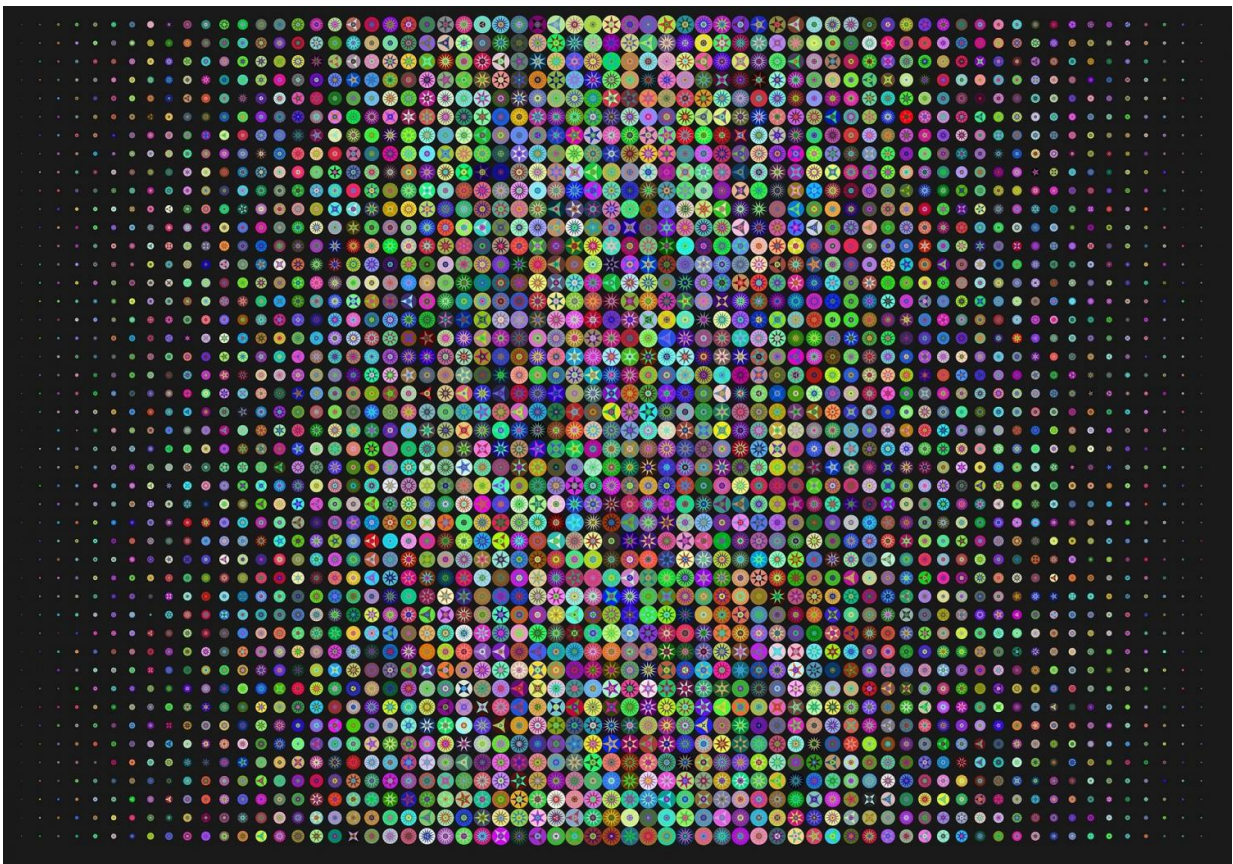# New big data algorithms improve earthquake detection; monitor livestock health and agricultural pests

November 27 2019, by Holly Ober

Credit: CC0 Public Domain

Two new algorithms could help earthquake early warning systems buy

you a few extra seconds to drop, cover, and hold on before the ground begins to shake.

Computer scientists at the University of California, Riverside have developed two algorithms that will improve earthquake monitoring and help farmers protect their crops from dangerous insects, or monitor the health of chickens and other animals. The algorithms spot patterns in enormous datasets quickly, with less computing power and lower cost, than other methods and have been used to improve earthquake detection, monitor the insect vector Asian citrus psyllid, and evaluate the feeding behavior of chickens.

Big data, big problems

Sensors, such as seismic sensors, which automatically record events that happen repeatedly over a period of time, have a problem. They gather so much data that it's hard to spot patterns. Time series analysis remedies this by looking for other examples of a sample sequence within a dataset, usually using graphics processing units, or GPUs. But for very large datasets this becomes impractical because it requires too many GPUs, which increases the cost.

