Improving EarthquakeMagnitude Prediction Using DataMining with Artificial Intelligence Techniques

Mrs.R.Kavitha, Department ComputerScience Engineering and Information Technology,University College Of Engineering(BIT Campus) Anna University ,Tiruchirappalli.India. rkavitha.autt@gmail.com

Dr.AValarmathi,Department of Master of ComputerApplications, University College Of Engineering(BIT Campus)Anna University, Tiruchirappalli. India. valar1030@yahoo.com

Abstract— The Objective of this study is to evaluate the performance of Naïve Bayes and Back Propagation neural network techniques in predicting earthquakes occurring in the region . The data set have the locations of 1000 seismic events of Magnitude > 4.0. The parameters of events occurred in a region are extracted from USGS catalogue comprised of all minor, major events and their aftershock sequences . This data warehouse contains event data, event time with seconds, latitude, longitude, depth and magnitude. These field data are converted into eight mathematically computed parameters known as seismicity indicators. These seismicity indicators have been used to train the Naive Bayes and BP Neural Network for better decision making and predicting the magnitude of the pre-defined future time period. We propose a three-layer feed forward BP neural network model to identify factors, with the actual occurrence of the earthquake magnitude. From these algorithm implementations, it can be proved that machine learning techniques yields better accuracy over existing algorithms.

Keywords—Naïve Bayes, Backpropagation, Seismic Energy, Gutenberg-Richter.

Introduction

Earthquakes are natural phenomena that occur with the influence of a lot of parameters such as Seismic activity due to Concentration of Radon Gas in soil, changing ground water motion, Change in the ground water's temperature etc[1]. Also, Earthquakes are one of the most destructive costly natural hazards faced by the nation in which they occur without an explicit warning [2]. More than 50% of area in India is considered prone to damaging Earthquakes. Main cause of earthquakes is the movement of Indian Plate towards Eurasian plate at the rate of 50mm/year. It causes loss of lives and lot of damages to properties. Prediction of earth quakes remains most frustrating issues to scientists and researchers of Seismology to predict the occurrence of major seismic events.

There are numerous difficulties in earthquake prediction including highly non-linear behavior of seismic activity and non-availability of reliable seismic precursors. The use of seismic facts and data in combination with artificial intelligence has become increasingly popular in the field of seismic activity and earthquake prediction. The process of seismic activity prediction includes identifying suitable mathematically calculated parameters and co-relating them to the actual earthquake occurrences from the past by means of intelligent data mining techniques [3]. Such a methodology of earthquake prediction relies upon temporal sequence of past earthquakes, i.e. earthquake catalog. Researchers have used a variety of approaches and methods to improve the accuracy of Earthquake Magnitude prediction.

II. LITERATURE SURVEY

Many studies are available for Earthquake Magnitude prediction using Data Mining Techniques with Artificial Neural Networks. They show how to build a model that can predict the next seismic event occurrence in an area more accurately [2].

Conventional Earthquake magnitude prediction system employs seismic data, sensor data and satellite data which can only be understood by the expert thus the prediction can be carried out only by the seismologist. A system was developed for forecasting seismic events by recognizing Patterns [3] from seismic data.

According to the employed methodologies, existing works on earthquake prediction can be mainly classified into four categories, like mathematical analysis, precursor signal investigation, machine learning algorithms like decision trees and support vector machines (SVM) and deep learning. In this the first type of work tries to formulate the earthquake prediction problem by using different mathematical tools, like the FDL (Fibonacci, Dual and Lucas) method, kinds of probability distribution or other mathematics proving and spatial connection theory. In the second type of work, researchers study earthquake precursor signals to help with earthquake prediction. The third type of work mainly explores data mining and time series analysis methods, such as J48, adaboost, multi-objective info-fuzzy network (M-IFN), k-nearest neighbors (kNN), SVM, and artificial neural networks (ANNs) , to predict the magnitude of the largest earthquake in the next year based on the previously recorded seismic events in the same region. In the fourth type of work, deep learning algorithms are utilized to predict both the magnitude and the time of major seismic events. Various kinds of neural networks have been adopted, such as multi-layer perceptron (MLP), backward propagation (BP) neural network, feed forward neural network (FFNN). recurrent neural network (RNN), which can work under certain particular circumstances[18].

