International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:1, 2013

A Hybrid Scheme for On-Line Diagnostic Decision Making using Optimal Data Representation and Filtering Technique

Hyun-Woo Cho

***Abstract*—**The early diagnostic decision making in industrial processes is absolutely necessary to produce high quality final products. It helps to provide early warning for a special event in a process, and finding its assignable cause can be obtained. This work presents a hybrid diagnostic scheme for batch processes. Nonlinear representation of raw process data is combined with classification tree techniques. The nonlinear kernel-based dimension reduction is executed for nonlinear classification decision boundaries for fault classes. In order to enhance diagnosis performance for batch processes, filtering of the data is performed to get rid of the irrelevant information of the process data. For the diagnosis performance of several representations, filtering, and future observation estimation methods, four diagnostic schemes are evaluated. In this work, the performance of the presented diagnosis schemes is demonstrated using batch process data.

***Keywords*—**Diagnostics, batch process, nonlinear representation, data filtering, multivariate statistical approach.

I. Introduction

THE operation of Manufacturing or production processes are, in general, subject to special or unexpected abnormalities of breakdowns and malfunctions. These process events or occurrence of statistical out-of-control states often impose a significantly negative impact on final quality of products produced [1], [2]. A fault is defined as an abnormal process event, and the goal of fault diagnosis is to determine the assignable causes of a fault. As an important topic of statistical process control, many researchers have developed diagnosis schemes suitable for specific industrial processes. Among them, batch process is difficult to handle because it has many challenging points such as nonlinearity, finite time duration, etc. The production of a batch includes tasks like charging ingredients, processing them under controlled conditions, and discharging final product [3].

Recent development of diagnosis methods applied to fault diagnosis area focused on the use of multivariate statistical methods, especially machine learning techniques such as nonlinear kernel versions of principal component analysis, partial least squares, and Fisher discriminant analysis [4]-[7]. These empirical modeling techniques for fault diagnosis have been widely used because of widespread sensor and data measurement technology.

As a feature extraction and classification technique, Fisher discriminant analysis provides a efficient representation of data where different classes can be separated as clearly as possible. Thus FDA has been shown to be the linear technique better than principal component analysis, which is attributed to the fact: FDA seeks directions efficient for discrimination, but PCA for explanation. The difference between FDA and PCA is illustrated in Fig. 1. In general, each fault group in diagnosis is equal to the data of a specific known fault. Then the task of fault diagnosis is to classify a new data into one of predefined fault groups. However, nonlinear patterns in data or processes cannot be explained by linear techniques. In order to overcome such a limitation, kernel methods have been used to develop a variety of nonlinear kernel techniques [8]-[10].

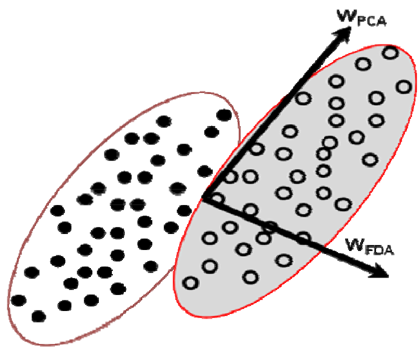

Fig. 1 A diagram for data representation

If a large number of features prevent exploring data patterns, it is necessary to eliminate unimportant ones. Unfortunately redundant or irrelevant features are likely to mask underlying patterns of data [4]. Thus one often tries to preprocess or filter raw data before main analysis. This work presents a classification approach to diagnosis, which is combined with nonlinear representation of raw process data along with the use of preprocessing or filtering techniques. The nonlinear dimension reduction is executed for nonlinear classification decision boundaries. The purpose of using nonlinear techniques in classification tree is to decrease the data suitable for discriminating several fault classes. Prior to building empirical diagnosis models, filtering of the data is performed to trim the irrelevant information of the process data. To compare the diagnosis performance of KPCA and KFDA, a total of four diagnosis schemes are evaluated, in which two filtering techniques are also tested. Due to the characteristics of batch data, the selection of estimation approaches for future

Hyun-Woo Cho is with the Department of Industrial and Management Engineering, Daegu University, 712-714 Kyungsan, Republic of Korea (phone: +82-53-850-6547; fax:+82-53-850-6549; e-mail: hwcho@daegu.ac.kr).

observation is also discussed. The performance of the presented diagnosis schemes is demonstrated using batch process data. This paper is organized as follows: an introduction of multivariate statistical techniques followed by a case study on a batch process. Finally, concluding remarks are given.

