Article citation info:
Strączkiewicz M, Czop P., Barszcz T. Supervised and unsupervised learning process in damage classification of rolling element bearings, Diagnostyka. 2016;17(2):71-80

DIAGNOSTYKA, 2016, Vol. 17, No. 2

ISSN 1641-6414
e-ISSN 2449-5220

SUPERVISED AND UNSUPERVISED LEARNING PROCESS IN DAMAGE CLASSIFICATION OF ROLLING ELEMENT BEARINGS

Marcin STRĄCZKIEWICZ, Piotr CZOP, Tomasz BARSZCZ

AGH University of Science and Technology, Department of Robotics and Mechatronics,
al. Mickiewicza 30, 30-059 Kraków, Poland, e-mail: tbarszcz@agh.edu.pl

Summary

Damage classification plays a crucial role in the process of management in nearly every branch of industry. In fact, is becomes equally important as damage detection, since it can provide information of malfunction severity and hence lead to improvement of a production or manufacturing process. Within this paper selected supervised and unsupervised pattern recognition methods are employed for this purpose. The attention of the authors is given to assessment of selection, performance benchmarking and applicability of selected pattern recognition methods. The investigation is performed on the data collected using an experimental test grid and rolling element bearing with deteriorating condition of an outer race.

Keywords**:** fault classification, pattern recognition, rolling element bearing, multiple classifiers comparison.

NADZOROWANY I NIENADZOROWANY PROCES UCZENIA W KLASYFIKACJI USZKODZEŃ ŁOŻYSK TOCZNYCH

Streszczenie

Klasyfikacja uszkodzeń odgrywa ważną rolę w procesie zarządzania w niemalże każdej gałęzi przemysłu. W rzeczywistości staje się ona równie istotna co samo wykrywanie uszkodzenia ponieważ pozwala określić stopień uszkodzenia, a co za tym idzie, poprawić efektywność zarządzania zakładem przemysłowym. W tym celu wykorzystano wybrane nadzorowane i nienadzorowane metody rozpoznawania wzorców. W artykule zwrócono uwagę na ocenę wyboru, porównanie wydajności oraz możliwości wykorzystania tych metod. Analiza przeprowadzona została na danych zgromadzonyh na eksperymentalnym stanowisku testowym, gdzie obserwowany jest stan łożyska tocznego z pogłębiającym się uszkodzeniem bieżni zewnętrznej.

Słowa kluczowe: klasyfikacja uszkodzeń, rozpoznawanie wzorców, łożyska toczne, porównanie klasyfikatorów.

1. INTRODUCTION

Nowadays, modern industrial plants are more flexible and cost-effective. Most of them are equipped with systems for remote monitoring of critical plant parameters or even for remote controlling of all plant parameters. This includes production or manufacturing management and maintenance processes. The process of monitoring is performed using highly configurable supervisory control and data acquisition units (SCADA – data-gathering orientated) and distributed control systems (process-orientated) which are interfaced to the plant via programmable logic controllers (PLCs) and measurement modules. Self-learning and easy-tuning PLCs operate to achieve plant demand targets. The controller settings, diagnostic and control signals are often available online via LAN and WAN computer networks, thus all of the information in the system might be processed by authorized personnel from any far place.

Simultaneously, there is a growing number of methods that improves the diagnostic process at various stages of fault development. Due to the constant advancement of computational and storage capabilities of modern condition monitoring systems (CMS) many of the proposed methods became entirely automatic and can be implemented e.g. in online CMS [1]. For the purpose of condition monitoring of some groups of objects, e.g. rotating machinery, it is especially important to assess the severity of damage since early symptoms of fault do not necessarily require replacement of the elements, but it might suggest that the replacement will be mandatory in foreseeable future and appreciate action can be performed ahead of time. Therefore it is crucial to distinguish arising malfunction and provide higher effectiveness of process management.

One may distinguish following failure modes, depending on the advancement of the systems used: (i) detection, (ii) isolation, (iii) assessment and (iv) classification [2].

The first one actives alarm once the established control limit is exceeded while the fault isolation provides a measured physical quantity that activates the alarm. These two modes are successfully implemented in many industrial applications, e.g. in SCADA system. Fault assessment informs of a weight factor for further operation – thus it is difficult to automatize and often rely on standards or opinion of the domain specialists. For the fault classification one estimates a reason for occurrence of malfunction and sorts a given set of information into particular category, like early warning, severe alarm, etc. The latter one may be performed using the theory of pattern recognition, which consists of a large group of techniques, including supervised and unsupervised ones, that derive the assessment on data with known and unknown class labels, respectively.

The authors decided to present widely used statistical methods, i.e. nearest neighbor and nearest mean classifiers, logistic regression, linear and quadratic discriminant analyses, Parzen classifier and on the other hands, soft computing methods, including radial basis function neural network, as examples of a supervised methods, and c-means fuzzy logic and k-means, as unsupervised ones. These approaches proved their effectiveness in various engineering and, in particular, condition monitoring applications, as pointed throughout the article. Their advantages can be found in relatively simple application and low computation time, thus it can be assumed that they can be directly applied in state-of-the-art CMS. This comparison constitutes the next step in the investigation performed by the research team, that is development of CMS for commercial applications in highly varying operational conditions, that includes signal processing methodology [3], fault detection techniques [4] and structure design and reasoning of such systems [5-7].

