



(12)发明专利

(10)授权公告号 CN 104615123 B

(45)授权公告日 2017.02.22

(21)申请号 201410810167.4

(22)申请日 2014.12.23

(65)同一申请的已公布的文献号

申请公布号 CN 104615123 A

(43)申请公布日 2015.05.13

(73)专利权人 浙江大学

地址 310027 浙江省杭州市西湖区浙大路38号

(72)发明人 杨春节 周哲 文成林

(74)专利代理机构 杭州求是专利事务所有限公司 33200

代理人 林松海

(51)Int.Cl.

G05B 23/02(2006.01)

(56)对比文件

- CN 103488561 A, 2014.01.01,
- CN 103576594 A, 2014.02.12,
- US 2012220875 A1, 2012.08.30,
- JP 2013152655 A, 2013.08.08,
- CN 104199441 A, 2014.12.10,
- ZHOU Zhe 等. Random projection based k

Nearest Neighbor rule for semiconductor process fault detection.《Proceedings of the 33rd Chinese Control Conference》.2014,第3169-3174页.

陈海彬 等.基于改进K_means聚类的kNN故障检测研究.《沈阳化工大学学报》.2013,第27卷(第1期),第69-73页.

张成 等.PC_KNN故障检测方法在半导体批次过程中的应用研究.《25th Chinese Control and Decision Conference(CCDC)》.2013,第4209-4214页.

周军伟 等.基于KNN分类算法的水下航行器声学故障检测识别研究.《噪声与振动控制》.2009,(第2期),第59-61,102页.

袁杰 等.基于KPC_kNN方法的批次过程故障诊断.《沈阳化工大学学报》.2014,第28卷(第2期),第170-174页.

徐涛 等.基于k均值聚类与K近邻的故障检测方法研究.《通化师范学院学报》.2013,第34卷(第3期),第38-40页.

审查员 张碧芸

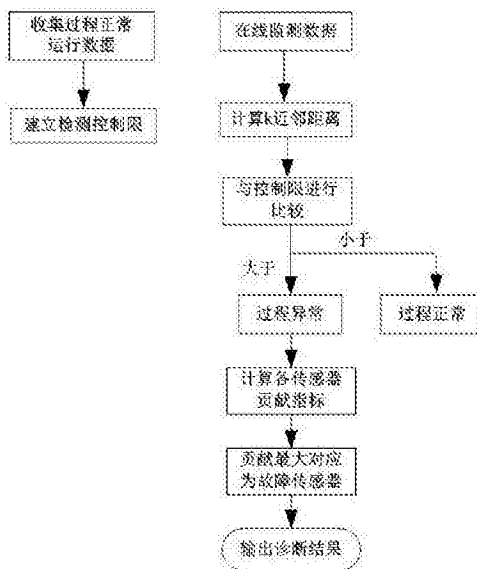
权利要求书2页 说明书5页 附图3页

(54)发明名称

基于k近邻的传感器故障隔离方法

(57)摘要

本发明公开了基于k近邻的传感器故障隔离方法,属于工业过程监控与诊断技术领域。本发明只需要正常数据用于建模,当在线测量数据显示过程异常时,首先,在正常工况数据集中找与该在线测量数据最近的k个数据样本;然后,分别计算在线测量数据的每个分量与k近邻样本的累积距离,将累积距离作为指标用于隔离;最后,指标超过设定阈值的传感器判定为故障传感器。相比于现有的其他方法,本发明方法可以保证正确地隔离传感器故障。



1.一种基于k近邻的传感器故障隔离方法,其特征在于该方法包括以下各步骤:

步骤一、故障检测:首先进行离线建模,即收集过程正常运行工况的数据,根据正常数据之间的k近邻距离建立监控控制限;然后在线检测,对于在线测量数据,计算在线数据与正常数据集的k近邻距离,然后与所建立的控制限比较,判断是否发生异常;

