Control Engineering Practice 68 (2017) 32–45

Contents lists available at ScienceDirect

Control Engineering Practice

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

Probabilistic fault diagnosis method based on the combination of nest-loop fisher discriminant analysis and analysis of relative changes

Yue Wang, Chunhui Zhao *

State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China

A R T I C L E I N F O

Keywords:
Bias of data location
Increase of data variations
Nested-loop fisher discriminant analysis (NeLFDA)
Relative changes
Probabilistic fault diagnosis

A B S T R A C T

Bias of data location and increase in data variations are two typical disturbances, which in general, simultaneously exist in the fault process. Targeting their different characteristics, a nested-loop fisher discriminant analysis (NeLFDA) algorithm and relative changes (RC) algorithm are effectively combined for analyzing the fault characteristics. First, a prejudgment strategy is developed to evaluate the fault types and determine what changes are covered in the fault process. Two statistical indexes are defined, which conduct Monte Carlo based center fluctuation analysis and dissimilarity analysis respectively. Second, for the fault data containing those two faults simultaneously, a combined NeLFDA-RC algorithm is proposed for fault deviations modeling, which is termed as CNR-FD. Fault directions concerning bias of data location are extracted by the NeLFDA algorithm and then corresponding fault deviations are removed from the fault data. Then RC algorithm is performed on these fault data to extract directions concerning increase of data variations. These fault directions are used as reconstruction models to characterize each fault class. Particularly, the compromise between these two algorithms is determined by the Monte Carlo based center fluctuation analysis. For online applications, a probabilistic fault diagnosis strategy based on Bayes' rule is performed to identify fault cause by discovering the right reconstruction models that can make the reconstructed monitoring statistics have the largest probabilities of belonging to normal condition. The motivation of the proposed algorithm is illustrated by a numerical case and the performance of the reconstruction models and the probabilistic fault diagnosis strategy are illustrated using pre-programmed faults from the Tennessee Eastman benchmark process and the real industrial process data from the cut-made process of cigarettes in some cigarette factory.

1. Introduction

With chemical processes becoming more and more complex, fault detection and diagnosis has been a popular research area over the last few decades (Du, Budman, & Duever, 2016; Hsu & Su, 2011; Kruger, Kumar, & Littler, 2007; Li, Zhao & Gao, 2015; Mahadevan & Shah, 2009; Muradore & Fiorini, 2012; Qin, Zhao, Wang & Gao, 2017; Qin, Zhao, Zhang & Gao, 2017; Simani & Patton, 2008; Zhang, Zhao, Wang, & Wang, 2017; Zhao & Gao, 2017a) as an important and necessary assurance of the process safety and product quality. Multivariate statistical analysis methods, such as principal component analysis (PCA) (Wold, Esbensen, & Geladi, 1987), partial least squares (PLS) (Burnham, Viveros, & MacGregor, 1996; De Jong, 1993) and fisher discriminant analysis (FDA) (Chiang, Kotanchek, & Kordon, 2004; Zhao & Gao, 2017b), have been widely applied to the field of statistical process monitoring (SPM) (Qin, 2003) in consideration of a large number of measured variables in process industries. They share the characteristics of reducing dimensionality and dealing with highly correlated variables, which can be implemented by projecting measurement data onto a latent space with fewer dimensions and orthogonal monitoring directions. PCA and PLS algorithms usually supervise process status using two monitoring statistics, the T^2 statistic, which reflects the major systematic variations resulting from the changes of distribution variance, and the *SPE* (Squared Prediction Error) statistic, which reflects the variations of residual information resulting from changes of variable correlations. For fault detection, if the monitoring statistics go beyond the desired regions established on the acceptable process variations, it can be concluded that some abnormal or faulty behaviors have occurred.

As summarized in the work of Katipamula and Brambley (2005), fault diagnosis is employed to evaluate the fault and determine its causes after some alarming signals have occurred in the procedure

* Corresponding author.
E-mail address: chhzhao@zju.edu.cn (C. Zhao).

http://dx.doi.org/10.1016/j.conengprac.2017.07.009
Received 23 March 2017; Received in revised form 6 July 2017; Accepted 25 July 2017
0967-0661/

of fault detection. Besides, it is desired to take necessary corrective actions to remove the abnormal behaviors, which means to bring out-of-control monitoring statistics back to the desired regions. For fault diagnosis, various methods have been reported, such as independent component analysis (ICA) (Lee, Yoo, & Lee, 2004; Widodo, Yang, & Han, 2007), structured residual-based approach (Gertler, Li, Huang, & McAvoy, 1999), and support vector machine (Widodo & Yang, 2007; Yang & Hou, 2016). The method of contribution plots (Alcala & Qin, 2009; Choi & Lee, 2005) is popular in the field of fault diagnosis, which is based on the assumption that faulty variables contribute significantly to the out-of-control monitoring statistics. However, it has been reported that this method involves fault "smearing" because of the correlations across variables, which might result in confusing results as the influence of contributing variables on the contribution might pass on to the noncontributing ones. (Dunia & Joe Qin, 1998) defined a fault reconstruction concept in the context of PCA monitoring systems so as to find the fault directions instead of the fault variables to model the fault deviations for further analysis and fault recovery. This method consisted of finding the reconstruction sample vector and bringing the sample vector back to the normal condition among fault subspace. The strategy of reconstruction is implemented by performing corrective actions on the data, and fault diagnosis via reconstruction ensures that the fault cause is identified by the right reconstruction model which can best recover fault-free data. A principal component of fault deviations (PCFD) algorithm Zhao & Gao (2013) is proposed on the phase nature of fault processes to characterize fault deviations related with out-of-control monitoring statistics and to develop reconstruction models based on the significant fault deviations in each phase. However, the conventional reconstruction method based on the PCA algorithm decomposed all the general distribution information in fault data, which might not distinguish fault patterns from normal. Based on the basic idea of eliminating the out-of-control monitoring statistics, (Zhao & Sun, 2013) proposed a two-step algorithm that focused directly on the alarming-responsible fault deviations in different monitoring subspaces based on the idea of the relative changes (RC) algorithm. The fault effects that could cause out-of-control monitoring statistics were modeled differently on the analysis of relative changes between each fault data and normal data in each monitoring subspace, which was more effective for fault reconstruction. This algorithm is effective in analyzing the changes of distribution variance as it compares variations between each fault data and normal data.

