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 system 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:

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

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

This iteration will stop when

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

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 $U^{(k)}$

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

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

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

Step 4 If $//U^{(k+1)} - U^{(k)}// < \varepsilon$ then STOP; otherwise return to step 2.

Results are presented in the form of dense clusters with the centroid, represented here by (1), at the center of each cluster formation.

Two cases were initiated and their results analyzed. For solid validation of results a SOM operation was carried out using neural network, this gave dimensional correlation of input data after undergoing batch weight/bias rules and performing the mean squared error operation. Results presented are for validation of the process for fault diagnostics of turbofan engines. Case 1 and 2 represent input data from a CF-34/10 engine. In Case 1, parameter registers abnormal or degraded engine performance; results are presented in Figure 1. Case 2 is the normal engine performance (fig. 2).

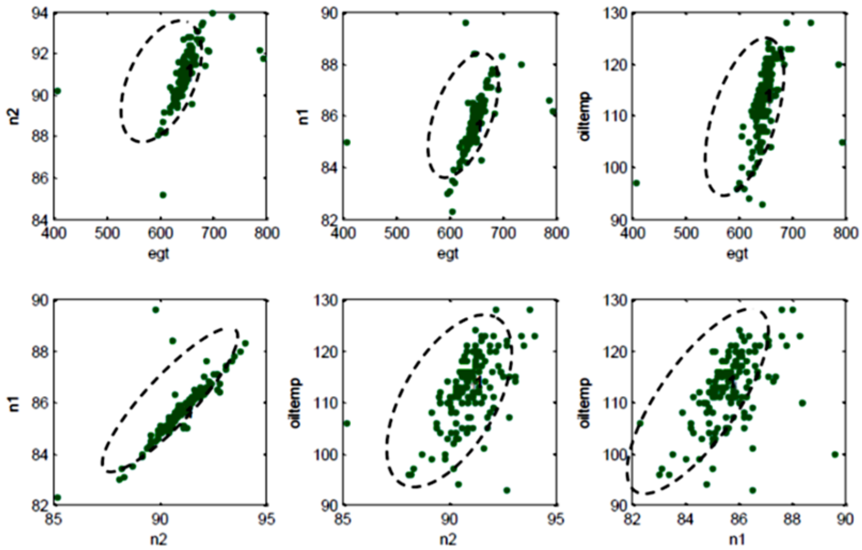

Fig. 1. Results from Case №1 engine parameters, showing clearly heavy scattering of clusters with dependence (oil/n2 and oil/n1), degraded engine performance (faulty)

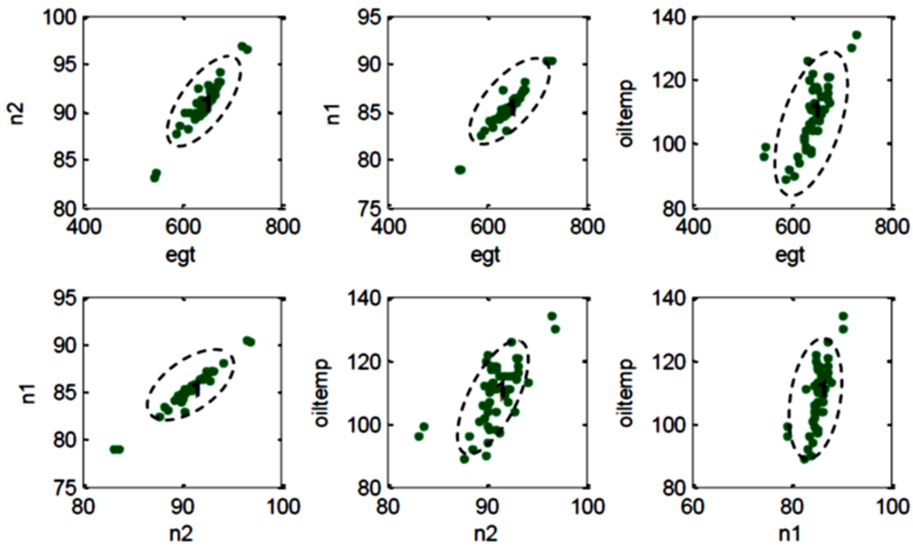

Fig. 2. Results from Case №2 engine parameters, showing clearly compact clusters but slight scattering of dependence oil/n2, performance (healthy)

