Editorial

Deep Learning Based Technology for Diagnosis of Production Equipment Abnormalities

CHIDA Yuichi*

1. Introduction

It is expected that abnormality diagnosis and prediction technologies will be developed for production equipment whose unexpected failure can significantly affect production schedules. Meanwhile, deep learning (DL) technologies have advanced rapidly in recent years and are expected to be applied to abnormality diagnosis and prediction for production equipment[1]. Very important points of these technologies are the quality and quantity of data used to build a diagnostic model. In other words, sufficient data must be obtained for both normal and abnormal times. Furthermore, if the differences between the two sets of data are clear in terms of characteristics, it would be easier to identify abnormalities. In many cases, however, abnormalities occur infrequently and data for abnormal times are usually not available. This makes it difficult to diagnose and predict equipment abnormalities. To solve this problem, it is necessary to choose an approach to artificially generate abnormality data by consulting the probability distribution of data for normal times or an unsupervised learning approach such as that based on an auto-encoder[2]. Our research group is also working on these approaches[3),4)]. This paper presents examples of the application of these approaches and real examples of their application at production sites.

2. Artificial Generation of Abnormality Data and Abnormality Detection by Multi-labeled Deep Learning Networks[3]

This section introduces how to detect abnormalities in the operation of parts cleaning equipment. In the parts cleaning process, parts are cleaned by two or more operations. Here, with a focus on any changes in the operating time of the processes, the possibility of abnormality detection is discussed. In other words, if the time taken to complete a work process deviates from its standard, this indicates that there is a problem in the process that should be detected earlier. In this case, it would be sufficient to focus on the operating time of each process if the individual processes were completely independent of each other. However, in cases where a process affects its related processes, a more appropriate abnormality detection may be possible by using a structure that can consider the mutual relevance, rather than recognizing them as independent processes. A multi-labeled deep neural network (ML-DNN) can then be applied[3].

ML-DNN is a binary classification method for data from two or more output layers of a deep neural network (DNN). The configured ML-DNN receives the input of the operating time for two or more operations in the parts cleaning process and represents the possibility of an abnormality in each operation by binary data output (normal or abnormal). When considering the output for all the operations, this is a matter of binary classification of multiple outputs to which ML-DNN can be applied. This allows us to detect which operation has the abnormality, while taking into account the mutual influence between the different operations in the parts cleaning process.

In this case, it is necessary to train the ML-DNN with a sufficient number of normal and abnormal data sets. In reality, abnormal data is difficult to obtain, while normal data is available from the data collected during normal operation. To solve this problem, we have selected the approach of obtaining the data distribution in normal times, artificially setting values that deviate from the distribution as abnormal data, and training the ML-DNN based on these data.

Fig. 1 shows an example of probability distribution results obtained from a histogram of the time taken to complete an operation within the cleaning process. Since the histogram is not as simple as a probability distribution, it is necessary to build a probability distribution model. This time, however, we decided to use a normal distribution as the most convenient way. Assuming it is a normal distribution, we determine the mean μ and the standard deviation σ. Since the

* Professor, Academic Assembly (Institute of Engineering), Shinshu University

probability of occurrence of data outside the range of $\mu \pm 3\sigma$ is about 0.3%, it is assumed that data within the range of $\mu \pm 3\sigma$ are normal and data in the range between $\mu \pm 3\sigma$ and $\mu \pm 8\sigma$ are abnormal. This probability distribution is used to artificially generate abnormal data. The abnormal data is then used to train the ML-DNN. The data collected during normal operation is used as normal data for the ML-DNN.

The data sets generated in this way were used to train the ML-DNN to build an abnormality detection system. As a result, we successfully confirmed that the ML-DNN can achieve a correct response rate of 99.6% for abnormality detection. On the other hand, a regular DNN showed a correct response rate of 98.2%. Therefore, the ML-DNN has proven its superiority[3]. We are considering applying a mixed Gaussian or other probability distribution model in the future.

