Mine Fire Prediction Based on WEKA Data Mining

Xin Wang 2,a, Jian Hao 1,2, *, Jun Chen 2,b, Weijia Cheng 2,c

1State Key Laboratory Breeding Base for Mining Disaster Prevention and Control, Shandong University of Science and Technology, Qingdao 266590, China
2College of Mining and Safety Engineering, Shandong University of Science and Technology, Qingdao 266590, China

*Corresponding author e-mail: 1345046483@qq.com, awoxingwangao@163.com, 943890757@qq.com, 410649664@qq.com

Abstract. Mine fires are prone to cause catastrophic accidents such as gas accidents and coal dust explosions in gas-exposed coal mines, which seriously threaten the safe production of mining enterprises. In order to effectively reduce the probability of gas explosion in mines, an accurate prediction of possible fire hazards in Linhua Coal Mine is made: collecting data on fire occurrence indicators; using Weka software on the basis of rough set theory, selecting SVM classifier, BP neural network and J48 decision tree to obtain the accuracy of the samples to be tested; analyzing the detailed accuracy, confusion matrix and node error rate, and obtain the optimal algorithm. On this basis, a mine fire predictive control model is proposed to determine and control the hazard grade of mine fire source and effectively reduce the probability of coal mine fire occurrence. The establishment of this model greatly ensures the safe production of gas outburst mine.

1. Introduction

After the coal seam mining in the working face, with the passage of time, the residual coal in the goaf comes into contact with oxygen, and the temperature rises and oxidizes to produce CO2, CO and so on [1]. When the heat does not spread in time and reaches the ignition point, the phenomenon of spontaneous combustion of residual coal occurs. The mine fire is particularly strong in the gas outburst mine. Due to the high CO concentration and temperature in the goaf and the improper underground control measures, the air leakage occurs and the probability of fire is greatly improved [2]. The safety of underground workers can not be guaranteed, and coal mining enterprises are also facing a severe threat to production safety.

In order to effectively reduce the probability of goaf fire and reduce the serious consequences of coal mine fire accidents, relevant scholars have carried out a lot of research on mine fire monitoring and pre-control, which greatly reduced the probability of mine fire. Sun Jiping [3] carries on the real-time monitoring of mine fire by using mature monitoring system, and Du Zhenyu [4] establishes the extension superiority evaluation model of mine fire safety evaluation according to matter element and extension theory. Geng Xiaowei [5] proposed a weight determination method based on Skowron discernibility matrix for coal mine fire risk analysis. Wang Chun’e [6] scientifically identifies the dangerous sources of underground fire accidents in coal mines by using statistical and system engineering methods, and finally comprehensively evaluates and early warns them; Through statistical analysis of 35 typical fire accidents in recent two years, Wang Guiming[7] summarized seven main factors that induced mine fire.
Above research has made abundant achievements in coal mine fire safety evaluation and prevention, management and control, which provides rich experience for this paper. However, the above research does not involve the determination of the location of mine fire sources, so the study of mine fire prediction becomes more urgent. This study basically realizes the accurate prediction of mine fire, processes the sample data of mine fire index by Weka, and compares three kinds of prediction algorithms with high precision: BP neural network, SVM classification machine and J48 decision tree, then determine the optimal algorithm; on this basis, the mine fire predictive control model is established, and the fire hazard level is evaluated, which can effectively control the probability of mine fire. At the same time, it also provide some reference ideas for the prediction research in different directions in the future.

2. Selection of influencing factors of mine fire in working face
Lin Hua Mine is located in the southwest of Jinsha County, Guizhou Province, 10km away from the center of the county. The administrative division is under the jurisdiction of Xinhua Township and Xi Luo Township. The 20917 working face of coal mine is located in the second mining area of 9# coal seam. The average thickness of the working face is 3.5 m; the strike length is 1200 m, the trend length is 190 m; the roadway width is 4.2 m and the height is 3.5 m; the 9# coal seam is of gas outburst. Fig.1 shows the location of the mine and the plane of the 20917 face mining project.

