Research on Intelligent Identification of Ore Minerals Based on CART Algorithm

Ying Shen and Yongyi Zhao
Shenyang Normal University, Software College, No.253 Huanghe North Street, Huanggu District, Shenyang City, Liaoning Province.
Email: 295211039@qq.com

Abstract. The identification of ores and minerals is essential in geology, and the use of data mining to identify and classify ores in the era of widespread use of big data will help its development more effectively. This article takes the pyrite simulation data as an example and based on the ore mineral color, hardness, relative density, and electrical conductivity as the characteristic values. It is built in the CHDC model to use the CART algorithm to classify and identify the pyrite ore. After a lot of training on the data, the recognition model tends to be stable. The probability of correct discrimination in the CHDC model based on the CART algorithm's confusion matrix is 93.8%. However, due to the complex composition of ore minerals, the component analysis cannot be performed in a short time, so the component characteristic value is not taken. The ore returned by the sensor is different in color under different light sources, which affects the accuracy of judgment. Using the CART algorithm in the CHDC model can judge the identification of the ore, thereby completing the intelligent identification of the ore mineral.

1. Introduction
In the identification of ore minerals, the color characteristics of ore minerals, the optical properties of ore minerals, and the mechanical properties of ore minerals are generally used to identify and classify ore minerals. General ore identification is dividing into two types. One is the preliminary judgment of the human eye when collecting ore minerals, and the discrimination is based on the shape and physical properties of the ore minerals. One is the use of instruments to analyse the types of ore mineral components in an indoor environment. There are many methods, such as polarized light microscope identification and chemical analysis. The appearance characteristics and physical properties of different ore minerals are very different. The identification of ores and minerals is the premise and basis of geological investigation in modern geology. For ore identification, the properties of the ore field and the distribution characteristics of regional mineral deposits can be determined. And determining the type of ore can find out the specific properties of the ore and the relationship between the minerals in the ore. Finding out the type of ore minerals has great guiding significance for rock mineral identification and ore dressing.

Therefore, mineral identification is the top priority in geology. In the traditional identification, geologists perform the identification outdoors and the appraisers of each appraisal center perform the appraisal one by one in the microscope. Easily fatigued, unable to maintain objectivity and may cause errors to affect the discrimination. The use of machines for intelligent judgment can largely avoid this situation and increase the objectivity of the results. The core application technologies of big data and mathematical geoscience include high-dimensional data dimensionality reduction, image data processing, infinite data stream mining, machine learning, association rule algorithms and
recommendation system algorithms, etc. This paper uses pyrite as an example to study the data mining of ore data information using the CARD algorithm.

2. Related Research
Data mining technology is a new technology based on machine learning. The essence of data mining is to mine useful knowledge and undiscovered laws from a large amount of unordered unstructured data. The CART algorithm is a commonly used algorithm in data mining that can perform classification and prediction. The biggest advantage of the CART algorithm over the neural network algorithm is that it has a fast convergence speed. Using the CART algorithm, users can mine data from the collected data. The required set of related rules can achieve the purpose of data mining. In addition, the computational complexity of the CART algorithm is relatively low, and it has relatively fast convergence. At present, the CART algorithm is widely used in various professions.

In recent years, relevant domestic scholars have carried out a lot of research work on using data mining technology to assist the development of geology. In 2017, Professor Zhou Yongzhang and Professor Li Peixing published an article "Research Background and Progress of Big Data and Intelligent Deposit Models" in the Bulletin of Mineralogy, Petrology and Geochemistry, detailing the connection between big data and geological deposits and using Bayesian network for Modeling solves problems encountered in geology.

Based on the CART algorithm, Shi Da et al. Used the eigenvalues to establish a more accurate classification and prediction model for road icing disasters. Based on the CART algorithm, Zhang et al. Established a more accurate classification and prediction model for typhoon turning and landing based on the data corresponding to whether the typhoon path turned and the typhoon path landed. Based on the inspiration of "Experimental Research on Intelligent Recognition of Ore Minerals Under the Mirror Based on Deep Learning", this paper proposes the use of CART algorithm for ore mineral identification research.

