Title: Self-supervised seismic data classification using auto-encoder and data clustering

Abstract

List of contents

1. Introduction

Improving the signal-to-noise ratio and quality of seismic signals is of great significance for high-precision seismic exploration. In the process of seismic data processing, scientists need to follow the three basic processing requirements of high signal-to-noise ratio, high resolution, and high fidelity, among which high signal-to-noise ratio is a necessary condition for realizing high resolution and high fidelity. Therefore, high-precision seismic exploration is inseparable from the improvement of the signal-to-noise ratio and quality of seismic signals.

Seismic noise suppression and data reconstruction are effective methods to improve the signal-to-noise ratio and quality of seismic data. In the process of seismic signal acquisition, due to the influence of various factors such as the field construction environment and acquisition instruments, such as the length and combination of geophones, wind and grass on the ground surface and animal activities, mechanical and human vibrations, etc., the collected earthquakes Various noise disturbances will appear on the recording. These noises will destroy the continuous effective reflection signal and reduce the signal-to-noise ratio of seismic data. According to the degree of coherence with the signal, seismic noise can be divided into coherent noise and incoherent noise. Among them, coherent noise interference usually has deterministic characteristics, which is highly distinguishable from effective signals and is easy to identify, while incoherent noise has irregular characteristics, random distribution, and sufficient aliasing with effective signals, making it difficult to suppress. Therefore, seismic noise needs to be suppressed to improve the signal-to-noise ratio of seismic signals. In addition, the lack of seismic data is also an important reason that affects the quality of seismic data and the signal-to-noise ratio. When the field acquisition conditions are not ideal, such as the presence of a large number of obstacles, faulty geophones, and insufficient number of geophones, there will be missing seismic traces in the collected data. The seismic data reconstruction method can be used to effectively recover missing seismic traces, thereby improving the signal-to-noise ratio and quality of seismic data [4-5].

The development of seismic data noise suppression and data reconstruction methods based on unsupervised feature learning can inject new vitality into the development of seismic noise reduction and monitoring. After decades of development, scholars at home and abroad have proposed many seismic noise suppression and data reconstruction methods to improve the signal-to-noise ratio and quality of seismic data, but most of the methods have shortcomings and require specific assumptions. In addition, conventional seismic noise suppression and data reconstruction methods often ignore the fidelity of the signal while pursuing a high signal-to-noise ratio, which in turn affects the accuracy of subsequent seismic imaging and inversion. Therefore, in order to continue to improve the effect of seismic noise suppression and data reconstruction, it is necessary to further study new methods of seismic data noise suppression and data reconstruction.

In recent years, the rapid development of artificial intelligence technology has laid a solid foundation for solving complex geophysical exploration problems. The intelligent method based on feature learning can effectively learn and mine the complex features of seismic data, which

brings dawn to the development and progress of seismic noise suppression and data reconstruction methods. Since both the suppression of seismic noise and the reconstruction of missing seismic data are inverse problems in seismic exploration, and unsupervised feature learning has better applicability in solving inverse problems than supervised feature learning, so the development is based on unsupervised Seismic noise pressure for feature learning
The manufacturing method and the seismic data reconstruction method are feasible. In addition, the study of seismic noise suppression methods and data reconstruction methods based on unsupervised feature learning can reduce the dependence on traditional assumptions, and realize the learning and extraction of seismic data features in a data-driven form, which is conducive to promoting the development of intelligent geophysical exploration.

2. Literature Review

2.1 Research status of seismic data reconstruction

When the field acquisition conditions are relatively harsh, there will be missing seismic traces in the acquired raw seismic data. In addition, the elimination of discarded shots and bad tracks in the process of seismic data preprocessing will also produce missing seismic tracks in seismic data. Seismic data reconstruction is an important processing technique for recovering missing seismic traces. In recent decades, domestic and foreign researchers have proposed a large number of seismic data reconstruction methods. According to the different theories, seismic data reconstruction methods can be divided into the following categories. The first category is the reconstruction algorithm based on the wave equation. Ronen (1987) [1] achieved interpolation of seismic traces by Fourier analysis of spatial aliasing and solving the wave equation. Fomel (2003) [2] proposed a finite-difference offset continuation filtering method, which can realize seismic data reconstruction under the premise of known velocity information. Such methods rely on known subsurface information and require a large amount of calculation, so it is difficult to be widely used in actual production.

