

Investigation of Ionospheric Disturbance and Seismic Events Based on Machine Learning

Yanchen Long, Qinghui Zhang, and Zhengquan Dai, Jian Rong*

College of the Big Data and Intelligent Engineering, Southwest Forestry University, Kunming, China

Institute of the Big Data and Intelligent Engineering, Southwest Forestry University, Kunming, China

Correspondence: swordrong@163.com, swordrong@swfu.edu.cn

Abstract. Earthquakes are natural disasters that endanger human life and cause the greatest loss of property. The study of anomalous disturbance of the ionosphere, one of the pre-earthquake anomalies, will help to further study the coupling effect on the ionosphere before the earthquake. This paper focuses on the analysis of the importance and influence of various parameters inside the seismic ionosphere under earthquake conditions, constructs a classification and prediction model of seismic ionospheric anomalies based on the gradient boosting decision tree GBDT algorithm, and analyzes the pre-earthquake ionospheric data. , the results show that NmE, nHe+, foF2 and TEC are important influencing factors, which lays a certain foundation for the further study of the internal structure and parameters of the pre-earthquake ionosphere

Keywords: Earthquake, Lonosphere, Deep Learning, GBDT.

1. Introduction

During and before earthquake events, various disturbance may be observed on satellite such as seismo-ionospheric coupling. In 1964 Alaska earthquake, ionospheric anomalies are reported associated with seismic events. After that, more and more seismo-ionospheric disturbance was reported with parameters such as the total electron content (TEC), the peak density and height of the F2 layer (NmF2 and hmF2), the magnetic and electric field and so fotrth. Some cases study focused on rapid disturbance found that amplitude disturbance was positive offset and some was negative offset, also some was not significant. In 2019 Ridgecrest (California) Earthquakes, the TEC anomalies were all positively correlated with the electron density (Ne) anomalies, and most of them were positive. The Ne and electron temperature (Te) anomalies were both positively and negatively correlated with mostly negative, while showing a strong linear negative correlation at nighttime. Earthquake pregnant is a complex process with many hypotheses including a chemical model, an acoustic model, a thermal model and an electromagnetic radiation model. The energy accumulates to a certain extent then explodes suddenly which leads to earthquake. Earthquake preparation is a long-term process, there should also be corresponding long-period changes in the ionosphere. The occurrence time of disturbance varies in many case study. A duration before earthquake, the ionospheric anomalies were demonstrated in many cases.

2. Data Preparation

This study has investigated 506 earthquakes during 2001 to 2020 with $M_s_level \geq 5.0$ occurred in China mainland land area. The parameters focus on Ne, Neutral Temperature (Tn), Ion Temperature (Ti), Te, Nitric Oxide ions (nO+), Atomic Hydrogen (nH+), Atomic Helium (nHe+), nCl, Atomic Nitrogen (nN+), Density of F2 peak (NmF2), Height of F2 peak (hmF2), Density of E peak (NmE), TEC and F2 plasma frequency (foF2) before 30 days at the height of 400km. In this study, the long term relative variation of the observation data with respect to the reference background is studied. The reference background is defined as mean value of 30 days cubic smoothing. The data is collected from web of International Reference Ionosphere(<http://www.irimodel.org/>)[1]. Some

days before earthquake, the class is labeled as 1 and else 0 for classifying by deep training model and decision forest[2]. The imbalance data is processed with Synthetic Minority Oversampling Technique (SMOTE) for increasing the number of cases in my dataset in a balanced way. The location of earthquakes is shown as figure 1.



