令和２年度修士論文要旨集（部内資料）
千葉大学大学院融合理工学府 地球環境科学専攻 都市環境システムコース　令和３年２月2,3日

機械学習を用いた最大速度の距離減衰式の構築と地点補正項に関する検討

CONSTRUCTION OF ATTENUATION RELATIONSHIP OF PEAK GROUND VELOCITY USING MACHINE LEARNING AND EXAMINATION OF STATION CORRECTION FACTOR

19WM1328 呉　俊傑
Wu Junjie

指導教員 丸山　喜久

SYNOPSIS

This study tries to develop new attenuation relationships of peak ground velocity using machine learning methods: random forest and neural network. In order to compare with the predictors obtained by machine learning, we have also constructed a new attenuation relationship of peak ground velocity using three-stage regression procedure proposed by Molas and Yamazaki (1995). In this study, 6,944 ground motion records at 1,184 seismic observation stations which were observed during the 32 earthquakes are employed to construct the attenuation relationships. Ground motion records from the 4 recent earthquakes are used as the test set. The test results show that when the shortest distance from the fault is small, the predictions by machine learning techniques are more accurate than those by the traditional equation. However, there is still a problem of overestimation in the predictors of machine learning, even if weights are considered for the training set. In addition, the station correction factors based on machine learning were derived and proved to be positively correlated with the average shear wave velocity.

1. Introduction

The attenuation relationship is a method to predict the ground motion intensity of earthquake that may occur in the future based on the ground motion records of past earthquakes. The attenuation relationships are used in both deterministic and probabilistic seismic hazard analyses. The attenuation refers to the phenomenon that the farther away from the epicenter, the weaker the earthquake intensity. The previous attenuation relationships are empirical equations that predict the level of ground shaking, based on the source characteristics (e.g., earthquake magnitude), the propagation path (e.g., the shortest distance from the fault), and the local site conditions, etc[1).

Since more ground motion records are obtained during recent earthquake events, the research of attenuation relationship has been greatly developed. However, due to the lack of ground motion records near the epicenter, it was found that previous attenuation relationships have low reliability under the small distances from the fault[2)]. Therefore, this study tries to develop new attenuation relationships of peak ground velocity (PGV) using machine learning methods.

In this study, random forest and neural network are used to predict the PGV. Previous studies have constructed attenuation relationships using random forest (Kubo 2018) and neural network (Derras 2012) [3)-6)]. In this study, we want to compare the predictors obtained by machine learning with a traditional equation. Therefore, we have constructed a new attenuation equation of PGV using three-stage regression procedure proposed by Molas and Yamazaki (1995).

2. Dataset

In this study, we use ground motion records obtained by K-NET (Kyoshin network) and KiK-net (Kiban Kyoshin network), which are strong-motion seismograph networks constructed by the National Research Institute for Earth Science and Disaster prevention (NIED) [7)]. In this study, 6,944 ground motion records at 1,184 K-NET and KiK-net seismic observation stations which were observed during the 32 earthquakes are employed as training data to construct the attenuation relationships[1)]. Table 1 shows the list of earthquake events used as training set. The training set consists of earthquake events from 1997 to 2011. In order to focus on the prediction of close-range data, we give the following weights to the training data when constructing the attenuation equation: 8 for 0-25 km, 4 for 25-50 km, 2 for 50-100 km, and 1 for larger than 100 km[2)]. Considering that all the 3 methods use the same explanatory variables and objective variable, we also give the same weights to the training data of the machine learning models. Moreover, ground motion records observed during the 4 recent earthquake events are used as the test set to evaluate the performances of predictors. Table 2 shows the list of earthquakes used as test data. In this study, PGV is used as the objective variable, and the moment magnitude (Mw), the shortest distance from the fault (r), the earthquake source depth (H), and the dummy variable (S_i), which is mentioned later, are used as the explanatory variables to construct the attenuation relationships.

