

大規模データを想定した異常振動検知手法

今村 謙汰* 横内 俊秀 近藤 稔 (鉄道総研)

Anomalous vibration detection method assuming large scale data

Kenta Imamura*, Toshihide Yokouchi, Minoru Kondo, (Railway Technical Research Institute)

It has been confirmed in the past that anomaly detection method by the vibration can detect and predict anomalies in railroad car equipment. However, it was difficult to use the method when processing large data sets because it required trial-and-error learning by humans checking the data. Therefore, we introduced principal component analysis, databasing, and learning methods for eliminating anomalous data, which are applicable to large-scale data, so that conventional methods can be applied to large-scale data as well.

キーワード：車両検修，状態監視，機械学習，近傍法

(Railroad car maintenance, condition monitoring, machine learning, nearest neighbor method)

1. はじめに

主電動機やディーゼルエンジン，変速機といった鉄道車両の駆動用機器は，故障を起こすと列車運行を妨げることになる重要な機器である。そのため，駆動用機器の異常を早期に検知することを目的に，振動のオクターブバンド分析と機械学習を用いた状態監視手法の開発に取り組んでいる⁽¹⁾。振動による駆動用機器の状態監視手法では，学習データとして用いる振動データを選定し，それらに対し主成分分析を行う。その後，近傍法を用いた異常検知を行う。

しかし，上記の手法では人間がデータを確認しながら試行錯誤的に学習させる過程が発生し，大規模データを処理する際には手法の適用が困難であるという課題があった。そこで，大規模データに適用可能な主成分分析，データベース化，および異常データを排除する学習手法を導入することで，従来の手法を大規模データへも適用可能となるようにした。

2. 状態監視システムの概要

〈2・1〉状態監視システムの構成 振動による駆動用機器の状態監視システム全体の構成を図 1 に示す。このシステムは，機器の振動を測定する振動加速度センサ，車上に設置した状態監視装置，および地上の PC に実装した診断プログラムで構成される。振動加速度センサで取得した振動データは，状態監視装置でオクターブバンド分析を行い 1 秒毎に各周波数成分の振動加速度として記録する。地上の診断プログラムでは，このデータを用いて異常を判定し，機器の状態を監視する。

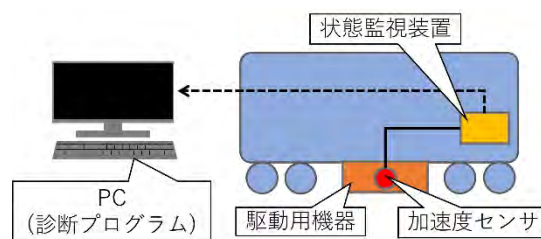


図 1 状態監視システムの構成図

Fig. 1. The condition monitoring system by vibration analysis.

〈2・2〉診断プログラム 駆動用機器は様々な回転数や出力トルクで動作するため，得られる振動データは複雑なものになる。したがって，振動の大きさに閾値を設定して異常判別するといった単純な方法での異常検知は難しい。そこで，診断プログラムでは機械学習の一種である近傍法を使用する。具体的には振動データをベクトル \vec{x} として多次元空間にプロットし，正常時のデータとの距離を計算する。その結果を予め設定した距離（基準距離）と比較し，異常度と呼ばれる値 $a(\vec{x})$ を求める。異常度は，判定を行いたいテストデータと正常データとの距離が設定した基準距離以下であれば $a(\vec{x}) \leq 0$ ，基準距離より大きければ $a(\vec{x}) > 0$ となるようにする。また，異常度の範囲は $-1 < a(\vec{x}) < 1$ となるようにする。この方法により，判定を行いたい振動データの異常度が負であれば正常，異常度が正であれば異常と判断し，異常振動を検知する(図 2)。

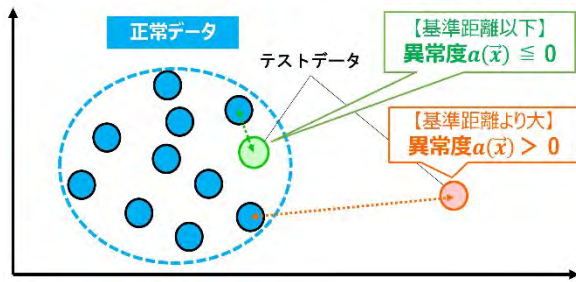


図 2 異常検知の方法

Fig. 2. Method of anomaly detection.

