LABORATORY EARTHQUAKE ANALYSIS

Olga Tanyuk[1], Daniel Davieau[1], Dr. Michael L. Blanpied[1], Dr. Charles South[1] and Dr. Daniel W. Engels[1]

[1] Southern Methodist University, Dallas TX 75205, USA
[2] *Add Los Alamos, USGS and or Kaggle here?*

Abstract. In 2017 the Los Alamos National Laboratory conducted an experiment which proved that laboratory earthquakes can be predicted with 90percent accuracy.[1]. The technologies used in the laboratory to simulate and collect earthquake data have since improved.
In this observational study we predict the remaining time before *laboratory* earthquakes occur using data collected with these improved technologies. We analyze data patterns using geophysical subject matter expertise, statistical methods and the latest technology available. We design a statistical algorithm to model the patterns and make a prediction. We compare predicted versus actual time remaining to determine our accuracy.
Our results prove that our model predicted impending laboratory earthquakes with 0.5680 Mean Absolute Error *null hypothesis, pvalue or confidence interval to be added*. The results are an improvement over the 2017 Los Alamos National Laboratory experiment[1]. The evidence in this observational study suggests *We have to find a way to measure against the prior study. We have to state the fact that we cannot prove that the technology improvements are better or that out algorithm is better because of confounding variables such as different collection methods and unavailble details about the 2017 model.*

1 INTRODUCTION

In August 2017 LANL conducted an experiment[1] which predicted the remaining time until *laboratory* earthquakes occur with 90% accuracy. This was accomplished using a random forest algorithm and data collected using *(additional facts to be gathered.)*

There have been improvements in the technology used to collect and measure laboratory seismic signal data since the 2017 experiment. The instruments used are ... *(additional facts to be gathered.)* LANL is now providing data collected with this improved technology to the public via a competition.

> "For this challenge we selected an experiment that exhibits a very aperiodic and more realistic behavior compared to the data we studied in our early work, with earthquakes occurring very irregularly.[2]"

There have also been improvements in computing capabilities. There are now publicly available software packages Keras, Tensorflow, Nvidia CUDA™ and

Turing^{TM} GPU architecture. These improvements enable us to apply complex deep learning algorithms which were more difficult or impossible to employ in the past.

The results of this experiment are potentially applicable to the field of real world earthquakes. Other potential applications include avalanche prediction or failure of machine parts.

> "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.[2]"

Given seismic signal data with considerably more a-periodic laboratory earthquake failures improved computing hardware and software can we improve on the Los Alamos study[1] to determine when laboratory earthquakes will occur?

A recurrent neural network (LSTM) in combination with wave-net algorithm predicts impending laboratory earthquakes with *(Mean absolute Error = 0.5680 additional facts to be gathered)* accuracy. The data, hardware and software allows us to predict impending earthquakes more accurately. However we only know milliseconds before failure. Therefore practical applications may bee limited. This may prove useful but only applies to laboratory experiments. *(How to collect data for real world earthquakes?)* This could be used in industry perhaps researching materials for wallboard, machine parts.

2 TUTORIAL MATERIAL

We hear about earthquakes mostly via news media when there is a large seismic event which is noticeable, causes death and destruction. These are stick–slip events that radiate seismic energy along the seams (fault lines) between tectonic plates. In this study we refer to these as *Regular Earthquakes*

Another type of earthquake we refer to in this study is a *Slow Slip Earthquake* (SSE). SSE's are fault behaviors that occur slowly enough to make them undetectable without instrumentation. They do not shake the ground and cause widespread destruction like regular earthquakes do. They occur near the boundaries of large earthquake rupture zones[3].

There is evidence to suggest that there is a relationship between slow slip earthquakes and more noticeable regular earthquakes[4].

This study analyzes the relationship between slow slip and regular earthquakes. We use this relationship information to predict regular laboratory earthquakes.

2.1 Temporary Title

Regular earthquakes are caused by a sudden slip on a fault. Tectonic plates are always slowly moving, but they get stuck at their edges due to friction. When the stress on the edge overcomes the friction, there is an earthquake that releases energy in waves that travel through the earth's crust and cause the shaking that we feel.[5]

Los Alamos National Laboratory researchers discovered a way to successfully predict earthquakes in a laboratory experiment that simulates natural conditions. In 2017, this team discovered a way to train a computer to pinpoint and analyze seismic and acoustic signals emitted during the movements along the fault to predict an earthquake. They processed massive amounts of data and identified a particular sound pattern previously thought to be noise that precedes an earthquake. The team was able to characterize the time remaining before a laboratory earthquake at all times.[6]

In the lab, the team imitated a real earthquake using steel blocks interacting with rocky material (fault gouge) to induce slipping that emitted seismic sounds. An accelerometer recorded the acoustic emission emanating from the sheared layers.[6]

