

ホワイトペーパー

予知保全・異常検知を可能にする特徴量

～ AI の実践的活用例に向けて ～

1. 機械学習の限界

AI・深層学習などのキーワードが注目を集める昨今、大量のデータさえ集めることができれば欲しい結果を算出してくれるといった印象があるが、様々な対象・用途に対して汎用的に使用できる手法は、残念ながら現存しないのが現実である。データの集め方、データの種類、サンプリング周波数などの諸条件が異なれば、前処理に必要な作業も適切に変更する必要がある。最適な機械学習アルゴリズムも異なる。

見逃されがちだが機械学習はノイズと意味ある重要な情報を自動で区別できるほど万能ではない。異常値を無視してよいのか、それとも、その異常値はモデルが考慮すべき現象（すなわち機械の故障や異常）を表しているのかを見極める必要もあるので、データの分析には必ず監視対象の機器に詳しい技術者の知見や経験が求められる。特に時系列データは、機械学習モデルに直接入力するのではなく、高いレベルの情報を捉えた「特徴量」に変換することが一般的であり成功の鍵を握る重要なステップである。本稿では時系列データから異常を見つける際に有効な特徴量の例を紹介する。

2. 故障予測を可能にする特徴量例

2.1 Structural Health Monitoring (SHM)

Los Alamos National Lab. の研究グループが公開する構造ヘルスマニタリング (SHM) で利用される振動データの解析を題材に、異常・正常の区別を可能にする特徴量を紹介する。本稿では技術的な詳細には触れず代表的な結果のみを紹介するにとどめるため、詳細は [1,2] を参照していただきたい。

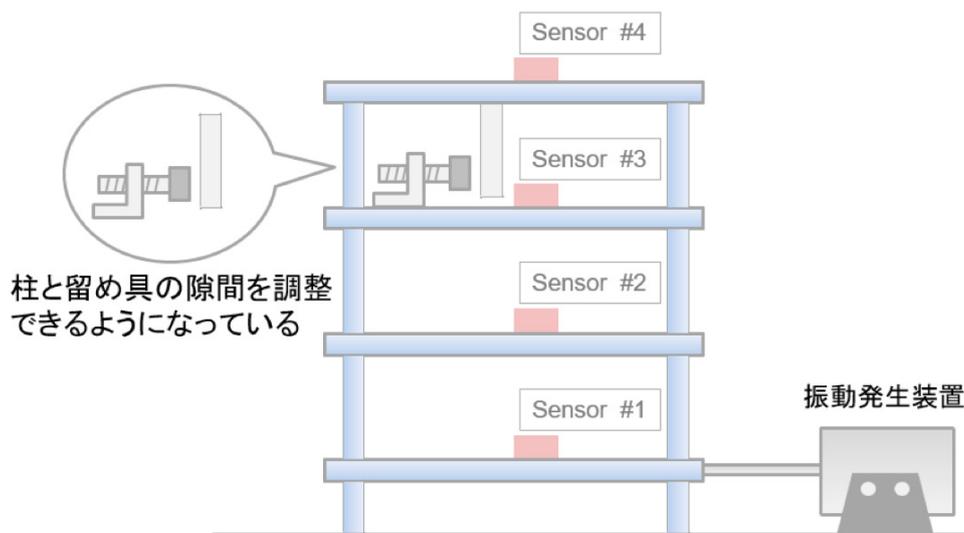


図1 実験セットアップイメージ図

図1に示すのは試験用に作られた3階建ての構造体のイメージ図である。こちらに取り付けられた加速度センサーから取得された時系列データのサンプルを図2に示す。ここでは1つの柱にバンパーが接触する仕組みにより正常時・異常時を模擬している。

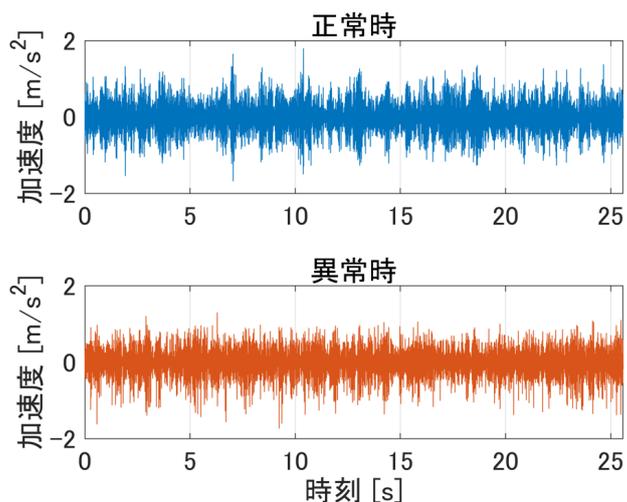


図2 振動データ波形：(上) 正常時 (下) 異常時

正常時・異常時に取得された2つの時系列データを例に、それぞれを区別するために有用な情報を簡易的に調査してみる。図3はそれぞれの加速度値の分布を表すヒストグラムだが、平均値や分散に決定的な違いがあるようには見られない。

