

データ分析プラットフォーム Alkanoを用いた異常検知

広島工業大学 前田研究室 畑井那津美

研究テーマ：Alkanoを用いた異常検知・分析

- Alkanoとは、NTTデータ数理システム様が開発した、ノーコードで容易にデータ分析ができるプラットフォームである。
- AI・機械学習や、クラスタリングなどを使用することで異常検知・分析をすることが可能である。



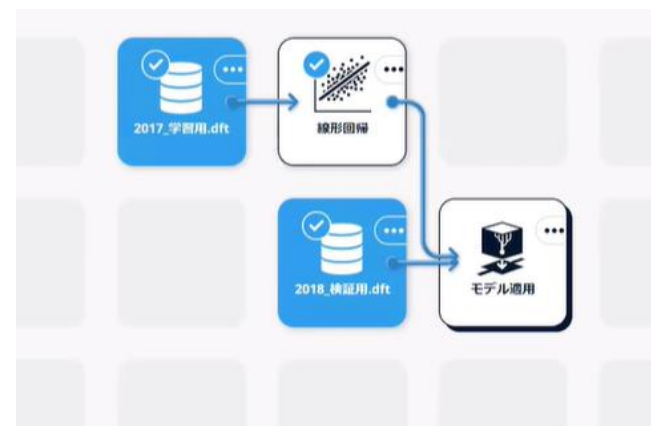
Alkano上の処理

- 予測分析やクラスタリング、ディープラーニング等様々な種類の分析が手軽に、素早くできる。
- Pythonでコードを組んで分析することもできる。



解析方法

- 使用するデータをアップロードし、それを始点に矢印をつなげていき、処理のフローを作成する。
- データをアップロードして矢印でつなげるだけなので、プログラミングができなくても簡単に操作できる。

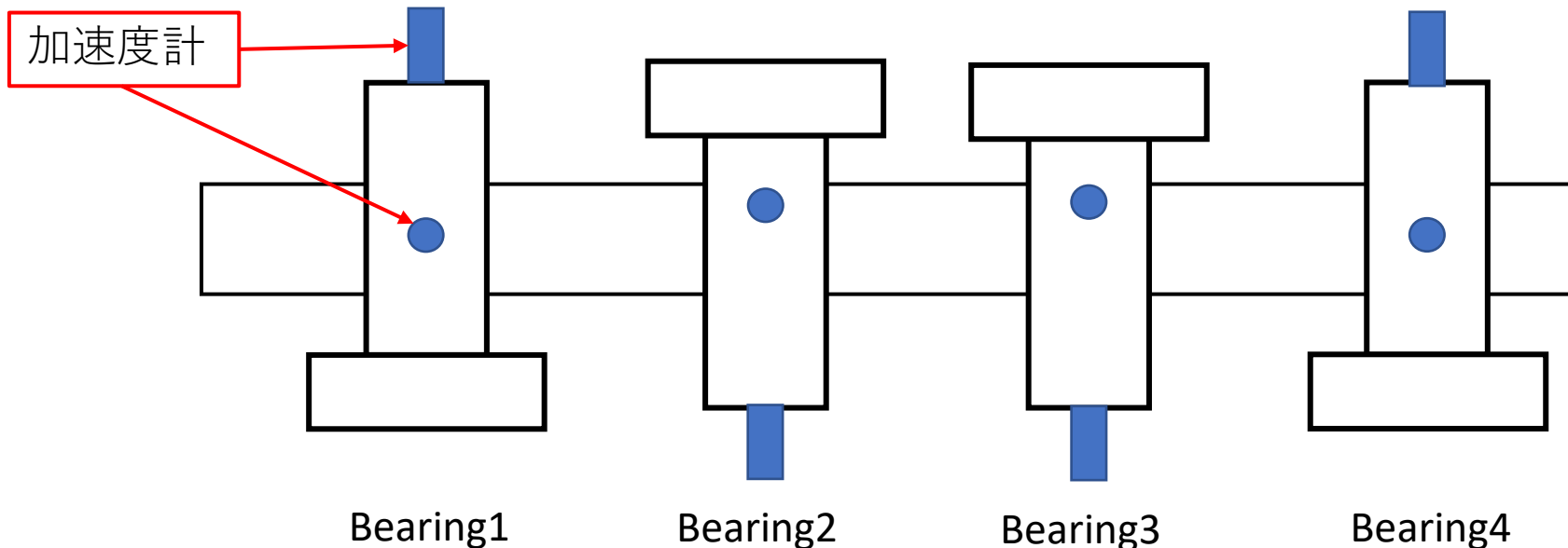


目的

- 本報告では、**Alkano**を用いた産業設備の異常診断方法を検討する。
- 公開されている**NASA**のベアリングデータで**Alkano**の基本機能を確認し、次に工場稼働しているコンプレッサを対象に異常診断手法を評価する。

