Applying Machine Learning to Crowd-sourced Data from Earthquake Detective

Omkar Ranadive
Department of Computer Science
Northwestern University
omkar.ranadive@u.northwestern.edu

Suzan van der Lee
Department of Earth and Planetary Sciences
Northwestern University
suzan@northwestern.edu

Vivian Tang
Department of Earth and Planetary Sciences
Northwestern University
vivian@earth.northwestern.edu

Kevin Chao
Data Science Initiative/NICO
Northwestern University
kchao@northwestern.edu

Abstract

We present the Earthquake Detective dataset - A crowdsourced set of labels on potentially triggered (PT) earthquakes and tremors. These events are those which may have been triggered by large magnitude and often distant earthquakes. We apply Machine Learning to classify these PT seismic events and explore the challenges faced in segregating such low amplitude signals. The data set and code are available online.[1][2]

1 Introduction

While the vast majority of earthquakes do not directly impact people, some earthquakes cause injuries, damage, or worse. Both types of earthquakes are part of the same statistical Pareto distribution, called the Gutenberg-Richter law (Gutenberg and Richter, 1954). This empirical law states that for every magnitude (M) 8 earthquake that occurs, 100 M6, 10,000 M4, and 1 million M2 earthquakes occur. Therefore, the abundance of low-magnitude earthquakes of the kind whose signals are often buried in background seismic noise, provides insights into the occurrence of the much more infrequent, damaging earthquakes.

Likewise, a class of so-called low-frequency earthquakes exists that represent slower slip between two blocks of rock than that during classical earthquakes, but nevertheless generates weak seismic signals that are often labeled as "tremor" (Obara, 2002), especially when many of such events occur quasi-simultaneously.

Both types of weak signals (from tremor and from low-magnitude local earthquakes) are both abundant (Rouet-Leduc et al., 2018; Hill and Prejean, 2015) and somewhat under-reported because they have been hard to detect. Past barriers to detection have included a sparse spatial distribution of

[1]https://www.zooniverse.org/projects/vivitang/earthquake-detective
[2]https://github.com/Omkar-Ranadive/Earthquake-Detective

AI for Earth Sciences Workshop at NeurIPS 2020.

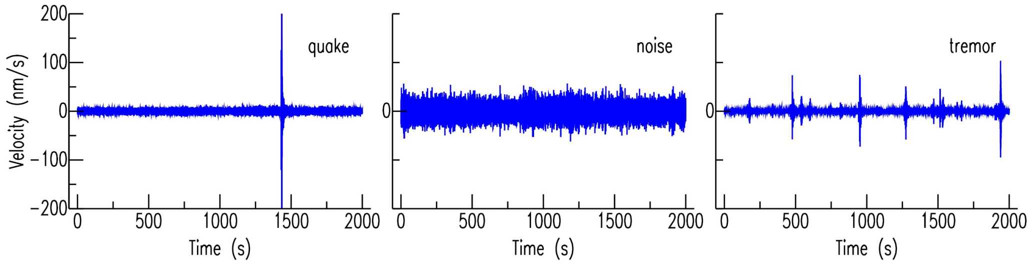

Figure 1: Example data plots shown to users on Earthquake Detective platform

seismic stations (instrumented with buried seismometers), and current barriers include the weakness of the seismic signals.

Our interests lie in detecting a special sub-class of these seismic events, namely local earthquakes and tremor that could have been triggered by slowly-oscillating large-amplitude seismic surface waves from large-magnitude and often distant earthquakes. Reporting and learning more about such Potentially Triggered (PT) events extends the spectrum of earthquake data available for study and adds information about how earthquakes might nucleate. In our quest to detect PT seismic events we face a number of additional challenges:

- The magnitudes (M) of PT earthquakes are typically below the M of completeness of earthquake catalogs, hence their signals are weak and often buried in ambient seismic noise signals, and there is no large database of template waveforms available for such low-M events. Signals from low-M events are not only lower in amplitude than those from higher-M events but also have narrower bandwidths.
- Even after high- or band-pass filtering, signals from PT tremor have different waveforms than those from typical tremor on account of the former signals being modulated by the surface waves (Chao et al., 2012). Therefore, a large database of template signals is not available for training or other purposes, although a catalog has been started (Kano et al., 2018).
- A signal from a PT event can arrive at any time during the surface wave signal, which is much longer in duration than the former signal. We have been considering up to 33 minutes of surface wave duration in labeling whether or not at least one local earthquake or tremor signal was recorded.
- Non-stationary noise signals are often comparable in amplitude and duration to the relatively weak signals of the PT events we are interested in.
- Optimal and accurate detection of signals from PT events requires a multi-scale, multi-band, multi-component interactive analysis that is labor-intensive. The formation of a labeled data set for training purposes is hence not straightforward.

2 Related Work

Using Machine Learning (ML) in the detection of signals from seismic activity is a relatively new and rapidly growing field. Several ML algorithms have successfully been trained to detect signals from local earthquakes (e.g. Ross et al. (2018), Tang et al. (2020b)) and Liu et al. (2019) used ML to detect tremor signals.

3 Earthquake Detective

Earthquake Detective is a crowdsourcing platform where volunteers are shown a graph of vertical ground velocity vs. time (Figure 1) along with a sonification of the signal, and are asked to classify the data as Earthquake/Tremor/Noise/None of the above. The platform currently has over 6000+ volunteers and 130k+ classifications. Tang et al. (2020a) present an analysis of how the volunteers and seismologists engage with the data.

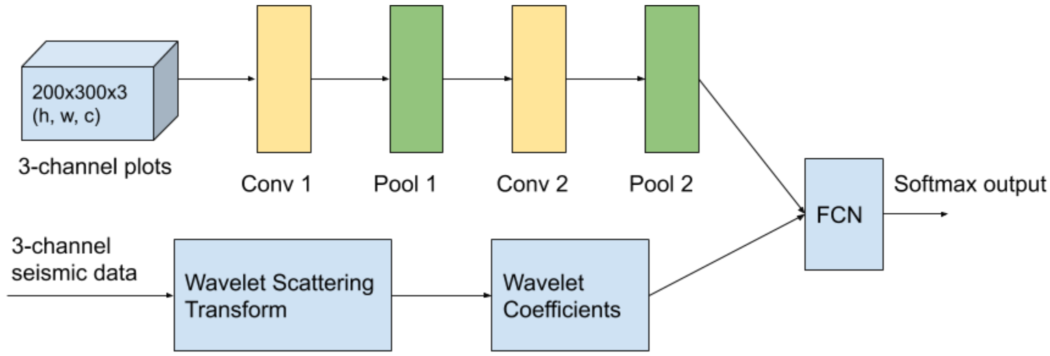

Figure 2: WavImg Model Architecture

All raw input data are time series of waveforms with durations of 2000 s, which are long enough to contain the time window needed by teleseismic surface waves to pass through. The raw data is demeaned, band-pass filtered between 2-8 Hz with a 2-pole Butterworth filter, deconvolved with the instrument response to convert digital counts to physical units of ground velocity, and resampled at 20 samples per second.

4 Approach

4.1 Wavelet Scattering Transform

The wavelet scattering transform decomposes a signal using a family of wavelets. This new representation is stable against deformations and is translation invariant. This family of wavelets does not need to be learned and hence, the features can be extracted without training which can then be passed on further to Machine Learning models (Oyallon et al., 2013). The Scattering Transform has been successfully used by Seydoux et al. (2020) for clustering earthquakes in an unsupervised fashion. The scattering transform works by successively convolving wavelets with the signal and applying modulus non-linearity at each step. This can be shown as follows:

$$S_{x(t)} = ||x * \psi_{\lambda 1}| * \psi_{\lambda 2}|...| * \psi_{\lambda m}| * \phi \tag{1}$$

where $S_{x(t)}$ is the set of scattering coefficients obtained at step m, $x(t)$ is the signal, $\psi_{\lambda_m}(t)$ denotes the set of wavelets at step m and $\phi(t)$ is a low-pass filter.