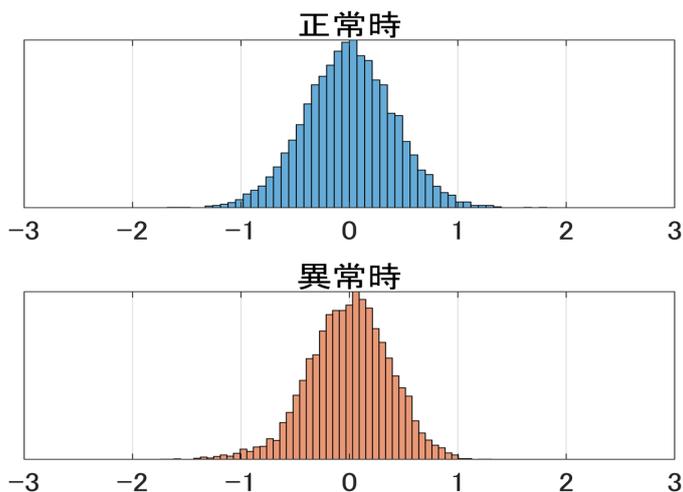


図3 加速度値の分布：(上) 正常時 (下) 異常時。いずれも似た分布であり平均値・分散などの特徴では区別が難しい。

次に周波数領域の情報を調査するためにスペクトル推定を実施したのが図4である。エネルギー密度が高い周波数帯では似た結果を示している。しかし100Hz以上の領域において、異常時のデータの方が高い値を示しており、異常であることを示す特徴量の候補として100Hz以上のバンドパワーが考えられる。

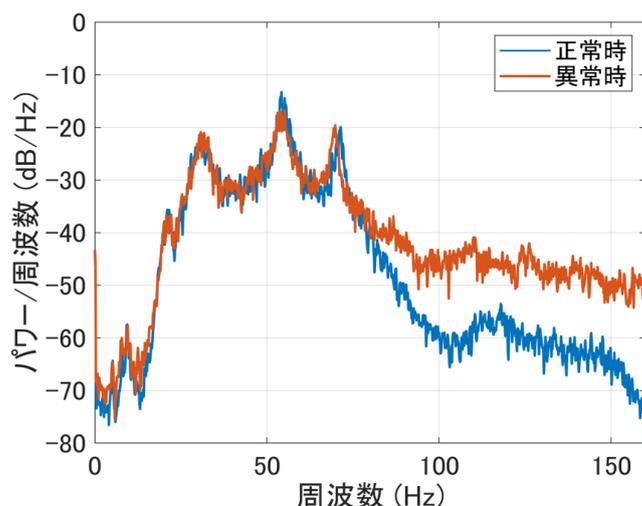


図 4 正常時と異常時のスペクトル推定比較：(青線) 正常時 (赤線) 異常時

自己回帰 (AR) モデルの係数を比較したのが図 5 であり、これまでの比較方法よりも明確に違いが表現されていることが分かる。この手法は時系列データそれぞれを自己回帰 (AR) モデルにフィッティングさせ、それぞれのモデルに表れる係数値を比較する。

$$X_n = a_0 + \sum_{i=1}^N a_i X_{n-i} + \varepsilon_n$$

X_n が時刻 n における計測値、 ε_n は残差項、 a_i が算出される AR 係数である。係数の変化はデータ発生源の特性変化ととらえることもできる。図 5 ではモデルの次数を 10 (すなわち $N=10$) とし、正常時・異常時のデータそれぞれから算出された係数を平行座標プロットで描いている。図 3 で表示した正常時・異常時のデータとは別に計測された正常時・異常時のデータから同様に算出された係数も薄い線で示した。正常時・異常時で係数の値に異なる傾向が存在し、異常時のデータでは係数値の分布も大きい傾向にある。