![Fig. 1 Lin Hua mine location and 20917 floor plan](image)

Mine fire is affected by many factors, including CO concentration, temperature, O2 concentration and wind speed. Because the object of study is in 9 # coal seam, the difference between the depth of coal seam is not very big. Based on the actual situation of coal mine and the relevant experience of mine fire control at home and abroad, the author determines the following indicators affecting mine fire:
1) Location of measuring points \( m \): The position of measuring point refers to the position where monitoring instrument is placed in the roadway, that is, the length between the instrument and the cutting hole.

2) Temperature\(^{\circ}C\), Temperature of goaf after mining in working face.

3) Residual gas content \( m^3/t \), Gas content monitored after mining in working face.

4) Wind speed \( M^3/h \), The wind speed is mainly at 20917 working face studied.

5) \( O_2 \) concentration \( \% \), Oxygen content in goaf.

6) \( CO \) concentration \( ppm \), CO content in goaf.

3. Mine Fire Prediction Based on Weka Software

3.1. Sample Collection

In this paper, 30 samples of mine fire in Linhua Coal Mine are collected, as shown in Table 1. According to the actual situation of coal mine working face fire, the dangerous degree is divided into four categories: none, weak, medium and strong, which are expressed by 1, 2, 3 and 4 respectively.

| Sample serial number | Location of measuring points (m) | Temperature (\(^{\circ}C\)) | Residual gas content \( m^3/t \) | Wind speed \( M^3/h \) | \( O_2 \) concentration \( \% \) | \( CO \) concentration \( ppm \) | Dangerous degree |
|----------------------|---------------------------------|-----------------------------|-----------------------------|----------------------|-----------------------------|------------------------|------------------|
| 1                    | 1000                            | 67.677                      | 4.23                        | 4.3                  | 15.8531                    | 286.0087               | 1                |
| 2                    | 960                             | 45.795                      | 5.3                         | 3.2                  | 19.2985                    | 83.1796                | 1                |
| 3                    | 1050                            | 34.060                      | 4.98                        | 4.85                 | 20.0641                    | 137.2674               | 1                |
| 4                    | 1100                            | 45.477                      | 5.35                        | 2.24                 | 16.2359                    | 15.5700                | 1                |
| 5                    | 900                             | 38.501                      | 5.13                        | 2                   | 20.4469                    | 150.7893               | 1                |
| 6                    | 200                             | 76.557                      | 5.26                        | 6.15                 | 8.1966                     | 556.4475               | 2                |
| 7                    | 60                              | 69.263                      | 6.03                        | 5.1                  | 19.2985                    | 759.2765               | 2                |
| 8                    | 30                              | 39.452                      | 6.46                        | 4.3                  | 3.22                       | 1029.715               | 2                |
| 9                    | 250                             | 30.889                      | 5.78                        | 3.7                  | 6.6654                     | 218.3990               | 2                |
| 10                   | 790                             | 69.263                      | 5.34                        | 6                   | 10.4936                    | 759.2765               | 2                |
| 11                   | 360                             | 90.194                      | 5.61                        | 3                   | 19.2985                    | 353.6184               | 2                |
| 12                   | 190                             | 126.35                      | 4.76                        | 7.3                  | 6.6654                     | 759.2765               | 2                |
| 13                   | 810                             | 40.403                      | 6.54                        | 14                  | 10.4936                    | 353.6184               | 3                |
| 14                   | 460                             | 31.206                      | 6.73                        | 4.2                  | 6.2825                     | 218.3990               | 3                |
| 15                   | 700                             | 53.723                      | 7.23                        | 12.7                 | 10.1108                    | 421.2281               | 3                |
| 16                   | 450                             | 24.546                      | 7.31                        | 13.8                 | 8.5795                     | 313.0526               | 3                |
| 17                   | 670                             | 92.731                      | 7.66                        | 3.7                  | 16.2359                    | 83.1796                | 4                |
| 18                   | 490                             | 123.81                      | 8.01                        | 8.7                  | 9.3451                     | 691.6668               | 4                |
| 19                   | 620                             | 126.34                      | 7.52                        | 7.1                  | 17.7672                    | 218.3990               | 4                |
| 20                   | 500                             | 126.34                      | 7.70                        | 7                   | 5.1341                     | 759.2765               | 4                |
| 21                   | 580                             | 45.477                      | 7.94                        | 3                   | 12.7905                    | 218.3990               | 4                |
| 22                   | 530                             | 88.608                      | 7.44                        | 9.8                  | 8.5795                     | 1097.32                | 4                |
| 23                   | 780                             | 77.191                      | 5.22                        | 6.3                  | 13.1733                    | 353.6184               | ?                |
| 24                   | 1130                            | 53.406                      | 5.62                        | 3.5                  | 17.3844                    | 83.1796                | ?                |
| 25                   | 430                             | 31.206                      | 6.54                        | 4                   | 6.2825                     | 218.3990               | ?                |
| 26                   | 540                             | 121.59                      | 7.62                        | 11                  | 5.1341                     | 624.0571               | ?                |
| 27                   | 70                              | 38.183                      | 5.25                        | 2                   | 20.4470                    | 150.7893               | ?                |
| 28                   | 750                             | 38.183                      | 7.21                        | 8                   | 12.4077                    | 394.1842               | ?                |
| 29                   | 600                             | 120.95                      | 7.66                        | 11                  | 5.1341                     | 556.4475               | ?                |
| 30                   | 390                             | 69.580                      | 5.57                        | 3.2                  | 9.3452                     | 421.2281               | ?                |
3.2. Sample Data Preprocessing
Due to the different units between the data in Table 1, the difference between each variable becomes larger, and there is a large deviation for subsequent calculations. Therefore, in order to ensure the accuracy of the calculation and prevent the loss of information, it is necessary to normalize the original data.

