Manuscript version
DOI original paper: 10.1007/978-3-031-07322-9_32

A new unsupervised learning approach for CWRU bearing state distinction

Xiao Wei*, Tingsheng Lee, and Dirk Söffker

Chair of Dynamics and Control, University of Duisburg-Essen. Duisburg
xiao.wei@uni-due.de

Abstract. As one of the most relevant components in rotary machinery, ball bearings play an important role in diverse areas. To research bearing health state and remaining useful lifetime, several datasets have been developed. Among these datasets, Case Western Reserve University (CWRU) dataset is the most commonly used for bearing diagnosis. A large variety of approaches are applied on CWRU dataset and generating good even the tendency of perfect results. However, most of these approaches are based on supervised learning approaches and focus on classification of bearing faults. In this contribution, in difference to well-known existing approaches, an unsupervised approach combining autoencoder with k-mean is applied on the CWRU dataset. Firstly, the original data are segmented into proper parts. Segments in time domain are transformed to time-frequency domain by adjusting the window length and window function using Short-Time Fourier Transform (STFT), and an associated spectrogram is generated. Spectrogram features are extracted using autoencoder and clustered using K-mean. Various metrics are used to evaluate the performance of the proposed approach. All metrics values demonstrate that this approach could distinguish CWRU bearing from fault-free state to faulty state. As a new result, the requirement of related training datasets of the other approach is – for fault detection – no longer necessary in the future.

Keywords: CWRU, STFT, autoencoder, k-mean

1 Introduction

As a tool which provide rotational and linear movements of the device, ball bearing plays an irreplaceable role in diverse areas not only in industry but also in fields of daily use like in the automobile sector. Once a bearing (or component in it) fails, other adjacent components and machines are effected in behavior up to failure. Several surveys regarding the likelihood of induction machine failure conducted by the IEEE Industry Application Society and the Japan Electrical Manufactures' association reveal that bearing fault is the most common fault type and is responsible for 30 to 40 % of all machine failures [1]. To avoid unplanned maintenance shutdowns and unsafe working conditions, detecting and identifying defects in bearing at an early stage is significant for rotary machinery. Several benchmark datasets are developed to evaluate bearings health state and forecast remaining useful lifetime. Among these datasets, Case Western Reserve University (CWRU) bearing dataset is the most cited one used to develop and verify bearing fault detection and diagnosis approaches. Approaches and their performance developed based on CWRU bearing dataset has been reviewed in [2]. Although almost all these approaches performed well with the tendency of perfect, most of these approaches are supervised learning. A few approaches apply unsupervised learning for distinguish CWRU bearing states. Even if these approaches use unsupervised learning, metrics for evaluating approaches are still accuracy, precision, recall, and specificity which usually used to evaluate supervised learning approaches.

Autoencoder is an unsupervised learning technique which leverage neural networks for the task of representation learning [3]. The aim of an autoencoder is to learn a lower-dimensional representation (encoding) for higher-dimensional data, typically for dimensionality reduction, by training the network to capture the most important parts of the input image [4]. A typical autoencoder consists of 3 parts: encoder, bottleneck, and decoder. Encoder is a set of convolutional blocks followed by pooling modules that compress the input to the module into a compact section. The bottleneck is a module that contains the compressed knowledge representations. Decoder is a module that decompresses the knowledge representations and reconstructs the data back from its encoded form. Autoencoders could be divided into different types, such as undercomplete autoencoders, sparse autoencoders, contractive autoencoders, denoising autoencoders, and variational autoencoders [5].

K-means clustering is a popular unsupervised machine learning algorithm. The objective of K-mean is grouping similar data points together and discover underlying patterns. To achieve this objective, a fixed number (K) of cluster in a dataset is important. For mining data, the k-means algorithm starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids [6]. Stopping criteria for k-means clustering are: centroids of newly formed cluster do not change; points remain in the same cluster, and maximum number of iterations are reached [7].

