Anna TIMOFIEJCZUK

KODOWANIE CECH SYGNAŁÓW W SYSTEMIE WNIOSKOWANIA DIAGNOSTYCZNEGO

SIGNAL FEATURE ENCODING IN AN INFERENCE DIAGNOSTIC SYSTEM

W referacie opisano część badań, których celem jest zastosowanie wnioskowania diagnostycznego z uwzględnieniem kontekstu. Zgodnie z opracowanym podejściem konteksty działania obiektu mogą być identyfikowane z zastosowaniem algorytmów ewolucyjnych. Jednym z najistotniejszych czynników zastosowania tych algorytmów jest odpowiednie przygotowanie danych wejściowych. W referacie opisano opracowane sposoby kodowania dla trzech rodzajów cech sygnałów. Przykłady zastosowania opisanych sposobów kodowania zastosowano dla sygnałów zarejestrowanych podczas działania stanowiska modelującego działanie maszyny wirnikowej.

Słowa kluczowe: *kodowanie, algorytm genetyczny, analiza sygnałów.*

In the paper a part of research focused on context based diagnostic inference has been presented. According to the elaborated approach contexts of machinery operation can be identified with the use of evolutionary algorithms. One of most important factors of the application of such algorithms is a proper form of input data. In the paper encoding of three different types of signal features have been discussed. Examples of elaborated notations have been applied to signals recorded during operation of a model of rotating machinery.

Keywords: *encoding, genetic algorithm, signal analysis.*

1. Introduction

Data acquired during observation of a technical object (e.g. vibration, noise) can be analyzed by means of different methods. As results one usually obtains huge sets of features. To extract the knowledge enclosed in the features different approaches are possible to be applied. Examples are evolutionary algorithms. In this case results of vibration analysis have to be encoded in forms of individuals.

In the paper a part of research focused on context based diagnostic inference has been presented. According to the elaborated approach contexts of machinery operation (e.g. particular properties, states or conditions) can be identified with the use of evolutionary algorithms.

It is important that results of analysis can be gathered in form of different functions. They are often not directly comparable. In the paper three types of features have been taken into account. They are respectively time functions (e.g. rms or mean values estimated in time), trajectories representing displacements of the centre of a shaft and time-frequency characteristics (e.g. spectrograms or scalograms). All of them are represented by means of series of elaborated approaches and they are considered and encoded within given time periods. The main assumption of elaborated algorithms was to preserve values of the features as well as characteristic changes.

In the paper procedures of signal feature encoding have been presented. Examples of encoding have been based on experiments performed with the use of signals recorded during operation of a laboratory stand that models operation of rotating machinery. Some ways of evaluation of encoding, its correctness as well as sensitivity to different phenomena have been indicated. Presented algorithms have been based on known approaches described in the next point. The applied encoding required special ways of evolutionary operations, which have been discussed in the last point.

2. Related encoding applied in evolutionary algorithms

In evolutionary algorithms individuals representing input data and potential solutions are encoded by individual genotypes (Fig.1). In most cases the genotype has a structure of a single chromosome. Such encoding is called haploidal [3,5,6,7]. Encoding of one genotype by means of two or more chromosomes is called diploidal.
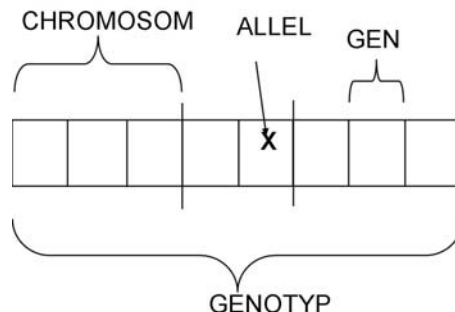

Fig. 1. Structure of a diploidal genotype

In case of simple genetic algorithms the most applied way of encoding is binary one. Data are replaced by values of two element set <0,1>. There are also different manners of encoding that are based on sets of real numbers or symbols. The following procedure of transformation of data into binary codes are being commonly applied [1–3, 6–8]:

1. binary notation being often used for natural number; numbers are transformed according to the following relation [8]:

$$x_i = a_i + \frac{b_i - a_i}{2^{n_i}} \sum_{n_i-1}^{j=0} 2^j x_{s(i)+j}$$

where: x_i is a value being encoded, such as $x_i \in [a_i, b_i] \in R$, n_i is a number of bits used for encoding, and s_i represents consecutive bit numbers,

2. according to the following relation [8]:

$$[\alpha, \beta, bin] = (-1)^{\beta} e^{(-1)^{\alpha} [bin]_{10}}$$

where: $[\alpha, \beta, bin]$ represents the code, α is a bit representing the sign of the logarithmic function, β is the sign of a power of the power function and *bin* is a value of the power encoded in the form of binary notation.

The logarithmic encoding is usually applied in order to decrease the length of genotype chains. Among other approaches one should enumerate:

- permutation encoding; data are represented by numbers and symbols included in an alphabet (e.g. [1,4,*,+,]) [7],
- encoding by means of real numbers; data are transformed into genotypes without any special processing; such operation can be expressed by, $x_i = v_i \in R^n$ where: x_i is a value of a single gene and v_i is a value of given data.

A way of data encoding depends on a kind and quantity of data as well as required accuracy of encoding. Genetic operations are strongly related to the applied notation. The more complicated encoding the more advanced genetic procedures are required to be applied. Taking into account these procedures one can state that the simplest notation is the binary one. However, in many examples such notation can not be applied because of huge sets of parameters that are necessary to be encoded. The application of the binary notation causes that genotypes are being very long. Decrease of encoding accuracy (lower number of bits used to encode a given value) leads to the loss of information about data. In the bibliography [1, 2, 7] there are two rules of data encoding:

- a notation should be selected in such the way that reflex data properties and genotypes of different data are significantly different,
- an alphabet used for coding should be possibly the smallest one; elements of this alphabet should express encoded data without changing their nature.

3. Types of input data

The data, which have been put through encoding were results of analysis of vibration signals. The analysis have been based on different methods. Three different types of signal features have been considered (Fig. 2):

- I type – one parameter function (e.g. time); features represented by means of vectors;

- II type – two parameter function (e.g. time and frequency); features represented by means of matrix,
- III type – trajectories of motion of a point in the established space; the trajectory is estimated within time function; feature recorded as two column matrix.

4. Encoding algrotihms

According to some assumptions, the way of encoding should let us to process obtained codes with the use of evolutionary algorithms. In case of the research related to analysis of signals by means of different domain methods, the fundamental difficulty is to elaborate one consistent notation of genotypes. Three individual approaches to estimated signal features have been elaborated. Their codes are elements of genotypes. Examples of elaborated procedures have been presented for features of signals recorded during operation of a laboratory stand which is a simple model of rotating machinery.

4.1. I type feature encoding

Examples of the I type feature are time functions. In this case, depending on the signal analysis as well as signal type (variability and non-stationarity) different encoding procedures are possible to be applied. This data encoding is described in the bibliography [1, 2, 8]. Two examples of features (mean values in time function) of this type have been shown in Fig. 3.

