Investigation of defects in roll contacts of machine elements with Acoustic Emission and Unsupervised Machine Learning

J Hillenbrand\textsuperscript{1*}, J Detroy\textsuperscript{1} and J Fleischer\textsuperscript{1}

\textsuperscript{1} wbk Institute of Production Science, Gotthard-Franz-Straße 5, 76131 Karlsruhe, Germany

*Corresponding author: jonas.hillenbrand@kit.edu

Abstract: In the age of Industry 4.0 and IIoT machines are becoming increasingly connected enabling continuous monitoring. A variety of information from machines and installed sensors is used to develop condition monitoring solutions. These systems are used to prevent premature failures and the follow-up costs due to machine downtime associated with them. Recent research in this area applies supervised machine learning, extracting features from captured signals and training classifiers. Supervised learning approaches require large amounts of labeled data, whose generation is time consuming and requires domain knowledge. For this reason, an unsupervised learning approach is being used in this work to distinguish between different defect and operation states of axial ball bearings. Within the scope of this work, acoustic emission (AE) measurements in the ultrasonic range are recorded and evaluated. Artificial defects are seeded in the rolling contact of axial bearings. From the AE signals a selection of state-of-the-art features is extracted. Then, the Laplacian Score, an unsupervised filter algorithm, is used to select the most significant features. Subsequently, the DBSCAN clustering algorithm is used to draw conclusions about the existing damage.

Keywords: Acoustic emission, Clustering, Bearing damage, DBSCAN

1. Introduction

Unsupervised learning algorithms gain more and more popularity with the proliferation of massive amounts of data, especially from unlabeled sources [1]. In particular, this popularity stems from the way unsupervised learning methods can be applied, where classical supervised learning techniques cannot be used. For example, if no dataset with ground truth in form of corresponding labels exists, supervised learning and its classification or regression capabilities cannot be applied. Problems where only the given input, but not the corresponding target values are given, are so called unsupervised problems. Goals of unsupervised learning are [2]:

- discover patterns or groups of data with similar structure (Clustering)
- determine data distribution
- reduce the dimensionality of the problem without relying on user input for labelling.

But applying unsupervised learning also comes with challenges, as it has not seen so much attention as the more prominent techniques in supervised learning [3]. Common challenges are tuning of model hyperparameters and selecting suitable features to describe the data properly.
By default, these features are handcrafted transformations applied to the raw data, they are domain specific and their design and selection requires human expertise. These facts limit the out-of-box application of these approaches and decrease scalability [4]. Still, especially in the context of condition monitoring (CM) and predictive maintenance, an abundance of these feature definitions exists for specific applications, such as monitoring machines and their components. While recent approaches move towards automatic feature learning using deep learning [5,6] and neglect the interpretability of the extracted features, we suggest using unsupervised approaches to automate the feature selection process in domains, where fruitful features are already available and known.

This paper introduces an approach employing unsupervised feature selection for acoustic emission (AE) monitoring of axial bearings. To learn the features without the use of labels, we apply the Laplace-score on a predefined set of features to identify the most promising ones for the online monitoring process. To validate our approach, we conducted experiments on a testbench in four different scenarios: normal condition, varying amounts of lubricant, contamination in the rolling contact and artificial pittings using an electric engraver. The measurements were captured with an acoustic emission sensor at a sampling rate of 2 MHz.

2. Machine Learning based monitoring of bearings
Traditionally, CM systems that use machine learning in a bearing context use handcrafted features, that have been shown to work effectively with supervised classifiers to detect failure states. In order to further automate the decision process supervised feature selection algorithms have been designed and applied [7,8]. However, these methods require labeled data.

