

# LinearSVC による機械音異常検知

大柴義永・富山哲男・西田健次\*・宮崎淳\*

## Abnormal machine noise detection with LinearSVC

Yoshito Oshiba, Tetsuo Tomiyama, Kenji Nishida\* and Jun Miyazaki\*

**Abstract:** With the advent of Industry 4.0, the use of machine sound abnormality detection systems has become increasingly important for the maintenance of industrial equipment used in fully automated plants and warehouses. In this paper, we conducted machine sound anomaly detection using supervised learning on bearings. An excellent learning model was constructed, achieving an F1 score of 0.95 and an AUC of 0.94 in the test validation.

**Keywords:** Abnormal machine noise detection, SVC

### 1. はじめに

環境音から得られる音響情報から機械やモノなどの異常状態を発見する異常音検知の研究は応用範囲が広い。先行研究 [1] を見ると、音響情報を用いた監視システム、特に事件事故に関連する銃声 [2,3]、悲鳴 [4]、爆発音 [5] などの検知。また、医療・メンタルヘルス分野では高齢者の転倒音 [6] や肺音からの異常検知 [7] などの研究が行われている。近年では、Industry4.0 に伴い全自動の工場や倉庫に用いられる産業機器の保守 (メンテナンス) への応用に需要が高まっている。従来は、熟練の検査作業員が機械音を聞くことでメンテナンスが行われているが、機械音異常検知を用いると検査作業員育成にかかる時間、メンテナンスに対する人的時間的労力を削減し、生産性、効率性の向上が可能になると期待されている。

機械音異常検知に関する従来研究では、wavelet 解析をベースとした異常音検知 [8,9] が行われていたが、近年の傾向として機械学習を用いた異常音検知の研究が進められている。機械学習に用いられる学習方法として、教師あり異常音検知と教師なし異常音検知の2種類に分類される。機械音異常検知の課題として、まず産業機器が故障する頻度が少ないことから異常データの取得が困難という点が挙げられる [10,11]。次に機器の故障するパターンは無数にあり異常状態を定義するのが困難であり大部分が未知である。さらに、産業機器によって正常状態であっても機器の運転状況によってパターンの変化なのか損傷であるのかの判断が困難である。

この課題を踏まえて教師なし学習では、異常時のデータを使わずに正常状態のみをデータとして用いた機械音異常検知の研究が行われている。基本的にはクラスタリング (混合

---

\* 株式会社オレンジテクラボ

ガウス分布モデル、自己組織化マップなど)による外れ値検知を行う事で通常とは違う音、つまり異常と定義され研究が行われていることが多い[10,11,13]。

また、教師あり学習に関しても異常検知に対する研究が行われている。ただ、機械音異常検知の課題を述べたように、異常時のデータの取得が困難であったり、異常状態の大半が未知であるという課題から学習に必要なデータが得られないという課題がある。これらの課題に対して、検知対象となる機械を運用していく中で僅かに得られた異常データや機器によって異常データを取得可能な事象を活用して、GMM[2]、HMM[5]など教師あり学習を用いて2値分類を行う研究が行われている。また、教師あり学習と教師なし学習双方を比べた時、教師あり学習の方が高い精度で異常検知できることが分かっている[12]。

以上を踏まえ実運用を想定すると、異常データを使わず正常状態のみで行う教師なし学習の方が実装はしやすい。一方で、機械を運用している中で僅かに得られた異常データなどを使った教師あり学習では異常状態(異常音)を明示的に定義することが可能であるという利点がある。さらに、教師あり学習の精度を上げるために、教師なし学習も交えた判定精度を高める研究[12]が行われている。

本稿では、機械音異常検知の実運用を想定し、僅かに得られた異常時の音響データを活用した教師あり学習を行った。一部の異常状態を明示的に定義する実験を行った上で、教師あり学習で得られた結果を教師なし学習へのデータとして活用することで教師なし学習の精度向上を目的としている。