近傍法では、1つのテストデータに対して全ての学習データとの距離を計算する必要がある。したがって、年間を通じて取得した車両の振動データを学習データとして使用する場合、多大な計算時間を要することとなる。そこで、状態監視システムにおける診断プログラムでは、「代表データ選択手法」⁽²⁾を使用している。図3に代表データ選択手法の概要を示す。代表データ選択手法では、まず、学習データ領域の中心近傍にある少数のデータを代表データとして定め、前述の基準距離を指標として学習データ自身を対象に異常判定を行う。その際、異常判定された学習データのうち異常度が最も小さいデータを既存の代表データに追加し、再度残りの学習データを用いて異常判定を行う。異常と判定される学習データが無くなるまでこの処理を繰り返し、代表データで構成される正常判定の領域を広げていく。これにより、類似したデータを間引きつつ、必要な代表的な学習データのみで構成された、異常判定を行うための分類器を作成することが出来る。

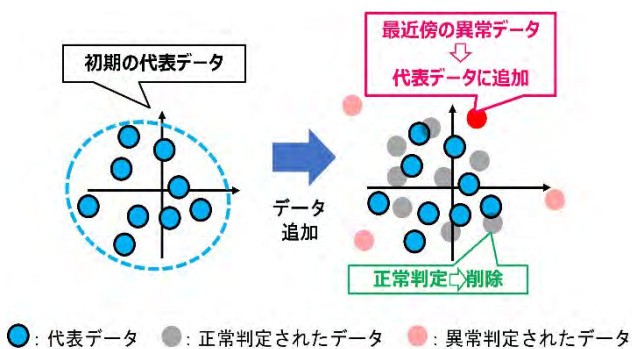


図 3 代表データ選択手法の概要

Fig. 3. Schema of prototype detection.

3. Incremental-PCA による主成分分析

振動データの前処理として、状態監視システムにおける診断プログラムでは、扱うデータ量を削減するために主成分分析(PCA: principal component analysis)により次元削減を行う。この時、振動の季節変動も考慮すると測定で得られた

すべての学習データを用いて主成分を求めることが望ましい。しかし、主成分分析では、対象となる振動データ全てを計算機のメインメモリに収め特異値分解を行う必要があり、数年分の全ての振動データを一度にメインメモリに収めることはメモリ容量上困難である。そこで、本研究では Incremental-PCA⁽³⁾を導入した。図4に Incremental-PCA を使用した主成分分析のフローを示す。Incremental-PCA ではメインメモリに収まる大きさに分割した振動データを逐次追加しながら主成分分析を行える。ただし、Incremental-PCA ではデータの追加順序によって計算結果が変わる可能性がある。そこで、振動データを追加している間は元の次元数を維持し、すべての振動データに対する Incremental-PCA が完了した時点で、寄与率の大きい順に必要な次元数の主成分を出力することとした。以上の処理によって、データの追加順序によって結果が変わることを防ぎつつ、年単位で取得した大規模な振動データにおいても PC のメモリ容量に制約を受けない主成分分析が可能になった。

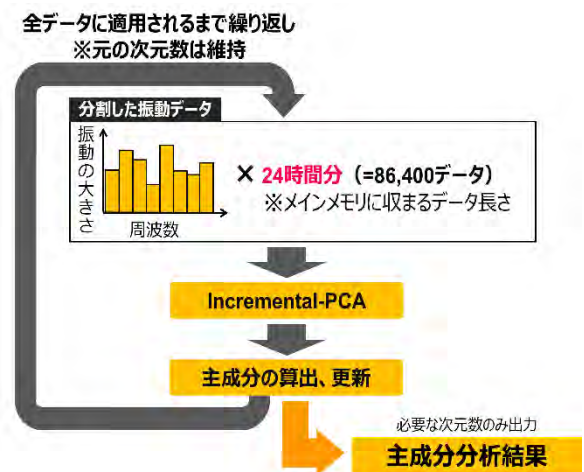


図 4 Incremental-PCA を使用した主成分分析

Fig. 4. Flow chart of PCA by applying Incremental-PCA.

4. データベース化および SQL を利用した処理

〈4.1〉 振動データのデータベース化 状態監視装置では測定した振動のオクターブバンド分析結果を 1 秒単位の CSV ファイルとして出力している。従来はこの CSV データを直接読み込み PC のメインメモリに収めた上で、主成分分析、代表データ選択といった処理を行う必要があった。この場合、メモリ容量に収まらないデータ量は処理出来ないため、年単位で取得した大容量の振動データを扱うことは不可能であった。そこで、CSV 形式データをデータベース化し、データベース言語 SQL (Structured Query Language) で取り扱えるようにした。これにより PC のメモリ容量に縛られることなく、数年間の学習データ全体から指定した割合の学

習データのみを読み出すといった処理が可能になる。また、複数のシステム操作者が同時にデータベースにアクセスすることも可能になり、将来の実用化も見据えたシステムとすることが出来る。

