IET 3rd International Conference on Intelligent Signal Processing (ISP 2017),
4-5 December 2017, London, UK.
DOI:10.1049/cp.2017.0349

Anomaly Detection of Aircraft Engine in FDR (Flight Data Recorder) Data

Chang-Hun Lee*, Hyo-Sang Shin*, Antonios Tsourdos*, and Zakwan Skaf †

**School of Aerospace, Transport and Manufacturing, Cranfield University, MK43 0AL, Cranfield, UK*
(email:lckdgns@gmail.com)
†Integrated Vehicle Health Monitoring Centre, Cranfield University, MK43 0AL, Cranfield, UK

Keywords: Data Analytics, Flight Data Recorder, Aircraft Engine, Maintenance Operation.

Abstract

This paper deals with detection of anomalous behaviour of aircraft engines in FDR (flight data recorder) data to improve airline maintenance operations. To this end, each FDR data that records different flight patterns is first sampled at a fixed time interval starting at the take-off phase, in order to map each FDR data into comparable data space. Next, the parameters related to the aircraft engine are only selected from the sampled FDR data. In this analysis, the feature points are chosen as the mean value of each parameter within the sampling interval. For each FDR data, the feature vector is then formed by arranging all feature points. The proposed method compares the feature vectors of all FDR data and detects an FDR data in which the abnormal behaviour of the aircraft engine is recorded. The clustering algorithm called DBSCAN (density-based spatial clustering of applications with noise) is applied for this purpose. In this paper, the proposed method is tested using realistic FDR data provided by NASA's open database. The results indicate that the proposed method can be used to automatically identify an FDR data in which the abnormal behaviour of the aircraft engine is recorded from a large amount of FDR data. Accordingly, it can be utilized for a high-level diagnosis of engine failure in airline maintenance operations.

1 Introduction

Recently, data analytics has received a great deal of attention in various industries because it can provide new information and value from data. In particular, the importance of data analytics in the aerospace industry is increasing [1]. In the aerospace industry, modern aircraft are typically equipped digital flight data recorders (FDRs) [2] and generate vast amounts of data in flight, such as altitude, pressure, acceleration, engine pressure, engine temperature, and so on. In addition, various types of data are generated during aircraft maintenance operations. In recent years, there have been many efforts to utilize this data to obtain new information.

In [3-4], FDR data was used to identify abnormal flights in order to improve aircraft operational safety. In reference [5], maintenance history data was utilized to identify unusual trends in maintenance demands, in order to improve airline maintenance operations. ACMS (aircraft condition monitoring systems) data and maintenance report data were used to predict failure modes in the aircraft [6]. In reference [7], maintenance history data was used to predict required man-hours for maintenance task. Also, based on text mining approach, air safety report [8] and maintenance text message [9] were used to enrich information. In reference [10], maintenance report data was utilized to identify recurring faults.

Among various types of data, utilization of FDR data is important from airline maintenance operations point of view because it can provide information on the evidence of potential issues inside the aircraft. For this reason, most airlines have dedicated departments for analysing this data and are constantly working to utilize the data. However, in general, it takes a lot of time and effort for human to consume huge amounts of FDR data. Accordingly, an automated data analysis algorithm is needed by using data mining technique and computing resources.

Among the parameters recorded in the FDR data, monitoring of parameters related to the aircraft engine is important for airline maintenance operations. The reasons for these are as follows. First, the aircraft engine is the most expensive part of the aircraft, and replacing the aircraft engine during maintenance work usually requires a lot of time. Second, the failure of the aircraft engine is directly linked to an aircraft accident.

