ISSN: 2688-8645

Comparative Study of Land Use Classification in Yan 'an City Based on Random Forest and Neural Network

Dan Wu[1, 2, 3, 4], Liangyan Yang[1, 2, 3, 4]

[1]Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd, Xi'an 710075, China

[2]Shaanxi Provincial Land Engineering Construction Group Co., Ltd, Xi'an 710075, China

[3]Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of natural Resources, Xi'an 710075, China

[4]Shaanxi Provincial Land Consolidation Engineering Technology Research Center, Xi'an 710075, China

Abstract

Accurate land use information is the basis of land resources monitoring and management. Random forest method has become an effective method in remote sensing classification machine learning. It is of great significance to explore the application of random forest classification method and neural network method for object-oriented land use classification based on Landsat OLI image. Based on Landsat8OLI multispectral data, this paper uses random forest classification method and neural network classification method to classify land use types in Yan 'an city. The results showed that :(1) the overall accuracy and Kappa coefficient of random forest classification in the study area were better than those of neural network classification, with the overall accuracy of 94.8043% and 90.7515%, and the Kappa coefficient of 0.9261 and 0.8573, respectively. (2) The classification accuracy of grassland and agricultural land under the neural network classification method in the study area is higher; Construction land, forest land and water area are more suitable for random forest classification. (3) In 2020, the land use of the study area is mainly woodland, followed by grassland, agricultural land and construction land, and the area of water area is relatively small. The object-oriented land use classification of medium resolution remote sensing images has good applicability and can provide reference for the extraction of regional land use information.

Keywords

Land use classification; Random forest; Neural network; Accuracy evaluation.

1. Introduction

Land use/land cover change has always been regarded as an important content of global environmental change and sustainable development research [1-3], and it is also a hot topic of current global change research, while land use/land cover classification is the basis of LUCC research [4-6]. Land use classification plays an important role in adjusting land use structure, rationally developing land resources and dynamically monitoring land use status [7]. At present, remote sensing technology is considered to the most rapid and most effective means to obtain information of land use and its spatial distribution [8], it has a quick update cycle, can study range, objective, and the advantages of mature technology development, has become a very useful to understand the surface of the earth changes the data sources, land use and land cover classification is a basic work of remote sensing. Land use classification combined with remote

ISSN: 2688-8645

sensing data and machine learning algorithm has always been a research hotspot of scholars at home and abroad, such as random forest, maximum likelihood, neural network, support vector machine and other methods have been widely used, and many scholars have perfected and improved the shortcomings of various classification methods, and achieved good classification effects [9-11]. Random forest algorithm is a new and efficient combinatorial classification method. Its superior performance has been widely used in many fields abroad, but there are few researches and applications in China. In recent years, the research results of land cover change mainly focus on land use model simulation and prediction [12], land use and cover change and its driving mechanism [13], and land use landscape pattern change [14]. Wang Shuyu et al. [15] studied the classification of honghe wetland remote sensing image based on random forest, and the results showed that it had high accuracy in wetland information extraction. Wang Na et al. combined random forest and univariate feature selection algorithm to carry out classification extraction of major crops in northern Jiangsu [16]. Based on the random forest algorithm and neural network algorithm, this paper classifies the land use types of Yan 'an city by using landsat8 remote sensing image data. Finally, the random forest is compared with the neural network to evaluate the practicality of the multi-source information comprehensive classification scheme based on the random forest algorithm in land use classification. It provides basis for monitoring land use and planning and managing land resources.

2. Data Source and Pre-processing

2.1. Data Sources

In this paper, Landsat8 multi-spectral data with a spatial resolution of 30m were adopted. In order to ensure the classification effect, images with cloud cover less than 3% in July-August 2020 were mainly used. The data came from geospatial Data Cloud Platform (http://www.gscloud.cn/) of Computer Network Information Center, Chinese Academy of Sciences. Data are processed on the basis of original data and can be applied to relevant research.

2.2. Data Preprocessing

Landsat8 OLI multispectral image data: Firstly, radiometric calibration was performed, and then atmospheric correction Flash model was used for atmospheric correction. The corrected image was Mosaic, and finally, the image was clipped according to the scope of the study area. Due to the inconsistent acquisition time of remote sensing images, in order to ensure the small color difference of images during Mosaic, image enhancement processing mainly adopts radiation enhancement and spectral enhancement. Radiation enhancement is achieved by transforming the gray value of a single pixel.

