

特徴抽出と物理モデルの適用で進化する異常検知

Application of feature extraction and physical models for anomaly detection

MathWorks Japan	王 暁星	Xiaoxing WANG	Member
MathWorks Japan	井上 道雄	Michio INOUE	Non-member
MathWorks Japan	竹本 佳充	Yoshinobu TAKEMOTO	Non-member

Abstract

本講演では振動データや音データの異常検知を例に、実践的な特徴抽出の方法や機械学習・深層学習への適用テクニックをご紹介します。また、異常時のデータが不足している際に、シミュレーションによって故障データを模擬する物理モデルの代替方法も説明します。

Keywords: 異常検知、故障予測、予知保全、特徴抽出、機械学習、深層学習、音データ、振動データ

1. はじめに

近年、製造業をはじめとする様々な業界においてメンテナンス費用の高騰化や機器の老朽化、保守・安全管理の熟練技術者の不足などの背景から「予知保全」への取り組みが注目を浴びている。予知保全では、機器が現在正常に機能しているかどうかを判断する「異常検知」から、機器の故障時期を予測し適切なタイミングでメンテナンスの実施を目指す「寿命予測」(Fig.1) までの広範囲かつ段階的な付加価値を有する機能的側面がある[1,2]。以下では、主に製造業において取り組みが活発になっている音データと振動データを対象にした異常検知と故障予測に関して紹介する。

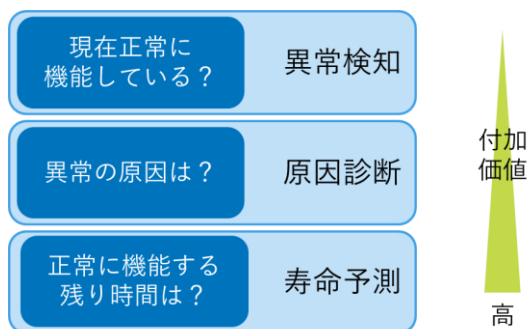


Fig.1 3つの「目的・機能レベル」別予知保全

2. 予知保全のワークフロー

予知保全を実現するために必要なステップを一般化してみると下記の4項目に大別される (Fig.2)。

1. 時系列信号へのアクセス
2. 欠損値処理や次元削減などのデータの前処理
3. 機械学習の適用と評価を実施し予測モデルを構築
4. システムへの統合とアプリケーション化

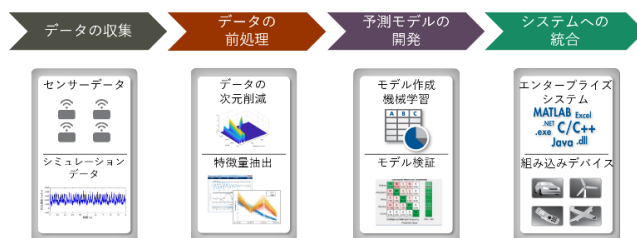


Fig.2 予知保全のための解析ワークフロー

Fig.2の一連の流れを俯瞰すると、故障予測モデルの構築は全ステップの一部分であることが分かる。モデルの構築は大量のデータさえあれば、機械学習や深層学習といったAIの技術が自動的に結果を算出してくれるといった印象があるが、残念ながら様々な対象や用途に対して汎用的なソリューションは存在しない。他方、故障予測モデルの開発を円滑に行う上では、データの前処理において、データの性質を効果的に抽出した良い特徴量を見つけ出すことが成功の鍵を握る最重要なステップとなる。次節では時系列信号から異常を見つける際に有効な特徴抽出を紹介する。

連絡先：王 暁星
住所：〒107-0052 東京都港区赤坂 4-15-1-7
所属先：MathWorks Japan
E-mail: xwang@mathworks.com

