Fourth International Workshop on Computational Intelligence & Applications
IEEE SMC Hiroshima Chapter, Hiroshima University, Japan, December 10 & 11, 2008

Multivariate Analysis for Fault Diagnosis System

Hanaa E.Sayed [+,*], Hossam A.Gabbar [-], Shigeji Miyazaki [+]

[+] *Department of Systems Engineering, Division of Industrial Innovation Science, Okayama University,1-1 Tsushima Naka 3-chome,700-8530 Okayama, Japan*

[-] *Faculty of Energy Systems and Nuclear Science, UOIT, ON L1H 7K4 Ontario, Canada*

Abstract— **Many multivariate techniques have been applied to diagnose faults such as Principal Component Analysis (PCA), Fisher's Discriminant Analysis (FDA), and Discriminant Partial Least Squares (DPLS). However, it has been shown that FDA and DPLS are more proficient than PCA for diagnosing faults. And recently applying kernel on FDA which is called KFDA (Kernel FDA) has showed outperformance than linear FDA based method. We propose in this research work an advanced KFDA for faults classification with Building knowledge base for faults structure using FSN. A case study is done on a chemical G-Plant process, constructed and experimental runs are done in Okayama University, Japan. The results are showing improving performance of fault detection rate for the new model over FDA.**

Keywords: KFDA, Fault Diagnosis, Genetic Algorithm, Process Monitoring

I. Introduction

Fault diagnosis of chemical processes is considered one of the most important tasks for safety and quality improvement. Real time faults diagnosis is considered an integral part of process design to enhance safety and control. Fault is an abnormal condition of a machine including dysfunction or malfunction of a part, an assembly, or the whole machine system. Another view of fault is system fault which is an unexpected change or malfunction in a system. Fault may not lead to a physical failure or breakdown.

Although FDA is an effective technique for classification and fault diagnosis, it is still linear method and cannot represent well the nonlinear behavior in data. [3]; and [7] have presented nonlinear kernel version to FDA, called KFDA to overcome the nonlinearity problem. KFDA has shown better diagnosis performance than FDA. Although, it is still limited in chemical processes fault diagnosis applications. Chemical processes have complex relations between variables and usually quick reaction, so it needs complex solutions to secure its safety. KFDA basic idea is that nonlinear input data is mapped into a kernel feature space by a nonlinear mapping function and then perform FDA to analyze the mapped data.

Although Multivariate Statistical Process Control (MSPC) is an excellent tool in fault diagnosis, it is still a numeric statistical tool and it does not include any process knowledge and information that allows automatic fault diagnosis. Many researchers have proposed approaches to overcome this problem by combining the statistical method with knowledge base technique especially for PCA methods such as in [1]; and [2]. In this paper, a new approach is used to combine KFDA with Fault Semantic Network (FSN) [4]; and [5]. So that interaction between statistical and intelligent methods is used to support fault diagnosis for improved maintenance and safe operation.

This paper is organized as follows. Section two is a brief introduction to KFDA. Then section three explains shortly Fault Semantic Network (FSN). The proposed model is explained in section four. In section five, a case study on G-Plant system is described and results are shown. A Conclusion is in section six.

II. Kernel Fisher's Discriminant Analysis

We will give in this section a short review for KFDA. Let us consider a set of m observations in an n-dimensional space. And the column vector is $x_i \in R^n$, where i=1, … , m. $x_i \in R^n$ is the transpose of the ith row of $X \in R^{m \times n}$, where X_h is the subset containing m_h samples that belong to the fault group h. Thus, $X = \bigcup_{h=1}^{c} X_h$ and $m = \sum_{h=1}^{c} m_h$. c is the number of fault classes. For a given nonlinear mapping Φ, input space R^n can be mapped into feature space F, Φ: R^n → F, x → Φ(x). F can have a much higher dimensionality could reach to infinity.

