Laboratory Earthquake Prediction Using TIME SERIES SEISMIC DATA

Aman Khandelwal[1,2], Amrendra Singh[1,3], Bharat Lodhi[1,4]

[1]B.Tech. (2016-2020), Computer Science and Engineering, Indian Institute of Technology Mandi
[2]B16007@students.iitmandi.ac.in
[3]B16010@students.iitmandi.ac.in
[4]B16015@students.iitmandi.ac.in

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 attempted to predict the time remaining before the laboratory earthquakes occurs, using the time series magnitude seismic data and Deep Learning based methodologies. We explored the time series data using LSTMs, GRU, Transformer Network and then using CNNs after obtaining spectrographs from the data. GRU based approach had the best results in our experiments.

***Index Terms*—Laboratory Earthquake, Time Series, Deep Learning, GRU, LSTM**

I. INTRODUCTION

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.

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.

A. *Subsection Heading Here*

Subsection text here.

1) *Subsubsection Heading Here*

Subsubsection text here.

II. MOTIVATION

Despite considerable research efforts by seismologists, currently scientifically reproducible predictions cannot yet be made to a specific day or month.

If this challenge is solved and the physics are 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.

Even if an earthquake is predicted a few seconds before, computers would be able to shut down gas, electricity and reactors in that time. This can save a lot of money and further damage as a leaked gas line could have caused a lot more damage.

III. CHALLENGES

- Correct form of data representations and feature extraction.
- The data is a complex stream of seismic wave magnitudes.
- A lot of portion of the data may not be actually useful for our network to understand the patterns.
- Little work has been done using Machine Learning to predict earthquakes.
- Extracting useful features from the given time-series seismic data
- Filtering the noise from useful data, which can be due to number of reasons like physical surroundings external to the setup, various irregularities in the instruments setup itself used for generating laboratory earthquakes.

IV. UNDERSTANDING EARTHQUAKES

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.

A. *Earthquake Mechanics*

A wave is a disturbance of a field in which a physical attribute oscillates repeatedly at each point or propagates from each point to neighboring points, or seems to move through space. The waves most commonly studied in physics are mechanical and electromagnetic. Earthquake waves called seismic wave are mechanical waves. Seismic wave are the wave which occurs due to sudden change in rocks in earth or explosion in earth. They are the energy which travels through the earth. There are several type of seismic wave which moves in different ways. The two main important type of waves are body wave and surface waves.Body waves can travel through the earth's inner layers, but surface waves can only move along the surface of the planet like ripples on water. Earthquakes radiate seismic energy as in both forms body and surface waves.

B. Laboratory Earthquake

Laboratory earthquakes created by by applying some shear and normal forces on some rock.The laboratory faults fail in repetitive cycles on some specific combination of forces. Laboratory experiment is simpler than a fault in Earth,but they shares many physical characteristics.

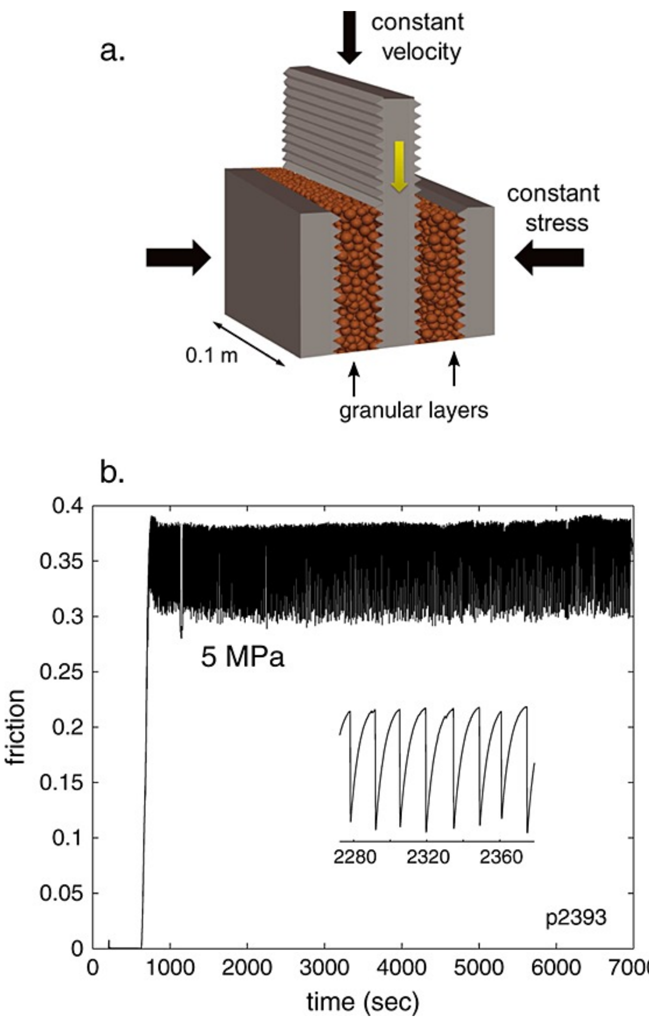

Fig. 1: a.) Laboratory Earthquake Setup b.) Force Generated with respect to time

V. Dataset

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.

