Deep Learning applications in machine condition monitoring under domain shifted conditions

Kenigsberger Genrikh (MSU master's st.), Sergey Stupnikov (cand. of technical sciences, sr. researcher IPI RAN), Ivan Shanin

***Abstract*—Deep Learning has been around for quite some time in machine condition monitoring (MCM) problems. On Google Scolar, there are dozens, if not hundreds, of articles on this area over the past 2-3 years. In addition to research articles, about ten detailed reviews of methods in this area have been written in recent years. The articles describe a variety of deep learning methods: autoencoders, convolutional neural networks, recurrent neural networks, restricted boltzmann machines, deep belief networks, which are used to detect anomalies in the behavior of bearings, hydraulic pumps, machine tools and even spacecrafts.**

Due to the large number of sufficiently recent detailed review articles, our goal was not to compile another review. Instead, the article pays little attention to existing modern methods of deep learning in the MCM problem, but focuses on the problem of domain adaptation in this area. Also, a separate part of the article is devoted to an overview of existing datasets and special attention is paid to articles published as part of the 2021 DCASE workshop.

***Index Terms*—deep learning, machine condition monitoring, domain adaptation, DCASE.**

I. Introduction

SOUND anomaly detection is the task of determining whether the sound produced by the device is normal or abnormal. Looking at the success of machine learning methods in image recognition, researchers have begun to apply the same techniques to MCM problems. There are many examples of the application of such models as CNN, RNN, autoencoders for monitoring the state of system components, for example, predicting the wear of machine tools, diagnosing the condition of bearings, and for managing the state of a complex system such as aircraft or even spacecraft.

This article provides a brief overview of the most modern techniques in MCM problems. The most close attention will be paid to methods based on deep learning methods. This decision is made due to the fact that DL-based methods perform best in MCM-themed competitions. Over the past years, the number of works aimed at applying DL methods to MCM problems has grown exponentially (as noted in [1]) and continues to grow. This popularity of DL methods in MCM is associated with the emergence of a large number of cheap sensors and sensors with Internet access.

As mentioned in [2], MCM models based on physical assumptions and laws are not suitable for complex dynamical systems. Modeling such systems is a very difficult task, in addition, such physical models cannot be corrected as new data are obtained. With the advent of inexpensive sensors available to everyone, data-driven models are gaining popularity. Machine learning technologies are used to process a large amount of data collected by sensors.

With the above in mind, DL technologies are good contenders for the MCM challenge.

II. Motivation

Automated detection of machine failures is an incredibly important technology for the 4-th Industrial Revolution [3], as is the AI-powered automation of factories. A quick detection of the abnormal behavior of a device by sound will be useful in the maintenance of machines, in devices for monitoring the state of complex mechanical systems (such as machine tools and conveyors).

Microphones are already being used as sensors for detecting anomalies in sound recordings. For example, ASD (anomalous sound detection) technology is used for audio surveillance [4], article [5] proposes methods for observing primates, in article [6] ASD is used to localize calls and shots in the streets.

As noted in the article [7], in the last 10 years, IoT technologies have revolutionized the industrial market, various methods of monitoring the state of devices are already being used:

- vibration-sensor methods [8][9]
- temperature-sensors methods [10]
- pressure-sensors methods [11]

Another method is anomaly detection using acoustic scene classification and event decoding technologies. In the past 5 years, there has been significant progress in this area: new publicly available datasets appear, competitions are held on various platforms.

The task of detecting abnormal behavior of mechanisms is more urgent than ever. For example, since 2015, the DCASE (Detection and Classification of Acoustic Scenes and Events) challenge has been held annually, one of the tasks of which is the task of machine condition monitoring.

III. Existing Datasets

For the development of research in the field of machine condition monitoring, it is necessary to develop high-quality public data sets that meet modern standards.

This section will list the most interesting and high-quality audio datasets that are freely available and used in solving problems on the topic ”machine condition monitoring”.

One of the important tasks of audio anomaly detection is to detect unknown anomalous behavior in a situation where only normal behavior is available as training data. In real factories, abnormal sounds are rare and very diverse and it is impossible to collect all of them, which is why it is so important to be able to detect previously unknown anomalies.

