Visual Analysis of World Earthquakes based on Data Science and Statistical Methods

Xing Xiangrong
International college, Beijing University of Posts and Telecommunication, Beijing, China
xxr@bupt.edu.cn

Abstract—Earthquake is a very serious geological disaster, often accompanied by serious casualties. Therefore, it is necessary to explore the law of earthquake occurrence and take corresponding measures in time to reduce losses and casualties. However, it is difficult to predict the mechanism and principle of earthquakes. In this study, we utilize the methods and tools of data science to explore the regularity of earthquake occurrence. This paper is a practice of enforcing the workflow of data science, which uses computer algorithm to make fast statistics of data and conduct visualization processing to study the law of earthquake occurrence. Compared with the traditional research methods, it has important significance.

1. INTRODUCTION
The advent of the era of big data has brought new opportunities to various disciplines, namely, the transformation of epistemology and research paradigm [7], and data science, the highly integrated discipline of data research in different professional fields has been derived [8].

The number of people dying in natural disasters is lower today than it was in the past, the world has become more resilient. Earthquakes, however, can still claim a large number of lives. Whilst historically floods, droughts and epidemics dominated disaster deaths, a high annual death toll now often results from a major earthquake and possibly a tsunami caused by them. However, the law of earthquake occurrence is difficult to predict, and the academic community has not yet made clear its occurrence mechanism and principle. The answer to which places have high probability and which place has low probability is also difficult to determine. Therefore, we need to use the seismic data set, use Python for visual analysis, find rules from the data, and conduct scientific research with data methods, so as to reduce the harm of earthquake to human and all material culture and other life of the earth.

It is well known that Python is the most popular programming language at present. It has the characteristics of powerful function and simple operation. It can meet the needs of software designers. It is an important auxiliary tool for data analysis. Python data visualization includes a Matplotlib class library, which involves rich data visualization resources, including four parts: Matplotlib’s basic figure type, adding map annotation (including coordinate axis range, length width ratio or coordinate axis, etc.), adjusting the style and color of figure, and other complex graphics, so as to make data visualization more intuitive and elegant. It is the presentation of data analysis results [2]. We use this technology to display a large number of data in a clear form. According to the trends indicated by the graphs, we find some laws that are difficult to be comprehensively studied by traditional methods.
The remainder of this paper is organized as follows. In section II, visualization of seismic data is presented. In section III, data analysis from six different perspectives is conducted. In section IV, related work is reviewed before we conclude in Section V.

2. **Visualization of Global Seismic Data**

In this section, we study and visualize the global seismic data.

2.1. *Geographical Distribution of Earthquakes with \( m \geq 5.5 \) in the world*

The system visualized every earthquake occurred in the world in 52 years according to the epicenter location, earthquake magnitude and earthquake time. As shown in Figure 1, each dot represents a sequence of earthquakes, the color of the dots indicates the magnitude, the radius of the dots represents the focal depth, and the area with dense dots represents the frequent earthquakes in the region.

According to the data, earthquakes mainly occurred between the circum Pacific plate, the Eurasian plate and the Indian Ocean plate. The frequency of earthquakes in the circum Pacific seismic belt and the Himalayan seismic belt is relatively frequent. This also confirms the well-known scientific common sense: the interaction in the process of the earth's plate movement is the main cause of earthquakes. At the same time, we also notice that the frequency of deep earthquakes occurring at the edge of the seismic zone is significantly higher than that at the center of the seismic zone.

It can be seen in the Fig. 1 that 80% of the world's earthquakes occur in the Pacific Rim seismic belt, and Indonesia, the Philippines and Japan in Asia are in the worst hit areas. Secondly, it starts from Indonesia, passes through the western part of Indochina Peninsula and Yunnan Guizhou Sichuan Tibet region in China to the north coast of the Mediterranean Sea and extends to the Himalayan seismic belt of the Atlantic Ocean.