In our experiments, we use the Kymatio library(Andreux et al., 2020) to perform the scattering transform and get a set of features which are passed on further to our supervised neural network model.

4.2 Convolution over 3-channel plots

As the volunteers classify the data based on plots and associated audio records of vertical ground motion, we decided to try out a similar approach with our model. Along with the wavelet coefficients, we additionally provide the model with 3-channel plots (BHZ, BHE, BHN) as input. Image convolution is applied over the 3-channel plots and the resultant features are concatenated with the wavelet coefficients and then passed through a fully connected neural network (FCN).

5 Experiments

For all experiments, we use the following hyper parameters: learning rate = 1e-5, batch size = 100, epochs = 300. We perform a 3-way classification between Earthquakes/Tremors/Noise. As the number of tremor samples is considerably less, we apply a weighted cross-entropy loss where the weights are calculated as follows:

$$w_i = N_{largest}/N_i \tag{2}$$

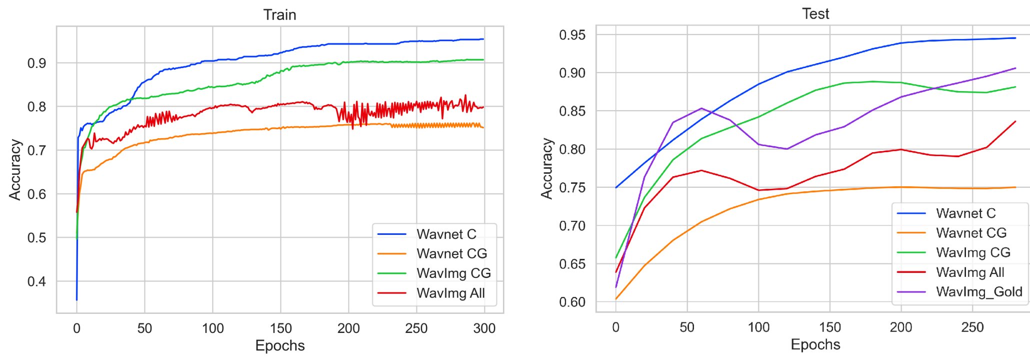

Figure 3: Model comparison chart. (C = Clean, CG = Clean+Gold, All = Clean+Gold+Volunteer)

where w_i is the weight assigned to class i, $N_{largest}$ is the number of samples of the largest class and N_i is the number of samples of class i. For all experiments, we perform a 80-20 stratified train/test split of the data.

We test two models:

- **WavNet:** In this model, we perform a wavelet scattering transform on 3-channel seismic data (BHZ, BHN, BHE) and extract relevant features from it. These features (wavelet coefficients) are then passed on to a 2-layer fully connected network.
- **WavImg:** This model combines the Wavelet Scattering Transform with 3-channel convolution over the image plots. The combined features are then passed on through a 2-layer fully connected network. (Figure 2)

5.1 Training the Machine Learning model

Training on clean data: To test the efficacy of wavelet transform, we first run a simple experiment with WavNet. As an upper baseline, we first ran the experiment on clean data (data cleaned and filtered by our seismologists, refer Appendix A). The model converges to 95.2% training and a 94.4% testing accuracy.

Training on clean + gold users data: Compared to the clean data, the data from the Earthquake Detective is difficult to segregate due to its low amplitude signals and larger time window ($\sim$33 mins). For this experiment, we consider data from gold users (Earthquake Detective data labeled by our experts) and combine it with the clean dataset from the previous experiment. When this data was trained on the WavNet, it gave a train accuracy of 75.4% and a test accuracy of 74.9%. Next, we trained the same data using the WavImg model. This model produces a 91.4% train and 89.6% test accuracy.