In this paper the authors concentrate on fault classification of rolling element bearings (REBs) in different stage of its development. The study is not intended to be performed on the methodological aspects of particular techniques, but on their effectiveness for REB’s outer ring fault classification. The reason for investigation of REBs is that they are one of the most commonly used kinematic elements and are used nowadays in almost every industrial branch there is. Unfortunately, in the same time, according to the study mentioned in [8], they are also most fault susceptible – about 80% of total failures in industry are caused by these elements.

This paper evaluates supervised and unsupervised classification methods considering uncertainty and imprecision of diagnostic reasoning. Uncertainty is considered by means of bias, stability, and linearity variance components. Nevertheless, the diagnostic reasoning process involves imprecision which is related to missing or incomplete knowledge about the object, related measurements (e.g. other bearings), historical data, and detail specification. The paper reviews a few clustering and classification algorithms to represent the uncertainty and impression of measurement data.

The article is organized as follows. After the introductory part the pattern recognition methods are described in Section 2 with distinction to supervised and unsupervised ones. In Chapter 3 the case study is presented on the experimental test rig along with the data acquisition and feature extraction algorithms. The consecutive stages of the rolling bearing are introduces. Next, the validation of classification algorithms is presented and their effectiveness in terms of uncertainty and imprecision is discussed. Finally, the paper is concluded with the remarks for further research.

2. PATTERN RECOGNITION

Generally, as pattern recognition one may understand a group of methods that focus on classification of objects or observations in a number of categories or classes. In the literature (e.g. [9]) there is established general model of pattern recognition and can be simply applied to any type of method from this group. It is presented in Fig. 1. The model consists of the preprocessing module, the feature selection/extraction module and the classification module. The input data are collected in the initial stage, before the pattern recognition process and refers to measurement or observation on the object to be classified. It can be represented by e.g. images, acoustic or vibration signals, but it is highly desired to acquire data that can be understood as acceptable pattern of the observed phenomena. Any disturbances in the measurement may be filtered in the preprocessing stage, which in general aims at improvement of the data quality. It is also in charge of converting the raw measurements, which could come from different sources and be provided in different standards, into unified format suitable for further operations.

The feature extraction/selection module prepares the feature space for use by classification module. The feature extraction/selection module prepares the feature space for use by classification module. The preparation involves selecting the best set of features coming from the preprocessing module and many kinds of linear or non-linear transformations of features carrying in order to obtain feature space with the most promising properties, i.e. that enhance classification and reduce overfitting of data [10]

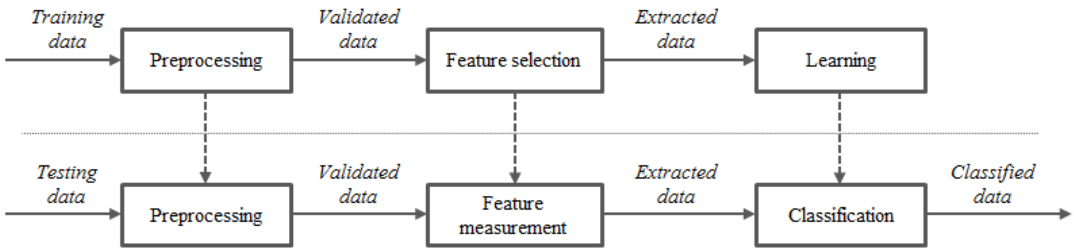

Fig. 1. General pattern recognition structure

There are more than a few methods introduced for statistical feature extraction and selection, e.g. principal component analysis (PCA), linear discriminant analysis or projection pursuit (for extraction) and sequential forward/backward floating search or exhaustive search (for selection) [11]. In case of vibration-based fault assessment, feature extraction stage may rely on e.g. time-series models, frequency or time-frequency analysis [12], while the selection of indicators is often dependent on the object to be observed and available methods for detection and identification of damage.

The task of the classification module is to determine the probability of the object belonging to the particular class based on the vector of features provided in the previous step.

In the presented example, the structure consists of two layers with training or testing data as inputs. In the first one the training data is accompanied with known output, and thanks to that the learning process may be executed. This type of proceeding is related to supervised methods. In the contrary, in unsupervised ones, the user provides as inputs examples of an unknown class and desire from the system to determine on its output to which of a set of classes the example belongs. In former mode, parameters and even structure of each module can change in order to provide better results, while in latter one, parameters and structure are constant [9].

2.1. Supervised methods

In the supervised approach, the classification is performed based on the relationship between the input explanatory independent vector of features and the dependent class or cluster. As stated previously, each explanatory observation should be labeled with the corresponding class. Such training set can be used for teaching of a selected method until the relation between inputs and category is established. The obtained pattern is than used for the unseen testing data. The label are not known to classifier until the verification step, when the obtained results are compared with the actual indications.

One of the most commonly used supervised classifier is a *k*-nearest neighbor (*k*-NN) one. It classifies observations by the distance estimation between the given observation and *k* nearest neighbors from the training set, regardless their class labels. The number of neighbors should be chosen to be odd for a two class problem, and in general not to be a multiple of a number of classes [13]. Next, out of these *k* samples, it identifies the number of instances that belong to each class. Finally, it assigns the observation to the class with the maximum number of samples assigned. The simplest version of the algorithm is for k=1, known as the nearest neighbor rule. Its efficiency for fault classification in mechanical engineering was verified in numerous studies, like for rolling element bearings [14], gears [15] or for various types of faults in induction motors [16].

