

# 雑音下異常検知における前処理としての NMF 音源抽出手法の検討

相場 亮人<sup>1,a)</sup> 吉田 実<sup>1</sup> 後藤 理<sup>1</sup> 北村 大地 高道 慎之介<sup>2</sup> 猿渡 洋<sup>2</sup>

**概要:** 収録された動作音をもとに、外れ値検知手法によって設備の異常を検知するシステムについて検討した。実環境では、周囲雑音に対して監視対象の動作音の SNR が低く、異常を検知するには雑音に埋もれたわずかな音の変化を捉えなければならないという課題がある。本研究ではこの課題に対して、非負値行列因子分解 (NMF) による音源抽出処理を、外れ値検知処理の前段に設ける構成を提案する。提案システムによって低 SNR 環境における異常検知の性能がどれだけ改善されるか、評価実験を行った。今回の実験では、NMF を前処理に設けることで、異常検知可能な SNR が大幅に改善された。

**キーワード:** 非負値行列因子分解, 半教師あり NMF, 音源抽出, 外れ値検知, 異常音検知

## NMF Source Extraction as Pre-processing of Anomaly Detection under the Noise Environment

AIBA AKIHITO<sup>1,a)</sup> YOSHIDA MINORU<sup>1</sup> GOTO SATORU<sup>1</sup> KITAMURA DAICHI TAKAMICHI SHINNOBUKE<sup>2</sup>  
SARUWATARI HIROSHI<sup>2</sup>

**Abstract:** We studied anomaly detection system for equipments by outlier detection method based on recorded sounds. In the real environment, SNR of the target sound against background noise is low, and there is a problem that it is necessary to catch the slight change in sound buried in noise. In this research, we propose a configuration in which a sound source extraction process by nonnegative matrix factorization (NMF) is provided at the preliminary stage of outlier detection process. We evaluated the improvement of the anomaly detection performance in the low SNR environment. In this experiment, SNR capable of detecting anomaly was greatly improved by providing NMF for pre-processing.

**Keywords:** Non-negative matrix factorization, Semi-supervised NMF, Source extraction, Outlier detection, Anomaly sound detection

### 1. はじめに

給排水設備や空調設備などの状態を人間の代わりに常時センシングして、異常動作時に通知する自動監視システムがいくつか提案されている [1], [2]. ここでは、マイクロ

フォンによって収録した動作音をもとに、設備の異常を検知するシステムを考える。

異常検知アプリケーションにおいて大きな問題のひとつが、異常時のサンプルを入手するのが困難なことである。特に、高価で施設等の運営管理に重要な設備類は、サンプル収集のために意図的に異常状態を作り出すようなことは難しい。また、自然に異常が発生するのを待っても、数か月や数年に一度しか発生しないような種類の異常も存在する。そのため、入手の容易な正常時のサンプルのみをあらかじめ収集してその特徴をモデリングし、その特徴から外

<sup>1</sup> 株式会社リコー 研究開発本部  
Ricoh Company Ltd., Ebina, Kanagawa 243-0460, Japan  
<sup>2</sup> 東京大学 大学院情報理工学系研究科  
The University of Tokyo, Bunkyo, Tokyo 113-8656, Japan  
<sup>3</sup> 香川高等専門学校 電気情報工学科  
National Institute of Technology, Kagawa College, Takamatsu, Kagawa 761-8058, Japan  
a) akihito.aiba@jp.ricoh.com

