Big data research is a developing field that is bringing change to people’s lives. Big data is an inevitable outcome of the development of science: scientific research has resulted in the accumulation of a large amount of data and conventional methods cannot handle such massive amounts of data. Therefore, alternatives such as big data, cloud computing, artificial intelligence, and blockchain have emerged.

As a new strategic resource for human beings, big data has become a strategic highland in this era of the knowledge economy. It has caused the transformation of scientific methodology and is becoming the new driver of scientific discovery. Scientific big data is an important branch of big data. As used in computational analysis, scientific big data is characterized by its non-reproducibility, high degree of uncertainty, high dimensionality, and high complexity. External characteristics include the data type, data volume, data acquisition, and data analysis. All these characteristics make big data a new challenge for conventional data processing techniques and methods. Thus, we make four recommendations for new methods of processing scientific big data: (1) the scientific recognition of scientific big data, (2) the construction of infrastructure to handle scientific big data, (3) the establishment of a scientific data research center, and (4) the construction of a scientific big data academic platform.

Human beings have entered the era of big data. The objective of big data research is to exploit data using the computer as its tool. Big data research progresses via the determination of correlations between data and is characterized by decision-making based on high probability.

1. Theory-driven mode and data-driven mode

The development of science can be divided into four stages: the era of empirical science, the era of theoretical science, the era of information science, and, finally, the era of big data and artificial intelligence. Traditional research relies on two methods: the deductive method (from general to individual) and the inductive method (from individual to general). For example, magma separation crystallization theory uses the deductive method while the basalt discriminant map uses the inductive method (Zhang, Jiao, & Lu, 2018).

The data-driven model uses the big data method and is the third new scientific method that has been proposed. The characteristics of big data can be simply summarized as the 4Vs: a large amount of data (volume), having many types of data (variety), a low value density (value), and rapidity and time efficiency (velocity). Traditional research employs a theoretically driven model, while big data research employs a data-driven model. Big data research usually involves the use of the full data...
model, which is different from that used in traditional research. The big data method has three important technical orientations: (1) emphasis on efficiency rather than accuracy, (2) emphasis on the whole rather than on sampling, and (3) emphasis on correlation rather than causality.

Theoretically driven models require good theoretical preparation, accurate data, and clear causality. Without good theoretical preparation, research cannot be conducted. The data-driven model, however, does not have such requirements and research using this model has, thus, surpassed the ceilings of scientific research. Geological data are characterized by diversity, multidimensionality, multi-source availability, correlation, randomness, uncertainty, and temporal and spatial inhomogeneity; thus, big data has brought unprecedented opportunities as well as challenges to research in the field of geology. The data-driven scientific model has brought a new perspective to geological research (Zhai, Yang, Chen, & Chen, 2018). In this model, machine learning is the core of artificial intelligence, affording computers a fundamental intelligence. Deep learning is a subclass of machine learning and, here, the convolutional neural network algorithm is one of the most commonly used algorithms (Zhou et al., 2018). The study by Wang et al. (2017) on basalt discriminant maps is a tentative study using data-driven models. Jiao et al. (2018) used machine learning to study global geochemical data and produced groundbreaking results, thus showing that the data-driven model has infinite applicability in this field of research (Liu et al., 2018; Zhang, Jiao, Li, & Chen, 2017; Zhang et al., 2018).

2. Opportunities for research in geology

2.1. Causality and correlation

Traditional research focuses on causality, while big data research focuses on correlation. Practice has shown that the pursuit of causality inevitably involves interference from human factors, while big data research involves the study of the correlations between data and, consequently, is able to largely avoid this interference (Zhang et al., 2018).

Geological research is most concerned with causal problems such as rock genesis, mineral deposit genesis, metamorphic rock genesis, sedimentary rock genesis, and so on. For example, basalt magma undergoes separation and crystallization. In this process, on the one hand, the magma breaks up into separate minerals and forms cumulates; on the other hand, the residual magma evolves into an iron-rich substance. Thus, the cumulates, minerals, and residual magma have a genetic relationship, while the relationship between magma separation and crystallization is a blood relationship. Another example is the wolframite in tungsten deposits. Tungsten is obtained from deep ore-related fluids which have risen to a shallow depth in the crust, leading to the formation of mineral deposits from the metal mineral in the fluid after complex physical and chemical reactions. Here, the wolframite can be said to have a causal relationship with the ore-related fluids.

However, the genesis of deposits is very complicated. For instance, the formation of the deposit associated with magma is affected by the rock mass and the alterations in the structure and composition of the stratum. The genesis of a deposit is influenced by several factors that are not clear in many cases. These factors are usually involved in
the process of mineralization and, in many cases, there may be a related proposition. However, geological research based on big data uses correlations and not causality. Therefore, big data research methods are appropriate for research on mineral deposits.