3. Construction of attenuation relationships

In order to compare with the predictors obtained by machine learning, we have constructed a new attenuation equation of PGV using three-stage regression procedure proposed by Molas and Yamazaki (1995). The resulting equation is

$log\,PGV = -1.541 + 0.648Mw - 0.00153r - log(r + 0.0033 * 10^{0.5*Mw}) + 0.00299H + C_i$ (1)

Figure 1 shows the comparisons of the predicted PGV of the attenuation equation developed in this study with those of Si and Midorikawa (1999), Joyner and Boore (1981), and Molas and Yamazaki (1995) when the magnitude is set to be 7; the earthquake source depth is set to be 5 km; the station coefficient is set to be 0. According to Fig. 1, the trends of the 4 equations are similar. Although there are variations among the 4 equations, the attenuation relationship proposed by this study has an intermediate value.

Table1. List of earthquake events used as the training set.

No.	Date	Mw	Depth	Number of Records
1	1997.03.26	6.0	8	58
2	1997.05.13	5.9	8	52
3	2000.10.16	6.6	11	245
4	2000.10.31	5.4	44	141
5	2001.03.24	6.9	51	274
6	2001.04.25	5.4	42	93
7	2003.05.26	7.0	71	392
8	2003.07.26	6.2	12	194
9	2003.09.26	8.0	42	342
10	2004.09.05	7.4	44	369
11	2004.10.23	6.5	13	294
12	2004.10.27	5.8	12	214
13	2004.11.29	6.8	48	180
14	2004.12.14	5.9	9	70
15	2005.03.20	6.6	9	192
16	2005.07.23	5.8	73	193
17	2005.08.16	7.1	42	395
18	2006.04.21	5.6	7	66
19	2006.05.02	5.1	15	64
20	2006.06.12	5.9	146	221
21	2006.08.31	4.8	76	107
22	2007.03.25	6.7	11	234
23	2007.07.16	6.7	17	304
24	2008.05.08	6.9	51	202
25	2008.06.14	6.9	8	250
26	2008.09.11	6.8	31	144
27	2009.08.11	6.2	23	287
28	2010.02.27	6.7	37	8
29	2011.03.09	7.3	8	234
30	2011.03.11	9.0	24	685
31	2011.04.11	6.6	6	306
32	2011.04.12	6.4	26	134

Table 2. List of earthquakes used as the test set.

No.	Date	Mw	Depth	Number of Records (r < 100km)
1	2016.04.14	6.5	11	116
2	2016.04.16	7.3	12	128
3	2018.06.18	6.1	13	120
4	2018.09.06	6.7	35	69

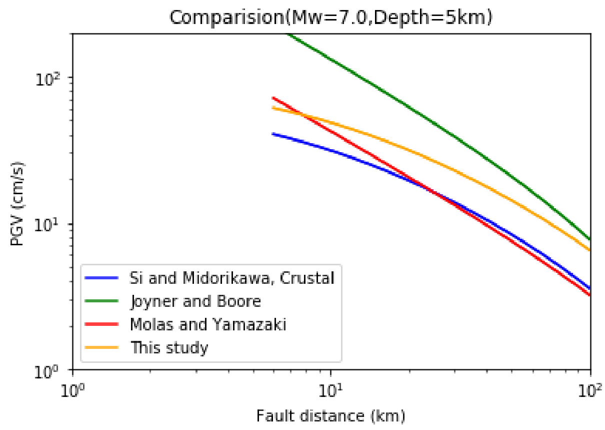

Fig. 1. Comparison of attenuation equations for PGV of this study and previous studies for magnitude 7.0 earthquake with the depth of 5 km.

In the 2 machine learning models, the objective variable is PGV, and the moment magnitude (M_w), the shortest distance from the fault (r), the earthquake source depth (H) and the 1,184 dummy variables (S_i) are used as the explanatory variables. The dummy variables, S_i, are configured such that the mean of station correction factor is zero. For the jth data recorded at station k, $S_{i=k,j} = 1$ and $S_{i\neq k,j} = 0$ except if the data is recorded at the last (1184th) station, then S_i is taken as -1 for i=1 to 1,183[8].

Random forest is an ensemble model consisting of many decision trees. Predictions are made by averaging the predictions from each decision tree. Alternatively, as a forest is a collection of trees, a random forest model is a collection of decision trees. The core idea behind random forest is to generate multiple decision trees from random subsets of the dataset.

Neural network model, more properly referred to as artificial neural network (ANN), is a forecasting method based on simple mathematical model of the brain[4]. It allows complex nonlinear relationships between the response variable and its predictions.