Based on these aspects, this paper aims to suggest an automated data analysis algorithm in order to detect an FDR data in which the abnormal behaviour of the aircraft engine is recorded. In the proposed method, since each FDR data records different flight patterns, all time series data recorded in each FDR data are first sampled at a fixed time interval starting at the specific flight event (i.e., take-off phase), allowing each FDR data to be comparative. And then, among the sampled parameters in each FDR, the parameters related to the aircraft engine are only selected down. After performing correlation analysis between the selected parameters, we identify the correlated parameters. Then, the correlated parameters are compressed into a single parameter: mean value of correlated parameters in order to reduce redundant information. The feature points are defined as the mean value of each parameter within the sampling interval. The feature vector for each FDR is determined by arranging all feature points of each parameter. In order to compare the feature vectors of all FDR data, a density-based clustering

This paper is a postprint of a paper submitted to and accepted for publication in
IET 3rd International Conference on Intelligent Signal Processing (ISP 2017) and is subject to Institution of Engineering and Technology Copyright.
The copy of record is available at the IET Digital Library

technique [12] with Euclidian distance is applied. This algorithm allows us to automatically identify an FDR data in which the abnormal behaviour of the aircraft engine is recorded from huge amounts of FDR data. By using realistic FDR data obtained from NASA's open database, the proposed data analytics is tested.

This paper is constructed as follows. In Section 2, the proposed data analytics is explained. In Section 3, data analytics results are shown. Finally, we conclude our study in Section 4.

2 The Proposed Data Analytics

2.1 Overall Process

This section explains the overall process of proposed algorithm. Fig. 1 provides the block diagram of proposed method. As shown in Fig. 1, for each FDR data, as data pre-processing, data sampling process is first conducted to match up the sampling rate of each time series data in FDR data as well as to make different FDR data become comparative. Sequentially, data selection is performed to pick out only relevant parameters from each FDR data. In the feature generation, correlation analysis using Pearson coefficient [11] is first performed in order to check informative parameters. The feature aggregation process is then carried out to compress correlated parameters into a single parameter and compute the mean value of each parameter within sampling interval. Next, the feature transformation is performed to uniformalise value ranges of feature points. Finally, a density-based clustering algorithm called DBSCAN (density-based spatial clustering of applications with noise) [12] is applied.

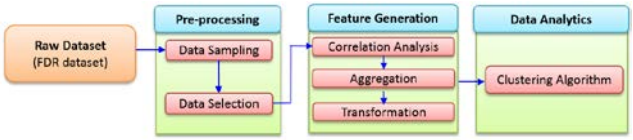


Figure 1: The overall process of proposed data analytics.

3.2 Data Pre-processing

The FDR data used in this paper is obtained from NASA's open database [13]. Each FDR data consists of 186 different flight parameters including altitude, speed, pressure, acceleration, engine speed, engine vibration, fuel flow, and so on with the purpose of condition monitoring of the aircraft.

Based on initial investigation of a raw FDR data, we identify two characteristics. First, each time series data in an FDR data has different sampling rate. Second, each FDR data contains time series data of different flight patterns because each aircraft has a different flight route. As an example, Fig. 2 shows the altitude profiles of two sample FDR data. These FDR data are flights of two different aircraft (aircraft A and aircraft B in Fig. 2). From this figure, we can readily observe that aircraft *A* and aircraft *B* have considerably different flight altitudes according to their flight routes. It implies that the time series data in different FDR data are not directly comparative.

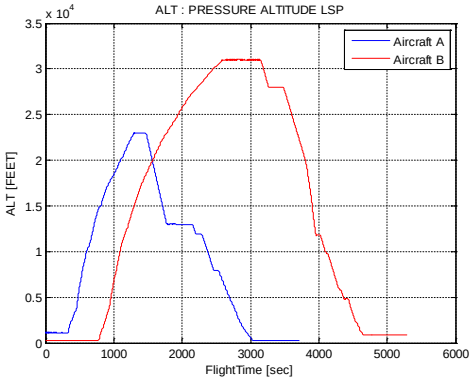


Figure 2: Altitude profiles of different flights

Therefore, before getting in to the feature generation, we first need to take data sampling process. It means each FDR data should be mapped into comparable data space. To this end, we first match up the sampling rate of each time series data in an FDR data, by using the linear interpolation method. That is, for a time series data with a slow sampling rate, we create a new data point using linear interpolation for a portion without a data point compared to other time series data with a high sampling rate.