2.2. New topics for geology

2.2.1. Geology requires datamation

In the era of big data, the use of data is an indicator of how scientific a discipline is. The implication is that any discipline that can be expressed in terms of data is called a science. Marx once said: “Any science can only achieve perfection by fully utilizing mathematics.” Geology has always been considered unscientific on this basis, but it is comforting that geology is different from physics and mathematics in that it is an observational science. However, in the era of big data, whether geological observations and field investigations can be digitized is the key to whether geology can be called a science. Despite explosive growth in the amount of data due to the popularity of high-performance computers, statistical analysis methods have largely broken through the limitations of data volume. Combined with the advantages of using machine learning algorithms for data processing, these methods have made it possible to push geoscience into new realms of quantitative research. In 2017, the China Geological Survey Geological Cloud 1.0 was launched, which opened the path to geological datamation. However, the realization of geological datamation is still a long way away.

2.2.2. New topics in geological research

Geological research has achieved great results, but exploration needs to continue. In geology, some theories derived from observation and experimentation are credible and some are not. Big data research is a new research method and its impact on geological theories is striking.

For example, the basalt discriminant map was an important scientific discovery in plate tectonics. In the 1970s and 1980s, Pearce (1976) explored basalt tectonic settings, opening up a new method of basalt geodynamics research. Based on the composition of basalt in various tectonic environments, such as mid-ocean ridges, island arcs, and ocean islands, they discussed the influence of the magma formation depth, partial melting degree, magma evolution process, and plate subduction on the geochemical properties of basalt. Examples of the application of this theory included many regional studies, and many basalt tectonic environment discriminant maps were proposed based on this theory and found to be widely applicable (Capedri et al., 1980; Galoyan, Rolland, Soisson, Corsini, & Melkonyan, 2007; Glassley, 1974; Harris, Pearce, & Tindle, 1986; Meschede, 1986; Mullen, 1983; Pearce, 1975, 1976, 1982, 1983, 2003; Pearce & Cann, 1973; Pearce & Gale, 1977; Pearce, Harris, & Tindle, 1984b; Pearce, Lippard, & Roberts, 1984a; Pearce & Norry, 1979; Pearce & Peate, 1995; Wood, 1980; Wood, Joron, & Treuil, 1979; Workman & Hart, 2005). However, as more data accumulated, it was found that the new data did not always fit these discriminant maps. Although some scholars have continued to attempt to explore and adopt new methods of constructing viable discriminant maps (Vermeesch, 2006, 2013), these are not widely acknowledged or supported by the
academic community involved in this field. In the new studies, researchers applied the big data method to derive a new basalt discriminant map based on which a discriminant graph theory was developed to solve the problem (Liu et al., 2018; Liu, Zhang, & Zhang, 2017; Wang et al., 2017; Zhang et al., 2017, 2018). In addition, the comparisons between TTG and adakite and between komatiite and picrite also gave interesting results (Luo et al., 2018; Zhang, Jiao, & Lu, 2018).

2.3. New topics in the research on mineral deposits

Research on the formation, genesis, and mechanisms related to and distribution of mineral deposits in the crust is often referred to as “economic Geology”. The genesis of mineral deposits is considered to be the core of mineral deposit research. Researchers hope that the problem of ore prospecting can be solved through the study of the genesis of mineral deposits. This is a reflection of the excessive attention given to causality. Metallogenic regularity and metallogenic models reveal the basic characteristics of mineralization, providing relatively broad information for ore prospecting. Detailed and specific prospecting information can be obtained mainly from geophysical research, geochemical exploration, and the use of drilling technology. In particular, metal deposits contain information that is significantly different from that which can be obtained from the surrounding rocks (stratigraphic, intrusion, volcanic, alteration, fracture zone, etc.). Many factors affect the location of a metal deposit; several of these are as yet unknown. This is undoubtedly a good way of determining the cause; however, the effect is difficult to determine. Therefore, geological prospecting relies primarily on information, not on metallogenic regularity.

In many cases, the relationship between mineralization and the elements surrounding a mineral deposit (such as the rock mass, stratum, structure, alteration, etc.) is correlative rather than causal. Mineral deposits are difficult to locate unless they are combined with information about ore deposits and geophysical and geochemical techniques are used. The reason for this is very simple: mineral deposit research is not mineral exploration.

