Chinese Journal of Chemical Engineering 23 (2015) 1357–1363

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Process Systems Engineering and Process Safety

An aligned mixture probabilistic principal component analysis for fault detection of multimode chemical processes☆

Yawei Yang, Yuxin Ma, Bing Song, Hongbo Shi *

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

ARTICLE INFO

Article history:
Received 3 September 2014
Received in revised form 5 January 2015
Accepted 25 January 2015
Available online 20 March 2015

Keywords:
Multimode process monitoring
Mixture probabilistic principal component analysis
Model alignment
Fault detection

ABSTRACT

A novel approach named aligned mixture probabilistic principal component analysis (AMPPCA) is proposed in this study for fault detection of multimode chemical processes. In order to exploit within-mode correlations, the AMPPCA algorithm first estimates a statistical description for each operating mode by applying mixture probabilistic principal component analysis (MPPCA). As a comparison, the combined MPPCA is employed where monitoring results are softly integrated according to posterior probabilities of the test sample in each local model. For exploiting the cross-mode correlations, which may be useful but are inadvertently neglected due to separately held monitoring approaches, a global monitoring model is constructed by aligning all local models together. In this way, both within-mode and cross-mode correlations are preserved in this integrated space. Finally, the utility and feasibility of AMPPCA are demonstrated through a non-isothermal continuous stirred tank reactor and the TE benchmark process.

1. Introduction

In modern industries, it is no longer a trivial task to improve process operation standards and product quality levels, especially to reduce the variation of the main quality features around their target values. With the increasing availability of data through faster and more informative sensors, multivariate statistical process monitoring (MSPM) has been applied to various processes and gained great success, among which principal component analysis (PCA) is the most everlasting algorithm used [1]. Due to various market demands for different products and the changes in feedstock or manufacturing strategies, data characteristics among different operating modes are easily distinguished [2]. Hence, evaluated confidence levels of traditional monitoring statistics such as Hotelling's T^2 and squared prediction error (SPE) will be drastically skewed, and the monitoring performance will be discounted and more seriously, the monitoring approach will completely lose effect. In order to deal with multimode characteristics, some progresses have been made in three different modeling manners.

Intuitively, several individual monitoring models can be built for multiple operating modes, but two essential problems need to be addressed. One is how to cluster the training data according to operating modes. Obviously, traditional clustering algorithms such as k-nearest neighbors and fuzzy c-means method can be directly employed. To improve the clustering accuracy, Zhu *et al.* [3] utilized a bilayer clustering approach, which consists of an ensemble moving window strategy and an ensemble clustering solutions strategy to obtain partitioned subgroups. Zhao [4] proposed a concurrent phase partition and between-mode statistical modeling strategy for batch processes with multiple phases and operating modes. The other problem is how to perform the online monitoring approach. If only one sub-model is applied at one sampling moment, specific rules must be developed to determine the most suitable sub-model. The distance between the sample and the center of local models can be used as an indicator [5] and the minimum SPE value is also effective for choosing a proper local PCA model [6]. Feital *et al.* [7] proposed a component-wise identification approach to estimate the current operating condition through optimizing a maximum likelihood objective function. If all sub-models are trustworthy, their monitoring results can be involved with respect to their posterior probabilities. Based on Bayesian inference, Yu and Qin [8] developed a global probabilistic index while Xie and Shi [9] developed an integrated Mahalanobis distance monitoring statistic. Similarly, Ge and Song [10, 11] combined the results achieved by some mixture modeling strategies such as MPPCA and mixture factor analysis.

Meanwhile, some adaptive methods have been improved so that they can tackle both time-varying and multimode behavior. Jin *et al.* [12] added a set of if–then rules into the monitoring scheme of recursive PCA from some extracted process factors for online identification of mode changes. With sufficient data of possible operating modes collected beforehand, Ge and Song [13] proposed a real-time modeling procedure with the online regression model which is built based on the relevant data set picked out from the historical database. Xie and Shi [14] and Yu [15] built Gaussian mixture models in the offline phase and performed the online updating in two different dynamic fashions

☆ Supported by the National Natural Science Foundation of China (61374140) and Shanghai Pujiang Program (12PJ1402200).
* Corresponding author.
E-mail address: hbshi@ecust.edu.cn (H. Shi).

http://dx.doi.org/10.1016/j.cjche.2015.01.014
1004-9541/

with a moving window strategy and a particle filter re-sampling approach. By considering the non-Gaussian characteristic, Ma *et al.* [16] proposed a moving window local outlier factor algorithm.

Recently, building a global model has been proven to be another effective way to cope with multimode problems. Ma *et al.* [17] introduced a local standardization strategy into the modeling phase to normalize the distance between samples and its neighbors. Ghosh and Srinivasan [18] put forward a process monitoring scheme based on the immune-inspired negative selection principle where samples from normal operating conditions and known faults are utilized to construct a non-self-space in the form of a collection of spherical detectors.

