A R C H I V E S O F M E T A L L U R G Y A N D M A T E R I A L S
Volume 57 2012 Issue 3
DOI: 10.2478/v10172-012-0091-7

Z. GŁOWACZ*, J. KOZIK*

FEATURE SELECTION OF THE ARMATURE WINDING BROKEN COILS IN SYNCHRONOUS MOTOR USING GENETIC ALGORITHM AND MAHALANOBIS DISTANCE

SELEKCJA CECH CHARAKTERYSTYCZNYCH DLA PRZERWY W UZWOJENIU SILNIKA SYNCHRONICZNEGO Z WYKORZYSTANIEM ALGORYTMU GENETYCZNEGO ORAZ ODLEGŁOŚCI MAHALANOBISA

The paper describes a procedure for automatic selection of symptoms accompanying the break in the synchronous motor armature winding coils. This procedure, called the feature selection, leads to choosing from a full set of features describing the problem, such a subset that would allow the best distinguishing between healthy and damaged states. As the features the spectra components amplitudes of the motor current signals were used. The full spectra of current signals are considered as the multidimensional feature spaces and their subspaces are tested. Particular subspaces are chosen with the aid of genetic algorithm and their goodness is tested using Mahalanobis distance measure. The algorithm searches for such a subspaces for which this distance is the greatest. The algorithm is very efficient and, as it was confirmed by research, leads to good results. The proposed technique is successfully applied in many other fields of science and technology, including medical diagnostics.
Keywords: feature selection, genetic algorithm, synchronous motor, faults diagnostics

Artykuł opisuje procedurę automatycznego wyboru symptomów towarzyszących przerwie w uzwojeniach twornika silnika synchronicznego. Procedura ta, nazywana selekcją cech, prowadzi do wyboru spośród pełnego zestawu cech opisujących dany problem takiego podzbioru, który pozwalałby na jak najlepsze odróżnienie stanu bezawaryjnego od stanu awaryjnego. Poszczególnymi cechami są amplitudy składowych widm sygnałów prądowych silnika. Spektra sygnałów prądowych są traktowane jako pełne przestrzenie cech, z których następnie wybierane są podprzestrzenie z zastosowaniem algorytmu genetycznego. Jakość każdej podprzestrzeni sprawdzana jest z użyciem miary odległości Mahalanobisa. Algorytm poszukuje takich podprzestrzeni, dla których odległość ta jest największa. Zastosowany algorytm jest bardzo wydajny i jak potwierdziły badania prowadzi do dobrych wyników. Proponowana technika jest z powodzeniem stosowana w wielu innych dziedzinach nauki i techniki, w tym w diagnostyce medycznej.

1. Introduction

Electrical machines, especially of synchronous type have a very wide application in the metallurgical industry. This is because of their main advantages – the constant angular velocity of the rotor and high efficiency. They can be found in rolling extrusion systems [1], mill stands [12], vacuum pumping [2] and also testing stands, i.e. the rotating hammer for testing the effect of deformation velocity on the properties of metal sheets [16]. In most of these applications a high reliability is expected, and any interruptions usually lead to very expensive breaks in production process. Thus the possibility of fault of every part of the process including motors needs to be reduced to a minimum. The motors similarly to the other electrical and mechanical devices, are constantly exposed to both external and internal influences which cause the changes to their physical state and functionality [14]. These gradually developing irreversible changes worsen the efficiency of the motor and thus the industrial process it is the part of, and finally can lead to complete damage to the machine. To minimize these adverse effects the state of the machine should be periodically checked allowing the maintenance works to be done before the fault develops. Rapid development of a digital measurement techniques in the last decades created new possibilities of machines condition diagnostics, which do not require any disintegration to the machine, and can be undertaken even without stopping the process it is serving. These new methods include: spectral analysis of currents [4], [5], [6], [7], [15], [18], axis flux [8], vibration and sound analysis [8], [9], thermal

* AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY, DEPARTMENT OF ELECTRICAL MACHINES, 30-059 KRAKÓW, 30 MICKIEWICZA AV., POLAND

pattern recognition. This paper describes a recognition of short-circuit faults in a synchronous motor stator windings based on analysis of stator currents and their Park's vector.