Earthquakes and Volcanos are very important and often fatal phenomena taking place around the globe on an unpredictable time scale which makes it more dangerous. Earthquakes occur due to collision of tectonic plates deep underground which send out vibrations of great magnitude through the layers of the earth. These vibrations are termed as seismic waves which are an important class of signals for analysis of their magnitude, frequency which also consequently help in prediction of occurrence and study of deep seismic activity. A novel SVM based kernel model has been proposed in this research paper for predicting seismic activity on a time scale. This method is computationally efficient and a frequency domain method has been used for analysis of the frequency content. The experimentations have been done on sample seismic activity observation inputs and results justify the efficiency of the proposed approach[19].

Data Mining

[12].previously unknown and potential useful information from the data and the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends. Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of this data mining process is to extract information from a data set and transform it into an understandable structure for further use.

PREDICTION METHODS

Prediction by Investigating, determining, and unearthing the relationship between precursors and the earthquake. It can be broadly classified into two categories. Traditional Methods and Knowledge based Methods. Traditional Methods Historical Data from External sensors. Knowledge based method employ features and available data from previous Earthquake at a time interval to predict the probability of a seismic occurrence. Recent Technologies collects data from Precursors and Remote sensing Devices. Any parameter which changes before the earthquake is called Precursor. Precursors can be the changes in the amount of radon gas in underground water, changes in temperature in groundwater, ground water change, foreshock before the main earthquake, the magnitude of foreshock. Here the magnitude is dependent variable, it depends on the number of foreshocks, the lack of an seismic occurrence in areas including foreshock due to the presence of fault, clouds due to the reaction of atmospheric gases with leaded ions released with radon gas changes, changes in air temperature and pressure, wind speed, relative humidity changes bird fluttering, and earthworm outflow and animal behaviors have been widely observed the date even several days before an occurrence an earthquake. Knowledge based Prediction Methods collects data form various Data Sources. The source of the underlying data for In this study, Implementation of Machine Learning Techniques Naïve Bayes Classifier and Back Propagation Compared[20].

Data Sources

The earthquake catalogue and database makes the first essential input for the description of seismic source zones and their characterization. Preparation of homogeneous catalogue for India is an important task. An uniform earthquake catalogue data have been compiled from historical to recent times. The earthquake catalogue has been prepared by extracting data from different sources, such as Guaribidanur Array (GBA), Indian Meteorological Department (IMD), National Geophysical Research Institute (NGRI) Hyderabad, Indira Gandhi Centre for Atomic Research (IGCAR) Kalpakkam, Incorporated Research Institutions for Seismology (IRIS), International Seismological Centre (ISC), United States Geological Survey (USGS) etc and also from various literatures [Oldham (1883); Kelkar (1968); Tandon and Srivastava (1974); Chandra (1977, 1978); Kaila and Sarkar (1978); Rao and Rao (1984); Srivastava and Ramachandran (1983); Bilham, R. (2004[12].

Seismicity of the Indian Region

Tectonic framework of Indian subcontinent covering an area of about 3.2 million sq. km is spatio temporally varied and complex. The rapid drifting of Indian plate towards Himalayas in the north eastern direction with a high velocity along with its low plate thickness (Kumar, 2007) might be the cause for an increase in the seismicity of the Indian region. Indian plate is moving northward at about 5 centimeters per year and in doing so, collides with the Eurasian Plate. Upon the Eurasian Plate lie the Tibet plateau & central Asia[13]. When continents converge, large amounts of shortening and thickening take place, like at the Himalayas and the Tibet. Due to this massive collision, the Himalayas are thrust higher and large numbers of earthquakes are generated due to this process. This is the major cause of earthquakes from the Himalayan regions to the Arakan Yoma. The similar process, involving the Indian Plate and the Burmese micro-plate, results in earthquakes in the Andaman & Nicobar Islands. Sometimes earthquakes of different magnitudes occur within the Indian Plate, in the peninsula and in adjoining parts of the Arabian Sea or the Bay of Bengal. Most earthquakes occur along narrow zones that follow the edges of tectonic plates[6].

