

研究论文

基于多模型外部分分析和 Greedy-KP1M 的多工况过程监控

王晓阳¹, 王 昕^{2,3}, 王振雷¹, 钱 锋¹

(¹ 华东理工大学化工过程先进控制和优化技术教育部重点实验室, 上海 200237;

² 上海交通大学电工与电子技术中心, 上海 200240;

³ 东北大学流程工业综合自动化国家重点实验室, 辽宁 沈阳 110004)

摘要: 传统的基于多元统计过程监控方法都是假设过程处于单一工况下, 而随着进料负荷、产品组分等过程参数的改变, 生产过程的工况也随之改变, 传统方法便不再适用。针对工业过程中的多工况监控问题, 提出了一种基于多模型外部分分析和 Greedy-KP1M 的多工况过程监控方法。首先针对传统外部分分析方法描述能力不足的问题, 用多模型局部建模代替单一模型来获得更好的描述能力, 同时获得监控残差, 通过对残差进行监控从而去除多工况的影响, 进而将核单簇可能性聚类 (KP1M) 用于对残差的监控上。该方法拥有和支持向量数据描述 (SVDD) 相当的监控效果, 但计算复杂度却远远小于 SVDD。同时, 采用 Greedy 方法提取特征样本, 进一步降低了算法计算复杂度。最后将上述方法应用在 TE 模型和乙烯裂解炉的监控上, 结果证明了该方法的有效性。

关键词: 多工况; 多模型外部分分析; Greedy 特征样本提取; 核单簇可能性聚类

DOI: 10.3969/j.issn.0438-1157.2012.09.031

中图分类号: TP 277

文献标志码: A

文章编号: 0438-1157 (2012) 09-2869-08

Multiple models external analysis and Greedy-KP1M based process monitoring with multiple operation modes

WANG Xiaoyang¹, WANG Xin^{2,3}, WANG Zhenlei¹, QIAN Feng¹

(¹ Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China;

² Center of Electrical & Electronic Technology, Shanghai Jiao Tong University, Shanghai 200240, China;

³ State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110004, Liaoning, China)

Abstract: Multivariable statistical process control (MSPC) is developed in order to extract useful information from process data and utilize them for process monitoring. But when the change in process feed load or product composition happens, the conventional MSPC method does not function well for the process with multiple operation modes. In order to solve these problems, a novel process monitoring method

2012-06-14 收到初稿, 2012-06-21 收到修改稿。

联系人: 王振雷。第一作者: 王晓阳 (1988—), 男, 硕士研究生。

基金项目: 国家自然科学基金项目 (61134007, 61174118); 国家高技术研究发展计划项目 (2012AA040307); 上海市基础研究重点项目 (10JC1403500); 上海市重点学科建设项目 (B504); 流程工业综合自动化国家重点实验室开放课题基金项目。

Received date: 2012-06-14.

Corresponding author: Prof. WANG Zhenlei, wangzhen_l@ecust.edu.cn

Foundation items: supported by the National Natural Science Foundation of China (61134007, 61174118), the High-tech Research and Development Program of China (2012AA040307), the Major State Basic Research Development Program of Shanghai (10JC1403500), the Shanghai Leading Academic Discipline Project (B504) and the State Key Laboratory of Synthetical Automation for Process Industries.

is proposed based on multiple model external analysis and Greedy-KP1M. First, based on the traditional external analysis, multiple models modeling method is introduced to have a better performance. The multiple operation modes of process are eliminated by multiple models external analysis, and the residual error is got for monitoring. Then, to monitor the residual error, the method called Kernel Possibilistic one-Mean clustering (KP1M) is proposed. KP1M has a good ability to monitor nonlinear process. Its performance is similar with support vector data description (SVDD). But the computation complexity of KP1M is far less than SVDD's. Moreover, to reduce the computation complexity furthermore, Greedy method is adopted to extract the feature samples for KP1M modeling. In the end, the proposed method is applied to monitor the TE (Tennessee Eastman) process and the ethylene cracking furnace to show its efficiency.

Key words: multiple operation modes; multiple models external analysis; Greedy feature samples extraction; KP1M

引 言

现代化学工业过程生产规模逐年扩大,设备复杂度越来越高,整个生产过程和整个社会越来越重视化工生产的安全,一旦过程的某个环节发生故障,不仅会造成财产和人员的损失,而且会对周围的生活环境造成巨大的破坏,因此有必要对生产装置进行过程监控,以便及时发现故障并采取相应措施^[1]。近十几年,经典多元统计过程监控方法得到了广泛的应用,但是这些方法大都假设整个过程处于单一工况,而实际工业过程的工况会随着进料负荷、产品组分等的改变而变化。因此,将传统的多元统计过程监控方法应用到实际工业过程时误报率高,得不到很好的监控效果。