$$x(t^*) = x(t_{k-1}) + \frac{x(t_k) - x(t_{k-1})}{t_k - t_{k-1}}(t^* - t_{k-1}) \qquad (1)$$

with $t_{k-1} < t^* < t_k$. In Eq. (1), t_{k-1} and t_k represents the time interval of a time series data. $x(t_{k-1})$ and $x(t_k)$ are the corresponding data points. t^* represents the point in time at which a new data point is desired to be obtained. $x(t^*)$ represents a new data point.

Next, after matching up the sampling rate, for all FDR data, each time series data in each FDR data should be sampled at a fixed time intervals starting from the common anchoring event, in order to make each FDR data to be comparative. In this study, we consider the take-off phase as common anchoring event of FDR data because this phase is the most important phase during the flight and most of engine failures occur at this flight phase.

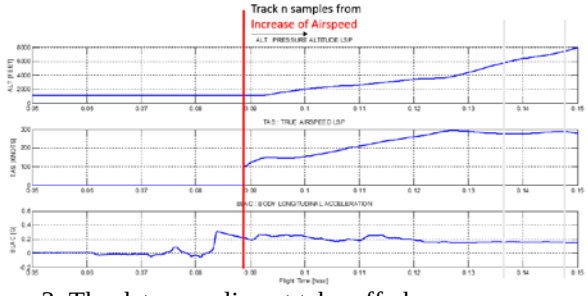
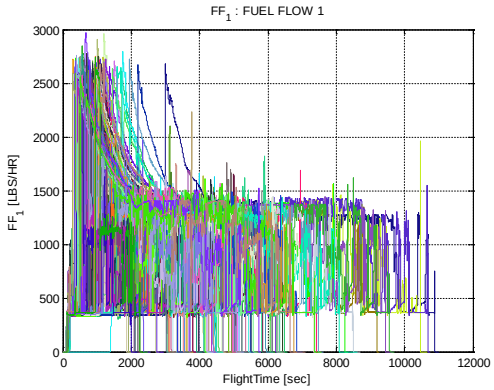


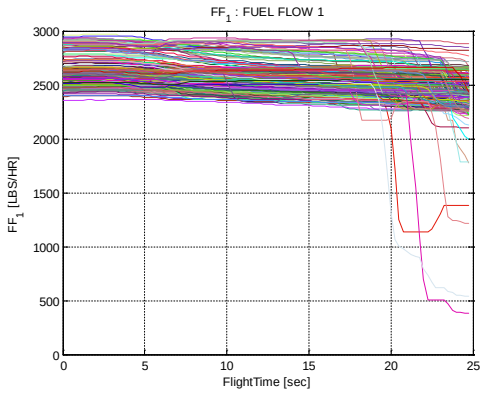
Figure 3: The data sampling at take-off phase.

During the take-off phase, the anchor is defined to be the point in time when the airspeed starts to build up shown in Fig. 3, where x-axis represents the flight time in hour and y-axis are altitude, airspeed and axial acceleration in an FDR data, respectively. In this step, for each time series data in an FDR data, sampled time series data is produced by tracking

n samples from the anchor point forwardly, where n is the desired number of samples in the time unit. As an example of data sampling, Fig. 4 (a) and 4 (b) show the original time series data and the sampled time series data related to the aircraft engine (i.e., fule flow) for all FDR data. In Fig. 4, we can readily observe that the original time series data are not comparative, however the sampled time series data are mapped into comparable data space.



(a) Original time series data



(b) Sampled time series data

Figure 4: Example of data sampling: original and sampled time series data.

In the next step, among sampled parameters in each FDR data, the parameters related with the aircraft engine are selected down. The list of the parameters selected is provided in Table 1.

