Chinese Journal of Chemical Engineering 24 (2016) 775–786

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Process Systems Engineering and Process Safety

Online process monitoring for complex systems with dynamic weighted principal component analysis☆

Zhengshun Fei [1,*], Kangling Liu [2]

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

ARTICLE INFO

Article history:
Received 31 October 2015
Received in revised form 18 February 2016
Accepted 3 May 2016
Available online 31 May 2016

Keywords:
Principal component analysis
Weight
Online process monitoring
Dynamic

ABSTRACT

Conventional multivariate statistical methods for process monitoring may not be suitable for dynamic processes since they usually rely on assumptions such as time invariance or uncorrelation. We are therefore motivated to propose a new monitoring method by compensating the principal component analysis with a weight approach. The proposed monitor consists of two tiers. The first tier uses the principal component analysis method to extract cross-correlation structure among process data, expressed by independent components. The second tier estimates auto-correlation structure among the extracted components as auto-regressive models. It is therefore named a dynamic weighted principal component analysis with hybrid correlation structure. The essential of the proposed method is to incorporate a weight approach into principal component analysis to construct two new subspaces, namely the important component subspace and the residual subspace, and two new statistics are defined to monitor them respectively. Through computing the weight values upon a new observation, the proposed method increases the weights along directions of components that have large estimation errors while reduces the influences of other directions. The rationale behind comes from the observations that the fault information is associated with online estimation errors of auto-regressive models. The proposed monitoring method is exemplified by the Tennessee Eastman process. The monitoring results show that the proposed method outperforms conventional principal component analysis, dynamic principal component analysis and dynamic latent variable.

1. Introduction

Advanced manufacturing systems rely on an efficient process monitoring to increase the quality, efficiency and reliability of existing technologies [1,2]. Manufacturing process is usually highly complicated and lacks accurate models, which makes the model-based methods [3–5] unsuitable. However, floods of data can be obtained on-line through sensors embedded in the process. This situation facilitates the development of multivariate statistical process monitoring method based on principal component analysis (PCA) [6] that utilizes process data and requires no explicit process knowledge. PCA is widely used in many applications because of its advantage of handling high dimensional and correlated process variables [7–10]. For process monitoring, PCA partitions the process data space into a principal component subspace and a residual subspace, and uses T^2 and Q statistics to monitor the two subspaces respectively.

On the other hand, manufacturing applications are generally dynamic processes and process variables exhibit auto-correlations because of controller feedback and disturbances. Here, auto-correlation means that current observation is correlated with previous ones. As a result, conventional multivariate statistical methods, which rely on assumptions that (1) the process is time invariant and (2) variables are serially uncorrelated, have the tendency to generate false alarms or missed detection [11]. This mismatch suggests that a dynamic method analyzing serial correlations is needed [11–14]. Some speech recognition approaches, such as hidden Markov model [15] and dynamic time warping [16], were developed for off-line diagnosis. These approaches rely heavily on known fault information, obviously, it is often not complete since we cannot ensure that all possible faults are pre-defined in complex systems. Ku *et al.* [11] proposed a dynamic PCA (DPCA) that constructs singular value decomposition on an augmented data matrix containing time lagged process variables, which increases the size of variable set and has difficulty in model interpretation [17–19]. With the similar idea, some subspace methods based on canonical variate analysis [20] and consistent DPCA [17] were proposed. Bakshi [12] introduced a multi-scale PCA that integrates PCA with wavelet analysis, which is an effective tool to monitor auto-correlated observations without matrix augmentation. Multi-scale PCA first decomposes process data into several time-scales using the wavelet analysis and then establishes

☆ Supported by the National Natural Science Foundation of China (61174114), the Research Fund for the Doctoral Program of Higher Education in China (20120101130016), the Natural Science Foundation of Zhejiang Province (LQ15F030006), and the Science and Technology Program Project of Zhejiang Province (2015C33033).
* Corresponding author.
E-mail address: zsfei@zju.edu.cn (Z. Fei).

