Dynamic Kernel Scatter-difference-based Discriminant Analysis for Diagnosis of Tennessee Eastman Process

Sumana C., Venkateswarlu Ch., Ravindra D. Gudi*, Mani Bhushan

Abstract - A dynamic kernel scatter-difference-based discriminant analysis (DKSDA) method, that addresses overlapping and auto-correlated data resulting from different types of abnormal situations, is proposed here for fault diagnosis of nonlinear chemical processes. The proposed method is based on scatter-difference-based discriminant analysis performed in a high dimensional nonlinear feature space that is obtained via nonlinear kernel transformation of a suitably lagged, dynamic representation of the variables. The DKSDA overcomes the singularity problem of within-class scatter matrix that is encountered in kernel Fisher discriminant analysis (KFDA), by considering scatter difference form of the Fisher criterion. Fault diagnosis is performed by scores classification using the nearest neighbor classifier in DKSDA space. The performance of the proposed method is evaluated by applying it for the isolation of complex faults in the Tennessee Eastman process. The results demonstrate the superiority of the DKSDA over other recently reported nonlinear classification methods.

I. INTRODUCTION

Efficient fault diagnosis of chemical and biochemical processes is essential to ensure their safe and optimal operation. The task of fault diagnosis is to identify an assignable cause of out-of-control status of the process which can then enable corrective actions required to bring the process back to its normal status. Several methods for fault diagnosis based on multivariate statistical analysis have been developed in recent years in order to meet the demands for better product quality and operation safety in process industries. These methods include principal component analysis (PCA) [1], partial least squares (PLS) [2], and Fisher discriminant analysis (FDA) [3]. The increasing demand for such data driven techniques for fault diagnosis is due to the availability of large amounts of multivariable data as a result of recent advances in automated measurement techniques as well as easy availability of data storage devices. Among these methods, PCA is the most popular method for extracting information from measured data. PCA involves representing the high dimensional process data in reduced dimensions by finding components that are efficient for describing the data. Fault diagnosis by PCA and PLS is carried out by using contribution plots and reconstruction techniques [4]. However, the contribution plots may not effectively identify the cause of an abnormal condition [5], as the effect of fault is observed in more than one variable. FDA is widely used technique for pattern classification [6], which has been recently explored for process fault diagnosis [7]. The basic idea of FDA is to find directions in which the Fisher criterion function is maximized. These directions provide an optimal lower dimensional representation in terms of discrimination among classes, which is very useful for fault diagnosis. Though FDA is discrimination oriented, and hence in general a better tool for fault diagnosis than PCA, it is still linear in nature and may not represent the nonlinear behavior in the data [8].

As most industrial processes are nonlinear, the above presented linear analysis methods may not perform effectively for fault diagnosis of real world problems. For such problems, the measurement data can be mapped via nonlinear transformations into a high dimensional space known as feature space (F). The transformed features in F now vary linearly and well established linear methods can be applied. Towards this end, recently, many kernel methods such as support vector machine (SVM) [9], kernel PCA (KPCA) [10], kernel PLS (KPLS) [11], and kernel FDA (KFDA) [12] have been proposed. These kernel methods are nonlinear methods that are designed by means of linear techniques in implicit feature spaces induced by kernel functions. KPCA is a nonlinear extension of PCA [13], which computes the principal components in a high dimensional feature space \( F \) that is nonlinearly related to the input space. KPCA aims at reconstruction but not classification, hence it is not well suited for the task of fault diagnosis. KFDA maps the input sample data into a kernel feature space by a nonlinear kernel function and then performs linear FDA [14] in the nonlinearly mapped feature space to find the discriminant feature vectors for classification. KFDA is well proven tool for nonlinear classification, but it suffers from the inversion problem arising due to the singularity condition of within-class covariance matrix (required for computing optimal Fisher directions) in the feature space due to reasons such as small sample size, high dimensionality and dynamic correlations present in data. In KFDA, the dimensionality in the feature space equals the total number of samples in the training data set. In case of multi class problems, as the number of classes...
increases, the number of samples in training set increases resulting in larger dimension of the feature space. Thus, the increase in dimension leads to the within-class-covariance matrix becoming singular, a problem which becomes severe especially when the data is highly auto correlated.