No.	*Name*	*Description*
1	EGT_#1,2,3,4	Exhaust Gas Temperature
2	FF_#1,2,3,4	Fuel Flow
3	FQTY_#1,2,3,4	Fuel Quantity Tank
4	N1_#1,2,3,4	Fan Speed
5	N1C	N1 Command
6	NIT	N1 Target
7	N2_#1,2,3,4	Core Speed
8	OIP_#1,2,3,4	Oil Pressure
9	OIT_#1,2,3,4	Oil Temperature
10	PLA_#1,2,3,4	Power Lever Angle
11	VIB_#1,2,3,4	Engine Vibration

Table 1: The parameters related with aircraft engine.

3.3 Feature Generation

This section provides the feature generation of proposed data analytics. First, we perform the correlation analysis using Pearson correlation coefficient [11] in order to determine redundant parameters in Table 1. Fig. 5 provides the heat-map of correlation coefficient between the selected parameters in Table 1. As shown in Fig. 5, we can readily observe that some parameters are highly correlated each other. This is because there are four time series data corresponding four aircraft engines in EGT, FF, FQTY, N1, N2, OIP, OIT, PLA, VIB shown in Table 1, and the patterns of time series data for four engines are almost same. It implies that these parameters contain redundant information.

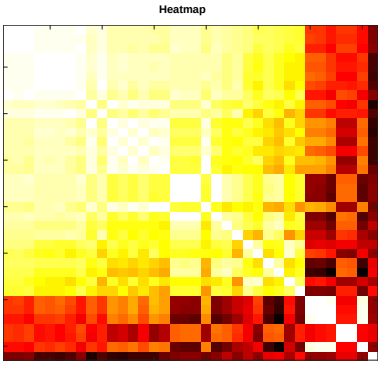


Figure 5: The heat-map of Pearson correlation coefficients between selected parameters.

Therefore, in this study, we compress the redundant parameters into a new single parameter by taking average value of them, in order to reduce redundant information. And then, we obtain 11 parameters related to the aircraft engine.

For 11 parameters, the feature points are selected as the mean value of each parameter within the sampling interval, as follows.

$$f_{i,j} = \sum_{k=1}^{n} x_{i,j}^{k} / n \qquad (2)$$

where $x_{i,j}^{k}$ represents the k-th sample point for the j-th selected parameter in the i-th FDR data. $f_{i,j}$ denotes the feature point of the j-th selected parameter in the i-th FDR data. n is the number of samples for j-th selected parameter in the i-th FDR data. For each FDR, the feature points are computed using Eq. (2).

Here, the value ranges of each feature point are considerably different because each parameter contains different physical quantities. If these feature points are directly used for data analysis, incorrect analysis results can be obtained because value ranges of each feature point are quite different so that the similarity measure between feature points are not correctly computed. Therefore, all feature points are needed to be transformed to uniformalise their value ranges using $\mu - \sigma$ standardization technique [11].

$$v_{i,j} = \frac{f_{i,j} - \mu_j}{\sigma_j} \qquad (3)$$

where $v_{i,j}$ denotes the transformed the feature point of the j-th selected parameter in the i-th FDR data. μ_j and σ_j are the mean value and the variance of the j-th feature point for all FDR data. These values are determined as follows.

$$\mu_j = \frac{\sum_{i=1}^{n_T} f_{i,j}}{n_T} \tag{4}$$

$$\sigma_j = \frac{\sum_{i=1}^{n_T} \left(f_{i,j} - \mu_j \right)^2}{n_T - 1} \tag{5}$$

where n_T is the total number of all FDR data. Then, for the i-th FDR data, the feature vector $1 \times n_j$, where n_j is the number of selected parameters (in this case $n_j = 11$), can be determined by arranging all transformed feature points shown in Eq. (3).

$$y_i = \left[v_{i,1}, v_{i,2}, \cdots, v_{i,j}, \cdots v_{i,n_j} \right] \tag{6}$$