Figure 1. Distribution of seismic detection systems in China

The epicenter of earthquakes in China

3. Deep Learning Classification on Parameters

Seismo-ionospheric phenomena have been frequently reported in recent years. The rapid ionosphere disturbance observed before an earthquake is considered to be related to the earthquake preparation[3]. The statics on seismo-ionospheric abnormality varies in different site, and it is difficult to summary the characteristics. Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to “learn” from large amounts of data[4]. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy. Also, deep learning drives much artificial intelligence (AI) applications and services that improve automation, performing analytical and physical tasks without human intervention. In this study, a deep learning training model is established based on the 'DenseBlock' layer and the 'ReLU' activation function with hyperparameter tuning. In this model, the parameters are normalized with $y = \left| \frac{(x-\bar{x})}{\max(x)-\min(x)} \right|$, where x minus the mean value of x, the statistical properties of x has not changed and eliminated dimensional effects between metrics, also the positive offset and negative offset are all unified into positive offset[5]. The class of 7 days before earthquake to outbreak time is labeled as 1 and else 0 by imbalance processing with SMOTE. The train dataset include 506 earthquakes occurred in China mainland with validation split is 0.3. The training model is shown as figure [fig2].

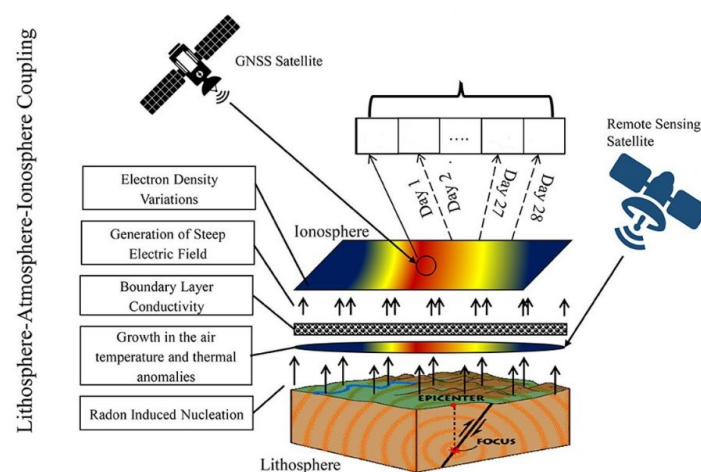


Figure 2. The training Process on seismo-ionospheric parameters

After training about 400 epochs, the accuracy gets reach about 70% and the loss decreases to about 0.56 without overfitting. The train accuracy and the loss are shown as figure [fig4a] and figure [fig4b] respectively[6].

4. GBDT Trees on Parameters

The accuracy of deep training gets reached about 70%, but it can not discover the importance of parameters. Some earthquakes have two or more parameters anomalies, but others have one or less. The parameters of rapid ionosphere disturbance may focus on one or more parameters, and also it is different for each earthquake. Simple parameter disturbance relationship are not enough to describe earthquake process.

Gradient Boosting Decision Tree (GBDT) is made for evaluating the impact of parameters. Validation data is split 30 percent from all 506 earthquake randomly. The input parameters are n_e , T_n , T_i , T_e , nO^+ , nH^+ , nHe^+ , nCl , nN^+ , $NmF2$, $hmF2$, NmE , TEC and $foF2$ according to the heatmap of grey relation analysis shown . The parameters named TEC , n_e , nO^+ have same feature, and they can be seen as one. The smaller relationship parameters can be used for GBDT trees shown as Figure3.