2. Measuring stand and testing procedures

As the diagnostic object a specially designed and constructed 2-pair pole synchronous machine was used. The stator of the machine consists of 48 slots with double layer windings inside (80 turns per phase). The rotor is of salient pole type with 4 poles and 145 turns per pole. All the ends of every group of windings are accessible allowing simulation of many fault conditions to be done. For the purpose of this paper a break in stator parallel branch was performed. The connection diagram of stator windings with the fault done is presented in the Fig. 1.

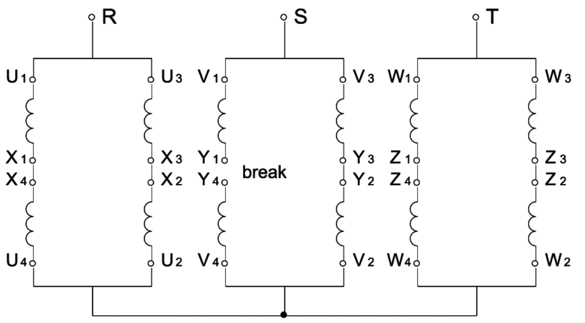

Fig. 1. Diagram of stator windings arrangement with a break fault present

The break fault in stator parallel branch is a quite common emergency state in synchronous machines. Armature coils are usually connected together with soldered joints, which can be damaged by vibration or deformations caused by currents of a high value. This leads to the break in the whole parallel branch. The machine in this state may still work, but there is increased value of current in the other parallel branches belonging to the same phase and in consequence it leads to further damage due to excessive heat generated in these circuits. The measurement was carried out while the machine was rotating with synchronous speed.

Investigated synchronous motor was mechanically loaded by separately excited DC machine as can be seen in the diagram shown in the Fig. 2. Three phase currents and the field current were recorded using PC with installed data acquisition card. Currents were converted to voltages having ranges acceptable by an acquisition card using three current transducers (LEM LTS 6-NP). To avoid an aliasing phenomena three antialiasing filters were used. In that role three Butterworth 4th order low-pass filters were adopted. The cutoff frequency was 5 kHz. Recorded signals were sampled with a frequency of 25 kHz.

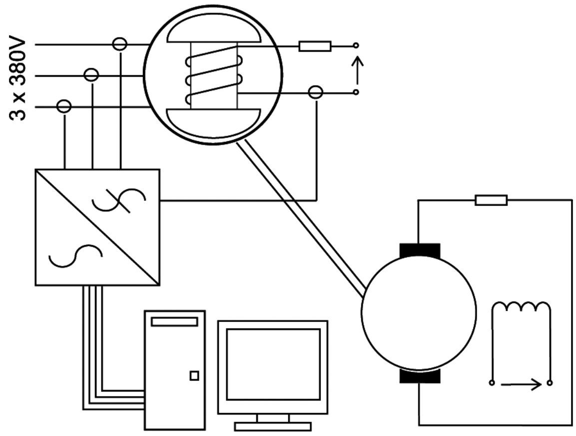

Fig. 2. Schematic diagram of the measuring stand

To reduce the impact of ripples in supply voltage on the results of analysis a great number of samples was recorded – 200 for each motor condition (break fault and healthy state) giving the overall number of 400 samples. Recordings for each condition were undertaken in 5 different days (40 recordings per condition per day). Additionally a different devices were supplied from the same power source as the investigated motor. These devices included power inverter supplying induction motor and two faulty induction motors. This caused the analysis even more resistant on the interference from the power supply.

