Chemometrics and Intelligent Laboratory Systems 145 (2015) 72–83

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Phase division and process monitoring for multiphase batch processes with transitions

Xiaochu Tang [a], Yuan Li [b,*], Zhi Xie [a]

[a] College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning Province 110819, China
[b] College of Information Engineering, Shenyang university of Chemical Technology, Shenyang, Liaoning Province, 110142 China

ARTICLE INFO

Article history:
Received 27 February 2015
Received in revised form 10 April 2015
Accepted 11 April 2015
Available online 20 April 2015

Keywords:
Multiphase batch process
Phase division
Process monitoring
Transition

ABSTRACT

In this paper, a new repeatability factor is introduced to achieve phase division for multiphase batch processes. Then a two-step feature vector selection based kernel variable correlation analysis (TSFVS-KVCA) method is proposed for batch processes with transitions monitoring. The TSFVS-KVCA method not only considers the within-phase nonlinear variable correlation but also between-phase one by extracting the common bases and the specific bases between two neighboring phases. The TSFVS method first selects feature vectors from each steady phase as within-phase subsets. Then between-phase subsets are selected from the within-phase subsets of two neighboring phases. These selected feature vectors have well capacity of description on original data so that they can substitute for original data in a feature space *F*. Thus the TSFVS method reduces the computational complexity and the instability of high-dimensional kernel in subsequent KVCA method. Moreover, each within-phase subset can be used as sub-bases in a feature space *F*, which simplify the objective function of KVCA method. In this way, the common bases can be extracted in an easier way. Based on the common bases, two neighboring steady phases can be separated into the common subspace and the specific subspace, in which nonlinear process monitoring method is carried out. Furthermore, an online monitoring method for transitions is proposed based on a just-in-time model. The model can be described by the common bases of two neighboring phases and specific bases of the dominant phase at current time interval. The dominant phase can be dynamically determined according to the correlation between current transition sample and neighboring phases. The results of simulation demonstrate effectiveness and superiority of the proposed method.

1. Introduction

Batch processes have been a core type of production in industrial process [1–5]. In order to ensure the safety of operation environment and meet the requirement of product quality, it is essential for batch processes to carry out the process monitoring. Multivariate statistical analysis, as an effective tool, has attracted increasing attention on applying this type of methods to batch processes monitoring. Multi-way principal component (MPCA) and multi-way partial least squares (MPLS) methods have laid a foundation for introducing multivariate statistical analysis method to batch processes monitoring [6–8]. After that, some researches and improvements for batch processes monitoring have been developed [9–18]. However, it is noted that there exits an important information on multiplicity in batch processes, which can not be mined by traditional MPCA and MPLS method. Therefore, process monitoring for multiphase batch processes should be well developed.

There are different process characteristics for different phases in multiphase batch process. So, different phase may exhibit different underlying behaviors. To improve the performance of process modeling, some multiphase statistical modeling and monitoring methods have been developed. Lu et al proposed a sub-PCA modeling and monitoring method, in which phase are divided by clustering weight loading of each time slice and within-phase model is built [19]. This algorithm was further developed, in which a new improvement is that separating phase based on PLS parameter matrix corresponding to each time slice [20]. Other methods were proposed by Camacho and Pico [21–23]. However, these methods utilized a strict hard-partition and neglected the transitions between two neighboring phases. This may compromise the accuracy of isolated sub-model during the transition period and it can cause a high false alarm rate in transitions. It illustrates that transition should not be ignored in process modeling and monitoring for multiphase batch processes. Therefore, it is necessary to investigate identification, modeling and monitoring for transitions.

Zhao et al. firstly proposed a soft transition algorithm to deal with hard phase division [24]. In this method, class radius and kernel radius are defined to give an ambiguous phase boundary. The transition regions are separated and the corresponding model is expressed as a weight form of two neighboring phase models. Although this

* Corresponding author at: College of Information Engineering, Shenyang University of Chemical Technology, Shenyang, Liaoning Province, 110042, China. Tel./fax: +86 02489383358.
E-mail address: li-yuan@mail.tsinghua.edu.cn (Y. Li).

http://dx.doi.org/10.1016/j.chemolab.2015.04.007
0169-7439/

method has made some improvements for the performance of transition monitoring, the subjective parameters selection and the calculation of the membership degrees based on Euclidean distances may affect transition identification and modeling accuracy. Yao and Gao identified the transitions automatically without the specified parameters, which obtained better monitoring results [25]. However, the above methods for transition modeling based on a fixed prior relationship of transition with two neighboring phases. This kind of predefined and determinate transition model may be difficult to characterize transition pattern enough. Also, above mentioned researches on multiphase modeling only focused on how to isolate each individual phase and neglected the interrelationship between phases. Recently, Zhao et al. proposed a new method for transition identification and modeling based on a reconstruction idea, overcoming the shortcomings of above methods [26]. Each transition sample is reconstructed based on the extracted common bases and specific bases between two neighboring phases. The transition can be found by tracking the change of contribution coefficients of target phase model. The transition behaviors can be detected by checking the residual of model. However, there exit some shortcomings for this method. First, the phase center was defined before modeling so that modeling still based on the specific number of phase and the hard partition. Furthermore, the transition identification is not complete since the only start of each transition can be found. In addition, the common bases and specific bases are extracted based on the multiset variable correlation analysis (MsVCA) [27]. A two-step extraction may lead to excessive extraction so that the bases obtained may be not optimal. Zhang et al. developed the variable correlation analysis (VCA) method based on the objective function on covariance information and introduce VCA method to the kernel space for multimode process monitoring [28]. However, high-dimensional kernel matrix can bring to the difficulties in obtaining the eigenvalues and eigenvectors. Besides, transition information is not considered and a detailed illustration on statistics is not given. Recently, Ge et al proposed a new phase division method using a repeatability factor based on Euclidean distance and transition information is considered [29]. However, their work mainly focused on quality prediction, process monitoring has not been considered.