The objective of KFDA is to find certain directions in the original variable space, along with latent groups or clusters in R^n are discriminant as clearly as possible. Kernel FDA performs FDA in the feature space F, which is nonlinearly related to the input space R^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 matrix S_b^{Φ} while minimizing total scatter matrix S_t^{Φ}, which are defined in F as follows:

$$S_b^{\Phi} = \frac{1}{m} \sum_{i=1}^{C} c_i \left(m_i^{\Phi} - m^{\Phi}\right)\left(m_i^{\Phi} - m^{\Phi}\right)^T \quad (1)$$

$$S_t^{\Phi} = \frac{1}{m}\sum_{i=1}^{m} c_i \left(\Phi(x_i) - m^{\Phi}\right)\left(\Phi(x_i) - m^{\Phi}\right)^T \qquad (2)$$

where:

m_i^{Φ}: represents the mean vector of the mapped observations of class i,

m^{Φ}: the mean vector of the mapped m observations, c_i the number of observations of class i,

and

C: is the total number of class of x_h, h = 1, … , m

The idea of KFDA is to solve the problem of FDA in the feature space F. This can be achieved by maximizing the following Fisher criterion [3]:

$$J^{\Phi}(\varphi) = \frac{\varphi^T S_b^{\Phi} \varphi}{\varphi^T S_i^{\Phi} \varphi}, \Psi \neq 0. \qquad (3)$$

The optimal discriminant vectors can be expressed as a linear combination of the observations in the feature space F, i.e.

$$\varphi = \sum_{i=1}^{m} \alpha_i (x_i) = Q\alpha \qquad (4)$$

where $Q = [\Phi(x_i), \dots, \Phi(x_m)]$ and $\alpha = (\alpha_1, \dots, \alpha_m)^T$. For a given kernel matrix K ($k_{i,j} = <\Phi(x_i), \Phi(x_j)> i, j = 1, \dots m$). Substituting Eq. (4) into Eq. (3), we have:

$$J^{\Phi}(\alpha) = \frac{\alpha^T (KWK)\alpha}{\alpha^T (KK)\alpha} \qquad (5)$$

where $W = diag(W_1, \dots, W_c)$ and W_1 is an n_1 x n_1 matrix with terms all equaling to $1/n_1$. Then, the optimal discriminant vectors in feature space are given by

$$\varphi_i = Q\alpha_i = QP\boldsymbol{\Lambda}^{-1/2}\beta_i, \mathrm{i} = 1, \dots, \qquad (6)$$

where $P = (p_1, \dots, p_r)$ and $\boldsymbol{\Lambda} = (\lambda_1, \dots, \lambda_r)$. $p_1, \dots, p_r$ are orthogonal eigenvectors of the matrix K corresponding to r nonzero eigenvalues $\lambda_1 \geq \dots \geq \lambda_r > 0$. $\beta_1, \dots, \beta_d$ are actually the eigenvectors of $\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^T\boldsymbol{W}\boldsymbol{P}\boldsymbol{\Lambda}^{1/2^{-1}}\boldsymbol{\Lambda}$ corresponding to $d(d \leq c-1)$ largest eigenvalues. Consequently, given a new sample $\mathbf{x}_{\mathrm{new}}$ and its mapped observation $\Phi(\mathbf{x}_{\mathrm{new}})$, its KFDA discriminant score vector can be obtained as follows:

$$\boldsymbol{z} = \varphi^T \Phi(\mathbf{x_{new}}) = (\beta_1, \dots, \beta_d)^T \left(\frac{p_1}{\sqrt{\lambda_1}}, \dots, \frac{p_r}{\sqrt{\lambda_r}}\right)^T \mathbf{k}_{\mathrm{new}} \qquad (7)$$

where:

$$\mathbf{k}_{\mathrm{new}} = \mathrm{Q}^T \Phi(\mathbf{x_{new}}) = [\mathbf{k}(\mathbf{x_1}, \mathbf{x_{new}}), \dots, \mathbf{k}(\mathbf{x_n}, \mathbf{x_{new}})]$$

III. Fault Semantic Network

FSN is considered the network of the knowledge base design for the fault model. Equipments are connected via ports. Each equipment can be associated with function, fault, symptom, cause, consequence, and other environmental factors.

