Chinese Journal of Chemical Engineering 23 (2015) 981–991

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Process Systems Engineering and Process Safety

Adaptive partitioning PCA model for improving fault detection and isolation☆

Kangling Liu [1], Xin Jin [1], Zhengshun Fei [2], Jun Liang [1,*]

[1] State Key Lab of Industrial Control Technology, Institute of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China
[2] School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China

ARTICLE INFO

Article history:
Received 25 April 2014
Received in revised form 29 August 2014
Accepted 5 September 2014
Available online 21 March 2015

Keywords:
Adaptive partitioning
Fault detection
Fault isolation
Principal component analysis

ABSTRACT

In chemical process, a large number of measured and manipulated variables are highly correlated. Principal component analysis (PCA) is widely applied as a dimension reduction technique for capturing strong correlation underlying in the process measurements. However, it is difficult for PCA based fault detection results to be interpreted physically and to provide support for isolation. Some approaches incorporating process knowledge are developed, but the information is always shortage and deficient in practice. Therefore, this work proposes an adaptive partitioning PCA algorithm entirely based on operation data. The process feature space is partitioned into several sub-feature spaces. Constructed sub-block models can not only reflect the local behavior of process change, namely to grasp the intrinsic local information underlying the process changes, but also improve the fault detection and isolation through the combination of local fault detection results and reduction of smearing effect. The method is demonstrated in TE process, and the results show that the new method is much better in fault detection and isolation compared to conventional PCA method.

1. Introduction

Proper fault detection and isolation have become an interesting research area in recent years, with growing demands for high performance, efficiency, reliability and process safety. As a data-driven methodology for extracting process information from massive data and interpreting them, PCA is a widely applied dimension reduction technique for capturing strong correlation underlying in the process measurements under normal operation condition, expressed as a statistical model [1–3]. A process fault is detected if the measured process behavior violates the control limits defined by the model. Typically, the auto-correlated principal component values or scores and the squared prediction error (SPE) are applied for fault detection, named as T^2 statistics and SPE statistics, respectively. Contribution plots are popularly adopted to identify and isolate the faulty variables causing the process to violate its normal operation region by determining the contribution of each variable to the fault detection statistics [4]. The contribution plot approach is based on the assumption that the faulty variables have the largest contribution to the fault detection index. Although the contribution plot approach does not need prior fault information, it leads to obscure diagnosis results because the faulty variables increase the contribution of variables not affected by the fault, called "smearing" effect [5].

PCA gradually presents some limitations for fault detection and isolation. One major limitation is poor physical interpretation. With PCA model, each extracted principal component is a linear combination of almost all observed variables, making each principal component difficult to interpret physically and, in turn, providing ambiguous information for fault isolation. Various approaches have been proposed to address this drawback of PCA [6–9]. Some approaches such as sparse PCA [6] and shrinking PCA [7] have been proposed to construct optimization problem, minimizing the variance of principal components and the number of nonzero elements in the loadings. Thus, sparse loading vectors with a few nonzero elements are obtained and easier to interpret. Multi-block methods [8,9] have been proposed to improve both the fault detection performance and fault interpretation, but these methods need some process knowledge for block partitioning, which is not always available, especially in complex chemical processes. To overcome this, some improved multi-block PCA approaches have been proposed, totally based on process data [10–12].

Another major limitation of PCA based fault detection and isolation approach is that the PCA model is time-invariant. Chemical processes are often slow-varying and with normal disturbances, so the application of a fixed model based fault detection and isolation might lead to mistake and fail to report the warning of fault. Huang *et al.* [13] proposed mixture discriminant monitoring, integrating supervised learning and statistical process control charting techniques, which also utilizes both normal and faulty historical data in process modeling. Lee *et al.* [14] developed

☆ Support by the National Natural Science Foundation of China (61174114), the Research Fund for the Doctoral Program of Higher Education in China (20120101130016) and Zhejiang Provincial Science and Technology Planning Projects of China (2014C31019).
* Corresponding author.
E-mail address: jliang@iipc.zju.edu.cn (J. Liang).

