ECE381K - Term project proposal

Anomalous sound detection for industrial machine with supervised and unsupervised learning algorithms

Group 9
Adhika Retnanto, Min-Gyu Kim, Ryan Roby, Sophia Shi

Problem Description

Industrial machinery can break down for various reasons: contamination, environmental strains, human errors, device degradation, etc. The malfunctioning of a single piece of machinery can affect the equipment and operations upstream and/or downstream, and potentially impose devastating effects on the entire industrial process. Detecting machine malfunctions in a timely and accurate manner is essential to avoid an excessive amount of operation downtime, resource waste, productivity reduction, maintenance costs, long-term damages, and most importantly, safety concerns. Active research has been developing methods in response to the growing demand for malfunctioning industrial machinery investigation and inspection (MIMII). Several possible decision characteristics of machinery have been considered, e.g. pressure, temperature, vibration, and acoustics. The latter approach, anomalous sound detection (ASD), is detecting anomalies through sound. The sound frequency spectrum behaves differently for normal and abnormal operations, in noisy environments, as well as in different malfunctioning events. Sound detection is a technique that holds the potential for low costs and high efficiency and will be the focus of our project.

There are a few **challenges** to be addressed regarding MIMII using acoustic detection. **(1)** While sound data under normal operations can be easily obtained, available data for the wide range of malfunctioning events is limited, and there are unprecedented events that we cannot obtain data from [1][2]. **(2)** The normal dataset is more challenging to work with for non-stationary machinery, where the statistical properties of the sound data change with time, sometimes in a seasonal pattern (e.g. a valve might open fully or to a certain percentage depending on the demand) [1]. **(3)** Changes to the factory operation or to the ambient environment result in domain shifts in the dataset, meaning the data features shift to a different distribution. Normal sounds could potentially be monitored as abnormal if such shifts are not addressed in the model, thus reducing the model performance significantly [2].

The **objective** of this proposed project is to produce *a robust machine-learning model that detects machine malfunctioning from operational sound recordings that are masked by background noise* while addressing the aforementioned challenges.

Data Available

Audio data from the Malfunctioning Industrial Machine Investigation and Inspection (MIMII) dataset made available by Hitachi, Ltd. will be used for the purposes of our project [1]. The dataset consists of sound samples that are generated from four types of industrial machines, including valves, pumps, fans, and slide rails. The dataset was collected using a circular microphone manufactured by *System in Frontier Inc.*, TAMAGO-03. The microphone is a circular array that consists of eight distinct microphones with a sampling rate of 16kHz and 16 bits per sample, while also enabling single- and multi-channel-based approaches. The microphone was placed 50 cm from the pumps, fans, and slide rails and 10 cm from the valves. 10-second sound intervals were collected.

Using the measurements, the authors recorded seven sound files for four individual product models for each of the industrial machines, of which four are publicly available. These sound files reflect realistic scenarios in a typical factory. More specifically, within the individual models, the data in each model consists of up to 10,000 seconds of normal sound and 1,000 seconds of abnormal sound. Anomalous sounds include contamination, leakage, rotating unbalance, etc. In addition, the normal and abnormal sounds were mixed with background noise from various factories. In total, 26,092 normal and 6,065 abnormal sounds were recorded.

Possible Approaches

Both supervised and unsupervised can be applied to anomaly detection based on the sound data. However, in this study, we will first focus on developing the supervised machine learning model and finding the appropriate feature extraction methods and CNN model structure. Then we will apply the unsupervised methods to see if they return better results.

The proposed workflow for **supervised learning** is as follows:

1. **Feature extraction:** Various feature extraction techniques can be considered to generate the spectrogram or other graphic features of the sound data for each WAV file in the dataset. We will compare the performance of the classification model according to various feature extraction methods. The candidate methods are:
    a. **STFT** (Short-Time Fourier Transform)
    b. **MFCCs** (Mel-Frequency Cepstral Coefficients)
    c. **GFCCs** (Gammatone-Frequency Cepstral Coefficients)
    d. **Mel spectrogram**
2. **Label the dataset** (1: abnormal, 0: normal)
3. **Construct of the CNN model and fit/prediction**
    a. Model structure design: CNN layers include input layer, convolution layer, activation layer, pooling layer, flatten layer, and dense layer.
    b. We will explore the impact of the layers and optimize the model structure and parameters for these layers to improve the model performance.
    c. Model compiling and fitting: we should determine the best batch size, number of epochs, input sample size, etc. to optimize the model performance.
4. **Threshold method**
    a. The CNN model provides real numbers between [0,1] as predicted results. Thus, we need to find the optimal threshold to distinguish the machine's status
5. **Model evaluation and discussion**
    a. Because this is a classification problem, we evaluate the result using the accuracy, ROC curve, AUC, PR curve, AP, etc., which are covered in the lecture.
    b. We will discuss the model performance by various machine categories, feature selection methods, the CNN model structures, and over/under fitting issues.