They are differ in a change character and range of values hey are equal to. The left plot is an example of two state signal. In such case the simplest form of binary notation can be used, such as $[b1,b2]$ where: $b1$ is the first and $b2$ the last value within selected interval and $[b1,b2] \in <1,0>$. Taking into account results of analysis of vibrations recorded during operation of rotating machinery one may state that signal features of two state signal type appear very rarely. Such signal are often observed during observation of different objects, especially such machines that were also a subject of investigations described within the first point. The binary notation presented above is often not enough to encode some data. In the previous point there have been presented more advanced cases of data transformation into binary chromosomes. The notation in the form of two values can be only applied in the case of two state signals or signals that can be transformed to such signals without significant loss of information. However, in this case of encoding lengths of considered intervals was crucial.
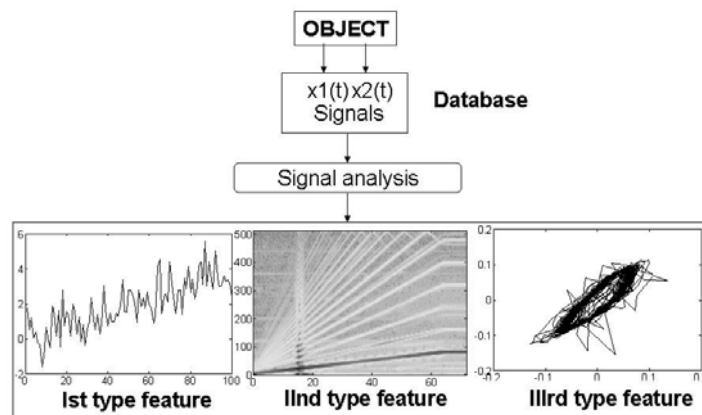

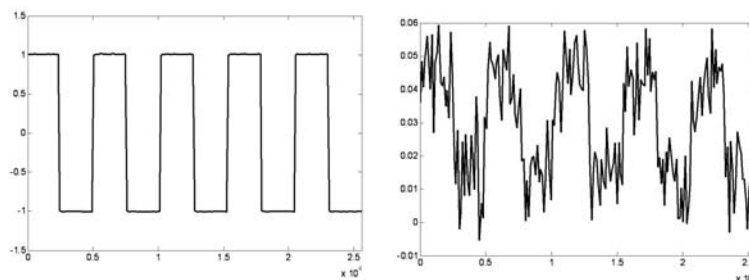*Fig. 2. Signal feature encoding*

Fig. 3. Examples of I type features

In case of the signal shown on the right side of Fig. 3 more complex notation was necessary to be applied. The most important thing was to determine such notation that reflex change character and tendency within encoded interval of signal feature.

As a result of the investigations related to numerous ways of data encoding a notation based on four real values has been accepted. A general scheme of such notation has been shown in Fig. 4. In this case the first, second, maximal and minimal values have been assumed to be consecutive gene values of the chromosome representing the considered interval. Some examples of such encoding have been presented in Fig. 5.
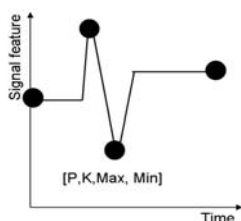

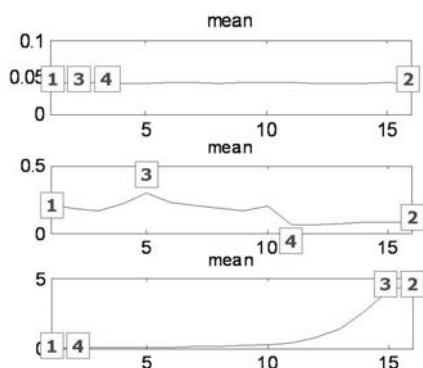*Fig. 4. Scheme of real numbers based encoding*

Fig. 5. Examples of real number based encoding of mean values estimated as time functions

Such notation makes it possible to preserve the information about the magnitude of the change as well as it character. The chromosome has the following form $[r1=x(0), r2=x(T), r3=\max(x), r4=\min(x)]$ where $r1, r2, r3, r4 \in R$. Such notation did not exhaust all possibilities to be applied. However, the performed encoding study has shown that such code is the shortest from different notation which are able to reflect characteristics of analyzed data.

