Improving earthquake detection with data mining and machine learning

Karianne Bergen[1] and Gregory Beroza[2]

[1] Institute for Computational and Mathematical Engineering, Stanford University
[2] Department of Geophysics, Stanford University

Extracting weak earthquake signals from continuous waveform data recorded by sensors in a seismic network is a critical and challenging task in seismology. In recent years, the availability of massive seismic data sets and new algorithmic advances in "big data" and artificial intelligence have created opportunities for new research to advance the state-of-the-art in earthquake detection algorithms. Machine learning is not new to seismology; techniques like neural networks and hidden Markov models have been used in seismic signal analysis studies for decades. But recent advances in machine learning have transformed fields like image recognition, natural language processing and recommendation systems. This has sparked renewed interest among the seismology community in integrating modern machine learning and data mining into earthquake detection and monitoring pipelines on a larger scale.

In this talk, I will present a brief overview of machine learning and discuss current research and opportunities for improving earthquake detection. I will focus on how machine learning and data mining techniques can enable new data-driven discovery in large seismic data sets. As an illustrative example, I will describe a new computationally efficient method for large-scale earthquake detection that takes inspiration from technology for rapid audio identification. Our method, Fingerprint and Similarity Thresholding (FAST), enables waveform-similarity-based earthquake detection over a seismic network in long duration continuous seismic data (Yoon et al., 2015; Bergen and Beroza, 2018). FAST leverages locality-sensitive hashing, a data mining technique for efficiently identifying similar items in large data sets, to detect similar waveforms without templates waveforms (training data). FAST offers a new detection capability: the ability to discover new events with unknown sources in more than 10 years of continuous data using waveform similarity (Rong et al., 2018). I will close the talk with recommendations for the research community, based on our experience developing the FAST algorithm, to drive progress and enhance adoption of new data mining and machine learning-based detection methods.

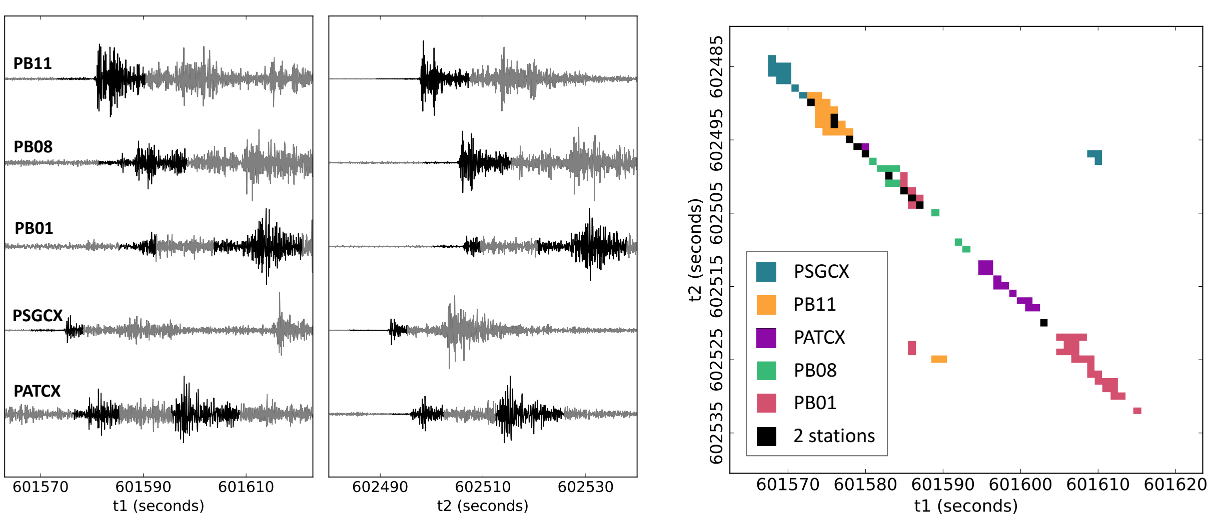

Fig. 1: In the multi-station extension of the FAST earthquake detector, the task of identifying the set of single-station observations that correspond to arrivals from a single source event, usually referred to as "association," is performed on pairs of events (with similar waveforms) and is equivalent to identifying offset diagonal clusters in a sparse matrix (Bergen and Beroza, 2018).