Aims and scope

Geophysical Research Abstracts (GRA) is the conference series publishing the abstracts accepted for the General Assemblies of the European Geosciences Union (EGU). It links the annual conference programmes listing programme groups, included sessions, and their contributions. The abstracts underwent an access review by the session conveners.

Licence and copyright agreement

The following licence and copyright agreement is valid for any abstract submitted to the EGU General Assembly 2018 or later. For abstracts submitted to earlier assemblies, the CC BY 3.0 License was used.

Author’s certification

Copyright

Creative Commons Attribution 4.0 License

You are free to:

Share — copy and redistribute the material in any medium or format
Adapt — remix, transform, and build upon the material for any purpose, even commercially.

Identification of Possible Velocity Pulses in Earthquake Near Fault Regions by Using Machine Learning

Deniz Ertuncay, Giovanni Costa

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

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

Seismological Research and Monitoring Group

Data

In this study, near field stations (epicentral distance < 100 km) of crustal earthquakes (Epicentral depth < 40 km) with significant magnitudes, $M_w \geqslant 6.0$ are used. We covered most of the active tectonic regions which are producing high seismic risk (Figure 1). Total number of earthquake signals are more than 19,000 recorded as a result of more than 200 earthquake with various source mechanisms.

In image recognition algorithms we used velocity waveform since it is easier to detect pulse shape signals on velocity [2] [3] and spectrogram graphs (Figure 2) since it is also used on previous research on this topic [5] of each channel of seismic stations.

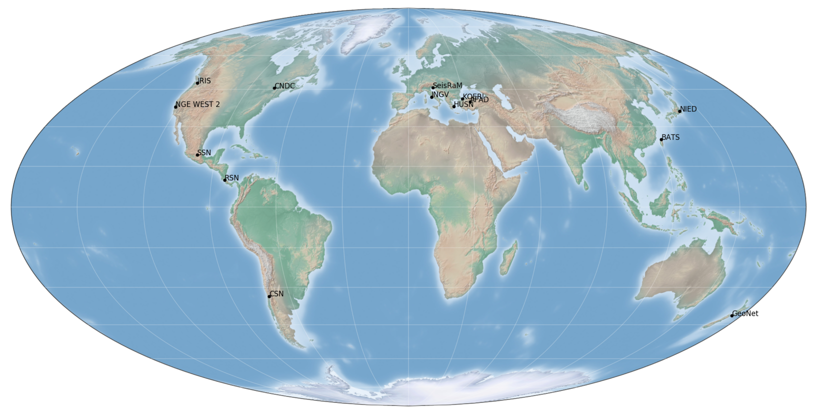

Figure 1: Seismological Networks & Databases That We Collect Data.

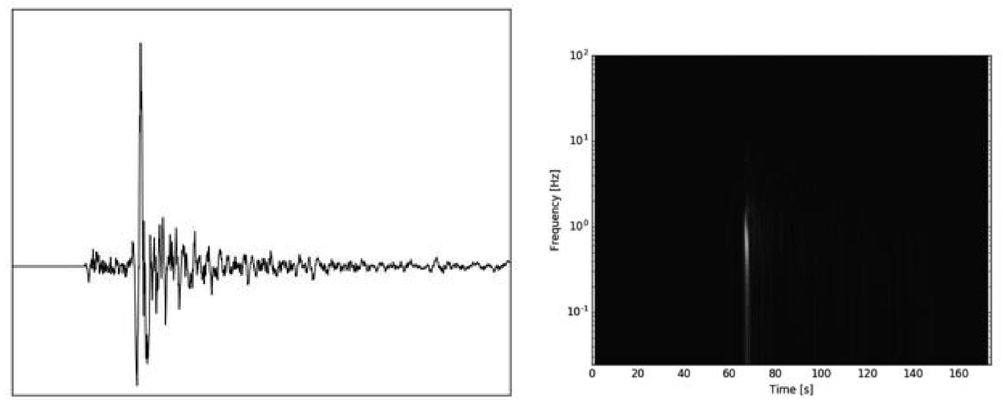

Figure 2: An example of velocity waveform and spectrogram of acceleration waveform of 5th of October 2015 M_w 6.1 at C100 station of University of Chile, National Seismological Center, respectively.