Recently, a kernel scatter-difference-based discriminant analysis (KSDA) was introduced in pattern recognition literature to avoid the inverse problems associated with KFDA [15]. This method considers the scatter-difference-based form of Fisher criterion instead of the quotient form, a modification that eliminates the computation of inverse while having an objective that is consistent with that of KFDA, i.e. minimizing within-class scatter while maximizing between-class-scatter. However, KSDA cannot effectively deal with the overlapping and auto-correlated data typically associated with chemical process fault diagnosis. In this work therefore, a dynamic kernel scatter-difference-based discriminant analysis (DKSDA) is proposed for fault diagnosis of processes involving nonlinear, overlapping and auto-correlated data resulting from different types of abnormal situations. This method incorporates dynamic lags in the data to eliminate the effects of auto-correlations so as to improve the classification efficiency. We demonstrate the superiority of the proposed method through the benchmark Tennessee Eastman case study.

II. KERNEL SCATTER-DIFFERENCE-BASED DISCRIMINANT ANALYSIS (KSDA)

The Kernel Fisher discriminant analysis (KFDA), as discussed earlier, is the kernel version of linear FDA but suffers from inverse computation problem in the feature space due to large dimensionality and small sample size of the data. The modified Kernel scatter-difference-based discriminant analysis (KSDA) overcomes this problem by solving the classification problem in $F$, but with a modified form of Fisher’s criterion. It successfully eliminates the inverse problem of KFDA, while meeting the same objective of minimizing within-class scatter and maximizing between-class-scatter.

For the diagnosis problem being considered, let the historic data collected from a nonlinear chemical process operating at different faulty conditions be stacked together into a matrix $X \in \mathbb{R}^{N \times m}$, where $N$ is the total number of samples collected for each of the dimensions 1,2,...m; we further assume that this matrix includes $n$ samples from each of $c$ fault classes labeled $C_1, C_2,...,C_c$. Let the sample vector be represented by

$$x_i = [x_{i1}, x_{i2}, ..., x_{in}]^T$$

(1)

The nonlinear transformation, $\psi$ through which $X$ can be transformed into a high or even an infinite dimensional feature space $F$ is given by [13],

$$\psi : x \in \mathbb{R}^n \rightarrow \psi(x) \in F$$

(2)

For simplicity of notation, the mapping of $x_i$ will be denoted by $\psi(x_i) = \phi$. Let the grand mean of all the transformed samples $\phi_i$ be $\bar{\phi}$ and the mean of $i^{th}$ class samples in the feature space be $\bar{\phi}_i$. The within-class scatter and between-class-scatter matrices in the feature space are then given as,

$$S^w_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (\phi_{ij} - \bar{\phi}_i)(\phi_{ij} - \bar{\phi}_i)^T$$

(3)

$$S^b_i = \sum_{j=1}^{n_i} (\phi_{ij} - \bar{\phi})(\phi_{ij} - \bar{\phi})^T$$

(4)

The conventional KFDA computes the optimal discriminating vectors by maximizing the Fisher’s criterion in high dimensional feature space $F$, which is given by,

$$J(\nu) = \max \left\{ \frac{\nu^T S^w \nu}{\nu^T S^b \nu} \right\}$$

(5)

The solution of the above optimization problem is to find the leading eigen vectors of $(S^w)^{-1} S^b$. Since the dimensions in feature space $F$ can be large and often direct computation of $\psi$ is not feasible, working with the mapped data explicitly becomes difficult. A kernel trick is introduced [13] due to which the explicit calculation of the nonlinear function $\psi$ is not required. This involves specification of a suitable kernel function $k$ that calculates the dot product of two vectors in the feature space, without explicitly calculating the transformation $\psi$:

$$k(x, y) = (\psi(x), \psi(y))$$

(6)

Since most classification algorithms can be posed in a form which requires only the dot product calculation, the kernel trick is very useful. Some of the nonlinear kernel functions typically used are given below,

**Polynomial:**

$$k(x, y) = (x^T y + \beta)^\delta, \ \beta \in \mathbb{R}$$

**Gaussian or Radial basis:**

$$k(x, y) = \exp \left( -\frac{||x - y||^2}{\delta} \right), \ \delta \in \mathbb{R}$$

(7)

**Sigmoid Kernel:**

$$k(x, y) = \tanh(\beta_0 (x^T y) + \beta_1)$$

where, $\delta, \beta, \beta_0$ and $\beta_1$ are the kernel design parameters to be specified a priori by the user. A specific choice of kernel function implicitly determines the mapping $\psi$ and the feature space $F$. Since various kernel functions correspond to different feature spaces $F$, it is important to choose an appropriate kernel function that delivers the best performance in terms of efficient classification for the problem at hand. Any solution vector $\nu \in F$ of Fisher criterion in (5) must lie in the span of all training samples [12]. Then there exist coefficient vector $\alpha$ such that $\nu$ can be expressed as,

$$\nu = \sum_{i=4}^{N} \alpha_i \psi(x_i) = \phi \alpha \cdot \alpha = [\alpha_1, \alpha_2, ..., \alpha_N]^T$$

(8)

where,

$$\phi = [\psi(x_1), \psi(x_2), ..., \psi(x_n)] = [\phi_1, \phi_2, ..., \phi_1, ..., \phi_n]$$
By projecting $\phi_i$ onto a vector $\nu$, one can get
$$v^T \phi = \alpha^T \phi^T \phi_i$$
(9)
where,
$$\alpha^T [\phi_1^T, \phi_2^T, ..., \phi_i^T, ..., \phi_N^T] \nu = \alpha^T \xi_i$$
(10)
By using the kernel trick, the sample vector $x_i$ in the input space is transformed to the kernel vector $\xi_i$ in the feature space, which is the kernel dot product of the $i^{th}$ sample with all $N$ samples in the training data set, and can be defined as follows,
$$\xi_i = [k(x_1, x_1), k(x_2, x_1), ..., k(x_N, x_1)]^T$$
(11)
The kernel matrix $K$ is constructed by considering kernel vectors $\xi$ corresponding to all $N$ samples, and is given by,
$$K_{NN} = [\xi_1, \xi_2, ..., \xi_N]^T$$
(12)
The Fisher discriminant criteria in the feature space from (5) is equivalent to the function expressed in terms of kernel matrices and is given as,
$$J(\alpha) = \max_{\alpha \neq 0} \frac{\alpha^T K_{\nu} \alpha}{\alpha^T K_{\mu} \alpha}$$
(13)
$$v^T S_{\mu}^{-1} v = \alpha^T K_{\alpha} \alpha$$
(14)
The between-class and within-class kernel matrices $K_{\alpha}$ and $K_{\mu}$ can be computed as follows [15],
$$K_{\mu} = \sum_{i=1}^{N} \frac{1}{\eta_i} \sum_{j \in C_i} (\xi_j - \mu) (\xi_j - \mu)^T$$
(15)
$$K_{\alpha} = \sum_{i=1}^{N} (\mu - \mu) (\mu - \mu)^T$$
(16)
Here, $\mu$ is the overall mean of all kernel vectors $\xi$ and $\mu_i$ is the mean of individual class $C_i$ in the kernel space, and is given by,
$$\mu_i = \frac{1}{\eta_i} \left[ \sum_{j \in C_i} k(x_1, x_j), \sum_{j \in C_i} k(x_2, x_j), ..., \sum_{j \in C_i} k(x_N, x_j) \right]^T$$
(17)
In KFDA, in order to find the optimal discriminating vectors in the feature space, one has to maximize $J(\alpha)$, whose solution is to diagonalise the matrix $(K_{\mu})^{-1} K_{\alpha}$ and find its leading eigen vectors $\alpha$, which are the optimal Fisher directions in the feature space. But the matrix $K_{\mu}$ cannot be guaranteed to be nonsingular and its inverse computation is not always mathematically feasible. This problem arises due to small sample size and high dimensionality of training data.

