Journal of Process Control 33 (2015) 14–24

Contents lists available at ScienceDirect

Journal of Process Control

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

Principal components selection for dimensionality reduction using discriminant information applied to fault diagnosis

A. Prieto-Moreno[a], O. Llanes-Santiago[a,*], E. García-Moreno[b]

[a] Automatic and Computing Department, CUJAE, Havana, Cuba
[b] Systems and Automatic Engineering Department, UPV, Valencia, Spain

ARTICLE INFO

Article history:
Received 2 April 2014
Received in revised form 28 May 2015
Accepted 3 June 2015
Available online 20 June 2015

Keywords:
Fault diagnosis
Feature extraction
PCA
Dimensionality reduction
Discriminant analysis

ABSTRACT

The Principal Component Analysis is one of most applied dimensionality reduction techniques for process monitoring and fault diagnosis in industrial process. This work proposes a procedure based on the discriminant information contained in the principal components to determine the most significant ones in fault separability. The Tennessee Eastman Process industrial benchmark is used to illustrate the effectiveness of the proposal. The use of statistical hypothesis tests as a separability measure between multiple failures is proposed for the selection of the principal components. The classifier profile concept has been introduced for comparison purposes. Results show an improvement in the classification process when compared with traditional techniques and the StepWise selection. This has resulted in a better classification for a fixed number of components, or a smaller number of required components to obtain a prefixed error rate. In addition, the computational advantage is demonstrated.

1. Introduction

Fault diagnosis of manufacturing processes in the oil, chemical and food industries, is one of the most environmentally-friendly, safe, efficient and economically viable tasks. Recent advances in sensors and data measurement technologies allow to store a large amount of information. Therefore, it demands the transformation of this information into useful information and knowledge.

The task of the fault diagnosis is the identification of an attributable deviation cause with respect to the normal behavior. The goal is to provide early warning for out-of-specification products or process shut-down with the purpose of applying corrective actions. When the diagnosis is conducted on the basis on historical information, it can then be seen as a knowledge extraction procedure, consisting of the iterative sequence of the following steps [1]:

1. **Data cleaning** (to remove noise and inconsistent data).
2. **Data integration** (where multiple data sources may be combined).
3. **Data selection** (where data relevant to the analysis task are retrieved from the database).
4. **Data transformation** (where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations).
5. **Data mining** (where intelligent methods are applied to extract data patterns).
6. **Pattern evaluation** (to identify the truly interesting patterns representing knowledge based on performance measures).
7. **Knowledge presentation** (where visualization and knowledge representation techniques are used to present mined knowledge to users).

Fig. 1 shows a data-based fault diagnosis block diagram. The preprocessing block includes data cleaning, integration, and selection steps. The feature extraction and selection block corresponds to the transformation step. The data mining and pattern evaluation steps are represented by the classifier design and optimization, and the classifier evaluation blocks, respectively. The fault diagnosis block performs the knowledge presentation task.

In fault diagnosis based on historical data – like in most data analysis problems – limitation on the size of the available training data is common. Often, the small size of the training set – small number of samples – limits the performance of classifiers. Typically, for a ten-attribute system to be used for classification purposes, the required sample size for optimum training falls into the order of

* Corresponding author. Tel.: +53 7 2663204; fax: +53 7 2663204.
E-mail addresses: albprieto@electrica.cujae.edu.cu (A. Prieto-Moreno), orestes@electrica.cujae.edu.cu (O. Llanes-Santiago), emoreno@isa.upv.es (E. García-Moreno).

http://dx.doi.org/10.1016/j.jprocont.2015.06.003
0959-1524/

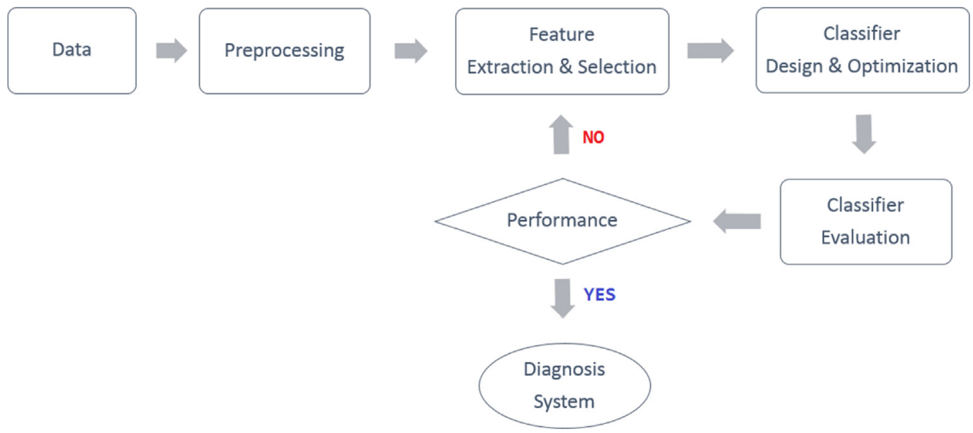

Fig. 1. Fault diagnosis data-based diagram.