The results presented in Figures 1 and 2, show that clustering method based on c-means is capable of presenting clear visual representations of input data in the form of clusters with unique patterns that distinguish a healthy engine from an engine with degraded performance. Diagnostic criteria for these results in the cluster compactness.

The SOM procedure carried out in the Neural Network shows result of Case №1 input data. After 200 iterations SOM input plane presented the following result in Figure 3. The same procedure was carried out for Case №2. The map forms a compressed representation of the inputs space, reflecting both the relative density of input vectors in that space, and a two-dimensional compressed representation of the input-space topology.

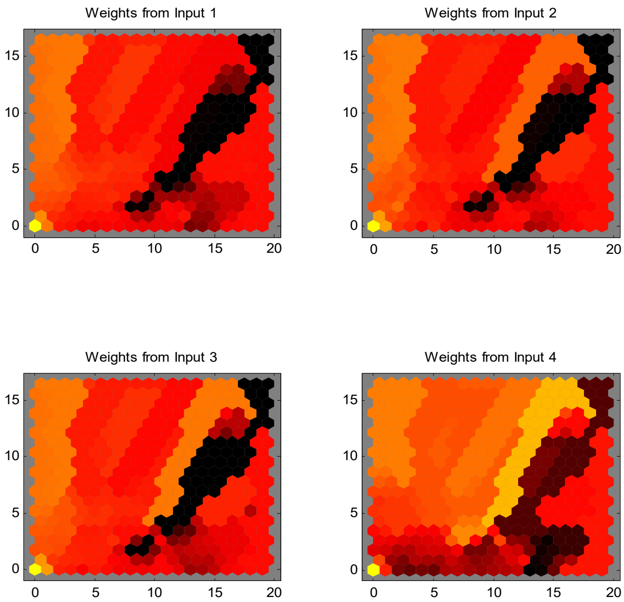

Fig. 3. Visualization of weights for Case №1 (CF34/10) with degraded engine performance (faulty)

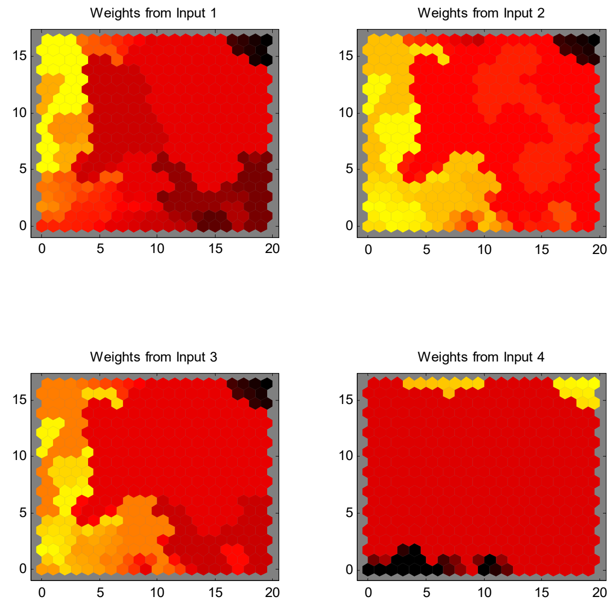

Fig. 4. Visualization of weights for Case №2 (CF34/10) healthy engine (healthy)

The Figures above can also visualize the weights themselves using the weight plane (fig. 3 and 4). There is a weight plane for each element of the input vector (four, in this case). They are the visualizations of the weights that connect each input to each of the neurons. (Lighter and darker colors represent larger and smaller weights, respectively). If the connection patterns of two inputs are very similar, you can assume that the inputs were highly correlated.