In many cases, the relationship between mineralization and surrounding factors (such as the rock mass, stratum, structure, alteration, etc.) is mostly correlative rather than causal. Mineral deposit research is a science that explores the regularity in the occurrence of existing deposits; locating new deposits is the task of mineral exploration. Luo et al. (2017) achieved good results in this regard and they deserve greater acknowledgment for this. Geological big data are used in mineral deposit research, including in the construction of big data-based intelligent identification systems designed for detecting metallogenic deposits and in prospecting. At present, intelligent geology is still at an early stage. The construction of big data-based intelligent identification systems used for detecting metallogenic deposits and in mineral prospecting is an important part of intelligent geology (Zhou et al., 2018). Therefore, the development of artificial intelligence technology in this field is also necessary.
3. Conclusions

At present, developments in science are encountering bottlenecks. Traditional methods are based on causality, but it is not possible to understand all phenomena in the world on this basis. Causality is inseparable from reasoning and reasoning is influenced by human factors. However, since data is objective and the ways in which it is used for research are scientific, there is basically no interference from human factors in the field of big data research. In the era of big data, many traditional perspectives are being challenged and subverted. Big data is not a straightforward rejection and denial of traditional research; on the contrary, big data technology is introduced into traditional research to advance progress in traditional scientific research.

Currently, the knowledge structure used in geology is unreasonable. Big data research is a cross-disciplinary field that requires two kinds of experts: (1) experts in big data and artificial intelligence and (2) geology professionals. Therefore, the youth will become the main force driving big data research. We should be glad to have the opportunity to become involved in this new era and must strive to make ourselves remain relevant in these times.

Funding

This work was supported by the Natural Science Foundation of China [41862009] and the Key Laboratory of Geological Information Technology of Ministry of Land and Resources [2017323].

References

Capedri, S., Venturelli, G., Bocchi, G., Dostal, J., Garutti, G., & Rossi, A. (1980). The geochemistry and petrogenesis of an ophiolitic sequence from Pindos, Greece. Contributions to Mineralogy and Petrology, 74(2), 189–200.

Galoyan, G., Rolland, Y., Sosson, M., Corsini, M., & Melkonyan, R. (2007). Evidence for superposed MORB, oceanic plateau and volcanic arc series in the Lesser Caucasus (Stepanavan, Armenia). Comptes Rendus Geoscience, 339(7), 482–492.

Glassley, W. (1974). Geochemistry and tectonics of the crescent volcanic rocks, Olympic Peninsula, Washington. Geological Society of America Bulletin, 85(5), 785–794.

Harris, N. B. W., Pearce, J. A., & Tindle, A. G. (1986). Geochemical characteristics of collision zone magmatism. Geological Society, London, Special Publications, 19(1), 67–81.

Jiao, S. T., Zhang, Q., Zhou, Y. Z., Chen, W. F., Liu, X. Y., & Gopalakrishnan, G. (2018). Progress and challenges of big data research on petrology and geochemistry. Solid Earth Sciences, 3, 105–114.

Liu, X. L., Zhang, Q., Li, W. C., Yang, F. C., Zhao, Y., Li, Z., . . . Jiao, S. T. (2018). Applicability of large-ion lithophile and high field strength element basalt discrimination diagrams. International Journal of Digital Earth, 11(7), 752–760.

Liu, X. Y., Zhang, Q., & Zhang, C. L. (2017). A discussion on the tectonic setting of global Cenozoic andesite. Chinese Journal of Geology, 52(3), 649–667.

Luo, J. M., Zhang, Q., Song, B. T., Wang, X. W., Yang, Z. M., Zhao, Y. Q., & Liu, S. Y. (2017). Application of integrated geophysical and geochemical data processing to metallogenic target zone quantitative prediction and optimization. Bulletin of Mineralogy, Petrology and Geochemistry, 36(6), 886–891.

Luo, Y., Yuan, F. L., Jin, W. J., Liu, X. L., Zhang, C. Z., Liu, X. Y., . . . Yu, H. J. (2018). The comparison of global picrite and Archean komatiite: Implications from the whole data model. Chinese Journal of Geology, 53(4), 1267–1284.
Meschede, M. (1986). A method of discriminating between different types of mid-ocean ridge basalts and continental tholeiites with the Nb, Zr, Y diagram. Chemical Geology, 56(3–4), 207–218.

Mullen, E. D. (1983). MnO-TiO₂-P₂O₅: A minor element discriminant for basaltic rocks of oceanic environments and its implications for petrogenesis. Earth and Planetary Science Letters, 62(1), 53–62.

Pearce, J. A. (1975). Basalt geochemistry used to investigate past tectonic environments on Cyprus. Tectonophysics, 25(1–2), 41–67.

