Forecasting Laboratory Earthquakes using Machine Learning

Sukriti Singh[1]

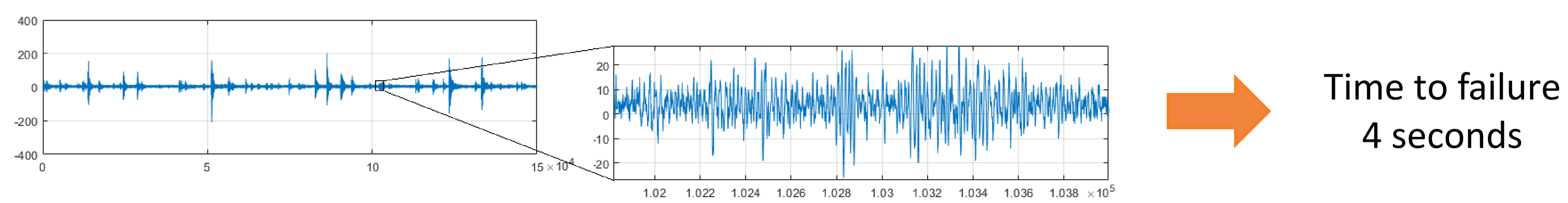
[1] Princeton University, Department of Computer Science, Princeton, NJ

ABSTRACT

- Los Alamos National Laboratory (LANL) has developed an apparatus to study earthquake physics in a laboratory setup. LANL has hosted a Kaggle challenge [1] and provided a dataset with real-time seismic data with approximate time to the next lab earthquake
- We benchmark various machine learning algorithms to predict the time to next quake from 1D acoustic signals.
- Instead of standard statistical features used in literature, we use features derived from wavelet decomposition of the raw signal. Using controlled experiments, we show that wavelet based features provide performance boost by ~3%.

BACKGROUND AND RELATED WORK

LANL's initial work [2] showed that the prediction of laboratory earthquakes from continuous seismic data is possible in the case of quasi-periodic laboratory seismic cycles. [2] use random forest regressor to predict the duration to the next failure from ~100 statistical features computed on the raw acoustic signals over time windows

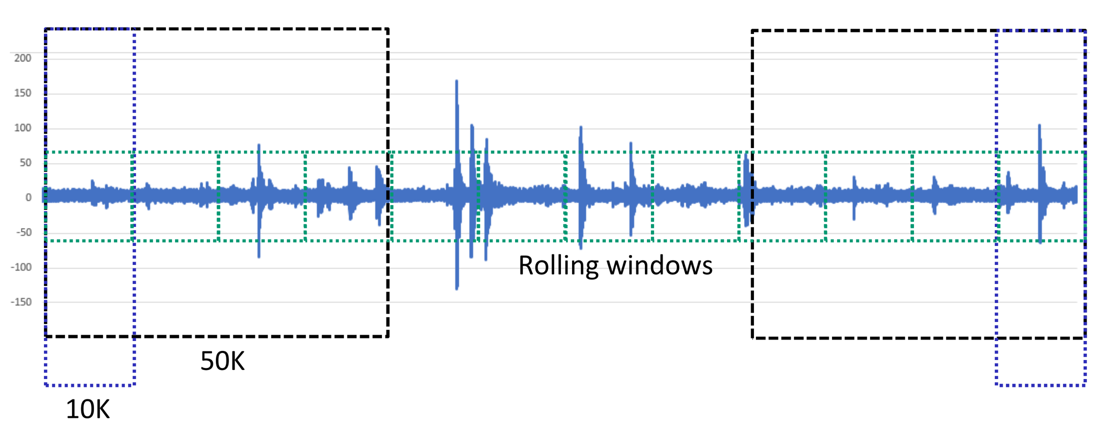


Challenge: Can we predict the labquakes which are considerably aperiodic, which are closer to the real world?

