Identifying Causes of Industrial Process Faults Using Nonlinear Statistical Approach

Hyun-Woo Cho[1*]

[1]**Department of Industrial and Management Engineering, Daegu University**

공정 이상원인의 비선형 통계적 방법을 통한 진단

조현우[1*]

[1]대구대학교 산업경영공학과

Abstract Real-time process monitoring and diagnosis of industrial processes is one of important operational tasks for quality and safety reasons. The objective of fault diagnosis or identification is to find process variables responsible for causing a specific fault in the process. This helps process operators to investigate root causes more effectively. This work assesses the applicability of combining a nonlinear statistical technique of kernel Fisher discriminant analysis with a preprocessing method as a tool of on-line fault identification. To compare its performance to existing linear principal component analysis (PCA) identification scheme, a case study on a benchmark process was performed to show that the fault identification scheme produced more reliable diagnosis results than linear method.

요 약 산업체 공정의 실시간 공정 모니터링과 진단은 생산 제품의 품질과 안전을 보장하는데 반드시 필요한 활동들의 하나이다. 그중에서 공정 진단은 공정에 발생된 특정 이상상황의 원인을 밝혀내는 것으로서 조업자들이 이상상황의 근본원인을 보다 효과적으로 도출하는데 도움을 줄 수 있다. 본 논문에서는 비선형 KFDA 기법과 데이터 전처리기법을 이용한 이상원인 진단방법을 적용하고 이의 진단 성능을 기존 선형 기법에 기반한 PCA 진단방법과 비교한다. 실제 공정을 모사한 Tennessee Eastman 공정 시뮬레이터의 공정 데이터를 통한 사례연구를 수행한 결과 기존 선형 진단 방법론 대비 신뢰할 수 있는 진단 결과를 얻을 수 있었다.

Key Words : Process Monitoring, Fault Identification, Nonlinear Statistical Method, Principal Component Analysis, Data Preprocessing

1. Introduction

To ensure safety and stability of industrial processes, we need to continuously monitor the process operations, detect and diagnose process abnormalities, and take appropriate remedial actions. The main reason for this is that the impact of abnormal process operations is enormous both on safety and cost. With the use of process data, for example obtained from plant information systems, automated on-line data collection has been popular. The availability of on-line process data has motivated real-time process monitoring and diagnosis based on multivariate statistical approaches.[1]

A production system may be subject to unexpected malfunctions or quality problems. They may lead to an equipment damage, plant shutdowns, or sometimes cause explosions or environmental issues. Thus, it is crucial that such abnormalities are detected and diagnosed promptly and reliably so that corrective actions can be taken to maintain a high level of performance of a production

*Corresponding Author : Hyun-Woo Cho
Tel: +82-10-4922-0051   email: hwcho@daegu.ac.kr

system.[2] In general, while monitoring is to determine the occurrence of a fault, diagnosis is to identify its assignable cause. These two operational tasks are the major components of the traditional statistical process control.[3]

In this work we assess the applicability of combing a nonlinear kernel technique of kernel Fisher discriminant analysis (KFDA) with a preprocessing method as a tool for on-line fault identification. Using a simulation process this work compare its performance to existing linear principal component analysis (PCA)-based identification technique.

In general, linear FDA has been reported to provide an efficient lower dimensional representation of data in that several groups or classes can be discriminated as clearly as possible.[4] Nonetheless, the use of linear techniques for nonlinear processes, however, will produce a lot of inaccurate identification results. Thus we utilize a nonlinear fault identification of KFDA, and combine it with an orthogonal filtering method, called orthogonal signal correction, to eliminate unwanted variation or noise of data.

We can obtain clues or information on which process variables caused the process to be out-of-control. For this purpose, we execute a pair-wise KFDA modeling on normal process data and each of fault data sets. It can characterize the change of each process variable's contribution relative to normal process data. The preprocessing step should be performed prior to KFDA modeling.

This paper is organized as follows. First, a review of kernel Fisher discriminant analysis (KFDA) is given, which is followed by a simulation case study to demonstrate the fault identification performance. Finally, concluding remarks are given.

