Chemometrics and Intelligent Laboratory Systems 138 (2014) 203–211

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Semi-supervised Fisher discriminant analysis model for fault classification in industrial processes

Shiyong Zhong [a], Qiaojun Wen [a], Zhiqiang Ge [a,b,*]

[a] State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, Zhejiang, P. R. China
[b] Key Laboratory of Advanced Control and Optimization for Chemical Processes, Shanghai, 200237, P.R. China

ARTICLE INFO

Article history:
Received 22 May 2014
Received in revised form 31 July 2014
Accepted 25 August 2014
Available online 30 August 2014

Keywords:
Fault classification
Semi-supervised modeling
Fisher discriminant analysis
Principal component analysis

ABSTRACT

While Fisher discriminant analysis (FDA) has been widely used for classification, it highly relies on the label information of the training data, which means that its classification performance cannot be guaranteed if there are only a small number of labeled data samples available for use. For fault classification in industrial processes, unfortunately, there are always a much smaller number of faulty samples compared to the normal samples. In addition, even if the collected faulty samples are enough for modeling, it will still need expert experiences and prior process knowledge to label them into different types, which is time-consuming and costly. In this paper, a semi-supervised form of the FDA model is proposed and used for fault classification in industrial processes. The named semi-supervised FDA bridges the superior class separability of FDA and the unsupervised nature of principal component analysis (PCA). With the incorporation of additional unlabeled data samples for modeling, the fault classification performance has been greatly improved by the new model. Both of the semi-supervised modeling efficiency and the fault classification performance are evaluated through two case studies.

1. Introduction

To enhance process safety and improve product quality, process monitoring methods including qualitative model-based methods and quantitative model-based methods have become more and more popular in industrial processes. Among the quantitative process monitoring methods, data-driven methods have raised an intensive interest in both industry and academy [1–4]. Distributed control systems, which have been widely used in industrial processes, are providing large amounts of data for the data-driven method. Those collected data can be used for analyzing whether the process is at normal operation mode or a fault has occurred.

While many research works have been done on the fault detection issue, the task to determine the type of the detected fault is a bit far beyond the ability of traditional data-based monitoring approaches. In recent years, lots of machine learning methods have been proposed for the fault classification purpose, based on which the fault classification problem can be transformed to a multi-class classification problem [3,5–11]. For example, Choi et al. [12] proposed a discriminant analysis-Gaussian mixture model (DA-GMM) based fault isolation method. Chiang et al. [13] introduced Fisher discriminant analysis (FDA) for fault diagnosis and figured out that the superior fault diagnosis performance provided by FDA is inherent. Zhu and Song [14] proposed the Gaussian mixture model (GMM) and k-nearest neighbor (kNN) based methods for fault detection and isolation, which have been developed in the kernel FDA subspace. Among all existing methods that have been proposed for fault classification, FDA may be the most widely used one which is easy to implement in practice.

Generally, FDA is a well-known linear technique for dimensionality reduction and feature extraction [24], which belongs to the type of supervised learning methods. As a result, the performance of the FDA model highly depends on the information offered by the labeled dataset. It makes use of the label information to find the optimal projection vectors, through which the labeled training data in different classes can be well separated. However, when only a small number of labeled samples are available, the traditional FDA method may perform poorly.

For fault classification in industrial processes, unfortunately, there are always a much smaller number of faulty samples compared to the normal samples. In addition, even if the collected faulty samples are sufficient, it will still need expert experiences and prior process knowledge to label them into different types, which is time-consuming and costly. Another drawback is that modeling with insufficient labeled samplings may also cause the overfitting problem in the supervised modeling methods. For this reason, the semi-supervised learning based pattern recognition methods can work in a supplementary way to the traditional supervised method. This is because the semi-supervised methods can handle both labeled and unlabeled datasets at the same time [23,25,26]. Recently, the semi-supervised learning method has been introduced for fault detection and diagnosis. For example, Monroy et al. [27] proposed a semi-supervised chemical process fault diagnose method which combined independent component analysis

* Corresponding author. Tel.: +86 571 87951442.
E-mail address: gezhiqiang@zju.edu.cn (Z. Ge).

http://dx.doi.org/10.1016/j.chemolab.2014.08.008
0169-7439/

(ICA) with GMM. Feng et al. [28] applied principal component analysis (PCA) as a semi-supervised form for process monitoring. Yan et al. [29] developed a semi-supervised mixture discriminant monitoring method for chemical batch processes.

In this paper, a semi-supervised form of the FDA method (SFDA) is proposed for fault classification purpose, which can utilize both labeled and unlabeled data samples. Different from the traditional FDA method, SFDA simultaneously takes the superior class separability of FDA and the unsupervised nature of PCA which can capture the global structure of the whole process data. SFDA aims to find optimal projection vectors which capture the discriminant structure inferred from the labeled data, as well as the intrinsic global structure inferred from both labeled and unlabeled data [15]. Therefore, when there are only a few labeled samples available, the fault classification method which developed upon SFDA can perform better than the traditional FDA based fault classification method.

The rest of this article is organized as follows. In Section 2, a brief review of the traditional FDA and PCA methods is given. Section 3 introduces the main characteristic of the SFDA method. A novel SFDA based fault classification method is proposed in Section 4. Section 5 contains a numerical example and detailed comparative studies between SFDA and FDA based approaches for fault classification. Finally, conclusions are made.