The second category is the reconstruction method based on predictive filtering. Such methods exploit the predictability of continuous events to build predictive filters and then reconstruct missing signals. Spitz (1991) proposed f−x domain prediction error filtering method for the first time and used it in the reconstruction of seismic data, but it has shortcomings in anti-aliasing [3]. Porsani (1999) improved the computational efficiency of the predictive filtering method by introducing a half-step predictive filter [4]. Wang (2002) extended the Spitz method to 3D, first estimated the linear predictor from the given frequency slice, and then predicted the high-frequency information, so as to reconstruct the signal with high quality [5]. Naghizadeh and Sacchi (2009) proposed an adaptive seismic trace interpolation method in the f−x domain to further deal with the lateral change of dip [6]. Liu and Chen (2018) proposed an unsteady autoregressive method in the f−x domain to efficiently reconstruct missing traces in complex seismic data [7]. This kind of method has gradually developed and matured, and has played an excellent role in solving complex seismic data reconstruction problems with aliasing frequency, high dimensionality and high missing rate, and has been widely recognized in actual production.

2.2

3. Methods

In seismic signal processing, feature learning can be used to mine low-dimensional hidden layer features in multidimensional seismic data and reconstruct and express original seismic data. Feature learning can be divided into supervised feature learning and unsupervised feature learning according to whether labeled data other than known data is required. Among them, supervised feature learning requires a large amount of manually labeled data to participate in the training of the model. On the contrary, unsupervised feature learning can capture the hidden layer features of the data only by knowing the input data without relying on additional manually labeled data. In the process of noise suppression or data reconstruction of actual seismic data, artificially labeled noise-free data or complete data are not accurate. At this time, unsupervised feature learning that does not rely on manually labeled data has better applicability than supervised feature learning. . Unsupervised feature learning includes algorithms such as principal component analysis, autoencoder network, matrix decomposition, and cluster analysis. Among them, autoencoder network can be used to suppress random noise interference in seismic data, and low-rank matrix approximation based on principal component analysis and matrix decomposition Theory can be used for seismic data reconstruction. This chapter will introduce the basic theory of autoencoder network, an unsupervised feature learning method.

3.1 Auto-encoder

The auto-encoder network is a neural network that aims at reconstructing the input and automatically learns the hidden layer features of the data from unlabeled data. The number of network layers is usually 3 or more. As shown in Figure 2, taking the three-layer self-encoder network as an example, the network encodes the input x into the hidden layer feature h, and then decodes the hidden layer feature h into the output y, by constructing a loss function that makes the output as consistent as possible with the input Realize the learning and extraction of hidden layer features in an unsupervised manner. It can be seen that the dimensions of the input x and the output y are always consistent, and the dimension of the hidden layer feature h can be determined according to the training goal of the network. When the hidden layer feature h dimension is smaller than the input x dimension, the autoencoder network realizes the compression of the input in the hidden layer, and further decompresses it to the output layer. Through the training of the network, the entire process can maximize the filtering of redundant information and keep the main characteristics of the data. Since the self-encoder network has the characteristics of better reconstructing the input by learning the hidden layer features of the data, this paper will focus on its ability to mine the hidden layer features of the seismic data, and suppress earthquakes by using it to reconstruct the seismic data. The processing target for random noise.

