Clustering Using Source Time Function

Harsha Vardhan Naidu (CE22B007)

1. Introduction

This report details an analysis workflow designed to cluster earthquake events based on their source time function (STF) data. The dataset includes various earthquake parameters such as time, duration, latitude, longitude, depth, moment (M0), moment magnitude (Mw), and focal mechanism angles (strike, dip, and rake). Additionally, each earthquake folder contains a CMT image that visually represents the source mechanism. The goal is to preprocess the data, visualize it on a world map, classify faulting mechanisms based on the rake angle, and finally apply several clustering techniques to assess their performance.

2. Data Preparation and Preprocessing

The analysis began by iterating through directories containing earthquake events. For each event, the code performed the following steps:

- **Data Extraction:**
  Each folder contained a text file (prefixed with "fctmoysource_") that provided the earthquake's metadata (date, time, latitude, longitude, depth, M0, Mw, strike, dip, and rake). The first two lines of these files were parsed to extract this information.
- **Image Processing:**
  The code searched for a `cmt.png` image within each folder. When found, it was read using OpenCV, ensuring the image was in the correct RGBA format before storing it alongside the numeric data.
- **Data Storage:**
  The extracted data (including the associated image) was stored in a Pandas DataFrame with clearly defined column names. The DataFrame was then exported to a CSV file ("Clustering_data_Cleaned.csv") and a pickle file ("Clustering_data_with_images.pkl") for subsequent analysis.

3. Visualization

Two primary visual outputs were generated from the processed data:

Figure 1: Earthquake CMT Images on Coloured World Map

A global map was created using Cartopy with the following features:

- **Coloured Base map:**
  The map was rendered with light green land and light blue ocean. Coastlines and borders were highlighted for clarity.
- **Image Overlay:**
  Each earthquake event was plotted at its corresponding latitude and longitude, with its

associated CMT image overlaid on the map. This provided a spatial view of the earthquake events and their focal mechanisms.

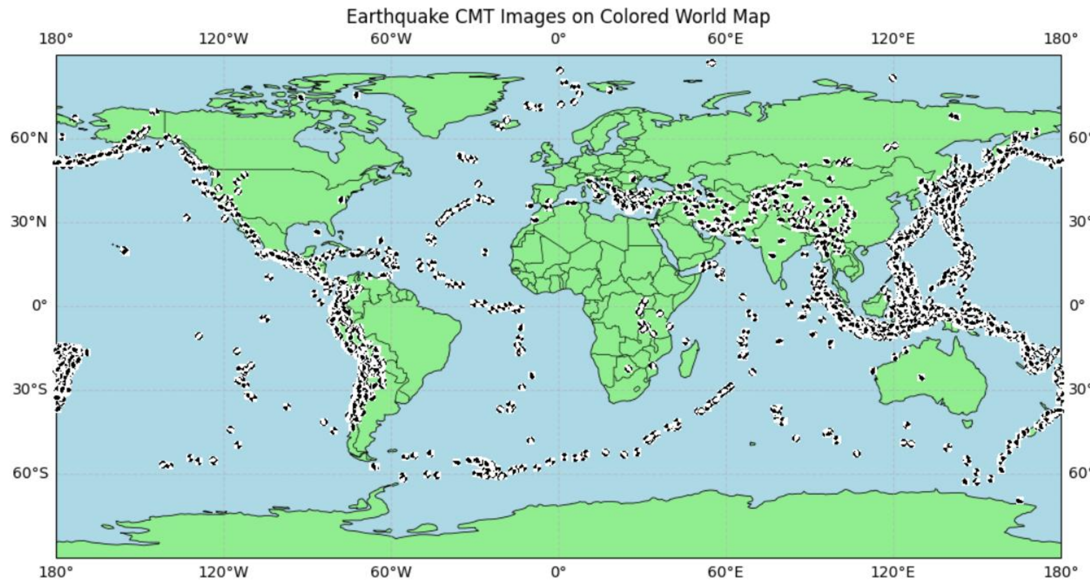

Fig.1: Earthquake CMT Images on Coloured World Map

Figure 2: Fault Types vs Magnitude (Mw)

Using Seaborn, a scatter plot was generated to show the relationship between the moment magnitude (Mw) of each event and its classified faulting mechanism. In this plot:

- **Fault Classification:**
  The faulting mechanism was classified using a custom function based on the rake angle:
  - "Strike slip" for rake angles near 0° or ±180°.
  - "Normal and normal oblique" for intermediate negative rakes.
  - "Reverse and reverse oblique" for intermediate positive rakes.
- **Visualization:**
  Data points were color-coded by fault type, providing an intuitive overview of how fault types distribute with respect to earthquake magnitude.

