Journal of Applied Mathematics and Computation, 2023, 7(3), 398-403
https://www.hillpublisher.com/journals/JAMC/
ISSN Online: 2576-0653
ISSN Print: 2576-0645

Research on Earthquake Discrimination and Magnitude Prediction Based on Fourier Power Spectrum Sample Entropy and Machine Learning Algorithm

Yuan Gao, Shuang Xu, Fei Wang*

Mudanjiang Normal University, Mudanjiang, Heilongjiang, China.

How to cite this paper: Yuan Gao, Shuang Xu, Fei Wang. (2023) Research on Earthquake Discrimination and Magnitude Prediction Based on Fourier Power Spectrum Sample Entropy and Machine Learning Algorithm. *Journal of Applied Mathematics and Computation*, 7(3), 398-403.
DOI: 10.26855/jamc.2023.09.011

Received: August 31, 2023
Accepted: September 30, 2023
Published: October 31, 2023

***Corresponding author:** Fei Wang, Mudanjiang Normal University, Mudanjiang, Heilongjiang, China.

Abstract

In this paper, the characteristics of seismic wave data are extracted mainly through windowed Fourier transform and power spectrum sample entropy. A support vector machine classification model and random forest regression model are respectively established to classify seismic events and predict the grade of the earthquake. The results showed that for earthquake discrimination, the accuracy of the test set reached 82.7%; For grade prediction, the MAE reached 0.58. Therefore, classification support vector machine and regression random forest can be used for earthquake identification and grade prediction. Additionally, the power spectrum sample entropy of the waveform signal can be used to measure the characteristics of the waveform. The earthquake discrimination and magnitude prediction methods based on Fourier power spectrum sample entropy and machine learning algorithms have potential applications in the field of earthquake monitoring and early warning. These research results provide a feasible technical tool for earthquake-related decision-making and help improve the prediction and response capability of earthquake hazards. However, further research and validation are still needed to further improve and optimize the performance and stability of the method.

Keywords

Fourier transform, Sample entropy, Support Vector Machine, Random Forest

1. Introduction

Earthquake is a relatively complex phenomenon of crustal movement, and countless seismic disasters occur in the world every year. However, with the rapid increase of urban engineering construction projects and the expansion of seismic network monitoring, unnatural seismic events such as explosions, mining earthquakes, weapons tests, and collapses occur, interfering with the accurate prediction of seismic events. Effective identification of natural seismic events in seismic monitoring and elimination of abnormal interference signals have important practical significance in earthquake warning and forecasting techniques for reducing seismic hazards. Magnitude level prediction is also one of the important objectives of earthquake forecasting. The accurate determination of seismic level depends on the feature mining of a large number of historical events and the estimation of seismic wave energy, which helps to develop targeted earthquake emergency plans and reduce damages [1].

In this paper, we will use window Fourier transform and power spectrum sample entropy methods to extract the characteristic variables of waves based on the seismic wave data recorded by instruments and establish vector machine classification models and random forest regression models to classify seismic waves and non-seismic waves with earthquake level prediction, hoping to provide a reference for the solution of earthquake differentiation and level prediction problems.

2. Wave feature extraction based on power spectrum sample entropy

The seismic signal data signal discrete waveform signal, this time, we select the data of the A question of the third national university modeling competition in Tianfu, which contains 120 natural seismic wave data and 30 non-natural seismic wave data, and all 120 seismic data are labeled with seismic level.

After observing the waveform plots of natural and artificial seismic data, it is found that the natural seismic waveform envelope transforms slowly and grows and falls back gradually, but for the non-natural seismic waves generated by blasting, etc., the waveform envelope changes abruptly at the initial position and grows suddenly from zero to a very large value in a short period. The graphs are shown in Figure 1. Therefore, the Fourier transform, a waveform signal processing tool, is used to extract the power spectrum, and the sample entropy is used to measure the power spectrum distribution and transformation to extract features from the waveform data [2].

