Seismic Waveform Classification Based on Multi-classification Model
Weighted Voting Method

Zhong-qiang GAO and Chuan-da QI
Xinyang Normal University, School of Computer and Information Technology, China

Keywords: Weighted voting, Sub-model, Waveform, Spectrogram, Three-component.

Abstract. The earthquake event is caused by the activity of the interior earth, which has the potential and can create great damage. Thus, in big data era, it is necessary to fast, efficient and universal automatic recognition and classification of seismic waveform. A well-performed recognition and classification algorithm should be sensitive to small and weak events with various waveform shapes, robust to background noise and non-seismic signals, and effective for processing a large number of data in different areas. In this paper, we introduce the weighted voting method in ensemble learning, taking 50000 seismic signals monitored by the capital circle and its nearby stations as training and testing data, we learn the main time-frequency characteristics of seismic positions in the three-component seismic signal data through the sub-model, and then identify and classify them. Finally, the classification score of the sub-model is weighted to confirm the final event type. Compared with single machine learning models, our result shows improved recognition and classification performance and reduced uncertainty of ensembles.

Introduction

Shallow fully connected neural networks (NN) is one of the earliest machine learning methods for the seismic signal recognition classification [1][2]. The neural network accepts the extracted feature vector as input and transforms it through a series of neurons in the hidden layer to predict the expectation in the output layer. In the model, each neurons is fully connected to the neurons in their adjacent layer, but, the neurons in each layer are independent of each other, and thus the weights are not shared. The nonlinear activation function inside the neurons makes the construction of relationship between complex input and output possible through the optimization process [3][4].

With the continuous emergence of new methods [5], Support vector machine (SVM) is used in seismic waveform recognition and classification for its fast and accurate classification effect. Experts at home and abroad combine various feature extraction algorithms with support vector machine to conduct identification and classification [6] [7]. Support vector machine (SVM) is a supervised learning method, which establishes an optimal decision hyperplane for extracted multidimensional features to, maximize the distance between two classes, and thus provides good generalization ability for classification problems. However, support vector machine also has some disadvantages, such as difficulty in using large-scale training samples and difficulty in analyzing multi-class classification problem.

The traditional time-frequency domain feature extraction algorithm produces, extracted features difficult to apply to fixed complex seismic signal, on the completeness and universality. With availability of big data and the improvement of computing power, CNN has been developed rapidly, Because of its strong parallel computing ability, fast running speed, good adaptive and fault-tolerant ability, and has made a great breakthrough in the field of pattern recognition. In addition to fully connected layers, the convolution neural network includes a convolution layers and pooling layers in addition to the fully connected layer. Convolution layer through the weights sharing, using the same convolution to check the size of a layer of output characteristics of convolution computation, the network will automatically learn low/mid/high-level features of data, which reduces pre-processing and manual engineering. The calculated feature graph is passed to the next layer through activation function, and the feature of adjacent regions at a certain position are calculated by pooling layer. In this way, convolution and pooling are continuously performed, and finally the
extracted feature map is subjected to dimensionality reduction operation at the full connection layer, so that the dimension of the learning feature can be reduced during data transmission. In recent years, convolution neural networks have also been applied to seismic event monitoring [8][9]. These studies show the promising performance of deep neural network in robust and efficient classification of seismic signals [10] [11].

Ensemble learning is the latest advancement in machine learning, it uses the existing data to generate multiple sub-model, and then merges these sub-model to achieve the final prediction[12][13]. Integrated learning can not only produce a more stable global model, but also ensure the reduction of uncertainty [14]. The effectiveness of ensemble learning is discussed in the published literature. Generally speaking, for the integrated learning framework, there are mainly three stages. The first stage is to generate a series of resampled subsets from the raw data, the second stage is the construction and optimization of the sub-model, and the third stage is the integration of the models, which combines the estimates generated by the sub-model to determine the overall estimate. Ensemble learning frameworks are divided into two broad categories, homogeneous ensembles and heterogeneous ensembles [15][16]. In homogeneous ensemble frameworks, the same resampling technique and the same nature model are used, and there is only one integration technology. Heterogeneous ensemble frameworks violate the definition of homogeneous ensembles, but maintain the three fundamental stages of ensemble learning [17]. In this work, heterogeneous ensemble frameworks are considered and, so all the individual models will have different input and output.