Evaluating the performance of an unsupervised learning algorithm is different with respect to the used metrics to supervised learning algorithms. Unsupervised learning evaluation metrics depend on the class of unsupervised algorithms such as dimensionality reduction algorithms, clustering algorithms, and generative models [8]. Clustering algorithms are evaluated based on some similarity or dissimilarity measure such as the distance between cluster points [9]. If the clustering algorithm separates dissimilar observations apart and similar observations together, it performs well. According to whether labels are available, two classes statistical techniques to validate results for cluster learning: internal validation and external validation. Internal metrics define the quality of a clustering algorithm without external labels by using the idea of cohesion and separation while external metrics can be understood as an equivalent the evaluation metrics of supervised algorithms [10]. Among external metrics, purity is a simple and transparent evaluation measure. Normalized mutual information (NMI) can be information-theoretically interpreted. The rand index (RI) penalizes both false positive and false negative decisions during clustering while adjusted rand index (ARI) measures the similarity of two assignments ignoring permutations. F-measure supports differential weighting of these two types of errors.

According to [2], most approaches applied to CWRU bearing dataset are based on supervised learning to classify different bearing states. According to authors knowledge, at present only a few approaches apply unsupervised learning to cluster bearing states. Besides, for unsupervised learning approaches applied to CWRU bearing dataset, metrics applied to evaluate their performance are accuracy, recall, and precision, which are usually applied for evaluating supervised learning approach. In this paper, unlike other approaches, a new unsupervised approach combing autoencoder and K-mean is applied to cluster different bearing states in CWRU dataset. Furthermore, metrics applied for proposed approach are external metrics like purity, RI, ARI, and NMI which are also different with other approaches. The first step of proposed approach is data selection - suitable measurements are selected among whole dataset. Then, measurements are divided into segments. Afterwards, segments in time domain are transformed from time domain to time-frequency domain by Short-Time Fourier Transform (STFT) and spectrograms are acquired. Finally, features of spectrograms are extracted by autoencoder and clustered by K-mean. The performance of the proposed approach is evaluated by external metrics. All results show that the proposed approach could distinguish faulty and fault-free states well.

The structure of this contribution is arranged as follows: CWRU bearing dataset and test rig are shown in Section 2. In Section 3 the proposed approach will be presented in detail. Results of the proposed approach are presented in Section 4. Finally, conclusions from calculating process are drawn in Section 5.

2 CWRU bearing dataset

Case Western Reserve University (CWRU) bearing data center provides access to ball bearing test data for normal and faulty bearings. As shown in Figure 1, the test stand consists of a 2 hp reliance electric motor, a torque transducer/encoder, a dynamometer, and control electronics [11]. Motor bearings are seeded with faults using electro-discharge machining (EMD). Faults ranging from 0.007 to 0.040 inches in diameter are introduced separately at the inner raceway, ball and outer raceway. Faulted bearings are installed into the test motor and vibration data was recorded for motor loads of 0 to 3 horsepower. Vibration data are collected using accelerometers, which are placed at the 12 o'clock position at both the drive and fan end of the motor housing. During some experiments, an accelerometer is attached to the motor supporting base plate as well. Data are collected for normal bearings, single-point drive end and fan end defects at the samples rate of 12 kHz. For drive end bearing, vibration data are also collected at 48 kHz.

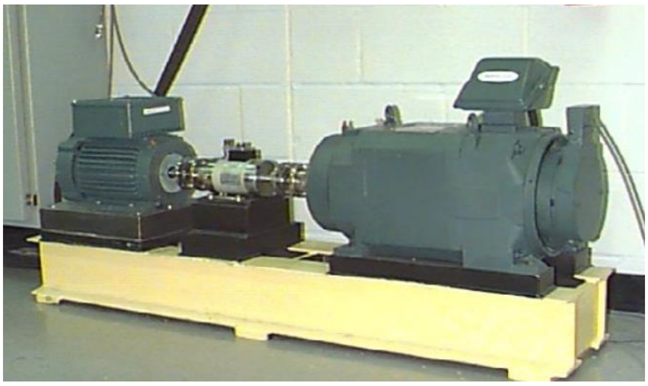

Fig. 1. Test rig of CWRU bearing dataset [11]

