Gas Turbine Fault Detection and Identification by using Fuzzy clustering methods

Amin Ollah Khormali¹, Mahdi Aliyari Sh².

Abstract— In this paper fault detection of gas turbine using signal based methods has been studied. First gas turbine data from a precise, validated and existent simulator of gas turbine acquired. This simulator is capable of simulating four common faults of gas turbine. There is one normal class and four different fault classes. Data from sixteen variables had stored. Each class has 780 samples of observations. Feature selection methods have been applied on the data set. Principal component analysis (PCA) and linear discriminant analysis (LDA) used as feature selection methods. Fuzzy C-means and Gustafson-Kessel clustering algorithm used to fault detection and identification of the faults of the gas turbine. Confidence matrix and correct rate methods used to evaluate the clustering performance. The simulation results show that clustering methods has acceptable performance for fault detection and identification of the gas turbine faults. The combination of LDA and Gustafson-kessel clustering algorithm had the best performance as 96.96% correct rate.

Key words: Fault Detection and Identification-Gas Turbine-Signal analysis-Feature selection -Fuzzy clustering

I. INTRODUCTION

Gas turbines (GT) are used widely in industry. The gas turbine is a main part of a power plant, which produces a great amount of energy for its size and weight. The GT is a system that provides torque with adjustable speeds that can be used to rotate generators. It is made of lots of stationary and rotary parts which makes the GT a very complex system with a dynamical performance. The gas turbine has established growing service in the past 40 years in the power industry both among utilities and merchant plants as well as the petrochemical industry, and utilities throw the world [1].

Due to high complexity and high price of the gas turbine importance of maintenance and fault detection and identification of the GT increases. Preventive maintenance can be performed periodically in fixed time intervals. Fault detection and identification will optimize the maintenance and repair time of the gas turbine. This affects the performance of the gas turbine and also economical aspects [2]. Fig.1 shows the cost distribution of a gas turbine in its operation cycle. It can be seen that only 10 percent of the total costs of gas turbine’s operation depends on its initial investment, whereas 15 percent depends on maintenance costs and the remaining 75 percent is the exploitation costs.

Experimental results of gas turbine plants proves that, only half percent increasement in initial investment costs due to a fault detection and identification system leads to 4 percent decrease in the maintenance costs.

Clustering is an unsupervised learning method to partition a collection of multivariate data points into meaningful groups, where all members within a group represent similar characteristics and data points between different groups are dissimilar to each other. It has been an important technique for pattern recognition, image processing and data mining. It has also been applied successfully in many fields such as marketing that finds groups of customers with similar purchasing behaviors, biology that groups unknown plants/animals into species, and medical image processing that divides an image into a few meaningful regions for diagnosis. The similarity criterion for distinguishing the difference between data points is generally measured by distance. Two data points belong to the same group if they are close to each other. They are evidently from different groups if the distance between them is distinctly large [3].

Therefore, clustering methods are useful techniques to group an unknown data set to some meaningful subgroups. In the term of fault detection and identification, the subgroups are the normal and the different fault groups.

In this paper, in order to fault detection and identification of a gas turbine, fault-affected data set of a gas turbine has been clustered in to a normal and some different fault groups.

The paper is organized as follows: in section II, the simulated model of an industrial gas turbine is taken into account. In section III feature selection methods are described. Section IV is specified to the description of the common fuzzy clustering methods. Results of clustering the gas turbine data is presented in section V. The confidence matrix and the correct rate are presented in the same section to evaluate the performance; and finally the section VI concludes the paper.
II. DESCRIPTION OF THE SIMULATED MODEL OF GAS TURBINE

Recently a simulated model of an industrial gas turbine has been presented in [4]. This model is capable to simulate the gas turbine precisely. This model has been validated with a single-shaft industrial gas turbine so its performance is matching the real gas turbines performance. Thus this simulated model could be a useful instrumental for academic studies. Fig.2 shows the simulated model.