The most suitable datasets for unsupervised audio anomaly detection are ToyADMOS and MIMII Dataset. Both datasets contain not only many normal sound samples for training and testing, but also examples of device failures for testing.

Another serious problem is the detection of anomalies in the context of shifted domain conditions, that is, when the acoustic characteristics differ between training and monitoring phase. It must be ensured that normal operation will not be erroneously judged as abnormal due to changing conditions. A common example of changing conditions would be a constant change in engine speed on a conveyor belt transporting products in a manufacturing facility. However, during the monitoring phase, anomaly detection system must continue to monitor the conveyor at any time, at any engine speed, including those that differ from the training data. In addition to changing operating conditions, the ambient noise level can change, the signal-to-noise ratio can change. The most suitable dataset for such tasks is the MIMII DUE dataset.

The most interesting of the listed datasets for machine condition monitoring tasks are the MIMII DUE and ToyADMOS datasets. These datasets have the largest number of records and are created directly for training models that detect anomalies in the operation of mechanical devices. MIMII DUE, unlike ToyAdmos, has been published to apply domain adaptation techniques for audio anomaly detection. The following is a brief description of the different datasets.

A. *MIMII dataset*

The dataset was presented in 2019 and, as the authors note, is the first of its kind dataset containing examples of normal and abnormal operation of industrial machines, such as a pump, fan and guide rail. In total, MIMII contains records of 4 devices, with 7 different models selected for each of the devices. For each of the mechanisms, different types of breakdowns were reproduced. Normal records of device operation were separated to the training part of the dataset, and in the test part of the dataset, equally normal and abnormal behavior are presented.

In total, the dataset contains 26'092 segments with normal system behavior and 6'065 with anomalous ones.

B. *MIMII DUE dataset*

As part of the annual challenge for the detection of acoustic scenes and events in 2021 (DCASE-2021), a dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions).

Standard models for detecting audio anomalies face difficulties due to the difference in environmental conditions between the training phase and the actual operating phase of the device. This dataset solves the problem of checking the reliability and stability of the model under various environmental conditions. Before the advent of MIMII DUE in 2021, such datasets did not exist.

The dataset consists of normal and abnormal audio examples of 5 different devices operating under 2 different types of operating / environment conditions - source domain and target domain. The data for each type of device consists of 6 ”sections”, each section is related to a specific type of product and consists of data for two domains (source and target). The total length of the audio records of the dataset is over 420'000 seconds. In total, the dataset contains about 42'000 sound clips, each of which is 10 seconds long. Each section contains about 1000 clips for the source domain and about 3 clips for the target domain. All training data represent normal system behavior. The Figure 1 shows a graph of the distribution of acoustic features during domain shifts for normal and abnormal operation of devices.

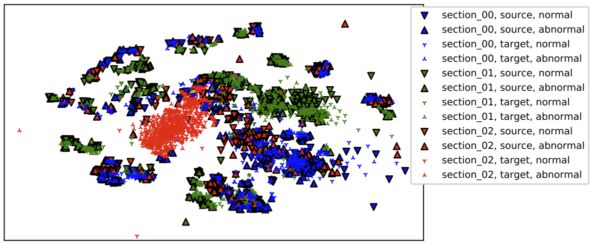

Fig. 1. Distribution of acoustic features during domain shifts for normal and abnormal operation. This figure is taken from [4]

C. *ToyADMOS dataset*

This dataset is one of the first large enough data sets for detecting abnormal sounds in device operation. The ToyADMOS dataset (dataset for anomaly detection in machine operating sounds) was presented in 2019 at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2019). To obtain such a large dataset (about 700'000 seconds of records), the authors recorded the work of mechanical toys and deliberately damaged the mechanisms for recording anomalous behavior. This dataset contains both normal and abnormal sounds of device operation.

One of the problems of research in the field of detecting sound anomalies was the fact that previously there was no sufficiently extensive data set on this topic. In this regard, many ASD models were trained on small datasets, which did not contribute to the development of research in this area.

