Automatic calculation of thresholds for load dependent condition indicators by modelling of probability distribution functions – maintenance of gearboxes used in mining conveying system

Jacek Wodecki[1], Pawel Stefaniak[2], Anna Michalak[3], Agnieszka Wylomanska[4]

[1, 2, 3]KGHM Cuprum Ltd, R&D Centre, Sikorskiego 2-8, 53-659 Wroclaw, Poland
[4]Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland
[1]Corresponding author
E-mail: [1]*jwodecki@cuprum.wroc.pl*, [2]*pkstefaniak@cuprum.wroc.pl*, [3]*amichalak@cuprum.wroc.pl*, [4]*agnieszka.wylomanska@pwr.edu.pl*

Check for updates

Abstract. Limit values for gearbox vibration-based condition indicators are key to determine in order to be able to estimate moment when object is in a need of maintenance. Further decision making process usually might utilize simple if-then-else rule using established threshold values. If diagnostic data takes the values from the Gaussian distribution, finding the decision boundaries is not difficult. Simplistically, that comes down to standard pattern recognition technique for "good condition" and "bad condition" based on probability density functions (PDFs) of diagnostic data. This situation is becoming more and more complicated when distribution is not Gaussian. Such cases require to develop much more advanced analytically solution. In this paper, we present the case of belt conveyor's gearbox for which PDFs of diagnostic features overlap each other because of strong influence of time varying operating conditions on spectral features. New approach to automatic threshold recognition has been proposed based on modeling diagnostic features with Weibull distribution and using agglomerative clustering to distinguish classes of technical condition, which leads to determination of thresholds separating them.

Keywords: gearbox, Weibull distribution, condition monitoring, statistical analysis, clustering.

1. Introduction

Condition – Based Maintenance (CBM) is the subject of growing interest in the industry. Initially, such approach was used only in case of the most critical machines based on simple statistics of vibration signals like RMS, skewness or kurtosis. Over time, monitoring systems were developed with special dedication to specific groups of machines both in online as well as periodic acquisition form and were applied to more and more objects of machinery park. Today, CBM comes down to data fusion – dozens of variables are acquired simultaneously from each object in real time. For maintenance and management purposes the key issue is to propose such set of indicators calculated from measured time series which will enable complete and objective evaluation of objects in technical, economic and organization aspects as well as estimation of their residual life time. On industrial scale, it very often takes the form of so called big data solution [1]. Another challenge is related to established thresholds for these condition indicators what usually requires to use a data-driven approach. In this paper, we will present procedure for setting optimal thresholds for diagnostic features proposed to compound diagnosis of gearboxes used in mining conveying system. In the literature, this problem is well-known. [2] discussed that there exist some kind of limit values which can be identified based on statistics. In simple terms, distribution of vibration data takes the values from the Gaussian distribution. In such classical case decision making regarding the technical condition of object is simple. That requires only usage of standard pattern recognition technique for "good condition" and "bad condition" based on probability density functions of features. [3] pointed out the fact that in case of time varying

operating conditions, alarm threshold for spectral features should be determined using load susceptibility characteristics (LSCh) of monitored object which can be estimated by linear regression model, while load of conveyor has been deeper analyzed in [4]. Classical approach is not sufficient because the PDFs of diagnostic features overlap each other. [5-7] propose methodology for recognition of decision boundaries based on LSCh for large scale monitoring system including spatially distributed machinery park. They explained that measured features have significantly different probability function from Gaussian. The necessity of data modelling to determine alarm threshold has been shown by [8], where he considered threshold setting using Chebyshev's inequality, Weibull and Pareto distributions. On the other hand, arguments in favor of other distributions, especially heavy-tailed ones, has been made in [9].

Majority of aforementioned cases require to choose the most appropriate distribution of diagnostic data before limit value can be estimated. In [5] it has been shown a goodness-of-fit test which allows to choose the most adequate one. In this paper authors extend the previous work regarding technical condition assessment, however until now condition classes have been defined manually after visual inspection of empirical tail distribution (see [6, 7]). Presented methodology allows to define them automatically in a data-driven manner.

