Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Review of the Application of Big Data and Artificial Intelligence in Geology

To cite this article: Lirong Chen *et al* 2020 *J. Phys.: Conf. Ser.* **1684** 012007

Review of the Application of Big Data and Artificial Intelligence in Geology

Lirong Chen[1], Liang Wang[1], Jinli Miao[1,*], Huan Gao[3], Yue Zhang[2], Yao Yao[3], Ming Bai[1], Lisi Mei[1] and Jing He[4]

[1]Development and Research Center of China Geological Survey (National Geological Archives of China), Beijing, China

[2]Faculty of Information Engineering, China University of Geosciences(Beijing), Beijing, China

[3]Faculty of Geography and Information Engineering, China University of Geosciences(Wuhan), Wuhan, China

[4]Fourth Institute of Oceanography, Ministry of Natural Resources, Beihai, China

*Corresponding author e-mail: mjinli@mail.cgs.gov.cn

Abstract. Big data and artificial intelligence (AI) have provided new methods and opportunities for many applications in geology. Nevertheless, big data and AI-based geoscience applications are still in their infancy, and the methods and objectives are still scattered, lacking a unified theoretical and application framework. This study reviews the application of big data and AI in geology and analyzes and summarizes the current status, development trends and existing problems of ongoing studies from the perspectives of geology. This study can provide a reference for the future research and development of the application of big data and AI in the field of geology.

1. Introduction

Geology is the discipline of the Earth's evolution and development characteristics. The main object of geology is the lithosphere below the surface, and its research content includes the composition, internal structure and evolution of the Earth’s history, among other subjects [1]. Because of the large and complex nature of geological research objects, traditional geological research often has problems related to complex data sources and low precision [2]. With the rapid development of science and technology, huge numbers of new technical methods have been cited in geological studies, such as geophysical exploration methods[3], geochemical exploration methods[4-5], isotope geology methods[6], remote sensing geology methods[7], etc. These technologies improve the accuracy of geological data and expand the data volume.

Artificial intelligence (AI) is a branch of computer science. It refers to the intelligence displayed by a system made by humans [8]. AI can be divided into two categories: strong AI and weak AI [8]. The perspective of weak AI is that intelligent machines that can reason and solve problems cannot be built [9]. The strong AI view is likely to create truly intelligent machines that are capable of reasoning and problem solving; such machines would be considered sentient and have a sense of self [9]. The research fields of AI include expert systems, machine learning, natural language recognition, computer vision and recommendation systems [10]. Among them, machine learning is the branch that has developed the fastest and can best reflect intelligence [11].

Machine learning is the algorithmic study of how computers simulate or implement human learning behavior, characterized by the ability to automatically improve through experience. Through machine learning algorithms, we can acquire new knowledge or skills. In order to solve the bottlenecks encountered in the development of machine learning, such as multiple model types, extensive training, difficulty in determining parameter weights and a large range of parameters, deep learning technology has emerged and been widely used, and it has become the key field of AI development[12].

Deep learning is a kind of machine learning technology. It uses deep neural networks to solve the learning process of feature expression. The main goals of a "deep learning" neural network are to simulate the human brain’s learning process through a deep neural network algorithm and combine low-level features into a higher-level abstract representation through the nonlinear relationship between input and output to finally reach the level of mastery and application[13].

The richness of data plays a very important role in the training effect of an AI model [14]. The concept of big data emerged with the explosive growth of information data and the rapid development of network computing technology, which provides the data foundation support for AI [15]. Big data is a large-scale data collection, and it is difficult for traditional database software tools to collect, store, manage and analyze it [16]. It refers not only to the large amount of data but also to the wide variety of data types and low value density [15]. Big data technology can help companies from all walks of life to dig out the user's demand information from mass data with low value density, and as a result, the data can be changed from quantitative to qualitative and truly generate value[15].