〈4・2〉 SQLによる基準距離の算出 2・2項で述べた通り、近傍法に基づく異常度計算では基準距離を事前に決定しておく必要がある。その際、従来の手法では基準距離の算出に適切なデータであるか試行錯誤的に検証しながら、PCのメモリに収まる1日分程度の学習データを選定しなければならなかった。

そこで本研究では、振動データのデータベース化により使用可能となる、SQLによるランダムサンプリングを導入した。正常データであることが確認できている測定開始後数ヶ月程度のデータから、PCのメモリ容量の制約下で処理可能なデータ数をランダムに選択し、基準距離の算出を行った(図5)。この基準距離の算出に使用したデータは、代表データ選択の起点となる初期の代表データとしても使用する。

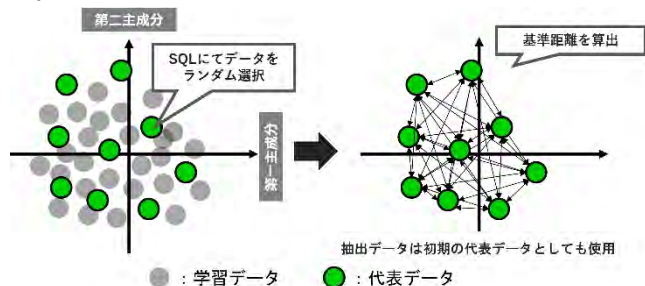


図5 ランダムサンプリングによる基準距離算出

Fig. 5. Calculating reference distance by applying random sampling.

5. 異常データを排除した学習手法

状態監視システムでは、代表データ選択手法の導入により異常度計算の高速化と負荷軽減が図れ、一般的な性能のPCでも異常判定が可能な分類器を作成できる。一方で、学習させる振動データは駆動用機器が正常状態にあるときに取得したデータである必要がある。異常な振動データが学習データに含まれる場合、駆動用機器に不具合が発生しても異常検知できない可能性がある。長期間取得した大容量の振動データを、車両の検修履歴や故障、不具合の申告実績を確認しながら選別することは困難であり、また、運行上支障のない軽微な不具合による異常な振動データは取り除けない可能性もある。そこで、異常度に基づいて代表データを追加していく過程において、追加する学習データの異常度に閾値を設けた。

2・2項で述べた通り、異常判定される振動データの異常度 $a(\vec{x})$ は $0 < a(\vec{x}) < 1$ の値となる。代表データを新た

に追加する過程において、この範囲内で閾値を設定し、異常度が最も小さい異常データであっても、閾値以上であれば代表データに追加されないアルゴリズムとした。これにより、分類器として作用する代表データに異常データが選択されてしまうことを防止する。

6. 大規模データを使用した検証

実際に営業走行時の車両で取得した駆動用機器の振動データを用いて、今回提案した手法で分類器として作用する代表データの作成が可能であるか検証した。状態監視システムにおける診断プログラムと同様に、本研究ではプログラミング言語のPythonを用いて検証した。

〈6・1〉 Incremental-PCAによる主成分分析 Incremental-PCAを使用した場合も、通常の主成分分析と同じ結果が得られるかどうか検証した。Pythonライブラリであるscikit-learnを導入し、Incremental-PCAおよび通常の主成分分析を行った。主成分分析の比較における計算条件を表1に示す。主成分分析により40次元ある振動データを6次元に削減する。

表1 主成分分析の計算条件

Table 1. Calculation condition in PCA.

項目	内容
振動データの種類	エンジン振動データ
使用したデータ数	30日分 (2019年10月取得)
Incremental-PCAで追加するデータ単位	1日分 (最大86,400秒分)
Incremental-PCAでのデータ追加順序	ランダムに5パターン
白色化	実施
主成分の数	6

主成分分析を行い得られた、第一主成分における中心周波数と主成分ベクトルの要素の関係を図6示す。茶線のプロット(PCA)が通常の主成分分析で得られた主成分である。その他のプロット(Inc PCA1~Inc PCA5)は、振動データの追加日順序をランダムに変えてIncremental-PCAを行い得られた主成分である。横軸は左から昇順で振動データの各中心周波数を並べたときの値であり、縦軸は各中心周波数における主成分の要素値である。グラフ上部には寄与率(Explained varratio)を記している。主成分分析の手法、計算順序に関わらず主成分ベクトルがよく一致しており、振動データの追加順序に関係なく、Incremental-PCAにおいても通常の主成分分析結果が得られることが分かった。これは、次元削減後の全6主成分において同様の結果であった。

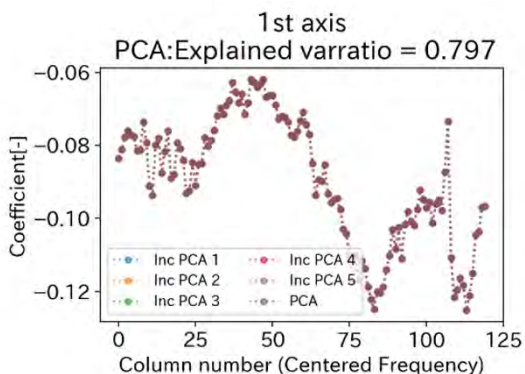


図 6 主成分分析の結果

Fig. 6. Result of PCA.

