Journal of Autonomous Intelligence (2021) Volume 4 Issue 1.
doi: 10.32629/jai.v4i1.479

Original Research Article

Machine Learning, Deep Learning and Implementation Language in Geological Field

Yongzhang Zhou[1,2,3], Jun Wang[1,2,3], Renguang Zuo[4], Fan Xiao[1,2,3], Wenjie Shen[1,2,3], Shugong Wang[1,2,3]

[1] *Center for Earth Environment ＆ Resources, Sun Yat-sen University, Guangzhou 510275, China*

[2] *Guangdong Provinical Key Laboratory of Mineral Resources and Geological Processes, Sun Yat-sen University, Guangzhou 510275, China*

[3] *School of Earth Sciences ＆ Engineering, Sun Yat-sen University, Guangzhou 510275, China*

[4] *China University of Geology, Wuhan 430074, China*

ABSTRACT

Geological big data is growing exponentially. Only by developing intelligent data processing methods can we catch up with the extraordinary growth of big data. Machine learning is the core of artificial intelligence and the fundamental way to make computers intelligent. Machine learning has become the frontier hotspot of geological big data research. It will make geological big data winged and change geology. Machine learning is a training process of model derived from data, and it eventually gives a decision oriented to a certain performance measurement. Deep learning is an important subclass of machine learning research. It learns more useful features by building machine learning models with many hidden layers and massive training data, so as to improve the accuracy of classification or prediction at last. Convolutional neural network algorithm is one of the most commonly used deep learning algorithms. It is widely used in image recognition and speech analysis. Python language plays an increasingly important role in the field of science. Scikit-Learn is a bank related to machine learning, which provides algorithms such as data preprocessing, classification, regression, clustering, prediction and model analysis. Keras is a deep learning bank based on Theano/Tensorflow, which can be applied to build a simple artificial neural network.

Keywords: *Geological Big Data; Machine Learning; Deep Learning; Artificial Neural Network; Intelligent Geology; Python*

ARTICLE INFO
Received: Jan 23, 2021
Accepted: Apr 4, 2021
Available online: Apr 10, 2021

***CORRESPONDING AUTHOR**
Yongzhang Zhou
zhouyz@mail.sysu.edu.cn;

CITATION
Zhou Y, Wang J, Zuo R, *et al.* Machine learning, deep learning and implementation language in geological field. Journal of Autonomous Intelligence 2021; 4(1): 6–12. doi: http: 10.32629/jai.v4i1.479

1. Introduction

Geological big data is growing exponentially. They are largely produced in the investigation, exploration and corresponding geological scientific research of basic geology, mineral geology, hydrogeology, engineering geology, environmental geology and disaster geology, the development and utilization of energy and minerals, the monitoring and prevention of environmental and geological disasters, and various space-based and space-based remote sensing observation activities. The ways to obtain geological big data include geophysics, geochemistry, drilling exploration wells, remote sensing and telemetry, sensing monitoring, and they can also come from various expanded applications, such as map compilation, analysis and calculation, simulation, prediction and evaluation, intelligent management and control etc. Geological big data can be structured, such as data obtained from geochemical analysis and geophysical exploration. There are more unstructured and semi-structured data, such as paleontology, minerals, rocks, ore deposits, core photos, tsunami audio, seismic video, structure, remote sensing spectral maps, specimens, field records, geological charts etc.

In this context, any individual handling geological big data in a traditional way is just like a person racing against a car, plane and rocket. The farther he goes, the greater the gap, and finally he is abandoned by advanced means of transportation. Only by developing intelligent data processing methods can we catch up with the extraordinary growth of big data. Therefore, it can be said that artificial intelligence geology should be an important development direction.

Machine learning is considered to be the core of artificial intelligence and the fundamental way to make computers intelligent. At present, the universal view of various basic problems of machine learning and artificial intelligence is forming (**Figure 1**).

Although scientists with a sense of historical mission are exploring seriously and diligently (Mayer-Schonberger & Cukier, 2013; Carranza & Laborte, 2015; de Mulder *et al.*, *2016;* Aryafar and Moeini, 2017; Ross *et al.*, 2018), artificial intelligence geology based on big data is far from mature (Zhang & Zhou, 2017; Zhou *et al.*, 2018a).