In this study, we use the Scikit-learn, which is a machine learning library for the Python programming language, to construct the both random forest model and neural network model. Figure 2 shows the predicted attenuation curves for PGV of the attenuation equation, random forest model, and neural network model for the moment magnitudes of 5.0, 6.0 and 7.0 earthquakes with the depth of 5 km. The predictions of machine learning seem to be reasonable and stable like the attenuation equation as demonstrated in Fig. 2. It can also be seen that PGV and the moment magnitude have a clear correlation, although the curve of random forest model is relatively volatile.

4. Evaluation and discussion

After constructing the traditional attenuation equation and the attenuation relationships based on machine learning, we evaluate their predictive abilities. Ground motion records observed by K-NET and KiK-net during the 4 recent earthquake events are used as the test set.

The observations and predictions for the 4 recent earthquakes in the test dataset are compared to demonstrate the prediction performance of the 3 models in Fig. 3. Figure 3 shows that the overall feature of the observation is reproduced by the machine learning models. In the case of the April 14 foreshock of 2016

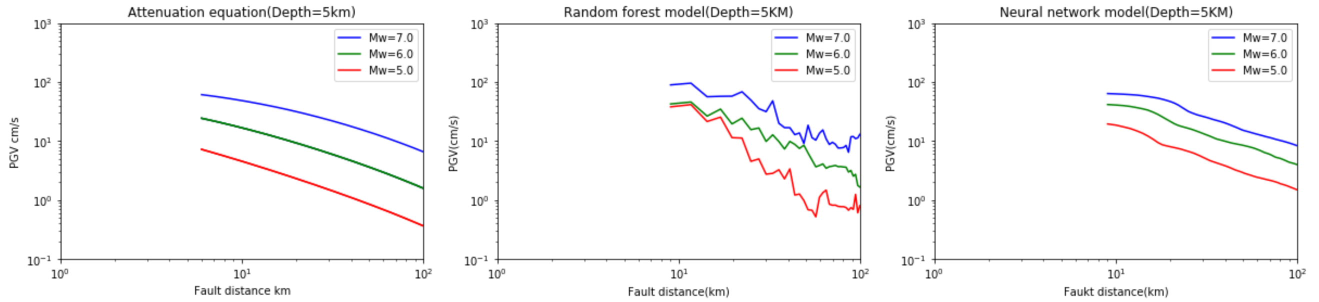

Fig. 2. Predicted attenuation curves for PGV obtained by the attenuation equation, random forest, and neural network for the magnitudes of 5.0, 6.0 and 7.0 earthquakes with depth of 5 km.