Due to the fact that the dimensions of mechanical toys are small, the authors managed to record the operation of devices in the laboratory under idealized conditions. The authors of the dataset note that this dataset can be used not only for the usual unsupervised ADMOS, but also for the initiation of domain adaptation, suppression of noise on records.

The dataset is created for 3 types of problems:

- product inspection (inspection of the device, a toy car is taken as a device)
- fault diagnosis for a fixed machine
- fault diagnosis for a moving machine

The Figure 2 shows the listed devices.

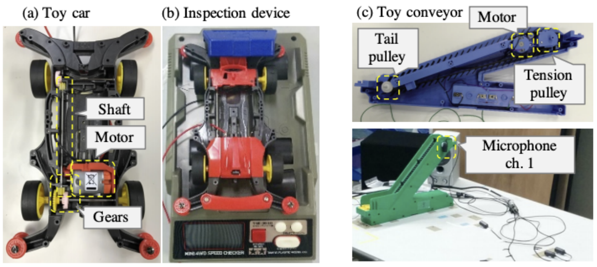

Fig. 2. Images of toys, parts, and microphone arrangement. This figure is taken from [12]

For each of the 3 types of devices, the dataset is divided into 3 sets - normal system behavior, abnormal system behavior, and separately recorded ambient noise. In the process of recording a set of data, various types of breakdowns were inflicted on the devices, for example, the brakes of the machine were bent, melted, the voltage of the battery varied, the shafts of the machine were bent.

D. ***AudioSet*** *dataset*

Inspired by the success of the ImageNet dataset in image recognition, Google researchers decided to create a similar dataset for sound events and scenes to train a universal audio event recognizer. The AudioSet consists of approximately 2 million 10 second YouTube recordings, categorized into 632 categories. This dataset is so depleted that one can find, for example, about 10’000 samples of engine operation at different frequencies, engine idling, engine starting, and so on. This dataset is not intended specifically for MCM tasks, but can be used in conjunction with other datasets.

E. ***IDMT-ISA-ELECTRIC-ENGINE*** *dataset*

This is a small dataset (1 hour of records) of real car engines operation under various conditions. Each of the 10-second files is marked by the state of the engine (the engine can be in one of three states (good, heavy load, broken). The dataset is presented in 2019.

F. ***CWRU*** *dataset*

As noted in one of the most recent reviews of DL applications in the MCM domain [2], the CWRU dataset [13] is the standard reference for validating models in this area. The article also emphasizes that this dataset is the most popular among all similar ones. Over the past 3 years, there are several hundred uses of the CWRU data in various articles.

We have not found other studies confirming such a great popularity of CWRU, but it is definitely worth mentioning it. The dataset was created to assess the quality of IQ PreAlert, a motor bearing condition assessment system developed at Rockwell. After that the experimental program has expanded to provide a motor performance database which can be used to validate and improve a host of motor condition assessment techniques. The dataset includes both normal engine behavior and abnormal behavior. Motor bearings were seeded with faults using electro-discharge machining (EDM). The dataset contains records of the operation of bearings at different breakdown levels at different engine loads.

IV. Brief summary of DL MCM methods

Authors of the most recent review article [2] have managed to find at least 11 articles devoted to the applications of DL in MCM. As the largest and most detailed review, it is worth highlighting the work that attracted more than 800 citations - the work of Zhao et al [1]. In this paper, authors propose to divide the available MCM approaches into the following categories: autoencoders and their variants, Restricted Boltzmann Machines and their variations, CNN and RNN.

A. ***Autoencoders***

Autoencoder is a special architecture convolutional neural network that allows unsupervised learning using backpropagation method. Any autoencoder consists of two parts: an encoder that converts input data into an intermediate representation of a smaller dimension and decoder, whose task is to restore the original one from the intermediate representation.

The metric that reflects the difference between the reconstructed result and the original representation is called reconstruction error.

During the learning process of the autoencoder, the reconstruction error (see Eq. 1) is minimized, where ψ_t - original representation and $r_\theta(\psi_t)$ - reconstructed representation:

$$A_\theta(X) = \sum_{t=1}^{T} \|\psi_t - r_\theta(\psi_t)\|_2^2 \tag{1}$$

The neural network learns to reproduce the most efficient intermediate representation and remembers the features of the input data structure.

