Research on Correlation between Ubiquitous Telecommunication Data and Earthquake Intensity

Wanjiang Han1,*, Lixin Jiang2, Di Zou1, Jiaqi Nie1, Xin Jin1, Feifei Gong1 and Shuwen Chen1

1School of Software Engineering, Beijing University of Posts and Telecommunication, Beijing 100876, China
2Department of Information Technology Support, China Earthquake Networks Center, Beijing 100036, China

*Corresponding author email: hanwanjiang@bupt.edu.cn

Abstract. Based on a large number of literatures and reports, this paper takes communication systems in lifeline engineering as an example, collects as many disaster-related communication data as possible to form a training data set, and analyzes the relationship between disaster-related telecommunications data and earthquake intensity. Then, use visual means to preprocess the collected data set. This paper uses different classification algorithms in machine learning, such as Naive Bayes, K-nearest neighbors, logistic regression and support vector machines for classification training of the formed data set. Optimize it, and finally it proposes a reference model that can be used to predict earthquake intensity. Finally, experiments show that the model has good accuracy.

Keywords: Machine learning; Algorithm optimization; Earthquake intensity prediction; Telecommunications data; Big data.

1. Introduction

With the rapid development of big data, machine learning and other technologies, it has been applied to all parts of life. In the past ten years, earthquake disasters have frequently occurred, and each earthquake disaster will cause major casualties and property losses. It is extremely important to carry out effective rescue and relief work in the golden age. However, the intensity information (such as intensity) of earthquakes in a short period of time cannot be obtained immediately. Therefore, if a model can be constructed to quickly infer earthquake disaster information, it will be reasonable to arrange rescue and relief. Strategies make meaningful contributions. Based on ubiquitous disaster information, this paper studies disaster prediction models and provides decision-making information for emergency response to earthquakes.

2. Data Preprocessing Process

The data pre-processing process of this paper includes data collection, data feature determination, data pre-analysis, Data normalization, etc.

2.1. Data Collection

The data collection phase is the first phase of big data mining analysis technology, and it is also the basic phase of the data analysis process. The accuracy and representativeness of the collected data will
determine the quality of the data mining process. This paper mainly collects data from the following three ways: Internet, Disaster assessment report, Papers.

2.2. Feature Analysis
By selecting different eigenvalues and combining them with the intensity of the earthquake, the correlation analysis is carried out\cite{15}\cite{16}\cite{17}. After a series of analyses, the eigenvalues are determined as follows: Magnitude, Focal depth, Transmission cable break length, Number of base stations returned, Number of collapsed communication poles, Communication equipment damage level, intensity.

3. Data Distribution Analysis
In this section, the entire data set will be graphically described. Different feature values will be individually selected to count the number of occurrences of data items in a certain range. From the general distribution of the data set, the different feature values. The number of occurrences is counted and displayed to understand the overall situation of the data set.

![Figure 1. Data item counts for different seismic intensities](image)

Figure 1 shows the number of data items with different intensities in the entire data set, taking seismic intensity as the statistical item. It can be intuitively seen from the figure that the data items of the study are mainly concentrated in earthquakes with intensity of 6-9, among which the seismic data items with intensity of 6 account for the largest proportion, reaching more than 40 items.

It can be known from the intensity analysis that an earthquake with an intensity of 6 is more obvious and will cause slight damage. The proportions of intensity 7, intensity 8, and intensity 9 are similar, all of which are about 20 items, and their corresponding seismic sensations are more Is strong, and the degree of damage is more serious; the remaining data items are intensity -4, intensity - 5, and intensity -11. The earthquakes of intensity -4 and intensity -5 cause almost no damage, so here only a small part of the data is collected for comparison and explanation, and an earthquake with an intensity of 11 will cause considerable damage, but since the number of earthquakes of such intensity in recent decades is too small, more data items cannot be collected.

4. Modeling Process
This paper uses different classification algorithms in machine learning, such as Naive Bayes, k-nearest neighbors, logistic regression, and support vector machines to perform classification training on the formed data set, and evaluates and compares the prediction accuracy of different methods. Finally, a trusted model is proposed.
4.1. Model Analysis and Evaluation

First, using Gaussian Naive Bayes classifier. The evaluation results is shown in Figure 2, and by adjusting the parameter of the Laplacian correction coefficient of the classifier, it is found that it has no effect on the evaluation result.

![Figure 2. Evaluation results of Gaussian Naive Bayes classifier](image)

Then, the evaluation results of the model by using k nearest neighbors is shown in Figure 3.

![Figure 3. KNN classifier evaluation results](image)

In the process of parameter tuning, the selected changed parameters are neighbors and weights, which have a greater impact. The classification accuracy is compared by different values. When we use the initial parameter 'l' and 'uniform', the accuracy rate of model classification is the lowest. When adjusted to 5 and 'uniform', the accuracy rate of model classification is improved by nearly 0.02.

And then, model training is performed through a logistic regression classifier, after running the program, the training accuracy rate, recall rate, and f1 value of the logistic regression classifier are shown in Figure 4.

![Figure 4. Logistic regression classifier evaluation](image)

From Figure 3, it can be seen that the accuracy of the results of training using the logistic regression classifier reached 0.9186, which is about 0.003 higher than the KNN classifier, but much higher than the Gaussian Naive Bayes classifier, but the recall rate and F1 value are slightly lower KNN classifier.

When the parameters are adjusted to C = 1.2, penalty = 'l1', class_weight = 'balanced', the accuracy rate is 0.9329, which is 0.01 higher than the initial parameter, which further improves the performance.