4.2. II type feature encoding

Examples of the II type signal features are results of short time Fourier transform. They are usually presented in the form of time-frequency characteristics [11, 13] (Fig. 6). Such features are being represented by means of matrix in which the row number corresponds to spectra, and column number is equal to frequency bands.
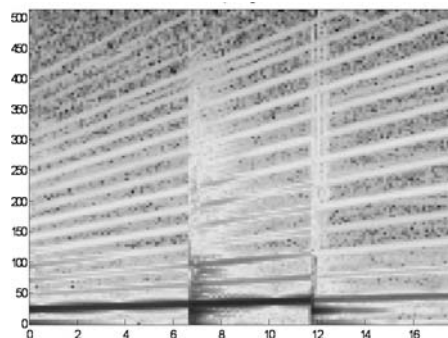

Fig. 6. Example of II type feature

The interpretation of time-frequency characteristics leads mainly in identification and estimation of such particular areas as:
- narrowband harmonics characterized by constant frequencies (e.g. components generated by elements rotating with constant rotation speed),
- broadband harmonics characterized by varying frequencies (e.g. components generated by elements rotating with varying rotating speed)
- narrowband noise (e.g. effect of operation of other objects within neighborhood of the observed one),
- resonances appearing as increases of magnitudes of narrowband harmonics,
- periodic impacts appearing as several regular wideband components,
- random impacts that appear as single wideband components.

Identification of enumerated areas can be based on two approaches:
- quantitative procedures, which result in estimation of values of frequencies and magnitudes of identified components,
- qualitative procedures that lets us only to identify the presence of these components and a degree of their dominance within the set of all distinguished areas of the characteristics.

In most cases the second approach can be applied in order to determine changes of technical state of the object. It was assumed that the time-frequency characteristics has been treated as an image [12]. According to this assumption all procedures are aimed at identification of lines and some areas that fulfill

determined requirements. Placement and orientation of the lines and areas have been also estimated. In order to make it possible to apply methods of image analysis some additional procedure of image processing had to be applied. They were equalization of histogram of color resolution (in order to distinguish unclear components; the left side of Fig. 7) and transformation of the characteristics into black and white image (threshold operation were applied; the right side of Fig. 7). The threshold value was established experimentally and was equal to 80% of maximal of power determined for whole characteristics.

Images being results of listed procedures were through three different methods of image analysis. Results of the methods are three different sets of features. The methods have been based on scanning operation. Directions of scanning have been shown in Fig. 8.

Scanning was performed along:
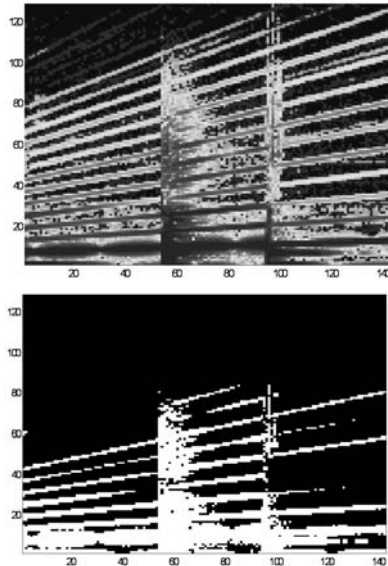- time axis; arrow number 1,
- frequency axis; arrow number 2,

Fig. 7. Results of processing and transformation of II type features

- sections whose ends were point laying on time axis (the point have been indicated by markers); arc number 3.

4.2.1. Algorithm I

In case of this algorithm the following results have be obtained (Fig. 9):
- averaged intersection of characteristics (arrow number 1),
- averaged intersection of characteristics (arrow number 2),
- averaged intersection of characteristic estimated along sections determined by the arc number 3.

In order to determine numbers of narrow and wide components these intersections have been analyzed A chromosome resulting from this algorithm has the following form $[t1, t2, f1, f2, s1, s2]$ $t1, t2, f1, f2, s1, s2 \in <0,N>$ where: $t1, f1, s1$ are numbers of wideband components, and $t2, f2, s2$ numbers of narrowband components estimated on the basis of intersection from Fig. 9.

