CS671: Deep Learning and its Applications

Project Title: Earthquake Prediction using Time Series Seismic data

Group-13

Aman Khandelwal (B16007)
Amrendra Singh (B16010)
Bharat Lodhi (B16015)

Indian Institute of Technology Mandi

June 4, 2019

Contents

1. Introduction
   - Problem Statement
   - Earthquakes
2. Motivation and Challenges
3. Dataset
4. Proposed Methodology
5. Methodology Explored
   - Method 1: Tranformer Network
   - Method 2: GRU
   - Method 3: WaveNet as discriminative model
   - Method 4: Spectogram and CNN
   - Method 5: Stateful LSTM
6. Conclusion
7. References

Problem Statement

Forecasting earthquakes is one of the most important problems in Earth science because of their devastating consequences. Current scientific studies related to earthquake forecasting focus on three key points: when the event will occur, where it will occur, and how large it will be.

We will attempt to predict the time remaining before laboratory earthquakes occur from time series magnitude seismic data.

Earthquakes

An earthquake is the shaking of the surface of the Earth, resulting from the sudden release of energy in the Earth's lithosphere that creates seismic waves.

Earthquakes are caused mostly by rupture of geological faults, but also by other events such as volcanic activity, landslides, mine blasts, and nuclear tests. An earthquake's point of initial rupture is called its focus or hypocenter. The epicenter is the point at ground level directly above the hypocenter.

For understanding, there are some standard experiments using which earthquakes are simulated in laboratories using rocks. Data is collected from such experiment and used for analysis and then same learning is used to understand real earthquake behaviours. Collecting data of real earthquakes is both difficult and noise prone.

Motivation and Challenges

Despite considerable research efforts by seismologists, currently scientifically reproducible predictions cannot yet be made to a specific day or month.
If solved and the physics is ultimately shown to scale from the laboratory to the field, researchers will have the potential to improve earthquake hazard assessments that could save lives and billions of dollars in infrastructure.
But earthquake prediction poses a lot of challenges:

- Very little work has been done using seismic data. So firstly the proper feature engineering of the data had to be done.
- Dealing with acoustic data has been done in terms of for eg., audio data so the important work will be to translate the work done there to the corresponding seismic data.
- Filtering the noise from data, which can be due to physical surroundings external to the setup, irregularities in the instruments setup.

Dataset I

The train data consists of magnitude of *acoustic_data*(time series seismic data) sampled at 4MHz and the *time_to_failure* which is the time remaining before the failure occurs
This is the data obtained from laboratory earthquake experiments. The data consists of a series of magnitude values for the seismic readings.
The training data consists of 629 million of data points(time steps of acoustic data).
The test data consists of multiple files with segments of the time-series seismic data. We are to predict the time to failure for each of the file.

The following image shows the visualizations of a few test segments:

Dataset II

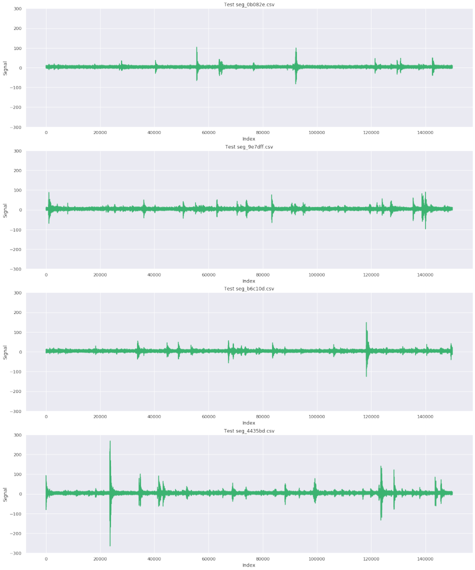

Figure: Data Visualization

Proposed Methodology I

The data we have is continuous and the output not only depend on the immediate input but also on the previous inputs which can be as far as nearly 150000 time steps so we need some model which can encode the important features from such large range of inputs.

