Abnormal Sound Detection for Rotary Parts in Noisy Environment by One-class SVM and Non-negative Matrix Factorization

Taichi Kitamura[1], Naoya Takeishi[2], Takehisa Yairi[3], and Koichi Hori[4]

[1,2,3,4]Department of Aeronautics and Astronautics, The University of Tokyo, 113-8656, Japan

kitamura-taichi@g.ecc.u-tokyo.ac.jp
takeishi@ailab.t.u-tokyo.ac.jp
yairi@ailab.t.u-tokyo.ac.jp
hori@g.ecc.u-tokyo.ac.jp

ABSTRACT

In this paper, we introduce a data-driven method for detecting abnormal sound from rotary machines, which is due to small scratches on the surface of bearings, in a manufacturing plant. Since it is difficult to obtain a sufficient amount of anomalous data beforehand, we assume only normal data are available for training a model. The challenge in our situation is the presence of high-level ambient noise, which makes detecting small anomalous sound very hard. In the proposed method, feature vectors are extracted by applying short-time Fourier transform, and one-class SVM is trained on normal data and used to discriminate normal and anomalous data. In addition, ambient noise is removed using nonnegative matrix factorization before extracting features to overcome the problem of noise superimposition and to improve the discrimination precision. We show the results of an experiment using actual sound data obtained from rotary machines with ambient noise.

1. INTRODUCTION

Nowadays, a variety of numerical data from systems, such as plant and infrastructure, are available due to the development of sensors. Accordingly, data-driven anomaly detection techniques based on machine learning are attracting much attention (see, e.g., Chandola et al (2009)). In this paper, we focus on rotary parts of machinery, which exist in a various types of machines. Rotary parts emit abnormal sound when some irregularities, such as defects on the surface of bearings, occur. The purpose of this research is developing a system that immediately detects such abnormal sound. We consider the characteristics of such problem summarized as follows.

First, the difficulty of abnormal sound detection depends on the surrounding environment. For example, it is considerably difficult to detect small abnormal sound in an environment with much ambient noise. In practice, however, a perfectly noise-free environment is very rare in industrial plants. In most cases, there is a certain level of noise, which prevents us from detecting small abnormal sound. Hence, we have to presume a noisy environment. Another important issue is that collecting anomalous data is even more difficult than gathering normal data because abnormal events do not occur as often as normal events do. Hence, we assume that anomalous data (and labels on them) are not contained in the available training data. Due to the above-mentioned reasons, we set our goal to developing an abnormal sound detection method that is robust to ambient noise and is trained using only normal data.

The proposed method is built on one-class support vector machine (OCSVM) as an algorithm for anomaly detection (Schölkopf et al. (2001)) and an extension of nonnegative matrix factorization (NMF), namely, semi-supervised NMF (SSNMF), as a way for noise separation (see, e.g., Lee et al. (2010), Smaragdis et al. (2007), and Nakano et al. (2010)). We combine these methods to develop an abnormal sound detection for rotary parts of machines.

The rest of this paper is organized as follows. First, Section 2 introduces the proposed method and explains how it detects abnormal sound in noisy environments. Next, Section 3 presents the experimental result on actual sound data obtained from rotary parts of machines in a plant. Finally, this paper is concluded in Section 4.

2. PROPOSED METHOD

2.1. Overview

In the proposed method, we use short-time Fourier transform (STFT) as a method for feature extraction from sound data. In addition, we utilize SSNMF as a method for noise reduction (i.e., denoising), and the anomaly detection is finally performed using OCSVM. The overall framework of the proposed method is depicted in Figure 1, wherein data flows of training and test data are manifested. The feature extraction and denoising are performed for both training and test datasets and OCSVM is learned using only training data.

