Automatic Analysis And Intelligent Information Extraction Of Remote Sensing Big Data

Ge Li1, 2, Li Jiajun1
1 Management Department, Northwestern Polytechnical University; Xi'an Shaanxi, 710072, P.R.China
2 Computer College, Weinan Normal University; Weinan Shaanxi, 714099, P.R.China;

Abstract: Aiming at the problems of low analysis accuracy, incomplete information extraction and low quality of traditional remote sensing data automatic analysis and extraction methods, a new remote sensing big data automatic analysis and intelligent information extraction method is proposed. Automatic analysis of remote sensing big data is realized through the expression, retrieval and understanding of remote sensing big data, and intelligent information extraction is carried out on the basis of automatic analysis of remote sensing big data. According to the procedures of remote sensing data mining, data preprocessing, feature acquisition, target recognition and evaluation, intelligent information extraction can be realized by utilizing the good learnability of convolutional neural network. In order to verify that the proposed method is superior to the traditional method, six different experimental areas are used as experimental objects, and a comparative experiment is designed. Experimental results: The proposed method automatically analyzes and extracts the remote sensing data, which is superior to the traditional method in terms of integrity, correctness and quality.

1. Introduction
With the development of remote sensing technology, the spatial resolution, temporal resolution, spectral resolution and radiation resolution of remote sensing data are getting higher and higher, and the data types are getting richer and richer. At the same time, the amount of data is getting larger and larger, which poses a new challenge to the traditional automatic analysis of remote sensing data and intelligent information extraction. Large data has the characteristics of large volume, complex types, strong timeliness, difficult to distinguish between true and false, and great potential value. In the case of no contact with objects, remote sensing technology obtains the relevant characteristics of research targets through sensors and extracts effective information from them, such as artificial buildings, land use types, vegetation, temperature, and other interested targets [1].

Remote sensing information extraction aims to obtain thematic information from image data to meet the needs of industry applications. Its research object is geographic entities and related phenomena existing in surface space. With the development of earth observation technology, in the field of human remote sensing and earth observation, the comprehensive observation ability of the earth has reached an unprecedented level. Different imaging methods, different bands and resolution data coexist, making the remote sensing data increasingly diversified; The amount of remote sensing image data increased exponentially. The speed of data acquisition is accelerated, the update cycle is shortened, and the timeliness is stronger and stronger. Remote sensing data show obvious "big data" characteristics. However, in sharp contrast to the ability of remote sensing data acquisition, the ability...
of remote sensing information processing is low. There is a serious imbalance between remote sensing information processing technology and data acquisition capability. Remote sensing information processing still stays in the stage from “data to data”. It is obviously insufficient in realizing the transformation from data to knowledge, and the utilization rate of remote sensing big data is low, caught in the paradox of "big data, small knowledge" [2]. What's more, because a large amount of accumulated data cannot be effectively utilized, massive data will occupy limited storage space for a long time, which will cause a certain degree of "data disaster". The ultimate goal of using remote sensing big data is to mine the hidden knowledge in remote sensing big data.

2. Automatic analysis and intelligent information extraction of remote sensing big data

According to the requirements of current development of remote sensing big data, a new method for automatic analysis and intelligent information extraction of remote sensing big data is proposed. The block diagram of automatic remote sensing big data analysis and intelligent information extraction is shown in figure 1. On the basis of realizing automatic analysis of remote sensing big data through data expression, retrieval and understanding, remote sensing big data mining is carried out, and then the intelligent information extraction is completed by using convolutional neural network.

Fig.1 Block diagram of automatic analysis and intelligent information extraction of remote sensing big data

2.1 Automatic analysis of remote sensing big data

Automatic analysis is the premise of remote sensing big data information extraction and data to knowledge transformation. Therefore, the establishment of a unified, compact and semantically clear automatic analysis system can lay a good foundation for the follow-up remote sensing big data mining and utilization [3]. Automatic analysis of remote sensing big data mainly includes data expression, retrieval and understanding.

