An earthquake prediction model based on precursor window detection

AI Yong1,2*, WANG Xin-an1, YONG Shan-shan1, SUN Chong2
1School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, P.R. China
2College Of Computer Science, South-Central University for Nationalities, P.R. China
aiy_scuec@qq.com

Abstract. The research of the earthquake precursor signal anomaly is one of the main research directions of short-term and imminent earthquake prediction. An earthquake prediction method based on time precursory window is proposed in this paper, which is based on the low-frequency electromagnetic signals collected by AETA. Firstly, the prediction model of historical low-frequency electromagnetic signals is constructed by machine learning method. The model is used to detect whether the current time period is in the window of earthquake precursors. Furthermore, two algorithms based on single-site and group-site position prediction is proposed in this paper. The algorithm filters three or more stations within the effective distance range, and uses the probability of earthquake occurrence as the weight to locate the earthquake center, so as to predict the position of earthquake occurrence. Finally, the real data set is tested on the earthquake of Qingchuan County, Guangyuan City, Sichuan Province, on February 18, 2018. The experimental results show that the proposed model has a good prediction effect.

1. Introduction
China is located at the junction of the Pacific Rim Seismic Zone and the Mediterranean Himalayan Seismic Zone. It is a country with strong and frequent seismic activity. According to the statistics of China Seismic Network Center1, since 2012, there have been more than 270 earthquakes with m ≥ 5 in China. For example, the 1976 Tangshan earthquake (m=7.8), the 2008 Wenchuan earthquake (m=8.0), the 2010 Yushu Earthquake (m=7.1), the 2013 Lushan earthquake (m=7.0) and the 2017 Nine-village Valley earthquake (m=7.0) occurred successively in China. And the 1960 Chilean earthquake (m=9.5), the 2004 Indonesia Sumatra earthquake (m=9.0), the 2011 East Japan earthquake (m=9.0) are also occurred internationally. These destructive earthquakes caused a large number of casualties and huge property losses [1].

Earthquakes are the result of tectonic activity on the earth. It has a certain randomness. It may also be affected by some unknown factors, which makes the current earthquake prediction work still in a low-level exploration stage [2]. The train of thought for earthquake prediction can be divided into three categories: seismogeology [3], earthquake statistics [4][5] and earthquake precursors [6][7]. The research of earthquake prediction based on precursory signals is a hot research direction in modern seismology, which mainly predicts earthquake events by studying the anomalies of precursory signals such as crustal deformation [8], electromagnetic [10], gravity field change [9] and geosound [11].

1 http://news.ceic.ac.cn/index.html
In order to carry out the prediction research based on earthquake precursory signals, AETA (Acoustic and Electromagnetic Testing All in one system) is developed by Earthquake Monitoring and Prediction Technology Research Center. Electromagnetic disturbance and geosound signal are selected as research objects in AETA [12][13]. Through this system, researchers can collect, process and analyze the electromagnetic disturbance and geosound signals in real time.

An earthquake prediction model based on precursor window detection is proposed in this paper. In this model, the precursory anomaly interval of low-frequency electromagnetic signal data is used as a class marker for learning, so as to predict the time of earthquake occurrence. Then, combined with the prediction results of multiple stations within the effective distance, the prediction probability is used as the weight to locate the earthquake position. Finally, this model is verified on the real data set in order to obtain a better effect of impending earthquake prediction.

2. Related works

The short-term and imminent earthquake prediction method is based on the precursory changes observed before the earthquake, or in other words, suspected precursory changes. At present, there are many researches on the changes of geomagnetic field, gravity field, crustal deformation, electromagnetic disturbance, geosound, underground resistivity, geoelectric field and so on.

The research methods based on geomagnetic field variation include geomagnetic low-point displacement anomaly [14] and daily variation ratio method [15]. By comparing the geomagnetic field vertical sub-daily fluctuations, if the low-point time of a magnetic station in the region and the low-point time of the adjacent region are more than two hours apart, there will be an obvious abrupt boundary on the map, which is called low-point displacement boundary [16]. The geomagnetic low-point displacement boundary before the 2011 Kunlun Mountain M8 earthquake is described in [16], which is used to predict earthquakes with m=5 or more.