Algorithms for the SOM operation went through a weight and bias process and updates according to its learning function after each epoch (one pass through the entire set of input vectors). Training stops when any of these conditions is met: The maximum number of epochs (repetitions) is reached. Performance is minimized to the goal. The maximum amount of time is exceeded. Validation performance has increased more than max fail times since the last time it decreased (when using validation). These steps enable the SOM to classify and cluster the input data correctly.

The results attained from processing the input data from a CF34/10 turbofan engine using fuzzy c-means and self-organizing-maps algorithms, both of which are based on Neural Network, have shown that such methods can be reliably used for fault diagnostics and classifying engine performance.

CONCLUSION

1. The results presented in this paper show that for fault diagnostics of turbofan engine CF34/10, the method c-means clustering algorithm is very capable of capturing and visualizing fault signatures from parametric input data, and representing these results in the form of compact clusters (for healthy) and clusters with higher degree of scattering for engine with degraded performance (faulty).

2. Captured fault signatures were visualized in the form of simple clusters, showing dense regions to mean healthy engine and clusters with a higher degree of scattering to form engine with the degraded engine performance.

3. The reliability of the results is consistent and corresponds with results from the self-organizing-maps which show high level of input weight correlation.

4. Using c-means clustering and SOM to identify engine characteristics was achieved.

REFERENECES

1. **Buraimah, I.J.** (2015). *Primenimosti neyrosetevykh algoritm k otsenke tekhnicheskogo sostoyaniya aviatsionnykh dvigateley* [Application of the neural network algorithm for determining technical condition of aircraft engine]. Nauchnyye chteniya po aviatsii, posvyashchennyye pamyati N.Ye. Zhukovskogo: sbornik dokladov XII Vserossiyskoy nauchno-tekhnicheskoy konferentsii [12th Russian Scientific and Technical Conference, (technical reading on aviation), Dedicated to the memory of Ghukovski, N.E.], pp. 104–106. (in Russian).

2. **DePold, H.R. and Gass, F.D.** (1999). *The application of expert systems and neural networks to gas turbine prognostics and diagnostics*. ASME. Journal of Engineering for Gas Turbines and Power, vol. 121, issue 4, pp. 607–612. DOI: https://doi.org/10.1115/1.2818515

3. **Polycarpou, M.M.** (1994). *An on-line approximation approach to fault monitoring, diagnosis, and accommodation*. SAE Technical Paper 941217. DOI: https://doi.org/10.4271/941217

4. **Merrigton, G.L. (**1994). *Fault diagnosis in gas turbines using a model-based technique*. ASME Journal of Engineering for Gas Turbines and Power, vol. 116, issue 2, pp. 374–380. DOI: https://doi.org/10.1115/1.2906830

5. **Gorinevsky, D., Nwadiogbu, E. and Mylaraswany, D.** (2002). *Model based diagnostics for small-scale turbomachines.* Proceedings of the 41st IEEE Conference on Decision and Control, Las Vegas, NV, USA, vol. 4, pp. 4784–4789. DOI: 10.1109/CDC.2002.1185136

6. **Buraimah, I.J.** (2014). *O primenimosti neyrosetevyye algoritm k otsenke tekhnicheskogo sostoyaniya aviatsionnykh dvigateley* [Using neural network based algorithm to determine technical condition of aircraft engines]. Aviatsiya: istoriya, sovremennost, perspektivy razvitiya: sbornik trudov IV Mezhdunarodnoy nauchno-prakticheskoy konferentsii [4th International Technical and Scientific Conference, 'Aviation: history, current technology, perspective development']. Minsk, Belarus. (in Russian)

7. **Wang, X. and Syrmos, V.L.** (2008). *Fault detection, identification and estimation in the electro-hydraulic actuator system using EKF-based multiple-model estimation*. 16th Mediterranean Conference on Control and Automation, pp. 1693–1698. DOI: 10.1109/MED.2008.4602248

