Identification of Near Fault Pulse Shaped Signals With Machine Learning Algorithms

Deniz Ertuncay[1], Andrea De Lorenzo[2], Giovanni Costa[1], Eric Medvet[2]

1.Universtiy of Trieste, Department of Mathematics and Geosciences, SeisRaM Working Group, Trieste, Italy

2.Universtiy of Trieste, Department of Engineering and Architecture, Machine Learning Lab, Trieste, Italy

deniz.ertuncay@phd.units.it, andrea.delorenzo@units.it, costa@units.it, emedvet@units.it

Photography not allowed

Data

$17\,581$ of waveform are used in this study (Figure 1). Recorded waveform are result of an 388 of crustal earthquakes which are depth less than $50\,\mathrm{km}$. Arrival of the P wave of each signal picked manually and signals are trimmed between the P wave arrival and $40\,\mathrm{s}$ ahead of the P wave arrival. Sampling rate of the data is fixed to $20\,\mathrm{Hz}$. 456, 407 and 458 of the signals are identified as impulsive signals by Shahi and Baker (2014); Chang et al. (2016); Ertuncay and Costa (2019) respectively. Signals that are classified as pulse shaped varies between the studies depending on the method that used on data. We also manually picked the impulsive signals and train our model with that label. We then compare the manually picked results with previous studies.

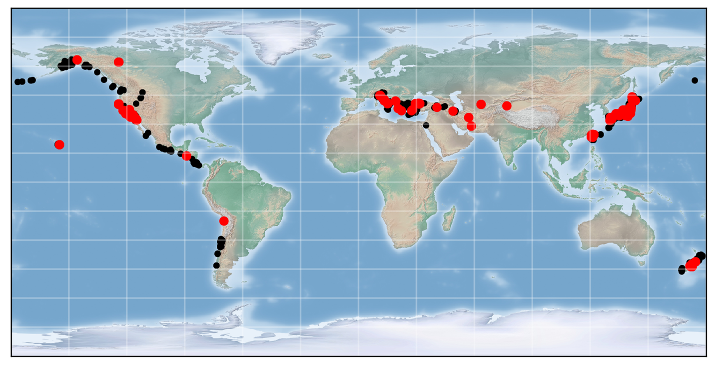

Figure 1: Spatial distribution of the seismic stations that are used in this study. Red colored stations are recorded impulsive signals according to any of the previous studies explained in Previous Studies whereas black color indicates the stations that are not contain any impulsive feature.

Data Augmentation

442 out of $17\,581$ waveform are labelled as impulsive by visual investigation. Since the ratio between impulsive (positive) and non impulsive (negative) signals are too low we incremented the number of impulsive signals by generating artificial ones. For each positive signal we generate as many signals as the ratio between positives and negatives. Artificial impulsive signals are created adding a zero mean Gaussian noise with 0.1 standard deviation.

Previous Studies

Shahi and Baker (2014)

Classification algorithm uses two criteria to determine whether the signal has impulsive or non-impulsive behavior. First criterion is the hazardousness of the signal. If PGV is less than $30\,\mathrm{cm/s}$, it is considered as non-hazardous signal. Second criterion is that the pulse indicator (PI) values should be higher than 0.85. PI is calculated as Eq. 1:

$$PI = \frac{1}{1 + e^{-23.3+14.6(PGV\,Ratio)+20.5(\text{energy ratio})}} \qquad (1)$$

Chang et al. (2016)

The algorithm determines a region around the PGV and determines the energy ratio between the pulse region and the total energy of the signal by taking the squared values on both signals. The region around PGV is calculated by using a least-square fitting for various pulse periods, then the one with the smallest residual is used for the pulse region. The energy ratio is then calculated as Eq. 2:

$$E(t) = \frac{\int_{t_s}^{t_e} v^2(\tau)d\tau}{\int_0^{\infty} v^2(\tau)d\tau} \qquad (2)$$

Ertuncay and Costa (2019)

Ertuncay and Costa (2019) used a combination of wavelet analysis and energy function of the waveform. Criteria for impulsive signals are explained in Eq. 3. If PGV value is greater than $30\,\mathrm{cm/s}$ and threshold in Eq. 3 is exceeded, signal is identified as impulsive.

$$\frac{\left(\frac{\int_{t_s}^{t_e} v^2(\tau)d\tau}{\int_0^{\infty} v^2(\tau)d\tau} + \frac{\int_{t_s}^{t_e} WPS(\tau)d\tau}{\int_0^{\infty} WPS(\tau)d\tau}\right)}{2} \geq 0.30 \qquad (3)$$

An example of a signal which is determined by all previous studies that we explained can be seen in Figure 2.