代表的な公開ベアリングデータ

- NASAのデータが公開されている。図のように4つのベアリングが取り付けられており、それぞれのベアリングに加速度計が取り付けられている。



NASAベアリングデータの仕様

- 一定時間ごとに記録された振動信号のスナップショットのファイルによって構成されている。
- 各ファイルは、サンプリングレート20kHzの1秒間データからなる。
- 2004/2/12/10:32:39～2004/2/19/06:22:39の10分間隔で記録されており、948個のファイルで構成されている。
- **test-to-failure**実験の終了時ベアリング1が破損している。

NASAベアリングデータ

データ読み込みの際、前処理として各列ごとに平均値と絶対値をとった。

	Bearing 1	Bearing 2	Bearing 3	Bearing 4
2004-02-12 11:02:39	0.061455	0.073844	0.084457	0.045081
2004-02-12 11:12:39	0.061361	0.075609	0.082837	0.045118
2004-02-12 11:22:39	0.061665	0.073279	0.084879	0.044172
2004-02-12 11:32:39	0.061944	0.074593	0.082626	0.044659
2004-02-12 13:02:39	0.061630	0.075284	0.080999	0.044688
...
2004-02-18 15:42:39	0.094821	0.089184	0.090852	0.054423
2004-02-18 15:52:39	0.093979	0.087966	0.090682	0.054724
2004-02-18 17:12:39	0.115109	0.091404	0.094160	0.054266
2004-02-18 17:42:39	0.133617	0.094244	0.096894	0.055240
2004-02-18 17:52:39	0.134302	0.095112	0.098138	0.056938

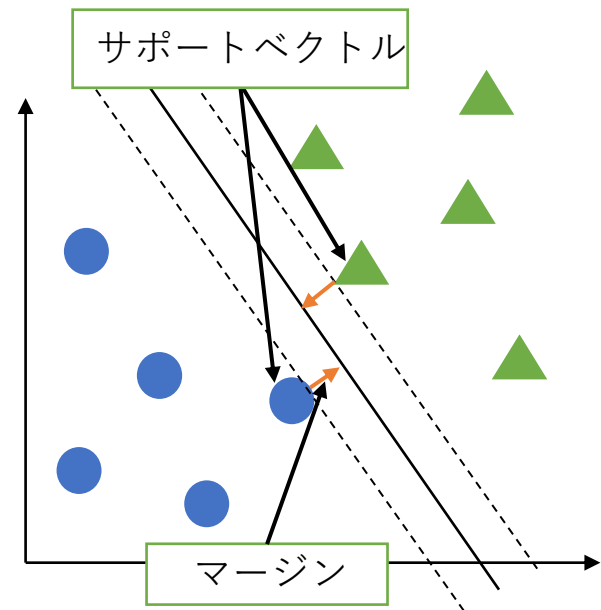
115 rows × 4 columns

課題

- NASAのデータは異常が明確にどこから始まっているか明らかにされていないため、教師データの必要がない外れ値検出手法を使いたい。
- よって今回はクラスタリング手法の一つであるOneClassSVMを使用して解析する。

SVM(サポートベクトルマシン)とは

- サポートベクトル (データを分割する境界に最も近いデータ) を用いてクラス分類を行う解析法。
- マージンが小さいと、少しのデータの揺れで誤判定してしまう。
- カーネルトリックを使用している。



SVM(サポートベクトルマシン)とは

- マージンを最大化する境界線を引くことがSVMの目的である。
- 最大マージンはこの式で与えられる。

$$D_{max} = \frac{1}{\|w\|} = \frac{1}{\sqrt{w_0^T w_0}} = \frac{1}{\sqrt{\sum_{i=1}^N \tilde{a}_i}}$$