8. **Luo, J., Namburu, M., Pattipati, K., Qiao, L., Kawamoto, M. and Chigusa, S.** (2003). *Model-based prognostic techniques [maintenance applications]*. Proceedings AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference, Anaheim, CA, USA, pp. 330–340. DOI: 10.1109/AUTEST.2003.1243596

9. **Ofsthum, S.C. and Wilmering, T.J.** (2004). *Model-driven development of integrated health management architectures*. IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720). Big Sky, MT, vol. 6, pp. 3692–3705. DOI: 10.1109/AERO.2004.1368185

10. **Bezdek, J.C.** (1981). *Pattern recognition with fuzzy objectives function algorithm*. Springer US, New York, 272 p. DOI: 10.1007/978-1-4757-0450-1

11. Dunn, J.C. (1973). *A fuzzy relative of the ISODATA process and its use in detecting compact, wellseparated clusters*. Cybernetics and System Journal, vol. 3, issue 3, pp. 32–57. DOI: https://doi.org/10.1080/01969727308546046

INFORMATION ABOUT THE AUTHOR

Engr Buraimah Igbafe Jubril, Principal Engineer in the Department of Engineering Space Systems, National Space Research & Development Agency (NASRDA), jubril.buraimah@nasrda.gov.ng

ИСПОЛЬЗОВАНИЕ СОВРЕМЕННЫХ МЕТОДОВ КЛАСТЕРИЗАЦИИ ДЛЯ ПАРАМЕТРИЧЕСКОЙ ДИАГНОСТИКИ НЕИСПРАВНОСТЕЙ ТУРБОВЕНТИЛЯТОРНЫХ ДВИГАТЕЛЕЙ

I.J. Buraimah[1]
[1]*Национальное агентство космических исследований и разработок*
Абуджа, Федеративная Республика Нигерия

Авиационные и аэрокосмические технологии XXI века эволюционировали и стали более сложными и техническими. Турбореактивные двигатели, а также их собратья, ракетный двигатель (жидкий/твердый) прошли через несколько конструктивных улучшений и усовершенствований в ходе их проектирования и эксплуатации. Эти технологические усовершенствования сделали двигатели очень сложными и дорогими машинами, которые нуждаются в постоянном контроле во время их рабочей фазы. По мере того как спрос и использование таких двигателей неуклонно растут как в гражданском, так и в военном секторах, становится необходимым отслеживать и прогнозировать поведение параметрических данных, генерируемых этими сложными двигателями во время их рабочих фаз. В работе для определения неисправности двигателя используются такие параметры полета, как температура выхлопных газов (EGT), частоты вращения вентиляторов двигателя (N1 и N2), расход топлива (FF), температура масла (OT), давление масла (OP), вибрация и другие. Все турбовентиляторные двигатели проходят через несколько отчетливо различающихся рабочих фаз: взлета, круиза и посадки. Запись генерируемых параметрических данных на этих различных этапах приводит к огромному количеству бортовых данных и отчетов о техническом обслуживании, что делает задачу проектирования и разработки системы диагностики неисправностей чрезвычайно перспективной. Становится необходимым использовать современные методы анализа данных, позволяющих обрабатывать большие объемы генерируемых данных и давать четкие визуальные результаты для определения технического состояния двигателя, являющегося объектом исследования/мониторинга. Эти современные методики должны быть способны дать четкую и объективную оценку исследуемому объекту. Для построения кластеров были использованы методы кластерного анализа, основанные на нейронных сетях таких, как c-means, k-means, самоорганизующиеся карты и алгоритм DBSCAN. Различия в кластерных группировках/паттернах между исправным двигателем и двигателем с пониженной производительностью сравниваются и используются в качестве основы для определения неисправностей. Диагностика неисправностей играет очень важную роль в управлении авиационными двигателями. Своевременное и точное обнаружение неисправностей является основой, на которой базируются сроки выполнения технического обслуживания, эксплуатационные расходы и безопасность полетов. Данные, использованные в работе для анализа, были получены с бортового самописца в течение одного полетного цикла. Окончательное решение о неисправности принимает инженер.