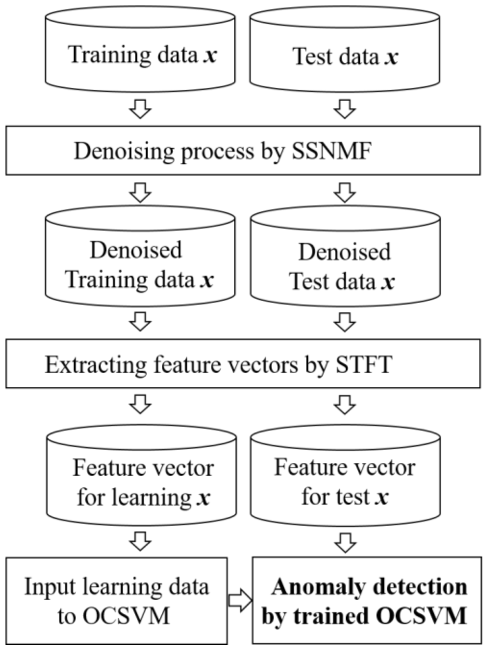

Figure 1. Overall framework of the proposed method.

2.2. Feature Extraction using STFT

We utilize the well-known STFT to extract feature vectors from original sound data. STFT is defined as the following equation.

$$X(\omega, mS) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t+mS) e^{-i\omega t} dt \tag{1}$$

In the above equation, $x(t+mS) \in \mathbb{R}$ represents the original target sound data at timestamp of $t+m$. In addition, $S \in \mathbb{R}$ represents the shift of the window width, and $m \in \mathbb{Z}$ represents the order of the window. Moreover, $X(\omega, mS)$ represents Fourier transformation of $x(t+mS)$.

After applying STFT, we calculate common logarithms of the amplitude spectrum of $X(\omega, mS)$. This procedure is expressed in the following equation.

$$F(\omega, mS) = \log_{10}(|X(\omega, mS)|) \tag{2}$$

We calculate the logarithm of amplitude spectrum because the original amplitude spectrum often becomes very large, which is undesirable from the computational point of view. Starting at the original data $x(t+mS)$, we finally obtain the features $F(\omega, mS)$, which are used in the subsequent procedures of noise reduction by SSNMF and anomaly detection by OCSVM.

2.3. Denoising using SSNMF

We focus on the fact that gathering *pure ambient noise data* is as easy as gathering normal sound data. Hence, suppose we have sufficient amount of pure ambient noise data as well as normal sound data. In addition, we assume that ambient noise has distinctive characteristics because such noise is generated from certain operations in a plant; in other words, we assume the ambient noise is *not* completely random. Then, we use SSNMF (Lee et al. (2010)) for noise reduction. SSNMF is an extension of NMF, which is often applied to sound source separation. Using SSNMF, we extract ambient noise from noise-mixing sound data, whose procedures are described in the following.

First, we extract a basis matrix from pure ambient noise data. This is expressed in the following equation.

$$\boldsymbol{X} \cong \boldsymbol{F}\boldsymbol{G}' \tag{3}$$

In the above equation, $\boldsymbol{X} \in \mathbb{R}_+^{i \times j'}$ is a non-negative matrix that represents the amplitude spectrogram of the original ambient noise data. In addition, $\boldsymbol{F} \in \mathbb{R}_+^{i \times k}$ is a non-negative basis matrix, and $\boldsymbol{G}' \in \mathbb{R}_+^{k \times j'}$ is a non-negative activation matrix. Here, i denotes the number of features (i.e., resolution of STFT), j' denotes the length of ambient noise data, and k is a tunable hyperparameter that represents the number of bases of NMF.

Next, we decompose amplitude spectra of the target sound of rotary parts using SSNMF and basis matrix $\boldsymbol{F}$ extracted from pure ambient noise data. This procedure is expressed in the following equation.

$$\boldsymbol{Y} \cong \boldsymbol{F}\boldsymbol{G} + \boldsymbol{H}\boldsymbol{U} \tag{4}$$

In the above equation, $\boldsymbol{Y} \in \mathbb{R}_+^{i \times j}$ is a non-negative matrix that represents the amplitude spectrogram of the target sound. In addition, $\boldsymbol{G} \in \mathbb{R}_+^{k \times j}$ is a non-negative activation matrix for the target data, and $\boldsymbol{H} \in \mathbb{R}_+^{i \times \ell}$ is a non-negative matrix corresponding to the decomposed basis. Moreover, $\boldsymbol{U} \in \mathbb{R}_+^{k \times j}$ is a non-negative matrix corresponding to the activations of bases in $\boldsymbol{H}$. Note that j denotes the length of the target sound data. Here, decomposed matrix $\boldsymbol{HU}$ corresponds to denoised sound amplitude spectrogram, which we utilize in the sequel.

