN 的故障监测流程

Fig. 2 Fault monitoring flowchart based on MI-MBkNN

- 1) 获取正常训练数据, 并对其进行标准化处理;
- 2) 计算两两变量间互信息, 根据 2.1 节所述方法对变量进行分块, 得到各个子块;
- 3) 对每个子块分别建立 kNN 模型, 利用核密度估计方法确定各自的故障控制限;
- 4) 对于新来的测试样本, 同样对其进行标准化处理和分块处理;
- 5) 对每个子块进行 kNN 故障监测, 获得每个子块的监测结果;
- 6) 通过贝叶斯推断方法, 利用式 (10) 将各个子块的统计量组合成为一个新的 BIC 统计量, 并根据置信度确定控制限, 当 BIC 超过控制限时则判断发生了故障, 否则正常;
- 7) 监测到故障后计算数据样本中各变量与其均值的马氏距离, 确定故障变量及故障块, 分离出对故障影响最大的变量。

3 仿真实验

3.1 TE 过程仿真

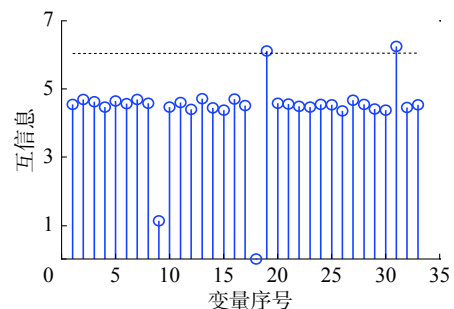
TE 仿真平台是基于实际工业过程的仿真平台, 它由反应器、冷凝器、压缩机、分离器和汽提塔 5 个主要操作单元组成^[24-25]。该过程包含的

变量数目多, 且变量与变量之间的关系复杂。主要分为 12 个操作变量, 41 个测量变量, 21 个预设定的故障。本文选取 22 个过程测量变量和 11 个操作变量 (不包括搅动速度) 用于监测方法建模和监测性能测试, 具体变量描述见文献 [16]。对于每种故障, 训练集用于训练建立模型, 测试集用来检验模型监测性能。训练集和测试集均采用 960 个样本, 测试集中故障从第 161 个样本点引入。

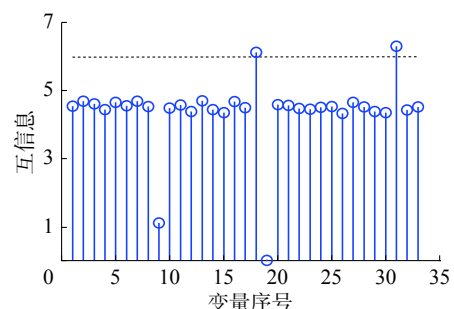
为了建立多块模型, 对选取的过程变量和操作变量进行互信息的计算并进行分块, 分块结果如表 1 所示。图 3 分别展示了变量 18、变量 19、变量 31 与其他 32 个变量间的互信息, 图中的虚线表示为 1.3 倍互信息中值, 互信息超过虚线的变量即为与该变量具有较大互信息的变量。因此将变量 18、变量 19 和变量 31 放到相同的子块中。图 4 分别展示了变量 10、变量 17、变量 28、变量 33 与其他 32 个变量间的互信息, 因此将它们组成一个子块。

表 1 TE 过程子块的划分
Table 1 Blocks division in the TE process

子块编号	变量
1	x_1, x_{25}
2	x_{12}, x_{29}
3	x_{15}, x_{30}
4	x_{18}, x_{19}, x_{31}
5	$x_{10}, x_{17}, x_{28}, x_{33}$
6	$x_7, x_{13}, x_{16}, x_{20}, x_{27}$
7	$x_2, x_3, x_4, x_5, x_6, x_8, x_9, x_{11},$ $x_{14}, x_{21}, x_{22}, x_{23}, x_{24}, x_{26}, x_{32}$