![Simulated model of an industrial gas turbine](image)

This model is also able to simulate some of the common faults of an industrial gas turbine. Four common faults which have been simulated in this model are:

- Fault 1 = Compressor contamination
- Fault 2 = Thermocouple sensor fault
- Fault 3 = High pressure turbine seal damage
- Fault 4 = Fuel actuator friction wear

These four faults are effect on the 16 variable of the model. TABLE I shows the variables.

<table>
<thead>
<tr>
<th>variable</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>Mass flow</td>
</tr>
<tr>
<td>m3</td>
<td>Mass flow</td>
</tr>
<tr>
<td>m8</td>
<td>Turbine outlet mass flow</td>
</tr>
<tr>
<td>m9</td>
<td>Pressure valve mass flow</td>
</tr>
<tr>
<td>p2</td>
<td>Compressor inlet temperature</td>
</tr>
<tr>
<td>p3</td>
<td>Compressor exit pressure</td>
</tr>
<tr>
<td>p7</td>
<td>Turbine back pressure</td>
</tr>
<tr>
<td>pt</td>
<td>Power</td>
</tr>
<tr>
<td>qt</td>
<td>Torque</td>
</tr>
<tr>
<td>t3</td>
<td>Compressor exit temperature</td>
</tr>
<tr>
<td>t4</td>
<td>Turbine cooling air temperature</td>
</tr>
<tr>
<td>t5</td>
<td>Compressor exit temperature</td>
</tr>
<tr>
<td>t7</td>
<td>Pressure valve up stream temperature</td>
</tr>
<tr>
<td>wt</td>
<td>Turbine speed</td>
</tr>
<tr>
<td>t3a</td>
<td>Thermocouple sensor</td>
</tr>
<tr>
<td>ff</td>
<td>Fuel flow</td>
</tr>
</tbody>
</table>

By analyzing of these variables we will be able to detect if gas turbine is working in normal condition or any of these faults are occurred.

Data from the variables named above stored in 5 condition of normal and fault 1 to fault 4 and stack in a matrix and form the data matrix X.

\[
X = \begin{bmatrix}
    x_{11} & \cdots & x_{1N} \\
    \vdots & \ddots & \vdots \\
    x_{M1} & \cdots & x_{MN}
\end{bmatrix}
\]

Where \(N=16\) is number of the variables and \(M=3900\) is number of the observations.

III. FEATURE SELECTION METHODS

A. Principal Component Analysis

PCA is a linear dimensionality reduction technique, optimal in terms of capturing the variability of the data [5]. It transforms a set of correlated random variables with zero mean value into a small number of de-correlated variables called principal components while saving as much information as possible from the original variables [6].

Data set with \(N\) variable and \(M\) observation is represented by \(X\) shown below:

\[
X = \begin{bmatrix}
    x_{11} & \cdots & x_{1N} \\
    \vdots & \ddots & \vdots \\
    x_{M1} & \cdots & x_{MN}
\end{bmatrix}
\]

PCA involves a data normalization procedure which leads to variables with zero-mean and unitary standard deviation.

PCA provides a linear mapping of data from the original dimension \(N\) to a lower dimension \(a\) using the transformation:

\[
T = XP
\]

Where \(T \in \mathbb{R}^{M \times a}\) is called the score matrix and \(P \in \mathbb{R}^{N \times a}\) the loading matrix. The dimension \(a\) represents the number of principal. The loading matrix may be found from the main \(a\) eigenvectors of the covariance matrix of \(X\). In real, PCA is often computed by Singular Value Decomposition (SVD) of the covariance matrix, i.e.

\[
XX^T = V \Lambda V^T
\]

Where \(V\) present an orthonormal matrix whose columns define the principal components (PCs) and form a subspace spanning the data. The order \(a\) of the system is determined by selecting the first \(a\) singular values in \(\Lambda\) which have a significant magnitude (‘energy’) as described in [8].

