Machine learning for earthquake early warning

Yuyan Chen, Santhoshi Ravichandran, Ashwini Rajaram, Nikhil Reddy

Sept. 17 2023

1 Introduction

Early-warning alerts are essential for earthquake mitigation. Having a solid early warning system can save lives, reduce injuries, and better coordinate emergency response organizations. Currently, early-warning alerts are sent around three to four seconds after an earthquake starts [1]. In this project, we aim to use machine learning to predict the occurrence of earthquakes using waveform traces and improve the early warning system by sending the warning before its occurrence. We chose a 10-second window as our training samples only consist of 120-second waveforms. In future work, we will want to use time series data of a much longer duration and predict the occurrence of an earthquake at an even earlier stage.

2 Dataset

We chose INSTANCE, the Italian seismic dataset for machine learning as our training set [2]. It includes 54,008 earthquakes for a total of 1,159,249 3-channel waveform recordings and 132,330 3-channel noise waveform recordings from 620 seismic stations. Each waveform has a length of 120 seconds, sampled at 100 Hz. The dataset comprises five distinct station channels, namely EH, EN, HH, HL, and HN. Each waveform recording is exclusively from one of these station channels. It is essential to differentiate these station channels from 3-channels contained in each waveform recording. For our use case, we only use the first 10 seconds of each recording since for most earthquake recordings, the P-wave onset, or the start of the earthquake, happened after the first 20 seconds.

3 Methodology and results

We trained three models, two random forest (RF) models, and a neural network (NN). One RF model used recordings from the HN channel only. The other RF model and the NN used recordings from different channels. Since recordings from different channels have different ranges, we normalized the data when training the NN. We achieved 87% accuracy for RF trained with only the HN channel and 61% accuracy for RF trained with mixed channels. We can only achieve 51% accuracy for NN, since our training set is extremely small, consisting of only 800 sample points for each class. Also, due to the time limit, we are unable to fine-tune our model and make use of a pre-trained neural network for transfer learning.

4 Future work

Due to time and resource limits, we are only able to train two types of models instead of having a comprehensive comparison of models with different architectures. Right now, we only have data that consists of 120-second waveforms. For future work, we want to investigate the dataset to see if we can concatenate the waveforms to have a longer time frame, for example, 10 minutes. If so, we would like to use the sequence that is at least 1 minute before the P-wave onset. In this sense, we will have a much earlier warning system and a much longer response time for earthquakes.

5 Pathway to impact

Building a machine learning-driven earthquake early warning system requires collaboration from different domains, including seismologists, computer scientists, government, and emergency management organizations.

Besides early warning systems, other aspects including magnitude and epicenter estimation and tsunami warning are also essential for earthquake mitigation strategies. We want to further investigate the dataset, collaborate with domain experts, and combine the waveforms with other features to build a more comprehensive dataset that can be used for the prediction or estimation of the aforementioned aspects.

References

[1] Richard M Allen et al. "The status of earth-
     quake early warning around the world: An
     introductory overview". In: *Seismological Re-
     search Letters* 80.5 (2009), pp. 682–693.

[2] Alberto Michelini et al. "INSTANCE–the Ital-
     ian seismic dataset for machine learning".
     In: *Earth System Science Data* 13.12 (2021),
     pp. 5509–5544.