thousands [2]. In order to avoid this problem – known as "curse of dimensionality" – many dimensionality reduction techniques such as: Principal Component Analysis (PCA) [3–5], Independent Component Analysis (ICA) [6–8], Fisher Discriminant Analysis (FDA) [9–11], and Partial Least Squares (PLS) [12,13] have been developed and applied to fault diagnosis. These techniques allow to reduce all the attributes into fewer meaningful dimensions, and capture the dominant information in the original set of variables.

In order to develop a better classification process, an special treatment of the input data is necessary. This work focuses on the steps 4, 5 and 6 of the knowledge extraction procedure. The first objective is to illustrate the importance of identifying the main attributes that allow to diminish the complexity of the classification problem, and to demonstrate that a proper variable selection classification will result in accurate fault diagnosis. The second objective of this paper is to propose a modification in the process of dimensionality reduction using PCA. An alternative approach to the traditionally used ones for the selection of the principal components (PCs) to be retained, based on the use of discriminant analysis, is presented. This, in order to reach a greater separability between the classes and, consequently, reduce the classification error rate.

In the specialized literature, many classifiers have been tried with different degrees of success for a variety of classification problems [14–16]. In this work the Maximum A Posteriori Probability (MAP) [17,15,18] and Artificial Neural Networks (ANN) [19–21], classifiers with different operating principles, are used to illustrate the advantages of the proposed procedure. Both classifiers will be applied to the industrial benchmark Tennessee Eastman Process (TEP) [22,23]. The reason for using this classifier is none other than its ease implementation, since the proposed procedure is independent of the classifier used. The proposal aims to retain the PCs that increases the between-class distance of the clusters that characterize each operating state. This can have a direct impact on every classification method, thus improving the performance of any of the classifiers used.

This article has been structured as follows: Section 2 shows the PCA and FDA dimensionality reduction methods, and introduces problems in traditionally used approaches of these techniques in the diagnostics. Section 3 describes the proposed procedure to conduct the selection of PCs showing the highest discriminant power and, in this way, improving the classification performance of a diagnostic system. A description of the case study and the classification methods used, is presented in Section 4. Section 5 includes the experimental outcomes, as well as a discussion of the same, based on the classification performance and the computational cost. Finally, conclusions are presented.

2. Data-driven fault diagnosis

Data-driven process monitoring or Statistical Process Monitoring (SPM) apply multivariate statistics, such as PCA and FDA, as well as machine-learning methods to fault diagnosis in industrial process operations and production results. This has become one of the most fruitful areas in research and practice over the last two decades [24–26]. SPM methods have wide applications in various industrial processes, including: chemicals, polymers, microelectronics manufacturing, iron and steel, pharmaceutical processes, and power distribution networks. Due to the data-based nature of the SPM methods, it is relatively easy to apply them to real processes, despite their large scale, compared with other methods based on system theory or rigorous process models, because they do not require first principle knowledge of the monitored process.

2.1. Procedures for the dimensionality reduction

In many cases, the dimension of a measurement vector – i.e. the number of sensors – can be very high, and lots of its elements can be redundant or even irrelevant with respect to classification process.

There is more than one reason for reducing the dimensionality of the measurement vector to a sufficient minimum. Computational complexity is the obvious one. Another major reason is that an increase of the dimension ultimately causes a decrease of performance of classifiers. Fig. 2 illustrates this statement.

Here, we have a measurement space with N dimension ranging between 1 and 13. There are K classes having equal prior probabilities, each one made up by N_s observations. The minimum error rate E_{min} would be the one obtained if all class densities of the problem are fully know. That is, if the number of observations in each class tends to infinity. Likewise, the minimum error rate is a non-increasing function of the number of sensors. Once an element has been added with discriminant information, the addition of another element cannot wipe out this information. Therefore, class information accumulates when increasing the dimension.

However, in practice, densities are not usually known completely. Often, classifiers must be designed using a training set instead of using density knowledge. This implies that the number of samples in the training set must be very large. If not, overfitting occurs and the trained classifier will become too much adapted to the noise in the training data. Fig. 2 shows that if the training set size is $N_s = 20$, the optimal dimension of the measurement vector is $N = 4$; accounting for the lowest E error rate. Increasing the sample size allows a dimension increase. With $N_s = 80$ the optimal dimension is $N = 6$.