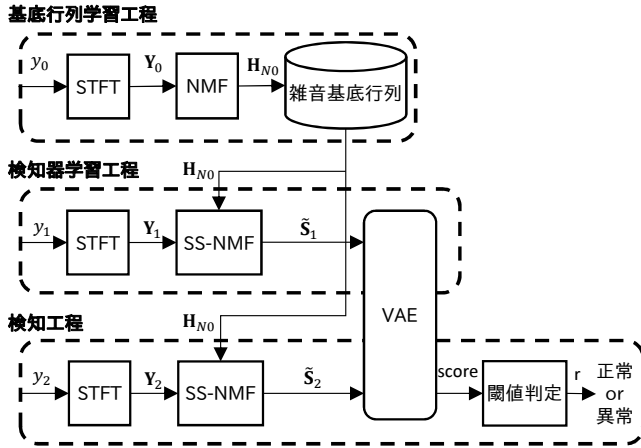


図 1 提案システムの概要図  
Fig. 1 Overview of proposed system

れたサンプルが得られた際にそれを異常と判定する，外れ値検知のアプローチがよく採られている [3], [4]．本研究でも，この外れ値検知のアプローチを用いる．

音データを入力として用いる場合には，実環境における周囲雑音の大きさも問題となる．大規模な施設においては，複数の設備機器が1つの部屋の中に設置され，非常に大きな動作音を発生させている状況がよく見られる．そのような状況では，1つの機器の動作音が異常な特徴を示したとしても，その他の機器の動作音に対してSNRが低いために，検知難易度が非常に高くなってしまふ．このような問題に対するアプローチのひとつに，外れ値検知などの識別処理の前段で，何らかの音源抽出処理を行うことで，検知対象音のSNRを改善し，検知をより容易にする方法がある [5], [6]．本研究では，この音源抽出処理として非負値行列因子分解 (NMF) [7] を用いる異常検知システムを提案し，その効果を検証する．

## 2. 提案システム

### 2.1 システム概要

図 1 に提案システムの概要を示す．提案システムは，大きく以下の3つの工程から構成される．

- (1) 周囲雑音スペクトログラムの基底行列を学習する基底行列学習工程
- (2) 抽出した正常動作時の検知対象音を検知器に学習させる検知器学習工程
- (3) 抽出した検知対象音が異常かどうかを検知する検知工程

以降，各工程について説明する．

### 2.2 基底行列学習工程

本システムではまず，NMFによる音源抽出処理のために必要な，周囲雑音の基底行列を求める．そのためには，周囲雑音のデータを収集することが必要である．

そこで，検知対象の動作音が発生していない状況の音  $y_0$  を収録し，短時間フーリエ変換 (STFT) してスペクトログラム行列  $\mathbf{Y}_0 \in \mathbb{R}_{\geq 0}^{\Omega \times T_0}$  を得る． $\mathbf{Y}_0$  は検知対象の動作音を含まないため，周囲雑音の非負スペクトログラム行列  $\mathbf{N}_0 \in \mathbb{R}_{\geq 0}^{\Omega \times T_0}$  に該当する．この非負行列を，NMFを用いて基底行列  $\mathbf{H}_{N0} \in \mathbb{R}_{\geq 0}^{\Omega \times K_N}$  及び非負のアクティベーション行列  $\mathbf{U}_{N0} \in \mathbb{R}_{\geq 0}^{K_N \times T_0}$  に分解する．

$$\mathbf{Y}_0 = \mathbf{N}_0 \simeq \mathbf{H}_{N0} \mathbf{U}_{N0} \quad (1)$$

$\mathbf{H}_{N0}$  および  $\mathbf{U}_{N0}$  は，全要素が非負の制約条件化での行列間距離最小化問題より推定できる．

$$\mathbf{H}_{N0}, \mathbf{U}_{N0} = \arg \min_{\mathbf{H}_{N0}, \mathbf{U}_{N0}} \mathcal{D}_{IS}(\mathbf{Y}_0 | \mathbf{H}_{N0} \mathbf{U}_{N0}) \quad (2)$$

$\mathcal{D}_{IS}$  は次式に示す板倉斎藤距離を表す．

$$\mathcal{D}_{IS}(\mathbf{B} | \mathbf{A}) = \sum_{m,n} \left\{ \frac{b_{m,n}}{a_{m,n}} - \log \frac{b_{m,n}}{a_{m,n}} \right\} \quad (3)$$

また， $a_{m,n}$  および  $b_{m,n}$  はそれぞれ非負行列  $\mathbf{A}$  と  $\mathbf{B}$  の要素である．以上より，周囲雑音のスペクトルパターンを含む基底行列  $\mathbf{H}_{N0}$  を学習することができる．