(a) 变量 18



(b) 变量 19

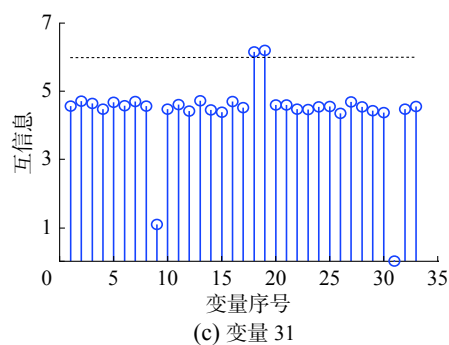


图3 子块4中各变量间的互信息

Fig. 3 Mutual information between variables in block 4

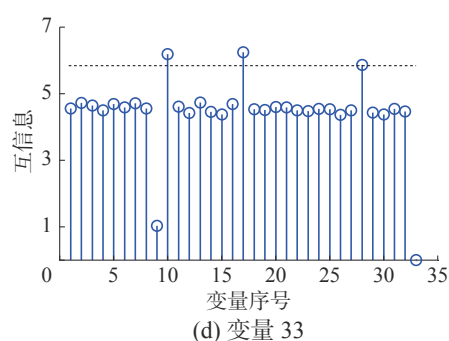
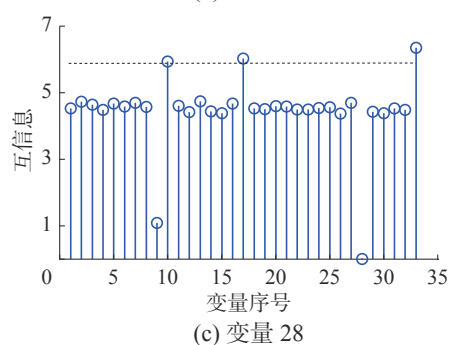
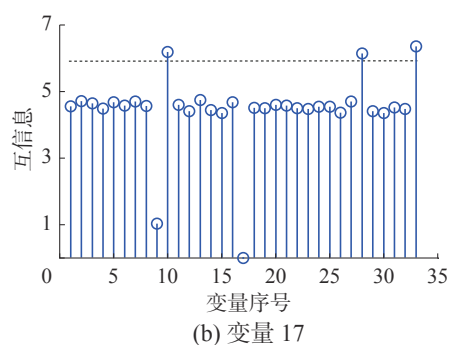
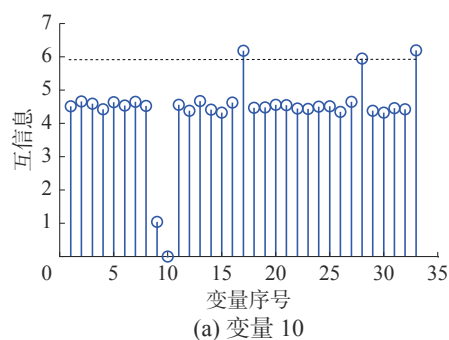


图4 子块5中各变量间的互信息

Fig. 4 Mutual information between variables in block 5

表2给出了7个子块对21种故障的报警率、平均报警率和平均误报率。从报警率来看,对于大多数故障类型,子块7的监测结果要优于其他6个子块。子块5的平均报警率很低,但是对于某些故障(如故障5),子块5的报警率达96%,对整个的监测起到了关键的作用。对于不同的故障,由于某些子块拥有较高的报警率和较低的误报率,使得最终融合的BIC统计量表现了良好的监测性能。从对21种故障的监测结果来看,对于大部分故障,融合后的监测性能有了明显的提高。

表2 TE过程各故障报警率
Table 2 Alarm rates of TE process

故障 编号	子块							BIC
	1	2	3	4	5	6	7	
1	0.989	0	0	0.970	0.103	0.373	0.98	0.998
2	0.054	0	0	0.898	0.986	0.729	0.951	0.986
3	0.003	0	0.001	0.003	0.008	0.029	0.011	0.033
4	0.003	0	0	0	0.008	0.019	1	0.994
5	0.090	0	0	0.196	0.960	0.230	0.194	0.951
6	1	0	0	0.966	0.944	0.996	0.990	1
7	0.208	0	0	0.311	0.153	0.415	1	1
8	0.580	0	0.003	0.720	0.721	0.936	0.839	0.976
9	0.003	0	0	0.001	0.009	0.023	0.013	0.021
10	0.079	0	0	0.715	0.004	0.348	0.141	0.758
11	0.005	0	0.001	0.020	0.011	0.046	0.735	0.660
12	0.446	0	0.004	0.879	0.420	0.916	0.968	0.991
13	0.448	0	0	0.933	0.574	0.875	0.890	0.949
14	0.005	0	0.001	0	0.003	0.004	1	1
15	0.013	0	0	0.034	0.003	0.074	0.010	0.089
16	0.019	0	0	0.801	0.005	0.180	0.070	0.804
17	0.010	0	0	0.073	0.011	0.103	0.943	0.900
18	0.828	0	0	0.881	0.859	0.880	0.896	0.899
19	0.008	0	0.001	0.011	0.004	0.541	0.105	0.468
20	0.018	0	0.003	0.050	0.008	0.651	0.114	0.630
21	0.003	0	0	0.490	0.003	0.376	0.334	0.449
平均故障报警率	0.229	0	0.001	0.426	0.276	0.416	0.580	0.741
平均故障误报率	0.002	0.001	0	0.002	0.005	0.017	0.003	0.017

表3给出了TE过程21种故障在不同监测方法下的报警率和误报率,主要方法包括传统PCA、SVDD、kNN和本文提出的MI-MBkNN。从仿真结果可以看出,对于绝大多数故障类型,MI-MBkNN能取得优越于其他3种方法的监测结果,尤其是对故障5、故障10、故障16、故障19的监测。图5以故障5为例展示了详细的监测过程与结果。