步骤二、故障隔离:对于步骤一判定为异常的在线测量数据样本,进一步辨识造成异常的故障传感器,将该在线测量数据样本与近邻的累积距离进行分解,并将每一个分量作为隔离指标用于辨识故障传感器,最大指标对应的传感器为故障传感器;

步骤一所述的故障检测过程具体如下:

1)利用多传感器数据采集系统收集过程正常工况运行下的监测数据构成数组 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$,其中,m表示过程监控变量的个数,n表示所收集正常工况下样本个数, $\mathbf{x}_i \in \mathbb{R}^m, i = 1, \dots, n$ 表示第i个正常样本,每个过程监控变量的样本方差为 σ_i^2 ;

2)计算每个训练样本的k近邻距离,并确定检测控制限

2.1)从数据集X中找每个样本 \mathbf{x}_i 的k个近邻

$$d_{i,j} = \|\mathbf{x}_i - \mathbf{x}_j\|_2, j = 1, \dots, n, j \neq i \quad (1)$$

其中, $\|\cdot\|_2$ 表示 l_2 范数,即欧式距离;

2.2)计算每个样本 \mathbf{x}_i 与其k近邻之间的平均距离

$$D_i^2 = \frac{1}{k} \sum_{j=1}^k d_{i,j}^2, i = 1, \dots, n \quad (2);$$

2.3)确定检测控制限 D_α^2

$$D_\alpha^2 = D_{\lfloor n(1-\alpha) \rfloor}^2 \quad (3)$$

其中, $D_{(i)}^2, i = 1, \dots, n$ 是对公式(2)中 $D_i^2, i = 1, \dots, n$ 按降序重新排列的序列, $\lfloor n(1-\alpha) \rfloor$ 表示取 $n(1-\alpha)$ 的整数部分,公式(3)中 $i = \lfloor n(1-\alpha) \rfloor, \alpha$ 表示置信水平;

3)对于在线采集测量数据 $\mathbf{y} \in \mathbb{R}^m$

3.1)根据式(1)从数据集X中找y的k近邻;

3.2)根据式(2)计算y与k近邻之间的平均距离 D_y^2 ;

3.3)比较 D_y^2 与式(3)中所建立的检测控制限 D_α^2 之间的大小,如果 $D_y^2 > D_\alpha^2$,则说明过程异常并报警;如果 $D_y^2 < D_\alpha^2$,则说明过程正常运行;

步骤二所述的故障隔离过程具体如下:

1)在线采集测量数据 $\mathbf{y} \in \mathbb{R}^m$,它的k近邻累积距离 D_y^2 为

$$D_y^2 = \sum_{i=1}^m \sum_{j=1}^k [\xi_i^T (\mathbf{y} - \mathbf{x}_j)]^2 \quad (4)$$

其中, ξ_i^T 是 ξ_i 的矩阵转置, ξ_i 表示单位矩阵的i列;

2)将y的k近邻累积距离 D_y^2 分解m个分量之和,第i个分量为

$$c_i^2 = \sum_{j=1}^k [\xi_j^T (\mathbf{y} - \mathbf{x}_j)]^2, \quad i = 1, \dots, m \quad (5);$$

3) 计算每个分量贡献的阈值 γ_i

$$\gamma_i = 2\sigma_i^2 \cdot \chi_{\alpha}^2(k), \quad i = 1, \dots, m \quad (6);$$

4) 辨识故障传感器, 贡献量 c_i^2 超过设定阈值 γ_i 的传感器为故障传感器

$$F = \{i \mid c_i^2 > \gamma_i\} \quad (7);$$

根据式(7)的结果, 判定集合F包含的所有传感器为故障传感器。

基于k近邻的传感器故障隔离方法

技术领域

[0001] 本发明属于工业过程监控与故障诊断领域,特别涉及基于k近邻的传感器故障隔离方法。

背景技术

[0002] 对于过程监控和故障诊断问题,在传统的多元统计过程监控(Multivariable Statistical Process Monitoring, MSPM)方法的框架下,主要采用贡献图(Contribution Plot, CP)和基于重构的贡献(Reconstruction-based Contribution, RBC)方法进行故障隔离。然而,这两种方法都存在故障干扰现象,即故障传感器可能会影响非故障传感器对应的贡献指标,从而导致误隔离。另外,这两个主要故障隔离方法都无法隔离多传感器故障。本发明方法将解决这些问题。