2. Method

Principal component analysis (PCA) is used to decompose correlated original variables into an un-correlated set of principal components in a reduced space. If the original variables are highly correlated, several principal components are enough to explain the data. Then the remaining principal components explain

the variability of noise. Thus noise-filtering effects can be achieved by discarding them.[4]

Kernel FDA (KFDA) is the nonlinear kernel version of FDA for the feature extraction and the classification of nonlinear data. KFDA is to solve the problem of linear FDA in nonlinear feature space.[5] Linear FDA aims at finding linear projections such that classes or groups in input data are discriminated as clearly as possible, which can be achieved by maximizing the Fisher criterion

$$J(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_b\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_t\mathbf{w}}, \mathbf{w} \neq \mathbf{0} \tag{1}$$

where $\mathbf{S}_b$ and $\mathbf{S}_t$ are the between-class and total scatter matrices of input data. KFDA performs linear FDA in the feature space, yielding a set of nonlinear discriminant vectors in the input space. The discriminant vectors are determined by maximizing the Fisher criterion:

$$J^{\Phi}(\mathbf{\psi}) = \frac{\mathbf{\psi}^{\mathrm{T}}\mathbf{S}_b^{\Phi}\mathbf{\psi}}{\mathbf{\psi}^{\mathrm{T}}\mathbf{S}_t^{\Phi}\mathbf{\psi}} \tag{2}$$

where

$$\mathbf{S}_b^{\Phi} = \frac{1}{M}\sum_{i=1}^{C} c_i (\mathbf{m}_i^{\Phi} - \mathbf{m}^{\Phi})(\mathbf{m}_i^{\Phi} - \mathbf{m}^{\Phi})^{\mathrm{T}} \tag{3}$$

$$\mathbf{S}_t^{\Phi} = \frac{1}{M}\sum_{i=1}^{M} (\Phi(\mathbf{x}_i) - \mathbf{m}^{\Phi})(\Phi(\mathbf{x}_i) - \mathbf{m}^{\Phi})^{\mathrm{T}} \tag{4}$$

The optimal discriminant vectors can be expressed as a linear combination of the observations in F. Therefore, we can find that there exist coefficients such that

$$\mathbf{\psi} = \sum_{k=1}^{M} a_k \Phi(\mathbf{x}_k) = \mathbf{H}\mathbf{\alpha} \tag{5}$$

$$\mathbf{H} = [\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_M)]$$

Then, the use of the M×M kernel matrix $\mathbf{K}$ and the substitution of equation (5) into equation (2) produces the following equation[5]:

$$J^{\mathrm{K}}(\mathbf{\alpha}) = \frac{\mathbf{\alpha}^{\mathrm{T}}(\mathbf{KWK})\mathbf{\alpha}}{\mathbf{\alpha}^{\mathrm{T}}(\mathbf{KK})\mathbf{\alpha}} \tag{6}$$

Finally, the optimal discriminant vectors in F, which is the optimal solutions for equation (2), are given by

$$\mathbf{\psi}_j = \mathbf{H}\mathbf{\alpha}_j = \mathbf{HP}\Lambda^{-1/2}\mathbf{\beta}_j \tag{7}$$

As shown in Fig. 1, the use of a kernel function enables us to compute dot products in F without nonlinear mappings Φ.
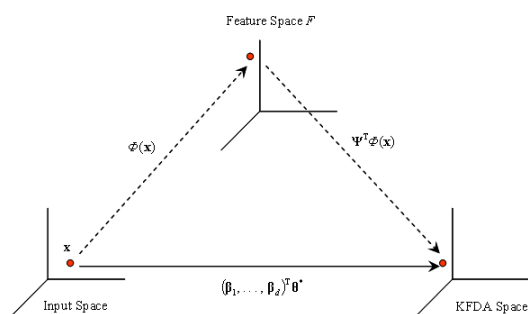
[Fig. 1] Nonlinear KFDA

The KFDA identification needs to perform a pairwise KFDA on normal and fault data. This pair-wise KFDA building is able to represent the index of process variable's contribution relative to normal data. Similar to PCA-based identification, the KFDA identification index value can be utilized, and we need to solve a reconstruction problem to calculate an approximate values.[6] Here they can be obtained by minimizing the squared distance like

$$\hat{\mathbf{x}} = arg \ \min_{\hat{\mathbf{x}}} \left\| \mathbf{\Omega} - \Phi(\hat{\mathbf{x}}) \right\|^2. \tag{8}$$