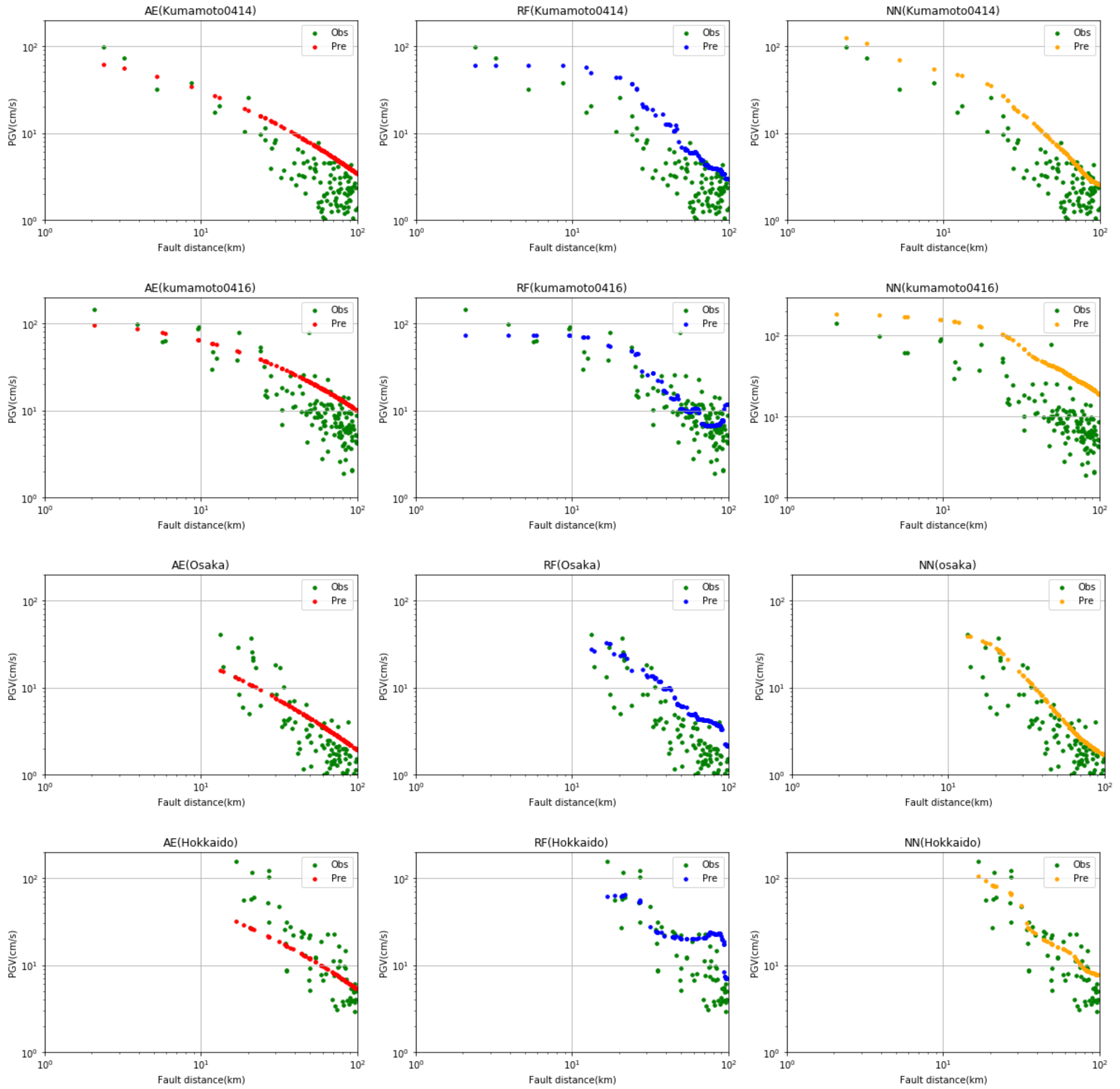

Fig. 3. Observations (Obs) and predictions (Pre) of the 4 recent earthquakes by predictors of attenuation equation (AE), random forest (RF), neural network (NN). Red dots represent the prediction by the attenuation equation, blue ones represent the prediction by the random forest predictor, orange ones represent the prediction by the neural network predictor, and green ones represent the observed data, respectively.

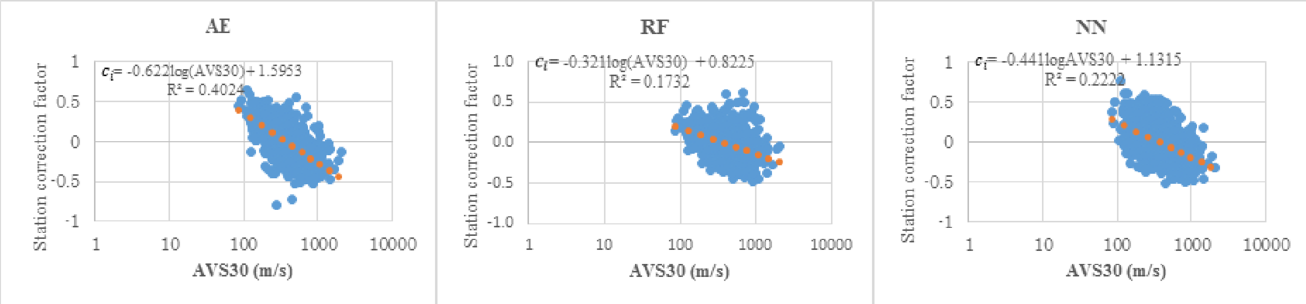

Fig. 4. Relationships between the station correction factors and AVS30 based on the attenuation relationship, random forest model, and neural network model, respectively.