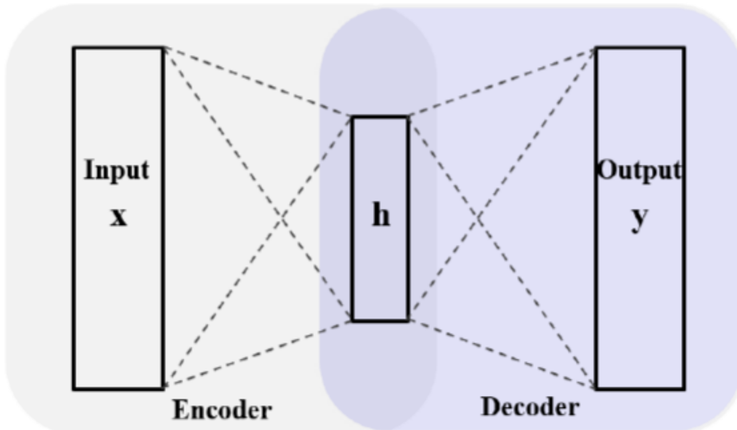

Figure. 1 Schematic diagram of autoencoder network

3.2 k-means cluster

The K-means algorithm is an unsupervised clustering algorithm based on Euclidean distance. The clustering process is to randomly extract K data points from the original sample set as the initial cluster center point, and then calculate the relationship between each data point in the sample set and K Then each data point belongs to the cluster center point corresponding to the minimum distance. The cluster center point and all the data points it belongs to are a category, and every time a new data point is added to the category, the position of the cluster center point is recalculated (that is, the cluster center point is determined by all data points in the category), continuously iterate the above process until the algorithm satisfies one of the following convergence criteria: (1) no new data points are assigned; (2) the center point of the category does not change; (3) the sum of squared errors is locally minimized. The specific process of the algorithm is described as follows:

Although the K-means algorithm has a short running time, the clustering results are greatly affected by the initial position of the centroid.

3.3 Proposal Method
(1) Seismic data preprocessing and network model building: Firstly, noise reduction and standardization processing are performed on seismic data, and stratum slices are made along horizon lines to obtain seismic waveform images. Then, according to the characteristics of the data, a deep self-encoding model based on convolutional neural network is built to extract the deep features of the seismic waveform image;
(2) Network model training and feature extraction: Seismic waveform images are used to train a deep autoencoder model based on a convolutional neural network. The seismic signal is reconstructed through the extracted deep features. When the error between the reconstructed seismic signal and the actual input is less than a certain threshold, the network model can be

considered to have been trained. Input these seismic waveform images again, and the depth features can be extracted after the pooling layer;
(3) Pattern recognition and clustering of features: Pattern recognition is performed on the extracted depth features, and the seismic data are classified through cluster analysis.

4 Experiment Detail

4.1 Data preprocessing

Data preprocessing is a very important step in the entire process of building a deep convolutional neural network. Whether the data is effectively processed affects the quality of the entire model training result. Due to the differences between different models and the characteristics of the data set itself, the preprocessing method should be adjusted accordingly according to actual needs. The following will introduce the two most commonly used data preprocessing methods in practical applications.

(1) Zero mean
Zero-meanization is to calculate the average value of each dimension in the original data, and then subtract all the data in each dimension from the average value of this dimension to form a new array. Zero-mean processing can eliminate the errors caused by different dimensions to a certain extent, and use the method of reducing the same part to highlight the differences between elements in the data.

(2) Normalization
Normalization is to transform the features of different dimensions in the original data to the same scale, so as to enhance the comparability between features in different distribution ranges. Normalization can eliminate the influence of outliers, improve the convergence speed of the entire network model and the accuracy of the classifier. The commonly used linear function normalization is to subtract the sample mean from the sample data and then divide it by the difference between the maximum value and the minimum value of the sample data to achieve the scaling of the original sample data.

4.2 Model parameter initialization

In deep learning, model parameter initialization is to reduce the dependence of parameter updates on the initial value of the objective function during stochastic gradient descent. The initial value setting of the model weight and bias is directly related to the convergence speed of the entire network. If the initialization is ideal, it can not only improve the stability of the model, but also prevent the optimal solution of the non-convex objective function from falling into the local extreme prematurely. value.

Common initialization methods include all-zero initialization (all parameters are initialized to 0 or the same constant), random initialization, MSRA, and the use of various function distributions to obtain parameters such as weights, where all-zero initialization ignores the update of parameters during error back propagation, resulting in the death of neurons throughout the network model. Usually in a deep convolutional neural network, we generally use random initialization to assign values to the network weight matrix, and the bias matrix is directly set to zero.

In order to avoid problems such as reduced convergence speed of the deep neural network model due to unsatisfactory parameter initialization, we generally follow the following empirical principles:
(1) The activation mean is zero;
(2) The activation variance remains constant.

4.3 Model training

The data is split into training data and validation data. The ratio of training and validation is 8 to 2.