http://dx.doi.org/10.1016/j.cjche.2014.09.052
1004-9541/

adaptive multi-block PCA, adaptive consensus PCA, and adaptive multiscale PCA algorithms for updating the model structure to deal with changing process. Zhao and Sun [15] presented relative PCA and multiple time region based fault reconstruction modeling algorithms for fault subspace extraction and online fault diagnosis. However, these approaches are ineffective when the historical fault information is shortage and deficient, which is a problem commonly present in complex chemical processes.

To address the above issues, this work proposes an adaptive partitioning PCA (APPCA) algorithm. The advantages of APPCA method are as follows. (1) It is easier for modeling and interpretation of the collected data matrix from process by decomposing it into several blocks. (2) The block partition step is carried out entirely based on the process data, which makes use of both historical normal data and online operation data. It is of particular interest when the process knowledge is deficient. (3) A series of sub-block PCA models are constructed adaptively according to the operating process. It not only reflects the local behavior of process change but also enhances fault detection and isolation through the combination of individual fault detection results and reduction of the smearing effect. In this paper, the APPCA algorithm is demonstrated and applied to fault detection and isolation schemes. A case study of TE process is used to illustrate the application of APPCA.

2. Adaptive Partitioning Principal Component Analysis

Consider two process measurement spaces as the analysis subjects, $\boldsymbol{X}_{\mathrm{n}}(N_{\mathrm{n}}\times J)$ and $\boldsymbol{X}_{\mathrm{f}}(N_{\mathrm{f}}\times J)$, each consisting of the same number of variables J and possibly different numbers of samples N_{n} and N_{f}. $\boldsymbol{X}_{\mathrm{n}}$ is the normal data matrix and $\boldsymbol{X}_{\mathrm{f}}$ is the faulty data matrix collected from one fault case. Subscript n denotes the normal case and subscript f denotes the fault case. Therefore, a global PCA model can be built based on the normal data matrix $\boldsymbol{X}_{\mathrm{n}}(N_{\mathrm{n}}\times J)$ as follows.

$$\begin{cases}\boldsymbol{T}_{\mathrm{n}}=\boldsymbol{X}_{\mathrm{n}}\boldsymbol{P}_{\mathrm{n}}\\ \boldsymbol{X}_{\mathrm{n}}=\boldsymbol{T}_{\mathrm{n}}\boldsymbol{P}_{\mathrm{n}}{}'+\boldsymbol{E}_{\mathrm{n}}=\hat{\boldsymbol{X}}_{\mathrm{n}}+\boldsymbol{E}_{\mathrm{n}}\end{cases}\tag{1}$$

where $\boldsymbol{T}_{\mathrm{n}}(N_{\mathrm{n}}\times R)$ and $\boldsymbol{P}_{\mathrm{n}}(J\times R)$ stand for the PCs and loadings, respectively, matrices $\hat{\boldsymbol{X}}_{\mathrm{n}}$ and $\boldsymbol{E}_{\mathrm{n}}$ denote modeled variances of $\boldsymbol{X}_{\mathrm{n}}$ and residuals (un-modeled variances of $\boldsymbol{X}_{\mathrm{n}}$), respectively, and R is the number of retained PCs. The retained PCs should represent most of the normal variability of process in an optimal way. As a simple and effective approach for selecting the number of principal components, CPV is popularly used. In this paper, CPV approach with 85% of normal variability can basically balance the amount of parsimony and comprehensiveness of retained PCs, so it is employed.