In order to address this issue, we consider a very recently reported method known as kernel scatter-difference-based discriminant analysis (KSDA) [16]. It overcomes the matrix singularity problem while still solving the nonlinear classification problem. The kernel trick is first employed to construct an implicit feature space $F$, and then scatter-difference-based discriminant rule is defined to analyze the data in feature space and produce nonlinear discriminating features, while avoiding the inverse computation. The primary idea is to maximize the following objective function
$$J(w) = \max_{w \neq 0} \left( w^T (S_{\nu}^{-1} - M S_{\mu}^{-1}) w \right)$$
(18)
where $M$ is a positive constant. Maximizing the objective function $J(w)$ is also equivalent to maximizing between-class-scatter and minimizing within-class scatter in the feature space, which is consistent with Fisher’s discriminant objective as in (5). $J(w)$ can be expressed in terms of kernel matrices, and can be calculated explicitly as follows,
$$J(\alpha) = \max_{\alpha \neq 0} \frac{\alpha^T (K_{\nu} - P K_{\mu} \alpha)}{\alpha^T (K_{\mu} - P K_{\mu} \alpha)}$$
(19)
Assuming $||\alpha|| = 1$, maximization problem of $J(\alpha)$ is equivalent to solving for the maximum of Lagrange function
$$L(\alpha, \lambda) = J(\alpha) - \lambda \sum_{i=1}^{N} w_i$$
(20)
Setting $\frac{\partial L(\alpha, \lambda)}{\partial \alpha} = 0$, leads to the form
$$(K_{\nu} - P K_{\mu}) \alpha = \lambda \alpha$$
(21)
The solution of the optimization problem is then translated to computing the leading eigen vectors of $(K_{\nu} - P K_{\mu})$, which does not involve any matrix inverse computation and hence successfully avoids the numerical problem. The factor $P$ is a nonnegative parameter, which is used to balance the within-class and between-class-scatter.

For a new sample $x_{new}$ its projection on to $\nu$ in $F$ can also be calculated by,
$$v^T \phi_{new} = \sum_{i=1}^{N} \alpha_i k(x_i, x_{new})$$
(22)

III. DYNAMIC KERNEL SCATTER-DIFFERENCE-BASED DISCRIMINANT ANALYSIS (DKSDA)

Though the static KSDA performs the nonlinear fault diagnosis by successfully eliminating the inverse problem in the feature space, it may not be effective for diagnosis purposes when the data contains dynamic auto correlations or when the classes representing the data overlap. In order to cope with such nonlinear dynamic systems, a dynamic kernel scatter-difference-based discriminant (DKSDA) analysis is proposed here, which incorporates dynamic lags in the data and performs scatter-difference-based discriminant analysis in a high dimensional nonlinear feature space.

A. Elimination of Data Correlations - Incorporation of Dynamic Lags

The static KSDA assumes implicitly that observations at one time instant are statistically independent of observations at past time instants, which could be justified for large sampling times in the order of hours. But for typical dynamic nonlinear chemical processes where sampling periods are usually of the order of minutes and seconds, this assumption is generally not valid. Here, this issue is addressed for the purpose of fault diagnosis by augmenting each observation vector with the previous $l$ observations, which accounts for the correlations present in the data. The procedure adopted to address this issue is similar to that of used by Chiang et al.[17] and Ku et al.[18] for the fault detection studies of dynamic systems. This concept is
recently used in case of dynamic KPCA [19] for capturing nonlinear process dynamics and eliminating the auto correlations in the data.

