

Unsupervised learning used in automatic detection and classification of ambient-noise recordings from a large-N array

Michał Chamarczuk, Yohei Nishitsuji^{2,3}, Michał Malinowski¹, Deyan Draganov²

¹Institute of Geophysics PAS, Warsaw, Poland, ²Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands, ³Petro Summit E&P Corporation, Tokyo, Japan.

We present a method for automatic detection and classification of seismic events from continuous ambient-noise (AN) recordings using an unsupervised machine-learning (ML) approach. We combine classic and recently developed array-processing techniques with ML enabling the use of unsupervised techniques in the routine processing of continuous data. We test our method on a dataset from a large-number (large-N) array, which was deployed over the Kylylahti underground mine (Finland), and show the potential to automatically process and cluster the volumes of AN data. Automatic sorting of detected events into different classes allows faster data analysis and facilitates the selection of desired parts of the wavefield for imaging (e.g., using seismic interferometry) and/or monitoring.

First, using array-processing techniques, we obtain directivity/location, velocity, and frequency representations of AN data. Next, we transform these representations into vector-shaped matrices. The transformed data are input into a clustering algorithm (called k-means) to define groups of similar events, and optimization methods are used to obtain the optimal number of clusters (called elbow and silhouette tests). We use these techniques to obtain the optimal number of classes that characterize the AN recordings and consequently assign the proper class membership (cluster) to each data sample. For the Kylylahti AN, the unsupervised clustering produced 40 clusters. After visual inspection of events belonging to different clusters that were quality-controlled by the silhouette method, we confirm the reliability of 10 clusters with a prediction accuracy higher than 90%. The obtained division into separate seismic-event classes proves the feasibility of the unsupervised ML approach to advance the automation of processing and the utilization of array AN data. Our workflow is very flexible and can be easily adapted for other input features and classification algorithms.