These events are known as Inter-Plate or Plate Boundary earthquakes. These earthquakes are the direct result of the interaction between two or more tectonic plates. Sometimes earthquakes occur far away from plate boundaries. These arise due to localized systems of forces in the crust sometimes associated with ancient geological structures such as in the Rann of Kachchh (http://asc- india.org/menu/seismi.htm).

Seismic occurrence of this nature contributes 1% of the annual seismic energy release globally. All earthquakes in peninsular India fall within this category. Geographical statistics of India show that almost 54% of the land is vulnerable to earthquakes. The seismic risk in the country has been increasing rapidly in the recent years as there were large magnitude earthquakes in recent times even in the stable continental region[3].

The latest version of seismic zoning map of India given in the earthquake resistant design code of India [BIS 1893 (Part 1) 2002] assigns four levels of seismicity for the entire India in terms of different zone factors. In other words, the earthquake zoning map of India divides India into 4 seismic zones (Zone II to V) as shown in Figure1, unlike its previous version which consisted of five or six zones for the country. According to present zoning map, Zone - V expects highest level of seismicity whereas Zone - II is associated with lowest level of seismicity. The Global Seismic Hazard Assessment Programme (Bhatia et al., 1999) also classifies the region in to five zones. Probabilistic seismic hazard map of North West Himalaya (Mahajan et al., 2009) and that of South India (Vipin et al., 2009) are available in the literature. Spatial-temporal variability of seismic hazard in peninsular India (Jaiswal and Sinha, 2008) and seismicity analysis of North East India (Thingbaijam et al., 2008) are also found in the literature.

Earthquake Parameters

1. Location Rank

Represents the rank of the location based on the occurrence of earthquake from past to present years.

2. State Rank

Represents the rank of the state based on the number of earthquakes occurred in state from past 500 years.

3. Elevation

Represents the height above or below a referenced point in a geographical location based on earthquake places.

4. Population

Represents the population of a particular area where the earthquake occurred.

5. Year

Represents the year of the location where the earthquake occurred.

6. Month

Rrepresents the month of the location where the earthquake occurred

7. Day

Represents the Day of the location where the earthquake occurred

8. Hour

Represents the time of the location in hours where the earthquake occurred.

9. Minutes

Represents the time of the location in minutes where the earthquake occurred

10. Second

Represents the time of the location in seconds where the earthquake occurred

11. Latitude

The location of a point north or south of the equator. Latitude is shown on a map or globe as east-west lines parallel to the equator.

12. Longitude

The location of a point east or west of the prime meridian. Longitude is shown on a map or globe as north-south line left and right of the prime meridian, which passes through Greenwich, England.

13. Depth

The vertical distance below a surface; the degree to which something is deep

14. Magnitude

The Richter magnitude scale assigns a magnitude number to quantify the energy released by an earthquake. Richter scale is a base-10 logarithmic scale, which defines magnitude as the logarithm of the ratio of the amplitude of the seismic waves to arbitrary, minor amplitude. If magnitude is equal to 7 or more, large areas are damaged depending on their depth. If magnitude is equal to 3 or less, the probability of occurrence of an earthquake is weak[21].

Naive Bayes classifier

Naïve Bayes Classifier is a supervised classification technique with a focus on predicting the outcome. Naïve Bayes Classifier is based on Bayes Theorem and studies found that Naïve Bayes Classifier outperforms many classification techniques like Decision Trees and Neural Network Classifiers.

Bayesian Classifier is calculating the posterior from Likelihood and Prior Probabilities, as stated at (1) and (2)

Posterior=(Likelihood*Prior)/Evidence (1)

$P(C \mid X)=P(X \mid C)P(C)/P(X)$ (2)

where, $P(C \mid X)$ - posterior probability of the target class, is calculated using P(C) - Class Prior Probability , P(X) – Prior Probability predictor Class and $P(X \mid C)$ - Likelihood. In Naïve Bayes Classifier, the assumption that the probabilities of each attribute on a class are independent of all other attribute values is very strong, and it is to simplify the calculation of probabilities.

The Earthquake Dataset is divided in to 70:30 ratio of Training and Test Dataset.