To illustrate, suppose the measurement vector $x_i$ is two-dimensional of the form $[x_{i1}, x_{i2}]^T$ at time $t$, then the total data matrix $D(l)$ containing all the samples after incorporating $l$ number of time lags becomes,

$$D(l) = \begin{bmatrix}
x_{11} & x_{12} & x_{13} & \cdots & x_{1l}
x_{21} & x_{22} & x_{23} & \cdots & x_{2l}
\vdots & \vdots & \vdots & \ddots & \vdots
x_{11} & x_{12} & x_{13} & \cdots & x_{1l}
\end{bmatrix}$$ (23)

When sufficient number of dynamic lags $l$ are included in the data matrix, then the rows could become statistically independent between time instants due to the explicit modeling of the auto correlations. Also, the proposed DKSDA enables the nonlinear diagnostic method to use more information contained in the augmented observation vector, which is the super set of the information contained in a single observation vector. The augmented data matrix, whose samples are statistically independent, is projected into the feature space and the scatter-difference-based rule is applied. This method therefore eliminates the problems associated with data correlations and ill-conditioning of the inverse while addressing the nonlinear fault diagnosis problem.

**B. Computation of Misclassification**

One measure of the similarity between the samples is the distance between them. In our approach, the misclassification analysis is performed in the scores space of DKSDA. The scores are subjected to a nearest neighbor classifier with an appropriate distance metric and used to compute the misclassification quantitatively. Initially, the DKSDA model is built using the training data. The training data is projected onto the DKSDA feature space, and the scores of each fault class and its individual means are computed. Next, the test data is projected on to the same DKSDA scores plane sample wise, and its distance from the mean of each fault class is computed. The test sample is assigned to the class to which it is found to be the nearest. The distance metric, that is used in nearest neighbor classifier is given by the Mahalanobis distance:

$$d_{i,\text{Mahalanobis}} = \sqrt{([\xi_{\text{test}} - m_i]\Sigma_i^{-1}([\xi_{\text{test}} - m_i])^T}$$ (24)

where $\xi_{\text{test}}$ is the discriminant score vector of test sample in the DKSDA plane, $m_i$ is the mean and $\Sigma_i$ is the covariance of scores of the fault class $C_i$ for $i=1,2,\ldots,c$.

**IV. TENNESSEE EASTMAN PROCESS**

The Tennessee Eastman Process (TEP) was formulated as a challenge problem, representative of an industrial nonlinear chemical process. Further details regarding the process can be obtained from [20]. It has been widely used as a challenging test bed for evaluating many new approaches for chemical process diagnosis [21]. TEP is a complex nonlinear chemical process, which includes a reactor, a condenser, a stripper, a vapor liquid separator and a recycle compressor, the schematic of which is given in Fig.1 [20]. There are 52 measurements which consist of 41 manipulated variables such as valve settings for feed and coolant streams and 11 process measurements.

![Fig.1 Schematic of Tennessee Eastman process [20]](http://www.brahms.scs.uiuc.edu)

Though, there are 21 different faulty conditions along with the normal operating conditions, our present study is focused only on a particular class of faults, which are complex, non-linear and overlapping in nature and which cannot be easily separated either by linear or nonlinear methods [14]. They are reported in literature as fault 4, fault 9 and fault 11, and tested with several fault diagnostic methodologies [14, 21, 22]. Fault 4 involves a step change in the reactor cooling water inlet temperature. Fault 9 is the random variation induced in the feed temperature and Fault 11 is the random variation induced in the reactor cooling water inlet temperature. Faults 4 and 11 are associated with reactor cooling water inlet temperature, but they are different in terms of the type of fault. All these 3 faults belong to the reactor and affect its temperature.