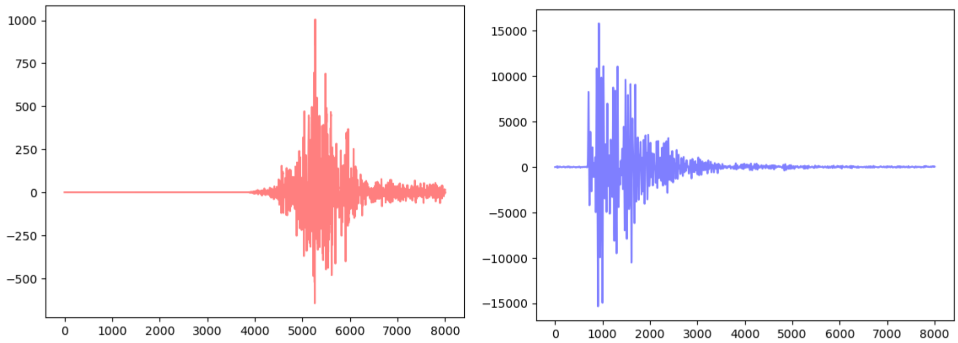

Figure 1. Feature extraction flowchart.

2.1 Adding windows

The purpose of windowing is to extract time series segments. If the data is truncated directly, spectral leakage occurs when the Fourier transform is performed, so Gu adopts the windowing process to reduce the spectral leakage. This time, the Hamming window is used:

$$\omega(n) = a_0 - (1 - a_0) \cdot \cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \le N-1 \tag{1}$$

where $a_0 = 0.53836$, the window length N is based on the fundamental period of the analyzed data, and an integer power of 2 slightly larger than the fundamental period is used as the window length.

2.2 Fourier Transform

Fourier transform can convert the signal from the time domain to the frequency domain, to obtain the amplitude and phase information of each frequency component of the waveform signal to facilitate the analysis of the signal characteristics [3]. The current data are discrete, so it is necessary to use the discrete Fourier transform, and the positive inverse transform formula for signal $x(n)$ and its spectrum $X(\mathrm{k})$ is:

$$X(\mathrm{k}) = DFT\big(x(n)\big) = \sum_{n=0}^{N-1} x(n) W_N^{bn}, k = 0,1, \dots, N-1 \tag{2}$$

$$x(n) = IDFT\big(X(\mathrm{k})\big) = \sum_{n=0}^{N-1} X(\mathrm{k}) W_N^{-kn}, n = 0,1, \dots, N-1 \tag{3}$$

In the processing, the fast Fourier transform algorithm is used to improve the efficiency of the computation by taking advantage of the periodicity of W_N^{kn} where the convolution period N is the same as the length of the add window.

2.3 Calculating the power spectrum

After obtaining the spectral variation map, a Mayer filter bank is selected to segment the frequency domain, and the frequency band is subjected to log energy statistics to obtain the log power spectrum of each window frame over time, calculated by the formula:

$$s(m) = \ln\left(\sum_{k=0}^{N-1} |X_a(k)|^2 H_m(k)\right), 0 \le m \le M \tag{4}$$

Where M is the number of frequency bands, 26 is chosen here, $H_m(k)$ is the band filter function, and N is the number of frequency points in the band. An example of the power spectrum calculation results for spectrum $X_a(k)$ is shown in Figure 2.

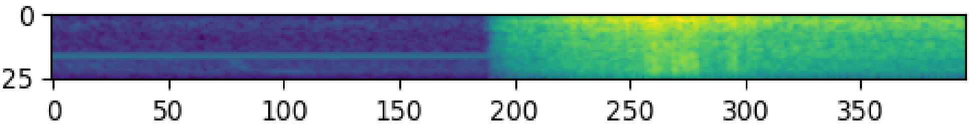

Figure 2. Segmented power spectrum.

