Hyperspectral Remote Sensing Image Classification Based On Deep Learning

Rui Jiang*
School of Information Engineering, Baise University, Baise 533000, Guangxi, China

*Corresponding author e-mail: jiangrui@bsuc.edu.cn

Abstract. In recent years, with the development of science and technology, in order to further explore the world, we have explored and studied remote sensing technology. As a new discipline, teleology has been widely studied and applied in the fields of spectroscopy, informatics, geography, environmental science and urban construction, and has become one of the most active fields of science and technology. However, both classifier and feature extraction are in the shallow level. How to extract deep features and make them more abstract and easy to classify is a hot issue in machine learning field. Therefore, through the study of different classifiers and various special cases, we find that the post-processing method proposed in this paper can alleviate this situation to a certain extent. When the number of hidden layers is 2, the overall classification accuracy of DBN model is higher. Experiments show that this method is better than the traditional deep learning method.

Key words: Deep Learning, Hyperspectral Remote Sensing Image, Remote Sensing Image Classifier, DBN Model

1. Introduction
For example, some domestic scientific research teams acquired hyperspectral images of Inner Mongolia Desert Steppe through UAV hyperspectral remote sensing platform, and established a desert steppe species classification model based on feature band and convolution neural network. The results provide data support for species coverage estimation, degradation succession judgment and grazing of desert steppe in Inner Mongolia, and provide research basis for ecological restoration of desert steppe [1-2]. Other research groups have found that hyperspectral image target detection performance is better because it mainly uses spectral information rather than spatial information [3-4]. Wang experts also found that there are six commonly used band selection algorithms for hyperspectral image data, including ABS and Mi algorithm based on information content, ID and FSD algorithm based on sorting, WA Lu Di and K-means algorithm based on clustering. Through the experimental data, they evaluated the effectiveness of the commonly used frequency band selection algorithms, and compared their advantages and disadvantages. A band selection method based on clustering and information quantity is proposed. Firstly, K-means clustering method is used to cluster similar bands. Euclidean distance, correlation coefficient, spectral information divergence and spectral angle are used to measure the similarity. Then, the improved ABS method is used to select the representative clustering
bands. In order to improve the information content and classification accuracy of clustering band selection method [5]. Although the research results of hyperspectral remote sensing image classification technology are abundant, there are still some deficiencies. For example, many classification methods only use the spectral information of hyperspectral images and ignore the spatial information.

In order to study the hyperspectral remote sensing image classification technology based on deep learning, this paper studies the overall accuracy of hyperspectral remote sensing image and the influence of different classifiers on the classification effect of hyperspectral remote sensing image. The experimental results show that sdae softmax has better classification effect than other classifiers, and the network model can obtain better classification accuracy. At the same time, it shows that the classification method based on DBN has better classification effect than SVM and traditional neural network.

2. Method

2.1. Deep Learning

Deep learning is a branch of machine learning. It is based on a series of nested layer based data extraction algorithms, which use complex structures or multiple nonlinear transformations to form multiple network layers. Deep learning is an expression learning method in machine learning. Features learned from data are often easier to classify than manually designed features [6-7]. There are many kinds of deep learning framework, such as search self coding neural network, convolution neural network, deep confidence network and so on. Intelligent self coding neural network is a self coding network based on single layer unsupervised training. It can stack multiple self coding networks into multi-layer networks, and can be used as the initial weight of ordinary multilayer perceptron. Scaffolding self coding can be said to be a better algorithm for training multilayer perceptron. Convolution neural network is mainly used for two-dimensional data, and the hidden layer is connected by convolution. Deep trust network is based on probability theory. It includes unsupervised learning method and supervised learning method.

2.2. Hyperspectral Remote Sensing Image Classification Technology

As the basic research of hyperspectral images, the main purpose of classification is to use the spectral information and geometric information of the target to divide each pixel into different graphic categories. There are many different hyperspectral image classification algorithms, and traditional methods only use the spectrum of the image. The existing size reduction methods are usually divided into function extraction and frequency band selection. The band selection algorithm mainly selects several representative bands from hundreds of bands to remove the remaining bands to achieve the effect of dimensionality reduction. It does not have the projection transformation of the data. Understand the physical meaning of the data. The most important frequency band selection algorithm is based on the amount of information and the separability of classes and subspaces [8-10].