The multinomial logistic regression deals with situation when the observed outcome of the input variables is modelled in the training mode using a linear predictor function that uses a set of weights or regression coefficients associated with the outcome that are linearly combined with input explanatory vector for observation [17]. The result of such combination is a value representing association of a particular observation with each class. In this case the predicted classification is based on the highest value obtained for each of the available classes or categories. The exemplary use of regression modeling was presented in [18], where in was used for fault estimation in wind turbines.

Another interesting approach is normal density-based linear classifier, also called linear discriminant analysis (LDA). It can be used to describe a linear combination of explanatory data that are most suitable for distinguishing of two or more categories of objects [19]. Here, the predictor is solely based on the vector of observations x, hence outcomes y for a training set are not necessary. It is possible because the LDA approach assumes that the probability density functions have normal distributions and the same covariance. It can be shown, that the required probability depends only on the scalar product of difference between mean values and the input. This means that the probability of the input belonging to particular class is a function of a linear combination of its known characteristics. The probability factor is a foundation for the class justification. In the industrial applications it was used e.g. for classification of engine oils [20] or sensor failure in air handling units [21].

Similar to LDA, for the quadratic discriminant analysis (QDA) it is assumed that the observation vector is normally distributed for each category.

The difference between the two methods can be found for covariances, which are supposed to be different. For QDA the classification is made using the likelihood ratio, that calculates probability that a given observation belongs to particular class [11]. Good examples of industrial applications are provided in [22].

Another classifier analyzed within this paper is called nearest mean or minimal distance. It provides a very simple optimal decision rule. To classify a feature vector, there is a need to measure the Euclidean distance between each observation to each of the mean vectors, and assign the input to the category of the nearest mean. Due to its simplicity it found a large variety of applications, especially in image processing [23] and fault assessment [24].

Parzen classification technique depend on the estimation of the probability for each class based on acquired explanatory examples. It integrates the contribution of the entire training set to the calculated output variable with modelling using a kernel function that is influenced by the smoothing parameter, i.e. the kernel width [25]. The examined unclassified variable is assigned to particular category based on the maximal posterior probability. The classification properties of this methods found its use, e.g. for self-adapting alarm level adjustment [26] or for structural fault identification [27].

An interesting approach is radial basis function neural network. The network is similar as for classical neural network and the inputs are formed using explanatory variables. At the input of each neuron, the distance between the neuron and the input vector is obtained. The outputs are calculated as weighted sum of the hidden layers and the unity bias. Most commonly, the basis function is the Gaussian bell one [28]. The neural networks were used for bearings diagnostics in [29], while in [30] the authors compared radial basis functions-based NN with back-propagation networks, showing their superiority based on the fast training time.

2.2. Unsupervised methods

Unsupervised methods are effectively used in operational diagnostics, when it is not feasible to explicitly and a priori categorize the potential malfunction classes. In such case, the class structure in the data needs to be discovered without the support of a priori knowledge. Unsupervised methods allow to transform and reduce the data using two main techniques: (i) subspace structure of data and its (ii) clustering characteristics. The first approach summarizes the objects using a smaller number of features than the original number of measurements; the second summarizes the data set using a smaller number of objects than the original number. Subspace structure is often interesting for visualization purposes. Clustering provides similar results and interpretation, but also data reduction. When very large amounts of data are available, it is often more efficient to work with cluster representatives instead of the whole data set.

There are different clustering methods, such as hierarchical clustering, k-mean, mixture of Gaussians, mixture of probabilistic principal component analysis or fuzzy c-means clustering [31]. For example, the k-mean clustering is performed in 4 steps: (i) assign each object randomly to one of the clusters, (ii) obtain the means of each clusters, (iii), reassign each observation to the cluster with the closest mean (iv) return to step ii until the means of the cluster do not change within the assumed tolerance. The unsupervised algorithms provide potential for state conditions reduction which is an essential stage in diagnostic reasoning from operational point of view. There are external and internal approach to evaluate the clustering performance results [11]. An external measure is an agreement between two partitions where the first partition is the a priori known clustering structure, and the second results from the clustering procedure. Internal measure is used to measure the goodness of a clustering structure without external information. The optimal number of clusters is usually determined based on an internal validity measures reported for example by Rousseeuw [31] or Thalamuthu et al. [32].

Within this paper two *k*-mean-based clustering approaches are employed, namely fuzzy logic and crisp logic. In the first one the resulting clusters are best analyzed as probabilistic distributions [33] while the second can be referred as a hard assignment of labels. Liu et al. [34] provided an expert system based on fuzzy logic for detection of REBs faults. In [35] authors proposed to employ this method to integrate bearing fault indicators and therefore simplify the diagnostic process.

3. CASE STUDY

The performed comparison of the pattern recognition methods was performed on laboratory test rig that enabled acquisition of vibration signals of rolling element bearings in four conditions: while no fault is observed and with three levels of bearing degradation, each with deepening crack of the outer race. The preprocessing of measurements, feature extraction and learning/classification stages were performed as discussed in the following subsections.

