Sound-Based Motor Defect Detection with Machine Learning

Tomoki Shirai s1240210 Supervised by Prof. Qiangfu Zhao

Abstract

In this study, motor abnormality detection is performed by machine learning. A microphone was attached to the inspector's hat to record the actual working sound. In the abnormality detection procedure, we first divide the sound data into sound fragments of 1 millisecond, and calculate the Mel frequency cepstral coefficient (MFCC) for each fragment. Using the MFCC as the feature vector, the next step is to predict whether the motor is running using random forest [2]. Thereby, the motor sound can be recognized. Finally, whether there is any abnormality is detected using LSTM [3], Auto-encoder [4], and Variational-auto-encoder [5]. According to the results, it is possible to detect the motor sounds with an accuracy of 99.7%. Anomaly detection accuracy is 89%, assuming that resulting in reducing the number of motors inspected by inspectors to 1/4.

1 Introduction

In the final step of a motor production line, it is usually necessary to detect motor abnormalities by skilled human inspectors based on auditory, visual, tactile and other information. In this case, the accuracy of defect detection depends on the health and fatigue state of the inspector. Machine learning can be used to solve this problem. By using AI, which does not get tired, quantification and stabilization of the test can be expected.

The purpose of this study is to create a machine learning model that can obtain the same test results as a human inspector. In addition, by focusing only on the motor operation sound in the machine learning data, we aim to obtain a general-purpose classifier that does not need to consider the motor types. If AI with a learning function is introduced, even if the type of motor changes, the inspection machine can flexibly adapt to it, so the development of new products will be easier than before. In addition, this study tries to detect anomalies in a noisy environment, and to create a general-purpose abnormality detection model that is resistant to noise.

2 Proposed Method

2.1 Experimental environment

Table1: Development environment used in this study

CPU	Intel Core i7-8700 3.20GHz
GPU	RTX2080
OS	Ubuntu
development	Anaconda
Machine learning	Scikit-learn , keras
Sound analysis	Librosa , praat

Table2: Standard of microphone used for recording

width	102mm
height	20mm
depth	7.4mm
weight	29g
Directivity	Undirectorial
Range of frequency	95Hz ~ 18,000Hz
Impedance	16Ω
Voltage	DC3.7V

In the environment of Table 1, the experiment was able to proceed smoothly.

2.2 Data Collection

In this experiment, the microphone of Table 2 was attached near the ear of the worker's hat. We recorded two hours of working sound daily for six weeks. Since the sound of other machines in the factory was also recorded at the same time, the sound data was recorded in a noisy environment.

2.3 Labeling of Dataset

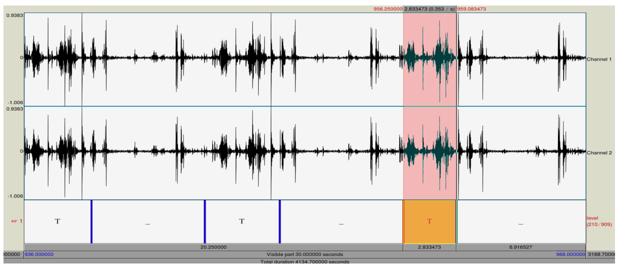

Figure 1: praat labeling

Labeling was performed as shown in Figure 1 using the voice analysis software praat. In the figure, |-| indicates environmental sound, | T | indicates normal motor sound, and | F | indicates abnormal motor sound. The following data were obtained after labeling.

・Number of environmental sound: 31600
・Number of Normal motor sound: 31499
・Number of Abnormal motor sound: 220

In most cases, the lengths of the data vary from 2 to 10 seconds. Some data can be as long as 30 seconds. If the length is different, the size of the frequency component included in the data changes, so that the length of the data must be unified. For this purpose, we sample 1 second data from $t = k \times \delta t$, where k=0,1,…, and δt=100 millisecond. Since the number of bad motors was small, we used $\delta t = 7$ms for bad motor sounds. The number of data obtained this way were as follows.

・Number of environmental sounds: 966428
・Number of normal motor sounds: 521956
・Number of abnormal motor sounds: 77517

The next step was to extract the features of the data. Sound data were recorded at a sampling rate of 44100. One second of sound data was represented by 44100 numerical values. The sound data can be recognized by humans. However, when performing machine learning, it is better to transform the voice data into some form so that it can be easily understood by the machine learning model. As a typical example, FFT was used to convert time-series data to frequency data. In addition, Mel frequency cepstral coefficient (MFCC), which is a feature value of speech in consideration of human hearing, is often used in the field of speech recognition. In this experiment, the MFCC was used as the feature value of the sound, considering that the inspector uses the auditory sense to detect abnormalities. In previous experiments, when an abnormality was detected in a motor in a noise-free environment, the MFCC was used to detect an abnormal motor with 96% accuracy. We calculated the MFCC for 17 dimensions for each data and take the average and standard deviation for each dimension. A total of 34 values were used as sound characteristics.

