Earthquake Magnitude Prediction using Probabilistic Classifiers

Mahshid Fahandezhsadi (fahandezhsaadi20@gmail.com)
Graduate University of Advanced Technology https://orcid.org/0000-0002-2154-7561

Hamed Fahandezh Sadi
University of Tehran

Methodology

Keywords: Earthquake prediction, Tree Augmented Naïve Bayes, Hidden naive Bayes

Posted Date: June 25th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-36094/v1

Earthquake Magnitude prediction using Probabilistic Classifiers

[1]Fahandezhsadi, M.* and [2]Fahandezh Sadi, H.

*Correspondence: fahandezhsaadi20@gmail.com

Abstract

Background: Earthquake prediction plays an important role in preventing catastrophic damage. We present here a scheme for earthquake prediction using an improved version Tree Augmented Naïve Bayes (TAN). This approach can achieve an appropriate performance by extracting dependencies among seismicity features effectively. At first we use the simplest Discretization Method, Equal Interval Width, for discretization of six seismicity features (Time, Mean Magnitude, Energy, Slope, deviation, Magnitude deficit) drive from National Earthquake Information Center (NEIC), then in order to the magnitude of earthquake we utilize an improve version of Tree Augmented Naïve Bayes that is based on Decomposable Models.

Results: Finally, we test our method by using two schemes and compare it to Tree Augmented Naïve Bayes and Hidden naive Bayes classifier. F-measure metric is used for evaluation of our results.

Conclusion: Experimental results demonstrated proposed approach based on an improved version of Tree Augmented Naïve Bayes achieves a higher performance compared to two other methods.

Keywords: Earthquake prediction, Tree Augmented Naïve Bayes, Hidden naive Bayes

Introduction:

Earthquake, one of the deathliest natural phenomena occurs as a consequence of a sudden movement of transition to the ground owing to the release of elastic energy in a few seconds. The impact of this sudden and unpredictable event is catastrophic as it affects a large area. Physical modeling is not

able to provide accurate earthquake predictions because of the complex nonlinear behavior of seismicity. In the recent years, research community have started to utilize the statistical methods for earthquake magnitude prediction. In these methods usually the earthquakes information is summarized to some features and then classification process is done.

A number of earthquake magnitude prediction schemes are shown in the previous literatures [1-7]. For example, the prediction of earthquakes of medium-large magnitude for the city of Tokyo and surroundings has been addressed by means of artificial neural networks. The outcomes point out that these networks are able to achieve the accuracy greater than 70 % for all the five consecutive datasets considered [1]. In [2], three-layer feedforward artificial neural networks have been constructed to predict earthquakes in four seismogenic areas in Chile. These networks use information related to the b-value, the Bath's law and the Omori/Utsu's law as input parameters and they predict the occurrence of earthquakes for a five-day horizon with great reliability. In [3], the writers decided to apply these mathematical models based on Artificial Neural Networks (ANNs) models to predict earthquakes. The goal of this experimental study was to examine the capability of advanced ANNs and machine learning to estimate the magnitude of the events recorded daily. Features that described each event were: origin time (UTC), latitude, longitude, depth, and magnitude. The results that the ANNs have very good performances both in functional approximation and in pattern recognition when the training set represents a sample of worldwide earthquakes. In [4], four machine learning techniques including pattern recognition neural network, recurrent neural network, random forest and linear programming boost ensemble classifier have been used to predict earthquakes in the subducting Hindukush region, which is one of the most active seismic regions of the world. Eight mathematically calculated seismicity parameters including the time elapsed over a predefined number of events, the mean magnitude of the last n events, the rate of release of square root of energy, the slope of the Gutenberg-Richter inverse power-law curve,