4.2.2. Algorithm II

The algorithm has been also based on the same scanning operations. However, in this case only two intersections (1 and 2) have been taken into consideration. Basing on them some statistical measures have been estimated. Series of values such as change, shape, kurtosis, pick, impulse coefficients as well
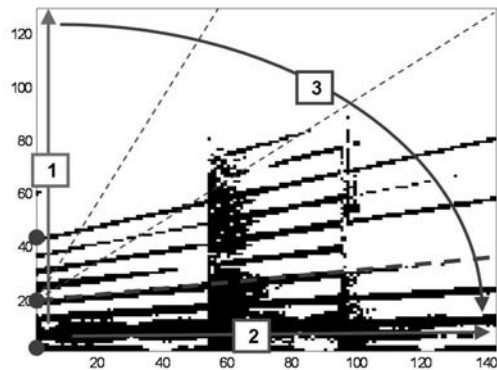

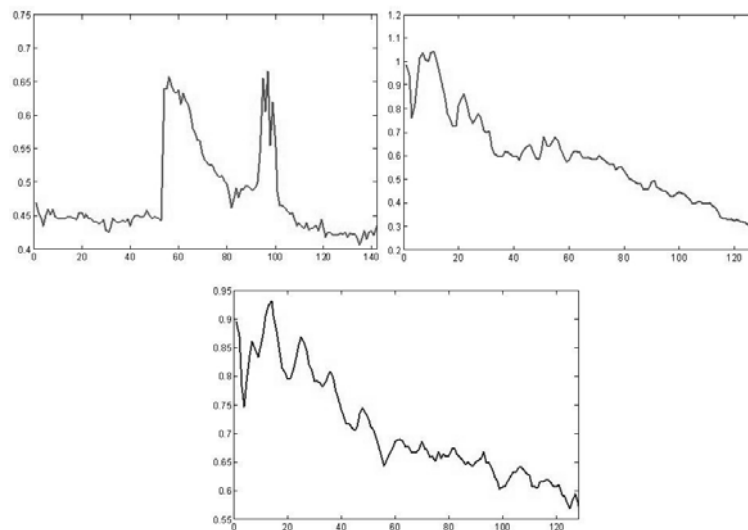*Fig. 8. Schema of scanning procedures*

Fig. 9. Results of scanning application Algorithm I

as standard deviation have been tested. They are often used to estimate a degree and kind of variability. The sensitivity analysis has been performed. On the basis of its results the set of enumerated coefficients has been decreased into four values estimated for one intersection. Exemplary result of such analysis has been shown in Fig. 11. Values of non-dimensional measures have been presented as bar plots. Standard deviation has been placed above corresponding plots.

In this case a chromosome has the form of the code: form $[k1, s1, kr1, os1, k2, s2, kr2, os2]$ $k1, s1, kr1, os1, k2, s2, kr2, os2 \in R$ where: k, s, kr, os are shape, pick, kurtisis factors and standard deviation values, numbers 1 and 2 are related to two analyzed intersections.
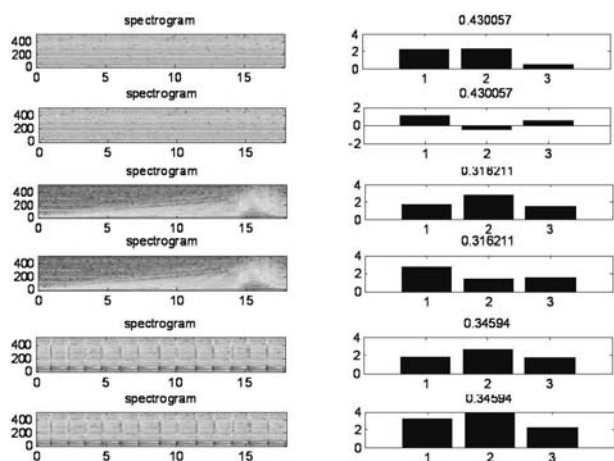