3. Automatic feature selection

Selection of symptoms considering each harmonic separately, and selecting those of them which form separable groups for different motor states (healthy or faulty state) does not always lead to good results of later classification. There is a possibility that the remaining harmonics, which do not form a separated groups by themselves, if considered together, as the coordinates of points in spaces of higher dimensionality, will produce groups with a high degree of separability. This type of situation is illustrated in the figure 3, where two sets of features (x and y), taken individually, do not give the separated clusters, but used as coordinates of vectors in two-dimensional space, form two linearly separable groups. Therefore, it is profitable to take into account all the harmonics altogether and search for such a subset, that in the sense of assumed criterion would give the greatest separability. For such a purpose, a procedure called feature selection is used [3],[11], [13].

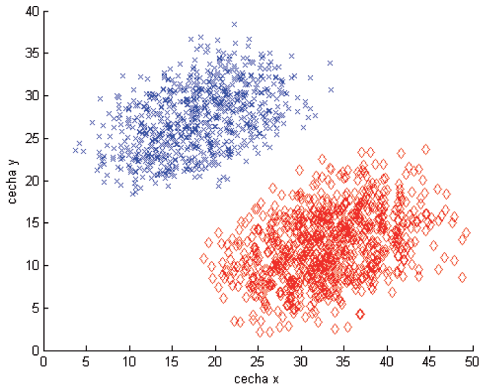

Fig. 3. Example feature vectors which form two linearly separable groups in two-dimensional space

Feature selection process consists of four basic elements [13]: 1. search procedure to generate the next candidate subset, 2. evaluation function to evaluate the subset under examination, 3. stopping criterion to decide when to stop, 4. validation procedure to check whether the subset is valid.

Feature selection algorithm starts by generating a subset of features for which the value of evaluation function is calculated. If this value does not meet the stopping criterion, another subset is generated and stopping criterion is checked again. These steps are repeated until the stopping criterion is met. Then the selection algorithm terminates and a validation of the best subset is carried out.

For selection of symptoms that are specific for investigated fault in the synchronous motor a filter feature selection technique was used. As the algorithm for searching the feature space and generating its subsets a classical genetic algorithm [10], [14], [17] was adopted. The subsets were evaluated using Mahalanobis distance between the two classes. The problem came down to search for such a feature subset for which the Mahalanobis distance would have the greatest value.

The genetic algorithm starts its operation from generating the initial population, consisting of artificial organisms (individuals) represented by the coded strings called chromosomes. All individuals from initial population are evaluated by calculating the values of their fitness functions. Then stopping criterion is checked, which may be achieving the assumed fitness or achieving the required number of iterations (generations). If the stopping criterion is met the algorithm terminates and the individual with best fitness is used as the output, otherwise the reproduction of individuals and forming the mating pool are conducted. Then, the individuals from mating pool are subjected to genetic operators: crossover and mutation leading to a forming of new chromosomes generation, which replaces the current generation. This new generation is evaluated and the stopping criterion is tested. The algorithm continues to run until the stopping condition is met. The diagram of genetic algorithm is shown in the Fig. 4.

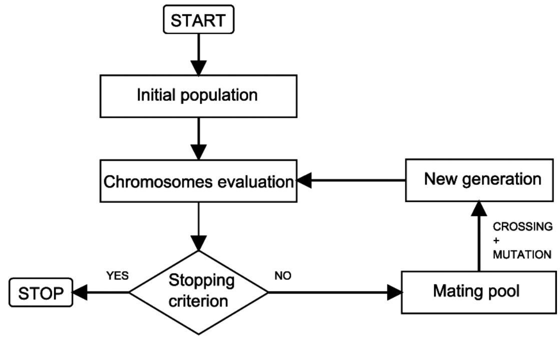

Fig. 4. Diagram of genetic algorithm operation

Particular feature vectors have been presented in the form of chromosomes, which are strings of zeros and ones (Fig. 5). Every feature (coordinate of the feature vector) was coded binary using 10 genes which gives 1024 combinations. The calculations were conducted for vectors consisting of two, three and four features giving chromosomes comprising of 20, 30 and 40 genes. An example of coding the chromosome consisting of three features is shown in the Fig. 5.