So the model suited we came across for this purpose which is neither complex not computationally expensive was TRANSFORMER NETWORK [**?**]. In this network similar to autoencoder we've encoder and decoder unit where encoder encodes the input but here the input decoder receives is different.

Here for generating our output which is also continuous in nature we have as our input to decoder the encoded input as well as the previously generated output.

Proposed Methodology II

Use of previously generated output helps in more accurate prediction of future outputs and now we don't have to recur to previous inputs one by one and check which affects our output the most but we can our input all at once.

In our model firstly we have divided our data into the segments such that the between each segment the earthquake occurs. Our network architecture is Figure 5

Proposed Methodology III

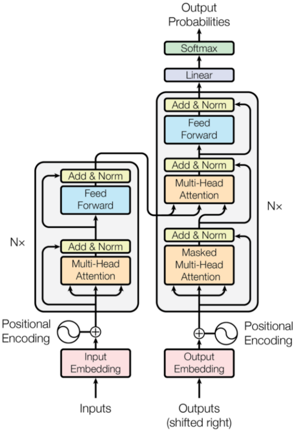

Figure 1: The Transformer - model architecture.

Figure: Architecture of the transformer network from the paper "Attention is all you need"

Proposed Methodology IV

In transformer architecture the left half is encoder and right half is decoder and the three inputs to the decoder can be seen. Resemblance of our architecture from the transformer architecture -:
Here our encoder part is the architecture upto the last convolutional layer is behaving like encoder and three inputs are fed to the decoder part which is after the Convolutional layer upto the last layer of the network. Also the time distributed layer was used so that to keep the remaining important info of the time steps separate as this wrapper applies a layer to every temporal slice of an input. Also as output is continous so softmax activation was used. Also at last global average pooling was used instead of more dense layers.

Methodology Explored 1: Transformer Network I

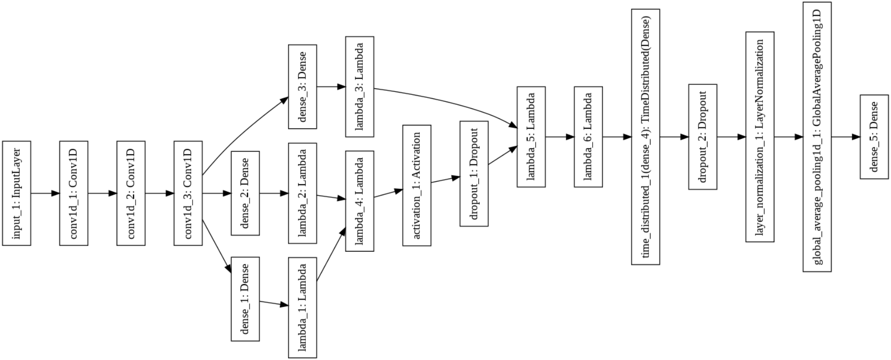

Figure: Network Architecture

Methodology Explored 1: Transformer Network II

Results
Public Mean absolute error = 1.6011
Private Mean absolute error = 2.6610

Methodology Explored 1: Transformer Network III

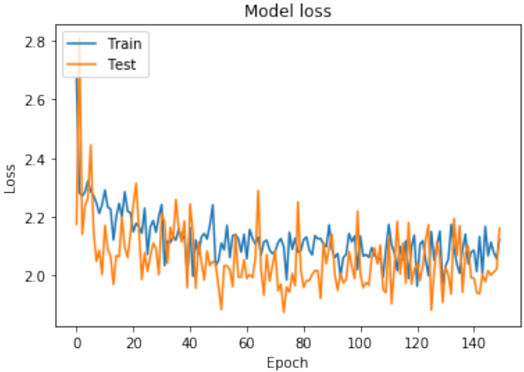

Figure: Model Loss

Methodology Explored 2: GRU I

In this approach, the train data was devided into sections of 150 time-steps. For each grouping 4 moments, viz. mean, min, max, std. dev. were computed and used.
Over these features, a GRU layer (cuDNNGRU) and two dense layers were applied.
In further experiments for the GRU(same for LSTM) we tweaked the model architecture and then the features in two ways. Firstly in the original paper it was mentioned that they have used the higher moments of the data as the features so insprired by that we applied the Conv-1D layers on the data so that network learns the important features(if moments are one of them) by itself and the MAE we got through it is foloowing: Public dataset MAE = 1.69
Private dataset MAE = 2.92
In another network we deliberately added the moments as the features and varied them in number and the best model MAE is following: Public Dataset MAE = 1.754