目前已有部分学者对多工况条件下过程监控进行了研究。文献[2]提出了一种鲁棒聚类方法,该方法假设每种工况条件下的数据具有相同的协方差矩阵;文献[3]提出了一种多层聚类 Super-PCA 的方法,该方法假设所有的聚类都有相同的特征向量。虽然基于聚类的方法能在一定程度上完成工况的划分,但其往往需要过程数据满足分布上的一些假设。而在实际过程中,数据难以满足上述假设条件,聚类的最终效果不能得到很好的保证。为此,文献[4]提出了一种外部分析的方法,将影响工况变化的量作为外部变量,剩下的变量作为主要变量,通过回归的方法去除主要变量中外部变量的影响,从而达到去除多工况影响的目的。该方法无须假设过程数据分布,增加了过程监控方法的适应性。但该方法采用全局模型来描述外部变量对主要变量的影响,当过程比较复杂时,全局模型无

法准确刻画外部变量和主要变量之间的关系,以致对过程的监控效果大打折扣。

在对过程进行了外部分析后,建立残差序列的监控模型实现过程监控。支持向量数据描述(SVDD)^[5]是一种非线性处理能力很强的单分类算法,近年来被用到过程监控上,获得了很好的效果^[6-8]。但是 SVDD 算法最终转化为一个二次规划问题,计算复杂度为 $O(N^3)$,因此大规模的样本集可以引发核函数矩阵计算时的“维数灾难”。同时,SVDD 对噪声和野点比较敏感,在建模前必须对数据集进行处理。

为此,本文提出了基于多模型外部分析和 Greedy-KP1M 的多工况过程监控。并将该方法应用在乙烯裂解炉监控上,以验证该方法的有效性。

1 基于多模型局部建模的外部分析

1.1 外部分析

在对多工况过程进行监控时,常规方法往往无法区分正常工况改变和过程故障,从而产生误报,因此,必须对工况改变和过程故障进行区分。外部分分析方法按照过程先验知识将过程变量分为两部分:外部变量和主要变量。过程正常工况的变化(进料负荷的变化、产品组分的变化等)由外部变量引起,剩下的变量则是主要变量。主要变量的变化由外部变量和过程的扰动及过程故障引起。当主要变量发生变化时,去除外部变量的影响,对剩下的部分再进行监控,则避免了正常工况变化对监控的影响,从而获得较好的监控结果。外部分析如图 1 所示。

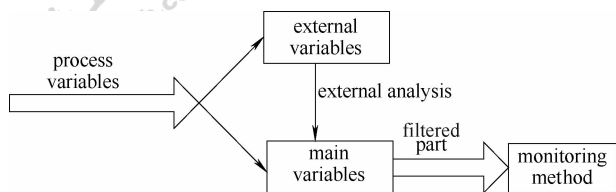


图1 外部分析

Fig. 1 External analysis

设原始数据矩阵 $\mathbf{X} \in R^{N \times r}$ ，其中 N 为数据样本个数， r 为每个样本的维数。其中，外部变量有 r_1 个，主要变量有 r_2 个，则 $\mathbf{X}$ 可分解为

$$\mathbf{X} = [\mathbf{U} \quad \mathbf{M}] \quad (1)$$

其中， $\mathbf{U} \in R^{N \times r_1}$ 为外部变量， $\mathbf{M} \in R^{N \times r_2}$ 为主要变量， $r_1 + r_2 = r$ 。 $\mathbf{M}$ 可以表示为

$$\mathbf{M} = f(\mathbf{U}) + \mathbf{\Xi} \quad (2)$$

其中， $f(\cdot)$ 表示外部变量和主要变量之间的关系； $\mathbf{\Xi}$ 为要监控的残差，是主要变量去除外部变量影响后的部分，主要受过程中的扰动和过程发生的故障影响，则残差可以表示为

$$\mathbf{\Xi} = \mathbf{M} - f(\mathbf{U}) \quad (3)$$

1.2 多模型局部建模

对于上述求取残差的过程，当过程比较复杂时，用单一全局模型无法准确描述外部变量和主要变量之间的关系。因此，本文将即时学习思想引入外部分析，用多模型局部建模方法来替换全局模型，以获得更加准确的系统描述。

即时学习^[9-11]多模型局部建模方法基于“相似输入产生相似输出”的原则，不建立全局模型，而是选取当前输入点附近的 K 组信息，构成学习集，并进行预测，可以获得很好的模型适应性和预测精度。本文将已存储的采样点看作向量，采用距离和角度要素作为近邻点的选择标准。

