

AI Quest

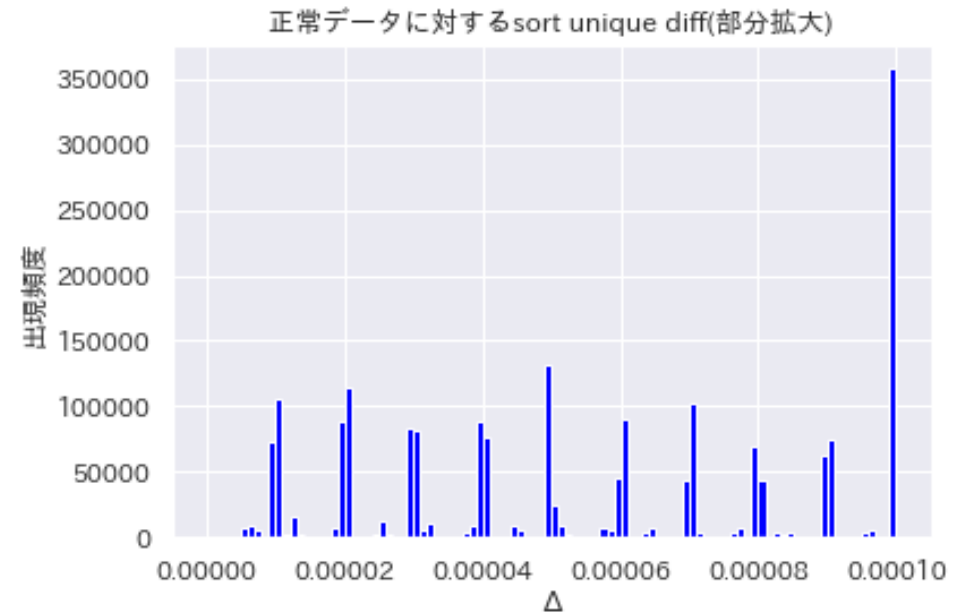
PBL06 機械設備の予知保全 製造

はじめに

- slackのb_pbl_06_機械設備の予知保全_製造を読んでいる方はご存じと思いますが、与えられたデータに癖があります

正常データ 分解能が0.00001
異常データ 半端な数字がある

せっかくなので役に立つ話を



自己紹介

フーリエ変換をさんざんやらされた

-
- 大学生の時の専門は画像認識、VR
- 同級生はニューラルネットワークをしていた
 - あんな遅いもの実用になるはずはないと思ってた
- 社会人になって最初の仕事はCNC旋盤の予知保全
- 現在 研究職

課題を見て最初に思ったこと

いので、当然デメリットもありま
由の周りのベアリングが壊れ、そ
たら気づくこともあります。著

+

エンベロープ解析 (1/2)