In recent years, with the rapid progress of big data and AI, there is a certain degree of application in geology [17]. With the improvement of computing performance, especially the rapid development of GPU high-performance computing technology, the constraint of computing power on big data and AI has been largely solved [18]. This relieves the pressure of big data and AI technology at the computing level and broadens the application space and development prospects of both in the fields of geology [19]. AI usually serves a purpose for geological reconnaissance and resource exploration, such as mineral identification and geochemical anomaly detection [20]. This paper discusses big data and AI in geology.

2. Application of big data and artificial intelligence in geology

Geological big data is a new concept resulting from the introduction of big data theory to the field of geology [17]. In geological research, geological data come from a wide range of sources. With the application of various new technologies, the reduction of data storage costs and the accumulation of historical data, geological data have begun to have the characteristics of big data in terms of quantity, value, diversity and timeliness [2]. The traditional data processing methods are gradually becoming unable to meet the requirements of geological data processing methods and processing speed [17]. Big data and AI theory have provided the geological industry with an effective method to deal with geological big data, especially the development of high-performance computing technology, which means that the AI models have a broader application space in the geological industry.

Geological science is an observational discipline; it is performed through the summarized analysis of observed geological phenomena by reasoning on the whole face [18]. Because of the difficulty in obtaining comprehensive data and the limitation of researchers' own experience and knowledge, it is difficult to perform analyses and make judgments that are quite consistent with the real situation. Compared with traditional geological research methods, big data and AI can take advantage of the vast amounts of geological data to summarize geological characteristics; explore the rules of geological activity; analyze geological phenomena objectively, impartially and quickly; and provide more scientific results for geological work [19]. Therefore, geological big data technology research will inject new vitality into the development of geology.

Geological big data is characterized by having multiple sources, multiple types and large volumes, which make the application of big data theory and related technologies in the geological industry possible, but this also causes many problems [2]. For example, the multi-source nature of geological big data makes data mining more difficult. Because there are different sources, types and quantities of data, data processing methods and analysis models are also different, so it is difficult to define the weight of the impact on the results. The value density of geological big data is relatively low, and the final mining results may not be reliable, which has a huge impact on mineral resource exploration, geological disaster monitoring and other aspects. There are not many studies on geological big data in China, so it is difficult to guarantee the quality of geological data with low value density, and the accuracy and preservation of the data face many problems[19].

To date, many scholars have tried to apply technological methods related to big data and AI to the research field of geology. In the field of geology-related scientific research, because of the increasing use of new technologies, the field of resource exploration has more economic benefits, more data sources, and greater data volume. Therefore, this field is more suitable for the application of artificial intelligence and theoretical methods of big data. The artificial neural network (deep learning model) is one of the most widely used artificial intelligence models in this field. Artificial neural networks (deep learning models) have proven to be powerful tools for classifying and identifying minerals [20,23]. Cracknell and Reading compared five machine-learning algorithms for geological mapping, proving that random forest was the preferred classifier of lithology [24]. In recent years, the research on key technologies of machine learning applied to the field of Mineral Prospecting and Mapping (MPM) has received widespread attention. Several supervised classification methods have been used in the mapping of mineral prospects. Rodriguez-Galiano et al. (2015) used an artificial neural network, decision tree, random forest and support vector machine separately for mineral exploration mapping. By comparing the four results, they found that random forest was superior to the other three machine learning methods [25].

In addition, because of the complexity of the geological environment and the unknown distribution of geochemical data, traditional mathematical and statistical methods are not effective in identifying geochemical anomalies. Thus, some machine learning methods have been applied experimentally to geochemical anomaly identification [4-5]. These methods have the advantage of not making assumptions about the distribution of data and dealing well with the nonlinear relationship between geochemical data. For example, Twarakavi et al. (2006) used a support vector machine and a robust least squares support vector machine to plot the concentration of arsenic trioxide using the concentration distribution of gold present in Alaskan sediments [4]. Beucher et al. used artificial neural networks to map soils in the downstream Sirppujoki region of southwest Finland [26]. Chen et al. (2014) used southern Jilin, China, as the study area and successfully identified geochemical anomalies in the area using the continuum-limited Boltzmann machine method [27]. Gonbadi et al. (2015) used supervised machine learning to identify Cu-related porphyry geochemical anomalies using Kerman Province, Iran, as a study area [5]. Chen et al. (2019) presented a spatially constrained multi-autoencoder approach-based NN model to effectively improve the ability of a neural network/deep learning to carry out multi-dimensional anomaly identification in exploration geochemistry [28]. Chen et al. (2019) presented a multi-convolutional auto-encoder approach that adopted a non-interactive net structure to accurately recognize geochemical anomalies [29]. The above studies show that the machine learning model is an effective tool for the recognition of multivariate geochemical anomalies.