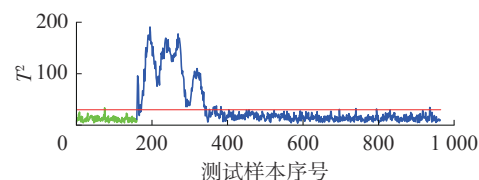
表3 几种现有的监测方法性能比较

Table 3 Comparison of some state of monitoring methods

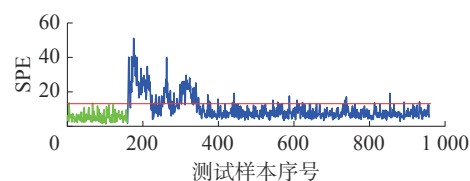
故障编码	报警率			
	PCA	SVDD	kNN	MI-MBkNN
1	0.999	0.993	0.996	0.998
2	0.984	0.983	0.983	0.986
3	0.026	0.036	0.013	0.033
4	1	0.791	0.975	0.994
5	0.243	0.275	0.260	0.951
6	1	1	1	1
7	1	1	1	1
8	0.969	0.975	0.976	0.976
9	0.018	0.03	0.020	0.021
10	0.299	0.449	0.418	0.758
11	0.749	0.599	0.683	0.660
12	0.984	0.986	0.989	0.991
13	0.953	0.944	0.946	0.949
14	1	1	1	1
15	0.030	0.061	0.029	0.089
16	0.274	0.284	0.289	0.804
17	0.954	0.878	0.919	0.900
18	0.901	0.898	0.896	0.899
19	0.125	0.046	0.099	0.468
20	0.498	0.459	0.495	0.630
21	0.473	0.419	0.425	0.449
平均报警率	0.642	0.624	0.639	0.741
平均误报率	0.004	1.756	0.006	0.017

TE过程中的故障5涉及冷凝器冷却水入口温度的变化,当故障发生时,从冷凝器到气/液分离器的出口流量增加,使温度升高。使用传统PCA、kNN、SVDD方法和本文提出的MI-MBkNN的监测结果如图5所示。从图5(a)~(d)可以发现,在故障开始时就可以检测出故障,但是在大约350个样本的时候,统计量出现低于控制限

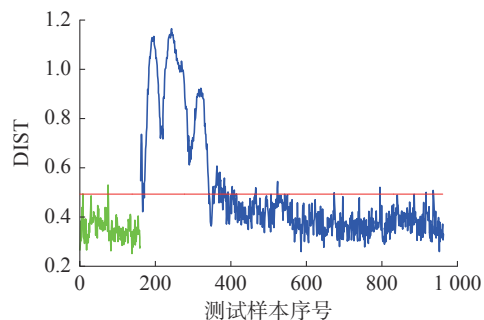
的情况,导致故障的漏报。由于该故障是局部故障,因此很难在全局模型中检测到,为了更好地找出故障的原因,图6给出了数据样本在第161个样本点(故障最开始处)的各变量与其均值中心的马氏距离。



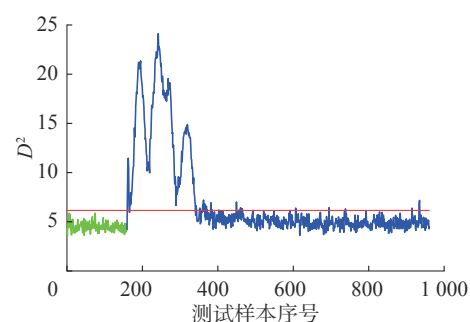
(a) PCA- T^2



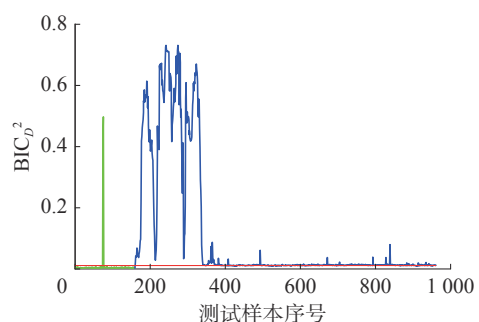
(b) PCA-SPE



(c) SVDD



(d) kNN



(e) MI-MBkNN

图5 故障5监测结果

Fig. 5 Monitoring result of fault 5

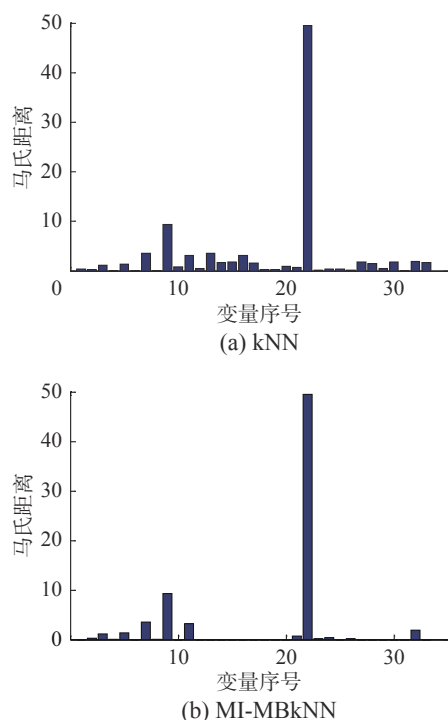


图6 第161样本点故障5的变量识别结果

Fig. 6 Variables identification results of fault 5 on the 161th point

可以看出这两个模型都能正确识别变量在过程中的变化,如分离器冷却水出口温度的变化(变量22),反应器温度(变量9),产品分离器温度(变量11)和反应器冷却水流量(变量32)。但是在350个样本点后,从图7(第400个样本点)可以看出,kNN无法识别出冷凝器冷却水流量的变化(变量33),但是MI-MBkNN模型可以成功识别,因此MI-MBkNN对故障5表现出了优越的监测效果。