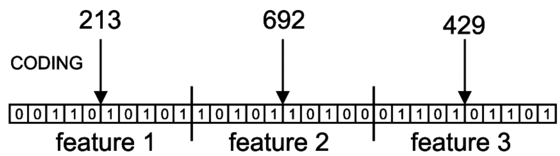

Fig. 5. Coding of three feature chromosome

The reproduction selects from the whole population of individuals these chromosomes whose values of fitness function are the highest [14]. The chosen chromosomes form the mating pool and are then subjected to crossover and mutation operations. Reproduction process is repeated several times until the offspring reaches the assumed abundance (usually it is the same number as the size of the mating pool). The most popular methods of reproduction are roulette wheel method, the ranking method and the tournament method.

In the roulette wheel method the individuals are drawn from a population with probability proportional to their fitness function. The method name is taken from its analogy to draw using the roulette wheel whose segments represent particular chromosomes, and the width of each segment is proportional to the value of chromosome fitness function. The disadvantage of this method is that individuals with very low values of fitness function may be too soon eliminated from the population, while numerously occurring individuals with medium value of fitness can dominate the population, leading to premature convergence of genetic algorithm.

In the ranking method the chromosomes in the population are sorted in descending order of fitness function values. The position in the sorted list is called the rank of the chromosome. Number of copies of each chromosome, which will be taken to the mating pool, is determined using the function, whose variable is the rank of chromosome M(r). For that purpose the most commonly are used linear functions. To the mating pool there is selected the number of first k chromosomes from the ranking list, that satisfies the condition:

$$\sum_{i=1}^{k} M_i(r_i) \leqslant n \qquad (1)$$

where:

n – the size of mating pool.

The ranking method is more resistant to premature convergence than the roulette wheel method.

In the tournament reproduction method the chromosomes in the population are divided into a certain amount of subgroups (usually two), and then in each subgroup the individual with the best fitness is chosen to the mating pool. This process is repeated until the desired number of individuals is reached. Copies of the selected chromosomes are usually returned to the tournament pool, so their re-election is possible.

The crossover is an operation in which two individuals swap the parts of their gene sequences between each other. The most classic type of crossover is a one-point crossover. In order to carry it out two individuals are drawn from the mating pool with the probability belonging to the range *[0.5,1]*. Then the crossover point l_k is randomly selected, which is a number from the range *[1, L-1]*, where L is the number of genes in the chromosome. The two offspring individuals are then created, the first of which consists of the first parent's genes from positions *1* to l_k and the second parent's genes from positions l_k to L. The second offspring is built from remaining parts of the chromosomes, that is the genes in positions *1* to l_k of the second parent and the positions l_k to L of the first parent. The mutation is an operation in which the values of individual genes are changed to opposite values (from 0 to 1 or from 1 to 0) and occurs with a probability which is usually the number from the range of *[0,0.1]*. The probability of mutation is much lower than the probability of crossover, so mutation plays a minor role in relation to crossover. In the conducted calculations the probability of mutation was reduced in successive generations (in a linear fashion). This is advantageous because in the initial phase of genetic algorithm high value of probability of mutation provides a wide variety of individuals (the dominance of better individuals is confined), and at the end of the algorithm, low probability of mutation improves the convergence.

As the evaluation function which calculates the fitness of individuals in population, and hence, the goodness of feature subsets, the Mahalanobis distance was used. It can be calculated using the following formula:

$$d^M\left(\overline{x_i}, \overline{x_j}\right) = \sqrt{\left(\overline{x_i} - \overline{x_j}\right)^T C_{ij}^{-1} \left(\overline{x_i} - \overline{x_j}\right)} \qquad (2)$$

where:

$\overline{x_i}, \overline{x_j}$ – the vectors of the means of i-th and j-th groups,

C_{ij} – the matrix of weighted covariance,

The matrix of weighted covariance is given by:

$$C_{ij} = \frac{1}{n_i n_j}\left(n_i C_i + n_j C_j\right) \qquad (3)$$

where:

n_i, n_j – the numbers of vectors belonging to i-th and j-th groups,

C_i, C_j – the covariance matrices in i-th and j-th groups.