In addition, AI and big data also have great potential in geological research fields such as geological sample identification, ore mineralogical study and earthquake monitoring. In the field of geological sample identification, the relevant theoretical techniques of big data can be well applied. The data classification standard of geological sample identification is clear, and the data sources are wide-ranging. Historical identification data have good reference value for current geological sample identification [19]. Using historical data, an AI model can be trained with geological big data, and the model can be used to discriminate the types of samples, which can not only greatly improve the efficiency but also increase the accuracy and avoid the errors caused by artificial discrimination[24]. To date, the deep learning model has been applied in the field of mineral sample identification, and the results have been good [30]. It is believed that in the future, the application of big data and artificial intelligence methods in geological sample identification will also have a broad development prospect.

The geological big data theory has a broad application prospect in mineralogical studies. The use of big data technology and AI models to analyze and mine geological big data can help geologists to understand the deposit model, analyze the deposit formation process, discover the metallogenic rule and realize the purpose of assisting mineral exploration and development[24]. Earthquake monitoring is related to the national economy and people's livelihood, and seismic exploration is also an important technical method in the geological industry. The development of science and technology, the application of new technology in earthquake monitoring areas and seismic exploration areas have produced a huge amount of data. Using data with the characteristics of big data and using big data and AI technology with high-performance computing technology will greatly improve the accuracy and efficiency of seismic data analysis. It will be of great help to seismic monitoring and field exploration operations.

3. Opportunities and Challenges

AI and big data have been widely used in computer science and data science, but they are still in their infancy in geology. There is no complete set of theories and systems that combine the theoretical problems of geology with machine learning.

The relevant achievements of geology include geochemical anomaly detection [4-5], mineral identification [20, 23] and many other areas. This proves that big data and AI have a very wide range of research prospects and application value in geology. Spatial big data have the advantage of reflecting the fine-grained "Earth–time" relationship [30], and the "from down to up" approach can be used to model the traditional geology problem. If AI is used as the means, big data are used as the input and problems are used as the purpose, results with higher accuracy can be effectively obtained compared with traditional statistical methods. However, the uncertainty of AI, especially deep learning, makes development uncertain in the fields of geology [22].

Multi-scale effects, spatiotemporal heterogeneity and spatiotemporal correlation are important features of geology [31]. The research of geology sciences has developed rapidly from single-scale to multi-scale and from static to dynamic [31]. In the process of development, the need for new models and new data is very urgent. AI and big data provide opportunities to meet this need. Therefore, data validity, data bias and model reliability have become the focus of future research.

In view of the four characteristics of big data—a massive data scale, diverse data types, low value density, and rapid data generation [15]—the analysis of big data is different from the traditional technical means of data analysis and research, requiring greater computing power and more efficient processing methods. As the bottom layer of AI and big data analysis, high-performance computing has greatly improved the computing capacity [107], which plays a very important role in the analysis and processing of big data and the realization of complex models of AI. Powerful computing power will be the basis for effectively constructing the spatiotemporal database of geology and solving the big data features of multi-source spatial data, which are massive, heterogeneous, multi-source and multi-temporal. In addition, the issue of how to effectively build an artificial intelligence platform based on a spatiotemporal database to support the big data environment and thus carry out the analysis and research of multi-source spatial data is a scientific and technical problem to be solved in the future.

