September 26, 2023

Scalable subsequence similarity search in seismic time series

Context of the internship

The objective of seismic monitoring is to detect and characterize the lowest seismic events, usually in real time, and trigger an alert when necessary (e.g., for tsunami early-warning, strong earthquakes, etc.). To perform this task automatically, fast and reliable data processing is required to extract as much information as possible from all the available time series. Such processing includes quality control (hole or spike detection, etc.), event detection, event localization, measurement of features (computation of magnitude, etc.) or classification (e.g., anthropic versus natural events), and others. To fulfill their mission, seismology institutes - like CEA - collect, store and analyze large amounts of time series, growing exponentially in size. However, due to the mismatch between data volume and software performance, it becomes increasingly difficult to manage, process and analyze all these data within an acceptable time frame.

Objective of the internship

To perform detection, seismic signal comparison -or template matching– is commonly used: the principle is to sequentially search for a known reference signal in the continuous flow of signals, by comparing time series using the cross-correlation function. A signal being like a “fingerprint” for earthquakes, if it matches a past observation, it means that a similar event has occurred.

In order to speed up the detection without hurting accuracy, a new algorithm has been developed, as a part of a post-doctoral job directed jointly by the LIPADE and the CEA. The method, called L-MESSI, is an efficient indexing method, derived from the MESSI [1] algorithm, to perform similarity search on long seismic signals. It has been evaluated on a limited seismic dataset composed of earthquakes recordings, and showed promising results.

The objective of the internship would be:

- Add support for index updates
- Enable concurrent query answering
- Build a user interface
- Test on very large seismic data coming from the CEA earthquakes database, to assess processing performance

Applicant Profile

Master's, or Engineering school student in applied mathematics / data science / computer science. Very good knowledge of C. Knowledge of Python and web application development would be a plus.

Contract duration

6-months internship, with a start date anywhere between October 2023 - March 2024.

Contract location

The internship will take place in the facilities of the CEA (6 months), with frequent displacements to the LIPADE.

CEA - Centre DAM Ile de France, Département Analyse surveillance Environnement (http://www-dase.cea.fr), Laboratoire Détection et Expertise des Evénements Géophysiques, Bruyères-le-Châtel, 91297 Arpajon Cedex

Application

Send CV and application letter to M. Yoann Cano yoann.cano@cea.fr and M. Themis Palpanas themis@mi.parisdescartes.fr

References

[1] Botao Peng, Panagiota Fatourou, Themis Palpanas. Fast Data Series Indexing for In-Memory Data. International Journal on Very Large Data Bases (VLDBJ) 2021 (https://helios2.mi.parisdescartes.fr/~themisp/publications/vldbj21-messi.pdf)