The paper is organized as follows: a short technical and operating aspects of machinery park will be described; then remarks and assumptions about automatic threshold finding will be formulated and the methodology will be proposed; industrial data and procedure to calculation of diagnostic features will be shown; finally, application of the method will be provided and results will be discussed.

1.1. Mining conveying system and proposed monitoring system

Investigated case study is belt conveyor transportation system using in one of the Polish underground mines of copper ore. Whole conveying system consists of over 80 technical objects combined in transportation network which has the total routes length of 50 km. Their reliability is critical – serious failure of single conveyor might stop operation of whole conveyor division in mine as well as cause of long-term breakdowns of mining processes in mining area or processing plant. One of the most critical conveyor components are gearboxes. As part of proactive tasks, a large-scale monitoring system for drive units has been developed. Because of number of technical objects and their spatial distribution, application of advanced approaches operating online is very difficult and too expensive. For this reason, portable solution has been proposed. Measurements have been performed by using three accelerometers placed orthogonally on the housings of gearboxes. Duration of measurements was equal to 60 seconds. Its sensors layer includes 3 accelerometers assembled for gearbox body and tachometric probe directed toward gearbox input shaft (see Fig. 1). Quick measurement delivers 3 vibration signals and tachometric signal.
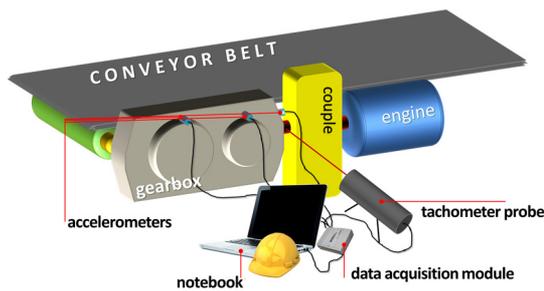
a) Diagram of the vibration data acquisition module with an exemplary arrangement of the measurement points

b) Real object in underground mine during measurement

Fig. 1. Object under investigation

2. Methodology

In this chapter methodology is described. After acquiring raw vibration signal, it is transformed into diagnostic feature carrying information about bearings' technical condition as described in Section 2.1. After that, for each measurement its empirical tail is calculated, and the outliers are rejected based on fitting Weibull distribution to the tails (see Section 2.2). In the next step, central points of tails are determined and clustered along with shaft rotational speed, that allows to determine classes of technical condition (see Section 2.3). Key aspects of the method are described in the following sections.

2.1. Diagnostic feature

Processing of raw vibration signals comes down to diagnostic features extraction from vibration signals and calculation of rotational speed of gearbox input shaft from tachometric data in order to identify operating condition. Developed feature extraction procedure is based on segmentation of raw signal dividing it into 60 equal segments without overlapping. Next, each single 1 sec. segment of the signal is transformed into frequency domain iteratively and all components are summed in given spectrum frequency bands (for shafts: 10-100 Hz, for gears: 100-3500 Hz, for bearings: 3500-10000 Hz). Finally, 60 sec. time series of three diagnostic features are extracted: DF1 (shafts condition), DF2 (gears condition) and DF3 (bearings condition). In this work attention is focused on the analysis of DF3 feature.