2.2. *Earthquake Prone Areas*

![Geographical distribution of earthquakes with \( m \geq 5.5 \) in the world](image)

Fig. 1. Geographical distribution of earthquakes with \( m \geq 5.5 \) in the world

3. **Seismic Data Analysis**

3.1. *The Magnitude Distribution of the Earthquake*

As shown in Fig. 2 It can be seen from the figure that the number of earthquakes with \( M = 5.5-6.0 \) is relatively large, the number of earthquakes decreases significantly with the increase of magnitude, and there are only single digits for earthquakes with \( m \geq 8.0 \). The number of large earthquakes in the world is less, and the number of earthquakes with moderate and strong magnitude can cause damage is more.
3.2. The Focal Depth

As shown in Fig. 3, it is obvious that the shallow earthquakes (focal depth ≤ 60km) account for more than 75% of the earthquakes with m ≥ 5.5, while the number of moderate earthquakes (60km < focal depth ≤ 300km) decreases linearly with the deepening of focal depth. In deep source earthquakes (focal depth > 300km), earthquakes with focal depth of 500 ~ 650km are the main ones.

The shallower the source is, the greater the damage is, but the smaller the scope is. Destructive earthquakes are generally shallow earthquakes. Therefore, we can see from the figure that most of the earthquakes in the world are destructive ones.

3.3. The Time Distribution of Earthquakes

Next, we analyze the time distribution of earthquakes, including annual distribution, monthly distribution and one-day time distribution.

First, the annual distribution is analyzed, as shown in Fig. 4. It can be seen that the number of earthquakes will occur once every 2-3 years, and the annual number of earthquakes is increasing slowly.
According to the official data of the national seismological administration of the United States, the annual average number of earthquakes in the world is increasing, and the trend is rapid. The number of earthquakes from the first century to the 12th century was only 2 digits, but by the 18th and 19th centuries, the number of earthquakes directly broke through from 3 digits to 4 digits (the number of earthquakes in the 19th century was 3.3 times of that in the 18th century), and in the 20th century, it was more than the total number of earthquakes in the past 5000 years. Even taking into account the omission of earthquake records in the past, earthquakes are now unprecedented [3].

According to the data set, from 1965 to 2016, there were 28 earthquakes with magnitude 8.1 or above, accounting for the majority of 18 earthquakes in the 21st century. In 2007, there were 4 earthquakes with magnitude 8 or above in the world, which is the largest number of earthquakes with magnitude 8 or above since the 21st century. But because scientists still don't know the mechanism and principle of the earthquake, it is impossible to accurately predict where and when the next earthquake will occur [4].

As shown in Fig. 5 and Fig. 6, the number of earthquakes has no obvious relationship with the month and the time of day, and the distribution is relatively uniform. At the same time, according to statistics, earthquakes occur 1750-2000 times a month in the world, with an average of 500-600 times per hour in a day.
3.4. Equations or Fitting Functions

As shown in the Fig. 7, the magnitude of an earthquake follows the Power Law distribution, and the fitting function is $4 \times 10^9 e^{-2.46x}$. According to this formula, the larger the magnitude, the smaller the frequency of occurrence, which is the same as that of Zheng Jianzhong, Institute of Geophysics, Chinese Academy of Science [6].

The seismicity parameters of various seismic regions in China or the dominant time interval $E[T]$ of recurrence of large earthquakes within the same location range vary with the magnitude and region. The principle is that the magnitude is determined by the amount of energy released by each earthquake activity measured by seismograph.

The greater the energy released by the earthquake eruption, the less likely it will occur. The relationship between the energy released by an earthquake and its burst probability is that if the energy released by class A earthquake is twice as much as that of class B earthquake, then the frequency of class A earthquake is one fourth of that of class B earthquake.

3.5. Use of Earthquake Catalogue

Earthquake catalogue is the only reliable data that can be used to predict large earthquakes (landmark earthquakes). The monitoring data used for earthquake prediction is earthquake catalogue. From the analysis of data sources, it can be seen that the earthquake catalogue used before 1972 was mainly calibrated by ISC-GEM survey, while the earthquake catalogue used since 1973 was calibrated by the United States. At the same time, the data of world seismic survey still rarely carried out by other countries and institutions will be used.
3.6. **Relationship between Aftershock and Focal Depth of Main Shock**

According to Fig. 8, we can see that the incidence of aftershocks of the main shock of a shallow earthquake with a distance of about 0-60km is relatively high, and the occurrence rate of aftershocks varies strongly with the change of the focal depth of the main shock. The aftershock occurrence rate of moderate earthquakes from 60km to 200km is obviously decreased, which is about half of that of shallow earthquakes.

When the focal depth is in the range of 200km-500km, the aftershock occurrence rate is particularly low. The statistics here is 0%, which may be due to the incomplete data. Only the world earthquakes with $M = 5.5$ and above are counted. However, the occurrence rate of aftershocks at depths below 500 km increases suddenly, which is about twice that in the depth range of 60km to 200km.

