NTGE 2018

NEW TRENDS IN GEOPHYSICS AND ENGINEERING

International Symposium

JEOSTATİK

İnşaat - Maden - Zemin Etüt Grubu Ltd. Şti.

Identification of Pulse Shape Signals on Near Fault Stations With Convolutional Neural Network Algorithms: Preliminary Results

Deniz Ertuncay, Giovanni Costa

University of Trieste, Department of Mathematics and Geoscience, Trieste, Italy

deniz.ertuncay@phd.units.it, costa@units.it

Introduction

Increasing number of seismic stations that are located near the active faults make it possible to investigate near-fault behavior of the seismic signals. Occasionally, these stations recorded earthquakes with peculiar seismic pattern at various major earthquakes. Such signals, called henceforward pulse-shape signals, can be seen at the early stages of the earthquake signal in velocity records.

There are various algorithms to detect pulse-shape signals in earthquake waveform. Mavroeidis and Papageorgiou (2003) proposed a wavelet analysis to construct a mathematical representation of the pulse, which depends on amplitude, period, duration and phase shift. Shahi and Baker (2014) used a 4th-order Daubechies wavelet to determine pulse-shape signals. The method has some constraints such as a minimum PGV amplitude, a pulse arrival located at the beginning of the signal and arbitrary thresholds for the energy function. Mena and Mai (2011) used windowed Fourier transform analysis for the pulse shape signal and its position with certain energy thresholds. Chang et al. (2016) used the energy function with certain thresholds to determine the pulse-shape signal position and period. Kardoutsou et al. (2017) used a cross-correlation between the potential pulse-shape signal and the wavelet functions to determine the pulse shape.

High amount of data in earth science and improvements on computers allowed earth scientist to use machine learning algorithms on their databases. Asencio-Cortés et al. (2018) used seismicity database of California to forecast earthquakes. DeVries et al. (2018) used deep learning methods to forecast aftershocks. Florido et al. (2018) implement several machine learning algorithms to discover seismic patterns. In Shahi and Baker (2011), generalized linear models (GLM) were used in combination with a model fitting for predicting the probability of near-fault earthquake ground motion pulses without using the time series data.

Data

Analyzed ground motions are selected from NGA-West2 (Ancheta et al., 2012), GeoNet, Itaca (Pacor et al., 2011; Luzi et al., 2016) and K-Net databases, which contain data from crustal earthquakes. Earthquake signals that are recorded due to $M_w \geqslant 5.5$ earthquakes with a maximum distance range of 115 km from the epicenter, are selected. In order to study pulse-shape signals, East and North components are rotated to radial and transverse components. In total, our database contains 2739 waveform recorded due to more than 100 earthquakes with various source mechanisms.

Waveform are used with the fixed length of 60 seconds. Records started where the P arrival is picked and ended 60 seconds later than the pick. If the signal has less duration, then zeros added at the end of the real signal.

Velocity waveform are used since it is easier to detect pulse shape signals on velocity records and spectrogram. Impulsive signals. Spectrogram are used since impulsive part of the waveform contains most of the seismic energy.

