Indian Journal of Science and Technology, Vol 8(16), DOI: 10.17485/ijst/2015/v8i16/62016, July 2015
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Paleo Earthquake Analysis for Sumatra Region in Order to Forewarn the Possible Occurrence of Future Earthquake

V. Britto Prabhu[1*], N. Venkatanathan[2] and S. Krishna Anand[1]

[1]School of Computing, SASTRA University, Thanjavur - 613401, Tamil Nadu, India; bpbrittovictor6@gmail.com, skanand@ cse.sastra.edu

[2]School of Electrical and Electronics Engineering, SASTRA University, Thanjavur - 613401, Tamil Nadu, India; venkatanathan@eee.sastra.edu

Abstract

By analysing the previous earthquakes in Sumatra region the long term earthquake prediction may help the Government and NGOs for necessary emergency preparedness. In order to predict chaotic behaviour of nature K-Means clustering algorithm is employed over the previous earthquake data for clustering the regions based on the coordinates of the earthquakes occurred places and Laws of Haversine is used further to compute the distance between the main earthquakes and secondary shocks. The 14 years of earthquake data of 8400 samples were collected from various sources. Each cluster results in different radius of the cluster. Based on the laws of Haversine results the data were loaded and analysed. The analysis results show the pattern of earthquakes and its changes. It gives the information about next larger earthquake pattern and place of occurrence.

Keywords: Aftershock, Cluster, Earthquake, Foreshock, Laws of Haversine, Seismic Event

1. Introduction

Every year more than 150,000 quakes are happening in our earth. Above 6.5 magnitude earthquakes are very dangerous. Only few quakes are crossing more than 7 magnitudes. It is happening less than trice in a decade. Predicting the smaller earthquakes are not useful, it will happen uniformly. Expert system, sensor networks are common methods for predicting earthquake. This paper proposes the method for predicting and forewarning the earthquakes in Sumatra region. By analysing the previous recorded earth quake data, this work attempts to predict the larger earthquake.

For an individual earthquake quantifying the activity of foreshock is difficult and it will increase but which gives clues when combining many the different events results[1]. From these combined observations, the main shock is observed which is linked to stress increase in the region[2]. A smaller earthquake occurs after large earthquake, within the same location. When magnitude of aftershock is higher than the seismic event, that event is considered as the main seismic event and the original main seismic event is considered to be a foreshock. Aftershocks are adjusts to main shock effects and which are made as the crust around the plate area which are displaced.

Artificial intelligence purpose is to enlarge the computers application. The expert systems development is ever centred on the knowledge base system. Figure 1 denotes the expert systems components and their integrations[3].

Aqdas Ikram and Usman Qamar presented a method for predicting the earthquakes worldwide. This method

*Author for correspondence

includes frequent pattern mining and predicate logic. Frequent pattern mining was used for gathering the related and frequent data from the database. Further, predicate logic used for enhancing the rules. It takes more time to predict the quakes and predict only smaller earthquakes[4]. Artificial neural network method for earthquake prediction gives 80.55% accuracy for M>4.0 and gives 58.02% for M>5.2. This method uses feed forward multilayer perceptron type neural network using back propagation. This method used for predicting the earthquakes only in Greece. Therefore, less number of data used for train the neural networks that are the main disadvantage of this system[5]. Probabilistic neural network methods predict the quakes within the time. This model tested in southern California region. Bayern classifier and parzen windows classifier used for creating statistics and classifications. It gives greater prediction accuracy from magnitude 4.5 to 6.0. The main drawback of this method was predict only magnitude, it was not predict the time and location of the earth where the earthquake going to occur[6]. Intelligent prediction model was used for predicting the debris flow. This method used multivariate analysis, multiple linear regression back propagation neural network and multiple linear regression. It gives 70% accuracy for predict the flow. This method used few data for train the neural network[7]. P. K. Dutta, O. P. Mishra and M. K. Naskar presented analysis for earthquake prediction. This method used fuzzy inference algorithm for analysing the earthquake data. This method proposed four step data retrieval process for any earthquake prediction algorithm[8]. Precursory patterns finding method used for predicting the earthquake in Chile. K-means classification clustering data method used for cluster the data. Silhouette index applied to clustered data. This method gives 70% accuracy for magnitude greater than 4.4. This method was only predicted in Chile region and giving low accuracy for some region[9]. Artificial neural network method was presented for predicting the earthquake in northern red sea region. This method used feed forward network with multi hidden layer. This gives 32% increased accuracy than other methods10. Artificial neural network method was applied to 200 records of ground motion parameters. It also determined the ground motion efficient parameters were sufficient for the analysis[17]. Three Dimensional FEA software was used for assessment of seismic safety but strengthening the historical masonry building is very difficult. Simplified kinematic limit analysis was utilized for finding building safety when the earthquake occurrence and anlysed the building safety in Italy region[18]. Conditional probability used for estimate the time intervals of Indian region data. The Weibul distribution gave more accurate results with conditional probability compared to poisonian probability[19].