发明内容

[0003] 本发明的目的在针对现有技术的不足,提供一种基于k近邻的传感器故障隔离方法,在原始测量空间根据样本与近邻之间累计距离定义故障贡献指标,从而避免故障干扰问题,实现故障传感器准确地隔离。

[0004] 本发明提出的基于k近邻的传感器故障隔离方法,包括以下各步骤:

[0005] 步骤一:故障检测。收集过程正常运行工况的数据,根据正常数据之间的k近邻距离建立监控控制限;对于在线测量数据,计算在线数据与正常数据集的k近邻距离,然后与所建立的控制限比较,判断是否发生异常。

[0006] 步骤二:故障隔离。对于步骤一判定为异常的在线测量数据样本,进一步辨识造成异常的故障传感器。将该在线测量数据样本与近邻的累积距离进行分解,并将每一个分量作为隔离指标用于辨识故障传感器,最大指标对应的传感器为故障传感器。

[0007] 步骤一所述的故障检测过程如下:

[0008] 1)利用多传感器数据采集系统收集过程正常工况运行下的监测数据构成数组 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$, 其中, m 表示过程传感器(监控变量)的个数, n 表示所收集正常工况下样本个数, $\mathbf{x}_i \in \mathbb{R}^m, i = 1, \dots, n$ 表示第 i 个正常样本;

[0009] 2)计算每个训练样本的k近邻距离,并确定检测控制限

[0010] 2.1)从数据集 X 中找每个样本 \mathbf{x}_i 的 k 个近邻

[0011] $d_{i,j} = \|\mathbf{x}_i - \mathbf{x}_j\|_2, j = 1, \dots, n, j \neq i$ (1)

[0012] 其中, $\|\cdot\|_2$ 表示 l_2 范数,即欧式距离;

[0013] 2.2)计算每个样本 \mathbf{x}_i 与其 k 近邻之间的平均距离

[0014]
$$D_i^2 = \frac{1}{k} \sum_{j=1}^k d_{ij}^2, i = 1, \dots, n \quad (2)$$

[0015] 2.3)确定检测控制限 D_α^2

$$[0016] \quad \mathcal{D}_\alpha^2 = \mathcal{D}_{\lfloor n(1-\alpha) \rfloor}^2 \quad (3)$$

[0017] 其中, $\mathcal{D}_{(i)}^2, i=1, \dots, n$ 是对 $\mathcal{D}_i^2, i=1, \dots, n$ 按降序重新排列的序列。 $\lfloor n(1-\alpha) \rfloor$ 表示取 $n(1-\alpha)$ 的整数部分, α 表示置信水平。

[0018] 3) 对于在线采集测量数据 $\mathbf{y} \in \mathbb{R}^m$

[0019] 3.1) 根据式(1)从数据集 X 中找 y 的 k 近邻;

[0020] 3.2) 根据式(2)计算 y 与 k 近邻之间的平均距离 \mathcal{D}_y^2 ;

[0021] 3.3) 比较 \mathcal{D}_y^2 与式(3)中所建立的检测控制限 \mathcal{D}_α^2 之间的大小, 如果 $\mathcal{D}_y^2 > \mathcal{D}_\alpha^2$, 则说明过程异常并报警; 如果 $\mathcal{D}_y^2 < \mathcal{D}_\alpha^2$, 则说明过程正常运行。

[0022] 步骤二所述的故障隔离过程如下:

[0023] 1) 在线采集测量数据 $\mathbf{y} \in \mathbb{R}^m$, 它的 k 近邻累积距离 D_y^2 为

$$[0024] \quad \mathcal{D}_y^2 = \sum_{i=1}^m \sum_{j=1}^k [\xi_i^T (\mathbf{y} - \mathbf{x}_j)]^2 \quad (4)$$