A recent research trend is using deep learning for feature extraction. Here autoencoder networks are used to compress and subsequently reconstruct the training data with minimal loss. The autoencoders are trained using normal state data. This results in high reconstruction errors when new unknown states are observed [5,6]. Similarly, [9] tried to detect and classify AE events for wear monitoring in sliding bearings. To accomplish this a two-stage process consisting of an autoencoder for anomaly detection and a convolutional neural network for classification were used. Labeled AE signals were used to train the convolutional neural network. [10] proposes a semi-supervised learning scheme for bearing anomaly detection using a variational autoencoder. Even though this method can provide an accuracy enhancement over both supervised and unsupervised methods, it still requires a small amount of correctly labeled data.

Although, these methods move towards being fully unsupervised, they rely on deep neural network architectures, where interpretability of extracted features is lost. Previous knowledge about useful features is not applied. Using already established features from CM applications ensures a better interpretability of the monitored object. Since a meaningful catalog of features for CM of rotating machinery already exists, we focus our work on the task of automatically selecting those features without prior knowledge of labels or system behavior.

3. Experimental Setup
Our experiments were conducted on an axial bearing testbench. The setup provides defined load and speed operation of the axial bearing. It consists of a rotary and a stationary part. The lower rotary part is driven by a stepper motor for position or speed-controlled operation, whereas the upper stationary part provides the axial force by means of a pneumatic cylinder (see figure 1). Loads up to 12 kN and speeds of 2000 U/min can be realized in the current setup. In this work NTN Type 51108 thrust bearings were used. These bearings have separable grooves and therefore simplify the visual inspection of the bearing’s condition during testing.

For the acoustic emission recordings, a Mistras wideband differential sensor and corresponding preamplifier with a frequency range of 125-1000 kHz were used. The sensor is located on top of the upper bearing ring in the stationary part of the testbench.
In order to simulate different operating conditions and wear scenarios, we induced different artificial defects on the bearing grooves and superposed different operating conditions (in terms of speed and load). Table 1 shows the different scenarios that were investigated on the testbench. For the inspection of abrasion wear, the lubricant was artificially contaminated (a similar procedure was used in [11]). As contaminant lubricant mixed with laser dust was used (mixing ratio of 8 : 1). The grain size of the particles used was not further specified with smaller than 10 µm. To simulate increased adhesion and tribo-chemical wear, tests with varying amounts of uncontaminated lubricant were carried out. According to [12], limited grease leads to more direct contacts between the rollers and the races, which results in stronger and more unstable AE signals. Fatigue was simulated in form of pitting defects, introduced on the groove surface, similar to experiments carried out in [13]. Here, we used an electrical engraver to apply artificial pitting damage. In order to differentiate the various artificial damages under different operating conditions, the experiments were recorded at different rotational speeds and different loads.

| Name                        | Type                        |
|-----------------------------|-----------------------------|
| contaminated lubricant [11] | abrasion wear               |
| varying amount of lubricant | adhesion and tribochemical wear |
| pittings on groove [13]     | fatigue                     |
| speed variation             | changing operating conditions |
| load variation              | changing operating conditions |

4. Unsupervised Monitoring workflow

The workflow we propose to process the acoustic emission data is displayed in figure 2. It consists of three mayor components. Firstly, a feature extraction step, where the features presented in table 2 are computed for the raw sensor data. In a second step an unsupervised feature selection algorithm is used to select the most significant features. This step can be done once at the start or online in certain intervals,
if the initially selected features fail to characterize new states. This has the advantage that the change of the best features can be monitored over time. However, each feature would have to be calculated at every timestep, which requires more computing power. The selected features are then used to cluster the raw data with DBSCAN [14]. The generated cluster index can subsequently be used to warn about new states encountered. Except the feature catalog, this procedure does not require any input from the user since the features and model hyperparameters are determined automatically. Therefore, it is possible to deploy this approach in a system with no a-priori knowledge.

