1. Introduction

In general, railway traction machines are rotating machines such as traction motors and reduction gears used in electric cars, and diesel engines and transmissions used in diesel cars. Since these traction machines are essential parts of a railway vehicle, their failure would lead to traffic disruption. Thus, early detection of abnormalities, thereby preventing malfunctions, will improve railway reliability and safety.

Consequently, the authors of this paper are developing a condition monitoring system (CMS) that monitors the state of the traction machines at all times using vibration sensors.

Among various methods used for condition monitoring, vibration monitoring is widely used for rotating machines. However, traction machine vibrations change in a complex way due to operating conditions and vibrations associated with travel. It is difficult therefore to detect abnormalities simply through vibration amplitude. As a result, anomaly detection methods which perform frequency analysis and statistical analysis of vibrations have been proposed [1, 2, 3].

At the same time, anomaly detection is not only for machine failure, since there are various fields in which anomaly detection is needed. In machine learning, which is a field of information engineering, researchers are studying anomaly detection as a general problem and proposing various algorithms that can be applied to various anomaly detection problems. An anomaly detection system applicable to various rotational machines could be achieved by applying these algorithms to machine vibrations. Machine learning is a technology whereby a computer predicts unknown events by learning from past data. In recent years, since it has become feasible to handle large amounts of data thanks to the advances in information and communication technologies, machine learning applications are progressing in various fields, and further developments are expected.

The authors of this paper aim to develop a CMS that combines machine-learning anomaly-detection algorithms and vibration monitoring. This report first gives an overview of the CMS that the authors are developing, and then proposes an anomaly detection method that applies machine learning. Finally this paper presents results obtained from running tests with a diesel car to show the validity of the system.

2. Condition monitoring system using vibration for traction machines

2.1 Overview of the proposed system

When performing vibration monitoring, measured vibration signals are processed and its results are analyzed for anomaly detection in the CMS. In this case, in order to use an ordinary machine learning algorithm, it is necessary to generate a set of the numeric values (feature vectors) that represents characteristics of the vibration by converting the measured vibration signal.

Among various methods for obtaining the feature vectors from the vibration signal, the authors selected the octave band analysis for the first step of the process, for the following reasons:

In general, it is difficult to predict all possible faults and corresponding abnormal vibrations in advance, because there are many different parts and faults that should be considered in complicated machines such as engines.

Nevertheless, machine abnormalities are often found as abnormal noise by train crews or passengers. It means that one can distinguish abnormal sound from normal one and can tell that some failure has occurred, although he does not know what failure has occurred. In other words, from the point of view of detecting the abnormality, it is not necessary to know all of the possible faults and the vibration caused by them in advance. All that is necessary is the vibration in normal condition.
In many cases, a noise is recognized as being abnormal when it sounds unusual and continues for some time. Ideally therefore the processed data for feature vectors should express the features of a steady tone. The results of octave band analysis, which is used in the vibration and the acoustic field, satisfy this requirement. Octave band analysis results were thus selected to be the source of feature vectors in this CMS.

In conventional methods to detect abnormal vibrations, the frequency of the vibration to be detected is commonly determined in advance, while the vibration in the frequency band is extracted with a filter. However, there is a possibility that the frequency of an unexpected abnormal vibration may fall outside the frequency band determined in advance.

In contrast, the method using octave band analysis covers all the frequency bands that have been measured, and it is not necessary to specify the frequency of the vibration to be detected in advance. Therefore, it is applicable even to complicated machines such as engines and should thus be suited to build a versatile condition monitoring system.

Such a condition monitoring system must include a process for measuring vibrations and performing octave band analysis, as well as a process for detecting anomalies. In the current system being developed, it is assumed that the former process is achieved with an “On-board condition monitor” while the latter is performed with a “Diagnostic program.” The assumed configuration of the CMS is shown in Fig. 1.

---

**Table 1 Specifications for signal processing in the on-board condition monitor**

| Item                  | Specification               |
|-----------------------|-----------------------------|
| Octave band analysis  | Band width: 1/12 octave     |
|                       | frequency: 10 Hz – 5 kHz    |
| Sampling period       | 100 ms                      |
| Recording period      | 1 s                         |
| Output file format    | CSV                         |

---

3. Diagnostic program

3.1 Preprocessing

3.1.1 Extraction of steady state

Octave band analysis results recorded in the on-board condition monitor are analyzed in the diagnostic program. Before carrying out anomaly detection, the program performs preprocessing for converting vibration data into feature vectors, which is used in anomaly detection. The preprocessing of the diagnostic program firstly performs steady-state extraction, and then performs principal component analysis of the extracted data. The process for steady-state extraction is described below.