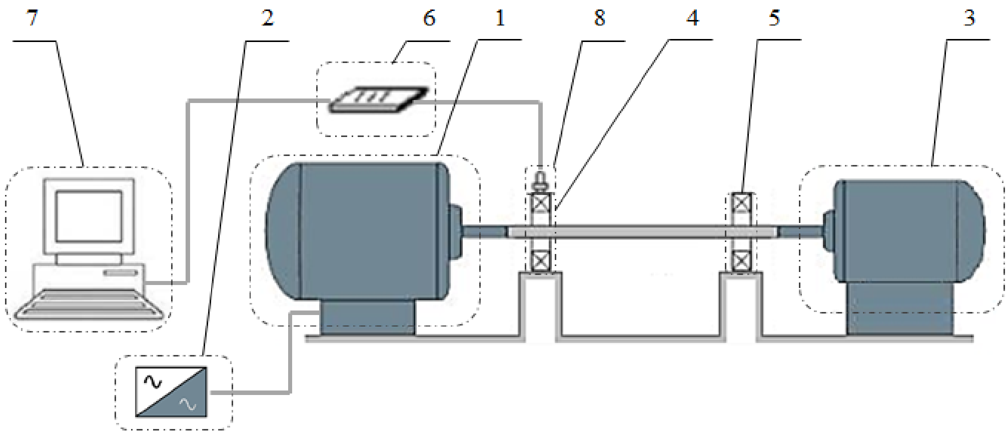

Fig. 2. Test rig used in the experiment

3.1. Test rig

In Fig. 2 one may find the test rig employed in the experiment. In consisted of three-phase asynchronous motor (marked as 1), that drove the mechanical system with rotational speed controlled by the inverter (2), power generator (3) imposing load to the system, two identical SKF 1205 EKTN9 double-row ball bearings (4 and 5), out of which one (4) was under investigation with introduced incrementing defect of outer ring.

The data was collected with 4-channel NI card 9233 (6), with 24bit resolution, 50 kHz of sampling frequency and maximal range +/- 5V and a stationary computer (7). It was acquired using VIS-311B accelerometer (8) with maximal range of 30 kHz and sensitivity equal to 100mV/g. The accelerometer was placed perpendicularly to the axis of shaft on the upper surface of the bearing. The rotational speed was set to a constant value of 45Hz. The power generator imposed load to the system equal to 130W and it remained constant during the entire enterprise.

During the experiment for each of the fault stages there were acquired 60 measurements of 10s duration in order to provide frequency resolution 0.1 Hz. It was captured with 25kHz sampling frequency. The faulty bearing was analyzed in four stages as described in Tab. 1. The damage enlargement was followed by the bearing mounting together with shaft alignment and the measurement. This procedure was repeated for each size of damage.

Table 1. Fault development stages of examined rolling element bearing

State	Depth of crack [mm]
Normal	0
Acceptable	0.5
Warning	1
Alarm	1.5

3.2. Preprocessing

The preprocessing of acquired vibration signals followed methodology proposed in [36]. It was examined against human mind-wise valid vibration signal characteristics (from a continuously running machine) i.e.: mandatory visual continuity, required certain complexity (as opposed to a computer-generated sine wave), rational amplitude levels, imperceptible quantization (for sufficiently long time period), sufficient sharpness of the time waveform shape (due to expectation of high frequency components), sudden signal changes present only to a degree allowable by the machine real behavior, expected mean value accuracy (zero in case of acceleration).

In the discussed study, aforementioned properties of signals were satisfied and all of the collected signals were classified as correct and valid, and therefore accepted for further feature extraction.

3.3. Feature extraction

The existence of a localized fault in REB results in the periodical excitation of elements' resonance frequencies. This manifests as amplitude modulation of the vibration signal (exemplary response was demonstrated in Fig. 3) once the rolling element encounters the defect (marked as 'Point of Entry' and 'Starting position of the defect' on the on right and left plot of Fig. 3, respectively). The point of impact represents the moment of excitation caused by the rolling element leaving the area of defect. For numerous balls mounted in cage this phenomenon is observed periodically over the entire cycle. In order to reveal these symptoms for each signal the envelope spectrum was obtained and the particular characteristic frequencies were observed.

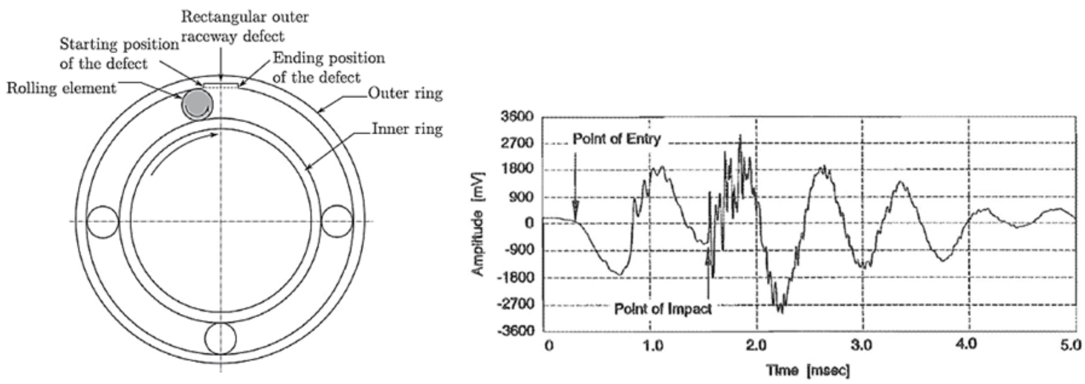

Fig. 3. Experimentally measured acceleration response (right) of a ball bearing with an outer race defect (left) [37]