Fig. 10. Variability measures estimated according to Algorithm II

4.2.3. Algorithm III

According to the algorithm results of scanning have been not averaged intersection but distinguished intersection. Examples of intersections which correspond to dominant components and areas have been presented in such Fig. 11.
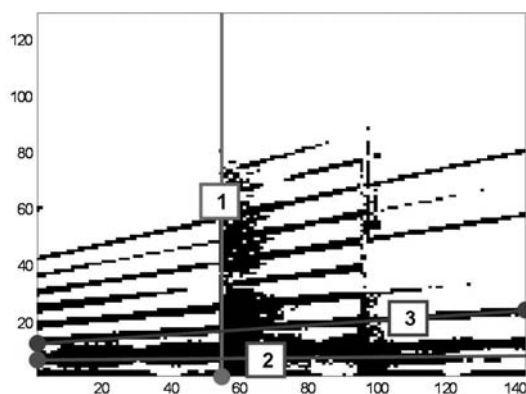

Fig. 11. Distinguished intersection of the characteristics

Number 1 indicates a narrowband components of constant frequency which has been encoded be means of two values, frequency and mean magnitude. Broadband components (result of impact) has been indicated by number 2. It is encoded by me-

ans of time moment, and mean magnitude. Number 3 indicates component of varying frequency, which has been represented by initial and final values of frequency and men magnitude. A chromosome which is composition of the notation has the following form $[f1, a1, t2, a2, f31, f32, a3]$, $f1, a1, t2, a2, f31, f32, a3 \in R$ where: $(f1,a1)$, $(t2,a2)$ and $(f31,f32,a3)$ are values corresponding to consecutive components. Consideration of only one components in each intersection can not make it possible to determine all phenomena which are reflected in characteristics (e.g. periodic impacts, resonances). However, taking into account high number of components such codes lead to significant decrease of the length of chromosomes.

4.3. III type feature encoding

An example of the III type feature is a trajectory of movements of a point estimated as time functions. The trajectory was evaluated on the basis of values of displacements recorded as two channel signal. Displacements were reiterated in two mutually perpendicular directions. Trajectories are being interpreted by means of different measures. Typical estimators are [4, 14] dimensions, spin direction, angular placement related to assumed axes, period and shape.

Trajectories estimated during observation of rotating machinery are usually averaged. An approach proposed in the paper assumed that they were not averaged what makes possible to reveal some additional changes in machinery operation. Trajectories have been represented by four values, which correspond to diagonals (the shortest and longest) for each of two perpendicular directions. In order to estimate these values trajectories were put to operation shown in Fig. 13. Numbers 1-4 carry information about trajectory dimensions and shape. The chromosome representing a trajectory has the form of four number code $[a1, a2, b1, b2]$ $a1, a2, b1, b2 \in R$ where: $(a1,a2)$, $(b1,b2)$ are the longest and shortest diagonals in horizontal and vertical positions.
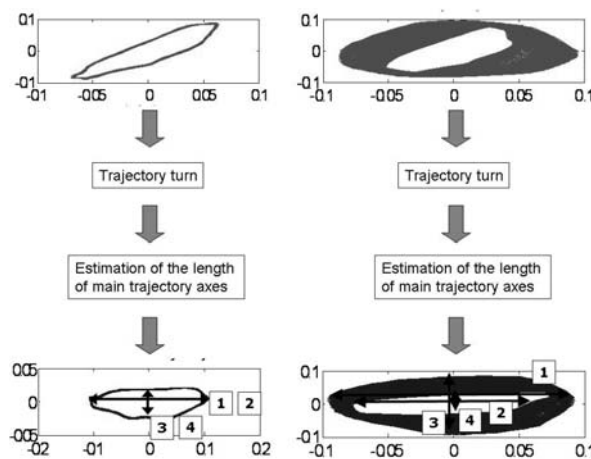

Fig. 12. Examples of the III type features and their notations

5. Final genotype structures and their further processing

During the experiments two vibration signals were registered. The signals were analyzed within selected time periods. For these intervals mean values as time functions, time-frequency characteristics and trajectories were evaluated. Signal features

were encoded by means of algorithms described in the paper. The final genotype structure has the following form:

$OV=[FI1 \mid FI2 \mid FII1 \mid FII2 \mid FIII]$, where OV means observation vector, $FI1$ and $FI2$ are codes of the I type features obtained for two signals, $FII1$ and $FII2$ are codes of the II type feature and $FIII$ is a code obtained for trajectories, which is an example of mutual feature of two signals. In this case the length of the whole genotype is 24 genes. It should be stressed that such notation let us to preserve specific characteristic changes and trends of estimated features as well as extract them out of the features.