### 2.3 検知器学習工程

検知器学習工程では，前節の基底行列学習工程で得られた基底行列  $\mathbf{H}_{N0}$  を用いて，周囲雑音と検知対象の動作音が混合した音から検知対象音のみを抽出し，推定された検知対象音を用いて異常検知器の学習を行う．まず，検知対象の動作音が発生している状況の音  $y_1$  を収録し，これをSTFTしてスペクトログラム行列  $\mathbf{Y}_1 \in \mathbb{R}_{\geq 0}^{\Omega \times T_1}$  を得る．ここで  $\mathbf{Y}_1$  は，検知対象音の非負スペクトログラム行列  $\mathbf{S}_1 \in \mathbb{R}_{\geq 0}^{\Omega \times T_1}$  と，周囲雑音の非負スペクトログラム行列  $\mathbf{N}_1 \in \mathbb{R}_{\geq 0}^{\Omega \times T_1}$  の和からなると近似できる．

$$\mathbf{Y}_1 \simeq \mathbf{S}_1 + \mathbf{N}_1 \quad (4)$$

本工程では，異常検知器の学習データ (検知対象の正常動作時の音) として， $\mathbf{Y}_1$  に含まれる検知対象音  $\mathbf{S}_1$  を用いる．その事前処理として，事前学習した周囲雑音の基底行列  $\mathbf{H}_{N0}$  を用いて，観測音  $\mathbf{Y}_1$  から  $\mathbf{S}_1$  のみを抽出する．ここでは，次式に示す半教師あり NMF (Semi-supervised NMF, SS-NMF) [8], [9] を適用する．

$$\mathbf{Y}_1 \simeq \mathbf{H}_{S1} \mathbf{U}_{S1} + \mathbf{H}_{N0} \mathbf{U}_{N1} \quad (5)$$

$$\tilde{\mathbf{S}}_1 = \mathbf{H}_{S1} \mathbf{U}_{S1} \quad (6)$$

$$\tilde{\mathbf{N}}_1 = \mathbf{H}_{N0} \mathbf{U}_{N1} \quad (7)$$

ここで， $\mathbf{H}_{S1} \in \mathbb{R}_{\geq 0}^{\Omega \times K_S}$ ， $\mathbf{U}_{S1} \in \mathbb{R}_{\geq 0}^{K_S \times T_1}$  はそれぞれ検知対象音の基底行列およびアクティベーション行列， $\mathbf{U}_{N1} \in \mathbb{R}_{\geq 0}^{K_N \times T_1}$  は周囲雑音のアクティベーション行列である． $\mathbf{H}_{S1}$ ， $\mathbf{U}_{S1}$ ，および  $\mathbf{U}_{N1}$  は，学習済みの基底  $\mathbf{H}_{N0}$  が与えられた下での，次式の非負制約付き行列間距離最小

化より推定できる。

$$\mathbf{H}_{S1}, \mathbf{U}_{S1}, \mathbf{U}_{N1} = \arg \min_{\mathbf{H}_{S1}, \mathbf{U}_{S1}, \mathbf{U}_{N1}} D_{\beta}(\mathbf{Y}_1 | \mathbf{H}_{S1} \mathbf{U}_{S1} + \mathbf{H}_{N0} \mathbf{U}_{N1}) \quad (8)$$

なお、 $\mathbf{H}_{N0}$  は上式の最適化において固定され、更新を行わない。以上の処理により、検知対象音の非負スペクトログラム行列  $\tilde{\mathbf{S}}_1$  を  $\mathbf{Y}_1$  から抽出する。

次に、抽出した  $\tilde{\mathbf{S}}_1$  を学習データとして用い、異常検知器の学習を行う。ここでは、 $y_1$  の収録時には全区間で検知対象は正常動作していたという仮定を置いている。

本システムでは Variational Autoencoder (VAE) による外れ値検知手法 [4] を用いる、ここで VAE はスペクトログラム行列  $\tilde{\mathbf{S}}$  が入力されると、以下のような流れで各行列要素に対応したガウス分布の平均と分散に対応する値を出力する。

- (1) エンコーダ  $f$  が、入力  $\tilde{\mathbf{S}}$  に対応する潜在変数  $z$  の分布パラメータとして平均  $\mu_z$  と分散  $\sigma_z$  を出力する。

$$\mu_z(\tilde{\mathbf{S}}), \sigma_z(\tilde{\mathbf{S}}) = f(\tilde{\mathbf{S}}) \quad (9)$$

