# Big Data Applications for Earthquake Predictions

Haridas M

Scientific communities from all parts of the world have been trying for some time now to design an earthquake prediction mechanism. Though they have not been able to make clear headway and give a model, the way the scientific study is going about these specific areas of Research and Development (R&D), the day is not far when loss to life and property due to earthquakes will become a thing of the past. Most of us know that earthquakes may occur, but just do not know where or when. Traditional methods of predicting earthquakes have largely been inefficient, as is evident from the many lives and properties that have been lost over the last century, most recently, in Nepal.[1] How useful it would be if scientists could predict earthquakes with greater accuracy so that the people in the particular region could be prepared for them? The recent earthquake in Nepal that claimed more than 8,000 lives is testimony to the fury of such a natural disaster and the lack of preparedness of the people to cope with it. Scientists are divided on the development of a model for earthquake prediction, with some claiming that it will be possible to do so and others arguing that this research is an exercise in futility.

Colonel **Haridas M** is Senior Fellow, Centre for Land Warfare Studies, New Delhi.

**N.B.** The views expressed in this article are those of the author in his personal capacity and do not carry any official endorsement.

**Scientists are divided on the development of a model for earthquake prediction, with some claiming that it will be possible to do so, while others arguing that this research is an exercise in futility.**

## Estimating Losses

We know that earthquakes can occur. We just don't know how to tell people early enough to avoid the catastrophe ahead. Around the world, more than 13,000 people are killed each year in earthquakes, and almost five million have their lives affected by injury or loss of property. Add to that $12 billion a year in economic losses to the global economy (the average annual toll between 1980 and 2008).[2] Earthquakes (and the tsunamis they cause when they occur at sea) are devastating in their cost to human lives and the economy – the 2011 Tohoku earthquake off Japan's east coast killed almost 16,000 people, destroyed over 127,290 buildings and made almost a quarter of a million people homeless. An increase in the accuracy of prediction will save lives, by enabling evacuation and disaster relief efforts to allocate resources more effectively.[3] Disasters like earthquakes and volcanoes not only crumble the economy of the nations but leave behind a saga of human disaster, sorrow and tragedy which for the affected becomes difficult to forget.

## Predicting Earthquakes in the Past: Case Studies

The United States Geological Survey (USGS) is of the opinion that scientists will never be able to predict earthquakes accurately. They can at most say that an earthquake will hit a place some time in the future, but not exactly when. However, big data analytics is providing a leap in the accuracy of earthquake predictions, with recent predictions being 90 percent accurate. A warning was issued on February 22, 2015, that an earthquake of a magnitude of around 6.5 on the Richter scale would shortly hit the Indonesian island of Sumatra. On March 03, the island was rocked by a 6.4

**Big data analytics is providing a leap in the accuracy of earthquake predictions, with recent predictions being 90 percent accurate.**

magnitude earthquake. Companies using satellite data are increasingly able to forecast major earthquakes from one to 30 days before they occur in all key seismic-prone areas by analysing satellite imageries and past earthquake data patterns. What are measured are abnormalities in the atmosphere caused by the release of energy and gases, which are often detectable well before the earthquake occurs. Prediction of future Nepal earthquakes was tried to some extent by the USGS, which monitors earthquakes across the world, after the massive earthquake which struck Nepal, and it is estimated , that a few more earthquakes, ranging from a magnitude 5 to 6 on the Richter scale, were likely to occur and that the chance of at least one such aftershock was about 62 percent. USGS expected the event to occur within the month from May 27 to June 26, 2015. Aftershocks are a normal occurrence after large earthquakes, and are expected to continue in Nepal but, with the passage of time, will occur less often, USGS has said. However, there is no way to predict the exact date or time of an earthquake or aftershock. USGS has produced a statistical analysis of expected aftershocks based on past earthquakes and the aftershocks recorded in Nepal. Apart from that, USGS has predicted aftershocks based on the nature of earlier earthquakes and aftershocks. They have warned the people of Nepal that they need to be alert, but not to panic. Although aftershocks may occur less often, people should remain aware of their possible occurrence in the coming weeks and months, especially when working in or around vulnerable structures or in landslide-prone areas.[4]

Chinese scientists, based on some precursors, have in the past been able to make very short-term predictions, as seen during the 7.3 magnitude earthquake that hit the Haicheng-Yingkow region on February 04, 1975. The prediction was submitted on the day

**The main method of the Chinese prediction was to look for the interval between medium-force earthquakes.**

that the earthquake occurred. The main method of the Chinese prediction was to look for the interval between medium-force earthquakes. If this is shorter than 47 days and more than five such earthquakes occur consecutively, another one with a magnitude above 7 is likely to strike within six months in the region. Chinese scientists were able to accurately predict the 6.1 magnitude Afghanistan earthquake on February 04, 1998. But the same method did not deliver on the 7.8 magnitude Tangshan earthquake in northern China that killed over 240,000 people in 1976. While acknowledging the predictions, experts asked the Chinese State Seismological Bureau to submit the documents or publications used for the predictions, but nothing has been made public till date.

In India, similar or low tremors, as cited in the Chinese method, had been recorded preceding the Nepal earthquake. Since April 01, as many as 11 mild tremors had been registered by the Indian Meteorological Department (IMD) in Assam, Manipur and Uttarakhand, all in a highly seismic zone. However, the Indian authorities did not see any connection. Further, the head of the Department of Geology at Pune University, told the *Indian Express* that it was difficult to say if the smaller tremors were precursors to the Nepal earthquake or not. He further elaborated that earthquakes are beyond predictability but what can be said for sure is that the Himalayan region is active in terms of earthquakes.[5]

A few changes noticed have also been recorded to have given certain kinds of indication, but there is no recorded study of such phenomena. What will be required is a mechanism to monitor these changes routinely which can then be considered earthquake precursors. Some of these are as follows:[6]

- **Ground Water Levels:** Changing water levels in deep wells are recognised as a precursor to earthquakes. This happens a few days before the earthquake strikes.
- **Chemical Changes in Ground Water:** The chemical composition of ground water is affected by seismic events. Researchers at the University of Tokyo tested ground water after the earthquake and the results of the study showed that the composition of the water had changed significantly during the period around the earthquake area. They observed that the chloride concentration was almost constant, but the levels of sulphate showed a rise.
- **Radon Gas in Ground Water Wells:** An increase in the level of radon gas in wells is a precursor of earthquakes, recognised by research groups.

There is no sensor mechanism or a grid of sensors to record such changes and alert agencies to forewarn people when such changes are noticed. It is a costly proposition, and governments have not moved in this direction because of various reasons. Dogs, snakes and birds can sense an earthquake much before humans can. The common toad (*Bufo bufo*) can detect seismic activity days in advance of an earthquake, as recorded during the 2009 L'Aquila, Italy, earthquake. A 2010 study published in the *Journal of Zoology* found that 96 percent of male toads in a population abandoned their breeding site five days before the earthquake struck almost 74 km away. Can humans borrow detecting techniques from the animal world.[7] Only the future will only answer this question.

## Technological Evolution Related to Earthquake Prediction

### *Prediction Based on Empirical Data*

Earthquake precursor data so far accumulated in Japan have shown that the larger the main shock magnitude, the larger the distance between the epicentre and an observation point where a precursor is observed. It is possible to empirically establish an approximate relationship between maximum detectable distance, **Dmax**, and main shock of magnitude **M**, the relationship being different from discipline to discipline of the precursor. When a number of precursors are observed, we may draw circles with a radius equal to **Dmax** peculiar to each precursor discipline, centred at the respective observation points on the condition that **M** takes on a certain value. In that case, the epicentre of a future earthquake can be located in the area which is common to all of the circles. If **M** is too small, some of the circles do not overlap. If it is too large, an epicentral area which is too wide to be realistic will have to be assumed. In this way, an approximate epicentre location and a rough value of the main shock magnitude can be assessed. Applying the procedure to the precursors of the Izu-Oshima Kinkai (M = 7.0, 1978) and the Niigata (M = 7.5, 1964) earthquakes in Japan, earthquake prediction was achieved with remarkable success.

The probability of an earthquake occurring in a specified time interval can be evaluated as a function of time when a precursor is observed. For such an evaluation, the Japanese relied either on the log **T**-**M** relationship, with a prescribed value of **M**, or on the frequency distribution of log **T** which was empirically obtained depending on precursor disciplines, **T** being the precursor time. When a number of precursors were observed one by one, changes in the synthetic probability of earthquake occurrence were estimated according to the existing formula (Utsu, 1977). As an example, such probability change with time was estimated, making use of the precursor data for the Izu-Oshima Kinkai earthquake. It turned out that the synthetic probability generally increased with time. When a long-term precursor happened to be observed, the probability dropped discontinuously, being followed by a gradual rise again. The temporal change in probability, therefore, showed considerable fluctuations. This was especially so when a short interval for probability evaluation was specified. Because of the uncertain preliminary probability involved in the evaluation and an unknown rate of false precursor signals against true ones, the absolute value of probability, thus, evaluated was not perfectly reliable. The temporal change in probability provided some clues for the actual publication of earthquake prediction.[8]

**Data mining can be applied in finding the consequences of earthquakes and, hence, alerting the public, and continues to be used to find relationships in data patterns.**

Since 1983, continuous monitoring of the electro-telluric field has been carried out using an array of measuring stations located at various sites in Greece. The basic physical properties of the transient changes—Seismic Electric Signals (SES)—in the electro-telluric field that are forerunners of earthquakes were first described six years ago. Since then, a large body of data has been collected, resulting in new insight into various aspects of the method. Based on this, predictions were officially issued in Greece during three years (January 01, 1987–November 30, 1989). Public warnings were issued well before the most destructive seismic activity.[9] The degree of success has not been described in the study.

## Data Mining from Past Earthquake Data Bases

Advanced technologies in networks have enabled the communication of large volumes of data across the world. This resulted in a need of tools and technologies for effectively analysing the scientific data sets with the objective of interpreting the underlying physical phenomena and insights in terms of patterns. Data mining applications in geology and geophysics have achieved significant success in areas such as weather prediction, mineral prospecting, ecology, modelling, etc and finally predicting the earthquakes from satellite maps. Data mining consists of an evolving set of techniques that can be used to extract valuable information and

**With advancements in computing speed, individual agent modelling systems have become capable of simulating human behaviour or reactions to given stimuli or scenarios.**

knowledge from massive volumes of data. The basic problem in this class of systems is the unobservable dynamics with respect to earthquakes. Data mining can be applied in finding the consequences of earthquakes and, hence, alerting the public, and continues to be used to find relationships in data patterns. The relationships in data patterns can be analysed via two types of models: descriptive and predictive models.

- **Predictive Model:** This encompasses a variety of statistical techniques from modelling, machine learning, and data mining that analyse current and historical facts to make predictions about future, or otherwise unknown, events. Predictive models are models of the relation between the specific performance of a unit in a sample and one or more known attributes and features of the unit. The objective of the model is to assess the likelihood that a similar unit in a different sample will exhibit the specific performance. This category encompasses models in many areas, such as marketing, where they seek out subtle data patterns to answer questions about customer performance, or fraud detection models. Predictive models often perform calculations during live transactions, for example, to evaluate the risk or opportunity of a given customer or transaction, in order to guide a decision. With advancements in computing speed, individual agent modelling systems have become capable of simulating human behaviour or reactions to given stimuli or scenarios. Based on the similarities of the pattern, the predictions are made.
- **Descriptive Model:** This quantifies relationships in data in a way that is often used to classify customers or prospects into groups. Unlike predictive models that focus on predicting a single

customer behaviour (such as credit risk), descriptive models identify many different relationships between customers or products. Descriptive models do not rank-order customers by the likelihood of their taking a particular action the way predictive models do. Instead, descriptive models can be used, for example, to categorise customers by their product preferences and life stage. Descriptive modelling tools can be utilised to develop further models that can simulate a large number of individualised agents and make predictions. For earthquake predictions too, the data can be segregated into groups and the severity of the earthquakes can be predicted.

**The results indicated that lithology had the strongest influence on landsliding, particularly when the rock is highly fractured, such as in shale, slate, caustic sediments, and limestone and dolomite.**

## Movement of Tectonic Plates and Earthquake Prediction

Scientists today predict where major earthquakes are likely to occur, based on the movement of the plates in the Earth and the location of fault zones. They calculate earthquake probabilities by looking at the history of earthquakes in the region and detecting where pressure is building along fault lines. These can go wrong as a strain released along a section of the fault line can transfer strain to another section. This is what happened in the recent Nepal earthquake, say French scientists, noting that the 1934 earthquake on the eastern segment had moved a part of the strain to the eastern section where the latest earthquake was triggered. Some scientists have proposed studying electromagnetic fields, which, they suggest, change just before an earthquake. Gas seepage and tilting of the ground are some other areas of study.

## GIS Technology-Based Earthquake Predictions

The October 08, 2005 Kashmir earthquake triggered several thousand landslides throughout the Himalayas of northern Pakistan and India. These were concentrated in six different geomorphic–geologic–anthropogenic settings. A spatial database, which included 2,252 landslides, was developed and analysed, using Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER) satellite imagery and Geographical Information System (GIS) technology. A multi-criterion evaluation was applied to determine the significance of event-controlling parameters in triggering the landslides. The parameters included lithology, faults, slope gradient, slope aspect, elevation, land cover, rivers and roads. The results showed four classes of landslide susceptibility. Furthermore, they indicated that lithology had the strongest influence on landsliding, particularly when the rock is highly fractured, such as in shale, slate, caustic sediments, and limestone and dolomite. Moreover, the proximity of the landslides to faults, rivers, and roads was also an important factor in helping to initiate failures. In addition, landslides occurred particularly in moderate elevations on south facing slopes. Shrub land, grass land, and also agricultural land were highly susceptible to failures, while forested slopes had few landslides. One-third of the study area was highly or very highly susceptible to future landsliding and required immediate mitigation action. The rest of the region had a low or moderate susceptibility to landsliding and remained relatively stable. This study supports the view that (1) earthquake-triggered landslides are concentrated in specific zones associated with event-controlling parameters; and (2) in the Western Himalayas, deforestation and road construction contributed significantly to landsliding during and shortly after earthquakes.

## Big Data-Based Applications for Earthquake Prediction and Minimising Loss

Academics often put forward arguments that accurate earthquake prediction is inherently impossible, as conditions for potential seismic disturbance exist along all tectonic fault lines, and a build-up of small-scale seismic activity can effectively trigger larger, more devastating quakes at any point. However, all this is changing. Big data analysis has opened up the game to a new breed of earthquake forecasters using satellite and atmospheric data combined with statistical analysis. And their striking results seem to be proving the naysayers wrong. Predictive analytics is the practice of extracting meaningful information from present as well as historical data sets. The practice in the predictive analytics market determines the patterns of data and predicts the future outcomes and trends. Basically, predictive models and analysis are used to estimate future probabilities with an acceptable level of consistency. Predictive models are used to evaluate current data and historical facts of customers, products, and partners. The analysis helps to understand the probable risks and opportunities in a market.[10] Predictive analytics, based on the Bayesian framework for probabilistic assessment of the initiation of seismic soil liquefaction can be carried out. A database, consisting of post-earthquake field observations of soil performance, in conjunction with *in situ* “index” test results can be used for the development of probabilistically-based seismic soil liquefaction initiation correlations. The proposed stochastic model will allow full and consistent representation of all relevant uncertainties including (a) measurement / estimation errors; (b) model imperfection; (c) statistical uncertainty; and (d) inherent variability. Different sets of probabilistic liquefaction boundary curves are developed for the seismic soil liquefaction initiation hazard problem, representing various

**Predictive models are used to evaluate current data and historical facts of customers, products, and partners. The analysis helps to understand the probable risks and opportunities in a market.**

sources of uncertainty that are intrinsic to the problem. The resulting correlations will represent a significant improvement over prior efforts, producing predictive relationships with enhanced accuracy and greatly reduced overall model uncertainty.

## Particles in Space to Predict Earthquakes

Students of the Indian Institute of Technology (IIT), Madras, are working on a satellite-based project to validate a theory that pressure generated in the tectonic plates before an earthquake emits low frequency waves which interact with the Van Allen belt (the Van Allen belt is a radiation belt starting from an estimated 650 km from the Earth's surface. The belt contains high-energy particles and can absorb energy from the sun. It can damage satellites that pass through the region and may also affect communication systems or the electric power grid in extreme cases) leading to a sudden discharge of charged particles. The fluctuations in the particles could be due to various reasons and one among them is believed to be earthquakes. This theory, which is yet to be accepted by the scientific community, if proven, will go a long way in developing an earthquake prediction model. The satellite will detect particles when they precipitate towards the low earth orbit. Data gathered at the ground station, which will be located in the IIT-M campus, will be corroborated with any natural event that may have occurred in a specific location on Earth during that time. Even if they are not able to come out with a functional model to predict earthquakes, they will be able to gather a lot of data of immense value which can be used in the future for big data analytics.[11]

When factors related to earthquakes are monitored and analysed constantly with big data, even the smallest changes can send warning signals as big data has the capability to capture it with greater precision. Already, some companies are working on using big data to predict earthquakes. M/s Terra Seismic claims to predict earthquakes with 90

**To prevent catastrophic losses, it is important to harness the power of emerging technologies and big data to predict the time and location of earthquakes with greater accuracy.**

percent accuracy using satellite data and other cues from nature. For example, they had predicted the occurrence of a 6.5 magnitude earthquake in Sumatra on February 22 of this year. On March 03, an earthquake occurred in Sumatra that measured 6.4 on the Richter scale.[12] They also claim to have forecast Tarapaca, Chile's mega earthquake (magnitude 8.1), Guerrero, Mexico's 7.2 earthquake and the 6.4 magnitude earthquake in Indonesia nine days before it hit on March 03. They offer several levels of service to subscribing clients; Terra Seismic offers its current predictions for free through its Quakehunters.com portal. The predictions are very useful for insurance companies, which can use them to accurately assess risk, and ensure that assets are covered in an efficient way. Terra Seismic clients include government agencies, insurers, hedge funds and multinational corporations.

To prevent catastrophic losses, it is important to harness the power of emerging technologies and big data to predict the time and location of earthquakes with greater accuracy. Big data makes it easier than ever before to collect information about a range of different events that point to the likelihood of the occurrence of an earthquake. Nature is always giving cues about the occurrence of events, and it is simply up to us to tune in to these cues so that we can act accordingly. Since these cues are widespread, it is best to use big data collectively to bring in this data to a central location so that analysis and the resulting predictions are more accurate. Some common information that can be tracked by big data is the movement of animals and the atmospheric conditions preceding earthquakes

A case in point is the Haicheng earthquake.[13] It occurred in eastern China on February 04, 1975. Just prior to this earthquake, the

**Big data analytics has successfully proliferated into applications to support customer recommendations, customer value and churn management, campaign optimisation, and fraud detection.**

temperatures were high and the pressure was abnormal. Many snakes and rodents also emerged from the ground as a warning sign. With this information, the State Seismological Bureau (SSB) was able to predict an earthquake, thus, helping to save many lives. However, this prediction was issued on the day when the earthquake occurred, so it did cause heavy loss of property. Had this earthquake been predicted a few days earlier, it could have been possible to completely evacuate the affected cities, and this is exactly where big data fits in.

The future of earthquake predictions lies in applications based on big data analytics. The technology innovations in the field of big data are showing promise of consolidating and stabilising around predictive analytics-based applications. Nevertheless, the emerging market for big data applications has been sustained by professional services, service bureaus and profitable applications in verticals such as retail, consumer finance, telecommunications, travel and leisure, and related analytic applications. Big data analytics has successfully proliferated into applications to support customer recommendations, customer value and churn management, campaign optimisation, and fraud detection. On the product side, success stories in demand planning, just in time inventory and market basket optimisation are the staples of big data analytics. The problem of earthquake prediction is based on data extraction of precursory phenomena and is a highly challenging task. Various computational methods and tools are used for detection of precursors by extracting general information from noisy data. Finally, they are at different stages of growth in the life cycle of technology innovation. By using a common framework of clustering, we will be able to perform multi-resolution analysis of seismic data starting from the raw data events described by their magnitude spatio-temporal data space. This new methodology can be also used for the analysis of data from the geological phenomena e.g. clustering method can be applied to volcanic eruptions.[14]

## An Implementation Methodology

Other than M/s Terra Seismic, many other organisations are also collecting pertinent data to predict earthquakes. If all of this data is run through big data analytics, there is a greater probability of more accurate prediction. It is expected that in the future, more companies and governments will embrace big data analytics as a part of their earthquake prediction and recovery programmes. Real-time detection and classification of signals or events present in time series data is a fairly common need. Stereotypical examples include identifying high-risk conditions in ICU data streams or classifying signals present in acoustic data from diagnostic or monitoring sensors. Using a combination of stream processing and machine learning is an agile and highly capable approach. It can effectively scale to large, fast data streams, and adapt to evolving problem spaces. The United Nations and its organisations like the Office for the Coordination of Humanitarian Affairs (OCHA) in collaboration with the Inter-Agency Standing Committee (IASC) which is the arm of the United Nations responsible for bringing together national and international humanitarian providers to ensure a coherent response to emergencies, should come out with a framework within which each holder or data generator of earthquake-based data should contribute to the overall R&D effort for finding out a scientific solution to predict earthquakes more accurately and much ahead in terms of time so that all types of loss can be minimised.

## Conclusion

Earthquakes occur unexpectedly and usually last only seconds. Because there is not much time to react, it will be prudent to prepare a disaster plan, and practise it ahead of time. Predictive analytics and big data really give us hope for tomorrow. Being able to predict, and prepare for, an earthquake or any disaster, for that matter, would make a huge impact on the world. Though we may not be able to prevent natural disasters, utilising technology and advanced analytics to reduce the response time would give the necessary advantage to reduce collateral damage. Clearly, cities, agencies, governments and nations need to collectively gather and aggregate existing and future disaster data—from historical event information to sensor data, system status and video. And they need to not just collect the data, but also to apply sophisticated analytics to automatically analyse the data and provide intelligent insights into key performance indicators and trends. Though we may not be able to prevent these natural disasters, innovative emergency management solutions employing advanced analytics capabilities can help reduce the response time. By using existing resources productively, real-time access to quality data can reduce direct costs.[15] In man's quest to achieve the unimaginable, he has created some of the greatest wonders of the world in terms of tall skyscrapers, deep tunnels and large buildings. Education of field engineers, and undergraduate and graduate engineers in earthquake resistant design and construction practices which was imparted on a continued basis has made it possible to construct earthquake resistant buildings. Similarly, with funding from governments and agencies like the United Nations Organisation (UNO) the day is not far when a prediction mechanism for earthquakes will be in place. An active interaction among scientist, R&D institutes working in the field of earthquakes, earthquake engineers, structural engineers, seismologist, architects, government authorities, data holders, etc should take place which will greatly help in coming out with a scientific mechanism to forecast earthquakes.

## Notes


1. Lavanya Rathnam, "Using Big Data to Predict the Next Big Earthquake," *icrunchdata News,* May 14, 2015.
2. Bernard Marr, "Big Data: Saving 13,000 Lives A Year By Predicting Earthquakes? *FORBES*, April 21, 2015.
3. T. Rikitake, "Earthquake Prediction: An Empirical Approach," *Tectonophysics*, Vol. 148, Nos. 3–4, May 1988.
4. As cited at http://www.disaster-report.com/2015/05/nepal-earthquake-prediction-usgs.html
5. Jayalakshmi K, "Predicting Earthquakes: A Mixed Outcome, But Improving With Data Crunching," *International Business Times*, April 28, 2015.
6. According to a Research Paper by the Department Of Information Technology, Narasaraopeta Engineering College, Andhra Pradesh
7. As cited at https://en.wikipedia.org/wiki/Earthquake_prediction
8. Ibid.
9. P. Varotsos and M. Lazaridou, "Latest Aspects of Earthquake Prediction in Greece Based on Seismic Electric Signals," *Tectonophysics*, Vol. 188, Nos. 3–4, March 20, 1991.
10. Rikitake, n. 3.
11. Report by U Tejonmayam, cited in, *The Times of India*, October 07, 2015.
12. As cited from M/s Terra Seismic
13. Rathnam, n. 1.
14. n. 6.
15. As cited at http://www.ibmbigdatahub.com/blog/jen-q-public-applying-analytics-emergency-preparedness