Our visualization analysis on Python is generally consistent with the results of Steven E. Persh, Heidi Houston, Li Qinglin [5]. However, the difference is that in their study, the aftershocks of 0-100km are more, and the aftershocks of earthquakes with focal depth below 100 km are less. The aftershock occurrence rate of the main shock with focal depth between 300 km and 500 km is especially low.

Among them, the aftershock occurrence rate of the main shock with focal depth in the range of 300km to 500km is about 1/3 of the aftershock occurrence rate of the main shock between 100 km and 300 km depth (after standardized treatment, the aftershock occurrence rate of main shock from 300km to 500km is basically 0), and the aftershock occurrence rate of the main shock with depth below 500km is more than twice that of the depth of 100 km to 300 km.
4. RELATED WORK

Steven E. Persh, Heidi Houston, Li Qinglin [5] selected Harvard University moment tensor (CMT) catalogue to study the results. The time from January 1, 1979 to December 31, 2002, the focal depth of these earthquakes was $\geq 100$ km and their research results were analyzed by Wiens and Gilbert (1996), which is more traditional.

In comparison, in this paper, we use the significant earthquakes, 1965-2016 dataset [1] to study 23412 earthquakes with $m \geq 5.5$ in the world from 1965 to 2016, involving a wide range of geographical distribution and focal depth distribution.

Because of the use of computer programming language python, we can directly analyze the trend in a large number of data instantaneously, which saves a lot of time for selecting regular data for analysis, automatically analyzes the data, and visualizes the processing, making the information more intuitive.

Visual processing based on Python is easier to practice in real life, help people quickly extract useful information, and the scope of application is more common. But because the time of aftershock is uncertain, it is possible after several decades, and the location of aftershock is close to the location of main shock is only a relative concept.

It is difficult to determine the location and time of the aftershock with exact figures. Therefore, this paper only delimits a range of judgment for aftershocks, which is not particularly accurate, and some aftershock data may be ignored.

5. CONCLUSION

We get the following conclusions from the data set through visualization technology.

① The vast majority of earthquakes in the world are destructive.

② According to the research data of each century, the occurrence of earthquakes has become more and more frequent, and has been increasing exponentially since the 19th century.

③ There are 500 to 600 earthquakes per hour in the world.

④ According to the data set, the fitting function of earthquake magnitude from 1965 to 2016 is $4 \times 10^5 e^{-2.46x}$, which follows the long tail distribution, and the greater the magnitude number, the greater the occurrence time interval.

⑤ There is a correlation between the occurrence of earthquakes and the depth, in which the incidence of aftershocks of main shocks of shallow earthquakes is higher; the incidence of aftershocks of moderate earthquakes is low, in which the occurrence rate of aftershocks between 200-300km is almost zero; the depth of deep earthquakes increases sharply at 500 km.

It is difficult to predict the law of earthquake occurrence, and the mechanism and principle are difficult to explore. It is hoped that the visualization system can help managers to excavate the laws of earthquakes and take effective measures to minimize the losses and casualties caused by disasters. At the same time, it is also hoped that through the visualization technology and products, managers in various fields can recognize the laws and make efficient decisions.

REFERENCES

[1] Kaggle. Significant Earthquakes, 1965-2016[EB/OL]. https://www.kaggle.com/usgs/earthquake-database,2016.

[2] Hao Haiyan, Pan Ping. Application of Python technology in data analysis [J]. Electronic technology and software engineering, 2020 (12): 179-181

[3] Global seismic frequency and frequency data [J]. Statistical Science and practice, 2011 (03): 62

[4] Whether earthquakes are becoming more and more frequent [J]. Science Park, 2010 (09): 6

[5] Steven E. Persh, Heidi Houston, Li Qinglin; a Strong Correlation between the Aftershocks of Deep Earthquakes and the Focal Depth of Main Earthquakes[J]; South China Earthquake; Issue 03, 2005

[6] Zheng Jianzhong. Occurrence mode, time interval and probability distribution of large earthquakes in China [J]. Earthquake research, 1983 (S1): 439-447
[7] Kitchin R. Big Data, new epistemologies and paradigm shifts[J]. Big Data & Society, 2014, 1(1): 2053951714528481.

[8] Chao lemen, Xing Chunxiao, Zhang Yong. Current situation and trend of data science research [J]. Computer science, 2018, 45 (01): 1-13