



# (12)发明专利申请

(10)申请公布号 CN 109933040 A

(43)申请公布日 2019.06.25

(21)申请号 201711365157.4

(22)申请日 2017.12.18

(71)申请人 中国科学院沈阳自动化研究所  
地址 110016 辽宁省沈阳市东陵区南塔街  
114号

(72)发明人 李帅 周晓锋 史海波 潘福成  
李歆 张宜驰

(74)专利代理机构 沈阳科苑专利商标代理有限公司 21002  
代理人 李巨智

(51)Int.Cl.  
G05B 23/02(2006.01)

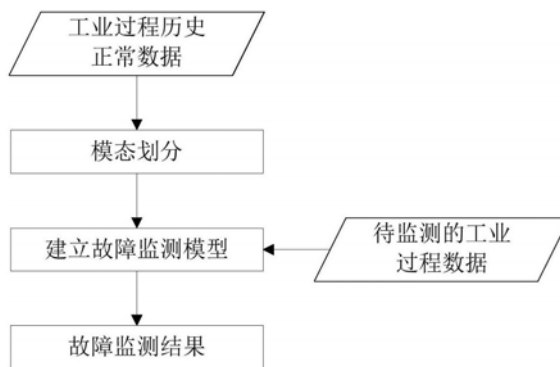
权利要求书3页 说明书6页 附图1页

## (54)发明名称

基于层次密度峰值聚类和最相似模态的故障监测方法

## (57)摘要

本发明涉及一种基于层次密度峰值聚类和最相似模态的故障监测方法,对工业过程历史正常数据进行模态划分,获取层次模态信息;利用层次模态信息对工业过程历史正常数据建立故障监测模型;将待监测的工业过程数据,得到最相似模态,输入到所述故障监测模型,进行故障监测。本发明利用已有的工业数据资源,考虑了复杂工业过程的多模态性和多模态动态性和不确定性,克服现有多模态故障监测方法依赖先验的模态信息,采用固定的模态划分及模型等局限,这对于及时发现工业过程的异常工况、保证生产安全、提高产品质量具有重要意义。



1. 一种基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在於:包括以下步骤:

步骤1:对工业过程历史正常数据进行模态划分,获取层次模态信息;

步骤2:利用层次模态信息对工业过程历史正常数据建立故障监测模型;

步骤3:将待监测的工业过程数据,得到最相似模态,输入到所述故障监测模型,进行故障监测。

2. 根据权利要求1所述的基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在於:所述模态划分包括以下过程:

步骤1:搜集工业过程历史正常数据并对其进行标准化处理;

步骤2:对标准化处理后数据进行层次密度峰值聚类,得到初始层次模态信息;

步骤3:根据模态连续特性调整初始层次模态信息,得到层次模态信息。

3. 根据权利要求2所述的基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在於:所述层次密度峰值聚类包括以下过程:

步骤1:计算标准化处理后数据 $X = \{x_1, x_2, \dots, x_N\}$ 各数据点之间的测地距离 $d_{geo}(x_i, x_j)$ ;

步骤2:选取不同的百分比参数P,得到截止距离 $d_c^{geo}$ ,计算公共近邻个数 $CNN(x_i, x_j)$ ,获取混合测地距离 $d_{hybgeo}(x_i, x_j)$ ;

步骤3:计算标准化处理后数据各数据点之间的局部密度 $\rho_i^{hybgeo}$ 和最小距离 $\delta_i^{hybgeo}$ ;

步骤4:人工观察选取不同的截止距离 $d_c^{geo}$ 对应的局部密度和最小距离均比一般数据点的局部密度和最小距离大且具有离群特性的数据点为类簇中心,将剩余非中心点划入比其局部密度高的数据点中距离最近数据点的类簇中,得到层次密度峰值聚类结果。

4. 根据权利要求3所述的基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在於:所述计算公共近邻个数 $CNN(x_i, x_j)$ ,是通过以下公式得到:

$$CNN(x_i, x_j) = \text{number} \left( d_{geo}(x_i, x_j) \leq d_c^{geo} \right) \quad (1)$$

所述截止距离,是通过以下公式得到:

$$d_c^{geo} = d_{N_d \times \frac{P}{100}}^{geo} \quad (2)$$

其中, $d_c^{geo} \in D = [d_1^{geo}, \dots, d_{N_d}^{geo}]$ , $d_c^{geo}$ 为截止距离,D是标准化处理后数据各数据点之间的测地距离的集合,N为数据点的个数, $N_d \times \frac{P}{100}$ 为 $d_{N_d \times \frac{P}{100}}^{geo}$ 的下标,P为百分比参数,取值范围为0.1-5;