4. Conclusions and Prospects

This paper summarizes the current research status of big data and AI in the field of geology. The following conclusions were reached.

Because of the diversification of geological data acquisition methods, the reduction of storage cost and the accumulation of historical data, geological data have gradually acquired the characteristics of big data. The introduction of big data and AI theory has provided more diverse data processing methods to the geological industry and the possibility for mining the geological rules hidden in the massive geological data. At present, resource exploration is a mature scientific research field of big data and AI technology in geology, and the artificial neural network is one of the most widely used AI models in this field, which has been proved to be a reliable tool in the classification and identification of minerals. In the field of geochemistry, some anomaly detection techniques based on machine learning have been put into practice and achieved good results. In the fields of geological sample identification, mineralogical study and earthquake monitoring, AI models can be trained with historical data and experience to improve efficiency, reduce errors and improve accuracy.

AI and big data are in their infancy in the geology. A complete set of theories and systems for AI and big data theory have not established in the field of geology. The application has shown advantages such as precision and efficiency in modeling. However, the uncertainty of AI, especially deep learning, makes development uncertain in the fields of geology.

The continuous development of big data and AI has offered many new technologies and methods, which provide a good opportunity for innovation in geology. In the development trend of "from single-scale to multi-scale, from static to dynamic", an important problem to be solved in the future is how to establish the spatiotemporal database and artificial intelligence platform of geology according to the research characteristics of geology. To solve this problem, geology scientists need to get rid of the shackles of traditional thinking, learn about the relevant theories of big data and AI and actively carry out interdisciplinary cooperation.

The application of big data and AI in the geology has become increasingly popular and achieved interesting results, but it has also encountered many opportunities and challenges. There is no royal road to scientific research, and interdisciplinary research is even more difficult. However, we have reason to believe that, with continuous in-depth research and analysis, geology sciences will certainly make great progress with the help of big data and AI.

References

[1] W. R. Muehlberger, Earth Science[J]. Science, 1966, 152 (3724):950-951.
[2] S. F. Huang, X. H. Liu, Thinking about the application of geological big data and geological information development[J]. China Mining Magazine, 2016, 25(08):166-170.
[3] X. S. Wang, B. L. Zhao, S. T. Dong, Y. Zhang, W. Q. Yi, G. C. Xu, Challenges and Strategies for Large Seismic Exploration Data of Oil and Gas Industry[J]. China Petroleum Exploration, 2014, 19(04):43-47.
[4] N. K. C. Twarakavi, D. Misra, S. Bandopadhyay, Prediction of Arsenic in Bedrock Derived Stream Sediments at a Gold Mine Site Under Conditions of Sparse Data[J]. Natural Resources Research, 2006, 15(1):15-26.
[5] A. M. Gonbadi, S. H. Tabatabaei, E. J. M.Carranza, Supervised geochemical anomaly detection by pattern recognition[J]. Journal of Geochemical Exploration, 2015, 157:81-91.
[6] E. I. Hamilton, Principles of isotope geology[J]. Earth Science Reviews, 1978, 14(2):190-191.
[7] B. S. Siegal, A. R. Gillespie, Remote sensing in geology[M]. Wiley, 1980.
[8] S. J. Russell, P. Norvig, Artificial Intelligence: A Modern Approach[M]. Prentice Hall, 1995: 2829-2833.
[9] P. Mccorduck, O. Selfridge, et al. History of artificial intelligence[A]. In: International Joint Conference on Artificial Intelligence, 1977[C].
[10] W. Zhu, H. H. Chen, S. Y. Tian, H. W. Wang, Artificial Intelligence: New Blue Ocean from a Scientific Dream——Analysis and Countermeasures of the Development Situation of AI Industry[J]. Science & Technology Progress and Policy, 2016, 33(21):66-70.
[11] M. I. Jordan, T. M. Mitchell, Machine learning: Trends, perspectives, and prospects[J]. Science, 2015, 349(6245):255-260.
[12] T. M. Mitchell, G. J. Carbonell, R. S. Michalski, Machine Learning[M]. McGraw-Hill, 2003.
[13] Y. Lecun, Y. Bengio, G. Hinton. Deep learning.[J]. Nature, 2015,521(7553):436.
[14] A. Braga, R. Logan, The Emperor of Strong AI Has No Clothes: Limits to Artificial Intelligence[J]. Information, 2017, 8(4):156.