http://dx.doi.org/10.1016/j.cjche.2016.05.038
1004-9541/

PCA on wavelet coefficients for different scales, and a moving window technique is used for online monitoring. A further analysis on multi-scale PCA was provided by Misra *et al.* [14]. Yoon [21] pointed out that MSPCA puts equal weights on different scales regardless of the scale contribution to overall process variation and then unreasonably increases the small contribution of high-frequency scales. Recently, Li and Qin [22] proposed a dynamic latent variable (DLV) model to extract auto-correlation and cross-correlations. In particular, some probability methods were developed for dynamic process monitoring [23–25]. Choi [23] constructed a Gaussian mixture model based on PCA and discriminant analysis for representing the distribution underlying dynamic data. Li and Fang [24] proposed an increasing mapping based on hidden Markov model for large-scale dynamic processes. Zhu and Ge [25] extended hidden Markov model to characterize the time-domain dynamics.

Inspired by these approaches, we propose a new monitoring method called dynamic weighted PCA (DWPCA), with the advantages that it is dynamic data driven and can detect faults in an automatic manner. The proposed method designs a hybrid correlation structure that simultaneously contains auto- and cross-correlation information of processes. The design includes two tiers. The first tier is to use the PCA method to extract the cross-correlation structure among process data, expressed by independent components, and the second tier is to estimate the auto-correlation structure among the extracted components as auto-regressive (AR) models. For online monitoring, we incorporate a weight approach into PCA. Actually, the weight approach is not new and has many applications such as correspondence search [26], face recognition [27] and process monitoring [28,29]. To the best of our knowledge, the weight method developed on a two-tier hybrid correlation structure is new for process monitoring. In this work, we use the weight approach to give different weights on different directions of components based on their contributions to a fault. Assume that fault information is associated with online estimation errors of AR models, a weight function is defined based on estimation errors for each component to take emphasis on directions of components, and its essential is that the directions are given high weight values if they have large estimation errors. The weight values are automatically computed when a new observation becomes available. Then, the computed weights can be used to dynamically partition the process data space into two new subspaces, namely an important component subspace and a remaining component subspace, and two new statistics are calculated to monitor them, with similar motivations of conventional PCA monitoring. But the differences are that (1) the proposed method makes use of online process operating information to actively perform subspace partition, (2) two new statistics take both auto- and cross-correlations into account while T^2 and Q statistics only consider cross-correlations, and (3) the contributions of component directions of the proposed method are not at the same degree while those of PCA are with the same value of 1.

The rest of this work is organized as follows. The conventional PCA is introduced briefly in Section 2. A simple process simulation is provided to illustrate problems of PCA monitoring based on T^2 and Q statistics. This gives rise to the motivations of DWPCA. In Section 3, DWPCA for process monitoring is detailed, including two new monitoring statistics. Tennessee Eastman process is employed to demonstrate the process monitoring performance of the proposed method in Section 4. The results show that the proposed method outperforms conventional PCA. Finally, Section 5 concludes the work.

2. Principal Component Analysis Monitoring

2.1. Principal component analysis

Suppose that a normal data set $\boldsymbol{X}\in\mathcal{R}^{N\times J}$ collecting N samples of J variables is scaled to have zero means and unit variances. The principal component analysis (PCA) decomposition is developed as $\boldsymbol{X}=\sum_{i=1}^{l}\boldsymbol{t}_i\boldsymbol{p}_i^{\mathrm{T}}+\sum_{i=l+1}^{J}\boldsymbol{t}_i\boldsymbol{p}_i^{\mathrm{T}}=\hat{\boldsymbol{T}}\hat{\boldsymbol{P}}^{\mathrm{T}}+\tilde{\boldsymbol{T}}\tilde{\boldsymbol{P}}^{\mathrm{T}}$. Here, $\boldsymbol{t}_i\in\mathcal{R}^{N\times 1}$ represents the ith component, and its direction $\boldsymbol{p}_i\in\mathcal{R}^{J\times 1}$ and variance $\lambda_i=\boldsymbol{t}_i^{\mathrm{T}}\boldsymbol{t}_i$ are eigenvector and eigenvalue of covariance matrix $\boldsymbol{S}_X=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}/(N-1)$. The components are in the order of variance decrease, *i.e.* $\lambda_1\geq\lambda_2\geq,\ldots,\geq\lambda_J$.

