ISSN: 2664-9640

# Challenges of Big Data Mining and Integration in Prospecting Based on Geological Anomaly

Zhihong Zong [1], Shouhua Lai [2,*]

[1]Hunan Nonferrous Metal Huangshapinng Mining Co., Ltd, Chenzhou 424421, Hunan, China

[2]Minmetals Exploration and Development Co., Ltd, Beijing 100010, China

## Abstract

**Mineral resources and related big data contain a lot of hidden information and useful knowledge, which are very important for regional geological exploration and metallogenic prognosis. However, if the useful information cannot be fully and effectively extracted, then these data will become meaningless digital symbols, which will result in a great waste of resources and greatly devalue the data obtained by spending a lot of manpower, material resources and financial resources. Big data mining technology, which has emerged in recent years, has the potential to solve the above problems and has become the frontier research direction in the field of spatial information. While, it's challengeable for big data mining and integration in geological prospecting on making the most of useful geological anomaly information. Taking big data mining as the starting point, and geological and mineral database as the application objects, the research status and challenges on big data mining in geological prospecting is discussed and analyzed, following the implementation process of geological anomaly big data mining in prospecting is expounded, which shows the significance and trend of big data mining in geological prospecting.**

## Keywords

**Geological prospecting, big data mining, geological anomaly.**

## 1. Introduction

Mineral resources are related to all aspects of the country's economic construction and national economy and people's livelihood. Minerals provide raw materials and power for various industries. Therefore, it is of great significance to understand and evaluate my country's resource distribution and reserves [1]. The mining and integration of the indication information in the deep-level prospecting is the key to the successful discovery of concealed and deep minerals [2]. The current focus of prospecting is shifting to covered areas and deep prospecting [3]. Under such complex geological conditions, mineral prospecting faces many difficulties [4]. Due to the influence of the overburden and the deep burial of the deposit, the prospecting information is weak, and the complex geological background and multiple phases of mineralization lead to mixed mineralization information, and the changes in the preservation environment of the deposit lead to lack or incomplete prospecting information [5]. The traditional mineral exploration theories and methods based on genetic model and prospecting indicator combination can no longer meet the needs of deep prospecting in the coverage areas characterized by weak information, mixed information and incomplete information [6]. There is an urgent need for innovative methods on deep prospecting and medium-deep prospecting in the coverage areas [7]. The hierarchical prospecting is a feasible way with great potential on the basis of information mining and integration [8].

Following the development of cloud computing, Internet and Internet of Things, big data technology has once again brought about great changes in information technology [9]. Data

ISSN: 2664-9640

mining technology has been widely used in financial industry, retail industry, catering industry, telecommunications and other industries, and has brought good economic benefits to people [10]. Based on the favorable information of geological prospecting reflected by airborne geophysical exploration, the calculation of metallogenic geology is complex, and with this as a measurement index, the geological anomaly areas favorable to polymetallic exploration are circled. In the face of massive, dynamic and uncertain data, how to realize data processing and how to quickly and real-time mine valuable information from complex big data, form a knowledge system, complete product release and realize data sharing can no longer be realized by traditional technologies. At the same time, contemporary geological disasters, groundwater pollution and sustainable exploitation of mineral resources are exacerbating geological anomaly. The application of big data technology is challengeable but conducive to the scientific prospecting with geological anomaly data, which can also promote the informatization construction of geological work.

## 2. Challenges faced by big data mining in geological prospecting

The basic theory of geological anomaly mainly refers to the geology or geological assemblage which is obviously different from the surrounding environment in mineral composition, material composition structure, structure or genetic order. China's regional geological anomaly zones and local geological anomaly zones are all distributed on the boundary between massive symmetric arc geotectonic systems or in the east and west wing outer fold zones and the east and west wing inner fold zones, which have the common characteristics of zonal distribution in space and concentrated occurrence in time. These geological structures control the output of various minerals. Practice has proved that the search for geological anomaly is of great guiding significance to the search for large and super-large deposits and metallogenic belts.