In order for the neural network to better remember the features of the structure of the input data, a term responsible for sparsity is added to the expression of the reconstruction error. To get a deeper network of several autoencoders, a technique called stacking is used. Stacking autoencoders (SAE) are repeatedly mentioned in the review [1] as a popular method. Paper [1] lists the applications of deep autoencoders in the field of MCM. SAE are used to classify engine breakdowns, to detect breakages in rotating machinery.

Special attention should be paid to the presentation of data supplied to the input of the neural network. Some of the articles use raw input data, which leads to too high dimensionality of the input representation, but mostly the frequency spectrum of the input signal or the log-mel spectrogram is used. To prevent gradient vanishing problem, many articles use the ReLU activation function and the dropout technique.

B. ***RBM***

RBM is a neural network, the graph of which must be bipartite. The task of the RBM is to learn the input distribution over a set of input data. RBM is trained to maximize the product of probabilities assigned to some training set V:

$$\arg\max_{W} \prod_{v \in V} P(v) \tag{2}$$

Applying the staking method to the RBM, we get a deep neural network called the deep belief network (DBN). Due to the structure peculiarity, the deep belief network can be trained layer-wisely.

The article describes that deep belief networks were used in the tasks of predicting the remaining life of bearings and other mechanical devices. [3] Deep Belief Networks were often used in conjunction with a linear regression layer or self-organizing map algorithm (SOM) to transform the representation learned by RBM to the health value. In some works, several RBMs were used for different types of data, for example, for sounds and vibrations, after which a random forest was applied to fuse these representations. In the paper [14] Cheng et al. compared different input representations on how they affect the quality of the trained model. The raw signal, time domain feature, frequency domain feature and time–frequency domain feature were compared. It turned out that the raw signal performs worse than the processed variants.

C. *CNN*

Convolutional Neural Network aims to learn abstract features by stacking convolutional and pooling layers. The convolutional layers convolve filters with input data and generate local features, the subsequent pooling layers extract features with a fixed-length over sliding windows following several rules such as average pooling or max pooling.

The idea of using convolutional neural networks for MCM tasks is pretty clear. Often the input data is a 2D time-frequency image, for example after FFT. Accordingly, 2D-CNN are used to solve the problems of classifying audio recordings. Janssens et al. used 2D-CNN to classify the behavior of a rotating device into 4 categories. [15] Data from two accelerometers were fed to the network input.

D. *RNN*

Recurrent Neural Networks are the deepest neural networks that can display the entire history of previous entries into target vectors and allows you to save the memory of previous entries in the internal state of the network.

The majority of machinery data belong to sensor data, which are in nature time series and RNN are one of the best options for working with this kind of data. However, RNNs have not found the same popularity as CNNs and AEs in the field of MCM. The most popular RNN models for MCM problems are LSTM and GRU networks. Yuan et al. has shown that they outperform conventional LSTM in quality [16]. The article also notes that the ensemble of networks did not give better quality compared to LSTM.

V. Methods comparison

In the previous section, the main types of models used for the MCM problems were listed. However, these models are rather difficult to compare with each other, since they are all used to solve various narrow problems and, moreover, use different data sets to validate the model.

In order to conduct a competitive comparison of the state of the art models, we chose the problem of unsupervised audio anomaly detection during the operation of mechanical devices. This choice is justified by the fact that a DCASE workshop is held annually on this topic, which allows us to assess the popularity and quality of various methods that have been validated on one specific dataset. We have analyzed the top 10 solutions to Problem 2 of the DSACE 2020 Challenge.

A. *Acoustic feature*

In this section, we'll look at the types of acoustics input to the models. Among the top 50 solutions, about 40 of them use log-Mel spectrograms as input data. The rest of the solutions use the result of the FFT. The Mel scale is a kind of non-linear transformation based on triangular filters applied to the frequency scale, which is based on a human subjective perception of the pitch.

