UDC 004.052.32:629.735.083.02(045)

S. DMITRIEV, O. POPOV, O. YAKUSHENKO, V. POTAPOV, O. PASHCHUK

National Aviation University, Kiev, Ukraine

GAS TURBINE ENGINES DIAGNOSING USING THE METHODS OF PATTERN RECOGNITION

The paper is dedicated to the relevant problem that pertains to gas turbine engines diagnosing. The issue considered in the paper is how to diagnose gas turbine engines using the methods of pattern recognition: in particular the method of "binary tree" and the "nearest neighbor" method. In computer science, a binary tree is a tree data structure in which each node has at most two children, which are referred to as the left child and the right child. A recursive definition using just set theory notions is that a (non-empty) binary tree is a triple (L, S, R), where L and R are binary trees or the empty set and S is a singleton set. Some authors allow the binary tree to be the empty set as well. I n computing, binary trees are seldom used solely for their structure. Much more typical is to define a labeling function on the nodes, which associates some value to each node. Nearest neighbor search (NNS), as a form of proximity search, is the optimization problem of finding the point in a given set that is closest (or most similar) to a given point. Closeness is typically expressed in terms of a dissimilarity function: the less similar the objects, the larger the function values. The initial data are provided and specific results are obtained.

Keywords: *diagnosing, gas turbine engines, binary tree, pattern recognition, nearest neighbor, classification, measure of distance.*

Analysis of research and publications

The paper has been included into a series of publications regarding the use of classification methods for a gas turbine engine (GTE) diagnosing [1, 2]. The research activities that this paper reviews deal with the GTEs diagnosing using the methods of "binary tree"[3] and the "nearest neighbor" method [4]. Reference [1] describes the algorithm of a numerical experiment in order to obtain the original sample which allows classifying conditions of gas turbine engines.

The mathematical formulation of the problem

As a result of the classification by the functioning parameters, a diagnosed object has to be assigned to a particular class of engine conditions. The term "class" should be understood as a group of objects that are characterized by a set of common properties. When solving the problems of diagnosing, examples of the classes of engine conditions can be given as "operative GTE", "non-operative GTE" or "GTE with the degraded flow path", etc.

To carry out the classification using both of the methods considered in the paper, a sample including the functioning parameters for all of the classes is required [1].

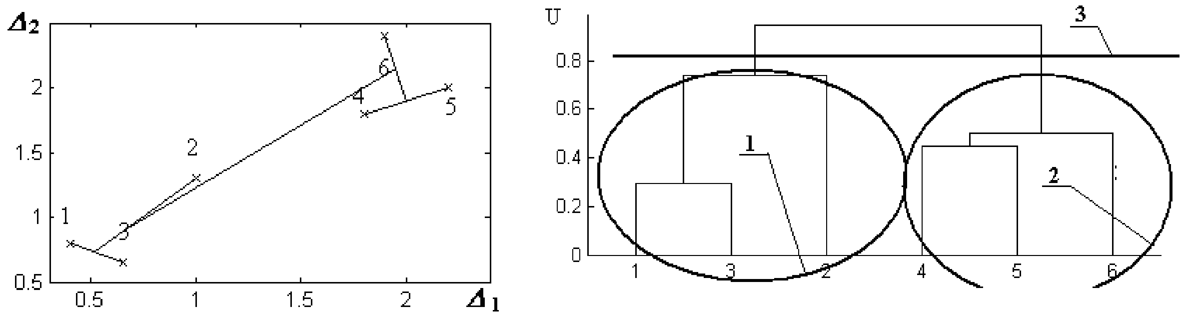
The classification procedure based on the method of "binary tree" is as follows [3]:

- to find the distance between all objects in the unclassified sample;
- to group the objects in a binary hierarchical cluster tree by means of connecting the closest pair of objects. Thus, the objects are joined together forming double clusters; formed clusters are grouped into large clusters, until a single hierarchical tree is obtained;
- to divide the hierarchical tree into clusters.

Figure 1 is an example that illustrates the approach mentioned above. This figure shows 6 objects characterized with two parameters Δ_1 and Δ_2 (see Fig. 1, a). Pertaining to a GTE, the characteristics of its conditions can be, for instance, the deviation of the parameter from its model value in standard atmospheric conditions. The lines in the figure connect the nearest objects. Figure 1, b shows the resulting dendrogram ("binary tree"). The horizontal axis in the dendrogram is indicated with the point numbers. U parameter specifies the Euclidean distance between the objects.