For all FDR data, we can determine n_T feature vectors. By arranging all feature vectors, $n_T \times n_j$ numeric matrix is obtained as

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_{n_T} \end{bmatrix} = \begin{bmatrix} v_{1,1} & & v_{1,j} & & v_{1,n_j} \\ & & & & \\ v_{i,1} & \cdots & v_{i,j} & \cdots & v_{i,n_j} \\ & & & & \\ v_{n_T,1} & & v_{n_T,j} & & v_{n_T,n_j} \end{bmatrix} \tag{7}$$

By using this data matrix denoted by Y given in Eq. (7), we will identify an FDR data in which the abnormal behaviour of the aircraft engine is recorded.

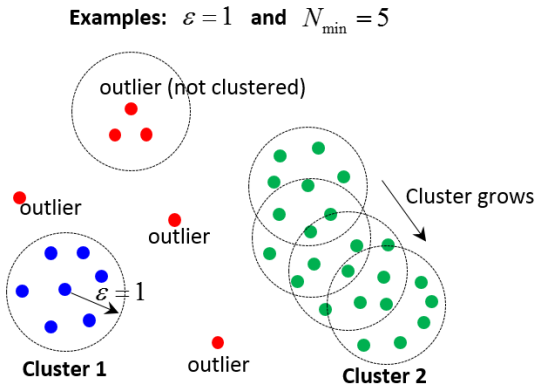


Figure 6: The illustration of DBSCAN clustering process [4].

3.4 Data Analysis Algorithm

In this study, we assume that a majority of feature vectors exhibit common patterns and a few feature vectors deviate from these common patterns. Additionally, it is assumed that these standalone feature vectors are considered as candidates of anomalies. In order to determine proximate feature vectors and standalone feature vectors, we adopt a clustering technique in conjunction with the similarity measure based on Euclidian distance. In this study, among the various options in the clustering algorithm, the density-based clustering algorithm called DBSCAN is chosen because it has the ability to identify outliers as well as cluster data points.

Fig. 6 describes the illustration of DBSCAN clustering process. The key feature of this algorithm is that clusters are identified based on the predetermined density criterion. In this algorithm, the two design parameters $N_{\min}$ and ε decide the density criterion, where $N_{\min}$ is the minimum number of points and ε is the radius of a circle. Thus, if at least $N_{\min}$ number of points are placed within ε radius of a circle, a cluster forms shown in Fig. 6. Then, a cluster expands by adding neighbourhood points which also satisfy the density criterion (see cluster 2 in Fig. 6). In this algorithm, if there are no other points to be added further, then a cluster stops expanding. A group of points that are not clustered is considered as outliers. Namely, the outliers are points that exhibit different patterns compared to other points. In the DBSCAN algorithm, according to selections of design parameters, the detection rate of outliers changes. In this study, appropriate design parameters are selected based on the sensitivity analysis, in order to attain the predetermined detection rate of outliers.

3 Data Analytics Results

In this section, the proposed method is tested with realistic FDR data obtained from NASA's open database [13]. In this study, 3,328 FDR dataset are used. In the DBSCAN algorithm, the design parameters are chosen as $N_{\min} = 5$ and $\varepsilon = 3.0$, in order to obtain 1% detection rate of outliers through sensitivity analysis.

Outlier	*Cluster #1*	*Cluster #2*	*Cluster #3*
2	*3292*	*21*	*13*

Table 2: The outputs of DBSCAN algorithm.