所述混合测地距离 $d_{hybgeo}(x_i, x_j)$ ,是通过以下公式得到:

$$d_{hybgeo}(x_i, x_j) = \frac{d_{geo}(x_i, x_j)}{(CNN(x_i, x_j) + 1)} \quad (3)$$

所述局部密度 $\rho_i^{hybgeo}$ ,是通过以下公式得到:

$$\rho_i^{hybgeo} = \sum_j \exp \left( -\frac{d_{hybgeo}(\mathbf{x}_i, \mathbf{x}_j)}{d_c^{hybgeo}} \right)^2 \quad (4)$$

所述最小距离 $\delta_i^{hybgeo}$ ,是通过以下公式得到:

$$\delta_i^{hybgeo} = \min_{j: \rho_j^{hybgeo} > \rho_i^{hybgeo}} (d_{hybgeo}(\mathbf{x}_i, \mathbf{x}_j)) \quad (5)$$

其中, $d_{geo}(x_i, x_j)$ 为各数据点之间的测地距离。

5. 根据权利要求2所述的基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在于:所述根据模态连续特性调整初始层次模态信息,是采用最小模态长度 $t_{min}$ ,取值范围为3-10个采样间隔,调整初始层次模态信息,通过将小于 $t_{min}$ 的模态数据点调整到最近的模态实现。

6. 根据权利要求1所述的基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在于:所述建立故障监测模型包括以下过程:利用层次模态信息,基于独立元分析建立故障监测模型,得到故障监测置信限。

7. 根据权利要求6所述的基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在于:所述基于独立元分析建立故障监测模型包括以下过程:

1) 计算协方差矩阵 $C(X)$ ;

2) 利用SVD分解 $C(X) = U \Lambda U^T$ ,计算协方差矩阵 $C(X)$ 的特征值对角矩阵 $\Lambda$ 和特征向量矩阵 $U$ ;

3) 计算白化矩阵 $Q = \Lambda^{-1/2} U^T$ ;

4) 得到白化变换矩阵 $Z = QX$ ;

5) 计算正交矩阵 $B$ ;

6) 计算预测的独立元矩阵 $S = B^T Z$ ;

7) 计算低维解混矩阵 $W = B^T Q$ ;

8) 计算故障监测 $I^2 = s^T s$ 统计量和 $SPE = (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}})$ 统计量,其中 $\hat{\mathbf{x}} = \mathbf{W}\mathbf{x}$ , $\mathbf{x}$ 表示 $X = \{x_1, x_2, \dots, x_N\}$ 的一个数据点。

8. 根据权利要求7所述的基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在于:所述计算正交矩阵 $B$ 包括以下过程:

步骤1:选择估计独立元的个数 $d$ ,设置迭代次数 $i \leftarrow 1$ ;

步骤2:以单位模向量给随机向量 $b_i$ 赋初值;

步骤3:令 $b_i \leftarrow E\{z g(b_i^T z)\} - E\{g'(b_i^T z)\} b_i$ ,其中 $g$ 和 $g'$ 分别为 $G = \tanh(\cdot)$ 的一阶导数和二阶导数;

步骤4:通过正交化来去除相关性, $\mathbf{b}_i \leftarrow \mathbf{b}_i - \sum_{j=1}^{i-1} (\mathbf{b}_i^T \mathbf{b}_j) \mathbf{b}_j$ ;

步骤5:归一化, $\mathbf{b}_i \leftarrow \frac{\mathbf{b}_i}{\|\mathbf{b}_i\|}$ ;

步骤6:假如 $b_i$ 没有收敛,返回步骤3;

步骤7:如果 $b_i$ 收敛,输出向量 $b_i$ ;

步骤8:若 $i \leq m$ ,则设定 $i \leftarrow i+1$ 且返回步骤2。

9.根据权利要求6所述的基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在于:所述得到故障监测置信限包括以下过程:

1) 计算全局故障监测置信限  $I_{\alpha}^2 = \frac{d(N^2 - 1)}{N(N - d)} F_{d, N-d, \alpha}$ , 其中 $\alpha$ 表示置信水平,取值为 $\alpha = 1\%$ , $d$ 为选择估计独立元的个数, $N$ 为数据点的个数, $F_{d, N-d, \alpha}$ 为置信水平为 $\alpha$ 的自由度为 $d$ 和 $N-d$ 的F分布临界值;