The procedure for envelope spectrum designation was described in numerous literature positions e.g. in [35]. Firstly, in the performed analysis, the unwanted frequencies related to shaft's rotating had to be cut-off using high-pass filtering. Secondly, the resulted signal was squared and the Fast-Fourier transform was computed in order to obtain envelope spectrum. The knowledge of geometrical parameters of the analyzed bearing allowed to identify four narrowband bearing indicators, i.e. ball-pass frequency of outer race (BPFO), ball-pass frequency of inner race (BPFI), ball spin frequency (BSF) and fundamental train frequency (FTF). The characteristic frequencies calculated for the given bearing were equal to 6.69 for outer race, 6.31 for inner race, 5.11 for rolling elements and 0.49 for cage. Each of the above values should be understood as given multiplication of a rotational speed of driven shaft.

3.4. Experimental validation

The conducted laboratory experiment provided data to evaluate both the uncertainty and imprecision terms in condition monitoring. Supervised and unsupervised methods were used in order to classify rolling bearing technical condition, reduce the amount of processing data, and find optimal number of data clusters representing diagnostic states [31]. The Matlab toolboxes, PR-Tools4.1 [38] and fuzzy clustering and data analysis [39], were used to cluster and classify the data sets.

Supervised methods used the learning set consists of 240 samples representing four categories, i.e. normal, acceptable, warning, and alarm conditions. The testing set consists of 80 samples, and it gives the learning/testing sets ratio as presented in Tab. 2. The applied algorithms and classification results are presented in Tab. 2. The correct classification performance indicator was used to validate the learning process based on the testing set. The classification was repeated 20 times for each algorithm to get a mean value which results from random shuffle of the samples for every classification case.

Table 2. Clustering and classification methods

Classifier type	Method type	Correct classification for testing set [%]
Logistic regression	supervised / parametric	93
k-nearest neighbor	supervised / non-parametric	87
Normal densities based linear classifier	supervised / parametric	98
Normal densities based quadratic classifier	supervised / parametric	97
Parzen classifier	supervised / non-parametric	96
Nearest Mean Classifier	supervised / non-parametric	88
RBF Neural Network	supervised / non-parametric	93
c-means fuzzy logic	unsupervised	89
k-means	unsupervised	94

For both supervised and unsupervised cases, the classification and clustering performance were also evaluated visually with the used of 2D diagrams (Fig. 4) that show partial classification features (dimensions), i.e. BPFO, BPFI, BSF, FTF components. Normal conditions area was characterized by minimum variance, while the alarm condition corresponding to the greater variation.

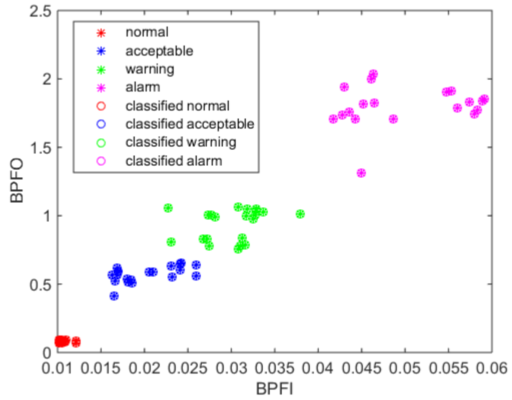

Fig. 4. Diagram of example of classification results (two features are visualized on x and y axes from the four total features)

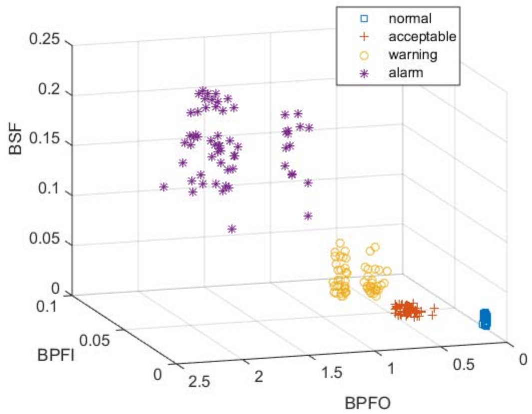

Fig. 5. Diagram of raw data clusters (three features are visualized on particular axes from the four total features)

Unsupervised clustering and classification process involves all data samples (i.e. 240 cases) to group data for a range of a priori given number of classes and it was performed in two steps. The optimal number of clusters were identified using cluster performance measures (Fig. 6), such as partition coefficient (PC), classification entropy (CE), Dunn index (DI), and alternative Dann index (ADI) [40]. The PC index indicates the average relative amount of membership sharing done between pairs of fuzzy subsets by combining into a single number, the average contents of pairs of fuzzy algebraic products [41]. The index values range in (1/c, 1), where c is the number of clusters. The closer to unity the PC, the smaller the sharing of the vectors in data set X among different clusters. The closer the value of PC to 1/c, the fuzzier the clustering is [42]. The PC index is a scalar measure of the amount of fuzziness in a given fuzzy set [39].