In the seismic waveform classification research, a learning framework of weighted voting method is proposed because of the uncertainly of single model and the lack of performance stability [13]. Using ANN and CNN as sub-model in ensemble learning, the eigenvalues, waveform and spectrogram after three-component seismic wave processing are identified and classified. Then the weighted results of three classification jobs were calculated to confirm the final type of vibration event. At the same time, the performance of ensemble models is compared with single machine learning models, and the effect of data availability on the models’ generalization ability is examined.

Method

Weighted voting method is one of the most common ensemble learning frameworks. It is a set of learning frameworks for any series of given machine learning models. While also following the three stages of ensemble learning. We first describe the sample generation process. Due to the special input format of the three sub-model (refer with: Fig.1). It is necessary to process the input data one by one. In weighted voting, we represent $h_i$ predicted output on sample $x$ as an N-dimensional vector ($h_i^1(x), h_i^2(x),..., h_i^n(x)$), where $h_i^j(x)$ is $h_i$ output on category mark $c_j$. as follows:
\[
H(x) = c_{\arg \max_j} \sum_{j=1}^T w_j h_j^j(x)
\]

(1)

Where \(w_i\) is the weight of \(h_i\), usually \(w_i \geq 0, \sum_i w_i = 1\).

For the artificial neural network model, the original data are extracted (refer with: Table 1), the waveform and phase are read from the obtained seismic (trace file), the signal-to-noise ratio (SNR) of P-wave is calculated, and the eigenvalues are extracted for the records satisfying the conditional (the default signal-to-noise ratio is greater than 2.5).

Table 1. Artificially extracted seismic features.

| Feature | Description |
|---------|-------------|
| F1-1    | Amplitude mean of the three values before P-wave |
| F1-2    | Maximum amplitude of the three values before P-wave |
| F2-1    | Maximum amplitude of the P-wave values |
| F2-2    | The ratio of the maximum amplitude of P-wave to its initial amplitude |
| F3      | Maximum amplitude of the S-wave values |
| F4      | Fluctuation degree of P-wave spectrum envelope |
| F5-1    | The dominant frequency of the first half of the P-wave |
| F5-2    | The dominant frequency of the posterior half of P-wave |
| F6      | The trajectory after S-wave phase |

In the traditional seismic waveform eigenvalue recognition and classification algorithm, the single component data of three components are used, which may cause the loss of information. Therefore, this paper splicing the three component eigenvalues corresponding to each event (9-dimensional single-channel eigenvalue data splicing into 27-dimensional three-channel eigenvalues), (refer with: Table 2):

Table 2. Vibration event feature.

| event     | features (dim = 27) | type | label |
|-----------|---------------------|------|-------|
| LC_SJS_201802120416.0001_SD.DSD.00 | [0.1181,0.029362,...,0.575,0.193033] | LC   | 2     |
| LC_SJS_201802120416.0001_SD.CH.00 | [0.04177,0.066312,...,0.28611,0.049496] | LC   | 2     |
| LP_SD_201608250027.0D.HAPY.00 | [0.032224,0.04267,...,0.08333,0.193033] | LP   | 1     |
| LP_IN_201712160436537.1GP_LN.XIN.00 | [0.051563,0.085965,...,0.094444,0.391915] | LP   | 1     |
| ...      | ...                 | ...  | ...   |
| L SD 2017082000121569.1GP_SD.SHH.00 | [0.038279,0.536360,...,0.9325,0.642105] | L    | 0     |
| L SX_201605180503001.1GP(SK.XAY.00 | [0.116754,0.15604,...,0.346667,0.396842] | L    | 0     |

For the waveform and spectrum data, considering the time difference of the P-wave and S-wave of the sample, the processed vibration waveform data is intercepted 3s before P-wave and the image is intercepted 40s after P-wave as the training sample of model. Data is affected by the idea of RGB three channels in pixels and the three component waveform is drawn into the form of three component waveform is drawn into the form of three-channel color map[18]. the image size is fixed to 1024×51 (refer with: Fig.2 Fig.3).