## 2. 方法論

学習に用いたデータセットは世界最大の音響認識分野における国際コンペティション DCASE2022 (Detection and Classification of Acoustic Scenes and Events) の task2 である(<https://dcase.community/challenge2022/index>)。出題されるタスクには音響シーン分類、音響イベント検出、音響イベント音源推定検知などが用意されている。2022 年の task2 には7種類の機械を対象とした異常音検知タスクが出題された。ファン、ギアボックス、ベアリング、スライドレール、おもちゃの車、列車、バブルが対象となり、それぞれ付属の情報が与えられている。今回は回転機器であるベアリングを対象とした異常音検知を行った。ベアリングに与えられるデータには3つの異なる条件が組み合わされて、1. 回転数が異なるもの。2. マイクの位置を変えたもの。3. ノイズが異なるものが用意されていた。1. 回転数が異なるもの、2. マイクの位置を変えたものを使い、正常時のデータ 200 個、異常時のデータ 200 個の計 400 個の音響データを学習データとした。400 個の内、200 個には12種類の回転速度、もう 200 個には2種類の回転速度が9種類の異なる位置にマイクをセットし音響データを取得したデータとなっている。これらのデータを用いて音響データを識別器に掛けた時に正常、異常の二値分類ができるのではないかと仮説の元、実験を行った(図1)。

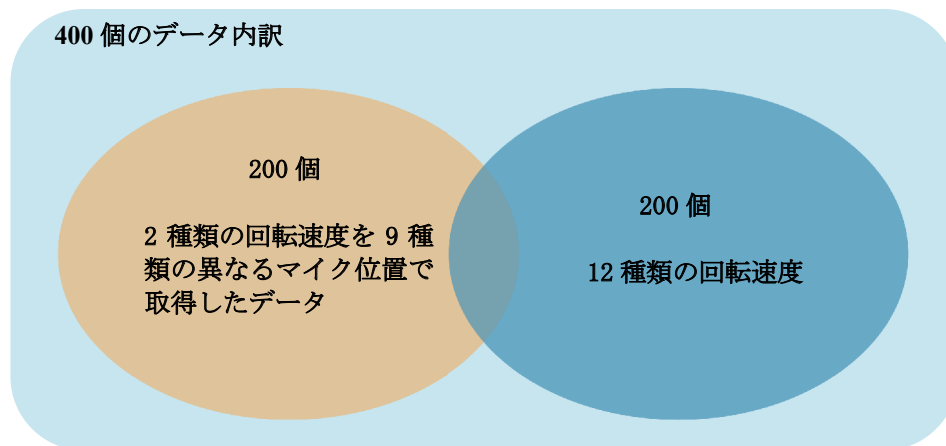


図 1. 使用したデータの内訳について

特徴量抽出には、まず音響データに対して短時間フーリエ変換を施した。入力信号を 1024 サンプルのフレームに分割し、各フレームに hamming 窓関数を適用した。hamming 窓関数を用いる事で信号の両端を滑らかにフェードイン・アウトさせることで、フレームの端で発生する波形の不連続性を低減する効果がある。そして、各フレームのサンプリングレートを 512 にして高速フーリエ変換を行う。得られたスペクトルを時間軸と周波数軸に分けた 2 次元グラフの出力を絶対値に直し、対数スケールにしたものを一次元に変換したものを学習データとした。400 個のデータは、訓練データ：240 個、検証データ：80 個、テストデータ：80 個に分割し学習・評価を行った。

二値分類をする機械学習モデルには、scikit-learn の LinearSVC を使用した。パラメータは、正規化パラメータを 15、収束の基準となる許容誤差の値を 0.01 に設定。その他は、デフォルトの値で実験を行った。まず、検証データを使ったモデルの評価は f1 スコアと混合行列を用いた。F1 スコアとは、(1) の式から導き出せるものであり *precision* (適合率) と *recall* (再現率) の調和平均として定義される値である。F1 スコアから分類器の精度を総合的に評価する事ができ、値が 1.0 (100%) に近づくほど *precision* と *recall* の両方の精度が高く効率よくバランスの取れたモデルと判断することが可能である。実験結果として、検証データを用いた f1 スコアは 0.96 という値が出力された。

$$F1\ score = 2 \times \frac{precision \times recall}{precision + recall} \quad (1)$$

次に混合行列を用いた評価を行った。混合行列とは分類器の予測結果と実際のクラスの対応関係を表した行列の事である。これは、True Positive (TP)、False Negative (FN)、False Positive (FP)、True Negative (TN) の 4 つの要素で構成されている。TP とは、分類器が陽性と判断し実際に陽性であったデータ。FN は、陰性と判断したが、実際に陽性であったデータ。FP は、陽性と判断したが実際に陰性であったデータ。TN は、陰性と判断し、実際に陰性であったデータを表している (図 2)。本稿の場合、Positive が異常時、Negative を正常時と定義する。