The DI is a metric type which is based on an internal evaluation scheme, where the results are based on the clustered data itself. This measure evaluates if sets of clusters are compact, with a small variance between members of the cluster, and well separated, where the means of different clusters are sufficiently far apart, as compared to the within cluster variance. For a given assignment of clusters, a higher Dunn index indicates better clustering (Fig. 5).

Alternative Dunn Index [43] is a variation of DI index, the difference between DI and ADI is on the inter-cluster distance measurement. The DI index tries to find the minimum distance between elements belonging to different clusters, while the ADI index tries to replace this calculation with the triangular inequality [45]. Fig. 8 shows the final classification results in 2D dimensions including the countour lines of fuzzy membership functions. Two features are visualized on 2D plain from the four total features (Fig. 7).

4. CONCLUSIONS

The conducted feasibility study on use of supervised and unsupervised methods in order to classify and cluster vibration measurement confirmed potentials for such approach under laboratory conditions. Nevertheless, the clustering analysis shows to be very efficient in determining the number of bearing state conditions to be considered in operational diagnostics were multiple-sensors data, monitored object specification, operational history including past failure/malfunctions, and other relevant information incompleteness is considered as a typical situation.

The supervised algorithms show good accuracy in the range between 88-97% considering the classification performance obtained with the use of testing set of four basic features of monitored rolling bearing, i.e. BPFO, BPFI, BSF, FTF frequency components. The unsupervised algorithms show also good accuracy in the range between 89-94% based on previously determined number of clusters equal to 4 condition states. This feasibility study allows to recommend all the classification methods to be used in a commercial diagnostic systems. Unsupervised methods are more suitable for operational data if it is not feasible to assign all malfunction classes to data clusters.

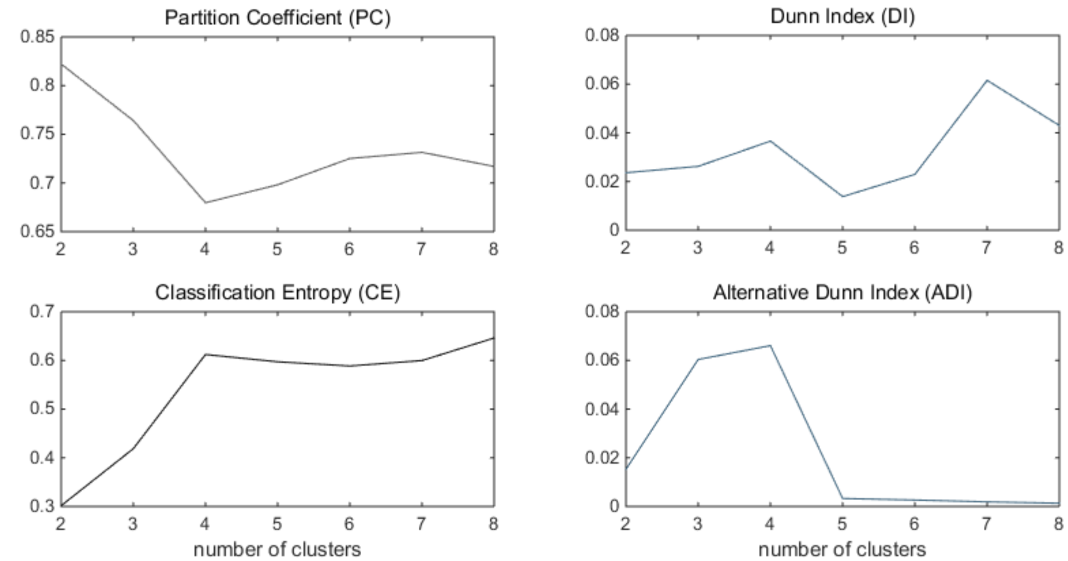

Fig. 6. Indices evaluating the cluster performance

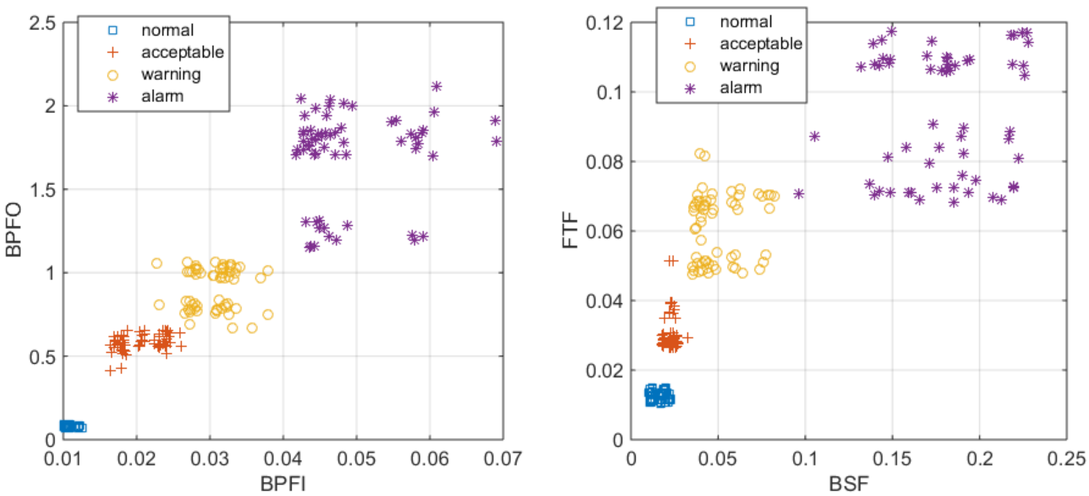

Fig. 7. Diagram of raw data clusters (two features are visualized on particular axes from the four total features)

