An Innovative Data Mining Approach for Determine Earthquake Probability Based on Linear Regression Algorithm

Thaharim Khan
Department of Computer Science & Engineering
Daffodil International University
Dhaka , Bangladesh
taharimkhan54@gmail.com

Masud Rabbani
Department of Computer Science & Engineering
Daffodil International University
Dhaka , Bangladesh
masud.cse@diu.edu.bd

Shah Md Tanvir Siddiquee
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
tanvir.cse@diu.edu.bd

Anup Majumder
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
anup.cse@diu.edu.bd

***Abstract*—Earthquake (EQ) causes many damages to our environment, this paper proposes a new approach to find the Earthquake probability (EQP) using linear regression (LR-EQP) is to minimize this extensive problem with the help of geological data by predicting the intended probability of the level of EQ. This approach is to take the all the geological information which are responsible for EQ such as to create pressure on the land like density of population (DP), soil type (ST), combustible elements (CE), specify the place use latitude and longitude (Lat-Long), distance from nearest tectonic plate (TP). Since TP is the divided part of earth which move independently and responsible earth's seismic activity and LR-EQP deploy these geological data, so LR-EQP approach can determine the EQP easily and accurately. Earthquake prediction using data mining is a process, which uses only three factors: (a) ground water level, (b) chemical changes and (c) radon gas in ground water. But this is a slow approach to determine EQP because in general these factors cannot be determined so easily. LR-EQP can be used for any region by targeting the TP area. This data model can be used also for mobile application for easily detecting the probability of earthquake.**

Keywords—Earthquake (EQ), Earthquake probability (EQP), Earthquake probability using linear regression (LR-EQP), combustible elements (CE), Latitude and Longitude (Lat-Long), Tectonic plate (TP), Soil type (ST), Density of population (DP).

I. INTRODUCTION

Natural Disaster (such as EQ) is an appearance that is occurred when compatibility of nature is broken and that causes destruction of lives and properties [1]; in this regard EQP is a big deal for our life. We can save many of our resources and life if we can be informed about EQ before. In April 2015 there occurred an EQ in Nepal with a magnitude of 7.8 or 8.1, which killed about 9000 and 22000 almost injured during this hazardous event [2]. Another hazardous EQ occurred in Indonesia on September 2018 at central Sulawesi province [3]. The EQ that occurred in Indonesia in September having magnitude of 7.5. This EQ killed 2000 people and 2500 people are seriously injured according to Indonesia disaster management agency .A heavy loss of life and property occurred there in both Nepal & Indonesia .The after effect of EQ cannot be prevented but it can be reduced if it can be possessed before. Our main motive for this work is: (i) to manage data mining for geological issues, (ii) to create strong connection between geological data and data mining; and (iii) to reduce the disastrous effect by getting information about disaster before.

The result of earthquake prediction using data mining [4] is not probably accurate more than 50% but our target is to find better result than earthquake prediction using data mining. This [4] earthquake prediction using data mining use 3 factors these are: (a) ground water levels (b) chemical changes in ground water; and (c) radon gas in ground water wells. Water level can be changed in the deep wells are considered as an indication for EQ. The water level can be gradually low or it can be rapidly high within a period of month. Chemical changes in ground water was found by the researchers of Tokyo University as they found the chloride concentration almost similar and also the level of sulphate was rise after an EQ . Another researcher found that increasing level of radon gas is another precursor for EQ. All these instances cannot be determined for a man easily without any analytical inquisition.

LR-EQP uses six the attributes which are related geologically: (i) Lat-Long is used to get the exact location of that area which area needs to verify. (ii) CE is such a kind of stuff as these are rubbing with each other, then it creates a huge temperature which is also responsible for shaking rock. (iii) DP can be used for measuring the EQP as population also creates a pressure on surface. (iv) Measuring distance of an area from nearest TP. Because where there is a TP there is a probable chance for EQ. (v) ST is also used by LR-EQP because with this we can get notify about how much elements that are responsible for earthquake is located that area. All these attributes are related with EQ. Anyone can grab the changes these attributes easily comparing with those 3 factors [4] which are used in earthquake prediction using data mining. If we can use our six factors with the earthquake prediction using data mining's [4] three factor, then it may give a better result for the measuring the probability of EQ.

The rest of the paper is organized as follows. Section 2 summarizes some related works. Section 3 describes the

978-1-5386-8158-9/19/$31.00©2019IEEE

methodology of LR-EQP. Section 4 illustrates the experimental studies. Finally Section 5 concludes the paper.