Fault diagnosis can also be regarded as a classification problem in some situations and this is also the case with the FDA algorithm. The FDA algorithm is a linear dimensionality reduction technique and famous for optimal separation from multiple classes. However, people have observed some limitations of the conventional FDA algorithm. They are singular problems of within-class scatter matrix, the rank of the between-class scatter matrix's influence on the amount of discriminant components and the non-orthogonality between discriminant components. Several improved methods have been proposed to solve the singular problem of within-class scatter matrix (Friedman, 1989; Raudys & Duin, 1998). Zhao and Gao (2015) proposed a nested-loop fisher discriminant analysis (NeLFDA) algorithm, which contained an inner-loop iterative process, an outer-loop iterative process and a class-specific data deflation technique, and could overcome the three limitations mentioned above. The NeLFDA algorithm is an appropriate tool to analyze measurement bias as it can effectively extract the directions along which the projections of measurement data can be separated from each other.

In the present work, two typical faults in the fault process are analyzed: bias of data location and increase in data variations. To determine what types of deviations exist in the fault process, a prejudgment strategy is developed by defining two statistical indices. For fault data that contain these two faults aiming at their different characteristics, it is natural to consider combining the NeLFDA algorithm and RC algorithm to extract fault deviations concerning biased location and fault deviations concerning increase of data variations respectively. Therefore, this paper proposes the combination of the NeLFDA and RC algorithms to extract fault deviations for comprehensive explanation of fault characteristics and more effective correction of fault effects. The proposed algorithm is referred to as the combined NeLFDA and RC based fault deviations (CNR-FD) modeling algorithm. In particular, the compromise between the NeLFDA algorithm and RC algorithm is achieved by ensuring that the center of fault data falls into the normal region of center fluctuation after removal of the biased location, which is defined by the Monte Carlo analysis. It is noted that for single biased location or increase of data variations in the fault data, the proposed algorithm converges to a single NeLFDA algorithm or a single RC algorithm. The directions extracted by the proposed combined algorithm are then used as reconstruction models. The conventional reconstruction based fault diagnosis methods attribute the observation to the class that can bring the monitoring alarms back to the normal region. However, this may lead to false diagnosis if multiple reconstruction models can get in-control monitoring statistics. Therefore a probabilistic fault diagnosis strategy based on Bayes' rule is used for online fault diagnosis. It identifies the fault cause by discovering the right reconstruction models that can make the reconstructed monitoring statistics have the largest probability of belonging to normal condition. The major contributions are specified as follows:

(1) A judgment strategy is developed by defining two statistical indices on the basis of the Monte Carlo based center fluctuation analysis and dissimilarity analysis, which can evaluate the fault types, so as to determine a proper algorithm for analyzing fault characteristics.
(2) For the fault process containing two typical fault types, the NeLFDA algorithm and RC algorithm are effectively combined to analyze different fault characteristics in which a threshold is defined which can definitely determine how many fault deviations should be extracted by the NeLFDA algorithm so that the remaining ones can be explained by the RC algorithm.
(3) A probabilistic fault diagnosis strategy is proposed to allocate a new observation to the fault class with the maximum a posteriori probability that the reconstructed monitoring statistics belong to the normal condition.

The rest of the paper is organized as follows: First, the motivation of combining the two algorithms is analyzed regarding three fault cases. Subsequently, the prejudgment strategy, the combined algorithm and the probabilistic fault diagnosis strategy are mathematically formulated. Then, in the illustration section, the proposed methods are verified by a numerical sample case, pre-programmed fault classes from the Tennessee Eastman process, and real process data from the cut-made process of cigarettes in some cigarette factory. At last, conclusions are drawn based on the results of this study.

2. Methodology

2.1. Motivation

The FDA algorithm is a famous fault diagnosis method. Fault directions are extracted by optimizing multiple objectives, concentration of samples in one data class, and dispersion of samples among different data classes. As an improvement, the NeLFDA algorithm (Zhao & Gao, 2015) is proposed to handle some problems of the conventional FDA, such as the non-orthogonality between discriminant components. For one fault class and normal class, the NeLFDA algorithm performs well in distinguishing the fault class from normal class because the projections of different classes on the extracted directions are far away from each other. Therefore it fits for analyzing biased location in fault data. However, besides biased location, there is another typical fault type, namely, the increase of data variations, in the fault process. The RC algorithm (Zhao & Sun, 2013) is effective in analyzing the