Figure 2. Three-component seismic waveform.
For the electronic equipment used in receiving the seismic waveform signal, the abnormal images is easily affected by the propagation path and the noise interference of the instrument itself. In this paper, the idea of signal-to-noise ratio (SNR) and high-pass filtering is introduced. Some low-frequency signals are filtered by high-pass filtering and the threshold of signal-to-noise ratio is set to threshold that, only the three-component minimum signal-to-noise ratio (SNR) above it is retained (refer with: Eq. 2).

\[
SNR(dB) = 10 \log_{10} \left( \frac{P_{signal}}{P_{noise}} \right) = 20 \log_{10} \left( \frac{A_{signal}}{A_{noise}} \right)^2
\]  

(2)

After data preprocessing, three separate machine learning models are generated to perform the second phase. The types of the individual models are predetermined, and they can be selected according to the nature of the problem. Each model will rely on the data created to train and create a relatively unique hypothesis. After all sub-model are generated and trained, the output of these individual models is derived from the output of these individual models by weighted voting.

Weight voting is a very intuitive method, which gives a high weight to the classifier with high classification performance. The voting result can often use the complementary function between classification models to reduce the error of the single classifier and improve the prediction performance and classification accuracy. First, calculate the correct rate for each classification models as follows:

\[
P_i^* = \frac{m_i}{m}
\]

(3)

Where \( i \) is the number of sub-model, \( s \) is the number of event types, and \( m \) is the number of correctly classified event types \( s \). Then the weight \( w_i \) is calculated according to the correct rate of the classification model, and the output probabilities of the detection results obtained by there respectively multiplied by the corresponding weights \( w_i \) to obtain \( sum_i \).

\[
w_i = \frac{t_i}{\sum_{i=1}^{3} t_i}, \quad i = 1, 2, 3 \quad t_i = \frac{\sum_{i=1}^{3} P_i^*}{3}
\]

(4)

Finally, \( sum_i \) of each sub-model is added to obtain \( sum \), which is the final type of the event according to the index number of the maximum value of \( sum \).

**Results and Discussion**

In the present study, the selection of the testing set as seismic events from completely different sites is important in order to test the model’s capability in the regional prediction problem. The selection and processing of research features and the marking of seismic events was performed manually. The model’s training and testing data are derived from raw data monitored in the monitors in the metropolitan area and surrounding stations. For the raw data, the natural data is masked as 0, the
blast mark as 1, and the collapse mark as 2, incomplete records in the original data set are not taken into account in the trained data set.

The preprocessing of the data involves normalizing the features. For the eigenvalue data, the minimum and maximum values are normalized between 0 and 1, and normalization will be performed independently according to each feature. For waveform and spectrum data, the global maximum value of three components is normalized to ensure that the generated images have same scale. We used the SMOTE method to equalize the range of data in the stratified sampling, which can analyze the samples in the minority and add new samples to the data set artificially in respect to the minority samples. For image-based classification, a mini-batch gradient descent training model was adopted to ensure category balance, and the same number of samples were randomly selected from different samples for each mini-batch.

We all know that the earliest convolution neural networks extract features by convolution with large convolution kernels. In this paper, we use the VGG depth network model (refer with: Fig.4). With input data to be the seismic three-component waveform or spectrogram. We replace the previous large-size convolution kernel by multiple 3×3 convolution kernels in each block, and maximize the extraction of features by repeatedly using 3×3 convolution kernel and 2×2 maximum pool layer. At the same time, the ReLU function is used as the activation function, which increases the BN normalization layer, which can effectively prevent gradient explosion and gradient dispersion, and it is also beneficial to the convergence of the network. We used a sigmoid binary activation function in the last layers of the network, and the output is a vector of predicted probabilities for each sample to contain an earthquake signal.

![Convolution neural network model.](image)

**Figure 4. Convolution neural network model.**

**Training and Testing of the Model**

We randomly split the data set into training (80%), validation (10%), testing (10%) sets. We trained the network on three GeForce RTX 2080 GPU. The experimental setup and simulations are carried out in Python3.6 environment. We used cross-entropy as the loss function and the ADAM algorithm for optimization [19] [20]. During training, the initial learning rate is set to 0.01, and the learning rate is reduce once every 5 epochs. If we found that the loss value of validation set remained unchanged and the validation accuracy did not improve in the last 15 epochs, the training of model is stopped (refer with: Fig.5). The callbacks function is used in the retraining process to monitor the loss of the model and save the optimal model.
Figure 5. Loss and accuracy of model training and testing.

