



ELSEVIER

Contents lists available at ScienceDirect

Engineering

journal homepage: [www.elsevier.com/locate/eng](http://www.elsevier.com/locate/eng)



Research  
Smart Process Manufacturing—Article

## 过程操作性能的在线评估与诊断

Shabnam Sedghi, Biao Huang\*

Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta T6G 1H9, Canada

### ARTICLE INFO

#### Article history:

Received 30 November 2016

Revised 26 January 2017

Accepted 2 February 2017

Available online 16 March 2017

#### 关键词

最优性能评估  
概率主成分回归  
多模态过程

### 摘要

随着时间的推移,过程性能会因为过程变动和不确定性从其初始状态偏移,这使系统性地发展基于日常过程操作数据的在线最优性评估方法变得必要。一些过程,由于其关键过程变量操作点的变动,会产生多个不同操作模态,以满足不同的产品要求。另外,每一个操作模态的操作区域会由于系统的不确定性产生变动。本文中,我们建立了一个用于多模态、多操作区域的典型过程的最优性评估,该方法也能够处理模态切换时的过渡状态的最优性评估。在该框架中,核密度方法被改进,并被用于操作模态检测。在线阶段,结合过程系统的先验知识,基于模型的聚类判别分析方法(model-based clustering discriminant analysis, MclustDA)被用于模态检测。此外,稳态模态的多模态特性由混合概率主成分回归方法(mixture probabilistic principal component regression, MPPCR)处理;动态主成分回归方法(dynamic principal component regression, DPCR)被用来探究不同模态间的过渡状态的性能评估。除此以外,一种基于序列浮动前向搜索(sequential forward floating search, SFSS)的概率因果关系检测方法被引入,用以检测系统的不良或非最佳性能。最后,本文提出的算法被应用于田纳西-伊斯曼(Tennessee Eastman)过程之中,用以评估本文算法的性能。

© 2017 THE AUTHORS. Published by Elsevier LTD on behalf of the Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. 引言

过程操作性能评估是过程工业的一个重要课题,在学界和工业界都引起了很高的关注。由于过程的性能会随着时间逐渐退化,操作模态会由于操作条件的变化或者过程的变动,随着时间逐渐偏离其初始设计模态,因此,有必要对过程的性能进行连续监控。这一类的分析是对传统控制性能评估的巨大改进,被称为“最优性能评估”。

近年来对于最优性能评估的一些研究[1–4]逐渐得

到发表。然而,这些研究并没有对通用的复杂过程提出解决方法。在本文中提出了一个系统性的最优性能评估方法,该方法能够提供一个系统性的框架,用于解决现代工业过程的最优性能评估问题。在这个框架中,首先,我们考虑到了由于操作条件和产品要求变化造成的多操作模态问题,其次,我们考虑了稳态模态中,系统的不确定性和扰动造成的每个单一模态的多操作区域问题,再次,我们考虑了不同操作模态间切换时的过渡状态问题。

为了解决上述问题,本文新提出了概率主成分回归

\* Corresponding author.

E-mail address: [biao.huang@ualberta.ca](mailto:biao.huang@ualberta.ca)

方法(probabilistic principal component regression, PPCR)。该方法首先在单一过程的性能评估中被提出(单一模态在实际情况下很常见),而后才被扩展应用于具有多个模态的过程中。对于单一模态的过程性能评估,PPCR方法包含两个阶段:离线训练阶段和在线评估阶段。在离线训练阶段,需要收集同时包含过程变量和最优性能指标(optimality index, OI)的稳态数据。需要注意的是OI由过程决定。例如, OI可以指系统的操作成本、利润、产品产量、环境指标等。为了获得一个在线的最优性能指标OI,需要基于过程变量,建立一个OI的预测模型。由于每个操作模态通常都具有多个操作区域, MPPCR被用于建立操作模态的模型。这个MPPCR模型描述了OI在每个操作区域的高斯分布特性。基于该模型,针对每个操作区域的局部OI可以被获得。通过比较每一个操作区域的局部OI可以分析它们的最优情况。在在线评估阶段,新的数据样本点的操作区域可以通过其后验概率进行估计。通过建立好的模型, OI可以通过贝叶斯推断来进行过程性能的评估。当过程不是最优时,诊断引起该故障的原因可以引导过程获得更好的性能。基于缺失变量分析[5]的概率分布分析技术被用来解决该问题。序列浮动前向搜索(SFFS)被用来代替分析界定法,用以降低计算代价,并且能够简化问题。