る事で、衝撃的な振動が発生します。



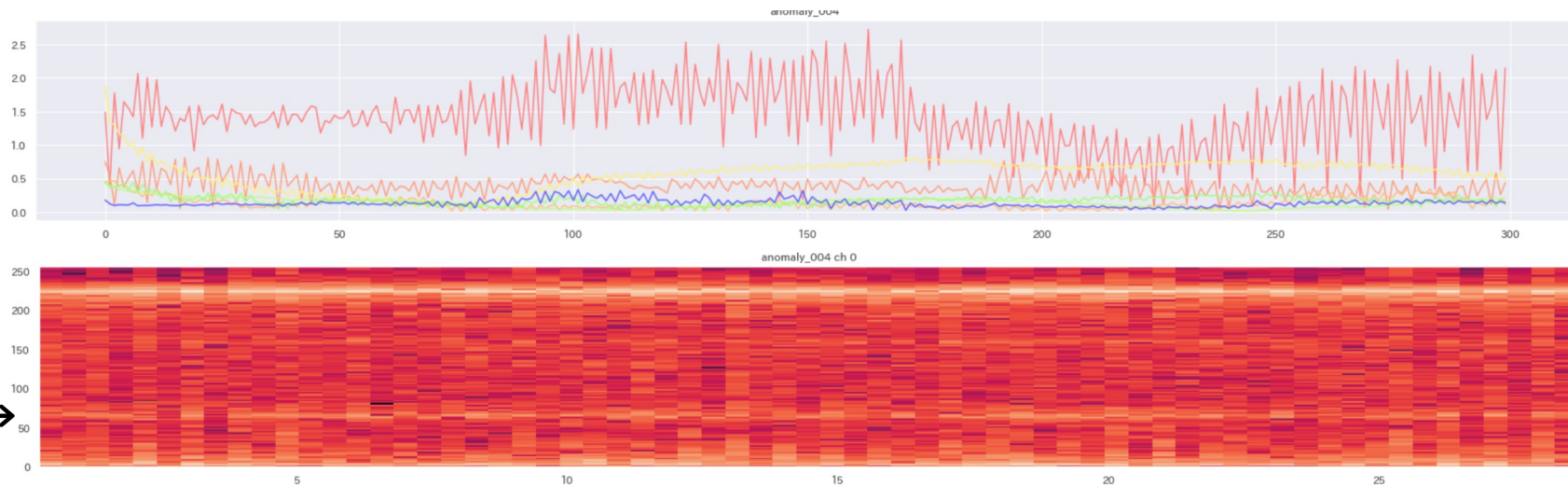
うなり音を検出すればいい



最初にしたこと

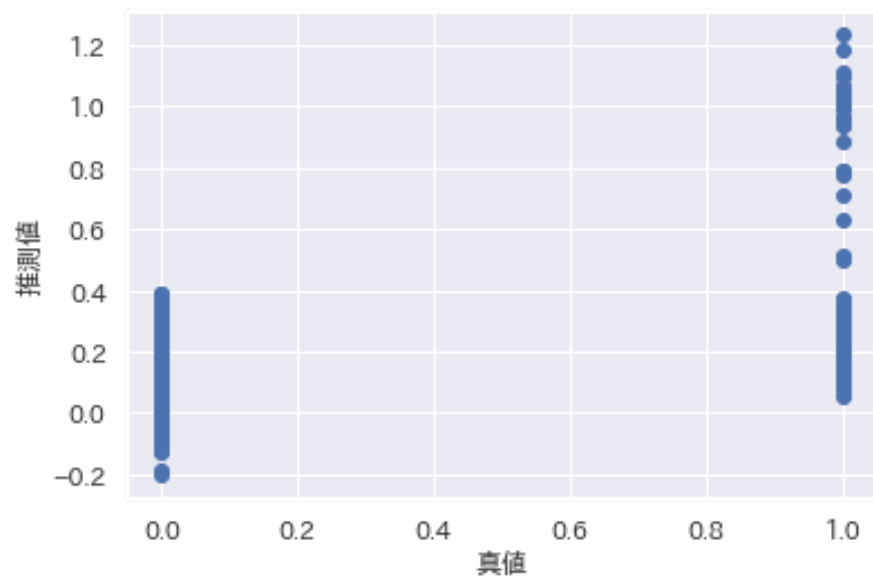
- 元波形とエンベロープを表示
- 短時間フーリエ変換も表示

ピークが
ありそう→



最初の提出

- 特徴量: エンベロープの周波数分析
- 分類アルゴリズム: 線形回帰



ベテランも3割しか気づけないと言ってるしこれでいいか

ところが

暫定評価

1.0000000

暫定評価が1の人が出たので
そんなはずはないと思いながら試行錯誤



試したこと

- 信号処理
 - フーリエ変換
 - ヒルベルト変換(エンベロープ検出)
 - 短時間フーリエ変換
 - mean, std
 - 自己相関(振動・うなりの解析によく使われる)
 - tsfresh
- 前処理
 - 特徴量選択(**SequentialFeatureSelector**)
 - データオーグメンテーション
 - $z=ax+(1-a)y$
 - 正規化
- 分類アルゴリズム
 - 線形回帰
 - RandomForest
 - SVC
 - オートエンコーダー(正常例再現)
 - CNN(後述)
 - メトリックラーニング