- (2) 分布パラメータ  $\mu_z, \sigma_z$  をもとに、潜在変数  $z$  をサンプリングする。

$$z \sim N(\mu_z(\tilde{\mathbf{S}}), \sigma_z(\tilde{\mathbf{S}})) \quad (10)$$

- (3) デコーダ  $g$  が、潜在変数  $z$  に対応する入力  $\tilde{\mathbf{X}}$  の分布パラメータ  $\mu_{\tilde{\mathbf{X}}}, \sigma_{\tilde{\mathbf{X}}}$  を出力する。

$$\mu_{\tilde{\mathbf{S}}}(z), \sigma_{\tilde{\mathbf{S}}}(z) = g(z) \quad (11)$$

学習は、この分布パラメータに基づく正規分布に対する、入力スペクトログラム行列  $\tilde{\mathbf{S}}_1$  の負の対数尤度が小さくなるように行う。学習時の損失関数  $\mathcal{L}$  は次式のようになる。

$$\mathcal{L}_{lat} = D_{KL} \left[ N(\mu_z(\tilde{\mathbf{S}}_1), \sigma_z(\tilde{\mathbf{S}}_1)) \parallel N(0, I) \right] \quad (12)$$

$$\mathcal{L}_{rec} = \frac{1}{L} \sum_{l=1}^L -\log p(\tilde{\mathbf{S}}_1; \mu_{\tilde{\mathbf{X}}}(z^{(l)}), \sigma_{\tilde{\mathbf{X}}}(z^{(l)})) \quad (13)$$

$$\mathcal{L} = \mathcal{L}_{lat} + \mathcal{L}_{rec} \quad (14)$$

ここで、 $l = 1, \dots, L$  は潜在変数  $z$  のサンプリング回数の番号を表す。 $D_{KL}[B||A]$  は分布  $A, B$  間の KL ダイバージェンスを表す。

以上により、 $\mathcal{L}_{lat}$  による正規化のもとで、正常動作中の検知対象音のスペクトログラムに対する  $\mathcal{L}_{rec}$  が最小化される。そのため、異常動作中のスペクトログラムのような外れ値に対して  $\mathcal{L}_{rec}$  は大きくなることが期待できる。

## 2.4 検知工程

検知工程では、雑音環境下で検知対象の動作音を収録した音  $y_2$  を STFT し、非負スペクトログラム行列  $\mathbf{Y}_2 \in \mathbb{R}_{\geq 0}^{\Omega \times T_2}$  を得る。そして、学習済みの周囲雑音の基底行列  $\mathbf{H}_{N0}$  を

用いた SS-NMF によって、 $\mathbf{Y}_2$  から検知対象音の非負スペクトログラム行列  $\tilde{\mathbf{S}}_2$  を抽出する。

$$\mathbf{Y}_2 \simeq \mathbf{H}_{S2} \mathbf{U}_{S2} + \mathbf{H}_{N0} \mathbf{U}_{N2} \quad (15)$$

$$\tilde{\mathbf{S}}_2 = \mathbf{H}_{S2} \mathbf{U}_{S2} \quad (16)$$

抽出した  $\tilde{\mathbf{S}}_2$  を VAE の入力とし、異常スコア  $score$  を次のように求める。

$$z \sim N(\mu_z(\tilde{\mathbf{S}}_2), \sigma_z(\tilde{\mathbf{S}}_2)) \quad (17)$$

$$score = \frac{1}{L} \sum_{l=1}^L -\log p(\tilde{\mathbf{S}}_2; \mu_{\tilde{\mathbf{X}}}(z^{(l)}), \sigma_{\tilde{\mathbf{X}}}(z^{(l)})) \quad (18)$$

$score$  が閾値  $\alpha$  以上であれば、その入力では異常が発生していると判断する。

## 3. 評価実験

前述の提案システムについて、異常識別性能を評価した。ここでは、NMF による前処理を行わないシステム（基底行列学習工程が無く、その他工程での SS-NMF を行わないシステム）を比較対象とする。