- 少ない学習データからでも一定の識別性能が得られる。
- ν というパラメータがあり、データに含まれている異常値の割合である。

One Class SVMとは

- 教師あり学習のSVMと違い、教師なし学習である。
- SVMと違い複数のラベルを付けて分類することなく、学習データをクラス1とし入力データがそのクラスに入るか入らないかのみ判断している。
- カーネルトリックを使用している。

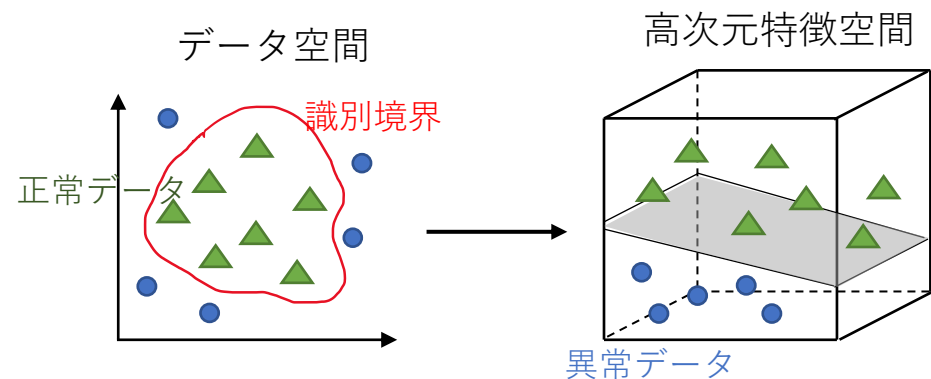
One Class SVMとは

- 外れ値か否かの識別関数は、

$$f(x) = \text{sgn} \left(\sum_{i=1}^N \alpha_i K(x_i, x) - \rho \right), \text{sgn}(a)$$
$$= \begin{cases} +1 (a > 0) \\ 0 (a = 0) \\ -1 (a < 0) \end{cases}$$

カーネルトリックとは

- カーネル法という高次元空間へデータを写像する手法により、例えば2次元のままでは複雑な境界になってしまうデータも、Z軸方向でうまく分けることができる。
- カーネルトリックでは、カーネル関数を使うことで、高次元空間上で評価したスコアを計算できる。
- このとき、異常データが原点の近くに集まる。



カーネルトリックとは

- 今回使ったのはガウスカーネルである。

$$K(x, x') = \exp\left(\frac{-\|x - x'\|^2}{2\sigma^2}\right)$$

- パラメーターとして σ^2 の数値を変化させる。

PythonでのOne Class SVM（比較確認）

- 今回はsklearnのOne Class SVMを使用した。

```
from sklearn.svm import OneClassSVM

X = merged_data['Bearing 1']
X = np.array(X)
X = X.reshape(-1, 1)
clf = OneClassSVM(nu = 0.3).fit(X)
clf = clf.predict(X)
clf
n = clf.shape[0]
Y = merged_data['Bearing 1']
Y = np.array(Y)
N = np.arange(n)

for i in range(n):
    plt.plot(N[i:i+2], Y[i:i+2], color='red' if clf[i] == -1 else 'blue', linestyle = 'None', marker='.')

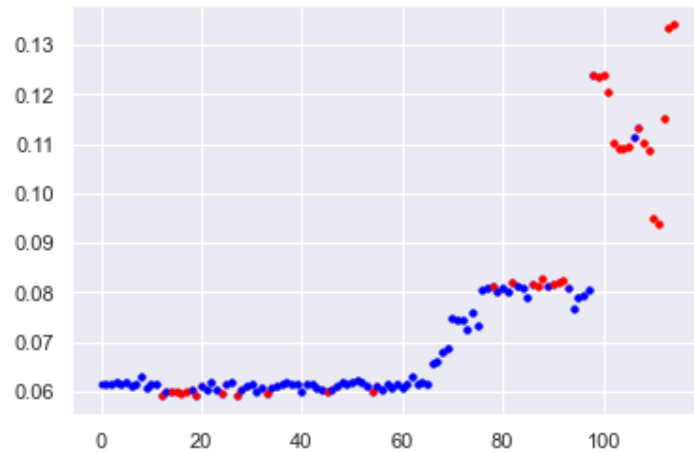
plt.figure(figsize=(5, 10))
plt.show()
```


