A Gas Prominence Prediction Model Based on Entropy-Weighted Gray Correlation and MCMC-ISSA-SVM

Liangshan Shao 1 and Yingchao Gao 2,*

1 Liaoning Institute of Science and Engineering, Jinzhou 121000, China; intushao@163.com
2 College of Business Administration, Liaoning Technical University, Huludao 125105, China
* Correspondence: 18041950831@163.com; Tel.: +86-18041950831

Abstract: To improve the accuracy of coal and gas prominence prediction, an improved sparrow search algorithm (ISSA) and an optimized support vector machine (SVM) based on the Markov chain Monte Carlo (MCMC) filling algorithm prediction model were proposed. The mean value of the data after filling in the missing values in the coal and gas prominence data using the MCMC filling algorithm was 2.282, with a standard deviation of 0.193. Compared with the mean fill method (Mean), random forest filling method (random forest, RF), and K-nearest neighbor filling method (K-nearest neighbor, KNN), the MCMC filling algorithm showed the best results. The parameter indicators of the salient data were ranked by entropy-weighted gray correlation analysis, and the salient prediction experiments were divided into four groups with different numbers of parameter indicators according to the entropy-weighted gray correlation. The best results were obtained in the fourth group, with a maximum relative error (maximum relative error, \( RE_{\text{max}} \)) of 0.500, an average relative error (average relative error, \( MRE \)) of 0.042, a root mean square error (root mean square error, \( RMSE \)) of 0.144, and a coefficient of determination (coefficient of determination, \( R^2 \)) of 0.993. The best predicted parameters were the initial velocity of gas dispersion (\( X_2 \)), gas content (\( X_4 \)), \( K_1 \) gas desorption (\( X_5 \)), and drill chip volume (\( X_6 \)). To improve the sparrow search algorithm (sparrow search algorithm, SSA), the adaptive t-distribution variation operator was introduced to obtain ISSA, and the prediction models of improved sparrow search algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm (MCMC-ISSA-SVM), sparrow search algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm (MCMC-SSA-SVM), genetic algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm (MCMC-GA-SVM) and particle swarm optimization algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm (MCMC-PSO-SVM) were established for coal and gas prominence prediction using the ISSA, SSA, genetic algorithm (GA) and particle swarm optimization (PSO) respectively. Comparing the prediction experimental results of each model, the prediction accuracy of MCMC-ISSA-SVM is 98.25%, the error is 0.018, the convergence speed is the fastest, the number of iterations is the least, and the best fitness and the average fitness are the highest among the four models. All the prediction results of MCMC-ISSA-SVM are significantly better than the other three models, which indicates that the algorithm improvement is effective. ISSA outperformed SSA, PSO, and GA, and the MCMC-ISSA-SVM model was able to significantly improve the prediction accuracy and effectively enhance the generalization ability.

Keywords: coal and gas prominence prediction; Markov Chain Monte Carlo (MCMC); improved sparrow search algorithm (ISSA); support vector machines (SVM); entropy-weighted gray correlation; variational arithmetic

1. Introduction

With the rapid development of industrial information and intelligence [1], coal output and consumption are growing day by day, and mining efforts are subsequently increasing,
making coal mine safety accidents frequent. Coal and gas protrusion is one of the most destructive hazards in coal mining, and the extent of the damage is not to be underestimated [2], which not only leads to casualties and affects the effective and safe production of the coal mining industry but also causes huge losses to the national economy. Therefore, the use of effective methods for predicting the risk of coal and gas protrusion not only provides the basis for the continued development of the coal industry but also provides some guidance for the safe production of coal mines.

Lu Yiyu et al. [3] developed a multifunctional physical simulation test system based on deep coal rock engineering and geological conditions to improve the accuracy of coal and gas protrusion prediction. The geological background of the coal seam environment generally includes four aspects: ground stress, geological and tectonic conditions, media structure, and gas, which are described in turn.

Ground stress is a contributing factor to the occurrence of coal and gas protrusions. The effect of ground stress is related to the magnitude of the ground stress; the greater the ground stress in the coal mining process, the greater the likelihood of coal and gas protrusion [4].

Geological and tectonic conditions are unstable due to the movement of the earth, and when the surrounding coal rock is affected, the equilibrium is disturbed, which has a direct impact on the coal seam and changes the state of the gas [5]. Geological and tectonic conditions are one of the most important factors in the occurrence of coal and gas protrusions.

The structure of the coal body is a manifestation of geological changes. The original form of the coal body can be broken by geological activity [6], and its original internal structure can be destroyed, reducing the strength of the coal body. The more severely the media structure is damaged, the more likely it is that coal and gas protrusions will occur.