[15] M. Hilbert, Big Data for Development: A Review of Promises and Challenges[J]. Development Policy Review, 2016,34(1):135-174.
[16] J. Manyika, M. Chui, B. Brown, et al. Big Data: The Next Frontier For Innovation, Competition, And Productivity[J]. Analytics, 2011.
[17] G. Yan, Q. Xue, K. Xiao, et al. An analysis of major problems in geological survey big data[J]. Geological Bulletin of China, 2015, 34(7):1273-1279.
[18] L. N. Huynh, Y. Lee, R. K. Balan, DeepMon: Mobile GPU-based Deep Learning Framework for Continuous Vision Applications[A]. In: The International Conference, 2017[C].
[19] Q. Zhang, X. Y. Jia, Z. Wu, J. R. Wang, S. T. Jiao, W. F. Chen, Big data will lead to a great change in Geological Science Research[A].2015 China Geoscience union annual meeting, Beijing, China, 2015 [C].
[20] S. Thompson, F. Fueten, D. Bockus, Mineral identification using artificial neural networks and the rotating polarizer stage[J]. Computers & Geosciences, 2001,27(9):1081-1089.
[21] Y. P. Yao, Y. C. Chai, F. C. Ma, New opportunities for the development of Geological Science: Reflections on the development trend of Geology[J]. Chinese Science Bulletin,2004(18):1919-1924.
[22] Yao Y, Zhang J, Hong Y, et al. Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data[J]. Transactions in GIS, 2018.
[23] N. A. Baykan, N. Yılmaz, Mineral identification using color spaces and artificial neural networks[J]. Computers & Geosciences, 2010,36(1):91-97.
[24] M. J. Cracknell, A. M. Reading, Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information[J]. Computers & Geosciences, 2014,63(1):22-33.
[25] V. G. Rodriguez, M. C. Sanchez, M. O. Chica, et al. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines[J]. Ore Geology Reviews, 2015,71:804-818.
[26] A. Beucher, P. Österholm, A. Martinkauppi, et al. Artificial neural network for acid sulfate soil mapping: Application to the Sirppujoki River catchment area, south-western Finland[J]. Journal of Geochemical Exploration, 2013,125(1):46-55.
[27] Y. Chen, L. Lu, X. Li, Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly[J]. Journal of Geochemical Exploration, 2014,140(4):56-63.
[28] L. Chen, Q. Guan, Y. Xiong, J. Liang, Y. Wang, Y. Xu, A Spatially Constrained Multi-Autoencoder Approach for Multivariate Geochemical Anomaly Recognition. Computers and Geosciences, 2019, 125: 43–54. DOI: 10.1016/j.cageo.2019.01.016.
[29] L. Chen, Q. Guan, B. Feng, H. Yue, J. Wang, F. Zhang, A Multi-Convolutional Auto-encoder Approach for Multivariate Geochemical Anomaly Recognition, Minerals, 2019, 9(5): 270. DOI:10.3390/min9050270.
[30] Y. M. Park, M. P. Kwan, Multi-Contextual Segregation and Environmental Justice Research: Toward Fine-Scale Spatiotemporal Approaches[J]. International Journal of Environmental Research & Public Health, 2017,14(10):1205.
[31] K. Regenauerlieb, M. Veveakis, T. Poulet, et al. Multiscale coupling and multiphysics approaches in earth sciences: Applications[J]. Journal of Coupled Systems & Multiscale Dynamics, 2013,1(3):281-323.
[32] V. V. Kindratenko, J. J. Enos, G. Shi, et al. GPU clusters for high-performance computing[A]. In: IEEE International Conference on CLUSTER Computing and Workshops, 2009[C].