実行結果

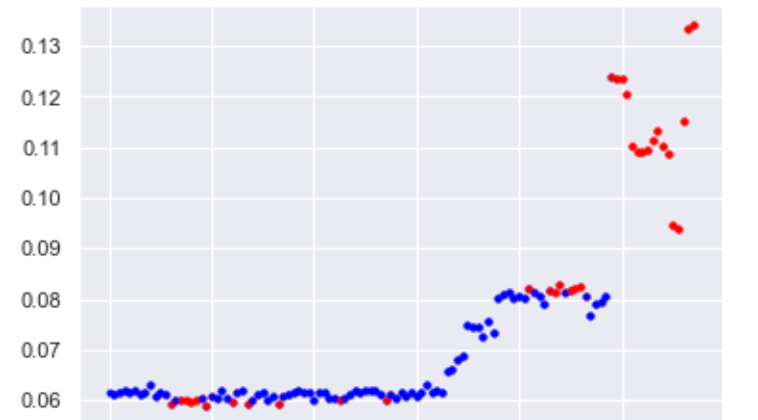
- 実行結果をエクセルファイルとして出力した。
- normalとabnormalに分けられている。
- 目的変数にBearing1のデータとした。
- Pythonの方は、 $v = 0.3$ に設定。
- Alkanoの方は $v = 0.3, \sigma^2 = 0.1$ に設定。

NASAベアリングデータ実行結果

- 赤がabnormal、青がnormal。
- Pythonで解析したものとAlkanoで解析したものを比較した。
- AlkanoでもPythonと同じように解析できる。



Python



Alkano

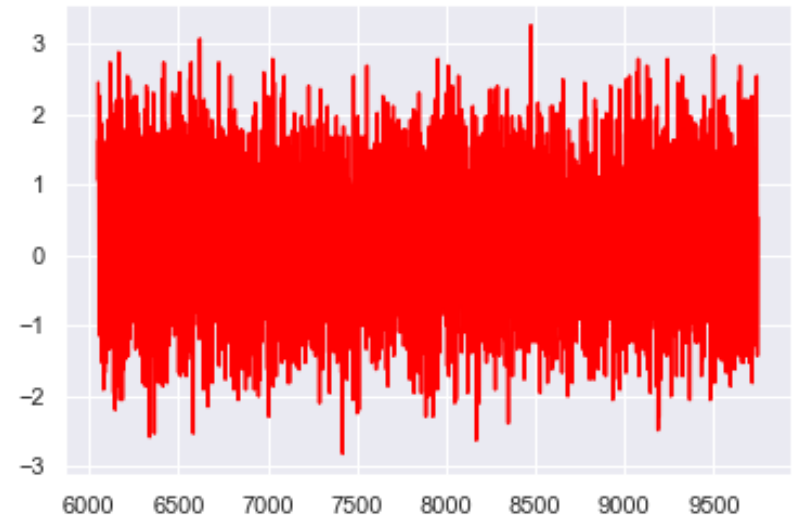
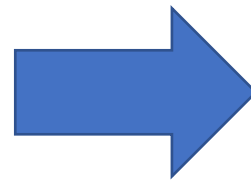
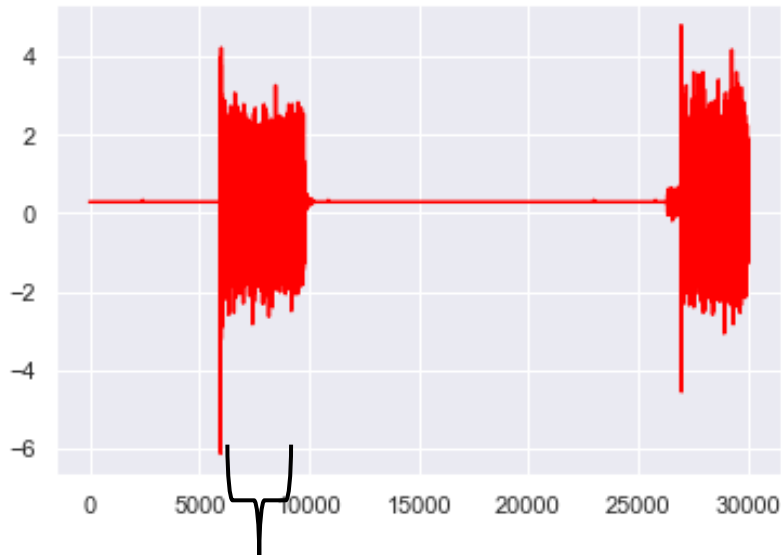
2004-02-12T11:02:39.000+09:00
2004-02-13T13:02:39.000+09:00
2004-02-14T16:~00
2004-02-16 ~00
2004-02-17 ~00
2004-02-18T08:22:39.000+09:00

