GC11-solidearth-45, updated on 25 May 2024
https://doi.org/10.5194/egusphere-gc11-solidearth-45
Galileo Conference: Solid Earth and Geohazards in the Exascale Era

Feasibility of Multiple Advanced Machine Learning Techniques for Synthetic Finite-fault Earthquake Ground Motion Data

Manuel Titos Luzon[1], Milad Kowsari[2], and Benedikt Halldorsson[2,3]
[1]Department of Signal Theory, Telematics and Communications, University of Granada, Granada, Spain
[2]Faculty of Civil and Environmental Engineering, School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland
[3]Volcanic activity, earthquakes and deformation dept., Service and research Division, Icelandic Meteorological Office, Reykjavik, Iceland

The overwhelming success of data-driven models to solve complex predictive real-world problems has made them an effective alternative to the simulation-driven models. In addition to their computational cost, simulation-driven approaches need to be calibrated by actual data that links both to the physical theory and thereby improves both our knowledge and existing models. Similarly, data-driven methods allow deepening the knowledge by analyzing existing data. Therefore, to develop improved predictive models, one needs to pursue a balance between both data-driven and simulation-driven approaches, keeping data as a common pivot. This can be done by using Machine Learning (ML) which is a powerful tool to extract knowledge directly from the data and provide complementary information to the previously developed physics-based models. The main advantage of ML methods is their ability to process massive and complex structure data sets that are difficult to be processed by traditional data-processing methods. Therein lies also the main disadvantage of ML methods i.e., they need massive amounts of data that often are not available. In this study however we take advantage of a new physics-based model of the earthquake fault system of the Southwest Iceland transform zone and generate synthetic, but physically realistic, finite-fault earthquake catalogues. For each earthquake in the catalogues we simulate seismicground motion parameters at a large hypothetical station network thus generating a massive parametric dataset of synthetic seismic data from earthquakes of magnitude 5 to 7. We will apply multiple types of new generation machine learning techniques such as deep neural network (DNN), deep Bayesian neural networks (DBNN) and deep Gaussian processes (DGP) to investigate the ability and efficiency of the methods in capturing the characteristics of the synthetic dataset in terms of key parameters e.g., ground motion amplitudes, ground motion attenuation versus source-to-site distance and site effects and independent parameters such as earthquake magnitude, fault extend and depth, etc. The ML methods will be trained using a procedure known as greedy layer-wise pretraining where each layer is initialized via the unsupervised pretraining, and the output of the previous layer can be used as the input for the next one. A typical advantage of these pre-trained networks compared to the other deep learning models is that the weights initialization renders the optimization process more effective, providing faster convergence by initializing the network parameters near a convergence region. This can help to avoid underfitting/overfitting problems when the training samples are highly correlated. The results will provide a new insight into the efficiency and usefulness of ML methods on synthetic seismic datasets with implications for their use on actual and more sparser datasets.