The genotypes, in the form presented above, are put through clasterization, which is based on PCM (Possibilistic C-Means) algorithm [8]. As results fuzzy clusters are being obtained that are treated as exemplary observations and are gathered within an example base. According to the assumption related to context identification and consideration [9, 10], these examples can be treated as associations which are most specific contexts of object operation. The application of specially elaborated evolutionary calculations makes it possible to identify other contexts. Since the structure of the genotype is complex genetic operation, especially crossover and mutation had to be elaborated. Identified contexts provide us with the information about so called "about context" knowledge and "in context" knowledge [10]. which are in some ways related to circumstances of the object operation. Such distinguishing the knowledge let us to improve diagnostic inference. The procedures of further genotypes processing briefly described above have been subjects of further works.

6. References

[1] Arabas J.: *Lectures on genetic algorithms*, WNT Warszawa 2001 (in Polish).

[2] Goldberg E. D.: *Genetic algorithms and their applications*, WNT 1998 (in Polish).

[3] Jacob C.: *Stochastic Search Methods, Intelligent Data Analysis*, p. 351-401, ed. Berthold M., Springer-Verlag Heidelberg – Berlin 2003.

[4] Jordan M.: *What are orbit plots, anyway*, Orbit 1993, vol. 14, no 4, p. 8-15.

[5] Kowalczuk Z., Białaszewski T.: *Genetic algorithms in multioptimalization of detection observers*, Process Diagnostics. Models, AI Methods, Applications, p. 465-512, ed. Korbicz J., Kościelny J. M., Kowalczuk Z., Cholewa W., PWN Warszawa 2002 (in Polish).

[6] Ponisch A., Azzaro-Pantel C., Domenech S., Pibouleau L.: *Some Guidelines for Genetic Algorithm Implementation in MINLP Batch Plant Design Problem*, [in] Advances in Metaheuristics for Hard Optimisation, p. 230-315, ed. Siarry P.. Michalewicz Z., Springer-Verlag Heidelberg - Berlin, 2008.

[7] Ronald S.: *Robust Encodings in Genetic Algorithms*: A Survey of Encoding Issues, IEEE International Conference on Evolutionary Computation, p. 43-48, April 1997.

[8] Rutkowski L.: *Methods and techniques of AI*, WNT 2005 (in Polish).

[9] Timofiejczuk A.: Identification *of associative rules with the use of evolutionary algorithm*. Symposium "Machinery Diagnostics", Węgierska Górka 2007 (in Polish).

[10] Timofiejczuk A.: *A concept of context-based inference in technical diagnostics*, CAMES no 12, p. 267 – 277, 2005.

[11] Timofiejczuk A.: *Methods of non-stationary signal analysis*, Publishing House of Silesian University of Technology, Gliwice 2004 (in Polish).

[12] Timofiejczuk A.: *Application of image analysis to identification of changes of features of diagnostic signals*, Symposium "Machinery Diagnostics", Węgierska Górka 2002. (in Polish)

[13] Timofiejczuk A.: *Visualization of time-frequency characteristics in layered plot form.*, Symposium "Machinery Diagnostics", Węgierska Górka 1996 (in Polish).

[14] Wysogląd B.: Methods *of representation of shaft vibration in diagnostic databases*. Publishing House of Silesian University of Technology, Gliwice 1996 (in Polish).

Dr inż. Anna TIMOFIEJCZUK

Silesian Univeristy of Technology
Faculty of Mechanical Engineering
Konarskiego Street 18a
44-100 GLiwice, Poland
e-mail: anna.timofiejczuk@polsl.pl