

時間的な変動を含まない特徴量を用いた機械稼働音の教師なし異常検知の検討

Study on the Machine Anomaly Sound Detection Using Features that do not Include Temporal Variation

金子 馨[†]
Kaoru Kaneko

平田 俊明[‡]
Toshiaki Hirata

1. はじめに

音声や音楽を中心とする音響解析において、メルスペクトログラムはニューラルネットワーク（以降 NN と略記）の入力として有効性が確認されている。画像解析の分野で大きく発展した NN にとって、音の波形データから周波数特性の時間変動を捉えて 2 次元に変換できるメルスペクトログラムは都合がよい。しかし、波形データを画像化するほかの方式と比較した際の優位性については議論がほとんど行われていない。

他方、機械稼働音の異常検知タスクにおける解析では一般に長期的な変化を捉える必要があるため、解析の単位となる短い時間の中で生じる変化は注目されていない。正常時や時には異常時も、短時間のうちには同様の音が続いて大きな変化がないためである。

メルスペクトログラムをこうした定常性をもつ機械稼働音に対して適用することは、必ずしも最適でないと考え、本研究ではグラミアン角場[1]およびリカレンスプロット[2]の 2 方式を波形データの画像化のための方式としてとりあげ、比較を実施する。いずれも本来は時系列データを入力とするが、周波数スペクトルを入力に算出することで、時間情報を持たない、2 次元のデータを得ることができる。

評価は MIMII DG データセット[3]を用いて実施した。グラミアン角場、リカレンスプロットを用いることで従来のメルスペクトログラムを用いる場合と同程度、一部のデータにおいては上回る異常検知精度を得ることができ、必ずしもメルスペクトログラムが最適な特徴量ではないことを示した。

2. 先行研究

2.1 機械稼働音の異常検知

機械稼働音を対象とした異常検知の研究は、実異常音の入手が難しいという特徴から教師なし異常検知の分野で広く進められており、とくに国際コンペティションタスクである DCASE Challenge task2 [4] で盛んに性能が競われている。DCASE 2024 では 28 チームから 96 システムの結果が投稿され、その多くがメルスペクトログラムを入力に NN を用いる方式であった。例えば最も優れた成績をのこした Lv らは事前学習済みのフレームワークである BEATs および EAT を用いているが[5]、これらのフレームワークは対数メルスペクトログラムを入力とする。また、メルスペクトログラム以外では波形データを直接入力とするものや、MFCC を用いるものがあつた。

[†] 株式会社マーブル Marble Corp.

[‡] 東京情報デザイン専門職大学

Tokyo Information Design Professional University

DCASE 以外では、金子ら[6]が機械稼働音を定常的なサウンドテクスチャとみなし、その再合成用の統計量を NN の入力として用いた。精度はメルスペクトログラムに劣るものの、正常と異常の識別が可能であり、短時間のうちの変化の情報が失われても異常判別が可能であることを示唆している。

2.2 グラミアン角場/リカレンスプロット

グラミアン角場 (Gramian Angular Field, 以降 GAF と略記) [1] は時系列データの時間相関を可視化する方式であり、1 次元の系列データを入力に 2 次元の画像を出力する。画像化することによって 2 次元畳み込みの入力とすることが可能になり、NN を活用した解析のために用いられている。ただし出力される画像の 1 辺の大きさが時系列長に等しいため、時系列が長くなるにつれて畳み込みの計算が高価になってしまう。これを回避するには、Piecewise Aggregation Approximation (PAA) によって時系列を圧縮する方法などがある[1]。

時系列データを直接画像化する以外に、Ding ら[7]は送電線データの周波数スペクトルを GAF によって画像化することで、CNN による故障検知の入力に用いている。周波数スペクトルに適用することで、周波数帯域間の相関を可視化できると考えられる。

なお本研究では GAF のうちグラミアン角和場 (Gramian Angular Summation Field) のみを用い、グラミアン角差場 (Gramian Angular Difference Field) は用いていない。

リカレンスプロット (Recurrence plot, 以降 RP と略記) [2]も GAF と同様に時系列データの特徴を視覚化する方式である。2 つの対応する時刻の状態が近ければプロットするというもので、画像の 1 辺の長さが時系列長と等しくなる点も GAF と同様である。

南ら[8]は RP を入力とする CNN で漏水音の検知を行っているが、10,000Hz のサンプリングレートで録音された音データを 100 分の 1 に区切ったものを RP に変換し、さらにバイキュービック補完法を用いることで、画像サイズを小さくしている。