Having been created, the hierarchical tree should be divided into clusters. It is necessary to determine the number of classes or maximum distance between objects in a cluster. If the distance equal to 0.8 is selected as the maximum one between objects in a cluster in the example given above, two clusters are formed (see Fig. 1, b).

The final stage of recognition requires verifying the fact that the resulting clusters combine having significant similarities objects. When this analysis for the



a b

Fig. 1. Clustering a set of two-dimensional data by constructing a "binary tree":
a – the initial data set and the shortest distance between the objects; b – the dendrogram divided into 2 clusters;
1, 2 – numbers of the clusters; 3 – the dividing line

gas turbine engine conducted, it is necessary to obtain additional information about the real conditions of objects constituting the data set.

The "nearest neighbor" method [4] is as follows: when a large representative sample of data, which adequately describes object classes is available, a single point from this sample closest to a classified object may be sufficient for definition of the diagnosed object class.

This method differs from the method of "binary tree" and the specifics of this difference are that the distance is calculated including the points only, but not both the points and clusters, and the training sample is classified in advance.

The procedure for such classification is as follows:

- to calculate the distance between a diagnosed point and all points in the set;
- to define a diagnosis by a class which includes a closest to diagnosed object point.

Figure 2 shows the classification diagram that has been constructed in accordance with the described method for the case of two classes which characterized with two parameters Δ1 and Δ2.

Initial data for classification

As seen from the description, the both methods use a sample of data (classified or unclassified sample) consists of parameters Δ characterizing the GTE conditions, and information about the class of each object in this sample. Both the methods are based on the calculation of the distances between objects. Reference [1] describes the basic algorithm of the numerical experiment conducted to obtain such a sample. To use parameters that are obtained in the numerical simulation in the described methods, these parameters ought to be further processed.

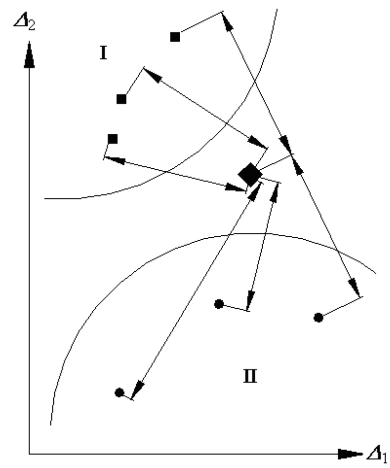


Fig. 2. The classification of GTE conditions using the "nearest neighbor" method:
■ – classified points assigned to the first class of engine conditions; ● – classified points assigned to the second class of engine conditions; ◆ – diagnosed point;
I, II – areas, points of which belong to the 1st and 2nd class of engine conditions

The need for an additional processing takes place due to the fact that the classification is performed using the selected metrics of distances between separate points of n-dimensional space of the GTE conditions indices and ranges of parameters are quite different. Thus, the formation of the distance between the points is significantly affected by the index which has the greatest difference between the maximal value Δmax and minimal value Δmin $(\Delta_{max} - \Delta_{min})_{max}$.

The index having the minimal value of this difference practically does not take part in the formation of clusters. To ensure an adequate contribution of i-index to the classification process, the value of this index should be multiplied by the coefficient, which is shown in the following equation:

$$k_i = \frac{(\Delta_{max} - \Delta_{min})_{max}}{\Delta_{max\,i} - \Delta_{min\,i}}. \qquad (1)$$

To prevent the loss of significant numbers or obtaining ultrahigh/ultralow intermediate values, there may be a necessity to use coefficient K generalized for all the indices of GTE conditions. Then, the value of i-index can be calculated in j-example as follows:

$$\Delta_{i,j}^{'} = K\, k_i\, \Delta_{i,j}\,, \qquad (2)$$

where $\Delta_{i,j}$ is the original value of the index of GTE conditions.

Obtaining the training sample for classification

In a numerical experiment [1] data containing the 6 classes of conditions of a two-shaft gas turbine engines with mixed flows was used. Table 1 shows the values of Δ_{max}, Δ_{min} и $\Delta_{max} - \Delta_{min}$ for each of the parameters that are included in the training sample.

Distance metrics between objects

One of the main issues of the proposed method is the selection of a metric to be applied to the distance between objects. The most commonly used metrics are shown in Table 2 (Δ_r, Δ_s - vectors of the coordinates of objects *R* and *S*, sign ' denotes transposition). Metric of distance selecting depends on the characteristics of a data set. It is recommended to select the metric, which provides the lowest level of errors occurring while diagnosing. The overall percentage of such errors can be accepted as a selection criterion.