[0025] 其中, ξ_i^T 是 ξ_i 的矩阵转置, ξ_i 表示单位矩阵的 i 列。

[0026] 2) 将 y 的 k 近邻累积距离 \mathcal{D}_y^2 分解 m 个分量之和, 第 i 个分量为

$$[0027] \quad c_i^2 = \sum_{j=1}^k [\xi_i^T (\mathbf{y} - \mathbf{x}_j)]^2, \quad i=1, \dots, m \quad (5)$$

[0028] 9) 计算每个分量贡献的阈值 γ_i

$$[0029] \quad \gamma_i = 2\sigma_i^2 \cdot \chi_\alpha^2(k), \quad i=1, \dots, m \quad (6)$$

[0030] 10) 辨识故障传感器。贡献量 c_i^2 超过设定阈值 γ_i 的传感器为故障传感器

$$[0031] \quad F = \{i | c_i^2 > \gamma_i\} \quad (7)$$

[0032] 根据式(7)的结果, 判定集合 F 包含的所有传感器为故障传感器。

附图说明

[0033] 图1. 是本发明方法的流程框图;

[0034] 图2. 是基于 k 近邻的变量贡献;

[0035] 图3. 是 SPE 贡献图;

[0036] 图4. 是 T2 贡献图;

[0037] 图5. 是 SPE 重构贡献图;

[0038] 图6. 是 T2 重构贡献图。

具体实施方式

[0039] 本发明提出的基于 k 近邻的传感器故障隔离方法, 其流程框图如图1所示, 包括以下各步骤:

[0040] 步骤一: 故障检测。首先进行离线建模, 即收集过程正常运行工况的数据, 根据正

常数据之间的k近邻距离建立监控控制限；然后在线检测，对于在线测量数据，计算在线数据与正常数据集的k近邻距离，然后与所建立的控制限比较，判断是否发生异常。

[0041] 步骤二：故障隔离。对于步骤一判定为异常的在线测量数据样本，进一步辨识造成异常的故障传感器。将该在线测量数据样本与近邻的累积距离进行分解，并将每一个分量作为隔离指标用于辨识故障传感器，最大指标对应的传感器为故障传感器。

[0042] 其中，步骤一所述的故障检测过程如下：

[0043] 1)利用多传感器数据采集系统收集过程正常工况运行下的监测数据构成数组 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ ，其中，m表示过程传感器(监控变量)的个数，n表示所收集正常工况下样本个数， $\mathbf{x}_i \in \mathbb{R}^m$ ， $i = 1, \dots, n$ 表示第i个正常样本；

[0044] 2)计算每个训练样本的k近邻距离，并确定检测控制限

[0045] 2.1)从数据集X中找每个样本 \mathbf{x}_i 的k个近邻

[0046] $d_{i,j} = \|\mathbf{x}_i - \mathbf{x}_j\|_2, j = 1, \dots, n, j \neq i$ (1)

[0047] 其中， $\|\cdot\|_2$ 表示 l_2 范数，即欧式距离；

[0048] 2.2)计算每个样本 \mathbf{x}_i 与其k近邻之间的平均距离

[0049]
$$\mathcal{D}_i^2 = \frac{1}{k} \sum_{j=1}^k d_{ij}^2, i = 1, \dots, n$$
 (2)

[0050] 2.3)确定检测控制限 \mathcal{D}_α^2

[0051]
$$\mathcal{D}_\alpha^2 = \mathcal{D}_{\lfloor n(1-\alpha) \rfloor}^2$$
 (3)

[0052] 其中， $\mathcal{D}_{(i)}^2, i = 1, \dots, n$ 是对公式(2)中 $\mathcal{D}_i^2, i = 1, \dots, n$ 按降序重新排列的序列， $\lfloor n(1-\alpha) \rfloor$ 表示取 $n(1-\alpha)$ 的整数部分，公式(3)中 $i = \lfloor n(1-\alpha) \rfloor$ ， α 表示置信水平。