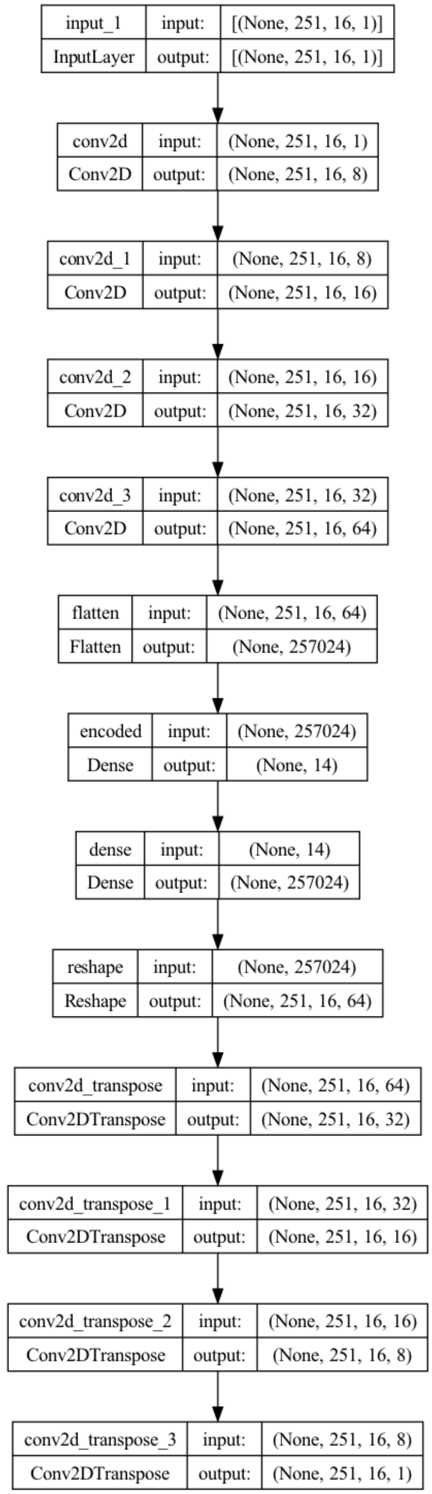

Figure 3. Structure of auto-encoder network

5 Result and Discussion

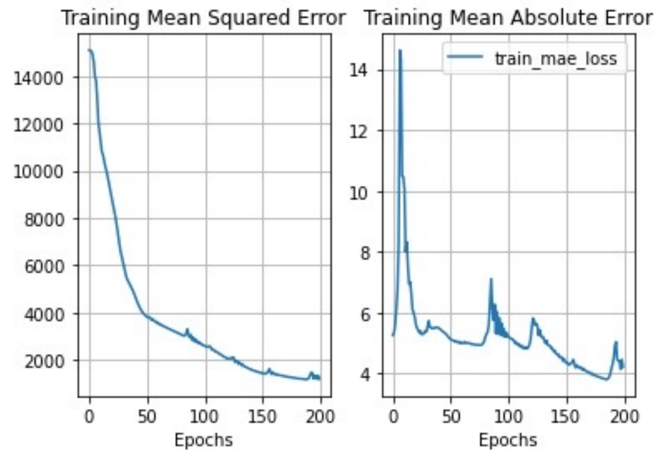

Figure 2. Training MSE and training MAE

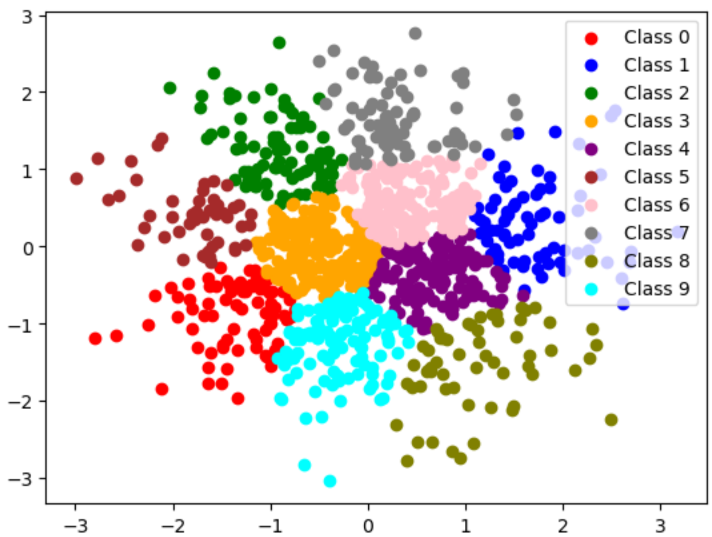

Figure 3. Cluster result

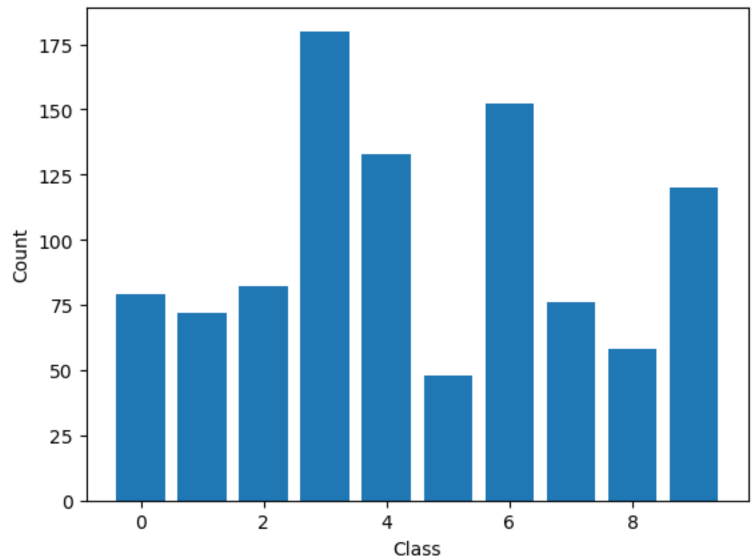

Figure 4. Cluster Statistics

Reference

[1] Ronen, J. Wave-equation trace interpolation[J]. Geophysics, 1987, 52(7): 973-984.
[2] Fomel S. Seismic reflection data interpolation with differential offset and shot continuation[J].
Geophysics, 2003, 68: 733-744.
[3] Spitz S. Seismic Trace Interpolation in the F-X domain[J]. Geophysics, 1991, 56(6): 785-794.
[4] Porsani M J. Seismic trace interpolation using half-step prediction filters[J]. Geophysics, 1999, 64(5): 1461-1467.
[5] Wang Y. Seismic Trace Interpolation in the f-x-y Domain[J]. Geophysics, 2002, 67(4): 1232-1239.
[6] Naghizadeh M, Sacchi M D. F-x adaptive seismic-trace interpolation[J]. Geophysics, 2009, 74(1): V9-V16.
[7] Liu G, Chen X. Seismic data interpolation using frequency domain complex nonstationary autoregression[J]. Geophysical Prospecting, 2018, 66(3): 478-497.