对于多操作模态的过程,假设样本点都是无标签的。换句话说,操作模态的数目和每个样本点所处的操作模态是未知的。为了估计样本集的标签,主导过程变化的关键过程变量被选择出来并且被称为调度变量(scheduling variables)。基于被选择的调度变量,一种基于局部核密度的方法[6]被改进并应用于检测样本点的标签。为了估计在线评估时的操作模态,基于有标签的数据集建立了混合判断式分析(mixture discriminant

analysis, MDA)方法。除此以外,为了提高在线模态检测的准确率,过程知识也被整合进MDA中。稳态模态的最优性能评估的评估方法与单一模态的最优评估相同。对于两个模态的过渡状态,动态主成分回归DPCR模型也被建立,其性能由(DPCR)的负载矩阵进行评定[7]。

本文以下内容的结构为:第2节讨论了研究的问题和提出的解决方法;第3节中,本文提出的最优性能评估策略在稳态模态中的实施方法被具体地讨论;第4节讨论了过渡状态的评估方法;第5节对多模态过程的模态检测方法进行了讨论;第6节中,本文提出的方法被应用于田纳西-伊斯曼(Tennessee Eastman, TE)过程;最后进行了总结。

## 2. 问题描述以及本文提出的新方法

一个综合的过程操作通常有多模态特点,并且每一个模态的稳态具有非高斯特征。过程系统概览如图1所示。它考虑到了由于已知关键因素(如产品要求)的变化而造成的操作模态的变化。此外,每一个稳态操作模态通常会含有多个操作区域,这是由不确定的系统变化引起的。因此,系统的最优性等级会根据操作的不同区域而变化。

本文中提出的方法的目标是基于正常操作过程进行在线估计操作过程的性能评估。换言之,通过估计操作模态和操作区域(或者过渡状态等级),然后预测OI值,最后诊断过程性能指标低的原因。本文提出的框架包括离线训练和在线评估阶段。其框架和算法如图2和图3所示,在接下来的章节中将对细节进行详细讨论。

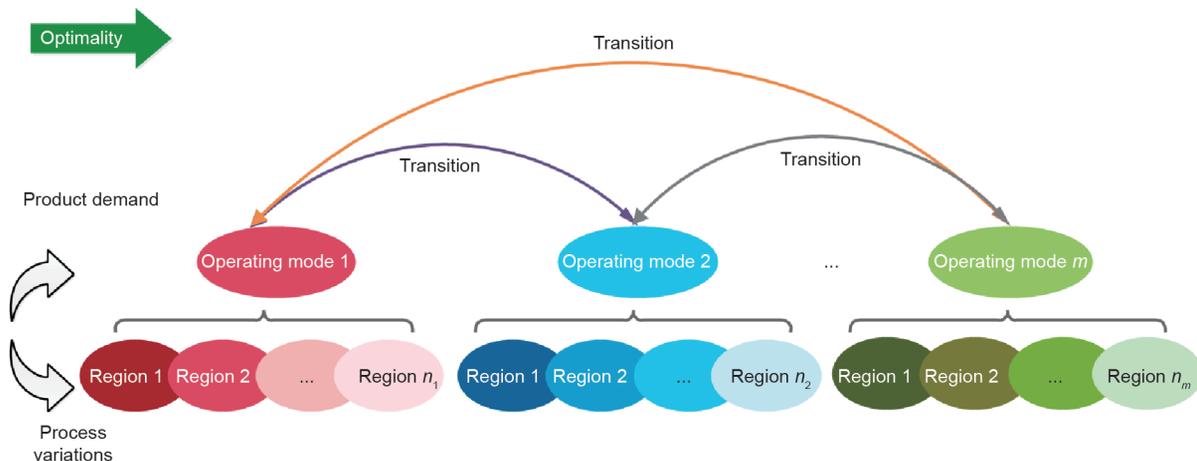


图1. 综合过程操作概览。

### 3. 稳态模态：定义

当过程中的过程变量、流程设计、产品要求等没有发生重大变化时，稳态模态是过程的主要模态。MPPCR模型可以利用训练数据建立模型。然后，可以利用建立好的监测模型获得每一个操作区域的局部OI。之后，基于过程知识可以定义几类最优值，并且，计算获得的操作区域可以被分配至相应的不同类。