The correlation between hydrogen isotope variations in groundwater near the Longmenshan active fault zone and earthquakes is studied in [17], and it is proposed that the anomaly of hydrogen isotope δD value in groundwater can reflect the seismicity of active fault zone, which can be used as a method for impending prediction of earthquakes with magnitude over 5.0[17]. After the Wenchuan earthquake in 2008, China Earthquake Administration paid much attention to the change of gravity field, and made some achievements in the medium-term earthquake prediction[18]. Crustal deformation has a certain effect in the medium-term and long-term earthquake prediction, but the effect of impending earthquake prediction is poor [8].

In a word, the prediction method based on precursory signal compares the predicted value with the true value, and predicts earthquake by anomaly. However, the precursory signal of earthquake is easily disturbed, so the prediction accuracy needs to be improved.

3. Earthquake prediction method based on precursory window

3.1. Introduction

The method uses low frequency electromagnetic signals to predict the time and position of earthquakes with corresponding magnitudes. The prediction method is divided into two stages. The first stage is to predict the earthquake time. According to the data of low-frequency electromagnetic signals collected by stations, the prediction model of precursory time window is used to determine whether there will be a potential earthquake of corresponding magnitude in a given period of time. If there is a possibility of an earthquake, it will enter the second stage. The second stage predicts earthquake position. According to the prediction of the same potential earthquake event by the stations distributed near the stations, the earthquake probability value predicted by multiple points is used to predict the position of the earthquake. A framework diagram of the prediction system is shown in figure 1.
The sensor data is collected from each station of AETA in real time. The sensor data processor preprocesses the sensor data, including missing data processing, noise reduction, and standardization processing. The output data is split into two copies. A copy is entered into the precursor window detector. The other one is input to the learning data manager. Learning data manager is mainly used to generate, save training data and archive the prediction results data. It provides prediction results and corresponding real situation for online model tuning. The on-line model tuning generates prediction models for the precursor window detector and performs model tuning periodically according to the difference between the prediction results and the corresponding real situation. The precursor window detector is used to predict the data uploaded by the sensor data processor, and the position detector is carried out once the anomaly is found. The position detector uses two different algorithms based on single-site and group-site to predict the position of the earthquake. Due to the limitation of space, this paper mainly introduces the core algorithms of the prediction of precursory time window and position.

3.2. Precursory time window prediction

The precursory time window of earthquake is a period of anomaly before the occurrence of seismic event, which can be shown as the anomaly of the data collected by sensors. By marking the history sensor data according to the time interval, the classification problem can be modeled as a classification problem, which can be divided into ordinary and precursory States. By constructing the prediction model, we can get the prediction of the earthquake. The specific description is as follows.

In this method, $d$ is represented as the number of days, $s$ is the start date (the time distance through January 1, 1900), and $I_{s,d} = [s, s + d)$ represents a date range. Let function $m(\cdot) : Z^+ \rightarrow R^d$, where $S \mapsto (x_s, x_{s+1}, \ldots, x_{s+d-1})$, then function $m$ means the mapping from the set of positive integers to $d$-dimensional real vector set, while $x_s$ means the average daily geomagnetic signal value of day $s$. Then $m(s)$ means the average daily geomagnetic signal value sequence vector of interval $[s, s+d)$.

Assume the earthquake precursory interval $I'$, given any interval $I_{s,d}$, suppose the probability event $A = I_{s,d} \cap I' \neq \emptyset$, the problem of earthquake precursory anomaly time window detection is equivalent to $P(A | m(s))$.

Set the sample set $M = \{(m(s_1), y_1), (m(s_2), y_2), \ldots, (m(s_n), y_n)\}$, where $Y_i$ belongs to $\{0 , 1\}$, $Y_i$ represents the class label corresponding to the $i$-th sample sequence while the value represents coincidence part between the sample and the precursor region. According to the need, the coincidence events can be further subdivided into two cases: the sample intersects with the precursory interval and the sample is the precursory interval. Because the algorithm is very easy to extend to the case of multi-probability
events, for the convenience of description, only \{0, 1\} is discussed in this paper, which means coincidence or not.

This paper uses logistic regression method to construct prediction model. Assume that event A follows the Bernoulli distribution, and that the probability of A occurring is \( p(A) = h_\theta(s; \theta) \). Through the sigmoid function, the function values are mapped between \([0,1]\), and the function is obtained as shown in Formula 1.

\[
h_\theta(s; \theta) = \frac{1}{1 + e^{-\theta m(s)}}
\]  

(1)

Next, we use the sample set to estimate the cost function, so that the cost function is an extreme value.

We use cross-entropy as the loss function, and the cost function is shown by Formula 2.

\[
J(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \left( y^{(i)} \log h_\theta(s^{(i)}) + (1 - y^{(i)}) \log (1 - h_\theta(s^{(i)})) \right)
\]

(2)

It is easy to prove that the function of Formula 2 is convex function. The gradient descent method can be used to solve it, and finally a group of optimal fitting coefficients can be obtained, which determines the boundary of the occurrence of A event.

3.3. Position prediction method

When there are sensors in the earthquake precursor time window, the position detector will activate and predict the position. We designed two different working modes based on single-site and group-site. Let \( i (1 \leq i \leq N) \) as the \( i \)-th detector, \( O_i \) as the position of \( i \)-th detector, \( p_{[t,t+\Delta t]} \) as the probability of \( i \)-th detector in the precursory position in the time window \([t, t + \Delta t]\), \( r_i \) as the maximum effective radius of \( i \)-th detector, and \( \text{dist}(O_i, O_j) \) as the distance between detector \( i \) and \( j \).

3.3.1. Single-site prediction algorithm. Firstly, the algorithm finds a detector \( i \) with the largest earthquake occurrence probability (the probability value is greater than 0.5) from the set of stations. Then it finds two detectors \( j \) and \( k \) which are nearest to the station \( i \) and whose precursory probability is greater than 0.5. So a weighted positioning center point \( x \) could be calculate by \( \text{tri locate}(i, j, k) \). The product of the distance between the center point and the station and the predicted earthquake probability is the minimum, which is \( \arg \min_{\tilde{x}} \sum_{m \in [i,j,k]} \text{dist}(\tilde{x}, O_m) p_{m[t,t+\Delta t]} \).

Algorithm 1: Single-site prediction algorithm

1: let \( p_{[t,t+\Delta t]} = 0.5 \), let \( \text{dist}(O_0, O_i) = \max_{1 \leq i \leq N} \{ \text{dist}(O_i, O_j) \} \)
2: repeat (1)
3: wait \( \Delta t \);
4: find \( i \) \( \left\{ \text{argmax}_{i} p_{[t,t+\Delta t]} \right\} \) such that \( i \in [0, N] \);
5: if (\( i = 0 \)) then continue;
6: end if
7: find \( j \) \( \left\{ \text{argmin}_{j} \text{dist}(O_i, O_j) \right\} \) such that \( j \in [0, N], j \neq i, p_{j[t,t+\Delta t]} \geq 0.5 \);
8: find \( k \) \( \left\{ \text{argmin}_{k} \text{dist}(O_i, O_k) \right\} \) such that \( k \in [0, N], k \neq i, k \neq j, p_{k[t,t+\Delta t]} \geq 0.5 \);
9: if (\( i \neq 0 \) \& \( k \neq 0 \)) then
10: \( r = \text{tri locate}(i, j, k) \);
11: output \( r \);
12: end if
13: \( t = t + \Delta t \);
14: end repeat
3.3.2. Group-site prediction algorithm. Compared with the single-site position prediction method, the group-site position prediction method adopts a delayed prediction strategy. The earthquake position prediction is activated when the probability of the prediction of the precursory time window of more than three stations exceeds 0.5. Therefore, the single-site position prediction is a special case of this algorithm. The algorithm is shown as algorithm 2.