![Diagram](image)

**Figure 2.** Unsupervised monitoring workflow.

| Time domain features | Frequency domain features |
|----------------------|--------------------------|
| **RMS:** $\text{RMS} = \frac{1}{n} \sum_{i=1}^{n} x_i^2$ | **Mean Frequency:** $MF = \frac{1}{N} \sum_{i=1}^{N} X_i$ |
| **Zeroscrossings:** $\text{ZC} = \frac{1}{\text{len}(x)} \cdot \text{SampleRate} \sum_{t_x=0} t_x$ | **Frequency mean-squared:** $\text{FRMS} = \frac{\sum_{i=1}^{N} f_i^2 \cdot X_i}{\sum_{i=1}^{N} X_i}$ |
| **Formfactor:** $\text{Ff} = \frac{\text{RMS}}{1/n \sum_{i=1}^{n} |x_i|}$ | **Frequency center:** $\text{FC} = \frac{\sum_{i=1}^{N} f_i \cdot X_i}{\sum_{i=1}^{N} X_i}$ |
| **Crestfactor:** $\text{Cr} = \frac{\max\{x_i\}}{\text{RMS}}$ | **Skewness:** $\text{Sk} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s}\right)^3$ |
| **Marginfactor:** $\text{Mf} = \frac{\max\{x_i\}}{\left(\frac{1}{n} \sum_{i=1}^{n} |x_i|\right)^{1/2}}$ | **Energy:** $E = \int |x(t)|^2 dt$ |
| **Impulsfactor:** $\text{If} = \frac{\max\{x_i\}}{1/n \sum_{i=1}^{n} |x_i|}$ | **Binned Entropy:** $\text{BE} = -\sum_{i=1}^{n} p_i \ln p_i$ |

With $x_i$ being the $i$th sample in the raw data, $p_i$ the probability that the measurement is in the $i$th discrete interval and $X_i$ being the calculated observed frequencies.

### 4.1. Feature Design

In order to limit the complexity of the models used and their calculation time, feature selection algorithms are used to choose the features best suited to the problem. This is particularly important,
since there are irrelevant features, which do not provide a significant increase to the model quality or redundant features, which contain the same information as other features.

In this work features from [15] are adopted and extended. A list of the features used is depicted in table 2. These features show good correlations with degradation mechanisms in bearings and have been used in several earlier works [16–18].

4.2. Laplacian score
Unsupervised feature selection can be solved by using the Laplacian score [19]. It ranks the features regarding their locality preserving power and is calculated as follows:

In a first step a nearest neighbor graph is constructed with m nodes, where every data point is one such node. If two nodes are connected the entry of the weight matrix S is calculated with the following equation:

\[ S_{ij} = e^{-\frac{||x_i - x_j||^2}{\epsilon}}, \text{ if nodes } i \text{ and } j \text{ connected} \]

\[ 0, \text{ otherwise} \] (1)

In a next step a transformed feature vector \( \tilde{f}_r \) is calculated.

\[ \tilde{f}_r = f_r - \frac{f_r^T \mathbf{D} \mathbf{1}}{\mathbf{1}^T \mathbf{D} \mathbf{1}} \]

(2)

with \( \mathbf{1} = [1 \ldots 1]^T \) and \( f_r = [f_{r1} \ldots f_{ri} \ldots f_{rm}]^T \), matrix L is called graph Laplacian \( L = \mathbf{D} - \mathbf{S} \), where \( \mathbf{D} = \text{diag}(\mathbf{S} \mathbf{1}) \). Finally, the Laplacian score for a feature \( L_r \) is calculated.

\[ L_r = \frac{f_r^T \tilde{L} f_r}{f_r^T \mathbf{D} f_r} \] (3)

The Laplacian score relies on the idea that similar datapoints belong to the same cluster. A feature is assigned a high weight, if close datapoints have similar feature values. A feature is ranked as unimportant if datapoints are far apart, but the feature value is similar. The Laplacian score is a univariate filter approach, therefore it must be extended in order to work for multivariate feature spaces. We solve this issue by combining the Laplacian score with a binary correlation comparison. In a first step the Laplacian score is calculated for all features. In the next step the feature with the highest score is selected and compared with all remaining features regarding the correlation. Features with a correlation coefficient larger than a predefined threshold (here \( \tau = 0.7 \)) are removed. Then the steps are repeated until either the specified number of features has been selected or no more features can be selected because there are no uncorrelated features left.