Training on clean + gold + volunteer's data: Finally, we combine the previous data with data from two volunteers. For each volunteer, we calculated a reliability score which includes a precision, recall and f1 score for each class. These scores were calculated by comparing volunteer's classification with gold-set labeled by our experts. To handle the unreliability introduced in labels, we add an additional gold-test set which consists of samples labeled from our gold users that were not used for training. WavImg produces a 80.1% train accuracy, 83.6% test accuracy and **90.4%** gold-test accuracy.

5.2 Results and Analysis

The WavNet model was able to perform extremely well on the clean data (95% accuracy) which proves that wavelet scattering transform extracts relevant features which can then be trained using a simple 2-layer FCN. However, due to the greater complexity of Earthquake Detective data, WavNet by itself is insufficient. The WavImg model overcomes this problem by using information from 3-channel plots. One interesting case was the last experiment in which there was variance in training due to label uncertainty (80% accuracy) but the model still performed well on the gold-test set (90.4% accuracy). (Figure 3) This shows that the model still ends up learning useful representations despite the uncertainty of the labels. For more in-depth analysis over select misclassified samples refer to Appendix B.

6 Conclusion

The Earthquake Detective dataset is the first crowdsourced dataset for PT local earthquake and tremor signals. This is also the first time that PT tremor signals were used in and detected by ML. Our experiments provide ML baselines for the data. However, we trained our models on a small subset of the 130k+ samples available. Therefore, when all of the data is considered, better techniques to incorporate reliability scores into the model will be required. We hope that this dataset will be used by researchers as a catalog for potentially triggered seismic events.

Broader Impact

This work might facilitate work by research seismologists, seismic network operators, and seismic monitoring agencies and ultimately result in reducing risks posed by damaging earthquakes. Other impacts include the engagement of volunteer scientists, via Earthquake Detective, with informal STEM learning and analyzing authentic data.

Acknowledgments and Disclosure of Funding

This research was funded by the Integrated Data-Driven Discovery in Earth and Astrophysical Sciences (ID^{3}EAS) program under National Science Foundation grant NSF-NRT 1450006. We thank Josiah Evans and Blaine Rothrock for helpful discussions.

References

Mathieu Andreux, Tomás Angles, Georgios Exarchakis, Roberto Leonarduzzi, Gaspar Rochette, Louis Thiry, John Zarka, Stéphane Mallat, Joakim Andén, Eugene Belilovsky, et al. Kymatio: Scattering transforms in python. *Journal of Machine Learning Research*, 21(60):1–6, 2020.

K. Chao, Z. Peng, C. Wu, C.C. Tang, and C.H. Lin. Remote triggering of non-volcanic tremor around taiwan. *Geophys. J. Int.*, 188:301– 324, 2012. doi: doi:10.1111/j.1365âĂŘ246X.2011.05261.x.

B Gutenberg and CF Richter. Seismicity of the earth 2nd ed., 310, 1954.

David P. Hill and Stephanie Prejean. Dynamic triggering. *Treatise on Geophysics*, 4, Earthquake Seismology:273–304, 2015. doi: doi.org/10.1016/B978-0-444-53802-4.00078-6.

Masayuki Kano, Naofumi Aso, Takanori Matsuzawa, Satoshi Ide, Satoshi Annoura, Ryuta Arai, Satoru Baba, Michael Bostock, Kevin Chao, Kosuke Heki, Satoshi Itaba, Yoshihiro Ito, Noriko Kamaya, Takuto Maeda, Julie Maury, Mamoru Nakamura, Takuya Nishimura, Koichiro Obana, Kazuaki Ohta, Natalia Poiata, Baptiste Rousset, Hiroko Sugioka, Ryota Takagi, Tsutomu Takahashi, Akiko Takeo, Yoko Tu, Naoki Uchida, Yusuke Yamashita, and Kazushige Obara. Development of a Slow Earthquake Database. *Seismological Research Letters*, 89(4):1566–1575, 06 2018. ISSN 0895-0695. doi: 10.1785/0220180021. URL `https://doi.org/10.1785/0220180021`.