For convenience, the location of the measuring point, temperature, residual gas content, wind speed, \(O_2\) concentration and CO concentration are set to \(x_1, x_2, x_3, x_4, x_5, x_6\), respectively, and the mine fire risk grade is set to \(Y\). The strong representation 4, the middle representation 3, the weak representation 2, and the none representation 1, are used as the category attributes of the classifier model to construct the model.

The first 22 samples were used as the training samples of the model, and the last 8 samples were used as the test samples. Firstly, the training samples are normalized, and the normalized data are discretized again by Weka software in order to remove one of the duplicates in the data. Then rough set is used for attribute reduction. When any index is removed, there are repeated data in the remaining data, then \(\text{ind}(A-A1) \neq \text{ind}(A)\), changes the original relationship in the decision table. So keep attributes \(x_1, x_2, x_3, x_4, x_5, x_6\), so the core of their property set is \(\{x_1, x_2, x_3, x_4, x_5, x_6\}\). Then, the final training sample is obtained by compiling the data by UltraEdit, the predicted sample is only normalized and discretized, and the final predicted sample is compiled by UltraEdit data. Finally, the two samples are imported into Weka software to predict different algorithms.

3.3. Three Algorithms for Mine Fire Prediction
Eight groups of samples were predicted and analyzed. Three algorithms, SVM, BP neural network and J48 decision tree, were selected to open the processed training data and tested by 10 fold cross validation.