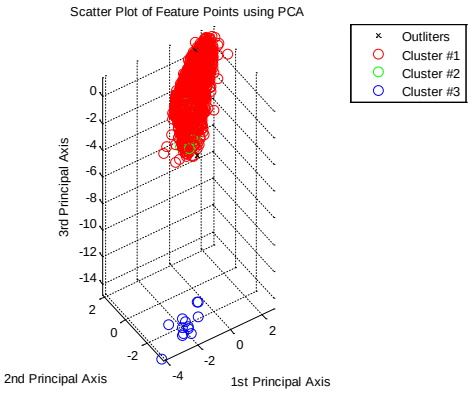


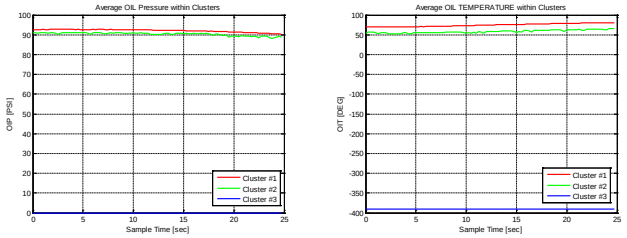
Figure 7: The scatter plot of feature points using PCA.

Table 1 provides clusters and outliers (anomalies) identified by the DBSCAN clustering algorithm. Fig. 7 shows the scatter plot of feature vectors. In this figure, since the dimension of each feature vector is 11, the first three principal

components are selected down in order to map high-dimensional feature vectors into three-dimensional space after performing the principal component analysis (PCA).

As shown in Table 1 and Fig. 7, we can readily observe that most of feature vectors belong to the Cluster #1. Therefore, the Cluster #1 can be considered as the normal behaviour of the aircraft engine in FDR data. Accordingly, a minority of feature vectors such as Cluster #2, Cluster #3, and outliers can be regarded as the abnormal behaviours of the aircraft engine.

Fig. 8 and Fig. 9 provide the average patterns of selective engine parameters in Cluster #2 and Cluster #3, respectively. As shown in Fig. 8, the readings of oil pressure and oil temperatures are abnormal. In Fig. 9, the reading of fan speed command is abnormal. These results imply that there are potential issues with sensors that measure these signals.



(a) Oil pressure (b) Oil temperature

Figure 8: The average patterns of Oil pressure and temperature in Cluster #2.

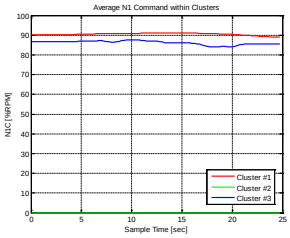
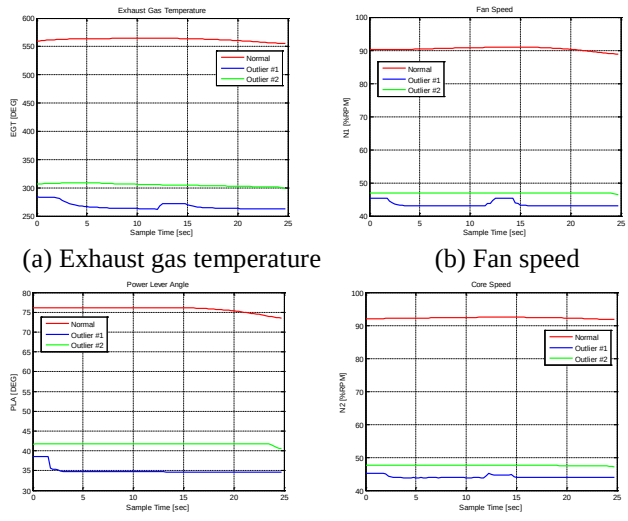


Figure 9: The average pattern of fan speed (N1) command in Cluster #3.



(a) Exhaust gas temperature (b) Fan speed

(c) Power lever angle (d) Core speed

Figure 10: Normal behaviour and abnormal behaviour of engine parameters.

Fig. 10 provides the comparison between the normal pattern and the abnormal pattern of engine parameters. In Fig. 10, we can readily observe that the values of EGT, N1, PLA, and N2 in the case of outliers are significantly lower than those in the case of normal. It implies that the aircraft engine was abnormally operated in that case.