A large number of training samples are required for the deep learning model to get better mineral prediction results. However, mineralization is a rare geological event, and the number of known deposits is scarce, which cannot meet the requirements of supervised learning for training samples, resulting in low accuracy and poor generalization ability of the model. How to build a large number of training samples is a challenge to deep prospecting based on deep learning. Data enhancement and transfer learning methods are often used in deep learning models to solve the problem of insufficient training samples. Among them, a large number of training samples are obtained by shooting limited targets from different angles. In the field of image recognition, affine transformation, image processing methods and data enhancement methods such as adding noise are often used to construct training samples. Geological prospecting big data has specific spatial characteristics, such as stratigraphic dip, structural strike, etc., which makes the transformation methods such as rotation and displacement possibly change the geological connotation of the original data. Therefore, it is challengeable to choose suitable data enhancement methods for constructing the training set samples of geological anomaly big data in prospecting.

## 3. Integration method of big data mining in geological prospecting

### 3.1. Technical process of data mining integration

Economic development and advancement in science and technology have promoted the development and progress of all walks of life, which will inevitably generate more and more data. Faced with massive amounts of data, people expect to be able to unearth useful information from them, with the goal of increasing the utilization rate of existing data and reducing blindness in future work. A lot of researches on data mining have been conducted at home and abroad, and big data mining and traditional data mining have been discussed in depth through various aspects such as data complexity, statistical basis, discoverable knowledge

ISSN: 2664-9640

types and algorithm processes. Big data mining is different from conventional transaction data mining, which increases the spatial scale dimension than the usual transaction data mining. The geometric features and spatial relationships of graphics in large databases are generally not directly stored in the database, but hidden in the graphic data of multiple layers. Special spatial operation and analysis are needed to summarize and learn useful attributes. The multi-source geological and mineral point source database model for geospatial data mining is shown in Figure 1.

Due to the particularity of the geological industry, the geological environmental data collected in various regions in reality usually has the characteristics of wide data sources, heterogeneity, ambiguity, redundancy, incompleteness, noise, randomness, and complex data. Although the development history of big data mining is not long, its development speed is very fast. In order to solve the practical problems faced by various research fields, researchers have invented many new algorithms or made a lot of improvements and innovations on the basis of the original algorithms, which promotes the research and development of big data mining algorithms. Typically, the technical process of spatial data mining includes regression analysis, clustering and classification, correlation analysis, principal component analysis, space analysis, visualization, and so on. However, there are also many differences in the algorithms that need to be used in the research and the effects of the algorithms against different application fields.

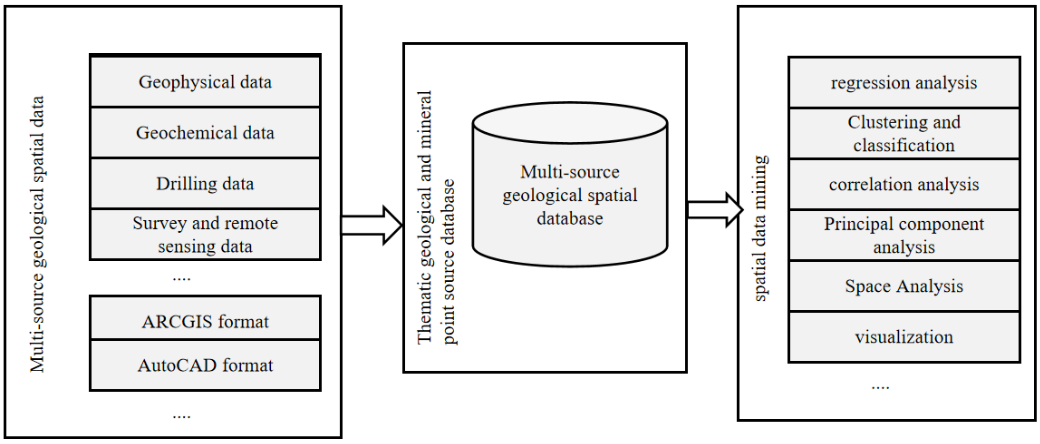


Fig. 1 Multi-source geological and mineral point source database model for geospatial data mining

## 3.2. Data application selection

Geological big data selection refers to extracting the target data set of geological big data mining for mining tasks in multi-source geological big database. The purpose is to identify the data set to be analyzed and narrow the processing range, which cannot only improve the efficiency of geological big data mining, but also ensure the quality of mining. If we can analyze the correlation among geological anomaly environment data through data mining technology and summarize its weight indicators, it will promote the accuracy of environmental assessment and prospecting based on geological anomaly data.

