Rough Set-Fisher Discriminant Analysis Method for Prediction of Classification of Rock Burst Risk

Jian Ping*
Teaching and Research Department of Social Construction and Ecological Civilization, CPC Liaoning Provincial Party School, Shenyang China

*Corresponding author email: 1187076493@qq.com

Abstract. This paper introduces the rough set theory, combined with the Fisher discriminant method to more accurately predict the risk level of rock burst. Firstly, based on the example of Yanshitai coal mine in Chongqing, we determine the influencing factors of rock burst risk; then, the evaluation indicators are reduced by rough set theory to obtain the key attributes; finally, the reduced key attributes are used as the discriminant factors of the rough set-Fisher discriminant model to test. According to the research results, the rough set-Fisher discriminant analysis method eliminates redundant indicators and reduces the number of indicators after reducing the evaluation indicators, achieving efficient prediction of the risk level of rock burst.

1. Introduction

Rock burst is an extremely changeable phenomenon of dynamic instability of “coal-rock” system[1]. How to accurately predict the risk of rock burst in time is the most urgent consideration in coal mine safety production. Scholars at home and abroad have put forward empirical analogy analysis method, drilling chip method, etc. [2]. Besides, many scholars present rock burst prediction model, such as the Generalized Regression Neural Network (GRNN) prediction model[3], AdaBoost integrated neural networks[4], PSO-SVM models[5], Fisher discriminant analysis model[6] etc. There are also scholars based on multiple indicators to study rock burst, such as Yin Guangzhi etc. [7] selected the 10 main factors in the mining process from many aspects, combined with neural network to establish a rock burst prediction and hazard classification model.

But above methods have some shortcomings, it is difficult to obtain high accuracy classification effects and so on[8]. The rock burst of coal mine is a combination of multiple factors, and the data is limited, so, the author intends to construct a multi-factor, small sample, and non-linear prediction model. This paper aims at constructing the model and checking the accuracy, and the outcome shows that the model has high accuracy. The novelty of this paper is introducing the rough set theory, enhancing the prediction accuracy, which can not work in a separate method of FDA.

2. Fisher Discriminant

The central idea of Fisher discriminant is projection, which keeps classes and classes as far away as possible, then the discriminant analysis functions are obtained based on the premise that the inner-class distance is the smallest and the inter-class distance is the largest. Fisher discriminant method has no strict requirements for data samples, only a few discriminant functions can directly identify, which is simple and widely used. However, when using the Fisher discriminant method, the redundant indicators independent of classification can not be eliminated before the key indicators are extracted, which makes the final prediction accuracy insufficient.
3. Rough Set Theory
Rough set theory can simplify and solve problems efficiently by analyzing random and fuzzy data which are unstable, imprecise and incomplete.

3.1. Attribute Reduction
Attribute reduction is the core of rough set theory. On the premise of ensuring the same classification ability of knowledge base, the irrelevant and redundant information is reduced to obtain accurate and efficient results.
Suppose $U$ is a domain of discourse with an equivalent relation cluster $P$ on it, $r \in P$, if there is a $\text{ind}(P) = \text{ind}(P-r)$, the $P$ is redundant, otherwise it is dominant. If the $Q = P - \{ r \}$ is independent, there is a reduction $Q$.

Based on the Equal Frequency Binning in the Rosetta software, we can discrete the selected data samples and obtain the crucial attribute.

3.2. Reduction Algorithm Description
If $U/C = \{ x_1, x_2, \cdots, x_n \}$, $U/D = \{ Y_1, Y_2, \cdots, Y_m \}$, define the support $k$ of the decision attribute $D$ to the condition attribute $C$ as:

$$k = \gamma_c(D) = \frac{|\text{POS}_c(D)|}{|U|} = \frac{1}{U} \sum_{i=1}^{m} |\gamma_c(D)|, \quad Y_i \in U/D$$  \hspace{1cm} (1)

Formula (1) represents the degree of dependence of decision attribute $D$ on condition attribute $C$.

