Journal of Process Control 40 (2016) 24–34

Contents lists available at ScienceDirect

Journal of Process Control

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

On the use of reconstruction-based contribution for fault diagnosis

Hongquan Ji [a], Xiao He [a], Donghua Zhou [b,a,*]

[a] Tsinghua National Laboratory for Information Science and Technology (TNList) and Department of Automation, Tsinghua University, Beijing 100084, PR China

[b] College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, PR China

ARTICLE INFO

Article history:
Received 10 September 2015
Received in revised form 13 January 2016
Accepted 25 January 2016
Available online 15 February 2016

Keywords:
Multivariate statistical process monitoring
Fault diagnosis
Reconstruction-based contribution
Chi-square contribution
Principal component analysis

ABSTRACT

In the multivariate statistical process monitoring (MSPM) area, principal component analysis (PCA) and reconstruction-based contribution (RBC) are two commonly used techniques for fault detection and fault diagnosis problems, respectively. This paper starts with a review of the two methods. It is then pointed out that, when the dimensionality of the principal component subspace or the residual subspace in the PCA model is equal to 1, several fault detection indices based RBC will be invalid for fault diagnosis. Corresponding geometric interpretations of the invalidation cases are illustrated intuitively according to the definition of RBC. In order to perform effective fault diagnosis in such invalidation cases, three methods including the available combined index based RBC, the derived Mahalanobis distance based RBC, and the proposed chi-square contribution (CSC) are introduced. The CSC is constructed by employing a moving window and the effect of the window width on its diagnosis performance is investigated. The failure cases of the RBC, the effectiveness of the proposed CSC, as well as the comparison of these three methods for fault diagnosis are demonstrated by case studies on two numerical examples and a simulated three-tank system.

1. Introduction

The past decades have witnessed the rapid development of the fault detection and diagnosis (FDD) technique, which is of vital importance to ensure safe and efficient operation of industrial processes. Compared with model-based FDD approaches, data-driven FDD techniques have attracted increasing research attention [1–3]. This is due to that modern industrial processes are getting more and more complex and exact mathematical models for complex and large-scale systems are generally hard to obtain [4]. As a main branch of data-driven FDD, multivariate statistical process monitoring (MSPM) utilizes the data information such as correlation and applies multivariate analysis techniques to accomplish fault detection and fault diagnosis tasks [5,6]. During the last two decades, many theoretical results and successful applications to industrial processes have been reported in the literature [7–10].

In the field of MSPM, principal component analysis (PCA) has been successfully applied to model the data collected under normal operating conditions and subsequently, determine whether abnormal situations occur or not in new measurements using the established model [11,12]. In PCA-based techniques, measurements are typically partitioned into two subspaces, i.e., the principal component subspace (PCS) and the residual subspace (RS). Fault detection can then be implemented in the subspaces by comparing the fault detection indices with their corresponding thresholds. Though several fault detection indices have been proposed based on the PCA model, the most popular and widely used ones are the Hotelling's T^2 statistic and Q statistic (or squared prediction error, SPE). These indices can be unified and represented by a general form since they are all quadratic functions of the measurement [6,13].

Once a fault is detected by certain fault detection indices, a diagnosis tool should be implemented to identify the most responsible variables and further, investigate the possible root causes of the occurred fault. There have been a great many fault diagnosis methods reported in statistical process monitoring [10,13], among which the contribution plot approach [7,14] is the most popular one and has been widely used for diagnosis. However, it is known that contribution plots suffer from the smearing effect that can lead to misdiagnosis even for the simplest single sensor faults [15]. Utilizing the fault direction information, Dunia and Qin proposed the reconstruction-based approach for fault identification [16]. A benefit of this method is that the true fault can be isolated from a set of candidate faults without ambiguity, but the a priori knowledge

* Corresponding author at: College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, PR China.
Tel.: +86 053286058103.
E-mail address: zdh@tsinghua.edu.cn (D. Zhou).

http://dx.doi.org/10.1016/j.jprocont.2016.01.011

of the fault direction is needed. Recently, the reconstruction-based contribution (RBC) [15] using the PCA model was proposed for fault diagnosis, which inherits the merits of conventional contribution plots and the reconstruction-based approach. Though the smearing effect is not eliminated in the RBC method, it has been pointed out that RBC can guarantee correct diagnosis in the case of single sensor faults with large magnitudes.

