

時系列データの解析

Background

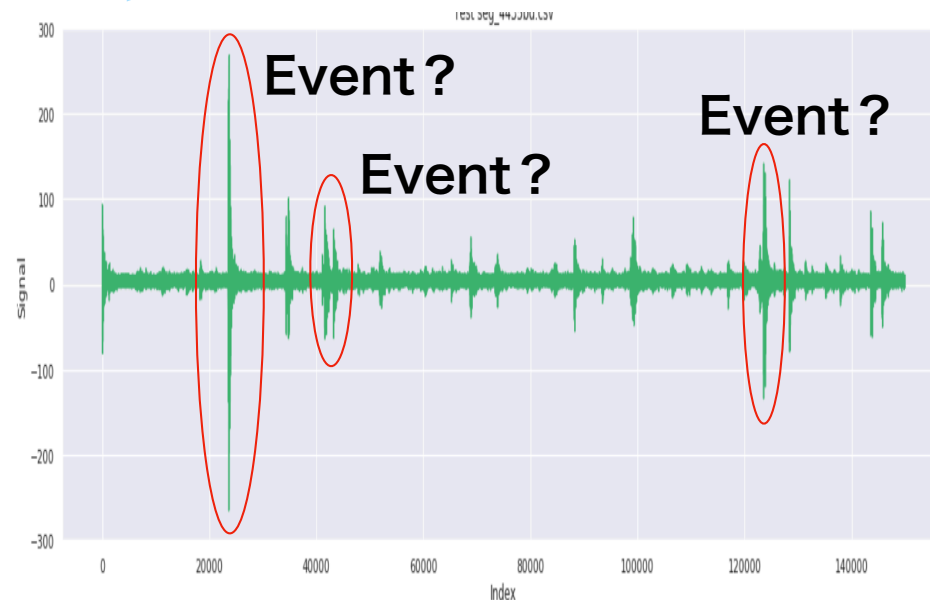
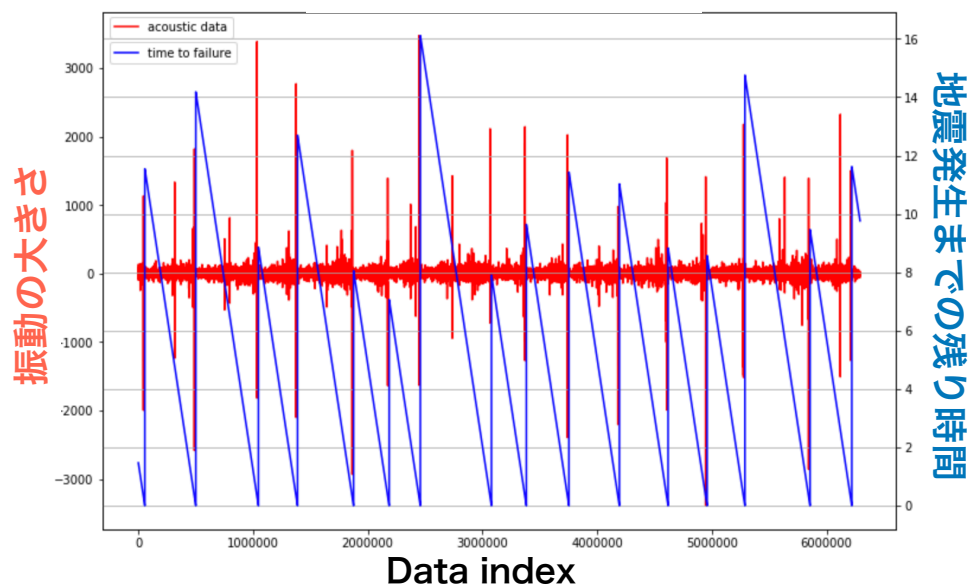
Kaggle LANL Earthquake predictionにチャレンジ！