2.4 Motor Sound Detection

In order to automate motor abnormality detection with AI, a model that can recognize whether the motor is running is required first. This model is assumed to operate in real time. Therefore, it is appropriate to use the fastest method possible. In this experiment, we used a random forest that can be predicted at high speed. The parameters used to design the random forest are as follows.

・Number of decision tree 10 to 80 (10 increments)
・Division criteria gini or entropy
・Maximum depth of decision tree 1 to 15

Create 320 random forests using grid search and adopt the most accurate model.

2.5 Motor Anomaly Detection

The motor can be recognized by the random forest designed in 2.3. Check the accuracy of abnormality detection using the following three models for each motor.

1. LSTM

Long short-term memory (LSTM) is a supervised machine learning model that is used when learning time series data.

Below are the parameters to design the LSTM

・Input 34 dimensions
・Intermediate layer 17 dimensions (sigmoid)
・Output 34 dimensions (soft-max)
・Loss function is binary cross entropy
・Number of epochs for learning 100

In supervised learning, abnormal motor data must be included in the training data. In this study, the number of bad motors is small, so the test data for abnormal motors was significantly reduced. Both supervised and unsupervised learning models were designed and the accuracy was verified. The following two models are used for unsupervised learning.

2. Auto-encoder

The auto-encoder is a model used for dimension reduction of data, but this is also used in the field of abnormality detection as unsupervised learning. In this experiment, a three-layer auto-encoder was used as follows.

・Input 34 dimensions
・Intermediate layer 17 dimensions (sigmoid)
・Output 34 dimensions (sigmoid)
・Loss function is mean square error
・Number of epochs for learning 100

3. Variational Auto-encoder

Variational auto-encoder is a kind of variational Bayes estimation. The loss function of the middle layer is special.

The used parameters are listed below

・Input 34 dimensions
・Intermediate layer 17 dimensions
・Output 34 dimensions

- Loss function mean square error
- Number of epochs for learning 100

2.6 Testing Method

Data was collected for 6 weeks. We used the first n weeks of training data and test with the remaining 6–n weeks of data. Learning and prediction were performed sequentially, assuming the operation at the factory. This paper contains the prediction results of the model trained for one week and the model trained for five weeks. In order to evaluate an unsupervised learning model, we must set a threshold. In this study, we evaluate the model assuming that we want to reduce the burden on inspectors to 1/4. Specifically, a threshold value was set so that the rate of false detection of a normal motor as abnormal for the test set was 25%. In order to compare the accuracy of supervised learning with unsupervised learning, LSTM was also measured with the same threshold set for soft–max output.

3 Results

3.1 Motor Sound Detection

The accuracy of determining whether the motor was running using random forest is as follows.

Table 3.1.1: Learn with one week of data, predict with other data

Random forest		predict		
		enviro	motor	acc
actual	enviro	512185	2582	0.995
	motor	9882	871242	0.984

Table 3.1.2: Learn with one week of data, predict with label

Random forest		predict		
		enviro	motor	acc
actual	enviro	28517	85	0.997
	motor	102	28483	0.996

Table 3.1.3: Learn with five weeks of data, predict with other data

Random forest		predict		
		enviro	motor	acc
actual	enviro	128643	708	0.994
	motor	2342	185950	0.987

Table 3.1.4: Learn with five weeks of data, predict with label

Random forest		predict		
		enviro	motor	acc
actual	enviro	6701	0	1.0
	motor	24	6667	0.996

Figure 2 shows automatic labeling by the learned Random forest.

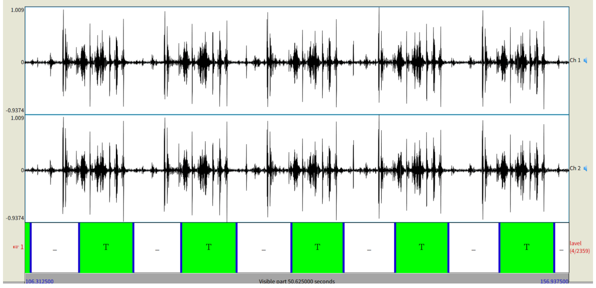

Figure 2

Detection of motor sound by random forest was sufficiently accurate. Even if we made a wrong prediction for each moment as shown in Table 3.1.3, it was possible to compensate for the wrongly predicted part as shown in Table 3.1.4 by taking a majority decision by referring to some predictions before and after. Some motors were unrecognized, but the unrecognized motors were placed far from the inspector, and their noise was not heard well. In other words, the motor near the inspector, which was supposed to be recognized, was recognized reliably.