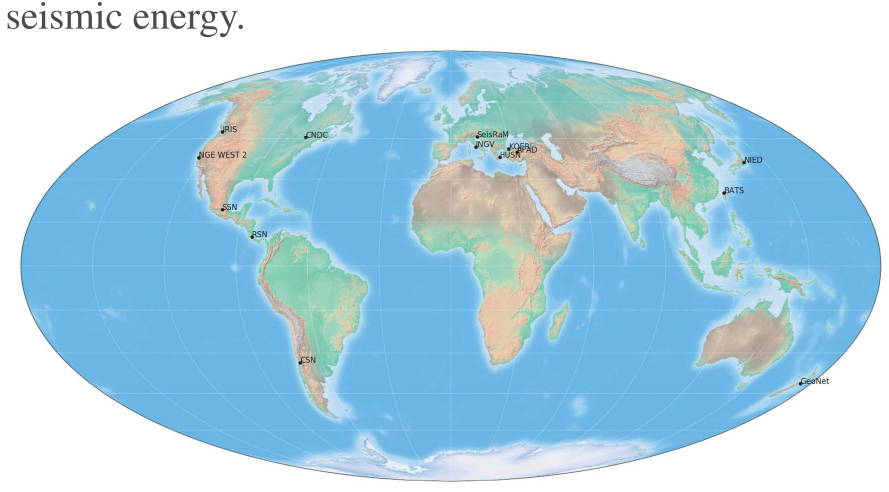

Figure 1: Seismological Networks & Databases That We Collect Data.

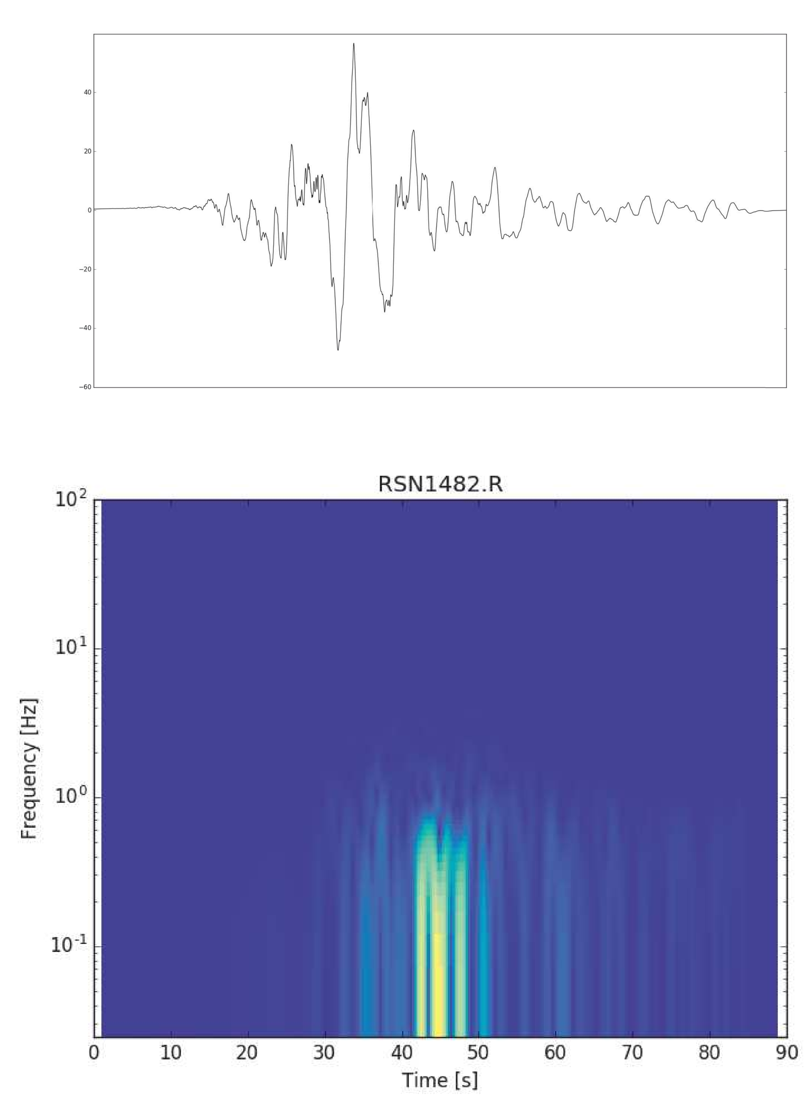

Figure 2: An example of velocity waveform and spectrogram of acceleration waveform of 20th of September 1999 M_w 7.6 at TCU039 station of Broadband Array in Taiwan for Seismology Institute of Earth Sciences (1996), respectively.