2) 计算局部故障监测置信限  $Q_{mi, \alpha} = g_{mi} \chi_{h_{mi}, \alpha}^2$ , 其中 $m_i$ 为层次模态信息,  $g_{mi} = S_{mi} / 2\mu_{mi}$ ,  $h_{mi} = 2\mu_{mi}^2 / S_{mi}$ ,  $\mu_{mi}$ 和 $S_{mi}$ 为层次模态信息为 $m_i$ 的SPE统计量的均值和方差。

10.根据权利要求1所述的基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在于:所述将待监测的工业过程数据,得到最相似模态,输入到所述故障监测模型,进行故障监测包括以下过程:

1) 得到待监测的标准化后工业过程数据 $x_{new}$ 的最相似模态;

2) 得到 $x_{new}$ 的统计量  $I^2 = \mathbf{s}_{new}^T \mathbf{s}_{new}$  和  $SPE = (\mathbf{x}_{new} - \hat{\mathbf{x}}_{new})^T (\mathbf{x}_{new} - \hat{\mathbf{x}}_{new})$ ;

3) 如果 $x_{new}$ 在最相似模态的 $I^2$ 或SPE统计量大于对应的故障监测置信限,则判定工业过程发生故障,否则为无故障。

11.根据权利要求10所述的基于层次密度峰值聚类和最相似模态的故障监测方法,其特征在于:所述得到待监测的标准化后工业过程数据的最相似模态,是通过以下公式得到:

$$\max_{mi} df(\mathbf{x}_{new}) = \rho_{new}^{mi} - d_{new, center}^{mi} \quad (6)$$

其中,  $\rho_{new}^{mi}$  为待监测的工业过程数据在层次模态信息中各个模态的局部密度,  $d_{new, center}^{mi}$  为待监测的工业过程数据点与层次模态信息为 $m_i$ 的类簇中心的距离。

## 基于层次密度峰值聚类和最相似模态的故障监测方法

### 技术领域

[0001] 本发明涉及故障监测与诊断技术领域,具体地说是一种基于层次密度峰值聚类和最相似模态的故障监测方法。

### 背景技术

[0002] 故障监测的目的是及时发现工业过程的异常工况、保证生产安全、提高产品质量。工业过程的操作条件通常因原材料,制造参数,产品规格等因素的改变而改变,从而产生多种操作模态。不同的模态具有相似的,各自特定的特性及持续时间。一个模态被定义为具有相似统计特性和一定持续时间的过程。此外,两个相邻的模态间还可能存在着逐渐改变的过渡过程。因此,研究有效可行的多模态过程故障监测方法具有重要意义。

[0003] 多元统计故障监测方法如主成分分析,偏最小二乘引起了广泛关注。然而,工业过程具有动态,非线性和多模态特性,因此传统的多元统计故障监测方法无法直接用于监测实际工业过程。因此,应该研究更好的多模态过程故障监测方法。目前,在现有的多模态过程故障监测算法中,针对不同模态建立局部模型的多模型方法未考虑模态间的全局相关性,影响故障监测的准确性;很多方法根据模态先验知识进行模态划分或采用固定的模态划分及模型,未考虑多模态动态性和不确定性,采用这些方法对实际工业过程进行故障监测时适应性低。

[0004] 密度峰值聚类是一种用于识别任意形状类簇的基于密度的聚类算法。在故障监测中,基于层次密度峰值聚类的模态划分方法可以获取复杂工业过程的层次模态信息,不需要模态先验知识。因而,采用层次模态信息建立故障监测模型是很有意义的。

### 发明内容

[0005] 针对现有技术的不足,本发明提供一种基于层次密度峰值聚类和最相似模态的故障监测方法,考虑了复杂工业过程的多模态性和多模态动态性和不确定性,克服现有多模态故障监测方法依赖先验模态信息的问题。

[0006] 本发明为实现上述目的所采用的技术方案是:

[0007] 一种基于层次密度峰值聚类和最相似模态的故障监测方法,包括以下步骤:

[0008] 步骤1:对工业过程历史正常数据进行模态划分,获取层次模态信息;

[0009] 步骤2:利用层次模态信息对工业过程历史正常数据建立故障监测模型;

[0010] 步骤3:将待监测的工业过程数据,得到最相似模态,输入到所述故障监测模型,进行故障监测。

[0011] 所述模态划分包括以下过程:

[0012] 步骤1:搜集工业过程历史正常数据并对其进行标准化处理;

[0013] 步骤2:对标准化处理后数据进行层次密度峰值聚类,得到初始层次模态信息;