设 $\mathbf{u}_p = [x_{p1}, x_{p2}, \dots, x_{pr_1}]^T$ 为数据库中已存储的某一时刻的外部变量值， $\mathbf{u}_q = [x_{q1}, x_{q2}, \dots, x_{qr_2}]^T$ 为当前时刻的外部变量采样点， $\mathbf{u}_p$ 和 $\mathbf{u}_q$ 的欧氏距离和夹角定义如下

$$\begin{cases} d(\mathbf{u}_p, \mathbf{u}_q) = \sqrt{\|\mathbf{u}_p - \mathbf{u}_q\|_2} = \sqrt{\sum_{j=1}^{r_1} (x_{pj} - x_{qj})^2} \\ \theta(\mathbf{u}_p, \mathbf{u}_q) = \cos^{-1} \left(\frac{\mathbf{u}_p^T \mathbf{u}_q}{\|\mathbf{u}_p\|_2 \cdot \|\mathbf{u}_q\|_2} \right) \end{cases} \quad (4)$$

按照以下规则在系统历史数据库中遍历选择 K 组信息构造局部建模学习集。

(1) 当 $\cos\theta(\mathbf{u}_p, \mathbf{u}_q) < 0$ 时，则认为此信息向量与当前工作点方向相反，放弃选入建模学

习集。

(2) 当 $\cos\theta(\mathbf{u}_p, \mathbf{u}_q) \geq 0$ 时，则综合考虑向量距离和向量间的夹角，采用如下准则进行选择

$$D(\mathbf{u}_p, \mathbf{u}_q) = \alpha e^{-d(\mathbf{u}_p, \mathbf{u}_q)} + (1 - \alpha) \cos(\theta(\mathbf{u}_p, \mathbf{u}_q)) \quad (5)$$

式中 α 为加权因子， $0 \leq \alpha \leq 1$ 。由式(5)可知，两个向量越相似， $D(\mathbf{u}_p, \mathbf{u}_q)$ 也就越大。选择 $D(\mathbf{u}_p, \mathbf{u}_q)$ 最大的 K 组数据构成学习集： $\{(\mathbf{u}_1, m_1), \dots, (\mathbf{u}_i, m_i), \dots, (\mathbf{u}_K, m_K)\}$ ，其中， m_i 为主要变量的一个样本。考虑到计算复杂度和模型精度， K 一般取为 $[30, 50]$ ，即建立局部模型最少需要 30 个样本，最多需要 50 个样本，其最优值采用文献[11]中的 PRESS 估计方法获取。

在选定了 K 组建模数据后，第 k ($k=1, \dots, K$) 个主要变量值可用下面的一阶线性回归多项式来近似表示

$$\hat{m}_k = \mathbf{u}_k^T \boldsymbol{\beta} \quad (6)$$

式中 $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_{r_1}]^T$ ，为模型参数。

以主要变量的真实值和模型近似值偏差的加权平方和为性能指标，构造准则函数

$$J = \sum_{k=1}^K (m_k - \hat{m}_k)^2 D(\mathbf{u}_k, \mathbf{u}_q) \quad (7)$$

最小化 J ，可以得到此时的模型参数 $\hat{\boldsymbol{\beta}}$ 。则当前工作点 $\mathbf{u}_q$ 处的主要变量模型预测值为

$$\hat{m}(t) = \mathbf{u}_q^T \hat{\boldsymbol{\beta}} \quad (8)$$

每个主要变量都进行了上述的多模型局部建模后，当前工作点处外部变量真实值和模型预测值的残差 $\phi(t)$ 为

$$\phi(t) = \mathbf{M}(t) - \hat{\mathbf{M}}(t) \quad (9)$$

式中 $\mathbf{M}(t) = [m_1(t), m_2(t), \dots, m_{r_2}(t)]$ 为真实输出， $\hat{\mathbf{M}}(t) = [\hat{m}_1(t), \hat{m}_2(t), \dots, \hat{m}_{r_2}(t)]$ 为局部模型估计值。这样，便通过多模型局部建模的方法更好地描述了外部变量对主要变量的影响关系，去除了由外部变量引起的工况变化，并获得了残差 $\phi(t)$ 用于后续的过程监控。

2 Greedy-KP1M 过程监控

2.1 核单簇可能性聚类 (KP1M) 过程监控

下面建立模型对残差 $\phi(t)$ 进行监控。本文采用一种新方法——核单簇可能性聚类 (KP1M) 来进行监控。由于 KP1M 在核空间进行分类，所以该方法拥有很好的非线性处理能力。

模糊 C 均值 (fuzzy C-means, FCM)^[12] 将模

模糊隶属度的思想引入到聚类分析中,收到了很好的聚类效果。FCM 认为每个样本对每个类的模糊隶属度和为 1,但是当聚类样本中存在野点时,由于该约束的存在,使得野点也有对每类的隶属度,导致 FCM 对野点不鲁棒。文献 [13] 通过放宽 FCM 中的约束条件,提出了可能性 C 均值 (possibilistic C-means, PCM) 聚类,使样本以某一概率隶属于某一类,从而提高了 PCM 的鲁棒性。单簇可能性聚类 (P1M) 作为 PCM 在进行单簇聚类时的一种特例,具有与 SVDD 相似的单类数据描述能力和很好的鲁棒性^[14]。本文将用于过程监控中,并引入核型 P1M (KP1M) 增强其非线性处理能力。