In this paper, new indexes are proposed to achieve phase division based on a new repeatability factor. For batch processes, repeatability between batches is an important characteristic. This characteristic can be used to achieve phase division since the repeatability in the steady phase is higher than in the transition. Therefore, the key is to find an index to reflect the repeatability information in the real sense so that phase division can be well achieved. This paper makes a further understanding on repeatability from a geometrical view. On the one hand, repeatability between two batches may depend on the similarity of space location corresponding data points. On the other hand, it also depends on the similarity of point attribution to the cluster. Based on this understanding, an improved repeatability factor, which based on the diffusion distance, is used to describe the repeatability between batches for phase division. Compared with the conventional repeatability factor, which only measure the similarity between data points, the improved repeatability factor can further reflect the intrinsic geometry of the dataset. This intrinsic geometry provides the underlying information on the local density distribution. The bigger density distribution is, the bigger repeatability is. Moreover, a pair of points in the same cluster has bigger density distribution than those in different cluster. Therefore, the improved repeatability factor can reflect the similarity of attribute to the cluster beside the similarity between two points. Based on the improved repeatability, the repeatability characteristic between batches can be described in the true sense, which improves the performance of phase division.

Based on the results of phase division, modeling and monitoring for multi-phase batch are developed. For multi-phase batch processes, modeling can include the steady phase modeling and the transition modeling. Due to different process characteristics between steady phases and transitions, different modeling schemes should be adopted. For steady phases, the TSFVS-KVCA method is proposed for nonlinear process monitoring. The first step feature vector selection is performed on each steady phase, and these selected feature vectors construct a subset of original phase data, which is called as the within-phase subset. Then, the second step feature vector selection is carried out on a dataset constructed by within-phase subsets of two neighboring phases and between-phase subset is obtained. Each within-phase subset or each between-phase subset forms the basis vectors, which has the well capacity of generalization on original data in feature space. Furthermore, a two-step feature selection can effectively reduce the size of original dataset avoiding the computational complexity and the instability of decomposition for high dimensional kernel. In addition, the TSFVS-KVCA method solves the excessive extraction problems and provides new sub-basis vectors to substitute for the traditional linear combination. It shows that the combination of TSFVS with KVCA is suitable and meaningful. Based on a series of subsets, the common bases and specific bases are extracted. The steady phase is separated into different subspaces. Then, the statistics derivation in different subspaces is given for nonlinear process monitoring.

For steady phases, process characteristic keeps similar, so a determinate model is suitable. However, for transitions, the process pattern is gradually changeable. Regularly, the process pattern of a transition sample may be quite alike the nearer phase. Because of dynamics and uncertainty, constructing a determinate model to describe the transition pattern is not reliable. Accordingly, it would be more desirable that modeling for transitions based on a dynamic model without the prior information. Although there exits the switch of transition pattern, it is certain that some inherent characteristics still remains unchanged. The extracted bases based on the TSFVS-KVCA method just can reflect the similarity and the dissimilarity between two neighboring phases. They not only achieve the modeling for steady phase, but also provide important information for transition sample analysis and modeling. The transition pattern can be described by the common bases between two neighboring phases and the specific bases in the starting and target phase, respectively. The change of transition is mainly reflected by the switch of specific bases, since the common bases between the two neighboring phases keeps unchanged during the transition. Although the transition covers two types of phase-specific characteristics, the fact is that the only one plays a dominant role to describe the transition pattern. It is why the transition pattern is more similar to the start phase at the beginning, whereas it behaves more similar to the target phase at the end. The dominant role of specific bases is switching from the start phase to the target phase as the transition progresses. The time interval at which the dominant role of specific bases switches from one phase to the other phase is called the switch point and the phase of which specific bases plays a dominant role is called the dominant phase. The transition pattern can be well described based on the common bases and the specific bases of dominant phase at current time interval since the variation caused by the specific bases of non-dominant is equal to the one that a disturbance within the normal limit caused. Therefore, the dynamic modeling for the transition pattern may mainly depend on how to choose the right specific bases to explain the transition pattern at each time interval. The dominant phase is determined based on which phase the nearest neighbor of the current transition sample is from. After the dominant phase is determined, a just-in-time model for transition sample is constructed based on the common bases and the specific bases of dominant phase for transition monitoring at the current time interval.

This paper can be composed as follows. In Section 2, the descriptions of proposed methods including phase division and process monitoring method are given. Then, the effectiveness and superiority of proposed method are demonstrated by applying it to the penicillin fermentation process. Conclusions are drawn in the last section.