[1] では他にも様々な特徴量が検討されているが、もしこのように区別したい事象間で大きく異なる特性を見つけることができれば、異常検知を実施するのは非常に単純な課題となる。場合によってはしきい値での判断だけで十分な可能性もあり、そうすると機械学習の必要性もなくなる。

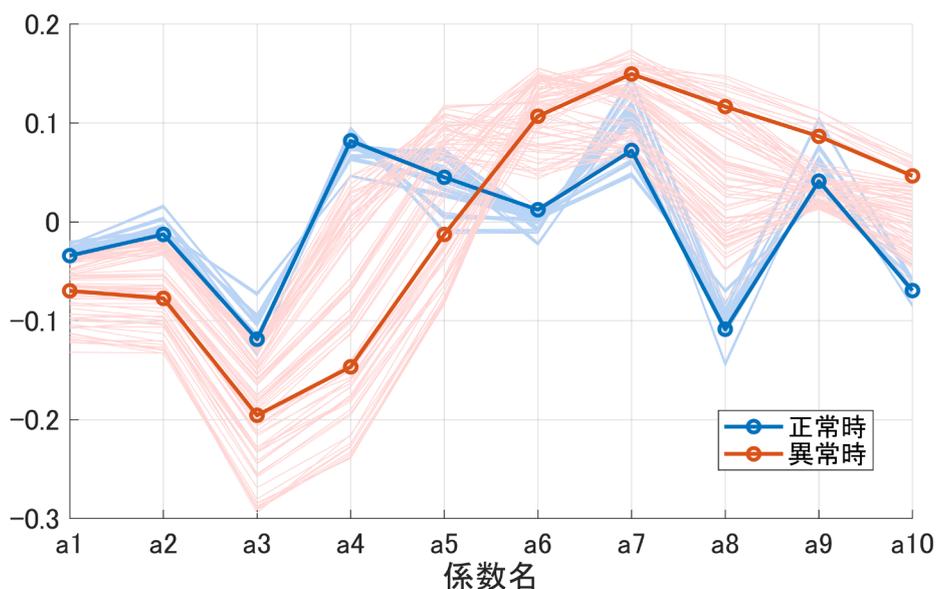


図 5 N=10 とした時の自己回帰モデル係数比較：
 (青) 正常時 (赤) 異常時。薄い線は同様の状況で計測されたその他のデータからの係数値を示す。
 *厳密には強制振動の時系列データを外部入力とした ARX モデルからの結果だが要点を伝えやすくするため文章中ではAR モデルとしている。

2.2 Human Activity Recognition (HAR)

次は Human Activity Recognition (HAR) のデータを紹介します。複数の被験者が歩く、寝る、階段を上るなどの Activities of Daily Living (ADL) を行っている際の加速度データである [3]。時系列データからそれぞれのADL を正確に予測する課題に取り組んだ [4] から一部を紹介する。

