

Report: Earthquakes precursor based on RTL algorithm with Machine Learning baseline

Braslavskiy

February 2, 2018

1 Introduction

At the moment, there is no way to build an accurate earthquake precursor because of the complex nonlinear behavior of seismicity. But since the 1960s our knowledge of the mechanism of earthquakes is constantly growing. Using data on large earthquakes, scientists managed to obtain several important empirical statistical relationships.

Let us note some of the most important relations. Gutenberg–Richter law [3] expresses the relationship between the magnitude and total number of earthquakes as follows

$$\log N = a - bM, \quad (1)$$

where N is the number of events having a magnitude greater than M , a and b (commonly referred to as the **b – value**) are fitting coefficients.

Omori–Utsu (O–U) law [4] represents the decay of aftershock activity with time

$$\dot{N}(t) = \frac{C_1}{(C_2 + t)^p}, \quad (2)$$

where t is time, N is earthquake count, C_1 , C_2 and decay exponent p (commonly referred to as the **p – value**) are fitting coefficients.

Recently, the approach using seismic data in combination with machine learning methods has become more popular. The main problem which appears in this approach is the data mining from earthquake catalogs. Authors of [5] considered the prediction of earthquakes as a problem of binary classification. They generated 51 meaningful seismic features calculated for our

dataset based on well-known seismic facts such as "Standard Deviation of b-value" or "Time (T) of n events". As models were used various ensemble methods such as Random Forest, Rotation Forest and RotBoost.

2 Description of the problem

Let t is the index of time. Y_t is a time series, where Y_i is a target event. We also observe features $X_{t,k}$. Our proposal is to predict the target event Y using feature description X .

So, We have a history $X_{t,k}, Y_t$ up to the time T , $t \in [0, T]$. We have to raise the alarm in the window $[T, T_{min}]$, in this case, the target event occurs in the window $[T + T_{min}, T + T_{max}]$.

Consider examples of such problems:

- Prediction of breakages in complex technical systems. Then k indexes the sensors in a system and $X_{t,k}$ is data from sensors.
- Prediction of natural disasters. In this case k can represent the coordinates (*latitude; longitude*) or a cluster number.

In this paper, we focus on the prediction of earthquakes. It is necessary to determine the type of the target event. There are several options:

- The prediction of the beginning of a series of aftershocks, i.e. sequence of earthquakes after the main shock. (**ETAS** models)
- Prediction of strong earthquakes in the long-term horizon (years). (**seismic gap** models)
- Prediction of strong earthquakes in the middle-term horizon (months).

We study the prediction of strong earthquakes in the middle-term horizon. Strong earthquake is the earthquake with the magnitude higher than $M_c = 5$. Predictions of earthquakes are related with the difficulties:

- The sample is very unbalanced. In Japan, from 1990 to 2016, there were 247,204 earthquakes. Consider a distribution by the magnitudes:

Magnitude	The number of earthquakes with greater magnitude
5	2346 (0.95%)
6	340 (0.14%)
7	37 (0.015%)
8	2

- There are artificial anomalies in the catalog due to changes in the network of seismic stations.
- There is a lag in time between available and desired prediction interval(?)

3 Introduction

4 RTL Precursor

The basic assumption of the **Region – Time – Length(RTL)** algorithm [1] is that the influence weight of each prior event on the main event under investigation may be quantified in the form of a weight. Weights become larger when an earthquake is larger in magnitude or is closer to the investigated place or time. Thus, **RTL** characterizes the level of seismicity in the point of space in the certain time.

The **RTL** takes into account weighted quantities associated with three parameters (time, place and magnitude) of earthquakes. A **RTL** parameter is defined as a product of the following three functions

$$\text{RTL}(x, y, z, t) = \text{R}(x, y, z, t) \cdot \text{T}(x, y, z, t) \cdot \text{L}(x, y, z, t), \quad (3)$$

where $\text{R}(x, y, z, t)$ is an epicentral distance, $\text{T}(x, y, z, t)$ is a time distance and $\text{L}(x, y, z, t)$ is a rupture length. They depends on the size of the space-time cylinder \mathcal{E} . Consideration defined by radius r_0 and time length t_0

$$\begin{aligned}
R(x, y, z, t) &= \left[\sum_{i \in \mathcal{E}} \exp \left(-\frac{r_i}{r_0} \right) \right], \\
T(x, y, z, t) &= \left[\sum_{i \in \mathcal{E}} \exp \left(-\frac{t - t_i}{t_0} \right) \right], \\
L(x, y, z, t) &= \left[\sum_{i \in \mathcal{E}} \left(\frac{l_i}{r_0} \right) \right],
\end{aligned} \tag{4}$$

For l_i we use an empirical relationship for Japan $\log l_i = 0.5M_i - 1.8$. We use only the earthquakes with magnitude at least M_0

RTL is very unstable statistics. Therefore in the article [2] author proposed to normalize the parameters on the variances. Also we can subtract the moving average. Thus a negative **RTL** means a lower seismicity compared to the background rate around the investigated place, and a positive **RTL** represents a higher seismicity compared to the background. We are interested in both types of anomalies.