对于残差数据集 $\Xi \in R^{N \times r_2}$, KP1M 可以表示为

$$\min J(\mathbf{U}, \mathbf{v}) = \sum_{i=1}^N u_i^l d^2[\phi(\xi_i) - \phi(\mathbf{v})] + w \sum_{i=1}^N (1 - u_i)^l \quad (10)$$

s. t. $0 \leq u_i \leq 1$

其中, $\mathbf{U} = [u_1, \dots, u_i, \dots, u_N]^T$, u_i 是第 i 个样本对目标类的隶属度; l 为模糊因子,这里取 $l=2$; $\phi(\cdot)$ 为非线性映射,将数据映射到高维空间; $\mathbf{v}$ 为聚类中心; w 为一个正的因子,防止式 (10) 出现隶属度全为 0 的解; $d[\phi(\xi_i) - \phi(\mathbf{v})]$ 为高维空间中残差 ξ_i 和 $\mathbf{v}$ 的欧氏距离。由于 KP1M 最小化样本与聚类中心距离的加权和,故 KP1M 比 SVDD 对野点更加鲁棒。采用文献 [13] 中的迭代算法,并引入核函数 $K(\xi_i, \mathbf{v}) = \phi(\xi_i)^T \phi(\mathbf{v})$,可以求得式 (10) 的解

$$u_i = \frac{1}{1 + \left| \frac{d^2(\phi(\xi_i) - \phi(\mathbf{v}))}{w} \right|} \quad (11)$$

$$\mathbf{v} = \frac{\sum_{i=1}^N u_i^2 K(\xi_i, \mathbf{v}) \xi_i}{\sum_{i=1}^N u_i^2 K(\xi_i, \mathbf{v})} \quad (12)$$

因子 w 可以控制样本隶属度的疏密程度,大的 w 可以使样本获得较大的隶属度,并且隶属度的值比较集中,反之亦然, w 根据经验确定,本文取为 0.01。

在得到了数据集 Ξ 对目标类的隶属度后,需要确定隶属度阈值 γ 。 γ 的选择与建模样本的噪声程度有关,同时 γ 的大小决定了最终的监控性能,较大的 γ 值可以获得较小的误报率,但是检测率也随之下降,反之亦然,因此根据经验判断来选取 γ 值。假设隶属度等于 γ 的点为边界点,由式 (11)

可得边界点到聚类中心的距离 R 为

$$R = \sqrt{w \left(\frac{1}{\gamma} - 1 \right)} \quad (13)$$

$\|\phi(\psi(t)) - \phi(\mathbf{v})\|^2$ 为当前时刻的残差 $\psi(t)$ 与聚类中心距离的平方,如果

$$\|\phi(\psi(t)) - \phi(\mathbf{v})\|^2 < R^2 \quad (14)$$

则认为属于正常状态,反之则发生了故障。将式 (14) 两边除以 R^2 得

$$\Omega = \|\phi(\psi(t)) - \phi(\mathbf{v})\|^2 / R^2 < 1 \quad (15)$$

则可以将 1 作为新统计量 Ω 的控制限。核函数选择高斯核函数 $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / 2\sigma^2)$, 则式 (15) 可以展开为

$$\Omega = 2[1 - K(\psi(t), \mathbf{v})] / R^2 < 1 \quad (16)$$

2.2 Greedy 方法进行数据特征提取

PCM 算法的计算复杂度是 $O(nc^2 dt)^{[15]}$, 其中 n 是样本个数, c 是聚类数, d 是样本维数, t 是迭代次数, 则 KP1M 的计算复杂度是 $O(ndt)$ 。当 $d \ll n$ 和 $t \ll n$ 时, KP1M 的计算复杂度主要由样本个数决定。因此,在进行 KP1M 建模前,提取建模数据集中的特征样本,用特征样本代替全部样本,可以减小计算量。下面采用 Greedy 方法求取特征样本。

Greedy 方法^[16]对数据进行特征提取的思想是:在给定的建模残差数据集 $\Xi = [\xi_1, \xi_2, \xi_3, \dots, \xi_N]^T \in R^{N \times r_2}$ 中,寻找到一个特征子集 $\Xi_t = [\xi_{t1}, \xi_{t2}, \dots, \xi_{ts}]^T$ ($s \ll N$),使由特征子集张成的空间近似等于由原建模数据集张成的空间,即 $\text{span}(\Xi) \approx \text{span}(\Xi_t)$ 。由于 KP1M 对数据的分类是在核空间进行的,这里通过非线性映射 ϕ 将 Ξ 和 Ξ_t 变换到特征空间,则在特征空间 Ξ 可以用 Ξ_t 近似线性表示为

$$\tilde{\phi}(\xi_i) = \sum_{j=1}^s \alpha_{ij} \phi(\xi_{tj}) \quad (17)$$

其中, α_{ij} 为待定系数。

用 Greedy 方法进行特征提取可以表示为如下最优化问题

$$\begin{aligned} \min e(\Xi_t, \Delta) &= \frac{1}{N} \sum_{i=1}^N \|\phi(\xi_i) - \tilde{\phi}(\xi_i)\|^2 = \\ &= \frac{1}{N} \sum_{i=1}^N (\phi(\xi_i) - \Phi(\Xi_t)^T \Delta_i)^T (\phi(\xi_i) - \Phi(\Xi_t)^T \Delta_i) \end{aligned} \quad (18)$$

