

9-17-2021

Fault Diagnosis of Industrial Process Based on LLE and K-means Clustering Algorithm

Li Yuan

Department of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China;

Zewei Geng

Department of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Fault Diagnosis of Industrial Process Based on LLE and K-means Clustering Algorithm

Abstract

Abstract: Because of the similarity of various types of data in the industrial process. The fault diagnosis using the K-means algorithm has a large error rate. *A K-means clustering algorithm based on Locally Linear Embedding (LLE) is proposed. the normal data is reduced by the LLE algorithm and the projection matrix is obtained. The projection matrix is used to map the original fault data to the low-dimensional space and the K-means algorithm clusters is used to carry out the data to establish a detection and diagnosis model.* The method is applied to the fault detection and diagnosis in the TE (Tennessee-Eastman) process and is compared with the traditional K-means algorithm and LLE algorithm. The results show that the proposed new method has a higher accuracy rate, and could effectively identify the unknown types of fault data.

Keywords

K-means clustering, locally linear embedding, TE process, fault diagnosis

Recommended Citation

Li Yuan, Geng Zewei. Fault Diagnosis of Industrial Process Based on LLE and K-means Clustering Algorithm[J]. Journal of System Simulation, 2021, 33(9): 2066-2073.

基于 LLE 与 K 均值聚类算法的工业过程故障诊断

李元, 耿泽伟

(沈阳化工大学 信息工程学院, 辽宁 沈阳 110142)

摘要: 工业过程中各类数据间具有一定的相似性, 单纯利用 K 均值算法对其进行故障诊断时, 存在很大的错误率。提出一种基于局部线性嵌入(Locally Linear Embedding, LLE)的 K 均值聚类算法, 将正常数据运用 LLE 算法降维并求出投影矩阵, 利用投影矩阵将原始故障数据映射到低维空间, 再利用 K 均值算法对其聚类, 建立检测与诊断模型。将此方法应用于田纳西-伊斯曼(Tennessee-Eastman, TE)过程中进行故障检测与诊断, 并同传统 K 均值算法及 LLE 算法对比, 结果表明: 提出的新方法具有更高的正确率, 同时可以有效地对未知类型的故障数据进行判别。

关键词: K 均值聚类; 局部线性嵌入; 田纳西-伊斯曼(Tennessee-Eastman, TE)过程; 故障诊断
中图分类号: TP277 文献标志码: A 文章编号: 1004-731X (2021) 09-2066-08

DOI: 10.16182/j.issn1004731x.joss.20-0362

Fault Diagnosis of Industrial Process Based on LLE and K-means Clustering Algorithm

Li Yuan, Geng Zewei

(Department of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China)

Abstract: Because of the similarity of various types of data in the industrial process. The fault diagnosis using the K-means algorithm has a large error rate. A K-means clustering algorithm based on Locally Linear Embedding (LLE) is proposed. the normal data is reduced by the LLE algorithm and the projection matrix is obtained. The projection matrix is used to map the original fault data to the low-dimensional space and the K-means algorithm clusters is used to carry out the data to establish a detection and diagnosis model. The method is applied to the fault detection and diagnosis in the TE (Tennessee-Eastman) process and is compared with the traditional K-means algorithm and LLE algorithm. The results show that the proposed new method has a higher accuracy rate, and could effectively identify the unknown types of fault data.

Keywords: K-means clustering; locally linear embedding; TE process; fault diagnosis

引言

工业生产技术以及产品不断更新, 化工过程变得越来越复杂, 所以对过程的监控、对数据的分析以及对故障检测诊断的研究越来越重要^[1-2], 基于数据驱动的方法受到广泛关注同时也取得了较好的效果, 此方法是以大量的历史数据为基础^[3-4]建立模型, 从而实现对化工过程的监测。

模式识别中的无监督聚类算法是工业过程故

障诊断的主要方法之一, 聚类就是利用相似性对数据进行分组的过程, 使同组数据相似度最高, 而组间数据差异性最大^[5]。因此, 当此类方法应用于工业过程故障检测与诊断时, 首先对历史数据进行挖掘, 得到诊断规则, 然后利用此规则对未知故障样本进行分类^[6]。各种聚类方法在工业过程故障诊断中也得到了较多应用。王英赫^[7]将模糊聚类应用于间歇过程故障诊断, 并取得了良好的效果。然而,

收稿日期: 2020-06-16

修回日期: 2020-07-10

基金项目: 国家自然科学基金重大项目(61490701); 国家自然科学基金(61673279)

第一作者: 李元(1964-), 女, 博士, 教授, 研究方向为基于数据驱动的工业过程故障诊断。E-mail: li-yuan@mail.tsinghua.edu.cn

