Seismology - Summer Internship 2022 at GFZ Potsdam

Moshe Beutel

September 6, 2022

1 Overview

Seismic earthquake research is full of annotated data. Open data sets from seismographic stations contains millions of manually annotated interesting events and is open to seismographic research community. Efficient deep learning models developed in the last 4 years to deal with the following main tasks: *Earthquake detection, Phase Identification* and *Onset Time Picking* which correspond to the general algorithmic tasks of: Detection, Classification and estimation respectively.

2 Tasks

The tasks we are examining are the following:

Task	Input	Output	Metric
Event Detection	30s window of seismic waveform	Contains First Arrival	AUC (or F1)
Phase Identification	10s window of seismic waveform	Detrmine P or S	MCC
Onset Time Picking	10s window contains one S or P wave (known)	Determine Onset Time	RMSE

3 Terminology

- **P and S Waves** - Short for primary ad secondary waves. *Body Waves* are energy travelling through solid volumes and *Surface Waves* travel through free surfaces. The Body waves travel faster hence called primary or P Waves and the - slower - surface waves are called Secondary or S Waves.

- **Seismic Wave Equation** -

- **Arrival Time** - The time of first discernible motion of a seismic phase.

- **Picking** - Measuring (Estimating ???) the arrival time

4   Models

The following models were tested in the benchmark:

- BasicPhaseAE (Woollam et al., 2019)

- CNN-RNN Earthquake Detector (CRED; Mousavi, Zhu, et al., 2019)

- DeepPhasePick(DPP; Soto & Schurr, 2021)

- Earthquake transformer (EQTransformer; Mousavi et al., 2020)

- PhaseNet (Zhu & Beroza, 2019)

5   Limitation

The noted models, although preformed well on the given tasks using the defined metrics, are still limited in the view of real life applications like early warning scenarios.

- **Datasets Limitations** -

  - Uncertainties and nonuniqueness of manual labels owing to limited resolution, presence of noise, **different levels of expertise**, cognitive biases, and inherent ambiguity of tasks is a limiting factor. In image object classification for example tasks this is generally non-issue because normally most annotators would agree about pictures of cars,cats,tables and other daily life objects.
  - Not all seismic signals classes and typical noise are covered in the datasets - e.g. data from nodal seismometers at local distances, mine blasts, or volcanic signals.

- **Tasks Limitation** - The tasks defined above does not exactly represent real life scenarios where:

  - There are no defined time windows
  - More than one event may occur in a given time frame
  - The metrics defined does not take into account how early the tested algorithms would be able to identify an event onset

- In continous time setup the false positive rate needs to be significantly lower than in post-processing.

- Transfer Learning - It is yet unclear which datasets are most suitable for pretraining models

6 Research Question Formulation

The research question subjeect for this internship term deals with quantifying the uncertainty of a given model performence. The 6-week time frame is ofcourse not suitable to solve that big problem so the exact task is yet to define