特徴量選択をした理由

- 計算能力が向上したので、訓練データに対して**100%**の精度を出す複雑なモデルを簡単に作れるようになった
- 過学習が起きやすい
- 特徴量選択によって過学習を防げるのではないか

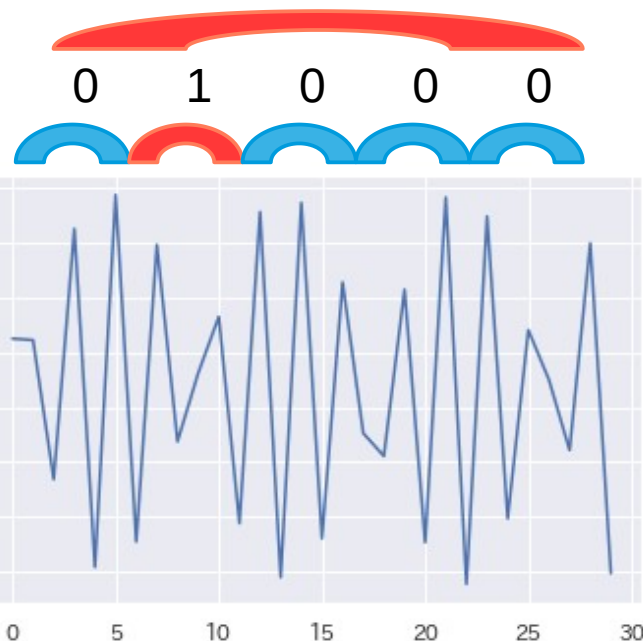
Convolutional Neural Network

- CNN: 画像認識でよく使われる
- 仮説
 - 正常ファイルは正常波形からなる
 - 異常ファイルには異常波形が含まれるが正常波形もある
- 狭い範囲(10サンプル)のコンボリユーションを取る
- 次の層でmax_poolingを取る
- 異常波形に対応するコンボリユーションだけ1を出力するはず
- このモデルは自信があったが異常の3分の1しか検出できない