Pearce, J. A. (1976). Statistical analysis of major element patterns in basalts. Journal of Petrology, 17(1), 15–43.

Pearce, J. A. (1982). Trace element characteristics of lavas from destructive plate boundaries. Andesites, 525–548.

Pearce, J. A. (1983). Role of the subcontinental lithosphere in magma genesis at active continental margins. Journal of the Electrochemical Society, 147(6), 2162–2173.

Pearce, J. A. (2003). Supra-subduction zone ophiolites: The search for modern analogues. Special Papers-Geological Society of America, 373, 269–294.

Pearce, J. A., & Cann, J. R. (1973). Tectonic setting of basic volcanic rocks determined using trace element analyses. Earth and Planetary Science Letters, 19(2), 290–300.

Pearce, J. A., & Gale, G. H. (1977). Identification of ore-deposition environment from trace-element geochemistry of associated igneous host rocks. Geological Society London Special Publications, 7(1), 14–24.

Pearce, J. A., Lippard, S. J., & Roberts, S. (1984a). Characteristics and tectonic significance of super-subduction zone ophiolites. Geological Society London Special Publications, 16(1), 77–94.

Pearce, J. A., Harris, N. B. W., & Tindale, A. G. (1984b). Trace element discrimination diagrams for the tectonic interpretation of granitic rocks. Journal of Petrology, 25(4), 956–983.

Pearce, J. A., & Norry, M. J. (1979). Petrogenetic implications of Ti, Zr, Y, and Nb variations in volcanic rocks. Contributions to Mineralogy & Petrology, 69(1), 33–47.

Pearce, J. A., & Peate, D. W. (1995). Tectonic implications of the composition of volcanic ARC magmas. Annual Review of Earth and Planetary Sciences, 23, 251–285.

Vermeeesch, P. (2006). Tectonic discrimination of basalts with classification trees. Geochimica Et Cosmochimica Acta, 70(7), 1839–1848.

Vermeeesch, P. (2013). Statistical significance does not equal geological significance: Reply to comments on “Lies, damned lies, and statistics (in Geology)”. Eos Transacions American Geophysical Union, 92(8), 66.

Wang, J. R., Chen, W. F., Zhang, Q., Jiao, S. T., Yang, J., Pan, Z. J., & Wang, S. (2017). Preliminary research on data mining of NMORB and E-MORB: Discussion on method of the basalt discrimination diagrams and the character of MORB’s mantle source. Acta Petrologica Sinica, 33(3), 993–1005.

Wood, D. A. (1980). The application of a Th-HF-Ta diagram to problems of tectonomagmatic classification and to establishing the nature of crustal contamination of basaltic lavas of the British Tertiary Volcanic Province. Earth and Planetary Science Letters, 50(1), 11–30.

Wood, D. A., Joron, J. L., & Treuil, M. (1979). A re-appraisal of the use of trace elements to classify and discriminate between magma series erupted in different tectonic settings. Earth and Planetary Science Letters, 45(2), 326–336.

Workman, R. K., & Hart, S. R. (2005). Major and trace element composition of the depleted MORB mantle (DMM). Earth and Planetary Science Letters, 231(1–2), 53–72.

Zhai, M. G., Yang, S. F., Chen, N. H., & Chen, H. L. (2018). Big data epoch: Challenges and opportunities for geology. Bulletin of Chinese Academy of Sciences, 33(8), 825–831.

Zhang, C. Z., Zhang, Q., Jia, S. T., Li, C. D., & Zhang, W. F. (2017). The application of a Th-HF-Ta diagram to problems of tectonomagmatic classification and to establishing the nature of crustal contamination of basaltic lavas of the British Tertiary Volcanic Province. Earth and Planetary Science Letters, 50(1), 11–30.

Zhang, C. Z., Zhang, Q., Jiao, S. T., Li, C. D., & Chen, W. F. (2017). Granite and continental tectonics, magma thermal field and metallogenesis. Acta Petrologica Sinica, 33(5), 1524–1540.

Zhang, Q., Jiao, S. T., & Lu, X. X. (2018). Discussion on causality and correlation in geological research. Acta Petrologica Sinica, 34(2), 275–280.
Zhou, Y. Z., Chen, S., Zhang, Q., Xiao, F., Wang, S. G., Liu, Y. P., & Jiao, S. T. (2018). Advances and prospects of big data and mathematical geoscience. *Acta Petrologica Sinica*, 34(2), 255–263.

Zhang Qi  
*Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China*

Liu Xuelong  
*Faculty of Land and resource Engineering, Kunming University of Science and Technology, Kunming, China*  
✉️ xuelongliu@foxmail.com  
🌐 http://orcid.org/0000-0003-1372-8642