3.2 Motor Anomaly Detection with LSTM

If the value output by soft-max was assumed to be larger than the predicted value, the accuracy was as follows.

Table 3.2.1: Learn with one week of data, predict with other data

LSTM		predict		
		normal	abnormal	acc
actual	normal	457478	13882	0.971
	abnormal	28151	15254	0.351

Table 3.2.2: Learn with one week of data, predict with label

LSTM		predict		
		正常	abnormal	acc
actual	normal	27688	698	0.975
	abnormal	68	51	0.428

Table 3.2.3: Learn with five weeks of data, predict with other data

LSTM		predict		
		normal	abnormal	acc
actual	normal	119964	1124	0.991
	abnormal	7335	926	0.112

Table 3.2.4: Learn with five weeks of data, predict with label

LSTM		predict		
		normal	abnormal	acc
actual	normal	6597	50	0.992
	abnormal	19	0	0.000

We compared these results with the accuracy with unsupervised learning model. The accuracy was measured by shifting the threshold so that the recognition accuracy of a normal motor was 0.75.

Table 3.2.5: Learn with one week of data, predict with other data

LSTM		predict		
		normal	abnormal	acc
actual	normal	353520	117839	0.75
	abnormal	16839	26566	0.612

Table 3.2.6: Learn with one week of data, predict with label

LSTM		predict		
		normal	abnormal	acc
actual	normal	21359	7026	0.75
	abnormal	19	100	0.840

Table 3.2.7: Learn with five weeks of data, predict with other data

LSTM		predict		
		normal	abnormal	acc
actual	normal	90816	30271	0.75
	abnormal	3752	4509	0.546

Table 3.2.8: Learn with five weeks of data, predict with label

LSTM		predict		
		normal	abnormal	acc
actual	normal	4997	1649	0.75
	abnormal	2	17	0.895

By allowing a certain amount of errors, erroneous detection of a normal motor increased, but the detection rate of an abnormal motor can be greatly improved.

Due to supervised learning, the number of abnormal motors in the test set has become extremely small.

3.3 Motor anomaly detection with Auto-encoder

The accuracy of abnormality detection by Auto-encoder is as follows.

Table 3.3.1: Learn with one week of data, predict with other data

Auto-encoder		predict		
		normal	abnormal	acc
actual	normal	353520	117839	0.75
	abnormal	30249	13156	0.303

Table 3.3.2: Learn with one week of data, predict with label

Auto-encoder		predict		
		normal	abnormal	acc
actual	normal	21391	6994	0.75
	abnormal	62	157	0.717

Table 3.3.3: Learn with five weeks of data, predict with other data

Auto-encoder		predict		
		normal	abnormal	acc
actual	normal	90816	30271	0.75
	abnormal	5541	2720	0.329

Table 3.3.4: Learn with five weeks of data, predict with label

Auto-encoder		predict		
		normal	abnormal	acc
actual	normal	5116	1530	0.75
	abnormal	33	186	0.849

Because it was an unsupervised learning, all bad motors could be used in the test set. Learning with more datasets can improve accuracy.

3.4 Motor anomaly detection with Variational auto-encoder

The accuracy of abnormality detection by variational auto-encoder (VAE) is as follows.

Table 3.4.1: Learn with one week of data, predict with other data

Variational Auto-encoder		predict		
		normal	abnormal	acc
actual	normal	353520	117839	0.75
	abnormal	18603	58913	0.760

Table 3.4.2: Learn with one week of data, predict with label

Variational Auto-encoder		predict		
		normal	abnormal	acc
actual	normal	21475	6990	0.75
	abnormal	26	193	0.881

Table 3.4.3: Learn with five weeks of data, predict with other data

Variational Auto-encoder		predict		
		normal	abnormal	acc
actual	normal	90816	30271	0.75
	abnormal	18413	59103	0.762

Table 3.4.4: Learn with five weeks of data, predict with label

Variational Auto-encoder		predict		
		normal	abnormal	acc
actual	normal	5125	1521	0.75
	abnormal	24	195	0.890

4 Discussion

The motor sound was recognized with good accuracy. The running sound of the motor is mainly concentrated at 5000Hz. The reason why the motor sound recognition was successful was that the noise in the factory was low in the frequency band as mentioned above. Figure 3 displays the result of FFT. The running sound is slightly visible at around 6 to 8 second with along the x-axis.