In this anomaly detection quasi-steady-state vibrations are expected. However, transient vibrations occur many times when the vehicle is traveling. Therefore, it is desirable to eliminate the data for the transient vibration.

As shown in Table 1, the octave band analyses are carried out every second and the results are stored in the CMS. Therefore, by comparing the octave band analysis results of interest with those before and after, this program...
determines whether or not it is a steady state, based on differences in the data.

In evaluating differences in octave band analysis output, the results are considered to be the coordinates of a point in a multidimensional space and differences are evaluated by using the distances between the points. In this program, transient data are eliminated by providing a threshold value for the ratio in distance between points of interest and the origin. However, since it is difficult to set an appropriate threshold value beforehand, an optimum value is determined later using test results.

3.1.2 Principal component analysis [4]

Principal component analysis is a method for reducing the dimension of vectors. The amount of data handled is greatly reduced by performing principal component analysis. In addition, if the data are converted to two-dimensional data, one can grasp the tendency of the data intuitively by having the feature vectors shown as points on a two-dimensional plane. Additionally, it is known that the performance of the nearest neighbor method described later is poor with high dimensional data, and it is preferable to reduce dimensions appropriately. Therefore, the program carries out principal component analysis as preprocessing.

In an anomaly detection method in machine learning, a diagnostic program is given normal data that has been measured in advance, as training data. Then, based on information obtained from the training data, the program performs diagnosis of test data. Therefore, this program performs principal component analysis of training data and obtains a transformation matrix for obtaining principal components and standardizing the components in such a way that the variance of the components is one. When making a diagnosis, it calculates the principal components for the training data and test data by using the transformation matrix. Then, the feature vectors used in anomaly detection are determined as vectors made of some principal components chosen from the direction of a greater contribution rate. At that time, since the number of principal components affects the performance of anomaly detection, it is necessary to clarify the optimal number of principal components and the number is also determined later using test results.

3.2 Anomaly detection

3.2.1 Anomaly detection methods in machine learning

In an anomaly detection method in machine learning, there are cases where only normal data are provided as training data and cases where abnormal data are provided together with normal data. In the case of traction machines, it is difficult to list all of the faults to be detected in advance and prepare abnormal data corresponding to their failure. Therefore, it is required to detect abnormality with only the given normal data. These problems are called one-class classification problems. Among methods for one-class classification problems, NNDD (Nearest Neighbor Data Description) [5], which applies the nearest neighbor method (NN), is one of the simplest methods. Although various alternative methods in machine learning fields have been proposed, the authors use only the NNDD in this paper. The application of other methods merits further research.

3.2.2 NNDD

The NN is a simple method used for classification problems. It classifies test data into classes of which training data match closest with test data, by comparing the test data with all the training data. NNDD is a method applying NN to anomaly detection.

In NNDD, each feature vector is considered to be the coordinates of a point on the multidimensional space as shown in Fig. 2. Then, using the locations of the test data X, of the closest training data A, and of the training data B, the degree of abnormality is determined by dividing the distance X ~ A by the distance A ~ B. This value is used to calculate the degree of abnormality, since it is expected that the distance between the test data and the training data will be greater than the distance between the two training data which are close to each other, if the test data corresponds to the abnormal state.

![Fig. 2 Anomaly detection with NNDD](image)

In this method however, when abnormal data is mixed into the training data, abnormal data is used for calculating abnormality which deteriorates detection performance. Therefore, in addition to the nearest training data, the same processing is applied to the second nearest data and the third nearest data, etc. and the final value for the degree of abnormality is calculated as the average value of the ratios of the distances calculated with each nearest data. This averaging should make calculation of the degree of abnormality more stable. At that time, setting the optimum number of nearest training data (number of nearest neighbors) used for the abnormality calculation, is a problem. As such, this parameter is also optimized later using test results.

The final degree of abnormality is calculated as the mean value of the ratios of the distances minus one, so that the degree of abnormality becomes a positive value when it is abnormal and a negative value when it is normal.

3.3 Performance evaluation using the ROC curve

To verify the anomaly detection performance, it is necessary to verify whether abnormal data are discriminated as abnormal. If test data appears abnormal when the degree of abnormality value calculated by the diagnostic program exceeds a certain threshold, reducing the threshold makes it easier to detect abnormal data. On the
other hand, reducing the threshold also makes it easier to misclassify normal data as abnormal. Therefore, in order to evaluate the anomaly detection performance, it is necessary to evaluate not only the ability to determine abnormal data as abnormal, but also ensure that normal data is not classified as abnormal.