Prior to obtaining reconstruction values, we need to find KFDA discriminant vectors. It is because the contribution of each process variable should be calculated using the reconstruction of the input data.[5] A KFDA discriminant score vectors is obtained by

$$\mathbf{s} = \mathbf{\Psi}^{\mathrm{T}} \Phi(\mathbf{x}) = (\mathbf{\psi}_1, \ldots, \mathbf{\psi}_d)^{\mathrm{T}} \Phi(\mathbf{x}) \tag{9}$$

Orthogonal signal correction (OSC) is one of preprocessing techniques for the removal of data noise. Wold et al. (1998) first developed OSC, and it removes systematic variation that is orthogonal or unrelated to predefined categories or class membership values.
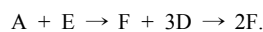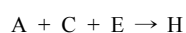
3. Simulation Results

This work is demonstrated using simulation data of Tennessee Eastman (TE) process, which is a simulation of a real industrial process where the components and operating conditions were modified for proprietary reasons.[7]
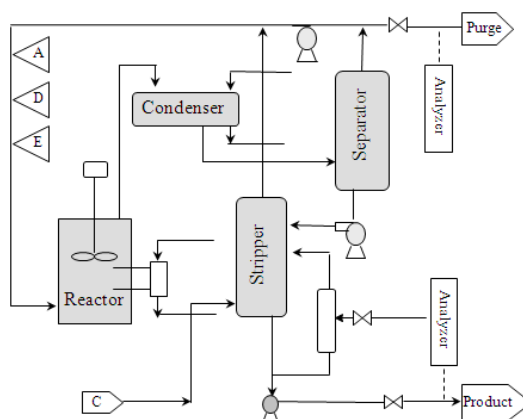

[Fig. 2] A schematic diagram of TE process

This simulator has been widely used as a test bed of continuous processes for various purposes. As shown in Fig. 2, it consists of five major units: a reactor, a product condenser, a separator, a recycle compressor, and a product stripper. This process also produces two products G and H from four reactants A, C, D, and E. Also there are an inert B and a byproduct F. A total of 53 process variables are measured on-line, and sampling interval was set to be three minutes. The gaseous reactants (i.e., A, C, D, E, and inert B) are fed to the reactor, where the liquid products G and H are formed. The reactions in the reactor are as follows:

A + C + D → G
A + C + E → H
A + E → F + 3D → 2F.

The simulation runs for TE process generated 900 normal data while fault dataset contains 450 observations. To compare the identification performance of KFDA with linear PCA, two faults of TE process are investigated. In this work, PCA monitoring method was used to detect faults based on the T^2 and SPE charts simultaneously, which is shown in Fig. 3. Here the monitoring statistics of T^2 and SPE are used to monitor the progress of the process. T^2 represents the systematic part of the process variation while SPE (squared prediction error) calculates the residual part of the process variation associated with the latest on-line measurement data. As shown in Fig. 3, unusual events are detected because the two statistics exceeded their control limits.
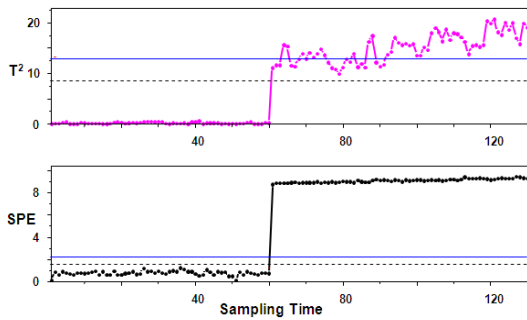
[Fig. 3] Monitoring charts