Yi-Hung Liu, Ting-Chen Yeh, Kate Huihsuan Chen, Yaochieh Chen, Yuan-Yi Yen, and Horng-Yuan Yen. Investigation of single-station classification for short tectonic tremor in taiwan. *Journal of Geophysical Research: Solid Earth*, 124(8):8803–8822, 2019.

K. Obara. Nonvolcanic deep tremor associated with subduction in southwest japan. *Science*, 296: 1679–1681, 2002. doi: 10.1126/science.1070378.

Edouard Oyallon, Stéphane Mallat, and Laurent Sifre. Generic deep networks with wavelet scattering. *arXiv preprint arXiv:1312.5940*, 2013.

Zachary E Ross, Men-Andrin Meier, Egill Hauksson, and Thomas H Heaton. Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, 108(5A): 2894–2901, 2018.

Bertrand Rouet-Leduc, Claudia Hulbert, and Paul A Johnson. Breaking cascadia's silence: Machine learning reveals the constant chatter of the megathrust. *arXiv preprint arXiv:1805.06689*, 2018.

Léonard Seydoux, Randall Balestriero, Piero Poli, Maarten De Hoop, Michel Campillo, and Richard Baraniuk. Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nature communications*, 11(1):1–12, 2020.

Vivian Tang, Boris Rösler, Jordan Nelson, JaCoya Thompson, Suzan van der Lee, Kevin Chao, and Michelle Paulsen. Citizen scientists help detect and classify dynamically triggered seismic activity in alaska. *Frontiers in Earth Science*, 8:321, 2020a.

Vivian Tang, Prem Seetharaman, Kevin Chao, Bryan A Pardo, and Suzan van der Lee. Automating the detection of dynamically triggered earthquakes via a deep metric learning algorithm. *Seismological Research Letters*, 91(2A):901–912, 2020b.

Appendix A: Data

Clean Data

The clean data used in the experiments was created as follows:

1. A set of 1000 s-long waveforms with confirmed PT tremor signals, labeled by the seismologists among us. The waveforms were resampled at 20 samples per second and band-pass filtered between 2 and 5 Hz. This data set is subdivided as follows:
   (a) Waveforms recorded by seismic stations in Taiwan or Japan, trimmed around surface waves from 6 large-M earthquakes in the eastern hemisphere (the Great Tohoku Earthquake, 1 more earthquake from Hokkaido, Japan, 1 from Qinghai, China, and 4 from Sumatra, Indonesia). These are the positive examples.
   (b) Waveforms recorded by seismic stations in Taiwan, Malaysia, Australia, and of the Global Seismographic Network (GSN), selected for having no significant signals from earthquakes or otherwise. These are the negative examples.
2. A set of 1000 s-long waveforms with confirmed PT local earthquake signals, labeled by the seismologists among us. All data was band-pass filtered between 2 and 8 Hz.
   (a) Waveforms recorded by seismic stations from USArray in the USA and the Hi-CLIMB array in Tibet, trimmed around surface waves from the 2010 M8.8 Maule Earthquake and 7 additional large-M earthquakes in the eastern hemisphere (1 from China, 1 from Japan, and 5 from Sumatra, Indonesia). Waveforms that showed signals from local earthquakes were labeled as positive.
   (b) The waveforms from this set (2a) that were not labeled positive - they were labeled as negative examples.
   (c) Additional waveforms recorded by the Hi-CLIMB array from 10 random local earthquakes with M<3.6. This auxiliary data set was used to expand the set of positive examples. An corresponding number of negative examples recorded by the same array was added to the negative examples.

