

異常時の機械から発生する稼働音の変化を説明するテキスト生成技術

#生産性向上 #AI・生成AI #IoT・データ活用 #研究開発

執筆者

川口 洋平 (Kawaguchi Yohei)

日立製作所 研究開発グループ 先端AIイノベーションセンタ メディア
知能処理研究部 所属
現在、音声・音響・時系列信号処理AIと基盤モデルの研究開発のマネ
ジメントに従事
博士（工学）
IEEEシニア会員、日本音響学会会員、電子情報通信学会会員、情報処
理学会会員

土肥 宏太 (Dohi Kota)

日立製作所 研究開発グループ 先端AIイノベーションセンタ メディア
知能処理研究部 所属
現在、時系列信号処理AIと基盤モデルの研究開発に従事
日本音響学会会員

西田 智哉 (Nishida Tomoya)

日立製作所 研究開発グループ 先端AIイノベーションセンタ メディア
知能処理研究部 所属
現在、音響信号処理AIと基盤モデルの研究開発に従事
日本音響学会会員

遠藤 隆 (Endo Takashi)

日立製作所 研究開発グループ 先端AIイノベーションセンタ メディア
知能処理研究部 所属
現在、音響信号処理AIと基盤モデルの研究開発に従事
日本音響学会会員、電子情報通信学会会員

ハイライト

従来、製造ラインでの品質検査やインフラ設備の点検など、多くの場面において熟練技術者の主観に依存した聴音点検が行われてきた。しかし、近年の労働人口減少による熟練者不足に伴い、人手での聴音点検は困難になりつつある。

そこで、日立は、聴音点検の自動化・効率化のためのソリューション提供や技術開発を進めている。従来技術は、機械の稼働音を入力として、正常時の音とどれだけ乖離しているかを示す異常度を出力する、あるいはその異常度が閾値を超過する場合にアラートを発出するものであった。本稿では、機械の稼働音というデータを異常時のアラートだけでなく、さらに具体的な保守作業につなげるため、異常検知の根拠となるテキストを生成するAI技術について紹介する。

1. はじめに

製造ラインでの品質検査やインフラ設備の点検など、これまで多くの場面において熟練技術者の主観に依存した聴音点検が行われてきた。しかし、近年の労働人口減少による熟練者不足に伴い、人手での聴音点検は困難になりつつある。これに対し、日立は聴音点検を自動化するソリューション¹⁾、²⁾ や、そのコアとなる異常音検知のAI技術を開発・提供してきた。こうした技術開発は、産業機械の稼働音データセットの公開³⁾～⁵⁾ および異常音検知の国際コンペティションの主催⁶⁾～¹⁰⁾ を通じたオープンイノベーションと、さまざまな自社AI (Artificial Intelligence) 技術の開発¹¹⁾～²²⁾ の両輪を回す活動であり、これらの取り組みによって、異常音検知の実用可能範囲を大きく広げてきたと考えている。

従来の異常音検知技術は、機械の稼働音を入力として、正常時の音とどれだけ乖離しているかを示す異常度を出力する、あるいはその異常度が閾値を超過する場合に異常を知らせるアラートを発出するものであった²³⁾。このアラートに加えて、AIがどのような根拠に基づいて異常と判断したのかの情報があれば、具体的な次の保守作業につなげる手がかりになると考えられる。

本稿では、AIが異常音を検知したのちに、あらかじめ収録されたデータベース上の正常音と新たに検知された異常音を比較し、音の特徴の変化について説明するテキストを出力するAI技術「音響変化キャプション」を紹介する。なお、本技術は国際会議The Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE) 2023にて発表したものであり²⁴⁾、詳細はそちらの文献を参照されたい。

2. 従来研究

「画像変化キャプションング」については、これまでも研究されてきた。これは、変化前後の二つの画像を入力としてその間の変化を説明するテキストを自動生成する技術である。2018年、前後画像間の変化領域をピンポイントで特定するために、ピクセル差分に基づく手法が提案された²⁵⁾。この手法では、説明対象として重要なシーン変化と、重要ではない視点や照明などの変動とを区別できなかったが、2019年には、この問題を解決するための、画像間の変化領域を識別する二重注意機構に基づく画像変化キャプションング手法が提案された²⁶⁾。ただしいずれの手法も、二つの入力の間の変化が、信号内の局所的な領域（画像ではその矩形領域）に現れることを仮定している。一方で機械の稼働音は、時間周波数領域の広い範囲に成分を有する場合もあり、この仮定は必ずしも成立しない。ゆえに、画像変化キャプションングの従来技術を音響変化キャプションングに適用することは困難である。そのため、日立は、二重注意機構に基づく画像変化キャプションング²⁶⁾を参考としつつ、音響変化キャプションングの技術を新たに開発した。