### 3.1 実験条件

検知対象音のサンプルとして、Case Western Reserve University Bearing Data Center [10] が公開しているデータセットから正常/異常の2種類のデータを用いた。このデータセットはモータの軸受の振動を加速度センサで取得した信号データで構成されており、正常な軸受のほか、傷がついた異常軸受の振動データも含まれている。正常データとして用いたのは”Normal Baseline Data”の”Normal\_2”，異常データとして用いたのは”48k Drive End Bearing Fault Data”の”OR007@6.2”である。

一方、周囲雑音としては株式会社リコー横浜仲町台事業所の設備室で収録した音を用いた。室内ではポンプなどの設備が複数稼動している。収録時間は全部で約 8[h] (28500[s]) である。

これらを重畳させるなどして、表1のように実験用データを準備した。基底行列学習用の周囲雑音と識別器学習用の周囲雑音は別区間のものである。また、識別テスト用の約 8[h] の周囲雑音には、基底行列学習と識別器学習に用いた計 4[h] の区間も含まれている。周囲雑音の音量は各データセットで一定である。すなわち、SNR が変わると軸受側の音量のみが変わっている。なお、軸受音データは正常異常どちらも 10[s] ほどの長さしかないので、周囲雑音データは 10[s] ごとに分割し、それぞれを軸受音データと混合している。

異常検知器として用いる VAE は、エンコーダ、デコーダともに 4 層の畳み込み層と Maxpooling 層から構成した。スペクトログラム行列を 1 チャネルの画像入力のように扱う。エンコーダの畳み込み層のカーネルサイズは  $5 \times 5$ ,

表 1 実験用データ

Table 1 Experimental data

	内訳	混合比 (SNR)	時間
基底行列学習	周囲雑音	-	1[h]
識別器学習	正常軸受+周囲雑音	-24[dB]	3[h]
識別テスト 識別テスト	正常軸受+周囲雑音	-24[dB]	8[h]
		-6[dB]	8[h]
	異常軸受+周囲雑音	-12[dB]	8[h]
		-18[dB]	8[h]
		-24[dB]	8[h]

表 2 評価実験条件

Table 2 Experimental conditions

サンプリング周波数		48000[Hz]
解析窓長/シフト幅		10[ms] / 5[ms]
FFT 点数		512
NMF	基底数 ( $K_N$ )	256
	基底数 ( $K_S$ )	1
	特徴量	振幅スペクトル (256 次元)
	$\beta$	0.0
NN	特徴量	パワースペクトル (256 次元 $\times$ 32 フレーム)
	構造	エンコーダ, デコーダともに 4 層の畳み込み層と Maxpooling 層 から構成された VAE
	エポック数	100

チャンネル数は 8 (1 層目), および 4 (それ以外の層), ストライド幅は 1 とした. デコーダでは逆の構成となる. 畳み込み層の出力には batch normalization[11] を行う. 活性化関数には ReLU 関数 [12] を用いた. ただし, 分布の分散が極端な値になるのを避けるため, 3 層目の出力のみ tanh 関数を用いる. また, 分散が負の値, または 0 になるのを避けるために, 分散を直接出力するのではなく, その対数値を出力させている.

その他の実験条件は表 2 に示す.

### 3.2 実験結果および考察

識別テスト用データに対する, 10 秒間ごとの異常スコアの平均値の推移を図 2, および図 3 に示す. 図 2 が NMF 処理なし, 図 3 が NMF 処理ありでの結果である.

SNR が低いほど, 異常データに対するスコアが正常データのスコアと近いことが見て取れる. NMF 処理なし, ありの二つの結果を比較すると, NMF を前処理に用いた場合は, より低い SNR の異常に対しても, 正常データとのスコアの差が大きくなっていることがわかる.