B. Linear Discriminant Analysis

Linear discriminant analysis (LDA) seeks an optimal linear transformation by which the original data is transformed to much lower dimensional space. The goal of LDA is to find a linear transformation that maximizes class separability in the reduced dimensional space. Hence the criteria for dimension reduction in LDA are formulated to maximize the between-class scatter and minimize the within-class scatter. The scatter are measured by using scatter matrices such as the between-class scatter matrix \((S_b)\), within-class scatter matrix \((S_w)\) and total scatter matrix \((S_t)\) [9].

Data set with \(N\) variable and \(M\) observation is represented by \(X\), shown below:
Each data point belongs to one of the C object classes \{X_1, ..., X_C\}. Each class \(i\) has \(M_i\) samples and the total number of data are \(M\).

\[
M = \sum_{i=1}^{C} M_i
\]

The between-class scatter matrix is defined as:

\[
S_B = \sum_{i=1}^{C} \sum_{j \in m_i} (x_j - \bar{x})(x_j - \bar{x})^T
\]

(7)

Where \(\bar{x}_i\) denotes the \(i\)th class mean and \(\bar{x}\) is the global mean of the entire sample. The number of samples in class \(X_i\) is denoted by \(M_i\) and \(m_i\) is the index set of data items in the class \(i\).

\[
\bar{x}_i = \frac{1}{M_i} \sum_{j \in m_i} x_j
\]

(8)

\[
\bar{x} = \frac{1}{M} \sum_{i=1}^{M} x_j
\]

(9)

The within-class scatter matrix is defined as:

\[
S_W = \sum_{i=1}^{C} \sum_{j \in m_i} (x_j - \bar{x}_i)(x_j - \bar{x}_i)^T
\]

(10)

The total-scatter matrix is equal to the sum of the between-class scatter matrix and the within-scatter matrix [10].

\[
T = XP
\]

(11)

The optimal dimension reducing transformation for LDA is the one that maximizes the between-class scatter and minimizes the within-class scatter in a reduced dimensional space [9].

LDA finds a matrix, \(U\), maximizing the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix as:

\[
U_{opt} = \arg \max_U \frac{|U^T S_B U|}{|U^T S_W U|} = [u_1, u_2, ..., u_N]
\]

(12)

The LDA vectors, \(\{u_i|i = 1, 2, ..., N\}\), are equal to the eigenvectors \(u_i\) of the generalized eigenvalue problem[11].

\[
S_B u_i = \lambda_i S_W u_i
\]

(13)

Where the eigenvalues \(\lambda_i\) indicate the degree of overall separability among the classes by projecting the data on to \(u_i\).

The first LDA vector is the eigenvector associated with the largest eigenvalue. The second LDA vector is the eigenvector associated with the second largest eigenvalue, and so on [5].

The linear transformation of the data \(X\), from \(M\)-dimensional space to \((C-1)\)-dimensional space is described by:

\[
T = XP
\]

(14)

where \(P\) is first \((C - 1 )\) LDA vectors as columns. LDA computes the matrix \(P\) such that data \(X\) for the \(C\) classes are optimally separated when projected in to the \((C - 1)\) dimensional space [5].

IV. CLUSTERING

Clustering is a procedure in which individual items are placed into groups on the basis of quantitative information regarding one or more characteristics inherent between different items [12]. The term “clustering” is used to describe the procedure of partitioning data into clusters such that data points in a cluster are more similar to each other than to other points belonging to different clusters [13]. There are many different clustering algorithm but fundamental of all of them are similar to each other and that is grouping the data by measuring the distance. The differences are because of the kind of the distance that every algorithm uses. If two data point are close to each other they belong to one group and if the distance between them is large they are from different groups. The performance of the clustering algorithms is dependent on the data structure including the cluster shape, cluster density and linear or nonlinear separability.

The use of different algorithms leads to different results, but there is no single best approach for selecting the best algorithm [13]. The choice of a suitable method is not an easy task, and there is no method that is obviously better than any other; for the results produced by each algorithm are always unreliable up to a point. The choice of the best algorithm for a particular task depends on the form of the problem, the structure of the data and the dominant geography [14].