模糊聚类需要计算各类别隶属度, 从而带来很大的计算量。张勇等^[8]通过 K-means 聚类算法进行变压器的故障诊断, 验证了该方法在变压器故障诊断中的优越性。管红立等^[9]提出一种将空间重构理论和 K-means 聚类相结合的算法, 并对电弧电压故障进行了诊断, 结果表明该方法能提高诊断的正确率, 但是样本的特征离散程度不能有效利用。目前得到广泛研究的还有降维算法, 如 PCA^[10], 但是此算法只能适用于线性结构数据; 非线性降维算法主成分分析 (Kernel Principal Component Analysis, KPCA)^[11], 此算法需要计算核函数以及高维空间数据, 增大了计算难度; 基于流形的非线性降维算法, 如 LLE (Locally Linear Embedding) 算法^[12]、NPE (Neighborhood Preserving Embedding) 算法^[13]等, 此类方法通过保持数据的局部线性关系, 得到低维特征数据。

根据工业过程故障诊断的要求, 以及传统 K 均值聚类算法的不足, 本文提出了 LLE-K 均值聚类算法, 通过 LLE 对正常数据降维时的投影矩阵将故障数据同样映射到低维空间, 放大各类数据间的差异性, 在低维空间中建立 K 均值聚类算法检测与诊断模型, 将局部线性嵌入 K 均值聚类算法应用于 TE 过程故障诊断中, 所得结果证明该算法具有一定的实用性和有效性。

1 局部线性嵌入算法

LLE 是一种流形学习算法, 此方法利用局部线性化构建超平面, 将高维数据映射到低维空间中, 使数据保持局部线性结构不变^[14]。运用 LLE 将高维数据样本 $\mathbf{X} = [x_1, x_2, \dots, x_m] \in R^{n \times m}$ 映射为低维数据 $\mathbf{Y} = [y_1, y_2, \dots, y_m] \in R^{d \times m}$ 的具体步骤为:

step 1: 采用欧式距离确定所有样本的 k 个近邻。

step 2: 计算权重矩阵 \mathbf{W} , 即每个样本与其近邻的线性关系, 使误差函数 $J(\mathbf{w})$ 最小,

$$J(\mathbf{w}) = \sum_{i=1}^m \left\| x_i - \sum_{j \in Q(i)} w_{ji} x_j \right\|^2, \quad \sum_{j \in Q(i)} w_{ji} = 1 \quad (1)$$

式中: $Q(i)$ 为样本 x_i 的 k 个近邻的集合; x_j 如果不

是 x_i 的近邻, 则其对应的 $w_{ji}=0$ 。

step 3: 计算降维后的矩阵 \mathbf{Y} , 通过 step 2 求出的权重矩阵 \mathbf{W} 使损失函数 $J(\mathbf{Y})$ 最小。损失函数与约束条件如下:

$$J(\mathbf{Y}) = \sum_{i=1}^m \left\| y_i - \sum_{j=1}^m w_{ji} y_j \right\|^2, \quad \sum_{i=1}^m y_i = 0; \quad \frac{1}{m} \sum_{i=1}^m y_i y_i^T = I \quad (2)$$

式(2)可进一步求解得:

$$J(\mathbf{Y}) = \text{tr}(\mathbf{Y}\mathbf{M}\mathbf{Y}^T) \quad (3)$$

式中: $\mathbf{M} = (\mathbf{I} - \mathbf{W})(\mathbf{I} - \mathbf{W})^T$, 因此 \mathbf{Y} 就是 \mathbf{M} 的特征向量构成的矩阵。

对于一个新的样本, LLE 算法不能直接将其投影到低维空间, 只能依靠映射矩阵 $\mathbf{A} = [a_1, a_2, \dots, a_d]^T \in R^{d \times n}$ 。可以通过式(4)求得映射矩阵^[15],

$$a_j = \arg \min_a \sum_{i=1}^k \left\| a^T x_i - y_{ij} \right\|^2 \quad (4)$$

解得 $\mathbf{A} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$, 新的样本 x_{new} 可以通过 $y_{\text{new}} = \mathbf{A}x_{\text{new}}$ 计算其新的嵌入坐标 y_{new} 。