The ROC (Receiver Operating Characteristic) curve [4] is an effective tool for this evaluation. When using an ROC curve for the anomaly detection performance evaluation, it represents the false alarm rate (percentage of the normal data misclassified as abnormal) on the horizontal axis, and the detection rate (percentage of the abnormal data classified as abnormal) on the vertical axis. Then the values for the degree of abnormality are calculated with the diagnostic program for test data consisting of normal data and abnormal data. After that, the false alarm rate and the detection rate are calculated for different threshold values. Then finally, plotting the results on the plane produces an ROC curve as in Fig. 3.

If an ROC curve is obtained, the AUC (Area Under the Curve) of the ROC curve is available as an evaluation index plainly representing the anomaly detection performance. The AUC is the area below the ROC curve. In the case of an ideal anomaly detector, the ROC curve goes through the upper-left corner (false alarm rate of 0% and detection rate of 100%) and becomes a poly-line that connects the origin, the upper-left corner, and the upper-right corner. The value of the AUC becomes unity in this case. Thus a high AUC value close to unity means higher anomaly detection performance.

4. Verification with running test results

4.1 Running test

In order to perform the verification of the CMS and the diagnostics program, the vibrations of traction machines were measured during railway vehicle running tests. In the tests, vibrations in normal condition were measured to constitute normal data. Vibrations in simulated abnormal conditions were also measured to produce abnormal data. Ideally, to verify the anomaly detection performance, the test train should run with a fault so that real abnormal vibrations can be measured. However, this is difficult to implement because of safety. It was decided therefore to simulate a fault using a compressor attached to the engine. Running tests were then conducted under the following two conditions:

Case A. with compressor operation (simulated abnormal condition)
Case B. without compressor operation (normal condition)

For both running conditions described above, the test train accelerated from a standstill to a speed of 45km/h; after coasting for a brief period of time, it then decelerated to a stop. The vibration of the traction engine was measured during the test-runs. A total of ten runs were made, five for each case. In case B, the measured data from three round trips was used as learning data, and the data from the remaining two round trips was used as normal test data. The data measured in case A was used as abnormal test data. Engine vibrations were measured with a piezoelectric vibration acceleration sensor on the mounting plate near the mating surface of the engine and the transmission. The measured vibration was processed and recorded in the on-board condition monitor described above. In addition, the signal of the power notch and the train speed were also recorded in the on-board condition monitor.

4.2 Verification of the diagnostic program

4.2.1 Steady-state extraction

The following passage examines the effects of the processes in the diagnostic program. Steady-state extraction, performed as pre-processing, is examined first: in steady-state extraction, a threshold value is used for the difference between serial data. The authors investigated how the end result of the anomaly detection changes with this threshold value. The ROC curves according to the change of the threshold value are shown in Fig. 4.

Since it is necessary to perform all subsequent processing in order to draw an ROC curve, the parameter values in these processes were fixed temporarily. Firstly, the num-
number of principal components was set at two to visualize the distribution of the data. The number of nearest neighbors used in NNDD was set in the simplest form, at one.

According to Fig. 4, the ROC curve with the threshold value of 0.2 has an ideal shape, which makes the AUC nearly one. Otherwise the AUC is of a lower value of 0.7-0.8.

To investigate the data distribution, the scatter plots of the processed data after the steady-state extraction and the principal component analysis are shown in Fig. 5 in two cases of the threshold values of 1.0 and 0.2.

As can be seen from Fig. 5, when the threshold value is 1.0 (the case in which steady-state extraction is hardly performed), since distribution of normal data and abnormal data overlaps, it is difficult to clearly distinguish the two classes. On the other hand, when the threshold value is 0.2, there is no overlap, and the training data and normal test data are distributed substantially in the same position. As a result, the abnormal data can be clearly distinguished, and an ideal ROC curve as shown in Fig. 4 is produced.

Incidentally, when confirming operation states corresponding to the data extracted by the steady-state extraction, it was found that all of them corresponded to idling states. In other words, in these test results, the simulated abnormal states were easily distinguishable only during idling, i.e. when vibrations are stable.

4.2.2 Principal component analysis

As for principal component analysis, anomaly detection performance is examined for changes in the number of principal components to be used. Figure 6 shows ROC curves when varying the number of principal components and fixing the threshold value of the steady-state extraction at 0.2 and the number of nearest neighbors at one.