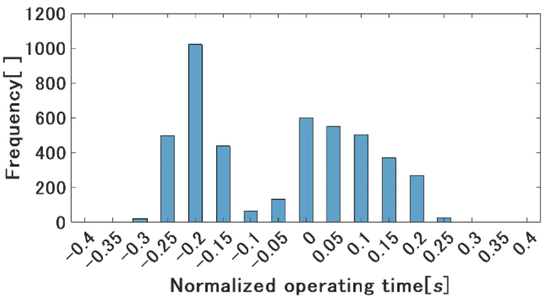

Fig. 1 Histogram of operating time[3]

3. Abnormality Detection Using Both Spectrum Characteristics and Auto-Encoder[4]

This section presents an example of considering the detection of abnormalities in a high-pressure pump in parts cleaning equipment[4]. Since the pump may have pressure data that is significantly affected by its drive frequency, it is effective to adopt an approach based on the frequency spectrum obtained by Fourier transform. We then subjected the time series data to Fast Fourier Transform (FFT) to determine the frequency characteristics and tried to detect abnormalities according to changes in the frequency spectrum. However, again, the problem is that it is not easy to obtain data during abnormal times. We then used an auto-encoder[3] based method, which is a type of unsupervised learning.

An auto-encoder does not necessarily need abnormality data. It is possible to train it only with normal time data to configure an abnormality identifier. Specifically, a network structure as shown in Fig. 2 can be set up to provide an $\hat{X}$ output that reproduces the input by encoding and decoding. In this case, the distance in data values between the input and the output is set as a loss function and the auto-encoder is trained so that the loss function is minimal. If the auto-encoder can be properly trained using this method, the loss function will be a somewhat small value when normal time data is input. Next, a threshold is set using the low loss function as a guide. Now the auto-encoder can judge the situation as normal if the loss function obtained from the input data is lower than the threshold, or abnormal if it is higher than the threshold. We have adopted this idea to simultaneously configure the auto-encoder and set a threshold to configure an abnormality identifier.

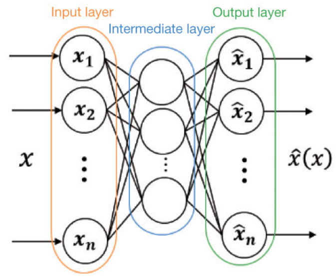

Fig. 2 Auto-encoder[2]

The input to the auto-encoder is the frequency spectrum data of 127 points obtained by FFT. The sum of the squares of the differences between the input and output of the auto-encoder is set as the loss function. On the other hand, the frequency spectrum data for abnormal times, which were required for the evaluation, were created artificially by processing part of the normal time data. In other words, the frequency spectrum data for abnormal times was created by increasing or decreasing some of the frequency spectrum values in the normal frequency spectrum data.

The auto-encoder was configured using the procedure above and its performance was verified. The results are shown in Fig. 3. The grey graph (right scale) represents the frequency spectrum and the red graph (left scale) the abnormality detection rate. The horizontal axis indicates the scaled frequency. The abnormality detection rate for the entire frequency range in Fig3. is approximately 83%, which means that the auto-encoder can detect most abnormalities. However, the abnormality detection rate is locally poor at a frequency of about 0.8. Similar poor performance can also be seen around frequencies of approximately 1.5, 1.8, and 2.3. At these frequencies, the frequency spectrum of the normal time data has a larger amplitude, which may lead to the poor accuracy in detecting amplitude changes at these frequencies.

This may be because at frequencies with a large amplitude, even the normal time data contained shifted frequency values as shown in Patterns 2 and 3 in Fig. 4, rather than converging to a single frequency value as shown in Pattern 1, making it difficult to detect changes in amplitude values for the spectrum in these frequency ranges. To solve this problem, it is appropriate to treat the different frequency spectrum patterns shown in Fig. 4

as the same pattern. We therefore decided to add another abnormality identification function to the auto-encoder for the specific frequency ranges where the abnormality detection rate was low.