As a big data album, this issue focuses on the modeling and application of machine learning (including deep learning) (Xu, 2018; Zhou, 2018; Han *et al.*, 2018; Jiao *et al.*, 2018; Liu *et al.*, 2018; Wang *et al.*, 2018). It also reflects that machine learning has become one of the important hotspots of current geological big data research. The author believes that machine learning will make geological big data winged, effectively process massive data, mine valuable and rich information behind them, and change geology accordingly.

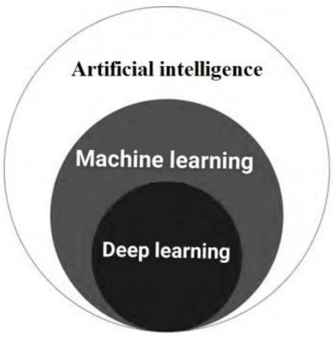

Figure 1. Relationship between artificial intelligence,machine learning and deep learning

2. Classification of Machine Learning

In essence, machine learning is a training process of model derived from data, and eventually gives a decision oriented to a certain performance measurement. Machine learning can be divided into Supervised Learning and Unsupervised Learning. Under supervised learning, each group of training data has an identification value or result value. When establishing the prediction model, Supervised Learning establishes a learning process, compares the prediction results with the actual results of training data, and constantly adjusts the prediction model until the prediction results of the model reach the expected accuracy. The common methods of Supervised Learning are shown in **Figure 2**. In Unsupervised Learning, the data is not specially identified, and the learning model is to infer some internal structures of the data. Common Unsupervised Learning methods are shown in **Figure 3**. In machine learning, SVM (Support Vector Machine) is a representative method. It is based on binary classification algorithm, and its core of thinking is ascending dimension and linearization. Many sample sets that cannot be processed linearly in low-dimensional sample space can be linearly divided (or regressed) through a linear hyperplane in high-dimensional feature space. SVM maps the sample space into a high-dimensional feature space through a nonlinear mapping P, so that the nonlinear separable problem in the original sample space is transformed into a linear separable problem in the feature space.

3. Deep Learning

Deep learning is a subclass of machine learning research. Its purpose is to establish and simulate the neural network of human brain for analysis and learning, and simulate the mechanism of human brain to interpret data, such as image, sound and text. The essence of deep learning is to learn more useful features by building machine learning models with many hidden layers and massive training data, so as to finally improve the accuracy of classification or prediction (Hinton *et al.*, 2006, 2012; Brenden *et al.*,2015; Lecun *et* al., 2015; schmidhuber, 2015; Bianco *et al.*, 2017) "Depth model" is the means and "feature learning" is the purpose.

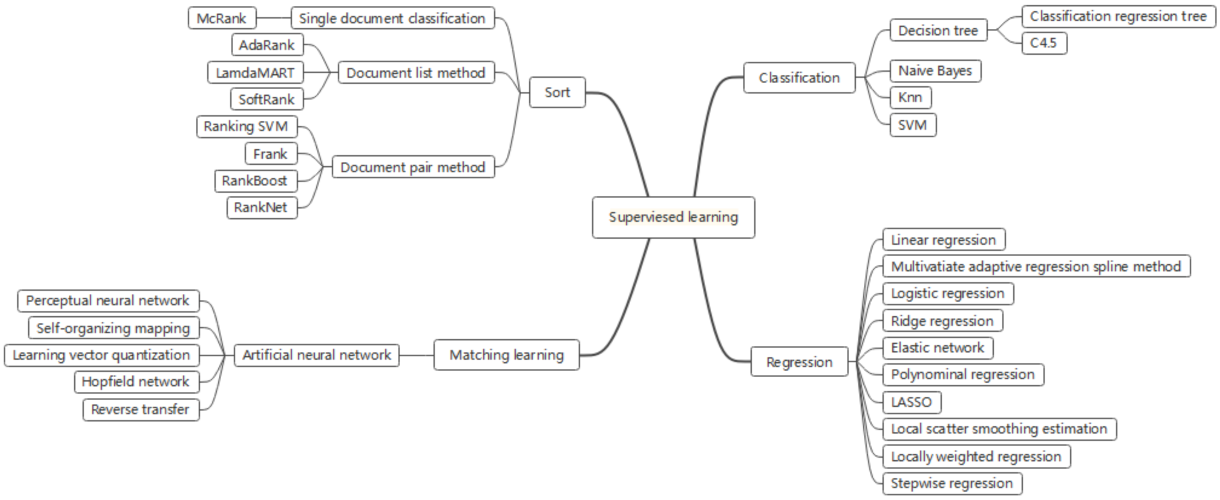

Figure 2. Common supervised learning algorithms (Zhou *et al.*, 2018b).