the summation of the mean square deviation about the regression line based on the Gutenberg-Richter inverse power law and the magnitude deficit or the difference between observed and expected magnitudes based on the Gutenberg-Richter inverse power law were utilized as features. Their outcomes demonstrate that Linear Programming Boost ensemble classifier is able to achieve better results in terms of sensitivity, while Pattern Recognition Neural Network tends to produce the least false alarms in comparison with the other classifiers. In [5], the different characteristics of blasts and seismic events were investigated by comparing three different classifiers including Fisher classifier, Naive Bayesian classifier and Logistic Regression (LR). Their research was conducted on three databases from Australia and Canada. Cross-validated results have represented that discriminator LA has an appropriate discriminating performance. The writers in [6] develop a novel Hidden Markov Model for earthquake modeling and forecasting by introducing a latent Markov process to model the unobservable state of the underground dynamics. Their proposed model is capable of forecasting the change-in-state of the hidden Markov chain, and thus can predict the arrival time and magnitude of future earthquake occurrences simultaneously. In [7], to forecast the earthquake magnitude, several algorithms including SVM (support-vector machine), M5P (Multivariate Regression prediction), Naive Bayes, KNN (k-nearest neighbors), J48, Random Forest, LP Boost Ensemble are considered. The results indicate that performance of artificial neural networks is a lot better than other algorithms. In this paper we proposed to apply an improved version of Tree Augmented Naïve Bayes (TAN) based on Decomposable Models [8] for earthquake prediction. In our work the classification of seismicity samples is based on maximum likelihood which predicts the magnitude of earthquake. For this purpose we approximate the joined probability distribution using Decomposable Models. Our approach is evaluated by using the data sat that is available in National Earthquake Information Center (NEIC). Experimental result showed that our proposed modeling approach based on Decomposable Models provides an appropriate performance

compared to Tree Augmented Naïve Bayes (TAN) and Hidden naive Bayes (HNB). Because our modeling unlike models used in TAN and HNB allows to contribute three- order marginal distributions along one- and two- order to approximate the joint probability distribution.

The proposed method:

As mentioned previously, many algorithms have been developed to improve the accuracy of Naive Bayes (NB) by relaxing its strong conditional independence assumption. Here, we propose to utilize an algorithm that is developed in [8] based on decomposable models to improve the performance of Tree Augmented Naive Bayes (TAN) classifier that is an improved version of NB and called a 1-dependence classifier.

By examining the basic relations behind TAN, we can see that it is actually a simple decomposable model that uses only one- and two- dimensional marginal distributions to approximate a high dimensional joint probability distribution. Here, we apply a simple decomposable model in which the three-dimensional marginal distributions are also utilized for modeling and therefore the obtained model is able to effectively capture the interdependencies among different features.

Decomposable models are a special type of graphs in which each cycle with a length greater than three contains an edge that connects its two non-adjacent vertices. Under this circumstances, we can calculate the joint probability distribution that describes a decomposable model, easily. For this purpose, we firstly present some preliminary definitions:

Complete graph: Every two vertices in a complete graph are connected by an edge.

Clique: A clique is a subset of vertices that their corresponding subgraph is complete.

Maximal clique: Maximal clique is a clique that is not a subset of a large clique.

Weighted clique intersection graph: A decomposable model has a weighted clique intersection graph in which maximal cliques correspond to vertices and their vertices share an edge if they

intersect. In this graph, the weight of each edge is determined by the number of variables shared between each two vertices.

For a decomposable model indicated by a junction tree including a set of maximal cliques (c) and a set of separators (s), the joint probability distribution of n discrete variables $X = \{X_1, X_2 ...., X_n\}$ can be calculated as the following:

$$p(\mathrm{X}) = \frac{\prod_{C \in c} p(x_c)}{\prod_{s \in s} p(x_s)^{\nu_s - 1}} \tag{1}$$

Where each clique C in the set of cliques c is corresponding to the $|C|$-dimensional random variable X_c, each separator S in the set of separator s is corresponding to the $|S|$-dimensional random variable X_s and ν_s is the number of cliques that include all of the variables in S.

Please note that the term 'separator' refers to the variables shared between maximal cliques in the junction tree.

To form an appropriate decomposable model, following steps are utilized:

1) The conditional mutual information between every two features (or variables) is calculated as following:

$$I\left(A_i ; A_j \middle| K\right) = \sum_{a_i, a_j, k} p(a_i, a_j, k) \log \frac{p(a_i, a_j \mid k)}{p(a_i \mid k) p(a_j \mid k)} \tag{2}$$

Where

2) For each pair of feature A_i and A_j, the weight $I(A_i : A_j | K)$ is assigned in to the edge between A_i and A_j.

3) The edges are sorted in descending order of weights, i.e. the edges e_1 and $\frac{e_{n(n-1)}}{2}$ are corresponding to the maximum and minimum weights, respectively.

