УДК 629.7.036.4
DOI: 10.26467/2079-0619-2020-23-6-20-27

USING MODERN CLUSTERING TECHNIQUES FOR PARAMETRIC FAULT DIAGNOSTICS OF TURBOFAN ENGINES

I.J. BURAIMAH[1]
[1]*National Space Research and Development Agency*
Abuja, Federal Republic Nigeria

The 21st century aviation and aerospace technologies have evolved and become more complex and technical. Turbofan jet engines as well as their cousins, the rocket engines (liquid/solid) have gone through several design upgrades and enhancements during the course of their design and exploitation. These technological upgrades have made engines very complex and expensive machines which need constant monitoring during their working phase. As the demand and use of such engines is growing steadily, both in the civilian and military sectors, it becomes necessary to monitor and predict the behavior of parametric data generated by these complex engines during their working phases. In this paper flight parameters such as Exhaust Gas Temperature (EGT), Engine Fan Speeds (N1 and N2), Fuel Flow (FF), Oil Temperature (OT), Oil Pressure (OP), Vibration and others where used to determine engine fault. All turbo fan engines go through several distinctly different working phases: Take-off phase, Cruise phase and Landing phase. Recording generated parametric data during these different phases leads to a massive amount of in-flight data and maintenance reports, which makes the task of designing and developing a fault diagnostic system highly challenging. It becomes imperative to use modern techniques in data analysis that can handle large volumes of generated data and provide clear visual results for determining the technical status of the engine under investigation/monitoring. These modern techniques should be able to give clear and objective assessment of the object under investigation. Cluster analysis methods based on Neural Networks such as c-means, k-means, self-organizing maps and DBSCAN algorithm have been used to build clusters. Differences in cluster groupings/patterns between healthy engine and engine with degraded performance are compared and used as the bases for defining faults. Fault diagnosis plays a crucial role in aircraft engine management. Timely and accurate detection of faults is the foundation on which maintenance turnaround times, operational costs and flight safety are based. The data used in this paper for analysis was obtained from flight data recorder during one flight cycle. The final decision on a fault is taken by an engineer.

Key word: engine fault diagnostics, parametric data, turbofan jet engines, monitoring, in-flight data handling, neural network, c-means, k-means, dbscan, clustering analysis, cluster pattern, clustering techniques, algorithm, flight parameters, exhaust gas temperature, data analysis, self-organizing maps.

The author expresses his gratitude to Prof. (Dr.) Boris A. Chichkov of the Moscow State Technical University of Civil Aviation, Russia, for his guidance, support and assistance.

INTRODUCTION

Current research in the development of engine fault diagnostics methods have effectively advanced in several directions with the two most popular. The first method is a combination of traditional rule-based diagnostic method (e.g., expert system) with other AI (artificial intelligence) techniques, such as neural network and fuzzy logic [1 – 3]. The other approach uses models of engine performance and is known as model-based fault diagnostics [4, 5]. Aircraft engine health deterioration can be attributed to a variety of reasons including extreme operating conditions, system malfunction, component or sub-system failure and aging. The most pressing challenges faced by most if not every airline are safety, airworthiness, and operational cost effectiveness. To achieve the above-mentioned goals an effective diagnostic method based on data from the engine should form the bases for designing a diagnostic method. A condition-based maintenance system can be divided into two categories: model based and data driven. Dynamic systems or sub-systems under investigation can sometimes be defined by developing accurate mathematical models [6 – 9]. Mathematical models are used to determine the relationship between different measured signals, interpreting trends and using advanced signal processing methods to detect faults. However, in some cases it is difficult to accurately represent a complex sys-

tem using mathematical models and the data from the system has to be used as the basis for analysis. That is why Data Driven Fault Diagnosis Scheme based on statistical methods, machine learning, and statistical pattern recognition approaches are used as the basis for developing new advanced fault diagnostic system for aircraft engine health management. The proposed content has two basic objectives: to show the implementation of clustering algorithms in detecting engine with degraded performance over healthy engine, and compare cluster patterns formed.

The huge amount of data generated during the use of modern turbojet engines and other propulsion systems, having onboard data recorders, demands for progressive new techniques in processing these parametric data generated during operation. Therefore, it is necessary and appropriate to use neural network-based methods such as fuzzy c-mean, self-organizing maps and others presented in this work for fault diagnosis of turbofan jet engine and propulsion systems with engine parameter monitoring systems onboard.

PROBLEM DESCRIPTION

Clustering is a Machine Learning Technique that involves the grouping of data points. In this paper data points are data got from the engine monitoring system, monitoring the different working parameters of the engine. Given a set of data points, we can use clustering algorithms such as C-means, K-means, DBSACN and Self-Organizing Maps to classify each data point into a specific group. In theory, data points that are in the same group should have similar properties and features (in the case of self-organizing maps), while data points in different groups should have highly dissimilar properties and features. This explanation is the most important in this work as it clearly defines that the data from healthy engines will be similar and have features distinctly unique to them. While data from engine with the degraded engine performance will exhibit highly dissimilar features when compared to those from a healthy engine. Clustering is a method of unsupervised learning and is a common technique for statistical data analysis.