II. Related Works

Earthquake prediction using data mining is approach to get the probability of EQ [4]. If the ground water level is become low or high then it may be taken as sign of probable earthquake. Water level of a deep wells cannot be grown or reduce rapidly as it happens naturally. Growth of Some chemical elements are found under the water level like chloride, sulphate etc. Increasing level of radon gas is also consider as a precursor of EQ.

A prediction method was proposed [5], as each transport that carry a line sensor with magnetic forces, GPS position detector and a data transmitter observe those area which transmit magnetic field data which are targeted for EQ prediction. Telluric current for those observed area received through the data for predicting EQ.

Another method for prediction of EQ [6] was proposed based on gases under the surface like He, Rn, Hg and so on gases are found nearly to the active fault. And several changes in the quantity of those gases are found before a large EQ.These changes can be also occurred because of meteorological, hydrological and some other circumstances of environment but the major changes are found before an EQ.

III.Earthquake Probability (EQP) using Linear Regression (LR-EQP)

For finding a better offshoot about the probability of EQ, LR-EQP use some attributes; these are: (i) Lat-Long (ii) CE (iii) DP (iv) Distance from nearest TP (v) ST.LR-EQP consists of several steps which can be described as Fig. 1.

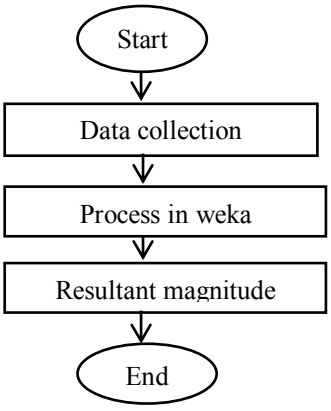

Fig. 1. Flow chart of LR-EQP

Step 1:

Data collection: LR-EQP deploys five attributes. These data are collected from different sources and used for EQP for an area.

i. Lat-Long are used to get the exact location of that area which we choose, with this we can also get exact situation of the tectonic plate which's movement is the main reason for EQ.

ii. **CE** is also responsible as they are rubbing with each other and create a huge temperature and if the temperature is created around a fault which is also responsible for EQ. The temperature which is created around the fault is also responsible for temperature rise on surface. This rise of temperature is also felt by the people or other living animals on the surface. As the humidity is increased that time on the surface. People on the surface feel suffocation [7].With the Lat-Long we get that information about where the fault is and what is that current situation.

iii. **DP** is another instances for LR-EQP. DP is needed to measure the pressure on surface. When there is a pressure on the layer of the surface then there is a probable chance of getting the CE come closer and create a temperature. Which is the reason behind the EQ.

iv. LR-EQP measure distance of an area from nearest **TP**. Because where there is TP there is most probable chance for EQ. [8].

v. **ST** is only an attribute for which we used non-numeric value. Here the non-numeric value is taken randomly. Because ST can tells us how much CE can it preserve or not. We assume a number against the soil type for our calculation. The types are – (a) alluvial-1 (b) loamy-4 (c) clay-1(both clay and alluvial are same type) (d) black soil-3 (e) lateritic-2.

All these values are correlated and gradually growing of these attributes or lowering of these attributes tells us what the Probability of EQ.LR-EQP give more accurate result if real time data can be collected. LR-EQP takes all the data set from various online sources.

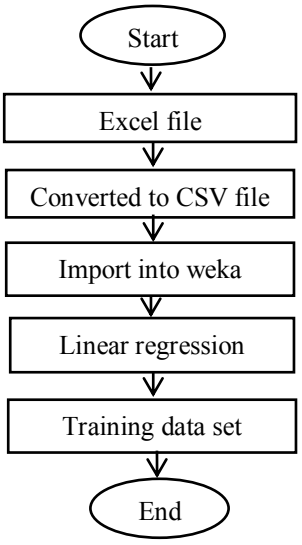

Fig. 2. Flow chart for Processing steps in Weka

Step 2:

Process in weka: The flow chart for processing the data in weka is shown in Fig. 2.Weka is a Data mining tool which is based on machine learning algorithm for processing data. Here all the data that is collected need to prepare, classify,

clustering because all the data is collected from various resources of web. So the data need to be fit for the project work.

