EARTHQUAKE ANALYSIS USING MACHINE LEARNING

ABSTRACT: This paper presents work done on finding earthquake magnitude, P-wave and S-wave arrival times, Strong motion duration and relations between various seismic events around the globe (limited to us for now) and also includes finding correlations between earthquakes occurring over time.

INTRODUCTION: Earthquakes are one of the most powerful natural phenomena, these pose significant risk to human lives and infrastructure. Understanding the characteristic properties and behaviours of seismic events is crucial for improving predictive models, early warning systems, and to decrease the destructive impacts of earthquakes. This paper aims to contribute to these understandings by focusing on three critical aspects of earthquake analysis: the determination of earthquake magnitude, the P-wave and S-wave arrival times, the determination of strong motion duration, and the correlations between various seismic events.

Accurately determining the magnitude of an earthquake is fundamental for assessing earthquakes impact. Magnitude quantifies the energy released during an earthquake. Identifying the arrival times of P-waves and S-waves is essential for locating earthquake epicentres and for early warning systems. P-waves [Primary Waves] are the fastest seismic waves, followed by slower S-waves [Secondary Waves], the time difference between these waves is very useful for calculation of distance to the earthquake source.

Another critical aspect is the prediction of strong motion duration, it is the time interval during which the ground shaking exceeds a certain threshold of acceleration. This duration is important in assessing the potential damage to structures and for designing buildings and infrastructure that can withstand seismic forces.

In addition to these, we analyze the correlations between earthquakes over time. Understanding these correlations is of at most importance for advancing our knowledge of seismic activities and for developing more robust strategies to anticipate future earthquakes.

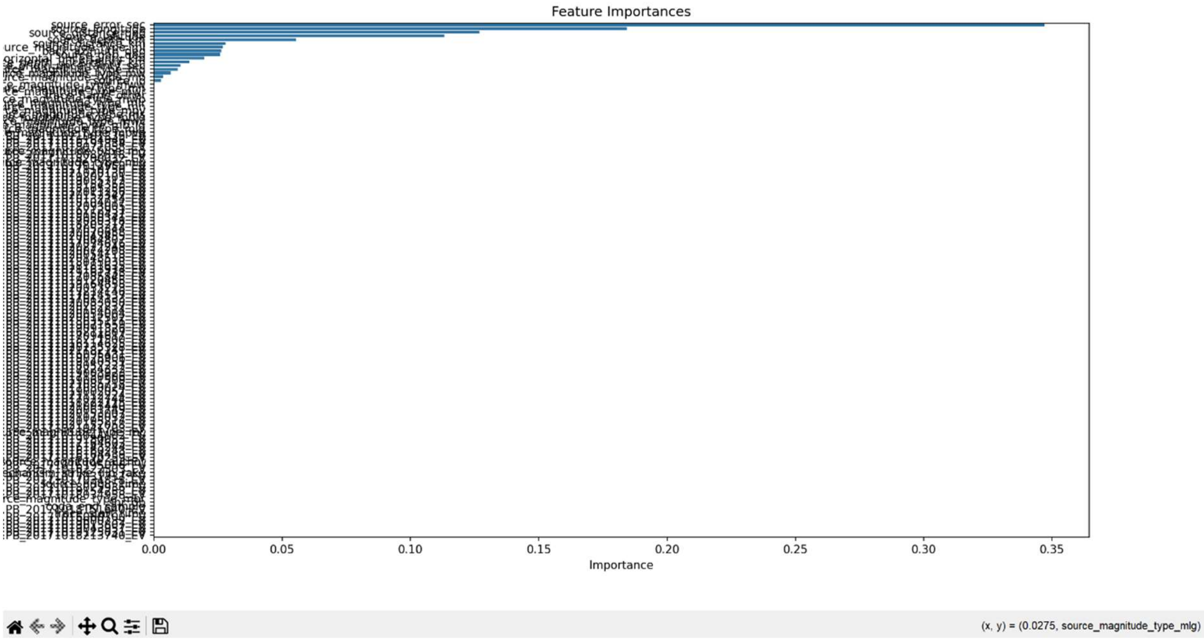
RELATED WORK: 1) Dataset of tremors recorded by seismic stations around the globe across the years 2011 – 2018 which was created by Stanford university called 'STEAD' [1]. Although being a reliable source containing more than sufficient and accurate metadata of the seismic signals, it does <u>not</u> contain any data regarding strong motion duration, correlations between attributes of earthquake data nor relations between any two earthquakes.