図 2, 図 3 のスコアに対して, 閾値を設定して正解率を求めた結果を図 4, 図 5 に示す. 正解率は, 正常データに対してはスコアが閾値未満となったデータの割合, 異常データに対してはスコアが閾値以上となったデータの割合とする. 正常データの正解率と異常データの正解率はトレード

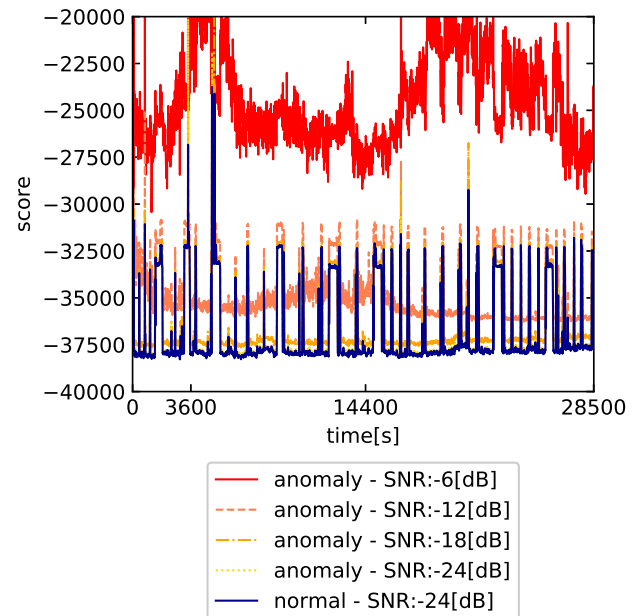


図 2 NMF 処理なしでの異常スコア

Fig. 2 Anomaly score without NMF processing

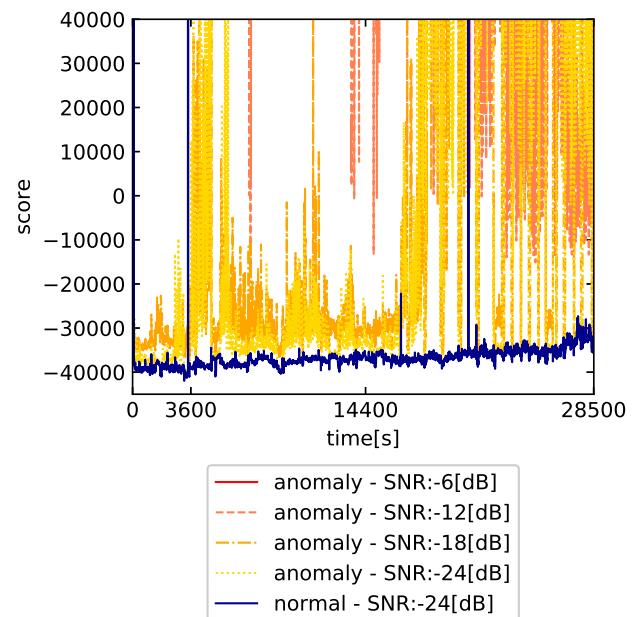


図 3 NMF 処理ありでの異常スコア

Fig. 3 Anomaly score with NMF processing

オフの関係にあるが, NMF 処理ありの結果ではどちらもある程度高いレベルで保つことができている.

代表値として, SNR=-24[dB] の正常データと, SNR=-18[dB] の異常データに対する正解率がおおよそ等しくなる閾値での正解率を表 3 に示す.

この結果からも, NMF を前処理に用いたことで低 SNR での異常検出性能が向上していることがわかる.