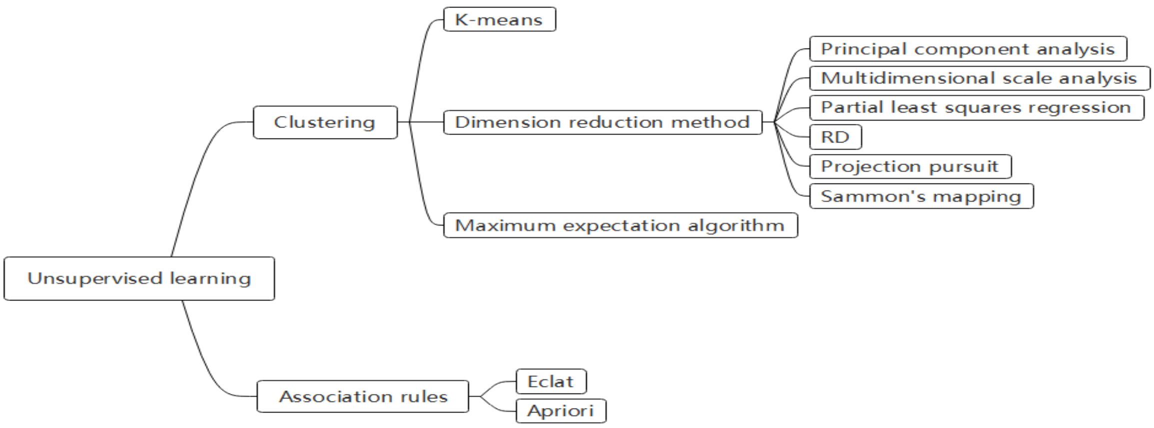

Figure 3. Common unsupervised learning algorithms.

Table 1 Lists the current common deep learning models or methods. Among them, convolutional neural network algorithm is the most commonly used deep learning algorithm. It is widely used in image recognition and speech analysis. It is essentially an input-output mapping, which can learn a large number of mapping relationships between input and output without any accurate mathematical expression between input and output. As long as the convolution network is trained with a known mode, the network has the mapping ability between input and output pairs. Convolutional neural network is a multi-layer neural network, each layer consists of multiple two-dimensional neural networks.

Table 1. Common deep learning models and algorithms

Common models or methods	Algorithm description
Autoencoder	An unsupervised neural network model. The implicit features (ENCODING) of the input data can be learned, and the original input data can be reconstructed (DECODED) with the learned new features. Automatic coders are used for dimension reduction or feature learning.
Sparse coding	An unsupervised learning method. It is used to find a set of "super complete" basis vectors to represent the sample data more efficiently. The method has spatial locality, directionality and band-pass in frequency domain. It is an adaptive image statistical method.

Table continued.

Restricted Boltzmann Machines	RBM, a randomly generated neural network that can learn probability distribution from input data set Restricted Boltzmann machine has been applied in dimension reduction, classification, collaborative filtering, feature learning and topic modelling. According to different tasks, the restricted Boltzmann machine can be trained by supervised learning or unsupervised learning.
Deep beliefnetworks	Dbns, a probability generation model composed of multiple restricted Boltzmann machine layers Compared with the neural network of traditional discriminant model, generative model is to establish a joint distribution between observation data and labels. It can be extended to convolution dbns (CDBNS).
Convolutional neural network	CNN is a kind of artificial neural network. Its weight sharing network structure makes it more similar to biological neural network, which reduces the complexity of network model and the number of weights cnns has become a research hotspot in the field of speech analysis and image recognition.

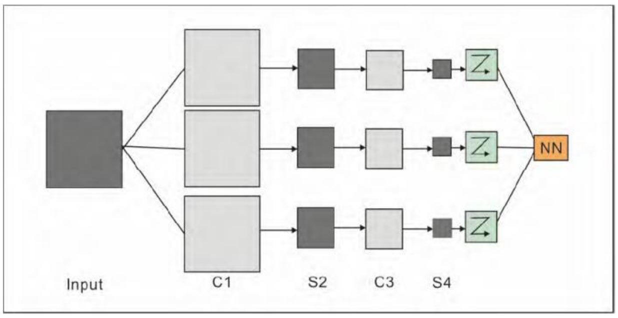

Figure 4. Conceptual diagram of convolution neural network.