In the existing approaches, within-mode correlations have been intensively analyzed, so that a high resolution can be offered by the constructed local models. However, cross-mode correlations are often inadvertently neglected. Zhang *et al.* [19,20] have claimed that this significant issue should be taken into consideration. As effective improvements, they have proposed two approaches to extract the common information between different modes and calculated the left specific information of each individual mode accordingly. Nevertheless, the question which monitoring model is adopted online still must be answered because multiple models are built offline.

Inspired by the work of Teh and Roweis [21], correlations among different operating modes can be preserved by integrating all local models together into a global one. In this study, to take full account of within-mode and cross-mode correlations for multimode process monitoring, a novel method named aligned mixture probabilistic principal component analysis is proposed. The main contribution of this paper is that a new angle of solving the multimode problem is introduced through aligning all constructed models. First, multiple local probabilistic models are constructed by using MPPCA and responsible levels of each sample are estimated with respect to each local model. Then, by minimizing the information loss of the alignment, all local models are merged into a low-dimensional coordinated space. Finally, remaining monitoring procedures including the definition of statistics and the online evaluation of new samples are conducted in the coordinated space.

2. Fault Detection Based on AMPPCA

For exploiting both within-mode and cross-mode correlations, MPPCA is employed first to build multiple local probabilistic models corresponding to the known operating modes. Then an automatic alignment procedure is conducted, in which a global coherent coordinate system for the original data is learned by mapping the disparate local representations together.

2.1. Construction of multiple local models based on MPPCA

To achieve the final low-dimensional coordination $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N] \in \boldsymbol{R}^{d \times N}$, MPPCA is first employed to automatically estimate several local models based on training data $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N] \in \boldsymbol{R}^{D \times N}$ collected from the known operating modes. Let the number of operating modes be K and the overall distribution of the mixture latent variable models should take the following form.

$$p(\boldsymbol{X}) = \sum_{\omega=1}^{K} \int p(\boldsymbol{X}|\boldsymbol{Z}, \omega) p(\boldsymbol{Z}|\omega) p(\omega) \mathrm{d}\boldsymbol{Z} \tag{1}$$

where $p(\omega)$ is constrained as $\sum_{\omega=1}^{K} p(\omega) = 1$, which represents the prior mixing probability of local models. $p(\boldsymbol{X}|\boldsymbol{Z}, \omega)$ and $p(\boldsymbol{Z}|\omega)$ can be calculated from $N(\boldsymbol{\mu}_\omega, \boldsymbol{\Lambda}_\omega \boldsymbol{\Lambda}_\omega^{\mathrm{T}} + \sigma_\omega^2 \boldsymbol{I})$ and $N(0, \boldsymbol{I})$, respectively. To be more specific, the aim of applying MPPCA is to estimate parameter sets $\boldsymbol{\theta}_\omega = \{\boldsymbol{\mu}_\omega, \boldsymbol{\Lambda}_\omega, \sigma_\omega^2, p(\omega)\} (\omega = 1, 2, \ldots, K)$ and hence a standard expectation and maximization (EM) algorithm is derived by considering the complete-data log likelihood.

2.1.1. E-step

The posterior mixing probability $p(\omega|\boldsymbol{x}_i, \boldsymbol{\theta}_\omega^{\mathrm{old}})$ and the latent variable posterior probability $p(\boldsymbol{z}_i|\boldsymbol{x}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}})$ of each sample can be determined as follows.

$$p\left(\omega \middle| \boldsymbol{x}_i, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) = p\left(\boldsymbol{x}_i \middle| \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) p\left(\omega \middle| \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) / p\left(\boldsymbol{x}_i \middle| \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) \tag{2}$$

$$p\left(\boldsymbol{z}_i \middle| \boldsymbol{x}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) = \frac{p\left(\boldsymbol{x}_i \middle| \boldsymbol{z}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) p\left(\boldsymbol{z}_i \middle| \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right)}{p\left(\boldsymbol{x}_i | \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right)} \tag{3}$$

where $p(\boldsymbol{x}_i|\boldsymbol{z}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}})$ follows the Gaussian distribution $N(\boldsymbol{\Lambda}_\omega^{\mathrm{old}} \boldsymbol{z}_i + \boldsymbol{\mu}_\omega^{\mathrm{old}}, (\sigma_\omega^{\mathrm{old}})^2 \boldsymbol{I})$ and thus the expectation of latent variables and its second moment can be calculated.

