Big Data Case Study

Race with Earthquake Waves:

Big Data in Earthquake Prediction and Early Warning

Student: XIA Yuchen
Student ID: 18076594g

1. Research Background

1.1 Concept Illustration

(1) Earthquake

An earthquake (also known as a quake, tremor or temblor) 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. Earthquakes can range in size from those that are so weak that they cannot be felt to those violent enough to toss people around and destroy whole cities. The seismicity, or seismic activity, of an area is the frequency, type and size of earthquakes experienced over a period of time. The word tremor is also used for non-earthquake seismic rumbling.[1]

(2) Big Data

Big data is a field that treats of ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software.[2]

Big data is high-volume, high-velocity and high-variety (and high-value) information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.

(3) Earthquake Early Warning

Earthquake early warning (EEW) systems use earthquake science and the technology of monitoring systems to alert devices and people when shaking waves generated by an earthquake are expected to arrive at their location. The seconds to minutes of advance warning can allow people and systems to take actions to protect life and property from destructive shaking. [3]

(4) Earthquake Prediction

Earthquake prediction is a branch of the science of seismology concerned with the specification of the time, location, and magnitude of future earthquakes within stated limits, and particularly "the determination of parameters for the next strong earthquake to occur in a region. [4]

(5) Earthquake data is a kind of big data

- **High Volume**

The seismic observation output data capacity is huge, the daily output is more than 200TB, and the historical data storage capacity reaches EB level.[5]

- **High Variety**

There are many seismic monitoring methods, such as volcanic and reservoir micro-seismic observations, dense array observations, seismic positioning networks, strong ground motion observation networks, and deformation disciplines (such as gravity, tilt, stress, aerial observation, GPS) , electromagnetic science, fluid science, flow observation points and macroscopic anomaly statistics, the number of seismic stations, sampling rate, and types of instruments are increasing. The earthquakes involve many fields, which are related to the evolution of material motion inside the Earth, as well as meteorological and celestial movements

• **High Velocity**

Not only the data generation of each subject is fast, but also the continuous, real-time query data flow is realized.

• **High Value**

Earthquake data can be applied to numerous subjects (geology, meteorology and astrophysics) except seismic research.

The earthquake data generally meets the four-V requirements of big data so it is supposed to be a kind of big data.

1.2 Subject Significance

Earthquake early warning can be roughly predicted in practice, and theoretically one accurate prediction does help to reduce the seismic damage to great extent. There were a large number of history prediction records, including Liaoning Haicheng Earthquake in 1975, Yunnan Longling Earthquake in 1976, Qinghai Gonghe earthquake in 1994 and Yunnan Dayao Earthquake in 2003 etc.

Meanwhile, earthquake early warning is the key point for citizens to survive a major earthquake. Theoretical studies have shown that if the warning time is 3 seconds, the casualties can be reduced by 14%; if it is 10 seconds, the casualties are reduced by 39%; if it is 20 seconds, the casualties are reduced by 63%. So the earlier warning released, the more possibility to survive. Actually, there were lots of existing case of earthquake early warning, on August 8, 2017, a magnitude 7.0 earthquake occurred in Jiuzhaigou, Sichuan, and Chengdu received an earthquake warning 71 seconds in advance; On 3 January, 2019, Sichuan Yibin Earthquake and Chengdu received warning 10 seconds in advance.[6]

The purpose of this case study is to analyse earthquake prediction as well as earthquake early prediction using big data techniques. Trying to find out the inner schedule as well as potential pros and cons of big data application among these two research orientations.

2. Big Data in Earthquake Prediction

2.1 Tree Traditional Seismic Prediction Methods

There are three main traditional seismic prediction methods: geological research method, seismic statistics method and earthquake precursor method.

(1) Geological Research Method

Studying the geological structural characteristics of the major earthquakes that have occurred should help determine where the geological background of the major earthquakes will occur. However, before some earthquakes, the geological structure was often unclear. After the earthquake, a fault was discovered and it was considered to be related to the earthquake.

(2) Seismic Statistics Method

For earthquakes that have occurred in the past, use mathematical statistics methods to discover the laws of earthquake occurrence, especially the laws of time series, and to speculate on the future based on the past.