The results obtained after methods using

The resulting "binary tree" obtained after processing the data (Figure 3) is shown in Figure 4. The Euclidean distance is used as a distance metric.

Table 1

Range of variations of the initial indices of GTE conditions in the training data set

GTE condition index Δ provided by	Δ_{min}	Δ_{max}	$\Delta_{max} - \Delta_{min}$
- the rotational speed of the high pressure rotor Δn_{hp}	-0.0050	0.0207	0.0258
- the total pressure in the external duct ΔP_f	-0.0098	0.0360	0.0458
- the total pressure of the compressor ΔP_c	-0.0528	0.0929	0.1457
- the total temperature of the compressor ΔT_c	-0.0401	0.0187	0.0588
- the total temperature of the turbine ΔT_t	-0.1441	-0.0125	0.1316
- the relative pressure of the turbine $\Delta \overline{P_t}$	-0.0123	0.0363	0.0486
- fuel consumption ΔG_f	-0.1894	0.0118	0.2012
$(\Delta_{max} - \Delta_{min})_{max}$			0.2012

Sample parameters $\Delta_{i,j}^{'}$ obtained with K = 1000 is shown in Fig. 2.

Table 2

The most commonly used metrics of distance between objects

Measure of distance	Calculation	Notes
Euclidean	$\Delta_{rs}^2 = (\Delta_r - \Delta_s)(\Delta_r - \Delta_s)^{'}$,	
Normalized Euclidean	$\Delta_{rs}^2 = (x_r - x_s)D^{-1}(x_r - x_s)^{'}$	D – diagonal matrix constructed using dispersions of the respective coordinate components that are calculated on all objects of the training sample
Mahalanobis	$\Delta_{rs}^2 = (x_r - x_s)V^{-1}(x_r - x_s)^{'}$	V – a simple covariance matrix
Total absolute deviation	$\Delta_{rs} = \sum_{j=1}^{n} \lvert x_{rj} - x_{sj} \rvert$	
Minkowski	$\Delta_{rs} = \left\{ \sum_{j=1}^{n} \lvert x_{rj} - x_{sj} \rvert^{P} \right\}^{1/p}$	P – parameter in the range from 1 to 2