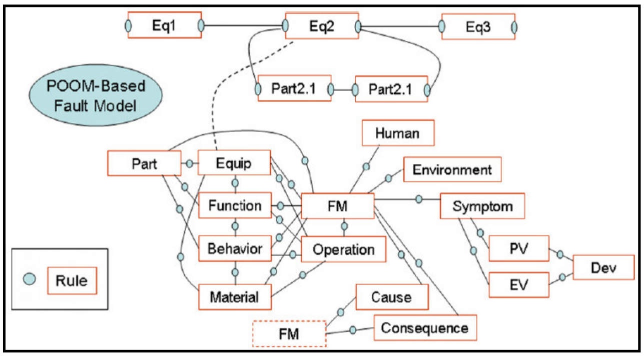

Figure 1: POOM-based Semantic Network of fault propagation (Gabbar, 2007)

Behavior and operation are defined and linked to equipment, functions, and faults. Process variables and equipment variables are used to describe the symptoms. Rules are defined and linked with each edge of the semantic network wherever applicable. FSN is dynamically updated where it captures the domain expert's feedback and knowledge to tune the fault models systematically [5]. Another way we used FSN to be updated when KFDA method capture a new fault it also updates the FSN. FSN model is shown in Figure 1.

IV. Proposed Model Framework

As KFDA is a multivariate statistical process control method, it is a pure data driven technique. So when dealing with complex relationships and structures it is needed to add to it a structural knowledge base previously defined on the process. Structural knowledge base is needed to easily connect the fault with its relation with other entities in the knowledge base such as consequences, causes, operators….So we used KFDA with FSN to achieve this target. Interaction is done when building the knowledge base in the startup run and at real time experimental run. The knowledge base is for faults and their links to plant, equipments, operations, and functions. It enables early fault diagnosis and taking the proper actions in abnormal situations.

Figure 2 shows the proposed integrated model flow chart where real time data are analyzed and faults structure are constructed in qualitative manner in FSN module that can be easily understood by human and process systems.

Distributed Control System (DCS) is a control unit used to monitor and measure the process variables of the plant. DCS is used also as alerter for higher or lower control variables values. Process variables real time data is the input of KFDA classifying method. If KFDA detected a fault it checks the FSN for already existing faults groups. In case the fault is classified in one of the existing groups the model reports the list of possible causes, consequences, and proper actions should be taken. In case the fault is new, the model updates FSN by creating a new fault group. This is constructing a full knowledge base for faults structures, causes, consequences, and actions as explained in Figure2.