2.2. Fitting Weibull distribution to ECDF tails, MSE rejection

First step is to fit translated Weibull distribution to diagnostic features. Density function of translated Weibull distribution is defined as follows:

$$f(x) = \frac{\tau}{\beta} \left(\frac{x - m}{\beta} \right)^{\tau - 1} e^{-\left(\frac{x-m}{\beta} \right)^{\tau}}, \qquad x \geq m, \tag{1}$$

where $m \in \mathbb{R}$ is a shift parameter, $\beta > 0$ is scale parameter and $\tau > 0$ is a shape parameter [10]. In the field of condition monitoring, Weibull distribution found interesting applications e.g. in time-to-failure modelling [11-13]. The idea of estimation the parameters is described in [6]. Next, we analyze the quality of fit by calculate mean square error (MSE) between empirical tail of diagnostic features and theoretical one given by:

$$T(x) = P(X > x) = e^{-\left(\frac{x-m}{\beta} \right)^{\tau}}, \qquad x \geq m. \tag{2}$$

We choose such diagnostic features for which the calculated MSEs exceed the given threshold and we reject it.

2.3. Central point and clustering

Core idea of automatic distinction of wear levels based on distribution tails incorporates tails clustering based on their distribution along the diagnostic feature value. Since tails take values between 0 and 1, it is reasonable to estimate tail location as its central point, being the argument of tail value equal to 0.5 (see Fig. 5).

When locations of tails are determined, they are clustered into three clusters using agglomerative clustering algorithm [14, 15]. The number of clusters is determined by three expected condition states: healthy, warning and alarm (approaching failure). Agglomerative hierarchical clustering is a bottom-up clustering method where clusters have sub-clusters, which in turn have sub-clusters, etc. The classic example of this is species taxonomy. Gene expression

data might also exhibit this hierarchical quality (e.g. neurotransmitter gene families). Agglomerative hierarchical clustering starts with every single object (gene or sample) in a single cluster. Then, in each successive iteration, it agglomerates (merges) the closest pair of clusters by satisfying some similarity criteria, until all of the data is in one cluster (see Fig. 2).

An alternative top-down hierarchical clustering method is less commonly used. It works in a similar way to agglomerative clustering but in the opposite direction. This method starts with a single cluster containing all objects, and then successively splits resulting clusters until only clusters of individual objects remain.
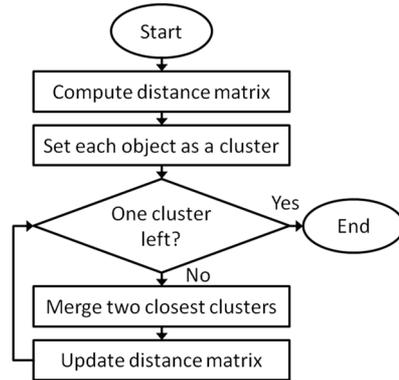

Fig. 2. Flowchart of agglomerative hierarchical algorithm

3. Results

Firstly, diagnostic feature of interest has been obtained from raw vibration signal according to the description in Section 2.1 (see Fig. 3).
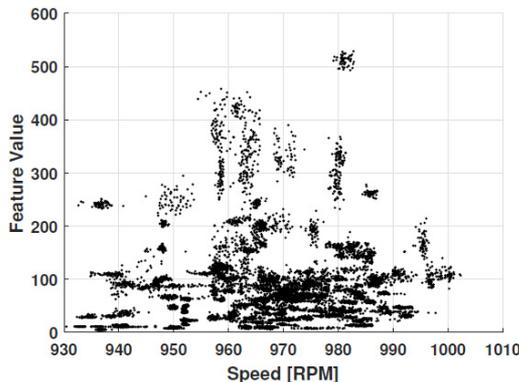

Fig. 3. Raw diagnostic feature data; dataset consists of 155 measurements, each represented as a point cloud of 60 points corresponding to 60 seconds of a measurement

In the next step a tail of ECDF was calculated for DF3 points of each measurement. Weibull distribution was fitted to each tail, and tails with the greatest fit errors have been disregarded (see Fig. 4). For remaining tails, middle point value was calculated and complete set of middle points was provided to the clustering algorithm (see Fig. 5).

As a result, tails have been classified into three clusters indicating different health states (see Fig. 6(b)). This information was then translated into the two-dimensional plane of DF3 values vs. RPM, which has been divided into three sectors (see Fig. 6(a)). Sectors' edges were determined as weighted means between linear fits of two adjacent clusters, where weights were the variance values of DF3 coordinates of clusters' members. For each pair of adjacent clusters, weights were

normalized by their sum.

