Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 32 (2007) 534–542
Expert Systems with Applications
www.elsevier.com/locate/eswa

Nonlinear feature extraction and classification of multivariate process data in kernel feature space

Hyun-Woo Cho *

Department of Industrial and Information Engineering, University of Tennessee, Knoxville, TN 37996, USA

Abstract

Batch processes have played an essential role in the production of high value-added product of chemical, pharmaceutical, food, biochemical, and semi-conductor industries. For productivity and quality improvement, several multivariate statistical techniques such as principal component analysis (PCA) and Fisher discriminant analysis (FDA) have been developed to solve a fault diagnosis problem of batch processes. Fisher discriminant analysis, as a traditional statistical technique for feature extraction and classification, has been shown to be a good linear technique for fault diagnosis and outperform PCA based diagnosis methods. This paper proposes a more efficient nonlinear diagnosis method for batch processes using a kernel version of Fisher discriminant analysis (KFDA). A case study on two batch processes has been conducted. In addition, the diagnosis performance of the proposed method was compared with that of an existing diagnosis method based on linear FDA. The diagnosis results showed that the proposed KFDA based diagnosis method outperforms the linear FDA based method.

Keywords: Fault diagnosis; Batch process; Kernel method; Fisher discriminant analysis; Feature extraction; Classification

1. Introduction

Fault diagnosis of manufacturing processes is one of the important tasks for productivity and quality improvement of final products. The task of the fault diagnosis is to identify an assignable cause of the detected out-of-control state. As a major component of statistical process control, it provides early warning for a fault and identification of its assignable cause. In particular, fault diagnosis has received much attention in process industry in that using the diagnostic decisions made on-line operating personnel need to take remedial actions to bring the process back to an in-control state (Tsung, 2000; Tsung & Apley, 2002).

Many multivariate statistical techniques had been developed and applied to fault diagnosis such as principal component analysis (PCA), partial least squares (PLS), Fisher discriminant analysis (FDA), and discriminant partial least squares (DPLS) (for example, see Akbaryan & Bishnoi, 2001; Alsbergav, Goodacrea, Rowlandb, & Kella, 1997; Chiang, Russell, & Braatz, 2000; Cho & Kim, 2004; Kemsley, 1996; Kourti, Nomikox, & MacGregor, 1995; Raich & Cinar, 1996). Such data-driven techniques for fault diagnosis have been extensively utilized due to the recent advances in sensors and data measurement technology. Given a number of datasets in the historical database, each associated with a different fault, the goal of fault diagnosis is equivalent to that of classification so that the out-of-control observations is assigned to the most similar or closely related fault group or class.

Fisher discriminant analysis (FDA) is a traditional statistical technique for feature extraction and classification. It provides a lower dimensional representation of data in that several groups or classes can be discriminated as clearly as possible. In terms of fault diagnosis or classification, each group or class corresponds to data obtained when a specific known fault occurs. Then the task of fault diagnosis is to classify a new data into one of predefined

* Tel.: +1 865 974 7655; fax: +1 865 974 0588.
E-mail address: hwcho@postech.ac.kr

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.12.007

fault groups when a fault is detected. FDA has been shown to be the good linear technique for fault diagnosis and to outperform PCA based diagnosis methods (Chiang et al., 2000; Cho & Kim, 2004). It is attributed to the fact that FDA seeks directions that are efficient for discrimination (but PCA for representation) (Müller, Mika, Rätsch, Tsuda, & Schölkopf, 2001).

A disadvantage of using "linear" FDA (LFDA) in multivariate classification or fault diagnosis problems is that nonlinear behavior in the data cannot be represented well. Although LFDA is an effective technique for feature extraction and classification, it is still a linear technique in nature. To overcome such a limitation, "kernel trick" has been used to develop a nonlinear kernel version of FDA, called kernel FDA (KFDA) (Baudat & Anouar, 2000). The basic idea of the kernel trick is that input data are mapped into a kernel *feature space* by a nonlinear mapping function and then these mapped data are analyzed. A number of powerful kernel based techniques have been developed, including support vector machines (Cortes & Vapnik, 1995), kernel PCA (Schölkopf, Smola, & Müller, 1998), kernel PLS (Rosipal & Trejo, 2001), and kernel FDA (Baudat & Anouar, 2000).

The selection of linear (i.e., LFDA) and nonlinear (i.e., KFDA) techniques for solving fault diagnosis or classification problems depends on the characteristics of target processes of interest. In linear case, fault groups or classes in data is separated easily using linear techniques. Such a linear case is the simplest problem, in which both linear and nonlinear techniques are expected to produce a good classification performance. The use of linear techniques in nonlinear case, however, may not classify most of data correctly. In this respect, a nonlinear technique such as KFDA will be useful for classification of batch processes because they have more nonlinear characteristics involved when compared to continuous processes. Thus it is essential to develop a more efficient diagnosis method for batch processes.