Finally, we obtain denoised sound data by applying inverse short-time Fourier transformation (ISTFT) to $\boldsymbol{HU}$. ISTFT is expressed in the following equation.

$$x'(t+mS) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} X(\omega, mS) e^{it\omega} d\omega \tag{5}$$

Here, $x'(t+mS)$ denotes denoised sound data. We use the phase information of the original sound data to surrogate the phase of $X(\omega, mS)$ in the above equation. This is expressed in the following equation.

$$X(\omega, mS) = \boldsymbol{HU} \times (\cos\theta + i \sin\theta) \tag{6}$$

$$\theta = \arg(x(t+mS)) \tag{7}$$

In the above equations, θ represents the phase of the original sound.

2.4. Anomaly Detection using OCSVM

We use OCSVM (Schölkopf et al. (2001)) as an anomaly detection algorithm because it needs only normal data for training. In OCSVM with nonlinear kernels, input data $\boldsymbol{x}$ are first mapped to some feature space $\phi(\boldsymbol{x})$, and then the model is trained so that it separates input data from the origin in the feature space. We use RBF kernel in this work. A hyperplane $\boldsymbol{w}$ that separates input data from the origin is expressed as the following equation.

$$\boldsymbol{w}^T\phi(\boldsymbol{x}) - \rho = 0 \quad (8)$$

In the above equation, $\rho \in \mathbb{R}$ denotes the bias in the feature space. By this separation, OCSVM models the regions in which most of data points live (i.e., regions of high-density), and test data are then classified whether they are included in the high-density region or not. This procedure is expressed by following equations.

$$f(\boldsymbol{x}) = \boldsymbol{w}^T\phi(\boldsymbol{x}) \quad (9)$$

$$F(\boldsymbol{x}) = \mathrm{sign}(f(\boldsymbol{x}) - \rho) \quad (10)$$

In the above equation, $F(\boldsymbol{x})$ works as an indicator that tells us whether a new data point $\boldsymbol{x}$ is included in the high-density region, i.e., $F(\boldsymbol{x}) = +1$ means the new data point is normal, whereas $F(\boldsymbol{x}) = -1$ means the new data point is anomalous.

3. EXPERIMENT

3.1. Details of Dataset

We used sound of a rotary bearing of a machine for an experiment and prepared three types of sound datasets: normal sound data, abnormal sound data, and pure ambient noise data. The normal sound data comprise sound of a rotary bearing working normally and ambient noise generated from the operation of an oil factory. The abnormal sound data comprise sound of a rotary bearing with scratches on its surface, again contaminated with the same sort of ambient noise. The pure ambient noise data comprise only the sound in a running oil factory.

We used a part of the normal sound data and the noise data as a training dataset and made two types of mixed sound data, which included normal sound data and abnormal sound data. One of the mixed sound datasets was used for validation to determine the values of hyperparameters, which was composed of the normal sound (for 20 seconds) and the abnormal sound (for 20 seconds). The other mixed sound dataset was used for testing, which was composed of the normal sound (for 60 seconds) and the abnormal sound (for 60 seconds).

In the following table, we summarize the configurations of the datasets we prepared in the experiment.

Table 1. Configuration of the datasets.

Dataset	Details
Normal sound data	Normal bearing sound + noise
Abnormal sound data	Abnormal bearing sound + noise
Ambient noise data	Only ambient noise
Validation dataset	Normal(20 sec)+Abnormal(20 sec)
Test dataset	Normal(60 sec)+Abnormal(60 sec)

3.2. Experimental setup

We conducted an experiment using the proposed method and the above-mentioned datasets. We applied the proposed framework (referred to as *denoised*) to the test data and measured accuracy of anomaly detection. In addition, we conducted the same sort of experiment by applying the proposed method without the denoising step (referred to as *simple*) to the test data for comparison. We verified the effectiveness of the denoising process by comparing the accuracy of *denoised* and *simple* methods.

We used the validation data to find proper values of the parameter of OCSVM. Parameters of OCSVM to be tuned are γ and ν; γ is the width of RBF kernel, and ν is the presumed proportion of abnormality in training data. We decided those parameters by grid search.