METHODOLOGY: Earthquake magnitude prediction and P- and S- wave arrival times were predicted using STEAD dataset by Random Forest Regressor Model from scikit learn, there are 35 features of an earthquake in STEAD dataset of which 10 important features were selected based on feature importance and model got trained on these 10 features, P-wave arrival time and S-wave arrival time were calculated from actually predicted P arrival sample and S arrival sample by using sampling rate[2] and the following formula[3].

s_arrival_time = s_arrival_sample / sampling_rate

p_arrival_time = p_arrival_sample / sampling_rate

Output containing model evaluation, feature importances and the magnitudes of P, S-wave arrival times along with earthquake magnitude.



Graph visually representing feature importances of all the 35 features.

After prediction of important features of 2 earthquakes (which may have occurred at any place and at any time) we then proceed to find their correlation as the next step:

Correlate.py: Analyzing Earthquake Event Interrelations

This script is designed to analyze the potential relationship between two detected seismic events (earthquakes or noise), evaluating whether the events are related or independent. The workflow involves multiple steps and criteria that examine the magnitudes, temporal occurrence, and spatial proximity of the events. These steps are summarized below:

1. **Initial Classification of Events**: The first step is to classify whether the detected events are seismic (earthquakes) or noise. If both events are earthquakes, the script proceeds by retrieving information about their respective magnitudes, arrival times, and locations from the prediction files. If either event is classified as noise—comprising approximately 20% of the dataset—no correlation analysis is performed.

2. **Initial Event Comparison**: The next step is to calculate the distance between the two seismic receiver stations using the Haversine formula. This distance is then compared alongside the start times of the two events to determine if they occurred within a time and spatial window conducive to potential interaction. If the events are too distant (in relation to their magnitudes) or temporally disconnected, they are likely independent. This scenario is common when analyzing two randomly chosen earthquakes from a large dataset.

3. **Assessment of Same Earthquake Detection**: The script then checks if the two events originate from a similar epicenter and occur within a close temporal window. If so, this suggests that both seismic stations may be detecting the same earthquake. Although this scenario is rare, it remains a critical step in the workflow to ensure redundancy is minimized.

4. **Direct Interaction Between Earthquakes**: When two earthquakes occur in proximity to one another, several factors can influence their interaction:

    - **Stress Transfer**: If two earthquakes are sufficiently close spatially, the larger event may influence the smaller one through stress transfer. Seismic waves from the primary earthquake can alter the local stress field, potentially exacerbating or triggering a nearby fault slip.

    - **Energy Attenuation**: As seismic waves propagate, their energy dissipates, governed by the attenuation coefficient. A low attenuation coefficient indicates that a significant portion of the earthquake's energy remains over a considerable distance, increasing the likelihood of interaction with nearby seismic events.

    - **Combined Seismic Impact**: If the two earthquakes occur within close temporal and spatial proximity, their combined wave effects can modulate the seismic response, influencing the magnitude of the secondary earthquake or creating new aftershock patterns.

To quantify the relationship between the two events, the script computes a correlation score based on their magnitudes, the distance between them, and predefined standard coefficients (commonly used in seismic studies). A higher correlation score suggests a stronger interaction between the two events.

5. **Potential Earthquake Triggering**: If direct interaction is ruled out, the script checks whether the stronger earthquake may have triggered the weaker one through two potential mechanisms:

    - **Mainshock Triggering**: If a larger earthquake (mainshock) precedes a smaller one, it may trigger a subsequent earthquake (aftershock) if the stress transfer from the mainshock exceeds the fault's critical stress threshold. This critical stress threshold represents the amount of stress necessary to rupture a fault line, thereby initiating a new seismic event. The potential for triggering is evaluated using seismic moment and static stress change calculations.

- **Aftershock Triggering**: If the temporal gap between the two earthquakes is large enough to preclude mainshock triggering but within the expected aftershock period, the script assesses whether the first earthquake could have induced an aftershock event. This analysis also takes into account the spatial proximity and magnitude of the preceding earthquake.

6. **Correlation Score Recalculation**: If the two events are found to be related either through mainshock triggering or aftershock generation, a revised correlation score is calculated. This score reflects the interaction between the events, considering factors such as stress transfer, seismic wave attenuation, and the critical stress threshold of the region surrounding the secondary event.

By employing this workflow, the script systematically analyzes potential relationships between seismic events, encompassing various scenarios such as direct interaction, triggering mechanisms, and redundant detections. This approach provides a foundation for further studies aimed at understanding the complex dynamics of earthquake interactions, contributing to the broader understanding of seismic event correlations.