故障10是流2(C进料)中温度的随机变化,从图8中可以看出350~650样本,传统的监测方法很难监测到故障,但是MI-MBkNN却能很好地检测出来。

为了更好地找出引发故障的源变量,图9(a)和(b)分别给出了使用kNN和MI-MBkNN方法时数据样本在第400样本点处的各变量与其均值的马氏距离,可以发现MI-MBkNN在寻找故障源变量方面提供更重要的指导,即汽提塔温度(变量18),汽提塔蒸汽流量(变量19),汽提塔蒸汽阀的变量(变量31)是引起故障10的原因,因此子块4的监测效果明显优于其他子块。通过贝叶斯融合后,使得整体的监测效果得到了很高的提升。因此本文提出的方法对故障10的监测效果优于其他几种传统的监测方法。图10对比了PCA、kNN、本文方法的子块4和MI-MBkNN的

对故障16的监测结果,传统kNN方法在统计量上只能从450~500样本和790~840样本之间做到相对持续的报警,而本文监测方法从故障引入点处开始就能做到大范围的持续报警。

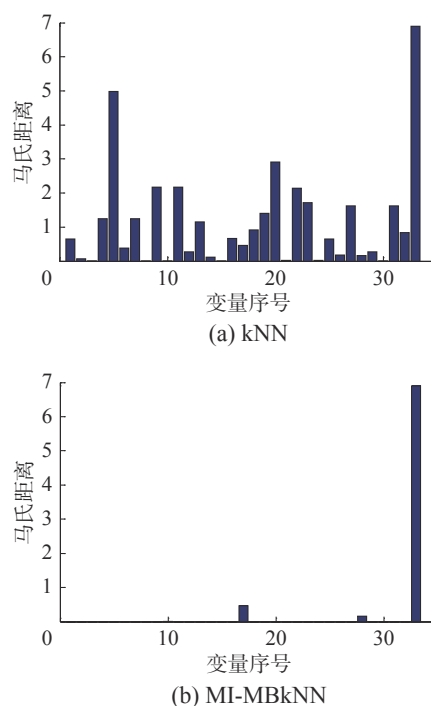
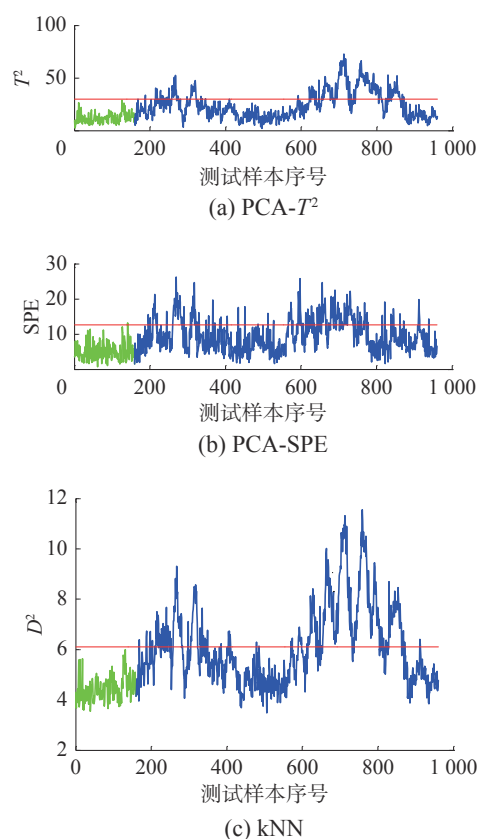
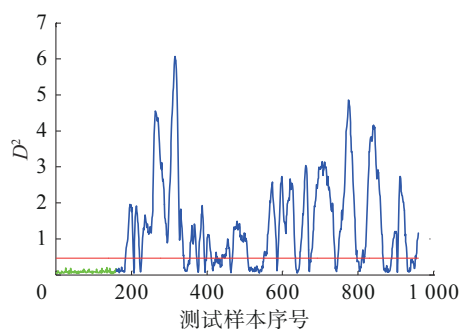


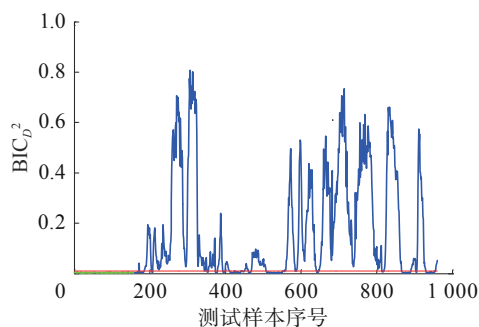
图7 第400样本点故障5的变量识别结果

Fig. 7 Variables identification results of fault 5 on the 400th point





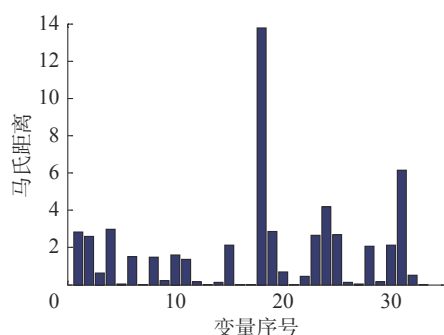
(d) 子块 4



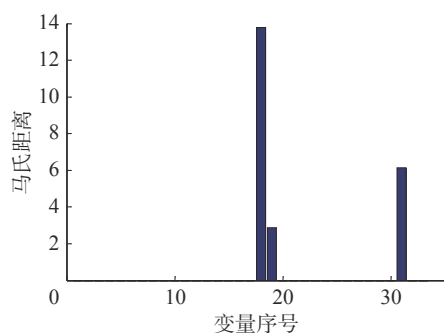
(e) MI-MBkNN

图 8 故障 10 监测结果

Fig. 8 Monitoring result of fault 10



(a) kNN



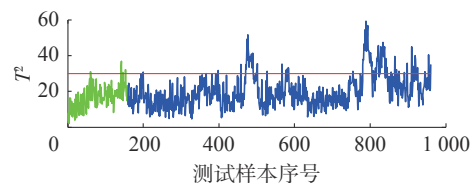
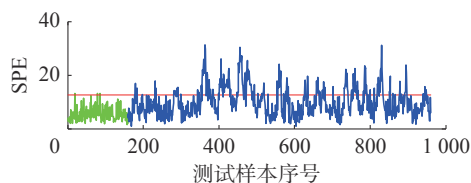
(b) MI-MBkNN