Gas is present in coal seams. It is an essential factor in coal and gas protrusion. Conditions such as gas content and gas pressure are important factors in the occurrence of coal and gas protrusions; the higher the gas content and gas pressure, the greater the risk of coal and gas protrusions [7].

Scholars at home and abroad have carried out a lot of research work on coal and gas prominence [8–10], and based on this, some new prediction methods have been proposed. Research on the prediction of coal and gas prominence was first conducted using the static prediction method. Yuan Liang [11] et al. developed a coal and gas prominence simulation test system based on a comprehensive hypothesis, which improved the accuracy and repeatability of the prediction. Lei Yang [12] explored the mechanism of coal and gas protrusion and conducted experiments on coal and gas protrusion by constructing numerical models. Gui Fu [13,14] et al. analyzed the influencing factors of protrusion through coal and gas protrusion accidents and proposed methods to predict protrusion and prevent accidents. The dynamic forecasting method was later applied to coal and gas protrusions. Xueqiu He et al. [15] proposed a prominent dynamic real-time monitoring and early warning technique by dynamically and continuously monitoring electromagnetic radiation and other data to mine its relationship with prominent omens. Jupeng Tang et al. [16] for coal and gas protrusion power hazard prediction by acoustic emission monitoring technique; for improving the accuracy of protrusion prediction. Junhui Mou [17] analyzed the law of coal and gas prominence in coal mines and proposed a new method of sensitivity of prominence prediction index to monitor the dynamic phenomenon of coal mine outbursts. Xusheng Zhao [18] established an online integrated system and database for coal and gas prominence prediction based on characteristics such as gas and geological conditions to achieve real-time prediction of coal and gas prominence. In recent years, machine learning prediction methods have been applied to coal and gas protrusions. Baohe Zhu [19] identified the main influencing factors of prominence based on the combined action hypothesis and used a non-linear support vector machine prediction method to accurately predict whether coal and gas prominence would occur. Yuhong Wang [20] used gray correlation analysis to screen coal and gas prominence influencing factors for dimensionality reduc-
tion and established particle swarm optimization algorithms based on sub-dimensional evolution and quantum gate node neural network models for prediction experiments. Kai Wang [21,22] et al. explored the mechanism of coal and gas protrusion, analyzed typical cases, and used big data and deep learning techniques to achieve coal and gas protrusion prediction. Wang Wei [23] constructed a prediction model after determining the subjective and objective weights of the predictors and conducting experiments on hazard prediction in mines. Yaqin Wu [24] developed an improved genetic simulated annealing algorithm (genetic simulated annealing algorithm, GASA) and optimized back propagation (back propagation, BP) neural network algorithm prediction model and used the model for coal and gas prominence prediction to improve the prediction speed.

The data was collected using the Mining Engineering Collaborative Big Data Cloud Platform system. Analysis of the data revealed that there were missing values and that prediction models that did not account for missing data would compromise the accuracy of the model predictions. Accordingly, the use of the Markov chain Monte Carlo algorithm (MCMC) is proposed to fill in the missing values in the coal and gas prominence dataset in order to expand the valid dataset. Entropy-weighted gray correlation analysis was performed on the complete data set after filling. The entropy-weighted gray correlation for each indicator was calculated, ranked according to its magnitude, and divided into four groups of experiments with different numbers of indicator parameters. The data from each group of experiments was input into a support vector machine (SVM) model for prediction. The best predictor parameters were identified through screening. To prevent the sparrow search algorithm (SSA) from falling into local optimality, an adaptive t-distribution adaptive operator was introduced to improve the SSA, and the parameters of the SVM were optimized with the improved sparrow search algorithm (ISSA). The model of an improved sparrow search algorithm optimized support vector machine based on the Markov chain Monte Carlo filling algorithm (MCMC-ISSA-SVM) was established to improve the coal and gas prominence prediction capability, and its prediction results were compared with those of other models in order to verify the model’s performance. The main methods used are literature research, numerical simulation, and comparative experimental methods.

2. Data Collection and MCMC Fills in the Missing Data

A coal mine in Shanxi Province, China, is a high-concentration gas mine, and coal and gas protrusions occur frequently during the construction of this mine. Data on coal and gas prominence was collected using the Mining Engineering Collaborative Big Data Cloud Platform system to monitor the 405 and 406 excavation faces in the No. 3 mineable coal seam of this coal mine. The parameter indicators and data sets obtained are shown in Table 1.