We then switched to the abnormality identification flow shown in Fig. 5. Here, if the auto-encoder produces a normal result, an additional judgement is performed to determine if there is a possibility of an abnormality in the frequency ranges with poor detection accuracy. A frequency bandwidth has been set for the poor accuracy frequency ranges to cover all neighboring frequency peaks even for Patterns 2 and 3 in Fig. 4. The sum of the power spectra in the bandwidth was used to determine the presence or absence of an abnormality. This procedure begins by determining the sum of the power spectra in the set bandwidth for the normal time data. The next step is to determine the distribution of the power spectral values using the histogram. Assuming a normal distribution, the mean μ and the standard deviation σ are determined. As in Chapter 2, a threshold value was set assuming that data values within the range of μ ±3σ are normal, based on which a discrimination between normal and abnormal would be made. This method was applied to each of the frequency ranges (including approximately 0.8, 1.5, 1.8, and 2.3 in Fig. 3) where the auto-encoder often made an incorrect judgement. In other words, once the auto-encoder judges a normal situation, another judgement takes place around the frequency of 0.8. If that is judged normal, another judgement is made in the frequency range around 1.5, and so on. The target frequency range was shifted sequentially in this way for discrimination purposes. The results of the performance check using the method above are shown in Fig. 6. Using the same data as in Fig. 3, the abnormality detection rate for the entire frequency range was improved from 83% in Fig. 3 to approximately 92%. Fig. 6 shows that the abnormality detection accuracy in the frequency ranges with poor detection rate in Fig. 3 has also been improved.

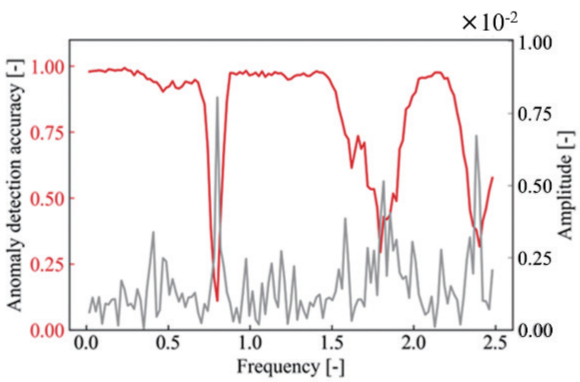

Fig. 3 Frequency data and abnormality detection rate (before improvement)[4)]

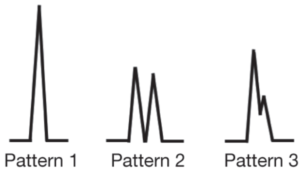

Fig. 4 Typical examples of frequency peak patterns[4)]

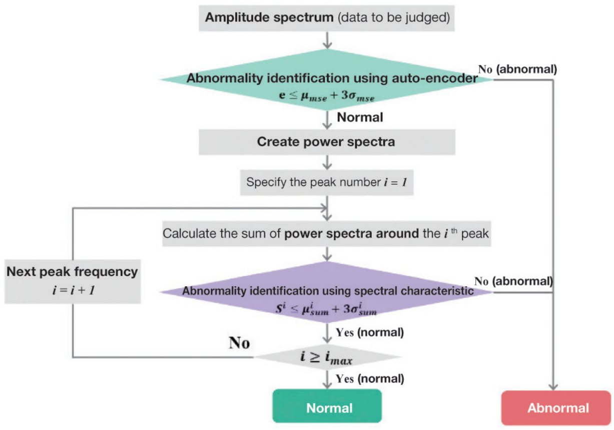

Fig. 5 Abnormality detection procedure[4)]

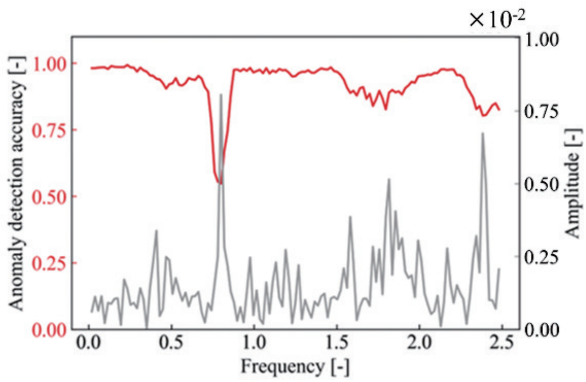

Fig. 6 Frequency data and abnormality detection rate (after improvement)[4)]

4. Examples of Application at Production Sites

A verification test was conducted to determine whether the abnormality detection system using the ML-DNN mentioned in Chapter 2 was effective for actual operation of production equipment. The test was conducted by the System Development Office, Digital Transformation Improvement Div., KYB Corporation. Specifically, operating time thresholds that can be easily controlled in the field were set and used to monitor any events that exceeded any of the thresholds. When such an event actually occurs, the ML-DNN is performed. If the result shows an equivalent abnormality, the ML-DNN should be able to detect abnormalities during actual operation.

The number of events exceeding the thresholds was visualized for easy identification using a business intelligence (BI) tool called Tableau, as shown in Fig. 7. The data was reviewed periodically. As a result, the phenomenon of an increasing number of abnormalities in the "lateral release of the work clamp" action was observed twice in a given period. The fact that this abnormal phenomenon occurred twice during the period is also evident from the control chart shown in Fig. 8. The same data was used to run the ML-DNN as shown in Fig. 9 for comparison. This confirmed that the abnormality trend was identical between the two.

The actual equipment was checked against the data. Locations associated with the "lateral release of the work clamp" action were examined to find any abnormalities. As a result, Abnormality 1 was found to be air leakage from the air tube (Photo 1) and Abnormality 2 was found to be deterioration of one of the air tubes (Photo 2). The defective parts were replaced. It was also confirmed that the abnormality data was eliminated after replacement.
We are preparing for commercialization in FY2024 or later of the abnormality detection for high-pressure pumps using both the spectral characteristics of their pressure data and an auto-encoder. We also plan to continue verification of effectiveness for vibration and acoustic data.

5. Conclusions

This paper has introduced abnormality detection and prediction technologies for production equipment using DL data such as multi-labeled deep neural networks and auto-encoders. These technologies can be applied to cases where little abnormality data is available. Key elements in these cases include a signal processing technique used to extract characteristic values to be focused for abnormality detection. It is also important to identify abnormality factors of the equipment and discuss how their impacts appear. In this context, implementation of these technologies is equivalent to building a model of the target equipment and conducting failure diagnosis or prediction depending on how far the actual equipment is away from the model. These technologies are called system identification[5] in the field of control engineering. The knowledge of system identification technology can also be applied to the construction of technology for the diagnosis of abnormalities in dynamic systems.

References

1) The Japan Society of Mechanical Engineers (JSME): "Proceedings of the 21st Symposium on Assessment and Diagnosis" (2023).
2) OKATANI Takayuki: "Deep Learning (Revised Edition 2)", Kodansha (2022).
3) NISHIDA, et al.: "Fault Detection for Machine Parts Manufacturing Equipment by Using Multi-label Deep Neural Networks", JSME D&D2021 (2021).
4) SAKA, et al.: "Fault Detection of Production Equipment by Using Spectral Characteristics and Autoencoder", JSME D&D2023 (2023).
5) ADACHI Shuichi: "Basics of System Identification", Tokyo Denki University Press (2009).

Fig. 7 Number of abnormal events exceeding the threshold

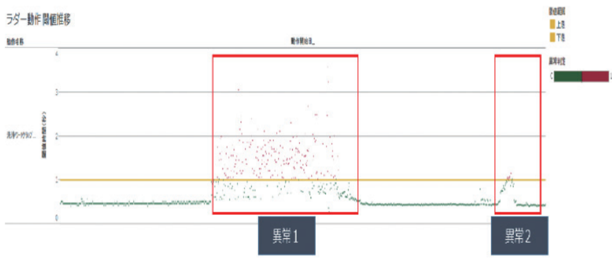

Fig. 8 Abnormality detection graph using control chart

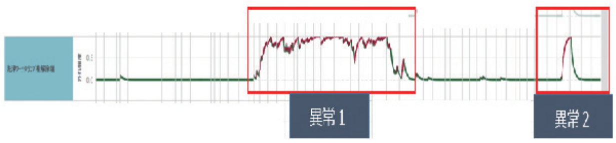

Fig. 9 Abnormality detection graph using ML-DNN

Photo 1 Abnormality 1 (air leakage)

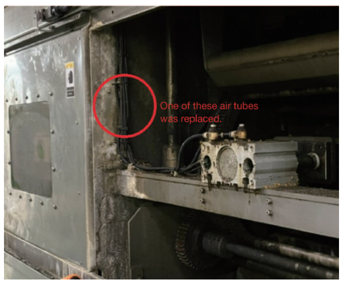

Photo 2 Abnormality 2 (deteriorated tube)