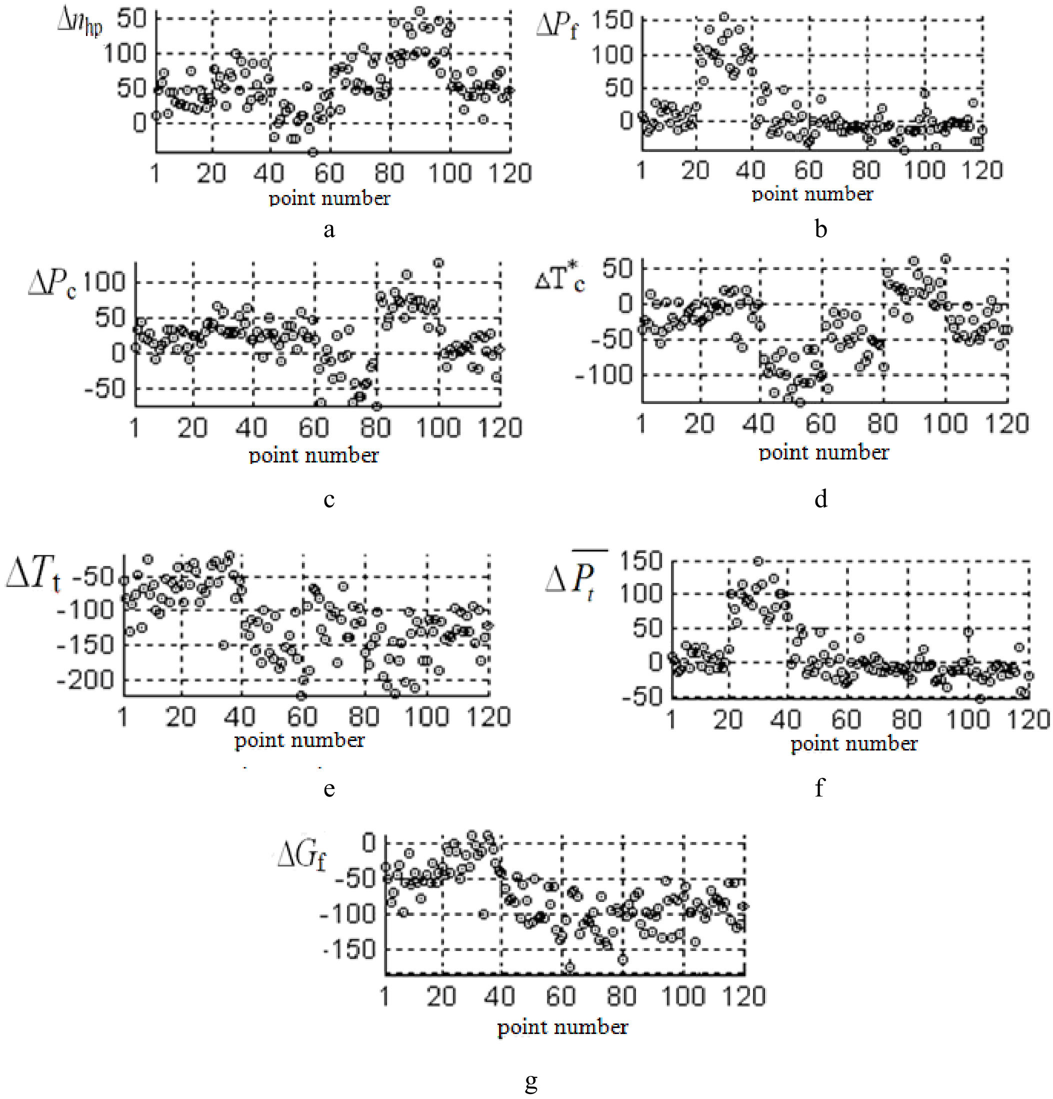


Fig. 3. The values of indices of GTE conditions after scaling by dependence (2).
Each of the six GTE condition classes represented by twenty points in the sample

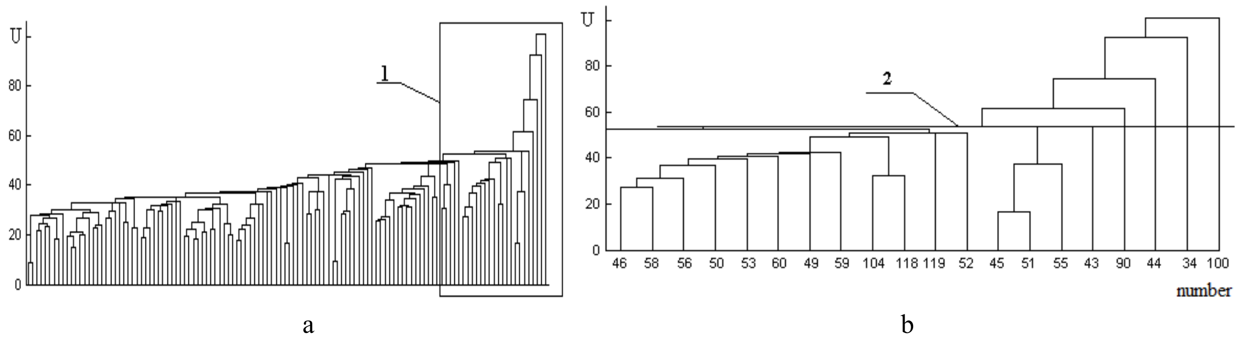


Fig. 4. Dendrogram obtained by means of clustering:
a – the general dendrogram; b – an enlarged view of area of the general dendrogram

Separation of the "binary tree" into 6 classes shows almost zero effectiveness of this method. At Figure 4b line 2 that divides the dendrogram into 6 classes are shown. Moreover, almost all the points are assigned to the same class and only 8 points are included into the remaining five classes. The reason for this is that GTE condition classes are not strictly separated in the sample, but on the contrary, these classes have the intersection areas. Change of distance between objects metric have not led to a significant improvement in the results.

The results of the "nearest neighbor" method using are shown in Fig. 5.