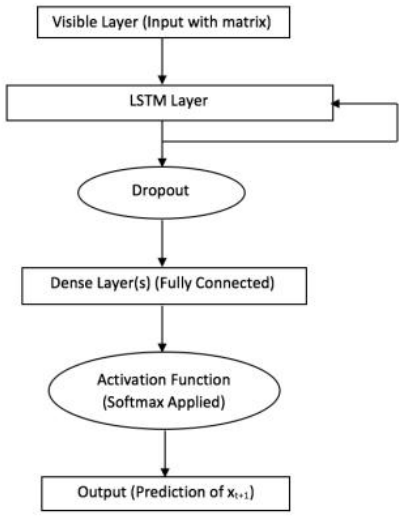
(3) Earthquake Precursor Method

Earthquakes are the rupture of the Earth's medium, so it is determined that the earthquake should belong to the physical process. Observing various parameters of the geophysical field and abnormal changes such as groundwater may find useful precursors to earthquakes. The biggest difficulty in precursor research is that natural and man-made disturbances are often encountered in observations, and the so-called precursors and earthquakes are often empirical.

2.2 Latest Seismic Method with LSTM

There was a new seismic prediction method using the Long Short-Term Memory: Earthquake Prediction based on Spatio-Temporal Data Mining: An LSTM Network Approach by Qianlong Wang, Yifan Guo, Lixing Yu, and Pan Li in 2017. [7]

In this method, long short-term memory (LSTM) networks to learn the spatio-temporal relationship among earthquakes in different locations and make predictions because earthquakes are spatially and temporally correlated because of the crust movement.



Graph 2.1 system flow chart

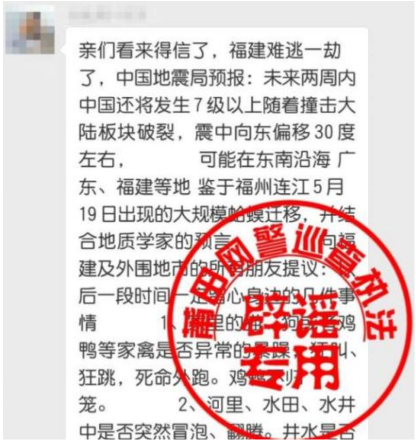
2.3 Chapter Summary

On the one hand, the prediction accuracy of latest LSTM network approach is actually not too bad (71.81%) compared with those traditional ones. Considering the particularity of big data, there is still much space for improvement as training data grows. On the other hand, however, there are constraints in both deep learning method and the old-school methods.

唐山地震预测

“1975年12月(距唐山地震半年多)，河北省地震地质大队1976年地震趋势意见上报国家地震局：从河北省乐亭至辽宁省敖汉旗—锦州—带及其东南渤海海域，可能发生大于6级地震。”

Graph 2.2 Unable to quantitatively predict, the specific earthquake magnitude, time and location is unclear. It is just a rough possibility and meanwhile wrong forecast bring detrimental effect of “wolf is coming”.



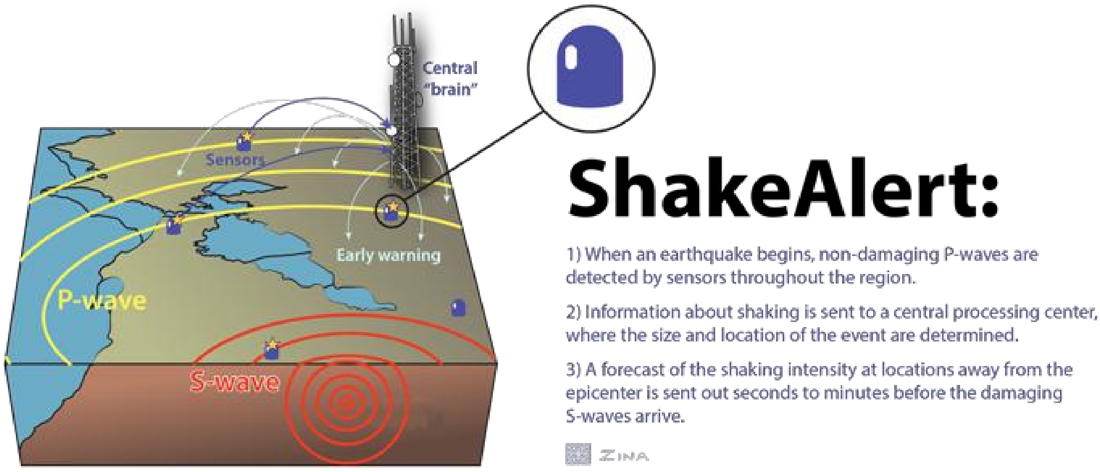
Graph 2.3 and Graph 2.4 Internet rumors about earthquake, which will undermine social stability to some extent.

