Computers and Chemical Engineering 77 (2015) 1–9

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

A combined canonical variate analysis and Fisher discriminant analysis (CVA–FDA) approach for fault diagnosis

Benben Jiang [a,b], Xiaoxiang Zhu [b], Dexian Huang [a], Joel A. Paulson [b], Richard D. Braatz [b,*]

[a] Dept. of Automation, Tsinghua University and Tsinghua National Laboratory for Information Science and Technology, Beijing 100084, China
[b] Dept. of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 11 September 2014
Received in revised form 7 January 2015
Accepted 3 March 2015
Available online 10 March 2015

Keywords:
Fault diagnosis
Canonical variate analysis
Fisher discriminant analysis
Dynamic FDA
Tennessee Eastman process
Process monitoring

ABSTRACT

This paper proposes a combined canonical variate analysis (CVA) and Fisher discriminant analysis (FDA) scheme (denoted as CVA–FDA) for fault diagnosis, which employs CVA for pretreating the data and subsequently utilizes FDA for fault classification. In addition to the improved handling of serial correlations in the data, the utilization of CVA in the first step provides similar or reduced dimensionality of the pretreated datasets compared with the original datasets, as well as decreased degree of overlap. The effectiveness of the proposed approach is demonstrated on the Tennessee Eastman process. The simulation results demonstrate that (i) CVA–FDA provides better and more consistent fault diagnosis than FDA, especially for data rich in dynamic behavior; and (ii) CVA–FDA outperforms dynamic FDA in both discriminatory power and computational time.

1. Introduction

Fault diagnosis, which is the determination of the root cause of faults, is important for efficient, safe, and optimal operation of an industrial process. The task of diagnosing the faults can be rather challenging when there are a large number of process variables that are highly correlated due to the process dynamics and controllers.

For fault diagnosis based on data-driven methods, data collected from the plant during specific faults are classified into multiple classes, where data in each class indicates a particular fault. Among the methods for classifying data of multiple classes, Fisher discriminant analysis (FDA) determines a set of projection vectors that minimize the scatter within each class while maximizing the scatter between the classes. While FDA has been used for decades in pattern classification (Duda et al., 2001), its application for analyzing chemical process data began to be explored only in the last 15 years (Chiang et al., 2000, 2001; He et al., 2005).

Due to process dynamics, observations are often serially correlated, that is, the observations at one time instant are correlated with observations at past time instants. In order to handle serial correlations in the data, the FDA method for fault diagnosis can be extended by augmenting the observation vector with lagged values of process variables. This method is referred to as dynamic FDA (DFDA) (Chiang et al., 2001), which enables dynamic information to be used in classifying the observations. Since the information contained in a single observation vector is a subset of the information contained in the augmented observation vector, the augmented vector approach can lead to improved fault diagnosis. The incorporation of time lags for autocorrelated variables benefits the fault classification by decreasing the degree of overlap among the augmented data (Chiang et al., 2001). However, data stacking in DFDA significantly increases the dimensionality of the problem, proportional to the number of lags included. Another drawback is that more data may be required to determine the mean vector and covariance matrix to achieve the same level of accuracy for each class. As a result, DFDA typically has high computational requirements that hinder its application to large-scale systems. This study will focus on developing a scheme that can not only better capture the dynamic information in the data but also has reduced computational cost.

Some past studies of fault diagnosis have employed dimensionality reduction techniques followed by discriminant analysis. Raich and Çinar (1995, 1996) presented a multivariate statistics approach for diagnosing abnormal behaviors by following principal component analysis (PCA) with discriminant analysis. A later study (Chiang, 2001) incorporated PCA and FDA for diagnosing both known and unknown faults. Employing partial least squares (PLS)

* Corresponding author at: 77 Massachusetts Avenue, Room E19-551, Cambridge, MA 02139, USA. Tel.: +1 617 253 3112; fax: +1 617 258 5042.
E-mail address: braatz@mit.edu (R.D. Braatz).

http://dx.doi.org/10.1016/j.compchemeng.2015.03.001
0098-1354/

for discrimination has been investigated extensively, such as in Chiang et al. (2000) and Barker and Rayens (2003). In general, PCA- and PLS-based discriminant analysis are limited in their ability to quickly diagnose faults for process data that contain significant serially 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 (Chiang et al., 2001; Negiz and Çinar, 1997a; Russell et al., 2000; Ku et al., 1995; Ricker, 1988).

Canonical variate analysis (CVA) is a dimensionality reduction technique in multivariate statistical analysis which utilizes state-space representations. CVA 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 (Larimore, 1997). This method takes serial correlations into account by employing this different augmented vector technique during the dimensionality reduction procedure.

Negiz and Çinar (1997a) 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 (Juricek et al., 1998; Negiz and Çinar, 1997b; Simoglou et al., 1999, 2002).

This article describes a combined CVA–FDA fault diagnosis scheme that employs CVA for pretreating the data and subsequently utilizes FDA for classifying faults. Employing CVA in the first step improves handling of serial correlations in the data, decreases the overlap among the data classes, and enables the pretreated datasets to have similar or even fewer dimensions compared with the original datasets.

The rest of this article is organized as follows. Section 2 provides some background knowledge on CVA, FDA, and DFDA. Section 3 elaborates the proposed CVA–FDA approach for fault diagnosis. The effectiveness of the CVA–FDA approach is demonstrated with the Tennessee Eastman process in Section 4, followed by conclusions in Section 5.

2. CVA and FDA

The relevant methods for fault diagnosis and classification are reviewed in this section. CVA is introduced as the dimensionality reduction method, followed by a brief review of FDA, which serves as the basis of the fault diagnosis.