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.

Baker's Algorithm

Baker [2] used wavelet analysis to distinguished pulse shape and non impulsive signals from each other. The largest velocity pulse was extracted using the wavelet decomposition. The Daubechies wavelet of order 4 is used as the mother wavelet because it approximates the shape of many velocity pulses and it was seen to perform well.

Chang's Algorithm

Chang [3] used energy ratio between a part of the signal where PGV is occurred and whole velocity waveform. As Baker's algorithm, 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:

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

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:

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

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

Machine Learning Algorithms

Scikitlearn [1], Keras [4] and Tensorflow [6] packages are used in order to process the data. Methods are explained in CNN and Bottleneck sections and work-flows can be seen in Figure 3.
Image data is processed by using classifier methods.

Convolutional Neural Network (CNN)

Model 1 Each layer use 2D convolution, first two layers have 32 filters and the third one has 64 filters. Then pooling and fully connected layer is applied with 128 neurons. Results of this process are shown in Table 1 in Model 1.

Model 2 Each layer use 2D convolution with 3×3 kernel size, 32 filters and relu activation function. Dense functions use 128 and 1 units and relu and sigmoid activation functions, respectively. Results of this process are shown in Table 1 in Model 2.

Bottleneck

We used Tensorflow's bottleneck to process our database. The first phase analyzes all the images on the database and calculates the bottleneck values for each of them. 'Bottleneck' is an informal term for the layer just before the final output layer that actually does the classification. This layer, the one before the final layer, layer has been trained to output a set of values that's good enough for the classifier to use to distinguish between all the classes it has been asked to recognize. That means it has to be a meaningful and compact summary of the images, since it has to contain enough information for the classifier to make a good choice in a very small set of values. Results of this process are shown in Table 1 in Model 3.