#### 3.1. 数据建模

假设有  $X=[x(1), x(2), \dots, x(n)]^T \in R^{n \times p}$  和  $Y=[y(1), y(2), \dots, y(n)]^T \in R^{n \times 1}$ ，这两个数据集分别是过程的变量和OI，其中， $n$ 是样本点的数目， $p$ 是过程变量的个数。由于操作模态会含有若干操作区域，MPPCR模型可以被用来建立一个预测性的模型，其中的输入是 $X$ ，输出是 $Y$ 。

#### 3.2. 最优性指数分析

每个操作区域 $k$ 的OI的高斯分布可以在建模阶段中被估计。结果是每个操作区域的局部OI[1]等于 $y$ 的高斯分布的均值。

$$OI_k = E(y) = \int f_k(y)y dy = \mu_{y,k} \quad (1)$$

#### 3.3. 非最优原因检测

为了获得导致系统处于非最优或者不良状态的根本原因，我们可以利用一种基于缺失变量方法的概率分布分析技术。该方法已经被利用在系统的故障检测上[5,8]、离群点检测[9]等诸多领域。本文中，我们将这种方法与一种改进的因果关系检测方法相结合，用于最优性评估。在这个改进的方法中，最优区域被称为参考区域或者标准区域。当一个样本点和参考区域的马氏距离 $M^2$ 超过置信限[该置信限由具有 $r$ 个自由度的卡方分布的 $\beta$ 分位数决定，即 $\chi^2 r(\beta)$ ]时，则可以被认为是一个非最优性能的样本点被检测到。该方法中，当被分析的变量被认为是缺失变量时，每一个变量的贡献度等于其马氏距离 $M^2$ 和对应的期望值 $E(M^2)$ 之间的差异。本文中，我们提出如下的算法，用以找出一组基于SFFS的因果关系变量[10]：首先，完全集 $x_{\text{new}}=Y=\{y_j | j = 1, 2, \dots, p\}$ 包括了 $x_{\text{new}}$ 中所有的被测变量。缺失变量分析的目的是找出数目最少的一组缺失变量，使经过重新计算获得的 $E(M^2)$ 小于置信限。假设我们选