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$X_8$</th>
<th>$X_9$</th>
<th>$X_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.00</td>
<td>15.50</td>
<td>0.36</td>
<td>7.78</td>
<td>0.29</td>
<td>3.20</td>
<td>351.00</td>
<td>2.40</td>
<td>36.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>15.70</td>
<td>0.36</td>
<td>7.76</td>
<td>0.19</td>
<td>3.40</td>
<td>366.00</td>
<td>2.40</td>
<td>62.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2.00</td>
<td>15.75</td>
<td>0.36</td>
<td>7.81</td>
<td>0.26</td>
<td>3.20</td>
<td>366.00</td>
<td>2.40</td>
<td>78.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2.00</td>
<td>15.56</td>
<td>0.36</td>
<td>7.30</td>
<td>0.34</td>
<td>3.80</td>
<td>367.00</td>
<td>2.40</td>
<td>47.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>15.80</td>
<td>0.36</td>
<td>7.05</td>
<td>0.23</td>
<td>3.13</td>
<td>364.00</td>
<td>2.40</td>
<td>21.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>15.58</td>
<td>0.36</td>
<td>7.68</td>
<td>0.23</td>
<td>3.27</td>
<td>364.00</td>
<td>2.40</td>
<td>3.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>17.40</td>
<td>0.36</td>
<td>7.98</td>
<td>0.26</td>
<td>3.33</td>
<td>365.00</td>
<td>2.40</td>
<td>35.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>17.52</td>
<td>0.36</td>
<td>8.68</td>
<td>0.26</td>
<td>3.13</td>
<td>365.00</td>
<td>2.40</td>
<td>35.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>17.43</td>
<td>0.36</td>
<td>8.77</td>
<td>0.22</td>
<td>3.07</td>
<td>368.00</td>
<td>2.10</td>
<td>11.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>15.32</td>
<td>0.36</td>
<td>8.35</td>
<td>0.25</td>
<td>3.33</td>
<td>365.00</td>
<td>2.60</td>
<td>43.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>15.56</td>
<td>0.36</td>
<td>7.92</td>
<td>0.28</td>
<td>3.40</td>
<td>363.00</td>
<td>2.60</td>
<td>37.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>15.56</td>
<td>0.36</td>
<td>7.86</td>
<td>0.21</td>
<td>3.27</td>
<td>365.00</td>
<td>2.50</td>
<td>9.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>15.42</td>
<td>0.36</td>
<td>7.84</td>
<td>0.25</td>
<td>3.53</td>
<td>365.00</td>
<td>2.50</td>
<td>20.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
The dataset contains one category label, prominence hazard ($X_{10}$), and nine parameter indicators, namely coal damage type ($X_1$), initial rate of gas dispersion ($X_2$), coal robustness factor ($X_3$), gas content ($X_4$, m$^3$/t), K1 gas desorption ($X_5$, mL/(g·min$^{0.5}$)), drill chip volume ($X_6$, kg/m), burial depth ($X_7$, m), coal thickness ($X_8$, m), and distance from tectonic zone ($X_9$, m), where $X_8$ and $X_9$ data are missing and need to be filled in so as not to affect the prediction.

MCMC is a chain-multiple-filling algorithm. MCMC can effectively analyze the variation of all parameters in a mathematical model and is suitable for more complex mathematical models [25]. The data collected belonged to a multivariate arbitrary missing pattern, a machine missing mechanism, and the number of missing items was less than 30%, for which the accuracy of using MCMC was higher than other methods such as the mean fill method. For better validation, the datasets were filled one by one using the mean fill method (Mean), random forest filling method (random forest, RF), and K-nearest neighbor filling method (K-nearest neighbor, KNN), and the results were compared. The mean and standard deviation corresponding to the different filling methods are shown in Figure 1.
3. Entropy-Weighted Gray Correlation Analysis

Steps in entropy-weighted gray correlation analysis.

(1) There are $n$ evaluation objects and $m$ evaluation indicators, and the evaluation matrix is $R'$.

$$R' = (X'_1, X'_2, \ldots, X'_n) = \begin{bmatrix} x'_1(1) & x'_2(1) & \cdots & x'_n(1) \\ x'_1(2) & x'_2(2) & \cdots & x'_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ x'_1(m) & x'_2(m) & \cdots & x'_n(m) \end{bmatrix}$$

(2) Determining the reference series.

$$R'_0 = (x'_0(1), x'_0(2), \ldots, x'_0(m))$$

(3) The dimensionless processing of the filled data gives the following matrix.

$$R = (X_1, X_2, \ldots, X_n) = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ x_1(2) & x_2(2) & \cdots & x_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(m) & x_2(m) & \cdots & x_n(m) \end{bmatrix}$$

(4) Calculate the absolute difference between the elements corresponding to the comparison sequence and the reference sequence. $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$.

$$|x_0(j) - x_i(j)|$$

(5) Determining the maximum and minimum difference.

$$\max_{i=1}^{n} \max_{j=1}^{m} |x_0(j) - x_i(j)|$$

$$\min_{i=1}^{n} \min_{j=1}^{m} |x_0(j) - x_i(j)|$$

(6) To determine the degree of association between the comparison series and the reference series, the gray correlation coefficient is calculated. $\rho$ is the resolution factor and takes values in the range $[0, 1]$. 