The data analytics results indicate that the proposed method can successfully identity an FDR data in which the abnormal behaviour of the aircraft engine is recorded from a vast amount of FDR data. This algorithm can be used to relieve the effort for human to consume a huge amount of FDR data. In addition, it can be utilized for a high-level diagnosis of engine failure in airline maintenance operations.

4 Conclusion

In this paper, we propose the data analysis method to detect anomalous behaviour of aircraft engine in FDR data in order to support airline maintenance process. In the data pre-processing step, the data sampling is performed to match up the sampling rate of each parameter in each FDR data and to make each FDR data to be mapped into comparable data space. The data selection is conducted to pick out only relevant parameters in each FDR data. The feature point is chosen as a mean value of each parameter within sampling interval. Finally, through clustering technique, anomaly detection is carried out in the proposed method. By using realistic FDR data provided by NASA's open database, the proposed algorithm is tested. The numerical analysis results show that the proposed method can be used to automatically identify an FDR data in which the abnormal behaviour of the aircraft engine is recorded from vast amount of FDR data. Thus, the proposed method can be used for a high-level diagnosis of engine failure in airline maintenance operations.

References

[1] T. Larsen, "Cross-platform aviation analytics using big-data methods", *Integrated Communications, Navigation and Surveillance Conference (ICNS)*, pp.1-9, (2013).

[2] E. A. Stephenson, "Aircraft flight data recorder data acquisition system", *US Patent 4656585*, (1987).

[3] S. Das, B. L. Matthews, A. N. Srivastava, and N. C. Oza, "Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study", *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.47-56, (2010).

[4] L. Li, S. Das, R. John Hansman, R. Palacios, and A. N. Srivastava, "Analysis of flight data using clustering techniques for detecting abnormal operations", *Journal of Aerospace Information Systems*, vol. 12, no. 9, pp.587-598, (2015).

[5] A. Dubrawski and N. Sondheimer, "Techniques for early warning of systematic failures of aerospace components", *Aerospace Conference*, pp.1-9, (2011).

[6] D. L. Mack, G. Biswas, X. D. Koutsoukos, D. Mylaraswamy, and G. Hadden, "Deriving bayesian classifiers from flight data to enhance aircraft diagnosis

models," *Annual Conference of the Prognostics and Health Management Society*, (2011).
[7] A. Mathur, "Data mining of aviation data for advancing health management," *AeroSense 2002*, pp. 61–71, (2002).
[8] J. M. Pena, F. Famili, and S. Létourneau, "Data mining to detect abnormal behavior in aerospace data," *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 390–397, (2000).
[9] A. McKenzie, M. Matthews, N. Goodman, and A. Bayoumi, "Information extraction from helicopter maintenance record as a springboard for the future of maintenance text analysis," *Trends in Applied Intelligent Systems*, pp. 590–600 (2010).
[10] A. N. Srivastava and B. Zane-Ulman, "Discovering recurring anomalies in text reports regarding complex space systems," *Aerospace conference*, (2005).
[11] T. A. Runkler, *Data Analytics*, Springer, (2012).
[12] M. Ester, H.-P. Kriegel, J. Sander, X. Xu et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," *Knowledge discovery and data mining*, vol. 96, no. 34, pp. 226-231 (1996).
[13] B. Matthews, "Flight Data for Tail 653," NASA DASH link (https://c3.nasa.gov/dashlink/resources/631).

https://dspace.org/

Anomaly detection of aircraft engine in FDR (flight data recorder) data

Lee, Chang-Hun; Shin, Hyo-Sang; Tsourdos, Antonios; Skaf, Zakwan

2018-05-21
Institution of Engineering and Technology (IET)
https://doi.org/10.1049/cp.2017.0349
Lee C-H, Shin H-S, Tsourdos A, Skaf Z. (2017) Anomaly detection of aircraft engine in FDR (flight data recorder) data. IET 3rd International Conference on Intelligent Signal Processing (ISP 2017), 4-5 December 2017, London, UK

Downloaded from a DSpace repository. https://dspace.org/