In order to improve the performance of genetic algorithm there was used a technique called fitness function scaling. In the initial phase of genetic algorithm scaling prevents premature convergence to a local minimum, caused by the situation in which non-optimal, but the best fitted individuals begin to strongly dominate and displace other individuals. However, in the final phase of genetic algorithm scaling prevents a situation in which a population remains a considerable variety, but the average value of the adaptation does not differ much from its maximum, and average individuals receive the same number of offspring as the individuals with best fitness. In the paper the two methods of scaling are investigated: linear and exponential.

4. Results of automatic feature selection

The results of feature selection for the break in the armature circuit are presented in Tables 1 and 2. The distribution of points in two and three dimensional feature space is shown in Figs 6 and 7 for the armature current and the Figs 11 and 12 for the field current.

TABLE 1

Results of feature selection based on armature current spectrum

number of features	2	3	4
feature 1	25 Hz	25 Hz	25 Hz
feature 2	150 Hz	30 Hz	30 Hz
feature 3	–	150 Hz	50 Hz
feature 4	–	–	150 Hz
Mahalanobis distance	12.5284	14.9119	16.7851

TABLE 2
Results of feature selection based on field current spectrum

number of features	2	3	4
feature 1	50 Hz	50 Hz	50 Hz
feature 2	100 Hz	100 Hz	100 Hz
feature 3	–	200 Hz	200 Hz
feature 4	–	–	150 Hz
Mahalanobis distance	18.2472	20.0908	20.5327

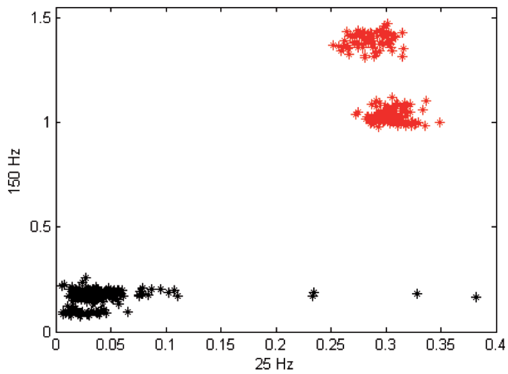

Fig. 6. Distribution of points in two dimensional feature space based on armature current spectrum (red – faulty state, black – healthy state)

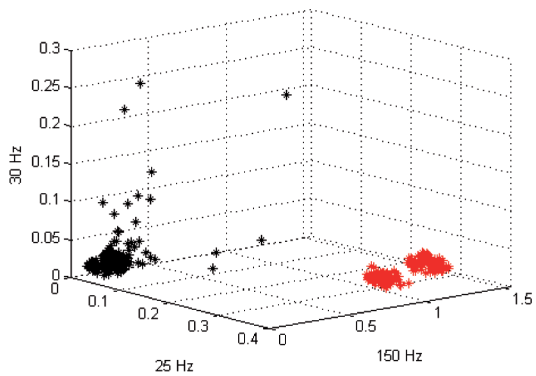

Fig. 7. Distribution of points in three dimensional feature space based on armature current spectrum (red – faulty state, black – healthy state)

Analysis of points scattering in the Fig. 6 leads to the conclusion that there are two distinctive groups of points corresponding to a healthy state and a break. It is caused mainly by a 150Hz component, but a component of 25Hz, if some of points are excluded, also creates separable clusters. The existence of points that are closer to the center of the opposite class (for 25Hz coordinate) may be due to the fact that the amplitude of 25Hz component is influenced by other factors, such as noise in the supply voltage. However, the great majority of points creates the two separable groups. Taking into account another component – 30Hz, which can be observed in the Fig. 7 does not improve the results significantly.