II. Methodologies

A. Principal Component Representation

A linear version of principal component representation, i.e., principal component analysis (PCA), is used to decompose correlated original variables into an uncorrelated set of linear principal components. In most cases, only several components are enough to explain the data variability. On the other hand, a nonlinear PCA, i.e., kernel PCA (KPCA), can be derived by solving $\lambda\mathbf{v} = \mathbf{C}^{F}\mathbf{v}$, where $\mathbf{C}^{F}$ indicates the covariance matrix in the feature space F. Furthermore, the eigenvalue equation can be written as [10]

$$\lambda\langle\Phi(\mathbf{x}_k),\mathbf{v}\rangle = \langle\Phi(\mathbf{x}_k),\mathbf{C}^{F}\mathbf{v}\rangle \tag{1}$$

and there exists coefficients α_i, $i = 1, \ldots, M$, such that

$$\mathbf{v} = \sum_{j=1}^{M}\alpha_j\Phi(\mathbf{x}_j)\cdot \tag{2}$$

Combining (1) and (2) yields the following:

$$\lambda\sum_{j=1}^{M}\alpha_j\langle\Phi(\mathbf{x}_k),\Phi(\mathbf{x}_j)\rangle = \frac{1}{M}\sum_{j=1}^{M}\alpha_j\left\langle\Phi(\mathbf{x}_k),\sum_{i=1}^{M}\Phi(\mathbf{x}_i)\right\rangle\langle\Phi(\mathbf{x}_i),\Phi(\mathbf{x}_j)\rangle\cdot \tag{3}$$

Finally, the principal components for $\mathbf{x}$ are given by:

$$t_k = \langle\mathbf{v}_k,\Phi(\mathbf{x})\rangle = \sum_{j=1}^{M}\alpha_j^k\langle\Phi(\mathbf{x}_j),\Phi(\mathbf{x})\rangle = \sum_{j=1}^{M}\alpha_j^k \mathrm{k}(\mathbf{x}_j,\mathbf{x})\cdot \tag{4}$$

B. Discriminant Analysis

The goal of linear Fisher discriminant analysis (FDA) is to find certain directions in original variables, along which hidden groups are discriminated as clearly as possible [8]. As an extension of linear FDA, nonlinear kernel FDA (KFDA) executes linear FDA in the feature space F. As a result, the discriminant weight vector is determined by maximizing between-class scatter matrix $\mathbf{S}_b^{\Phi}$ while minimizing total scatter matrix $\mathbf{S}_t^{\Phi}$, which are defined in F as follows:

$$\mathbf{S}_b^{\Phi} = \frac{1}{M}\sum_{i=1}^{C}c_i(\mathbf{m}_i^{\Phi}-\mathbf{m}^{\Phi})(\mathbf{m}_i^{\Phi}-\mathbf{m}^{\Phi})^{\mathrm{T}} \tag{5}$$

$$\mathbf{S}_t^{\Phi} = \frac{1}{M}\sum_{i=1}^{M}(\Phi(\mathbf{x}_i)-\mathbf{m}^{\Phi})(\Phi(\mathbf{x}_i)-\mathbf{m}^{\Phi})^{\mathrm{T}} \tag{6}$$

By maximizing the Fisher criterion

$$J^{\Phi}(\boldsymbol{\psi}) = \frac{\boldsymbol{\psi}^{\mathrm{T}}\mathbf{S}_b^{\Phi}\boldsymbol{\psi}}{\boldsymbol{\psi}^{\mathrm{T}}\mathbf{S}_t^{\Phi}\boldsymbol{\psi}} \tag{7}$$

and solving the eigenvalue problem of $\mathbf{S}_b^{\Phi}\boldsymbol{\psi} = \lambda\mathbf{S}_t^{\Phi}\boldsymbol{\psi}$ we can obtain the optimal discriminant vectors, which are actually the eigenvectors of $\mathbf{S}_b^{\Phi}\boldsymbol{\psi} = \lambda\mathbf{S}_t^{\Phi}\boldsymbol{\psi}$.

