Hierarchical exploration of single station seismic data with unsupervised learning

René Steinmann, Leonard Seydoux, and Michel Campillo

Universite Grenoble Alpes, ISTerre, France (rene-steinmann@web.de)

Seismic datasets contain an enormous amount of information and a large variety of signals with different origins. We usually observe signatures of earthquakes, volcanic and non-volcanic tremors, rockfalls, road and air traffic, atmospheric perturbations and many other acoustic emissions. More and more seismic sensors are deployed worldwide and record the seismic wavefield in a continuous fashion, generating massive volumes of data that cannot be analyzed manually in decent times. Therefore, identifying classes of signals in seismic data with automatic strategies is a crucial stage towards the understanding of the underlying physics of geological objects. For that reason seismologists have developed different tools to detect and classify certain types of signals. Recently, machine learning gained much attention due to its ability to recognize patterns. While supervised learning is a great tool for detecting and classifying signals within already-known classes, it cannot be used to infer new classes of signals, and can be strongly biased by the labels we impose. We here propose to overcome this limitation with unsupervised learning. In this study, we present a new way to explore single-station continuous seismic data with a dendrogram produced by agglomerative clustering. Our method is motivated by the idea that labels in a seismic data set follow a hierarchical order with different levels of details. For example earthquakes belong to the larger class of stationary signals and can be also divided into subclasses with different focal mechanism or magnitudes. We first use a scattering network (a convolutional neural network that makes use of wavelet filers) in order to extract a multi-scale representation of the continuous seismic waveforms. We then select the most meaningful features by means of independent component analysis, and apply an agglomerative clustering on this representation. We finally explore the dendrogram in a systematic way in order to explore the different signal classes revealed by the strategy. We illustrate our method on seismic data continuously recorded in the vicinity of the North-Anatolian fault, in Turkey. During this time period, a seismic crisis with more than 200 micro-earthquakes occurred, together with many other anthropogenic and meteorological events. By exploring the classes revealed by the dendrogram with *a posteriori* signal features (occurrence, within-class correlations, etc.) we show that the strategy is capable of retrieving the seismic crisis as well as signals related to anthropogenic and meteorogical activities.