Figure 1. The mean and standard deviation corresponding to different filling methods.
\[
\varepsilon_i(j) = \frac{\min_{i=1}^{m} \min_{j=1}^{n} |x_0(j) - x_i(j)| + \rho \max_{i=1}^{m} \max_{j=1}^{n} |x_0(j) - x_i(j)|}{|x_0(j) - x_i(j)| + \rho \max_{i=1}^{m} \max_{j=1}^{n} |x_0(j) - x_i(j)|}
\]

(7) \(P_{ij}\) is the weight of the characteristics of the \(j\)th evaluation object under the \(i\)th evaluation indicator, called the contribution degree.

\[
P_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}
\]

(8) \(e_j\) is the total contribution of the \(j\)th indicator, called the entropy value.

\[
e_j = -\frac{1}{\ln n} \sum_{i=1}^{n} P_{ij} \cdot \ln P_{ij}
\]

(9) \(\omega_j\) is the entropy weight of the \(j\)th indicator; calculate the entropy weight.

\[
\omega_j = \frac{1 - e_j}{\sum_{j=1}^{m} 1 - \epsilon_j}
\]

(10) Calculating the entropy-weighted gray correlation. The entropy-weighted gray correlation was calculated, and the results are shown in Table 2.

\[
r_i = \frac{1}{m} \sum_{j=1}^{m} \omega_j \cdot \varepsilon_i(j)
\]

(11)

Table 2. Entropy-weighted gray correlation results.

<table>
<thead>
<tr>
<th>Indicators</th>
<th>(X_1)</th>
<th>(X_2)</th>
<th>(X_3)</th>
<th>(X_4)</th>
<th>(X_5)</th>
<th>(X_6)</th>
<th>(X_7)</th>
<th>(X_8)</th>
<th>(X_9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r_i)</td>
<td>0.010</td>
<td>0.016</td>
<td>0.009</td>
<td>0.017</td>
<td>0.015</td>
<td>0.013</td>
<td>0.011</td>
<td>0.011</td>
<td>0.012</td>
</tr>
<tr>
<td>sort</td>
<td>8</td>
<td>2</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

The higher the \(r_i\) value, the greater the correlation between this parameter indicator and the risk of protrusion, and the greater the role it plays in the coal and gas protrusion process. It should therefore be given priority when conducting prominence prediction experiments.

4. Building MCMC-ISSA-SVM Prediction Model

4.1. Sparrow Search Algorithm

The sparrow search algorithm (SSA) is a new swarm intelligence algorithm [26,27]. It has been demonstrated that SSA requires fewer parameters, is stable, and has excellent global search capability and fast convergence. It has a wide range of application prospects due to its advantages [28–32].

The matrix of the set of sparrows with population size \(n\) and spatial dimension \(d\).

\[
X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,d}), i \in [1, n]
\]

(12)

Corresponding matrix of fitness values.

\[
F_x = [f(x_1), f(x_2) \cdots f(x_n)]^T
\]

(13)
\( f(x_i) \) indicates the individual fitness value. The primary goal of the discoverer is to explore new spaces randomly in search of food, and the discoverer’s location is updated by the following formula.

\[
X_{i,j}^{r+1} = \begin{cases} 
X_{i,j}^r & X_{i,j}^r \cdot \exp \left( \frac{-d}{r_{\text{max}}} \right), R_2 < ST \\
X_{i,j}^r + Q \cdot L, R_2 \geq ST 
\end{cases}
\]  

\( r \) represents the current number of iterations, \( j \in [1, d] \), \( r_{\text{max}} \) represents the maximum number of iterations, \( X_{i,j} \) represents the position of the \( i \)th sparrow in the \( j \)th dimension, \( \alpha \) is a random number taking values in the range \((0, 1]\). \( R_2 \) is the safety value and takes values in the range \([0, 1\]. ST stands for safety value and takes values in the range \([0.5, 1\]. \( Q \) is a random number that follows a standard normal distribution.