There are two main group of clustering methods, the crisp clustering method and the Fuzzy clustering method. Crisp clustering methods, allocate each data point to a unique cluster, on the other side fuzzy clustering algorithms result in membership values between 0 and 1 that illustrate the degree of membership for each data point to each of the clusters. The fuzzy clustering methods are studied in this paper.

A. Fuzzy C-means Algorithm

The fuzzy C-means algorithm was presented in its initials form by Dunn in 1973 and completed by Bezdek in 1981 [15]. FCM clusters the data set \(X\) in to \(C\) partition by minimizing the error terms of the distance of each data point \(x\) to all centroids of the \(C\) clusters.

\[
X = [x_{11}, ..., x_{1N}]
\]

(15)

Where \(N\) is number of the variables and \(M\) is the number of the observations. The membership matrix \(U\) presented as:

\[
U = \begin{bmatrix}
u_{11} & \ldots & u_{1C} \\
\vdots & \ddots & \vdots \\
u_{M1} & \ldots & u_{MC}
\end{bmatrix}
\]

(16)

The FCM minimizes the objective function as:

\[
J = \sum_{i=1}^{M} \sum_{j=1}^{C} u_{ij}^m \|x_i - Center_j\|^2
\]

(17)

The parameter \(m\) controls the fuzziness of the clusters. When \(m\) tends to a high values the algorithm sets all the member ships equal while selection of small values close to 1 for \(m\) leads the FCM to the crisp clustering algorithm.

\[
\sum_{j=1}^{C} u_{ij} = 1 \quad i = 1, ..., M
\]

(18)

Minimizing the objective function by using the Lagrange multipliers yields an iterative algorithm named as expectation-maximization (E-M) algorithm [3]. In expectation step the Euclidean distance \(d\) is computed and then the memberships are calculated as:

\[
d_{ij} = \|x_i - Center_j\|
\]

(19)
\[ u_{ij} = \left( \frac{\sum_{c=1}^{C} \left( \frac{d_{ij}}{d_{ic}} \right)^{2/m-1} }{ \Sigma_{c=1}^{C} \left( d_{ic} \right)^{2/m-1} } \right)^{-1} \]

\[ i = 1, \ldots, M \text{ and } j = 1, \ldots, C \]  \hspace{1cm} (20)

In the maximization step the centers of the clusters are updated as:

\[ \text{Center}_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \quad j = 1, \ldots, C \]  \hspace{1cm} (21)

When the convergence condition satisfies, the algorithm stops and at the end a soft clustering of the data is obtained.

### B. Gustafson-Kessel Fuzzy Clustering Algorithm

The Gustafson-Kessel Fuzzy clustering algorithm is proposed by Gustafson and Kessel in 1979 [16]. Gustafson and Kessel extended the FCM algorithm hiring an adaptive distance norm. The GK algorithm is able to identify clusters with different geometrical shapes, as ellipsoidal clusters [17]. To identify clusters with different shapes, each cluster has its specific norm-inducing matrix. This matrix is considered as the covariance matrix of the cluster \( \Sigma_i \), which effects the distance norm.

The GK clustering algorithm partitions the data set \( X \) in to \( C \) clusters.

\[ X = \begin{bmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{M1} & \cdots & x_{MN} \end{bmatrix} \]  \hspace{1cm} (22)

Where \( N \) is number of the variables and \( M \) is the number of the observations. The membership matrix \( U \) presented as:

\[ U = \begin{bmatrix} u_{11} & \cdots & u_{1C} \\ \vdots & \ddots & \vdots \\ u_{M1} & \cdots & u_{MC} \end{bmatrix} \]  \hspace{1cm} (23)

The adaptive distance is defined as:

\[ d_{ij}^2 = \| x_i - c_j \|^2 \Sigma_i \]  \hspace{1cm} (24)

The GK is based on iterative optimization of an objective function defined as:

\[ J = \sum_{i=1}^{C} \sum_{j=1}^{M} u_{ij}^m \cdot d_{ij}^2 \]  \hspace{1cm} (25)