First, a step change in the condenser cooling water inlet temperature was investigated. The KFDA-based identification results for the step change of the condenser cooling water inlet temperature are displayed in Fig. 4. We can select the process variables V17 (stripper underflow) and V52 (condenser cooling water flow) as the most assignable or contributing variables. It also shows that two process variables (V11 and V22) have relatively larger identification index values than the other process variables. PCA-based identification index value, on the other hand, fails to identify these contributing process variables, which is shown in Fig. 5. When a fault is detected by the PCA monitoring model, we obtained PCA-based identification index values that calculated the contributions of each variable to Hotelling's T^2 or squared prediction error (SPE) statistics.
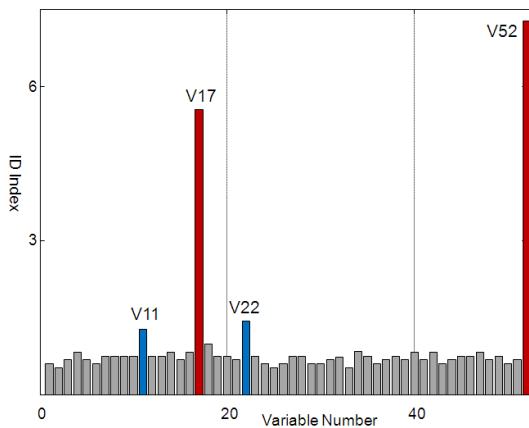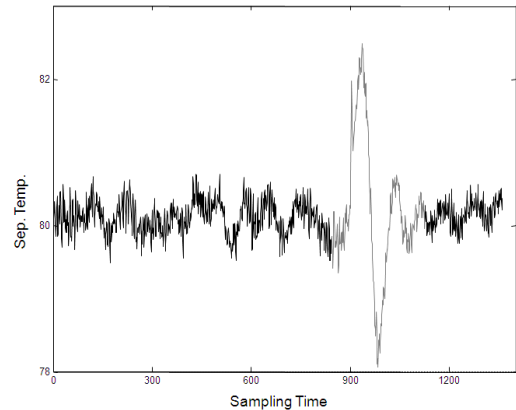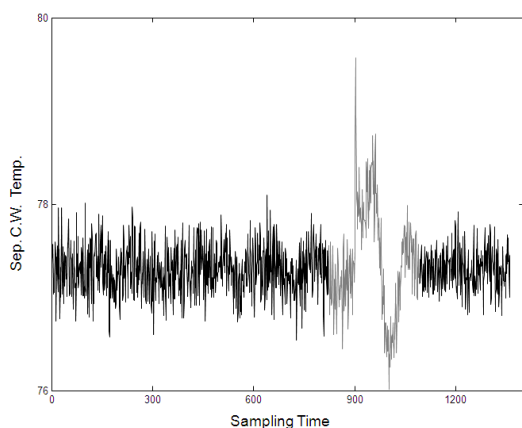

[Fig. 4] KFDA identification results
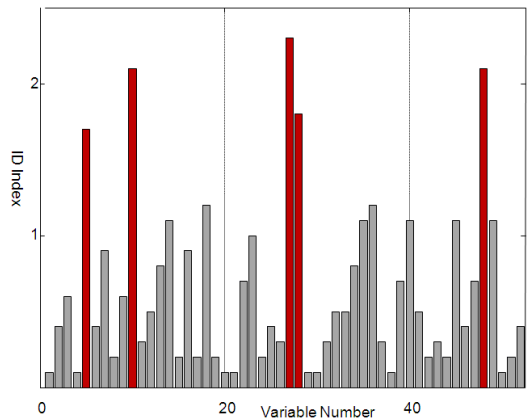
[Fig. 5] PCA identification results

Here, the identification of the two process variables (i.e., V17 and V52) makes sense because the fault of this case causes some fluctuation in these two variables. It can be seen by inspecting variable plots of Fig. 6 and Fig. 7.


[Fig. 6] Separator temperature plot

As shown in the figures, this fault results in some fluctuation in the separator temperature and in the separator cooling water outlet temperature accordingly. It is observed that the compensation of control actions cause the process variables of V11 and V22 to return to normal values after some fluctuations.