4.2. Improved Sparrow Search Algorithm

To improve the algorithm’s search optimization capability and convergence speed [33], when the population is foraging, the lower-adapted joiners are not grabbing food and need to fly close to the best locations to get food, usually with 10% to 20% of the sparrows acting as vigilantes. The position of the vigilantes is updated according to the following formula.

\[
X_{\text{best},j}^{r+1} = X_{\text{best},j}^r + \beta \cdot \left| X_{i,j}^r - X_{\text{best},j}^r \right|, f_i > f_b \]

\[
X_{i,j}^{r+1} = X_{i,j}^r + K \left( \frac{X_{i,j}^r - X_{\text{worst},j}^r}{f_i - f_{\text{worst}}} \right), f_i = f_b
\]  

\( X_{\text{best},j}^{r} \) represents the current best position. \( \beta \) represents a random number that follows a standard normal distribution. \( f_{\text{worst}} \) represents the worst adaptation of the population. \( K \) is a random number of \([-1, 1]\). \( \epsilon \) is a very small constant value.

The joiner will follow the finder and keep fighting for food with him. The location of the joiner is updated as follows:

\[
X_{i,j}^{r+1} = \begin{cases} 
X_{i,j}^r & X_{i,j}^r \cdot \exp \left( \frac{X_{\text{best},j}^r - X_{i,j}^r}{r^2} \right), i < n/2 \\
X_{i,j}^r + Q \cdot L, i \geq n/2 
\end{cases}
\]  

\( X_{\text{worst},j}^{r} \) represents currently in the worst position. \( X_{p}^{r+1} \) represents the current best position of the discoverer. \( A \) is a \( 1 \times d \) row matrix that takes the value 1 or -1.

\[
A^+ = A^T (AA^T)^{-1}
\]  

When the population is foraging, the lower-adapted joiners are not grabbing food and need to fly close to the best locations to get food, usually with 10% to 20% of the sparrows acting as vigilantes. The position of the vigilantes is updated according to the following formula.

\[
X_{i,j}^{r+1} = \begin{cases} 
X_{\text{best},j}^{r} + \beta \cdot \left| X_{i,j}^r - X_{\text{best},j}^r \right|, f_i > f_b \\
X_{i,j}^r + K \left( \frac{X_{i,j}^r - X_{\text{worst},j}^r}{f_i - f_{\text{worst}}} \right), f_i = f_b
\end{cases}
\]  

The probability density function for the distribution of \( t \) is given in Equation (18).

\[
f(x) = \frac{\Gamma \left( \frac{n+1}{2} \right)}{\sqrt{\pi n} \Gamma \left( \frac{n}{2} \right)} \left( 1 + \frac{x^2}{n} \right)^{-\frac{n+1}{2}}, x \in (-\infty, +\infty)
\]

\( \Gamma (n + 1/2) \) denotes the second type of Euler integral, and \( n \) denotes the t-distribution degree of freedom. The sparrow positions were updated according to Equation (19) after adding the t-distribution variance operator.

\[
X_{i,j}^{r+1} = X_{i,j}^r + \lambda t(r) X_{i,j}^r
\]


\[ r \text{ denotes the number of iterations. } r_{\text{max}} \text{ denotes the maximum number of iterations.} \]

\[ X_{i,j}^{t} \] represents the current position of the \( i-th \) sparrow. \( X_{i,j}^{t+1} \) represents the position of the \( i-th \) sparrow after \( t \)-distribution variation. \( \lambda \) is the variance control factor and takes values in the range \([0, 1]\).

\[ \lambda = \frac{1 - r}{r_{\text{max}} - 1} \quad (20) \]

\( \lambda(t) \) is the variation operator that enhances the diversity of the sparrow population and improves the algorithm’s local optimization-seeking ability.

The specific variation strategy is as follows: first generate a random number between \([0, 1]\), called \( \text{rand} \), then define the variation probability formula. 20 is the number of sparrow populations.

\[ p = (1 - \frac{r}{r_{\text{max}}})^{20} + 0.05 \quad (21) \]

Finally, the judgment condition is determined; when \( p > \text{rand} \), the variation operation is performed on the updated optimal sparrow position; otherwise, no variation is performed.

To improve the efficiency of the algorithm, the formula for updating the position of the vigilantes is improved.

\[ X_{i,j}^{t+1} = \begin{cases} X_{\text{best},j} + \beta \cdot (X_{i,j}^{t} - X_{\text{best},j}), & f_i > f_b \\ X_{i,j}^{t} + \beta \cdot (X_{\text{wold},j}^{t} - X_{\text{best},j}), & f_i = f_b \end{cases} \quad (22) \]

The specific steps for building the MCMC-ISSA-SVM prediction model are as follows: The main flow of building the MCMC-ISSA-SVM model is shown in Figure 2.

**Figure 2.** Flow chart for building MCMC-ISSA-SVM prediction model.
Step 1: Import the collected raw coal and gas prominence data and fill in the missing data using the MCMC algorithm to finally obtain a complete dataset with good results.