\( m \) is a positive integer grater then one which controls the fuzziness of the memberships. The objective function has been minimized under the following constraints:

\[ 0 \leq u_{ij} \leq 1 \quad i = 1, \ldots, M \text{ and } j = 1, \ldots, C \]  \hspace{1cm} (26)

\[ 0 \leq \sum_{j=1}^{C} u_{ij} \leq M \quad j = 1, \ldots, C \]  \hspace{1cm} (27)

\[ \sum_{i=1}^{C} u_{ij} = 1 \quad i = 1, \ldots, M \]  \hspace{1cm} (28)

Until satisfaction of the convergence condition, in each iteration the membership matrix and the center of the clusters are updated as: [18]

\[ u_{ij} = \left( \frac{\sum_{j=1}^{C} \left( \frac{d_{ij}}{d_{ic}} \right)^{2/m-1} }{ \Sigma_{j=1}^{C} \left( d_{ic} \right)^{2/m-1} } \right)^{-1} \]

\[ i = 1, \ldots, M \text{ and } j = 1, \ldots, C \]  \hspace{1cm} (29)

\[ \text{Center}_j = \frac{\sum_{i=1}^{M} u_{ij}^m x_i}{\sum_{i=1}^{M} u_{ij}^m} \quad j = 1, \ldots, C \]  \hspace{1cm} (30)

### C. Gath-Geva Fuzzy Clustering Algorithm

The Gath-Geva Fuzzy clustering algorithm is proposed by Gath and Geva in 1989 [19]. The GG clusters the data set \( X \) in to \( C \) partition by minimizing the error terms of the distance of each data point \( x \) to all centroids of the \( C \) clusters.

\[ X = \begin{bmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{M1} & \cdots & x_{MN} \end{bmatrix} \]  \hspace{1cm} (31)

Where \( N \) is number of the variables and \( M \) is the number of the observations. The membership matrix \( U \) presented as:

\[ U = \begin{bmatrix} u_{11} & \cdots & u_{1C} \\ \vdots & \ddots & \vdots \\ u_{M1} & \cdots & u_{MC} \end{bmatrix} \]  \hspace{1cm} (32)

In GG clustering algorithm clusters are described by covariance matrix which empowers the algorithm to identify more sophisticated clusters. Volume of the clusters is not limited in GG clustering and the algorithm dose not effect by the volume of each cluster. The GG minimizes the objective function as:

\[ J = \sum_{i=1}^{C} \sum_{j=1}^{M} u_{ij}^m \| x_i - \text{Center}_j \|^2 \]  \hspace{1cm} (33)

\( m \) is a positive integer grater than one which controls the fuzziness degree of the clusters. In order to identify the clusters with different densities and unequal numbers of data point in each cluster an exponential distance measure has been used in GG, 

\[ d_{ij} = \frac{P_i}{\sqrt{\det(\Sigma_j)}} \exp \left( \frac{1}{2} (x_i - c_j) \Sigma_j^{-1} (x_i - c_j)^\top \right) \]

\[ i = 1, \ldots, M \text{ and } j = 1, \ldots, C \]  \hspace{1cm} (34)

Where \( \Sigma_i \) is the covariance matrix of the ith cluster and \( P_i \) is the coefficient designed for eliminating the sensitivity of the algorithm to number of data points in different clusters and computed as [20] :

\[ P_i = \frac{\sum_{j=1}^{M} u_{ij}^m}{\sum_{j=1}^{M} \sum_{i=1}^{C} u_{ij}^m} \]  \hspace{1cm} (35)

Minimizing of the objective function took place in the following iterative process:

Calculating the distance:

Calculating the membership matrix using the following formula:

\[ u_{ij} = \left( \frac{1}{\sum_{j=1}^{C} \left( \frac{d_{ij}}{d_{ic}} \right)^{2/m-1} } \right)^{-1} \]

\[ i = 1, \ldots, M \]  \hspace{1cm} (36)

Computing the center of the clusters using the following equation:

\[ \text{Center}_j = \frac{\sum_{i=1}^{M} u_{ij}^m x_i}{\sum_{i=1}^{M} u_{ij}^m} \quad j = 1, \ldots, C \]  \hspace{1cm} (37)

Computing the covariance of the clusters as:

\[ \Sigma_j = \frac{\sum_{i=1}^{M} u_{ij}^m (x_i - \text{Center}_j)(x_i - \text{Center}_j)^\top}{\sum_{i=1}^{M} u_{ij}^m} \]  \hspace{1cm} (38)

Go to first step until satisfaction of the convergence condition.