Fig.1 Proposed supervised learning workflow

The domain generalization techniques are required to ensure the scalability of a machine learning model under various domain shifts. For this purpose, we will attempt to apply one of the unsupervised learning methods based on the feature extraction method that are mentioned in the supervised learning part. In unsupervised learning, the workflow will be a little different from the

supervised learning.
The proposed workflow of **unsupervised learning** is as follows:

1. **Feature extraction :** In this step, we can combine various feature extraction methods considering the input format of the selected classification models.
    a. **(for Autoencoder)** The same methods (STFT, MFCCs, GFCCs, Mel spectrogram) can be applied, which convert the audio files to a spectrogram.
    b. **(for other clustering algorithms)** Convolutional Neural Networks excel at extracting features from the 2D feature inputs. In this project, the pre-existing feature extraction methods using CNN such as ResNet-50, MobileNetV2, and AlexNet may be incorporated with the above spectrogram methods.
2. **Construct the model and fit/prediction**
    a. **(Autoencoder)** In these methods, the features would be embedded into a lower dimensional space and then subsequently reconstructed to its original dimension in the autoencoder. The loss function will be defined as the reconstruction error of the selected model using only the normal dataset and the anomaly scores will be calculated for all dataset for classification purposes.
        i. **Autoencoder (AE)**
        ii. **Deep Convolutional Autoencoder (DCAE)**
    b. **(Other clustering algorithms)** We expect the clustering algorithms to help us distinguish between the anomalous sounds from the normal sounds. Some of the clustering algorithms we propose to compare in our project are:
        i. **The Isolation Forest (IF)**
        ii. **Gaussian Mixture Model (GMM)**
        iii. **A One-Class Support Vector Machine (OCSVM)**
        iv. **Kernel Density Estimation (KDE)**
3. **Thresholding method:** In the autoencoder methods, we expect the anomalous dataset to have high reconstruction error and the normal dataset to have low error. An optimal threshold should be selected that accurately determines the machine's condition. **4. Model evaluation and discussion:** the same method with supervised learning.

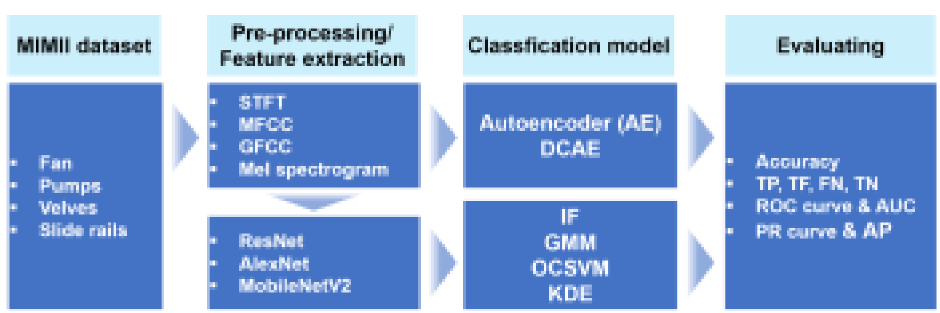

Fig.2 Proposed unsupervised learning workflow

References

[1] Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., & Kawaguchi, Y. (2019). MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. *arXiv preprint arXiv:1909.09347*.

[2] Tanabe, R., Purohit, H., Dohi, K., Endo, T., Nikaido, Y., Nakamura, T., & Kawaguchi, Y. (2021, October). MIMII DUE: Sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions. In *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*

(pp. 21-25). IEEE.
[3] Dohi, K., Nishida, T., Purohit, H., Tanabe, R., Endo, T., Yamamoto, M., ... & Kawaguchi, Y. (2022). MIMII DG: Sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task. *arXiv preprint arXiv:2205.13879*.
[4] Mnasri, Z., Rovetta, S., & Masulli, F. (2022). Anomalous sound event detection: A survey of machine learning based methods and applications. *Multimedia Tools and Applications*, 1-50. [5] Müller, R., Ritz, F., Illium, S., & Linnhoff-Popien, C. (2020). Acoustic anomaly detection for machine sounds based on image transfer learning. *arXiv preprint arXiv:2006.03429*. [6] Nejjar, I., Meunier-Pion, J., Frusque, G., Fink, O., & CentraleSupélec, G. S. Y. (2022, November). DG-Mix: Domain Generalization for Anomalous Sound Detection Based on Self-Supervised Learning. In *DCASE*.
[7] Nunes, E. C. (2021). Anomalous sound detection with machine learning: A systematic review. *arXiv preprint arXiv:2102.07820*.