2.1.1 Express remote sensing big data

With the continuous emergence of big data of remote sensing for earth observation, the new characteristics such as the complexity of its semantics, the richness of data dimension semantics, and the diversity of sensor semantics make the traditional way of expression unable to meet the needs of practical application. Observation data of different granularity, time phase, azimuth and level of the same feature can be regarded as the projection of the feature in different observation Spaces. Therefore, feature extraction of remote sensing big data needs to consider the unique feature expression model of multi-source and multi-resolution images, as well as the relationship between features and the mutual transformation of models [4]. The core problem of remote sensing big data expression is to study the feature calculation method of remote sensing big data, extract the intrinsic representation of multiple features from the low-level features such as spectrum, texture and structure, and bridge the semantic gap from local features to target characteristics, and then establish the integrated target expression model of remote sensing big data. The research contents mainly include:

Multivariate discrete feature extraction of remote sensing big data: in the framework of big data, it is necessary to study the feature extraction of remote sensing images with multiple resolutions,
multiple data sources and multiple space-time spectra, so as to form a discrete and multivariate feature extraction method of remote sensing big data in different sensor nodes.

Normalized expression of multivariate features of remote sensing big data: feature extraction of remote sensing big data needs to consider the fusion of multivariate discrete features and dimensionality reduction. Feature fusion aims to unify multiple features into the same distinguishing feature space and assimilate discrete features with different sources and resolutions into the application space of big data by means of data transformation [5]. At the same time, the purpose of multivariate feature dimension analysis is to reduce the dimensionality of high-dimensional mixed feature space of remote sensing big data and form normalized low-dimensional feature nodes and data manifolds, so as to improve the efficiency of big data processing.

2.1.2 Retrieval of remote sensing big data

The application of remote sensing big data is developing towards the direction of networking and integration. National plans for spatial data infrastructure have also been developed around the world to provide access to and download services for elevation, orthophoto, hydrology, administrative boundaries, transport networks, cadastre, land control and various thematic data by means of networks. For example, the space information portal established by the US government aims to establish a one-stop geographic space site to improve the work efficiency of the government and provide space information services for the public, which facilitates the access to information to a certain extent. However, this service mode mainly provides data download by means of directory search, which is far from enough for data processing and analysis, and it is difficult to realize on-demand services for user needs. The existing service chains of geographic information and remote sensing data are also difficult to carry out adaptive processing of task demand changes and dynamic environment changes, and it is also difficult to carry out service collaborative optimization under the condition of task concurrency [6].

In order to retrieve data that meets users' needs and interests from massive remote sensing big data, the similarity and heterogeneity between data must be measured. On this basis, the efficient organization, management and retrieval of remote sensing big data can realize the fast retrieval of interested targets from multi-source and multi-mode data and improve the utilization efficiency of remote sensing big data. At present, the retrieval of remote sensing scene data basically realizes image feature-based search. However, in the remote sensing big data, there is a large amount of redundancy and similarity in different observation data of the same feature. How to make use of the redundancy information, study the similarity or difference of images, fully mine the semantic information of images, and effectively improve the retrieval efficiency is the key problem in the use of remote sensing big data.

The traditional remote sensing image retrieval method only for a certain type of image has been difficult to be applied to the retrieval of remote sensing big data. As shown in figure 2, the development of knowledge-driven remote sensing big data retrieval method is one of the effective approaches, which mainly includes: establishment of scene retrieval service chain: Due to remote sensing image describes the surface information, there is no clear or single subject information, and diversity of the sensor and imaging conditions led to the diversification of remote sensing image, therefore, need to be in remote sensing image semantic feature extraction, object recognition, scene recognition and on the basis of autonomous learning, aiming at the characteristics of different types of remote sensing data, establish a scenario for the data type and is used to retrieve service chain, access to different types of remote sensing data that are Shared by geological knowledge, to provide a knowledge retrieval multi-source heterogeneous data [7].
Fig. 2 remote sensing big data scene retrieval strategy

Multi-source massive complex scene data intelligent retrieval system: the massive scene data intelligent retrieval system retrieves multi-source massive remote sensing data based on the information to be retrieved (text description, scene image, etc.) given by the user, and quickly returns the scene required by the user.

Knowledge updating method integrated with user perception information: as a supervised autonomous learning method, correlation feedback technology is an important means to improve image retrieval performance in content-based image retrieval. Relevance feedback is a kind of the through user feedback on the search results, the real-time of the associated low-level features and high-level semantic mechanism, its basic idea is: query, first by the system to provide the query results, then the user feedback to their satisfaction with the degree of the system, so as to exercise and improve the system of learning ability to simulate human perception of image, achieve the goal of high-level semantic retrieval.