**V. RESULTS**

The data corresponding to the overlapping fault classes fault 4, fault 9 and fault 11 was generated from the realistic standard model of Tennessee Eastman challenge problem which is available at [http://www.brahms.scs.uiuc.edu](http://www.brahms.scs.uiuc.edu). The data contains both train and test sets with sampling interval of 12 minutes. Each train data set consists of 480 samples while the test data set consists of 960 samples with fault introduced at sample 160. Each sample vector contains 52 variables, out of which the variables that are responsible for these faults are only considered for further analysis. The key variables of these nonlinear, overlapping type faults are identified by a genetic algorithmic (GA) based FDA variable selection scheme in [14] and, by a feature selection algorithm based on mutual information in [22]. It is to be noted that variables 9 (reactor temperature) and 51 (reactor cooling water valve position) are important in terms of
distinguishing the three fault classes. Only variable 51 is important for fault 4 and variables 9 and 51 for fault 9 and fault 11. The rest of the 50 variables do not provide much information in terms of classification.

The original data of the three fault classes is plotted in the two dimensional plane corresponding to these two variables in Fig. 2. It is clearly noted from this figure that fault 4 dataset can be separated from fault 9 dataset using variables 9 and 51, but these two data sets overlap with the fault 11 data set. The study of variable selection methodology is currently not in the scope of our objective.

The time lags can be incorporated for these two auto-correlated variables to decrease the degree of overlap between the data with added dimensions. The training data matrix of each class with variables 9 and 51 is of dimension 480x2. With incorporation of time lags \( l \) for these variables the dimension of the augmented matrix becomes \( (480-l) \times 2(l+1) \). The performance of DKSDA as a function of dynamic lags is studied here, using 4 fold cross validation procedure. The average misclassification vs. data lags considered in DKSDA model building is shown in the bar chart displayed in Fig.3. The DKSDA misclassification for the data with no lags is found to be high, and as the number of lags increase, the misclassification rate is found to decrease and exhibits minima at a lag value of 3. The optimum number of dynamic lags that provide improved classification in the DSKDA space was therefore found to be 3. In the above analysis Gaussian kernel is considered for nonlinear transformation and Mahalanobis distance metric is considered for nearest neighbor method for finding the scores misclassification in DKSDA plane. The DKSAD scores plot shown in Fig. 4 displays clear discrimination between the classes when compared to the original overlapping data plot shown in Fig. 2.

The results obtained using the KSAD and DKSAD with optimal lags are compared with the recently reported results for the case of discriminating the overlapping type of faults of Tennessee Eastman challenge problem, and are displayed in Table 1.

<table>
<thead>
<tr>
<th>Method</th>
<th>% Misclassification</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDA</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>SVM</td>
<td>5.6</td>
<td>6.5</td>
</tr>
<tr>
<td>PSVM</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Invariant</td>
<td>-</td>
<td>6.0</td>
</tr>
<tr>
<td>SVMs</td>
<td>LDA</td>
<td>31.58</td>
</tr>
<tr>
<td></td>
<td>QDA</td>
<td>5.87</td>
</tr>
<tr>
<td></td>
<td>KSDA</td>
<td>8.8235</td>
</tr>
<tr>
<td></td>
<td>DKSDA</td>
<td>1.3305</td>
</tr>
</tbody>
</table>

It can be seen from Table 1 that the proposed Dynamic kernel scatter-difference-based discriminant analysis has shown much superior performance over other methods for both train as well as test data set.
VI. CONCLUSIONS

The proposed dynamic kernel scatter-difference–based discriminant analysis (DKSDA) based nonlinear fault diagnosis method is found to be superior in isolating the causes for the case of auto correlated data and overlapping classes. It efficiently eliminates the auto correlations in the data by incorporating time-lagged data, thus capturing dynamic characteristic of processes. The nonlinear transformations that are essential to classify faults are achieved via optimal kernel functions. It also avoids the inverse computation problem by considering the scatter-difference-based form of Fisher criterion and classifies the scores, with suitable distance metric to facilitate improved diagnosis for processes with severe nonlinearity and auto correlations. It is shown that the proposed DKSDA outperforms several nonlinear classifiers that are recently published in the literature for the case of overlapping faults of Tennesse Eastman Challenge Process.

REFERENCES