3. Related Method
The use of CART algorithm for ore mineral identification mainly includes data preprocessing, data modeling, data training, model evaluation and other links. Figure 1 is the flow chart of ore mineral identification.

3.1. Analysis Method and Process
The goal of this article is to achieve the intelligent identification of ore minerals. There are three indicators for identifying the types of ore minerals. The color characteristics of the ore minerals, the optical properties of the ore minerals, and the mechanical properties of the ore minerals correspond to
Color, Hardness, Density, and Conductivity for intelligent identification of ore minerals, referred to as the CHDC model.

In the CHDC model, the color of the ore returned by the sensor is different under different light sources, so after the ore color is converted to RGB format, the range of color values is appropriately extended. The corresponding value of the pyrite color in the RGB color is \# B87333, and its range is set with \# B87333 as the center. The left and right values are each set to \#feedcf and \# 383604. The hardness, relative density and conductivity of ore minerals are detected by the respective sensors to return data, and then pre-processed.

3.2. Data Preprocessing

Mineral colors are processed into RGB colors after computer processing. RGB colors are stored using hexadecimal numbers. Color values smaller than the range of color values are converted to 1, and values within the range of color values are converted to 2, which are greater than the range the value is converted to 3. There are three types of conductivity classification: conductor, semiconductor, and insulator, which are recorded as numbers 1, 2, and 3, respectively. The hardness range of pyrite is 6 ~ 6.5, and the relative density is 4.9 ~ 5.2. However, the hardness characteristic value and relative density characteristic value data have the characteristics of many decimal places in the data, and the value range is not uniform. Therefore, data transformation is required during the preprocessing process.

There are three main types of data normalization:

- Min-max Normalization
  Min-max normalization is a linear transformation of the original data, mapping the value of the value to \[0, 1\]
  \[
  x' = \frac{x - \text{min}}{\text{max} - \text{min}}
  \]

- Z-Score
  The z-score normalization also becomes the standard deviation standardization. The processed data has a mean value of 0 and a variance of 1.
  \[
  x' = \frac{x - \bar{x}}{\sigma}
  \]

- Decimal Scaling Normalization
  The attribute value is mapped to \([-1, 1]\] by moving the decimal point of the attribute value. The number of decimal places moved depends on the maximum value of the absolute value of the attribute value.
  \[
  x' = \frac{x}{10^k}
  \]

Use the above three methods to perform data conversion and normalization on the hardness data and relative density data of ore minerals.

3.3. CART Algorithm

The CART algorithm is an algorithm in the form of a binary tree. CART determines the sample classification rules by constructing a tree structure. A multi-level, multi-node tree structure is generated based on the breadth-first principle for the sample data, reflecting the generic relationship of the samples. CART modeling includes the following two key technologies:

CART algorithm flow:

- When the obtained result meets the conditions for stopping the splitting of the process, the splitting is stopped;
- When the result does not meet the stopping splitting conditions of the process, the minimum Gini index splitting is selected;
- Repeat the first two steps recursively until the binary tree stops splitting

CART binary splits the feature attributes and divides the simulated experimental data set into two sub-sample sets, T1 and T2. Then continue to use binary segmentation for the two subsets respectively. And so on and so on, the process of recursion is continuously repeated, and it is concluded that each non-leaf node generated in the end has two corresponding left and right branches generated. The
CART algorithm uses the minimum GINI index as its root node. For discrete samples, the CART algorithm checks all subsets and uses the subset with the smallest GINI index in all subsets as the GINI index for the corresponding attribute:

\[
\text{GINI}(T) = 1 - \sum_{i=1}^{m} p_i^2
\]

(1)

\[
\text{GINI}(T) = \frac{T_1}{T} \text{GINI}(T_1) + \frac{T_2}{T} \text{GINI}(T_2)
\]

(2)

3.4. Simulation Results
The CART algorithm model constructed using specific features is used to train the simulated data after data preprocessing. The simulation results are shown in the confusion matrix shown in Figure 2. The confusion matrix formed by the CART algorithm yields a 93.8% chance of correct identification. This experiment shows that the CART algorithm model can have a good accuracy.