Ключевые слова: диагностика неисправностей двигателей, параметрические данные, турбовентиляторные реактивные двигатели, мониторинг, обработка полетных данных, нейронная сеть, c-means, k-means, DBSCAN, кластерный анализ, кластерный паттерн, методы кластеризации, алгоритм, параметры полета, температура выхлопных газов, анализ данных, самоорганизующиеся карты.

СПИСОК ЛИТЕРАТУРЫ

1. **Бураймах И.Д.** Применимости нейросетевых алгоритм к оценке технического состояния авиационных двигателей // Научные чтения по авиации, посвященные памяти Н.Е. Жуков-

ского: сборник докладов XII Всероссийской научно-технической конференции. Москва, 16–17 апреля 2015 г. М.: ООО «Экспериментальная мастерская НаукаСофт». С. 104–106.

2. **DePold H.R., Gass F.D.** The application of expert systems and neural networks to gas turbine prognostics and diagnostics // ASME. Journal of Engineering for Gas Turbines and Power. 1999. Vol. 121, iss. 4. Pp. 607–612. DOI: https://doi.org/10.1115/1.2818515

3. **Polycarpou M.M.** An on-line approximation approach to fault monitoring, diagnosis, and accommodation // SAE Technical Paper 941217. 1994. DOI: https://doi.org/10.4271/941217

4. **Merrigton G.L.** Fault diagnosis in gas turbines using a model-based technique // ASME. Journal of Engineering for Gas Turbines and Power. 1994. Vol. 116, iss. 2. Pp. 374–380. DOI: https://doi.org/10.1115/1.2906830

5. **Gorinevsky D., Nwadiogbu E., Mylaraswany D.** Model based diagnostics for small-scale turbomachines // Proceedings of the 41st IEEE Conference on Decision and Control. NV, USA, Las Vegas, 2002. Vol. 4. Pp. 4784–4789. DOI: 10.1109/CDC.2002.1185136

6. **Бураймах И.Д.** О применимости нейросетевых алгоритмов к оценке технического состояния авиационных двигателей // Авиация: история, современность, перспективы развития: сборник трудов IV международной научно-практической конференции. Минск, 22 мая 2013 г. Минск, 2014. С. 87–88.

7. **Wang X., Syrmos V.** Fault detection, identification and estimation in the electro-hydraulic actuator system using EKF-based multiple-model estimation // 16th Mediterranean Conference on Control and Automation, 2008. Pp. 1693–1698. DOI: 10.1109/MED.2008.4602248

8. **Luo J.** Model-based prognostic techniques [maintenance applications] / J. Luo, M. Namburu, K. Pattipati, L. Qiao, M. Kawamoto, S. Chigusa // Proceedings AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference. Anaheim, CA, USA, 2003. Pp. 330–340. DOI: 10.1109/AUTEST.2003.1243596

9. **Ofsthum S.C., Wilmering T.J.** Model-driven development of integrated health management architectures // IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720). Big Sky. MT, 2004. Vol. 6. Pp. 3692–3705. DOI: 10.1109/AERO.2004.1368185

10. **Bezdek J.C.** Pattern recognition with fuzzy objectives function algorithm. New York: Springer US, 1981. 272 p. DOI: 10.1007/978-1-4757-0450-1

11. **Dunn J.C.** A fuzzy relative of the ISODATA process and its use in detecting compact, well separated clusters // Cybernetics and System Journal. 1973. Vol. 3, iss. 3. Pp. 32–57. DOI: https://doi.org/10.1080/01969727308546046

СВЕДЕНИЯ ОБ АВТОРЕ

Бураймах Игбафе Джубрил, главный инженер отдела инженерных космических систем, Национальное агентство Космических исследований и разработок. Абуджа, Федеративная Республика Нигерия, jubril.buraimah@nasrda.gov.ng

Поступила в редакцию	27.07.2020	Received	27.07.2020
Принята в печать	26.11.2020	Accepted for publication	26.11.2020