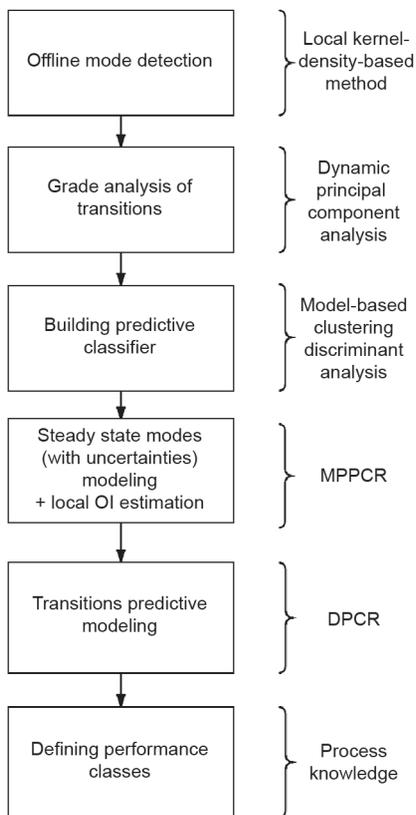


图2. 本文提出的用于离线训练的框架和算法。

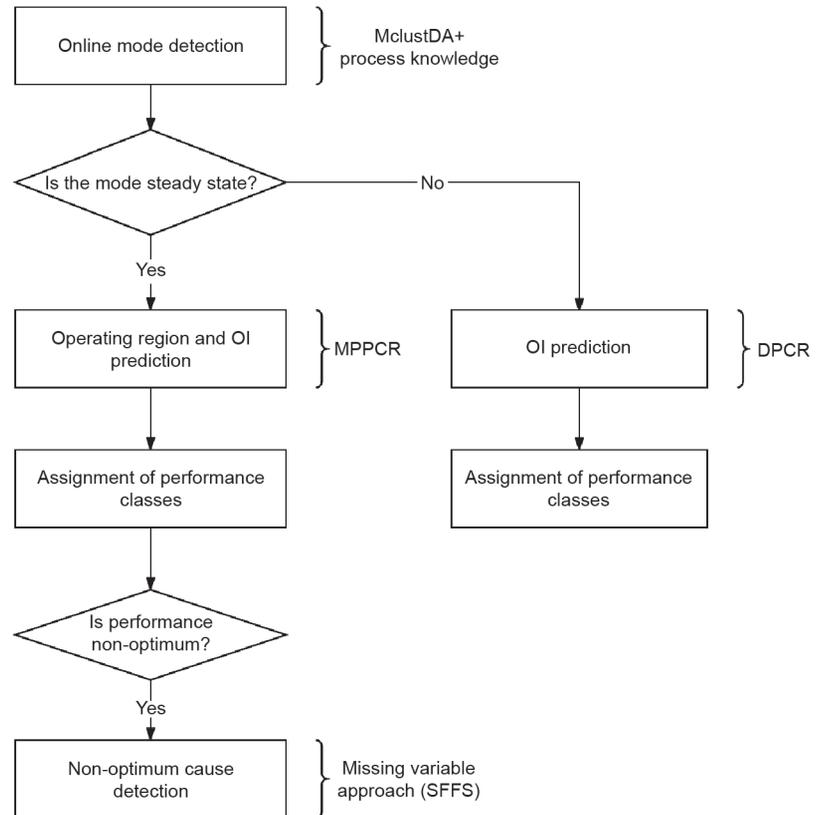


图3. 本文提出的用于在线评估的框架和算法。

择了 $k$ 的一个子集是一组缺失变量 $X_k$ ，其判别函数可由下式表达：

$$J(X_k) = M^2 - E_{Y|X_k}(M^2) \quad (2)$$

其中， $E_{Y|X}(M^2)$ 是 $M^2$ 的期望值，由缺失变量的选择方法决定，亦即 $X_k$ 。算法从 $k=1$ 启动，步骤如下。

步骤1：利用SFFS方法最大化 $J(X_k)$ ，从 $Y$ 中找出 $k$ 个特征，称为 $X_k$ 。

步骤2：如果 $E_{Y|X_k}(M^2)$ 小于置信限，则 $X_k$ 可以作为最终的原因变量集合；如果不小于置信限，则 $k = k+1$ ，之后重复步骤1。

## 4. 过渡状态的处理

过渡状态主要发生在多模态过程两个稳态模式的切换过程。本文中假设操作模式的切换是被监控的。因此，尽管在每个过程中都有噪声，但引起操作模式切换的关键操作变量是可以被监控的，通常被称为调度变量。

### 4.1. 过渡状态等级分析

动态主成分分析(DPCA)被利用于分析变量间的自相关性，同时也能够通过整合时滞信息和数据矩阵，分析它们的时滞特征。通过DPCA可以获得每一个基于调度变量的过渡状态的负载矩阵，并且可以计算获得相应的相似度指数。Srinivasan等[7]已经提过这种方法的细节。假设 $S$ 和 $T$ 分别是具有相同指数和最终状态的两个过渡过程。如果它们的相似性指数大于用户定义的控制限 $\theta_T$ ，则可以认为它们具有相同的过渡等级。

### 4.2. 过渡过程预测建模

基于完整的训练数据集对DPCR进行建模。对于每个过渡过程，当DPCA模型建立后，就可以利用回归来估计潜变量。在在线评估阶段，基于估计的过渡等级，OI的估计值与该等级中每个过渡状态的平均估计值相同。