VI. Literature Review

A. Attention is all you need

We read the paper "Attention is all you need". In this we got information about transformer network. The sequence models are based on complex recurrent convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. This paper described transformer network architecture for sequence translation from one language to another which is based on attention mechanism. Attention mechanisms have become an integral part of compelling sequence modeling and models in various tasks. It allows modeling of dependencies without regard to their distance in the sequences. So it can learn important feature from the dataset.

Model Architecture:

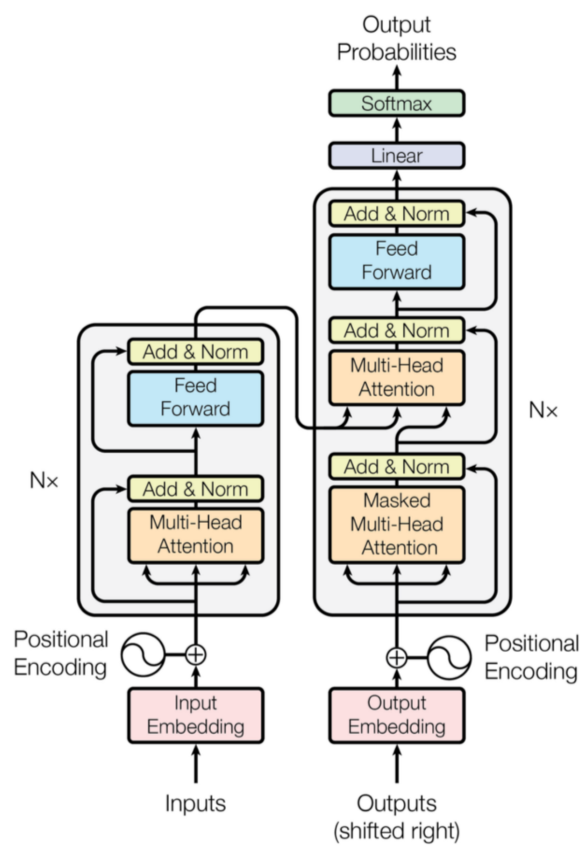

Figure 1: The Transformer - model architecture.

Fig. 2: Caption

It has both encoder and decoder network. Encoder composed of stack of some identical layers. Each layer consists

two sub-layers. The first is a multi head self-attention mechanism, and the second is position wise fully connected feed-forward network. There is residual connection around each sub layer, followed by layer normalization.
Decoder also consists stack of some identical layers. In addition to the two sublayers in each encoder layer, the decoder inserts a third sublayer, which performs multi-head attention over the output of the encoder stack. Decode also have residual connection followed by layer of normalization. Decoder has modefy the self-attention sublayer in the decoder stack to prevent positions from attending to subsequent positions, This ensures that output at i position will depend only on less than i position.

B. *Machine Learning Predicts Laboratory Earthquakes*

This paper shows that by listening to the acoustic signal emitted by a laboratory fault, machine learning can predict the time remaining before it fails with great accuracy. These predictions are based on the instantaneous physical characteristics of the acoustical signal and do not make use of its history. The experiment closely related to Earth faulting because they both share many physical characteristics, so the same approach may work in predicting timing, but not size, of an earthquake.
This predict failure predictions on testing data the acoustic signal corresponding to a sequence of events that the model has never seen. There is no past or future information considered when making a prediction, each prediction uses only the information within one single time window of the acoustic signal. Thus, by listening to the acoustic signal currently emitted by the system, it predict the time remaining before it failsa now prediction based on the instantaneous physical characteristics of the system that does not make use of its history.