2 基于 LLE 算法的 K 均值故障诊断

K-means 是依赖欧氏距离的方法^[16]。此方法具有易于理解、便与实现、理论可靠以及处理数据效率高等优点。它能把数据集 $\mathbf{X} = \{x_1, x_2, \dots, x_i, \dots, x_n\} \in R^{n \times d}$, 划分为 k 个子集 $C = \{c_k, k = 1, 2, \dots, K\}$, 子集就是类别。聚类中心 μ_i 为均值, 即:

$$\mu_i = \frac{1}{N_i} \sum_{x \in C_i} x \quad (5)$$

K-means 算法随机选取初始聚类中心, 计算样本与中心的距离, 将样本划分为最近的中心所属的类别中。将每一类数据的均值作为新的聚类中心, 多次迭代直至满足收敛条件^[17], 最后得到各个类别的数据。

K 均值聚类在故障诊断中的应用主要是对已知标签的正常数据与多类故障数据的混合数据进行聚类, 求出各个类别的聚类中心, 建立诊断模型。对未知类别的故障数据诊断时, 判断此故障样本与哪一类别的聚类中心更近, 此故障样本的属性就等

同于那一类别的故障属性。因此，建模的正确率直接影响未知数据诊断时的正确率。

工业过程的故障数据虽然是对正常数据的偏离，而且不同的故障发生在不同的变量上，但是各类数据间仍存在一定的相似性，以距离为指标对其划分也就是用 K-means 对其聚类时不能达到理想的效果，并不能建立理想的诊断模型。而 LLE 算法通过线性重构得到的低维空间数据可以保持高维数据的流形结构，即样本间的线性关系，所以利用正常数据降维所求出的映射矩阵对各类故障数据投影时，可以提取各类故障数据的特征结构，也就是可以放大各类故障数据间以及故障数据与正常数据间的差异性，从而利用欧氏距离为指标对数据进行划分时，可以提高准确率。

因此，本文提出基于 LLE 的 K 均值聚类算法，利用正常数据的映射矩阵将已知标签的故障数据映射到低维空间，同时在算法中加入故障检测的部分。

离线建模过程：

- (1) 正常数据通过 LLE 降维并求出投影矩阵 A ；
- (2) 各类故障数据利用矩阵 A 映射到相同维度的低维空间；
- (3) 利用 K 均值算法对降维后的混合数据聚类；
- (4) 找到正常数据的类别，并求出此类别中所有数据到聚类中心 m_1 的距离，从而确定控制限。

在线故障检测与诊断过程：

- (1) 待检测数据利用矩阵 A 映射到低维空间；
- (2) 计算待检测数据与建模结果中正常数据的聚类中心 m_1 的距离，并与控制限比较，判断其是否为故障数据；
- (3) 对于故障数据，判断其与建模时哪类故障数据的聚类中心最近，则其属性等同于那一类的故障数据。

对于故障数据的诊断结果有一个验证步骤：故障样本所被划分的那一类别中同正常数据一样确定一个控制限，而此故障样本与其聚类中心的距离如果小于此控制限，说明结果正确，如果大于此控制限，说明此故障样本不适用于此模型诊断，可以将其带入其他的模型中诊断其具体类别。基于 LLE 的 K 均值聚类算法进行故障诊断时的整体流程图如图 1 所示。

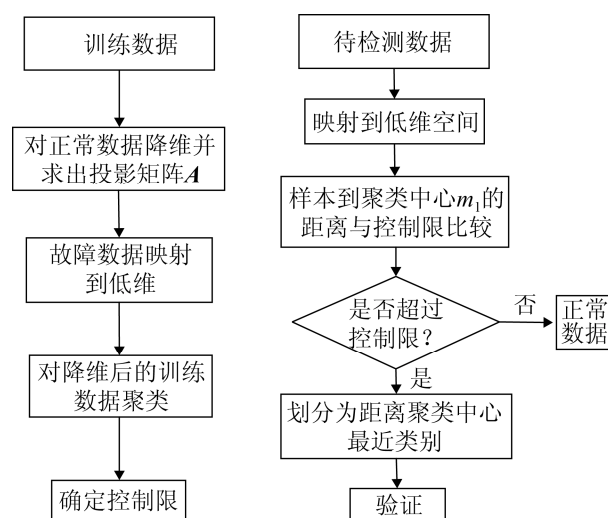


图 1 故障诊断流程图

Fig. 1 Flow chart of fault diagnosis

3 仿真实验

3.1 TE 过程数据

本文中的所有数据均来源于田纳西-伊斯曼 (Tennessee Eastman, TE) 仿真过程，此过程被广泛应用于故障检测与诊断的方法验证，TE 过程主要描述了装置、物料和能量之间的非线性关系^[18]。工艺流程图如图 2 所示。