[0014] 步骤3:根据模态连续特性调整初始层次模态信息,得到层次模态信息。

[0015] 所述层次密度峰值聚类包括以下过程:

[0016] 步骤1:计算标准化处理后数据 $X = \{x_1, x_2, \dots, x_N\}$ 各数据点之间的测地距离 $d_{geo}(x_i, x_j)$ ;

[0017] 步骤2:选取不同的百分比参数 $P$ ,得到截止距离 $d_c^{geo}$ ,计算公共近邻个数 $CNN(x_i, x_j)$ ,获取混合测地距离 $d_{hybgeo}(x_i, x_j)$ ;

[0018] 步骤3:计算标准化处理后数据各数据点之间的局部密度 $\rho_i^{hybgeo}$ 和最小距离 $\delta_i^{hybgeo}$ ;

[0019] 步骤4:人工观察选取不同的截止距离 $d_c^{geo}$ 对应的局部密度和最小距离均比一般数据点的局部密度和最小距离大且具有离群特性的数据点为类簇中心,将剩余非中心点划入比其局部密度高的数据点中距离最近数据点的类簇中,得到层次密度峰值聚类结果。

[0020] 所述计算公共近邻个数 $CNN(x_i, x_j)$ ,是通过以下公式得到:

$$[0021] \quad CNN(\mathbf{x}_i, \mathbf{x}_j) = \text{number}(d_{geo}(\mathbf{x}_i, \mathbf{x}_j) \leq d_c^{geo}) \quad (1)$$

[0022] 所述截止距离,是通过以下公式得到:

$$[0023] \quad d_c^{geo} = d_{N_d \times \frac{P}{100}}^{geo} \quad (2)$$

[0024] 其中, $d_c^{geo} \in D = [d_1^{geo}, \dots, d_{N_d}^{geo}]$ ,  $d_c^{geo}$ 为截止距离, $D$ 是标准化处理后数据各数据点之间的测地距离的集合, $N$ 为数据点的个数, $N_d \times \frac{P}{100}$ 为 $d_{N_d \times \frac{P}{100}}^{geo}$ 的下标, $P$ 为百分比参数,取值范围为0.1-5;

[0025] 所述混合测地距离 $d_{hybgeo}(x_i, x_j)$ ,是通过以下公式得到:

$$[0026] \quad d_{hybgeo}(\mathbf{x}_i, \mathbf{x}_j) = \frac{d_{geo}(\mathbf{x}_i, \mathbf{x}_j)}{(CNN(\mathbf{x}_i, \mathbf{x}_j) + 1)} \quad (3)$$

[0027] 所述局部密度 $\rho_i^{hybgeo}$ ,是通过以下公式得到:

$$[0028] \quad \rho_i^{hybgeo} = \sum_j \exp\left(-\frac{d_{hybgeo}(\mathbf{x}_i, \mathbf{x}_j)}{d_c^{hybgeo}}\right)^2 \quad (4)$$

[0029] 所述最小距离 $\delta_i^{hybgeo}$ ,是通过以下公式得到:

$$[0030] \quad \delta_i^{hybgeo} = \min_{j: \rho_j^{hybgeo} > \rho_i^{hybgeo}} (d_{hybgeo}(\mathbf{x}_i, \mathbf{x}_j)) \quad (5)$$

[0031] 其中, $d_{geo}(x_i, x_j)$ 为各数据点之间的测地距离;

[0032] 所述根据模态连续特性调整初始层次模态信息,是采用最小模态长度 $t_{min}$ ,取值范围为3-10个采样间隔,调整初始层次模态信息,通过将小于 $t_{min}$ 的模态数据点调整到最近的模态实现。

[0033] 所述建立故障监测模型包括以下过程:利用层次模态信息,基于独立元分析建立故障监测模型,得到故障监测置信限。

[0034] 所述基于独立元分析建立故障监测模型包括以下过程:

[0035] 1) 计算协方差矩阵 $C(X)$ ;

[0036] 2) 利用SVD分解 $C(X) = U \Lambda U^T$ ,计算协方差矩阵 $C(X)$ 的特征值对角矩阵 $\Lambda$ 和特征向

量矩阵U;

[0037] 3) 计算白化矩阵 $Q = \Lambda^{-1/2}U^T$ ;

[0038] 4) 得到白化变换矩阵 $Z = QX$ ;

[0039] 5) 计算正交矩阵B;

[0040] 6) 计算预测的独立元矩阵 $S = B^T Z$ ;