		モデルの予測値	
		Positive	Negative
実際のクラス (正解値)	Positive	TP	FN
	Negative	FP	TN

図 2. 混合行列の構成について

検証データ 80 個 (異常データ: 42 個、正常データ: 38) を使った二値分類結果の内訳を混合行列で確認すると、TP: 42 個、TN: 35 個、FP が 3 個という結果になった。つまり、80 個の内、77 個は正しく分類できており、また 3 個は異常と判断したが、実際には正常であるものがあつた (図 3)。

		モデルの予測値	
		Positive(異常)	Negative(正常)
実際のクラス (正解値)	Positive(異常)	42	0
	Negative(正常)	3	35

図 3. 検証データを使った混合行列

検証データでは、F1 スコアが 0.96 で混合行列を見ても比較的精度良く分類が出来ていると言える。

次に、テストデータ 80 個 (異常データ: 34 個、正常データ: 46 個) を用いてこの分類器の精度を確認する。テストデータでは、検証データでの評価と同様に F1 スコア、混合行列を用いたモデル評価を行う。まず、F1 スコアは 0.95 という値が得られた。混合行列を見てみると FP に 2 個、FN に 2 個の誤検知が確認できた (図 4)。

		モデルの予測値	
		Positive(異常)	Negative(正常)
実際のクラス (正解値)	Positive(異常)	32	2
	Negative(正常)	2	44

図 4. テストデータを使った混合行列

そして、テストデータでは、ROC (Receive Operating Characteristic Curve) 曲線を用いた AUC 評価を行った。まず ROC 曲線とは、二値分類モデルの性能を評価する為の指標であり True Positive Rate(TPR) と False Positive Rate(FPR) の関係を表したグラフである。TPR とは、Positive の中で正しく分類された割合を示し (2) の数式で求められる。FPR は、Negative の中で誤って Positive と分類された割合であり (3) の数式で導ける。

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$FPR = \frac{FP}{TN + FP} \quad (3)$$

つまり、TPR が高く、FPR が低いほどモデル性能が良いと言える。また、この ROC 曲線の下部に囲まれた面積を表す数値が AUC (Area Under the ROC Curve) であり、モデル性能を数値化した指標になる。0 から 1.0 までの値を取り 1.0 に近づくほどモデル性能が良いと判断する事ができ、テストデータを用いた ROC 曲線を用いた AUC 評価では、0.94 とモデル性能が高いと言える値が得られた (図 5)。