対象設備：コンプレッサ

- コンプレッサSRL-22DMPの振動データを使用。
- マイクロストーン社製のセンサによってX軸、Y軸、Z軸ごとの振動をそれぞれ計測した。
- センサは、コンプレッサ内の圧縮機と電動機の上に設置されており、サンプリング周波数は200Hzである。
- 異常が存在しないとわかっているデータなので、疑似的に異常を付与したデータを作り解析する。
- こちらもOneClassSVMで解析していく。

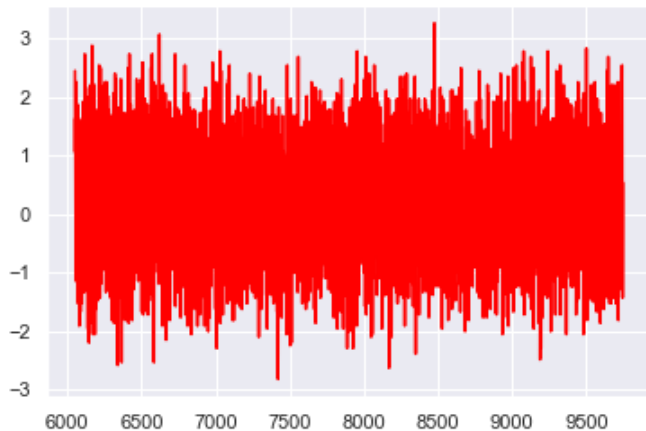
コンプレッサデータの前処理

- 波形から立上りと立下りの部分を除いた波形データを作成した。

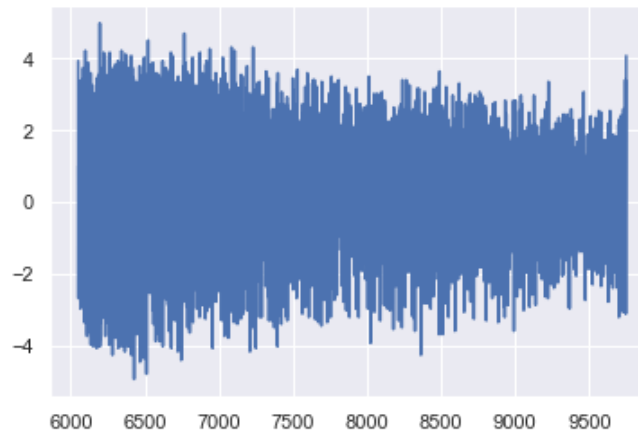


コンプレッサデータの前処理

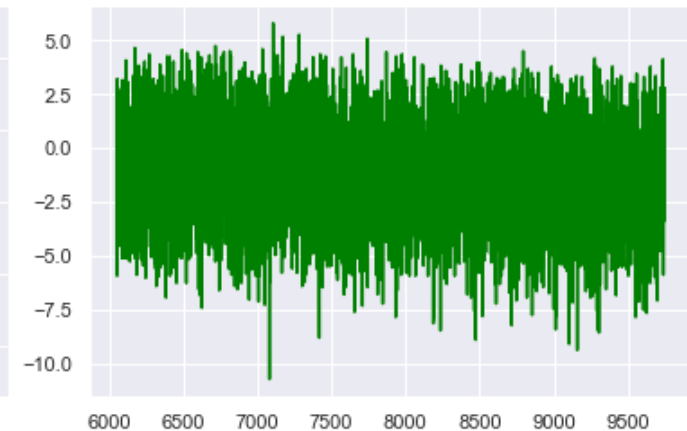
- X軸、Y軸、Z軸の波形データを求めたものがこれである。これらをもとに解析していく。