[0053] 3)对于在线采集测量数据 $\mathbf{y} \in \mathbb{R}^m$

[0054] 3.1)根据式(1)从数据集X中找y的k近邻；

[0055] 3.2)根据式(2)计算y与k近邻之间的平均距离 \mathcal{D}_y^2 ；

[0056] 3.3)比较 \mathcal{D}_y^2 与式(3)中所建立的检测控制限 \mathcal{D}_α^2 之间的大小，如果 $\mathcal{D}_y^2 > \mathcal{D}_\alpha^2$ ，则说明过程异常并报警；如果 $\mathcal{D}_y^2 < \mathcal{D}_\alpha^2$ ，则说明过程正常运行。

[0057] 步骤二所述的故障隔离过程如下：

[0058] 1)在线采集测量数据 $\mathbf{y} \in \mathbb{R}^m$ ，它的k近邻累积距离 D_y^2 为

[0059]
$$D_y^2 = \sum_{i=1}^m \sum_{j=1}^k [\xi_i^T (\mathbf{y} - \mathbf{x}_j)]^2$$
 (4)

[0060] 其中， ξ_i^T 是 ξ_i 的矩阵转置， ξ_i 表示单位矩阵的i列；

[0061] 2)将y的k近邻累积距离 D_y^2 分解m个分量之和，第i个分量为

[0062]
$$c_i^2 = \sum_{j=1}^k [\xi_i^T (\mathbf{y} - \mathbf{x}_j)]^2, i = 1, \dots, m$$
 (5)

[0063] 3)计算每个分量贡献的阈值 γ_i

$$[0064] \quad \gamma_i = 2\sigma_i^2 \cdot \chi_{\alpha}^2(k), \quad i = 1, \dots, m \quad (6)$$

[0065] 4) 辨识故障传感器。贡献量 c_i^2 超过设定阈值 γ_i 的传感器为故障传感器

$$[0066] \quad F = \{i | c_i^2 > \gamma_i\} \quad (7)$$

[0067] 根据式(7)的结果, 判定集合F包含的所有传感器为故障传感器。

[0068] 实施例

[0069] 下面结合田纳西-伊斯曼过程数据来说明本发明方法的有效性。该标准测试实验平台由Downs和Vogel根据伊斯曼化学公司的一个实际化工联合反应过程所开发, 其过程数据具有复杂非线性、强耦合和时变等特性, 为监控方法提供了一个真实的工业过程。该过程由连续搅拌式反应釜、分凝器、汽液分离塔、离心式压缩机和汽提塔五个操作单元组成, 包含了八种成分:A, B, C, D, E, F, G和H。其中, A, C, D, E四种气体进料成分和惰性气体成分B一起作为反应物, 形成产物G和H, 以及副产物F。田纳西-伊斯曼过程共有41个测量变量和12个控制变量, 这里选取16个过程变量作为监控变量, 如表1所示。正常训练样本数为960个, 这里选取第10种故障为例说明本发明方法, 与故障10直接相关的测量变量包括变量13和变量14。

[0070] 表1: 监控变量说明

变量序号	变量名称	变量序号	变量名称
1	物料 A 流量	9	分离塔温度
2	物料 D 流量	10	分离塔压力
3	物料 E 流量	11	分离塔底部流量
[0071] 4	总物料流量	12	汽提塔压力
5	循环流量	13	汽提塔底部流量
6	反应器进料流速	14	汽提塔温度
7	反应器温度	15	反应器冷却水出口温度
8	排空速率	16	分离塔冷却水出口温度

[0072] 接下来结合该具体过程对本发明的实施步骤进行详细地阐述:

[0073] 步骤一所述的故障检测过程如下:

[0074] 1) 利用多传感器数据采集系统收集过程正常工况运行下的监测数据构成数组 $\mathbf{X} \in \mathbb{R}^{16 \times 960}$;

