Prediction of Spontaneous Combustion Tendency Grade of Sulfide Ore Based on Decision Tree Combined Classifier

Nianping Liu, Siqi Huang and Xiaojun Xie
School of Environmental and Resources, Southwest University of Science and Technology, No. 59, Middle Section of Qinglong Avenue, Fucheng District, Mianyang, Sichuan Province, China.
Email: liunp@swust.edu.cn

Abstract. Based on the related theories of decision tree and its classifier, a decision tree combiner model for spontaneous combustion of sulfide ores was established. Based on the analysis of the influencing factors of spontaneous combustion of sulfide ore, the three most representative measured characteristics, such as oxidation rate, self-hot spot, and auto-ignition point, were selected as predictors of the decision tree classifier model. Using the actual measured data of spontaneous combustion of sulfide ore as a training sample, a decision tree model for establishing spontaneous combustion of sulfide ores was used to predict the spontaneous combustion tendency grades, and verified with other measured data that did not participate in training. The results show that the decision tree combined classifier model is simple and feasible, with high prediction accuracy and simplified data processing. It is one of the effective methods to solve the problem of predicting the spontaneous combustion grade of sulfide ore.

1. Introduction
With the rapid development of the social economy, China's mining in coal mines not only involves coal mining projects, but also includes the mining of many non-coal mines. In non-coal mines, high-sulfur mines are their main mine resources. In the development of outdoor non-coal mine production and production plans, the spontaneous combustion tendency of sulfide ore is an important indicator for the prevention and control of fire in high-sulfur mines, which can reduce the economic losses of mine fires [1]. The prediction of spontaneous combustion tendency of sulfide ore is an indicator for guiding the actual production operation of non-coal mines. It is especially important to study the prediction of spontaneous combustion tendency of sulfide ore. In the prediction of the degree of spontaneous combustion of sulfide ore, domestic scholars have adopted many research methods, such as Fisher discriminant analysis theory [2]; using GA-BP neural network model [3]; using SVM model prediction [4]; using RS-cloud model Analysis [5]. The decision tree algorithm is an inductive learning algorithm based on an actual algorithm example. The algorithm infers the classification of the decision tree representation in its disordered case. The decision tree is a tree structure similar to the flow chart. The algorithm is easy to understand, the flow operation is simple and efficient, the numerical calculation is easy to extract and analyze, and the application of data type and regular attributes can be processed simultaneously [6]. Therefore, this paper draws on the idea of decision tree combination classifier, and proposes a decision tree model for the spontaneous combustion tendency of sulfide ore. It is verified by the measured data to provide a new method for predicting the spontaneous combustion tendency of sulfide ore.
2. Decision Tree Related Concepts and Algorithms

2.1. Constitutive Decision Tree and Its C4.5 Algorithm

The decision tree is a tree structure similar to a flowchart. As a predictive model, there is a certain mapping between object values in the decision tree and its attributes. The decision node is a classifier representation of the attributes of each object, and its branches correspond to different prediction results of the respective attributes, and each leaf node corresponds to a test value of each type of identifier, which represents a possible result of the final presentation of the decision tree. The construction phase of the decision tree can be expressed as: the construction phase and the pruning phase. For the construction phase, the overall planning of the decision tree structure is mainly completed, and the other pruning phase is mainly to perform a more precise pruning process on the decision tree completed in the construction phase. By pruning the individual data of the measured data, it is possible to clarify an accuracy. High decision tree model [7]. At present, the algorithms for constructing decision trees mainly include ID3, C4.5, CART, and SLIQ. However, ID3 can only solve the data of discrete attributes, and C4.5 not only can solve the data of discrete attributes, but also the ability to process data with continuous attributes.

ID3 introduces a conceptual algorithm for information gain. The basic point is that in the process of constructing the decision tree, information entropy is used to discriminate the attribute categories of each node, so that the attribute information acquired by each non-node node is the largest. After the subsets can be divided into smaller subsets using the determined attributes, the entropy value of the decision tree as a whole is the smallest at this time, and the average depth representation is also small. The purpose is to expand the classification speed and improve the accuracy. Rate [8].

The calculation formula of the information entropy Info(S) of the set S is

$$\text{Info}(S) = -\sum_{i=1}^{k} \left( \frac{\text{freq}(C_i, S)}{|S|} \times \log_2 \left( \frac{\text{freq}(C_i, S)}{|S|} \right) \right)$$  \hspace{1cm} (1)

Wherein $C_i$ represents a base classifier; $S$ represents a set. For the subset of decision numbers $T$ is divided into $n$ subsets according to the result representation of an attribute check $x$, and the desired result can be obtained by using the weighted sum of the entropies of its corresponding subset, and the numerical calculation formula is

$$\text{Info}_x(T) = \sum_{i=1}^{k} \left( \frac{|T_i|}{|T|} \times \log_2 \left( \frac{|T_i|}{|T|} \right) \right)$$  \hspace{1cm} (2)

Where $T_i$ represents a subset corresponding to the decision tree; $T$ represents a subset of the decision numbers.