Method

Determination of the Pulse Shape Signal

The propagation of fault rupture toward a site at a velocity close to the shear wave velocity causes most of the seismic energy from the rupture to arrive in a single large pulse of motion that occurs at the beginning of the record. This pulse of motion represents the cumulative effect of almost all of the seismic radiation from the fault. The radiation pattern of the shear dislocation on the fault causes this large pulse of motion to be oriented in the direction perpendicular to the fault, causing the strike-normal ground motions to be larger than the strike-parallel ground motions.

Chang et al. (2016)'s Algorithm

Chang et al. (2016) used energy ratio between a part of the signal where PGV is occurred and whole velocity waveform. PGV should be bigger than 30 cm/s in order to consider the waveform as possible pulse shape signal producer. The pulse model is expressed as below:

$$v_p(t; A_p, T_p, N_c, T_{pk}, \theta) = A_p exp\left[-\frac{\pi^2}{4}\left(\frac{t-T_{pk}}{N_c T_p}\right)^2\right]\times\cos\left(2\pi\frac{t-T_{pk}}{T_p}-\theta\right) \quad (1)$$

where v_p is the extracted pulse, t time series, A_p means the amplitude of the extracted pulse, T_p is the pulse period, N_c is the number of the cycles in the pulse, T_{pk} means the location of A_p in the time axis, θ represents the phase of the pulse.

Occurrence of the pulse shape behavior of the signal is valid, if the energy ratio between the part of the signal where PGV occurs and the whole waveform exceeds the threshold level which is 0.3. Mathematical representation of this methodology is in below:

$$E_p = \frac{\int_{t_s}^{t_e} v^2(t)dx}{\int_0^{\infty} v^2(t)dx} \quad (2)$$

where t_s and t_e represent the pulse starting and ending points in the time domain, respectively.

Machine Learning Algorithms

Scikitlearn (Pedregosa et al., 2011) and Tensorflow (Abadi et al., 2016) packages are used in order to process the data. Methods are explained in Convolutional Neural Network (CNN) and work-flows can be seen in Figure 3.

Convolutional Neural Network (CNN)

In this study, both velocity waveform and spectrogram are processed with artificial neural networks. Architecture of the network has 4 layers. 1D and 2D convolutions are used for waveform and spectrogram, respectively. Each layer has 32 filters and the relu activation function. Architecture of the neural network can be seen in Figure 3.

Softmax cross entropy method is used to calculate the cost function. Then Adam optimizer is used for optimization. 50 batch are used in 2000 epochs with a learning rate of 0.001. Data split up to 30% percent of test and 70% of train the algorithm. 30% of the train data is reserved for validation. Data length of waveform is 6000 data points in 1D vector. On the other hand, spectrogram are hold in 2D matrices with $452x512$. Dimensions are related with time segments and frequency resolution, respectively.