Ax



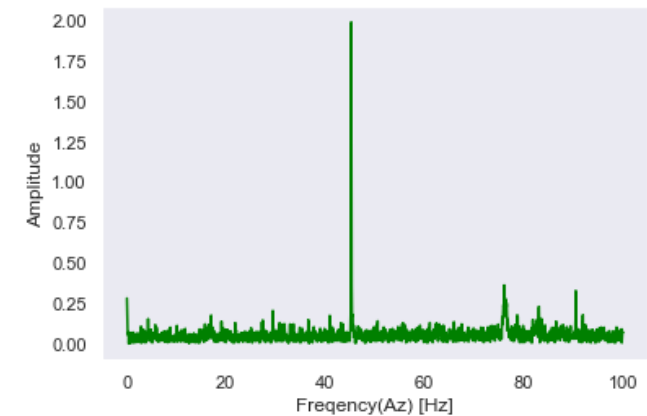
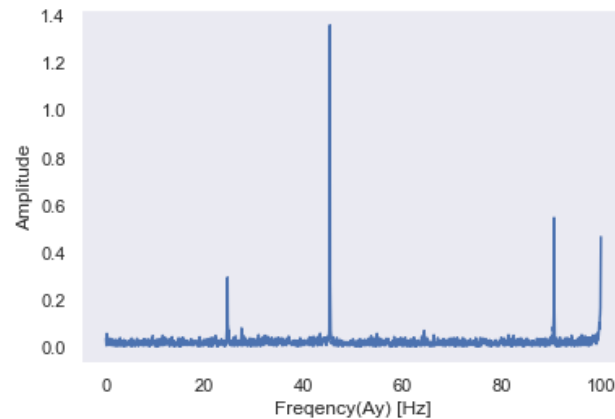
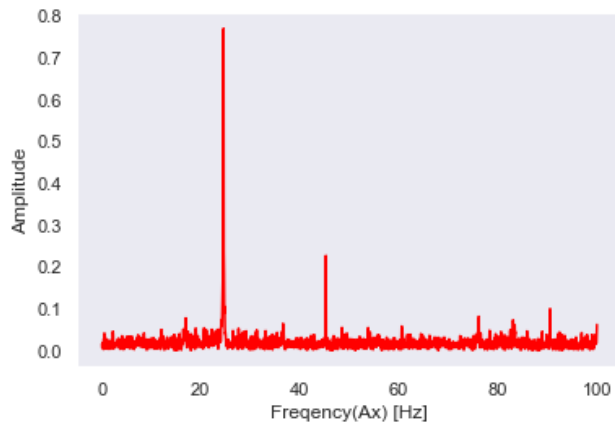
Ay



Az

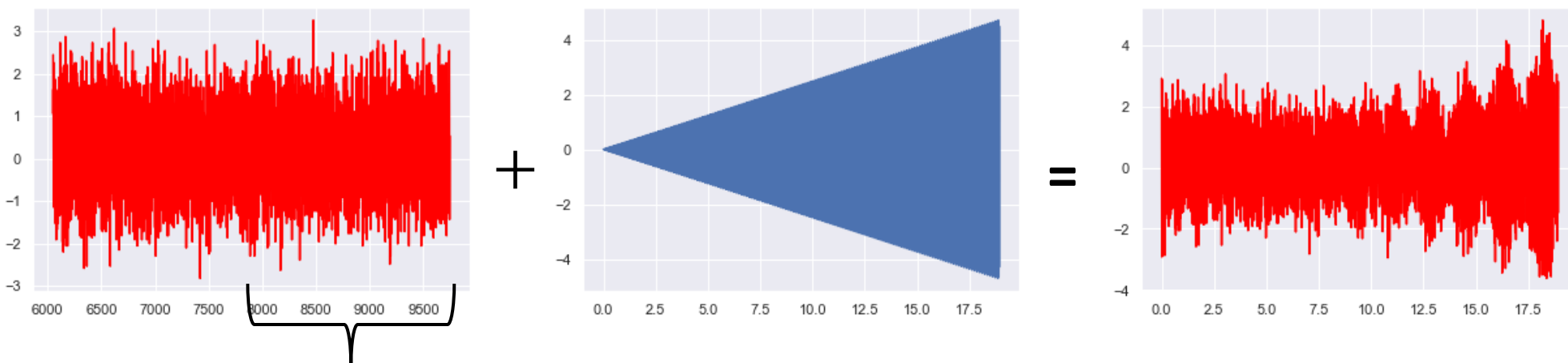
疑似異常データの作成

- 波形データに加算する異常波形を作成するため基本周波数を求める。
- フーリエ変換で求めたところ24Hz、45Hzという結果になった。
- メーカーに問い合わせた結果、実際の基本周波数は24Hz (モータ起因)、55Hz (圧縮機起因)なので今回はこれらを使用する。