As a result, technical condition evaluation map has been constructed. It can be used as a condition evaluation basis for future measurements to be obtained, upon which the map can be dynamically updated.
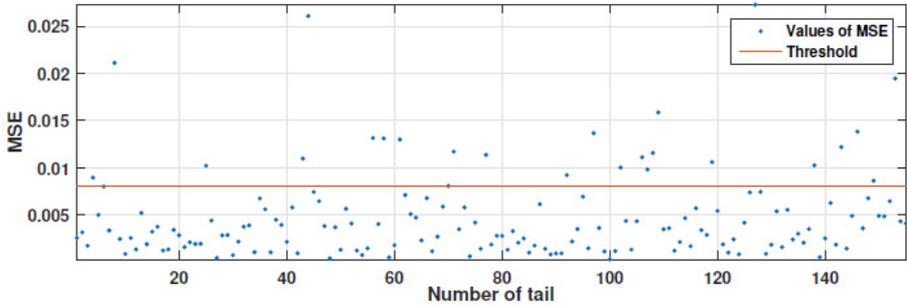
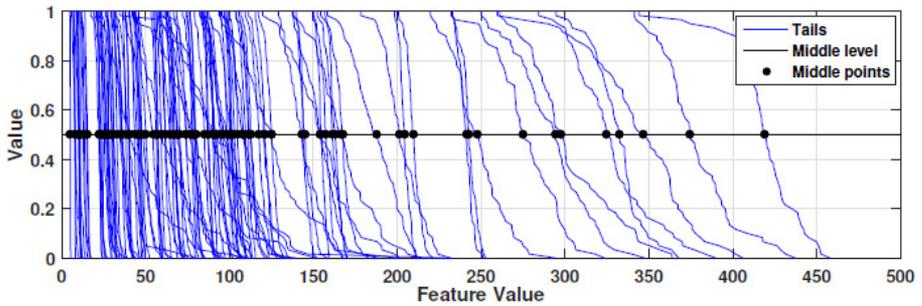

Fig. 4. MSE values for tail fitting

Fig. 5. ECDF tails with marked locations on half-magnitude level

a)   b)
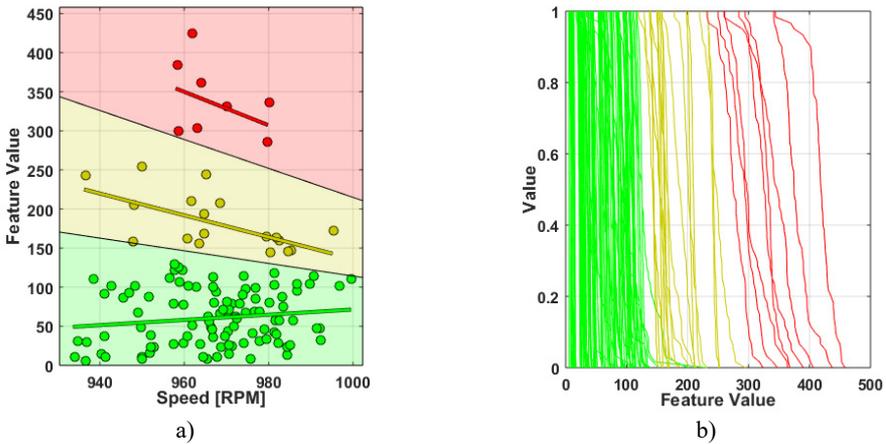
Fig. 6. Results of condition regimes classification. a) Measurements as a diagnostic feature in a function of rotational speed. Each dot represents a point cloud of a single measurement and is its centroid. Linear regression functions for each cluster allow to determine the borderlines between adjacent clusters. b) Empirical tails of measurements

4. Conclusions

The paper concerns significant issue of condition monitoring related to making diagnosis based on vibration signal, namely identification of thresholds for diagnostic features. The authors consider gearboxes used in mining conveying system for which strong influence of time varying

operating conditions on spectral features cause that well-known methods are ineffective. Proposed approach is based on statistical modeling of diagnostic data set from a single measurement. After disregarding outliers based on goodness of tails fit to Weibull distribution, dataset is clustered to distinguish separate classes of technical condition. Statistical analysis allows to determine boundaries of those classes after the separation. Results perfectly correspond with previously obtained ones that utilized manual classification and thresholding.