2.3. Realization Process of Deep Learning

The traditional neural network is composed of input layer, hidden layer and output layer. Features are extracted from the input layer layer by layer, and the feature expression of the layer is obtained by using the nonlinear activation function as the input of the next layer. Since the initialization weights of each layer of the model are randomly selected, a large number of weight parameters w and B need to be trained before the formal application of neural network, so that the overall performance of the network can meet the needs of specific projects.Suppose that the number of sample sets is m and the input is $X$, output is $h_{wb}(x)$,The expected output is $y$, use $a^{(l)}$ to represent the output value of the i cell in layer L. the relationship between them is shown in the following formula.
\[ a^{(l)} = f\left(\sum_{j=1}^{S_{l-1}} w^{(l)}_j x_j + b^{(l)}_j\right) \]  

(1)

\[ S_{l-1} \] represents the number of neurons in the L-1 layer. For the conventional 4-layer network, the error of the last layer is represented by \( \delta \)

\[ \delta^{(4)} = a^{(4)} - y \]  

(2)

The error value calculated by the formula is used to calculate the error of the upper layer:

\[ \delta^{(3)} = ((w^{(3)})^T \delta^{(4)}) f(z^{(3)}) \]  

(3)

\( f(z^{(3)}) \) is the derivative of the activation function. If the Sigmoid function is selected as the activation function, then the \( f(z^{(3)}) \) value is

\[ f(z^{(3)}) = a^{(3)}(1 - a^{(3)}) \]  

(4)

3. Experience

3.1. Extraction of Experimental Objects

The spectral dimension of hyperspectral remote sensing image is the spectral response value obtained by hundreds of continuous narrow-band reflections, so the adjacent spectral feature points have strong correlation and data redundancy, so we consider extracting low-dimensional robust features from high-dimensional data space. Whether it is linear transformation or nonlinear transformation, the dimension reduction of spectral features will destroy the physical characteristics of the original data to a certain extent, so band selection has been widely used in the field of remote sensing. Band selection, as the name implies, is to select an appropriate number of optimal bands from all bands as features. Because there are some overlaps between hundreds of bands, there is a certain correlation between adjacent bands. Not all bands are equally important for classification. Therefore, the purpose of this paper is to select the appropriate band for subsequent classification through certain band selection criteria.

3.2. Experimental Design

The CNN model structure extracted by feature method is divided into three steps: the first step is image preprocessing, including edge preserving relaxation filtering and pixel block extraction; the second step is spectral spatial feature extraction; the third step is feature fusion classification. Because SPE CNN is mining the effective information in the original spectrum, the lack of spatial information, if not denoising in advance, excessive noise interference will more or less affect. On the other hand, it can make up for the lack of spatial information of each pixel, or extract the spatial features of each pixel better than the single feature in the space. In order to verify the effectiveness of the model, two public data sets of Pines India and Pavia University were selected for comparative experiments. Then the noise interference features are denoised to improve the classification accuracy. If the noise is not preprocessed, the noise is too large, which makes the neural network unable to learn the features we want, so the classification accuracy is very low. Finally, a comparative study is carried out.

3.3. Statistical Analysis of Data

Mathematical statistics: use Excel data processing software to conduct statistical processing of relevant data, and present in the form of charts. The formula is as follows:

\[ \text{SUMIF} \left( \$A$2 : $G$2, $H$2, $A$3 : $G$3 \right) \]

4. Discussion
4.1. Influence of Network Depth and Hidden Layer Node Number on Overall Accuracy
For deep learning model, network depth plays an important role in classification. On the one hand, by increasing the number of layers of RBM, the modeling ability of DBN can be improved, and more abstract feature table systems can be found by higher level representation, so as to improve the classification performance. On the other hand, too many hierarchies will reduce the generalization performance of DBN, resulting in over fitting. As a result, the appropriate number of DBN layers is often related to specific applications and datasets.

Table 1. DBN parameter settings for experiments on SAR data

| Pretraining stage          |       |
|----------------------------|-------|
| Learning rate              | 0.02  |
| Number of epochs           | 60    |
| Size of minibatch          | 120   |
| Momentum                   | 0.5 for the first epochs, 0.9 thereafter |
| Weight decay rate          | 0.0004|
| Fine - tuning stage        |       |
| Learning rate              | 0.05  |
| Number of epochs           | 25    |

It should be noted that the network depth refers to the number of hidden layers; meanwhile, for convenience, the number of nodes in each hidden day is set to the same number. In the experiment, the number of hidden layer m increases from 1 to 4, and the number of hidden layer nodes increases from 100 to 600. It can be seen from the results in Fig. 2 that the DBN model with 2 hidden layers has high classification accuracy. At different depths, when the number of hidden layer nodes is 500, the classification accuracy is the highest.
Figure 2. The influence of network depth and hidden layer node number on the overall accuracy

4.2. Comparison with Other Classifiers

In order to compare the performance of classifiers, we compared them with several different classifiers on Indian pines and Pavia University datasets. In the experiment, all the spectral information of these two hyperspectral remote sensing data sets is used as input. Table 2 shows the comparison of stack noise reduction self encoder Softmax with RBF kernel support vector machine, stack automatic encoder logic regression, k-nearest neighbor and naive Bayes on Indian pines and Pavia University datasets. The overall classification accuracy and kappa coefficient table of the classifier are given.

