Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 18-19, 2023

A study on the design of ground motion database and processing for input seismic evaluation and nuclear power plant safety

Jin Koo Lee [a*], JeongBeom Seo [a], Sung Young Whang [a]
[a]*KITValley co., Ltd., 68, Digital-ro 9-gil, Geumcheon-gu, Seoul, Korea*
[*]*Corresponding author: jinguman@gmail.com*

1. Introduction

The 9.12 Gyeongju earthquake (Sep. 12, 2016, ML 5.8) and the Pohang earthquake (Nov. 15, 2017, ML 5.4) have occurred in the Korean Peninsula, and as a result, the stability of nuclear power plants has been emphasized. To confirm the stability against earth-quakes, ground motion processing is essential for many research topics such as structural design, input seismic evaluation, and ground motion attenuation. The time windows data for ground motion processing should have a single event to derive accurate results and it should be reasonably and scientifically processed.

We need to know P and S phase arrival times to find the event time window. However, the catalog of earthquakes in the Korean peninsula has P and S phase arrival times only for some stations used in the origin calculation. Therefore, we use deep learning-based phase (P & S phase) picker models to find event time windows from continuous data for all stations. Since 2018, various deep learning models have been studied around the world, showing high accuracy and efficiency compared to human analysis. General Phase Detector (GPD) [1], based on Convolutional Neural Networks (CNN), found many events missed by human analysts in the 2016 Bombay Beach, California Swarm. ARRU Phase Picker [2] was able to effectively find P and S phases at low Signal to Noise (SNR) ratio. Earthquake Transformer (EQTransformer) [3], which uses an attention mechanism, showed higher accuracy compared to manual picking for a Japanese earthquake.

Many methods of obtaining ground motion data have been proposed around the world. The NGA-east database generated ground motion data by processing earthquake events in the Central and Eastern North America (CENA) region. The processing flow of the NGA-east database [4] is 1) remove mean and instrument response, apply cosine taper 2) determine corner frequency by FAS 3) filter to reduce the noise 4) baseline correction. The ground motion processing method of RESOURCE [5], the reference database for seismic ground-motion, is 1) visual screening 2) remove mean and taper 3) determine corner frequency 4) apply 4-pole acausal Butterworth filter 5) fit a 6th order polynomial to the displacement trace 6) subtract the second derivative of the polynomial from the acceleration method. The method used by ITACA[6] in Italy is similar to the one described above.

In this study, we have developed a method for ground motion processing from continuous waveforms and a database scheme. We used deep learning models to determine the time window and ensemble the results to improve the accuracy under various conditions. The ground motion processing was developed using the NGA-east database as a reference. Considering the data structure of Seisbench [7], an open-source platform that provides standard datasets for machine learning, we designed the ground motion database and naming scheme.

2. Processing and Database schema

2.1 Processing

2.1.1. Phase picking based on multiple deep learning models

We used a deep learning-based picking and detection model to define the time window of the event in the continuous data and to know the time window of the noise, P, and S phases for ground motion processing. Münchmeye [8] evaluated EQTransformer, GPD, and PhaseNet [9] as the best performing models both in- and cross-domain. Therefore, we selected these models and ensemble the results.

The models and Ensemble (Voting) were evaluated using the Korean dataset by 3 tasks. 1) event detection 2) phase identification 3) P onset time determination. Fig. 1 shows the distribution of the Korean dataset. It consists of 1,394 earthquakes from 2010 to 2022. All earthquakes' magnitude is over 2.0.

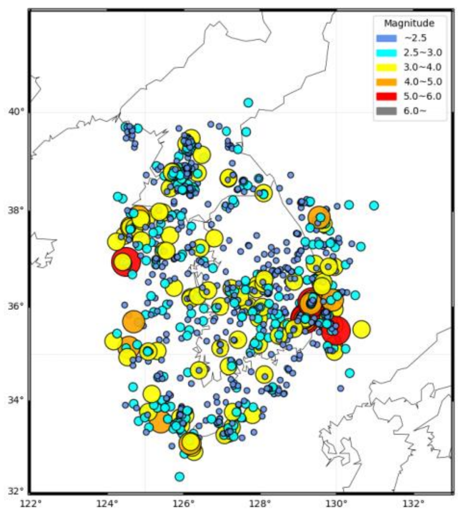

Fig. 1. The earthquake distribution of the Korean dataset

Fig. 2 shows the results of the performance comparisons. For event detection, EQTransformer (AUC 0.87) and Voting (AUC 0.87) showed the best performance. For Phase Identification, EQTransformer (AUC 0.76) and Voting (AUC 0.76) performed best. For P onset time determination, Ensemble (Voting) (MAE, Mean Absolute Error 1.74) and EQTransformer (1.83) performed well [10].

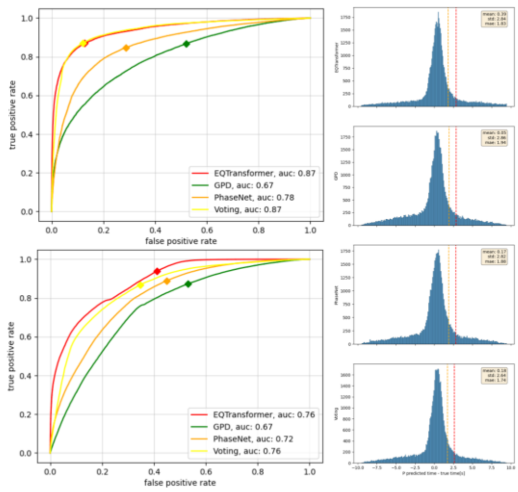

Fig. 2. Performance comparisons of each task. The left shows the Receiver operating characteristic for detection (Top) and identification (Bottom). The right shows histogram of P residuals.

2.1.2. Ground Motion Processing

For ground motion processing, we used Obspy [11] and TSPP [12]. Obspy is a Python-based open-source platform for seismic data processing, and TSPP is a set of Fortran programs for processing and manipulating time history.

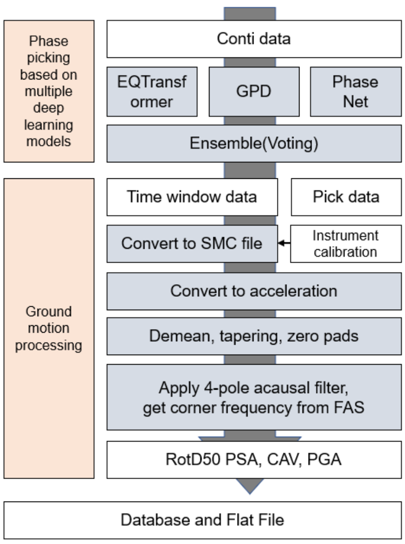

Fig. 3. Overall flow for processing ground motion data from Conti data

The continuous data is segmented into the time window data by Phase picking based on multiple deep learning models. It has a mini-seed file format, so it must convert to the SMC format used in TSPP. In the case of velocity records, it must convert to acceleration records. Finally, after mean, slope removal, and instrument calibration, the preprocessing is complete.
Apply tapering and zero-padding to preprocessed data. Compare the Fourier amplitude spectra (FAS) of noise and event parts or observe the drift of the displacement curve to determine the cutoff frequency. Adapt 4-pole acausal Butterworth and finally compute Peak Ground Acceleration (PGA), Cumulative absolute velocity (CAV) and Pseudo Spectral Acceleration (PSA). Fig. 3 shows the overall process.

2.2 Database Scheme

The naming scheme is organized in the form of "CATEGORY_PARAMETER_UNIT" as suggested by Seisbench. The categories are trace, source, station, atch, feat and path. The parameter describes the provided information, such as latitude, longitude or depth. The unit defines the unit in which the information is provided. It represents the parameter's physical value, like m, cm, counts, and samples.

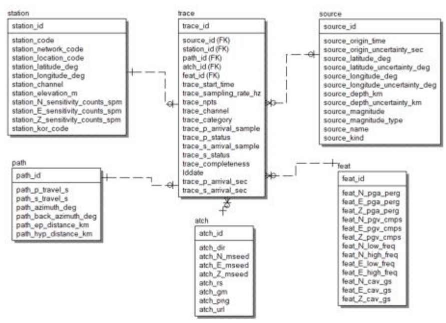

Fig. 4. Database scheme for ground motion database

Fig. 4 shows the database scheme for ground motion data. Station contains all information about the station that recorded the trace, such as the station's latitude, longitude, altitude, and calibration. Trace describes all information about the time history, such as the start time, end time, and P or S phase arrival time. Source contains earthquake related information. Path contains information about the relationship between source and station, such as azimuth, back azimuth, epicenter distance, and hypocenter distance. Feat contains a 1-dimensional array of information about features, such as PGA, PGV, and CAV that have only one value for each time history. Atch contains a 2-dimensional array information such as PSA, and FAS. Since it is difficult to include two-dimensional array information in a ground motion database, we use the flat file instead, and only its links are included in the database.

3. Conclusions

This study describes a methodology for ground motion database and processing. The results obtained are as follows:

1) We validated Phase Picking based on multiple deep learning models using Korean dataset and presented a method for generating ground motion data from continuous waveforms. This will be used to automatically generate ground motion data from real-time sensor.
2) Using Phase picking based on multiple deep learning models, we can process large amounts of historical data in a short time. It leads to a reduction in time costs and also reduces human error.
3) We presented a data processing method using the NGA-East database. This procedure can be used to increase the reliability of ground motion data in the future.
4) We proposed the direction for the scheme of the database. It was applied to the naming and database scheme of the Seisbench. We used flat file to include not only one-dimensional array information but also two-dimensional array information. This makes the database schema flexible so that it can store more data of different types that can be changed without major schema changes.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. RS-2022-00144482)

REFERENCES

[1] Z. E. Ross, M. Meier, E. Hauksson, and T. H. Heaton, "Generalized Seismic Phase Detection with Deep Learning," *Bull. Seismol. Soc. Am.*, vol. 108, no. 5A, pp. 2894–2901, Oct. 2018, doi: 10.1785/0120180080.

[2] W.-Y. Liao, E.-J. Lee, D. Mu, P. Chen, and R.-J. Rau, "ARRU Phase Picker: Attention Recurrent-Residual U-Net for Picking Seismic *P* - and *S* -Phase Arrivals," *Seismol. Res. Lett.*, vol. 92, no. 4, pp. 2410–2428, Jul. 2021, doi: 10.1785/0220200382.

[3] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, "Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking," *Nat. Commun.*, vol. 11, no. 1, p. 3952, Dec. 2020, doi: 10.1038/s41467-020-17591-w.

[4] C. A. Goulet *et al.*, "PEER NGA-East database," *Earthq. Spectra*, vol. 37, no. 1_suppl, pp. 1331–1353, Oct. 2014, doi: 10.1177/87552930211015695.

[5] S. Akkar *et al.*, "Reference database for seismic ground-motion in Europe (RESORCE)," *Bull. Earthq. Eng.*, vol. 12, no. 1, pp. 311–339, Feb. 2014, doi: 10.1007/s10518-013-9506-8.

[6] F. Pacor, R. Paolucci, G. Ameri, M. Massa, and R. Puglia, "Italian strong motion records in ITACA: overview and record processing," *Bull. Earthq. Eng.*, vol. 9, no. 6, pp. 1741–1759, Dec. 2011, doi: 10.1007/s10518-011-9295-x.

[7] J. Woollam *et al.*, "SeisBench -- A Toolbox for Machine Learning in Seismology," *ArXiv211100786 Phys.*, Nov. 2021, Accessed: Jan. 10, 2022. [Online]. Available: http://arxiv.org/abs/2111.00786

[8] J. Münchmeyer *et al.*, "Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers," *ArXiv211013671 Phys.*, Oct. 2021, Accessed: Jan. 10, 2022. [Online]. Available: http://arxiv.org/abs/2110.13671

[9] W. Zhu and G. C. Beroza, "PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method," *ArXiv180303211 Phys. Stat*, Mar. 2018, Accessed: Jan. 10, 2022. [Online]. Available: http://arxiv.org/abs/1803.03211

[10] J. K. Lee and J. Seo, "A study on input ground motion processing platform for evaluating seismic fragilities using Deep Learning Phase Determination Model," EGU General Assembly 2023, Vienna, EGU23-4709, Apr. 2023. doi: 10.5194/egusphere-egu23-4709.

[11] L. Krischer *et al.*, "ObsPy: a bridge for seismology into the scientific Python ecosystem," *Comput. Sci. Discov.*, vol. 8, no. 1, p. 014003, May 2015, doi: 10.1088/1749-4699/8/1/014003.

[12] D. M. Boore, "TSPP---A Collection of FORTRAN Programs for Processing and Manipulating Time Series," U.S. Geological Survey Open-File Report, Open-File Report, 2008.