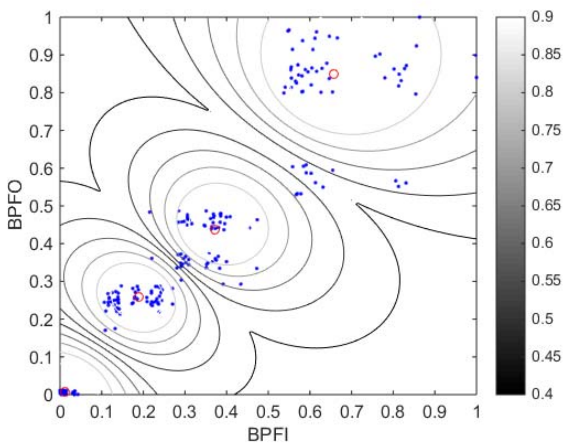

Fig. 8. Diagram of typical classification results

Acknowledgement

First author would like to acknowledge the financial support of AGH University of Science and Technology within the program of Dean's Scholarship (No. 15.11.130 987).

REFERENCES

[1] Barszcz T, Jabłoński A, Strączkiewicz M. New generation of condition monitoring systems for non-stationary machinery – proposal of the architecture, 10th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 2013, Krakow, Poland, pp. 1-10.

[2] Rytter A. Vibrational based inspection of civil engineering structures, fractures and dynamics [PhD thesis], defended publicly at the University of Aalborg; Department of Building Technology and Structural Engineering, Aalborg University, April 20, 1993, p. 206.

[3] Urbanek J, Barszcz T, Antoni J. A two-step procedure for estimation of instantaneous rotational speed with large fluctuations, Mechanical Systems and Signal Processing 38/1 (2013) 96-102.
[4] Barszcz T, Randall RB. Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine, Mechanical Systems and Signal Processing 23 (2009) 1352-1365.
[5] Barszcz T, Czop P. Presentation of a virtual power plant environment and its application with combined first-principle and data-driven models intended for the diagnostics of a power plant, Pt. 1, Simulation 88/2 (2012) 139-166.
[6] Strączkiewicz M, Barszcz T, Jabłoński A. Detection and classification of alarm threshold violations in condition monitoring systems working in highly varying operational conditions, Journal of Physics. Conference Series 628 (2015), doi:10.1088/1742-6596/628/1/012087.
[7] Barszcz T, Strączkiewicz M. Novel intuitive hierarchical structures for condition monitoring system of wind turbines, Diagnostyka, 14/3 (2013) 53-60.
[8] Tabaszewski M. Optimization of a nearest neighbors classifier for diagnosis of condition of rolling bearings, Diagnostyka 15/1 (2014) 37-42.
[9] Jain AK, Duin R PW, Mao J. Statistical Pattern Recognition: A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence 22/1 (2000) 4-37.
[10] Vachtsevanos G, Lewis F, Roemer M, Hess A, Wu B. Intelligent Fault Diagnosis and Prognosis for Engineering Systems, John Wiley & Sons, New York, 2006.
[11] Van Der Heijden F, Duin R P W, de Ridder D, Tax D M J. Classification, Parameter Estimation and State Estimation, An Engineering Approach using MATLAB, 2004, John Wiley and Sons, Chichester, England.
[12] Worden K, Staszewski WJ, Hensman JJ. Natural computing for mechanical systems research: a tutorial overview, Mechanical Systems and Signal Processing, 25 (2011) 4-111.
[13] Theodoridis S, Koutroumbas K. Pattern Recognition, Fourth Edition, Academic Press, 2008.
[14] Mechefske CK, Mathew J. Fault detection and diagnosis in low speed rolling element bearing. Part II: the use of nearest neighbor classification, Mechanical Systems and Signal Processing 6 (1992) 309-316.
[15] Lei Y, Zuo MJ, Gear crack level identification based on weighted K nearest neighbor classification algorithm, Mechanical Systems and Signal Processing 23/5 (2009) 1535-1547.
[16] Casimir R, Boutleux E, Clerc G, Yahoui A. The use of features selection and nearest neighbors rule for faults diagnostic in induction motors, Engineering Applications of Artificial Intelligence 19 (2006) 169-177.
[17] McCullagh P, Nelder JA. Generalized Linear Models, New York: Chapman & Hall, 1990.
[18] Schlechtingen M, Ferreira I. Santos, Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection, Mechanical Systems and Signal Processing 25/5 (2011) 1849-1875.
[19] McLachlan GJ. Discriminant Analysis and Statistical Pattern Recognition, Wiley, Hoboken, NJ, 1992.
[20] Kim Y, Kim NY, Park SY, Lee DK, Lee JH. Classification and individualization of used engine oils using elemental composition and discriminant analysis, Forensic Science International 230/1-3 (2013) 58-67.
[21] Du Zh, Jin X. Multiple faults diagnosis for sensors in air handling unit using Fisher discriminant analysis, Energy Conversion and Management, 49/12 (2008) 3654-3665.
[22] Houshmand AA, Kannatey-Asibu Jr E. Statistical process control of acoustic emission for cutting tool monitoring, Mechanical Systems and Signal Processing 3/4 (1989) 405-424.
[23] Sánchez J, Perronnin F, Mensink T, Verbeek J. Image Classification with the Fisher Vector: Theory and Practice, International Journal of Computer Vision, 105/3 (2013) 222-245.
[24] Głowacz A, Głowacz A, Korohoda P. Recognition of Monochrome Thermal Images of Synchronous Motor with the Application of Binarization and Nearest Mean Classifier, Archives of Metallurgy and Materials, 59/1 (2014) 31-34.
[25] Parzen E. On the estimation of a probability density function and mode, The Annals of Mathematical Statistics., 33 (1962) 1065-1076.
[26] Chen X, Xu G. A self-adaptive alarm method for tool condition monitoring based on Parzen window estimation, Journal of Vibroengineering 15/3 (2013) 1537-1545.
[27] Qian Y, Mita A. Structural damage identification using Parzen-window approach and neural networks, Structural Control and Health Monitoring, 14/4 (2007) 576–590.
[28] Joo Er M, Wu S, Lu J, and Toh H L. Face recognition with radial basis function (RBF) neural networks, IEEE Transactions on Neural Networks, 13/3 (2002) 697-710.
[29] Paya B A, Esat I I, Badi M N M. Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor, Mechanical Systems and Signal Processing 11/5 (1997) 751-765.
[30] Baillie DC, Mathew J. A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings, Mechanical Systems and Signal Processing 10/1 (1996) 1-17.
[31] Rousseeuw P J, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics 20 (1987) 53-65.
[32] Thalamuthu A, Mukhopadhyay I, Zheng X, Tseng GC. Evaluation and comparison of gene clustering methods in microarray analysis, Bioinformatics, 22/19 (2006) 2405-2412.
[33] Hullermeier E. Fuzzy methods in machine learning and data mining: Status and prospects, Fuzzy Sets and Systems 156 (2005) 387-406.
[34] Liu TI, Singonahalli JH, Iyer NR. Detection of roller bearings defects using expert system and fuzzy logic, Mechanical Systems and Signal Processing, 10 (1996) 595-614.
[35] Strączkiewicz M, Czop P, Barszcz T. The use of a fuzzy logic approach for integration of vibration-based diagnostic features of rolling element