TE 过程数据包括正常数据与 21 个故障数据，每个数据集都包括 52 个观测变量，部分已知具体类型的故障数据的描述如表 1 所示。

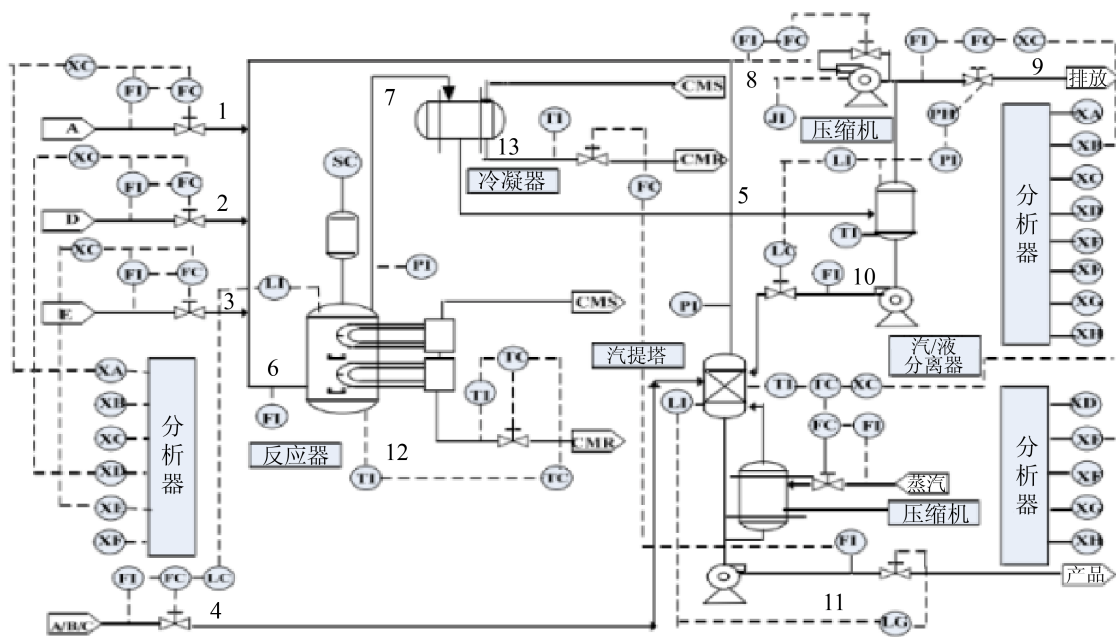


图 2 TE 过程工艺流程图

Fig. 2 TE process flow chart

表 1 故障描述及变化类型^[19-20]

Tab. 1 Fault description and change type

故障编号	性质描述	变化类型
IDV(1)	物料 A/C 进料比改变, 物料 B 含量不变	阶跃
IDV(2)	物料 A/C 进料比不变, 物料 B 含量改变	阶跃
IDV(3)	物料 D 进料温度改变	阶跃
IDV(4)	反应器冷却水入口温度改变	阶跃
IDV(5)	冷凝器冷却水入口温度改变	阶跃
IDV(6)	物料 A 进料损失	阶跃
IDV(7)	物料 C 压力损失	阶跃
IDV(8)	物料 A, B, C 的组成比例改变	随机变量
IDV(9)	物料 D 进料温度改变	随机变量
IDV(10)	物料 C 进料温度改变	随机变量
IDV(11)	反应器冷却水入口温度改变	随机变量
IDV(12)	冷凝器冷却水入口温度改变	随机变量
IDV(13)	反应动力学参数改变	慢偏移
IDV(14)	反应器冷却阀门	粘住
IDV(15)	冷凝器冷却阀门	粘住
IDV(21)	物流 4 阀门固定在恒定位置	恒定位置

3.2 数据采集

本文以 TE 过程为背景, 训练集采用正常数据

与已知标签的 3 种故障数据的组合共 1 200 个样本(每种数据各取 300 个样本), 52 个变量来建立模型。待检测数据为同样的 4 类数据共 400 个样本(每种数据各取 100 个样本), 将其作为未知标签的数据, 通过训练数据建立模型进行诊断, 待检测数据同训练数据一样是 52 个变量。对于建模聚类结果中正常数据所建立的故障检测模型可以直接针对故障整体数据(960 个样本, 52 个变量)进行检测。由于 K 均值算法的局限性, 本文只采用正常数据与 3 种故障数据建立模型。LLE 算法中近邻个数 $k=33$, 低维特征空间的维数 $d=14$ ^[21], 其中 k 是文献[21]求出的最佳近邻数。在下一小节中, 主要针对正常数据与故障 1, 4, 7 这 4 类数据建立模型详细说明, 并针对这 4 类数据的诊断结果与故障 4 和故障 2 的检测结果详细说明, 而对于其他类别的故障数据建模结果以及对于故障的检测效果基本相同的, 在本文中不再赘述。