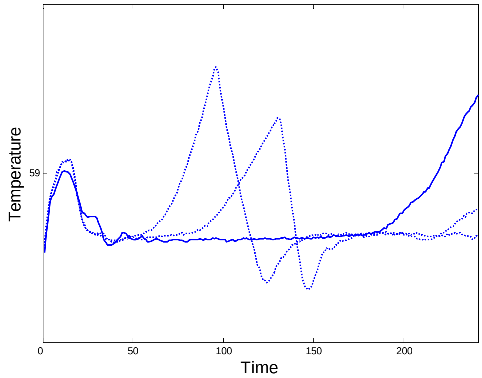

Fig. 2 Trajectories of a process variable

C. Classification-based Analysis

In general, classification tree classifier represents a statistical technique for the classification of data of interest. Basically it constructs trees by recursively partitioning predictor space. Such processes utilize training data sets in which classes are given and known in advance. During model building, a class is assigned to each of terminal nodes. For the new data processed, their predicted classes are the ones related to the terminal node to which the new data are assigned using their predictor values [11]. The class assigned to each terminal node t minimizes the misclassification cost, which is given by:

$$r(t) = \min_i \sum\nolimits_j c(i|j)\,p(j|t) \tag{8}$$

where $c(i|j)$ represents the cost misclassifying a class j as a class i and $p(j|t)$ the estimated probability of the class j in node t. The Gini diversity index is one of the commonly used functions for node impurity

$$i(t) = \sum\nolimits_{i\neq j} p(i|t)\,p(j|t) = 1-\sum\nolimits_j p^2(j|t). \tag{9}$$

The goodness of a split can also be evaluated by the deviance for a node t

$$d(t) = -2\sum\nolimits_j n_{tj}\log p(j|t) \tag{10}$$

III. Results and Performance Comparison

In this section, the diagnosis performance of the proposed scheme based on nonlinear kernel method combined with tree classifier is demonstrated. The test process is a polyvinyl chloride batch process, which represents a straight resin polymerization process initiated by vinyl chloride monomer. This process includes a polymerization reactor, reflux condenser, agitator, and cooling jacket. Eleven process variables are automatically measured on-line with optimally controlled operating condition. The trajectories of the temperature variable for several batches are shown in Fig. 2. For this diagnosis comparison purpose, 200 fault batches in seven fault groups are used as training data of nonlinear kernel diagnosis model for this batch process. Seven fault batches for each fault group were used as test data. In this work, they are referred to as D1 through D5. To compare the diagnosis performance between KPCA and KFDA, these two nonlinear kernel techniques were applied separately to find optimal representation of various fault patterns. That is, when a new batch is available on-line, diagnostic decision making step utilizes corresponding nonlinear kernel diagnosis model to identify an assignable cause of a fault. It is used to classify a new batch into one of predefined fault groups. Considering three-way characteristics of batch process data as shown in Fig. 3, it is necessary to estimate “future observations” of a new batch. It is due to the fact that the data of a new batch is not complete until the end of its operation.

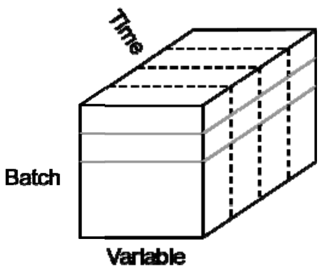

Fig. 3 Three-way batch data

An on-line KFDA vector of a new batch is given b projecting the observation onto discriminant vectors as follows:

$$\mathbf{s}_{new}(k^*) = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)^{\mathrm{T}} \left(\frac{\gamma_1}{\sqrt{\lambda_1}}, \ldots, \frac{\gamma_r}{\sqrt{\lambda_r}} \right)^{\mathrm{T}} \left[\mathrm{k}(\mathbf{z}_1, \mathbf{x}_{new}(k^*)), \ldots, \mathrm{k}(\mathbf{z}_I, \mathbf{x}_{new}(k^*)) \right]$$

As stated before, the selection of a kernel function for this application is done after testing various kernel functions. It was found that the second-order polynomial kernel is appropriate to capture nonlinearity of the data.