Each plane is composed of multiple independent neurons (**Figure 4**). Convolution network is a multi-layer perceptron specially designed to recognize two-dimensional shapes. This network structure is highly invariant to translation, scaling, tilt or other forms of deformation. In this paper, Xu Shuteng and Zhou Yongzhang (2018) took pyrite, chalcopyrite, galena, sphalerite and other sulfide minerals from Jiapigou gold mine in Jilin Province and Shihu gold mine in Hebei Province as examples to design a targeted Unet convolution neural network model to realize the automatic recognition and classification of ore minerals under the microscope based on deep learning algorithm. The experiment shows that the recognition success rate of the trained model for the ore mineral photos under the microscope of the test set is higher than 90%, indicating that the model established in the experiment has good image feature extraction ability and can complete the task of intelligent recognition of ore minerals under the microscope.

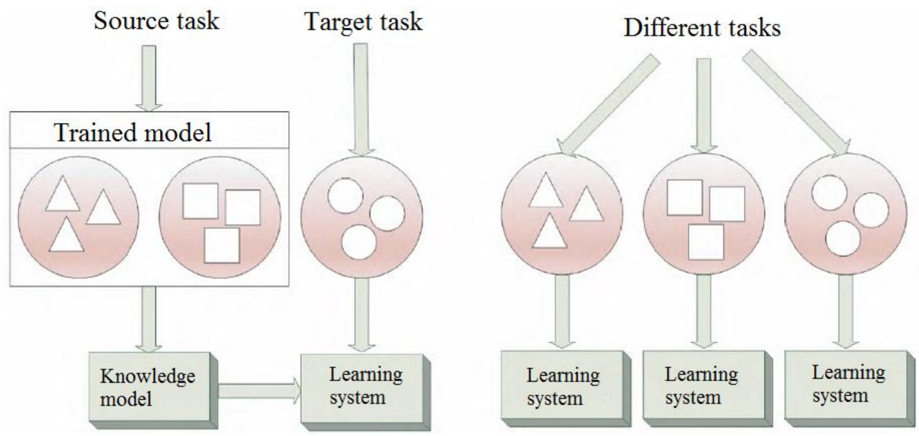

Figure 5. Comparison between traditional machine learning (LEFT) and transfer learning (RIGHT).

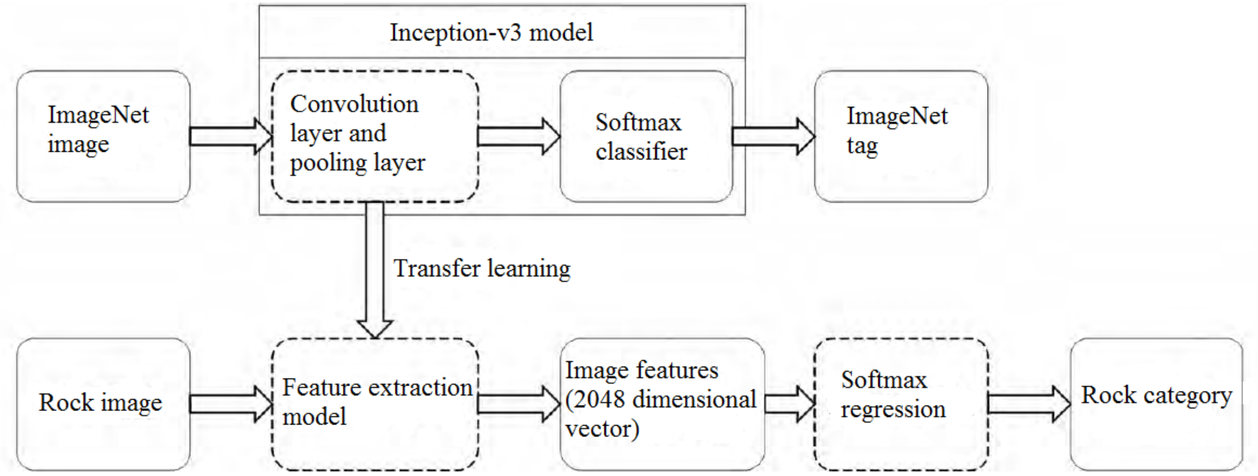

Figure 6. Transfer learning model for automatic identification and classification of minerals.