3. 特徴抽出と異常検知

Fig.1において、どんなデータが必要なのかは目的による。異常が発生していることが分かれば十分なのか、それともどんなアクションを起こすかなど意思決定に寄与する情報が求められるのか、どのレベルでの予知保全を実現すべきなのかは明確にする必要がある。一般的には、「寿命予測」を目指すほど付加価値は高くなるが、その分ハードルも高くなってしまう。そのため、多くの製造業において、対象機器に対して、現在正常に機能しているかを判断する異常検知からのスタートが散見される。いずれにしても部品や機械の微妙な変化を検知しようとするのであれば、ある程度の精密な計測データが重要となる。機器の故障に起因する現象を捉える際には、一般的に加速度センサーから取得される振動データが多用されるが、直近ではその代替データとして、音のデータの活用が増えている。そのメリットとしては、やはりデータの測定が容易であり、非接触で計測できるため、対象機器に外乱を与えないことや、センサーを比較的自由に設置できる点などが挙げられる。Fig.3ではコンプレッサーから取得された正常音と7種の異音の計8クラスの音データが示されている。それぞれのデータには既にラベリングが施されており、各クラス225個、計1800個のデータがある[3]。

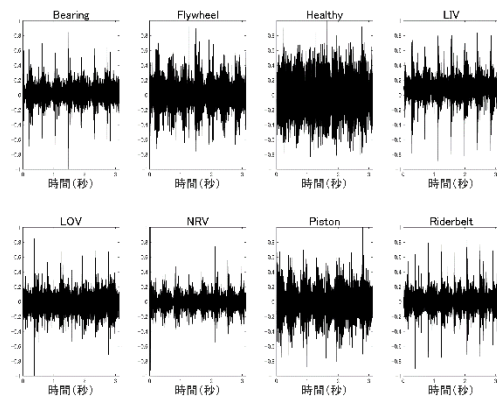


Fig.3 正常音と7種の異音の計8クラスの音データ

3.1 音の特徴抽出

Fig.3の音データの異常検知として、AIモデル（深層学習）を活用したコンプレッサーの異音判定を考える。アプローチの流れは既にFig.2で説明しているが、とりわけ重要となる特徴抽出と深層学習の適用に関してはFig.4で示している。その際、マニュアルの特徴抽出の代表例として、

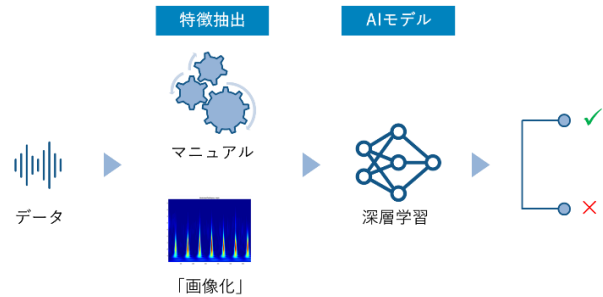


Fig.4 時系列信号に対して深層学習を活用した異常検知例

音声領域固有の特徴量であるメル周波数ケプストラム係数などが存在する[4]。今回の異音判定のアルゴリズムでは、メル周波数ケプストラム係数に加え、その時間変動の情報を持つデルタパラメーターの計26次元の特徴量をMATLAB®のAudio Toolbox™を介して抽出した。予め深層学習に適用するために7:2:1の割合でランダムに分割した学習用、検証用、テスト用のデータから上述の特徴抽出を行い、学習用（と検証用）の特徴量に対してLSTM（隠れ層=100）を使用した深層学習のモデルを構築する（Fig.5）。その際、各次元のデータは標準化している。生成されたモデルはテスト用の特徴抽出した26次元のデー

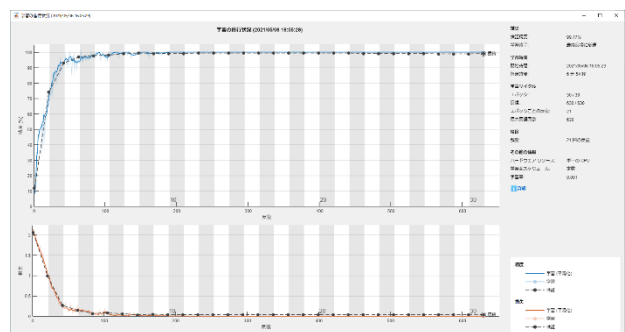


Fig.5 音の特徴量に対してLSTMを活用した学習曲線（上）と損失関数の推移（下）

真のクラス	Bearing	Flywheel	Healthy	LIV	LOV	NRV	Piston	Riderbelt
Bearing	100.0%							
Flywheel		100.0%						
Healthy			100.0%					
LIV				95.5%			4.5%	
LOV					95.5%		4.5%	
NRV						100.0%		
Piston							100.0%	
Riderbelt								100.0%