3 Proposed approach

Measuring time for each condition in CWRU bearing dataset is from 10 seconds to 40 seconds, besides, the data are collected at 12,000 or 48,000 samples per second. Therefore, it is impossible to calculate all data. To overcome this problem, the first step is to select suitable data. If one measurement is considered as one sample, samples are too little to be trained. To increase sample quality, measurements are divided into different segments. Segments in time domain are transformed to time-frequency domain by STFT, and associated spectrograms are generated. Then, spectrogram features are extracted using autoencoder and clustered using K-mean. Flowchart of proposed approach is presented in Figure 2.

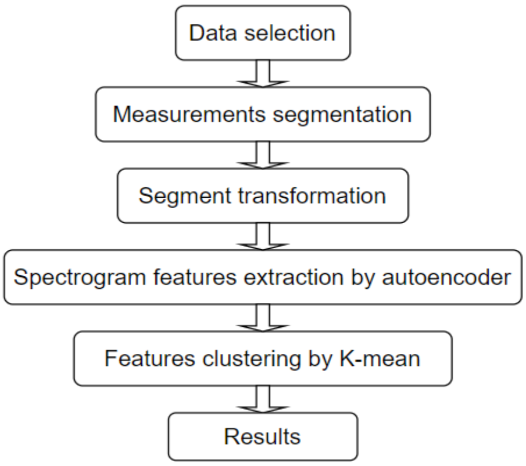

Fig. 2. Flow chart of proposed approach

3.1 Data selection

Measurements in CWRU represent three different conditions: fault-free baseline (BA), faulty bearing located at drive end (DE), and faulty bearing located at fan end (FE). When measuring data for baseline and faulty bearing at fan end, sampling rate is 12 kHz. When measuring data for faulty bearing located at drive end, sampling rate are in both 12 kHz and 48 kHz. Since sampling rate of 12 kHz is applied to all measurements, to keep samples under same sampling rate in subsequent calculations, only data with 12 kHz sampling rate are considered. In addition to that, another characteristic of CWRU dataset is that some measurements contain data from 1 channel (sensor located at DE) and some contains data from 3 channels (sensors located in basement, DE, and FE separately). In the proposed approach, both DE and FE channels data are applied for fault-free state. For faulty bearing located at drive end and at fan end, only the nearby accelerometers data are adopted. In other words, when faulty bearing is located at DE, data from DE channel are used, when faulty bearing is located at FE, data from FE channel are used.

3.2 Measurement segmentation

In CWRU bearing dataset, data for each operating condition are only measured once with 204 measurements in CWRU bearing dataset in total. Since data are at the core of machine learning, a large amount of training data plays a critical role in making the machine learning models successful. To train a machine learning model, the sample number must be suitably large. Measurement segmentation is an efficient technique for increasing samples. As bearings belong to rotating machinery, the segmentation length is designed by it's speed. Under different motor load, the motor speed is diverse, from 1797 rpm (0 HP), 1772 rpm (1 HP), 1750 rpm (2 HP), to 1730 rpm (3 HP). In other words, under different motor loads, data in each round are 400, 406, 411, 416 with an average of 408. To get a complete spectrogram of one round, the segmentation length should consider in the minimum one round. The segment length is settled at 408, 512 (1.25 time data for each round), and 1000 (almost 2.5 times of one round data) firstly. Final results shows that when segment's length is 1000, the result is the best. One segment is shown in the left of Figure 3 (left).