Data Distribution

- In the first experiment we used 551 earthquake samples, 570 noise samples and 39 tremor samples from the clean data.
- In the second experiment (clean + gold users data), we had the following data distribution: 1031 earthquake samples, 1014 noise samples and 48 tremor samples.
- In the final experiment (which additionally includes the chosen volunteer's data) had the following data distribution: 3013 earthquake samples, 2436 noise samples, 203 tremor samples.

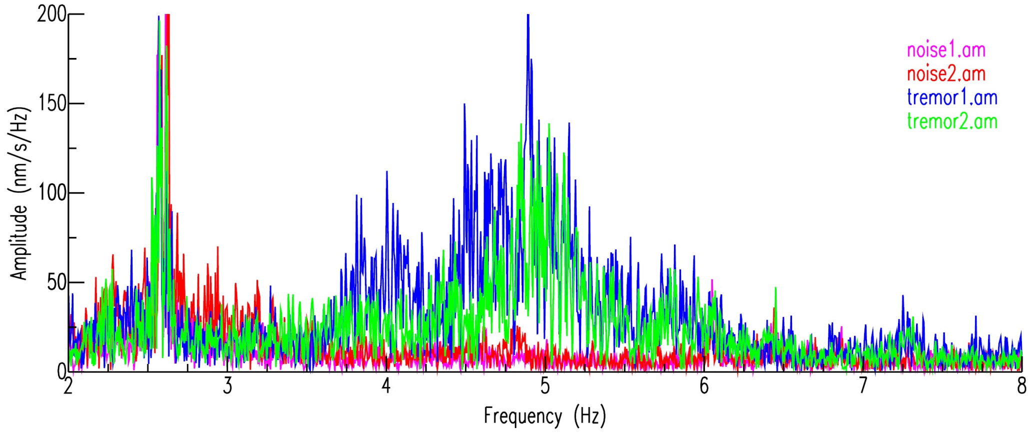

Figure 4: Amplitude spectra for four 200-s subsets of a 2000-s wave train labeled as tremor and misclassified as noise. Two subsets (red and magenta) represent noise and two (blue and green) contain a tremor-like signal.

Appendix B: Model Analysis

To gain further insights into how the WavImg model classifies, we looked at select examples. One example is where the ML algorithm classified a wave train with tremor as a noise wave train. This could be the result of

- There being two separate bursts of tremor,
- Both tremor signals being relatively short in duration,
- The tremor signals being weak,
- The tremor signals sounding different from more typical tremor signals
- The presence of non-stationary noise signals.

Extracting two 200-s tremor signals and two 200-s stationary noise signals from the 2000-s wave train reveals that the the weak tremor signals have some additional power between 3 and 6 Hz, compared to the noise (Figure 4). This is not entirely characteristic but still consistent for tremor signals. Although the wave train was likely labeled correctly, the wave train is not a role model for its class and hence may have confused the ML algorithm. In one other case labeled as tremor and classified as earthquake, the tremor signal was so brief that is easy to mistake for an earthquake signal. In another case, strongly peaked signals elsewhere in the wave train might have distracted the ML from the tremor signal.

The type of waveform data used in our study contains a wild variety of noise signals, for which we did not designate a single class. However such noise signals can interfere with the ability of volunteers, and sometimes experts to correctly label wave trains. In at least 5 cases, wave trains with noise signals, labeled as noise, were misclassified as earthquakes.

Several other cases of misclassification by the ML algorithm can be traced to a mislabeling of the original data. In three cases of wave trains with noise signals mislabeled as tremor, the ML algorithm classified the wave trains as earthquakes. The ML algorithm also classified a case of mislabeled tremor correctly as noise. In at least 4 cases, the ML algorithm correctly classified wave trains as noise, while they were labeled as earthquakes. In these four cases, listening to and viewing spectral properties of the wave trains confirmed in hindsight that these signals should have indeed been labeled as noise.