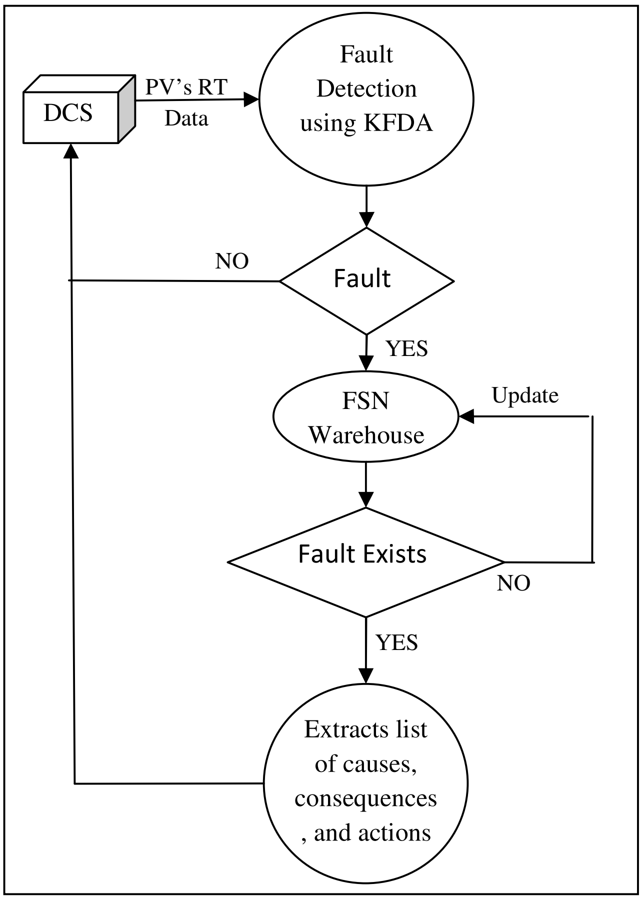

Figure 2: Proposed Model Flow Chart

V. Case Study & Results

We used a real chemical plant designed and constructed in Okayama University, Japan. Real time data are being extracted from the DCS of the plant. Plant process design is shown in Figure3. Process variables values and measured and we compared the KFDA with the proposed model and measured the faults detection success rates.

A. Operations & Process Variables

As shown in the plant process Figure3 we run in this experiment four types of operations and we monitored six process variables in order to get the results of the experiment. The four operations are:

(1) Cold Water Filling (TANK2 Filling): by opening the feed source valves MV1 (Manipulated Valve1) and MV2 and control valve CV2 (Control Valve2) water starts filling TANK2. In order to control the overflow we open also MV14 & MV12.

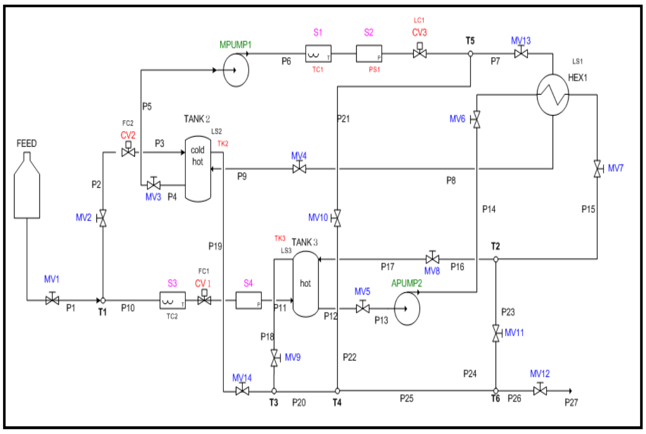

Figure 3: Plant Process Design

(2) Hot Water Filling (TANK3 Filling): by opening feed source valve MV1 and control valve CV1 water starts filling TANK3. In order to control the overflow we open MV9 & MV12. Water is still cold, but we called the process hot water filling to differentiate it from the first cold water filling and because we are filling the tank which will contain hot water in the rest of the operations.

(3) Heating Water (TANK3 Heating): by turning the heater ON.

(4) Cold & Hot Water Circulation: to exchange the temperature degree between hot and cold water, by turning the HEX1 ON, and opening all valves except MV10 & MV11.

Table 1: Process Variables List

TK2	Temperature in TANK2
TK3	Temperature in TANK3
TC1	Temperature in water circulation line between TANK2 and HEX1
TC2	Temperature in water circulation line between TANK2 and TANK3

B. Run Results

We run the model for about 500 runs as training set to build our knowledge base. And then after building the FSN knowledge base with the groups of faults categorized as per each set of causes, symptoms and consequences. A comparison between FDA and proposed model is shown in Table 3.