振動データから地震イベントを検出し発生時刻を予想する

Training Data

学習

Prediction

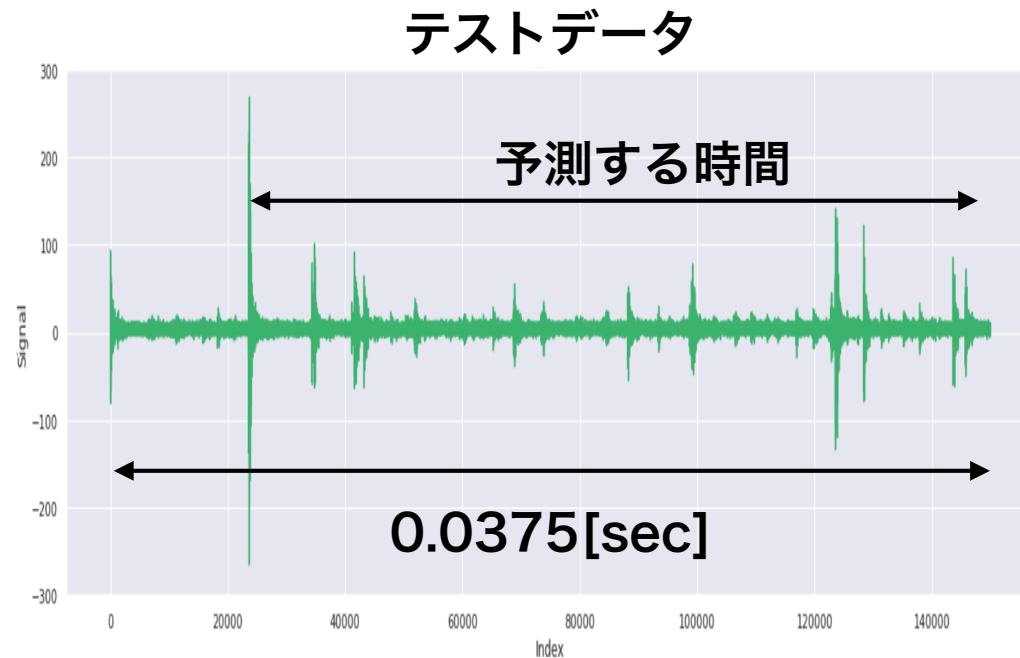
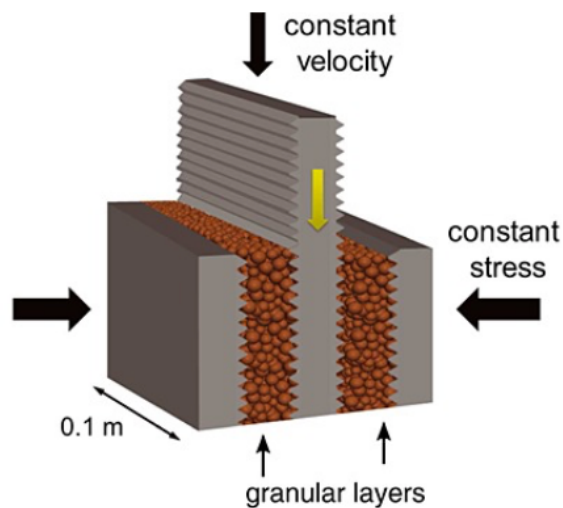


Goal

振動データから地震発生時刻を予測する！

Requirement

- 振動データから地震発生時間を予測する
 1. 予測する時間はTestデータの最後尾からイベントまでの時間
 2. Testデータは地震イベント前後の0.0375[sec]間のデータ



Data

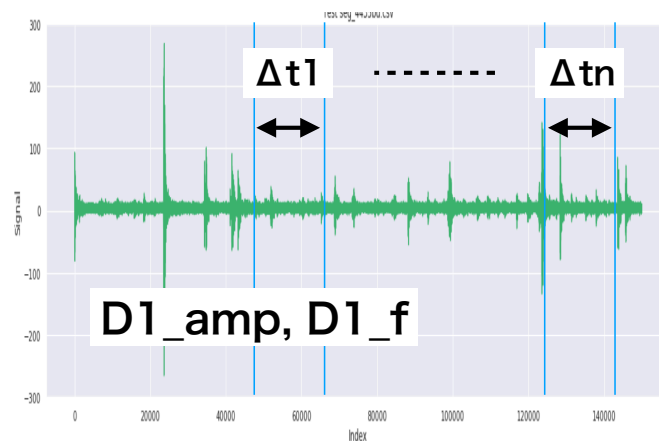
- Kaggle competitionページからDL

	Train data	Test data
File数 [file]	1	2496
Data数 [1/file]	$6.8 \times E8$	150000
Event数[1/file]	16	1
サンプリング周期[MHz]	4	4

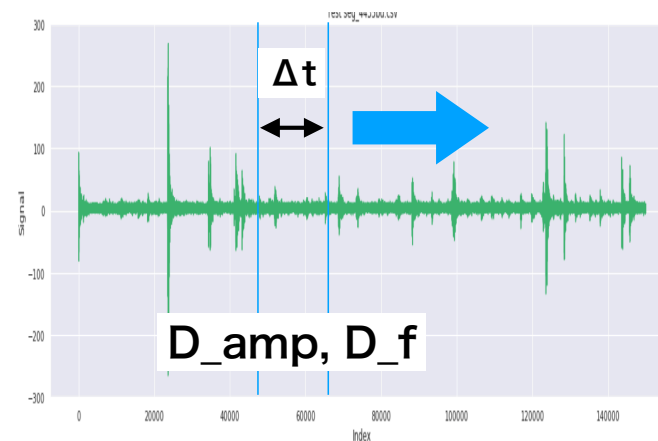
- EDA

振幅分布、周波数分布にイベント特有のパターンが現れるはず…

=>時間方向に領域を分けるorウィンドウをずらして各領域で振幅、周波数分布を比較



OR

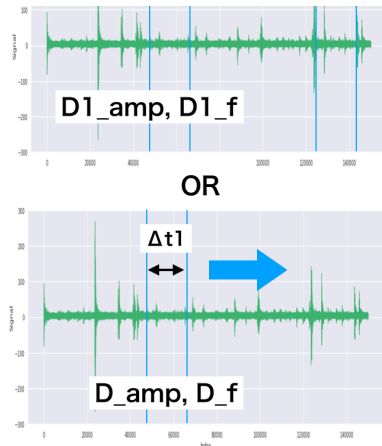


Method

- 古典的手法
 - 外れ値検出(ホテリングのT2法など) 今回はこれをやる！
- 機械学習
 - 抽出した特徴量から回帰分析を行う
- DNN
 - 時系列データなのでRNNで予測ができるはず…

Method

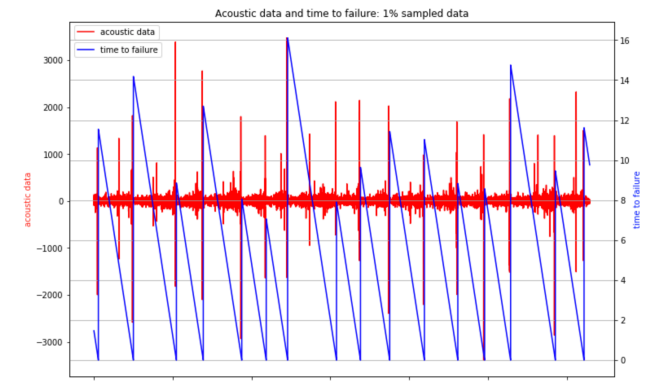
周波数,振幅分布
を特徴量としてインプット



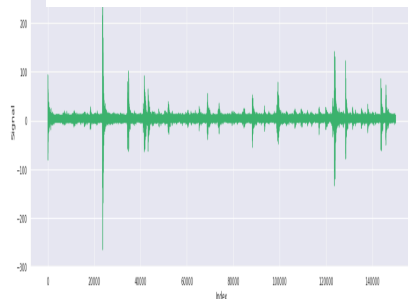
入力した特徴量からイベント発生までの
時間を回帰で求める

機械学習モデル

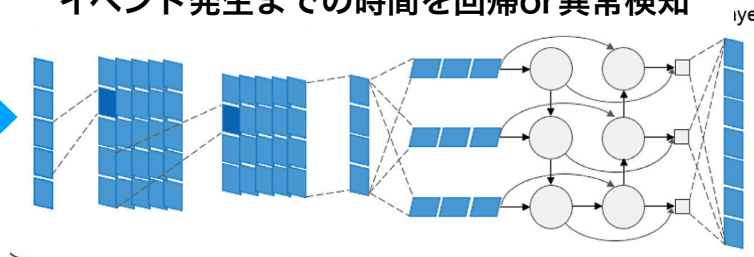
赤：振動データ
青：イベント発生までの時間（予測）



振動データをインプット



DNNで特徴量を抽出 &
イベント発生までの時間を回帰or異常検知



LSTMを用いたモデルを構築（調査中）

Evaluation

- 平均絶対誤差で評価を行う

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

y_j : 予測値
 \bar{y} : 正解

目標は MAE < 1.321 !!

現在のTop Score