疑似異常データの作成

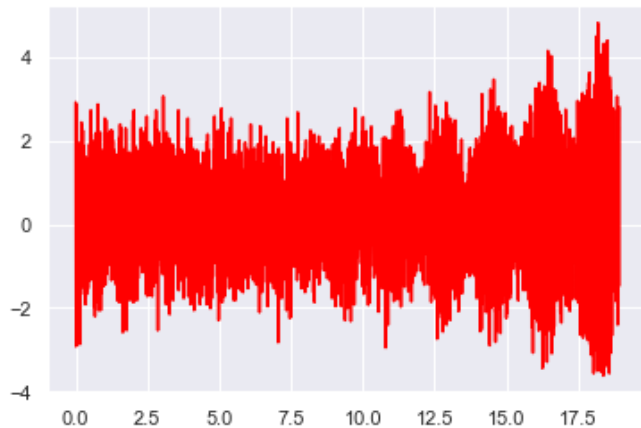
- 基本周波数の2倍の周波数で徐々に大きくなる正弦波データを元の波形の後半に加算し、徐々に異常が大きくなっていく疑似異常データを作成した。



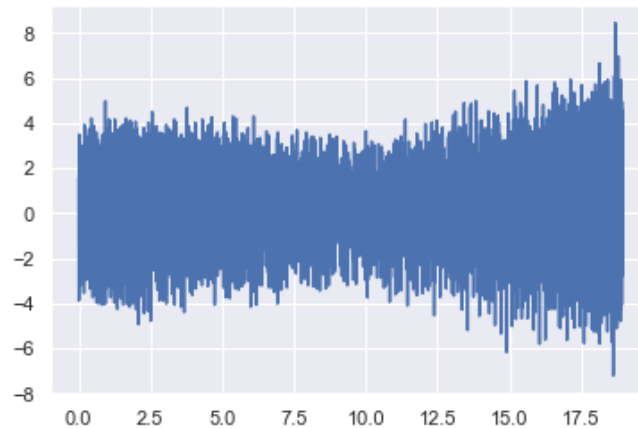
波形を加算した部分

作成した疑似異常データ

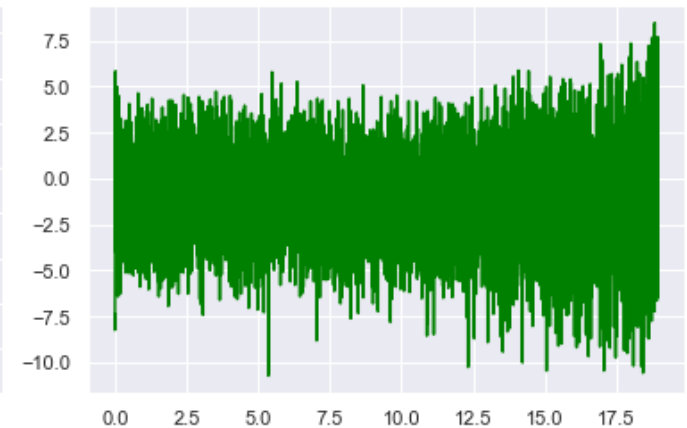
- x軸の波形には24Hzの倍、y軸とz軸の波形には60Hzの倍の周波数の正弦波データを加算した。



