

EGU2020-10937

<https://doi.org/10.5194/egusphere-egu2020-10937>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



Machine Learning based Multivariate Seismic Site Classification System for South Korea

Han-Saem Kim, Chang-Guk Sun, Hyung-Ik Cho, and Moon-Gyo Lee

Korea Institute of Geoscience and Mineral Resource, Korea, Republic of (adoogen@kigam.re.kr)

Earthquake-induced land deformation and structure failure are more severe over soft soils than over firm soils and rocks owing to the seismic site effect and liquefaction. The site-specific seismic site effect related to the amplification of ground motion has spatial uncertainty depend on the local subsurface, surface geological, and topographic conditions. When the 2017 Pohang earthquake (M 5.4), South Korea's second-strongest earthquake in decades, occurred, the severe damages influencing by variable site effect indicators were observed focusing on the basin or basin-edge region deposited unconsolidated Quaternary sediments. Thus, the site characterization is essential considering empirical correlations with geotechnical site response parameters and surface proxies. Furthermore, in the case of so many variables and tenuously related correlations, machine learning classification models can prove to be very precise than the parametric methods. In this study, the multivariate seismic site classification system was established using the machine learning technique based on the geospatial big data platform.

The supervised machine learning classification techniques and more specifically, random forest, support vector machine (SVM), and artificial neural network (ANN) algorithms have been adopted. Supervised machine learning algorithms analyze a set of labeled training data consisting of a set of input data and desired output values, and produce an inferred function which can be used for predictions from given input data. To optimize the classification criteria by considering the geotechnical uncertainty and local site effects, the training datasets applying principal component analysis (PCA) were verified with k-fold cross-validation. Moreover, the optimized training algorithm, proved by loss estimators (receiver operating characteristic curve (ROC), the area under the ROC curve (AUC)) based on the confusion matrix, was selected.

For the southeastern region in South Korea, the boring log information (strata, standard penetration test, etc.), geological map (1:50k scale), digital terrain model (having 5 m × 5 m), soil map (1:250k scale) were collected and constructed as geospatial big data. Preliminarily, to build spatially coincided datasets with geotechnical response parameters and surface proxies, the mesh-type geospatial information was built by the advanced geostatistical interpolation and simulation methods.

Site classification systems use seismic response parameters related to the geotechnical characteristics of the study area as the classification criteria. The current site classification systems in South Korea and the United States recommend Vs30, which is the average shear wave velocity

(Vs) up to 30 m underground. This criterion uses only the dynamic characteristics of the site without considering its geometric distribution characteristics. Thus, the geospatial information for the input layer included the geo-layer thickness, surface proxies (elevation, slope, geological category, soil category), average Vs for soil layer ($V_{s,soil}$) and site period (TG). The Vs30-based site class was defined as categorical labeled data. Finally, the site class can be predicted using only proxies based on the optimized classification techniques.