Step 2: Determine the model inputs and outputs based on the filled data set. Identify each parameter indicator as a salient predictive feature and salient hazards as category labels. The predictive features and category labels are used as input values to the model. Determine the predicted salient hazard category and prediction accuracy as the output values of the model. Consistency and standardization of the data using the corresponding methods based on the indicator characteristics.

Step 3: Select a training set and a test set for the prominent prediction experiment. A portion of the data is selected to form the training set to train the model, and the remaining portion forms the test set to test the predictive effectiveness of the model.

Step 4: Initialize the SVM penalty parameter $c$, the kernel function parameter $g$, and the relevant parameters of the MCMC-ISSA-SVM model. Set the sparrow population size to 20, the maximum number of iterations to 100, the proportion of discoverers to 70%, the proportion of followers to 30%, the proportion of vigilantes to 20%, and the safety value $ST$ to 0.6.

Step 5: Classification of training samples using cross-validation. The number of cross-validation folds is set to 5, and the cross-validation recognition accuracy of the SVM coal and gas prominence type is used as the fitness value of the individual sparrow. The optimal fitness value and location information are retained.

Step 6: Rank the fitness of all individuals in the sparrow population. Individuals with higher fitness levels are considered discoverers, and the remaining sparrows are considered followers.

Step 7: Update finder locations according to Equation (14). Sparrows can search extensively when they are in a safe state. If the warning value is greater than the safety value, the sparrow population will fly to a safe space to forage.

Step 8: Update the follower positions according to Equation (15). According to the ranking principle, when $i > n/2$, the individual fitness value is low, and these followers need to search for other locations to improve their individual fitness.

Step 9: Update the position of vigilantes aware of danger according to Equation (22). Sparrows at the periphery of the population will move closer to the safe area. Sparrows at the center of the population will walk randomly to get closer to other sparrows. The fitness value of this iteration is compared with the current best fitness value, and the global optimal sparrow position is updated.

Step 10: Compare the magnitude of the variation probability $p$ and the random number $rand$. When the value of the variation probability $p$ is greater than the random number $rand$, a t-distribution variation operation is performed on the updated optimal sparrow position. Conversely, the variation condition is not satisfied.

Step 11: If mutation occurs, calculate the adaptation values for the sparrow population before and after mutation and compare the two. If no mutation occurs, the fitness value of the new position of the individual sparrow is calculated, and the updated sparrow fitness value is compared with the original optimal sparrow fitness value. Eventually, the sparrow with the better fitness value is selected as the current sparrow location based on whether or not there is variation, and the global optimum information is updated.

Step 12: Determine whether the number of iterations satisfies the termination condition. If the condition is not satisfied, the number of iterations is increased by 1 and then returned to step 6. If the termination condition is satisfied, the iteration is stopped, and the optimal combination of parameters $c$ and $g$ is directly output. The MCMC-ISSA-SVM model is built, and the test set samples are input to the model. The final output is the result of the coal and gas prominence prediction.
5. Coal and Gas Prominence Prediction

5.1. Conduct Group Experiments to Select the Best Parameter Indicators

The entropy-weighted gray correlation analysis method was used on the coal and gas prominence dataset to screen parameter indicators for the coal and gas prominence prediction experiment. Different influencing factors were selected as parameter indicators for the coal and gas prominence prediction experiment according to the ranking of the entropy-weighted gray correlation. The greater the entropy-weighted gray correlation, the fewer parameter indicators are included. The experiment was divided into four groups, and the groupings are shown in Table 3.

Table 3. Experimental groupings.

<table>
<thead>
<tr>
<th>Name of Experimental Group</th>
<th>Included Parameter Indicators</th>
<th>Entropy-Weighted Gray Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_1 \sim X_9$</td>
<td>$r_i &gt; 0.008$</td>
</tr>
<tr>
<td>2</td>
<td>$X_2, X_4, X_5, X_6, X_7, X_8, X_9$</td>
<td>$r_i &gt; 0.010$</td>
</tr>
<tr>
<td>3</td>
<td>$X_2, X_4, X_5, X_6, X_9$</td>
<td>$r_i &gt; 0.012$</td>
</tr>
<tr>
<td>4</td>
<td>$X_2, X_4, X_5, X_6$</td>
<td>$r_i &gt; 0.013$</td>
</tr>
</tbody>
</table>

Different experimental results can be obtained by inputting different parameter index data into the MCMC-SVM prediction model. Coal and gas prominence predictions were made for each group of experimental data separately, and a comparison of the predicted and actual values of coal and gas prominence is shown in Figure 3.

Figure 3. Test set prediction results for each set of experiments.

