Computers and Chemical Engineering 61 (2014) 9–19

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Fault isolation using modified contribution plots

Jialin Liu[a,*], Ding-Sou Chen[b]

[a] Center for Energy and Environmemtal Research, National Tsing Hua University, No. 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan, Republic of China
[b] New Materials Research and Development Department, China Steel Corporation, 1, Chung Kang Rd., Hsiao Kang, Kaohsiung, Taiwan, Republic of China

ARTICLE INFO

Article history:
Received 3 September 2012
Received in revised form
17 September 2013
Accepted 14 October 2013
Available online 23 October 2013

Keywords:
Fault isolation
Principal component analysis
Contribution plots
Missing data analysis

ABSTRACT

Investigating the root causes of abnormal events is a crucial task for an industrial process. When process faults are detected, isolating the faulty variables provides additional information for investigating the root causes of the faults. Numerous data-driven approaches require the datasets of known faults, which may not exist for some industrial processes, to isolate the faulty variables. The contribution plot is a popular tool to isolate faulty variables without a priori knowledge. However, it is well known that this approach suffers from the smearing effect, which may mislead the faulty variables of the detected faults. In the presented work, a contribution plot without the smearing effect to non-faulty variables was derived. A continuous stirred tank reactor (CSTR) example and the industrial application were provided to demonstrate that the proposed approach is not only capable of locating different faulty variables when the fault was propagated by the controllers, but also capable of identifying the variables responsible for the multiple sensor faults.

1. Introduction

Investigating the root causes of abnormal events is a crucial task for an industrial process. In modern chemical processes, distributed control systems are equipped for regulating the processes, and the operating data are collected and stored in a historical database. However, information about process operations is hidden under the historical data. Therefore, it is more practical to develop methods that detect and investigate the root causes of process faults based on data-driven approaches, rather than to use other methods based on rigorous process models or knowledge-based approaches. Since the measured variables are correlated for a chemical process, principal component analysis (PCA) is a popular tool to extract the features of the process data that are applied to monitor the process variations. After a fault is detected, the faulty variables need to be isolated in order to diagnose the root causes of the fault. Contribution plots are the most popular tool for identifying which variables are pushing the statistics out of their control limits. Kourti and MacGregor (1996) applied the contribution plots of quality variables and process variables to find faulty variables of a high-pressure low-density polyethylene reactor. They remarked that the contribution plots may not reveal the assignable causes of abnormal events; however, the group of variables contributed to the detected events will be unveiled for further investigation. Westerhuis, Gurden, and Smilde (2000) introduced the confidence limits of the contribution plots to enhance the capability of identifying the behaviors of faulty variables departing from the normal operating condition (NOC). They reported that there must be a careful interpretation of the contribution plots, since the residuals of the PCA are smeared out over the other variables. Yoon and MacGregor (2000) comprehensively compared model-based and data-driven approaches for fault detection and isolation, and summarized that the contribution plots provide for the easy isolation of simple faults, but that additional information about operating the process is needed to isolate complex faults.Dunia and Qin (1998) developed the reconstruction-based approach to isolate faulty variables from the subspaces of faults. Their method has been applied to reconstruct the predictor data of faulty variables before performing a prediction for a soft sensor model (Qin, Yue, & Dunia, 1997). Yue and Qin (2001) combined the statistics Q and T^2 to develop an index that is minimized when isolating the faulty variables; therefore, a more feasible solution could be found than that from the original approach (Dunia and Qin, 1998). They also proposed to extract fault directions from faulty data until all reconstructed combined indices under the confidence limit. The fault subspace, formed by the fault directions, induced that all faulty data must share the same faulty variables. The limitation of the fault subspace approach can be observed when a process fault propagates to other variables by the process controllers. Recently, Zhao, Sun, and Gao (2012) used multiple PCA models to extract fault directions from faulty datasets in order to address the issue of fault propagation. It can be expected that numerous fault PCA models are created for a process with time-varying characteristics. This type of approach, extracting the fault subspace from faulty data, is not practical for an industrial process,

* Corresponding author. Tel.: +886 3 5735294; fax: +886 3 5725924.
E-mail address: jialin@che.nthu.edu.tw (J. Liu).

0098-1354/$ – see front matter
http://dx.doi.org/10.1016/j.compchemeng.2013.10.004

Nomenclature

$\mathbf{C}$	a matrix converting the measured data into Q statistic
c_i^Q	the contribution of the ith variable to the Q statistic
c_i^T	the contribution of the ith variable to the T^2 statistic
c_i^{RCI}	the contribution of the ith variable to the reduction of the combined index
$\mathbf{D}$	a matrix converting the measured data into T^2 statistic
$\mathbf{D}_F$	a diagonal matrix contains the singular values of $\mathbf{X}_F$
F	fault event
$\mathbf{I}$	identity matrix
K	number of principal components expand the PC subspace
M	number of observations in the training dataset
N	number of variables
nf	the number of faulty variables
$\mathbf{P}$	loading matrix of the PC subspace
$\tilde{\mathbf{P}}$	loading matrix of the residual subspace
Q	statistic Q of PCA
Q_α	the $(1-\alpha)$ confidence limits of the statistic Q
Q_k^*	the reconstructed Q along the kth variable direction
RBC_i^Q	the reconstructed-based contribution of the ith variable to the Q statistic
RBC_i^T	the reconstructed-based contribution of the ith variable to the T^2 statistic
RBC_i^φ	the reconstructed-based contribution of the ith variable to the combined index
$\mathbf{S}$	covariance matrix of the training data
T^2	statistic T^2 of PCA
T_α^2	the $(1-\alpha)$ confidence limits of the statistic
$\mathbf{t}$	PC scores of the test data
$\mathbf{U}_F$	the left-singular matrix for $\mathbf{X}_F$
$\mathbf{V}_F$	the right-singular matrix for $\mathbf{X}_F$
$\mathbf{X}$	normalized training data
$\mathbf{X}_F$	faulty data under fault F
$\mathbf{x}$	normalized test data
$\mathbf{x}$	systemic parts of the test data
$\mathbf{x}_{nf}$	the collection of the faulty variables
$\mathbf{x}_{nf}^*$	the reconstructed faulty variables