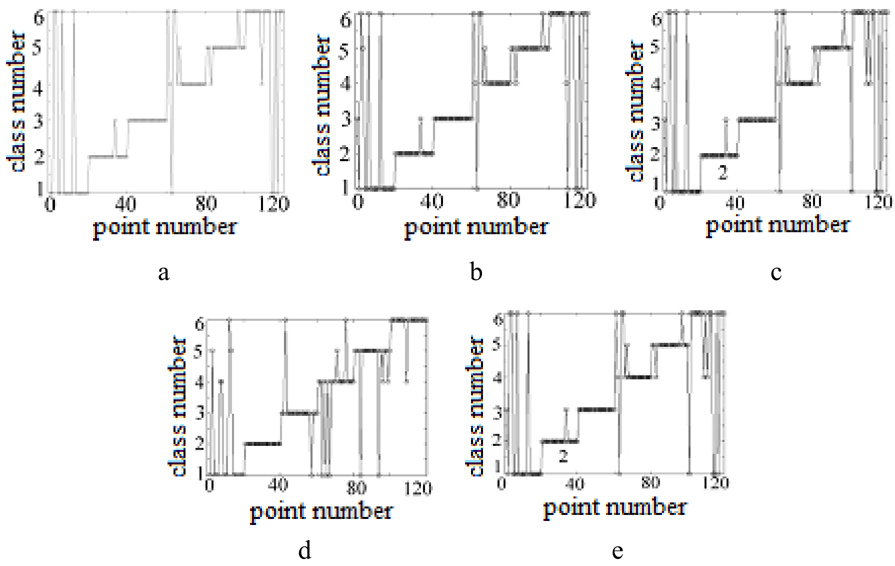


Fig. 5. The classification results of GTE conditions on the basis of the "nearest neighbor" method using different distance metrics:
a – Euclidean; b – normalized Euclidean; c – the total absolute deviation;
d – Mahalanobis; e – Minkowski (P = 1.5)

As seen from Figure 5, the best metrics for classification are: Euclidean distance metric and normalized Euclidean distance metric. They provide the lowest percentage of classification errors (about 13 %).

Conclusions

The disadvantage of the “binary tree” method is that this method does not work if there are overlapping classes at least at 1 or 2 points. This method can only be used with a clear separation of classes. To provide more stable and reliable methods of GTE conditions classification, it is expedient to use the "nearest neighbor" method which is less sensitive to the availability of overlapping classes. To improve the quality of classification The classification by some "neighbors" can be applied.

References (GOST 7.1:2006)

1. Method of formulating input parameters of neural network for diagnosing gas-turbine engines [Text] / M. Kulyk, S. Dmitriev, O. Popov at al. // Aviation : Press Technika. – Vilnius, Lithuania. – 2013. – Vol. 17, no. 2. – P. 52–56.

2. Yakushenko, O. Neural network for diagnosing GTE [Text] / O. Yakushenko // Bulletin of the National Aviation University Kyiv, Ukraine, 2004. – Vol. 2, no. 20. – P. 67-71

3. Potemkin, V. MATLAB System [Text] : handbook / V. Potemkin. – Moscow, Russia : Press ‘Dialog-MIFI’, 1998. – 350 p.

4. Edward, A. Patrick. Fundamentals of pattern recognition [Text] / A. Edward. – Prentice-Hall Inc., Englewood Cliffs, N.J., 1972. – 504 p.

References (BSI)

1. Kulyk, M., Dmitriev, S., Popov, O., Yakushenko, O. Method of formulating input parameters of neural network for diagnosing gas-turbine engines. *Aviation : Press Technika*, Vilnius, Lithuania, 2013, vol. 17, no. 2, pp. 52–56.

2. Yakushenko, O. Neural network for diagnosing GTE. *Bulletin of the National Aviation University,* Kyiv, Ukraine, 2004, vol. 2, no. 20, pp. 67-71.

3. Potemkin, V. *MATLAB System*. Mocow, Russia, Press Dialog-MIFI Publ., 1998. 350 p.

4. Edward, A. Patrick. *Fundamentals of pattern recognition*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1972. 504 p.

Поступила в редакцию 5.04.2017, рассмотрена на редколлегии 9.06.2017

Рецензент: д-р техн. наук, проф., т.в.о. ректора В. В. Панін, Державний університет інфраструктури та технологій, Київ.

ДИАГНОСТИРОВАНИЕ ГАЗОТУРБИННОГО ДВИГАТЕЛЯ МЕТОДАМИ РАСПОЗНАВАНИЯ ОБРАЗОВ

С. А. Дмитриев, А. В. Попов, А. С. Якушенко, В. Е. Потапов, А. В. Пащук

Статья посвящена актуальной проблеме, связанной с диагностикой газотурбинных двигателей. Вопрос, рассматриваемый в статье, заключается в том, как диагностировать газотурбинные двигатели с использованием методов распознавания образов: в частности, метод «бинарного дерева» и метод «ближайшего соседа». В информатике двоичное дерево представляет собой структуру данных дерева, в которой каждый узел имеет не более двух детей, которые называются левым дочерним и правым дочерними. Рекурсивное определение с использованием только определенных теоретических представлений состоит в том, что (непустое) двоичное дерево является тройным (L, S, R), где L и R - бинарные деревья или пустое множество, а S - одноэлементное множество. Типичным является определение функции маркировки на узлах, которая связывает некоторое значение с каждым узлом. Поиск ближайшего соседа (NNS), как форма поиска близости, является задачей оптимизации нахождения точки в заданном наборе, ближайшем (или наиболее близком) к данной точке. Близость обычно выражается через функцию несходства: чем меньше аналогичны объекты, тем больше значения функции. Исходные данные приведены и получены конкретные результаты.