5. Experimental Results
Running this algorithm on the recorded AE-data the following results are obtained: The Mean frequency was identified as the most significant feature, followed by the Root-Mean-Square and Formfactor. The results are presented in figure 3.

Using the three most significant features identified by the Laplacian score, the feature space of the different measurements can be graphed. Figure 4 shows the obtained feature space.

We observed that the selected features are sensitive to the damages introduced, as the different scenarios can be properly separated by boundaries in the feature space. In case of artificial pittings, the change in features between normal and degraded state is weak. This can stem from low depths of the introduced damages.

All features are sensitive to the rolling speed, whereas they are less dependent from the load applied. Contemplating the feature space displayed in figure 4a), we observed well distinguishable states for single speeds (graphic on the left). If, however too many states are present at the same time, these overlap and no suitable spatial separation can be achieved in the feature space (see Figure 4b)).
Figure 3. (a) Feature importance ranked by Laplacian score. (b) Covariance matrix between selected features.

Figure 4. (a) Shows the feature space with all measurements at 200 rpm. (b) Shows the feature space of all measurements.

In many applications the rotary speed of a machine is either known or captured by control systems. If that is the case, the speed can be used to increase the accuracy of the unsupervised learning approach. For instance, the speed can be used as an additional feature in clustering. Alternatively, the speed can be binned and only datapoints belonging to the same speed bin are clustered together. This results in multiple cluster results with similar rotational speeds and should therefore eliminate the large variance caused by speed differences.

The recorded AE-Signals were used to create a data stream of continuous measurement values. Subsequently, a sliding window was used to calculate the three, previously identified, most significant features: Mean frequency, Root-Mean-Square and Formfactor for each timestep. This set of features is then clustered using DBSCAN to differentiate the failure modes. The hyperparameters of the DBSCAN algorithm are determined automatically using the data distribution. The epsilon value $\epsilon$ is computed by finding the first valley of the sorted k-distance graph as suggested in the original paper introducing DBSCAN [14]. To determine the minPoints hyperparameter, the maximum of either two times the number of features used or the logarithm of the number of data points is used, so that minPoints also increases with an increasing number of data points, as suggested in [20].
In order to analyze the collected data with the method described above, exemplary files were selected. As a first example, files were read in at different rotational speeds. The load was kept constant at $6 \text{kN}$. The measurements were all taken under normal condition. The clustering result is shown in Figure 5. We observed that the DBSCAN algorithm is capable of identifying the seven different rotational speeds. Only at the highest rotational speed a single state is falsely split into two separate clusters.

![DBSCAN Clustering Result](image)

**Figure 5.** (a) Clustering result with different rotational speeds plotted in the 2D feature space. Each cluster represents a different speed. (b) Clustering purity for test with rotational speeds and different damage types.

In a second example the rotational speed and load was kept constant, but different damage scenarios were used (see Figure 4a). The algorithm can correctly differentiate most introduced damages but fails to differentiate the artificial pitting damage from baseline measurements. This aligns with the observation made in chapter 5, that measurements with artificial pittings have very similar features to a damage free state due to small damage dimensions.

6. Conclusion and Outlook
With our approach, we enabled unsupervised selection of features without prior knowledge. The features selected, then performed well in the following cluster operation. Different operating conditions and degradation states were separated and found by the cluster algorithm. If however, too many different states prevail, the DBSCAN algorithm fails to cluster them apart. Adding more features, e.g. the rotary speed itself, can help to solve this issue. In addition, the method fails for data, whose clusters do not have the same density, thus violating the assumptions of DBSCAN.

Hence, we want to investigate the workflow on other datasets and experiments in the context of CM and look into other cluster algorithms to deal with the shortcomings of DBSCAN.