4. Transfer Learning

Transfer learning is to transfer the learned and trained model parameters to the new model to help the new model training (Yosinski *et al.*, 2014). Considering that most of the data or tasks are relevant, the learned model parameters (also known as the knowledge learned by the model) can be shared with the new model in some way through migration learning, so as to speed up and optimize the learning efficiency of the model without learning from scratch like most networks. There are substantial differences between traditional machine learning and transfer learning, as shown in **Figure 5**.

Migration learning is used to study the automatic recognition and classification of minerals and rocks, which can provide a new means for the automatic classification of rock lithology. Zhang Ye *et al.* (2018) selected three rock images of granite, phyllite and breccia for test, recognition and analysis. The rock image samples used in the experiment are collected by different means such as photos, rock database and network search. The rock types are mainly composed of laboratory rock samples, on-site rock samples and on-site large-scale rock images. In order to make the whole process more intelligent, the scaling and clipping of rock image are automatically completed in the training. The input image only ensures a fixed format, and there are no specific requirements for image size, size and pixel. The author established a rock image depth learning migration model based on Inception-v3. As shown in **Figure 6**, the automatic recognition rate of granite, phyllite and breccia can reach higher than 80%, and some results can even reach higher than 95%. The training process has low requirements for the size, imaging distance and light intensity of rock image, which fully demonstrates its robustness and generalization ability.

5. Algorithm Implementation

With the development of NumPy, SciPy, Matplotlib, Pandas and other program libraries, Python plays an increasingly important role in the field of science (Zhou *et al.*, 2018a).

Scikit-Learn is a library related to machine learning and a powerful machine learning toolkit of Python. It provides a complete machine learning toolbox, including data preprocessing, classification, regression, clustering, prediction, model analysis, etc.

Artificial neural network is a model with powerful function but simple principle. It plays an important role in image recognition, language processing and other fields. Theano is also a python library. It was developed by a deep learning expert Yoshua Bengio. It is used to define, optimize and efficiently solve the simulation estimation problem of mathematical expression corresponding to multi-dimensional array data. It has the characteristics of efficient symbol decomposition, highly optimized speed and stability. The most important thing is that it also realizes GPU acceleration, so that the processing speed of intensive data is dozens of times that of CPU. Theano can build an efficient neural network model, but the threshold is relatively high for ordinary readers.

Therefore, Keras library can be used to build neural network. The application of Keras library can greatly simplify the steps of building various neural network models, and allow ordinary users to easily build and solve the deep neural network with hundreds of input nodes. Keras is not a simple neural network library, but a powerful deep learning library based on Theano. It can be used to build not only ordinary neural networks, but also various deep learning models, such as self encoder, cyclic neural networks, recursive neural networks, convolutional neural networks, etc. As it is based on Theano, it is also quite fast.

The process of building neural network model with Keras is quite simple and intuitive. It is just like building blocks. A very powerful neural network model, even a deep learning model, can be built through just a few dozen lines of code. In the author's teaching and scientific research, it is recommended to build a Python development platform, and the application of Python language can well realize the machine learning algorithm.

6. Conclusion

Through the above discussion, the following understanding can be formed.

(1). Geological big data is growing exponentially. Only by developing intelligent data processing methods can we catch up with the extraordinary growth of big data. Therefore, the development of artificial intelligence geology should be an important development direction.

(2). Machine learning is the fundamental way to make computers intelligent. It is essentially a model training process derived from data, and finally gives a decision oriented to a certain performance measurement.

(3). The purpose of deep learning is to establish and simulate the neural network of human brain for analytical learning, and simulate the mechanism of human brain to interpret data. Its essence is to learn more useful features by building machine learning models with many hidden layers and massive training data, so as to finally improve the accuracy of classification or prediction.

(4). The development of NumPy, SciPy, Matplotlib, Pandas and many other program libraries makes Python occupy an increasingly important position in the field of science. Scikit-Learn and Keras are important toolkits for building machine learning and artificial neural networks using Python.

