Journal of Process Control 58 (2017) 131–138

Contents lists available at ScienceDirect

Journal of Process Control

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

Research paper

Fault detection of process correlation structure using canonical variate analysis-based correlation features

Benben Jiang [a,b,*], Richard D. Braatz [b]

[a] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
[b] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 7 November 2016
Received in revised form 9 August 2017
Accepted 4 September 2017

Keywords:
Fault detection
Process correlations
Dimensionality reduction technique
Canonical variate analysis
Canonical correlation
Process monitoring

ABSTRACT

This paper proposes a canonical variate analysis (CVA) approach based on feature representation of canonical correlation for the monitoring of faults associated with changes in process correlations, which involves two new metrics, R_s and R_r, corresponding to the state and residual spaces. The utilization of the canonical correlation feature can improve the monitoring proficiency by providing more application-dependent representations compared with the original data, as well as a decreased degree of redundancy in the feature space. A physical interpretation is provided for the canonical correlation-based method. The effectiveness of the proposed approach for the monitoring of process correlation changes is demonstrated for both abrupt (step change) and incipient (slow drift) types of faults in simulation studies of a network system. In the simulation results, the canonical correlation-based method has superior performance over both the causal dependency-based method and the traditional variable-based method.

1. Introduction

With the increasing demand for production efficiency, energy savings, and environmental protection, many industrial processes are moving in the direction of large-scale and complicated structures. Industries can face large economic loss and security threats when faults occur. It is important to detect and diagnose abnormal events to improve the reliability and safety of industrial processes by utilizing fault monitoring techniques. A variety of the process monitoring approaches designed so far, such as multivariate statistical control charts [1,2], state-space/time-series approaches [3,4], and dimensionality reduction techniques [5–7], focus on the monitoring of process variables to detect faults [8].

In contrast to the high level of performance achieved in the fault monitoring of process variables, the important complementary problem of monitoring the faults of process correlation structures has not been extensively explored [9]. Process data has grown explosively with the increased scale of manufacturing plants, which represent a valuable resource for the analysis and operations of industrial processes, providing that methods are developed that can effectively extract information from this data. With the rise of the Big Data concept, increasing interest has been raised towards exploiting correlation structures [10]. In sight of these observations, this paper is devoted to developing a new data-driven approach for monitoring changes in process correlations.

Feature representation, which produces derived values (features) from an initial set of data to facilitate the desired task, is of great significance to process monitoring. Sarfraz et al. [11] reported that better fault monitoring results are achieved for features that are more application-dependent. In terms of application-dependence of the feature generation, features that measure the linkages between process variables (e.g., correlation coefficient) are more appropriate than features directly using the process variables themselves for the monitoring of process correlation structures.

The proficiency of monitoring faults from data can be improved using dimensionality reduction techniques such as principal component analysis (PCA) [12,13], partial least-squares (PLS) analysis [14,15], and canonical variate analysis (CVA) [3,16,17]. PCA- and PLS-based monitoring methods are not optimal for the detection of faults for process data that contain significant serial correlation, because the underlying PCA and PLS approaches do not generate the most accurate dynamic models, even when lagged values of process variables are augmented in the observation vectors [3,18,19]. CVA-based approaches [18,19] employ the CVA multivariate system identification method that generates accurate state-space models from serially correlated data. This method maximizes the correlation between the combinations of the "past" values of the process inputs and outputs and the combinations of the "future" values of the outputs of the system. CVA takes serial correlations into account by employing this different augmented vector technique during the dimensionality reduction procedure [18,20]. Negiz and

* Corresponding author at: College of Information Science and Technology, Beijing University of ChemicalTechnology, Beijing 100029, China
E-mail addresses: jiangbb@mail.buct.edu.cn, bbjiang@mit.edu (B. Jiang).

http://dx.doi.org/10.1016/j.jprocont.2017.09.003
0959-1524/© 2017 Published by Elsevier Ltd.

Çinar [3] discuss and demonstrate the higher accuracy of dynamic models constructed by CVA compared to dynamic PCA through application to numerical examples. CVA has been observed to have better numerical stability and parsimony than alternative identification methods, including balanced realization (BR), numerical algorithms for state-space subspace system identification (N4SID), and partial least squares (PLS), in many case studies [4,21].

The maximum correlations produced by the dimensionality reduction procedure of CVA are referred to as *canonical correlations* [18,19,22]. In terms of application dependence of feature representation, canonical correlation is an appropriate feature for monitoring process correlation structures. Canonical correlations are regarded as lower dimensional representations of process dynamic relations. As such, the performance of process monitoring can be enhanced by performing fault detection on the feature space of canonical correlation, owing to the higher sensitivity of the lower dimensional representations of the data [18].