A filtering or preprocessing of process data is performed first, which is followed by executing nonlinear kernel analysis and a classification tree. This work considers two filtering methods for the test process, called discriminant partial least squares and orthogonal signal correction. The orthogonal signal correction (OSC) is a PLS-based algorithm that eliminates the unwanted variation of **X** orthogonal to **Y** [12]. On the other hand, discriminant partial least squares (DPLS) represents the classical PLS applied to classification problems. In both methods applied, we introduce a coding having information about class memberships of samples where binary **Y** matrix has a structure:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_{n_k} & \mathbf{0}_{n_k} & \cdots & \mathbf{1}_{n_k} \end{bmatrix}$$

Finally, it is necessary to build a classification tree using nonlinear discriminant scores. Using theses scores fault classification should be performed to identify assignable causes of process faults. Here, the maximum size tree needs to be pruned back to select the optimal sized tree. During this process, cross validation is used, and then classification tree building processes are repeated around the optimum obtained in the previous stage.

TABLE I
DIAGNOSIS RESULTS OF TEST PROCESS

	DIAGNOSIS ACCURACY (%)			
	M1	M2	M3	M4
D1	85	89	95	97
D2	87	90	94	95
D3	86	90	91	92
D4	88	88	95	98
D5	85	89	94	94
D6	87	94	96	97
D7	84	87	88	89

Table I shows the diagnosis results for the seven test batches of D1 through D7. In Table I diagnosis accuracy values (% success of classification) are listed to compare the diagnosis performance of four diagnosis schemes denoted as M1 through M4. Here, diagnosis accuracy is defined as the proportion of the observations correctly diagnosed. M1 denotes the diagnosis scheme of using DPLS, KPCA, and tree classifier, but M2 shows the M1 using OSC instead of DPLS. On the other hand, M3 differs from M1 in that it utilizes KFDA rather than KPCA. The only difference M4 has is that it replaces DPLS with OSC. As shown in Table I, the M4 diagnosis scheme showed the best diagnosis performance in that it yielded the highest diagnosis accuracy for all test batches except D5’s equal value 94% for M3 and M4.

In terms of the average values of diagnosis accuracy over the test batches, the M4 diagnosis method based on KFDA and OSC produced the best diagnosis performance (i.e., average diagnosis accuracy of 94.6. On the contrary, average values of M1, M2, and M3 are 86, 89.6, and 93.3, respectively. It should be also noted that overall diagnosis performance of using KFDA (i.e., M3 and M4) outperforms that of using KPCA (i.e., M1 and M2), irrespective of filtering techniques used. The use of KFDA diagnosis methods has significantly improved the diagnosis performance. This is meaningful to operating personnel who need to take control actions using the diagnosis making results are important for quality of batch production

Similar to Table I, diagnosis results for the test batch process using M1 through M4 diagnosis schemes are reported in Table II and Table III. Here, only difference between Table I and the two tables is that different estimation methods for future observations for batch process are utilized. The fault library method of [13] was applied to obtain the diagnosis results of

Table I. On the other hand, PCA-based estimated future observations for batch data are used for Table II whilst current deviation method are adopted in resulting in Table III. The current deviation approach for future observations are illustrated as shown in Fig. 4.

TABLE II
DIAGNOSIS RESULTS USING PCA-BASED FUTURE VALUES

True Cause	DIAGNOSIS ACCURACY (%)			
	M1	M2	M3	M4
D1	84	87	94	94
D2	85	90	94	95
D3	86	88	90	91
D4	87	84	94	97
D5	82	85	93	93
D6	84	90	95	96
D7	80	84	87	87

TABLE III
DIAGNOSIS RESULTS USING CURRENT FUTURE VALUES

	DIAGNOSIS ACCURACY (%)			
	M1	M2	M3	M4
D1	81	83	91	92
D2	83	90	88	89
D3	83	84	87	88
D4	84	84	89	92
D5	81	82	87	92
D6	84	87	89	94
D7	80	83	84	85