Results and Discussions

Table1.Performance metrics

S.NO	Prediction Features	Prediction Measures
1	Accuracy	0.9527
2	95%CI	(0.93424,0.9672)
3	No Information Rate	0.8323
4	P-Value(Accuracy > NIR)	<0.000000000
5	Kappa	0.84583
6	Mcnemar's Test P-Value	0.00000109
7	Sensitivity	0.9829060
8	Specificity	0.9466437
9	Pos Pred Value	0.7876712
10	Neg Pred Val	0.9963768
11	Prevalence	0.1676218
12	Detection Rate	0.1647564
13	Detection Prevalence	0.2091691
14	Balanced Accuracy	0.9647749
15	'Positive' Class	high

The prediction accuracy of a classification model is given by the proportion of the total number of correct predictions. The accuracy for this model turns out to be 95%. We see that despite many simplifying assumptions, the Naive Bayes algorithm does reasonably well at predicting the correct prediction classes in R for this dataset.

Back Propagation Learning Algorithm

A Multilayer Perceptron is feed forward neural network (ANN), consisting of a number neurons linked together and attempts to create a desired relation in an input/output set of learning patterns. A Neural network consists of an input vector layers, more hidden layers and output vector layers. Each layer has its corresponding neurons weight connects. A single training pattern is equations I/O vectors of pairs of input and output values in the entire matrix of I/O training set. Here the Fiji Earthquake Dataset is divided into Test data and Training data in 80: 20 ratio.

Earthquake parameters, Latitude, Longitude, depth (km), magnitude(Richter Scale),Station Number(Number of stations reporting) are existing in the Fiji Earthquake dataset. The output obtained

Mean Squared Error = 0.001280

fig1 visualizes the computed neural network. Our model has 3 neurons in its hidden layer. The black lines show the connections with weights. The weights are calculated using the back propagation algorithm explained earlier. The blue line is the displays the bias term. Figure 2 shows the accuracy of predicted magnitude vs real prediction.

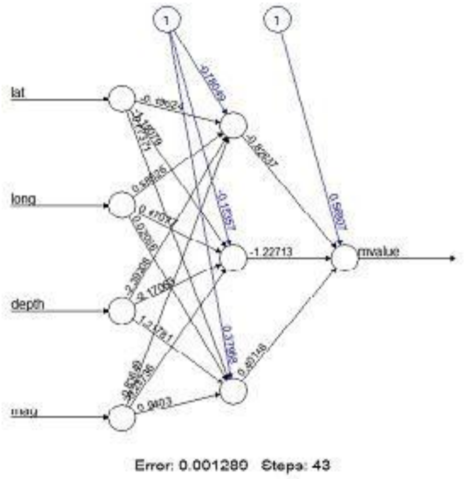

Fig. 1 3-Layered Back propagation Neural Network

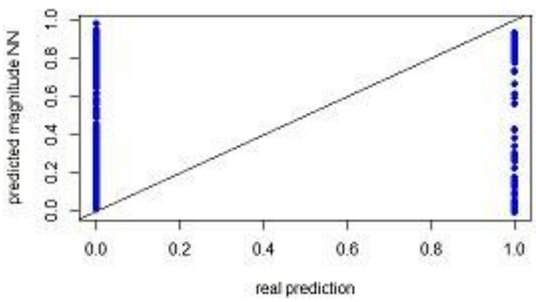

Fig.2 Accuracy of Predicted Magnitude over real Prediction

Conclusion and Future Work

In this work, the implementation of Naïve Bayes algorithm and Back propagation algorithm was done in R and the obtained results were analysed. The results showed these machine learning algorithms proved that improved accuracy than existing algorithms. In Future, We planned to implement Particle Swarm optimization algorithm in for this dataset to improve the accuracy. We also planned to develop a machine learning earthquake magnitude prediction model for Himalayan earthquake dataset.

REFERENCES

[1] A. Jiménez, A.M. Posadas, K.F. Tiampo, "Describing seismic pattern dynamics by means of using cellular automata", Lecture Notes in Earth Sciences 112 (2008) 273–290.*)*

[2] Ali Negarestani Saeed Setayash Bakashe "Layered neural networks based analysis of radon concentration and environmental parameters in earthquake prediction" published 2000 in Journal Environmaental radioactivity.