In this article, a new fault detection approach is proposed based on the canonical correlations induced by CVA for the monitoring of process correlations. In virtue of the interpretable correlative levels of canonical correlations, two novel indices, R_s and R_r, that correspond to the state space (that is, retained canonical correlations obtained via CVA) and the residual space (that is, the rest of the canonical correlations in the CVA model) respectively, are put forward to examine the utility of the two different measures for detecting process dynamics. The two types of monitoring statistics (state space and residual space) correspond to different characteristics of the process dynamics and can potentially provide more insights into the fault.

The rest of this paper is organized as follows. The CVA method is briefly reviewed in Section 2, where a geometric interpretation of CVA is also provided. Section 3 presents the canonical correlation-based measures for monitoring correlation structural faults. The effectiveness of the proposed method is demonstrated by a gene network system in Section 4, followed by conclusions in Section 5.

2. Canonical variate analysis (CVA) revisited

2.1. CVA theorem

CVA is a dimensionality reduction technique that aims to maximize a correlation statistic between two sets of variables. Considering the two variable vectors $\boldsymbol{x} \in R^{m}$ and $\boldsymbol{y} \in R^{n}$ with covariance matrices $\Sigma_{\boldsymbol{xx}}$ and $\Sigma_{\boldsymbol{yy}}$ and cross-covariance matrix $\Sigma_{\boldsymbol{xy}}$, the orthogonal basis $\boldsymbol{J} \in R^{m\times m}$ and $\boldsymbol{L} \in R^{n\times n}$ can be determined from

$$\begin{cases} \boldsymbol{J}\boldsymbol{\Sigma}_{\boldsymbol{xx}}\boldsymbol{J}^{\mathrm{T}} = \boldsymbol{I}_{\bar{m}} \\ \boldsymbol{L}\boldsymbol{\Sigma}_{\boldsymbol{yy}}\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{I}_{\bar{n}} \end{cases}, \tag{1}$$

and

$$\boldsymbol{J}\boldsymbol{\Sigma}_{\boldsymbol{xy}}\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{D} = \mathrm{diag}(\lambda_1, \cdots, \lambda_r, 0, \cdots, 0), \tag{2}$$

where $\bar{m} = \mathrm{rank}(\boldsymbol{\Sigma}_{\boldsymbol{xx}})$, $\bar{n} = \mathrm{rank}\left(\boldsymbol{\Sigma}_{\boldsymbol{yy}}\right)$, $r = \mathrm{rank}\left(\boldsymbol{\Sigma}_{\boldsymbol{xy}}\right)$, and $\boldsymbol{I}_{\boldsymbol{k}}$ is a block-diagonal matrix with a $k \times k$ identity matrix as the first block and a zero matrix as the second block [18].

The λ_i $(i = 1, 2, \cdots, r)$ are canonical correlations with $\lambda_1 \geq \cdots \geq \lambda_r$ [19]. The vectors of canonical variables $\boldsymbol{c} = \boldsymbol{J}\boldsymbol{x}$ and $\boldsymbol{d} = \boldsymbol{L}\boldsymbol{y}$ contain a set of independent variables with the covariance matrix $\boldsymbol{\Sigma}_{cc} = \boldsymbol{I}_{\bar{m}}$ and $\boldsymbol{\Sigma}_{dd} = \boldsymbol{I}_{\bar{n}}$, respectively.

By solving the singular value decomposition (SVD)

$$\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1/2}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-1/2} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}, \tag{3}$$

the matrix of canonical correlations $\boldsymbol{D}$ can be computed as

$$\boldsymbol{D} = \boldsymbol{\Sigma}, \tag{4}$$

and the projection matrices $\boldsymbol{J}$ and $\boldsymbol{L}$ can be obtained from

$$\begin{cases} \boldsymbol{J} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1/2} \\ \boldsymbol{L} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-1/2} \end{cases}. \tag{5}$$

The matrices $\boldsymbol{U}^{\mathrm{T}}$ and $\boldsymbol{V}^{\mathrm{T}}$ rotate the canonical variables to be pairwise correlated, and the matrices $\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1/2}$ and $\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-1/2}$ scale the canonical variables to be unit variance.

2.2. CVA algorithm

The CVA method was pioneered by Akaike [23,24]. Larimore [25] both provided a series of later results on and applications of CVA based on state-space representations, and is primarily responsible for convincing the academic and industrial control communities of its value. Given input variables $\boldsymbol{u}(t) \in R^{mu}$ and output variables $\boldsymbol{y}(t) \in R^{my}$, the linear state-space model is formulated as

$$\boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t) + \boldsymbol{v}(t), \tag{6}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}\boldsymbol{u}(t) + \boldsymbol{E}\boldsymbol{v}(t) + \boldsymbol{w}(t), \tag{7}$$

where $\boldsymbol{A}$,$\boldsymbol{B}$,$\boldsymbol{C}$,$\boldsymbol{D}$, and $\boldsymbol{E}$ are coefficient matrices, $\boldsymbol{x}(t) \in R^{d}$ is a d-dimensional state vector, and $\boldsymbol{v}(t)$ and $\boldsymbol{w}(t)$ are independent white noise sequences.