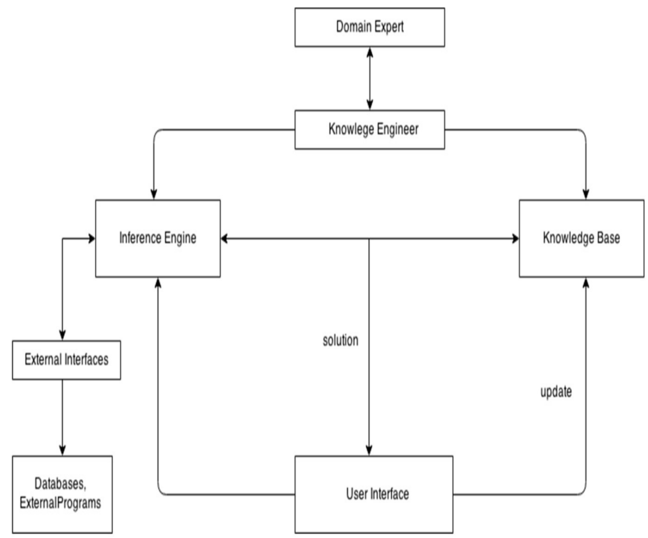

Figure 1. Expert System Components.

2. Data Acquisition and Method

For data mining problems, getting all the data is very important task. Various data sources were measured for earthquake investigation. Three data sources United States Geological Survey (USGS)[12], Advanced National Seismic System (ANSS)[13] and National Geophysical Data Centre (NGDC)[14] were chosen for data extraction. National Earthquake Hazards Reduction Program (NEHRP) contains the USGS Earthquake Hazards Program as a part. They are monitoring and reporting about earthquakes, impacts of earthquakes, hazards, find the causes and effects of earthquake. NGDC is a subsidiary of the National Oceanic and Atmospheric Administration (NOAA), National Environmental Satellite, US Department of Commerce (USDOC), Data and Information Service (NESDIS). It provides long-term scientific geophysical data.

2.1 USGS Database

USGS website provides earthquakes data publically and can be searched from URL http://earthquake.usgs.gov/earthquakes/ as shown in Figure 2. USGS provides entire world seismic events data. Table 1 represents number of earthquakes occurred in Sumatra region from 2000 to 2014. Based on this data, the larger magnitude earthquake will happen once in a decade and it will create tremendous devastation on the earth surface. Table 2, represents the number of less magnitude earthquakes in the Sumatra region from 2001 to 2014. It denotes the aftershocks and post shocks for the largest earthquakes. Every time the pre and post shocks will change. After shocks and fore shocks is depends upon the magnitude level of the big earthquake in the region. Sometimes more than magnitude 6.0 earthquakes occurred as a pre shock and post shock for the larger earthquakes. Data provided by USGS in many formats such as KML, Map and List, CSV, QuickML as shown in Figure 2. For this research data is retrieved in CSV file format. After database retrieval the entire database are transported into MySQL database. CSV file contains more than 10 different column values such as Date, Time, Latitude, Longitude, Magnitude, Magnitude Type, Depth, Region Name…etc. For this research we need important columns such as Date and Time, Latitude, Longitude, Magnitude. By removing irrelevant data, CSV files are normalised. Only relevant data are fed into the MySQL database. MySQL database creates a single database with large number of related data files.