So far, only the detection of state changes was considered. For a more profound, still unsupervised, decision of normal or faulty states, we want to investigate the information contained in the temporal evolution of clusters. Enabling the method to detect slow transitions into new cluster states.

Acknowledgements
This research work is funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project 388141462.

References
[1] Celebi M and EandAydin K 2016 *Unsupervised Learning Algorithms* (Springer Int. Publishing)
[2] Bishop C M 2009 Pattern recognition and machine learning (New York: Springer)
[3] Seif G 2019 An easy introduction to unsupervised learning with 4 basic techniques (https://towardsdatascience.com/an-easy-introduction-to-unsupervised-learning-with-4-basic-techniques-da7fb0c3adf) accessed 17 February 2021
[4] Michau G, Hu Y, Palmé T and Fink O 2020 Feature learning for fault detection in high dimensional condition monitoring signals Journal of Risk and Reliability 234 pp 104–115
[5] Yoon A S, Lee T, Lim Y, Jung D, Kang P, Kim D and Choi Y 2017 Semi-supervised Learning with Deep Generative Models for Asset Failure Prediction KDD Workshop on Machine Learning for Prognostics and Health Management 17
[6] Chen Z and Li W 2017 Multisensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network IEEE Transactions on Instrumentation and Measurement 66 pp 1693–1702
[7] Cerrada M, Li C, Sánchez R V, Pacheco F, Cabrera D and de Oliveira J V 2018 A fuzzy transition based approach for fault severity prediction in helical gearboxes Fuzzy Sets and Systems 337 pp 52–73
[8] Sugumaran V, Muralidharan V and Ramachandran K I 2007 Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing Mechanical systems and signal processing 21 (2) pp 930–942
[9] König F, Sous C, Chaib A O and Jacobs G 2021 Machine learning based anomaly detection and classification of acoustic emission events for wear monitoring in sliding bearing systems Tribology International 155 p 106811
[10] Zhang S, Ye , Wang B and Habetler T G 2019 Semi-Supervised Learning of Bearing Anomaly Detection via Deep Variational Autoencoders arXiv preprint 1912.01096
[11] Verl A, Heisel U, Walther M and Maier D 2009 Sensorless automated condition monitoring for the control of the predictive maintenance of machine tools CIRP Annals 58 pp 375–378
[12] Wang K, Liu X, Wu X and Zhu Z 2019 Monitoring the lack of grease condition of rolling bearing using acoustic emission Int. J. of Modelling, Identification and Control 31 (1) pp 94–102
[13] Lessmeier C, Kimotho J K, Zimmer D and Sextro W 2016 Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification European Conference of the Prognostics and Health Management Society vol 3 (1)
[14] Ester M, Kriegel H P, Sander J and Xu X 1996 A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96) vol 96 (34) pp 226-231
[15] Kimotho J K and Sextro W 2014 An approach for feature extraction and selection from non-trending data for machinery prognosis European Conference of the Prognostics and Health Management Society vol 2 (1)
[16] Zhang J M, Huang Y H and Wang S M 2012 Fault Diagnosis Method for the Rolling Bearing Based on Information Fusion and BP Neural Network Advanced Materials Research 538-541 pp 1956–1961
[17] Galar D, Kumar U, Lee J and Zhao W 2012 Remaining Useful Life Estimation using Time Trajectory Tracking and Support Vector Machines Journal of Physics: Conference Series 364 (1) pp 012063
[18] Di Maio F, Hu J, Tse P, Pecht M, Tsui K and Zio E 2012 Ensemble-approaches for clustering health status of oil sand pumps Expert Systems with Applications 39 (5) pp 4847–4859
[19] Xiaofei He, Deng Cai, Partha Niyogi 2006 Laplacian Score for Feature Selection Advances in Neural Information Processing Systems 18 pp 507–514
[20] Schubert E, Sander J, Ester M, Kriegel H P and Xu X 2017 DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN ACM Transactions on Database Systems (TODS) 42 (3) pp 1–21

8