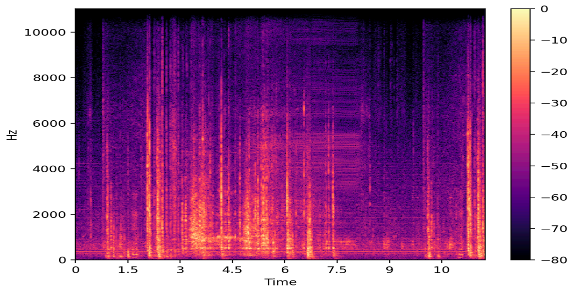

Figure 3

In this experiment, three types of models were used to detect motor abnormalities. First, we focused on LSTM for supervised learning. The accuracy was not good if we took the predicted value straightforwardly for Soft-max. Since the data used for this contained much noise noisy and the number of abnormal motors was small, the number of abnormal motor data might have been too may be small for the purpose. By setting a threshold that allows a certain amount of errors, prediction can be made with reasonable accuracy as shown in Table 3.2.4. The accuracy of the auto-encoder was also improved by increasing the training data, so it may be possible to improve the accuracy by increasing the number of data and re-testing. However, with regard to VAE, the difference in the accuracy between the data learned for one week and that for five weeks was only 0.9%, so it seems that the accuracy already leveled off with one week of data. VAE was the most promising of the three types of models we tested in this research. In fact, the accuracy was very good compared to Auto-encoder. We conclude that it is difficult to achieve higher accuracy than VAE if anomaly detection is performed by unsupervised learning using current features. For supervised learning, it would be worth experimenting with more data for abnormal motors to see how accuracy improves. We use MFCC as a feature, and we believed this was a good one in this experiment, but there was probably a better feature. Accuracy may be improved by adding a device that removes noise to the microphone, or by trying a noise removal procedure such as a Kalman filter on the audio after recording.

5 Conclusion

Recognition of the motor sound was sufficiently accurate. On the other hand, the accuracy of motor abnormality detection must be improved. In this experiment, it was possible to detect a abnormal motor with an accuracy of 89%, assuming that inspectors would re-examine 25% of the total. However, assuming that it was actually running in a factory, the abnormal motor detection rate must have at least 100% accuracy for test data. In addition, if the overall false positive rate cannot be reduced below 25%, the benefits of introducing machine learning will not be recognized much. In the previous experiment. Anomaly detection was performed with good accuracy in an experimental environment without noise. If appropriate noise removal is performed, anomaly detection will be possible even in an environment with large noise as in this experiment. Based on these, we need to review the pre-processing part of the data and try to improve the accuracy in our future study.

6 Future work

With the accuracy of the current motor abnormality detection, it is difficult to actually use it on site. The data used in this research contained much noise and the sound of the essential motor was not heard much. Nevertheless, since the abnormality of the motor was detected with a certain level of accuracy, even slightly reducing the noise may significantly improve the accuracy of the abnormality detection. With that in mind, we will record data in better condition and focus on data preprocessing in the next step, not on learning models. For example, the frequency band in which the abnormality of the motor is located is searched. If the feature is extracted by focusing on it, the accuracy should be higher than now. Before calculating the MFCC, we will consider going through some process to eliminate as much noise as possible without reducing the motor features. Applying well-known noise removal techniques such as microphone array techniques, spectral subtraction techniques, MAP estimation, and Kalman filtering to this anomaly detection should improve the accuracy even further.

Reference

[1] Breiman, Leo (2001). “Random Forests”. *Machine Learning* **45** (1): pp 5–32

[2] Sepp Hochreiter; Jürgen Schmidhuber (1997). “Long short-term memory”. *Neural Computation* **9** (8): 1735–1780.

[3] Geoffrey E. Hinton; R. R. Salakhutdinov (2006-07-28). “Reducing the Dimensionality of Data with Neural Networks”. *Science* **313** (5786): 504-507.

[4] B. P. Bogert, M. J. R. Healy, and J. W. Tukey: Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed) Chapter 15, 209-243. New York: Wiley (1963). "The quefrency alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking".

[5] D. G. Childers, D. P. Skinner, R. C. Kemerait. Proceedings of the IEEE, Vol. 65, No. 10, October 1977, pp. 1428-1443. "The Cepstrum: A Guide to Processing,"