Algorithm 2: Group-site prediction algorithm

1: repeat 
2: wait $\Delta t$;
3: let $S = \{i | p_i(t + \Delta t) > 0.5, 1 \leq i \leq N\}$;
4: if ($|S| < 3$) then continue;
5: end_if
6: let $f(\hat{x}) = \frac{\sum_{i \in S} (1 + s\text{ign}(t_i - \text{dist}(x, O_i)) \cdot \text{dist}(x, O_i)) \cdot p_i(t_{i+1} + \Delta t)}{\sum_{i \in S} (1 + s\text{ign}(t_i - \text{dist}(x, O_i)))}$,
7: find $\hat{x}_{opt}$, where $\arg\min_{\hat{x}} f(\hat{x})$;
8: output $\hat{x}_{opt}$;
9: $t = t + \Delta t$;
10: end_repeat

In algorithm 2, all detectors within the effective distance range are filtered out by $\frac{1 + s\text{ign}(t_i - \text{dist}(x, O_i))}{2}$. When the effective distance is exceeded, the value of 0 and filtered out from the set. After obtaining all the valid candidate sites, the algorithm calculates the extreme points through the formula of line 6, and then obtains the predicted positions.

4. Experiment

The support vector machine (SVM) and logistic regression (LR) are used as classification models. The prediction model of precursory time window is learned according to the algorithm in Section 3.2. Then, the position prediction algorithm is constructed according to the algorithm in Section 3.3.2. Finally, the earthquake of Qingchuan County, Guangyuan City, Sichuan Province on February 18, 2018 is used as a model to verify the validity of the method in this paper.

4.1. Experimental Environment

The experimental code is implemented in Python, and mainly using the third-party machine learning library scikit-learn, version 0.20.3. The experiment was run on Win10 enterprise edition operating system with Intel i7-6700 CPU and 16G RAM.

4.2. Experimental Data Set

The original data set of the experiment comes from the electromagnetic low frequency peak and average frequency monitoring values of the real seismic records of AETA system. The data were recorded by different monitoring stations at the same time (February 18, 2018) and at the same place (Qingchuan County, Guangyuan City, Sichuan Province, longitude and latitude (105.02, 32.29)), with an earthquake magnitude of 4.4. The specific raw data information is shown in table 1. The historical data of low frequency peak and average frequency in 3 months before the seismic events at each platform are selected as the data set.

| Seq. | Station(No.)                                    | Longitude | Latitude | Data Amount |
|------|------------------------------------------------|-----------|----------|-------------|
| 1    | Earthquake Prevention and Disaster Mitigation Bureau of Pingwu County (116) | 104.55    | 32.41    | 186         |
| 2    | Earthquake Prevention and Disaster Mitigation Bureau of Nine-village Valley (121) | 104.25    | 33.26    | 186         |
| 3    | Qingchuan County Yaodu Observation Station (141) | 105.42    | 32.78    | 186         |
4.3. Precursory period prediction

The seismic interval data of each station is regarded as a sample set according to the time window interval. Then the data set is divided into training set and testing set according to the ratio of 7:3. The training set is used for model training, and the test set is used for model quality verification. The goal of the experiment is to judge whether the target time window is in the earthquake precursor period, that is, to judge whether the target time window is in the earthquake precursor interval. In the experiment, we use SVM and LR classifier to compare the classification results.

The parameter C of SVM is the objective function, and the penalty coefficient is set to 1.0, which is used to balance the classification interval and misclassification. The kernel function is "rbf", which means radial basis function. The function coefficient gamma is 0.2. The multiple loss parameter solver of LR is set to "lbfgs".

Accuracy and F1-SCORE which are defined in Formula 3 and 4 are used to evaluate the quality of the model. We use micro-average because the model is a multi-class. The total precision and Recall of all classes are calculated first. Then this is used to calculate F-SCORE, that is, Micro-F1.

\[
\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}
\]

\[
F1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

\[
\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{TP+FN}.
\]

The model prediction quality on the test set is shown in Table 2.

| Interval   | SVM Accuracy | F1 | LSTM Accuracy | F1 |
|------------|--------------|----|---------------|----|
| 10 (5 days) | 0.75         | 0.75 | 0.74          | 0.74 |
| 12 (6 days) | 0.76         | 0.76 | 0.75          | 0.75 |
| 14 (7 days) | 0.77         | 0.77 | 0.74          | 0.74 |

4.4. Earthquake position prediction

On the basis of the prediction model of precursory period, we predict the place where the earthquake may occur. The data of stations 116, 121, 141 and 150 were used to train the prediction model of the precursory period. The LR classifier is used in the premonitory model. We then use data from stations 241, 169, 167, 183 to test the validity of our algorithm for predicting the position of an earthquake.