The first l components retained span a principal component subspace (PCS) and the remaining $J-l$ components represent a residual subspace (RS). The $\hat{\boldsymbol{T}}\in\mathcal{R}^{N\times l}$ and $\hat{\boldsymbol{P}}\in\mathcal{R}^{J\times l}$ are component and direction matrices in the PCS, respectively, and $\boldsymbol{T}\in\mathcal{R}^{N\times(J-l)}$ and $\tilde{\boldsymbol{P}}\in\mathcal{R}^{J\times(J-l)}$ correspond to the RS. To determine l, the cumulative percent variance (CPV) is widely used for its simplicity. For a particular observation $\boldsymbol{x}\in\mathcal{R}^{J\times 1}$, the T^2 and Q statistics are established for monitoring the two subspaces. In the PCS, $T^2=\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{P}}\boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{P}}^{\mathrm{T}}\boldsymbol{x}=\sum_{i=1}^{l}\boldsymbol{t}_i^2/\lambda_i\leq T^2_{\mathrm{lim}}$, where $\boldsymbol{\Lambda}=\mathrm{diag}\{\lambda_1,\lambda_2,\ldots,\lambda_l\}\in\mathcal{R}^{l\times l}$ is a diagonal matrix, and in the RS, $Q=\boldsymbol{x}^{\mathrm{T}}\tilde{\boldsymbol{P}}\tilde{\boldsymbol{P}}^{\mathrm{T}}\boldsymbol{x}=\sum_{i=l+1}^{J}\boldsymbol{t}_i^2\leq Q_{\mathrm{lim}}$. A fault is detected when the monitoring statistics violate their control limits T^2_{lim} and Q_{lim}.

2.2. Problems of PCA monitoring

Control limits for both statistics can be calculated from an F or weighted χ^2 distribution [30] with a confidence α, typically set $\alpha=$ 95% or 99%. In other words, a fault is detectable by PCA when its statistics must violate their corresponding control limits more than $(1-\alpha)\cdot 100\%$ times. The essential of PCA monitoring lies in detecting changes in the cross-correlation structure among components. PCA monitoring neglects dynamic information hidden in the data and it may be insensitive to changes in the component auto-correlation structure under the condition formulated in Fig. 1. In the PCS, T^2 is computed based on axes $\boldsymbol{t}_1/\sqrt{(\lambda_1)}$ and $\boldsymbol{t}_2/\sqrt{(\lambda_1)}$ that represent the directions $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ with maximum variances of λ_1 and λ_2, and in the RS, Q is

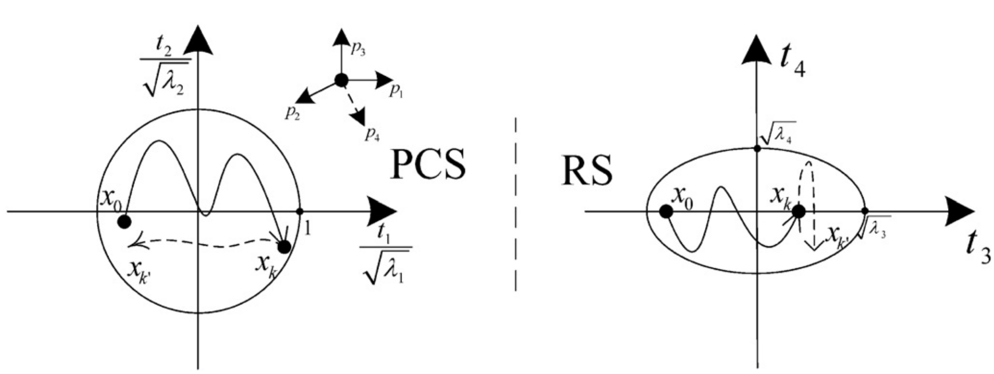

Fig. 1. Schematic illustration of problems of PCA monitoring.