2.4 Calculating the frequency band sample entropy

The sample entropy is used to measure the distribution complexity of the sequence. The distribution of seismic data and non-seismic data is significantly different, with seismic data initially distributed gently, but non-seismic data, initially generated at a location where there is an abrupt change, so the sample entropy is needed to measure the distribution complexity of the power data over the time series, as a way to identify the earthquake classification and predict the earthquake level.

The sample entropy of a vector sequence $X_{\mathrm{m}}(\mathrm{i})$ of length m is calculated as follows:

(1) Define the distance function

$$d[X_{\mathrm{m}}(\mathrm{i}), X_{\mathrm{m}}(\mathrm{j})] = \max_{k=0,\ldots,N-1}(|x(i+k) - x(j+k)|) \tag{5}$$

and set the distance threshold r. Calculate the number of vectors with distance less than r from $X_{\mathrm{m}}(\mathrm{i})$, denoted as

(2) Define:

$$B_i^m(r) = \frac{1}{N-m-1} B_i \tag{6}$$

and for all i, calculate its value. Then calculate

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \tag{7}$$

(3) Add 1 to the value of m to obtain:

$$A^m(r) = B^{m+1}(r) \tag{8}$$

(4) Calculate the sample entropy:

$$SampEn(m, r, N) = -\ln \frac{A^m(r)}{B^m(r)} \tag{9}$$

Here, m=3 and r=1 are chosen to calculate the sample entropy of the power spectrum of each seismic waveform data, and attach the label of whether it is a natural earthquake to the latter column of the sample entropy to construct the classification data, and if the magnitude data is attached, the magnitude regression data is constructed, and 10 entries of the classification data are selected.

3. Build machine learning classification and regression models

In dealing with such tasks as classification and regression, the current classical and effective methods are mainly machine learning classification and regression models [4]. Machine learning algorithms can learn the intrinsic laws contained in the data from the data and thus build corresponding mathematical models to characterize the relationship between the corresponding features and the target results, which can achieve the solution of the corresponding tasks. For the classification prediction of true and false earthquakes and earthquake level regression prediction, this paper mainly uses the support vector machine classification model and random forest regression model.

3.1 Support vector machine classification model

The support vector machine model is a binary classification model whose basic idea is to find the correct classification of the data set into two.

The basic idea is to find the hyperplane that correctly divides the data set into two geometrically spaced maxima. Simply put, the training data set is divided into two categories by a line (hyperplane in n dimensions), and the data can be classified by this line when new data is added. In a support vector machine model, the problem of solving the optimal decision boundary line (hyperplane) is usually transformed into the problem of solving the maximum interval between two classes of data sets, and then the center of the interval is taken as the decision boundary of this model:

$$\arg \max_{w,b} \frac{2}{\|w\|} \tag{10}$$

$$s.t.\, y_i(w^T x_i + b) \geq 1, i = 1,2, \ldots, m \tag{11}$$

When the problem to be handled is a nonlinearly divisible classification problem, the decision hyperplane under the original dimension cannot correctly explain the data [5]. The linearly indivisible data set under the original space is spatially transformed to linearly divisible by the spatial mapping transformation function, and the decision hyperplane is solved again in the new dimensional space.

The mapping transformation function makes it difficult to find the inner product, so the kernel function can be used instead of the original dimension under the

direct calculation of the vector dot product. Commonly used kernel functions are polynomial kernel, radial basis function kernel, etc. In this case, a linear kernel function with the highest subterm of 3 is chosen.

3.2 Random forest regression model

Random forest is a parallel integrated learning approach (bagging) of the decision tree model as the basis of a weak classification model.

Decision tree regression is a kind of weak classifier similar to classification decision tree, which mainly contains the computational process of the CART (classification and regression tree) algorithm, which is mainly based on loss evaluation decision to find the optimal cut point to form a bifurcation node, and then recursively solve the process for each branch region, and finally form a regression tree with tree structure.