To evaluate the performance of the model for detection, we used 5781 test samples. After selecting a decision threshold value (tr) for output probabilities. So far, we have used accuracy in evaluating a model. Precision is generally used to evaluate the accuracy of global models, with only a number, not reflect too much detail. Therefore, the Kappa coefficient is used to measure the classification accuracy of seismic events [21].

In order to evaluate the classification effect of the model better, the confusion matrix of three classifiers is made, the results of mode prediction errors, the wrong prediction categories and the correct prediction quantity are all show in one matrix, which is convenient and intuitive to evaluate the classification results of the model. At the same time, a feature’s relative importance test is carried out for each of the individual models (refer with: Table. 3, Table. 4, Table. 5, Table. 6).

|                | Earthquake(pred) | Blasting(pred) | Collapse(pred) |
|----------------|------------------|----------------|----------------|
| Earthquake(gt) | 1801             | 130            | 82             |
| Blasting(gt)   | 55               | 1783           | 157            |
| Collapse(gt)   | 35               | 79             | 1659           |

Table 3. Confusion matrix of eigenvalue classification.

|                | Earthquake(pred) | Blasting(pred) | Collapse(pred) |
|----------------|------------------|----------------|----------------|
| Earthquake(gt) | 1899             | 89             | 25             |
| Blasting(gt)   | 16               | 1845           | 134            |
| Collapse(gt)   | 13               | 55             | 1705           |

Table 4. Confusion matrix of waveform classification.

|                | Earthquake(pred) | Blasting(pred) | Collapse(pred) |
|----------------|------------------|----------------|----------------|
| Earthquake(gt) | 1920             | 74             | 19             |
| Blasting(gt)   | 15               | 1889           | 91             |
| Collapse(gt)   | 9                | 53             | 1711           |

Table 5. Confusion matrix of spectrum classification.

|                | Earthquake(pred) | Blasting(pred) | Collapse(pred) |
|----------------|------------------|----------------|----------------|
| Earthquake(gt) | 1943             | 43             | 27             |
| Blasting(gt)   | 9                | 1952           | 34             |
| Collapse(gt)   | 5                | 35             | 1733           |

Table 6. Confusion matrix of Weighted Voting classification.
The stability of the investigated models is an important aspect that should be investigated in the comprehensive analysis of machine learning models. The confusion matrix of the sub-model and the calculated accuracy and Kappa coefficients are described form Table 3 to Table 6. Under the same test data set, whether it is single event classification or global classification accuracy, the seismic recognition classification model with spectrum map as data is superior to the other two models. The classification effect for the F1 model is slightly lower than that of the other two models. It may be because the artificial extraction of seismic wave characteristics is less, so that some features of the original data are lost, and the comprehensive information of the seismic waveform cannot be well represented. Models F2 and F3 use a convolutional neural network, which feature is learned during model training.

Since the classification effect of the sub-model on a single event is different, it is not possible to judge the quality of the model. Therefore, the sub-model is given different weights according to the global classification accuracy of the model and the Kappa coefficient, and the final prediction is obtained based on the weighted voting method. Under the same test data set, the classification effect of the comprehensive model was tested experimentally. The relevant results are listed in Table 7 and 8. The classification accuracy of the integrated model is improved compared with the sub-model in the single event and global classification accuracy. The model has certain generalization performance in seismic classification and recognition.

Summary
Seismology is a data-driven subject. Accurate identification of event types from massive raw waveform data is the first step in seismology and is therefore the basis for seismological research. In this paper, a multi-classification model weighted voting method is applied to weight the sum of the effects of the “base classifier” to obtain a strong classification model framework, which can detect the types of earthquake events with high efficiency and high precision. Compared with the traditional classification method of signal or multiple feature functions, the convolution neural network in the sub-model extracts abstract features from the training data, which means that the model has stronger generalization ability and can match with more waveform features. However, in practical applications, there will be a “bottleneck” problem, which is difficult to continue to improve after the accuracy of training and testing reaches a certain level. Therefore this paper applies a weighted voting method to prove the usability of weighted voting in seismic waveform recognition classification through various targeted test schemes. Although there is still room for improvement in this method, waveform and spectrogram as feature parameters combined with convolution neural network provide some reference for the application of seismic signals. Traditional seismic waveform classification tends to focus only on natural earthquakes and blasting. Most of the data uses single-component data from three components to train, resulting in the loss of partial dimensional information. Therefore, this paper increase collapse events in the event type, and splice the three-component data in data preprocessing to ensure the completeness of the data, which is a new challenge in this filed. For future studies, more research is needed to fully address...
the integrated learning problems in this field. The explicit use of diversity concepts in the
development of integrated learning models is expected to provide further improvements for
predictions.