图 9 变量识别结果

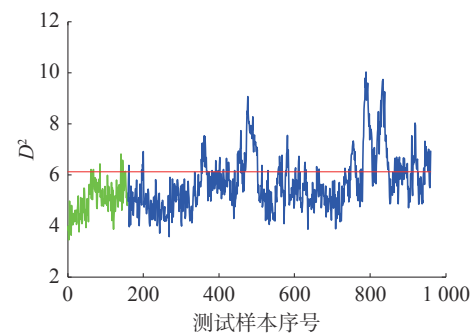
Fig. 9 Variables identification results

图 11 给出了对故障 16 的诊断结果,可以看出,相比于 kNN,本文方法对故障提供更清晰的识别结果,即引起该故障的源变量是变量 x_{18} 、 x_{19}

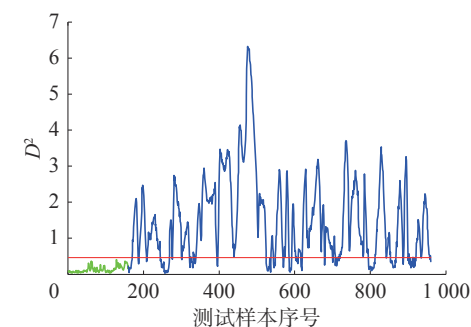
和 x_{31} 。因此子块 4 对故障 10 的敏感程度远远大于其他子块,其表现了良好的监测性能。通过贝叶斯推断融合后提升了整体的监测效果,再次验证了本文所提方法的有效性。

(a) PCA- T^2 

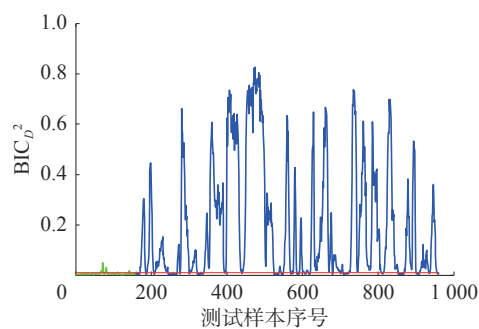
(b) PCA-SPE



(c) kNN



(d) 子块 4



(e) MI-MBkNN

图 10 故障 16 监测结果

Fig. 10 Monitoring result of fault 16

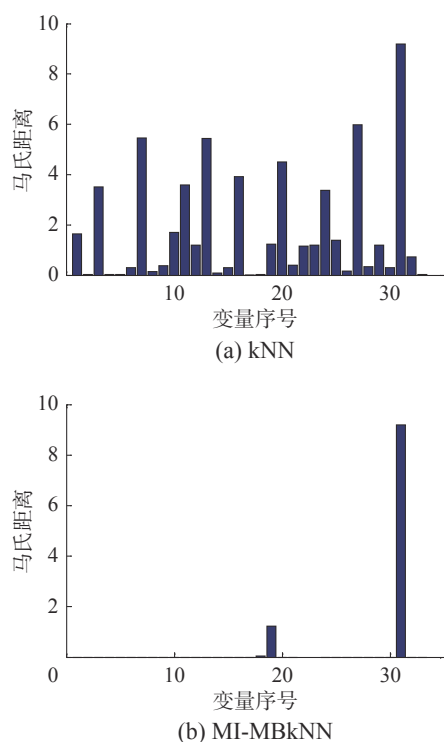


图 11 变量识别结果

Fig. 11 Variables identification results

3.2 高炉炼铁实际过程应用

为了达到高炉炼铁过程节能降耗的目的,必须保证铁水的生产质量和产量。当气体流动不稳定时会影响碳的燃烧,最终导致炉腹架空,产生悬挂故障。若没有及时检测出悬挂故障,将会导致热应力和内部的气体压力过大,使得顶部结构受到严重的破损。本节考虑了实际情况中悬挂故障的存在,采集正常工况下的 8 个过程变量的 2 000 个样本作为训练样本,同时采集了悬挂故障下的 1 900 个样本作为测试样本。在悬挂故障下,炉内的温度和压力增加,炉顶的一氧化碳和二氧化碳浓度上升,氢气的浓度下降。为了更好地表现变量的特性,表 4 给出了 8 个过程变量的描述,图 12 给出了 8 个变量的变化曲线图,其中前 2 000 个样本为正常样本,后 1 900 个样本为故障样本。

表 4 悬挂故障监测中选择的 process 变量

Table 4 Process variables selected for monitoring of the hanging fault

过程变量	变量描述	过程变量	变量描述
u_1	风量	u_5	富氧量
u_2	风温	u_6	炉顶煤气 CO_2 含量
u_3	风压	u_7	炉顶煤气 CO 含量
u_4	炉顶压力	u_8	炉顶煤气 H_2 含量