(5). Although the artificial intelligence geology relying on big data is far from mature, the breakthrough and development of machine learning algorithms make it possible to quickly process massive geological big data and mine valuable and rich information behind them, which will change geology accordingly.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Aryafar A, Moeini H. Application of continuous restricted Boltzmann machine to detect multivariate anomalies from stream sediment geochemical data, Korit, East of Iran. Journal of Mining and Environment 2017; 8(4): 673–682. doi: 10.22044/JME.2017.966.
2. Bianco S, Buzzelli M, Mazzini D, *et al.* Deep learning for logo recognition. Neurocomputing 2017; 245: 23–30. doi: 10.1016/j.neucom.2017.03.051.
3. Brenden M, Ruslan S, Joshua B. Human-level concept learning through probabilistic program induction. Science 2015; 350(6266): 1332–1338. doi: 10.1126/science.aab3050.
4. Carranza EJM, Laborte AG. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). Computers & Geosciences 2015; 74: 60–70. doi: 10.1016/j.cageo.2014.10.004.
5. de Mulder EFJ, Cheng Q, Agterberg F, *et al*. New and game-changing developments in geochemical exploration. Episodes 2016; 39(1): 70–71.
6. Han S, Li M, Ren Q, *et al*. Intelligent determination and data mining for tectonic settings of basalts based on big data methods (in Chinese). Acta Petrologica Sinica 2018; 34(11): 3207–3216.
7. Hinton GE, Osindero S, Teh Y. A fast learning algorithm for deep belief nets. Neural Computation 2006;

18(7): 1527–1554. doi: 10.1162/neco.2006.18.7.1527.

8. Hinton GE, Deng L, Yu D, *et al*. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine 2012; 29(6): 82–97. doi: 10.1109/MSP.2012.2205597.
9. Jiao S, Zhou Y, Zhang Q, *et al*. Study on intelligent discrimination of tectonic settings based on global gabbro data from GEOROC (in Chinese). Acta Petrologica Sinica 2018; 34(11): 3189–3194.
10. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521 (7553): 436–444.
11. Liu Y, Zhu L, Zhou Y. Application of Convolutional Neural Networkin prospecting prediction of ore deposits: Taking the Zhaojikou Pb-Zn ore deposit in Anhui province as a case (in Chinese). Acta Petrologica Sinica 2018; 34(11): 3217–3224.
12. Mayer-Schonberger V, Cukier K. Big data: A revolution that will transform how we live, work and think. New York: Houghton Mifflin Harcourt Publishing Company. 2013.
13. Ross ZE, Meier MA, Hauksson E. P-wave arrival picking and first-motion polarity determination with deep learning. Journal of Geophysical Research: Solid Earth 2018; 123(6): 5120–5129. doi: 10.1029/2017JB015251.
14. Schmidhuber J. Deep learning in neural networks: An overview. Neural Networks 2015; 61: 85–117. doi: 10.48550/arXiv.1404.7828.
15. Wang H, Luo J, Wang J, *et al*. Quantitative classification and metallogenic prognosis of basic-ultrabasic rocks based on big data: Taking its application in Beishan area for example (in Chinese). Acta Petrologica Sinica 2018; 34 (11): 3195–3206.
16. Xu S, Zhou Y. Artificial intelligence identification of ore minerals under microscope based on deep learning algorithm (in Chinese). Acta Petrologica Sinica 2018; 34 (11): 3244–3252.
17. Yosinski J, Clune J, Bengio Y, *et al*. How transferable are features in deep neural networks? Advances in Neural Information Processing Systems 2014; 27: 3320–3328. doi: 10.48550/arXiv.1411.1792.
18. Zhang Q, Zhou Y. Big data will lead to a profound revolution in the field of geological science (in Chinese). Scientia Geologica Sinica 2017; 52 (3): 637–648. doi: 10.12017/dzkx.2017.041.
19. Zhang Y, Li M, Han S. Automatic identification and classification in lithology based on deep learning in rock images (in Chinese). Acta Petrologica Sinica 2018; 34(2): 333–342.
20. Zhou Y, Chen S, Zhang Q, *et al*. Advances and prospects of big data and mathematical geoscience (in Chinese). Acta Petrologica Sinica 2018; 34(2): 256–263.
21. Zhou YZ, Zhang LJ, Zhang AD, *et al.* Big data mining & machine learning in geoscience (in Chinese). Guangzhou: Sun Yat-sen University Press. 2018. p. 1–360.