2.1.3 Understand remote sensing big data

The main goal of remote sensing big data science is to realize the transformation from data to knowledge, so the semantic understanding of remote sensing big data scenes is crucial. At present, the processing of remote sensing scene data has basically realized the transition from "pixel oriented" to "object oriented" processing mode, and can realize object layer - target layer target extraction and recognition [8]. However, due to the semantic gap between the underlying data and high-level semantic information, the lack of cognition of the relationship between the target and the target, and the relationship between the target and the scene, leads to the insufficient ability to use the scene information acquired in the process of target recognition. As shown in figure 3, in order to achieve high-precision extraction of high-level semantic information of remote sensing big data scene, on the basis of feature extraction and data retrieval of remote sensing big data, the following contents should be studied:

Feature-target-scene semantic modeling: in order to realize the scene semantic understanding of remote sensing big data and overcome the semantic gap in scene understanding. It is necessary to study the semantic model of feature-target-scene from three directions: target-scene relationship model, feature-visual vocabulary model, and feature-target-scene integration model.

Scene multivariate cognition of remote sensing big data: taking multi-source, multi-scale and other multi-features as input and feature-target-scene semantic model as the basis, this paper studies the
2.2 Intelligent information extraction from remote sensing big data

On the basis of automatic analysis of remote sensing big data, intelligent information extraction is carried out. The target of remote sensing information extraction is to extract the useful information contained in the image data, including the recognition of geographic targets and the correlation between geographic targets. Remote sensing information extraction generally involves two types of objects. One is the target to be studied, that is, geographical entities distributed in the surface space and related geological phenomena are information sources, that is, remote sensing image data acquired. The essence of remote sensing information extraction is to find out the relationship between these two types of objects and establish an effective information transformation model [9].

Remote sensing information extraction as the inverse process of remote sensing data acquisition, the key is found on the image to express a variety of characteristics of geographical entity, through the characteristics of form a complete description of geographical entity, thus achieving of geographical entity recognition, its process as shown in figure 4, including remote sensing data mining, data preprocessing, feature acquisition, target identification, five steps.

Fig. 4 remote sensing information extraction flow chart

In the data preprocessing stage, various errors and gaps in the original image data are eliminated. In addition, data should be prepared according to the requirements of feature extraction. For example, in order to highlight the image feature information, principal component transformation can be carried out to extract the main feature information band.

Feature acquisition is to obtain all kinds of feature information required by target recognition, such as vegetation index transformation, mean variance, etc. through attribute extraction or transformation from remote sensing data. The goal of feature acquisition is to obtain features with the following properties. That is, the characteristics of different samples from the same category should be very
similar, while the characteristics of samples from different categories should be very different, and it is best to show the correlation of categories with the smallest feature combination.

Target recognition is based on access to features, according to the features or combination will stay on the image object recognition is converted to a specific entity objects depends on two factors of difficulty of target recognition came from the same category features between different individual eigenvalues of fluctuations, such as remote sensing images obtained because of the influence of atmospheric or sensor, causing the same feature in image do not match the spectrum characteristics, on the other hand is from the differences of characteristics between different category feature size, characteristics of difference, the greater the features, the better the discriminant.

Evaluation is to compare the recognition results with the knowledge described by experts and the knowledge of geotechnical regularity, analyze its reliability and accuracy, so as to guide and improve the process of target recognition.

2.2.1 Mining remote sensing big data information

The whole process of big data mining includes data acquisition and storage, data processing and analysis, data mining, data visualization and data fusion, etc., all of which have the characteristics of big data, as shown in figure 5.