Ключевые слова: диагностика, газо-турбинный двигатель, бинарное дерево, метод ближайшего соседа, мера расстояния.

ДІАГНОСТУВАННЯ ГАЗОТУРБІННИХ ДВИГУНІВ МЕТОДАМИ РОЗПІЗНАВАННЯ ОБРАЗІВ

С. О. Дмитрієв, О. В. Попов, О. С. Якушенко, В. Є. Потапов, О. В. Пащук

Стаття присвячена актуальній проблемі, пов'язаній з діагностикою газотурбінних двигунів. Питання, яке розглядається в статті, полягає в тому, як діагностувати газотурбінні двигуни з використанням методів розпізнавання образів: в зокрема, методом "бінарного дерева" та методом "найближчого сусіда". В інформатиці двоїчне дерево представляє собою структуру даних дерева, в якій кожен вузол має не більше двох дітей, які називаються левами дочірніми та правовими дочками. Рекурсивне визначення з використанням лише певних теоретичних представлень полягає в тому, що (не пусте) бінарне дерево є потрійним (L, S, R), де L і R - бінарні дерева або пустої множини, а S - одноелементна множина. Типовим є визначення функції позначення на вузлах, яка зв'язує деяку важливість з кожним вузлом. Пошук найближчого сусіда (ННС), як форма пошуку близькості, є завданням оптимізації розташування точки в заданому наборі, найближчого (або найближчого) до даної точки. Близькість звичайно виражається через функцію відмінності: чим менше аналогічні об'єкти, тим більше значення функції. Вхідні дані приведені і отримані конкретні результати.

Ключові слова: діагностика, газо-турбінний двигун, бінарне дерево, метод найближчого сусіда, міра відстані.

Дмитриев Сергей Алексеевич – д-р техн. наук, проф., зав. каф. сохранение летной годности авиационной техники, Учебно-научный аэрокосмический институт, НАУ, Киев, Украина, e-mail: sad@nau.edu.ua.

Попов Александр Викторович – канд. техн. наук, доц. каф. сохранение летной годности авиационной техники, Учебно-научный аэрокосмический институт, НАУ, Киев, Украина, e-mail: popche75@gmail.com.

Якушенко Александр Викторович – канд. техн. наук, ст. науч. сотр. каф. авиационных двигателей, Учебно-научный аэрокосмический институт, НАУ, Киев, Украина, e-mail: yyysss@mail.ru.

Потапов Валентин Евгеньевич – аспирант каф. сохранение летной годности авиационной техники, Учебно-научный аэрокосмический институт, НАУ, Киев, Украина, e-mail: valentine.potapov@gmail.com.

Пащук Александр Валерьевич – аспирант каф. сохранение летной годности авиационной техники, Учебно-научный аэрокосмический институт, НАУ, Киев, Украина.

Dmitriev Sergey – Doctor of Technical Science, Professor, Head of the Department of maintenance, Educational & Research Aerospace Institute, National Aviation University, Kyiv, Ukraine, e-mail: sad@nau.edu.ua.

Popov Olexander – Candidate of Technical Sciences, Docent of the Department of maintenance, Educational & Research Aerospace Institute, National Aviation University, Kyiv, Ukraine, e-mail: popche75@gmail.com.

Yakushenko Olexander – Candidate of Technical Sciences, Senior scientist of the Department of aircraft engines, Educational & Research Aerospace Institute, National Aviation University, Kyiv, Ukraine, e-mail: yyysss@mail.ru.

Potapov Valentin – Postgraduate at the Department of maintenance, Educational & Research Aerospace Institute, National Aviation University, Kyiv, Ukraine, e-mail: valentine.potapov@gmail.com.

Pashchuk Olexander – Postgraduate at the Department of maintenance, Educational & Research Aerospace Institute, National Aviation University, Kyiv, Ukraine, e-mail: max2360@rambler.ru.