Firstly, we use algorithm 1 for position prediction. When the data of one station is found to be abnormal (the probability is greater than the threshold value 0.5), two other stations with similar distance and abnormal data are selected. The longitude and latitude of the three anomalous stations are mapped to a plane coordinate system by using Miller projection method. Then, based on the probability weight, the longitude and latitude are calculated on the basis of the inverse process of Miller projection method. The forecast results are shown in Table 3.

| Interval   | Predicted Longitude | Predicted Latitude | Forecast Center Deviation (km) |
|------------|---------------------|--------------------|-------------------------------|
| 10 (5 days)| 104.91              | 33.07              | 67.4                          |
| 12 (6 days)| 104.77              | 32.97              | 64.2                          |
| 14 (7 days)| 104.75              | 32.96              | 64.1                          |
Then, the prediction is carried out based on the group station algorithm 2. If we have more than one station, the minimum value of the distance sum between them is calculated by using the probability as the weight within the allowable range of the station. When more than three stations are found to have abnormal data, we use Miller projection method to convert the longitude and latitude of these locations into plane coordinates. By calculating the station distance to each effective range, the minimum value of the distance sum is optimized. The forecast results are shown in table 4.

| Interval  | Predicted Longitude | Predicted latitude | Forecast Center Deviation (km) |
|-----------|---------------------|--------------------|--------------------------------|
| 10 (5 days) | 104.46              | 32.36              | 62.8                           |
| 12 (6 days)  | 104.45              | 32.37              | 64.1                           |
| 14 (7 days)  | 104.69              | 32.35              | 36.9                           |

Table 4. Experimental results of group-site algorithm prediction

The prediction data is very close to the real position of the earthquake (105.02, 32.29), which shows that the algorithm is effective.

5. Summary and Future Works
An earthquake prediction method based on time precursory window is proposed in this paper, which is based on the low-frequency electromagnetic signals collected by AETA. Firstly, temporal prediction of earthquakes can be transformed into the question of whether a given time interval is in a precursory window period of an earthquake. Taking the precursory anomaly interval of low-frequency electromagnetic signal data as the class mark and the historical real data as the example, the problem is modeled as a classification problem. Through learning the training set of seismic signals, we can obtain the abnormal situation judgment of prediction data.

Secondly, this paper presents position prediction algorithms based on single-site and group-site to predict the position of the earthquake. By selecting three or more stations within the effective distance range, the probability of earthquake occurrence can be predicted. The probability of earthquake occurrence is used as the weight to locate the earthquake center. The longitude and latitude of these places are converted into the coordinates of plane coordinate system by Miller projection method, so as to calculate the coordinates of the place where the earthquake occurs.

Finally, the real data set is tested on the earthquake of Qingchuan County, Guangyuan City, Sichuan Province, on February 18, 2018. The experimental results show that the minimum deviation is 36.9 km from the earthquake center, which means the proposed model has a good prediction effect.

The results of earthquake prediction can be further optimized, and we think there may be two problems. One is the accumulation of errors in the prediction model of the precursor period, which leads to the errors in the prediction of the position. The other one is the error caused by the transformation between longitude and latitude and plane coordinate system using Miller projection method. In the future work, we can continue to improve and optimize the algorithm proposed in this paper. In addition, we need to collect more valid data to further study and verify the effectiveness of the model and algorithm.

Acknowledgments
This work is supported by the fundamental research funds of Shenzhen Science and Technology Innovation Committee (No. JCYJ20180503182125190), Shenzhen Science and Technology Application Demonstration Project (No. KJYY201707211151955849) and the Fundamental Research Funds for the Central Universities, South-Central University for Nationalities (No. CZY19014).