[0041] 7) 计算低维解混矩阵 $W = B^T Q$ ;

[0042] 8) 计算故障监测 $I^2 = s^T s$ 统计量和 $SPE = (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}})$ 统计量,其中 $\hat{\mathbf{x}} = W\mathbf{x}$ , $\mathbf{x}$ 表示 $X = \{x_1, x_2, \dots, x_N\}$ 的一个数据点。

[0043] 所述计算正交矩阵B包括以下过程:

[0044] 步骤1:选择估计独立元的个数d(一般保留累计贡献率为0.85-0.95的独立元),设置迭代次数 $i \leftarrow 1$ ;

[0045] 步骤2:以单位模向量给随机向量 $b_i$ 赋初值;

[0046] 步骤3:令 $b_i \leftarrow E\{zg(b_i^T z)\} - E\{g'(b_i^T z)\} b_i$ ,其中g和 $g'$ 分别为 $G = \tanh(\cdot)$ 的一阶导数和二阶导数;

[0047] 步骤4:通过正交化来去除相关性, $b_i \leftarrow b_i - \sum_{j=1}^{i-1} (b_i^T b_j) b_j$ ;

[0048] 步骤5:归一化, $b_i \leftarrow \frac{b_i}{\|b_i\|}$ ;

[0049] 步骤6:假如 $b_i$ 没有收敛,返回步骤3;

[0050] 步骤7:如果 $b_i$ 收敛,输出向量 $b_i$ ;

[0051] 步骤8:若 $i \leq m$ ,则设定 $i \leftarrow i+1$ 且返回步骤2。

[0052] 所述得到故障监测置信限包括以下过程:

[0053] 1) 计算全局故障监测置信限 $I_\alpha^2 = \frac{d(N^2 - 1)}{N(N - d)} F_{d, N-d, \alpha}$ ,其中 $\alpha$ 表示置信水平,

取值为 $\alpha = 1\%$ ,d为选择估计独立元的个数,N为数据点的个数, $F_{d, N-d, \alpha}$ 为置信水平为 $\alpha$ 的自由度为d和N-d的F分布临界值;

[0054] 2) 计算局部故障监测置信限 $Q_{mi, \alpha} = g_{mi} \chi_{h_{mi}, \alpha}^2$ ,其中 $m_i$ 为层次模态信息, $g_{mi} = S_{mi} / 2\mu_{mi}$ , $h_{mi} = 2\mu_{mi}^2 / S_{mi}$ , $\mu_{mi}$ 和 $S_{mi}$ 为层次模态信息为 $m_i$ 的SPE统计量的均值和方差。

[0055] 所述将待监测的工业过程数据,得到最相似模态,输入到所述故障监测模型,进行故障监测包括以下过程:

[0056] 1) 得到待监测的标准化后工业过程数据 $x_{new}$ 的最相似模态;

[0057] 2) 得到 $x_{new}$ 的统计量 $I^2 = \mathbf{s}_{new}^T \mathbf{s}_{new}$ 和 $SPE = (\mathbf{x}_{new} - \hat{\mathbf{x}}_{new})^T (\mathbf{x}_{new} - \hat{\mathbf{x}}_{new})$ ;

[0058] 3) 如果 $x_{new}$ 在最相似模态的 $I^2$ 或SPE统计量大于对应的故障监测置信限,则判定工业过程发生故障,否则为无故障。

[0059] 所述得到待监测的标准化后工业过程数据的最相似模态,是通过以下公式得到:

$$[0060] \quad \max_{mi} df(\mathbf{x}_{new}) = \rho_{new}^{mi} - d_{new,center}^{mi} \quad (6)$$

[0061] 其中,  $\rho_{new}^{mi}$  为待监测的工业过程数据在层次模态信息中各个模态的局部密度,  $d_{new,center}^{mi}$  为待监测的工业过程数据点与层次模态信息为mi的类簇中心的距离。

[0062] 本发明具有以下有益效果及优点:

[0063] 本发明通过将层次密度峰值聚类,独立元分析和最相似模态结合,先利用基于层次密度峰值聚类的模态划分方法获取多模态工业过程的层次模态信息,再利用层次模态信息建立故障监测模型,最后利用最相似模态进行故障监测,考虑了多模态动态性和不确定性,克服现有故障监测方法依赖先验的模态信息,采用固定的模态划分及模型等局限,这对于及时发现工业过程的异常工况、保证生产安全、提高产品质量具有重要意义。