The calculation formula of the information gain $\text{Gain}(X)$ is

$$\text{Gain}(X) = \text{Info}(T) - \text{Info}_x(T)$$  \hspace{1cm} (3)

The C4.5 calculation method is an improved algorithm for ID3, and its decision tree pruning is divided into pre-pruning and pruning. It is also possible to solve data for discrete attributes and continuous attributes.

If there is a training set $T$, the $n$ different values of the discrete attribute data $x$ are divided into $T_1$, $T_2$, ..., $T_n$ different training sets, and the information gain rate $\text{Gain ratio}(x)$ is determined by using $x$ to $T$ respectively.

$$\text{Gain ratio}(x) = \frac{\text{Gain}(x)}{\text{Split}_x \text{Info}(x)}$$  \hspace{1cm} (4)

Among them,

$$\text{Split}_x \text{Info}(x) = -\sum_{i=1}^{n} \left( \frac{|T_i|}{|T|} \times \log_2 \left( \frac{|T_i|}{|T|} \right) \right)$$  \hspace{1cm} (5)
The difference between \textit{C4.5} and \textit{ID3} is that \textit{C4.5} uses the gain ratio \text{Gain ratio}(x) to construct the decision tree model, while \textit{ID3} uses \text{Gain}(x) to construct the decision tree model. The constructiveness of \textit{C4.5} is greater than \textit{ID3}, while distinguishing for each node attribute, the criterion for detecting the attribute is also determined by determining the attribute sample having the largest gain ratio.

2.2. Decision Tree Combiner
The combination of decision tree combiners can increase the efficiency of decision classification, improve its branch attributes and numerical accuracy. The basic idea of combinatorial learning is to construct multiple decision combination classifiers using raw data samples, and then predict the results of unprocessed sample data when it is discriminatively classified \cite{9}.

The general process of combining learning methods:
(1) Let \( D \) be the original data sample set, \( k \) be the decision combination classifier number, and \( Z \) be the test sample set;
(2) For \( i = 1 \) to \( k \) do;
(3) Create a training set \( D_i \) from \( D \);
(4) Create a base classifier \( C_i \) by \( D_i \);
(5) End for;
(6) For each test sample \( x \in Z \) do;
(7) \( C^*(x) = \text{Vote}(C_1(x), C_2(x), \ldots, C_k(x)) \);
(8) End for.

Currently, the most commonly used decision-making combination classifier is the Bagging algorithm.

The process of Bagging is shown in Figure 1: In the model test phase, for the test sample set \( S \), \( S_i \) is generated by the Bootstrap phase, and the decision combination classifier \( C_i \) is constructed with \( S_i \) as the sample set; the sample data is tested with the respective combination classifier for its \( x \) prediction, and based on the final screening result, determines the attribute classification of the training test set \( x \).

\begin{figure}
\centering
\includegraphics[width=\textwidth]{bagging_process.png}
\caption{Bagging combination process.}
\end{figure}

3. Bagging-based Decision Tree Combination Classifier
Bagging stabilizes data by combining multiple weak decision classifiers into a strong decision combiner \cite{10}. Therefore, this paper uses the Bagging algorithm to construct the decision combiner, and its subset decision classifier is realized by the \textit{C4.5} algorithm.

4. Prediction Model of Spontaneous Combustion Tendency of Sulfide Ore Based on Decision Tree Combiner
4.1. Analysis of Factors Affecting the Spontaneous Combustion Tendency of Sulfide Ore
The spontaneous combustion tendency of sulfide ore is an important indicator for fire prevention in high sulfur mines. In this paper, with reference to the correlation research on the spontaneous combustion of domestic sulfide ore, the three characteristics of sulfide ore low temperature oxidation
weight gain rate ($X_1$), self-hotspot temperature ($X_2$) and auto-ignition temperature ($X_3$) are selected for the spontaneous combustion tendency of sulfide ore. The decision-making factors of classification $^{[11]}$.

4.1.1. Low temperature oxidation weight gain rate of ore samples. Sulfide ore is prone to oxidation reaction under low temperature and humid environment conditions, which increases the quality of the actual measurement and affects the correctness of the sample data. The contact of the ore sample with air in a humid environment will result in an increase in its mass; the absorption of oxygen per unit time will increase and the oxidation progress will be accelerated.