Relying on the strong generalization ability of integrated learning, the random forest has a good processing ability on the corresponding tasks [6]. The growth process of the decision tree in the random forest regression model is random, with a random selection of samples from the training set, and a random selection of features from these samples for training, resulting in different training results. The advantage of the random forest regression model is that because the magnitude of the impact of each feature on the results is not known initially, the use of a random process can reduce the impact of this on the classification results, and the final decision result of the random forest regression model is determined by the vote of each decision tree.

Here, the number of decision trees is selected as 100, the maximum depth of the tree is 10, the maximum number of leaf nodes is 50, and the node classification loss evaluation strategy uses MSE.

4. Experimental results and analysis

The experiments use the power spectrum sample entropy data extracted from the dataset to slice the training data and the test data, and the slice ratio is 7:3. The corresponding data are used to train the support vector machine classification model and the random forest regression model respectively, and then the test data are used to test the obtained models.

The support vector machine classification model was tested, and the test results are shown in Table 1. Both the training set data and the test set data performed more than 82% accuracy on the model, indicating that the obtained model better reflects the characteristics of the original data and has good generalization and prediction ability.

Table 1. Classification model evaluation results

	Accuracy Rate	Recall Rate	Accuracy Rate	F1
Training set	0.822	0.822	0.676	0.742
Test set	0.827	0.827	0.684	0.749

The random forest regression model was tested again, and the results of the 15 tests selected were plotted in Figure 3, as well as the evaluation results of the model shown in Table 2. From the graph of the prediction results, we see that the model was able to make good predictions for most of the data, as well as the MAE of the test set was 0.589, indicating that the average absolute value of the error in prediction was about 0.6, which is about a little more than half a magnitude, indicating that the model has a good prediction for earthquake magnitude.

Table 2. Evaluation results of random forest regression model

	MSE	RMSE	MAE	MAPE	R^2
Training set	0.104	0.323	0.261	4.353	0.929
Test set	0.601	0.776	0.589	9.701	0.652

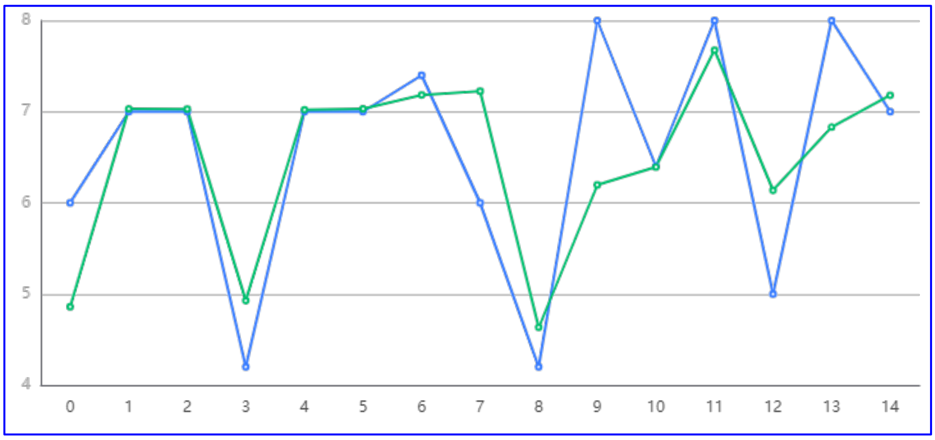

Figure 3. Test results of random forest regression model.

5. Conclusion

From the experimental results, the two selected support vector machine and random forest machine learning models can achieve good result indicators for the corresponding tasks, so they can be used to do earthquake discrimination and magnitude prediction, and also show that for seismic waveform signals, the power spectrum sample entropy can characterize the seismic waveform well. However, as seen in the test results of the random forest regression model, the performance of the test set is worse than the training set, so there is an overfitting phenomenon. The next work can further deal with the overfitting problem, such as regularization, or improving the quantity and quality of the data set to reduce or eliminate the overfitting problem.