Acknowledgements
This work is supported by Xinyang Normal University and Nanjing Innovation Data Technologies, inc.

References
[1] Bian Y J. 2002. Application of genetic BP network to discriminating earthquake and explosions
[J]. Acta Seismologica Sinica, 24 (3): 516-524.

[2] Mousavi SM, Horton SP, Langston CA, Samei B. Seismic Features and Automatic
Discrimination of Deep and Shallow Induced-Microearthquakes Using Neural Network and
Logistic Regression. Geophysical Journal International. 2016; 207: 29-46. doi: 10.1093/gji/ggw258.

[3] Wang T T, Bian Y J, 2011. Criterion selection of earthquake and explosion recognition[J].
Seismological and Geomagnetic Observation and Research, 32(6): 62-67.

[4] Tian Ye, Huang Han-Ming, Bian Yin-Ju, Et Al. Visualization Methods for Discriminating
between Earthquakes and Explosions[J]. Advances in Geosciences, 2014(4):136-146.

[5] Zhang B, Bian Y J, Wang T T. 2014 Discrimination of earthquake and explosions by SAMC
decision method. Acta Seismologica Sinica, 36(2): 233 - 243.

[6] Bi M X, Huang H M, Bian Y J. 2011. A study on seismic signal HHT features extraction and
SVM recognition of earthquake and explosion[J]. Progress in Geophysics, 26(4): 1157-1164.

[7] Li X H, Liu G, Liu D Z, Qin Q Q , 2009. Discrimination of unclear explosions and earthquakes
using the support vector feature line fusion classification algorithm[J]. Chinese Journal of
Geophysics, 52(7): 1816-1824.

[8] Chan Run-Hang, Huang Han-Ming, Chai Hui-Min. Study on discrimination of seismic
waveform signals between earthquake and explosion events by convolutions neural network.
progress in Geophysics, 2018, 33(4):1331-1338.

[9] T. Perol, M. Gharbi, and M. Denolle. Convolutional neural network for earthquake detection
and location. Science Advances, 4(2): e1700578, Feb.2018.

[10] Zhao M, Chen S, Dave Yuen. 2019. Waveform classification and seismic recognition by
convolution neural network. Chinese J. Geophys. 62(1): 374-382.

[11] Chen Run-Hang, Huang Han-Ming, Chai Hui-Min. 2108. Study on the discrimination of seismic
waveform signals between earthquake and explosion events by convolutional neural
network. Propress in Geophysics, 33(4): 1331-1338.

[12] Zhao Gang, Huang Han-Ming, Lu Xin-Xin, etal. Research on the Classification of Seismic
Wave Signals of Earthquakes and Explosion Events Based on BP-Adaboost [J]. China Earthquake
Engineering Journal, 2017, 39(3)

[13] Mohammad H. A, Mohamed A. M, Fateh C, Predicting seismic-induced liquefaction through
ensemble learning frameworks. 2019, Scientific Reports 2019 9:11786. doi:
10.1038/s41598-019-48044-0.

[14] Bishop, C. M. Machine learning and pattern recognition. Information Science and Statistics.
Springer, Heidelberg (2006).
[15] Zhang C. Ma, Y. Ensemble machine learning: methods and applications. (Springer, 2012).
[16] Brown, G. In Encyclopedia of Machine Learning 312-320.(Springer, 2011).
[17] Polikar, R. In Ensemble machine learning 1-34(Springer, 2012).
[18] Mousavi S. M, Zhu W q, Sheng Yi Xiao, A Deep Residual Network of Convolutional and Recurrent Units for Earthquake Signal Detection. Scientific Reports 2019 9:10267. doi: 10.1038/s41598-019-45748-1.
[19] Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization arXiv:1412.6980 (2014).
[20] R. Sebastian, “An overview of gradient descent optimization algorithms,” arXiv:160704747.
[21] Cohen J. A coefficient of agreement for nominal scales. Educational and psychological measurement. 1960;20:37-46. doi:10.1177/001316446002000104.