3. 提案手法

3.1 音響変化キャプションングのためのAIモデル

日立は、変化前後の二つの音響信号を入力としてその間の変化を説明するテキストを自動生成する、音響変化キャプションングのためのAIモデルを提案する。このモデルは、変化前後の二つの音響信号を可変長ベクトルとして受け付け、一つの固定次元ベクトルへと符号化する「音響エンコーダ」と、符号化されたベクトルをテキスト系列へと復号する「デコーダ」の二つから成る、エンコーダーデコーダ型のニューラルネットワークアーキテクチャである。

まず、変化前後の二つの音それぞれに対して、短時間フーリエ変換、パワースペクトログラム計算、および周波数軸の対数メル尺度への変換を行い、二つの対数メルスペクトログラムを求める。次に、それぞれの対数メルスペクトログラムを共通の音響エンコーダに入力すると、変化前の音響信号を符号化した固定次元ベクトル X_{before} と、変化後の音響信号を符号化した固定次元ベクトル X_{after} が生成される。さらに、2019年に提案された画像変化キャプションング手法と同様に、AIモデルが差分に注目しやすいよう、 X_{after} と X_{before} のベクトル同士の差分 X_{diff} を計算し、 X_{before} 、 X_{after} 、 X_{diff} を結合した X_{all} を得る。そして、 X_{all} をデコーダに入力すると、変化を説明するテキストが出力される（図1参照）。

音響エンコーダには、標準的なトランスフォーマーエンコーダ²⁷⁾と、二重注意機構に基づく画像変化キャプションング²⁶⁾で用いられた「空間的注意」の2種類のエンコーダを用いる。トランスフォーマーエンコーダは、一般的な音響キャプションングで有効性が実証されている。その一方、空間的注意は、二つの音響信号の間の変化が時間周波数領域において局在していることを仮定し、畳み込みニューラルネットワークを利用して、時間周波数領域の局所的な範囲に注意を向けるものである。前述のとおり、この仮定は、機械の稼働音すべてにおいて成立するものではないが、定常音では多くの場合に周波数方向での局所性が成立するため、空間的注意が有効に機能すると予想される。このように、稼働音の変化を説明する観点によって最適なアーキテクチャが異なることを想定し、アーキテクチャを使い分けることにした。すなわち、稼働音の変化を説明する観点として「定常音」、「周期音」、「非周期音」の3種類を定義し、音響エンコーダとして定常音には空間的注意を、周期音と非周期音（いずれも非定常音）にはトランスフォーマーエンコーダを採用することにした。

デコーダには、一般的な音響キャプションングで有効性が実証されている、標準的なトランスフォーマーモデルを用いる。このモデルは、テキスト系列に対するマルチヘッド自己注意層とマルチヘッドエンコーダーデコーダ注意層から構成されている。

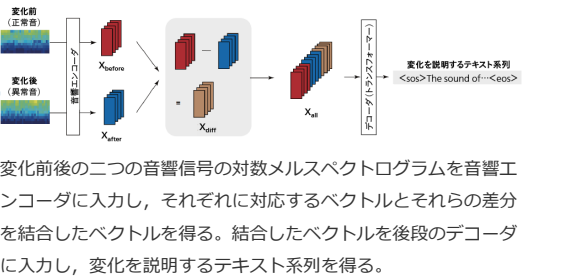
3.2 データセットおよびモデル訓練方法

音響変化キャプションングにおいては、訓練や評価のためのデータセットがないことから、日立は新たにデータセットを作成した。このデータセットは、異常音検出の訓練や評価のために開発したMIMII-DG（Malfunctioning Industrial Machine Investigation and Inspection for Domain Generalization）データセット¹⁷⁾をベースに作成したものであり、MIMII-Changeと呼ばれる。MIMII-Changeデータセットは、5種類の機械（ベアリング、ファン、ギアボックス、スライダ、バルブ）の稼働音を対象とする。機械1種類当たり、正常音と異常音の一つずつの音響信号のペアが300個含まれており、さらに各ペアには正常な状態から異常発生に至るまでの稼働音の変化を説明するテキストを付与した。