4.1.2. Self-hotspot temperature of the ore sample. The self-hot spot of the ore sample is the temperature at which the temperature of the ore sample itself exceeds the ambient temperature by 5°C to 10°C. Sulfide ore is exposed to air in a humid environment and undergoes an oxidation reaction, which easily releases heat. Under low temperature conditions, the oxidation reaction rate is slow, and it is not easy to generate heat or less heat, which is not easy to affect the ambient temperature. Under high temperature conditions, the oxidation reaction rate of the ore sample increases rapidly, which is highly prone to high heat. Its own temperature will eventually be higher than the ambient temperature, resulting in spontaneous combustion. Whether the ore sample spontaneously ignites easily depends on the hot spot temperature, so it is necessary to control its self-heating temperature.

4.1.3. Auto-ignition temperature of the sample. The spontaneous ignition point of the ore sample is the ambient temperature at which the flue gas combustion occurs when the ore sample reacts. When high-temperature ore heat is generated during the oxidation reaction, the self-ignition phenomenon occurs when the temperature reaches the temperature of the self-ignition point of the mineral itself. Therefore, whether the self-ignition of the ore sample is related to the hot spot temperature is also closely related to its own auto-ignition temperature.

4.2. Case Analysis of Spontaneous Combustion Tendency Prediction of Sulfide Ore Based on Decision Classifier

4.2.1. Sulphurized ore spontaneous combustion tendency classification. Referring to the results of the literature on the correlation of sulfide ore, the self-ignition propensity grade is divided into three levels after comprehensive analysis $^{[12]}$.

4.2.2. Decision tree prediction of spontaneous combustion tendency of sulfide ore. The relevant measured data of the ore sample is selected as a sample $^{[11]}$. See Table 1 for details.
Table 1. Spontaneous combustion tendency of sulfide ore classification of training samples.

| Ore sample number | Oxidation weight gain rate $x_1/$% | From hotspot $x_2/$C | Spontaneous ignition $x_3/$C | Comprehensive judgment | Decision tree decision result |
|-------------------|-----------------------------------|----------------------|-----------------------------|------------------------|-------------------------------|
| 1                 | 9.50                              | 60                   | 296                         | Spontaneous combustion | Spontaneous combustion (I)    |
| 2                 | 1.20                              | 127                  | 453                         | Easy to self-heat and not easy to spontaneously ignite | Easy to self-heat and not easy to spontaneously ignite(II) |
| 3                 | 2.10                              | 270                  | 422                         | Easy to self-heat and not easy to spontaneously ignite | Easy to self-heat and not easy to spontaneously ignite(II) |
| 4                 | 14.70                             | 140                  | 270                         | Spontaneous combustion | Spontaneous combustion (I)    |
| 6                 | 0.90                              | 243                  | 301                         | Easy to self-heat and not easy to spontaneously ignite | Easy to self-heat and not easy to spontaneously ignite(II) |
| 7                 | 3.20                              | 400                  | 414                         | Not easy to spontaneously ignite | Not easy to spontaneously ignite(III) |
| 8                 | 1.60                              | 400                  | 425                         | Not easy to spontaneously ignite | Not easy to spontaneously ignite(III) |
| 9                 | 9.30                              | 56                   | 233                         | Spontaneous combustion | Spontaneous combustion (I)    |
| 10                | 0.45                              | 400                  | 435                         | Not easy to spontaneously ignite | Not easy to spontaneously ignite(III) |
| 11                | 1.56                              | 220                  | 408                         | Easy to self-heat and not easy to spontaneously ignite | Easy to self-heat and not easy to spontaneously ignite(II) |

The decision judgment result is shown in Figure 2. First, by judging the sample data from the hot spot $x_2$, when $x_2$ is greater than or equal to 348.5, the sulphuric ore spontaneous combustion tendency level decision tree is judged to be less susceptible to spontaneous combustion (III). If $x_2$ is less than 348.5, the sample data of the oxidation weight gain rate $x_1$ is discriminated. When $x_1$ is less than 6, the sulfide ore spontaneous combustion tendency level decision tree is determined to be easy to self-ignite and not self-igniting (II), if $x_1$ is greater than or equal to 6 o'clock, the decision tree of the spontaneous combustion propensity grade of sulfide ore is judged to be easy to self-ignite (I).

Through the analysis of the classification of 11 groups of ore samples, the determination results of the spontaneous combustion tendency of sulfide ore are consistent with the results of the comprehensive judgment method. Through the decision tree to mine the data, it can reduce the error of the traditional data discrimination method, enhance the accuracy and practicability of its level judgment, and improve the accuracy of the spontaneous combustion tendency level prediction of sulfide ore.
Figure 2. Sulfide ore self-ignition tendency grade decision tree decision result.

In order to verify the accuracy of the decision tree decision results, the two sets of ore samples are selected for correlation verification. See Table 2 for details.