4) An initial graph is built by using of two edges which have the maximum weights (e_1 and e_2). Then, some edges are gradually selected among e_3, e_4, e_5 to add to the graph. An edge is added to the set of previously selected edges, if and only if, the length of the longest path between the vertices of that edge does not exceed 2. The output of this step is a decomposable graph.
5) A weighted clique intersection graph is built based on the decomposable graph formed in the step 4.
6) The junction tree corresponding to the weighted clique intersection graph is built.
7) Finally, the joint probability distribution of n discrete variables is calculated by using maximal cliques and separators.

For more details related to decomposable model please refer to [8].

Experiments with seismic data

In this section, we evaluate the performance of our scheme to forecast earthquake magnitude using a seismic dataset. Accurate earthquake prediction can have an effective role in preventing catastrophic damage. Hence, the research community has highly focused on the scheme related to earthquake prediction.

DATASET:

In this study, we carry out some experiments on a subset of a seismic dataset provided by the National Earthquake Information Center (NEIC). This dataset includes a record of the date, time, location, depth, magnitude, and source of every earthquake with a reported magnitude 5.5 or higher since 1965. These subset is categorized into two and three categories by considering two schemes as follows:

- ✓ Scheme A: subset is divided into two classes by considering the threshold value 6.5 on magnitude. In other words, samples with the magnitude between 5.5 and 6.5 and larger than 6.6 are labeled as class 1 and class 2, respectively. In this scheme, 3000 samples are selected including 1000 training samples (e.g. 500 samples for each class) and 2000 test samples in a random way.
- ✓ Scheme B: subset is divided into three classes by considering the threshold values 6 and 7 on magnitude. In other words, samples with the magnitude in the range of 5.5 to 6, 6 to 7 and larger than 7 are labeled as class 1, class 2 and class 3, respectively. In this scheme, 5400 samples are selected including 900 training samples (e.g. 300 samples for each class) and 4500 test samples in a random way.

Discretized Features:

In two schemes, each sample is represented by eight mathematically-defined seismic features [9]. These features are listed in the following:

1) The time elapsed over a predefined number of events (T).
2) The mean magnitude of the last n events.
3) The rate of square root of seismic energy released over time T.
4) Slope of the log of the earthquake frequency versus magnitude curve.
5) Summation of the mean square deviation from the regression line based on the Gutenberg Richter inverse power law.
6) The difference between the largest observed magnitude and the largest expected magnitude
7) Based on the Gutenberg-Richter relationship the average time or gap observed between characteristic or typical events among the last n events.

8) Coefficient of variation of the mean time between characteristic events known as the aperiodicity of the mean.

Please note that six out of the eight features mentioned above are finally utilized since the last two features produced a large number of zeros based on http://github.com/sajeed786/Earthquake-Prediction. Furthermore, the simplest discretization method, Equal Interval Width, is employed to discretize continuous features as discretized features can improve the performance of some algorithms rather than continuous features.

EVALUATION METRICS:

In order to evaluate the performance of our scheme, F-measure is utilized [10]. This criteria is defined as follows:

$$F-measure = \frac{2 \times precision \times recall}{precision + recall}$$

$$recall = \frac{TP}{TP + FN}$$

$$precision = \frac{T_p}{TP + FP}$$

Where we have:

- True Positive (TP) = the number of samples correctly classified as belonging to class X.
- True Negative (TN) = the number of samples correctly classified as not belonging to class X.
- False Positive (FP) = the number of samples incorrectly classified as belonging to class X.
- False Negative (FN) = the number of samples incorrectly classified as not belonging to class X.

Table 1: The performance of three algorithms based on F-measure per- lass for scheme A

Class number	TAN	HNB	Proposed scheme
1	84.88	88.91	**90.06**
2	85.31	88.48	**90.14**
Ave	85.10	88.70	**90.1**

Table 2: The performance of three algorithms based on F-measure per- lass for scheme B

Class number	TAN	HNB	Proposed scheme
1	78.80	.79.26	**87.51**
2	80.04	78.19	**88.35**
3	80.73	79.41	**88.64**
Ave	79.85	78.95	**88.17**

Results:

To provide a comprehensive evaluation, we choose TAN and Hidden Naïve Bayes (HNB) for comparison with our presented scheme for earthquake prediction. TAN classifier is an extension of Chow-Liu algorithm that utilizes conditional mutual information to find an MST efficiently as a classifier. The Chow–Liu algorithm approximates a joint probability distribution by use of first- and second- order distributions optimally. In HNB, a hidden parent is created for each feature which combines the influences from all other attributes. However, this model uses only first- and second-order distributions to approximate the joint probability distribution, too. In comparison, our scheme allows to contribute three order distributions along first- and second- order distributions for approximation.