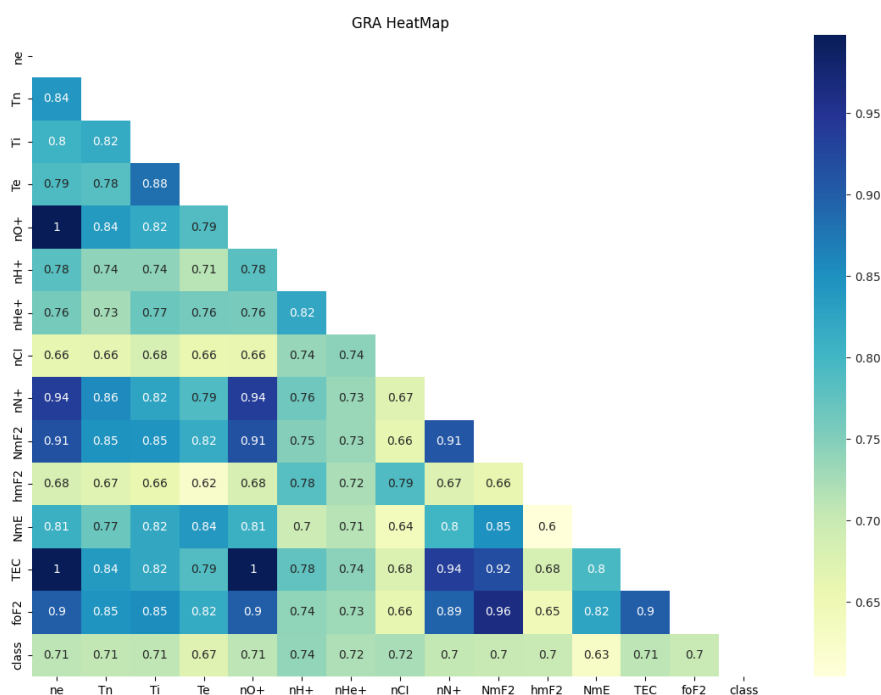


Figure 3. The heatmap of seismo-ionospheric parameters

The validation accuracy of the output GBDT tree get reach 70%. The green path denotes the condition is true, and the red path denotes the condition is false. See Figure 4.

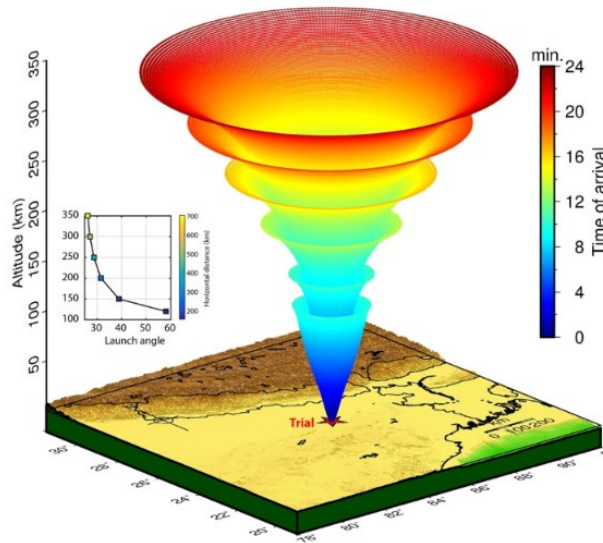


Figure 4. Computer modeling analysis research

The GBDT trees of seismo-ionospheric parameters

According to the number as root, the variable importance is shown in table 1.

Variable Importance

Table 1 NUM_AS_ROOT

No	name	score
1	nHe+	61
2	NmE	48
3	hmF2	40
4	foF2	32
5	nH+	32
6	Tn	25
7	Ti	18
8	nH+	12
9	Te	11
10	TEC	9
11	nO+	5
12	NmF2	4
13	nCI	3

Table 2. NUM_NODES

No	name	score
1	NmE	1331

Table 2. NUM_NODES

2	nHe+	1021
3	Ti	811
4	hmF2	754
5	Tn	740
6	Te	737
7	nN+	714
8	foF2	682
9	nN+	669
10	NmF2	509
11	TEC	268
12	nO+	258
13	nCI	141
14	ne	141

Variable Importance

Table 3. SUM_SCORE

No	name	score
1	NmE	113550.26
2	nHe+	56181.96
3	Ti	49153.52
4	foF2	35659.48
5	hmF2	32484.13
6	Te	32456.67
7	nH+	32133.64
8	Tn	30810.95
9	nN+	29449.98
10	NmF2	20847.36
11	TEC	16090.90
12	nO+	11984.25
13	ne	5886.00
14	nCI	4634.66