2.2 K-Means Clustering

In this research, for clustering the regions K-Means clustering is used. K-Means clustering algorithm is very easy, numerical, iterative, non deterministic and simplest for clustering. This method is very suitable for clustering the regions. It's very useful when the large number of variables used in the clustering process. This method is faster than hierarchical clustering[15]. The K-means algorithm gives a partitioned data clusters, with no particular structure within them. Initial number of clusters need to be specified. The Centroid (C) values may be fixed or random values. K-means algorithm changes the Centroid (C) values in iteration. The sums of squares of distances within the clusters are used to measure different clustering. The smaller sum solution within cluster distance is an optimal solution. If optimal solution is found more than one time, then the algorithm has found overall optimal solution[11]. Figure 3, represents the operation of K-Means clustering algorithm. The objective function by K-Means is,

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

Where, $\left\| x_i^{(j)} - c_j \right\|^2$ is a distance between a $x_i^{(j)}$ data point and c_j centroid.

Table 1. Number of earthquakes in Sumatra region from 2000 to 2014

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
>9	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
7	1	-	1	-	-	-	-	1	1	1	3	-	-	-	-
6	1	-	-	-	1	2	-	1	2	2	-	1	1	2	-

Table 2. Number of earth quakes more than Magnitude 4.0 from 2000 to 2014 in Sumatra region

Magnitude	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
>4.0	471	227	190	168	615	2702	713	919	629	389	521	338	308	307	345

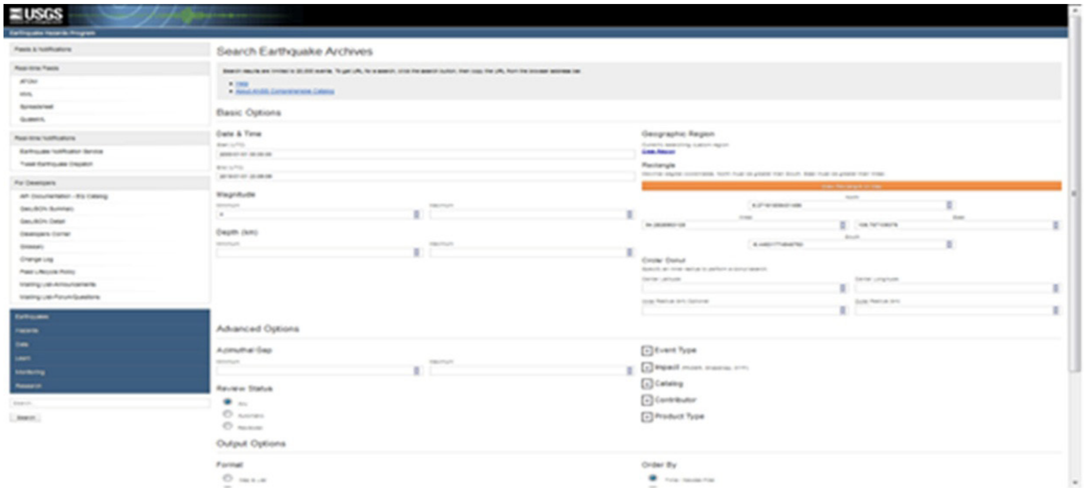

Figure 2. USGS website for downloading earthquake data.