![Fig. 5 remote sensing big data information mining](image)

Remote sensing image data mining technology is a technology that uses data image analysis technology, pattern recognition, artificial intelligence, geographic information system and spatial data mining to discover and mine information hidden in remote sensing images. It is an application of data mining technology in the field of remote sensing images. It involves many new technologies such as image processing, spatial database, spatial data analysis, expression and visualization, spatial data mining, geographic information technology, information extraction, machine learning and software design. It is an important branch of the research field of spatial data mining technology. The difficulty in the design and implementation of remote sensing image data mining lies in the complexity of remote sensing image data. The research of remote sensing image data mining needs to consider not only the visual features of images, but also the extraction, storage and management of semantic features and related knowledge. Previous studies have focused on the comprehensive extraction and analysis of image semantic and visual features, emphasizing the use of structured analysis methods, dividing the information reflected by remote sensing images into levels according to the degree of information abstraction, and using corresponding technical methods at different levels for feature extraction, query and analysis. Hierarchical modeling is also a commonly used system analysis and design idea in the field of information technology. Its advantages lie in simplifying the modeling of complex objects and layered design and implementation, which can improve the relative stability, adaptability, flexibility and maintainability of the system. At present, object-oriented method is often used to model remote sensing image information. The remote sensing image information is divided into four information levels: image representation layer, image object layer, professional target layer and professional event layer. The latter two levels support domain objects and support image information query according to image features and domain object semantic features.

2.2.2 Intelligent extraction using convolutional neural network

Remote sensing image contains abundant natural and social attributes of the surface. From the application point of view, in addition to the extraction of single object information, in many cases, the extraction of multiple image target complexes [10].

Convolutional neural network (CNN) has a very wide range of applications, because it can directly input the original image and avoid the complex pre-processing. The convolutional neural network
composed of multiple convolutional layers and alternating sampling layers has good target information expression ability and can automatically extract road features for learning. Moreover, the learned features can better reflect the features of the original image, which is more conducive to information extraction.

The formula for calculating the convolution layer is as follows:

\[ X_j^l = f \left( \sum_i X_i^{l-1} * K_{ij}^l + b_j^l \right) \]  \hspace{1cm} (1)

where \( X_j^l \) and \( b_j^l \) represent the output characteristic graph and bias of the Jth neuron in the current layer. \( K_{ij}^l \) represents the convolution kernel of the JTH neuron in the current layer when convolving the input ith feature graph, \( X_i^{l-1} \) represents the ith feature graph input by the current layer, function f is the nonlinear activation function, and the nonlinear is introduced for the neural network. The remote sensing data are processed by convolutional neural network and the visualized data are analyzed. Compared with the historical data, according to the demand, the convolution neural network is used to complete the intelligent information extraction of remote sensing big data by virtue of its good target information expression ability. Specific steps are as follows: first, according to the uniform distribution of convolution initialized weights of neural network, then the remote sensing data input to the neural network and multiple convolution operation, after the calculation of the neural network output and the relative error, and then adjust the network weights and calculated using the error reverse bias, when the loss function of neural network light rain the preset threshold, record the network parameters, realize intelligent information extraction of remote sensing of big data.

3. Contrast experiment

In order to verify that the proposed automatic analysis and intelligent extraction method of remote sensing big data is superior to the traditional automatic analysis and extraction method, a comparative experiment was designed.

Six different experimental areas were selected to ensure the diversity of data collected within the divided experimental areas, such as rivers, roads, trees and houses. Using the same acquisition equipment and the same remote sensing sensor, the automatic analysis and extraction experiments of remote sensing data are carried out under the condition that all other conditions are the same.

The automatic analysis and intelligent extraction method of remote sensing big data proposed in this paper was used as the experimental group, while the traditional automatic analysis and extraction method of remote sensing data was used as the control group. Through computer software, the integrity, accuracy and quality of the two methods for remote sensing data analysis and extraction are compared.

| Table 1 integrity results of the two methods (%) |
|-------------------------------|
| remote sensing image | Experience group | control group |
|---------------------|-----------------|---------------|
| 1                   | 89.58           | 76.21         |
| 2                   | 88.78           | 75.37         |
| 3                   | 87.25           | 76.95         |
| 4                   | 90.29           | 73.95         |
| 5                   | 88.35           | 77.01         |
| 6                   | 90.82           | 76.25         |
Table 2 analysis of extraction quality by two methods (%)

| remote sensing image | Experience group | control group |
|----------------------|------------------|--------------|
| 1                    | 82.69            | 65.22        |
| 2                    | 81.31            | 61.95        |
| 3                    | 80.18            | 62.93        |
| 4                    | 83.78            | 61.26        |
| 5                    | 82.72            | 54.47        |
| 6                    | 83.96            | 61.61        |

Fig. 6 results of accuracy of the two methods

As can be seen from table 1, the integrity of automatic analysis and extraction of remote sensing data in 6 experimental areas in the experimental group was higher than that in the control group. In addition, from the perspective of numerical analysis, the method of the experimental group has been greatly improved in the integrity index.