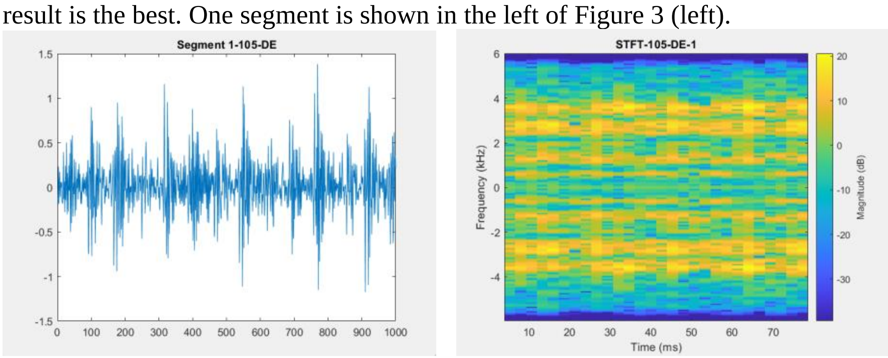

Fig. 3. Left: one segment; Right: spectrogram of one segment

3.3 Segment transformation

For each segment in time domain, time-frequency transform is an efficient way to generate features in time-frequency domain. To analyze non-stationary signals, Short-time Fourier transform is one of the most straightforward approaches for performing time-frequency analysis. Short-time Fourier transform partitions the time-domain input signals into several disjointed or overlapped blocks by multiplying the signal with a window function and applying the Discrete Fourier Transform (DFT) to each block. For random signals, Hanning window is often the preferred window function choice to manipulate a portion of the signal to reduce errors in the discrete Fourier analysis. The window length can help to balance time resolution and frequency resolution. Other parameters like the number of overlapped samples, numbers of DFT points, and STFT frequency range also affect spectrograms. By adjusting these parameters, spectrograms (Figure 3 (right)) are obtained.

3.4 Feature extraction

Spectrograms are not easy to distinguish as a large number of pixels are contained. Features extraction and compression are important steps to be realized before clustering spectrograms. Autoencoder can capture the most important features present in data. A whole encoder consists of three parts, namely: encoder, bottleneck, and decoder. In the encoder step, the model reduces input data dimensions and compress them into an encoded representation. Bottleneck is the lowest possible dimension which contains the compressed representation of input data. In the decoder step, the model reconstructs data from the encoded representation to be as close to the original input as possible [12]. In the proposed approach, only encoder part is applied since the main goal is spectrograms' feature extraction and compression. Compare with principal component analysis (PCA) which can also reduce dimensionality of large datasets, encoder could form nonlinear dimensionality reduction. As original spectrogram size is very large to maintain features, in the proposed approach, the bottleneck size is assumed as 200. Meanwhile, the epoch is assumed as much as 1000 for model training.

3.5 Feature clustering

Features of spectrograms are not easy to distinguish. Clustering is an exploratory data analysis technique used for mining data structure. The task of clustering is identifying data subgroups in the way that data in the same cluster are very similar while data in different clusters are very different. A cluster contains a centroid and a boundary. To define the centroid and boundary, iterative calculation is needed until desired or appropriate result is achieved. K-mean algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinction non-overlapping clusters and each data belongs to only one group [13]. According to CWRU bearing dataset, the first step is to cluster vibration signals between fault-free and faulty. So the value of K is determined by two. Initial two centroids are selected randomly and the sum of squared distance between data points and centroids are computed. Data points are assigned to the closer cluster. Afterwards, new centroids are computed by taking the average of all data points that belong to each cluster. The iterative calculation is stop when two centroids of newly formed cluster are stable.

4 Results

To evaluate the performance of proposed approach, external metrics like purity, rand index, adjusted rand index, normalized mutual information, and F-measure are applied in this contribution. Purity is the percent of the total number of objects (data points) that are classified correctly ranging from zero to a tolerance limit. The disadvantage of purity is that it is impossible to tradeoff between the quality of the clustering and the number of clusters. Normalized mutual information which measures the similarities between two labels of the same data can balance the cluster quality and cluster numbers. One advantage of NMI is that it can be applied to compare different cluster models that have different number of clusters. Another commonly used metric for clustering algorithms is the rand index. Rand index is a similarity measure comparing two clusters by considering all pairs of samples and counting pairs that are assigned in the same or different clusters in the predicted and true clusters. However, the major problem with the RI is that the expected value of rand index of two random cluster or partition does not take a constant value. To solve the problem, adjusted rand index is introduced where the generalized hypergeometric distribution considered as a random model. Besides the metrics above, F-measure is used for evaluating the proposed approach.