3. Proposed Work and Results

The proposed research work consists of two phases. In phase 1, the normalized data given as input for the k-means clustering. The random values are taken as centroid from the database. The distance between the cluster modes are calculated and clusters are evaluated. Further the clusters are randomly chosen. From each clusters nodes are picked. Further, two values has been taken from each cluster. Based on laws of haversines, distance between two points has been calculated as shown in Figure 5. The haversin formula is,

$$haversin\left(\frac{d}{r}\right) = haversin(\emptyset 2 - \emptyset 1) + \cos(\emptyset 2)\, haversin(\lambda 2 - \lambda 1)$$

Where, r is sphere radius, d is distance between two points in the sphere, ϕ1 and ϕ2 are point 1 latitude and point 2 latitude, λ1 and λ2 are point 1 longitude and point 2 longitude16 From the radius and latitude, longitude get the database of a particular region data.

In phase 2, new database has been given as an input to the shock calculation phase. In shock calculation phase, every node in a larger earthquake clusters compared

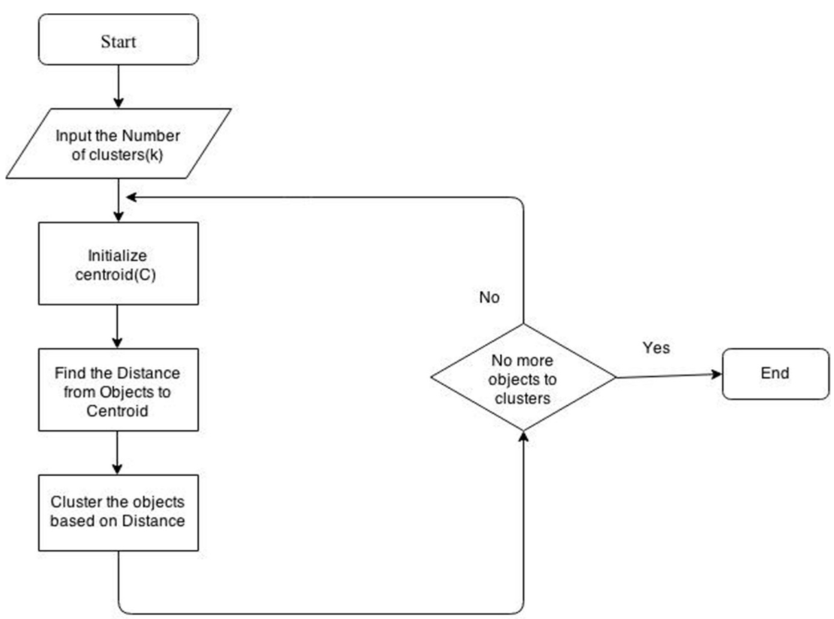

Figure 3. Flowchart architecture for K-Means clustering.