[0075] 2) 计算每个训练样本的k近邻距离, 并确定检测控制限。本例中, 近邻数 $k=50$, 置信水平 $\alpha=0.95$, 计算得到检测控制限 $D_{\alpha}^2 = 18.6459$;

[0076] 3) 对于在线采集测量数据 $\mathbf{y} \in \mathbb{R}^{16}$

[0077] 3.1) 根据式(1)从数据集X中找y的k近邻;

[0078] 3.2) 根据式(2)计算y与k近邻之间的平均距离 D_y^2 ;

[0079] 3.3) 比较 D_y^2 与式(3)中所建立的检测控制限 D_{α}^2 之间的大小。

[0080] 本例中, $D_y^2 > D_{\alpha}^2$, 说明过程异常并报警;

[0081] 步骤二, 检测出故障后, 进行故障隔离的过程如下:

[0082] 4)计算测量数据 y 的 k 近邻累积距离 D_y^2

[0083] 5)将 y 的 k 近邻累积距离 D_y^2 分解 m 个分量之和,每个分量为 c_i^2 ;

[0084] 6)计算每个分量贡献的阈值 γ_i ;

[0085] 7)辨识故障传感器。贡献量 c_i^2 超过设定阈值 γ_i 的传感器为故障传感器

$$F = \{i | c_i^2 > \gamma_i\}。$$

[0086] 本实施例中,集合 F 中包含变量13和变量14。基于 k 近邻的变量贡献如图2所示,变量13和变量14贡献量明显大于其他变量,因而可以正确地判定变量13和变量14为故障变量。图3和图4分别是传统的SPE和 T^2 贡献图,从图中可以看出,均不能正确地隔离真实的故障变量。图5和图6分别是SPE和 T^2 重构贡献图,从图中可以看出SPE重构贡献图无法正确地隔离真实的故障变量,尽管 T^2 重构贡献图中变量13和变量14的贡献最大,但是受到故障干扰的影响使得变量1、7、11和12的贡献量也较大,区分度不明显。说明了本发明方法具有较好的故障隔离性能。