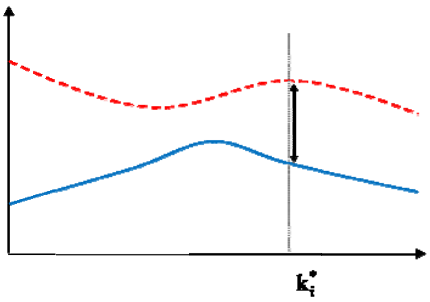

Fig. 4 Estimation using current deviation approach

Table II and Table III showed the diagnosis results similar to Table I in that the M4 produced the best diagnosis accuracy for the batch test data in all the fault data. For example, as shown in Table II, the M4 yielded maximum values except same diagnosis accuracy in D1, D5, and D7 (obtained by M3). When we compare the effect of using different estimation methods, it is obtained that the results of Table I outperformed Table III and Table III. More specifically, the difference between PCA-based method and current deviation method as shown in Fig. 4 can be shown in these tables. Also the use of KFDA improved diagnosis performance significantly when compared to using KPCA. This is believed to stem from the fact that the algorithm or philosophy of FDA is suitable for classification rather than for capturing maximum variation.

IV. CONCLUSION

A nonlinear representation based diagnosis was presented with the filtering techniques used for the elimination of unimportant variation of data. The use of a nonlinear technique of KFDA was able to represent nonlinear behavior in the input data of batch processes. Considering the widespread availability of nonlinear industrial batch type production, such nonlinear technique based diagnosis schemes would be helpful to make diagnostic decision. It is also due to the fact that they have much nonlinear characteristics involved inherently. Compared to other diagnosis approaches, furthermore, empirical model based schemes for diagnosis can be implemented easily. A case study on the batch process has shown that the use of KFDA combined with OSC yielded reliable diagnosis results. It turned out that the M4 outperforms the other diagnosis schemes of M1, M2, and M3. In terms of future observations handling, we tested the three estimation methods using the same process. Though not shown here, it was demonstrated that the nonlinear kernel method outperforms the linear methods of PCA and FDA. The performance of the four diagnosis schemes would be directly affected by the quality of historical batch data. In this case, it would help to gather as many batch data as possible, but this inevitably results in a computational problem. Another issue is concerned with the case of a new type of fault. It is expected to be highly important, particularly when the batch process has frequent operational changes over time.

REFERENCES

[1] S. J. Qin, "Statistical process monitoring: basics and beyond," *Journal of Chemometrics*, vol. 17, pp. 480–502, 2003.
[2] S. Bersimis1, S. Psarakis, and J. Panaretos, "Multivariate statistical process control charts: an overview," *Quality and Reliability Engineering International*, vol. 23, pp. 517–543, 2007.
[3] X. Meng, A. J. Morris, and E. B. Martin, "On-line monitoring of batch processes using PARAFAC representation," *Journal of Chemometrics*, vol. 17, pp. 65-81, 2003.
[4] V. A. Sotiris, P. W. Tse, and M. G. Pecht, "Anomaly detection through a bayesian support vector machine," *IEEE Transactions on Reliability,* vol. 59, pp. 277-286 , 2010.
[5] R. Lombardo, J.-F. Durand, and A. P. Leone, "Multivariate additive PLS spline boosting in agro-chemistry studies," *Current Analytical Chemistry*, vol. 8, pp. 236-253, 2012.
[6] 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*, vol. 50, pp. 243-252, 2000.
[7] J. C. Wong, K. A. Mcdonald, and A. Palazoglu, "Classification of abnormal plant operation using multiple process variable trends," *Journal of Process Control*, vol. 11, pp. 409-418. 2001.
[8] G. Baudat, and F. Anouar, " Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, pp. 2385-2404, 2000.
[9] R. Rosipal, and L. J. Trejo, "Kernel partial least squares regression in reproducing Kernel Hilbert space," *Journal of Machine Learning Research*, vol. 2, pp. 97–123, 2001.
[10] B. Schölkopf, A. J. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299-1319, 1998.
[11] A. Boulesteix, G. Tutz, and K. Strimmer, "A CART-based approach to discover emerging patterns in microarray data," *Bioinformatics*, vol. 19, pp. 2465-2472, 2003.
[12] J. Westerhuis, S. de Jong, and A. K. Smilde, "Direct orthogonal signal correction," *Chemometrics and Intelligent Laboratory Systems*, vol. 56, pp. 13-25, 2001.
[13] H. W. Cho, and K. J. Kim, "A method for predicting future observations in the monitoring of a batch process," *Journal of Quality Technology*, vol. 35, pp. 59-69, 2003.