B. *Model types*

The task of this challenge was to detect anomalies during the operation of such mechanical devices as a pump, gearbox, fan. The complexity of the problem is that only the normal behavior of the system presented in the training dataset.

Among the top 20 solutions to the problem, approximately 15 solutions used CNNs as classifiers, the remaining 5 solutions were implemented using various types of AEs.

Giri et al. (the winner of the challendge) used an ensemble of models, consisting of masked autoencoders for density estimation (MADE) and 2 CNNs: MobileNetV2 and ResNet-50. [17]

1) AE models: A masked autoencoder is a simple modification of an autoencoder that allows you to use them to estimate distributions. Using a masked autoencoder, authors approximated a probability distribution that simulates normal audio recordings. Approximation was done with a mixture of Gaussian models. During the training stage, a negative log likelihood was taken as a loss function [17].

Daniluk et al. [18] used variational autoencoders. A typical variational AE comprises two networks: Encoder (E) which maps an input feature vector X to a normal distribution over the latent space $\mathcal{N}(\mu, \sigma^2)$ and Decoder (D) which maps a latent vector Z to a reconstructed point in the feature space. Loss function is composed of two terms:

$$D_{K_L}(\mathcal{N}(\mu, \sigma)||\mathcal{N}(0, 1)) + MSE(X, D(Z)) \tag{3}$$

Where the first term is explained by the addition of sparsity and is called Kullback–Leibler divergence, and the second is a typical reconstruction error.

2) CNN models: The MobileNetV2 neural network is the most popular network used in this challenge. It has relatively small number of parameters, in addition, this network has only 2 hyperparameters and is very easy to learn. In article [17], this neural network is trained to classify the type of mechanical device. The negative probability of an object belonging to the correct class is used as the anomaly score.

C. Data augmentation

In order to make the models more robust and to expand the training dataset, the contestants used a variety of data augmentation methods.

To generate new sound samples, Giri et al [17] used a mixup of sound spectrs with different weights. In such a case, it was expected that the model would yield probabilities proportional to the weights of samples. Another popular method for data augmentation is time stretching and pitch bending.

VI. Domain adaptation

One of the major problems of MCM is the performance of ASD (anomalous sound detection) in a domain-shifting environment, that is, when the acoustic characteristics differ between the training and monitoring phases. It must be ensured that sounds under altered conditions are not classified as anomalous. Real cases are often associated with different operating conditions of devices.

In this part of the article, we will analyze the main methods of domain adaptation used in MCM context and consider the methods used during the DCASE-2021 challenge, one of the tasks of which was to detect anomalies of mechanical devices in a shifted domain conditions.

Jose et al. [19] found that blending the predictions of various anomaly detectors, rather than relying on well-known domain adaptation techniques alone, gave them the best performance under domain shifted conditions In their submission, they used an ensemble of three models. The first model was an autoencoder network which used 1D convolutions. The input to this model was a spectrogram. This model did not apply any preprocessing and augumentation to the input data. The second model builds on the well-known WaveNet architecture by adding an x-vector classification after the dilated convolutions. For this model they used the Teager-Kaiser energy operator to preprocess the audio, this operator provides noise suppression. The third model attempts to learn several distributions of some Mel spectrogram bins conditioned on the remaining bins. They attempted to model the probability density function of the Mel spectrograms using normalizing flows. For the last system they combined these three models by first standardizing the training data scores and then searching over a grid of convex combinations.

Kazuki et al. [20] used the conventional detection methods but focused on feature extraction using CNN. By using spectrograms they trained a CNN such as MobileNetV2 and MobileFaceNet, they used Additive Angular Margin as a loss function. To calculate anomaly score they used an output of Local Outliers Factor (LOF) for source domain and kNN for target domain.

Kevin [20] presented novel anomalous sound detection system based on sub-cluster AdaCos. This system is trained to extract embeddings whose distributions are estimated in different ways for source and target domains.

There are two main strategies for detecting abnormal sound in MCM context: the first approach is based on training autoencoders. The main assumption is that recovering the abnormal data will result in a high reconstruction error. Thus reconstruction error can be used as anomaly score. The second approach is to train neural networks to distinguish classes of devices. Next, you can use a trained neural network to extract a representation of the data.