Since the proposal of traditional RBC, it has been extended to be used with other data models besides PCA. Moreover, modified RBC methods have been developed to satisfy the needs of different processes or improve the diagnosis performance for more complex faults. To diagnose faults in nonlinear and dynamic processes, two extended methods of traditional RBC based on kernel PCA and dynamic PCA were proposed, respectively [17,18]. By extending traditional RBC to the case of known fault directions, a generalized RBC method with total projection to latent structures was proposed to diagnose output-relevant faults [19]. To reduce the complexity of fault diagnosis in large scale processes, a multi-block diagnosis strategy was presented [20]. More recently, a new fault diagnosis method based on weighted RBC was proposed to reduce the smearing in fault estimates and therefore improve the diagnosis accuracy [21]. Based on traditional RBC, the average residual-difference reconstruction contribution plot (ARdR-CP) was proposed to cope with the incipient sensor fault diagnosis problem [22]. Considering that a fault in practice may be in multidimensional directions, Mnassri et al. extended traditional RBC to the multidimensional case using PCA [23]. In addition, an alternative method called RBC ratio (RBCR) was proposed to diagnose more complex faults when no a priori knowledge of the fault direction information is available.

In most cases, traditional RBC and its variants or modifications can provide satisfying results when they are employed for fault diagnosis. In this paper, however, an exceptional situation of the effectiveness of traditional RBC is revealed. It is shown that when the dimensionality of the PCS in the PCA model is equal to 1, the T^2 statistic-based RBC may fail to diagnose the fault. Similar cases can be found when the dimensionality of the RS is equal to 1. Corresponding geometric illustrations of the invalidation cases are presented. To perform effective fault diagnosis, three methods including the available combined index based RBC, the derived Mahalanobis distance based RBC, and the proposed chi-square contribution (CSC) method are introduced. The reason why the combined index and Mahalanobis distance based RBC methods do not suffer from the invalidation cases is detailed. The CSC method belongs to a univariate fault diagnosis method. The moving window technique is utilized in CSC and the effect of the window width on its diagnosis performance is investigated.

The remainder of this paper is organized as follows. Basic theories that will be used in this paper are reviewed briefly in Section 2. The main results are presented in Section 3, which includes the revelation of three invalidation cases of RBC, corresponding geometric illustrations, introduction of the combined index and Mahalanobis distance based RBC methods, and the proposal of CSC. Section 4 presents case studies to illustrate the proposed theoretical results, followed by concluding remarks in Section 5.

2. Preliminaries

In this section, the PCA-based fault detection and RBC-based fault diagnosis approaches are reviewed briefly.

2.1. PCA for fault detection

In a process with m variables, a PCA model can be built using N measurements $\mathbf{x}(k) \in \mathbb{R}^m$, $k \in \{1, \ldots, N\}$, collected under normal operating conditions. A data matrix denoted by $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)]^T \in \mathbb{R}^{N\times m}$ can be formed with each row representing a measurement. Then, $\mathbf{X}$ is scaled to zero mean and unit variance for correlation-based PCA modeling. To obtain the principal and residual loadings of the model, the covariance matrix $\mathbf{S}$ of $\mathbf{X}$ is eigen-decomposed as follows

Table 1
Fault detection indices in PCA.

