AI-based Earthquake Signal Detection and Processing

S.Mostafa Mousavi, PI, Stanford University
Gregory C. Beroza, CoPI, Stanford University

Seismology – the study of earthquakes - is a data rich and data-driven science. Every day there are about fifty earthquakes worldwide that are strong enough (magnitude > 2.5) to be felt locally and a multitude of smaller earthquakes (magnitude < 2.5), that are too weak to be felt, but that are readily recorded by modern instruments. However, earthquakes are not the only sources that generate seismic waves. Many other sources such as explosions, landslides, oceanic waves, planes, helicopters, trains, wind, thunderstorms, traffic, and people, generate ground motions that are recorded by thousands of seismic instruments that are continuously operated by seismic monitoring networks around the world. Hence, there is an enormous amount of seismic data generated every day. The massive amount of data highlights the need for more efficient and powerful tools for data processing and analyzing. The main challenge is the efficient extraction of as much useful information as possible and by doing so to gain new insights from seismic data. This is where rapidly evolving machine learning techniques have the potential to play a key role to reach a deeper understanding of coupled processes on the earth through data-driven discoveries.

IMPACT:

How does your project align with AI for Earth areas of focus?

Some of the effects of the climate change such as the seasonal changes in the thickness of the ice sheet and rapid movements of glaciers can be tracked using seismic waves. Glaciation is related to tectonic processes, and as a result, to earthquakes. The retreat of a glacier can reduce stress loads on Earth's crust underneath, impacting the movement of subsurface materials and faults. Such responses to global warming can be tracked through seismic monitoring of glacial earthquakes. On the other hand, even small stress changes due to climate-like forcings can causes micro-earthquakes. The main focus of our study is to enable monitoring of smaller earthquakes and subtle changes in seismic data through building more robust and efficient AI-powered algorithms.

How does your project transform the way we address environmental challenges?

Tracking the climate-related phenomena through unconventional data such as seismic data, not only can provide us a new monitoring approach, but also can bring us a new understanding of coupled earth processes. This can help us to understand the other factors that might have an influence on earthquakes. Massive amounts of seismic data continuously being recorded all around the world, state-of-the-art machine learning techniques, and Microsoft Azure's cutting-edge resources can bring us an unprecedented opportunity to explore what is usually left unused in seismic recording and hopefully gain new insights. Our project aims to improve seismic data processing through building AI models using a massive labeled dataset. These powerful models can help us to address environmental challenges such as man-made (induced) earthquakes due to wastewater disposal by providing a more efficient and effective monitoring tool as well.

How will the results of your project be leveraged and beneficial to a community of users?

The outputs of the project and produced models can be used for routine earthquake monitoring purposes as well. This would be important from a seismic hazard point of view. Currently, we are helping USGS, North California Seismic Datacenter, and Texas Seismic Network implementing AI-based models for automatic processing of their seismic data. The proposed project would result in more powerful models that hopefully would have applications in other national and international seismic monitoring agencies.

1-2 SENTENCE SUMMARY:

In overview, work will consist of applying a currently developed network to collected local seismograms by the offerors to the problem of detection of small seismic events. Current methods include the application of a very deep neural network, for lowering the detection threshold of seismic events within local distances and automatic data labeling through transudative learning.

PROJECT SUMMARY:

What challenges are you addressing? What is your proposed solution?

Currently, only earthquakes larger than magnitude 2 are completely documented by most of the monitoring agencies. This is due to the difficulty of detecting and locating of smaller events manually and ineffectiveness and inefficiency of available automatic algorithms. However, the numerous micro-earthquakes that left unused contain valuable information about the environment and climate-related phenomena. This project we will develop efficient and effective algorithms capable of monitoring such small earthquakes using AI. Our previous studies suggest that the AI-based method can significantly improve seismic data processing and earthquake monitoring. We plan to use these results and successful prototypes to build large-scale models and APIs using Microsoft Azure resources.

How does your project build on existing research?

We have studied individual elements of the pipeline and built successful prototypes for seismic signal discrimination (Mousavi et al. 2016 and Mousavi et al 2019a), detection (Mousavi et al 2019b), denoising (Zhu et al 2019), magnitude estimation (Mousavi and Beroza, 2019a), and location estimation (Mousavi and Beroza, 2019b).

What datasets will you use for your analyses?