As in the previous section, when the number of principal components is two, the ROC curve becomes ideal. In contrast, when gradually increasing the number of principal components to 5, the value of the AUC is reduced. However, the value of the AUC is almost 1.0 in the case where the number of principal components is six or above. This is because the principal components from the third to the fifth do not contribute to anomaly detection while the sixth principal component contributes to the anomaly detection in this case.

4.2.3 NNDD

The number of nearest neighbors is a parameter of NNDD. The impact of the parameter on the ROC curve is examined in this section. When the threshold of the steady-state extraction is 0.2, the value of AUC becomes unity regardless of the number of nearest neighbors. Therefore, the threshold value was fixed at 0.4 to make anomaly detection harder by contaminating the training data and the test data with transient vibration data, when generating the ROC curves shown in Fig. 7. The number of principal components was fixed at six in this example.

Figure 7 illustrates that the value of the AUC increases as expected when the number of nearest neighbors increases. However, since the shapes of the ROC curves are
4.2.4 Calculation of the degree of abnormality

Finally, Fig.8 shows the calculated degrees of abnormality of the test data when combining the optimized parameter values. The threshold value of the steady-state extraction, the number of principal components, and the number of nearest neighbors were 0.2, six, and four, respectively.

Figure 8 shows that the normal data and the abnormal data can be clearly distinguished with the degrees of abnormality calculated using the diagnostic program. The above results demonstrates that the proposed method can correctly diagnose the condition of traction machines.

5. Conclusion

In developing a CMS for railway traction machines, a versatile anomaly detection method is needed. Therefore, the authors proposed an anomaly detection method that applies machine learning to the octave band analysis result of the vibration of the machines. The effectiveness of the proposed method was verified using the data obtained from running tests with a diesel car.

The authors proposed to conduct steady-state extraction and principal component analysis as pre-processing. As a result of the verification, it was found that the proposed preprocessing with optimal parameter values was effective in improving the detection performance. The optimal number of principal components was six. It was also found that the diagnostic program can clearly distinguish abnormal data from normal data with NNDD, which is a method of machine learning. In order to perform stable anomaly detection with NNDD even when the training data is contaminated with abnormal data, it is preferable to increase the number of nearest neighbors to more than four.

From above results, it was confirmed that the proposed method has sufficient anomaly detection performance as expected. The authors are going to continue research and development in the future to realize a high-performance CMS.

Acknowledgement

The authors would like to express their sincere gratitude to the related members of the West Japan Railway Company for their indispensable work with the vehicle running tests featured in this paper.

References

[1] Pennacchi, P., Bruni, S., Chatterton, S., Borghesani, P., Ricci, R., Marinis, D., Didonato, A., Unger-Weber, F., "A Test Rig for the Condition-Based Maintenance Application on the Traction Chain of Very High Speed Trains," WCRR2011, 2011.

[2] Oba, T., Yamada, K., Okada, N., and Soma, H., "Condition Monitoring for Shinkansen Bogies Based on Vibration Analysis," Transactions of the Japan Society of Mechanical Engineers, Series C, Vol.75, No.757, pp.93-101, 2009.

[3] Oba, T., Yamada, K., Okada, N., and Tanifuji, K., “Condition Monitoring for Shinkansen Bogies Based on Vibration Analysis (2nd Report, Comparison of Vibration States Between Two Bogies in the Same Cars and Examination of Continuity of Vibration Peaks)," Transactions of the Japan Society of Mechanical Engineers, Vol. 76, No. 769, pp. 27–35, 2010.

[4] Hackeling, G., "Mastering Machine Learning with scikit-learn," Packt Publishing, 2014.

[5] Tax, D.M.J., "One-class classification," Ph.D. thesis, Delft University of Technology, 2001.
Authors

Minoru KONDO, Dr. Eng.
Senior researcher, Drive Systems Laboratory,
Vehicle Control Technology Division
Research Areas: Condition Monitoring, Traction System

Shinichi MANABE
Researcher, Drive Systems Laboratory,
Vehicle Control Technology Division (Former)
Research Areas: Condition Monitoring, Running Resistance

Tatsuro TAKASHIGE
Assistant Senior Researcher, Drive Systems Laboratory, Vehicle Control Technology Division
Research Areas: Condition Monitoring, Diesel Engine

Hiroshi KANNO, Dr. Eng.
Researcher, Drive Systems Laboratory,
Vehicle Control Technology Division
Research Areas: Diesel Engine