[3] G. Asencio-Cortés, F. Martínez-Álvarez, A. MoralesEsteban, and J. Reyes, "A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction," Knowledge-Based Systems, vol. 101, pp. 15-30, 2016.

[4] M.Last, N. Rabinowitz, and G. Leonard, "Predicting the Maximum Earthquake Magnitude from Seismic Data in Israel and Its Neighboring Countries," PloS one, vol. 11, p. e0146101, 2016

[5] H. Adeli and A. Panakkat, "A probabilistic neural network for earthquake magnitude prediction," Neural networks, vol. 22, pp. 1018-1024, 2009.

[6] A. Panakkat and H. Adeli, "Neural network models For earthquake magnitude prediction using multiple seismicity indicators," International Journal of Neural Systems, vol. 17, pp. 13-33, 2007.

[7] J.Reyes, A. Morales-Esteban, and F. MartínezÁlvarez, "Neural networks to predict earthquakes in Chile," Applied Soft Computing, vol. 13, pp. 1314- 1328, February 2013

[8] A.Morales-Esteban, F. Martínez-Álvarez, and J. Reyes, "Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence," Tectonophysics, vol. 593, pp. 121-134, 2013.

[9] B.Gutenberg and C. Richter, Seismicity of the earth and associated phenomena: Princeton University Press, 1954.

[10] A. S. N. Alarifi, N. S. N. Alarifi, and S. Al-Humidan, "Earthquakes magnitude predication using artificial neural network in northern Red Sea area," Journal of King Saud University - Science, vol. 24, pp. 301-313, 2012.

[11] T. Utsu and Y. Ogata, "The centenary of the Omori formula for a decay law of aftershock activity," Journal of Physics of the Earth, vol. 43, pp. 1-33, 1995.

[12] S. Wiemer and M. Wyss, "Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the western United States, and Japan," Bulletin of the Seismological Society of America, vol. 90, pp. 859-869, 2000.

[13] K. Asim, F. Martínez-Álvarez, A. Basit, and T. Iqbal, "Earthquake magnitude prediction in Hindukush region using machine learning techniques," Natural Hazards, pp. 1-16.

[14] A. Zamani, M. Sorbi, and A. Safavi, "Application of neural network and ANFIS model for earthquake occurrence in Iran," Earth Science Informatics, vol. 6, pp. 71-85, 2013/06/01 2013.

[15] C.-X. Zhang and J.-S. Zhang, "RotBoost: A technique for combining Rotation Forest and AdaBoost," Pattern Recognition Letters, vol. 29, pp. 1524-1536, 2008.

[16] Sreevalsa Kolathayar, K. S. Vipin and T. G. Sitharam " Recent Seismicity in India and Adjoining Regions",International Journal of Earth Sciences and Engineeering, February 2012, P.P. 52-60

Peninsula based on computational intelligence," Tectonophysics, vol. 593, pp. 121-134, 2013.

[17] K Netti* and Y Radhika CSIR-National Geophysical Research Institute, Uppal Road,Hyderabad-500007, India "A Model for Accurate Prediction in GeoRSS Data Using Naive Bayes Classifier" Journal of Scientific & Industrial Research Vol. 76, August 2017, pp. 473-476

[18] Qianlong Wang, Yifan Guo, Lixing Yu, and Pan Li "Earthquake Prediction based on Spatio-Temporal Data Mining: An LSTM Network Approach"

[19] Yongyang Zheng" A Novel Seismic Wave Analysis and Prediction Model using SVM Kernel and Time Series Prediction" Nanchang University, Nanchang, China

[20] Pouria Hajikhodaverdikhan, Mousa Nazari, Mehrdad MohsenizadeShahaboddin Shamshirband & Kwok-wing Chau "Earthquake prediction with meteorological data by particle filter-based support vector regression" Engineering Applications of Computational Fluid Mechanics 2018 Vol 12

[21] Dr.K Srujan Raju, Kandukuri Rajesh"Data Mining Techniques for Earthquake Frequency-Magnitude analysis and Seismic Zone Estimation" International journal of Computer Trends and Technology, July2015