2.1. CVA

CVA is a dimensionality reduction technique in multivariate statistical analysis which maximizes the correlation between two selected sets of variables. Hotelling initially proposed the CVA concept for multivariate statistical analysis, which was employed to system identification by Akaike for autoregressive-moving-average model (ARMA) models (Larimore, 1997; Akaike, 1974). The CVA method was further developed for identifying state-space models by Larimore (1997).

Given time series output data $\boldsymbol{y}_t \in R^{m_y}$ and input data $\boldsymbol{u}_t \in R^{m_u}$, the linear state-space model is (Russell et al., 2000)

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{v}_t \tag{1}$$

$$\boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{x}_t + \boldsymbol{D}\boldsymbol{u}_t + \boldsymbol{E}\boldsymbol{v}_t + \boldsymbol{w}_t \tag{2}$$

where $\boldsymbol{x}_t$ is a state vector, $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ are independent white noise processes, and $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, and $\boldsymbol{E}$ are coefficient matrices.

The CVA algorithm uses the concept of past and future vectors. At a particular time instant $t \in (1, \ldots, n)$, the past vector $\boldsymbol{p}_t$ containing the past outputs and inputs is

$$\boldsymbol{p}_t = [\boldsymbol{y}_{t-1}^{\mathrm{T}}, \boldsymbol{y}_{t-2}^{\mathrm{T}}, \ldots, \boldsymbol{u}_{t-1}^{\mathrm{T}}, \boldsymbol{u}_{t-2}^{\mathrm{T}}, \ldots]^{\mathrm{T}} \tag{3}$$

and the future vector $\boldsymbol{f}_t$ comprising of the outputs in the present and future is

$$\boldsymbol{f}_t = [\boldsymbol{y}_t^{\mathrm{T}}, \boldsymbol{y}_{t+1}^{\mathrm{T}}, \ldots]^{\mathrm{T}} \tag{4}$$

For an assumed state order k, the CVA algorithm computes a constant matrix $\boldsymbol{J}_k$ that linearly relates the past vector $\boldsymbol{p}_t$ to the memory $\boldsymbol{m}_t \in R^k$,

$$\boldsymbol{m}_t = \boldsymbol{J}_k \boldsymbol{p}_t \tag{5}$$

where the term "memory" is used instead of "state" since the vector $\boldsymbol{m}_t$ may not necessarily contain all of the information in the past (Larimore, 1990). The optimal matrix $\boldsymbol{J}_k$ is calculated via the singular value decomposition (SVD) to minimize the average prediction error

$$E\{||\boldsymbol{f}_t - \hat{\boldsymbol{f}}_t||^2_{\Lambda^\dagger}\} = E\{(\boldsymbol{f}_t - \hat{\boldsymbol{f}}_t)\Lambda^\dagger(\boldsymbol{f}_t - \hat{\boldsymbol{f}}_t)\} \tag{6}$$

where E is the expectation operator, $\hat{\boldsymbol{f}}_t$ is the prediction of $\boldsymbol{f}_t$, and the weighing $\Lambda^\dagger$ is the pseudo-inverse of Λ. Selecting $\Lambda = \Sigma_{ff}$ nearly maximizes the likelihood function for the state-space system (1) and (2) (Larimore, 1990), where Σ_{ff} is the covariance of $\boldsymbol{f}_t$.

The SVD algorithm calculates the optimal value for $\boldsymbol{J}_k$ as

$$\Sigma_{pp}^{-1/2} \Sigma_{pf} \Sigma_{ff}^{-1/2} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}} \tag{7}$$

where Σ is the diagonal matrix of nonnegative singular values with descending order, $\boldsymbol{U}$ and $\boldsymbol{V}$ are matrices of the right and left singular vectors, and the matrices $\boldsymbol{J}_k$ are obtained by

$$\boldsymbol{J}_k = \boldsymbol{U}_k^{\mathrm{T}} \Sigma_{pp}^{-1/2} \tag{8}$$

where $\boldsymbol{U}_k$ contains the first k columns of $\boldsymbol{U}$ in (7).

2.2. FDA

For fault diagnosis, data collected from the process during specific faults are categorized into classes, where each class contains data indicating a particular fault. FDA is widely used as a technique of pattern classification. The basic idea of FDA is to determine a set of projection vectors that optimize the Fisher criterion (He et al., 2005). A brief mathematical description is provided here.

Given n observations of m measurement variables, an n by m matrix $\boldsymbol{X}$ is constructed to stack the training data for all classes, and the ith row of $\boldsymbol{X}$ is represented as the column vector $\boldsymbol{x}_i$, the total-scatter matrix is given by (Duda and Hart, 1973; Chiang et al., 2004)

$$\boldsymbol{S}_t = \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{x}_{mean})(\boldsymbol{x}_i - \boldsymbol{x}_{mean})^{\mathrm{T}} \tag{9}$$

where $\boldsymbol{x}_{mean}$ indicates the total mean vector, elements of which are the means of the columns of $\boldsymbol{X}$. Define $\boldsymbol{X}_j$ as the set of vectors $\boldsymbol{x}_i$ belonging to the class j, then the within-scatter matrix for class j is defined by

$$\boldsymbol{S}_j = \sum_{\boldsymbol{x}_i \in \boldsymbol{X}_j} (\boldsymbol{x}_i - \boldsymbol{x}_{j,mean})(\boldsymbol{x}_i - \boldsymbol{x}_{j,mean})^{\mathrm{T}} \tag{10}$$