## 附图说明

[0064] 图1是本发明的方法流程图。

[0065] 图2是本发明的层次密度峰值聚类人工观察步骤示意图。

## 具体实施方式

[0066] 下面结合附图及实施例对本发明做进一步的详细说明。

[0067] 如图1所示为本发明的方法流程图。

[0068] 基于层次密度峰值聚类和最相似模态的故障监测方法,搜集工业过程数据,在得到层次模态信息后,利用独立元分析与最相似模态进行故障监测。

[0069] 本发明的具体步骤如下:

[0070] 步骤1:搜集青霉素发酵过程历史正常数据,包括通风率,搅拌器功率,底物加料速度,反应产生的热量,培养基体积,二氧化碳浓度,PH值,温度,溶氧饱和度,并对其进行标准化处理,处理后的数据各变量均值为0,方差为1;

[0071] 步骤2:计算标准化处理后数据  $X = \{x_1, x_2, \dots, x_N\}$  各数据点之间的测地距离  $d_{geo}(x_i, x_j)$ ,测地距离能比欧氏距离更能反映数据的真实结构;

[0072] 步骤3:选取不同的百分比参数P(取值范围为0.1-5),得到截止距离

$$d_c^{geo} = d_{N_d \times \frac{p}{100}}^{geo} \in D = [d_1^{geo}, \dots, d_{N_d}^{geo}], D \text{ 是标准化处理后数据各数据点之间的测地距离的集合,}$$

$N$ 为数据点的个数,  $N_d \times \frac{p}{100}$  为  $d_{N_d \times \frac{p}{100}}^{geo}$  的下标,计算公共近邻个数  $CNN(x_i, x_j)$  和混合测地距离  $d_{hybgeo}(x_i, x_j)$ ,公共近邻反映数据点间的局部距离特性,能更好地区分相同类簇与不同类簇的数据点,公式如下:

$$[0073] \quad CNN(\mathbf{x}_i, \mathbf{x}_j) = \text{number}(d_{geo}(\mathbf{x}_i, \mathbf{x}_j) \leq d_c^{geo}) \quad (1)$$

$$[0074] \quad d_{hybgeo}(\mathbf{x}_i, \mathbf{x}_j) = \frac{d_{geo}(\mathbf{x}_i, \mathbf{x}_j)}{(CNN(\mathbf{x}_i, \mathbf{x}_j) + 1)} \quad (2)$$

[0075] 步骤4:计算各数据点之间的局部密度  $\rho_i^{hybgeo}$  和最小距离  $\delta_i^{hybgeo}$ ,不同的截止距离计



算得到的局部密度和最小距离也不同,公式如下:

$$[0076] \quad \rho_i^{hybgeo} = \sum_j \exp\left(-\frac{d_{hybgeo}(\mathbf{x}_i, \mathbf{x}_j)}{d_c^{hybgeo}}\right)^2 \quad (3)$$

$$[0077] \quad \delta_i^{hybgeo} = \min_{j: \rho_j^{hybgeo} > \rho_i^{hybgeo}} (d_{hybgeo}(\mathbf{x}_i, \mathbf{x}_j)) \quad (4)$$

[0078] 步骤5:人工观察选取不同的截止距离 $d_c^{geo}$ 对应的类簇中心,对于一个数据点,与其距离小于截止距离 $d_c^{geo}$ 的数据点越多,局部密度越大,因此类簇中心点的局部密度为最大值或局部最大值,此外,类簇中心点与比其局部密度大的数据点的最小距离为最大值或局部最大值,如图2所示为本发明的层次密度峰值聚类人工观察步骤示意图,横轴为局部密度,纵轴为最小距离,“\*”标记的两个数据点的局部密度和最小距离均比其余数据点大且具有离群特性,通过人工观察被选取为类簇中心,将剩余非中心点划入比其局部密度高的数据点中距离最近数据点的类簇中。其它截止距离 $d_c^{geo}$ 对应的密度峰值聚类结果也可按此过程获取,最终得到层次密度峰值聚类结果,即青霉素发酵过程通过层次密度峰值聚类可划分为2-6个模态( $P=[0.1, 0.2, 0.25, 0.3, 1.0]$ )。

[0079] 步骤6:根据模态连续特性最小模态长度 $t_{min}=3$ 调整初始层次模态信息,将小于 $t_{min}$ 的模态数据点调整到最近的模态,得到层次模态信息,对实际多模态工业过程,模态先验知识是不完备的且具有不确定性,层次模态信息可以从多模态工业过程数据中提取更多可用于过程建模的信息,并改善故障监测方法对多模态不确定性的适应性;

[0080] 步骤7:计算协方差矩阵 $C(X)$ ,利用SVD分解 $C(X)=U\Lambda U^T$ 计算协方差矩阵 $C(X)$ 的特征值对角矩阵 $\Lambda$ 和特征向量矩阵 $U$ ;

[0081] 步骤8:计算白化矩阵 $Q=\Lambda^{-1/2}U^T$ ,得到白化变换矩阵 $Z=QX$ ;

[0082] 步骤9:计算正交矩阵 $B$ ,过程如下:

[0083] 1) 选择估计独立元的个数 $d$ (一般保留累计贡献率为0.85-0.95的独立元),设置迭代次数 $i\leftarrow 1$ ;

[0084] 2) 以单位模向量给随机向量 $b_i$ 赋初值;

[0085] 3) 令 $b_i\leftarrow E\{zg(b_i^T z)\}-E\{g'(b_i^T z)\}b_i$ ,其中 $g$ 和 $g'$ 分别为 $G=\tanh(\cdot)$ 的一阶导数和二阶导数;

[0086] 4) 通过正交化来去除相关性,  $\mathbf{b}_i \leftarrow \mathbf{b}_i - \sum_{j=1}^{i-1} (\mathbf{b}_i^T \mathbf{b}_j) \mathbf{b}_j$ ;

[0087] 5) 归一化,  $\mathbf{b}_i \leftarrow \frac{\mathbf{b}_i}{\|\mathbf{b}_i\|}$ ;

[0088] 6) 假如 $b_i$ 没有收敛,返回步骤(3);

[0089] 7) 如果 $b_i$ 收敛,输出向量 $b_i$ ;

[0090] 8) 若 $i\leq m$ ,则设定 $i\leftarrow i+1$ 且返回步骤(2)。

[0091] 步骤10:计算预测的独立元矩阵 $S=B^T Z$ 和低维解混矩阵 $W=B^T Q$ ,故障监测 $I^2=s^T s$ 统计量和 $SPE=(\mathbf{x}-\hat{\mathbf{x}})^T(\mathbf{x}-\hat{\mathbf{x}})$ 统计量,其中 $\hat{\mathbf{x}}=W\mathbf{x}$ , $\mathbf{x}$ 表示 $X=\{x_1, x_2, \dots, x_n\}$ 的一个数据点;

[0092] 步骤11:置信水平 $\alpha = 1\%$ ,计算全局故障监测置信限 $I_\alpha^2 = \frac{d(N^2-1)}{N(N-d)} F_{d,N-d,\alpha}$ ,计算

局部故障监测置信限 $Q_{mi,\alpha} = g_{mi} \chi_{h_{mi},\alpha}^2$ ,其中 $m_i$ 为层次模态信息, $g_{mi} = S_{mi} / 2\mu_{mi}$ ,  
 $h_{mi} = 2\mu_{mi}^2 / S_{mi}$ , $\mu_{mi}$ 和 $S_{mi}$ 为层次模态信息为 $m_i$ 的SPE统计量的均值和方差,不同的层次模态对  
 应相同的全局故障监测置信限和不同的局部故障监测置信限,考虑了多模态工业过程的全  
 局相似性和局部非相似性;

[0093] 步骤11:得到待监测的标准化后工业过程数据 $x_{new}$ 的最相似模态,公式如下:

$$[0094] \quad \max_{mi} df(\mathbf{x}_{new}) = \rho_{new}^{mi} - d_{new,center}^{mi} \quad (5)$$

[0095] 其中, $\rho_{new}^{mi}$ 和 $d_{new,center}^{mi}$ 为待监测的工业过程数据在层次模态信息中各个模态的局部  
 密度和与其类簇中心的距离,基于静态模态信息的故障监测可能会影响性能,动态确定待  
 监测的工业过程数据的最相似模态能改善故障监测方法对多模态动态性的适应性;

[0096] 步骤12:得到 $x_{new}$ 的统计量 $I^2 = \mathbf{s}_{new}^T \mathbf{s}_{new}$ 和 $SPE = (\mathbf{x}_{new} - \hat{\mathbf{x}}_{new})^T (\mathbf{x}_{new} - \hat{\mathbf{x}}_{new})$ ;

[0097] 步骤13:如果 $x_{new}$ 在最相似模态的 $I^2$ 或SPE统计量大于对应的故障监测置信限,则  
 判定工业过程发生故障,否则为无故障。

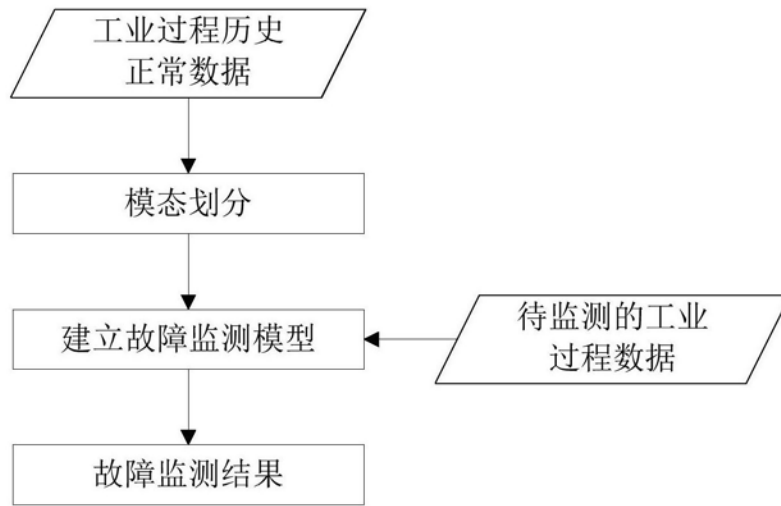


图1

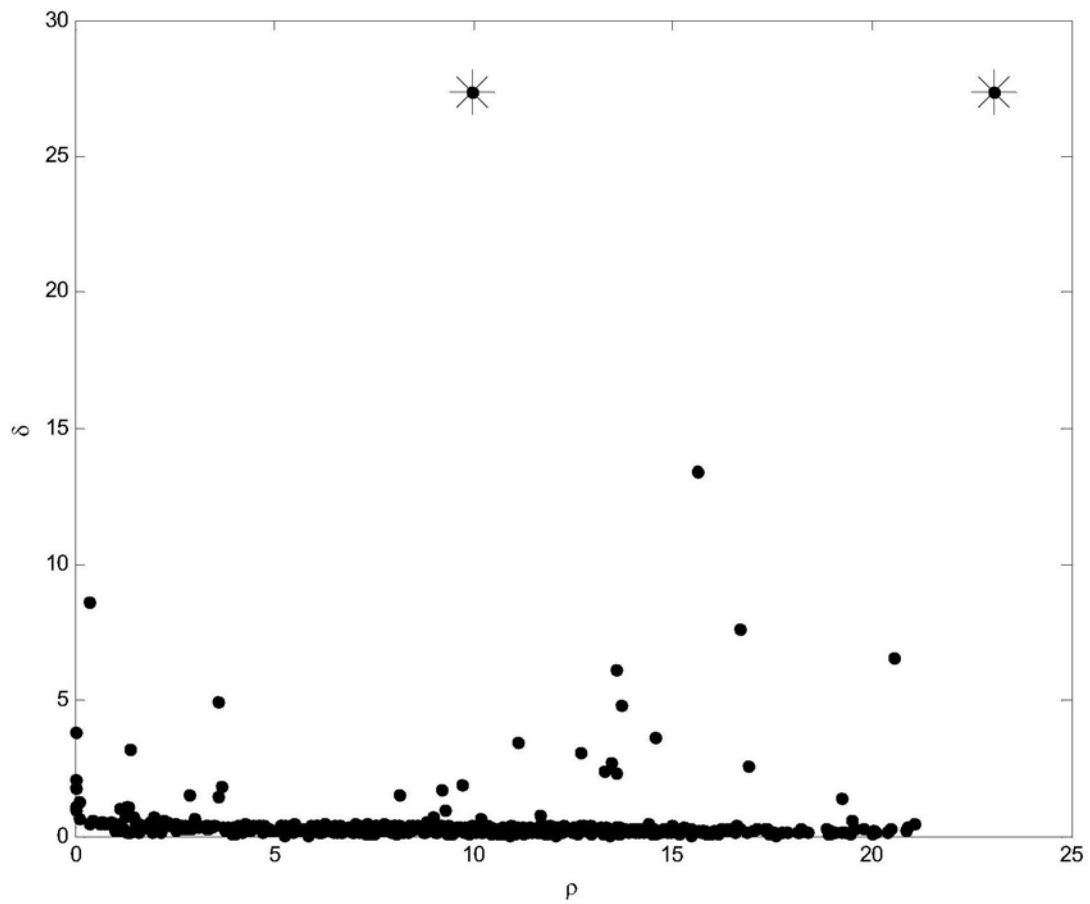


图2