3. Research Methods

3.1. Random Forest

Random forest is a machine learning method proposed in 2001. N samples are randomly put back and extracted from the original training samples by bootstrapping method to form a training sample set. Each training sample set can construct a decision tree model, so that N decision tree models can be generated. M feature variables are randomly selected at the split node of each decision tree as predictive variables, and the optimal feature variable is selected as the split feature of the split node according to the minimum principle of node impurity, until the decision tree stops growing at the minimum impurity of each split node [17]. After all decision trees are generated, the classification results of all decision trees are integrated by voting method and the final result is obtained [18]. Every tree in the forest depends on a

ISSN: 2688-8645

random vector, and all vectors in the forest are distributed independently. The prediction accuracy of random forest is related to the strength of individual trees and the correlation between trees. Two key parameters need to be set when applying random forest model classifier :(1) number of spanning trees, which determines the overall size of random forest. The value is generally between 100 and 10000. The larger the value is, the more convergent the model will be, and the running time of the model will increase. In addition, when the number of trees keeps increasing, the model will be oversaturated. (2) Random extraction of the maximum number of features (Max features), which represents the maximum number of features randomly extracted from the feature space when each decision tree is generated. The larger the value is, the stronger each decision tree in the model will be, but the correlation between decision trees will also increase. Therefore, Max Features needs to be tuned according to the OOB error rate to achieve certain accuracy.

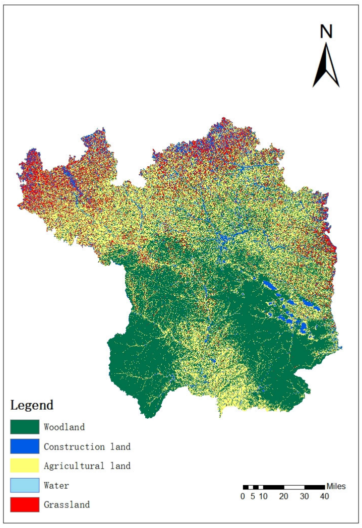
3.2. Neural Network

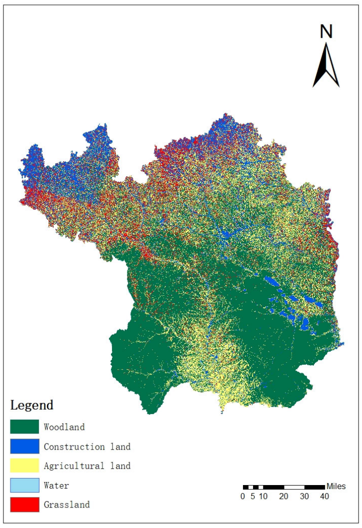
Neural network refers to the use of computer to simulate the structure of human brain, with many small processing units to simulate biological neurons, with algorithms to achieve the recognition of human brain, memory, thinking process. Neural Net Classification is a kind of supervised Classification and has been applied in many fields of remote sensing technology [19]. At present, there are two main types of neural networks, which are feedforward neural networks: convolutional neural network (CNN) and recurrent neural network (RNN). Although neural network classification has ideal anti-interference and adaptability, its algorithm training cost is very high. With the progress of computer technology and the enhancement of computing power in the era of big data, neural network can achieve ideal classification effect once it meets the strict requirements of effective data and computing resources.

4. Result Analysis

4.1. Comparison of Classification Results

According to the selected samples and ENVI5.3 software, this paper classifies land use types in Yan 'an city by using random forest classification method and neural network classification method respectively. Based on The Classification of Land Use Status (GB/T21010-2017) [20], land use types are divided into five types: forest land, construction land, agricultural land, water body and grassland. The specific classification results are shown in Figure 1.





a.Random forest classification results b.Neural network classification results

Figure 1. Comparison of land use classification results

According to the comparison in Figure 1, The spatial distribution of the ground objects in result 5 of the two classification methods is similar, but there are some differences in the distribution area of the same ground objects. Compared with the neural network classification method, the classification accuracy of random forest classification for grassland and agricultural land is relatively poor. The results of forest land and water body classification are similar and the classification accuracy is high. In construction land classification, neural network fault the grass can be divided into construction land area is larger, lead to neural network's overall classification accuracy is not high, the classification effect is not ideal, so the random forest classification method is more suitable for yan 'an land use classification, better classification effect, classification of feature separability, lower degree of confusion.

4.2. Comparison of Classification Accuracy

The precision analysis is obtained by testing the classification results obtained from the training samples with the validation samples. The evaluation of the overall classification accuracy of land use is to test the classification results. There are two commonly used methods. One is to compare the classification results with the actual land use status classification data; the other is to use high-resolution remote sensing images to select a certain proportion of training samples. To test the accuracy of Landsat8 classification data. The most commonly used evaluation coefficients are the overall accuracy coefficient (Overall Accuracy) and the Kappa coefficient. In the image selection of the supervised classification method, the image maps of the study area where the ground objects are easy to distinguish, and the image maps of the study area with large differences in the types of ground objects are selected. On the other hand, in terms of resolution selection, the image map of the study area with higher resolution is selected. These external factors all contribute to the classification results. The random forest and neural network classification methods are used for classification, and the confusion matrix is calculated according to the classification results to obtain the overall accuracy and Kappa coefficient. The classification accuracy table is shown in Table 1.

Table 1. Classification accuracy comparison

Classification	Overall Accuracy	Kappa Coefficient
Random Forest	94.8043%	0.9261
Neural Networks	90.7515%	0.8573

It can be seen from Table 1 that the overall accuracy of the two classification methods is above 90%, and the overall classification accuracy is 94.8043% and 90.7515% respectively, indicating that the overall classification accuracy is relatively high; the Kappa Coefficient of random forest classification reaches 0.9261 , which is nearly 7% higher than that of the neural network, indicating that the random forest classification method is more suitable for the classification of land use in Yan'an area.