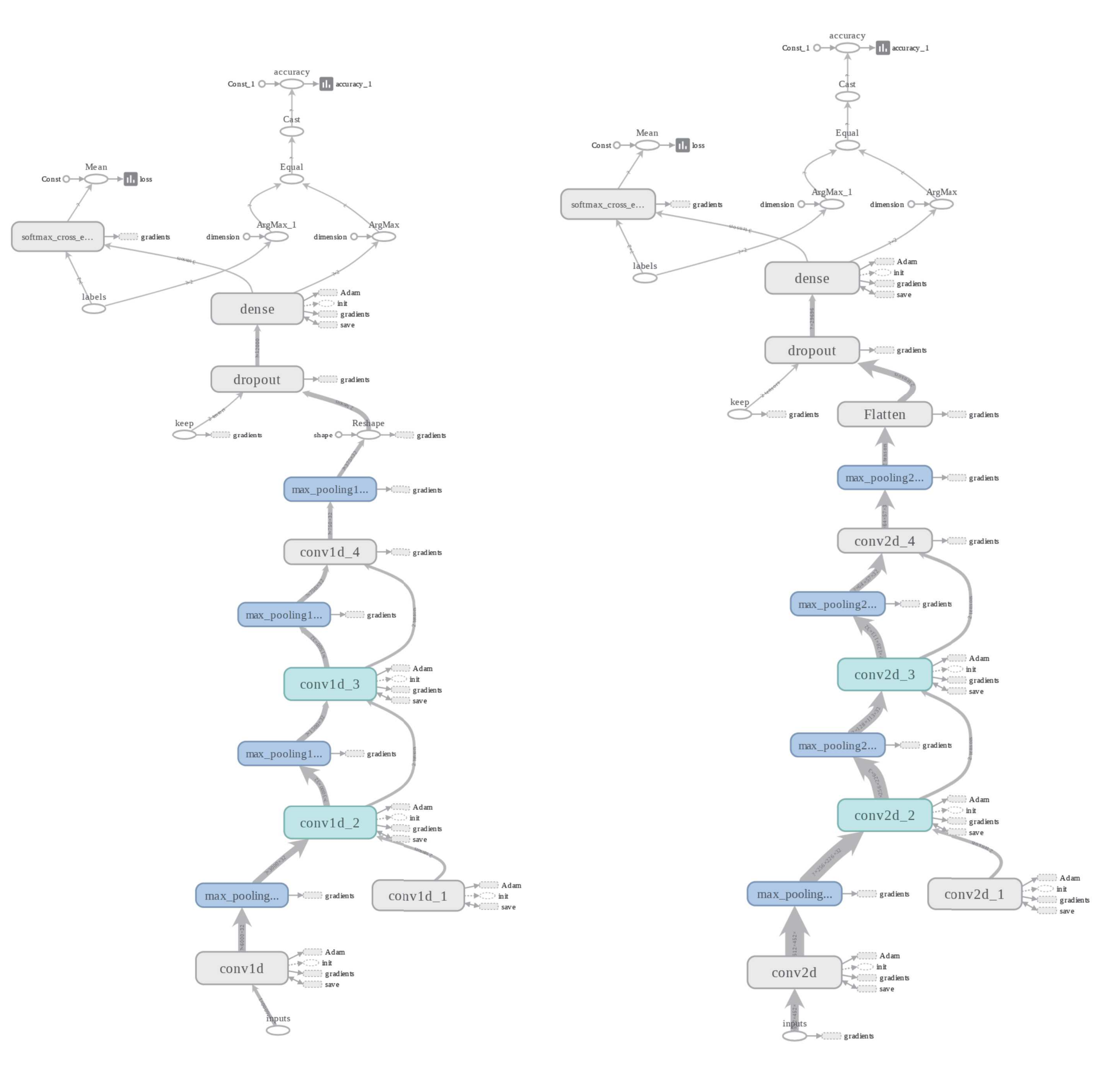

Figure 3: Structures of the neural network. Structure is the same for both waveform and spectrogram analysis except that in waveform analysis 1D convolution is used (left), whereas in spectrogram analysis 2D convolution is used (right).

Results

In our database we have 229 pulse shape signals. In order to train the model, we keep 1:1 ratio between impulsive and non impulsive signals and used randomly selected non impulsive signals in our database to keep the ratio. Result are in average of various random choices of impulsive and non impulsive signals for test, train and validations. The results can be seen in Table 1.

	Waveform	Spectrogram
Accuracy (%)	84	81

Table 1: Results of averaged randomly selected impulsive and non impulsive signals.

Shortcomings

- Requires high amount of memory to process
- Algorithms cannot detect the exact location of the impulsive part of the waveform

Conclusions

- Both waveform and spectrogram analysis have promising accuracy rates, which are 84% and 81%, respectively.
- Depending of layer architecture, waveform and spectrogram can analyzed more efficiently.

Future Plans

- Determining the position of the impulsive part of the waveform.
- Instead of using the spectrogram and waveform directly, one can extract high level information high can describe impulsive and non impulsive behaviours more accurately.
- Using the trained model in real time observations.