There were six, four, two, and one outlier sample point in the results of the four sets of experiments. Group 4 experiments had the highest number of correct predictions and the highest prediction accuracy of the test set. In order to observe the experimental results more clearly, the relative errors of the four groups of experiments were compared, and the results are shown in Figure 4.
Different experimental results can be obtained by inputting different parameter index data into the MCMC-SVM prediction model. Coal and gas prominence predictions were made for each group of experimental data separately, and a comparison of the predicted and actual values of coal and gas prominence is shown in Figure 3.

![Figure 3](image.png)

There were six, four, two, and one outlier sample point in the results of the four sets of experiments. Group 4 experiments had the highest number of correct predictions and the highest prediction accuracy of the test set. In order to observe the experimental results more clearly, the relative errors of the four groups of experiments were compared, and the results are shown in Figure 4.

![Figure 4](image.png)

The relative error (RE) is used to reflect the degree of confidence in the predicted value. $y_i$ denotes the true value, and $y_p$ denotes the predicted value.

$$RE = \frac{|y_i - y_p|}{y_i}$$  \hspace{1cm} (23)

The mean relative error (MRE) represents the average of the relative errors.

$$MRE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - y_p|}{y_i}$$  \hspace{1cm} (24)

The root mean square error (RMSE) is used to measure the error between the predicted and actual values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_p)^2}$$  \hspace{1cm} (25)

The coefficient of determination ($R^2$) is a measure of how well the prediction model fits. $y_m$ represents the average of the true values.

$$R^2 = 1 - \frac{(y_i - y_p)^2}{(y_i - y_m)^2}$$  \hspace{1cm} (26)

$RE_{\text{max}}$ denotes the maximum relative error. Using the above four indicators to evaluate the experimental results, the smaller the value of the first three indicators, the better the prediction result, and the opposite for $R^2$, the larger the value, the better the experimental result. The results of evaluation indicators for each experimental group are shown in Table 4.

<table>
<thead>
<tr>
<th>Experimental Group</th>
<th>$RE_{\text{max}}$</th>
<th>MRE</th>
<th>RMSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.500</td>
<td>0.111</td>
<td>0.226</td>
<td>0.981</td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
<td>0.083</td>
<td>0.323</td>
<td>0.945</td>
</tr>
<tr>
<td>3</td>
<td>0.500</td>
<td>0.063</td>
<td>0.177</td>
<td>0.989</td>
</tr>
<tr>
<td>4</td>
<td>0.500</td>
<td>0.042</td>
<td>0.144</td>
<td>0.993</td>
</tr>
</tbody>
</table>

### 5.2. Prediction of Coal and Gas Prominence by MCMC-ISSA-SVM Model

The prediction models of improved sparrow search algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm (MCMC-ISSA-SVM),
sparrow search algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm (MCMC-SSA-SVM), genetic algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm (MCMC-GA-SVM) and particle swarm optimization algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm (MCMC-PSO-SVM) were established using the ISSA, SSA, genetic algorithm (genetic algorithm, GA) and particle swarm optimization algorithm (particle swarm optimization, PSO) respectively. The best results were obtained from Group 4 experiments, and the data corresponding to the four main parameter indicators of gas dispersion initial velocity (X2), gas content (X4), K1 gas desorption (X5), and drill cuttings (X6) included in Group 4 experiments were input into each model for coal and gas protrusion prediction. As a result, the training and test sets were randomly divided, and the results had a certain randomness. In order to be able to further analyze the prediction effect more accurately, each model was subjected to repeated prediction experiments 50 times; the prediction results were recorded each time, and the experimental results were tallied, as shown in Figure 5.

To obtain a more intuitive picture of the prediction effectiveness of the models, the average of the prediction results of 50 experiments for each model was calculated as its prediction accuracy. The experimental error of each model was then calculated based on the prediction accuracy. The results for each model are shown in Figure 6.

The best-fit and mean-fit convergence curves allow a better comparison and analysis of the dynamic process of finding the optimum and the speed of convergence of each model.
iteration. The best fit and average fit convergence curves for each model optimization process are shown in Figure 7.

*Figure 7.* Fitness curves for each model.