1) SVM is implemented by SMO algorithm in Weka. SMO algorithm is mainly used to solve the optimization problem of object function of support vector machine [8]. It is equivalent to the maximum value of solving the following quadratic linear programming problem:

\[
\begin{align*}
\max_{\alpha} & \quad -\frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j (x_i \cdot x_j) \alpha_i \alpha_j + \sum_{i=1}^{l} \alpha_i, \\
\text{s.t.} & \quad \sum_{i=1}^{l} y_i \alpha_i = 0, \\
& \quad \alpha_i \geq 0, i = 1, \ldots, l
\end{align*}
\]

Its core is to introduce the kernel function \(K(x_i, x_j)\) and give a training set after it is introduced:

\[
T = \{(x_i, y_i), \ldots, (x_l, y_l)\} \in (R^n \times Y)^l
\]

In style \(x_i \in R^n, y_i \in Y = \{1, -1\}, i = 1, \ldots, l\). Then the corresponding decision function is:

\[
f(x) = \text{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b^*\right)
\]

The implementation path of SVM algorithm is weka.classifiers.function.smo. The predicted results are shown in Table 2.

| Serial number | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---------------|----|----|----|----|----|----|----|----|
| Mine fire     |     |     |     |     |     |     |     |     |
| predicted value | weak | nothing | during | strong | nothing | during | strong | weak |
| actual value  | weak | nothing | during | strong | nothing | strong | strong | weak |
| Result comparison | right | right | right | right | right | wrong | right | right |

2) BP neural network is realized by Multilayer Perceptron algorithm in Weka. BP network is a multi-layer feedforward network trained by error back propagation algorithm, It consists of input layer, hidden layer and output layer. [9]. Its structure is shown in Fig 2.
Fig. 2 Topology of BP Neural Network

Fig. 2 $x_1, x_2, ..., x_M$ is the input value of BP neural network, $y_1, y_2, ..., y_p$ is the predicted value of BP neural network. $w_{im}$ and $w_{ip}$ are weights of BP neural network. The relations between the input layer and the hidden layer and the output layer are as follows:

$$\text{hidden}[j] = f \left( \sum_{i=1}^{n} w_{ij}a_i - \theta_j \right)$$

(4)

$$\text{out}[l] = f \left( \sum_{j=1}^{m} w_{lj}b_j - r_i \right)$$

(5)

In the formula, $a_i$ is the input of point $i$; $b_j$ is the output of the node of the $j$ hidden layer; $w_{ij}$ is the weight of the input layer to the hidden layer; $r_i$ is the weight of the hidden layer to the output layer; and $\theta$ is the threshold of the hidden layer.

The implementation path of BP neural network algorithm is weka.classifiers. Multilayer Perceptron. The predicted results are shown in Table 3.

### Table 3. Analysis of BP Neural Network Prediction Results

| Serial number | 23   | 24   | 25   | 26   | 27   | 28   | 29   | 30   |
|---------------|------|------|------|------|------|------|------|------|
| Mine fire     | predicted value | weak | nothing | during | strong | nothing | during | strong | weak |
| Result comparison | right | right | right | right | right | wrong | right | right |

3) Decision tree is mainly C4.5 algorithm, which is implemented by J48 decision tree algorithm in weka. The algorithm is based on Occam razor: The smaller the decision tree, the better the larger the decision tree [10]. Nevertheless, the algorithm does not always generate the smallest tree structure, it's a heuristic algorithm. Occam's razor expounds a concept of information entropy:

$$I_k(i) = -\sum_{j=1}^{m} f(i, j) \log_2 f(i, j)$$

(6)

The implementation path of J48 decision tree algorithm is weka.classifiers.trees.j48. The predicted results are shown in Table 4.

### Table 4. J48 Decision Tree Prediction Results Analysis

| Serial number | 23   | 24   | 25   | 26   | 27   | 28   | 29   | 30   |
|---------------|------|------|------|------|------|------|------|------|
| Mine fire     | predicted value | weak | nothing | nothing | during | strong | nothing | during | strong | weak |
| Result comparison | wrong | right | right | right | right | strong | strong | nothing |

3.4. Prediction Accuracy Analysis

Observe the results of mine fire prediction by three algorithms in Table 2-Table 4: There are many contingency factors, which lead to the increase of errors. Further data processing is needed to select the
most accurate prediction methods. Therefore, the author focuses on three aspects: detailed accuracy, confusion matrix and node error rate.