For the faulty data matrix $\boldsymbol{X}_{\mathrm{f}}(N_{\mathrm{f}}\times J)$, variable scores $\boldsymbol{T}_{\mathrm{f}}$ and residuals $\boldsymbol{E}_{\mathrm{f}}$ can be obtained as

$$\begin{cases}\boldsymbol{T}_{\mathrm{f}}=\boldsymbol{X}_{\mathrm{f}}\boldsymbol{P}_{\mathrm{n}}\\ \boldsymbol{E}_{\mathrm{f}}=\boldsymbol{X}_{\mathrm{f}}-\boldsymbol{T}_{\mathrm{f}}\boldsymbol{P}_{\mathrm{n}}{}'\end{cases}\tag{2}$$

The residual space typically describes the material and energy balances, which is much sensitive to process faults [7]. When variable residuals diverge significantly, the existence of a fault associated with the breakdown of the correlation among variables. Since an interpretable residual space can provide useful information for isolation to facilitate the detection and isolation of nascent faults, the variable partitioning method is performed based on the PCA residuals of operation data. First, Pearson's correlation coefficient $\boldsymbol{R}_{\mathrm{f}}$ of variable residuals $\boldsymbol{E}_{\mathrm{f}}$ is calculated, and the element of row i and column j in $\boldsymbol{R}_{\mathrm{f}}$ shows the correlation coefficient between variables i and j, represented as [16]

$$\boldsymbol{R}_{\mathrm{f}}(i,j)=\frac{\boldsymbol{E}\left[\left(\boldsymbol{E}_{\mathrm{f},i}-\overline{\boldsymbol{E}_{\mathrm{f},i}}\right)\left(\boldsymbol{E}_{\mathrm{f},j}-\overline{\boldsymbol{E}_{\mathrm{f},j}}\right)\right]}{\sqrt{\sum_{i=1}^{N_{\mathrm{f}}}\left(\boldsymbol{E}_{\mathrm{f},i}-\overline{\boldsymbol{E}_{\mathrm{f},i}}\right)^{2}}\sqrt{\sum_{j=1}^{N_{\mathrm{f}}}\left(\boldsymbol{E}_{\mathrm{f},j}-\overline{\boldsymbol{E}_{\mathrm{f},j}}\right)^{2}}}\tag{3}$$

where $\boldsymbol{E}_{\mathrm{f},i}$ and $\boldsymbol{E}_{\mathrm{f},j}$ denote residuals of the ith and jth variables, respectively, $\overline{\boldsymbol{E}_{\mathrm{f},i}}$ is the mean value of $\boldsymbol{E}_{\mathrm{f},i}$ and $\bar{\boldsymbol{E}}_{\mathrm{f},j}$ is the mean value of $\boldsymbol{E}_{\mathrm{f},j}$. Then the t-test is applied to $\boldsymbol{R}_{\mathrm{f}}$, resulting in significance matrix $\boldsymbol{S}_{\mathrm{f}}$, each s_{f}-value in the matrix $\boldsymbol{S}_{\mathrm{f}}$ denotes the probability of getting a correlation when the true correlation is zero, ranging from 0 to 1. Smaller s_{f}-value means that the correlation of the two variables is more significant. If s_{f}-value is sufficiently small, say less than 0.05 for 95% confidence interval, the correlation between the two variables is significant. Lastly, based on matrix $\boldsymbol{S}_{\mathrm{f}}$, the variables are partitioned into some sub-blocks using the complete linkage algorithm [9], and variables within a predefined positive threshold are grouped into a sub-block.

How to determine the number of sub-blocks or the pre-defined positive threshold will greatly influence the performance of fault detection and isolation. Ideally, each sub-block should be well-connected and any two sub-blocks should be well-separated. Furthermore, the number of sub-blocks should be sufficiently small to ensure manageability, while the number of sub-blocks should be sufficiently large to warrant the superiority of APPCA. Therefore, the following optimization problem is constructed to determine the number of sub-blocks b and the value b_{f} is the solution, which results in sub-blocks C_i $\{i=1,2,\ldots,b_{\mathrm{f}}\}$:

$$\begin{aligned}b_{\mathrm{f}}&=\arg\min\frac{b}{b_{\max}}+\lambda\sum_{i=1}^{b}f(T_i)\\ \text{s.t.}T_i&=\max_{j,k\in C_i,j\neq k}\boldsymbol{S}_{\mathrm{f}}(j,k)\\ T&=\max_{i}T_i\end{aligned}\tag{4}$$

where T denotes the positive threshold, within which all variables are grouped, so b is determined by T; $b_{\max}$ denotes the maximum allowable number of sub-blocks ($b\le b_{\max}$) through lots of trial, which can be 10 in the case study of this work. $\lambda>0$ is employed to tune a compromise between the number of sub-blocks and the maximum distance T_i between two variables in the ith sub-block. In this work, $b_{\max}=10$ and $\lambda=1$. $f(T_i)$ is a piecewise linear function of T_i as follows:

$$f(T_i)=\begin{cases}T_i-T_{\max} & T>T_{\max}\\ 0 & T\le T_{\max}\end{cases}\tag{5}$$

where $T_{\max}$ is a threshold value, which defaults to 0.05 in this work. Thus, the normal data matrix $\boldsymbol{X}_{\mathrm{n}}(N_{\mathrm{n}}\times J)$ is partitioned into b_{f} sub-blocks $\boldsymbol{X}_{\mathrm{n},i}(N_{\mathrm{n}}\times J_i)$ $(i=1,2,\ldots,b_{\mathrm{f}})$, normal data matrix of J_i variables in the ith sub-block C_i and $J_1+J_2+\ldots+J_b=J$. Therefore, the b_{f} sub-block PCA models based on $\boldsymbol{X}_{\mathrm{n},i}(N_{\mathrm{n}}\times J_i)$ $(i=1,2,\ldots,b_{\mathrm{f}})$ is

$$\begin{cases}\boldsymbol{T}_{\mathrm{n},i}=\boldsymbol{X}_{\mathrm{n},i}\boldsymbol{P}_{\mathrm{n},i}\\ \boldsymbol{X}_{\mathrm{n},i}=\boldsymbol{T}_{\mathrm{n},i}\boldsymbol{P}_{\mathrm{n},i}{}'+\boldsymbol{E}_{\mathrm{n},i}=\hat{\boldsymbol{X}}_{\mathrm{n},i}+\boldsymbol{E}_{\mathrm{n},i}\end{cases}\tag{6}$$

where $\boldsymbol{T}_{\mathrm{n},i}(N_{\mathrm{n}}\times R_i)$ and $\boldsymbol{P}_{\mathrm{n},i}(J\times R_i)$ stand for PCs and loadings of sub-block C_i, respectively, matrices $\hat{\boldsymbol{X}}_{\mathrm{n},i}$ and $\boldsymbol{E}_{\mathrm{n},i}$ denote the modeled variances of $\boldsymbol{X}_{\mathrm{n},i}$ and residuals, and R_i is the number of retained PCs of sub-block C_i. CPV with 85% is used to select the number of PCs.

3. APPCA Based Fault Detection and Isolation

3.1. Fault detection

PCA partitions the measurement space into two orthogonal spaces: principal component subspace and residual subspace. Each measurement is identified by its score distance to the principal component subspace and the model error on the residual subspace. With the APPCA method, fault detection is carried out as follows.

Considering the data set $\boldsymbol{X}_{\mathrm{noc}}\in R^{Nc\times M}$ in normal operation condition (NOC) and new measurements $\boldsymbol{X}_{\mathrm{mea}}\in R^{Nm\times M}$, a series of models can be built, including one global PCA model and b sub-block PCA models. Besides, process variables are partitioned into b sub-blocks and $\boldsymbol{X}_{\mathrm{mea}}$ can be rearranged as

$$\boldsymbol{X}_{\mathrm{mea}}=[\boldsymbol{X}_1\quad\boldsymbol{X}_2\quad\cdots\quad\boldsymbol{X}_b]\tag{7}$$