Existing machine learning pipeline – Extract features over the time window of 150000 signal length, use regression methods (Random Forests [2], Gradient boosting (Kaggle)) for prediction

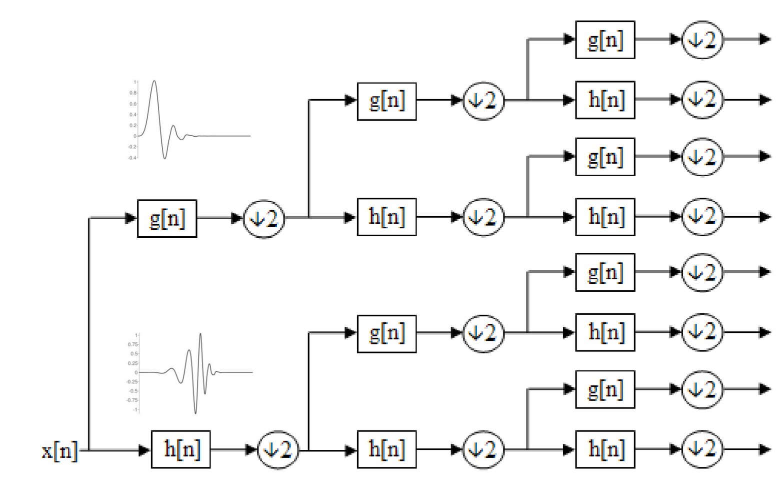
Feature Extraction

- Global statistical features (min, max, median, mean, g/h mean, …) computed using Python tsfresh library
- Statistics over window segments (first and last 1K, 10K, 50K, … measurements) and rolling time window
- Fast Fourier Transform based features