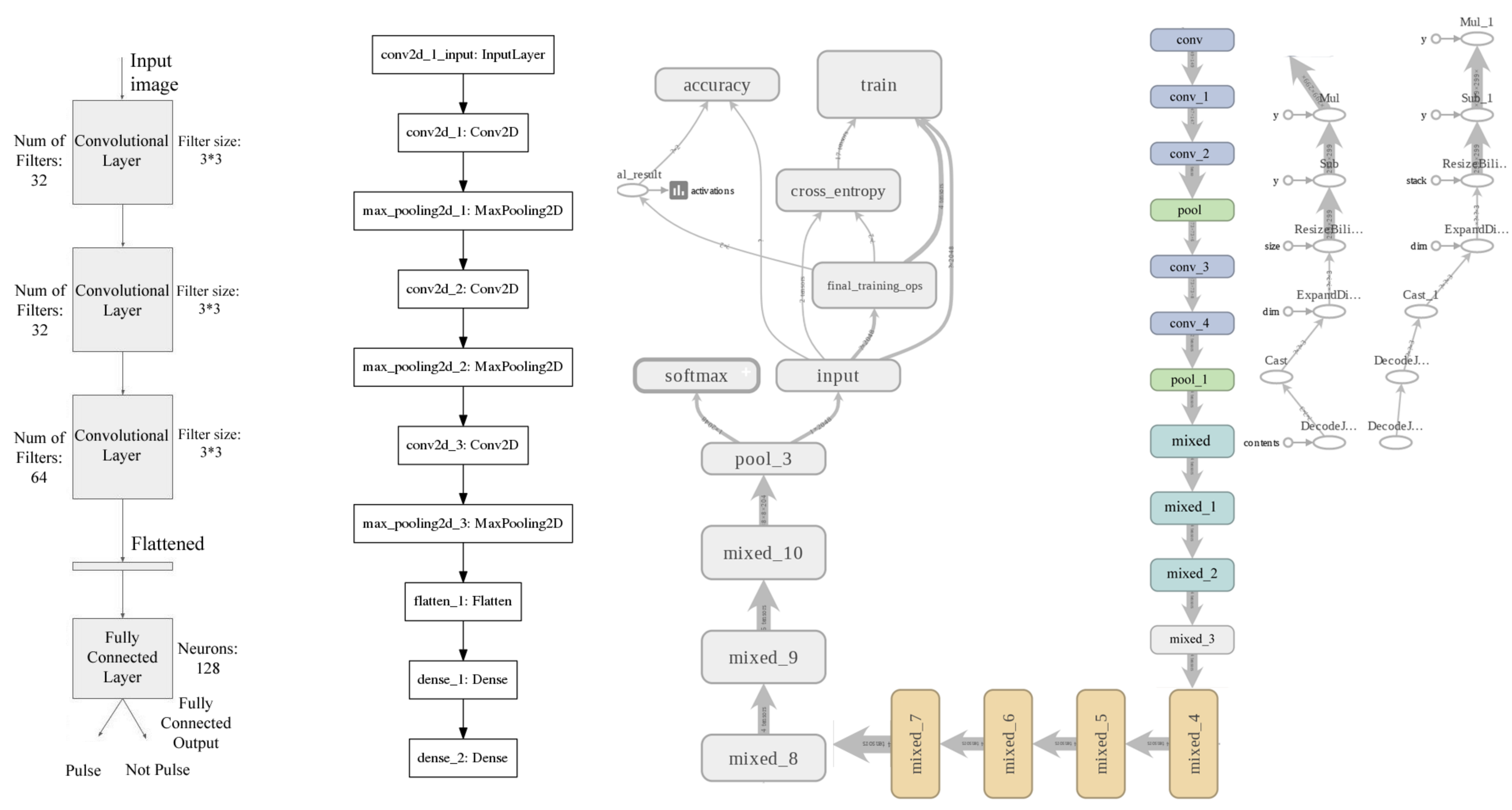

Figure 3: Work-flow of Models 1 to 3, respectively.

Results

		Model 1			Model 2			Model 3
		Not Pulse (%)	Pulse (%)	Acc. (%)	Not Pulse (%)	Pulse (%)	Acc. (%)	Acc. (%)
Baker 2007	Waveform	97.83	43.48	71.60	99.95	2.19	97.58	84.4
	Spectrogram	100	0	50	99.22	43.80	97.86	83.9
Chang 2016	Waveform	99.24	12.12	55.68	100	4.38	97.69	82.6
	Spectrogram	100	0	50	98.66	43.07	97.30	83.4

Table 1: Results of different models. Not pulse and Pulse are the percentages correctly determined signal records, respectively. Accuracy is the total accuracy of the process.

Shortcomings

- Requires high amount of memory to process
- Processing earthquake waveforms as images is not efficient
- Data should be trained with vast amount of previous waveforms

Conclusions

- Machine learning algorithms for image processing are not the best solution for determining the pulse shape signals.
- Image recognition approach can be useful, if the layers are tuned carefully.
- Depending of layer architecture, waveforms and spectrograms can analyzed more efficiently.
- Model 1 is relatively good for velocity waveform classification with respect to Model 2, whereas Model 2 is relatively good for spectrogram classification with respecto to Model 1.
- Model 3 is the best way of classifying pulse shape signals both in velocity waveform and spectrogram.

Future Plans

- Handling data as time series
- Determining high-level information about the waveforms to create simpler machine learning processes
- Combining these high-level information with PGA, PGV, M_w, azimuth, fault mechanism etc. information of waveforms for creating classification algorithms by using continuous data.

Acknowledgement

We would like to thank to Dr. Stefano Parolai from Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS) for his solid advices, 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) for her advices about machine learning algorithms.

References

[1] 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.

[2] Jack W Baker. Quantitative classification of near-fault ground motions using wavelet analysis. *Bulletin of the Seismological Society of America*, 97(5):1486–1501, 2007.

[3] 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.

[4] François Chollet et al. Keras. https://github.com/keras-team/keras, 2015. Accessed: 2018-02-28.

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

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.