Acknowledgement

We would like to thank to Assistant Prof. Dr. Zhiwang Chang from School of Civil Engineering, Southwest Jiaotong University for sharing his algorithm with us and Dr. Emine Kucukbenli from Department of Condensed Matter Theory, International School for Advanced Studies (SISSA) and Dr. Eric Medvet from Universtiy of Trieste, Machine Learning Lab. for their advices about machine learning algorithms. We also would like to thank the National Research Institute for Earth Science and Disaster Resilience (NIED) for their websitef or K-NET and KiK-net (http://www.kyoshin.bosai.go.jp/, last accessed October, 2018) that makes it possible to access strong-motion seismographs in Japan, and GeoNet (www.geonet.org.nz/, last accessed October, 2018) for providing New Zealand strong-motion data.

References

Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.

TD Ancheta, Y Bozorgnia, R Darragh, WJ Silva, B Chiou, JP Stewart, DM Boore, R Graves, NA Abrahamson, KW Campbell, et al. Peer nga-west2 database: A database of ground motions recorded in shallow crustal earthquakes in active tectonic regions. In *Proceedings, 15th World Conference on Earthquake Engineering*, 2012.

G Asencio-Cortés, Antonio Morales-Esteban, Xueyi Shang, and Francisco Martínez-Álvarez. Earthquake prediction in california using regression algorithms and cloud-based big data infrastructure. *Computers & Geosciences*, 115:198–210, 2018.

Zhiwang Chang, Xiaodan Sun, Changhai Zhai, John X Zhao, and Lili Xie. An improved energy-based approach for selecting pulse-like ground motions. *Earthquake Engineering & Structural Dynamics*, 45(14):2405–2411, 2016.

Phoebe MR DeVries, Fernanda Viégas, Martin Wattenberg, and Brendan J Meade. Deep learning of aftershock patterns following large earthquakes. *Nature*, 560(7720):632, 2018.

Emilio Florido, G Asencio-Cortés, José Luis Aznarte, Cristina Rubio-Escudero, and Francisco Martínez-Álvarez. A novel tree-based algorithm to discover seismic patterns in earthquake catalogs. *Computers & Geosciences*, 115:96–104, 2018.

Taiwan Institute of Earth Sciences, Academia Sinica. Broadband array in taiwan for seismology, 1996.

Vassiliki Kardoutsou, Ioannis Taflampas, and Ioannis N Psycharis. A new pulse indicator for the classification of ground motions. *Bulletin of the Seismological Society of America*, 107(3):1356–1364, 2017.

Lucia Luzi, Rodolfo Puglia, Emiliano Russo, Maria D'Amico, Chiara Felicetta, Francesca Pacor, Giovanni Lanzano, Ulubey Çeken, John Clinton, Giovanni Costa, et al. The engineering strong-motion database: A platform to access pan-european accelerometric data. *Seismological Research Letters*, 87(4):987–997, 2016.

George P Mavroeidis and Apostolos S Papageorgiou. A mathematical representation of near-fault ground motions. *Bulletin of the seismological society of America*, 93(3):1099–1131, 2003.

Banu Mena and P Martin Mai. Selection and quantification of near-fault velocity pulses owing to source directivity. *Georisk*, 5(1):25–43, 2011.

Francesca Pacor, Roberto Paolucci, Lucia Luzi, Fabio Sabetta, Andrea Spinelli, Antonella Gorini, Mario Nicoletti, Sandro Marcucci, Luisa Filippi, and Mauro Dolce. Overview of the italian strong motion database itaca 1.0. *Bulletin of Earthquake Engineering*, 9(6):1723–1739, 2011.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

Shrey K Shahi and Jack W Baker. Regression models for predicting the probability of near-fault earthquake ground motion pulses, and their period. *Applications of Statistics and Probability in Civil Engineering*, 30(4):459, 2011.

Shrey K Shahi and Jack W Baker. An efficient algorithm to identify strong-velocity pulses in multi-component ground motions. *Bulletin of the Seismological Society of America*, 104(5):2456–2466, 2014.