According to the data in table 2, the quality of remote sensing data analysis and extraction in the experimental group was much higher than that in the control group, that is, the proposed automatic analysis and intelligent extraction method of remote sensing big data was significantly superior to the traditional automatic analysis and extraction method in terms of its completion quality.

Accuracy is the most important factor in remote sensing data analysis and extraction. Figure 6 shows the correct rate of the two methods' analysis and extraction of complex remote sensing data in six experimental areas. It can be seen that the accuracy of the experimental group method is significantly higher than that of the control group method. In addition, the accuracy rate of the experimental group remained relatively stable without major fluctuations.

In conclusion, the proposed automatic analysis and intelligent extraction method of remote sensing big data is superior to the traditional analysis and extraction method of remote sensing data.

4. Peroration

At present, big data has become a vital resource in social production and life. With the development of satellite remote sensing technology in China, remote sensing big data will gain new development space in the breadth and depth of earth observation. However, the development of satellite remote sensing technology is only the first step to realize remote sensing big data collection, and we still have
a long way to go to realize automatic data analysis and deep mining. Based on this, the remote sensing technology in the future study, thinking, we should take advantage of big data will technology research focus on the practice of technical analysis and data mining, improve the remote sensing data in terrain analysis, to explore resources, ecological monitoring, the application value of the industrial development of urban planning, to fully reflect the role of big data applications in the modern social development.

Acknowledgement
The study was supported by “2018 Scientific Research Plan Projects of Education Department of Shaanxi Province 18JK0276” ; “2019 Young Talent Research Scholar Program of Shaanxi Province (2019 KJXX-027)”; “2019 Graduate School-Enterprise Collaborative Innovation Fund Project of Northwestern Polytechnical University(XQ201914)” ; “Program of Major Theoretical and Practical Issues in Shaanxi Province(2020Z388)”.

References
[1] H.W. Hui, C.C. Zhou, S.G. Xu, F.H. Lin, A Novel Secure Data Transmission Scheme in Industrial Internet of Things, China Communications, vol. 17, no. 1, 2020, pp: 73-88.
[2] F.H. Lin, Y.T. Zhou, X.S. An, I.You, K.R. Choo, Fair Resource Allocation in an Intrusion-Detection System for Edge Computing: Ensuring the Security of Internet of Things Devices, in IEEE Consumer Electronics Magazine, vol. 7, no. 6, 2018, pp: 45-50. doi: 10.1109/MCE.2018.2851723.
[3] J.T. Su, F.H. Lin, X.W. Zhou, X. Lv, Steiner tree based optimal resource caching scheme in fog computing, China Communications, vol. 12, no.8, 2015, pp: 161-168.
[4] Anonymous. Design of dynamic image information intelligent extraction system based on Internet [J]. Television technology, 2018, 42(09):54-58.
[5] Shen quanfei, cao min, shi zhaoliang, et al. Intelligent classification of remote sensing images based on cuckoo algorithm [J]. Bulletin of surveying and mapping, 2017(1):65-68.
[6] Li shengyang, zhang wanfeng, Yang song. Intelligent fusion of multi-source high-resolution remote sensing images [J]. Journal of remote sensing, 2017, 21(3):415-424.
[7] Zhang chunmin, mu tingkui, yan tingsu, et al. Development and prospect of hyperspectral remote sensing technology [J]. Space return and remote sensing, 2018, 39(3):108-118.
[8] Feng wenqing, sui haigang, tu jihui, et al. Random forest change detection method for high-resolution remote sensing images [J]. Journal of surveying and mapping, 2017, 46(11):90-100.
[9] Liu ruyi, song jianfeng, quan yining, et al. An automatic road extraction method for high-resolution remote sensing images [J]. Journal of xi 'an university of electronic science and technology, 2017, 44(1):100-105.
[10] Zhan yating, julli, sun yonghua, et al. Automated extraction of spectral angle-distance similarity growth model of remote sensing for coastline [J]. Journal of remote sensing, 2017, 21(3):458-469.