Fig.6 Fig.5の学習済みモデルに対する精度検証

タに対して検証を行い、特に過学習が起きていないことが確認できる (Fig.6)。

3.2 時系列信号の「画像化」

時系列信号に対して、深層学習を活用するもう一つの代表的な方法が信号を「画像化」することである。画像化においては、例えば時間周波数表現が有効となる場合が多く、時系列信号に対して、短時間フーリエ変換を施すことでスペクトログラムが求まり時間周波数表現を可能とする。Fig.7はSignal Processing Toolbox™より求めた8クラスの内の2種の音データのスペクトログラムを示している。一部の周波数帯で両者の強度に差があることが確認できる。なおこのような時間周波数表現は、他にもウェーブレット解析ベースのスカログラムや定Q変換、更にはHilbert-Huang変換などがある[5]。求められたスペクトログラムは画像として保存できるため、深層学習への再利用が簡単となる点も「画像化」のメリットの一つである。

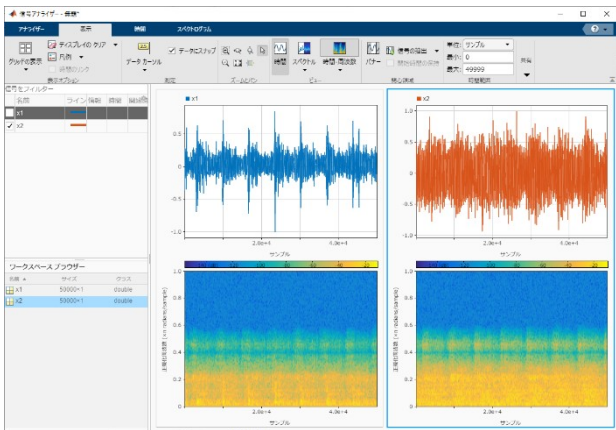


Fig. 7 2種類の音データのスペクトログラム例

Fig.8では8クラスのスペクトログラムを画像データとして、音データの判別を目的に公開されている事前学習済みネットワーク YAMNet[6,7]を活用した転移学習の学習曲線を示している。転移学習したモデルに対してテス