We will use the STanford EArthquake Dataset (STEAD) (Mousavi et al. 2019c) for building our preliminary models. STEAD is a large-scale global dataset of earthquake and non-earthquake signals. It contains more than 1.2 million waveform associated with about 450,000 earthquakes and noise that occurred between January 1984 and August 2018. All of the waveforms were recorded within 300 km from the source and the majority of them are small earthquakes (magnitude <2.5). Waveforms are 1 minute long with a sampling rate of 100 HZ. 35 labels (mostly manually determined) are provided for each earthquake waveform including information about the recording station, recorded earthquake, and station-source parameters. We will use STEAD for building preliminary models that will be applied for automatic labeling of signals associated with larger earthquakes (magnitude > 2.5) recorded in larger distance ranges (> 300 km). This will be done through a combination of transductive learning and auxiliary QC algorithms. The expanded dataset will be used for building final models.

Approximate timeline for key project milestones:

We propose to carry out this research over a period of one year. Work can be broken down into three specific tasks as follows:

1. *April – June 2020:* Setting and writing up codes for implementing prototype networks on the Azure environment. This includes testing simplified models on a subset of the dataset, hyperparameter tuning, and building preliminary models.

2. *July 2020– December 2020:* Applying preliminary models for expanding the dataset through automatic labeling and QC. We have developed a station data bank containing information on more than 27,000 stations. We will retrieve event and phase information from North California, Southern California, New Madrid, Pacific Northwest, USGS, IRIS, ISC and a few temporary experiments. These data will be used to calculate station-event parameters and form the metadata for the global data set. Based on this metadata, waveform data we will retrieve data from the IRIS data center and develop and apply a processing pipeline to label and pre-process the data.

3. *January 2021– March 2021:* Scaling up the models and conducting the training/evaluations on the full dataset. This includes carrying out precision and sensitivity experiments and applying the trained models on a few months of continuous data recorded in Alaska and Greenland.

TECHNICAL COMPONENTS

What Azure and machine learning offerings will you use?

We propose a grant of Azure services to use Azure machine learning, enabling us to significantly improve the efficiency training large-scale models. The services we request include Azure Batch AI, VM with GPU, and Storage.

Justify the grant amount your project will need in Azure credits:

We request an Azure grant of $10,000. Please see the detailed budget justification below:

Service type	Custom name	Region	Description	Estimated Cost per month
Storage		West US	Block Blob Storage, Blob Storage, LRS Redundancy, Hot Access Tier, 5 TB Capacity, 100,000 Write operations, 100,000 List and Create Container Operations, 100,000 Read operations, 1 Other operations. 1,000 GB Data Retrieval, 1,000 GB Data Write	$107.64
Storage		West US	File Storage, Standard, General Purpose V1, LRS, 3 TB, 500 Snapshots, 1000 Write operations, 1List, 1 Other operations.	$229.34
VM with GPU	ND6	West US	ND6 * 200 hours	$498.60
			Support	$0.00
			Monthly Total	**$835.58**
			Annual Total	**$10,026.96**

Do you have the resources and technical skills to complete your project successfully?

PI has technical skills for conducting the project. We have experience in building and implementing various advanced neural network architecture and performing large-scale training processes. We also have experience working with other cloud platforms such as GCP. Moreover, we are open to collaborate with Microsoft AI team for implementing the models into Azure.

References:

Mousavi, S. M., Horton, S. P., Langston, C. A., & Samei, B. (2016). Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression. Geophysical Journal International, 207(1), 29-46.

Mousavi, S. M., Zhu, W., Ellsworth, W., & Beroza, G. (2019a). Unsupervised clustering of seismic signals using deep convolutional autoencoders. IEEE Geoscience and Remote Sensing Letters, 16(11), 1693-1697.

Mousavi, S. M., Zhu, W., Sheng, Y., & Beroza, G. C. (2019b). CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection. Scientific reports, 9(1), 1-14.

Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and decomposition using deep neural networks. IEEE Transactions on Geoscience and Remote Sensing, 57(11), 9476-9488.

Mousavi, S. M., & Beroza, G. C. (2019a). A Machine-Learning Approach for Earthquake Magnitude Estimation. Geophysical Research Letters.

Mousavi, S. M., & Beroza, G. C. (2019b). Bayesian-Deep-Learning Estimation of Earthquake Location from Single-Station Observations. arXiv preprint arXiv:1912.01144.

Mousavi, S. M., Sheng, Y., Zhu, W., & Beroza, G. C. (2019c). STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. IEEE Access.