From the table1 and the GDBT trees, the parameters with high score focus on the NmE, nHe+, foF2, TEC and so forth according to the score of 'NUM_ASROOT', 'NUM_NODES' or 'SUM_SCORE'. Besides, there are many other parameters along the tree's leaves. The accuracy of GDBT trees gets reached about 70% same as that of deep training model, which says the earthquake

is really related with the ionospheric disturbance, and it can predict or classify the earthquake occurrence from ionospheric disturbance to some extent. This view is consistent with the case study that the abnormality can be found in some site but not all. The number of data on the path from $nHe+$ to TEC is much larger than the number to foF2, which means the TEC disturbance is more significant than foF2. On the leaf from $nHe+$ to foF2, the path with $nHe+ \rightarrow foF2 \rightarrow Ti \rightarrow \dots$ can be considered as the main path. On this path, the parameters Tn, Te and hmF2 affect the leaves on the decision tree. On the leaf from $nHe+$ to TEC, the path with $nHe+ \rightarrow TEC \rightarrow nHe+ \rightarrow \dots$ can be considered as the main path. On this main path, in addition to $nHe+$ and TEC, there are NmE and nH+ on the leaves. The number of data on the leaves $nHe+$ is much larger than that of the others and the root node is $nHe+$ means that the $nHe+$ is the most significant parameter. The number of data on the node TEC is much larger than on the node foF2, which means the TEC is much significant than foF2. The NmE and other parameters are the child node of the $nHe+$, TEC and foF2. So significance can be ordered as $nHe+ > TEC > foF2 > NmE \dots$ according to the main paths.

5. Discussion

The accuracy of the deep training model and the GDBT trees get reach about 70%, which means to some extent they can be used as the precursors of the seismic event from the present seismo-ionospheric parameter but unreliable although parameters such as geomagnetism or ozone have been added in many studies. The result is consistent with the fact in some case study the ionospheric abnormality are significant, but others are not obvious. The GDBT trees show the NmE, $nHe+$, foF2 and TEC are the important factor, so the TEC and foF2 disturbance is studied in many case study is reasonable, but the rules are hard to figure out, the deep learning model may be used for analyzing the abnormality. The seismo-ionospheric coupling is a complex process and which component of the disturbance come from the seismic events is unknown, so it is still challenging to detect such precursors in seismic analysis.

6. Conclusion

In this study, 506 earthquakes during 2001 to 2020 is analyzed with 14 parameters collected 30 days before occurrence. The deep training model and the GDBT trees demonstrate the earthquake is really relation with ionospheric disturbance. The GDBT trees computed on seismo-ionospheric parameters show the NmE, $nHe+$, foF2 and TEC are the important factor. There are two main paths from the root to the last node, that is the path from $nHe+ \rightarrow foF2 \rightarrow Ti \rightarrow \dots$ and the path from $nHe+ \rightarrow TEC \rightarrow nHe+ \rightarrow \dots$. The significance can be ordered as $nHe+ > TEC > foF2 > NmE \dots$ according to the main paths.

References

- [1] He Yufei, Yang Dongmei,. Analysis and summary based on the research results of earthquake ionospheric phenomena. China Earthquake. Vol. 02 (2020) No. 11, p.122-126
- [2] Shen Xuhui, Huang Jianping. Research project and research progress of geophysical exploration satellite data analysis and processing technology and earthquake prediction application. Advances in Earthquake Science.Vol. 01 (2022) No. 06, p.75-78
- [3] Zhang Xuemin.Research progress on seismic application of very low frequency/low frequency artificial source radio wave signal.Acta Seismological. Vol. 05 (2021) No. 08, p.221-223
- [4] Chen Tao, Zhang Xiaoxin,Using the regional atmospheric electrostatic field monitoring network to predict earthquake disasters. Journal of Geophysics. Vol. 04 (2021) No. 10, p.225-229
- [5] Zhang Xuemin, Qian Jiadong Progress and future development of electromagnetic satellite seismic applications. Earthquake. Vol. 02 (2020) No. 12, p.36-38