Since different conditional attributes $C_i$ have different importance to the total set of conditional attributes $C$, the importance of the conditional attribute subsets $C_i$ to the decision attributes $D$ is defined as:

$$\sigma_{CD}(C_i) = \gamma_c(D) - \gamma_{C-C_i}(D)$$  \hspace{1cm} (2)

Firstly, calculate the frequency of attributes: $f(a) = \frac{\lambda_{ij}}{D_{ij}}$, and $a \in D_{ij}, i, j = 1, 2, \cdots, n$. $n$ is the number of samples; then, for each attribute of $a \in E$, calculate the attribute importance of $\sigma_{CD}(a)$ according to formula (1)~(3), and select the largest attribute of it, adding it to $R$; based on $R = R + \{ a \}, E = E - \{ a \}$, calculate whether $\text{POS}_R(D) = \text{POS}_C(D)$ is correct, if so, end; otherwise, turn to the (2); finally, output the $R, R$ is attribute reduction. The set of the $R$ of attribute reduction sets can be expressed as: $R = \{ R : R \in C, \text{POS}_R(D) = \text{POS}_C(D) \}$, so the termination condition of the algorithm is $\text{POS}_R(D) = \text{POS}_C(D)$, to ensure the minimum attribute set of classification quality. Similar approach has been implemented in the study of network percolation, which can be saw in the seminal paper Multi-type directed scale-free percolation.

3.3. Weight Coefficient Calculation
Select the measured data sample, define the conditional attribute set $C$ and the decision attribute set $D$, establish the initial decision table, apply the rough set theory to reduce the redundant index, and determine the optimal index combination; then calculate the support of the decision attribute and the importance of the attribute based on formula (1)~(3); finally calculate the weight coefficient of the $\alpha_i$ based on formula (4).

$$\alpha_i = \frac{\gamma_{C-C_i}(D)}{\sum_{i=1}^{n} \gamma_{C-C_i}(D)}, (i = 1, 2, \cdots, n) \hspace{1cm} (4)$$

4. Rough Set-Fisher Discriminant Model
The index weight coefficient matrix $W$ of sample $X$ is obtained, by calculating the weight coefficients, then it is substituted into the Fisher discriminant analysis model, after the importance of different evaluation indexes is obtained, and the rough set-Fisher discriminant model is established:
\[ d^2_p(X, G) = (X - u)^T W \sum^{-1} W (X - u) \]  

In the formula, the weight matrix of the \( W = \text{diag}(W_1, W_2, \cdots, W_m) \) is diagonal matrix, and \( W_i \in [0, 1] (i = 1, 2, \cdots, m) \) is the weight factor of each evaluation index in the Fisher discriminant function. Combined with the relevant principles of Fisher discriminant, the discriminant criteria of the rough set-Fisher discriminant model are as follows:

\[ d^2(X, G_i) = \min_{1 \leq i \leq k} \{ d^2(X, G_i) \} \quad x \in G_i \]

If the back-substitution discrimination of training samples is high and the discrimination of the new sample is accurate, the criterion is feasible.

5. Rough Set-Fisher Discriminant Analysis Model for Prediction of the Hazard Level of Rock Burst and Application

5.1. Sample Selection of Coal Mine Engineering Examples

Taking Yanshitai coal mine in Chongqing as an example, based on the coal mine's authigenic geological structure and mining conditions, the coal thickness \((X_1, X_2, \cdots, X_5)\), structure condition \((X_6)\), inclination angle \((X_7)\), inclination angle change \((X_8)\), coal thickness change \((X_9)\), gas concentration \((X_10)\), and sound of coal cannon \((X_{10})\) are selected as the prediction indicators of rock burst. Such indicators as coal thickness \((X_1)\), inclination \((X_2)\), buried depth \((X_3)\), gas concentration \((X_4)\), structural condition \((X_5)\), inclination change \((X_6)\), coal thickness change \((X_7)\), pressure relief \((X_8)\), and sound of coal cannon \((X_{10})\) are state parameters, which need to be quantified before prediction, and the quantification rules are shown in table 1. Based on the years data of rock burst in Yanshitai Coal Mine, 35 samples with reference were extracted, of which 23 samples were used as training samples of rough set Fisher discriminant model (Table 2), and the remaining 12 samples were used as samples to be tested (Table 3).