The steps in weka are:

i. All the data (collected by step 1) of the attributes are collected in a excel file and implement into weka for processing. Fig. 3 shows a data set for LR-EQP.

	A	B	C	D	E	F	G
1	Places	Latitude	Longitude	Magnesium	Density of population	Distance from nearest Techtonic plate	Soil type
2	Srinagar	34.0836	74.7973	32.5	1.57	113	clay 1

Fig. 3. A screenshot of Data file (excel file) for LR-EQP

ii. Before implemented into weka the excel file is converted into CSV file, as weka processed next steps with this CSV file.

iii. Weka use this CSV file for preprocessing data (remove outliers, replace missing values etc.), select attribute, choose coherent attribute, and remove incoherent attribute, classification and assessment [9]

iv. After preprocessing data file LR-EQP is implemented through linear regression. Linear regression gives a probable value for magnitude on the other hand Naïve Bayes or other algorithm gives an exact value not fraction.

v. The training data set which is get from weka is used for further measurement of magnitude for EQ. Fig. 4 shows some training data set. As linear regression is used for the process so the data set should be appeared linearly. Here the training data set is not exact linear as shortage of data. Because is to is tough to collect geological data for each and every places.

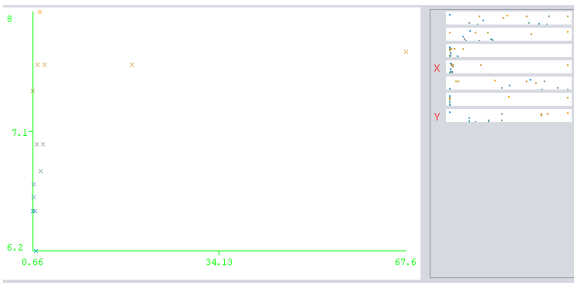

Fig. 4. Training data set of the collected data for LR-EQP

Step 3:

Resultant magnitude: After getting the training data set from weka (from step 2) it is used as a model for getting the magnitude for EQ. Before getting the magnitude of an area LR-EQP needed to be given another magnitude of the relevant area which is the magnitude of previous EQ. Without this manually given magnitude the resultant magnitude cannot be acquired. Fig. 5 shows the magnitude *vs* Probable magnitude, here X-axis represent the magnitude while Y-axis represent the probable magnitude. From Fig.5 we get that right portion is red and it indicates that those are critical places. There is most probable chances for EQ.Left portion of the places is green it indicates less critical places for probable magnitude.

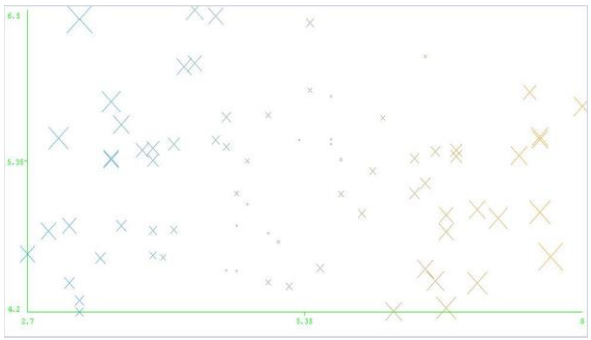

Fig. 5. Magnitude vs. Probable magnitude from the data set of LR-EQP

IV. Experimental Analysis

A. Experimental Setup:

LR-EQP has been developed under the environment on Intel Core i3-2.30 GHz processor with 4.0 GBytes of RAM running on Windows 8.1 operating system. LR-EQP has been developed in Weka [10] for preparing a model which is further implemented in Android studio [11] we get an APK file from android studio. APK file is an application for android operating system. Here android studio is used to implement the model get from weka. The APK file give the probable magnitude for that relevant area.

B. Experimental Result and Comparisons:

The whole experimental result is shown in the Fig. 6. This figure show the magnitude of Jammu region of India. The probable magnitude can be get after selecting the region. There was an EQ in Jammu (Indian region) in 2013 and the magnitude of that EQ was 5.7 [12]. LR-EQP uses data of Jammu (Indian region) earlier 5 years of 2013 and give the result of magnitude 5.5. After selecting a region that APK show the probable magnitude for an EQ.

Fig. 6. EQP using LR-EQP

The experimental results presented in the above section can be measure the magnitude and give output near about original magnitude. Table I shows the comparison of the magnitude.