4.1 Determination of critical values

There are several strategies for determining anomalous values:

- One-class SVM trained for a fixed interval $[0, T]$,
- Critical quantiles of model tails $RTL_t = f(RTL_{t-1}, \dots, RTL_{t-k}) + \epsilon_t$,
- Empirical histogram quantiles.

5 Description of the model

To solve classification problem we need to create a matrix of input features \mathbf{X} and target labels \mathbf{Y}

So, we want to make labels of the target events Y for features X^j , each of which is indexed in space-time by vector (x, y, t) . Thus, we will create a labels for the indices (x, y, t) .

The target earthquakes are everything that falls into the space-time cylinder, i.e. for the statistics calculated for the index (x, y, t) the value 1 is

assigned if there is at least one earthquake of magnitude $M > Mc$ with coordinates (x_e, y_e, t_e) satisfying the following constraints:

$$\|(x, y) - (x_e, y_e)\|_2 \leq R_c, \quad \delta_c < t_e - t < T_c \quad (5)$$

Thus, we can determine the optimal working conditions for the predictive model by choosing the hyperparameters (R_c, δ_c, T_c) of the labels. Also, it should be noted that for a **RTL** precursor, We can obtain a series of precursors by specifying various hyperparameters $(r_0$ and $t_0)$.

Let's describe our approach to earthquake prediction:

- Make a grid of hyperparameters (r_0^i, t_0^i) . For each set of parameters counts RTL statistics for each earthquake
- Make a grid of hyperparameters $(M^i, R_c^i, \delta_c^i, T_c^i)$. For each parameter set, form the labels of the target events
- Select the anomalous values of the statistics, accordingly there is obtained the binary matrix of feature **X**
- Take the optimal for the task label as the target variable **Y**
- Solve the problem of binary classification for the sample (\mathbf{X}, \mathbf{Y})

6 Classifiers and quality metrics

6.1 Classifiers

As classifiers we used the following machine learning methods:

- **Logistic regression.** This is a statistical model used to predict the probability of occurrence of an event. The model is based on the following assumption:

$$Pr(y = 1|x) = f(\theta^T x), \quad (6)$$

where $f(z) = \frac{1}{1+e^{-z}}$

- **Random Forest.** This is an ensemble classifier that is developed on the basis of majority voting of decision trees. Various number of decision trees are generated over bootstrap samples of the training dataset. The final decision is made by aggregating the predictions obtained by

all the decision trees. Thus, a Random Forest allows to find complicated relationships in the data, but at the same time more resistant to retraining.

7 Numerical results

7.1 Data

There are earthquakes in the vicinity of Japan from 1990 to 2002. Target earthquakes are earthquakes with a magnitude greater than 5. The training sample contained 576 target events and 99423 earthquakes of small magnitude.

7.2 Hyperparameters and quality metrics

the following grid was created for **RTL**:

$$\mathbf{r_0} = [10, 25, 50, 100], \mathbf{t_0} = [30, 90, 180, 365]$$

Thus, 16 features were generated. The label parameters for which the best results were previously obtained were the following:

$$M_c = 50, R_c = 50, \delta_c = 10, T_c = 180$$

Precision, **Recall** and **F1** were selected as quality metrics.

7.3 Result

Algorithm	Precision	Recall	F1
choice by "OR"	0.613	0.497	0.548
The Best RTL	0.786	0.453	0.574
Logistic Regression	0.812	0.572	0.671
Major vote	0.889	0.425	0.575
XGBoost	0.941	0.632	0.756
Random Forest	0.945	0.690	0.796

References

- [1] Sobolev G. A. and Y. S. Tyupkin 1997: Low-seismicity precursors of large earthquakes in Kamchatka. Volc. Seismol., 18, 433-446
- [2] Huang Q. Seismicity pattern changes prior to large earthquakes-An approach of the RTL algorithm //TERRESTRIAL ATMOSPHERIC AND OCEANIC SCIENCES. – 2004. – T. 15. – №. 3. – C. 469-492.
- [3] B. Gutenberg and C. Richter, Seismicity of the earth and associated phenomena: Princeton University Press, 1954.
- [4] 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.
- [5] K. M. Asima, A. Idrisb, F. Martínez-Álvarezc, T. Iqbala, Short Term Earthquake Prediction in Hindukush Region using Tree based Ensemble Learning, 2016 International Conference on Frontiers of Information Technology