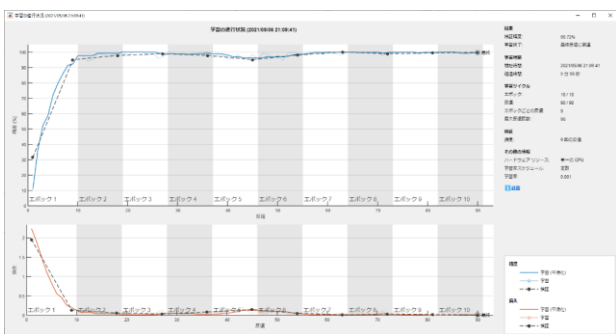


Fig. 8 スペクトログラムに対して転移学習した学習曲線

ト用のスペクトログラムを検証したところ、高い分類精度を有しており、特に過学習が起きていないことが確認できる (Fig.9)。

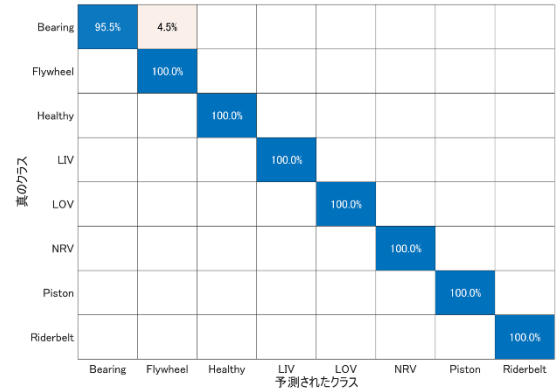


Fig. 9 Fig. 8の転移学習済みモデルに対する精度検証

3.3 振動データの特徴抽出と故障予測

音データに続き、回転機等の異常検知でよく使用される加速度センサーから取得される振動データの特徴抽出を以下に紹介する。

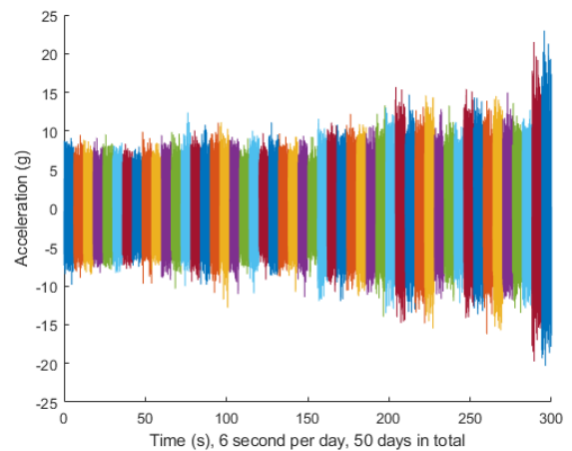


Fig. 10 インパルス特性が増加傾向にある振動データ

Fig.10は2 MW 風力タービンの高速シャフトの振動データ[8] (サンプリング周波数約 100 kHz) を1日あたり6秒間、50日分連続して表示しているが、信号のインパルス特性が増加傾向にあることが分かる。50日の期間中に内輪に不具合が発生し、劣化が徐々に進み、対象機器が故障する様子を示している。

この振動データから特徴抽出をするにあたり、まず前半25日間のデータは劣化の可能性が低い「正常」のクラスとして、後半25日間のデータは劣化の可能性が高い「異常」のクラスとしてラベリングする。目的は時間経過で変

化する特徴量を見つけ出すことであり、上述のような大胆なラベリングは、劣化は時間経過とともに徐々に起きているという仮定をおいたデータに対して有効な処理となる。その際、時間経過と相関があるような特徴量を見つけ出せれば、その特徴量は劣化を定性的に表現できていると考えられる。次に、クラス分けした振動データの各日に対して、標準偏差・尖度・波高率 (Crest Factor) などの時間領域の特徴量やピーク値・ピーク周波数・バンドパワーなどの周波数領域の特徴量を求め、正常と異常の特徴量のヒストグラムを描く (Fig.11)。ここで、個々の特徴量において、正常と異常のヒストグラムが重なってしまう場合は、その特徴量は両者を識別できないことを示唆し、一方で、正常と異常のヒストグラムが分かれていますらば、その特徴量は両者を識別できる良い特徴量となる。

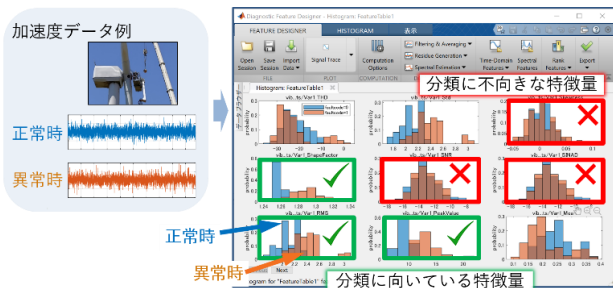


Fig. 11 良い特徴量と悪い特徴量の一例

上述の複数の特徴量から正常と異常の2クラスに有意差があるかを見つけるために、2標本t統計量の特徴量の良し悪しを決める一つのスコアとして計算する。即ち、特徴量の重要度をランキングする (Fig.12)。こうすることで、重要度が高い特徴量を選択し、機械学習のアルゴリズムに手渡すことができる。Fig.13 では Statistics and Machine Learning Toolbox™を介した9つの機械学習のモデル構築を行っているが、ここでは分類木が一番良い精度 (100%)