| Assignment | Qualitative impact indicators |
|------------|-----------------------------|
| Structure condition \((X_4)\) | Inclination angle change \((X_5)\) | Coal thickness change \((X_6)\) | Roof management \((X_8)\) | Pressure relief \((X_9)\) | Sound of coal cannon \((X_{10})\) |
| 0 | Simple | No | No | No or poor | No | No |
| 1 | General | Less | Less | General | General | Less |
| 2 | More complex | Greater | Greater | Better | Better | More |
| 3 | Complexity | Huge | Huge | Good | Good | - |

5.2. Attribute Reduction

Based on the Equal Frequency Binning in the Rosetta software, we should discretize large irregular data before attribute reduction, then by computer simulation we can obtain the least attribute reduction of 7 elements of \( RED(C) = \{X_1, X_2, X_4, X_5, X_7, X_9, X_{10}\} \).

5.3. Establish rough set-Fisher Discriminant Model

The rock burst is divided into four grades: micro impact \((G_1)\), weak impact \((G_2)\), medium impact \((G_3)\) and strong impact \((G_4)\), considering the degree of rock burst. Firstly, based on the rough set reduction, the 7 highly correlated indicators are used as the discriminant factors of the rough set-Fisher discriminant model, combined with the Fisher discriminant theory to derive the Fisher discriminant function; then the distance between the function value of the samples of the hazard level of rock burst to be measured and the central value of the 4 types of hazard grade categories is compared; finally determine which
category the sample to be tested belongs to.
In order to verify the feasibility of rough set-Fisher discriminant model, 23 groups of training sample
data were tested by the established model and the risk grade of rock burst was written into Table 2,
which was highly consistent with the actual classification results, and the prediction error rate is 0. As a
result, it is feasible to predict the hazard grade of rock burst by rough set-Fisher discriminant model in
practical engineering.