Table 2. Comparison of test sample determination results with decision judgement.

| Ore sample number | Decision discriminating factor | Comprehensive judgment | Decision tree decision result |
|-------------------|--------------------------------|------------------------|------------------------------|
|                   | Oxidation weight gain rate $x_1$/% | From hotspot $x_2/^\circ$C | Spontaneous ignition $x_3/^\circ$C | Spontaneous combustion (I) Easy to self-heat and not easy to spontaneously ignite (II) |
| 1                 | 10.80                           | 146                    | 239                          | Spontaneous combustion |
| 2                 | 2.20                            | 230                    | 437                          | Easy to self-heat and not easy to spontaneously ignite |

The two sets of ore samples to be verified are brought into the decision tree process of Figure 2, and the decision tree decision results as shown in Table 2 are obtained, which is found to be consistent with the results of the comprehensive decision method. Therefore, the decision tree decision method can be applied to the prediction of the spontaneous combustion tendency level of sulfide ore, which has high application value.

5. Conclusions

(1) Using decision tree combined classifier correlation analysis theory, the three characteristics of sulfide ore low temperature oxidation weight gain rate, self-hotspot temperature and spontaneous combustion point temperature are used as the decision factors for the classification of sulfide ore spontaneous combustion tendency, which can improve the accuracy of its classification. Sexuality, its sample data is tested, and its test results are consistent with the comprehensive judgment results.

(2) According to the decision-making rules, the data from the hot spot $x_2$ sample is first analysed. When $x_2$ is greater than or equal to 348.5, the decision tree of spontaneous ignitability of sulfurized ore is judged to be difficult to self-ignite (III); if $x_2$ is less than 348.5, it is increased by oxidation. The weight ratio $x_1$ is discriminated. When $x_1$ is less than 6, the decision is judged to be easy to self-ignite and not self-igniting (II). If $x_1$ is greater than or equal to 6, the decision is judged to be easy to self-ignite (I).

(3) The classification of spontaneous combustion tendency of sulfide ore is determined by decision tree analysis method. The discriminant result is consistent with the actual comprehensive judgment result of the ore sample. The decision tree is easy to operate and provides a new method for the classification and discrimination of the spontaneous combustion tendency of sulfide ore. A new way to judge easy operation.

6. Acknowledgments

The authors thank the supported by Sichuan Science and Technology Program (Project ID: 19GJHZ0103) and the grant of the China Scholarship Council.
7. References

[1] Changnian Xia and Wu Chao. Study of risk assessment technique for evaluating the spontaneous combustion of sulfide ore in mining stope [J]. Fire Safety Science, 2005(02): 106-110+58

[2] Hanhua Hu, Zheng Liu, Zijun Li and Tiantian Cui. Fisher discriminant analysis to the classification of spontaneous combustion tendency grade of sulphide ores [J]. Journal of China Coal Society, 2010, 35(10):1674-1679

[3] Fuqiang Yang, Guangning Liu and Lele Guo. GA-BP neural network model and its application to spontaneous combustion tendency classification of sulfide ores [J]. Journal of Natural Disasters, 2015, 24(04): 227-232

[4] Feng Guo. Parameter optimization for support vector machine model of spontaneous combustion tendency classification of sulfide ores [J]. Hu Nan Non Fereous Metals, 2017, 33(02): 5-7+14

[5] Zhichao Li, Keping Zhou and Yun Lin. Comprehensive evaluation on spontaneous combustion tendency of sulfide ore based on RS-cloud model [J]. Journal of Safety Science and Technology, 2017, 13(09): 126-131

[6] Peiling Chen. Research on the algorithm optimization of decision trees classification [D]. Central South University, 2007

[7] Xin Shi. Research of decision tree classification and its application to tax assessment [D]. Ocean University of China, 2004

[8] Junlei Li, Shaohua Teng and Wei Zhang. Prediction of atmospheric temperature based on multi-classifiers of decision tree [J]. Journal of Guangdong University of Technology, 2014, 31(04): 54-59

[9] Li Song. Research on classifier ensemble based on decision tree [D]. Xidian University, 2012

[10] Xiaodong Ma. Optimizing voting process of random forests algorithm based on weighted decision trees [D]. Central China Normal University, 2017

[11] Fuqiang Yang, Chao Wu, Zijun Li, Wei Pan and Hui Liu. Grade classification of spontaneous combustion tendency for sulfide ores based on distance discriminant analysis theory [J]. Journal of China Coal Society, 2010, 35(12): 2111-2115

[12] Zijun Li. Investigation on the mechanism of spontaneous combustion of sulphide ores and the key technologies for preventing fire [J]. Central South University, 2007, 10:21-25.