

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

金子 馨†

コンピュータロン株式会社†

平田 俊明‡

東京情報デザイン専門職大学‡

1. はじめに

工場の機械稼働音をはじめとした環境音の異常検知を実用化の際、多数に分散したエッジコンピューティングにおけるハードウェアコストの低減が課題である。より安価なエッジ端末で運用できるよう、計算コストが小さい異常検知手法が求められている。

本研究では、データの特徴量抽出に McDermott らのサウンドテクスチャ統計量[1] を用いることで、異常検知の計算コスト低減を検討した。

サウンドテクスチャ統計量は定常音からサウンドテクスチャを合成する過程で得られる統計量であり、メルスペクトログラムと同様に人の聴覚を模した計算によって算出される。

サウンドテクスチャ統計量の特徴量として用いた音声解析の先行研究としては、Ellis ら[2] や、Villamizer ら[3] が分類タスクに適用したのがある。本研究ではサウンドテクスチャ統計量の特徴量として用いるにあたり、分類タスクのみならず、異常検知については教師なし学習においても同様に機能することを示すとともに、先行研究[4][5] のニューラルネットワーク（以降 NN と略記）による特徴量抽出手法と比較して計算コストを低減できることを確認した。

2. 実験方法

2.1 使用データ

本研究では、DCASE 2022 Challenge task2 データセット[6] の Development データセットを用いた。各 10 秒のサウンドファイルを McDermott のツールボックスを使用してテクスチャ統計量に変換した。Villamizar ら[3] に倣い 1,707 個のパラメータを得た。これを分類・異常検知の各モデルの入力にした。

2.2 実験の流れ

先行研究では McDermott らのサウンドテクスチャ統計量は自然環境音に適用されることが多く、工場環境音に適用した研究は少ない。工場環境音の正常と異常、ある

いはその他のラベルの違いを識別できる特徴がサウンドテクスチャ統計量に含まれているのかを確かめる必要があった。第一に分類モデルを作成してその精度を評価することで工場環境音へのサウンドテクスチャの適用可能性を検討した。第二にサウンドテクスチャ統計量を用いた教師なし学習による異常検知モデルを作成し、先行研究と比較・評価した。

3. 実験結果

3.1 分類タスク

サウンドテクスチャ統計量を工場環境音に適用する妥当性を確認するべく、工場環境音の分類を試みた。

分類は以下の 2 通りで実施した。

- ① 機械種類の分類
- ② 機械種類ごとの正常／異常分類

第一に機械種類の分類を実施した。データセットに含まれる 7 種の機械それぞれごとに、対象の音声がその機械から発せられたものであるか否かの 2 分類を行った。

分類モデルにはサポートベクタマシン（以降 SVM と略記）を用い、1,707 次元のサウンドテクスチャ統計量を入力とした。また比較のためにメルスペクトログラムを入力とするモデルも作成した。学習には train データを用い、うち正例は全ファイルを、負例は正例と同数になるよう無作為に抽出したファイルを用いた。

分類を test データの全ファイルを対象に実施し、その正解率で結果を評価した。サウンドテクスチャ統計量およびメルスペクトログラムを入力とするモデル、それぞれの結果を表 1 に示す。

表 1：機械種類分類の正解率

Table 1 : Accuracy of Machine Types Classification.

	サウンドテクスチャ統計量	メルスペクトログラム
bearing	99.57 %	99.90 %
fan	96.76 %	97.93 %
gearbox	98.36 %	98.26 %
slider	98.21 %	95.21 %
ToyCar	99.74 %	100.00 %
ToyTrain	99.98 %	100.00 %
valve	99.67 %	98.45 %
Total	98.90 %	98.54 %

Study on unsupervised anomaly detection of environmental sounds using texture statistics.

† KAORU KANEKO, COMPUTRON CO., LTD.

‡ TOSHIKI HIRATA, Tokyo Information Design Professional University

第二に機械ごとの正常／異常分類を実施した。

分類モデルは機械種類の分類と同様に SVM を使い、機械それぞれに 2 分類を行うモデルを作成した。入力も同様に、1,707 次元のサウンドテクスチャ統計量、および比較用のメルスペクトログラムをそれぞれ用いた。train データには異常のデータが含まれないため、test データを学習用と評価用に分割して用いた。

分類を評価用に分割された test データを対象に実施し、その正解率で結果を評価した。サウンドテクスチャ統計量およびメルスペクトログラムを入力とするモデル、それぞれの結果を表 2 に示す。

表 2：正常／異常分類の正解率

Table 2 : Accuracy of Normal/Anomaly Classification.

