Finding Patterns in Earthquake Data

Frank Cusano - Computer and Systems Engineer

May 3, 2017

0.1 Introduction

Every year there are more disasters happening all around the world causing massive amounts of damage that takes lives and costs people large amounts of money. Wouldn't it be nice to be able to predict when these events were going to occur and with high accuracy so that people in the area can evacuate.

Predicting events such as an earthquake is theoretically impossible because an earthquake occurs instantly and without warning. Predicting events such as a hurricane is a little easier to accomplish due to the fact we can monitor weather patterns and watch hurricanes form using satellite images. This is why earthquakes cause so much damage, since there is no physical warning of when they occur, people don't have time to prepare. Earthquake data can be put to use in order to attempt to predict when the next earthquake will occur by analyzing the patterns that show up within the data. The data sets are large and the algorithms for finding patterns within the data are computationally expensive which is why parallel computing can be used in order to expedite the process which will allow for instantaneous and accurate predictions.

Figure 1: Earthquakes in April 2017

Figure 1 shows a map of all of the earthquakes in the past month. The size of the circle represents the magnitude of the earthquake. There were over 9,000 earthquakes in April of 2017 alone. There is, on average, one earthquake per month that is larger than a 5.5 in magnitude. This is a lot of data to work with, with a lot of information hidden within it.

There are a lot of ways that people attempt to predict earthquakes, but none of them are perfect. The main goal of this project is to bring multiple methods together to create a more accurate way of predicting earthquakes. In an article taken from WiseGeek.com they mention that many small earthquakes can suggest that there is potential of a large one happening soon due to the fact that there is a large amount of energy stored within the tectonic plates.

> *"Small earthquakes suggest a certain level of fault line activity and pressure building up, and scientists regularly suggest that we must all prepare for the Big One, since it could occur at any time."*
>
> \- WiseGeek

A method for determining where earthquakes occur is by looking at fault lines. Figure 2 shows a map of all of the fault

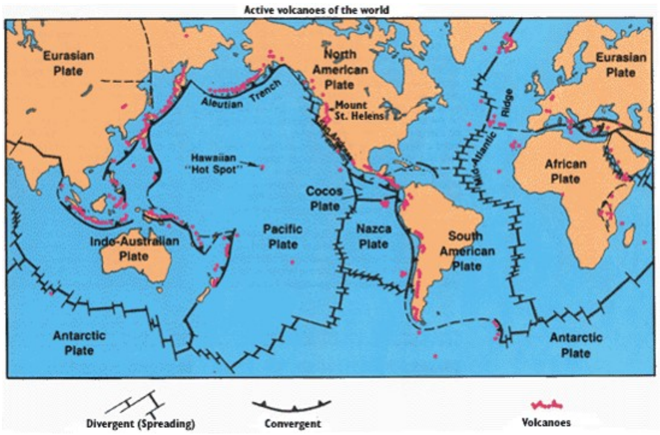

Figure 2: Fault Lines

lines on the earth which can be compared to Figure 1 which shows that most of the earthquakes do take place along the fault lines. Another method that can be used to determine when an earthquake will occur is by monitoring volcanic activity. The movement of magma causes the pressure to build up between the plates which eventually leads to them snapping, causing an earthquake.

*0.2 **Implementation***

After finding the data I used javascript and the Mapbox API to map the data shown in Figure 1 so that I can get my own visual representation of the earthquakes around the world. After mapping the data I switched to c++ for predicting where the next earthquake would occur. I used OpenMP to parallelize my program.

After parsing the data I kept track of the crucial data such as the magnitude, location, and date. I grouped the data by region so that I can analyze that area to determine if it was vulnerable for an earthquake. Regions were determined by latitude and longitude and were split by degrees of 10. If there was a large earthquake (> 5.5) in that region than that region would be removed form the list because the odds are good that it would not have another large earthquake soon. This is because all of the energy that was built up in the ground was released. If there were many small earthquakes in a region (> 100) than I would save the region because that indicates that a large earthquake might happen soon. Once I found the area with the most earthquakes that were small, I averaged out the location of all of those earthquakes to make my prediction of where a future earthquake might occur. In order to determine the magnitude of my predicted earthquake I averaged out the magnitudes of all of the large earthquakes that occurred that month.

There are a few parts of the program that I was able to run in Parallel. Since a lot of the data was independent it made it easier to run the program in parallel. First off, parsing the done was done in parallel because it doesn't matter the order in which the data is stored in the vector. Finding the location was also able to be done in parallel because it was two nested for loops that does not rely on previously calculated data. Each earthquake location had to be compared which each other earthquake location only once and it didn't matter the order in which this was done.

0.3 Results

Figure 3: Output of Program

The results of my implantation were somewhat to be expected. The blue circle shown in Figure 3 indicates the prediction. This seems to make sense due to all of the small earthquakes that have occurred around Alaska in the past month. Unfortunately, I did not receive the same answer when running my program in parallel but I did receive a speedup of about 2x.

	Serial	Parallel
Time(s)	0.031	0.016

0.4 Problems Encountered

Although this seems like an easy thing to implement and accomplish there were many setbacks that I faced. When I first began testing my program I used a smaller data sample which was working fine. After attempting my program on the larger data it was not working. I later figured out that not all of the data that I was using was formatted correctly so I modified my data parser to ignore incorrectly formatted inputs. Even though this removed some of the earthquakes from the calculation of the prediction, I don't believe that this effected the answer in any way because there were so many earthquakes that losing a few of them would only have a very small impact on the final outcome.

Another issue I faced was using with OpenMP. For some reason my program would stop running after the first #pragma parallel for loop. I was able to solve this issue by separating the #pragma parallel and the #pragma for so that they weren't on the same line. Ultimately my program did not produce the correct answer when running with OpenMP, which is an issue that I was unable to resolve. I think that I there are a few ways in which I would be able to make this work better, including, the use of proper critical statements, serializing the program where necessary, and using better memory management techniques.

My last issue that I faced was finding volcano data to use to aid my prediction algorithm. Unfortunately I was unable to find useful volcanic data in the form that I needed so I wasn't able to implement that part of my program. If I were able to implement volcano data I would have kept track of the volcano location and activity. If the volcano had was having significant activity in the past month than I would consider that region to be more likely to have an earthquake.

0.5 Ways to Improve

Going forward I believe that this project has very big potential but there are many ways in which it can be improved upon. Right now it only can predict when the next earthquake will occur, and only in one area. In the future I would like to implement a way to predict more than just one earthquake by analyzing multiple clusters of earthquakes in different areas. Also I would like to be able to predict a region of time which can warn people about when the earthquake might occur and not just where.

As stated earlier I would like to improve my algorithm by adding in volcano data since that information can be helpful for predicting earthquakes. In addition to volcano data, using older earthquake data could also improve the accuracy of the predictions. Unfortunately I was unable to find older earthquake data but I'm sure it is available somewhere. With older earthquake data it would be possible to detect trends in certain earthquake prone regions.

The last idea for an improvement would be to implement this idea with all natural disasters. Although other disasters such as hurricanes and tornadoes are easier to predict by observing weather patterns, using data can greatly improve the accuracy. Patterns in the data from previous events can store a lot of information as to what might occur in the future.

0.6 Conclusion

With some additional time and effort put into this project I believe it can turn out to be a very useful tool in the future. Having an accurate way to predict natural disasters could potentially save thousands of lives. Not only could it save lives but it can also improve the lives of many countries who are forced to start over every time tragedy strikes causing them to remain at a low standard of living because of their unfortunate location on earth. An example of this would be Haiti, which was hit with a massive earthquake in 2010. Haiti was a poor country to begin with and their location leaves them vulnerable to many earthquakes and tsunamis. Had this been able to be predicted, preventative measures could have been taken to avoid so much damage. A country like Haiti has to start over every time a disaster like this strikes causing them to remain and poor country.

Earthquakes and tsunamis are also a large threat here in the United States. California sits on one of the largest faults in the world and it will give way one day, but we don't know when. When it does strike some experts say that the resulting earthquake can be as large as a magnitude 9 on the Richter scale and which is represented below in Figure 4.

Figure 4: Potential California Earthquake

An earthquake of this size would absolutely devastate California and would be felt form places in central United States. Using my program, if we are able to get an accurate prediction as to when this will strike we will be able to be better prepared for when it does. To get the most accurate prediction we would have to use tons of historical data that will cause the program to run slowly. This is why parallel computing can be used in order to be able to make these predictions instantaneously every time new information arises or new data is added.

0.7 References

1. Ellis-Christensen, Tricia. "Is It True That Little Earthquakes are Precursors to Big Earthquakes?" Wisegeek. Ed. O. Wallace. N.p., 9 Apr. 2017. Web. 28 Apr. 2017.

2. "Earthquake Hazards Program." Earthquake.usgs.gov. N.p., n.d. Web. 20 Apr. 2017.

3. Morrison, Jessica. "7 Biggest Earthquakes in California History." National Geographic. National Geographic, 25 Aug. 2014. Web. 30 Apr. 2017.

4. "Adelaide Para-Fault Line." Word Press. Parafaultlinecassandraurgl, 8 June 2015. Web. 30 Apr. 2017.

5. Coding Challenge #57: Mapping Earthquake Data. Perf. Daniel Shiffman. YouTube. N.p., 13 Feb. 2017. Web. 28 Apr. 2017.

6. Lignier, Roxanne, and David Moore. "10 Major Natural Disasters Predicted In The Near Future." List Verse. N.p., 25 May 2015. Web. 28 Apr. 2017.

7. Gautier, Michael. "Pivot US Geological Survey Earthquake Data with C." Word Press. N.p., 7 Mar. 2015. Web. 28 Apr. 2017.