TABLE I. COMPARISON OF MAGNITUDE

Year	Region	Actual magnitude(x_a)	Magnitude by LR-EQP(x_p)
2013	Jammu	5.7 [12]	5.5
2006	Gujarat	5.5 [13]	6.67

The fluctuation of magnitude from the original magnitude is given below finding the standard deviation.

$$sd = \sqrt{\frac{\sum(x_a - x_p)^2}{n-1}}$$

$$= \sqrt{\frac{(5.7-5.5)^2+(5.5-6.67)^2}{2-1}}$$

$$= 1.4089$$

Where,

x_a= actual magnitude

x_p=probable magnitude.

n = number of attributes

Standard deviation is a term which measures the diffraction of data set [14]. Higher value in standard deviation implies that data points are extended over a wider range of value and the lower value indicated that data points are closed to the expected values. Here in LR-EQP the standard deviation is 1.4089 which indicates that the data set for LR-EQP is closed to the expected value or the value of magnitude which we want to find out as a probable magnitude for the specific area.

C. Discussions:

LR-EQP process data of various Indian state like Srinagar, Jammu, Patna, Guwahati etc. all these data set for this region is collected from various online sources. But LR-EQP gives a better feedback if all the data set can be collected as a real-time data. The fluctuation of data from the original value is only getting lower if it is possible to manage real time data set for this approach.

V. CONCLUSION

According to science EQP is improbable. Success for predicting probability for EQ from first physical principles not happen yet. LR-EQP can be used for getting the information about the probability of EQ. It is a very laborious work as many other tools are used to get the better consequence from all the noisy data set. EQP cannot be totally metering without comparing the many more geological data rather than only water level, chemical changes, radon gas. The effect of these 3 factors can be detected very slowly as the result of all these cannot be finding without empirical analysis. To escape or to know about EQ people need some phenomena which they can realize easily without any analytical report. For further work this methodology can be modified with more geological data as the probability of EQ can more precisely get by any one as like as if there is cloud in the sky then people can easily realize that there will be rain.

REFERENCES

[1] Wikipedia: The Free Encyclopedia. Wikimedia Foundation Inc. Updated 31 August 2018, at 09:48 UTC. Encyclopedia online. Available from:https://en.wikipedia.org/wiki/ Natural disaster.

[2] Wikipedia: The Free Encyclopedia. Wikimedia Foundation Inc. Updated 31 August 2018, at 09:48 UTC. Encyclopedia online. Availablefrom:https://en.wikipedia.org/wiki/April_2015_Nepal_earthquake.

[3] World vision staff,"2018 Indonesia earthquake", Updated 12 October 2018, Available from : https://www.worldvision.org/disaster-relief-news-stories/2018-indonesia-earthquake-facts.

[4] Radhika Kulkarni, Rashmi Kulkarni, "Earthquake Prediction Using Data Mining," IJSR, vol. 6, 2017.

[5] Tomonari Yomoda, Toshio Yoshida, "Earthquake prediction method and system thereof,"vol.7, 2007.

[6] Chi-Yu King, "Gas geochemistry applied to earthquake prediction: An overview,"1986.

[7] Catherine Zandonella, "Temperature rises hint at earthquake prediction,"2001,https://www.newscientist.com/article/dn1696-temperature-rises-hint-at-earthquake- prediction/.

[8] Wikipedia: The Free Encyclopedia. Wikimedia Foundation Inc. 16 August 2018, at 16:21 (UTC). Encyclopedia online. Available from: https://en.wikipedia.org/wiki/Plate_tectonics.

[9] Jasim, Dalia, "main steps for doing data mining project using weka," Book, 2016.

[10] "Weka",available at:https://Sourceforge.net/projects/weka/.

[11] "Android_Studio",available:https://developer.android.com/studio/2018.

[12] Wikipedia: The Free Encyclopedia. Wikimedia Foundation Inc. 12 July 2018, at 10:19 (UTC).Encyclopedia online. Available from: https://en.wikipedia.org/wiki/List_of_earthquakes_in_India.

[13] Wikipedia: The Free Encyclopedia. Wikimedia Foundation Inc. 12 July 2018, at 10:19(UTC).Encyclopedia online. Available from: https://en.wikipedia.org/wiki/List_of_earthquakes_in_India.

[14] Wikipedia: The Free Encyclopedia. Wikimedia Foundation Inc. 8 October 2018, at 16:21 (UTC). Encyclopedia online. Available from: https://en.wikipedia.org/wiki/Standard_deviation