VII. Methodologies Tried

A. *Methodology1: Transformer network*

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. The MAE we got is following:
Public Mean absolute error = 1.6011
Private Mean absolute error =2.6610

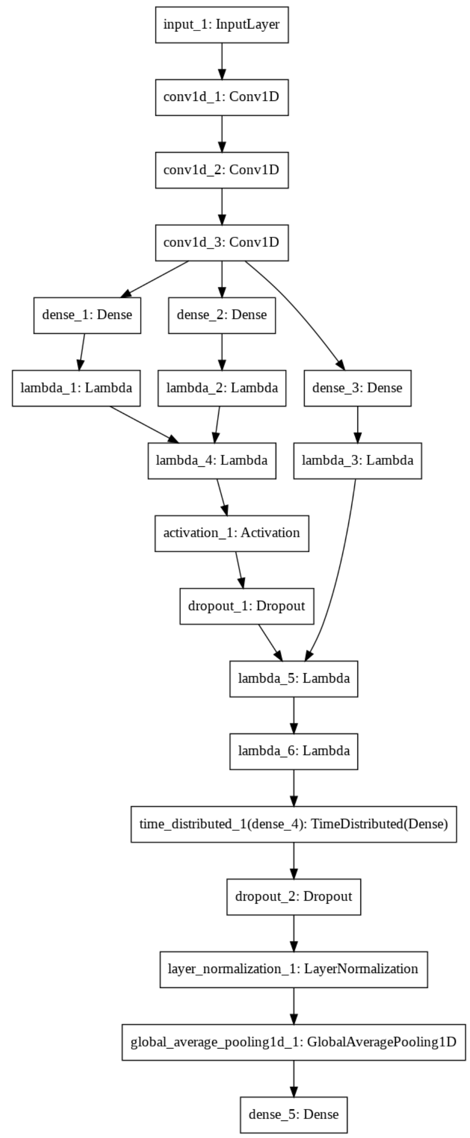

Fig. 3: GRU Based Approach: Model Architecture

B. *Methodology2:GRU*

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 inspried 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
Private Dataset MAE = 2.784

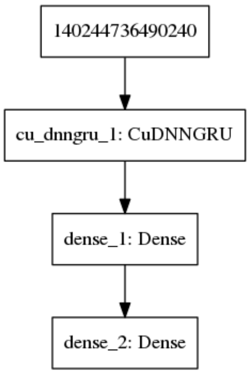

Fig. 4: Caption

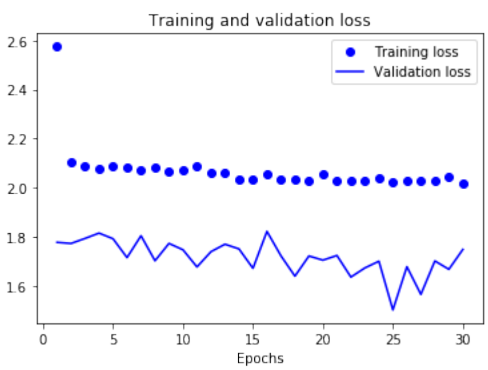

Fig. 5: GRU loss

C. Methodology3: wavenet

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.

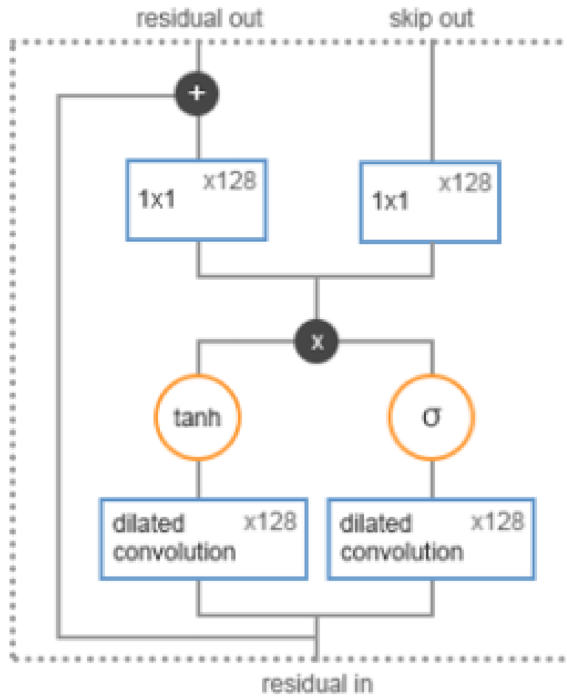

Fig. 6: Caption

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 absolute error achieved by the same was 2.384.

D. Spectrogram and CNN

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. Spectrograms evaluate a 1-dimensional signal input, and produce a 2-dimensional output which is a Fourier transform moving through time. The approach is to divide the signal into frames. On this 2d data we can apply neuaral networks.

E. Stateful LSTM

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.

VIII. Result Comparisons

Approach	Public Score	Private Score
Transformer Network	1.601	2.66
GRU based approach 1	1.69	2.92
GRU based approach 1	1.754	2.784
Spectrogram Approach	3.409	2.88

TABLE I: Result Comparisons

REFERENCES

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