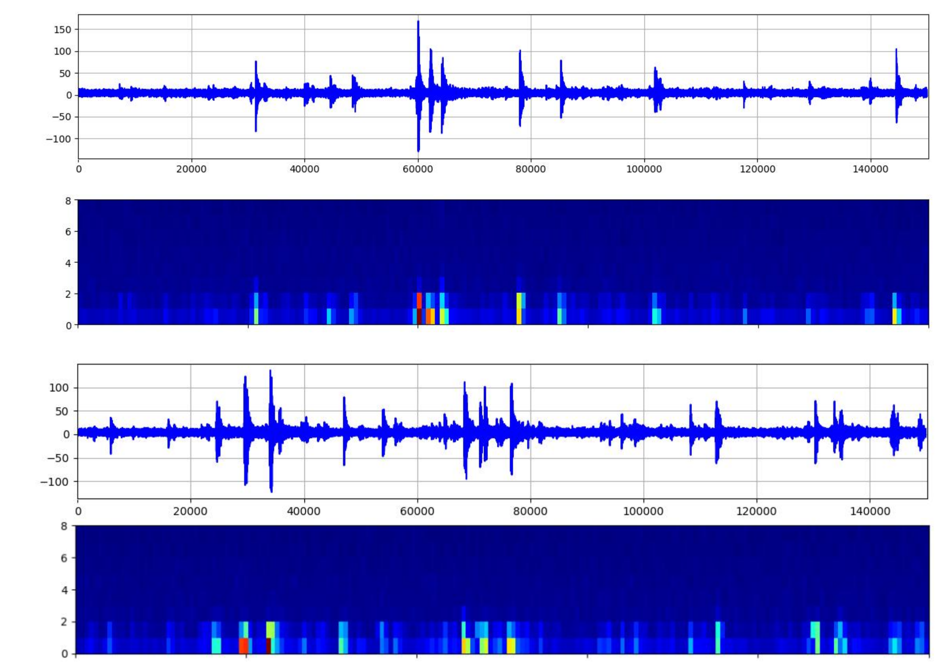
WAVELET BASED FEATURE EXTRACTION

Step 1: Wavelet Packet decomposition

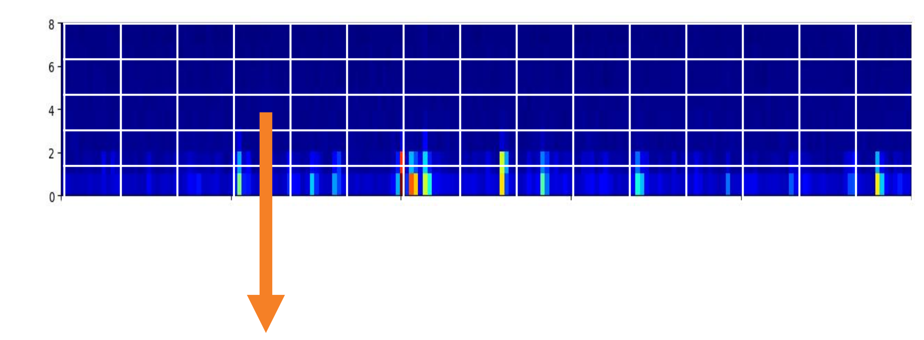


g[n] is the low-pass approximation coeffs.
h[n] is the high-pass detail coeffs.

wavelet coefficients



Step 2: Statistics over wavelet coefficients



min, max, median, mean, number of peaks, …

EXPERIMENTS

Dataset Description

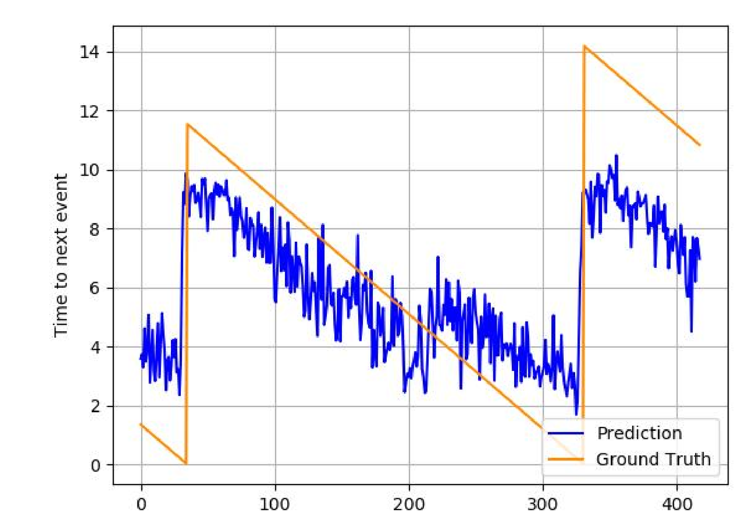
- Available as one long seismic signal with time to next quake at each instant. Test sequences – 150K time steps, time to next quake is not released.
- Dataset generated by sampling windows of size 150000 at intervals of 50000
- Total 12574 samples. Use first 90% for training (11320) and rest for testing (1254)

Regression Analysis

Method	*MAE*	*80%*	*90%*
FFT (1419 feats) + RF	4.2	6.78	9.07
Wavelets (280 feats) + RF	**4.14**	**6.68**	**8.82**
Wavelets + XGB	4.21	6.84	9.1
Wavelets + Linear SVR	4.31	6.91	9.37

Observations

- Regression fails to predict distant or immediate events (see figure on right)
- Statistical normalization did not improve performance



ACKNOWLEDGEMENTS

We thank Los Alamos National Lab (https://www.lanl.gov/) for providing the dataset for scientific study of lab earthquakes

REFERENCES

1. LANL Earthquake Prediction. https://www.kaggle.com/c/LANL-Earthquake-Prediction/overview/. Accessed: 2019-04-23.
2. Bertrand Rouet-Leduc, Claudia Hulbert, Nicholas Lubbers, Kipton Barros, Colin J.Humphreys, and Paul A. Johnson. Machine learning predicts laboratory earthquakes. Geo-physical Research Letters, 44(18):9276–9282, 2017

sukritis@princeton.edu