3. 実験方法

本研究では MIMII DG データセットの機械稼働音データを GAF および RP を用いて画像形式に変換する。これらのデータは 24kHz、10 秒の音声であり、そのまま変換するのでは画像サイズの問題がある。音データを短時間に区切り、それらを離散フーリエ変換によって周波数スペクトルに変換したのちに、メル尺度を用いてバンドパスフィルタリングすることで各データサイズを圧縮した。コードの実装上はメルスペクトログラムを求めたのちに、各時刻を 1 入力として取り出している。

ベースラインとして、DCASE 2022 Challenge task2 [4] を取り上げる。このタスクは 5 種類の機械の稼働音を多様なドメインで収録したデータセットを対象に、異常検知を実施するタスクである。

このベースラインは全結合 NN を用いたオートエンコーダ (AE) の再構成誤差による異常検知と、畳み込み NN (CNN) を用いたセクション分類の成否による異常検知の 2 つが用意されているが、これに畳み込み NN を用いたオートエンコーダ (CAE) による再構成誤差を行う方式を加えて、3 種類の解析手法でそれぞれ比較した。CAE のモデル構造は Keras のサンプル実装[9]に倣った。ただし各層にバッチノーマライズを追加し、誤差関数を MSE に変更している。

メルスペクトログラムを入力として用いているベースラインのコードを一部変更して、入力に GAF, RP を入力して用いるコードをそれぞれ作成した。また、GAF についてはこの移動平均を入力として用いるコードも作成している。AE および CAE においては正常音データの再構成精度が異常判定精度に寄与すると考えられるので、移動平均をとることによって単純に GAF を求めたものを用いるよりも高い AUC スコアが得られることが期待される。

前述のとおり GAF, RP の計算はメルスペクトログラムの各時刻を取り出して入力としており、このメルスペクトログラムを求めるパラメータはベースラインのものと同一である。

学習・評価を異なる環境で計 2 回実施し、得られたスコアの平均を求めることで最終的なスコアを得た。

4. 実験結果

表 1~3 にそれぞれ AE, CNN, CAE の結果を示す。a~c はそれぞれ異なる機械種別ごとの結果である。

各表の ROC-AUC (AUC), Precision (Prec) および Recall (Rec) はそれぞれ異常をポジティブ、正常をネガティブとした場合の各指標で、括弧内の s, t はそれぞれ source ドメインと target ドメインを対象とした結果であることを示している。source ドメインは対象機械の稼働に伴うパラメータが同じになっている音が学習用のデータセットに多量に含まれるグループであり、target ドメインはそれが少量しか含まれていないグループである。いずれのドメインも異常と正常が同数の 150 ファイルずつ、計 600 ファイルを機械種別ごとにテストデータセットに含んでいる。Recall はドメイン別の結果でなく単一の結果のみを記載しているが、これは DCASE のベースラインで実装されている方式に倣い、異常ラベルのデータについてはドメインを問わず対象に含んだことによって、ドメイン別で結果が変わらないためである。なお、Precision および Recall の算出に用いるしきい値はベースラインの実装に従い、学習用データを異常判別して得られたスコアから求めたガンマ分布の上位 10% に設定された。用いたデータの個数ではなく予測された分布の面積に従うため、極端な外れ値が紛れ込んで分布がゆがむとしきい値が非常に高く設定されてしまう場合があり、本実験でもいくつかその事象が発生した。これにより Precision, Recall が共に 0 となったケースは表中に N/A と記載している。

入力の別において Base, GAF, Mean, RP にそれぞれ結果を示している。Base は DCASE のベースラインとして公開さ

れているコードを我々の環境で動作させた結果であり、メルスペクトログラムを入力とする。Mean は GAF の移動平均であり、ほかは GAF, RP を入力とする。

4.1 オートエンコーダ (AE)

各機械種別において最も高い AUC が得られた特徴量 (各表で太字表記) は bearing なら GAF, fan ならメルスペクトログラムと、それぞれ異なっている。またいずれも AUC の値が次点に高い機種との差は小さい。このことから使用した特徴量によって大きな違いは生じなかったと言える。GAF の移動平均による改善は fan および slider で見る事ができたが、valve はスコアの差が小さく、bearing ではわず

表 1a : 異常検知精度 (bearing)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.56	0.59	0.50	0.51	1.00
GAF	0.60	0.55	0.62	0.45	0.49
Mean	0.59	0.61	0.51	0.53	0.98
RP	0.56	0.57	0.50	0.51	0.98

