Seismic Data Analysis for Earthquake Prediction: A Machine Learning Approach

Adqas Ikram[1] and Usman Qamar[2]

Computer Engineering Department,
College of Electrical and Mechanical Engineering,
National University of Sciences and Technology (NUST),
Pakistan

Dec 03, 2016

Summary

About two hundred distinct quakes of various magnitudes are recorded on our planet every day. Forecasting an earthquake is a difficult job which gets even more difficult because of our lack of understanding of the phenomenon. Data mining, is the process of discovering hidden patterns in the data. Machine learning is part of data mining, which can be considered as the process of automatic pattern recognition based on training data. This research focuses on analyzing which machine learning techniques are most suitable for earthquake prediction.

KEYWORDS: Earthquake, Data mining, Machine learning, Seismic data analysis, Prediction

1. **Introduction**

Earthquake prediction is an important topic. Extensive research is being done on it such as papers written by Asencio-Cortés et al. (2016), Aqdas et al. (2015), Uyeda (2015), Florido et al. (2015) and Dutta et al. (2016). Machine learning techniques have recently become prominent in this field. Machine learning algorithms can be categorized into supervised learning and unsupervised learning as mentioned by Idowu et al. (2016). In supervised learning, pairs of input and target output are given to train a function. In unsupervised learning, no label is given in sampled data.

2. Seismic Data Acquisition

The most important aspect of data mining is the quality of data. Therefore the data was acquired from Advanced National Seismic System (ANSS). For ANSS, data from 1970 to 2012 was retrieved with 2773727 distinct quake records. The dataset was scrubbed against missing magnitude, latitude, longitudes and depth values. Since data is bimodal, therefore replacing any missing value against any mode is not a good option; therefore, all examples with missing values were discarded.

3. Analysis of Machine Learning Techniques

The aim of this research is to evaluate various machine learning techniques for predicting earthquakes. We have considered seven most popular techniques i.e. Linear Regression, Polynomial Regression, Local Polynomial Regression, Vector Linear Regression, Gaussian Process, Support Vector Machine (SVM) and Neural Networks. The performance measures we have considered are root-mean-square error, mean-absolute error, root-relative-squared error, relative-absolute error, correlation coefficient. To study these techniques we have restricted the data to four attributes, i.e. latitude, longitude, magnitude and depth. After normalization data, it is divided exactly in two halves. One is defined as training set and the other is testing set. We used holdout method for evaluation. This method is simple and is still compatible with other methods like cross validation as mentioned by Muñoz et al. (2015).

[1] aqdas27sep@yahoo.com
[2] usmanq@ceme.nust.edu.pk

3.1 Latitude. Latitude is a virtual line that cuts the earth horizontally. If latitude is predicted successfully, considerable intelligent guess work can be made about prospective longitude. Performance output of latitude prediction against different machine learning techniques is shown in Table 1.

Table 1 Performance measures for predicting latitude.

L A T I T U D E					
Techniques	**Performance Measures**				
	Root-mean-square error	**Mean-absolute error**	**Root-relative-squared error**	**Relative-absolute error**	**Correlation coefficient**
Linear Regression	**30.062**	**25.433**	**1.006**	**0.998**	0.097
Polynomial Regression	80.7311	67.492	2.703	2.648	0.051
Local Polynomial Regression	7438.954	415.676	249.022	16.309	0.008
Vector Linear Regression	30.064	25.435	**1.006**	**0.998**	0.097
Gaussian Process	31.610	26.372	1.058	1.035	-0.005
Support Vector Machine	31.091	26.159	1.041	1.026	-0.015
Neural Network	37.933	30.724	1.270	1.206	**0.110**

Following results are concluded from table 1:

- Latitudes are modeled with minimum errors using linear regression models. This reinforces our understanding that all latitudes are parallel lines; therefore have same perpendicular distance between any two latitudes from all points. Vector linear regression & SVM is a close match.
- Neural networks produce relatively better model in terms of similarity but other performance parameter suggest the modeled attribute has considerable drift from original.
- Non-linear models can’t predict the latitude well.

3.2 Longitude. Longitude is a virtual line that cuts the earth vertically. Again if longitude is predicted successfully, considerable intelligent guess work can be made about latitude. However, predicting longitude is trickier because longitudes converge towards pole and diverge towards equator. Two quakes 5000 km apart horizontally on equator will have less variations of longitude as compare to two quakes 5000 km apart horizontally in arctic circle. Performance parameters output of latitude prediction against different machine learning techniques is shown in Table 2.

Following results are concluded from table 2:

- No single models has produced consistent results.
- All polynomial regression and Gaussian process models can’t handle the longitude.
- Root-mean-square error values are higher as compare to root-mean-square error values for latitude. These large discrepancies support the hypothesis of longitude convergence towards pole and divergence towards equator.

Table 2 Performance measures for predicting longitude.