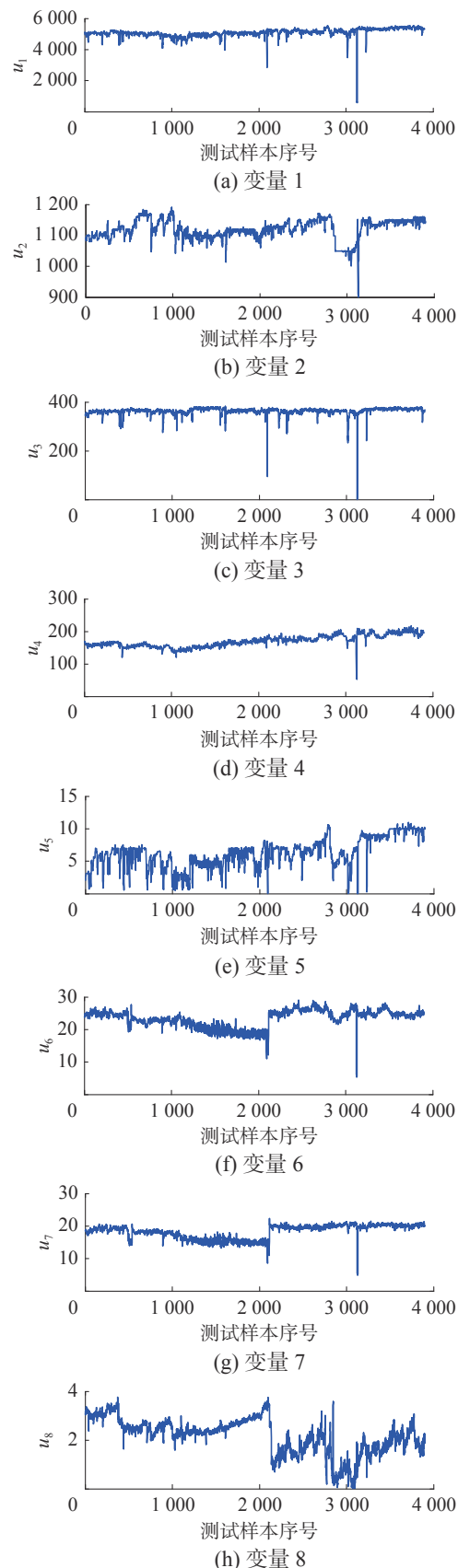


图 12 高炉过程各变量曲线图

Fig. 12 Curves of Blast furnace process variable

利用上述所提分块方法将 8 个变量分成两个子块,子块 1 为 u_1 、 u_3 ,子块 2 为 u_2 、 u_4 、 u_5 、 u_6 、 u_7 、

u_8 , 图13给出了子块1中各变量间的互信息。表5给出了不同监测方法的监测结果, 图14展示了子块1和子块2的监测结果。可以看出子块2在2200样本点后可以达到持续报警, 其监测效果明显好于子块1, 结合图15给出的故障诊断结果, 可以看出变量4和变量8是引起故障的主要原因, 由于本文所提方法对变量进行了合理分块, 把结构相似且对故障最为敏感的变量放在同一个子块中, 使得整体的监测性能得到了提升, 再次验证本文所提方法的有效性和优越性。

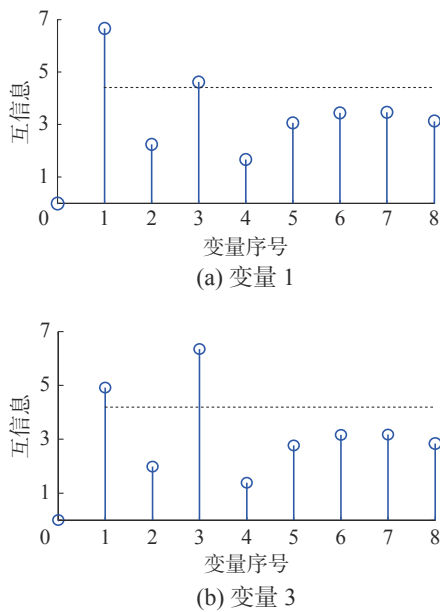


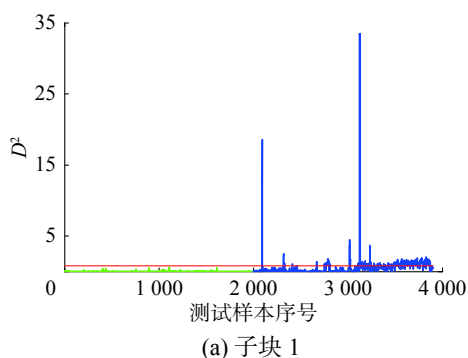
图13 高炉过程子块1变量间的互信息

Fig. 13 Mutual information between variables in blast furnace process block 1

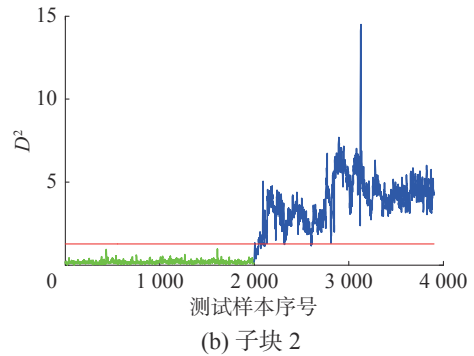
表5 不同方法的监测性能比较

Table 5 Comparison of monitoring performance of two methods in blast furnace process