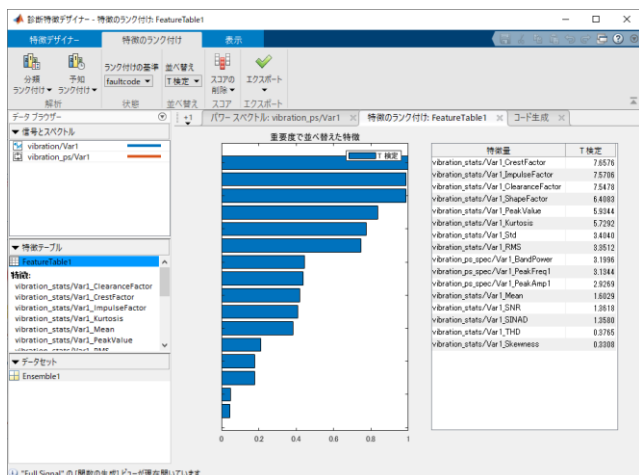


Fig. 12 抽出された特徴量の重要度ランキングの一例

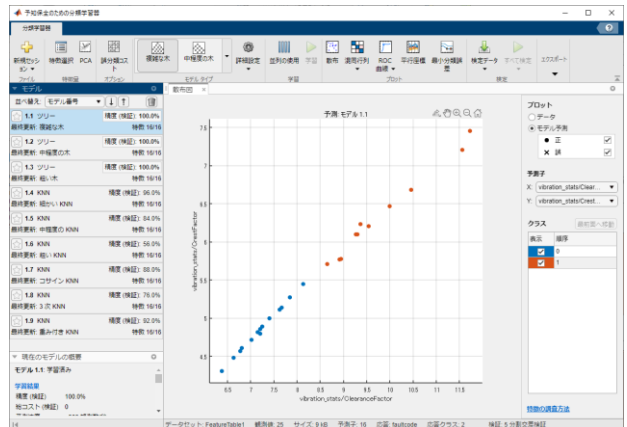


Fig. 13 機械学習による振動データの分類

を示している。なお Fig.12 の特徴量のランキングは時間軸上でも重要な意味をなす。とりわけ、一步先の異常検知として、対象機器が故障してしまう時期を事前に予測する残寿命予測 (Fig.14) では、対象機器の健康 (或いは劣化) 状態を示す健康 (劣化) インジケータの構成が重要となるが、Fig.12 で示されている特徴量 (のランキング) は、正常と異常を有意に区別することから、劣化の度合いまたは健康インジケータの役割も担っている。上位に占める特徴量の中でも、時間に対して (単調に) 増加または減少するものを選択し、更には選択した複数の特徴量を主成分分析などを用いて次元削減することにより健康インジケータを構成する方法もある。構築した健康インジケータはパラメトリックモデルなどを適用することで残寿命を回帰的に求めることが可能となる [9]。Fig.14 では、故障が偶発的に起こるものではなく、「疲労や損傷の蓄積により対象機器の内部状態が少しずつ劣化し、ある閾値を超えると故障が発生する」というメカニズムを数式的に模擬する指数劣化モデル (Predictive Maintenance Toolbox™) を導入している。その際、モデルのパラメータ値は観測値とのベイズ推定に基づきながら更新され、対象機器の残寿命の信頼区間を計算している [10,11,12]。

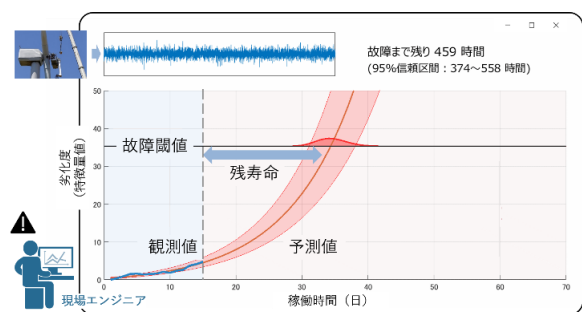


Fig. 14 機器に取り付けたセンサーデータから機器の残寿命を予測する例

4. 物理モデリングの活用

異常検知や故障予測モデルの開発は手元にあるデータに対して適切な処理を行いながら構築していく。一方で、故障が発生しないよう頻繁にメンテナンスを行っている場合や、セーフティクリティカルなシステムのために故障を発生させるわけにはいかない場合などでは故障データが不足していたり、存在しないことが多々ある。故障データがない場合に試行できる1つの方法に物理モデリングがある。これは第一原理的に各専門領域の物理法則に基づき機器をモデル化し、故障データをシミュレーションで再現する方法がある。機器の物理的構成要素がどのように互いに影響しあうかを考慮し、適切なツールを介してこれらの相互作用の機能的関係を高度に設計、可視化していく。その際、様々な故障シナリオを想定して物理モデル内にそれらの故障の原因となる劣化のモードを組み込むことで擬似的に故障データを生成する。ここでは、あるピストンポンプの物理モデルをSimulink®/Simscape™を使用して構築した例をFig.15に示している[13]。