表 1b : 異常検知精度 (fan)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.80	0.48	0.67	0.51	0.63
GAF	0.52	0.56	0.51	0.15	0.12
Mean	0.65	0.58	0.67	0.57	0.51
RP	0.78	0.50	0.78	0.62	0.35

表 1c : 異常検知精度 (gearbox)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.69	0.65	0.67	0.65	0.61
GAF	0.61	0.55	0.68	0.47	0.14
Mean	0.69	0.61	N/A	N/A	N/A
RP	0.67	0.66	0.67	0.60	0.42

表 1d : 異常検知精度 (slider)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.79	0.49	0.76	0.56	0.42
GAF	0.59	0.50	0.72	0.69	0.11
Mean	0.77	0.55	N/A	N/A	N/A
RP	0.71	0.45	0.71	0.50	0.30

表 1e : 異常検知精度 (valve)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.52	0.50	0.54	0.47	0.30
GAF	0.49	0.52	0.44	0.68	0.07
Mean	0.51	0.52	0.58	0.65	0.12
RP	0.55	0.48	0.55	0.42	0.14

かではあるが移動平均をとらないほうが高いスコアが得られる結果となった。

4.2 畳み込みニューラルネットワーク (CNN)

AE と同様に各機械種別で異なった結果が得られているが、次点との差は slider, valve の 2 機種で大きく開いており、それぞれ GAF, メルスペクトログラムを用いた場合のスコアが最も高かった。

GAF とその移動平均を比較すると、いずれの機械種別においても単純に GAF を求めたものを用いたほうが高いスコアを得ることができた。

表 2a : 異常検知精度 (bearing)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.69	0.55	0.73	0.68	0.86
GAF	0.64	0.55	0.75	0.71	0.76
Mean	0.58	0.57	0.46	0.47	0.64
RP	0.65	0.54	0.74	0.71	0.81

表 2b : 異常検知精度 (fan)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.68	0.43	0.96	0.60	0.30
GAF	0.77	0.33	0.96	0.54	0.41
Mean	0.60	0.42	0.55	0.42	0.39
RP	0.76	0.38	0.95	0.58	0.50

表 2c : 異常検知精度 (gearbox)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.70	0.59	0.89	0.81	0.47
GAF	0.69	0.52	0.88	0.68	0.28
Mean	0.63	0.54	0.67	0.59	0.58
RP	0.66	0.51	0.85	0.69	0.23

表 2d : 異常検知精度 (slider)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.70	0.48	0.72	0.51	0.39
GAF	0.86	0.59	0.95	0.80	0.35
Mean	0.70	0.57	0.56	0.48	0.61
RP	0.80	0.57	0.77	0.64	0.34

表 2e : 異常検知精度 (valve)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.68	0.63	0.98	0.90	0.32
GAF	0.52	0.51	0.79	0.76	0.36
Mean	0.60	0.60	0.49	0.50	0.27
RP	0.50	0.52	0.77	0.71	0.43

4.3 畳み込みオートエンコーダ (CAE)

他の解析と同様に各機械種別で異なった結果が得られた。次点との差は fan で大きく開いているが、他の機械種別においては大きくない。CNN では slider, valve はそれぞれ GAF, メルスペクトログラムでスコアが最も高かったが、CAE ではそれとは逆にメルスペクトログラム, GAF でスコアが高いという結果になった。

表 3a : 異常検知精度 (bearing)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.54	0.50	0.52	0.58	0.19
GAF	0.57	0.65	0.63	0.74	0.19
Mean	0.59	0.63	0.70	0.69	0.44
RP	0.51	0.54	0.55	0.60	0.34

表 3b : 異常検知精度 (fan)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.54	0.47	0.67	0.49	0.14
GAF	0.36	0.52	0.19	0.00	0.09
Mean	0.41	0.60	0.62	0.00	0.20
RP	0.45	0.60	0.18	0.00	0.19

表 3c : 異常検知精度 (gearbox)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.52	0.54	0.58	0.34	0.24
GAF	0.57	0.52	0.64	0.52	0.10
Mean	0.53	0.51	0.45	0.58	0.11
RP	0.48	0.51	0.45	0.53	0.13

表 3d : 異常検知精度 (slider)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.52	0.42	0.55	0.36	0.09
GAF	0.49	0.47	0.69	0.59	0.12
Mean	0.47	0.48	0.65	0.24	0.15
RP	0.43	0.45	N/A	N/A	N/A