性能	PCA	kNN	MI-MBkNN
报警率	0.941	0.935	0.970
误报率	0.018	0.001	0



(a) 子块1



(b) 子块2

图14 高炉过程子块1和子块2的监测性能比较

Fig. 14 Comparison of monitoring performance of block1 and block2 in blast furnace process

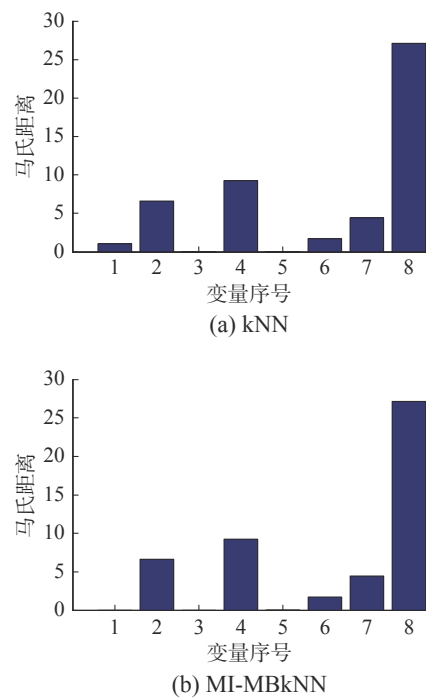


图15 高炉过程变量识别结果

Fig. 15 Variables identification results of the blast furnace process

4 结束语

本文提出了一种基于互信息的多块 k 近邻故障监测方法, 使用互信息对过程变量进行划分, 并在每个子块中建立基于 kNN 的故障监测模型, 所提方法反映了过程的更多局部特征, 所以更易于故障的监测和诊断。将所提方法应用于TE过程和实际高炉炼铁过程中, 均取得了比较好的监测效果。本文所提方法是一种完全基于数据驱动的监测方法, 可以考虑将其与先验知识或专家经验相结合, 并且可以考虑变量间关系的更多细节, 提出其他分块方法, 是本文进一步研究的工作之一。

参考文献:

- [1] CHIANG L H, RUSSELL E L, BRAATZ R D. Fault detection and diagnosis in industrial systems[J]. *Measurement science and technology*, 2001, 12(10): 1745.
- [2] GE Zhiqiang, SONG Zhihuan, GAO Furong. Review of recent research on data-based process monitoring[J]. *Industrial & engineering chemistry research*, 2013, 52(10): 3534–3562.
- [3] QIN S J. Statistical process monitoring: basics and beyond[J]. *Journal of chemometrics*, 2003, 17(8/9): 480–502.
- [4] JIANG Qingchao, YAN Xuefeng, ZHAO Weixiang. Fault detection and diagnosis in chemical processes using sensitive principal component analysis[J]. *Industrial & engineering chemistry research*, 2013, 52(4): 1635–1644.
- [5] KRESTA J V, MACGREGOR J F, MARLIN T E. Multivariate statistical monitoring of process operating performance[J]. *The Canadian journal of chemical engineering*, 1991, 69(1): 35–47.
- [6] NOMIKOS P, MACGREGOR J F. Multivariate SPC charts for monitoring batch processes[J]. *Technometrics*, 1995, 37(1): 41–59.
- [7] HE Q P, WANG Jin. Principal component based k -nearest-neighbor rule for semiconductor process fault detection[C]//*Proceedings of 2018 American Control Conference*. Seattle, USA, 2008: 1606–1611.
- [8] HE Q P, WANG Jin. Fault detection using the k -Nearest neighbor rule for semiconductor manufacturing processes[J]. *IEEE transactions on semiconductor manufacturing*, 2007, 20(4): 345–354.
- [9] 陈海彬, 郭金玉, 谢彦红. 基于改进 K-means 聚类的 kNN 故障检测研究 [J]. *沈阳化工大学学报*, 2013, 27(1): 69–73.
- CHEN Haibin, GUO Jinyu, XIE Yanhong. kNN fault detection based on improved K-means clustering[J]. *Journal of Shenyang University of Chemical Technology*, 2013, 27(1): 69–73.
- [10] 魏域琴, 宋丹丹, 翁正新. 基于 DPCA-KNN 的工业过程故障诊断方法研究 [C]//*第三十八届中国控制会议论文集 (7)*. 广州, 2019: 14–19.
- WEI Yuqin, SONG Dandan, WENG Zhengxin. Research on fault diagnosis method of industrial process based on DPCA-KNN[C]//*Proceedings of the 38th China Control Conference*. Guangzhou, 2019: 14–19.
- [11] 郭小萍, 徐月, 李元. 基于特征空间自适应 k 近邻工业过程故障检测 [J]. *高校化学工程学报*, 2019, 33(2): 453–461.
- GUO Xiaoping, XU Yue, LI Yuan. A feature space adaptive k -nearest neighbor method for industrial fault detection[J]. *Journal of chemical engineering of Chinese universities*, 2019, 33(2): 453–461.
- [12] 冯立伟, 张成, 李元, 等. 基于标准距离 k 近邻的多模态过程故障检测策略 [J]. *控制理论与应用*, 2019, 36(4): 553–560.
- FENG Liwei, ZHANG Cheng, LI Yuan, et al. Fault detection strategy of standard-distance-based k nearest neighbor rule in multimode processes[J]. *Control theory & applications*, 2019, 36(4): 553–560.
- [13] MACGREGOR J F, JAECKLE C, KIPARISIDES C, et al. Process monitoring and diagnosis by multiblock PLS methods[J]. *AIChE journal*, 1994, 40(5): 826–838.
- [14] 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 Zhenlei, JIANG Wei, WANG Xin. Plant-wide process monitoring based on multiblock MICA-PCA[J]. *Control and decision*, 2018, 33(2): 269–274.
- [15] 石怀涛, 王雨桐, 李颂华, 等. 基于多块相对变换独立主元分析的故障诊断方法 [J]. *控制与决策*, 2018, 33(11): 2009–2014.
- SHI Huaitao, WANG Yutong, LI Songhua, et al. Fault diagnosis approach based on relative transformation ICA of multiblock[J]. *Control and decision*, 2018, 33(11): 2009–2014.
- [16] GE Zhiqiang, SONG Zhihuan. Distributed PCA model for plant-wide process monitoring[J]. *Industrial & engineering chemistry research*, 2013, 52(5): 1947–1957.
- [17] JIANG Qingchao, YAN Xuefeng. Plant-wide process monitoring based on mutual information–multiblock principal component analysis[J]. *ISA transactions*, 2014, 53(5): 1516–1527.
- [18] HUANG Junping, YAN Xuefeng. Quality relevant and independent two block monitoring based on mutual information and KPCA[J]. *IEEE transactions on industrial electronics*, 2017, 64(8): 6518–6527.
- [19] JIANG Qingchao, YAN Xuefeng. Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA[J]. *Journal of process control*, 2015, 32: 38–50.
- [20] 童楚东, 史旭华. 基于互信息的 PCA 方法及其在过程监测中的应用 [J]. *化工学报*, 2015, 66(10): 4101–4106.
- TONG Chudong, SHI Xuhua. Mutual information based PCA algorithm with application in process monitoring[J]. *CIESC journal*, 2015, 66(10): 4101–4106.
- [21] 范雪莉, 冯海泓, 原猛. 基于互信息的主成分分析用于声场景分类 [J]. *声学技术*, 2013, 32(3): 222–227.
- FAN Xueli, FENG Haihong, YUAN Meng. Mutual in-