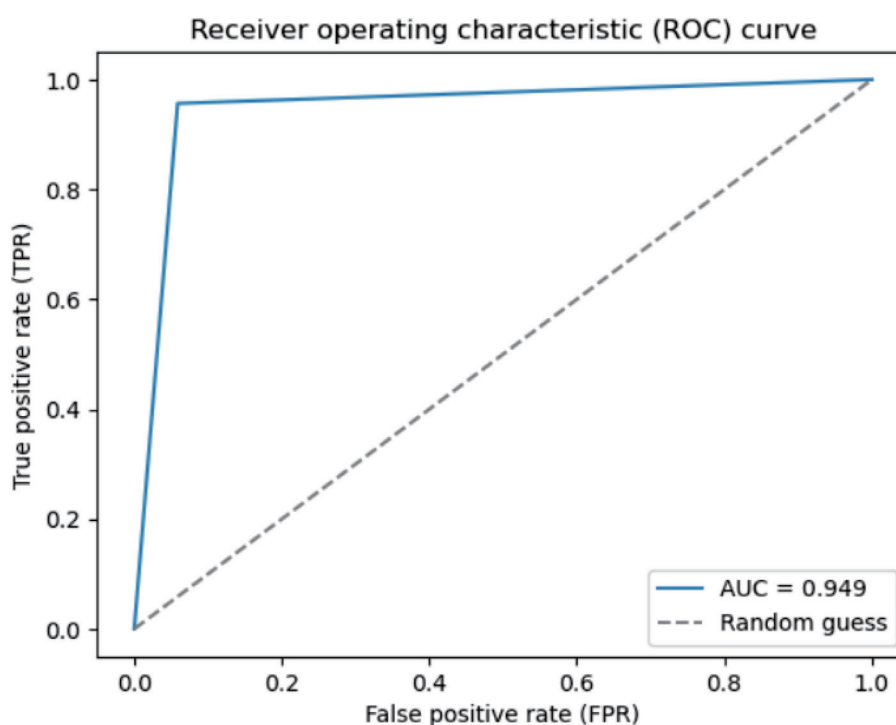


図 5. テストデータでの ROC 曲線を用いた AUC 評価

### 3. 考察

実験結果として、テストデータでの評価が F1 スコア 0.95、AUC 評価が 0.94 と二値分類モデルとして優れたモデルが得られた。学習コストとして、Windows CORE-i7 の CPU を用いた学習にかかった時間は 33 秒。最適なモデルパラメータを求める為の反復回数であるイタレーション数は、235 であった。本稿の研究では、12 種類の回転速度と 9 種類のマイクの位置の 2 つの条件に絞り実験を行ったが実際の機器での運用を想定すると、動作パターンはいくつもあり、ノイズなど様々な条件を考慮しなければならない。

#### 4. 結論

本稿での実験結果から、特定の条件下で扱える 2 値分類モデルとして期待する結果が得られる学習モデルを構築することが出来た。次のステップとして製造機器にこの学習モデルを実適用する為に、学習データに使う音響情報を広げ ( 検知対象以外の異なるノイズ音、様々な回転速度など ) このモデルを教師なし学習に応用させる実験を行う事で実用化を進めていきたい。

#### 参考文献

- [1] 伊藤彰則 (2019)," 環境音から異常を検知する統計的手法 ", 日本音響学会誌 75 巻 ,pp.538-543
- [2] C. Clavel, T. Ehrette, and G. Richard(2005), "Events detection for an audio-based surveillance system," Proc. IEEE International Conference on Multimedia and Expo (ICME),pp.1306–1309.
- [3] K.M. Jeon, D.Y. Lee, H.K. Kim, and M.J. Lee, "Acoustic surveillance of hazardous situations using nonnegative matrix factorization and hidden Markov model," Proc. 137th Audio Engineering Society (AES) Convention, 2014.
- [4] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti(2007), "Scream and gunshot detection and localization for audio-surveillance systems," Proc. IEEE Advanced Video and Signal Based Surveillance (AVSS), pp.21–26.
- [5] S. Ntalampiras, I. Potamitis, and N. Fakotakis(2009), "On acoustic surveillance of hazardous situations," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.165–168.
- [6] C.N. Doukas and I. Maglogiannis(2011), "Emergency fall incidents detection in assisted living environments utilizing motion, sound, and visual perceptual components," IEEE Trans. Inf. Technol. Biomed., vol. 15, no. 2, pp.277–289.
- [7] R.L.H. Murphy, A. Vyshedskiy, V.-A. PowerCharnitsky, D.S. Bana, P.M. Marinelli, A. Wong-Tse, and R. Paciej(2004), "Automated lung sound analysis in patients with pneumonia," Respiratory Care, vol. 49, no. 12, pp.1490–1497.
- [8] T. Anmen, T. Murosaki, N. Takaha, H. Ito, and Y. Kondo(2004), "Fricative inspection with wavelet analysis," SICE Annual Conference, pp.807–812.
- [9] Masahiro TSUCHIYA, Michiyuki TAKAGI(1998),"Vibration and Acoustic Fault Diagnosis of Rotational Machine", 日本機械学会論文誌 64, pp.465-472
- [10] 緒方淳, 村川正宏 (2015)," 風力発電スマートメンテナンスのための機械学習技術に基づく異常検知システムの構築 ", 日本機械学会シンポジウム講演論文集 14
- [11] 植松 尚, 小泉 悠馬, 齊藤 翔一郎, 中川 朗, 原田 登 (2017)," 動作音から機器の異常を検知する異常音検知技術 ",NTT 技術ジャーナル 29,pp.24-26
- [12] Y. Kawachi, Y. Koizumi, and N. Harada, "Complementary set variational autoencoder for supervised anomaly detection(2018)," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),pp. 2366–2370.
- [13] 江原 史朗, 武藤 義彦 (2021),"One Class SVM を用いたベアリング異常音検知システム ", 産業応用工学学会論文誌 ,9 巻 1 号 p.31-37

大柴義永 東京国際工科専門職大学 工科学部 情報工学科 4 年生  
 富山哲男 東京国際工科専門職大学 副学長・工科学部長 情報工学科教授

西田健次 東京工業大学システム制御系 特任准教授、オレンジテクラボ CTO  
宮崎 淳 東京国際工科専門職大学 客員教授、オレンジテクラボ CEO