![Figure 2](image)

4. Experimental Data and Results Analysis

4.1. ROC Curve Verification
According to the training research of the experimental simulation data, it is concluded that the CART algorithm model has a higher recognition and classification accuracy rate. In order to further evaluate the recognition and classification performance of the constructed model in the study, the ROC evaluation method was used to evaluate the simulated test samples based on the CART model using the eigenvalue data, as shown in the ROC evaluation curve shown in Figure 3.
As shown in Figure 3, the ROC evaluation curve under the CART algorithm is close to the top of the unit square, which has a good effect on the identification of ore minerals, and can obtain a set of recognition rules, which is of a certain scientific nature. CART algorithm is a classic and efficient decision tree algorithm in data mining. The use of CART algorithm to study the classification and classification of ore minerals can be better applied to the classification and classification of ore minerals. It also provides a non-linear classification model. New applied research ideas.

5. Conclusion
This paper uses the CART model to carry out data mining research on the identification and classification of ore minerals to achieve the effect of ore minerals identification and classification. Based on the characteristics of identification and classification, the division criteria are established to identify the existing simulation data, and the correct evaluation of the identification and classification of ores and minerals is obtained, which indicates that the big data technology has played a good role in large geological research. With the rapid development of big data technology, according to mining technology, more and more fields are involved in various professional fields, and the problems that help various industries to solve are getting wider and wider. On the basis of the continuous development and improvement of geological theory, data mining technology will have a greater role. With the continuous enrichment and accumulation of data in the field of geology, data mining technology will play a greater role in the field of geology.

6. References
[1] Yu Xiaotong, Xiao Fan, Zhou Yongzhang, Wang Yu, Wang Kail. Mining and Extraction of Silver-Gold Geochemical Anomaly Information in Panxijing Region, Western Guangdong [J]. Geology and Prospecting, 2019, 55 (01): 77-86.
[2] Xu Shuteng, Zhou Yongzhang. Experimental research on intelligent identification of ore minerals under the microscope based on deep learning [J]. Acta Petrologica Sinica, 2018, 34 (11): 3244-3252.
[3] Zhou Yongzhang, Wang Jun, Zuo Renguang, Xiao Fan, Shen Wenjie, Wang Shugong. Machine Learning, Deep Learning and Implementation Language in Geological Field [J]. Acta Petrologica Sinica, 2018, 34 (11): 3173-3178.
[4] Jiao Shoutao, Zhou Yongzhang, Zhang Qi, Jin Weijun, Liu Yanpeng, Wang Jun. Intelligent discrimination of the tectonic environment of the global gabbro big data based on the GEOROC database [J]. Acta Petrologica Sinica, 2018, 34 (11): 3189-3194.
[5] Liu Yanpeng, Zhu Lixin, Zhou Yongzhang. Convolutional neural network and its application in ore prospecting prediction: A case study of Zhaojikou lead-zinc deposit in Anhui Province [J]. Acta Petrologica Sinica, 2018, 34 (11): 3217-3224.
[6] Zhang Qi, Zhou Yongzhang. Big Data Helps Geology Take Off: The 11th Big Data Special "Order" of Acta Petrologica Sinica [J]. Acta Petrologica Sinica, 2018, 34 (11): 3167-3172.

[7] Chang Xu, Li Yijie, Liu Wanjun. Decision tree pruning optimization algorithm combining CDC and REP [J]. Computer Engineering, 2012, 38 (14): 32-34.

[8] Xu Linzhang, Zhao Qiang, Zhang Yanning. A Method of Constructing Decision Tree Based on FP_Tree Algorithm [J]. Computer Engineering, 2009, 35 (08): 53-55.

[9] Liu Kun, Liu Yezheng. Medical Data Analysis Based on Decision Tree [J]. Computer Engineering, 2002 (02): 41-43 + 78.

[10] Zhang Liang, Ning Yan. Two Improvements and Applications of CART Decision Trees [J]. Computer Engineering and Design, 2015, 36 (05): 1209-1213.