Greek letters

α	significance level for statistic testing
$\mathbf{\Phi}$	a matrix converting the measured data into the combined index
$\mathbf{\Gamma}$	a diagonal matrix, the elements are one for the faulty variables and zero for the non-faulty ones.
$\mathbf{\Lambda}$	diagonal matrix of the significant eigenvalues
$\tilde{\mathbf{\Lambda}}$	diagonal matrix of the residual eigenvalues
η_k	sensor validity index for the kth variable
φ	the combined index
φ_k^*	the reconstructed combined index along the kth variable direction
φ_{nf}^*	the reconstructed combined index along nf faulty variables
Ξ_F	the fault subspace for fault F
ξ_i	a column vector in which the ith element is one and the others are zero

since the known event lists might not exist for some industrial processes. In addition, an incorrect fault isolation result will be induced when encountering a new fault.

The reconstruction-based contribution (RBC; Alcala and Qin, 2009) approach has been derived based on missing variable approach, and it was reported that RBC still suffers the smearing effect, as the contribution plots of the PCA are enduring. Therefore, the magnitude of RBCs was used to isolate the faulty variable for a single sensor fault. More recently, Kariwala, Odiowei, Cao, and Chen (2010) integrated the branch and bound (BAB) method with the missing variable approach of probabilistic PCA (PPCA) to locate faulty variables. The concept of the approach is similar to the reconstruction-based method (Dunia and Qin, 1998; Yue and Qin, 2001), but the known event datasets are not needed. Since the BAB method searches for faulty variables by minimizing the monitoring statistic of PPCA, it can be expected that the solutions of the faulty variables will be inconsistent when the fault is propagating or when the controllers try to bring the process back to NOC. In the presented work, a contribution plot without smearing effect to non-faulty variables were derived. In this approach, it is not necessary to prepare the known event datasets, which might not exist for some industrial processes, and the time-consuming task of continuously optimizing the mixed-integer programming problem for every sampling data until reaching a stable solution is also not required.

The remainder of this paper is organized as follows. Section 2 gives an overview of PCA, the contribution plots of statistics, and the fault isolation methods based on missing data approach. The contribution plots without smearing effect to non-faulty variables are detailed in Section 3. In Section 4, the continuous stirred tank reactor (CSTR) with feedback controllers and the industrial application are utilized to demonstrate the effectiveness of the proposed approach and the comparisons with the alternatives of missing data approaches, RBC and fault subspace extraction, are also provided. The first example illustrates that a fault subspace, which extracted the fault directions from faulty data, misidentified faulty variables for a process with the feedback controllers. In addition, the faulty variables of the simple and the complex faults from the work of Yoon and MacGregor (2001) are identified using the proposed approach. In the industrial application, the contribution plots normalized with the corresponding control limits suffering the smearing effect are demonstrated and the fault subspace misled by the faulty dataset, which contained multiple abnormal events, is detailed. Finally, conclusions are given.

2. Basic theory

2.1. Principal component analysis

Consider the data matrix $\mathbf{X} \in R^{N\times M}$ with N rows of variables and M columns of observations. Each row is normalized to zero mean and unit variance. The covariance of the reference data can be estimated as:

$$\mathbf{S} \approx \frac{1}{M-1}\mathbf{X}\mathbf{X}^{\mathrm{T}} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{\mathbf{T}} + \tilde{\mathbf{P}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{P}}^{\mathrm{T}} \tag{1}$$

where Λ is a diagonal matrix with the first K terms of the significant eigenvalues and $\mathbf{P}$ contains the respective eigenvectors. The $\tilde{\mathbf{\Lambda}}$ and $\mathbf{P}$ are the residual eigenvalues and eigenvectors respectively. The statistic Q is defined as a measure of the variations of the residual parts of data:

$$Q = (\mathbf{x}-\hat{\mathbf{x}})^{\mathrm{T}}(\mathbf{x}-\hat{\mathbf{x}})^{\mathrm{T}} = \mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x} \tag{2}$$

where $\mathbf{C} \equiv \sim\tilde{\mathbf{P}}\tilde{\mathbf{P}}^{\mathrm{T}}$. In addition, another measure for the variations of systematic parts of the PC subspace is the statistic T^2:

$$T^2 = \mathbf{t}^{\mathrm{T}}\mathbf{\Lambda}^{-1}\mathbf{t} = \mathbf{x}^{\mathrm{T}}\mathbf{D}\mathbf{x} \tag{3}$$

where $\mathbf{D} \equiv \sim\mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^{\mathrm{T}}$ and $\mathbf{t}$ are the first K term scores. The confidence limits of Q and T^2 can be found in Jackson (1991).