Clustering analysis in this work should allow us to gain a valuable insight from the input data by presenting a visual representation of the dynamic processes occurring in the object under investigation. Data points should fall into groups after clustering algorithm is initiated. It should be noted that a clustering algorithm must meet some requirements which are listed below.

The main requirements that a clustering algorithm should meet are:

- scalability;
- dealing with different types of attributes;
- discovering clusters with arbitrary shape;
- minimal requirement for domain knowledge to determine input parameters;
- ability to deal with noise and outliers.

The significance of the clustering algorithm is to extract value from large quantities of structural and unstructured data. It allows us to segregate the data based on their properties/features and groups them into different clusters depending on their similarities.

FUZZY C-MEANS CLUSTERING

Fuzzy C-means clustering was principally introduced by J.C. Dunn and improved upon by J.C Bezdek [10, 11], and has been built upon for various applications. Fuzzy c-means is a partitive clustering approach where a given data set is divided into K clusters in a way such that each data sample belongs to one of the clusters to some degree. In this paper fuzzy c-means is used to cluster flight parameter data in a way that the data clusters formed are representative of the nominal and fault values of the selected parameter.

In this paper four flight data such as: EGT (exhaust gas temperature), N1 (rotor speed), N2 (rotor speed), and OT (oil-temperature) were used.

The algorithm divides the available data into spherical clouds of data samples in a p-dimensional space. Each cluster is then in its turn represented by its cluster center. The Euclidian distance is used to measure the distance between the cluster centre and all the points that fall within the cluster boundary. General objective of c-means is to obtain a partitioning of a given data set which minimizes an objective function for a fixed number of clusters. Represented by *J* below:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad (1)$$

where m – is any real number greater than 1

u_{ij} – is the degree of membership of x_j in the cluster j

x_j – it the i^{th} of the d-dimensional measured data

c_j – is the d-dimension center of the cluster

$\| \ \|$ – is any norm expressing the similarity between measured data and the center.

Fuzzy partitioning begins through an iterative optimization of the objective function J shown above, with the update of the membership u_{ij} and the cluster center c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\| x_i - c_j \|}{\| x_i - c_k \|} \right)^{\frac{2}{m-1}}}, \quad (2)$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}. \quad (3)$$

This iteration will stop when

$$max_{ij} \left\{ u_{ij}^{(k+1)} - u_{ij}^{(k)} \right\} < \varepsilon, \quad (4)$$

where ε – is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point J_m.

INITIALIZING FUZZY C-MEANS ALGORITHM

To initialize c-means clustering for a set of data, input data is organized into a *M*N* matrix called *U*. Flight data from a CF-34/10 turbofan engine was used in the paper. Data set was organized into four parametric data groups: EGT, N1, N2 and Oil temperature. Below there are the initialization steps of the algorithm:

Matlab was used for running fuzzy c-means. Educational Version

Step 1 Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$

Step 2 At k-step: calculate the centers vector $C^{(k)} = [c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

Step 3 Update $U^{(k)}$, $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\| x_i - c_j \|}{\| x_i - c_k \|} \right)^{\frac{2}{m-1}}}$$

6WHSU$^{(k+1)}$,I–U$^{(k)}$ ¡ WKHQ 6723 RWKHUZLVH UHWXUQ WR VWH
5HVXOWV DUH SUHVHQWHG LQ WKH WHWUDKHGUDO RU GLVSODFHPHQW
WKH FHQWHU RI HDFK FOXVWHU IRUPDWLRQ
7ZR FDVHV ZHUH LQLWLDWHG DQG WKH GDWD IURP WKH VWUHVV DQDO
DWLRQ ZDV FDUULHG RXW XVLQJ QHXUDO QHWZRUN ZLWK LQSXW
JRLQJ EDWFK ZHLJKW ELDV UXOHV DQG SHUIRUPDQFH RI WKH VXE
IRU YDOLGDWLRQ RI WKH SURFHVV RI WKH VWUXFWXUH EDVHG LQ
IURP D &) HQJLQH ,Q &DVH SRZHU GHJUDGHG HQJLQH VS
VXOWV DUH SUHVHQWHG LQ)LJXUH IRU &DVH LV WKH QRUPD

)LJ 5HVXOWV IURP &DVH WHVWHU XVLQJ WKH IRXU FOXVWHUV ZL
RLO Q DQG RLO Q GHJUDGHG HQJLQH SHUIRUPDQFH

)LJ 5HVXOWV IURP &DVH WHVWHU XVLQJ WKH FOXVWHUV FRPSDFW ZLJKW V
RI GHSHQGHQFH RLO Q SHUIRUPDQFH KHDOWK\