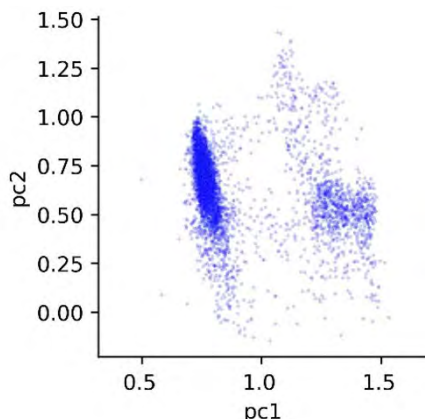


図 7 作成した代表データの一例

Fig. 7. An example of prototype data.

〈6・2〉 異常データを排除した代表データ作成 データベース化した振動データを用いて、異常データを排除しつつ分類器として使用する代表データの作成を行った。異常データは、営業走行車両への状態監視システム搭載期間中に、測定装置の配線に断線が生じた時のデータを使用した。代表データの作成に必要な基準距離の算出は、SQL によるランダムサンプリングを使用している。計算条件は表 2 の通りである。

表 2 代表データ作成における計算条件

Table 2. Calculation condition in creating prototype data.

項目	内容
振動データの種類	エンジン振動データ
使用したデータ数	584 日分
基準距離算出に用いるデータ	測定開始から約 4 ヶ月間で取得したデータ
起点になる初期の代表データ	基準距離の算出に用いたデータ
代表データ選択時の異常度閾値	0.05, 0.2, 0.4, 0.6, 0.8, 1.0(異常度の最大値。異常度による正常・異常判定なしと同等)
異常振動データの定義	振動加速度実効値が 3.0m/s ² 以下のデータ

異常データを排除しつつ作成した代表データの一例として、取得した振動データの高周波帯(1kHz~10kHz)を用いて代表データを作成し、横軸を第一主成分(PC1)、縦軸を第二主成分(PC2)として 2 次元空間にプロットした結果を図 7 に示す。また、図 8 に代表データに含まれる正常データおよび異常データの内訳を示した。横軸は異常データを排除するために設定した閾値、縦軸はデータ点数とし、青色が正常データ、赤色が配線断線時の異常データのグラフである。閾値が 0.05, 0.2 の場合、代表データには断線時のデータは全く含まれておらず、閾値の設定により測定系の異常を自動で排除しつつ代表データを作成できていることを確認した。

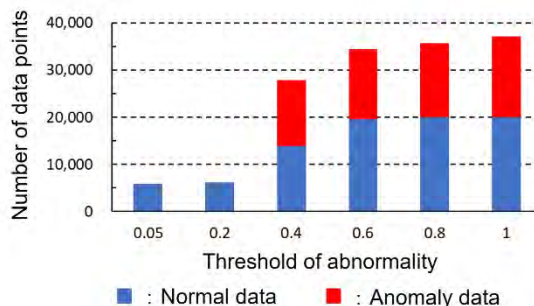


図 8 代表データの内訳

Fig. 8. Breakdown of prototype data.

7. おわりに

本研究では、振動による駆動用機器の状態監視手法において、Incremental-PCA、データベース化および SQL を使用した処理、閾値による異常データ排除を行い、その有効性を確認した。これにより、従来は人間がデータを確認しながら試行錯誤的に学習させていた作業を自動化し、数年間分の大規模データに対しても適用可能な異常検知の手法を実現することが出来た。

文 献

- (1) 西谷幸祐, 近藤稔, 高重達郎, 片岡祐太, 野口敬太: 「振動による駆動機器用状態監視システムの営業列車への適用」 鉄道総研報告 32 巻 8 号 pp.23-28 (2018)
- (2) 近藤稔: 「機械学習による車両機器状態監視のための代表 データ選択法」 交通電気鉄道・フィジカルセンサ合同研究会, TER-18-036 /PHS-18-011 (2018)
- (3) David A. Ross, Jongwoo Lim, Ruei-Sung Lin, Ming-Hsuan Yang: 「Incremental Learning for Robust Visual Tracking」, International Journal of Computer Vision 77, pp.125-141 (2008)