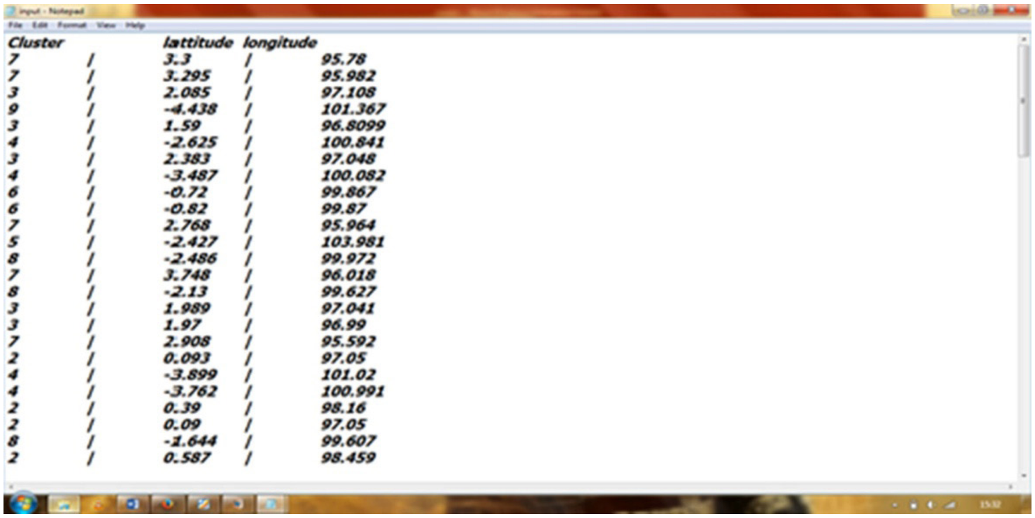

Figure 4. Clusters along with cluster numbers and nodes.

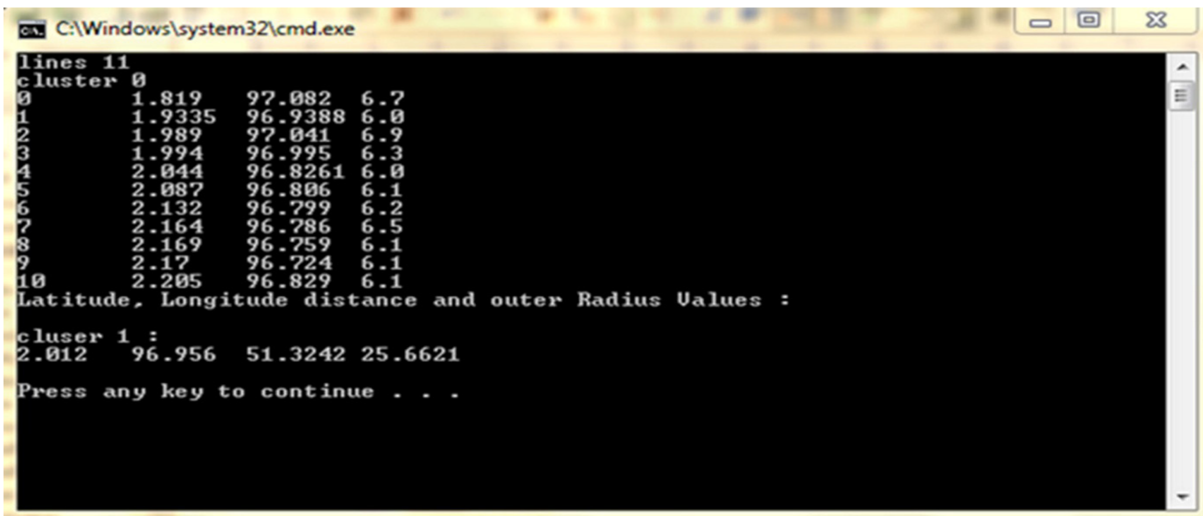

Figure 5. Distance between cluster values.

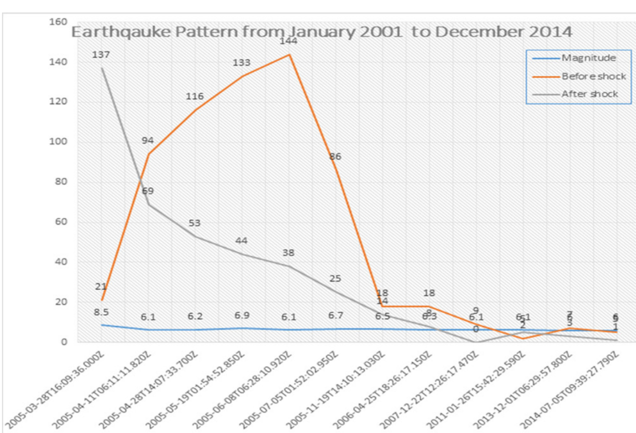

Figure 6. Earthquake pattern change.

with the new database data. Further, aftershocks and fore shocks are calculated for every cluster points in the large earthquake clusters. In this research work, the result of the earthquake pattern as shown in Figure 6. This pattern shows the seismic event occurrence changes in the regions after the large magnitude earthquake. This pattern is very important for scientists who are involved in earthquake prediction research.