Index	Kernel matrix $\mathbf{M}$	Threshold Γ^2
T^2	$\mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}^T \triangleq \hat{\mathbf{D}}$	τ^2
SPE	$\tilde{\mathbf{P}}\tilde{\mathbf{P}}^T \triangleq \tilde{\mathbf{C}}$	δ^2
T_H^2	$\tilde{\mathbf{P}}\tilde{\boldsymbol{\Lambda}}^{-1}\tilde{\mathbf{P}}^T \triangleq \tilde{\mathbf{D}}$	ϵ^2
ϕ	$\tilde{\mathbf{C}}/\delta^2 + \hat{\mathbf{D}}/\tau^2 \triangleq \boldsymbol{\Phi}$	ζ^2
MD	$\bar{\mathbf{P}}\bar{\boldsymbol{\Lambda}}^{-1}\bar{\mathbf{P}}^T \triangleq \mathbf{D}$	η^2

$$\mathbf{S} = \frac{1}{N-1}\mathbf{X}^T\mathbf{X} = \bar{\mathbf{P}}\bar{\boldsymbol{\Lambda}}\bar{\mathbf{P}}^T = [\mathbf{P}\,\tilde{\mathbf{P}}]\begin{bmatrix}\boldsymbol{\Lambda} & \mathbf{0}\\ \mathbf{0} & \tilde{\boldsymbol{\Lambda}}\end{bmatrix}[\mathbf{P}\,\tilde{\mathbf{P}}]^T \tag{1}$$

where $\bar{\mathbf{P}}$ consists of m eigenvectors of $\mathbf{S}$ and $\bar{\boldsymbol{\Lambda}} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_m\}$ contains the corresponding eigenvalues arranged in descending order. $\mathbf{P} \in \mathbb{R}^{m\times l}$ and $\tilde{\mathbf{P}} \in \mathbb{R}^{m\times(m-l)}$ are the principal and residual loading matrices, respectively. The number of principal components (PCs) l can be determined by a certain criterion such as cumulative percent variance, average eigenvalue, parallel analysis, variance of the reconstruction error and so forth [24,25]. The columns of $\mathbf{P}$ and $\tilde{\mathbf{P}}$ span the PCS and RS, respectively. A sample vector $\mathbf{x}$ can therefore be decomposed as $\mathbf{x} = \hat{\mathbf{x}} + \tilde{\mathbf{x}}$, where $\hat{\mathbf{x}} = \mathbf{P}\mathbf{P}^T\mathbf{x}$ is the projection of $\mathbf{x}$ to the PCS and $\tilde{\mathbf{x}} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T\mathbf{x}$ is the projection of $\mathbf{x}$ to the RS.

Based on the PCA model, several fault detection indices have been proposed in the literature, among which the most popular and widely used ones are the Hotelling's T^2 statistic and SPE that measure the variability in the PCS and RS respectively. Hawkins' T_H^2 statistic is another fault detection index that measures the variability in the RS. In addition, the combined index ϕ and the global Mahalanobis distance MD can also be employed for fault detection, which measure the variability in the PCS and RS simultaneously [6]. These indices are all quadratic functions of $\mathbf{x}$, which can be expressed in a unified form as [15,23]

$$\mathrm{Index}(\mathbf{x}) = \mathbf{x}^T\mathbf{M}\mathbf{x} = ||\mathbf{x}||_{\mathbf{M}}^2 \tag{2}$$

where the kernel matrix $\mathbf{M}$ depends on the specific fault detection index. These quadratic fault detection indices and their control limits or thresholds, as well as the corresponding values of $\mathbf{M}$ are summarized in Table 1. Under the assumption that $\mathbf{x}$ is multi-normal, the control limits of these indices can be calculated theoretically using the results of Box [26]. If the requirement of multi-normal distribution is not fulfilled, as an alternative, non-parametric estimation method like kernel density estimation can be employed to determine empirical control limits. The process under monitoring is considered normal if the fault detection index is within its normal region, i.e., $\mathrm{Index}(\mathbf{x}) \le \Gamma^2$.

2.2. RBC for fault diagnosis

The following fault model is assumed which indicates an additive fault occurring in variable x_j

$$\mathbf{x} = \mathbf{x}^* + \boldsymbol{\xi}_j f \tag{3}$$

where $\mathbf{x}^*$ denotes the normal part, $\boldsymbol{\xi}_j$ is the jth column of the identity matrix, and f denotes the fault magnitude. The RBC method makes reconstruction in turn along all m variable directions $\boldsymbol{\xi}_i$, $i \in \{1, \ldots,$