Figures 8-10 contain the results comparisons between different variants of genetic algorithm in the problem of selection respectively two, three and four features from the armature current spectra.

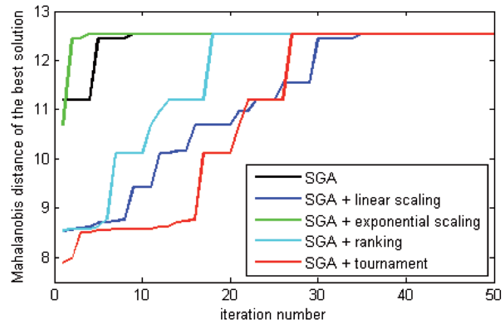

Fig. 8. Comparison of different variants of genetic algorithm in the problem of selecting two features from armature current spectrum

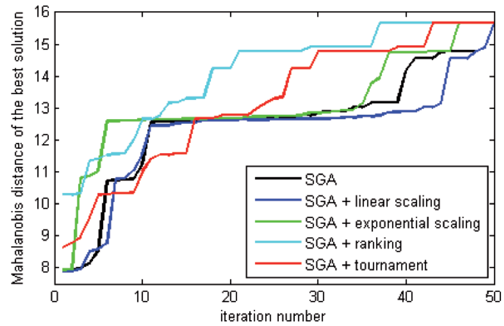

Fig. 9. Comparison of different variants of genetic algorithm in the problem of selecting three features from armature current spectrum

It can be noticed that for problems with a small number of features selected (two) algorithms based on roulette wheel reproduction more frequently lead to a best solution, however if the number of selected features is greater (three, four) algorithms with ranking or tournament reproduction have a better performance.

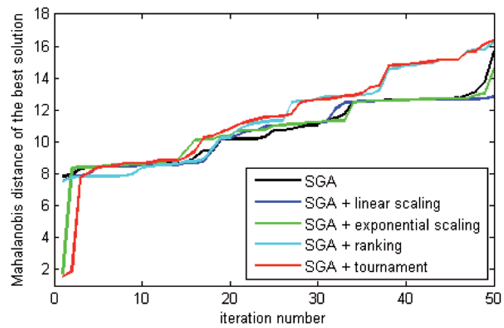

Fig. 10. Comparison of different variants of genetic algorithm in the problem of selecting four features from armature current spectrum

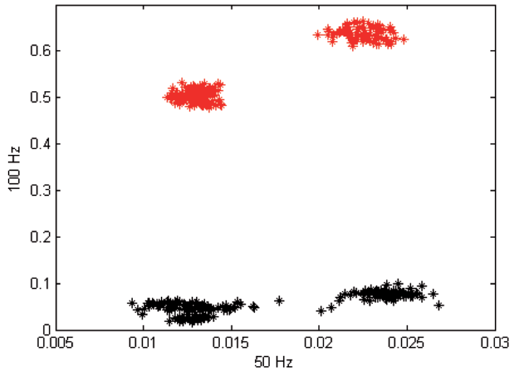

Fig. 11. Distribution of points in two dimensional feature space based on field current spectrum (red – faulty state, black – healthy state)

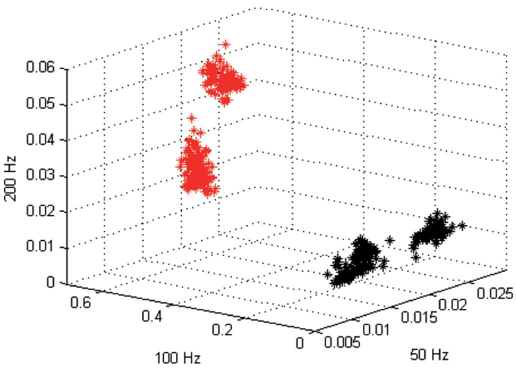

Fig. 12. Distribution of points in three dimensional feature space based on field current spectrum (red – faulty state, black – healthy state)

Figure 11 shows that the large separability of points belonging to the both classes depends mainly on the 100Hz component. The second component – 50Hz considered alone does not provide separability. A division of each group into two subgroups, which is primarily determined by the 50Hz component, suggests that other factors (not only the fault) may affect the amplitudes of these components.