[Fig. 7] Separator c.w. temperature plot

The second fault considered in this work is a feed loss of the reactant "A" in stream 1, KFDA identification result of which is shown in Fig. 8. This fault means that the TE process has an abrupt decrease or no reactant "A" in the feed. This fault can be seen from the variable plot of feed flow rate of reactant "A" in Fig. 9. In this case the control loop on the stream 1 must fully open the feed valve of "A". As expected, these two process variables, i.e., V1 (feed flow rate of "A") and V44 (feed valve of "A"), were identified as shown in Fig. 8.
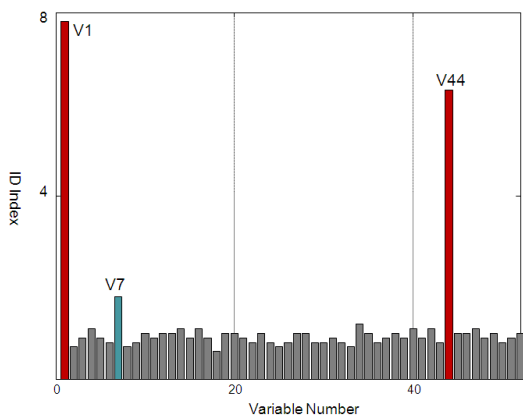
[Fig. 8] KFDA identification results
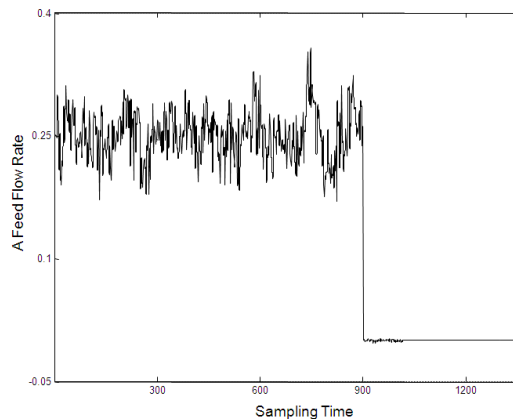
[Fig. 9] A plot for flow rate of "A"

At this point note that V7 process variable for reactor pressure has a slightly larger value. It may be related to the fact that "A" feed loss leads to the accumulation of D and E in the reactor. Resultantly, TE process reactor pressure increases and reaches a safety limit value. On the other hand, the linear PCA based identification did not select these right process variables except V1, which is not shown here. These better KFDA-based identification results may be due to the fact that mathematically FDA seeks the directions efficient for the discrimination between different groups of the data while PCA finds directions that capture the maximum variability for parsimonious representation.

4. Conclusion

In this work, nonlinear kernel version of FDA is presented to perform fault identification in an real-time basis. KFDA fault identification is based on the reconstruction of pre-images. By executing a pair-wise KFDA on normal data and each of fault classes we are able to characterize the change of each process variable's contribution relative to normal data during a specific fault.

A case study on the Tennessee Eastman process has shown that the method produces reliable identification results compared to linear PCA's identification index. Using a nonlinear technique of KFDA in a fault identification task was able to represent nonlinear behavior of the process of interest. This method requires,

on the other hand, representative data to describe different operation faults. Thus, process knowledge must be needed to provide detailed identification results from the data-driven approach.

There may be practical issues regarding the implementation of the identification scheme. An on-line data collection system should be available to gather the historical and on-line process data. In addition, the improvement of the diagnosis performance at the beginning of a fault would be an important issue in practice. This can be useful to practitioners who take some remedial actions using identification results.

Hyun-Woo Cho   [Regular member]

- Aug. 2003 : POSTECH., Industrial Eng., PhD
- Aug. 2003 ~ Aug. 2007 : GIT/UT, Research Associate
- Sep. 2007 ~ Feb. 2011 : SEC, Senior Engineer
- Mar. 2011 ~ Current : Daegu Univ., Dept. of Industrial. & Management Eng., Professor

<Research Interests>
Intelligent Process Monitoring, Data Mining

References

[1] S. J. Qin, "Statistical process monitoring: basics and beyond", Journal of Chemometrics, 17, pp. 480 – 502, 2003.

[2] S. Bersimis, S. Psarakis, J. Panaretos, "Multivariate statistical process control charts: an overview", Quality and Reliability Engineering International, 23 (5), pp. 517 – 543, 2007.

[3] A. Norvilas, A. Negiz, J. Decicco, and A. Cinar, "Intelligent process monitoring by interfacing knowledge-based systems and multivariate statistical monitoring", Journal of Process Control, 10, pp. 341 – 350, 2000.

[4] L. H., Chiang, E. L. Russell, and R. D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis", Chemometrics and Intelligent Laboratory Systems, 50, pp. 243 – 252, 2000.

[5] G. Baudat and F. Anouar, Generalized discriminant analysis using a kernel approach. Neural Computation, 12, pp. 2385 – 2404, 2000.

[6] A. K. Conlin, E. B. Martin, and A. J. Morris, "Confidence limits for contribution plots. Journal of Chemometrics", 14, pp. 725 – 736, 2000.

[7] J. J. Downs, and E. F. Vogel, "A plant-wide industrial process control problem", Computers and Chemical Engineering, 17, pp. 245 – 255, 1993.