$$\mathrm{E}\left(\boldsymbol{z}_i \middle| \boldsymbol{x}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) = \left(\sigma_\omega^2 \boldsymbol{I} + \left(\boldsymbol{\Lambda}_\omega^{\mathrm{old}}\right)^{\mathrm{T}} \boldsymbol{\Lambda}_\omega^{\mathrm{old}}\right)^{-1} \left(\boldsymbol{\Lambda}_\omega^{\mathrm{old}}\right)^{\mathrm{T}} \left(\boldsymbol{x}_i - \boldsymbol{\mu}_\omega^{\mathrm{old}}\right) \tag{4}$$

$$\begin{aligned} \mathrm{E}\left(\boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}} \middle| \boldsymbol{x}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) &= \mathrm{E}\left(\boldsymbol{z}_i \middle| \boldsymbol{x}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) \mathrm{E}\left(\boldsymbol{z}_i^{\mathrm{T}} \middle| \boldsymbol{x}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) \\ &\quad + \sigma_\omega^2 \left(\sigma_\omega^2 \boldsymbol{I} + \left(\boldsymbol{\Lambda}_\omega^{\mathrm{old}}\right)^{\mathrm{T}} \boldsymbol{\Lambda}_\omega^{\mathrm{old}}\right)^{-1} \end{aligned} \tag{5}$$

2.1.2. M-step

First, the prior mixing probability of a local model can be updated by summing up the posterior mixing probabilities of all samples relative to this model.

$$p(\omega|\boldsymbol{\theta}_\omega^{\mathrm{new}}) = \sum_{i=1}^{N} p\left(\omega \middle| \boldsymbol{x}_i, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) / N \tag{6}$$

By setting the derivative of the expectation of complete-data log likelihood to be zero with respect to $\boldsymbol{\mu}_\omega$, $\boldsymbol{\Lambda}_\omega$ and σ_ω^2, their updated values can be calculated.

$$\boldsymbol{\mu}_\omega^{\mathrm{new}} = \left(\sum_{i=1}^{N} p\left(\omega \middle| \boldsymbol{x}_i, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) \cdot \boldsymbol{x}_i\right) / \sum_{i=1}^{N} p\left(\omega \middle| \boldsymbol{x}_i, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) \tag{7}$$

$$\begin{aligned} \boldsymbol{\Lambda}_\omega^{\mathrm{new}} &= \left(\sum_{i=1}^{N} p\left(\omega \middle| \boldsymbol{x}_i, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) (\boldsymbol{x}_i - \boldsymbol{\mu}_\omega^{\mathrm{new}}) \mathrm{E}\left(\boldsymbol{z}_i^{\mathrm{T}} \middle| \boldsymbol{x}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right)\right) \\ &\quad \times \left(\sum_{i=1}^{N} p\left(\omega \middle| \boldsymbol{x}_i, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) \mathrm{E}\left(\boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}} \middle| \boldsymbol{x}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right)\right)^{-1} \end{aligned} \tag{8}$$

$$\begin{aligned} (\sigma_\omega^{\mathrm{new}})^2 = \sum_{i=1}^{N} p\left(\omega \middle| \boldsymbol{x}_i, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) \Big(&\|\boldsymbol{x}_i - \boldsymbol{\mu}_\omega^{\mathrm{new}}\|^2 - 2\mathrm{E}\left(\boldsymbol{z}_i^{\mathrm{T}} \middle| \boldsymbol{x}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) \cdot (\boldsymbol{\Lambda}_\omega^{\mathrm{new}})^{\mathrm{T}} (\boldsymbol{x}_i - \boldsymbol{\mu}_\omega^{\mathrm{new}}) \\ &+ tr\left(\mathrm{E}\left(\boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}} \middle| \boldsymbol{x}_i, \omega, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right) (\boldsymbol{\Lambda}_\omega^{\mathrm{new}})^{\mathrm{T}} \boldsymbol{\Lambda}_\omega^{\mathrm{new}}\right) \Big) \Big/ \left(D \cdot \sum_{i=1}^{N} p\left(\omega \middle| \boldsymbol{x}_i, \boldsymbol{\theta}_\omega^{\mathrm{old}}\right)\right) \end{aligned} \tag{9}$$

2.2. Alignment of local latent subspaces

As pointed out in [21], rather than considering the collected variables from different clusters to be single quantities, such as performing density estimation or dimension reduction separately, it is possible to view them as networks that convert high-dimensional inputs into a vector of internal coordinates from each sub-model, accompanied by responsibilities. Hence, the final coordination can be obtained by averaging the guesses using the responsibilities as weights $\boldsymbol{Y} = \boldsymbol{L}^{\mathrm{T}} \boldsymbol{U}$, where the integrated matrix $\boldsymbol{U} \in \boldsymbol{R}^{Km \times N}$ and its general element is $\boldsymbol{U}_{i,\omega} = p(\omega|\boldsymbol{x}_i, \boldsymbol{\theta}) \cdot \mathrm{E}(\boldsymbol{z}_i|\boldsymbol{x}_i, \omega, \boldsymbol{\theta})$. To evaluate the mapping matrix $\boldsymbol{L}$, a