VII. Summary

This article explores the main datasets used in MCM problems, lists the main methods and approaches. Special attention in the article is paid to the methods for making the model resistant to domain changes.

References

[1] R. Zhao, Z. Chen, K. Mao, and R. Yan, “Deep learning and its applications to machine health monitoring,” 2019. DOI: https://doi.org/10.1016/j.ymssp.2018.05.050.

[2] W. Wang, J. Taylor, and R. J. Rees, “Recent advancement of deep learning applications to machine condition monitoring part 1: A critical review,” 2021. DOI: https://doi.org/10.1007/s40857-021-00222-9.

[3] R. Tanabe, H. Purohit, K. Dohi, T. Endo, Y. Nikaido, T. Nakamura, and Y. Kawaguchi, “Mimii due: Sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions,” 2021.

[4] C. Clavel, T. Ehrette, and G. Richard, “Events detection for an audio-based surveillance system,” 2005. DOI: https://doi.org/10.1109/ICME.2005.1521669.

[5] S. Heinicke, A. K. Kalan, O. J. Wagner, R. Mundry, H. Lukashevich, and H. S. Kühl, “Assessing the performance of a semi-automated acoustic monitoring system for primates,” 2015. DOI: https://doi.org/10.1111/2041-210X.12384.

[6] G. Valenzise, L. Gerosa, M. Tagliasacchi, and F. Antonacci, “Scream and gunshot detection and localization for audio-surveillance systems,” 2007. DOI: http://dx.doi.org/10.1109/AVSS.2007.4425280.

[7] H. Purohit, R. Tanabe, K. Ichige, and T. Endo, “Mimii dataset: Sound dataset for malfunctioning industrial machine investigation and inspection,” 2019.

[8] E. P. Carden, “Vibration based condition monitoring: A review,” 2004. DOI: http://dx.doi.org/10.1177/1475921704047500.

[9] G. S. Galloway and V. M. Catterson, “Diagnosis of tidal turbine vibration data through deep neural networks,” 2013. DOI: https://doi.org/10.36001/phme.2016.v3i1.1603.

[10] G. Lodewijks, W. Li, Y. Pang, and X. Jiang, “An application of the iot in belt conveyor systems,” 2016. DOI: https://doi.org/10.1007/978-3-319-45940-0_31.

[11] R. F. Salikhov, Y. P. Makushev, and G. N. Musagitova, “Diagnosis of fuel equipment of diesel engines in oil-and-gas machinery and facilities,” 2019. DOI: https://doi.org/10.1063/1.5122152.

[12] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. Imoto, “Toyadmos: A dataset of miniature-machine operating sounds for anomalous sound detection,” 2019.

[13] C. W. R. U. B. D. Center, “Cwru bearing dataset (https://csegroups.case.edu/bearingdatacenter).”

[14] Z. Chen and S. Deng, “Deep neural networks-based rolling bearing fault diagnosis,” 2017. DOI: https://doi.org/10.1016/j.microrel.2017.03.006.

[15] O. J. amd Viktor Slavkovikj, “Convolutional neural network based fault detection for rotating machinery,” 2016. DOI: https://doi.org/10.1016/j.jsv.2016.05.027.

[16] M. Yuan, Y. Wu, and L. Lin, “Fault diagnosis and remaining useful life estimation of aero engine using lstm neural network,” 2016.

[17] R. Giri and S. V. Tenneti, “Unsupervised anomalous sound detection using self-supervised classification and group masked autoencoder for density estimation,” 2020.

[18] P. Daniluk, M. Gozdziewski, S. Kapka, and M. Kosmider, “Ensemble of auto-encoder based and wavenet like systems,” 2020.

[19] J. A. Lopez, G. Stemmer, and P. Lopez-Meyer, “Ensemble of complementary anomaly detectors under domain shifted conditions,” 2021.

[20] K. Morita, T. Yano, and K. Q. Tran, “Anomalous sound detection using cnn-based features by self supervised learning,” 2021.