[0087] 上述实施例用来解释说明本发明,而不是对本发明进行限制,在本发明的精神和权利要求的保护范围内,对本发明做出的任何修改和改变,都落入本发明的保护范围。

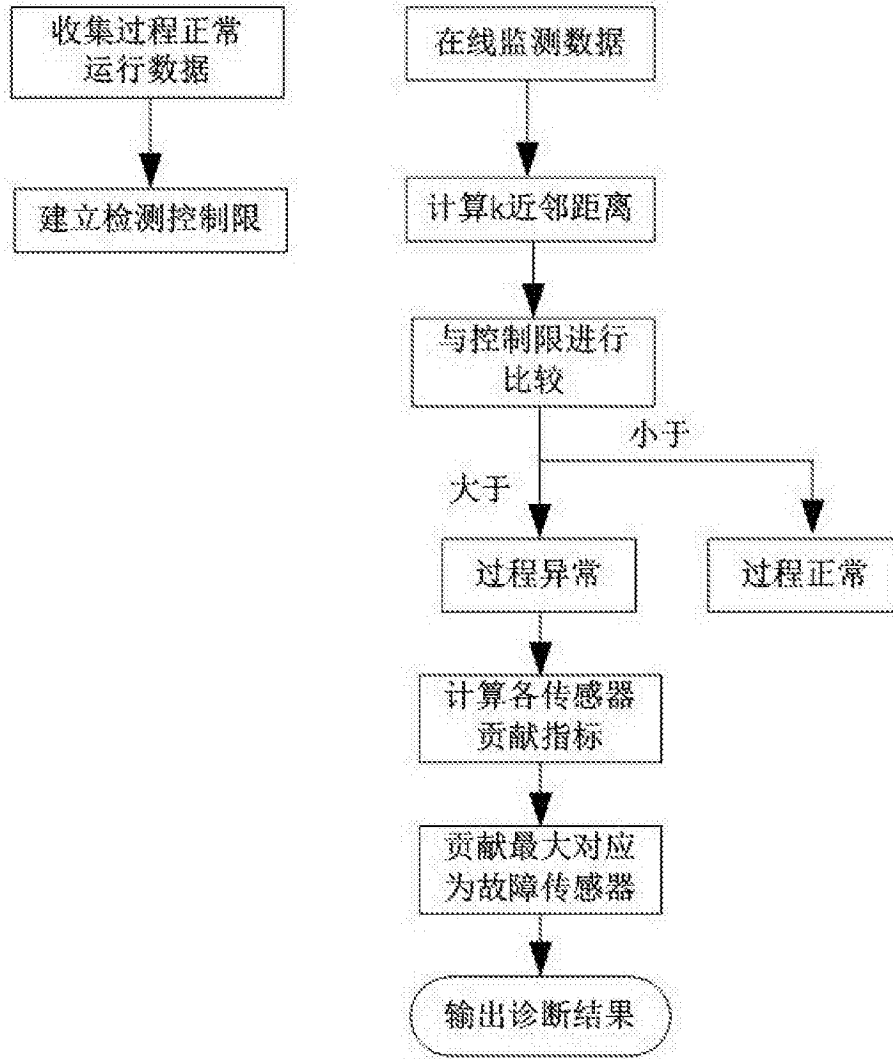


图1

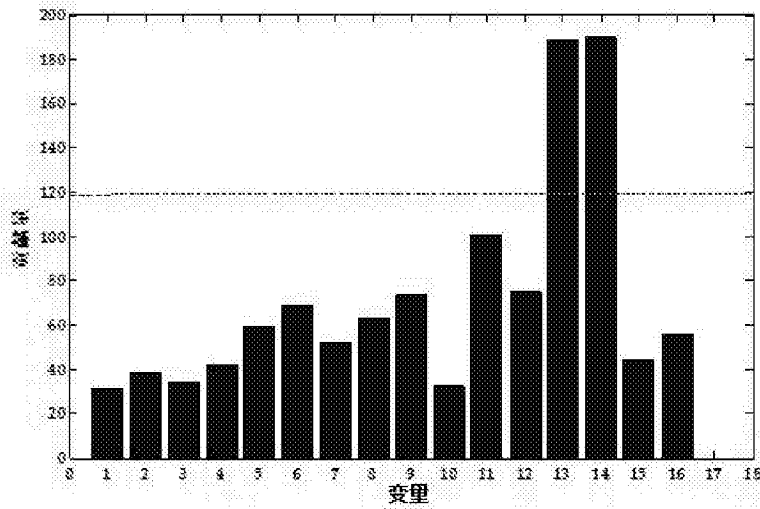


图2