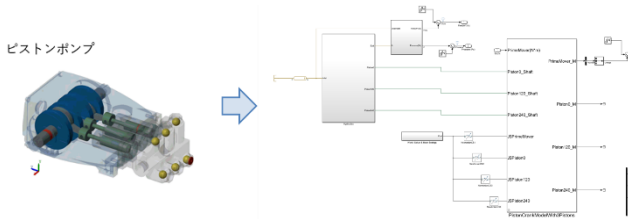


Fig. 15 Simulink/Simscape を使用した物理モデルの構築と故障データの生成

この物理モデルでは故障の要因として、モーターの劣化やベアリングの劣化、またパイプラインからの液体の漏れやパイプラインの目詰まりなどの4つの劣化モードを仮定している。各劣化モードに対しては、あらかじめ提供されている機能要素ブロックのパラメータを調整することで故障を模擬している。例えば、トルクのパラメータ値を正常値よりも小さくすることでモーターが劣化した状態を表し、パイプラインと外部大気が接する断面積の値を大きくすることで液体が外に漏れている状態を現象論的に再現している。ベアリングの劣化についてはジョイント部分の摩擦抵抗の増大として表現し、パイプラインの目詰まりについては断面積の値を小さくする。故障が発生または劣化が進展した際にどんな現象が発生するかを再現するものであり、モデルを実行すれば“自然”

と劣化が進展するというものではなく、劣化の度合いをパラメータで調整する点に注意されたい。実機の測定データをもとにモデルのチューニングを行う場合も多く、物理モデルを活用する際にも実測データの重要性は変わらないが、実測データがない場合であっても故障原因の予測という観点からの異常検知アルゴリズム開発に活かすことができる。

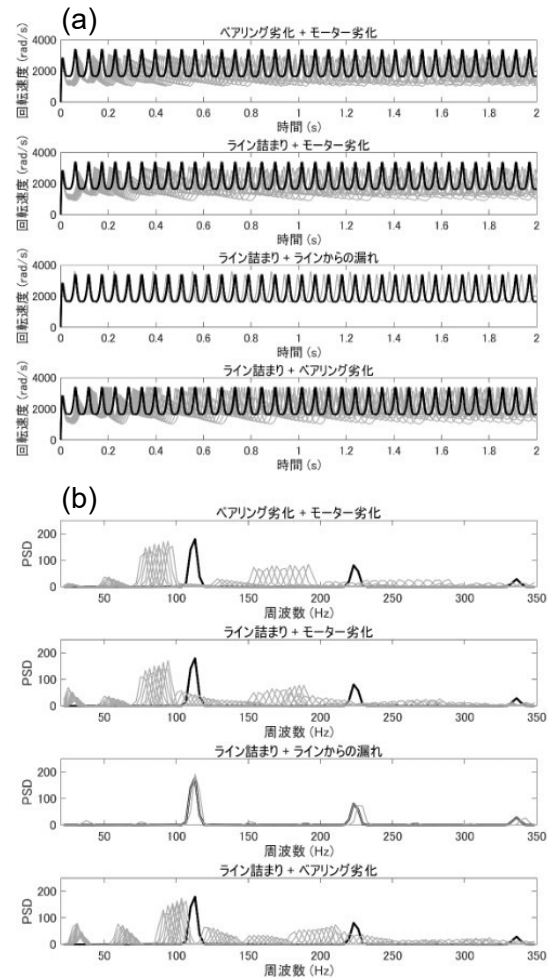


Fig. 16 (a) 擬似的に生成された正常・故障データ、(b) 周波数領域における正常データと故障データ。異なる故障形態では異なる周波数帯にピークが出現する。

Fig.16(a)に正常時(黒色)と2つの劣化モードが組み合わさったピストンポンプ故障時(灰色)の回転速度を示す。図から故障時のデータは正常時の周期とは異なる振る舞いをしている様子が観察できる。しかし、生データを目視しただけでは各々の故障形態を判別することは難しい。そこで Fig.16(b)が示すスペクトル解析を行った結果、正常時のスペクトルとの区別は勿論、特定の周波数帯で各々の故障形態を区別できる特徴的なピーク(特徴量)が

求まった。そこで、求められた一部ピーク値とピーク周波数 (Fig.17) に対して、前述の機械学習を利用することで故障形態判別のアルゴリズムを構築することが可能となる。この例でも、やはりデータの前処理の段階で良い特徴量を見つけ出すことが問題解決への近道となった。