Test Run1: Two earthquakes occurring nearby but after a large gap of 201days

```
22   test_trace_names = ['KAN16.GS_20150424061553_EV','KAN16.GS_20151111113626_EV']
23   ##########################################################################################

PROBLEMS (2)   OUTPUT   DEBUG CONSOLE   TERMINAL   PORTS   SEARCH ERROR

PS C:\Users\Siddharth\Desktop\IIITB\bashok_re(only code)> python corr_updated.py
earthquake
earthquake
magnitude1: 3.15383243560791
magnitude2: 2.0035433769226074
start date of eq1: 2015-04-24
start time of eq1: 06:15:54
start date of eq2: 2015-11-11
start time of eq2: 11:36:27
lat1 is 37.225617 ,lon1 is -98.064717, lat2 is 37.225617, lon2 is -98.064717
sourcelat1 is 36.7408 ,sourcelon1 is -97.5691, sourcelat2 is 37.148, sourcelon2 is -97.87433
distance between stations is:0.0
distance between sources is:52.781948626484194
gap in days is:201
The two earthquakes are independent and dont directly effect each other ,by days,
The two earthquake signals are not from the same epicenter
the region of KAN16.GS_20150424061553_EV is Kansas
the region of KAN16.GS_20151111113626_EV is Kansas
the value of do_they_interact is: False
value of is_MainTriggering_possible : False
the magnitude of the aftershock of the stronger eq ,AFTER travelling is: 1.5483870540580282
Even though the two earthquakes occur in reasonable distances they are independent and dont effect each other
PS C:\Users\Siddharth\Desktop\IIITB\bashok_re(only code)>
```

TEST RUN 2: Two earthquakes occur in an interval of 0.5 hour thus the aftershockwaves of the former triggered the latter earthquake

```
Even though the two earthquakes occur in reasonable distances they are independent and dont effect each other
PS C:\Users\Siddharth\Desktop\IIITB\bashok_re(only code)> python corr_updated.py
earthquake
earthquake
magnitude1: 0.804690420627594
magnitude2: 0.9642623662948608
start date of eq1: 2015-12-16
start time of eq1: 13:53:37
start date of eq2: 2015-12-16
start time of eq2: 14:22:06
lat1 is 39.710602 ,lon1 is -119.385399, lat2 is 39.710602, lon2 is -119.385399
sourcelat1 is 39.6151 ,sourcelon1 is -119.9589, sourcelat2 is 40.1094, sourcelon2 is -119.7935
distance between stations is:0.0
distance between sources is:56.74763786521207
gap in days is:0
The two earthquake signals are not from the same epicenter
the region of PAH.NN_20151216135336_EV is Nevada
the region of PAH.NN_20151216142205_EV is Nevada
the value of do_they_interact is: False
value of is_MainTriggering_possible : False
the magnitude of the aftershock of the stronger eq ,AFTER travelling is: 0.8656816223512384
The two earthquakes are related and the AFTER shockwaves of 0.9642623662948608 triggers 0.804690420627594
Correlation Score: 0.9841694469532398
PS C:\Users\Siddharth\Desktop\IIITB\bashok_re(only code)>
```

The duration of strong motion was predicted using Random Forest Regressor model and by taking a threshold for strong motion duration as 0.005 [4], the dataset used for this prediction is the Portland_cv dataset [5], hence these duration predictions are limited to the city of Portland, Oregon, USA. For any new Strong acceleration values given the duration of strong motion can be predicted.

```python
# Prepare new data for prediction
new_data = pd.DataFrame({
    'SA(0.010)': [-0.00363],
    'SA(0.020)': [-0.00363],
    'SA(0.030)': [-0.00363],
    'SA(0.050)': [-0.00364],
    'SA(0.075)': [-0.00365],
    'SA(0.100)': [0.00371],
    'SA(0.150)': [0.00548],
    'SA(0.200)': [0.00413],
    'SA(0.250)': [0.00431],
    'SA(0.300)': [0.00440],
    'SA(0.400)': [0.00519],
    'SA(0.500)': [0.00520],
    'SA(0.750)': [0.00793],
    'SA(1.000)': [0.00835],
    'SA(1.500)': [0.01102],
    'SA(2.000)': [0.01274],
    'SA(3.000)': [0.01023],
    'SA(4.000)': [0.00628],
    'SA(5.000)': [0.00301],
    'SA(7.500)': [0.00152],
    'SA(10.000)': [0.00083],
})
```

New data of strong acceleration for prediction.

```
Mean Squared Error: 0.047031979492187584
Predicted Strong Motion Duration: 3.7067500000000035 seconds
```

Model evaluation and Predicted values for above data.

REFERENCES:

(1) For reference on STEAD   dataset visit: https://ieeexplore.ieee.org/document/8871127

(2) Sampling rate is the frequency at which we sample seismic signal. For STEAD sampling rate was taken to be 100Hz.

(3)

(4) They do not record all the **time** but are triggered by a minimum level of **ground** acceleration, usually of the order of **0.005** to 0.01g in the vertical direction.

http://www.isesd.hi.is/ESD_Local/Documentation/documentation.htm

(5) aagaard_2023portland (strongmotioncenter.org)