We applied two types of systems (*denoised* and *system*) to the test data. In this experiment, the test data (120 seconds) were converted into 20,676 feature vectors, whose dimensionality was 1,024, by STFT, i.e., in the test data, the number of normal data point was 10,338, and the number of anomalous data point was 10,338. We calculated the precision and accuracy to evaluate the performances of two types of systems. Furthermore, we computed receiver operating characteristics (ROC) curves and their area under the curve (AUC). We used the distance from decision boundary of OCSVM insteacod of the probability of presumed labels to draw ROC curve.

3.3. Results

The result of the experiment is summarized in the following tables and figures.

Table 2. Precisions and accuracy of anomaly detection.

	denoised	*simple*
Precision for normal data	99 %	84%
Precision for anomalous data	88 %	83%
Accuracy	92.6 %	83.6%

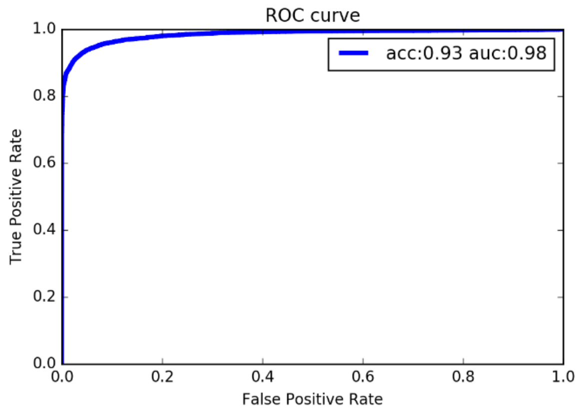

Figure 2. ROC curve for *denoised* method and AUC corresponding to this ROC curve.

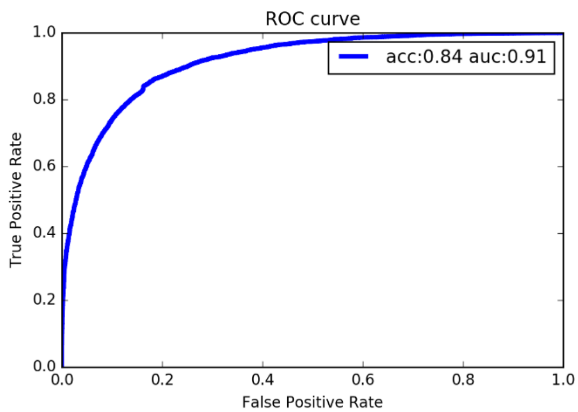

Figure 3. ROC curve for *simple* method and AUC corresponding to this ROC curve.

We can confirm the effectiveness of the proposed method by comparing the above results of *denoised* method and those of *simple* method. From viewpoint of every measure (precisions, accuracy, ROC, and AUC), *denoised* method is better than *simple* method, which indicates the validity of the noise reduction step of the proposed method.

4. CONCLUSION

In this paper, we have proposed an abnormal sound detection method for rotary parts of machines, which combines the anomaly detection by OCSVM and the ambient noise removal by SSNMF.

We confirmed the effectiveness of the proposed method. Denoising by SSNMF improves the accuracy of anomaly detection in noisy environment. As future work, applying our proposed method to data obtained from a variety of noisy environment is an important and necessary challenge.

ACKNOWLEDGEMENT

The part of this research was conducted as collaborative work with JX Nippon Oil & Energy Corporation. We would like to express our gratitude to Haruhisa Goto and Keisuke Nakayama for providing valuable experimental data and discussion.

REFERENCES

Chandola, V., Banerjee, A. & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41 (3), 1–58.

Schölkopf, B. et al. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13 (7), 1443–1471.

Lee, H., Yoo, J. & Choi, S. (2010). Semi-supervised nonnegative matrix factorization. *IEEE Signal Processing Letters*, 17 (1), 4–7.

Smaragdis, P., Raj, B. & Shashanka, M. (2007) Supervised and semi-supervised separation of sounds from single-channel mixtures. *Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation* (414–421).

Nakano, M. et al. (2010). Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with β-divergence. *Proceedings of 2010 IEEE International Workshop on Machine Learning for Signal Processing* (283–288).