Efficient blind search for similar-waveform earthquakes in years of continuous seismic data

Clara Yoon, Gregory Beroza
Stanford Geophysics

Kexin Rong, Hashem Elezabi, Peter Bailis, Philip Levis
Stanford Computer Science

Karianne Bergen
Stanford ICME

AGU Fall Meeting, 2017-12-12
Seismology has big data sets

Large-N: Dense seismic networks (1000’s of stations)

Large-T: continuous seismic data over long time periods

10 years
Seismology needs new scalable methods to extract information from massive data volumes.

- Computing power, parallel processing
- Memory
- Disk space
- Algorithms
Application: Detect small earthquakes missing from catalog

75 catalog earthquakes, $1.2 < M_L < 2.9$

14,604 detected earthquakes, $-1.5 < M_L < 2.9$

Yoon et al. (2017)
Waveform similarity allows detection of smaller earthquakes

*Template waveform:* known

*Similar event waveform:* previously unknown

**Informed search:** Template Matching
Comprehensive, exhaustive search for earthquakes with similar waveforms

Large-T: continuous seismic data over long time periods

Uninformed (blind) search
Efficient search for similar items in large databases

Music  Webpages  Videos

Fingerprint And Similarity Thresholding (FAST)

Earthquakes  Yoon et al. (2015)
              Bergen et al. (2016)
FAST Detection Pipeline

Data → Preprocessing → Feature Extraction → Efficient Similarity Search → Post-processing → Detection Results

Continuous seismic data (1 channel)

Fingerprints should be discriminative

Similar Waveform → Similar Fingerprint

Dissimilar Waveform → Dissimilar Fingerprint

Binary Fingerprint

(3 x 10^8 fingerprints for 10 years)
FAST Detection Pipeline

- **Data**
- **Preprocessing**
- **Feature Extraction**
- **Efficient Similarity Search**
- **Post-processing**
- **Detection Results**

Fast approximate similarity search

- **MinHash** and
- **Locality Sensitive Hashing**

Query

waveform

??

Similar
waveform

!!
FAST Detection Pipeline

Data → Preprocessing → Feature Extraction → Efficient Similarity Search → Post-processing → Detection Results

FAST output: sparse similarity matrix

9 x 10^{16} matrix
(<0.0001% elements filled)
Detect earthquakes over a **seismic network**

Earthquake pair at different stations: consistent inter-event time $dt$
Reduce false detections

Poster **S43A-0828**, Thursday PM, K. Bergen
Detecting Earthquakes over a Seismic Network using Single-Station Similarity Measures

*Bergen and Beroza (2017)*
Many active faults near Diablo Canyon nuclear power plant

Planned shutdown in 2025

Can uninformed large-T search find additional seismicity to constrain active fault structures?

M 6.8 possible on Shoreline Fault? (Hardebeck, 2013)
Blind search for similar-waveform earthquakes in a decade (2007/06-01 – 2017/10/24) of continuous seismic data

Sampling rate 100 Hz (500 GB) -> decimate to 25 Hz (100 GB)

Channel-specific bandpass filter, between 3 and 12 Hz: remove repeating noise

11 stations, 27 channels
Blind search detection results

3,957 catalog earthquakes, 597 new local earthquakes
We detect catalog earthquakes all over California.
San Simeon earthquake (M6.5, 2003) aftershocks?

San Andreas Fault (creeping)

Kettleman Hills blind thrust

Rinconada Fault

Hosgri Fault

Regional catalog seismicity: dominated by other sources >50 km away
New detected event locations*

- Within 50 km of network
- Offshore south of network

* VELEST (Kissling et al., 1994); 1D velocity model (McLaren and Savage, 2001); Vp/Vs = 1.66 (Hardebeck, 2010)
San Simeon earthquake (M6.5, 2003) aftershocks?

San Andreas Fault (creeping)

Hosgri Fault

Rinconada Fault

Kettleman Hills blind thrust

Power Plant

Stations (detection)

Stations (location)

Cities/Towns

Detected catalog events: Magnitudes

New detected local events: Magnitudes

-0.2 ≤ M_L ≤ 2.4

Magnitude of new detected events
274 detected catalog events:
Low rate of local earthquake activity
261 missed catalog events (not detected): Almost all outside seismic network
355 new detected local events: will help discern active fault structures
Yes, uninformed large-T similarity search finds unknown earthquake sources

Local seismicity 2007-2017

LEGEND
Detected catalog events: Magnitudes
Missed catalog events: Magnitudes
New detected local events: Magnitudes

Stations (detection)
Stations (location)
Power Plant
Cities/Towns
Faults (USGS Quaternary database)
New detected events should provide important constraints on active structures near Diablo Canyon.

Historical catalog seismicity, 1913-2007

LEGEND
- Historical catalog events: Magnitudes
- Stations (detection)
- Stations (location)
- Power Plant
- Cities/Towns
- Faults (USGS Quaternary database)
Rapid progress in FAST earthquake detection

Continuous data duration (years)

Calaveras Fault, California (1 week)
1.5 hours

Yoon et al. (2017), JGR
Guy-Greenbrier, Arkansas (3 months)
5 days

AGU 2017
Diablo Canyon, California (10 years)
16 hours (parallel, new Python-C++ software)
Summary

• New FAST software allows uninformed search for similar-waveform earthquakes in large-T (10 years) continuous seismic data
  – Applicable to any seismic network

• Discovered ~600 uncataloged earthquakes (-0.2 < $M_L < 2.4$) near Diablo Canyon nuclear power plant and offshore central California
  – More complete picture of seismicity on nearby active faults
Future Work

• Open source FAST software release with user guide: planned for 2018
  – We invite the seismology community to use FAST

• Explore earthquake detection in other continuous data sets (months to years, 6 to 40 stations)
  – Saudi Arabia
  – New Zealand
  – Ocean Bottom Seismometer deployments
  – Your data set here?