In the selection of geological big data, it is necessary to restrict the target data, select the data that meets the conditions, and retain the data that may be of interest by excluding the data that are not of interest. Therefore, it is necessary to determine the appropriate filtering strategy for the selection of geological big data, so as to ensure the quality of the target data participating in mining. After selection, the geological big data participating in geological big data mining can be divided into two types: graphic data and attribute data. Graphic data mainly includes stratum distribution layer, fault structure layer, mineral layer, remote sensing image layer, gravity and aeromagnetic anomaly layer, etc. Attribute data includes fault structure properties, fault

ISSN: 2664-9640

structure trends, mineral types, stratigraphic lithology, etc. The selected geological big data must be preprocessed and sorted into a suitable form before it can be used by the geological big data mining method.

# 4. Big data analysis and storage in Geological Prospecting

## 4.1. Big data analysis

For the distributed computing problems such as sorting and query of massive geological data, parallel computing technology is adopted for geological big data, including offline parallel computing technology, distributed resource management technology and memory computing technology. The whole task is divided into several sub tasks, which are completed by different nodes on the basis of ensuring sufficient fault tolerance of intermediate data in the computing process, and then the intermediate results generated by the subtask are integrated to generate the final query results of the whole task, so as to realize the distributed high-performance computing and automatic parallelization of massive geological data.

In addition, geological big data visualization technology has broad application prospects in the fields of engineering geological exploration, mineral resources exploration, mine design and development, geological disaster exploration and treatment, water conservancy and hydropower engineering design and national defense engineering construction. The key technologies of geological big data visualization include reasonable data structure, storage and rapid scheduling of geological big data, digital fast modeling technology, fast dynamic updating technology of digital geological bodies, fast free vector shearing technology, fast dynamic modeling technology and diversified spatial analysis technology. Complex geological structure expression and rapid dynamic modeling method of geological bodies are still the research focus of geological big data visualization in the future. Further research on knowledge-driven and ontology will be an effective way to solve these problems. The flow of geological big data combing mode is shown in Figure 2.

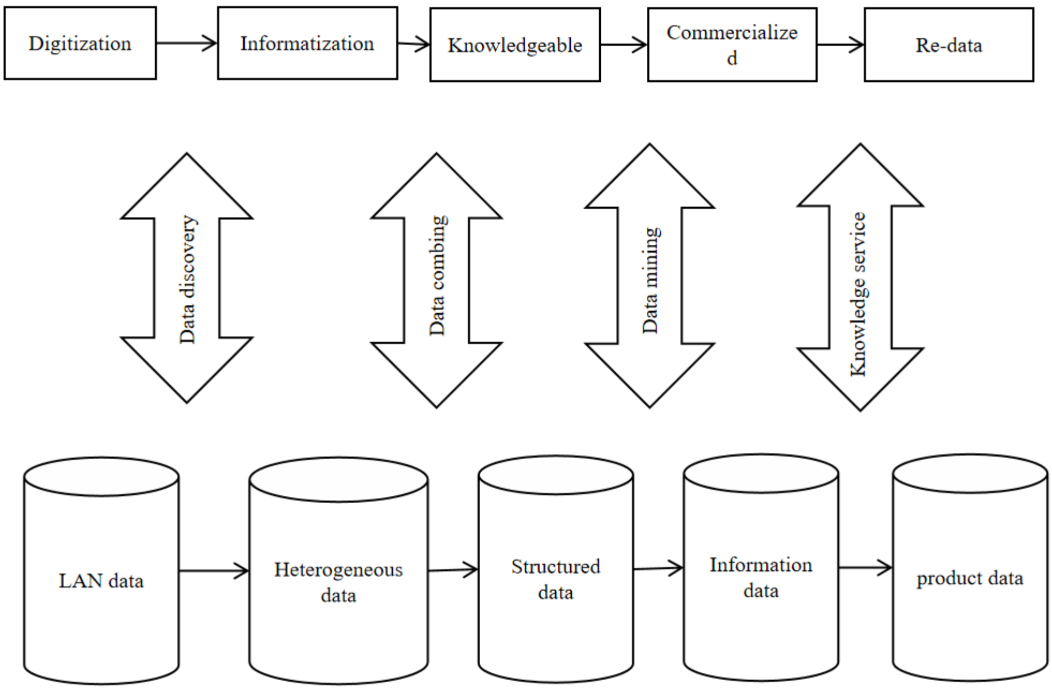


Fig. 2 Flow of geological big data combing mode

Geology is closely related to the problems faced by earth sciences. The long-term accumulation of various types of geological data and the large amount of digital land and resource-related data that can be collected on the Internet constitute the overall geological big data. Through the

ISSN: 2664-9640

data discovery module, local area network data is located and updated, and the local area network data is initially transformed into heterogeneous data. The data combing module designs the index data structure for the local area network geological data and establishes the index database. At the same time, it designs and realizes the file transfer according to the file transfer protocol. With comprehensive processing on the information obtained from the data mining module, it provides useful infoemation and data update storage for various application needs, and then completes knowledge services.