L O N G I T U D E					
Techniques	**Performance Measures**				
	Root-mean-square error	**Mean-absolute error**	**Root-relative-squared error**	**Relative-absolute error**	**Correlation coefficient**
Linear Regression	**116.978**	105.801	**0.996**	0.994	0.095
Polynomial Regression	140.071	120.563	1.192	1.133	0.031
Local Polynomial Regression	13572.873	1400.083	115.530	13.160	-0.003
Vector Linear Regression	116.991	105.908	**0.996**	0.995	0.095
Gaussian Process	124.192	112.018	1.057	1.053	0.008
Support Vector Machine	121.031	**103.764**	1.0302	**0.975**	0.094
Neural Network	120.483	110.404	1.026	1.038	**0.115**

3.3 Magnitude. Magnitude is a logarithmic scale associated with the shake of the ground. Performance parameters output of magnitude prediction against different machine learning techniques is shown in Table 3.

Table 3 Performance measures for predicting magnitude.

M A G N I T U D E					
Techniques	**Performance Measures**				
	Root-mean-square error	**Mean-absolute error**	**Root-relative-squared error**	**Relative-absolute error**	**Correlation coefficient**
Linear Regression	0.641	0.534	1.220	1.331	0.068
Polynomial Regression	68.699	59.591	130.719	148.589	-0.026
Local Polynomial Regression	110.086	7.389	209.469	18.425	0.004
Vector Linear Regression	0.640	0.533	1.219	1.329	0.067
Gaussian Process	10.374	4.267	19.740	10.639	-0.002
Support Vector Machine	**0.635**	**0.527**	**1.207**	**1.313**	0.069
Neural Network	1.154	1.065	2.196	2.655	**0.072**

Following result is concluded from table 3:

- SVM has produced the best results as compared with all other models. Correlation coefficient of SVM is just trailing the Neural Networks.

3.4 Depth. Depth defines the distance in kilometer between epicenter and the actual point of energy-release beneath the ground. Performance parameters output of depth prediction against different machine learning techniques is shown in Table 4.

Table 4 Performance measures for predicting depth.

D E P T H					
Techniques	**Performance Measures**				
	Root-mean-square error	**Mean-absolute error**	**Root-relative-squared error**	**Relative-absolute error**	**Correlation coefficient**
Linear Regression	123.200	79.700	1.001	1.053	0.039
Polynomial Regression	148.315	89.170	1.205	1.179	0.031
Local Polynomial Regression	10942.541	1013.314	88.924	13.394	0.007
Vector Linear Regression	123.209	79.752	1.001	1.054	0.038
Gaussian Process	146.012	78.944	1.187	1.043	0.014
Support Vector Machine	130.908	**59.409**	1.064	**0.785**	0.029
Neural Network	**123.062**	72.323	**1.000**	0.956	**0.050**

Following results are concluded from table 4:

- Root-mean-square error values suggest there are large fluctuations in predicted and true values.
- Unlike longitude with an apparent cause for variations; depth has some hidden cause forcing it to behave with large variations. These variations are large enough to be visible on mean-absolute error scale.
- It is difficult to choose between SVM or Neural Network as better predicting option than other, however, Neural Networks maintained unchallenged best correlation coefficient values for each and every attribute.

4. Conclusion

Earthquake prediction is a challenge. Recently machine learning techniques have also become prominent in this field. The aim of this investigation was to analyze different machine learning techniques for earthquake prediction. It was seen that for latitudes are modeled with minimum errors using linear regression models. All polynomial regression and Gaussian process models can’t well-define the longitude. SVM produced the best results for magnitude and Neural Networks for depth. These results could be used as reference for further studies for predication of earthquakes.

5. Biographies

Mr Ikram did his Maters from NUST. Dr. Qamar is the head of Knowledge and Data Engineering Research Centre at NUST. He has done his PhD from University of Manchester, UK.

References

Asencio-Cortés, G., Martínez-Álvarez, F., Morales-Esteban, A. and Reyes, J., 2016. A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction. *Knowledge-Based Systems*, *101*, pp.15-30.

Ikram, A. and Qamar, U., 2015. Developing an expert system based on association rules and predicate logic for earthquake prediction. *Knowledge-Based Systems*, *75*, pp.87-103.

Uyeda, S., 2015. Current affairs in earthquake prediction in Japan. *Journal of Asian Earth Sciences*, *114*, pp.431-434.

Florido, E., Martínez-Álvarez, F., Morales-Esteban, A., Reyes, J. and Aznarte-Mellado, J.L., 2015. Detecting precursory patterns to enhance earthquake prediction in Chile. *Computers & Geosciences*, *76*, pp.112-120.

Dutta, S.C., Nayak, S., Acharjee, G., Panda, S.K. and Das, P.K., 2016. Gorkha (Nepal) earthquake of April 25, 2015: Actual damage, retrofitting measures and prediction by RVS for a few typical structures. *Soil Dynamics and Earthquake Engineering*, *89*, pp.171-184.

Idowu, S., Saguna, S., Åhlund, C. and Schelén, O., 2016. Applied machine learning: Forecasting heat load in district heating system. *Energy and Buildings*, *133,* pp.478-488.

Muñoz, Mario A., Yuan Sun, Michael Kirley, and Saman K. Halgamuge. Algorithm selection for black-box continuous optimization problems: A survey on methods and challenges. *Information Sciences*, 317 (2015): 224-245.