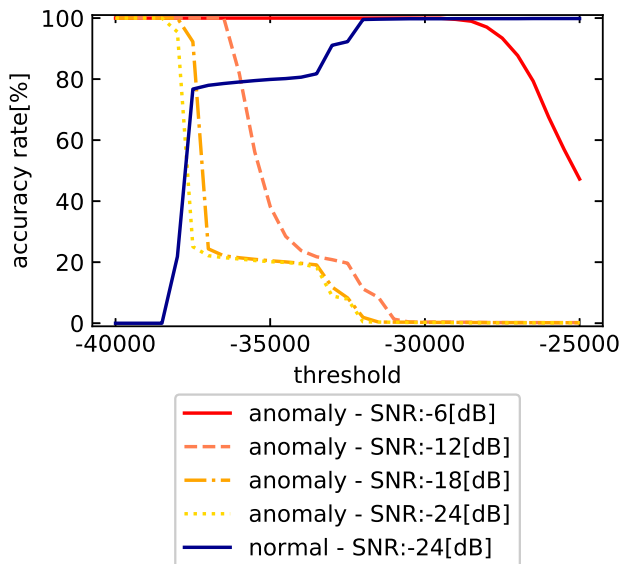


図 4 NMF 処理なしでの正解率

Fig. 4 Accuracy rate without NMF processing

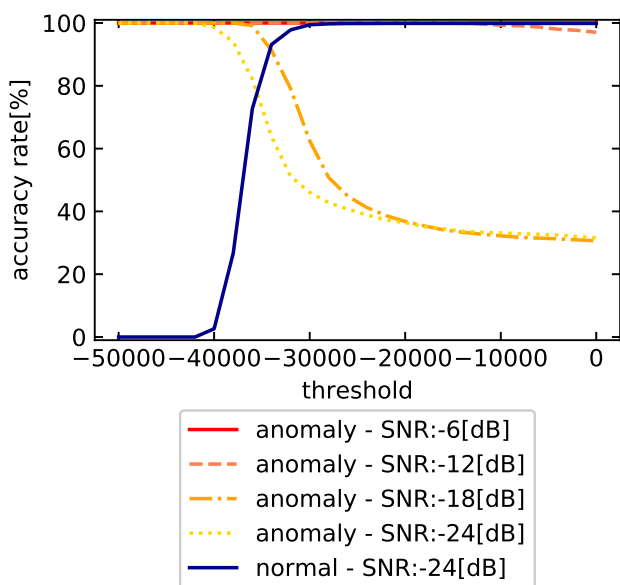


図 5 NMF 処理ありでの正解率

Fig. 5 Accuracy rate with NMF processing

表 3 正解率

Table 3 Accuracy rate

データ種類	正解率 [%]	
	NMF 処理無し	NMF 処理あり
正常 (-24[dB])	77.3	92.5
異常 (-6[dB])	100.0	100.0
異常 (-12[dB])	100.0	100.0
異常 (-18[dB])	77.2	92.4
異常 (-24[dB])	23.4	65.8

#### 4. まとめ

本研究では、低 SNR 環境における異常検知性能の改善のために、外れ値検知手法による異常検知の前処理に NMF を用いる構成を提案した。評価実験の結果、提案システムは低い SNR の異常音に対する検知性能が向上していることが確認できた。

今後の課題としては、別環境での性能確認、未知雑音への適応の検討などが挙げられる。

#### 参考文献

- [1] 吉澤亜耶, 橋本洋一: 異常検知技術の概要と応用動向について, INTEC TECHNICAL JOURNAL 17, INTEC (2016).
- [2] Chandola, V., Banerjee, A. and Kumar, V.: Anomaly detection: A survey, *ACM computing surveys (CSUR)*, Vol. 41, No. 3, p. 15 (2009).
- [3] Singh, K. and Upadhyaya, S.: Outlier detection: applications and techniques, *International Journal of Computer Science Issues*, Vol. 9, No. 1, pp. 307–323 (2012).
- [4] An, J. and Cho, S.: Variational autoencoder based anomaly detection using reconstruction probability, *SNU Data Mining Center, Tech. Rep.* (2015).
- [5] Qiu, H., Lee, J., Lin, J. and Yu, G.: Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, *Journal of sound and vibration*, Vol. 289, No. 4-5, pp. 1066–1090 (2006).
- [6] Kitaoka, N., Akahori, I. and Nakagawa, S.: Speech recognition under noisy environments using spectral subtraction with smoothing of time direction and real-time cepstral mean normalization, *International Workshop on Hands-Free Speech Communication* (2001).
- [7] Lee, D. D. and Seung, H. S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol. 401, No. 6755, p. 788 (1999).
- [8] Smaragdakis, P., Raj, B. and Shashanka, M.: Supervised and semi-supervised separation of sounds from single-channel mixtures, *Proceedings of the 7th international conference on Independent component analysis and signal separation*, Springer-Verlag, pp. 414–421 (2007).
- [9] Kitamura, D., Saruwatari, H., Yagi, K., Shikano, K., Takahashi, Y. and Kondo, K.: Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. 97, No. 5, pp. 1113–1118 (2014).
- [10] Case Western Reserve University: Bearing Data Center, <http://csegroups.case.edu/bearingdatacenter/home>.
- [11] Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv preprint arXiv:1502.03167* (2015).
- [12] Glorot, X., Bordes, A. and Bengio, Y.: Deep sparse rectifier neural networks, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315–323 (2011).