2. Fisher discriminant analysis and principal component analysis

2.1. Fisher discriminant analysis (FDA)

Fisher discriminant analysis (FDA) is a supervised linear dimensionality reduction technique [16,24]. The basic idea of FDA is to seek a transformation matrix which maximizes the between-class scatter and minimizes the within-class scatter simultaneously. Assume a labeled sampling dataset matrix $\mathbf{X}_L = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_{n_l}\}$, where the vector $\mathbf{x}_i$ is from m-dimensional space $\mathbf{R}^m$. And the sampled dataset are supposed to be normalized. All the n_l labeled samples are from K classes and there are $\mathbf{n}_k$ samples in the kth$(1 \le k \le K)$ class C_k.

Let $\mathbf{S}_b$ and $\mathbf{S}_w$ be the between-class scatter matrix and within-class scatter matrix:

$$\mathbf{S}_b = \sum_{k=1}^{K} n_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T \tag{1}$$

$$\mathbf{S}_w = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in C_k} (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \tag{2}$$

where $\boldsymbol{\mu}_k = \frac{1}{n_k} \sum_{\mathbf{x}_i \in C_k} \mathbf{x}_i$ is the mean vector of the kth class and $\boldsymbol{\mu} = \frac{1}{n_l} \sum_{i=1}^{n_l} \mathbf{x}_i$ is the mean vector of the whole sampling data. The Fisher discriminant vectors are determined by the following optimization problem

$$J_{FDA} = \arg \max_{\mathbf{v} \in R^m, \mathbf{v} \neq \mathbf{0}} \left\{ \frac{\mathbf{v}^T \mathbf{S}_b \mathbf{v}}{\mathbf{v}^T \mathbf{S}_w \mathbf{v}} \right\}. \tag{3}$$

It has been proven that this optimization problem is equivalent to the generalized eigenvalue problem below [17]:

$$\mathbf{S}_b \mathbf{t} = \lambda \mathbf{S}_w \mathbf{t} \tag{4}$$

where λ is the generalized eigenvalue and the vector $\mathbf{t}$ is the corresponding eigenvector which equals to one of the Fisher discriminant directions. The eigenvalue λ indicates the degree of the separability between different classes.

In FDA, the between-class scatter matrix and within-class scatter matrix can be rewritten as the pairwise forms for simplicity:

$$\mathbf{S}_b = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{W}_{i,j}^b \left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \tag{5}$$

$$\mathbf{S}_w = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{W}_{i,j}^w \left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \tag{6}$$

where

$$\mathbf{W}_{i,j}^b = \begin{cases} \frac{1}{n} - \frac{1}{n_k} & \text{if } \mathbf{x}_i \in C_k,\ \mathbf{x}_j \in C_k \\ \frac{1}{n} & \text{otherwise} \end{cases} \tag{7}$$

and

$$\mathbf{W}_{i,j}^w = \begin{cases} \frac{1}{n_k} & \text{if } \mathbf{x}_i \in C_k,\ \mathbf{x}_j \in C_k \\ 0 & \text{otherwise} \end{cases}. \tag{8}$$

2.2. Principal component analysis (PCA)

Different from FDA, PCA is an unsupervised dimensionality reduction method, which iteratively finds the maximum-variance direction of the data points [30]. Suppose the unlabeled sampling dataset matrix is given as $\mathbf{X}_U = \{\mathbf{x}_{n_l+1}, \mathbf{x}_{n_l+2}, \ldots, \mathbf{x}_n\}$, then the whole dataset matrix including both labeled and unlabeled data samples could be represented as $\mathbf{X} = [\mathbf{X}_L, \mathbf{X}_U]^T = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{n_l}, \mathbf{x}_{n_l+1}, \cdots, \mathbf{x}_n]^T_{n \times m}$. The orthogonal projection vectors determined by PCA are called the loading vectors $\mathbf{d}$ which can be optimized by the following problem: [1]

$$J_{PCA} = \arg \max_{\mathbf{d} \in R^m, \mathbf{d} \neq 0} \left\{ \frac{\mathbf{d}^T \mathbf{X}^T \mathbf{X} \mathbf{d}}{\mathbf{d}^T \mathbf{d}} \right\}. \tag{9}$$

To be compatible with FDA, the optimization function in PCA is represented in a different manner from the common PCA description. Here $\mathbf{S}_t$ is defined as the total scatter matrix:

$$\mathbf{S}_t = \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{u})(\mathbf{x}_i - \mathbf{u})^T \tag{10}$$

where $\boldsymbol{\mu} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$ is the total mean vector for all the sampling data. The total scatter matrix can be also reformed in pairwise form:

$$\mathbf{S}_t = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{W}_{i,j}^t \left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \tag{11}$$

where $\mathbf{W}_{i,j}^t = \frac{1}{n}$.

Similarly, the solution of loading vectors in PCA model is equal to the following optimization problem, since all the loading vectors are orthogonal to each other

$$J_{PCA} = \arg \max_{\widetilde{\mathbf{d}} \in R^m, \widetilde{\mathbf{d}} \neq 0} \left\{ \frac{\widetilde{\mathbf{d}}^T \mathbf{S}_t \widetilde{\mathbf{d}}}{\widetilde{\mathbf{d}}^T \mathbf{I}_m \widetilde{\mathbf{d}}} \right\} \tag{12}$$

where $\widetilde{\mathbf{d}}$ is the corresponding loading vector. The above optimization problem is equivalent to the following generalized eigenvalue problem

$$\mathbf{S}_t \mathbf{y} = \delta \mathbf{I}_m \mathbf{y}. \tag{13}$$