For the first time, researchers discovered a pattern that accurately predicted when a quake would occur. The team acknowledges that the physical traits of the lab experiment (such as shear stresses and thermal properties) differ from the real world but the application of the analysis to real earthquakes to validate their results is ongoing. This method can also be applied outside of seismology to support materials' failure research in many fields such as aerospace and energy.[6]

The team's lab results reveal that the fault does not fail randomly but in a highly predictable manner. The observations also demonstrate that the fault's critical stress state, which indicates when it might slip, can be determined using exclusively an equation of state.[6]

So far seismologists and Earth scientists have relied exclusively on catalogues of historical data to try to characterize the state of faults. These catalogues contain a minute fraction of seismic data, and remaining seismic data is discarded during analysis as useless noise. The authors discovered that hidden in this noise-like data there are signals emitted by the fault that inform them of the state of the fault much more precisely than catalogues.[6]

> "Our work shows that machine learning can be used to extract new meaningful physics from a very well studied system," said Bertrand Rouet-Leduc, Los Alamos Earth and Environmental Sciences Division scientist and the paper's lead author. "It also shows that seismogenic faults are continuously broadcasting a signal that precisely informs us of their physical state and how close they are to rupture, at least in the laboratory."

3 DATA

The data used in this study was provided by LANL via a Kaggle competition[2]. It consists of 629,143,480 seismic signal observations with an accompanying record of the time remaining before the next laboratory earthquake occurred.

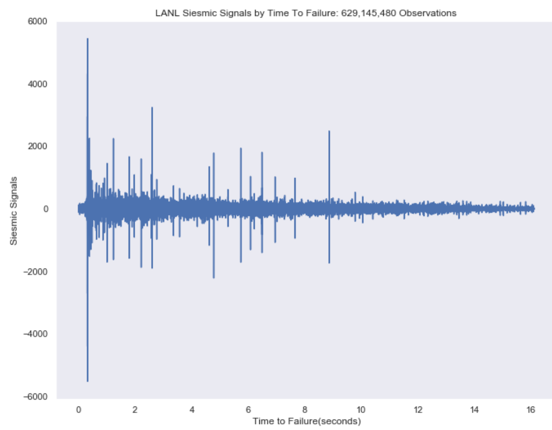

Fig. 1. The magnitude of each seismic signal and its related time remaining before the next laboratory earthquake.

Table 1. Sample of Data Provided

Index	**Seismic Signal**	**Time to Failure**
0	12	1.469099998474121
1	6	1.469099998474121
2	8	1.469099998474121
3	5	1.469099998474121
4	8	1.469099998474121

The data was generated by a standard laboratory earthquake physics set-up called *tbd*. The 629,143,480 observations are a continuous segment. The seismic signals are signed integer values ranging from *min max*. They are acoustic and *hopefully we find more information about this*. The time to failure recordings are floating point decimal ranging from *min max* in seconds.

Table 2. Seismic Signal Sample

Index	**Signal**
0	3
1	10
2	4
3	4
4	1

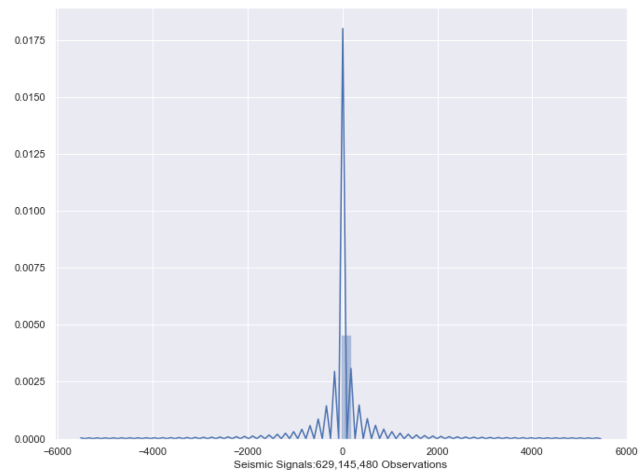

Fig. 2. The distribution of seismic signal measurements by LANL

Kaggle Competition Additional seismic signal data is provided with the time remaining excluded. The intention is to predict the time remaining and submit to Kaggle for scoring in the competition[2]. It is a collection of 2624 segments with 150,000 observations in each. The data in each segment is continuous *within* but *not* between.

There was one file in the test folder for each prediction (seg id) in sample submission:

Table 3. Competition Data

Index	Segment Id	Time to Failure
0	seg 00030f	0
1	seg 0012b5	0
2	seg 00184e	0
3	seg 003339	0
4	seg 0042cc	0

One huge csv file had all the training data, which is a single continuous experiment. There were only two columns in this file: Acoustic data (int16): the seismic signal; Time to failure (float64): the time until the next laboratory earthquake (in seconds). There were no missing values for both columns.