其中, $e(\Xi_t, \Delta)$ 为均方重构误差, $\Phi(\Xi_t) = [\phi(\xi_{t1}), \phi(\xi_{t2}), \dots, \phi(\xi_{ts})]^T$, $\Delta_i = [\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{is}]^T$ 。将式 (18) 展开并引入核函数可得

$$\min e(\mathbf{\Xi}_t, \mathbf{\Delta}) =$$

$$\frac{1}{N} \sum_{i=1}^N [K(\xi_i, \xi_i) - 2k_i(\xi_i)^T \cdot \mathbf{\Delta}_i + \mathbf{\Delta}_i^T \cdot K_i \cdot \mathbf{\Delta}_i] \quad (19)$$

其中, $k_i(\xi_i) = \phi(\xi_i) \cdot \Phi(\mathbf{\Xi}_t) = [K(\xi_i, \xi_{i1}), \dots, K(\xi_i, \xi_{is})]^T$, $K_i = \Phi(\mathbf{\Xi}_t) \cdot \Phi(\mathbf{\Xi}_t)^T$ 。当 $\mathbf{\Xi}_t$ 给定时, $\mathbf{\Delta}_i$ 可以通过使 $e(\mathbf{\Xi}_t, \mathbf{\Delta})$ 最小求得

$$\mathbf{\Delta}_i = (K_i)^{-1} k_i(\xi_i) \quad (20)$$

将式(20)代入式(19)可得

$$\min e(\mathbf{\Xi}_t) = \frac{1}{N} \sum_{i=1}^N [K(\xi_i, \xi_i) - k_i(\xi_i)^T K_i k_i(\xi_i)] \quad (21)$$

通过式(21)可以看出, Greedy 方法的最终目的就是要找到使 $e(\mathbf{\Xi}_t)$ 最小的 $\mathbf{\Xi}_t$ 。在 $\mathbf{\Xi}$ 中求解出子集 $\mathbf{\Xi}_t$ 总共有 C_N^s 种选择方法, 当 N 很大, s 很小时, 求解过程是不符合实际要求的, 为此 Greedy 方法采用了一种基于贪婪思想的近似解法。

(1) 初始化: $\mathbf{\Xi}_t = \emptyset$ 。

(2) For $g=1$ to s :

① 在 $\mathbf{\Xi}$ 中找出使 $e(\mathbf{\Xi}_t)$ 达到最小的样本 ξ_{ig} ;

② $\mathbf{\Xi}_t = \mathbf{\Xi}_t \cup \{\xi_{ig}\}$;

③ $\mathbf{\Xi} = \mathbf{\Xi} - \{\xi_{ig}\}$ 。

通过以上步骤便可以得到特征子集 $\mathbf{\Xi}_t$, 然后将 $\mathbf{\Xi}_t$ 代替 $\mathbf{\Xi}$, 通过式(10)~式(16)进行 KP1M 监控。

3 新方法监控步骤

本文提出的基于多模型外部分分析和 Greedy-KP1M 的过程监控算法可以分为以下两部分。

(1) 离线建模

① 将过程变量分为外部变量和主要变量, 并采集正常条件下的建模数据;

② 进行多模型外部分分析, 得到残差序列;

③ 使用 Greedy 算法得到残差序列中的特征样本;

④ 建立 KP1M 模型, 并以 1 作为统计量 Ω 的控制限;

(2) 在线监控

① 获得新的采样点, 进行外部分分析得到残差;

② 用建立的 KP1M 模型对残差进行监控, 并计算统计量 Ω 的值;

③ 监控 Ω 是否超限。

4 仿真和应用研究

4.1 TE 过程仿真

TE (Tennessee Eastman) 过程是一个标准的

(benchmark)、公认的用于比较、测试控制和监控方法的平台^[17], 包括 41 个检测变量和 12 个操作变量。选择检测变量中的 22 个连续变量作为过程监测变量。其中, 外部变量为 4 个: 物料 A 流量, 物料 B 流量, 物料 C 流量和物料 A、C 混合流量; 剩下的 18 个监测变量为主要变量。为了模拟多工况过程, 本文选择工况 1 和工况 3 过程。在工况 3 条件下, 采集 1400 个正常样本, 其中 800 个样本作为多模型局部建模已存储的历史数据集, 300 个用于建模, 300 个用于测试。然后将过程切换到工况 1 下, 采集与工况 3 同样的数据样本。此外, 为了验证算法的故障监控能力, 在工况 3 条件下, 采集 400 个正常样本, 并在第 401 个样本处引入故障, 再采集 400 个故障样本。然后切换到工况 1 条件下, 同样收集 400 个正常数据和 400 个故障数据。为了验证算法对多种故障的监控能力, 本文选择具有代表性的故障 2、故障 5、故障 10、故障 11 和故障 20。

仿真中, 多模型局部建模算法中近邻个数 K 的选择范围为 $[30, 50]$ 。Greedy 特征样本提取方法采用高斯核函数, 核宽选择为 17。Greedy 算法最终提取两种工况 600 个建模数据样本中的 70 个样本用于 KP1M 建模。KP1M 算法中, w 为 0.01, 核宽选择为 100, 阈值 γ 为 0.77。

本文方法用于 TE 过程故障监控的仿真结果如图 2 所示。表 1 比较了本文方法和单一模型外部分分析的 Greedy-KP1M 算法, 无 Greedy 特征样本提取方法的 KP1M 算法, 以及采用多模型外部分分析的 Greedy-SVDD 过程监控方法。从图中可以看出, 本文方法对正常工况和过程故障的监控都收到了较好的效果。本文方法和方法 1 比较后可以看出, 由于单模型不能很好地描述外部变量和主要变量之间的关系, 所以故障 10、故障 11、故障 20 的检测率比本文方法低。和方法 2 比较后可以看出, 虽然本文方法仅采用了不到 12% 的建模数据, 但是获得了和使用全部数据集进行建模一样的监控效果。SVDD 方法的计算复杂度为 $O(N^3)$, 而 KP1M 方法的计算复杂度为 $O(Ndt)$, 其中 N 为建模样本数, d 为建模数据维数, t 为算法迭代次数。本文 N 为 70, d 为 18, t 为 10, 则本文方法的计算复杂度约为 SVDD 方法的 4%。和方法 3 比较后可以看出, KP1M 方法在过程监控上可以获得和 SVDD 一样的监控效果, 但其算法复杂度大大小于

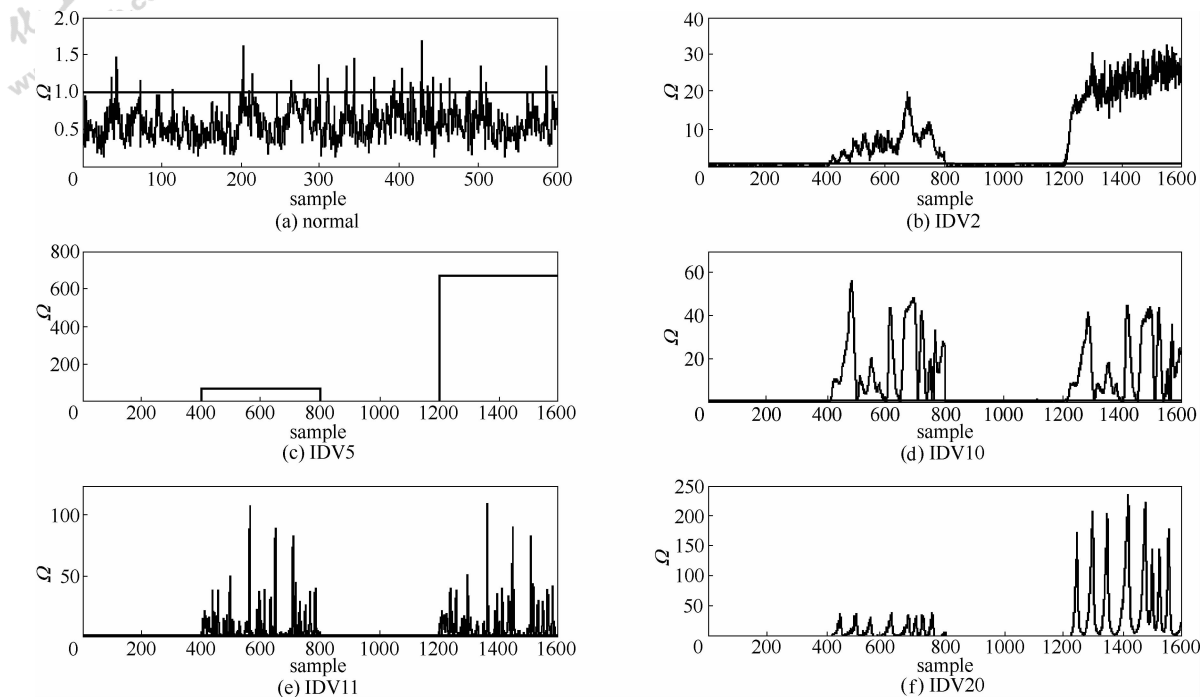


图 2 本文方法的正常工况、IDV2、IDV5、IDV10、IDV11、IDV20 监控结果

Fig. 2 Monitoring results of TE process in the cases of normal mode, IDV2, IDV5, IDV10, IDV11 and IDV20