1) The detailed accuracy of the three algorithms is obtained from the Detailed Accuracy By Class results. The main reference data are shown in Table 5 - Table 7. The closer the ROC Area is to 1, the better the diagnostic effect of the model is.

Table 5. Detailed accuracy of SVM algorithm

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---------|---------|-----------|--------|-----------|----------|-------|
| 1       | 0       | 1         | 1      | 1         | 0.875    | 1     |
| 1       | 0       | 1         | 1      | 1         | 0.894    | 2     |
| 0.5     | 0       | 1         | 0.5    | 0.667     | 0.875    | 3     |
| 1       | 0.167   | 0.667     | 1      | 0.8       | 0.917    | 4     |
| Weighted Avg | 0.875 | 0.042     | 0.917  | 0.875     | 0.867    | 0.948 |

Table 6. Detailed accuracy of BP neural network

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---------|---------|-----------|--------|-----------|----------|-------|
| 1       | 0       | 1         | 1      | 1         | 0.875    | 1     |
| 1       | 0       | 1         | 1      | 1         | 0.894    | 2     |
| 0.5     | 0       | 1         | 0.5    | 0.667     | 0.75     | 3     |
| 1       | 0.167   | 0.667     | 1      | 0.8       | 0.833    | 4     |
| Weighted Avg | 0.875 | 0.042     | 0.917  | 0.875     | 0.867    | 0.896 |

Table 7. J48 Decision Tree Classifier Detailed Accuracy

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---------|---------|-----------|--------|-----------|----------|-------|
| 1       | 0.333   | 0.5       | 1      | 0.667     | 0.833    | 1     |
| 0       | 0       | 0         | 0      | 0         | 0.75     | 2     |
| 0.5     | 0       | 1         | 0.5    | 0.667     | 0.917    | 3     |
| 1       | 0.167   | 0.667     | 1      | 0.8       | 0.917    | 4     |
| Weighted Avg | 0.625 | 0.125     | 0.542  | 0.625     | 0.533    | 0.854 |

As shown in Table 5-Table 7: TP Rate; SVM = BP Neural Network > J48 Decision Tree; FP Rate; J48 Decision Tree > BP Neural Network = SVM; Precision; SVM = BP Neural Network > J48 Decision Tree; Recall; SVM= BP Neural Network > J48 Decision Tree; F-Measure; SVM = BP Neural Network > J48 Decision Tree; The weighted sum of sample classification accuracy of the three algorithms is 0.948, 0.896 and 0.854. It can be concluded that the values of the three algorithms are all above 85%, which shows that the selected algorithm is suitable for mine fire prediction.

2) Table 8 shows the confusion matrix of three algorithms, and the diagonal value is the correct number of examples.

Table 8. Confusion Matrix of Three Algorithms

| SVM | BP Neural Network | J48 Decision Tree |
|-----|-------------------|-------------------|
| a b c | a b c | a b c |
| 2 0 0 | 2 0 0 | 2 0 0 |
| 0 2 0 | 0 2 0 | 0 2 0 |
| 0 0 1 | 0 0 1 | 0 0 1 |
| 0 0 2 | 0 0 2 | 0 0 2 |

The results of 8 groups of predicted samples in Table 8 were cross-validated by 10 folds. The correct examples of SVM, BP neural network and J48 decision tree are 7, 7 and 5 respectively. The accuracy expression relationship was 87.5% = 87.5% > 62.5%. SVM algorithm and BP neural network are more accurate than J48 decision tree.

3) Table 9 shows the node error rates of three algorithms.
Data from table 9 show that: Mean absolute error: SVM > J48 Decision Tree > BP Neural Network; Root mean squared error: J48 Decision Tree > SVM > BP Neural Network; Relative absolute error: SVM > J48 Decision Tree > BP Neural Network; Root relative squared error: J48 Decision Tree > SVM > BP Neural Network. Comprehensive analysis shows that SVM and J48 decision tree algorithm are better than BP neural network.