3.3 仿真结果

首先对正常数据进行 LLE 降维, 得到投影矩阵 A , 其次将故障 1, 4, 7 的数据利用投影矩阵映

射到相同维度，最后用 K 均值算法对降维后的混合数据聚类，同时利用 K 均值算法对降维前的混合数据聚类得到结果作对比，结果如图 3~4 所示。

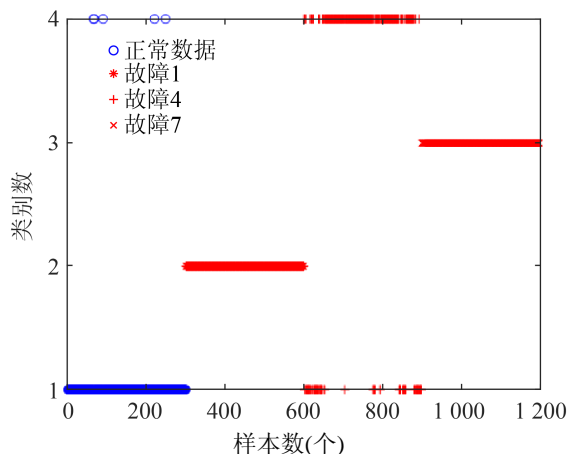


图 3 LLE-K 均值聚类结果
Fig. 3 LLE-K-means clustering results

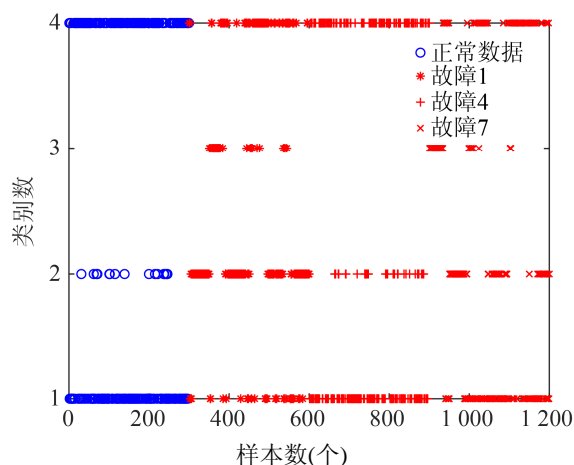


图 4 K 均值聚类结果
Fig. 4 K-means clustering results

建模数据一共包括 1 200 个样本，其中前 300 个是正常样本，后 900 个数据是 3 类故障样本。由图 3 结果可以看到 LLE-K 均值算法将数据分成了 4 类，并且与已知的数据信息对比发现此结果对于大部分数据是正确的，只有正常数据中 4 个样本以及故障 4 数据的一小部分样本结果错误。而对于图 4，未降维直接运用 K 均值算法的结果中有大量的数据并不能准确地划分，可以看出故障 1 与故障 7 的数据在各个类别中都存在，说明这个结果中所有的数据是混为一起的。聚类应用于故障诊断中

所得的诊断结果的正确率是直接受建模结果影响的，LLE-K 均值算法能够得到近似正确的结果，能够提取有效的诊断规则，也就是能够对未知的故障样本进行诊断。

得到聚类结果后，找到正常数据所处的类别，即图 3 中类别数是 1 的数据，将所有数据与聚类中心的距离排列并确定 95% 的控制限，建立故障检测模型。因为故障 4 是微小故障，从图 3 中也可以看出，一部分故障 4 的数据被划分为了正常数据，所以此检测模型并不能将故障 4 中所有的故障数据检测出来，本文将此检测模型所得结果与 LLE 故障检测得到的结果进行对比，如图 5~6 所示。