bearings, Journal of Vibroengineering, 17/4 (2015), 1760-1768.
[36] Jabłoński A, Barszcz T. Validation of vibration measurements for heavy duty machinery diagnostics, Mechanical Systems and Signal Processing 48/1 (2013) 248-263.
[37] Epps IK. An investigation into Vibrations Excited by Discrete Faults in Rolling Element Bearings, PhD Thesis, School of Mechanical Engineering, The University of Canterbury, Christchurch, New Zealand, 1991.
[38] Rawashdeh M, Ralescu A. Fuzzy cluster validity with generalized silhouettes, Proceedings of the 23rd Annual Midwest Artificial Intelligence and Cognitive Science Conference, Cincinnati, Ohio, USA, 2012, pp. 11–18.
[39] Bezdek J C. Cluster validity with fuzzy sets, Taylor & Francis Group 3/3 (1973) 58-73.
[40] Sheng Li A. The Improved Partition Coefficient, Procedia Engineering, 24 (2011) 534–538.
[41] Yeh JH, Joung FJ, Lin JC. CDV Index: A Validity Index for Better Clustering Quality Measurement, Journal of Computer and Communications 2/4 (2014) 163-171.
[42] Wang W, Zhang Y, On fuzzy cluster validity indices, Fuzzy Sets and Systems 158/19 (2007) 2095-2117.
[43] Dudoit S, Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset Genome Biology 3/7 (2002) 1–21.

Received 2016-03-31
Accepted 2016-05-25
Available online 2016-06-04

Marcin STRĄCZKIEWICZ MSc
Graduate of Cranfield University, United Kingdom (2011) and AGH University of Science and Technology in Krakow (2012). He specializes in signal and image processing and vibroacoustics (vibro-diagnostics). He is currently a PhD student researching the area of diagnostic features integration in the field condition monitoring systems of rotating machinery.

Piotr CZOP PhD DSc
He is Assistant Professor at AGH University of Science and Technology in Cracow. His main areas of research interest are modeling, identification and optimization of mechatronics systems.

Tomasz BARSZCZ DSc
Graduate of Electronic Eng at Politechnika Gdańska (1993). Received PhD in mechatronics in 1997 and DSc in 2009 (AGH). Has long experience of working with companies like ABB Zamech and ALSTOM Power. Author of over 200 papers and books on machinery diagnostics. Monitoring systems developed under his supervision were installed on several hundred machines worldwide.