| classifier       | Indian Pines                        | Pavia University                       |
|------------------|-------------------------------------|----------------------------------------|
|                  | Overall accuracy classification    | Kappa coefficient | Test time(s) | Overall accuracy classification | Kappa coefficient | Test time(s) |
| SDAE_Softmax     | 88.08                               | 0.8668                          | 0.0404       | 95.05                            | 0.9353               | 0.2824       |
| RBF-SVM          | 87.60                               | 0.8525                          | 1.4745       | 93.72                            | 0.9173               | 14.5833      |
| SAE_LR           | 87.09                               | 0.8450                          | 0.321        | 94.50                            | 0.9330               | 0.2331       |
| KNN              | 74.22                               | 0.6701                          | 2.4051       | 85.66                            | 0.8111               | 29.0531      |
| Native Bayes     | 50.01                               | 0.4063                          | 0.1749       | 66.55                            | 0.5730               | 0.5626       |

As can be seen from table 2, SDAE_U2;SOFTMAX, SAE_LR AND RBF-SVM are better than KNN and native Bayes. Softmax and SAE_ As a deep learning network, LR can learn deep nonlinear features through activation function. Compared with the nonlinear RBF kernel SVM classifier with optimal parameters, the Sdae Softmax used in this paper can achieve better classification accuracy and kappa coefficient. In terms of test time, Sdae is the best choice. SOFTMAX and SAE_ The test speed of the three methods is better than that of other methods. SDAE_U-SOFTMAX is not only superior to other algorithms in overall classification accuracy, but also superior to other methods in almost all specific categories. Because only spectral information is used for classification, there are many noise points in the classification result map. Compared with LR, RBF-SVM, KNN and native Bayes, the results of the first three classifiers are significantly close to the actual ground markers.
5. Conclusion
Hyperspectral remote sensing image classification is an important way to obtain the information of ground objects. Although hyperspectral remote sensing images have hyperspectral components, the classification of hyperspectral remote sensing images has been widely affected. However, due to the correlation between adjacent bands, there is a high-dimensional spectral function. In addition, the image acquisition process is also affected by many external factors, resulting in a lot of noise in the original spectral information. Therefore, in the early classification technology, classification technology based on spectral size information can not achieve good classification results. Compared with some mainstream deep learning methods, this paper uses the denoising method to denoise the image before the model training, which reduces the excessive attention of the model to the image noise, ignores the real distribution of pixels, and improves the convergence time of the model. Aiming at the problem of over fitting of deep learning with few training samples in the field of hyperspectral remote sensing image classification, a post-processing method is proposed. To some extent, this situation has been alleviated. Experimental results show that this method makes full use of spectral spatial information in the process of feature extraction, preprocessing and post-processing, and is superior to the traditional deep learning method.

Reference
[1] Bob V D M, Kooistra L, Suomalainen J, et al. Remote sensing of plant trait responses to field-based plant-soil feedback using UAV-based optical sensors[J]. Biogeoences, 2017, 14(3):733-749.
[2] Ning Z, Xiaoli Z, Guijun Y, et al. Assessment of defoliation during the Dendrolimus tabulaeformis Tsai et Liu disaster outbreak using UAV-based hyperspectral images[J]. Remote Sensing of Environment, 2018, 217:323-339.
[3] Li H, Li C, Zhang C, et al. Hyperspectral Image Classification with Spatial Filtering and $\ell^2,1$ Norm[J]. Sensors, 2017, 17(2):314.
[4] Lingyan R, Yanning Z, Wei W, et al. A Hyperspectral Image Classification Framework with Spatial Pixel Pair Features[J]. Sensors, 2017, 17(10):2421-.
[5] Wang Q, Lin J, Yuan Y. Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 27(6):1-11.
[6] Shen D, Wu G, Suk H I. Deep Learning in Medical Image Analysis[J]. Annual Review of Biomedical Engineering, 2017, 19(1):221-248.
[7] Litjens G, Kooi T, Bejnordi B E, et al. A Survey on Deep Learning in Medical Image Analysis[J]. Medical Image Analysis, 2017, 42(9):60-88.
[8] Liu Y, Wang C, Wang Y, et al. Differential weights-based band selection for hyperspectral image classification[J]. International Journal of Wavelets Multiresolution & Information Processing, 2017:S0219691317500655.
[9] Tang Y D, Huang S C, Xue A J. Sparse Representation Based Band Selection for Hyperspectral Imagery Target Detection[J]. Tien Tzu Hsueh Pao/Acta Electronica Sinica, 2017, 45(10):2368-2374.
[10] Villmann T, Bohmsack A, Kaden M. Can Learning Vector Quantization be an Alternative to SVM and Deep Learning? - Recent Trends and Advanced Variants of Learning Vector Quantization for Classification Learning[J]. Journal of Artificial Intelligence & Soft Computing Research, 2017, 7(1).