### 4.2. Big data storage

The storage mode of big data includes traditional data storage mode and distributed storage mode. Structured, semi-structured, and unstructured data are stored in column or row-column hybrid storage mode to realize distributed storage. Big data storage based on memory instead of disk, and strict sequence, makes use of the uncertain data management system and its technology, and constructs the uncertain relationship model, so as to realize the direct storage of dynamic and uncertain data.

The choice of storage database and the distributed storage and parallel computing of geological big data are the cores of the geological big data processing process. It is necessary to choose a suitable data model and data storage method according to the different application requirements of geological big data. Compared with common transaction data storage, big data storage increases the spatial scale dimension with higher complexity, and the geometric features and spatial relationships of graphics in large databases are generally not stored directly in the database, but are implicitly stored in multiple layers of the graphics data. As thus, special spatial operations and analysis are required to summarize and learn useful attributes.

## 5. Conclusion

Big data mining is a new field that has developed rapidly with more and more attention. The application of big data mining and integration on geological prospecting faces many challenges, but no doubt has great potential to improve the efficiency of geological prospecting. In the process of geological anomaly big data mining in prospecting, firstly the data enhancement and data standardization are needed to build a data mining source database, then carry out data mining modeling, model evaluation and selection are carried out to deploy the model, and obtain useful information from the data mining of geological anomaly big data. The regional metallogenic prediction and deep prospecting based on geological anomaly big data mining makes full and effective use of massive multi-source geological big data, improves the efficiency and accuracy of mineral prediction, and reduces the cost of mineral exploration. By use of geological big data mining technology to discover potential and valuable information, rules and knowledge from massive geological data is of great significance to prospecting decision-making.

## References


[1] Zuo Renguang. Deep mineralization information mining and integration based on deep learning [J]. Bulletin of Mineral and Rock Geochemistry, 2019, 38(1):9.

[2] Li Min, Fu Jie, Chen Anshu, et al. Connotation of geological big data mining based on cloud computing [J]. China Mining, 2019, 28(A01):5.

[3] Huang Shaofang, Liu Xiaohong, Sun Ling, et al. Preliminary discussion on geological data information integration and service in the era of big data [J]. China Mining, 2016, 25(2):3.

[4] Yang Yi, Zhi Qingquan, Deng Xiaohong, et al. Anomalous field extraction based on layered earth response and TEM inversion technology of current-carrying loop-well [J]. Advances in Geophysics, 2018, 33(5):8.

ISSN: 2664-9640

[5] Shu Ningkai, Wang Xinwen, Song Liang, et al. Multi-attribute fusion fracture prediction of glutenite based on seismic imaging -- A case study of the third member of Shahejie Formation in the northern steep slope of Chexi area in Jiyang Depression [J]. Oil and Gas Geology and Recovery, 2016, 023(002):57-61.

[6] Zuo Renguang. Exploration geochemical data mining and weak anomaly identification [J]. Earth Science Frontiers, 2019, 26(4):9.

[7] Wu Chonglong, Liu Gang, Wang Lizhe, et al. Ideas and methods of urban geological ambient intelligence supervision based on big data [J]. Geological Bulletin of Science and Technology, 2020, 39(1):7.

[8] Huang Shaofang, Liu Xiaohong. Thoughts on the application of geological big data and the development of geological informatization [J]. China Mining, 2016, 25(8):5.

[9] Sun Zhenchuan, Li Fengyuan, Chu Changhai. Analysis method of the relationship between shield tunneling and geology based on big data [J]. Tunnel Construction (Chinese and English), 2020(2):162-169.

[10] Shi Jianmin, Shi Shaoshan, Jiangshan, et al. Correlation analysis between annular structure and geochemical anomaly and its prospecting significance [J]. Geology and Resources, 2016, 25(2):5.