## 5. 模式选择

为了将本文提出的方法扩展到多模态领域之中，需要添加一个步骤用于稳态模式和过渡状态的检测，而后，在第3和第4节中讨论的方法可以被继续用于过程评估。模

态检测包含操作模式的识别以及根据估计出的模式建立预测模型。

### 5.1. 操作模式的标记方法

Quiñones-Grueiro等[6]最近基于局部核概率密度估计的监控方法提出了一种离线模式检测方法。该方法是一种基于密度的聚类方法(DENCLUE)方法，其集成了过程的序列信息，使得分类的准确性得以提高。在本文中，上文提到的方法首先被修改，而后被应用于最优性评估。这种算法提出了一种非常有效的应用于操作模式标签的方法，该方法的特点是无需知道过程中存在的模式数目。对上文提到的方法的具体扩展如下：

(1) 为了找出过渡状态的准确起止时间，每一个过渡部分都被切割成更短小的时间窗，用于更清楚地分析其动态特性。

(2) 通过距离准则来比较每个过渡状态的稳态的初始和最终状态。如果这两个状态是相似的，说明过程工作在同一个模式中，因此，被检测到的过渡状态可以被认为是噪声引起的误报，而非一个真实的过渡状态。

(3) 通过距离准则比较稳态模式的初始和最终窗口，可以检测最终状态所处的模式和过渡状态。

### 5.2. 在线模式的监测

在在线评估阶段，根据新的数据样本点估计其对应的操作模式，从而选择合适的模型。所以，在离线模式监测阶段中，可以基于估计出的标签建立一个预测性的分类器。Fraleley和Raftery[11]通过将MDA的分类方法和基于模型的聚类方法(Mclust)结合成一个方法——Mclust判别分析(MclustDA)，来进行非高斯分布的分类。对于每一个操作模式，其中包括稳态模式和过渡状态，都通过Mclust-DA方法进行建模。需要注意的是，在本文中，假设所有的操作模式都是已知的。然而，在在线监测阶段中，仍然有可能出现原先并未出现过的新的模式。对于这种情况，一种可能的监测新模式的处理方式是计算每一个操作模式的新样本点的条件概率，然后计算其对应于每个操作模式的后验概率，得到条件概率和后验概率的联合概率。当其联合概率值非常小的时候，表明出现了一个新的模式[12]。

对于操作模式的在线预测，过程知识被整合进入预测过程以用于提高预测的准确性[4]。换句话说，代替计算每个样本点的所有操作模式的后验概率密度分布，相关操

作模态的后验概率可以根据如下步骤计算:

(1) 如果当前操作模态的数据样本点属于稳态模态 $i$ , 对于下一个样本点, 模态 $i$ 和所有对应于模态 $i$ 的过渡状态都可以被预测。

(2) 如果当前样本点处于过渡状态 $ij$ 的阶段 $p$ , 也就是 $\{ij\}_p$ , 则 $\{ij\}_p$ 的后验概率和其稳态模态 $j$ 可以被计算得到。最后, 样本点可以根据最高的后验概率, 被分类进不同的操作模态。需要注意的是, 在带噪声的环境中, 考虑到单个样本点可能会导致错误的解, 因此, 在这种情况下, 应该使用整个时间窗的样本点进行估计, 用以获得变化操作模态的更好的鲁棒估计。

## 6. 田纳西 - 伊斯曼 (TE) 标准测试模型

TE标准测试模型近年来被广泛应用于过程控制、软测量设计、过程监控等诸多领域。模型首先由Downs和Vogel[13]提出, 模型基于伊斯曼化学公司的实际工业过程

表1 稳定操作模态的设置情况

Mode	G/H	Reactor level (%)	Reactor temperature (°C)
1	50/50	65	122.9
2	10/90	50	130.0
3	40/60	55	135.0

“G” and “H” are the main products of the TE process.