formation based principal component analysis for acoustic environment classification[J]. Technical acoustics, 2013, 32(3): 222–227.

- [22] 周红标, 乔俊飞. 基于高维 k -近邻互信息的特征选择方法[J]. 智能系统学报, 2017, 12(5): 595–600.

ZHOU Hongbiao, QIAO Junfei. Feature selection method based on high dimensional k -nearest neighbors mutual information[J]. CAAI transactions on intelligent systems, 2017, 12(5): 595–600.

- [23] GE Zhiqiang, SONG Zhihuan. Multimode process monitoring based on Bayesian method[J]. Journal of chemometrics, 2009, 23(12): 636–650.

- [24] YIN Shen, DING S X, HAGHANI A, et al. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process[J]. Journal of process control, 2012, 22(9): 1567–1581.

- [25] CHIANG L H, RUSSELL E L, BRAATZ R D. Tennessee Eastman process[M]//CHIANG L H, RUSSELL E L, BRAATZ R D. Fault Detection and Diagnosis in Industrial Systems. London: Springer, 2001: 103–112.

作者简介:



郑静, 硕士研究生, 主要研究方向为过程故障检测。



熊伟丽, 教授, 博士生导师, 主要研究方向为复杂工业过程建模与监控、智能软测量技术。主持国家自然科学基金面上项目、国家自然科学基金青年项目、江苏省产学研等省部级以上纵向项目等, 获授权发明专利近20项。发表学术论文近百篇。

第三届国际高性能大数据暨智能系统会议

The 3rd International Conference on High Performance Big Data and Intelligent Systems

第三届国际高性能大数据暨智能系统会议(The 3rd International Conference on High Performance Big Data and Intelligent Systems, HPBD&IS 2021)拟于2021年12月5日至12月7日在中国澳门举办。会议旨在搭建高性能计算、大数据及人工智能领域高端前沿交流平台, 促进海内外专家学者的交流与合作, 推动智能技术进步和智能产业发展。

会议主旨

本次会议将汇聚全球顶级专家、学者和产业界优秀人才, 共同围绕国际研究热点、核心关键技术、产业发展及挑战等进行开放式研讨。会议由中国计算机学会(CCF)、中国人工智能学会(CAAI)联合主办, IEEE Computer Society技术支持, 澳门大学、中国科学院深圳先进技术研究院、中国科学院半导体研究所、CCF高性能计算专委会、CAAI神经网络与计算智能专委会、中国自动化学会(CAA)模式识别与机器智能专委会、北京联合大学、天津理工大学、深圳市龙岗区机器人协会、深圳国际机器人城产业园共同承办。会议论文集将由IEEE Xplore®出版, EI收录, 优秀论文将推荐至Computation Practice and Experience(CCPE)、Optoelectronics Letters等SCI、EI期刊发表。

热忱欢迎广大同仁踊跃投稿并莅临本届会议! 详情请访问HPBD&IS 2021国际会议官网: <http://www.hpbd-dis.org/>