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

Preliminary study of estimation of earthquake characteristics in unmeasured site through machine learning of earthquake observation records

Yongmoon Hwang [a], Junghoon Lee [b*]
[a]*Korea Atomic Energy Research Institute, Daedeok-daero 989 beon-gil, Yuseong-gu, Daejeon, 34057, Korea*
[b]*Korea Advanced Institute of Science and Technology, Daehak-ro 291, Yuseong-gu, Daejeon, 34141, Korea*
[*]*Corresponding author: jhlee07@kaist.ac.kr*

1. Introduction

Since the 2016 Gyeong-ju earthquake, the safety issue of infrastructure under earthquake has been increased. It is important to minimize the damage caused by the earthquake and to take prompt follow-up management. To this end, it is important to utilize real-time monitoring data, but the number is limited excepted for specific locations. Therefore, in this study, technique for estimating the seismic characteristics of unmeasured site using the data measured nearby is developed. The characteristics of the ground are one of variables that occur depending on the location of the seismograph, and to minimize this, machine learning was used to estimate the characteristics of the seismic load in the unmeasured site. In the future, method for estimating the earthquake characteristics in unmeasured site can be used in nuclear power plant structures using the seismic monitoring system.

2. Methods and Results

In this study, as a preliminary study to identify earthquake characteristics in unmeasured sites, various earthquake records were obtained and pre-processed, and the characteristics were confirmed through machine learning. In this section, the data collection and pre-processing of the measured earthquake records, the selection of learning model variables, and the learning method are mentioned, and the initial results were calculated.

2.1 Data Acquisition and Pre-processing

NECIS collected continuous waveform data from all seismic stations operated by the Korea Meteorological Administration. As for the collected data, a list of domestic earthquakes with a magnitude of 3.0 or higher from 2015 to 2021 was obtained. The obtained data were listed as location code of seismic stations, epicenter location (degree of latitude and longitude), and 3-directional acceleration gain data.

For each earthquake, the measurement data was limited to 5 minutes after the earthquake occurred, and the data were interpolated using the average value of the accelerometer data. The data obtained in the pre-processing was identified as the PGA value according to the distance from the epicenter to confirm the decrease trend according to the distance, but outliers were identified as shown in Fig. 1.

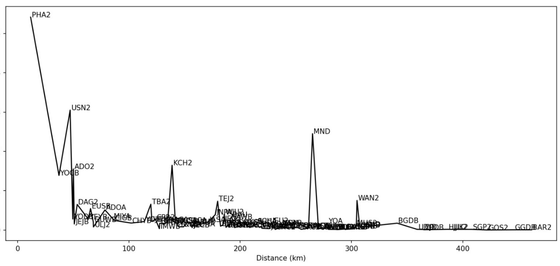

Fig. 1 PGA according to the distance of the Pohang earthquake

2.2 PGA estimation using Machine Learning

After the pre-processing process, the input data and output data were organized to perform the machine learning on the measurement data. The input variables are as follows: 1) latitude and longitude of the epicenter, 2) magnitude, 3) latitude and longitude of seismic stations, 4) distance between the epicenter and seismic stations, 5) orientation of the seismic station from the earthquake (based on north, west (-), east (+), degree). About 8,000 data were organized using PGA as an output variable.

As a machine learning method, a multi-perceptron method (MLP) Deep neural network model was used. In this study, it was determined that the method to be estimated is close to regression, which is a case where there are N input and output values, and learning using existing data to predict the output for a new input value, using a commonly used MLP method was adopted [1].

The structure of MLP model consists of one input layer, two hidden layers, and one output layer. The input layer consists of a total of 7 nodes, the hidden layers each consist of 100 nodes, and the output layer consists of 1 node. It is composed of dense layer connected between adjacent layers, and the rectified linear unit (ReLU) is used the activation function of the nodes [2].

The epoch that repeats the learning is performed 20 times, and the k-fold method is adopted to randomly separate and verify 80% of the learning set and the validation set for each epoch. The optimizer used RMSPorp, the target function to be minimized through learning was Mean Square Error (MSE), and the evaluation index was MSE and Mean Absolute Error (MAE), and the convergence and accuracy of learning were judged. Learning cases were performed in four ways as follows:

1. Learning with pre-processed input and output variables

MSE and MAE quickly approached 0 during execution, so learning seemed to be successful, but the difference between the actual value (seismic measurement data) and the predicted value (value predicted by the neural network) was significantly large, so it was confirmed that the learning was inaccurate.

2. Scaling PGA as log-scale

When learning progressed, the predicted value compared to the previous result was accurate, but some trends were confirmed to be reversed.

3. Learning after normalization using mean and standard deviation

After obtaining the average and standard deviation of each variable for input and output data, learning was conducted by normalizing to (normalization value) = (actual value – average value) / (standard deviation value).

4. Learning after normalization using maximum and minimum value

After obtaining the maximum and minimum values for each variable, the input and output data were normalized to (normalization value) = (actual value – average value) / (maximum value – minimum value).

The results of the first 250 sets among a total of 8,000 sets for the validation values and real values obtained as a results of the learning above are shown in Fig. 2.

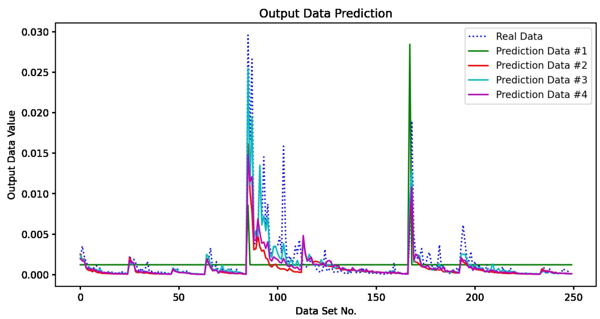

Fig. 2 Output variable prediction comparison for each training

3. Conclusions

In this study, the PGA of the unmeasured site was predicted using the measured earthquake data. In order to increase the prediction accuracy, it was performed based on 4 learning cases, and some limitation also occurred. In future, PGA as well as frequency characteristics of the unmeasured site can be identified by updating the constructed model.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grand funded by the Korea government (MSIT) (RS-2022-00144425 and 2021R1I1A1A01050935).

REFERENCES

[1] Shiblee, Md, Perm Kumar Kalra, and B. Chandra, “Time series prediction with multilayer perceptron (MLP): a new generalized error based approach,” Advances in Neuro-Information Processing: 15th International Conference, ICONIP 2008.

[2] Brownlee J., “A gentle introduction to the rectified linear unit (ReLU),” Machine learning mastery, 6, 2019.