表2 过程的不确定性

	Process variable	Type
1	B composition (stream 4)	Step
2	Reactor pressure	Step

进行设计。为了使过程维持稳定, Ricker[14]针对该开环系统设计了集散控制策略。

TE过程可以工作在三种不同模态中, 具体设置如表1所示。除此以外, 在每种不同的操作模态中, 都引入了两种不确定扰动, 这两种具体描述如表2所示。OI被用来表征过程的操作成本。离线训练数据中的两个变量[A和C的进料流股(流股4)和再循环的流股(流股8)]的变量映射关系如图4所示。具体来说, 每一个操作模态的近似边界如图4所示。

OI值预先定义好的等级如表3所示。需要注意的是, 等级越高, 最优性越差。局部OI值和等级如表4至表6所示。

### 在线评估

在在线模态检测阶段, 经过计算, 本文提出的方法的分类误差为0.0104, 这表明算法在模态检测上有很高的准确性。当OI值被预测之后, OI值的预测值和真实值的比较如图5所示。由于使用的模型随着过程而变动, 其所对应的模型如图5所示。此外, 经过计算, 均方根误差(RMSE)和 $R^2$ 值, 分别为0.3723和0.8475, 这表明了OI值预测的高准确性。离线模态的监测效果、在线模态的监测效果以及预测效果在表7中进行了总结。

OI等级的预估值如图6所示。根据图6, 过程从最优操作起步, 而后跳转到次优的等级2。举例来说, 以第1219个采样样本点为例, 以解释导致非最优性能的原因。基于先前的估计, 该数据样本点属于模态1的操作区域1, 因此, 该模态被选择作为参考的标准模态, 以

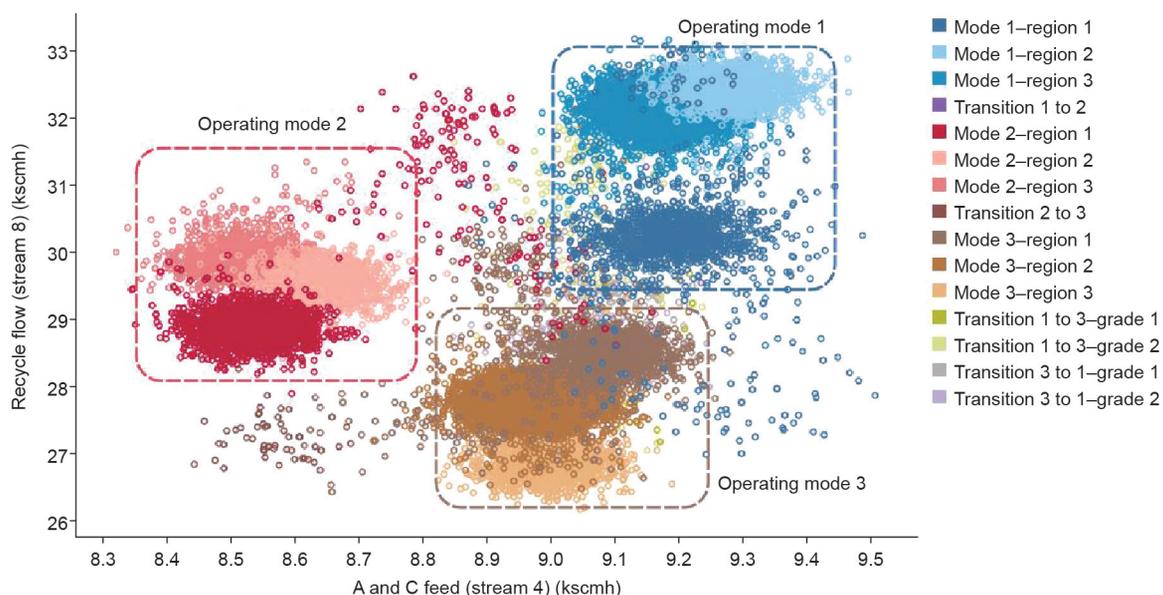


图4. 离线训练数据的二维散点图。

表3 定义OI等级

OI range (\$·h <sup>-1</sup> )	OI level
100–140	1
141–180	2
181–220	3
Above 221	4

表4 局部OI等级(模态1)

Operating region	Local OI (\$·h <sup>-1</sup> )	OI level
1	142.78	2
2	179.43	2
3	120.09	1

表5 局部OI等级(模态2)

Operating region	Local OI (\$·h <sup>-1</sup> )	OI level
1	205.67	4
2	259.01	3
3	186.62	3

表6 局部OI等级(模态3)

Operating region	Local OI (\$·h <sup>-1</sup> )	OI level
1	309.51	4
2	250.49	4
3	275.11	4