表 3e : 異常検知精度 (valve)

	AUC(s)	AUC(t)	Prec(s)	Prec(t)	Rec
Base	0.46	0.42	N/A	N/A	N/A
GAF	0.49	0.55	0.44	0.65	0.05
Mean	0.53	0.55	0.56	0.51	0.15
RP	0.53	0.50	0.51	0.49	0.23

5. 考察

各解析手法に共通して、機械種別ごとに最も高い AUC スコアが得られる特徴量が異なるという結果が得られた。また、その機械種別と最も高い AUC スコアが得られた特徴量の対応関係が解析手法間で異なっていた。したがって、どの特徴量がほかの特徴量よりも優れているあるいは特定の解析手法や機械種別に適しているということは言えなかった。これは必ずしもメルスペクトログラムだけが最適な特徴量ではないという仮説を支持する結果である。

ただし、いずれの特徴量においても異常判別の精度が低く、機械種別によっては無作為に判別を行った場合に得られる 0.5 に近いスコアが得られている点は考慮が必要である。すなわち、今回 AE, CNN, CAE の 3 解析手法を用いたが、これらの手法が適切でなかったために、特徴量が本来有していた異常判別に寄与する特徴が活かされなかった可能性がある。ドメインの別によって target は source よりも異常判別の精度が下がることが予想されるのに反し、いくつかの結果においては大きな差が開いていない、ないしは target がよりよい精度を示していることも、その可能性を示唆するものである。

6. おわりに

本研究では、時間的な変動を含まない特徴量として GAF および RP を工場機械の異音検知に適用する可能性を検討した。従来のメルスペクトログラムとおおむね同程度、一部では上回る結果も得られたことで、これらの特徴量を用いる一考の余地があることを示した。一方で、メルスペクトログラムより劣る結果も得られており、これらの特徴量が勝っているとは言えない。今後、精度検証を継続してこれらの特徴量が有効な条件を明らかにする。

参考文献

- [1] Wang, Z. and Oates, T. "Encoding Time Series as Images for Visual Inspection and Classification Using Tiled Convolutional Neural Networks", AAAI Conference on Artificial Intelligence (online), available from <<https://cdn.aaai.org/ocs/ws/ws0115/10179-46015-1-PB.pdf>> (accessed 2025-05-26)
- [2] Eckmann, J. P., Kamphorst, S. O. and Ruelle, D. "Recurrence Plots of Dynamical Systems", Europhysics Letters, Vol. 4, No. 9, pp. 973-977 (1987).
- [3] 土肥 宏太, 西田 智哉, PUROHIT Harsh, 田邊 亮, 遠藤 隆, 山本 正明, 二階堂 悠貴, 川口 洋平, "MIMII DG: 稼働音診断におけるドメイン汎化手法向けデータセット", 日本音響学会研究発表会講演論文集, Vol.2022, No.秋, (2022).
- [4] DCASE.: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring Applying Domain Generalization Techniques, available from <<https://dcase.community/challenge2022/task-unsupervised-anomalous-sound-detection-for-machine-condition-monitoring.html>> (accessed 2024-01-10)
- [5] Lv, Z., Jiang, A., Han, B., Liang, Y., Chen, X., Liu, J. and Fan, P. "AITHU System for First-Shot Unsupervised Anomalous Sound Detection", DCASE(online), available from<https://dcase.community/documents/challenge2024/technical_reports/DCASE2024_Lv_110_t2.pdf> (accessed 2025-05-21)
- [6] 金子 馨, 平田 俊明, "テクスチャ統計量を活用した環境音の教師なし異常検知の検討", 情報処理学会全国大会講演論文集, Vol.86, No.2, (2024).
- [7] Ding, C., Wang, Z., Ding, Q. and Yuan, Z. "Convolutional neural network based on fast Fourier transform and gramian angle field for fault identification of HVDC transmission line", Sustainable Energy Grids and Networks, Vol. 32, No. 1, 100888 (2022).
- [8] 南 泳旭, 荒井 康裕, 國實 誉治, 小泉 明, "リカレンスプロットの活用と畳み込み NN による漏れ判別モデルの構築", 土木学会論文集 G (環境), Vol.76, No.65, (2020).
- [9] Keras documentation: Convolutional autoencoder for image denoising, available from <<https://keras.io/examples/vision/autoencoder/>> (accessed 2025-06-05)