Ax



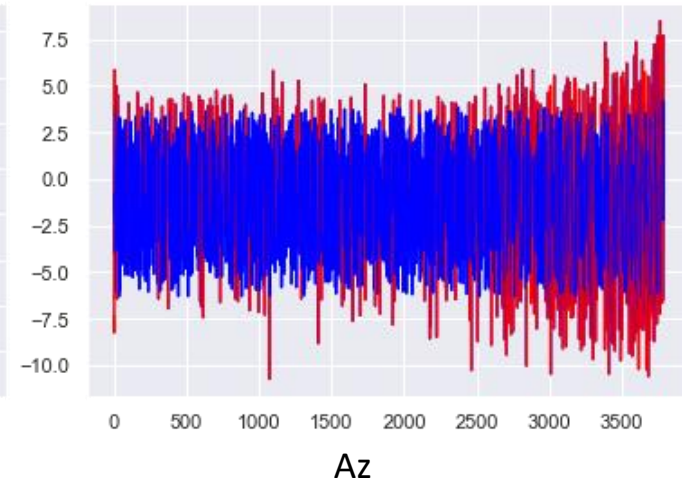
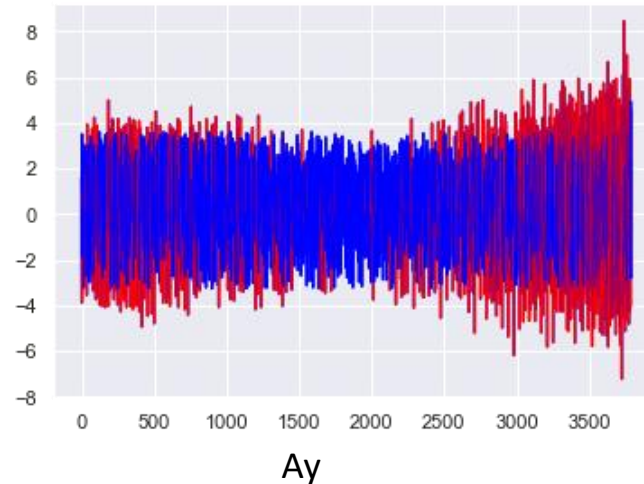
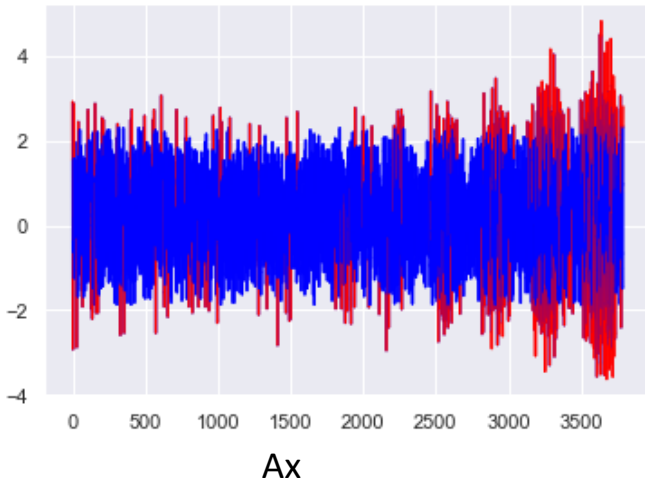
Ay



Az

OneClassSVMを使用した解析結果

- NASAのデータと同様OneClassSVMで解析。
- $\nu = 0.15, \sigma^2 = 0.05$ に設定した。
- 赤の部分が検出された異常である。



考察

- X軸の後半に現れた脈動データは異常データを加えたことにより発生したものである。ONeClassSVMによりこれを異常と判定していることがわかる。
- モータ、圧縮記は間欠動作をしており、起動の衝撃で剛性の弱いY軸方向に低周波振動が見られる。
- Y軸、Z軸については、120Hzの異常波形を加えたのであるが、使用した加速度センサの応答が100Hz以下であるため、高周波の異常成分が若干抑圧されて観測されている。

結論と今後の課題

- 本報告では、Alkanoを用いた産業設備の異常診断方法を検討した。
- 異常波形を加算していない前半にも多少異常としてデータが出てきたが、後半に比べると少なく、異常データを判別できているといえる。
- 特にAyは元の波形が徐々に小さくなっていく形をしているので、異常波形を加算していない前半にも比較的大きく異常判定が出ている。
- Ayのような一定ではない形の波形だと異常のない部分でも異常判定が出やすいので、そういった形の波形でもうまく分析できる手法を考えたい。

Alkanoを使った感想

- ノーコードで処理することによってコードを組む時間が短縮でき、Alkanoに実装されている種々の解析手法を試すことができる。
- Pythonであらかじめ前処理したデータとAlkanoでの処理を組み合わせることによって、Alkanoでは実装されていない処理も使えてより複雑な処理ができる。
- Alkanoには相互情報量やKLD(カルバックライブラー情報量)などを求める機能がないので、そういった機能があると嬉しい。
- Alkanoはグラフ表示の際、アスペクト比を変化させたり折れ線グラフのマーカーを消したり等の操作ができないのが不便。

謝辞

- 今回、データを提供していただき、更にアドバイスをいただいた京セラSOC様には深く感謝いたします。