### V. Simulation

In this section results of the simulation are shown. At first data for each class acquired from the simulator of the gas
turbine. Clustering the data set into the number of the classes, five, took place in two main part. In first part data set clustered in to 5clusters without applying any feature selection methods on the data set. Fig.2 shows the scheme of the part one.

Fig. 3. Scheme of clustering the data set - part one

TABLE II, III AND IV shows the confidence matrix of the clustering result in part one.

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>CONFIDENCE MATRIX FOR FCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCM</td>
<td>Normal</td>
</tr>
<tr>
<td>Normal</td>
<td>47.88</td>
</tr>
<tr>
<td>Fault1</td>
<td>2.52</td>
</tr>
<tr>
<td>Fault2</td>
<td>0</td>
</tr>
</tbody>
</table>

In part two the feature selection methods applied on the data set and then the selected features clustered in to five clusters using fuzzy clustering methods mentioned in section V. Fig. 3 shows the scheme of part two.

Fig. 4. Scheme of clustering the data set - part two

TABLE V TO X shows the confidence matrix of the clustering result for this part.

<table>
<thead>
<tr>
<th>TABLE V</th>
<th>CONFIDENCE MATRIX FOR PCA+FCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA+FCM</td>
<td>Normal</td>
</tr>
<tr>
<td>Normal</td>
<td>59.69</td>
</tr>
<tr>
<td>Fault1</td>
<td>8.72</td>
</tr>
<tr>
<td>Fault2</td>
<td>8.72</td>
</tr>
<tr>
<td>Fault3</td>
<td>0</td>
</tr>
</tbody>
</table>

In order to compare the results of clustering the data, correct rate of the clustering in the confidence matrixes shown in TABLE XI.

<table>
<thead>
<tr>
<th>TABLE XI</th>
<th>COMPARISON OF THE CORRECT RATES OF DIFFERENT CLUSTERING METHODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering method Feature selection method Correct rate</td>
<td></td>
</tr>
<tr>
<td>FCM</td>
<td>-</td>
</tr>
<tr>
<td>GK</td>
<td>-</td>
</tr>
<tr>
<td>GG</td>
<td>-</td>
</tr>
<tr>
<td>FCM</td>
<td>PCA</td>
</tr>
<tr>
<td>GK</td>
<td>PCA</td>
</tr>
<tr>
<td>GG</td>
<td>PCA</td>
</tr>
<tr>
<td>FCM</td>
<td>LDA</td>
</tr>
<tr>
<td>GK</td>
<td>LDA</td>
</tr>
<tr>
<td>GG</td>
<td>LDA</td>
</tr>
</tbody>
</table>
VI. CONCLUSION

In this paper fault detection and identification of a gas turbine using different clustering methods is done. The correct rate method is used to compare the performance of the different clustering methods.

Firstly without applying any feature selection methods on the data set, different clustering methods utilized in order to clustering the data set in to a normal and 4 different groups. In this case, the FCM clustering method has better performance in comparison with GK and GG techniques. However, there is still underperforming and it is necessary to achieve better performances.

Therefore, at next step to achieve higher performances, feature selection methods applied on the data set and the more important features selected at first and then different fuzzy clustering methods applied on the data set with selected features and clustered in to a normal and 4 different groups. Results indicates that using feature selection methods improves the performance of the fuzzy clustering algorithms. Combination of the LDA feature selection and the GK fuzzy clustering algorithm performs a considerable performance. It is able to cluster the data set in to the exact clusters by more than 96 percent of correct rate.

REFERENCES