Generally speaking, among the three algorithms of SVM, BP neural network and J48 decision tree, SVM and BP neural network are close to each other in detail, which is better than J48 decision tree algorithm; On the confusion matrix, the accuracy of SVM and BP neural network is more than 85%, which is better than J48 decision tree algorithm; In terms of node error rate, SVM and J48 decision tree algorithm are better than BP neural network. Therefore, the SVM algorithm is the most accurate for coal mine fire prediction.

4. Establishment of prediction model
In order to better realize the accurate prediction and pipe hole of the mine fire in Linhua Coal Mine, the author establishes the coal mine fire control model to summarize its realization flow, and the flow chart is as follows:

1) step 1: analyze the influencing factors of coal seam fire in the working face, in order to ensure the accuracy of prediction, the selected factors should have a certain relationship with the occurrence of fire.
2) step 2: sort out the samples selected in the previous step and set the prediction level. For convenience, the influencing factor is set to X and the prediction grade is set to Y; then the selected sample data are processed by Weka software, and the samples are divided into training samples and prediction samples. The two samples are normalized, discretized and attribute reduced. Finally, the prediction methods are compared and analyzed.
3) step 3: analyze and compare the prediction results of the samples by the three algorithms, find out the differences, analyze the reasons, and formulate the fire extinguishing scheme for the areas with high level of prediction results.
5. Conclusion

1) By selecting the measurement samples of the mine fire in Linhua Coal Mine, the factors affecting the mine fire are classified, and the data processing is carried out on the WEKA platform. Finally, the accuracy of the three algorithms is 87.5\%, 87.5\% and 62.5\%, respectively.

2) Through the comparison of the detailed accuracy, confusion matrix and node error rate among the three algorithms, it is finally determined that the SVM classification machine under WEKA software is the most suitable for mine fire prediction.

3) Taking Linhua Coal Mine as the research object, this paper establishes a mine fire prediction and control model, which reduces the probability of mine fire occurrence in goaf, realizes the early warning and timely management of mine fire, and ensures mine safety.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant No.51804180, No.51574055), the key research plan of Shandong province (Grant No.2018GSF116007).

References

[1] Chen Chen Chen, Qi Qingjie. Development status and trend of mine fire prevention technology [J]. Coal technology, 2009, 28 (03): 3-5.
[2] Theory and Practice of Mine Fire Relief [M]. Coal Industry Press, Zhou Xinquan, Written by Wu Bing, 1996
[3] Sun Jiping, Sun Yanyu. Research on mine fire monitoring and trend prediction methods [J]. Industrial and mining automation, 2019, 45 (03): 1-4.
[4] Du Zhenyu, Yang Shengqiang, Wang Jun, et al. Evaluation method of mine fire safety based on extension superiority [J]. Mining safety and environmental protection, 2013, 40 (01): 107-110+114.
[5] Geng Xiaowei, Li Rong, Cui Tiejun. Weight analysis of influencing factors of coal mine fire [J]. Practice and understanding of mathematics, 2018, 48 (11): 109-117.
[6] Wang Chun'e. Risk early warning of underground fire accidents in coal mines [D]. Hebei University of Engineering, 2016.
[7] Wang Guiming, Zhang Jiaoyong, Gong Xuemin, et al. [J]. Research on evaluation methods of coal mine fire accidents [J]. Energy technology and management, 2010 (01): 52-54.
[8] Krol M, Florek A. Two-Stage Data Reduction for a SVM Classifier in a Face Recognition Algorithm Based on the Active Shape Model[M]// Computer Recognition Systems 4. 2011.
[9] Mazurowski M A, Habas P A, Zurada J M, et al. Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance.[J]. Neural Netw, 2008, 21(2-3):427-436.
[10] Wang Q, Garrity G M, Tiedje J M, et al. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy[J]. Applied and Environmental Microbiology, 2007, 73(16):5261-5267.