5. Conclusion

In this study, according to the current situation of land use in Yan'an City, according to the principle of random forest and neural network method, and supported by ENVI software, the classification results and accuracy of the method are found: 1) The classification accuracy of random forest and neural network method is high 80%, 92.61% and 85.73% respectively, but the random forest classification is nearly 7% more accurate than the neural network classification; 2) The spatial distribution of the 5 ground objects obtained in the classification of the two classifications is similar, but the distribution area of the ground objects is the same There are certain differences. The classification of land use in this paper also has great shortcomings. Only identification training In this study, due to the particularity of this area, it determines the limitations of its application, so the specific method needs to be further improved, but in the remote sensing extraction of land use classification, this classification The extraction methods and ideas are still feasible through practice verification.

Acknowledgments

This paper was supported by Shanxi Province Land Enginaeering Construction Group Project DJNY2022-28.

References

[1] Liu Jiyuan, Liu Mingliang, Zhuang Dafang, et al. Spatial pattern analysis of recent land use change in China [J]. Chinese Science: Series D, 2002, 32(12):13.

[2] Ii BLT , Lambin EF , Reen Be Rg A . The emergence of land change science for global environmental change and sustainability[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(52) :20666-20671.

[3] Liu Jiyuan, Zhang Zengxiang, Xu Xinliang, et al. Analysis of the spatial pattern and driving force of land use change in China in the early 21st century [J]. Chinese Journal of Geography, 2009.

[4] Tang Huajun, Wu Wenbin, Yang Peng, et al. Research progress of land use/land cover change (LUCC) model[J]. Chinese Journal of Geography, 2009, 64(4):13.

[5] Zhai Tianlin[1,3] Jin Gui[1,2,3] Deng Xiangzheng[2] Li Zhaohua[1,3] Wang Run[1,3]. Research on land use classification method in Wuhan based on multi-source remote sensing image fusion[1,2] J]. Yangtze River Basin Resources and Environment, 2016, 25(10):9.

[6] Shi Lijiang, Wang Shengyun, Yao Xiaojun, et al. Analysis of temporal and spatial variation characteristics and driving forces of land use in Shanghai from 1994 to 2006 [J]. Yangtze River Basin Resources and Environment, 2012, 21(12):1468.

[7] Lin Nan, Jiang Qigang, Yang Jiajia, et al. Classification of land use in agricultural areas based on high-resolution data of ZY-1[J].Journal of Agricultural Machinery,2015, 46(001):278-284.

[8] An Jianke. Research on land use change and landscape pattern in Suzhou District of Jiuquan City based on RS and GIS [D]. Northwest Normal University, 2011.

[9] Zhao Jianhua. Semi-supervised classification algorithm based on SOM neural network [J]. Journal of Xihua University: Natural Science Edition, 2015, 34(1):6.

[10] Meng Zhiguo. Application analysis of BP neural network in land use classification [D]. Jilin University, 2004.

[11] Ma Yue, Jiang Qigang, Meng Zhiguo, et al. Research on land use classification in farming areas based on random forest algorithm [J]. Journal of Agricultural Machinery, 2016, 47(1):7.

[12] Liu Miao, Hu Yuanman, Sun Fengyun, et al. Application of land use model CLUE-S in urban agglomeration planning in central Liaoning Province [J]. Journal of Ecology, 2012.

[13] Zhang Tao, Zhang Ying, Yang Lipeng, et al. Landscape pattern changes and their driving forces in Hohhot, Inner Mongolia Autonomous Region from 1990 to 2010 [J]. Bulletin of Soil and Water Conservation, 2018(2).

[14] Liu Juan. Research on the dynamic change and driving force of landscape pattern in Mu Us Sandy Land in Inner Mongolia [D]. Inner Mongolia University, 2016.

[15] Wang Shuyu, Zhang Yuwei, Yu Zhenhua. Classification of Honghe Wetland Remote Sensing Image Based on Random Forest [J]. Surveying and Mapping and Spatial Geographic Information, 2014, 000(004):83-85.

[16] Wang Na. Research on spatiotemporal effects of remote sensing identification features of major crops in Jiangsu Province [D]. University of Chinese Academy of Sciences (Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences).

ISSN: 2688-8645

[17] Jennifer D. Watts and Scott L. Powell and Rick L. Lawrence and Thomas Hilker. Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery[J]. Remote Sensing of Environment, 2011.
[18] Zhou Zhenglong, Sha Jinming, Fan Yuexin, et al. Remote sensing land use classification and landscape pattern analysis based on random forests [J]. Application of Computer Systems, 2020, 29(2):9.
[19] Chen Guifen, Zhao Shan, Cao Liying, et al. Maize plant disease identification based on transfer learning and convolutional neural network [J]. Smart Agriculture, 2019, 1(2):11.
[20] GB/T201010-2017. Classification of land use status [S]. 2017.