References
[1] Xu W, Liu J, Xu G, et al. Earthquake disasters in China[M]. Natural disasters in China. Springer, Berlin, Heidelberg, 2016: 37-72.
[2] Zhang J, Wu X, Yang X, et al. Observational evidence of anisotropic changes of apparent resistivity before strong earthquakes[J]. Geophysical Journal International, 2017, 210(3): 1323-1331.
[3] Wan Y G, Shen Z K, Shang D, et al. Crustal stress evolution over the past 700 years in North China and earthquake occurrence[J]. Earthquake Research in China, 2006, 20(3): 244-261.

[4] Ogata Y. Statistics of earthquake activity: Models and methods for earthquake predictability studies[J]. Annual Review of Earth and Planetary Sciences, 2017, 45: 497-527.

[5] Matcharashvili T, Hatano T, Chelidze T, et al. Simple statistics for complex Earthquake time distributions[J]. Nonlinear Processes in Geophysics, 2018, 25(3): 497-510.

[6] Lu K, Hou M, Jiang Z, et al. Understanding earthquake from the granular physics point of view- Causes of earthquake, earthquake precursors and predictions[J]. International Journal of Modern Physics B, 2018, 32(07): 1850081.

[7] Huang F, Li M, Ma Y, et al. Studies on earthquake precursors in China: A review for recent 50 years[J]. Geodesy and Geodynamics, 2017, 8(1): 1-12.

[8] Zhang J, Zhu Y, Wu Y, et al. Intermediate to Long-term Estimation of Strong Earthquake Risk Areas in Mainland China based on Geodetic Measurements [J]. Earthquake, 2018, 38(1):1-16. (in chinese)

[9] Xuan S, Shen C, Li H, et al. Characteristics of subsurface density variations before the 4.20 Lushan M S 7.0 earthquake in the Longmenshan area: inversion results[J]. Earthquake Science, 2015, 28(1): 49-57.

[10] Guiping Yuan, Hongyu Li, Guixia Zhang, et al. Daily Variation Ratio of Geomagnetic Z Component and its Relationship with Magnetic Storms and Earthquakes[J]. Earthquake, 2018, 38(1):139-146.

[11] Le Pichon A, Mialle P, Guilbert J, et al. Multistation infrasonic observations of the Chilean earthquake of 2005 June 13[J]. Geophysical Journal International, 2006, 167(2): 838-844.

[12] Wang X, Yong S, Xu B, et al. Research and Implementation of Multi-component Seismic Monitoring System AETA[J]. Beijing Da Xue Xue Bao, 2018, 54(3): 487-494.

[13] Yong S, Wang X, Pang R, et al. Development of inductive magnetic sensor for Multi-component seismic monitoring system AETA[J]. Beijing Da Xue Xue Bao, 2018, 54(3): 495-501.

[14] Huijiu Chang. Analysis On the Causes of Low-Point Displacement Anomaly in Daily Variation of Z Component of Geomagnetic Field[J]. Earthquake, 2011, 31(2):79-86.

[15] Guiping Yuan, Hongyu Li, Guixia Zhang, et al. Daily Variation Ratio of Geomagnetic Z Component and its Relationship with Magnetic Storms and Earthquakes[J]. Earthquake, 2018, 38(1):139-146.

[16] Jianhai Ding, Xuhui Shen, Weiyian Pan, et al. Seisme Electromagnetism Precursor Research Progress[J]. Chinese Journal of Radio Science, 2006, 21(5):791-801.

[17] Zhao Y, Bai J, Li X, et al. Correlation between hydrogen isotope in underground water near active fault and earthquakes[J]. Acta Petrologica Sinica, 2011, 27(6):1909-1915. (in chinese)

[18] Zhu Y, Shen C, Zhang G, et al. Rethinking the Development of Earthquake Monitoring and Prediction in Mobile Gravity[J]. Journal of Geodesy and Geodynamics, 2018, 38(5):441-446. (in chinese)