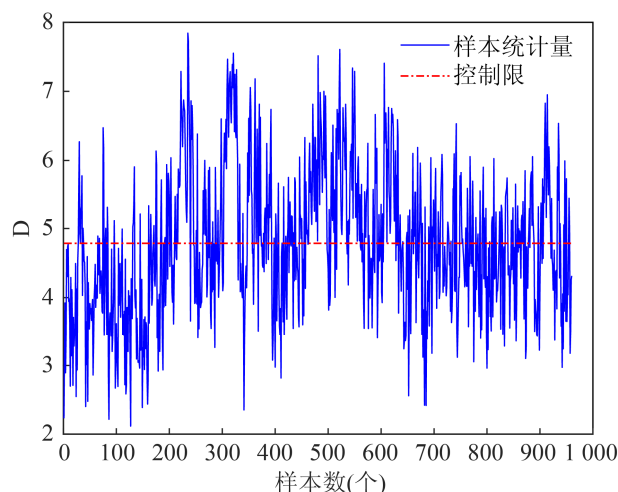


图 5 LLE-K 均值故障检测结果
Fig. 5 LLE-K-means Fault detection results

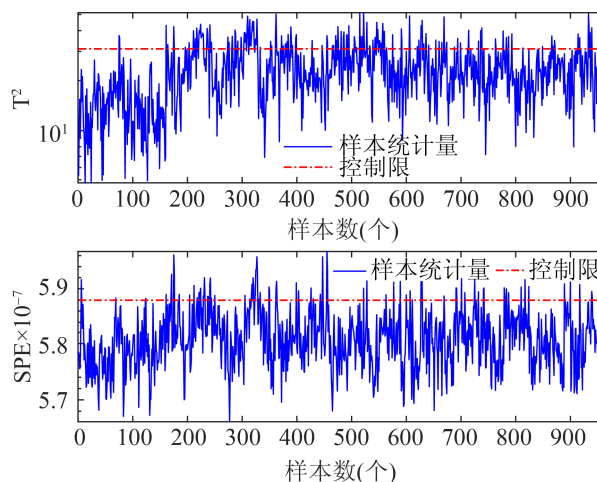


图 6 LLE 故障检测结果
Fig. 6 LLE fault detection results

通过图 5~6 对比可以看出本文提出的方法比传统的 LLE 算法具有更高的检测率, 而这 2 种方法中 LLE 降维的环节涉及到的参数的具体数值相同。并且本文方法对于其他的故障数据也同样具有更高的检测率, 如表 2 所示。

表 2 2 种方法检测率对比

Tab. 2 Comparison of detection rate between 2 methods

/%					
故障 编号	LLE-K 均值	LLE	故障 编号	LLE-K 均值	LLE
1	99.88	99.88	12	99.50	98.88
2	99.00	98.62	13	95.75	95.25
3	27.12	15.25	14	66.00	57.25
4	54.00	26.25	15	29.88	15.37
5	99.88	99.75	16	74.12	61.12
6	100	100	17	54.25	42.63
7	100	100	18	91.13	89.25
8	98.75	98.00	19	22.50	10.75
9	22.88	10.50	20	79.63	71.63
10	73.75	57.13	21	68.37	62.25
11	42.13	26.50			

对于只包含同建模数据相同类型的待检测数据, 可以将其直接带入诊断模型而不需要经过检测模型, 也就是通过 LLE 降维之后直接求待检测样本与建模结果中哪个类别的聚类中心更近, 将其划分为最近的那一类中, 如图 7 所示。

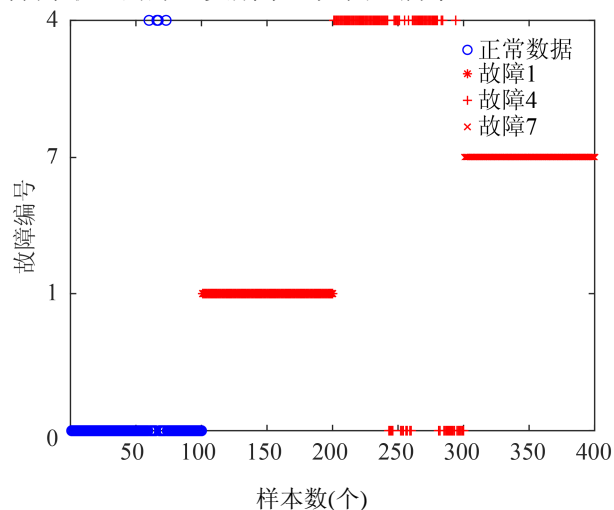


图 7 故障诊断结果

Fig. 7 Result of fault diagnosis

图 7 中的待检测样本就是正常数据与故障 1, 4, 7 各 100 个, 可以看出正常数据中有 3 个样本与故障 4 数据中的 10 多个样本划分错误, 正确率在 95% 以上。而如果待检测数据并不是建模数据相同类型的, 可以将其先进行故障检测阶段, 本文以故障 2 为例说明, 如图 8 所示为故障 2 的故障检测结果。