The objective of this paper is to present the use of kernel Fisher discriminant analysis for diagnosing batch processes. For this end, a new nonlinear KFDA based diagnosis framework is proposed to identify an assignable cause of a fault when an out-of-control state occurs. The proposed KFDA diagnosis framework has two phases of an off-line KFDA diagnosis model building and on-line diagnostic decision making. The KFDA diagnosis model is constructed by performing KFDA on unsuccessful data obtained from past batch operation, based on which diagnostic decisions are made on-line. The performance of the proposed KFDA diagnosis method is demonstrated using batch process data obtained from PVC and penicillin batch processes. In addition, the diagnosis performance of the proposed method is compared with that of a linear FDA diagnosis method. This paper is organized as follows. First, a review of kernel Fisher discriminant analysis (KFDA) is given in Sections 2 and 3. Then the proposed KFDA based diagnosis method for batch processes is presented in terms of an off-line KFDA diagnosis model building (Section 4) and on-line KFDA based diagnostic decision making (Section 5). A case study is conducted to demonstrate the proposed method in Section 6. Finally, concluding remarks are given.

2. Kernel Fisher discriminant analysis

Kernel FDA (KFDA) is the nonlinear kernel version of linear FDA to deal with the feature extraction and the classification of nonlinear characteristics. Let us consider a set of M observations in an n-dimensional space $\mathbf{x}_k \in \Re^n$, $k = 1, \ldots, M$. For a given nonlinear mapping Φ, the input space $\Re^n$ can be mapped into the *feature space* F, $\Phi : \Re^n \to F$, $\mathbf{x} \to \Phi(\mathbf{x})$. Note that the feature space F could have a much higher, possibly infinite, dimensionality.

The objective is KFDA is to find certain directions in the original variable space, along which latent groups or clusters in $\Re^n$ are discriminated as clearly as possible. Kernel FDA performs LFDA in the feature space F, which is nonlinearly related to the input space $\Re^n$. As a result, kernel FDA produces a set of nonlinear discriminant vectors in the input space. The discriminant weight vector is determined by maximizing between-class scatter matrix $\mathbf{S}_\mathrm{b}^\Phi$ while minimizing total scatter matrix $\mathbf{S}_\mathrm{t}^\Phi$, which are defined in F as follows:

$$\mathbf{S}_\mathrm{b}^\Phi = \frac{1}{M}\sum_{i=1}^{C} c_i(\mathbf{m}_i^\Phi - \mathbf{m}^\Phi)(\mathbf{m}_i^\Phi - \mathbf{m}^\Phi)^\mathrm{T}, \tag{1}$$

$$\mathbf{S}_\mathrm{t}^\Phi = \frac{1}{M}\sum_{i=1}^{M} (\Phi(\mathbf{x}_i) - \mathbf{m}^\Phi)(\Phi(\mathbf{x}_i) - \mathbf{m}^\Phi)^\mathrm{T}. \tag{2}$$

In Eqs. (1) and (2), $\mathbf{m}_i^\Phi$ represents the mean vector of the mapped observations of class i, $\mathbf{m}^\Phi$ the mean vector of the mapped M observations, c_i the number of observations of class i, and C the total number of class of $\mathbf{x}_k$, $k = 1, \ldots, M$. Similarly to the formulation of LFDA, this can be done by maximizing the Fisher criterion (Baudat & Anouar, 2000):

$$J^\Phi(\boldsymbol{\psi}) = \frac{\boldsymbol{\psi}^\mathrm{T}\mathbf{S}_\mathrm{b}^\Phi\boldsymbol{\psi}}{\boldsymbol{\psi}^\mathrm{T}\mathbf{S}_\mathrm{t}^\Phi\boldsymbol{\psi}}, \quad \boldsymbol{\psi} \neq \mathbf{0}. \tag{3}$$

The optimal discriminant vectors in feature space F can be obtained by solving the eigenvalue problem $\mathbf{S}_\mathrm{b}^\Phi\boldsymbol{\psi} = \lambda\mathbf{S}_\mathrm{t}^\Phi\boldsymbol{\psi}$ instead of Eq. (3). They are actually the eigenvectors of $\mathbf{S}_\mathrm{b}^\Phi\boldsymbol{\psi} = \lambda\mathbf{S}_\mathrm{t}^\Phi\boldsymbol{\psi}$.

3. Optimal discriminant vector of KFDA

The optimal discriminant vectors can be expressed as a linear combination of the observations in feature space F. Thus there exist coefficients b_i such that

$$\boldsymbol{\psi} = \sum_{k=1}^{M} b_k\Phi(\mathbf{x}_k) = \mathbf{H}\boldsymbol{\alpha}, \tag{4}$$