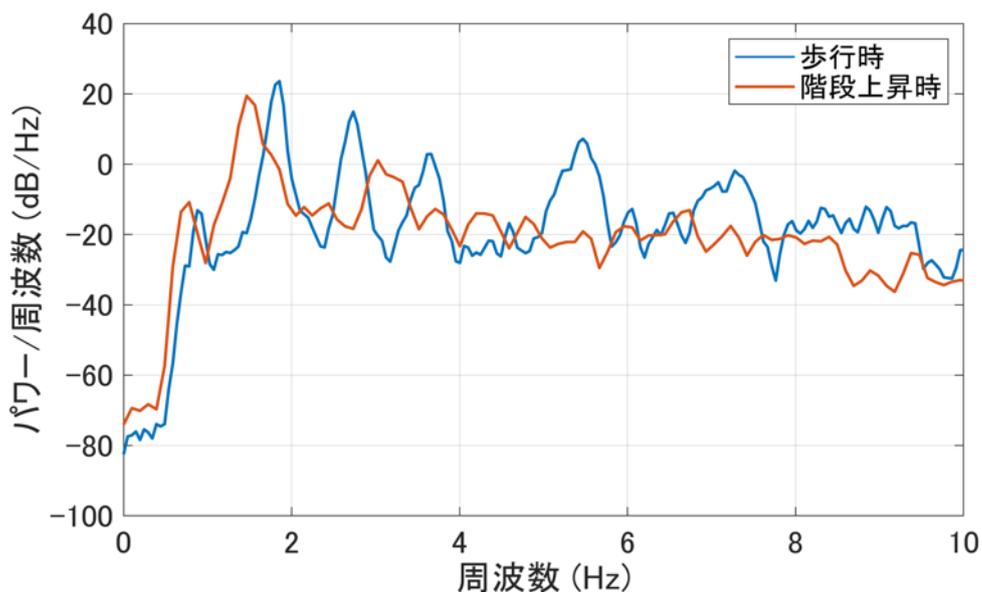


図 6 スペクトル推定結果：(青) 平地歩行時 (赤) 階段上昇時

図 6 はある被験者の2つの行動時にとられた加速度データからスペクトルを推定したものを示している。青線が平地を歩行時、赤線が階段を上っている時に観測されたスペクトル値である。歩行時は 2 Hz あたりに大きなピークがあるのに比べて、階段を上昇時は 1.5Hz 付近と比較的低い周波数でピークが観測されている。ここから階段を上っている時のほうが、動作が緩やかであったことが想像できる。また 4Hz より高い周波数帯では歩行時より階段上昇時のほうがエネルギー密度は小さく、よりハーモニックな、すなわちより一定のテンポで階段を上昇する姿も予測できる。

このように、直感的にも「なるほど」と感じるような特徴量をデータから導き出すことができれば、課題は半分以上解決したと考えてよい。直感に沿う特徴量を探すには予測対象の機器に詳しい技術者の勘・コツ・経験を活かした解析が必要である。平均、分散といった単純な統計量であっても十分な場合もある。特徴量と関くと身構えてしまうこともあるが、まずはヒストグラムを描いてみるだけでもヒントが見つかることも多い。

2.3 特徴量抽出の自動化

例で紹介したように複数の特徴量を算出して比較検討する試行錯誤は通常手間がかかる。ここでは作業を一部自動化するツールを紹介する。

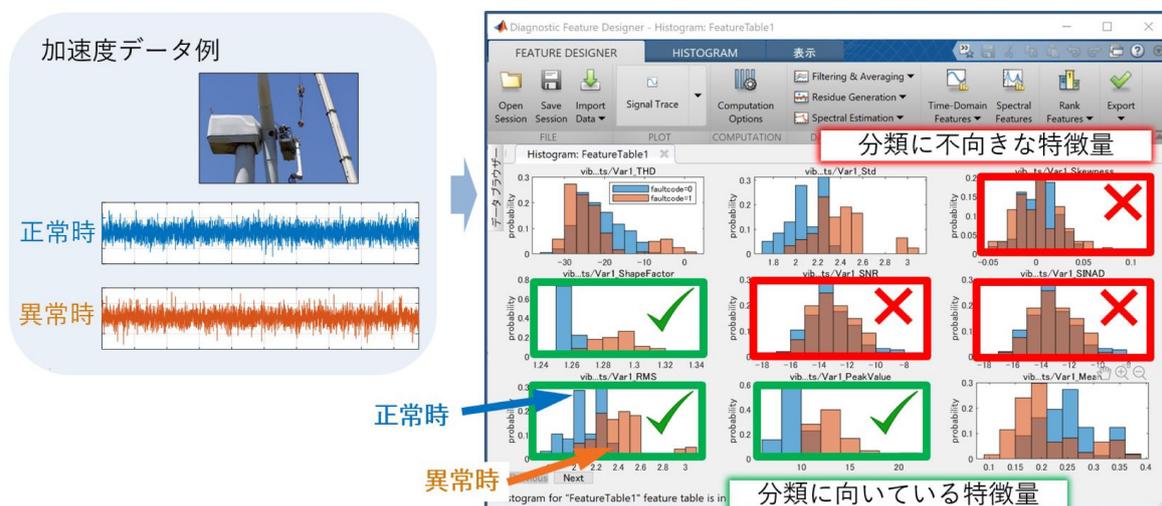


図 7 Diagnostic Feature Designer は読み込んだ時系列データから複数の特徴量算出・ヒストグラム表示、有効な特徴量を評価実行する GUI ツール。

Diagnostic Feature Designer (図 7) は時系列データから複数の特徴量を算出し、可視化、有益な特徴量の特定までをコードを書くことなく実現するGUIツールである。ノイズ除去などの前処理機能から、分散・尖度 (Kurtosis) ・波高率 (Crest Factor) などの時間領域の特徴量や、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サポートするため幅広いデータに対して適応が可能である。また機械学習モデルの検討を行うツールと組み合わせてアルゴリズム開発の工数削減を実現することができる。具体的な操作方法は [5] で紹介しているので、興味のある方は参照してほしい。