3. Big Data in Earthquake Early Warning

3.1 Earthquake Early Warning

Earthquake early warning is the rapid detection of earthquakes, real-time assessment of the shaking hazard, and notification of people prior to shaking. Warning times range from a few seconds to a few minutes depending on your location and how big the earthquake is. The further you are away from the epicenter, the more warning time. The bigger the earthquake, the stronger the shaking at greater distances. An early warning should tell you how strong the shaking will be at your location, and how long until that shaking starts (the warning time). [8]

There are many existing EEW system working in this world to protect citizens, including the ICL (Institute of Care-life) Earthquake Warning Technology System in China, ElarmS (Earthquake Alarms Systems) in America and Earthquake Early Warning System in Japan.



Graph 3.1 The mechanism of EEW system

3.2 ICL(Institute of Care-life)

This is the world's biggest earthquake warning system. (covers an area of 2.2 million square kilometers, covers 90% of the population of the earthquake-stricken areas in mainland China)

There are four module of this system:

(1) **Seismic monitoring**: High density distributed sensor network installed in areas where crustal

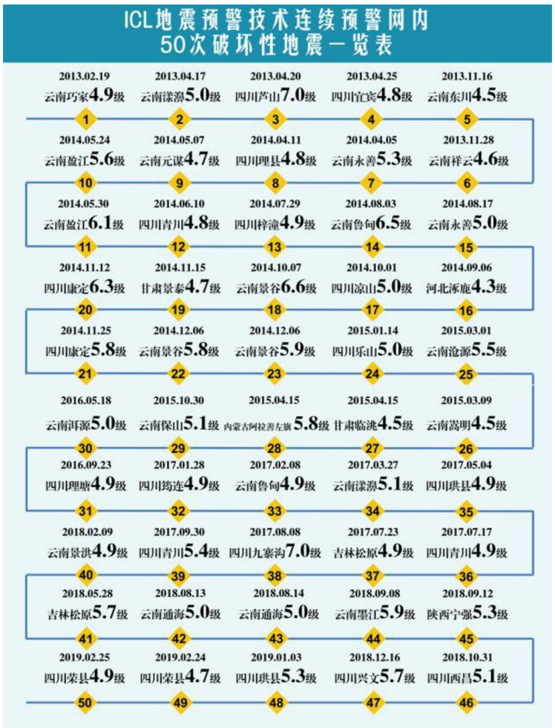
activity is frequent.

(2) **Information analysis and processing**: 24-hour data center to processing and analyse the data from seismic monitoring network.

(3) **Information release**: EEW center release the warning information after a series process of processing and analysis.

(4) **information reception and application**: Warning information will be sent to kinds of facilities(smart-phone, computer, radio, TV, factory, power plant, nuclear facility, police office, fire station)

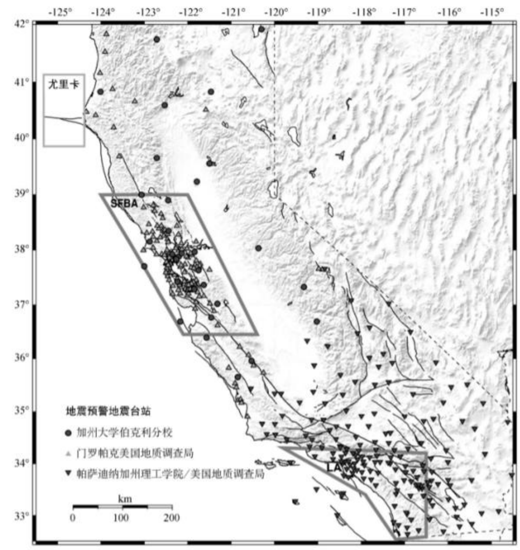
Performance of ICL system is actually good in the area of Sichuan and Yunnan provinces in China. According to the released data, the ICL early warning system has made more than 50 warnings (see the table below), including Sichuan Lushan 7.0 earthquake, Sichuan Jiuzhaigou 7.0 earthquake (71s in advance for Chengdu), Sichuan Gongxian 5.3 earthquake (0-27s in advance for)The early warning system has realized the simultaneous release of early warning information through smart phones(with specific APP), radio and television, micro-blogs, and earthquake warning information receiving servers.



Graph 3.2 List of the last 50 destructive earthquakes

3.3 ElarmS

ElarmS, short for Earthquake Alarms Systems, is a methodology designed to provide alerts in California, Oregon, Washington, and other earthquake prone regions around the world. It is developed by Professor Allen at the Seismology Laboratory at the University of California, Berkeley, tested offline using data from Japan, Taiwan, Italy, Alaska, the Pacific Northwest and California, real-time data from California Integrated Seismic Network(CISN) and is now being deployed in California.



Graph 3.3 Distribution of earthquake early warning stations

There are also four module of this system:
(1) **Sensor Network**: Seismic monitoring stations collect wave data(P-wave, frequency, amplitude) and send it to data center of CIT and UC Berkeley
(2) **Association Filter**: Retrieve input event in the local event base, comparing the consistency of p-wave then get feedback(if matching, it is a aftershock; if not, creating a new earthquake event).
(3) **Source Parameter Estimation**: Estimate earthquake magnitude and recheck the event(determining if it is a aftershock or not).
(4) **Alarm Filter**: Three conditions before releasing the warning: Magnitude is greater than; At least 4 stations are triggered; 40% stations besides the epicenter are triggered.

Performance of ElarmS is not bad, from the data point of view, the magnitude median error is -0.5 ± 0.39 and the average time value is -0.29 ± 1.16 while standard deviation time error is -0.10 ± 1.59. However, there were still five false reports since 2012 (neither of them in the center of California), also three missing earthquake reports were recorded.

4. The Other Side of Big Data
4.1 Constraint
(1) Unstructured Data
The structures of earthquake dataset vary between different government departments, which does cost a lot in data mining and data processing.
(2) Diverse Prediction Effect
The government has given part of the disaster prediction and system maintenance work to the enterprise, and the company will simplify the data mining work due to cost considerations, which makes the captured data seriously one-sided and affects the objectivity of the prediction results.
(3) Limited Sensor Network
The performance of EEW system and earthquake predicting system is constrained by the input data, which means sensor network deployment and maintenance will consume a lot of cost.
(4) Data Island Problem

Earthquake data collected by the authority is unwilling to be open, dare not open, cannot open and will not open, which form a information barrier.

4.2 Drawbacks

(1) Data Security Issue

The deployment of large-scale sensor networks may lead to the leakage of personal information.

(2) National Security Issue

Some seismic data may involve confidential national security issues. (such as data generated by underground nuclear tests and mapping data of military facilities)

(3) False Alarm Issue

Financial loss, life safety is threatened, distrust of the system.

4.3 Improvement

(1) Standard Warning Threshold

Widespread sensor network is the baseline of an effective system, and a scientific and reasonable early warning threshold standard system to avoid false alarms and missed alarms. Globally coordinated big data technology-based disaster warning interactive sharing system

(2) Big Data Alliance

International specialized agencies established by countries are responsible for formulating and implementing data sub-library construction specifications and standards.

Break the information barrier of each independent database by means of Internet distributed interoperability technology

(3) Legal security system

It is urgent to formulate and improve the legal norm system for specifying the specific rights and responsibilities distribution and traceability accountability schemes in various disaster management departments, many big data affiliates, other organizations and public public disaster prediction and police communication.

5. Conclusion

Big data does works in earthquake prediction and early warning, and compared with traditional prediction, there is a great improvement (more feasible and effective). There are also practical application in specific area, generally big data seismic prediction methods will become a trend. However, big data brings convenience with hazard, data security and national security issue are the most significant problems and sensor network constraint the performance of system.

From my point of view, regional cooperation can be the way out, global cooperation is not realistic now considering of political and national security problem, so construct a regional sensor network (Big data technology helps to reduce costs and increase efficiency) and structured database are the prerequisite for effective earthquake prediction networks.

6. Reference

[1] Wikipedia: https://en.wikipedia.org/wiki/Earthquake

[2] Wikipedia: https://en.wikipedia.org/wiki/Big_data

[3] USGS: https://earthquake.usgs.gov/

[4] Wikipedia: https://en.wikipedia.org/wiki/Earthquake_prediction

[5] 川崎一郎．２０１３．何谓慢地震：探索巨大地震预报的可能性［Ｍ］．北京：地震出版社
[6] 成都高新减灾研究所: http://www.365icl.com/
[7] Qianlong Wang, Yifan Guo, Lixing Yu, and Pan Li, Earthquake Prediction based on Spatio-Temporal Data Mining: An LSTM Network Approach
[8] Berkeley Seismology Lab: http://seismo.berkeley.edu/research/eew_basics.html