label 正常=0 異常=1 1

max_pool

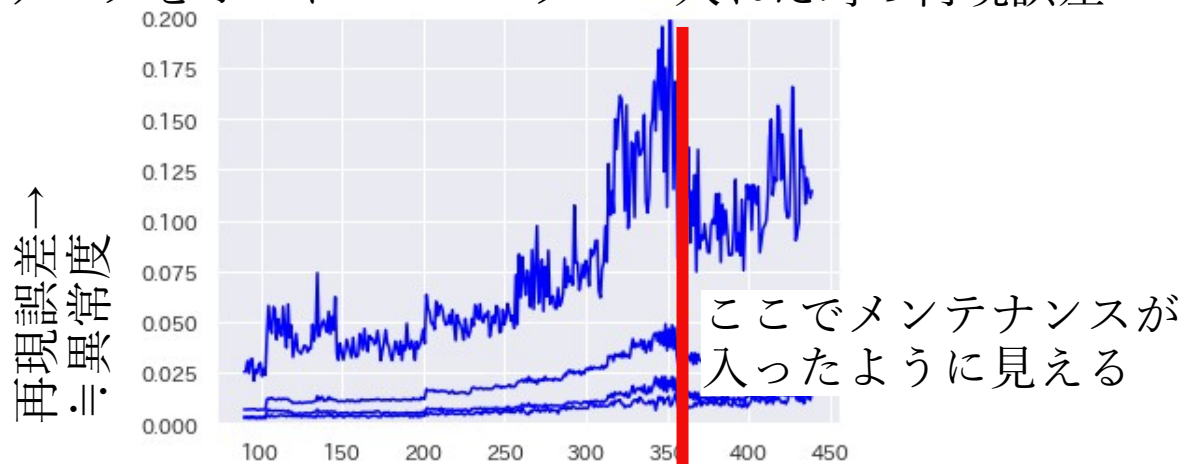
conv1d



分類問題なのか？

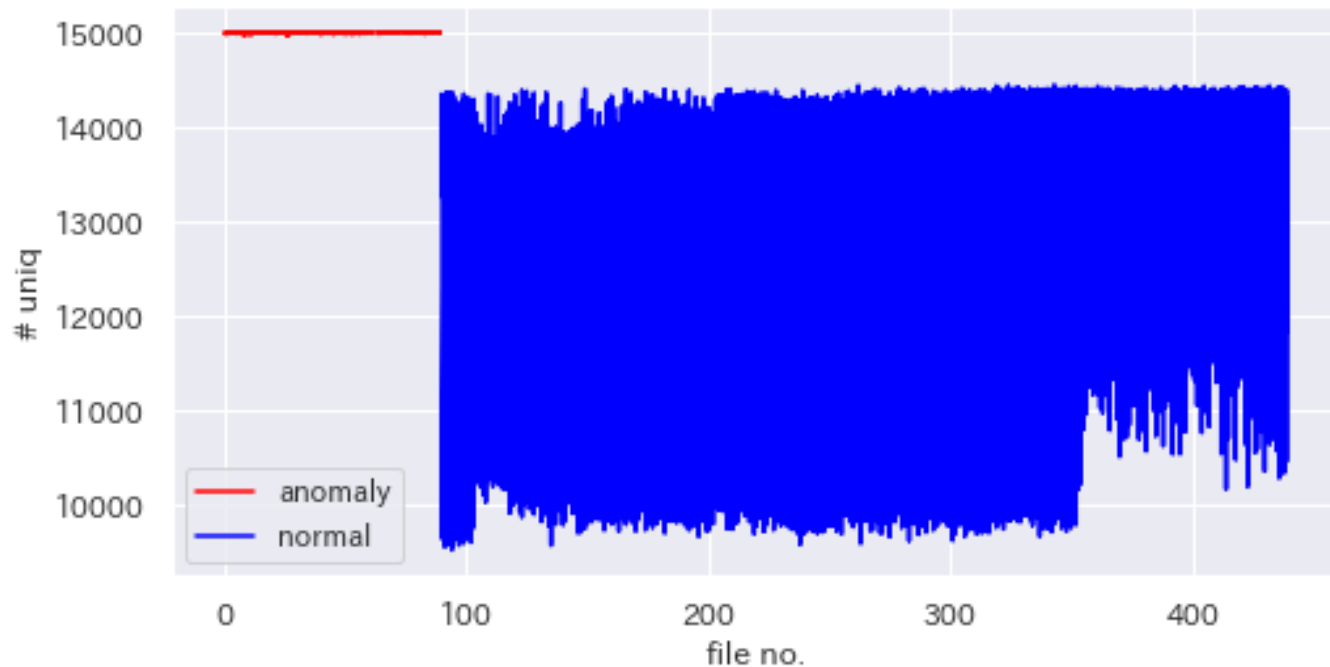
- 今回「正常と異常を見分ける」という課題だが分類アルゴリズムを使うのが正解なのか？
- 正常時でも経時劣化が進行している異常度を推測する回帰問題と見るべきでは

正常データをオートエンコーダーへ入れた時の再現誤差



結局は

- 特徴量検出ライブラリ tsfresh が出力する uniq

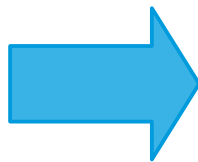


学んだこと

- まずベストプラクティスから始める
 - 広く使われているライブラリにはたくさんの人のノウハウが詰まっている
- 上手くいかなかったら自己流

いままで持っていた知識

chainer
PyTorch



AI Questで身につけた知識

Pandas
sklearn
tsfresh
stattools