変化説明のテキストは、3名のアノテータに人手で付与させた。アノテータには、音の変化を表現する際にはできるだけ擬音語を使うように指示した。擬音語は音を音韻的に模倣する文字列であり、多様な環境音の特徴を表現するのに有効である²⁸⁾。また、アノテータには「定常音」、「周期音」、「非周期音」の各観点から説明テキストを分けて付与させることで、異なる説明テキストを用いて異なるアーキテクチャのモデルを訓練し、説明の観点ごとの使い分けを可能とした。

モデルの訓練は、一般的な音響キャプションングと同様、クロスエントロピー損失に基づいて実現できる。すなわち、訓練データ内の変化前後の二つの音響信号から成るペアを入力として出力される説明テキストと、それらのペアに付与された、正解の説明テキストとの間ができるだけ一致するように、音響エンコーダとデコーダを訓練する。

図1 | 音響変化キャプションングのためのAIモデル



4. 評価実験

「定常音」，「周期音」，「非周期音」の各観点に対する評価結果を表1に示す。BLEU，METEOR，CIDEr，SPICE，SPIDErは，キャプション分野で一般的に使用される指標であり，これらの指標が大きいほど，出力テキストが正解の説明テキストと類似していることを示す。MPERは，擬音語の部分の正確さを評価するために新たに定義した指標である。文中のすべての擬音の組み合わせに対する音素誤り率の平均値であり，この指標が小さいほど，出力テキストの擬音語部分の音素列が正解の説明テキストと類似していることを意味する。

実験の結果を表1に示す。空間的注意に基づく音響エンコーダは，定常音の変化の説明に適していることが分かる。その一方で，周期音と非周期音（すなわち，非定常音）の変化の説明には，トランスフォーマーエンコーダが適していることが分かる。したがって，定常音の変化を説明するか，非定常音の変化を説明するかによって，これらの異なるエンコーダのモデルを切り替えるのが望ましいことが示唆される。

表1 | 実験結果

稼働音の変化を説明する観点	エンコーダ	BLEU 3	BLEU 4	METEOR	CIDEr	SPICE	SPIDEr	MPER
定常音	トランスフォーマー	0.616	0.542	0.427	0.969	0.34	0.655	0.281
	空間的注意	0.669	0.601	0.441	1.086	0.365	0.726	0.266
周期音	トランスフォーマー	0.464	0.387	0.39	0.946	0.255	0.601	0.338
	空間的注意	0.426	0.354	0.402	0.881	0.249	0.565	0.38
非周期音	トランスフォーマー	0.413	0.339	0.427	1.864	0.373	1.118	0.327
	空間的注意	0.328	0.269	0.411	1.441	0.304	0.873	0.321

「定常音」，「周期音」，「非周期音」という3通りの観点に対して，2種類のエンコーダを用いた場合の各評価指標を示す。MPER以外は値が大きいほど，MPERは値が小さいほど，正解の説明テキストと類似した結果であることを示す。太字の数値は，2種類のエンコーダ間でより良好な結果を示している。

5. おわりに

本稿では，AIが異常音を検知したのちに，正常音と異常音との間にある音の特徴の変化をテキストで説明する「音響変化キャプションング」技術を紹介した。今後も，聴音点検の自動化およびその適用可能範囲の拡大のための技術開発を進める。

本稿で紹介した技術は，現場の状況を解釈して作業判断に資する情報を現場作業者に提供するためのAI技術群の一つであり，それらの開発は，現場作業者の業務を対話的に支援するAIエージェントの実現へとつながっていくと期待される。

謝辞

本稿で述べた音響変化キャプションングの開発においては，同志社大学の椿 俊介氏（当時修士課程在学），井本 桂右准教授，および立命館大学の岡本 悠希氏（当時博士課程在学）にご協力いただいた。深く感謝の意を表する次第である。

参考文献など

1)

日立製作所，レトロフィット無線センサー（メーター自動読取り，異音検知）

2)

日立製作所，Hitachi Intelligent Platformユースケース，保守DX

3)