Acoustic Data The acoustic feature were integers in the range [-5515, 5444] with mean 4.52. The plot below is using a 1 percent random sample (6M rows):

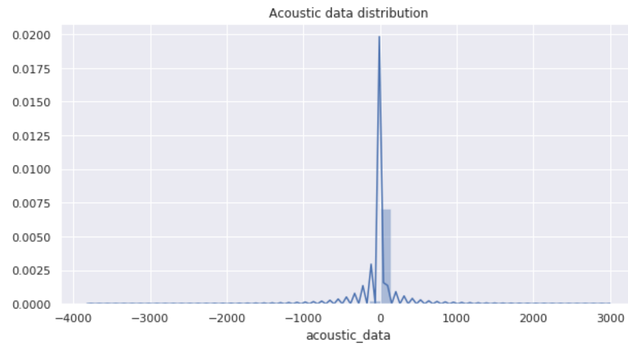

Fig. 3. 1% random sample from 629,143,480 observations

Time to Failure The target variable was given in seconds:

Table 4. Sample of Data Provided

Index	Seismic Signal	Time to Failure
count	6.29145480	
mean	4.47708428	
std	2.61278939	
min	9.55039650	
25	2.62599707	
50	5.34979773	
75	8.17339516	
max	1.61074009	

Timeseries

– We can see that usually acoustic data shows huge fluctuations just before the failure and the nature of data is cyclical
– Another important point: visually failures can be predicted as cases when huge fluctuations in signal are follows by small signal values. This could be useful for predicting "time_to_failure" changes from 0 to high values;

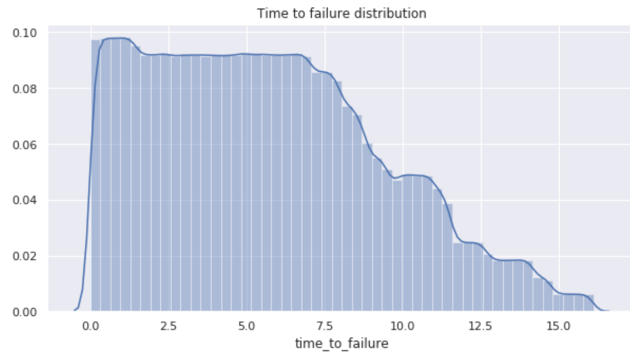

Fig. 4. The min value was very close to zero (around 10ˆ5) and the max was 16 seconds. The distribution for the random sample

Table 5. Sample of Data Provided

Index	**Seismic Signal**	**Time to Failure**
count	11325	
mean	0.64454830	
std	1.32147193	
min	0.31079626	
25	0.31549615	
50	0.31689683	
75	0.32029617	
max	8.86059952	

4 METHODS AND EXPERIMENTS

Describe the solution approach here.

4.1 Data Transformations

Explain what we did to the data here. Explain our train and test data here.

4.2 Recursive Neural Network

Mean absolute = 0.5680 Patrick Yam: LSTM cannot handle data with 150000 sequence length therefore we wavenet in the earlier layers as feature extraction and reduce the sequence length to 150.

LSTM is a type of RNN that helps us deal with vanishing or exploding gradients (incorrect slope). For example most of our observations are within x time to failure (insert zoomedInTimePlot.png) however more extreme occur just before failure. The extreme differences in the slope difference between these observations can skew the accuracy of more traditional RNN algorithms.

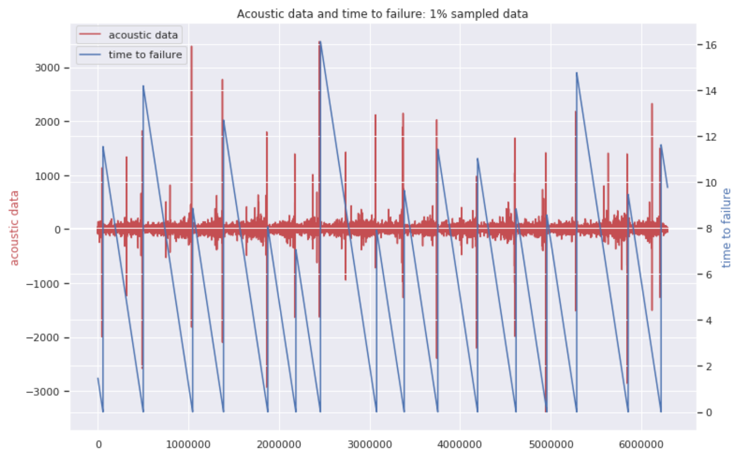

Fig. 5. We checked how both variables changed over time. The red line is the acoustic data and the blue one is the time to failure. On a plot above we can see, that training data has 16 earthquakes. The shortest time to failure is 1.5 seconds for the first earthquake and 7seconds for the 7th, while the longest is around 16 seconds.