Zachary Zimmerman, a doctoral student in computer science in the Marlan and Rosemary Bourns College of Engineering, built on an algorithm previously developed by co-author and professor of computer science Eamonn Keogh to handle extremely large datasets and ran it on 40 GPUs hosted on the Amazon Web Services cloud.

The algorithm, called SCAMP, sorted nearly two years of seismic recordings from California's Parkfield Fault, a segment of the San Andreas Fault located near the town of Parkfield, in just 10 hours, at a reasonable cost of about $300, and discovered 16 times more earthquakes than were previously known.

"It is difficult to overemphasize how scalable this algorithm is," Keogh said. "To demonstrate this, we did one quintillion—that's 1 followed by 18 zeros—pairwise comparisons of snippets of earthquake data. Nothing else in the literature comes within one-tenth of a percent of that size."

Identifying earthquakes isn't always easy

"The most fundamental problem in seismology is identifying earthquakes at all. There have been a number of methodological improvements by seismologists applying strategies from computer science to look for similar patterns," said co-author Gareth Funning, an associate professor of seismology. "The big advance here is that the dataset you can manage is way, way bigger. When we're looking at seismic data we used to think we were doing well comparing everything in a two-month time window."

Other methods of earthquake detection require the algorithm to find sequences that match a known earthquake. The UC Riverside method instead compares everything within a given time and thus can identify earthquakes that don't necessarily match one given as a model.

For example, their analysis of the Parkfield data discovered subtle, low-frequency earthquakes underneath the San Andreas fault. Sequences of these earthquakes, also known as nonvolcanic tremors, accompany deep, slow movements of tectonic plates.

Flurries of low-frequency earthquakes have occasionally preceded massive earthquakes, like the one in Japan 10 years ago. Better detection of low-frequency earthquakes could help improve forecasts of the largest earthquakes and also help scientists better monitor movements of tectonic plates.

From earthquakes to chickens and insect pests

The SCAMP algorithm can also detect harmful agricultural pests. Keogh attached sensors that recorded the motions of insects as they sucked juices out of leaves and used the algorithm to identify Asian citrus psyllid, the insect responsible for devastating citrus crops by spreading the bacteria that causes Huanglongbing, or citrus greening disease. He also used the algorithm to analyze a dataset from accelerometers, which measure various kinds of movements, attached to chickens over a period of days. SCAMP then identified specific patterns related to feeding and other behaviors.

SCAMP has one limitation, however.

"SCAMP requires you to have the entire time series before you search. In cases of mining historic seismology data, we have that. Or in a scientific study, we can run the chicken around for 10 hours and analyze the data after the fact," said co-author Philip Brisk, an associate professor of computer science and Zimmerman's doctoral advisor. "But with data streaming right off the sensor, we don't want to wait 10 hours. We want to be able to say something is happening now."

Faster real-time earthquake detection

Zimmerman used the billion datapoints, called a matrix profile, generated by SCAMP's analysis of the Parkfield fault data to train an algorithm he called LAMP. LAMP compares the streaming data to examples it has seen before to select the most relevant data as it comes off the sensor.

"Having the matrix profile available to you at the sensor means that you can immediately know what's important and what's not. You can do all

your checks in real time because you're just looking through the important bits," Zimmerman said.

The ability to more quickly interpret seismic data could improve earthquake warning systems that already exist.

"With earthquake early warning, you're trying to detect things at monitoring stations and then forward the information to a central system that evaluates whether or not it's a big earthquake," said Funning. "A setup like this could potentially do a lot of that discrimination work before it's transmitted to the system. You could shave time off the computation required to determine that a damaging event is in progress, buying people a couple extra seconds to drop, cover, and hold on."

"A couple of seconds is huge in earthquake early warning," he added.

The paper on SCAMP, "Matrix Profile XIV: Scaling Time Series Motif Discovery with GPUs to Break a Quintillion Pairwise Comparisons a Day and Beyond," was presented at the ACM Symposium on Cloud Computing November 20–23, 2019 in Santa Cruz. Authors are Zachary Zimmerman, Kaveh Kamgar, Nader Shakibay Senobari, Brian Crites, Gareth Funning, Philip Brisk and Eamonn Keogh.

The paper on LAMP, "Matrix Profile XVIII: Time Series Mining in the Face of Fast Moving Streams using a Learned Approximate Matrix Profile," was presented at the 2019 IEEE International Conference on Data Mining held in Beijing earlier in November. The authors are Zachary Zimmerman, Nader Shakibay Senobari, Gareth Funning, Evangelos Papalexakis, Samet Oymak, Philip Brisk, and Eamonn Keogh.

Provided by University of California - Riverside