Through the linear support vector machine classifier model training, the evaluation effect of the model is obtained, as shown in Figure 5.

![Figure 5. Support Vector Machine Classifier Evaluation](image)
It can be seen that its accuracy is as high as 0.9519, which is the most accurate classifier model among the four classification models we have selected. In addition, the recall rate and f1 value have also improved correspondingly, and the effect is even higher than the parameter tuning. After the remaining three models.

### 4.2. Model Optimization

We further optimize the model to explore which feature set training results are the most accurate. Through experiments, we conclude that when only using the magnitude and source depth of the earthquake as input, the classification effect is very poor. It shows that there is no obvious linear relationship between the source depth and the intensity, which is consistent with the conclusions we have drawn during the visual data analysis.

Next, divide all the telecommunication data feature values into a set for training, that is, the number of base stations withdrawn, the number of broken optical cables, the number of collapsed poles, and the level of damage to communication equipment. The classification models are constructed and trained, and the results are shown in Figure 6.

![Figure 6. Fractal training results](image)

### 5. Experiment

After the above analysis, the optimal model parameters and implementation methods are obtained. We use the training set and test set to test the classification prediction results of the model, and divide the entire data set into a 9:1 ratio. Experiments have proved that this model has made more accurate predictions.

### 6. Conclusion

Throughout the entire research process, the data collection and data feature engineering design stages are time-consuming. After determining the complete data set, we chose four different classification algorithm models to train the data set. The above four models are also based on the small size and small attributes of the data set. Under such conditions, it is obviously inappropriate to choose a neural network algorithm or a clustering regression algorithm.

After evaluating the accuracy of each classification algorithm and further parameter adjustments, the optimal classification model, that is, the support vector machine classification model, was selected. Based on this model, a visual system display was constructed, and relevant disaster-related
telecommunication data could be input. After that, a strong and weak information was fed back to judge the severity of the earthquake disaster so that rescue can be carried out in a planned manner.

There are still many areas for improvement in this research. In the future, we will choose a better method for data collection. In addition, we will also study more ubiquitous information as feature input and establish a more accurate prediction model in order to help earthquake relief. Provide more practical decisions for earthquake prediction.

Acknowledgement
This work was supported by the National Key R & D Program. (Grant No. 2018YFC1504502)

References
[1] Shao Dan, Sun Zhe. Statistics of historical earthquake intensity in Mainland China. Seismic and Geomagnetic Observation and Research. 39 (4). 2018, 04: 135-140.
[2] Zhang Yong, Zhao Xuezhi. Preliminary Judgment of Earthquake Impact Based on Communication Big Data. Science & Technology and Innovation. 20.2018, 10: 43-44.
[3] Han Ying, Li Shanshan. Machine learning-based seismic anomaly data mining model. Computer simulation. 31 (11) 2014, 11: 319-322. Li Zhengyuan, Xiong Daohui. Tracking of seismic precursory network observation data based on big data mining Analysis. Seismological and Geomagnetic Observation and Research. 37 (3). 2016, 06: 319-322.
[4] Yang Jianfeng, Qiao Peirui, Li Yongmei. Review of Machine Learning Classification Problems and Algorithms. Statistics and Decision. 6.2019, 03: 36-40.
[5] Guo Xuncheng. Application of Naive Bayes classification algorithm. Communication World. 1.2019, 03: 241-242.
[6] Liu Lizhi, Deng Jieyi, Wu Yuntao. Research on HBase-based Multi-Class Logistic Regression Algorithms. Computer Application Research. 10.2017, 10: 3007-3010.
[7] Pì Y. Research on application of K nearest neighbor classification algorithm. Communication World. 1.2019,01: 286-287.
[8] Sun Lu. Talking about Support Vector Machines. Computer Fans. 8.2018, 08.
[9] Zheng Tongyan, Zheng Yi. Review of Earthquake Disasters in Mainland China in 2012. Journal of Natural Disasters. 23 (3). 2014, 06: 166-170.
[10] Zheng Tongyan, Zheng Yi. Review of Earthquake Disasters in Mainland China in 2013. Journal of Natural Disasters. (1). 2015, 02: 239-246.
[11] Zheng Tongyan, Zheng Yi. A Review of Earthquake Disasters in Mainland China in 2016. Disaster Science. 03.2018, 07: 141-144.
[12] Lin Xiangyang, Zheng Tongyan. A Review of Earthquake Disaster Losses in Mainland China in 2017. Journal of the Academy of Disaster Prevention Science and Technology. 20 (3). 2018, 09: 52-58.
[13] Guo Endong, Sun Baitao, Liu Rushan, etc. GB / T 17742. GB / T 17742-2008. China Earthquake Intensity Table. China. China Earthquake Administration. 2008.
[14] Guo Endong, Sun Baitao, Liu Rushan, etc. GB / T 24336. GB / T 24336-2009. Classification of earthquake damage levels for lifeline engineering. China. Institute of Engineering Mechanics, China Earthquake Administration, 2009.
[15] Zheng Wenqing, Ji Zhi. Visual analysis in the field of data mining. Science and Technology and Information. 05.2018, 05: 161-162.
[16] Duan Huaming, He Yang. The constructive improvement of big data for disaster assessment. Disaster Science. 31 (1). 2016, 01: 188-192.
[17] Chen Xiaoyan, Qian Liping, Guo Wenping. Comparison and application of big data visualization tools. Computer Education. 06.2018, 06:97.