[8] Yu S, Ma J, Zhang X, et al. Interpolation and denoising of high-dimensional seismic data by learning a tight frame[J]. Geophysics, 2015, 80(5): V119-V132.
[9] Jia Y, Yu S, Ma J. Intelligent interpolation by Monte Carlo machine learning[J]. Geophysics, 2018, 83(2): V83-V97.
[10] Kaur H, Pham N, Fomel S. Seismic data interpolation using CycleGAN[C]. SEG Technical Program Expanded Abstracts, 2019: 2202-2206.
[11] Jia Y, Ma J. What can machine learning do for seismic data processing? An interpolation application[J]. Geophysics, 2017, 82(3): V163-V177.
[12] Wang B, Zhang N, Lu W, et al. Deep-learning-based seismic data interpolation: A preliminary result[J]. Geophysics, 2019, 84(1): V11-V20.
[13] Siahkoohi A, Kumar R, Herrmann F J. Deep-learning based ocean bottom seismic wave field recovery[C]. SEG Technical Program Expanded Abstracts, 2019: 2232-2237.
[14] Wang Y, Wang B, Tu N, et al. Seismic Trace Interpolation for Irregularly Spatial Sampled Data Using Convolutional Auto-Encoder[J]. Geophysics, 2020, 85(2): V119-V130.