To verify the suitability of the proposed approach on the CWRU dataset, samples are divided into training and test groups. The ratio between training and test samples number is 9:1. As shown in Table 1, all the metrics value for both training and test are 1 which means that samples can be clustered into two groups totally by proposed approach. In other words, the proposed approach can distinguish fault-free and faulty state in CWRU bearing dataset perfectly.

Table 1. Results

Training					Test				
Purity	RI	ARI	NMI	F-measure	Purity	RI	ARI	NMI	F-measure
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

5 Conclusion

As one of the most cited benchmark datasets, CWRU bearing dataset has been applied to verify new approaches and related performance results. The CWRU dataset data are labeled, consequently, more of the applied approaches are supervised approaches.

In this contribution, a new unsupervised approach is raised to cluster data in CWRU bearing dataset. Data are selected, segmented, and transformed from time domain to time-frequency domain. Afterwards, autoencoder is utilized for spectrograms' feature extraction and compression. These compressed features are clustered into two clusters: fault-free and faulty by K-mean.

Using the labeled data, external evaluation metrics of the proposed approach is possible. Results using purity, rand index, adjusted rand index, normalized mutual information, and F-measure are consequently 100 %. This means that the proposed unsupervised approach could distinguish fault-free and faulty state in CWRU bearing dataset. For further work, more bearing states in CWRU should be distinguished by adjusting the values of K. Furthermore, beside external metrics, internal metrics could also be employed to estimate the approach.

References

1. Zhang, S., Zhang, S., Wang, B., Habetler, T.G.: Deep learning algorithms for bearing fault diagnostics - a comprehensive review. IEEE Access. 8, 29857–29881 (2020)
2. Wei X, Söffker D. Comparison of CWRU Dataset-Based Diagnosis Approaches: Review of Best Approaches and Results. European Workshop on Structural Health Monitoring. Springer, Cham, 525-532 (2020).
3. Wang H, Pang G, Shen C, Ma C. Unsupervised representation learning by predicting random distances. Proceeding of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 2950-2956 (2019).
4. Sigal L. Human pose estimation. Computer Vision. Springer, Cham, 1-20 (2020).
5. Bose S. How stuff works: K-means clustering. https://medium.com/@souravboss.bose/how-stuff-works-k-means-clustering-8f318755750d. Last accessed 12.01.2022.
6. Mashayekhi H, Habibi J, Khalafbeigi T, Voulgaris S, Steen M. GDCluster: a general decentralized clustering algorithm. IEEE transactions on knowledge and data engineering, 27(7): 1892-1905 (2015).
7. Cui M. Introduction to the k-means clustering algorithm based on the elbow method. Accounting, Auditing and Finance, 1(1), 5-8 (2020)
8. Adadi A. A survey on data-efficient algorithms in big data ear. Journal of Big Data. 8(1), 1-54 (2021).
9. Shirkhorshidi A S, Aghabozorgi S, Wah T Y. A comparison study on similarity and dissimilarity measures in clustering continuous data. PloS one, 10(12), 2015.
10. Pfitzner D, Leibbrandt R, Powers D. Characterization and evaluation of similarity measures for pairs of clusterings. Knowledge and Information Systems, 19(3): 361-394 (2009).
11. https://engineering.case.edu/bearingdatacenter. Last accessed: 09.02.2022.
12. Toderici G, O'Malley S M, Hwang S J, Vincent D, Minnen D, Baluja S, Covell M, Sukthankar R. Variable rate image compression with recurrent neural networks. International Conference on Learning Representation 2016.
13. Ahmad A, Dey L. A k-mean clustering algorithm for mixed numeric and categorical data. Data & Knowledge Engineering, 63, 503-527 (2007).