Table 1 and 2 illustrate the results of these algorithms per class for scheme A and scheme B. In these tables, the higher values are shown in bold font.

From the results, it is observed that the presented scheme provided a better classification performance per-class compared with TAN and HNB classifiers. It is valid for all classes on both schemes.

Table 1 shows that for scheme A, TAN, HNB classifiers achieved the Ave F-measure values 85.10% and 88.70%, respectively. While our scheme provides an improvement of Ave F-measure value, 90.10%. The superiority of our method for earthquake prediction is also validated using scheme B. For this scheme, our method reaches 88.17% of Ave F-measure value. Corresponding values for TAN and HNB classifiers are 79.86% and 78.95%.

Conclusion

In this paper, we proposed to apply an improved version of TAN classifier for earthquake prediction. This classifier is able to effectively capture the interaction between different seismicity features. Our scheme was implemented and compared to two other methods such as TAN and HNB using a seismicity data set. The results shown that the superiority of our scheme compared to two other classifier.

Abbreviations

TAN: Tree Augmented Naïve Bayes; HNB: Hidden naive Bayes classifier; NEIC: National Earthquake Information Center

Ethics approval and consent to participate

Not applicable

Consent for publication:

Not applicable

Availability of data and materials:

All the testing data are available in the GitHub repository http://github.com/sajeed786/Earthquake-Prediction

Competing interests:

The authors declare that they have no competing interests.

Funding:

This work is supported by 1Graduate Geophysics, Earth Science Department, Graduate University of Advanced technology, Kerman, Iran.

Authors' contributions:

All the authors contributed in equal parts. All authors read and approved the final manuscript.

Acknowledgements:

Not applicable.

1. Asencio-Cortés G, Martínez-Álvarez F, Troncoso A, Morales-Esteban A. Medium–large earthquake magnitude prediction in Tokyo with artificial neural networks. Neural Computing and Applications. 2015;28(5):1043-1055.
2. Reyes J, Morales-Esteban A, Martínez-Álvarez F. Neural networks to predict earthquakes in Chile. Applied Soft Computing. 2013;13(2):1314-1328.
3. Buscema, P. M., Massini, G., & Maurelli, G. Artificial Adaptive Systems to predict the magnitude of earthquakes. Bollettino di Geofisica Teorica ed Applicata. 2015;56(2).
4. Asim K, Martínez-Álvarez F, Basit A, Iqbal T. Earthquake magnitude prediction in Hindukush region using machine learning techniques. Natural Hazards. 2016;85(1):471-486.
5. Dong L, Wesseloo J, Potvin Y, Li X. Discrimination of Mine Seismic Events and Blasts Using the Fisher Classifier, Naive Bayesian Classifier and Logistic Regression. Rock Mechanics and Rock Engineering. 2015;49(1):183-211.
6. Yip, C. F., Ng, W. L., & Yau, C. Y.A. hidden Markov model for earthquake prediction. Stochastic environmental research and risk assessment. 2019;32(5):1415-1434.
7. Florido, E., Aznarte, J. L., Morales-Esteban, A., & Martínez-Álvarez, F. Earthquake magnitude prediction based on artificial neural networks: A survey. Croatian Operational Research Review. 2016;7(2),159-169.
8. Ghofrani, F., Keshavarz-Haddad, A., & Jamshidi, A. A new probabilistic classifier based on decomposable models with application to internet traffic. Pattern Recognition. 2018;77,1-11.
9. Panakkat, A., & Adeli, H. Neural network models for earthquake magnitude prediction using multiple seismicity indicators. International journal of neural systems. 2007;17(01):13-33.
10. Zhang, J., Chen, C., Xiang, Y., Zhou, W., & Xiang, Y. Internet traffic classification by aggregating correlated naive bayes predictions, IEEE Trans. Inf. Forensics Secur. 2013;8(1): 5–15.