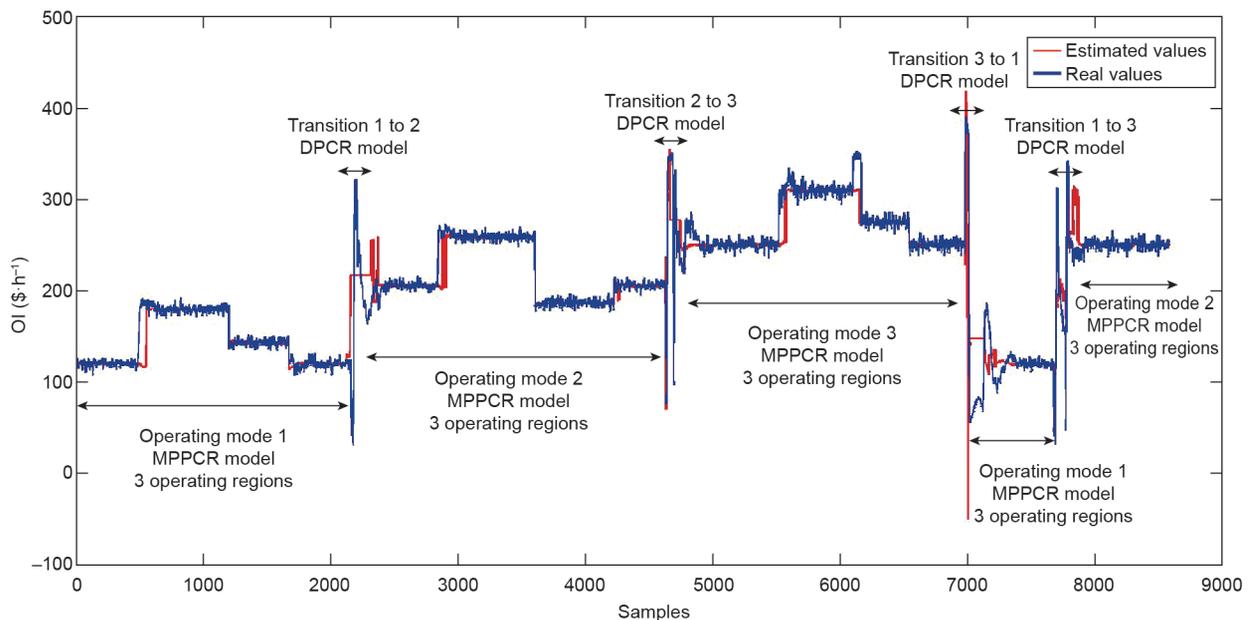


图5. OI的预测值和真实值的比较。

表7 算法结果总结

	Results
Offline mode detection	ARI = 0.9891 FM index = 0.9928
Online mode detection	Error = 0.0104
Prediction	RMSE = 0.3723 $R^2 = 0.8475$

ARI: adjusted Rand index; FM index: Fowlkes-Mallows index.

用于非最优性能时的原因检测。这时，该样本点距离参考模态的距离为195.11，这也就是说，距离远超过具有22个(过程的变量个数)自由度的卡方分布的0.95分位数 [ $\chi_{22}^2(0.95) = 33.924$ ]。9个导致非最优性能的变量可以被检测出，并且它们的贡献率如图7所示。当这9个变量被从过程中除去之后，样本点里参考模态的距离变为33.66，也就是低于 $\chi_{22}^2(0.95)$ ，这表明可以通过这9个变量将过程导向最优性能。