表 1 四种过程监控方法的对比

Table 1 Comparison of four process monitoring methods

No.	Type of external analysis	Monitoring method	Number of modeling samples	Normal false detection rates/%	Detection rates/%				
					IDV2	IDV5	IDV10	IDV11	IDV20
1	single model	Greedy-KP1M	70	8.5	97	99.875	47.5	26.375	78.75
2	multiple models	KP1M	600	4.1667	98.25	99.875	90.75	83.875	93.125
3	multiple models	Greedy-SVDD	70	6.5	98.75	99.875	92.125	85.625	93.5
4	multiple models	Greedy-KP1M	70	5.6667	98.5	99.875	92.375	85.875	93.25

SVDD 算法。

4.2 乙烯裂解炉应用

以某石化企业大型裂解炉为例，将本文方法应用到乙烯裂解炉过程监控中。

乙烯裂解炉是乙烯生产中的龙头设备，裂解原料（石脑油、乙烷等）通过裂解炉的高温裂解后变为乙烯、丙烯、丁二烯等化工基础原料。由于生产的需要，每台裂解炉的进料负荷经常进行调整。当进料负荷发生变化（上升或者下降）时，由于要保持裂解炉的出口温度不变，裂解炉的燃料气量会增加或者减少，废热锅炉的出口温度也会上升或者下降，同时炉膛温度和稀释蒸汽流量也会发生相应的变化。因此，不同的进料负荷对应着不同的工况。由于乙烯裂解是一个复杂过程，若采用常规的故障检测方法，当负荷发生变化时，极易产生误判和漏判。因此有必要对裂解炉进行多工况条件下的过程

监控。

选择对裂解炉运行影响较大的 36 个变量，其中外部变量为 6 组炉管的进料流量，主要变量为 6 组炉管的稀释蒸汽流量，裂解炉底部燃料气流量，侧壁燃料气流量，6 组炉管的出口温度，6 组炉管的横跨段温度，6 组炉管的出口压力，3 个废热锅炉的出口温度和炉膛温度。本文对 30 个主要变量建立过程监控模型，并以进料负荷为 $38\text{ t}\cdot\text{h}^{-1}$ 和 $40\text{ t}\cdot\text{h}^{-1}$ 的两种工况条件为例对本文方法进行说明。现场采集工况 1 条件下 900 个正常样本，其中 500 个用作多模型局部建模历史数据集，250 个用于建模，150 个用作检验模型的有效性；在工况 2 条件下采集与工况 1 同样的数据样本。另外，收集了 3 种裂解炉故障数据，每种故障数据 300 组，并在第 101 个采样点处出现故障。3 种故障见表 2。

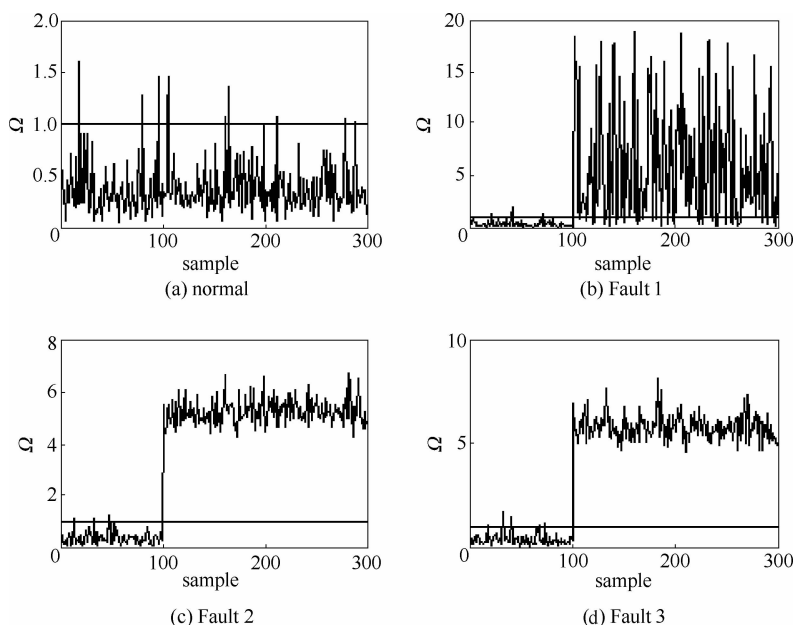


图 3 乙烯裂解炉正常工况、Fault 1、Fault 2、Fault 3 的监控结果

Fig. 3 Monitoring results of ethylene cracking furnace in cases of normal mode, Fault 1, Fault 2 and Fault 3