Kumamoto earthquake, even if the model of attenuation equation makes a better prediction in the most parts, the neural network predictor gets a prediction of over 100 cm/s, whose observation value is also greater than 100 cm/s. It also can be seen that in case of the April 16 main shock of 2016 Kumamoto earthquake and the 2018 Hokkaido Eastern Iburi earthquake, the neural network predictor makes the closest predictions on the large observed PGV whose values are greater than 100 cm/s. However, the neural network predictor overestimates the observed PGV, overall. Figure 3 also shows that the random forest model has higher reliability than the attenuation equation at close range. Apart from the case of the April 14 foreshock of 2016 Kumamoto earthquake, the random forest predictor has better performance than the attenuation equation, especially in terms of the prediction of the case of the April 16 main shock of 2016 Kumamoto earthquake and the 2018 Hokkaido Eastern Iburi earthquake. In the prediction of the maximum value of PGV for the 4 earthquake events, the random forest performs better than the attenuation equation except for the case of the April 16 main shock of 2016 Kumamoto earthquake. However, the random forest predictor trends to overestimate the observed PGV when the shortest distance from the fault is larger than 70 km.

In this study, we calculate the station correction factors based on the attenuation relationships of machine learning methods. The station correction factors are effective for evaluating site amplifications. Moreover, in the three-stage regression procedure of the attenuation equation, the station correction factors have been calculated in this study. Figure 4 shows the relationships between the station correction factors calculated in this study and the shear wave velocity averaged over the upper 30 m (AVS30) of the attenuation equation, random forest model, and neural network model. The AVS30 is used for soil classification in the seismic design code in the United States[1]. Figure 4 shows that the station correction factors based on machine learning are positively correlated with the AVS30, although the coefficients of determination are lower than those of the attenuation equation.

5. Conclusion

In this study, a traditional attenuation equation using three-stage regression procedure proposed by Molas and Yamazaki (1995) and the 2 attenuation relationships using random forest and neural network are constructed to predict the PGV. The objective variable and the explanatory variables of the 3 models are set to be the same. The prediction performance of the attenuation equation for the training data is not good as that for the predictors of machine learning. As the machine learning predictor is fully data-driven predictive models when the attenuation equation is formulated. In the case of test data, although the predictor of attenuation equation makes a better prediction for the April 14 foreshock of Kumamoto earthquake. Predictions by the 2 machine learning predictors are improved for the other 3 earthquake events at close range. It seems that machine learning methods improve the reliability of attenuation relationship at close range. In addition, the station correction factors based on machine learning are derived and proved to be positively correlated with AVS30. However, there is an overestimation problem for machine learning models.

References

1) Yoshihisa Maruyama, Masaki Sakemoto: Development of nationwide amplification map of response spectrum for Japan based on station correction factors, Earthquakes and Structures, Vol. 13, No. 1, pp17-27, 2017.
2) Hongjun Si, Saburoh Midorikawa: New Attenuation Relationships For Peak Ground Acceleration And Velocity Considering Effects Of Fault Type And The Condition, J. Struc, Constr. ENG, AIJ, No. 523, pp. 63-70, 1999.
3) Boumédiène Derras, Pierre-Yves Bard, Fabrice Cotton, Abdelmalek Bekkouche: Adapting the Neural Net-work Approach to PGA Prediction: An Example Based on the KiK-net Data, Bulletin of the Seismological Society of America, Vpl. 102, No. 4, pp. 1446-1461, August 2012.
4) Hisahiko Kubo, Takashi Kunugi, Shingo Suzuki, Wataru Suzuki, Shin Aoi: Construction of ground motion prediction equation using random forest, The 32nd Annual Conference of the JSAI, 2018.
5) Hisahiko Kubo, Takashi Kunugi, Shingo Suzuki, Wataru Suzuki, Shin Aoi: Attempt to reduce the effect of biased data-set on ground-motion prediction using machine learning, The 33rd Annual Conference of the JSAI, 2019.
6) Hisahiko Kubo, Takashi Kunugi, Wataru Suzuki, Takeshi Kimura, Shin Aoi: Construction of explainable random forest predictor for ground-motion intensity, The 34th Annual Conference of the JSAI, 2020.
7) Yoshimitsu Fukushima: Survey of Recent Studies on Attenuation Relation of Strong Ground Motion (Empirical Prediction Relation), The earthquake, Vol. 46, No. 2, pp. 315-328, 1993.
8) Gilbert L. Molas, Fumio Yamazaki: Attenuation of Equation Ground Motion in Japan Including Deep Focus Events, Bulletin of the Seismological Society of America, Vol. 85, No. 5, pp. 1343-1358, October 1995.