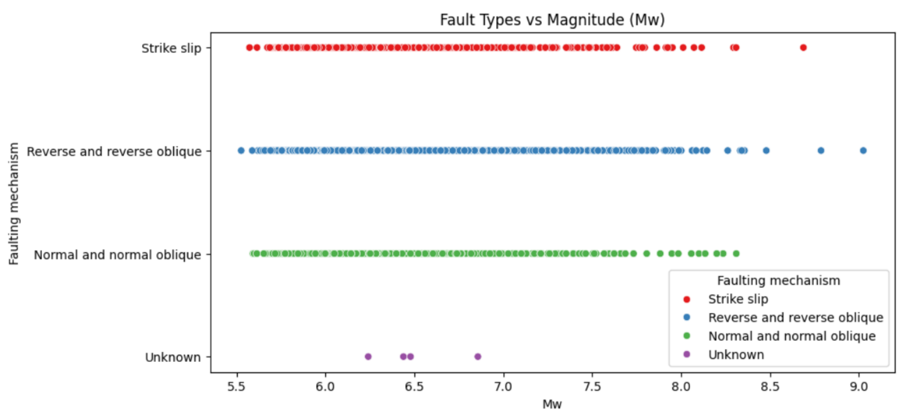

Fig.2: Fault Types vs Magnitude (Mw)

4. Faulting Mechanism Classification

A simple classification function was applied to the `rake1(°)` values:

- **Strike Slip:** For rake angles in the ranges (–180, –150], (–30, 30), and (150, 180].
- **Normal / Normal Oblique:** For rake angles between –150° and –30°.
- **Reverse / Reverse Oblique:** For rake angles between 30° and 150°.

This classification was added as a new column in the DataFrame ("Faulting mechanism"), which later assisted in the clustering analysis.

5. Clustering Analysis

Three clustering techniques were applied to the dataset, each using two different feature sets—one including epicentre coordinates (latitude and longitude) and one without:

Methods Applied

- **K-Means Clustering:**
  Standard K-Means was applied to the standardized features.
  - **Accuracies:**
    - Without epicentre: 68.88%
    - With epicentre: 69.52%
- **Gaussian Mixture Model (GMM):**
  A probabilistic approach that fitted Gaussian distributions over the features.
  - **Accuracies:**
    - Without epicentre: 62.44%
    - With epicentre: 51.87%
- **DBSCAN Clustering:**
  A density-based clustering method that identifies clusters based on local point density.
  - **Accuracies:**
    - Without epicentre: 58.08%
    - With epicentre: 96.00%

Observations

The clustering analysis revealed that while KMeans and GMM provided moderate performance, **"DBSCAN"** dramatically improved in accuracy when epicentre information was included. This suggests that spatial information plays a crucial role in distinguishing between earthquake fault types.

6. Conclusion

The analysis successfully demonstrated a complete workflow for processing earthquake data using the source time function. Key conclusions include:

- **Data Fusion:**
  Integrating numerical data with corresponding CMT images provided a comprehensive view of each event.
- **Visualization:**
  The world map and scatter plot visualizations offered intuitive insights into the spatial distribution and faulting mechanisms of the earthquakes.
- **Clustering Insights:**
  Incorporating epicentre data significantly enhanced clustering performance, particularly for DBSCAN, which achieved an accuracy of 96% with epicentre information.

Overall, this workflow can serve as a robust framework for earthquake data analysis, offering both quantitative clustering results and qualitative visual insights.

*Refer the code from the below github link:

https://github.com/Harsha-Vardhan00/CE6018_SDA_PROJECT/tree/main/week2

THANK YOU