Machine learning for the quality control of seismic data

Mathieu CHAMBEFORT – Mines Paristech – Centre de géosciences

Our goal in seismic imaging is to recover the velocity field and the reflectivity of the subsurface from seismic traces in order to have access to more rock properties via guided well inversion. Knowing our emitted source, we can describe recorded seismic traces as follows (figure 1):

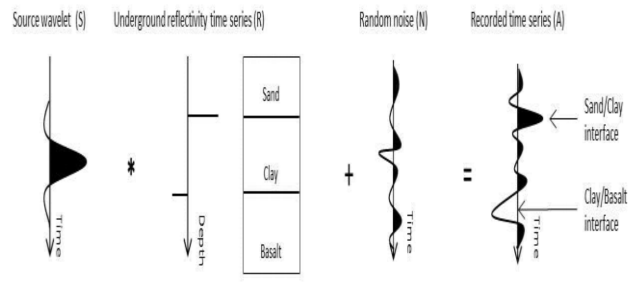

Figure1: Seismic trace description

We can rewrite this figure as:

$$d(t) = s(t) + n(t)$$

and

$$d(t) = o(t) * r(t) + n(t)$$

with d(t) the data, s(t) the signal, n(t) the noise (additive or destructive), r(t) the earth reflectivity (assumed as a comb of Dirac) and o(t) the emitted source. Earth reflectivity can be estimated by removing the noise.

In seismic imaging, the data are in a "shot points" collection (figure2)

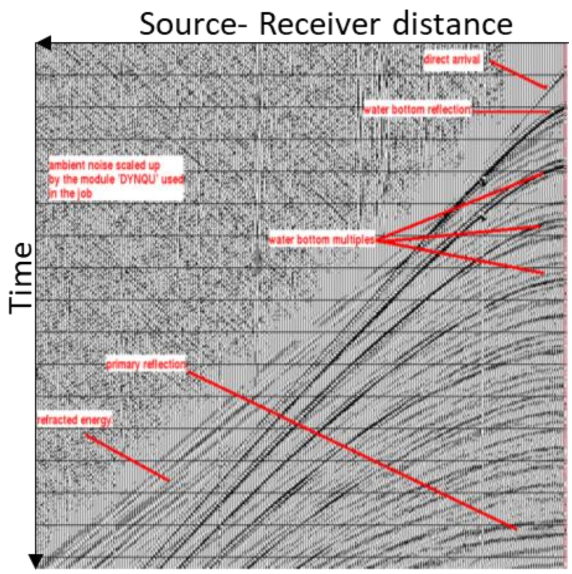

Figure2: Recorded seismic shot point

The different hyperbola follow the Normal Move Out equation (NMO):

$$t = 2\sqrt{\frac{{t_o}^2}{4} + (\frac{x}{V})^2}$$

with t_0 the recorded time at 0 offset (where source and receiver location are similar), x the offset distance and V the velocity of the media (that we need to retrieve). Signals that do not follow this equation are the noises coming from different sources than the seismic one.

Each shot point can be defined as an image with as many pixels as sample and receivers:

$$d(y, t|X_k)$$

with X_k the k-th shot point source, y the position of the receiver and t the time of the record.

The main types of noise that need to be removed to image the underground properly are spikes, swell noise, random noise, linear noise or interference noise.

Currently, the seismic processing is sequential, applying one process per kind of noise.

For a given project, we have many processing steps and for each step we need to ensure the good application of a process for the entire dataset (sometimes hundreds of Terabytes).

This is so far mostly a human work and my goal in this PhD thesis is to develop an algorithm based on machine learning techniques, which outputs multi labels for each processing step:

- 0: Process works well
- 1: Anomalies created by the process
- 2: Remaining noise (the process is too mild)
- 3: Primary leakage (the process is too harsh)

My idea is to associate a "trust" score to this label, which will greatly help geophysicists to control the quality of a dataset.