6. Discussion

Earthquake discrimination and magnitude prediction based on Fourier power spectrum, sample entropy, and machine learning algorithms is a promising research area that can be further investigated in the future from the following aspects:

Improving feature extraction methods: Fourier power spectrum and sample entropy are commonly used features in seismic signal analysis, but there is still room for improvement. The use of other spectral analysis methods, such as wavelet transform and time-frequency analysis, can be explored to obtain more accurate and comprehensive feature information. In addition, other seismological parameters, such as source mechanism and crustal structure, can be combined to construct more complex feature representations.

Introduce deep learning algorithms: Traditional machine learning algorithms have achieved certain results in earthquake discrimination and magnitude prediction, but deep learning algorithms have more powerful representation and feature learning capabilities. Future research can explore the use of deep learning algorithms, such as convolutional neural networks, recurrent neural networks, and attention mechanisms, to improve the accuracy and robustness of earthquake discrimination and magnitude prediction.

Data set and sample labeling: The importance of seismic data acquisition and labeling for research cannot be overstated. Future research can collect more seismic data and label the data more carefully and accurately. At the same time, the problem of classifying multiple seismic event types, such as crustal shaking, groundwater dynamics, and subsurface rock movement, as well as the problem of predicting different earthquake magnitudes, can be considered.

Multimodal data fusion: In addition to seismic signal data, seismic research can also combine other related multimodal data, such as surface deformation data, geomagnetic data, and seismic waveform data, for comprehensive analysis and prediction. By fusing information from multiple data sources, the accuracy and reliability of earthquake discrimination and magnitude prediction can be improved.

Real-time monitoring and early warning: Real-time monitoring and early warning of earthquakes are important for reducing earthquake disasters. Future research can apply earthquake discrimination and magnitude prediction based on Fourier power spectrum sample entropy and machine learning algorithm to the real-time monitoring system to achieve fast and accurate earthquake identification and early warning.

In conclusion, the research of earthquake discrimination and magnitude prediction based on Fourier power spectrum sample entropy and machine learning algorithm has a broad development prospect. The research has a broad development prospect. Future research can explore in-depth feature extraction, algorithm model, data set and sample labeling, and multimodal data fusion to improve the accuracy and reliability of earthquake monitoring and prediction and provide more effective support for earthquake disaster prevention.

Funding

Heilongjiang University Student Innovation and Entrepreneurship Training Program Project (alpha intelligent app and algorithm software company No.202010233025); Mudanjiang Normal University Scientific Research Project (Research and Improvement of Image Style Migration Algorithm No. QN2020006); Heilongjiang Provincial Education Department Project (Research on Face Recognition Related Algorithm No.1355MSYQN003).

References

[1] Renouard, A., Maggi, A., Grunberg, M., Doubre, C., & Hibert, C. (2021). Toward false event detection and quarry blast versus earthquake discrimination in an operational setting using semiautomated machine learning. *Seismological Research Letters*, *92*(6), 3725-3742.

[2] Yang, D. H., Zhou, X., Wang, X. Y., & Huang, J. P. (2021). Mirco-earthquake source depth detection using machine learning techniques. *Information Sciences*, *544*, 325-342.

[3] Saini, K., Kalra, S., & Sood, S. K. (2022). An Integrated Framework for Smart Earthquake Prediction: IoT, Fog, and Cloud Computing. *Journal of Grid Computing*, *20*(2), 17.

[4] Xiong, P., Marchetti, D., De Santis, A., Zhang, X., & Shen, X. (2021). SafeNet: SwArm for earthquake perturbations identification using deep learning networks. *Remote Sensing*, *13*(24), 5033.

[5] Toh, G., & Park, J. (2020). Review of vibration-based structural health monitoring using deep learning. *Applied Sciences*, *10*(5), 1680.

[6] Mukherjee, T., Singh, C., & Biswas, P. K. (2021). A Novel Approach for Earthquake Early Warning System Design using Deep Learning Techniques. *arXiv preprint arXiv:2101.06517*.