### Table 2. Initial data of training sample by rough set-Fisher discriminant model

| Serial number | Coal seam thickness $X_1$/m | Fall Angle $X_2$ (°) | Deep buried $X_3$/m | Structure $X_4$ | Inclination change $X_5$ | Change in coal thickness $X_6$ | Gas Concentration $nX_7$/m$^3$ | Roof management $nX_8$ | Pressure relief $X_9$ | Coal cann $nX_{10}$ | This method $G_1$ | Actual category |
|----------------|-----------------------------|---------------------|---------------------|-----------------|-----------------------|--------------------------|--------------------------|----------------|----------------|----------------|----------------|----------------|
| 1              | 1.3                         | 29                  | 530                 | 0               | 0                     | 0.07                     | 3                        | 3              | 0              | $G_1$          | $G_1$          |                |
| 2              | 1.2                         | 25                  | 542                 | 0               | 0                     | 0.24                     | 3                        | 3              | 0              | $G_1$          | $G_1$          |                |
| 3              | 1.4                         | 44                  | 560                 | 0               | 0                     | 0.09                     | 3                        | 3              | 0              | $G_1$          | $G_1$          |                |
| 4              | 3.0                         | 24                  | 573                 | 0               | 0                     | 0.36                     | 2                        | 3              | 0              | $G_1$          | $G_1$          |                |
| 5              | 0.8                         | 34                  | 553                 | 1               | 0                     | 0.15                     | 0                        | 2              | 1              | $G_2$          | $G_2$          |                |
| 6              | 1.2                         | 40                  | 490                 | 0               | 0                     | 0.20                     | 2                        | 0              | 1              | $G_2$          | $G_2$          |                |
| 7              | 1.4                         | 35                  | 480                 | 0               | 0                     | 0.36                     | 2                        | 0              | 1              | $G_2$          | $G_2$          |                |
| 8              | 1.2                         | 27                  | 490                 | 0               | 0                     | 0.64                     | 2                        | 2              | 1              | $G_2$          | $G_2$          |                |
| 9              | 2.6                         | 48                  | 752                 | 2               | 0                     | 0.48                     | 1                        | 1              | 1              | $G_3$          | $G_3$          |                |
| 10             | 2.8                         | 52                  | 730                 | 0               | 1                     | 1.54                     | 2                        | 2              | 1              | $G_3$          | $G_3$          |                |
| 11             | 3.0                         | 78                  | 560                 | 1               | 3                     | 1.14                     | 2                        | 3              | 1              | $G_3$          | $G_3$          |                |
| 12             | 6.0                         | 30                  | 465                 | 1               | 1                     | 1.30                     | 1                        | 0              | 2              | $G_3$          | $G_3$          |                |
| 13             | 1.5                         | 65                  | 570                 | 1               | 3                     | 0.28                     | 1                        | 2              | 2              | $G_3$          | $G_3$          |                |
| 14             | 3.0                         | 35                  | 612                 | 2               | 0                     | 0.56                     | 2                        | 0              | 2              | $G_3$          | $G_3$          |                |
| 15             | 2.0                         | 35                  | 614                 | 1               | 0                     | 0.56                     | 1                        | 0              | 2              | $G_3$          | $G_3$          |                |
| 16             | 3.0                         | 55                  | 855                 | 3               | 2                     | 0.075                    | 1                        | 1              | 2              | $G_4$          | $G_4$          |                |
| 17             | 4.0                         | 52                  | 675                 | 3               | 2                     | 1.88                     | 0                        | 0              | 2              | $G_4$          | $G_4$          |                |
| 18             | 1.3                         | 73                  | 486                 | 3               | 3                     | 0.43                     | 1                        | 0              | 2              | $G_4$          | $G_4$          |                |
| 19             | 2.1                         | 67                  | 498                 | 3               | 3                     | 1.89                     | 0                        | 0              | 2              | $G_4$          | $G_4$          |                |
| 20             | 2.5                         | 65                  | 450                 | 3               | 2                     | 0.67                     | 1                        | 1              | 2              | $G_4$          | $G_4$          |                |
| 21             | 1.7                         | 60                  | 314                 | 3               | 3                     | 1.30                     | 0                        | 0              | 2              | $G_4$          | $G_4$          |                |
| 22             | 1.1                         | 47                  | 485                 | 3               | 3                     | 0.43                     | 1                        | 0              | 2              | $G_4$          | $G_4$          |                |
| 23             | 1.8                         | 54                  | 238                 | 3               | 1                     | 1.00                     | 0                        | 0              | 2              | $G_4$          | $G_4$          |                |

5.4. Rough set-Fisher Discriminant Model Validity Test
Based on the trained rough set-Fisher discriminant model, the other 12 groups of samples to be tested
are tested, and the results are written in Table 3. According to the results, the prediction sample results
are consistent with the actual classification results, thus the rough set-Fisher discriminant model
prediction results are better, which has practical reference significance in coal mines and other fields.
Table 3. Initial data of samples to be tested by rough set-Fisher discriminant model