The concept of past and future vectors is important to the CVA algorithm. Given a particular time instant t, the past vector comprising of the past outputs and inputs is introduced as

$$\boldsymbol{p}(t) = \left[\boldsymbol{y}_{\mathrm{T}}(t-1), \boldsymbol{y}_{\mathrm{T}}(t-2), \cdots, \boldsymbol{u}_{\mathrm{T}}(t-1), \boldsymbol{u}_{\mathrm{T}}(t-2), \cdots\right]\mathrm{T}, \tag{8}$$

and the future vector containing the outputs in the present and future is defined as

$$\boldsymbol{f}(t) = \left[\boldsymbol{y}_{\mathrm{T}}(t), \boldsymbol{y}_{\mathrm{T}}(t+1), \cdots\right]\mathrm{T}. \tag{9}$$

By substituting the matrix $\Sigma_{\boldsymbol{xy}}$ with $\Sigma_{\boldsymbol{pf}}$, $\Sigma_{\boldsymbol{xx}}$ with $\Sigma_{\boldsymbol{pp}}$, and $\Sigma_{\boldsymbol{yy}}$ with $\Sigma_{\boldsymbol{ff}}$, the matrix of canonical correlations $\boldsymbol{D} = \mathrm{diag}(\lambda_1, \cdots, \lambda_r, 0, \cdots, 0)$ can be computed via the SVD as (4), and the matrices $\boldsymbol{J}$ and $\boldsymbol{L}$ can be computed as (5). In addition, the canonical variables $\boldsymbol{x}_p(t)$ and $\boldsymbol{x}_f(t)$ which corresponds to the past and future vectors $\boldsymbol{p}(t)$ and $\boldsymbol{f}(t)$ respectively, can be derived as

$$\begin{cases} \boldsymbol{x}_p(t) = \boldsymbol{J}_d\boldsymbol{p}(t) = \boldsymbol{U}_d^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{pp}}^{-1/2}\boldsymbol{p}(t) \\ \boldsymbol{x}_f(t) = \boldsymbol{L}_d\boldsymbol{f}(t) = \boldsymbol{V}_d^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{ff}}^{-1/2}\boldsymbol{f}(t) \end{cases}, \tag{10}$$

where $\boldsymbol{J}_d = \boldsymbol{U}_d^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{pp}}^{-1/2}$ and $\boldsymbol{L}_d = \boldsymbol{V}_d^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{ff}}^{-1/2}$, and $\boldsymbol{U}_d$ and $\boldsymbol{V}_d$ contains the first d columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ in (3) respectively [18].

Practically, owing to the finite quantity of data available, the vectors $\boldsymbol{p}(t)$ and $\boldsymbol{f}(t)$ are usually truncated to be

$$\boldsymbol{p}(t) = \left[\boldsymbol{y}_{\mathrm{T}}(t-1), \boldsymbol{y}_{\mathrm{T}}(t-2), \cdots, \boldsymbol{y}_{\mathrm{T}}(t-l), \boldsymbol{u}_{\mathrm{T}}(t-1), \boldsymbol{u}_{\mathrm{T}}(t-2), \cdots, \boldsymbol{u}_{\mathrm{T}}(t-l)\right]^{\mathrm{T}}, \tag{11}$$

and

$$\boldsymbol{f}(t) = \left[\boldsymbol{y}_{\mathrm{T}}(t), \boldsymbol{y}_{\mathrm{T}}(t+1), \cdots, \boldsymbol{y}_{\mathrm{T}}(t+h)\right]^{\mathrm{T}}, \tag{12}$$

where l and h are the numbers of lags in vectors $\boldsymbol{p}(t)$ and $\boldsymbol{f}(t)$, respectively. Moreover, the appropriate numbers of lags l and h and the state order d can be determined by the Akaike information criterion [18].

Remark 1. The canonical correlations generated by the CVA algorithm have a geometric interpretation. Calculating $\boldsymbol{J}_d^{(i)}$ and $\boldsymbol{L}_d^{(i)}$ to maximize the correlation between the combinations $\boldsymbol{x}_p^{(i)} = \boldsymbol{J}_d^{(i)}\boldsymbol{p}$ and $\boldsymbol{x}_f^{(i)} = \boldsymbol{L}_d^{(i)}\boldsymbol{f}$—where $\boldsymbol{x}_p^{(i)}$ and $\boldsymbol{x}_f^{(i)}$ are the ith element of canonical variables $\boldsymbol{x}_p$ and $\boldsymbol{x}_f$ respectively, and $\boldsymbol{J}_d^{(i)}$ and $\boldsymbol{L}_d^{(i)}$ denotes the ith row of the projection matrices $\boldsymbol{J}_d$ and $\boldsymbol{L}_d$ respectively—is equivalent to