4. Conclusion

Predicting and forewarning the earthquakes are still challenge for scientists. Whenever the large earthquakes occurred, they not only leave the trail of destruction but also vital clue on seismicity of the region. Understanding earthquake preparation process will help us to go for long term prediction of earthquake. The aim of this research we propose and implement an expert system to fore-warn the earthquake by analysing the pattern. This was achieved by applying K-Means clustering and Laws of Haversine. Results showed that the proposed system found the pattern of the earthquakes in Sumatra region from January 2001 to December 2014. Of course there is no direct help to the public from the long term prediction of earthquakes; it gives vital support to government and scientist involving in earthquake prediction research.

5. References

1. Kayal J. Microearthquake seismology and seismotectonics of South Asia. Springer Science and Business Media; 2008.
2. Maeda K. Time distribution of immediate foreshocks obtained by a stacking method. Pure and Applied Geophysics. 1999; 155(2–4):381–94.
3. Chen Y, Hsu CY, Liu L, Yang S. Constructing a nutrition diagnosis expert system. Expert Systems with Applications. 2012; 39(2):2132–56.
4. Ikram A, Qamar U. Developing an expert system based on association rules and predicate logic for earthquake prediction. Knowledge-Based Systems. 2015; 75:87–103.
5. Moustra M, Avraamides M, Christodoulou C. Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals. Expert Systems with Applications. 2011; 38(12):15032–9.
6. Adeli H, Panakkat A. A probabilistic neural network for earthquake magnitude prediction. Neural Networks. 2009; 22(7):1018–24.
7. Kung HY, Chen CH, Ku HH. Designing intelligent disaster prediction models and systems for debris-flow disasters in Taiwan. Expert Systems with Applications. 2012; 39(5):5838–56.
8. Dutta P, Mishra O, Naskar M. Decision analysis for earthquake prediction methodologies: Fuzzy inference algorithm for trust validation. International Journal of Computer Applications. 2012; 45(4):13–20.
9. Florido E, Martínez-Álvarez F, Morales-Esteban A, Reyes J, Aznarte-Mellado J. Detecting precursory patterns to enhance earthquake prediction in Chile. Computers and Geosciences. 2015; 76:112–20.
10. Alarifi AS, Alarifi NS, Al-Humidan S. Earthquakes magnitude predication using artificial neural network in northern Red Sea area. Journal of King Saud University-Science. 2012; 24(4):301–13.
11. Devi DMR, Thambidurai P. Similarity measurement in recent biased time series databases using different clustering methods. Indian Journal of Science and Technology. 2014; 7(2):189–98.
12. USGS Catalogue Search. Available from: http://earthquake.usgs.gov/earthquakes/
13. ANSS Catalogue Search. Available from: http://www.ncedc.org/anss/catalog-search.html
14. NOAA Catalogue Search. Available from: http://www.ngdc.noaa.gov/nndc/
15. K-Means clustering. Available from: http://www.improvedoutcomes.com/docs/WebSiteDocs/Clustering/K-Means_Clustering_Overview.htm
16. Haversine Laws. Available from: http://en.wikipedia.org/wiki/Haversine_formula
17. Kia A, Sensoy S. Assessment the effective ground motion parameters on seismic performance of R/C buildings using Artificial Neural Network. Indian Journal of Science and Technology. 2014; 7(12):2076–82.
18. Pouraminian M, Hosseini M. Seismic safety evaluation of Tabriz Historical Citadel using finite element and simplified kinematic limit analyses. Indian Journal of Science and Technology. 2014; 7(4):409–17.
19. Sharma M, Kumar R. Estimation and implementations of conditional probabilities of occurrence of moderate earthquakes in India. Indian Journal of Science and Technology. 2010; 3(7):808–17.