Figures 13-15 contain the results comparisons between different variants of genetic algorithm in the problem of selection respectively two, three and four features from the field current spectra.

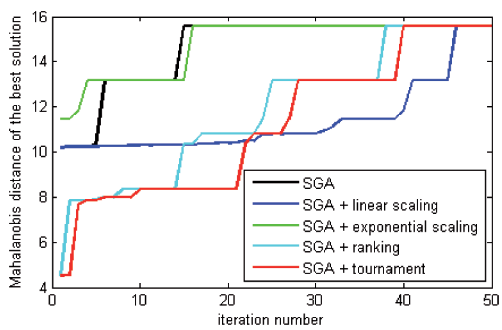

Fig. 13. Comparison of different variants of genetic algorithm in the problem of selecting two features from field current spectrum

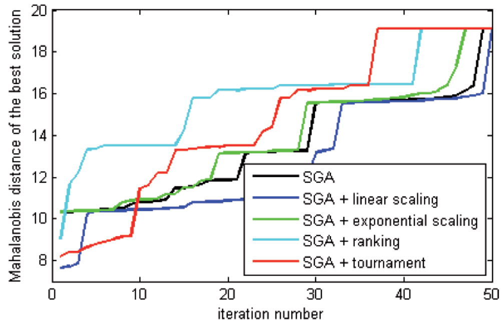

Fig. 14. Comparison of different variants of genetic algorithm in the problem of selecting three features from field current spectrum

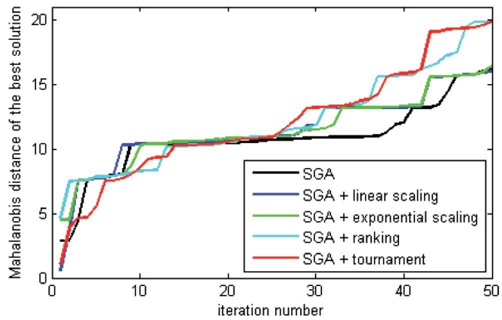

Fig. 15. Comparison of different variants of genetic algorithm in the problem of selecting four features from field current spectrum

Similarly to the problem of feature selection based on armature current in the selection of features contained in the field current the algorithms with tournament and ranking reproduction have a better performance for a greater number of selected features (three and four) and worse performance for a lower number of features (two) when compared to algorithm with reproduction based on roulette wheel method. Also algorithms with tournament and ranking reproduction had a very good convergence while algorithms based on roulette wheel method were oscillating in the vicinity of the best solution.

5. Conclusions

The results confirm that the feature selection is a very effective tool for determining the symptoms associated with faults in synchronous machines. In the case of determining the symptoms of the break in an armature parallel branch as symptoms of this damage can be considered components 150Hz and, to a lesser extent, 25Hz in the armature current and also 100Hz and 200Hz components in the field current. The results obtained during the fully automatic symptoms searching coincide with the results obtained by an expert assessment. However, the method based on automated feature selection is more effective, because it checks the combinations of multiple features, while an expert may examine the individual features alone. In addition, the automatic selection computation time is much shorter. Furthermore because the investigated feature selection algorithm searches for differences between faulty and healthy states, it functions well in the case of a data containing the noise. The comparison between different types of used algorithms shows, that the algorithm with tournament reproduction leads to the best results.

Acknowledgements

Research described in this paper was financially supported by the Polish Ministry of Science and Higher Education under the project no.: 3564/B/T02/2009/37.