Table 2: Set of Alarms Variables

TK2	Temperature in TANK2
TK3	Temperature in TANK3
TC1	Temperature in water circulation line between TANK2 and HEX1
TC2	Temperature in water circulation line between TANK2 and TANK3
LS1	Level in HEX1
LS2	Level in TANK2
LS3	Level in TANK3

FSN can show possible causes, symptoms and consequences related to each fault as it is shown in Table 4. A list of equipments exist in our experimental G-Plant is shown. Also some sample runs of faults we made in our training runs related to these equipments, with possible causes.

Actions or consequences which we applied during our experimental run such as closing pumps or even shutdown the plant or close the electricity.

It was clear during the experimental runs that the proposed fault diagnosis model enabled the plant operators and engineers to understand all possible faults, deviations, and associated behavior. For example if temperature is high in TANK3 (exceeds the limit set-point) for some operation, operator will be able to understand which set of possible causes reached the temperature this degree, and what action he can do to return back to normal zone. This is linked with operation design and verification for normal and abnormal situations.

Table 3: Comparisons of Fault Diagnosis Success Rates

	FDA	Proposed Model
G-PLANT1	85%	92%
G-PLANT2	90%	93%
G-PLANT3	83%	92%
G-PLANT4	80%	85%
G-PLANT5	87%	90%
G-PLANT6	80%	85%
G-PLANT7	88%	95%
G-PLANT8	83%	98%
Average	85%	91%

Table 4: FSN Sample runs

	Equipment	FM	Cause	Symptom	Consequence /Action
Example Samples	Pump	Overflow	Operator Error	Increase Temp	Close Valve
	Valve	Leak	Heater Failure	Decrease Temp	Open Valve
	Heat Exchanger	Cavitations	Pump \ Failure	Increase Level	Close Feed
	Tank	No Heating	Wrong SOP	Decrease Level	Open Pump
		Over Heating	Circulation Lines Damage	High Vibration	Close Pump
		Delayed Heating	Air in Pipes	High Sound	Stop Electricity
			Outlet Blockage	Drop Pressure	
			Controller Failure		
			Valve Failure		

VI. Conclusion

Chemical processes are considered very complex systems. Therefore it needs advanced solutions to be able to diagnose faults efficiently. Recently KFDA has been used in fault diagnosis and it showed better results than FDA, PCA. But due to its pure data-driven, it needs to be embedded within knowledge-base system so that to give a complete integrated fault diagnosis model. Our proposed model supports analysis of abnormal situations where fault semantic networks are constructed by representing faults, symptoms, causes, and consequences for all abnormal scenarios. KFDA and FSN are integrated together to build the knowledge base and get use of it at real time plant execution. Case study is being analyzed on G-Plant process using different operations runs, and results shown better faults diagnosis success rate. Also it enhances the operator to understand the fault causes and consequences which reduces the maintenance cost by eliminating redundant work.

REFERENCES

[1] David Leung & Jose Romagnoli. An integration mechanism for multivariate knowledge-based fault diagnosis. *Journal of Process Control,* 12, 15-26, 2002.

[2] Fu Xiao, Shengwei Wang, Xinhua Xu & Gaoming Ge. An isolation enhanced PCA method with expert-based multivariate decoupling for sensor FDD in air-conditioning systems. *Applied Thermal Engineering*, doi:10.1016/j.applthermaleng.2008.03.046, 2008.

[3] G. Baudat, & F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural Computation*, 12, 2385- 2404, 2000.

[4] H. A.Gabbar, R. Abdursul & M. Kamel. Learning mechanisms for intelligent fault diagnosis. *International conference of IEEE SMC*, Taiwan, 2006.

[5] H. A.Gabbar. Improved qualitative fault propagation analysis, *Journal of Loss Prevention in the Process Industries,* 20, 260–270, 2007.

[6] H.W. Cho. Nonlinear feature extraction and classification of multivariate process data in kernel feature space, *Expert Systems with Applications*, 32, 534–542, 2007.

[7] H.W. Cho & K.J. Kim. Fault diagnosis of batch processes using discriminant model, *International Journal of Production Research*, 42, 597-612, 2004.