LSTM separates the long term(fault slip) from short term (slow slip) Hidden state= memory from prior observations Typical RNN input+priorhiddenstate¿tanh (-1,1) activation function¿new hidden state LSTM is same but includes gates to determine what information is included in hidden state and what is not.

The gates are rnn's themselves.

input+priorhiddenstate¿sigmoid (0,1) activation function¿new hidden state. 0, 1 allows us to forget or remmeber

Forget gate decides what to include 1, or exclude 0. https://www.youtube.com/watch?v=2GNbIKTKCfE https://www.youtube.com/watch?v=8HyCNIVRbSU

4.3 Auto Regressive integrated Moving Average (ARIMA)

The data shows evidence of non-stationarity (the mean, variance change over time). We use an ARMIA model to analyze the changing means and variance. This this model also accounts for white noise.

ARIMA(0,0,0). Check Wikipedia.

4.4 Gradient Boosting Decision Tree

http://mlexplained.com/2018/01/05/lightgbm-and-xgboost-explained/ mean absolute error: 2.230923263623967 r2 score: 0.42923076680297423

5 RESULTS

Include evaluation methodology We measure this by Abolute Mean Square error

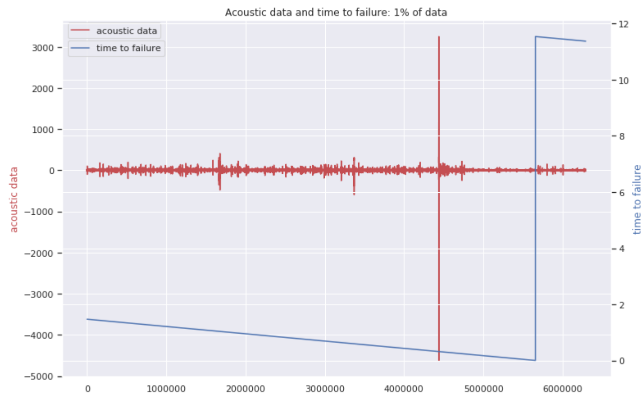

Fig. 6. On this zoomed-in-time plot we can see that actually the large oscillation before the failure is not quite in the last moment. There are also trains of intense oscillations proceeding the large one and also some oscillations with smaller peaks after the large one. Then, after some minor oscillations, the failure occurs. Interesting thing to check is the time between high levels of seismic signal and the earthquakes. We are considering any acoustic data with absolute value greater than 1000 as a high level

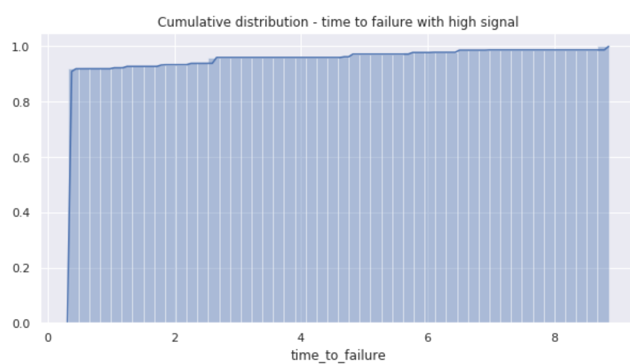

Fig. 7. More than 90% of high acoustic values are around 0.31 seconds before an earthquake

It is clear that the LSTM + Wave-net algorithms are superior performers. This is because it handles extreme variations in the slope (gradient descent) and we can tune it to remember not only the common signals but the extreme signals which occur only when failure is imminent. Results of experiments Use tables and graphs Use tables and graphs Use tables and graphs Don't forget explanations

6 ANALYSIS

Analyze results. These are NOT conclusions.

7 ETHICS

If people believe us and we are wrong; bad things can happen. If people believe us and we are right; good and bad things can happen.

8 CONCLUSION

Draw a minimum of 3 conclusions This is NOT a summary section.

We are able to predict with Mean absolute Error = 0.5680 however we only know milliseconds before failure. Therefore practical applications may bee limited. This may prove useful but only applies to laboratory experiments. How to collect data for real world earthquakes?This could be used in industry perhaps researching materials for wallboard, machine parts.

References

1. Bertrand Rouet-Leduc, Claudia Hulbert, N.L.K.B.C.J.H.P.A.J.: Machine learning predicts laboratory earthquakes
2. Kaggle, R.: Lanl earthquake prediction
3. Ikari Matt J, Marone Chris, S.D.M.K.A.J.: Slip weakening as a mechanism for slow earthquakes
4. Baptiste Rousset, Roland Burgmann, M.C.: Slow slip events in the roots of the san andreas fault
5. USGS: What is an earthquake and what causes them to happen? (March 2019)
6. Rouet-Leduc, B.: Los alamos machine learning discovers patterns that reveal earthquake fault behavior