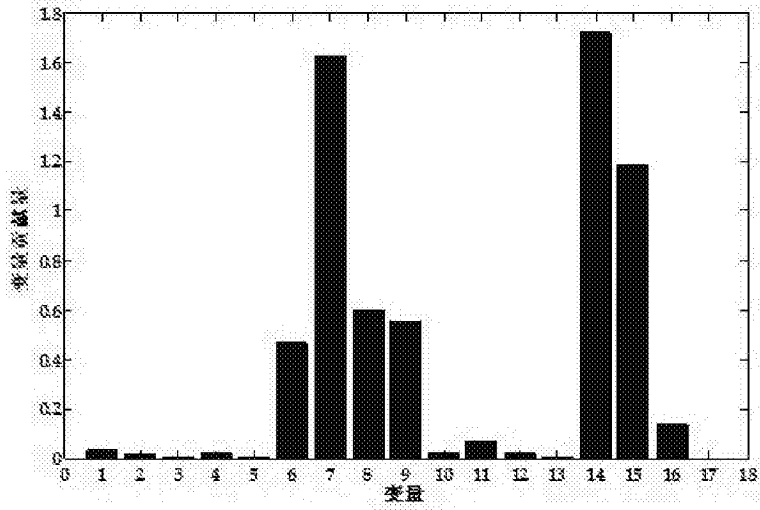


图3

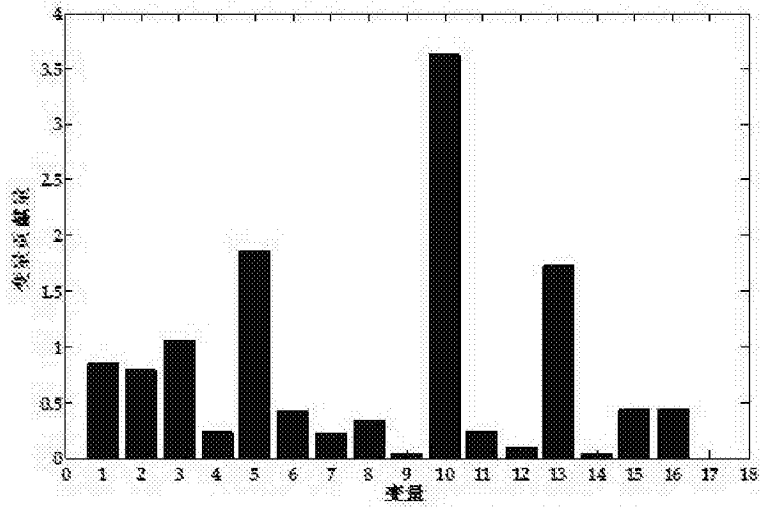


图4

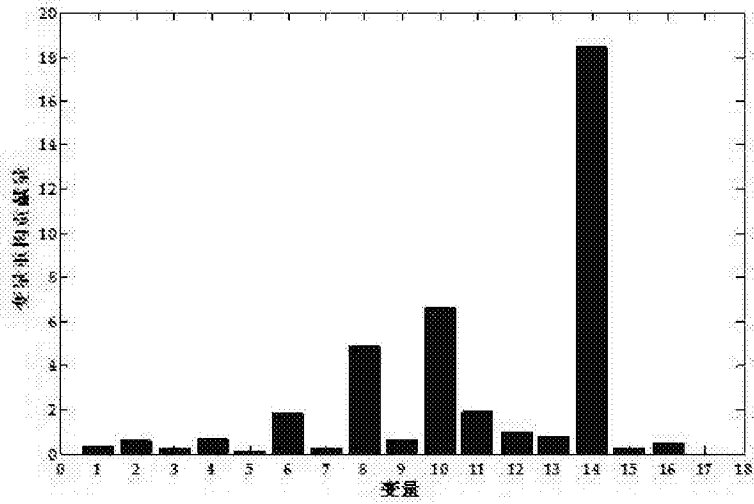


图5

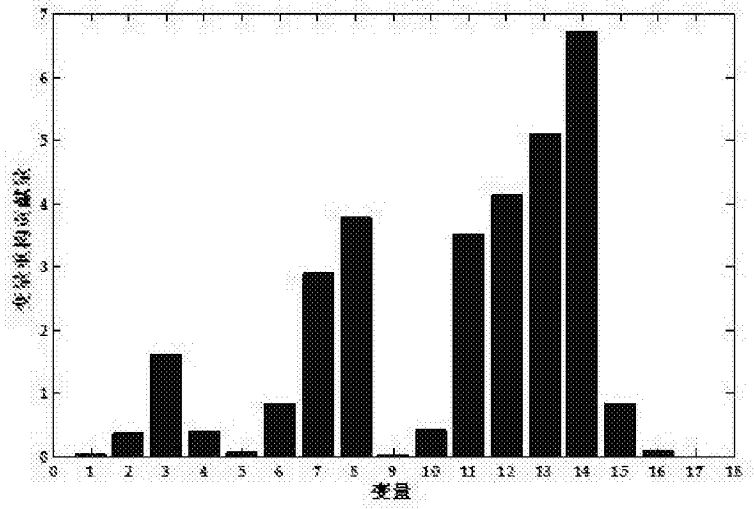


图6