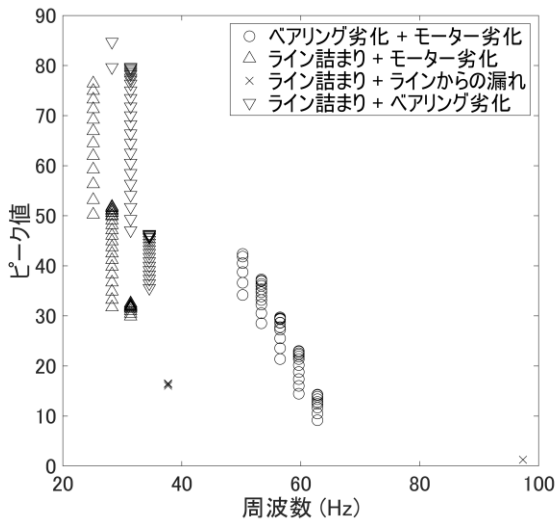


Fig. 17 Fig. 16 (b)の故障データの一部ピーク値とその周波数

終わりに、擬似的に生成した故障時のデータの信憑性がどこまであるかの議論がある。物理モデルを細部まで再現できれば当然現実により近い故障データを生成することが可能となる。しかし多くの場合、モデルを構築するには専門領域の高度な知識が必要とされる。物理モデルはあくまでモデルであり、現実のデータとは乖離があることを認識した上で、生成された正常・故障データがどれだけ直感的に正常時・故障時の特徴を掴んでいるかを把握することが大事となる。故障データがなく、何もできずに解析を諦めてしまうよりも、故障時の特徴を定性的なりとも掴んだデータを活用し一歩前に進み、後から修正を加えていくことが重要となる。

5. システムへの展開

最後にシステムへの統合について簡単に触れる。構築した予測アルゴリズムの実装先はデータの転送量や計算量、また実装先の運用方法によっても対象機器上かエッジデバイス上、またはクラウド上で様々な選択肢が存在する。データの転送量を極力抑えたい場合は、対象機器で必要な前処理を施し、エッジデバイスに予測に必要なデータだけを転送し、正常・異常を判別させる。判別の結果はクラウドに転送され、エンドユーザーと共有される。ま

た、算出された予測結果の表示は固定 PC 上で動くアプリが良いのか、それともポータブルタブレット上で動くもの良いのか、結果を確認する担当者は整備スタッフなのか、それとも機器のオペレーターなのか、などといった予測結果をどう運用に反映させるかについての議論を深める必要もある。例えば、3.1で紹介した異音判定のアルゴリズムを廉価なエッジデバイスとして活用できる Raspberry Pi に実装した例を Fig.18 に示している。Raspberry Pi 上では入力された任意の音データに対してスペクトログラムを求め、畳み込みニューラルネットワークにより正常・異常の判別を実施している。正常音 (異音) の場合は青色 (赤色) の LED を点灯させている。その際、MATLAB で構築したスペクトログラムの計算と深層学習への適用のアルゴリズムは書き換えることなく、そのまま MATLAB Coder™によって実装先の Raspberry Pi 上に C/C++コードがクリック一つで生成される。また、MATLAB では Raspberry Pi との通信や各センサーへのアクセスを容易にするサポートパッケージが提供されているため複雑な設定を必要としない。

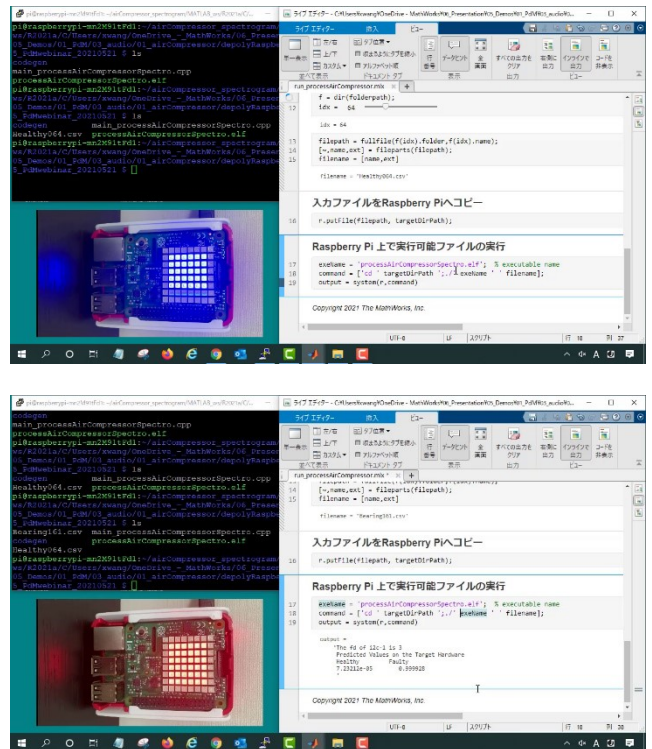


Fig. 18 Raspberry Pi における異音判定アルゴリズムの実装例。正常音に対して LED が青色に点灯 (上)、異音に対して LED が赤色に点灯 (下)