Methodology Explored 2: GRU II

Private Dataset MAE $= 2.784$

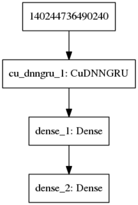

Figure: Network Architecture

Methodology Explored 3: WaveNet I

In WaveNet is typically a generative model for raw audio and can be used as discriminative one for phoneme recognition. Here, as we've some sort of pattern in data between each recurring earthquake and the temporal relationship is quite long(approximately 150000 time steps) so our best bet was to use architecture involving only convolutions and WaveNet was the one in which to increase the receptive field of our output we use dilated convolutions and also for barring the present output to depend on future inputs we've causal convolutions. Below is the architecture same as WaveNet that we've used.

Methodology Explored 3: WaveNet II

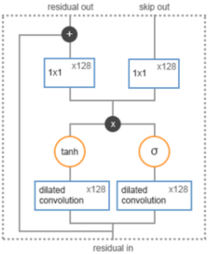

Figure: Network Architecture

Here we divided our time sequence in window of different time steps. We tried to two windows one of 4096 and other of 16384 time steps and with receptive field of 2048 and 8192 respectively with the help of dilations in order to so that network will learn some patterns But it didn't so much promising results and the lowest mean aboslute error achieved by the same was 2.384.

Methodology Explored 4: Spectogram and CNN I

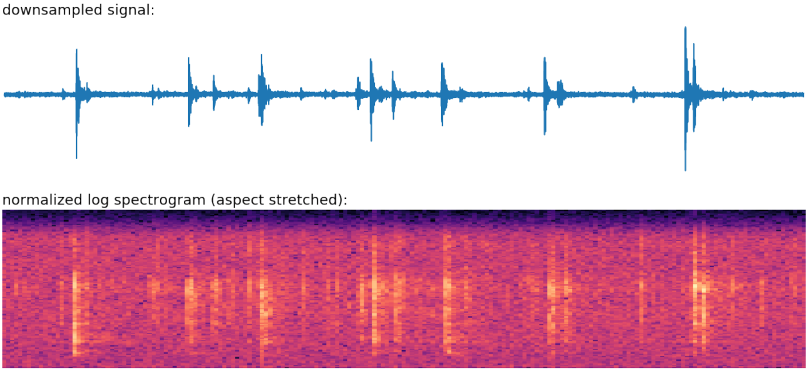

Figure: Spectral Analysis

Methodology Explored 4: Spectogram and CNN II

We divide the time-series data into segments and produce spectograms for the same. These in turn can be fed to a combination 2D CNN and RNN networks.

In one of the research paper we studied, inspired by the advantages of both CNNs and the CTC approach, they propose an end-to-end speech framework for sequence labeling, by combining CNNs with Connectionist Temporal Classification(CTC) directly without recurrent connections. So we can use this model because it is less expensive and it is computationally faster and data they have used in their experiment is resemble our data.

Methodology Explored 5: Stateful LSTM I

We used another feature of RNN/LSTM i.e the stateful LSTM in order to formulate the more long term relationship between the chunks of data but it neither helped and the validation mean accuracy was also above 2.0 in this case.

Conclusion I

We got our best results using the transformer network for the private data-set

References I

- Laboratory Earthquake
  https://www.youtube.com/watch?v=m_dBwwDJ4uo
- Kaggle LANL Competition
  https://www.kaggle.com/c/LANL-Earthquake-Prediction
- Wavenet
  https://arxiv.org/pdf/1609.03499.pdf
- Attention is all you need https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf