IJITR
Y.V.Srinivasa Murthy* et al.
(IJITR) INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH
Volume No.2, Issue No.5, August – September 2014, 1214 – 1218.
Research Article

Earth Quake Prognostication Using Data Mining and Curve Fitting Techniques

Y.V. SRINIVASAMURTHY
National Institute of Technology
Karnataka, Surathkal

MANDA V ARUN KUMAR
GITAM University
Visakhapatnam

KATURI V S SANDEEP
GITAM University
Visakhapatnam

SANTOSHA MANIKYA V J
GITAM University
Visakhapatnam

EEMANI VENUGOPAL REDDY
GITAM University
Visakhapatnam

Abstract—**The title "EARTHQUAKE PROGNOSTICATION" is a Global Earthquake prediction, that is used to predict that an earthquake of a specific magnitude will occur in a particular place at a particular time, we however cannot tell the exact time and date the earthquake is going to occur but we can well predict that an earthquake will affect a given location over a certain number of years. The "Gutenberg Richter power-law distribution of earthquake sizes" implies that the largest events are surrounded by a large number of small events, with this statement we collected the data sets of all the EARTHQUAKES of magnitude ranging from small to big since 1900 to 2010 all over the world. After collecting this data we performed clustering techniques to the datasets available with latitude, longitude and time as parameters, which helped to find similarities between them and discovered patterns using non-linear regression functions that helped to forecast earthquakes. This prediction is based on both the historical seismic catalogue and the structural zoning.**

Keywords- **Clustering, Earthquake prediction, Prognostication, non linear regression, Curve fitting.**

I. INTRODUCTION

When the earth shakes due to the movement of plates below the earth's crust it is known as Earthquake. Perhaps the Earthquake is the most fearful natural phenomenon in the human life. It is more so, because it is unpredictable and arrives without notice or without announcing its vigor and strength. Earthquakes are natural disasters, which kill thousands of people in an instant and can destroy cities and countries- human habitation across miles. The vibration during an earthquake has the potential to wreak havoc and destruction, which is beyond imagination. The destruction is maximum near the epicenter, the place from where the vibrations arise and spread. Of late there have been many such natural disasters, which can be associated with earthquakes like the Tsunami and razed down buildings and annihilated many lives. It seemed like Mother Nature was avenging herself on us who have used all her endowments to the fullest extent without caring to rejuvenate them.

A. Effects of Earthquakes:

Earthquakes produce various damaging effects to the areas they act upon. This includes damage to buildings and in worst cases the loss of human life. The effect of the rumbling produced by earthquakes usually leads to the destruction of structures such as buildings, bridges, and dams. They can also trigger landslides.

B. Geologic Effects on Shaking:

When we discussed earthquake intensity we discussed some of the basic factors that affect the amplitude and duration of shaking produced by an earthquake (earthquake size, distance from fault, site and regional geology, etc.) and as you are aware, the shaking caused by seismic waves can cause damage buildings or cause buildings to collapse. The level of damage done to a structure depends on the amplitude and the duration of shaking. The amplitudes are largest close to large earthquakes and the duration generally increases with the size of the earthquake (larger quakes shake longer because they rupture larger areas). Regional geology can affect the level and duration of shaking but more important are local site conditions. Although the process can be complicated for strong shaking, generally shaking in soft sediments is larger and longer than when compared with the shaking experienced at a "hard rock" site.

C. Primary and secondary effects:

Earthquakes have primary violent impacts followed by long-lasting secondary impacts. In a sparsely populated area, a large earthquake might not cause many casualties. In a big city, even a relatively small earthquake could kill thousands. While the primary effects are dangerous and deadly, the secondary effects, especially if help is not available or adequate can be much worse. Though thousands died in the earthquake in Haiti in early 2010, the longer lasting effects are proving just as deadly, according to Common Dreams.

D. Destruction of nature and proeperty:

Due to the shaking and rupturing of the ground from an earthquake, primary effects include buildings and surrounding nature being destroyed. Depending on the development of the affected area, the damage can be catastrophic. Buildings can fall, bridges can collapse, fires can start along with avalanches and mudslides. Earthquake engineering is important in cities that have a high risk of earthquakes. Buildings can be retrofitted and developed to have a better chance of withstanding an earthquake. Engineers consider the potential seismic performance of a building. Its structure should be able to withstand the violent shaking nature of an earthquake. Preparations in countries like China aim to reduce the casualty count if and when the next major earthquake arrives, according to NPR.

II. BACKGROUND WORK

Data mining is the extraction of hidden predictive information from large databases. It is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems.

E. Clustering

Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many areas and fields, including machine learning, data mining, pattern recognition, image analysis, information retrieval, bioinformatics and several other. Maintaining the Integrity of the Specifications

F. K-Means

K-means clustering is a method of cluster analysis which aims to partition *n* objects into *k* clusters so that the resulting intra-cluster similarity is high but the inter cluster similarity is low. Cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's center of gravity.

First it randomly selects k of the objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is most similar, based on the distance between the object and the cluster mean. It then computes the new mean for each cluster. This process iterates until the criterion function converges. It works well when the clusters are compact clouds that are rather well separated from one another. The method is relatively scalable and efficient in processing large datasets because the computational complexity of the algorithm is O(nkt), where n is the total number of objects, k is number of clusters and t is number of iterations

G. Multiple Regression

The multiple regressions can be used to model the relationship between several independent or predictor variables and a dependent or response variable. In the context of data mining, the predictor variables are the attributes of interest describing the tuple (i.e., making up the attribute vector). In general, the value of the predictor variables are known, the response variable is what we want to predict.

H. Non-linear regression:

This is to model data that does not show a linear dependence i.e., If a given response variable and predictor variable have a relationship that maybe modeled by a polynomial function.

Transformation of a polynomial regression model to a linear regression model:-

Consider a cubic polynomial relationship given by y = w0+w1x+w2x2+w3x3

To convert this equation to linear form, we define new variables: x1 = x x2 = x2 x3 = x3 then the above equation can be converted to linear form by applying the above assignments, resulting in the equation y = w0 +w1x1 +w2x2 +w3x3, which is easily solved by the method of least squares using software for regression analysis. Note that polynomial regression is a special case of multiple regressions. That is, the addition of high-order terms like *x*2, *x*3, and so on, which are simple functions of the single variable, *x*, can be considered equivalent to adding new independent variables.

I. Curve fitting:

Capturing the trend in the data by assigning a single function across the entire range is known as curve fitting. Curve fitting can involve either interpolation, where an exact fit to the data is required, or smoothing, in which a "smooth" function is constructed that approximately fits the data. Fitted curves can be used as an aid for data visualization, to infer values of a function where no data are available, and to summarize the relationships among two or more variables.

A first degree polynomial equation is given by:

This is a line with slope a. We know that a line will connect any two points. So, a first degree polynomial equation is an exact fit through any two points with distinct x coordinates.

If we increase the order of the equation to a second degree polynomial, we get:

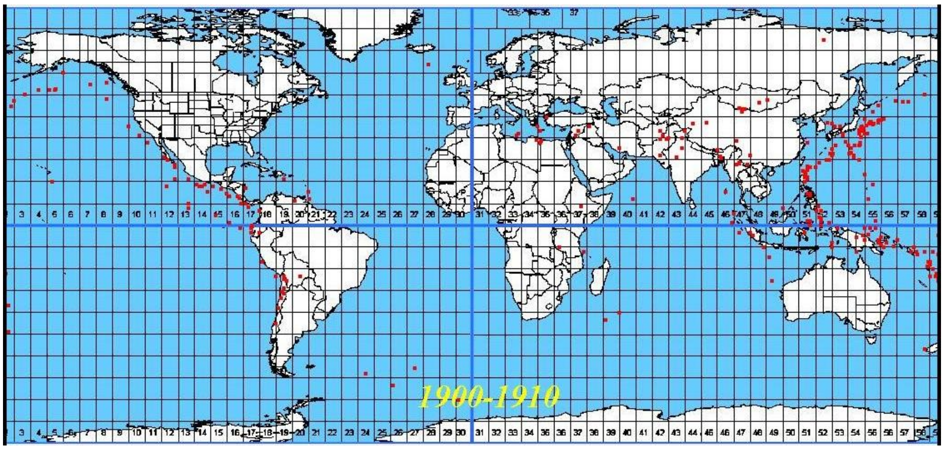
This will exactly fit a simple curve to three points.

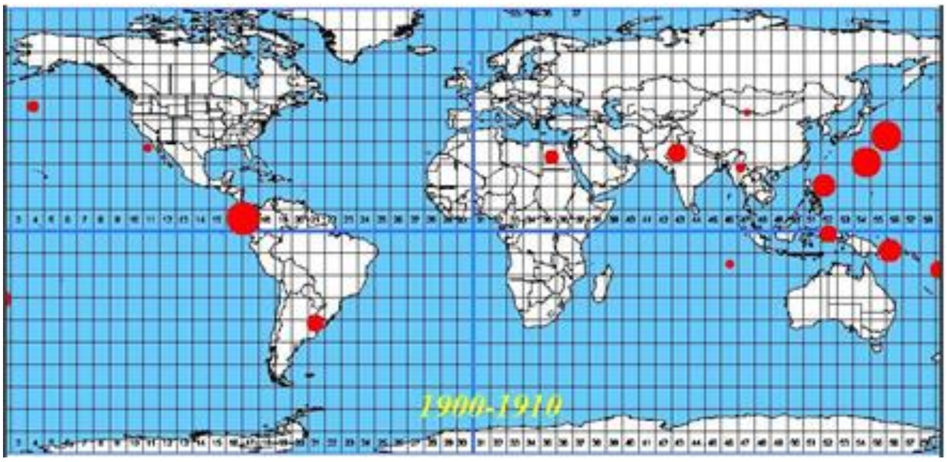
If we increase the order of the equation to a third degree polynomial, we get:

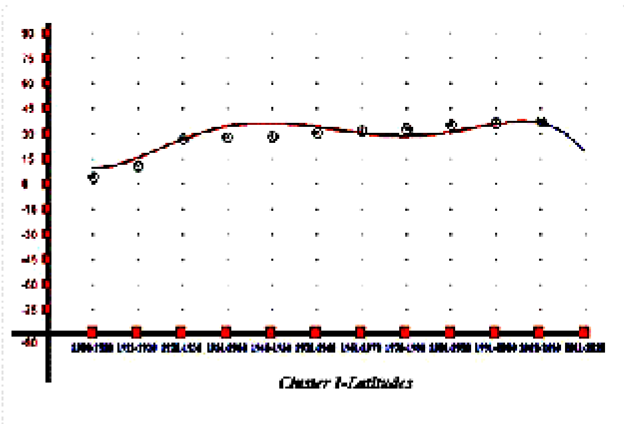
This will exactly fit four points. The first degree polynomial equation could also be an exact fit for a single point and an angle while the third degree polynomial equation could also be an exact fit for two points, an angle constraint, and a curvature constraint. Many other combinations of constraints are possible for these and for higher order polynomial equations.

If we have more than n + 1 constraints (n being the degree of the polynomial), we can still run the polynomial curve through those constraints. An exact fit to all the constraints is not certain. In general, however, some method is then needed to evaluate each approximation. The least squares method is one way to compare the deviations.

III. IMAGES







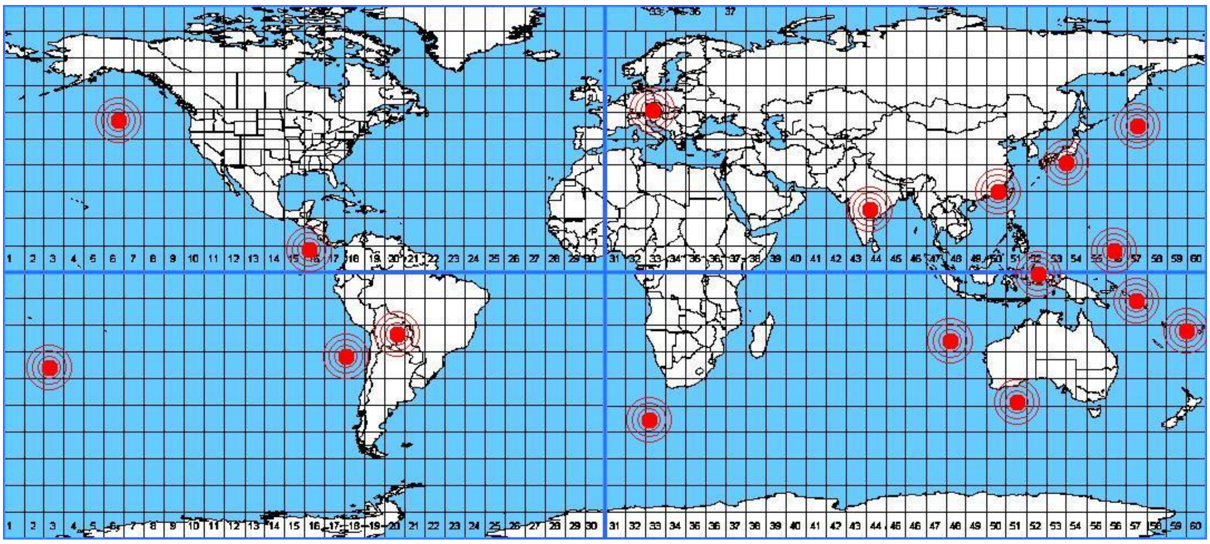
Y.V.Srinivasa Murthy* et al.
(IJITR) INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH
Volume No.2, Issue No.5, August – September 2014, 1214 – 1218.

IV. PROCESS

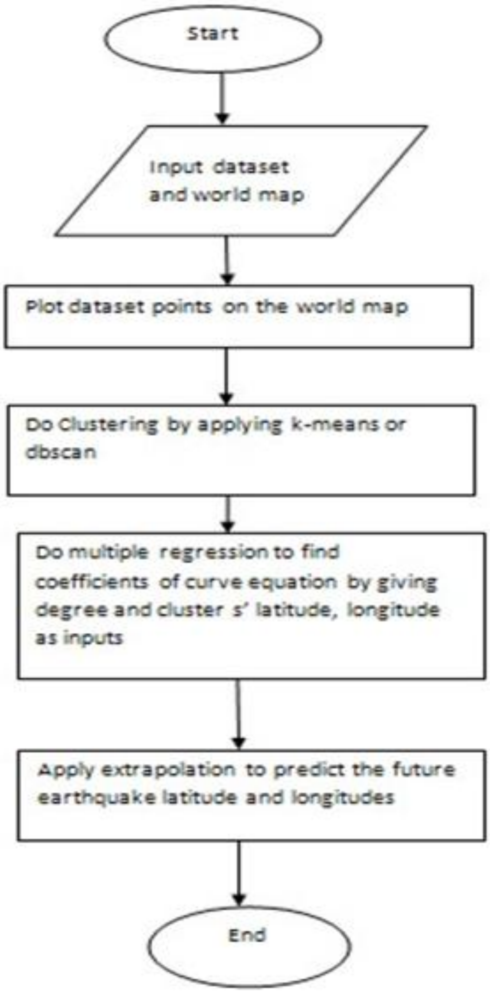
1. Initially from data sets points are plotted on world map by taking longitude and latitudes.
2. Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. Here clustering is done by using two techniques “k-means” and “db-scan”. These two clustering techniques are used to know the number of clusters.
3. Considering previous years clustering points, a graph is plotted using curve fitting.
4. Extrapolation is the process of constructing new data points outside a discrete set of known data points. A polynomial curve can be created through the entire known data.

V. RESULTS

ID	Latitude	Longitude
1	162	978
2	57	985
3	116	256
4	210	917
5	20	271
6	61	1011
7	24	835
8	49	910
9	174	1082
10	38	34
11	123	3
12	152	324
13	122	752
14	35	975
15	220	842
16	57	1031
17	75	934
18	48	756



VI. FLOW CHART FOR ENTIRE PROCESS



REFERENCES

[1] http://earthquake.usgs.gov

[2] http://en.wikipedia.org

[3] Baker, R.S.J.d. (in press) Data Mining for Education. To appear in McGaw, B., Peterson, P.,

[4] Baker, E. (Eds.) International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier

[5] K. Alsabti, S. Ranka, and V. Singh, ªAn Efficient k-means Clustering Algorithm,º Proc. First Workshop High Performance Data

[6] http://ifsc.ualr.edu/xwxu/publications/kdd-96.pdf

[7] http://www.ykhoanet.com/baigiang/Topic09.%20Multiple%20regression.pdf

[8] http://www.umich.edu/~gs265/society/earthquakes.htm

[9] http://eqseis.geosc.psu.edu/~cammon/HTML/Classes/IntroQuakes/Notes/earthquake_effects.html

[10] Joseph L. Kirschvink, April 2000, "Earthquake prediction by Animals: Evolution and sensory Perception" Bulletin of Seismological Society of America

[11] Sh o z o mAtSu m u rA, April 2009, "Trends and problems in Earth Quake Prediction Research", Science & Technology Trends.

[12] Aydin, M. Karakose, and E. Akin, 2009, "The Prediction algorithm based on Fuzzy Logic Using Time Series Data mining method., World Academy Science, Engineering & Technology.

[13] http://www.ce.ufl.edu/~kgurl/Classes/Lect3421/NM5_curve_s02.pdf

http://ieeexplore.iee.org