REFERENCES

[1] J. Bartnicki, Z. Pater, G. Samołyk, The Distributions of Wall-Thickness of Hollowed Parts in Rolling Extrusion Process, Archives of Metallurgy and Materials **52**, 2 (2007).

[2] S. Bruce, V. Cheetham, Recent Developments and Experiences in Modular Dry Mechanical Vacuum Pumping Systems for Secondary Steel Processing Archives of Metallurgy and Materials **53**, 2 (2008).

[3] M.E. El Alami, A filter model for feature subset selection based on genetic algorithm, Journal of Knowledge-Based Systems **22**, 356–362 (2009).

[4] T. Glinka, Badania diagnostyczne maszyn elektrycznych w przemyśle, Komel, Katowice 2000.

[5] Z. Głowacz, J. Kozik, Detection of synchronous motor inter-turn faults based on spectral analysis of Park's vector. Archives of Metallurgy and Materials **56** (2011).

[6] Z. Głowacz, J. Kozik, Diagnostyka silnika synchronicznego oparta na rozpoznawaniu prądów stojana z zastosowaniem FFT i logiki rozmytej, Czasopismo Naukowo-Techniczne Mechanizacja i Automatyzacja Górnictwa (MIAG), Nr 9, Katowice 2010.

[7] Z. Głowacz, Automatic Recognition of Armature Current of DC Motor with Application of FFT and GSDM, Archives of Metallurgy and Materials **56**, 1 (2011).

[8] A. Glowacz, Z. Glowacz, Diagnostics of Induction Motor Based on Analysis of Acoustic Signals with Application of FFT and Classifier Based on Words, Archives of Metallurgy and Materials **55**, 3 (2010).

[9] A. Glowacz, W. Glowacz, Diagnostics of Synchronous Motor Based on Acoustic Signals with Application of LPCC and Nearest Mean Classifier with Cosine Distance, Archives of Metallurgy and Materials **55**, 2 (2010).

[10] D.E. Goldberg, Algorytmy genetyczne i ich zastosowania, WNT, Warszawa 2003.

[11] I. Guyon, A. Elisseeff, An Introduction to Variable and Feature Selection, Journal of Machine Learning Research **3**, 1157-1182 (2003).

[12] J. Kempken, G. Kleinschmidt, K. Schmale, U. Thiedemann, H.P. Gaines, J.T. Kopfle, Short Route – Long-Term Success, Integrated Mini-Mill Solutions by Midrex and SMS Demag, Archives of Metallurgy and Materials **53**, 2 (2008).

[13] G.H. John, R. Kohavi, K. Pfleger, Irrelevant Features and the Subset Selection Problem, Proeedings of the Eleventh International Conference on Machine Learning, 121-129 (1994).

[14] J. Korbicz, J.M. Kościelny, Z. Kowalczuk, W. Cholewa, Diagnostyka procesów. Modele. Metody sztucznej inteligencji. Zastosowania. WNT, Warszawa, 2002.

[15] J. Kozik, Zastosowanie sieci neuronowych w detekcji uszkodzeń silnika synchronicznego ze zwartym uzwojeniem wirnika, Maszyny elektryczne: Zeszyty Problemowe nr 80/2008, PEMINE: Problemy Eksploatacji Maszyn i Napędów Elektrycznych, Rytro 2008.

[16] A. Niechajowicz, A. Tobota, Dynamic Tension Testing of DP600, DP800 Steel and Al 6061 T4 Alloy Sheets by Means of Rotary Hammer, Archives of Metallurgy and Materials **54**, 3 (2009).

[17] L. Rutkowski, Metody i techniki sztucznej inteligencji, Wydawnictwo Naukowe PWN, Warszawa 2006.

[18] K. Weinreb, M. Sułowicz, Nieinwazyjna diagnostyka wewnętrznych niesymerii uzwojeń maszyny synchronicznej, Zeszyty Problemowe Maszyny Elektryczne Nr 77, 2007.

Received: 15 February 2012.