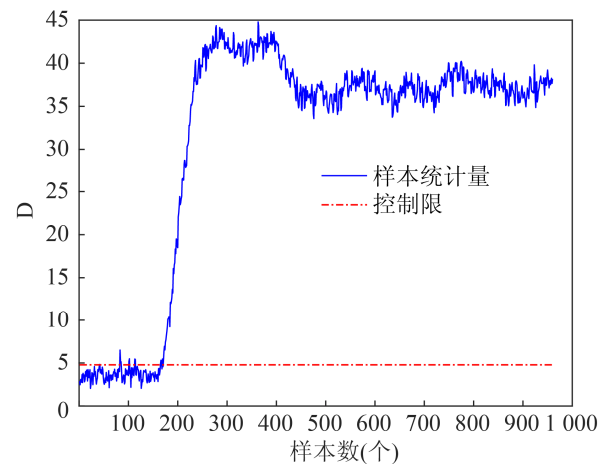


图 8 故障 2 检测结果

Fig. 8 Fault 2 detection results

通过图 8 可以看到大部分的故障数据都被检测出来了, 而对故障数据诊断时, 所有的故障数据都被划分为了故障 4 的类别中, 通过检验环节可以得出, 这些故障数据与聚类中心的距离大于故障 4 类别的控制限, 说明这些故障数据与建模数据的类型都不相同, 不适用于此诊断模型判别其具体类型, 如果想要得知其具体类型, 可以将其带入其他数据建立的模型中。而对于其他类型的故障数据按照上文所述的步骤检测与诊断时可以得到类似的结果。

4 结论

传统的 K 均值聚类算法并不能对 TE 过程的故障数据进行建模诊断, 因为这些故障数据间存在一定的相似性, 而本文提出的方法利用正常数据 LLE 降维时的投影矩阵, 将各类故障数据映射到低维空间后再运用 K 均值算法建立模型, 这样可以放大

各类数据间的差异性,从而可以得到更高正确率的建模结果,能够提高对未知数据划分的正确率。同时在 K 均值的建模结果中,加入了故障检测的模型。通过分析本文方法对 TE 数据的运行结果可以得知,此方法能够对未知数据进行检测与诊断,并且与 LLE 故障检测相比具有更高的正确率。然而对于微小故障的数据并不能将全部的故障检测出来,今后的工作是如何继续提高本文方法的检测率。

参考文献:

- [1] 周东华, 李钢, 李元. 数据驱动的工业过程故障诊断技术[M]. 北京: 科学技术出版社, 2011.
Zhou Donghua, Li Gang, Li Yuan. Data Driven Industrial Process Fault Diagnosis Technology[M]. Beijing: Science and Technology Press, 2011.
- [2] 李元, 马雨含, 郭金玉. 基于动态多向局部离群因子的在线故障检测[J]. 计算机应用研究, 2017, 34(11): 3259-3261, 3266.
Li Yuan, Ma Yuhan, Guo Jinyu. Online Fault Detection based on Dynamic Multidirectional Local Outliers[J]. Computer Application Research, 2017, 34(11): 3259-3261, 3266.
- [3] 刘强, 柴天佑, 秦泗钊, 等. 基于数据和知识的工业过程监视及故障诊断综述[J]. 控制与决策, 2010, 25(6): 801-807.
Liu Qiang, Chai Tianyou, Qin Sizhao, et al. Overview of Industrial Process Monitoring and Fault Diagnosis Based on Data and Knowledge[J]. Control and Decision, 2010, 25(6): 801-807.
- [4] Li Y, Zhang X M. Diffusion Maps based K-nearest-neighbor Rule Technique for Semiconductor Manufacturing Process Fault Detection[J]. Chemometrics and Intelligent Laboratory Systems (S0169-7439), 2014, 136(15): 47-57.
- [5] Yu N, Luo J R. Study on Magneto-Hydro- Dynamics Disturbance Signal Feature Classification Using Improved S-Transform Algorithm and Radial Basis Function Neural Network[J]. Sensors & Transducers (S2306-8515), 2014, 178(9): 219-225.
- [6] 毛茗茗. 基于 KPCA-FCM 的工业过程故障检测与诊断[D]. 北京: 北京化工大学, 2017.
Mao Mingming. KPCA-FCM-based Industrial Process Fault Detection and Diagnosis[D]. Beijing: Beijing University of Chemical Technology, 2017.
- [7] 王英赫. 基于模糊聚类的间歇过程故障诊断的研究[D]. 哈尔滨: 哈尔滨理工大学, 2018.
Wang Yinghe. Research on Fault Diagnosis of Batch Process Based on Fuzzy Clustering[D]. Harbin: Harbin University of Science and Technology, 2018.
- [8] 张勇, 钟建伟, 周文辉, 等. 基于 K-means 聚类分析的变压器故障诊断[J]. 湖北民族学院学报(自然科学版), 2018, 36(4): 452-455.
Zhang Yong, Zhong Jianwei, Zhou Wenhui, et al. Transformer Fault Diagnosis based on K-means Cluster Analysis[J]. Journal of Hubei University for Nationalities (Natural Science Edition), 2018, 36(4): 452-455.
- [9] 管红立, 李亚芳, 郑文栋, 等. 基于相空间重构理论和 k-means 聚类算法电弧故障诊断[J]. 电器与能效管理技术, 2017(17): 1-8.
Guan Hongli, Li Yafang, Zheng Wendong, et al. Arc Fault Diagnosis Based on Phase Space Reconstruction Theory and K-means Clustering Algorithm[J]. Electrical Appliance and Energy Efficiency Management Technology, 2017(17): 1-8.
- [10] 王晶, 刘莉, 曹柳林, 等. 基于核 Fisher 包络分析的间歇过程故障诊断[J]. 化工学报, 2014, 65(4): 1317-1326.
Wang Jing, Liu Li, Cao Liulin, et al. Fault Diagnosis of Batch Process Based on Nuclear Fisher Envelope Analysis[J]. Journal of Chemical Industry and Engineering, 2014, 65(4): 1317-1326.
- [11] Jiang Qingchao, Yan Xuefeng. Nonlinear Plant-wide Process Monitoring Using MI-spectral Clustering and Bayesian Inference- based Multiblock KPCA[J]. Journal of Process Control (S0959-1524), 2015, 32(1): 38-50.
- [12] Yang X F, Goh A, Qiu A Q. Locally Linear Diffeomorphic Metric Embedding (LLDME) for Surface-based Anatomical Shape Modeling[J]. Neuroimage (S1053-8119), 2011, 56(1): 149-161.
- [13] Miao A M, Ge Z Q, Song Z H, et al. Nonlocal Structure Constrained Neighborhood Preserving Embedding Model and Its Application for Fault Detection[J]. Chemometrics and Intelligent Laboratory Systems (S0169-7439), 2015, 142(1): 184-196.
- [14] Wang J. Real Local-linearity Preserving Embedding[J]. Neurocomputing (S0925-2312), 2014, 136(20): 7-13.
- [15] Li B W, Zhang Y. Supervised Locally Linear Embedding Projection (SLLEP) for Machinery Fault Diagnosis[J]. Mechanical Systems and Signal Processing (S0888-3270), 2011, 25(8): 3125-3134.
- [16] 王千, 王成, 冯振元, 等. K-means 聚类算法研究综述[J]. 电子设计工程, 2012, 20(7): 21-24.
Wang Qian, Wang Cheng, Feng Zhenyuan, et al. Review

- of K-means Clustering Algorithm Research[J]. *Electronic Design Engineering*, 2012, 20(7): 21-24.
- [17] 李元, 白岩松. 改进主成分分析的 KNN 故障检测研究[J]. *沈阳化工大学学报*, 2018, 32(4): 366-371.
Li Yuan, Bai Yansong. Research on KNN Fault Detection with Improved Principal Component Analysis[J]. *Journal of Shenyang University of Chemical Technology*, 2018, 32(4): 366-371.
- [18] 郭金玉, 韩建斌, 李元, 等. 基于局部 Fisher 判别分析的复杂化工过程故障诊断[J]. *计算机应用研究*, 2018, 35(4): 1122-1125, 1129.
Guo Jinyu, Han Jianbin, Li Yuan, et al. Fault Diagnosis of Complex Chemical Processes based on Local Fisher Discriminant Analysis[J]. *Computer Application Research*, 2018, 35(4): 1122-1125, 1129.
- [19] 张凯林. 基于主元分析和偏最小二乘的 TE 过程监测方法的研究[D]. 天津: 天津理工大学, 2015.
Zhang Kailin. Research on TE Process Monitoring Method based on Principal Component Analysis and Partial Least Squares[D]. Tianjin: Tianjin University of Technology, 2015.
- [20] 张姮. TE 过程故障诊断方法比较研究[D]. 沈阳: 沈阳理工大学, 2014.
Zhang Heng. Comparative Study of Fault Diagnosis Methods in TE Process[D]. Shenyang: Shenyang University of Science and Technology, 2014.
- [21] 薄翠梅, 韩晓春, 易辉, 等. 基于聚类选择 k 近邻的 LLE 算法及故障检测[J]. *化工学报*, 2016, 67(3): 925-930.
Bo Cuimei, Han Xiaochun, Yi Hui, et al. LLE Algorithm and Fault Detection Based on Cluster Selection K-nearest Neighbor[J]. *Journal of Chemical Industry and Engineering*, 2016, 67(3): 925-930.