H. Purohit et al., “MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection,” in Proc. Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)（2019）

4)

R. Tanabe et al., “MIMII DUE: Sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions,” in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics WASPAA）（2021）

5)

K. Dohi et al., “MIMII DG: Sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task,” in Proc. Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)（2022）

6)

Y. Koizumi et al., “Description and discussion on DCASE2020 Challenge Task2: Unsupervised anomalous sound detection for machine condition monitoring,” in Proc. Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)（2020）

7)

Y. Kawaguchi et al., “Description and discussion on DCASE 2021 Challenge Task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions,” in Proc. Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)（2021）

8)

- K. Dohi et al., "Description and discussion on DCASE 2022 Challenge Task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques," in Proc. Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE) (2022)
- 9) K. Dohi et al., "Description and discussion on DCASE 2023 Challenge Task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," in Proc. Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE) (2023)
- 10) T. Nishida et al., "Description and discussion on DCASE 2024 Challenge Task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," in Proc. Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE) (2024)
- 11) Y. Kawaguchi et al., "How can we detect anomalies from subsampled audio signals?," in Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP) (2017)
- 12) Y. Kawaguchi, "Anomaly detection based on feature reconstruction from subsampled audio signals," in Proc. European Signal Processing Conference (EUSIPCO) (2018)
- 13) Y. Kawaguchi et al., "Anomaly detection based on an ensemble of dereverberation and anomalous sound extraction," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019)
- 14) K. Suefusa et al., "Anomaly detection based on an ensemble of dereverberation and anomalous sound extraction," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2020)
- 15) H. Purohit et al., "Deep autoencoding GMM-based unsupervised anomaly detection in acoustic signals and its hyper-parameter optimization," in Proc. Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE) (2020)
- 16) K. Dohi et al., "Flow-based self-supervised density estimation for anomalous sound detection," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2021)
- 17) K. Dohi et al., "Disentangling physical parameters for anomalous sound detection under domain shifts," in Proc. European Signal Processing Conference (EUSIPCO) (2022)
- 18) T. Nishida et al., "Anomalous sound detection based on machine activity detection," in Proc. European Signal Processing Conference (EUSIPCO) (2022)
- 19) H. Purohit et al., "Hierarchical conditional variational autoencoder based acoustic anomaly detection," in Proc. European Signal Processing Conference (EUSIPCO) (2022)
- 20) K. Shimonishi et al., "Anomalous sound Detection based on sound separation," in Proc. Annual Conference of the International Speech Communication Association (INTERSPEECH) (2023)
- 21) K. Dohi et al., "Distributed collaborative anomalous sound detection by embedding sharing," in Proc. European Signal Processing Conference (EUSIPCO) (2024)
- 22) T.V. Ho et al., "Stream-based active learning for anomalous sound detection in machine condition monitoring," in Proc. Annual Conference of the International Speech Communication Association (INTERSPEECH) (2024)
- 23) 井本桂右, 外: 環境音分析・異常音検知の研究動向, 電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review (2022)
- 24) S. Tsubaki et al., "Audio-change captioning to explain machine-sound anomalies," in Proc. Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE) (2023)
- 25) H. Jhamtani et al., "Learning to describe differences between pairs of similar images," in Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP) (2018)
- 26) D.H. Park et al., "Robust change captioning," in Proc. IEEE/CVF International Conference on Computer Vision (ICCV) (2019)
- 27) A. Vaswani et al., "Attention is all you need," in Proc. Advances in Neural Information Processing Systems (NIPS) (2017)
- 28) Y. Okamoto et al., "Environmental sound extraction using onomatopoeic words," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2022)

日立評論

日立評論は、イノベーションを通じて社会課題に応える
日立グループの取り組みを紹介する技術情報メディアです。

日立評論Webサイトでは、日立の技術者・研究者自身の執筆による論文や、
対談やインタビューなどの企画記事、バックナンバーを掲載しています。ぜひご覧ください。

日立評論(日本語) Webサイト

<https://www.hitachihyoron.com/jp/>



Hitachi Review(英語) Webサイト

<https://www.hitachihyoron.com/rev/>



 日立評論メールマガジン

Webサイトにてメールマガジンに登録いただきますと、
記事の公開をはじめ日立評論に関する最新情報をお届けします。