6. まとめ

本稿では、音データや振動データの異常検知を行う際の特徴抽出に焦点を当てて解説を行ってきた。その際、音声領域固有の特徴量や時間周波数表現、また時系列信号から抽出される統計量を適切に機械学習・深層学習へ適用することで、精度の高い異常検知が可能となる。また一歩先の異常検知として、稼働する機器の残寿命を予測する故障予測・予知保全についても簡単に触れた。一方で、故障データがない場合の異常検知に実践できる1つのアプローチとして、物理モデリングがある。この方法では、マルチドメインの物理法則（例えば電気、機械、油圧等）に基づき機器のモデル化を行う。その際、異常の原因となる要素をモデル内に取り入れることで、故障データを再現することが可能となるが、その際もやはり特徴抽出が故障診断の要となる。

抽出された特徴量を基に故障予測モデルのアルゴリズムを構築する際には、入力データはどこから来て、最終的にどうシステムへと実装するのか、という全体像を踏まえた開発が重要となる。これに対してMATLABはデータの前処理からアルゴリズムの開発、そしてシステムへの展開までの全体のワークフローをカバーするプラットフォームとして活用できる。結果、開発効率の向上や設計資産の共有及び再利用が望まれ、メンテナンスコストの削減が期待できる。

MATLAB とは：

MATLABは35年以上の実績を持つソフトウェアであり、数値計算、可視化、プログラミングのための高水準言語による対話型の開発環境である。世界中で400万人を超えるエンジニアや科学者が、MATLABを使用し、データ解析やアルゴリズムの開発、モデルやアプリケーションの作成を行い、アイデアの共有や専門分野を超えた共同プロジェクトのための共通言語として使用している。MATLABの予知保全に関連する機能・ソリューションについて詳しく知りたい方は「予知保全 MATLAB」で検索。MATLABによる予知保全のビデオシリーズも公開されている[14]。

参考文献

[1] MATLAB EXPO 2019: 故障予測×IoT ～予知保全システムの構築～
<https://www.matlabexpo.com/jp/2019/proceedings.html#section3>

[2] 井上道雄 他, 保全学, Vol. 19, No. 4, pp. 22-26, 2021.

[3] J. F. Gemmeke *et al.*, "Audio Set: An ontology and human-labeled dataset for audio events," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 776-780.

[4] L. R. Rabiner and R. W. Schafer, "Theory and Applications of Digital Speech Processing," 1st edition, Pearson, 2011.

[5] <https://jp.mathworks.com/help/signal/ug/time-frequency-gallery.html>

[6] M. Plakal and D. Ellis, "YAMNet," Oct. 2020.
<https://github.com/tensorflow/models/tree/master/research/audioset/yamnet>

[7] <https://jp.mathworks.com/help/audio/ref/yamnet.html>

[8] E. Bechhoefer *et al.*, "Processing for improved spectral analysis," Annual Conference of the Prognostics and Health Management Society, New Orleans, LA, Oct. 2013.

[9] 今村誠, システム/制御/情報, Vol. 65, No. 4, pp. 119-125, 2021.

[10] N. Gebraeel, "Sensory-Updated Residual Life Distributions for Components with Exponential Degradation Patterns," in IEEE Transactions on Automation Science and Engineering, vol. 3, no. 4, pp. 382-393, Oct. 2006.

[11] N. Gebraeel, "Residual-life distributions from component degradation signals: A Bayesian approach," *IIE Transactions*, vol. 37, pp. 543-5557, 2005.

[12] <https://jp.mathworks.com/help/predmaint/ref/exponentialdegradationmodel.html>

[13] 王曉星, 計測技術, 6月号, 2020年.

[14] MATLAB/Simulink による予知保全ビデオシリーズ
<https://jp.mathworks.com/videos/series/predictive-maintenance.html>