| Serial number | Coal seam thickness X1 (m) | Fall Angle X2 (°) | Deep burial X3 (m) | Structure X4 | Inclination change X5 | Change in coal thickness X6 | Gas Concentration X7 (m³/min) | Roof management X8 | Pressure relief X9 | Coal cannon X10 | Discriminant results |
|---------------|-----------------------------|------------------|------------------|---------------|-----------------------|-----------------------------|-------------------------------|-------------------|-----------------|-----------------|------------------------|
| 1             | 1.6                         | 35               | 583              | 2             | 0                     | 3                           | 1.50                          | 3                 | 3               | 1               | G3, G3                 |
| 2             | 1.5                         | 35               | 530              | 0             | 0                     | 0                           | 0.56                          | 3                 | 3               | 0               | G1, G1                 |
| 3             | 1.6                         | 62               | 307              | 3             | 2                     | 2                           | 1.00                          | 0                 | 0               | 2               | G4, G4                 |
| 4             | 1.9                         | 59               | 542              | 1             | 2                     | 3                           | 0.25                          | 0                 | 0               | 1               | G3, G3                 |
| 5             | 1.8                         | 62               | 283              | 3             | 2                     | 3                           | 1.00                          | 0                 | 0               | 2               | G4, G4                 |
| 6             | 1.3                         | 44               | 570              | 0             | 0                     | 0                           | 0.66                          | 3                 | 3               | 0               | G1, G1                 |
| 7             | 2.2                         | 54               | 290              | 3             | 2                     | 2                           | 1.00                          | 0                 | 0               | 2               | G4, G4                 |
| 8             | 3.0                         | 34               | 475              | 2             | 2                     | 1                           | 0.42                          | 0                 | 0               | 2               | G3, G3                 |
| 9             | 3.2                         | 42               | 574              | 3             | 0                     | 0                           | 0.29                          | 0                 | 0               | 2               | G3, G3                 |
| 10            | 1.8                         | 62               | 283              | 3             | 2                     | 3                           | 1.00                          | 0                 | 0               | 2               | G4, G4                 |
| 11            | 1.3                         | 44               | 656              | 2             | 1                     | 3                           | 0.24                          | 1                 | 1               | 2               | G3, G3                 |
| 12            | 1.2                         | 40               | 553              | 2             | 2                     | 2                           | 0.49                          | 1                 | 2               | 2               | G3, G3                 |

6. Conclusions

In this paper, rough set theory is applied to reduce irrelevant redundant attributes, and it has a high accuracy and feasibility in predicting the rock burst hazard level in Yanshi coal mine. Based on the coal mine's own geological structure and mining conditions, this method is generally applicable to forecast the danger level of rock burst in fields such as coal mines. There is individual subjectivity in selecting evaluation indicators, and the selected data samples are typical and cannot fully express the nonlinear relationship between rock burst and evaluation indicators, so, in the future we need examine more ordinary samples, and it is necessary to further study how to predict the rock burst risk level efficiently and accurately.

References

[1] Wen Changping. Prediction and classification of impact ground pressure based on Bayes discriminant analysis [J]; and Journal of Natural Disasters 24(05):229-236. DOI:10.13577/j.jnd.2015.0526
[2] Shao Liangshu, Xu Bo. Prediction of Impact Ground Pressure Hazard Level Based on Fisher Discriminant Analysis [J]; and Metal mines,2015,(01):138-144.
[3] Wen Tingxin, Chen Xiaoyu, Yang Hongyu, Dou Rong, Tian Yuchen. Impulse ground pressure prediction [J] based on optimized Bagging-LSSVM model Chinese Journal of Safety Sciences,2017,27(06):121-126. DOI:10.16265/j.cnki.issn1003-3033.2017.06.021
[4] Sun Fengqi. The Application of AdaBoost Integrated Neural Network in Impulse Ground Pressure Prediction [J]; and Journal of Jilin University (Information Science Edition).
[5] Zhao Hongbo, Ru Zhongliang. PSO-SVM Model [J]. for Impulse Ground Pressure Prediction Journal of Rock Mechanics and Engineering,2007,(S1):3479-3483.
[6] Zhang Xiaoming, Lu Gang, Bi Jianwu. Prediction of impact ground pressure Fisher discriminant analysis [J]. Journal of Liaoqing University of Engineering and Technology (Natural Science Edition)(03):313-316.
[7] Yin Guangzhi, Dai Gao Fei, Yan he, Wei Zuonan. Genetic Neural Network Method for Impulse Ground Pressure Prediction [J]. Geotechnical Mechanics,2003,(06):1016-1020. DOI:10.16285/j.rsm.2003.06.034
[8] Bi Juan, Li Xijian, Chen Liu Yu. Fisher of R Factors for Prediction of Impact Ground Pressure Hazard Grade Chinese Journal of Safety Sciences, 2019, 29(12): 103-109. DOI: 10.16265/j.cnki.issn1003-3033.2019.12.017