	サウンドテクスチャ統計量	メルスペクトログラム
bearing	86.67 %	93.33 %
fan	66.67 %	63.89 %
gearbox	90.00 %	74.44 %
slider	85.56 %	81.67 %
ToyCar	78.89 %	75.00 %
ToyTrain	71.67 %	72.28 %
valve	72.22 %	51.11 %
Total	78.81 %	73.17 %

①機械種類の分類 および ②機械種類ごとの正常/異常分類、いずれにおいてもメルスペクトログラムを入力とした場合と比較して大きく劣ることなく、Total で比較すると高い精度で結果が得られた。この結果から、工場環境音に対してサウンドテクスチャ統計量を用いる妥当性を確認することができた。

3.2 異常検知

教師なし学習による異常検知を試みた。局所外れ値因子法 (LOF)、混合ガウスモデル、1 クラス SVM、k 近傍法の 4 出力をアンサンブルして異常度を求め評価した。

入力にはサウンドテクスチャ統計量を全結合 NN によって次元削減した特徴量を用いた。全結合 NN は機械種類の分類およびセクションの分類の誤差を最小化するよう特徴量抽出器として学習させた。一方で、Kuroyanagi ら[4] は Transformer ベースの NN を、Wilkinghoff [5] は ResNet ベースの NN をそれぞれ特徴抽出に用いており、いずれも末尾で全結合 NN により次元削減している。

test データを対象にして異常検知を実施した。AUC による評価を表 3 に示す。

サウンドテクスチャ統計量を用いたほとんどの場合で先行研究と比べて AUC の値が低く出ている。これは、実験環境の制約から先行研究と同等量のデータセットを揃えられなかったことによるもので、今後の課題とする。

表 3：テストデータを対象とした異常検知の AUC

Table 3 : AUC of Anomaly Detection on Test Dataset.

	Kuroyanagi ら[4]	Wilkinghoff [5]	本研究
bearing	77.81 %	74.10 %	48.91 %
fan	68.27 %	88.48 %	56.70 %
gearbox	81.25 %	86.79 %	55.14 %
slider	87.41 %	90.48 %	85.84 %
ToyCar	79.50 %	79.60 %	70.39 %
ToyTrain	69.20 %	73.80 %	57.28 %
valve	91.33 %	86.58 %	69.12 %

メモリ使用量の観点からも比較を行った。1 サウンドファイルを読み込んでから抽出された特徴を変数に格納するまでのメモリの増加量を調べた。結果を表 4 に示す。

表 4：特徴量抽出時のメモリ使用量

Table 4 : Memory Usage During Feature Extraction.

Kuroyanagi ら[4]	288.3 MiB
Wilkinghoff [5]	168.4 MiB
本研究	13.9 MiB

NN を用いた先行研究と比較してサウンドテクスチャの計算は期待通りメモリ使用量が少なかった。

4. おわりに

本研究では、サウンドテクスチャ統計量を工場機械の異常検知に適用する可能性を検討した。機械種類の分類や正常／異常分類では十分高い正解率を確保できた。一方で、教師なし学習による異常検知では精度検証を十分に実施できなかったが、少ない計算コストで実装できることを確認できた。今後、精度検証を継続してサウンドテクスチャ統計量の有効性を明らかにする。

参考文献

- [1] McDermott, J. H. and Simoncelli, E. P.: Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis, *Neuron*, Vol.71, No.5, pp.926–940 (2011).
- [2] Ellis, D. P., Zeng X. and McDermott, J. H.: Classifying soundtracks with audio texture features, *Proc. ICASSP 2011*, pp.5880-5883, IEEE (2011).
- [3] Villamizar, D., Battaglino, D., Muratore, D. G. and Hoshya, R.: Sound Classification using Summary Statistics and N-Path Filtering, *Proc. ISCAS 2019*, pp.1-5, IEEE (2019).
- [4] Kuroyanagi, I., Hayashi, T., Takeda, K. and Toda, T.: Two-stage anomalous sound detection systems using domain generalization and specialization techniques, DCASE(online), available from < https://dcase.community/documents/challenge2022/technical_reports/DCASE2022_Kuroyanagi_11_t2.pdf > (accessed 2024-01-10).
- [5] Wilkinghoff, K.: An outlier exposed anomalous sound detection system for domain generalization in machine condition monitoring, DCASE(online), available from < https://dcase.community/documents/challenge2022/technical_reports/DCASE2022_Wilkinghoff_2_t2.pdf > (accessed 2024-01-10).
- [6] 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)