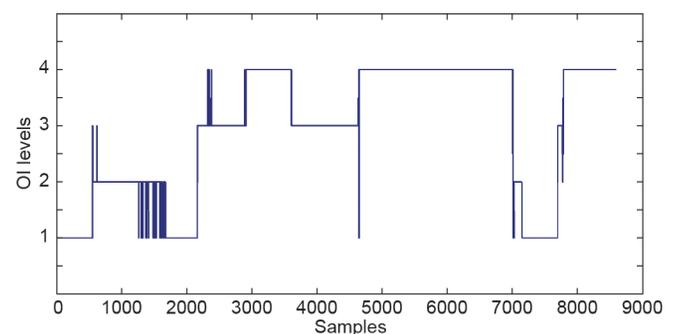


图6. OI等级的预估值。

## 7. 结论

在本文中，提出了一个新颖的用于在非高斯分布多模态过程中的最优操作性能评估框架。本文提出的方法能够用来检测操作模态、过渡状态和区域，并且提供了

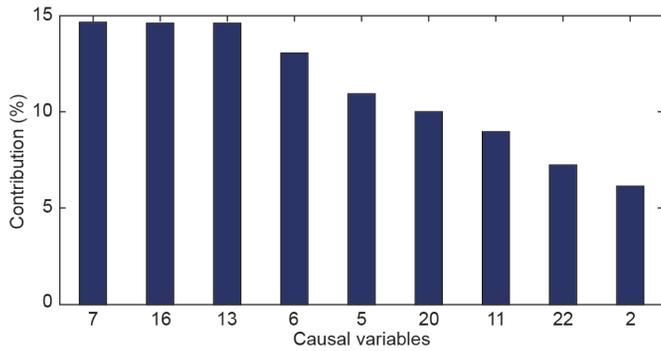


图7. 样本1219中因果变量的贡献百分比。

一种预测过程性能的模式。除此以外，对于不良或者非最优状态，本文也提供了因果关系检测方法。之后，将该方法应用于TE标准过程，确认了本文中提出的方法的有效性。

## Acknowledgements

This work is supported in part by the Natural Science Engineering Research Council of Canada and by Alberta Innovates Technology Futures.

## Compliance with ethics guidelines

Shabnam Sedghi and Biao Huang declare that they have no conflict of interest or financial conflicts to disclose.

## References

- [1] Ye L, Liu Y, Fei Z, Liang J. Online probabilistic assessment of operating performance based on safety and optimality indices for multimode industrial processes. *Ind Eng Chem Res* 2009;48(24):10912–23.
- [2] Liu Y, Chang Y, Wang F. Online process operating performance assessment and nonoptimal cause identification for industrial processes. *J Process Contr* 2014;24(10):1548–55.
- [3] Liu Y, Wang F, Chang Y, Ma R. Comprehensive economic index prediction based operating optimality assessment and nonoptimal cause identification for multimode processes. *Chem Eng Res Des* 2015;97:77–90.
- [4] Liu Y, Wang F, Chang Y, Ma R. Operating optimality assessment and nonoptimal cause identification for non-Gaussian multimode processes with transitions. *Chem Eng Sci* 2015;137:106–18.
- [5] Kariwala V, Odiowei PE, Cao Y, Chen T. A branch and bound method for isolation of faulty variables through missing variable analysis. *J Process Contr* 2010;20(10):1198–206.
- [6] Quiñones-Grueiro M, Prieto-Moreno A, Llanes-Santiago O. Modeling and monitoring for transitions based on local kernel density estimation and process pattern construction. *Ind Eng Chem Res* 2016;55(3):692–702.
- [7] Srinivasan R, Wang C, Ho WK, Lim KW. Dynamic principal component analysis based methodology for clustering process states in agile chemical plants. *Ind Eng Chem Res* 2004;43(9):2123–39.
- [8] Chen T, Sun Y. Probabilistic contribution analysis for statistical process monitoring: A missing variable approach. *Control Eng Pract* 2009;17(4):469–77.
- [9] Chen T, Martin E, Montague G. Robust probabilistic PCA with missing data and contribution analysis for outlier detection. *Comput Stat Data Anal* 2009;53(10): 3706–16.
- [10] Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection. *Pattern Recognit Lett* 1994;15(11):1119–25.
- [11] Fraley C, Raftery AE. Model-based clustering, discriminant analysis, and density estimation. *J Am Statist Assoc* 2002;97(458):611–31.
- [12] Ge Z, Song Z. Mixture Bayesian regularization method of PPCA for multimode process monitoring. *AIChE J* 2010;56(11):2838–49.
- [13] Downs JJ, Vogel EF. A plant-wide industrial process control problem. *Comput Chem Eng* 1993;17(3):245–55.
- [14] Ricker NL. Decentralized control of the Tennessee Eastman Challenge Process. *J Process Contr* 1996;6(4):205–21.