Acknowledgements

References

[1]   **Bartkowiak A., Zimroz R.** Dimensionality reduction via variables selection-linear and nonlinear approaches with application to vibration-based condition monitoring of planetary gearbox. Applied Acoustics, Vol. 77, 2014, p. 169-177.

[2]   **Cempel C.** Limit value in the practice of machine vibration diagnostics. Mechanical Systems and Signal Processing, Vol. 4, 1990, p. 483-493.

[3]   **Coles S., Bawa J., Trenner L., Dorazio P.** An Introduction to Statistical Modeling of Extreme Values. Springer, 2001.

[4]   **Jablonski A., Barszcz T., Bielecka M., Breuhaus P.** Modeling of probability distribution functions for automatic threshold calculation in condition monitoring systems. Measurement, Vol. 46, 2013, p. 727-738.

[5]   **Kaufman L., Rousseeuw P. J.** Finding Groups in Data: an Introduction to Cluster Analysis. John Wiley & Sons, 2009.

[6]   **Król R., Kisielewski W., Kaszuba D., Gładysiewicz L.** Testing belt conveyor resistance to motion in underground mine conditions. International Journal of Mining, Reclamation and Environment, Vol. 31, 2017, p. 78-90.

[7]   **Marhadi K., Hilmisson R.** Simple and effective technique for early detection of rolling element bearing fault: A case study in wind turbine application. International Congress of Condition Monitoring and Diagnostic Engineering Management, 2013, p. 94-97.

[8]   **Rokach L., Maimon O.** Clustering Methods. Data Mining and Knowledge Discovery Handbook, Springer, 2005, p. 321-352.

[9]   **Sikorska J., Hodkiewicz M., Ma L.** Prognostic modelling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing, Vol. 25, 2011, p. 1803-1836.

[10]   **Stefaniak P. K., Wyłomańska A., Obuchowski J., Zimroz R.** Procedures for decision thresholds finding in maintenance management of belt conveyor system-statistical modeling of diagnostic data. Proceedings of the 12th International Symposium Continuous Surface Mining-Aachen, 2015, p. 391-402.

[11]   **Stefaniak P. K., Wyłomańska A., Zimroz R., Bartelmus W., Hardygóra M.** Diagnostic features modeling for decision boundaries calculation for maintenance of gearboxes used in belt conveyor system. Advances in Condition Monitoring of Machinery in Non-Stationary Operations, 2016, p. 251-263.

[12]   **Wang W.** A model to predict the residual life of rolling element bearings given monitored condition information to date. IMA Journal of Management Mathematics, Vol. 13, 2002, p. 3-16.

[13]   **Wyłomańska A., Żak G., Kruczek P., Zimroz R.** Application of tempered stable distribution for selection of optimal frequency band in gearbox local damage detection. Applied Acoustics, 2016, https://doi.org/10.1016/j.apacoust.2016.11.008 .

[14]   **Zhou D.** Transformer lifetime modelling based on condition monitoring data. International Journal of Advances in Engineering and Technology, Vol. 6, 2013, p. 613-619.

[15]   **Zimroz R., Bartelmus W., Barszcz T., Urbanek J.** Diagnostics of bearings in presence of strong operating conditions non-stationarity-a procedure of load-dependent features processing with application to wind turbine bearings. Mechanical Systems and Signal Processing, Vol. 46, 2014, p. 16-27.