表 2 裂解炉 3 种故障

Table 2 Three kinds of ethylene cracking furnace's fault

Item	Fault
Fault 1	radiant section COP fluctuation of furnace tube A
Fault 2	XOT step change of furnace tube C
Fault 3	COT sensor gain fault of furnace tube F

在应用中，多模型局部建模算法近邻个数 K 的选择范围为 $[30; 50]$ ；Greedy 特征样本提取方法同样采用高斯核函数， σ 选择为 19。Greedy 算法最终提取 500 个建模数据样本中的 50 个样本用于 KP1M 建模。KP1M 算法中， w 为 0.01，核宽选择为 100，阈值 γ 为 0.5358。对乙烯裂解炉的监控如图 3 所示。

从图 3 可以看出，本文方法在对乙烯裂解炉进行监控时，可以很好地应对裂解炉进料负荷的变化，无论是正常情况还是 3 种故障情况，都取得了很好的效果。在建模时，本文方法仅用到了 500 组建模数据中的 50 组，减少了 KP1M 建模的计算复杂度。

5 结 论

本文针对多工况过程监控问题，首先基于多模型局部建模方法对系统进行外部分析，去除多工况的影响。然后用 KP1M 方法对外部分析后的残差建立非线性监控模型，采用 Greedy 方法降低 KP1M 方法建模时的计算复杂度。通过在 TE 标准

平台上的仿真对比，可以看出 KP1M 方法有着和 SVDD 方法相当的过程监控能力，但其计算复杂度和鲁棒性却优于 SVDD；通过乙烯裂解炉上的应用研究可以看出文本方法可以很好地完成实际过程的监控任务。

References

- [1] Chiang L H, Russell E L, Braatz R D. Fault Detection and Diagnosis in Industrial Systems [M]. London: Springer-Verlag, 2001: 2
- [2] Maestri M, Farall A, *et al.* A robust clustering method for detection of abnormal situations in a process with multiple steady-state operation modes [J]. *Computers and Chemical Engineering*, 2010, **34** (2): 223-231
- [3] Hwang D H, Han C. Real-time monitoring for a process with multiple operating modes [J]. *Control Engineering Practice*, 1999, **7** (7): 891-902
- [4] Kano M, Hasebe S, Hashimoto I, Ohno H. Evolution of multivariate statistical process control: application of independent component analysis and external analysis [J]. *Computers and Chemical Engineering*, 2004, **28** (6/7): 1157-1166
- [5] Tax D M J, Duin R P W. Support vector data description [J]. *Machine Learning*, 2004, **54** (1): 45-66
- [6] Zhang Shaojie (张少捷), Wang Zhenlei (王振雷), Qian Feng (钱锋). FS-SVDD based on LTSA and its application to chemical process monitoring [J]. *CIESC Journal (化工学报)*, 2010, **61** (8): 1894-1890
- [7] Wang Kun (王坤), Du Wenli (杜文莉), Qian Feng (钱锋). Non-Gaussian process fault detection based on wavelet kernel clustering [J]. *CIESC Journal (化工学报)*, 2011, **62** (2): 427-432

- [8] Zhu Zhibo (祝志博), Song Zhihuan (宋执环). Fault isolation: an FDA-SVDD based pattern classification algorithm [J]. *CIESC Journal* (化工学报), 2009, **60** (8): 2010-2016
- [9] Atkeson C G, Moore A W, Schaal S. Locally weighted learning [J]. *Artificial Intelligence Review*, 1997, **11** (1): 11-73
- [10] Bontempi G, Birattari M, Bersini H. Lazy learning for local modeling and control design [J]. *International Journal of Control*, 1999, **72** (7/8): 643-658
- [11] Pan Tianhong (潘天红), Li Shaoyuan (李少远), Wang Xin (王昕). A multi-model modeling approach to nonlinear systems based on lazy learning//Proceedings of the 24th Chinese Control Conference [C]. Guangzhou, China, 2005: 268-273
- [12] Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms [M]. New York: Plenum Press, 1981
- [13] Krishnapuram R, Keller J. A possibilistic approach to clustering [J]. *IEEE Transaction on Fuzzy Systems*, 1993, **1** (2): 98-110
- [14] Chen Bin (陈斌), Feng Aimin (冯爱民), Chen Songcan (陈松灿), Li Bin (李斌). One-cluster clustering based data description [J]. *Chinese Journal of Computers* (计算机学报), 2007, **30** (8): 1325-1332
- [15] Kolen J, Hutcheson T. Reducing the time complexity of the fuzzy C-means algorithm [J]. *IEEE Transaction on Fuzzy Systems*, 2002, **10** (2): 263-267
- [16] Franc V, Hlavác V. Greedy Algorithm for a Training Set Reduction in the Kernel Methods [M]. Berlin: Springer-Verlag, 2003: 426-433
- [17] Downs J J, Vogel E F. A plant-wide industrial process control problem [J]. *Computers and Chemical Engineering*, 1993, **17** (3): 245-255