次節では、システム全体を考えたアルゴリズム選定の考え方と特徴量抽出のメリットを紹介する。

3. システムへの統合の観点から見た特徴量

構築した予測アルゴリズムの実際の運用方法を考えると図 8 に示す通り様々な実装先の可能性とそれに伴う課題が見えてくる。ハードウェア側の制約からアルゴリズム選択の幅が狭まる可能性もあるので、事前に運用方法を検討しておくことはアルゴリズムを設計する際にも役に立つ。



図 8 アルゴリズムの実装先もデータ転送量、計算量、運用方法によってデバイス上、エッジ上、クラウド上など様々な選択肢が存在する。

センサーデバイスの消費電力や通信帯域を考慮する場合、ここでも効果的な特徴量の算出が重要である。例えば特徴量の抽出を予測対象となる機器上で行うことができれば、データ転送量を大幅に削減することができる。特に振動や音といった情報から異常を検知する場合には高いサンプリングレートが要求されるため、機器上で必要な前処理を実施し予測に必要なデータだけを転送するという仕組み（エッジコンピューティング）が効果的となる。異常検知や寿命予測は圧縮された情報をもとにクラウド上で集中管理してもよい。予測アルゴリズム自体を機器に組み込む場合 [8]、大量のメモリを必要とする計算負荷の高いアルゴリズムは適さないため、情報を圧縮するという意味で特徴量の計算が役に立つだろう。

ここまではデータを予測に結び付ける部分の話だが、予測結果の活用方法についても考察が必要である。予測結果を見るのは整備スタッフか、機器のオペレーターなのか。その場合PC上で動くアプリが適切な場合もあれば、タブレットで持ち運べる実装がよい場合もある。予測結果をどう運用に反映させていくのかについての議論を深める必要がある。

4. まとめ

本稿では異常検知や予知保全の仕組みを構築するうえで重要な要素である特徴量について紹介した。監視対象の機械から取得されるデータを活用する為に必要なステップを一般化してみると、下記の4項目に大別されるが(図9) 予測モデルの構築自体は全体の一部であることに注目してほしい。



図 9 データ解析全体像を構成する 4 ステップ。各項目の具体的な内容についてはMondi 社の実例 [6] をもとに [7] で解説している。

AI という機械学習を適用する部分だけが注目を集めがちだが、データはどこから来るのか、そして最終的にアルゴリズムはどう実装するのか、それらの点も踏まえてシステム全体を考えた開発が重要になる。

特に故障の兆候を示す特徴量の選定・評価には多くの試行錯誤が求められ、機器に対する深い理解が必要不可欠である。MATLAB をアルゴリズム開発からシステムへの統合まで、プラットフォームとして組織的に利用することにより、より一層資産共有及び再利用が加速され、開発効率向上及びコスト削減が期待できる。

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」で検索してほしい。今回紹介した処理内容は [2,4,5] にて MATLAB のプログラムも合わせて公開しているので興味のある方はダウンロードして試していただきたい。

MATLAB とは

MATLAB は、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環境であり、データ解析、アルゴリズム開発、モデルやアプリケーションの作成が可能である。35 年以上の実績を持つソフトウェアであり、世界中で 400 万人を超えるエンジニアや科学者が、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に、MATLAB を共通言語として使用している。

参考文献

- [1] Figueiredo, Eloi, Gyuhae Park, Joaquim Figueiras, Charles Farrar, and Keith Worden. “Structural health monitoring algorithm comparisons using standard data sets.” No. LA-14393. Los Alamos National Lab. (LANL), Los Alamos, NM (United States), 2009.
- [2] 太田英司, “センサーデータ解析と機械学習 ～振動データからの異常検出～” (Web セミナー)
- [3] Anguita, Davide, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. “A public domain dataset for human activity recognition using smartphones.” In Esann. 2013.
- [4] 竹本佳充, “機械学習のための信号処理” (Web セミナー)
- [5] 井上道雄, “予知保全を可能にする特徴量選択” (Web セミナー)
- [6] MathWorks, “*Mondi*, 機械学習を使用した統計ベースの状態監視および予知保全を製造プロセスに導入” (ユーザー事例)
- [7] MathWorks, “*実例に学ぶ予知保全向けデータ解析*” (ホワイトペーパー)
- [8] MathWorks, “*Respiri*, 喘鳴検出および喘息管理のモバイル アプリとサーバー ソフトウェアを開発” (ユーザー事例)