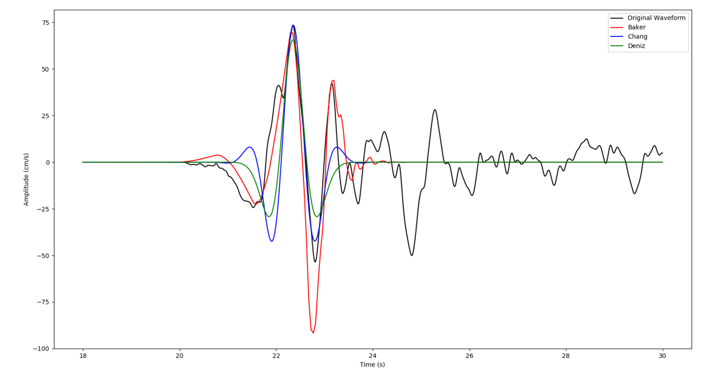

Figure 2: Velocity waveform of Amatrice Earthquake that is recorded at CLO station on October 30, 2016. Velocity waveform and extracted impulsive signals by Shahi and Baker (2014); Chang et al. (2016); Ertuncay and Costa (2019) are demonstrated with black, red, blue and green colors, respectively.

Convolutional Neural Network

We trained a Convulational Neural Network as described in Figure 3. We used as activation function ReLu for all layers except the last one in which we used a sigmoid. The loss function is defined as the binary cross-entropy and the learning rate is set accordingly with Adam optimization algorithm Kingma and Ba (2014). The weights of the neural network are initialized using the Glorot normal initializer. The network as been trained using a 10-fold cross-validation procedure which splits the whole dataset in two portions: the training and the testing one. The training set has been divided into two different portions: training and validation. The first one has been used in order to effectively train the network, whereas the second one as been used to stop the learning if the loss function start to grow up. Finally, we measured the False Positive Rate (FPR) and False Negative Rate (FNR) on the testing set. Accordingly with the cross-validation procedure, these steps have been repeated 10 times, varying the portions used as training and testing.

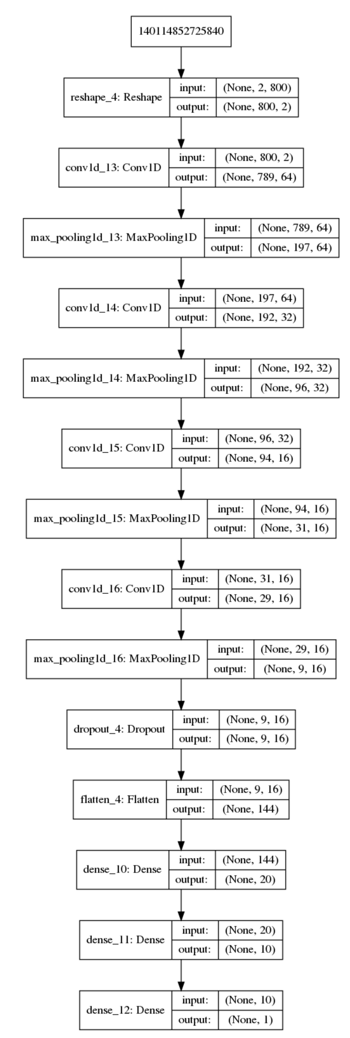

Figure 3: Structure of the Convolutional Neural Network used in this study.

Results

	FPR	FNR
Our proposal	0.023	0.249
Shahi and Baker (2014)	0.000	0.007
Chang et al. (2016)	0.500	0.008
Ertuncay and Costa (2019)	0.000	0.009
Zhai et al. (2018)	0.480	0.22

Conclusions

- In order to train the model, manually picking is necessary and picking the pulse shaped signals visually may cause disagreement with previous works. It is due to fact that identify the signal as pulse shape is subjective in some cases.
- Since the manual identification of impulsive signals is hard and subjective, FPR & FNR can be re-calculated depending on the manually picked dataset.
- Our method distinguish the impulsive and not impulsive signals with consicerebily high accuracy.

Future Plans

- Create a baseline with a combined results of previous works instead of using manually picked data.
- Use more sophisticated data augmentation methods such as rotating impulsive signals between 0^o - 90^o and adding Gaussian noise afterwards.
- Identifying the starting and ending points of the impulsive signals.

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 Jacopo Talamini from University of Trieste, Machine Learning Lab for his advises about machine learning algorithms.

References

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.

Deniz Ertuncay and Giovanni Costa. An alternative pulse classification algorithm based on multiple wavelet analysis. *Manuscript submitted for publication.*, 2019.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

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

Changhai Zhai, Cuihua Li, Sashi Kunnath, and Weiping Wen. An efficient algorithm for identifying pulse-like ground motions based on significant velocity half-cycles. *Earthquake Engineering & Structural Dynamics*, 47(3):757–771, 2018.