The best-fit curve of the MCMC-ISSA-SVM model reached a prediction accuracy of 98.25% after 5 iterations and remained stable. The average fit was 66% in the first generation, rising rapidly to 90% after 4 iterations, and then fluctuating up and down in the range of 85–95%. The best fit curve of the MCMC-SSA-SVM model reached a prediction accuracy of 92.41% after 10 iterations and remained stable. The average fit increased rapidly from 61% accuracy to 84% after 6 iterations and has been fluctuating up and down in the interval of 80–90% since then. The best-fit curve of the MCMC-GA-SVM model took 22 iterations before it reached a steady state. Its prediction accuracy was 87.25%, and the average fitness was overwhelmingly distributed, mainly in the interval of 75–85%. The best-fit curve of the MCMC-PSO-SVM model took 42 iterations to converge to a steady state. Its prediction accuracy was 82.66%, and the average fitness was overwhelmingly distributed, mainly in the interval range of 70–80%.
As can be seen in Figures 6 and 7, the MCMC-SSA-SVM model has a prediction accuracy of 92.41% with an error of 0.078, while the MCMC-ISSA-SVM has a prediction accuracy of 98.25% with an error of 0.018. Improvements to the algorithm have resulted in a 5.84% increase in prediction accuracy, a 0.060 decrease in error, a reduction in the number of iterations by 5, and a reduction in the number of iterations. The best fit and average fit increased by about 5%. This shows that the improvement of SSA by the adaptive t-distribution variation operator is effective. The prediction accuracy of MCMC-GA-SVM was 87.25% with an error of 0.128, and that of MCMC-PSO-SVM was 82.66% with an error of 0.183. The comparison revealed that MCMC-ISSA-SVM had the highest prediction accuracy and the lowest error. In addition, it has the fastest convergence speed, the lowest number of iterations, the highest best fit, and the average fit among the four models, and the prediction results are significantly better than the other three models. It indicates that ISSA has better optimization capability than SSA, PSO, and GA, and MCMC-ISSA-SVM can improve prediction performance.

6. Results

(1) The MCMC filling algorithm was used to fill the missing values and compared with the mean, RF, and KNN filling results. The mean value of MCMC-filled data is 2.28 and the standard deviation is 0.19, which is closer to the mean and standard deviation of the original dataset than the mean and standard deviation of other models, indicating that MCMC-filled data is the best. The results indicate that the MCMC algorithm performs well in filling missing values and can optimize the dataset effectively.

(2) The parameter indicators of the prominence data were ranked by entropy-weighted gray correlation analysis and divided into four experimental groups with different numbers of parameter indicators according to the entropy-weighted gray correlation, which were input into the MCMC-SVM model for prominence prediction experiments. The last group of experimental coal and gas prominence predictions has the best effect. The results show that the entropy-weighted gray correlation analysis can effectively filter the prediction parameters, reduce the interference of other parameters, and improve the prediction accuracy.

(3) An improved sparrow search algorithm is proposed to enhance sparrow population diversity by adding an adaptive t-distribution variation operator and improving the update position of vigilantes. The MCMC-ISSA-SVM model is established and compared with the MCMC-SSA-SVM model for experiments, and it is found that the improved prediction accuracy is improved by 5.84%, the error is reduced by 0.060, the number of iterations is reduced by 5, and the best and average fitness are increased by about 5%. The results show that the improvement of the algorithm is effective.

(4) The prediction accuracy of MCMC-ISSA-SVM is 98.25% with an error of 0.018, which is significantly better than the other three models. In addition, its convergence speed is the fastest, the number of iterations is the least, and the best fitness and average fitness are the highest among the four models. The results verified that ISSA outperformed SSA, PSO, and GA in terms of optimization seeking, and the proposed MCMC-ISSA-SVM model could indeed significantly improve the prediction accuracy and effectively enhance the generalization ability.

Other methods mainly choose one of both an optimized dataset and an improved algorithm for modeling, and this paper mainly deals with both an optimized dataset and an improved algorithm before building a model to be applied to coal and gas prominence prediction. There are some limitations to this method. The changes in ground stress, geological structure conditions, media structure, and gas at the construction location will affect the coal and gas prominence prediction, and the collected data set is suitable for this method when there is a deficiency.
Author Contributions: A gas prominence prediction model based on entropy-weighted gray correlation and MCMC-SSA-SVM: Use entropy-weighted gray correlation and machine learning algorithms; software, SPSS and Matlab; data curation, Y.G.; writing—original draft preparation, Y.G.; supervision, L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation Project (71771111). Project name: Research on the prediction method and application of coal and gas outbursts based on big data. Discipline classification: G0104. Prediction and evaluation. Project leader: L.S. Funding amount: 460,000 yuan.

Data Availability Statement: The data are shown in Table 1.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCMC</td>
<td>Markov chain Monte Carlo</td>
</tr>
<tr>
<td>SSA</td>
<td>sparrow search algorithm</td>
</tr>
<tr>
<td>ISSA</td>
<td>improved sparrow search algorithm</td>
</tr>
<tr>
<td>SVM</td>
<td>support vector machine</td>
</tr>
<tr>
<td>GA</td>
<td>genetic algorithm</td>
</tr>
<tr>
<td>PSO</td>
<td>particle swarm optimization</td>
</tr>
<tr>
<td>MCMC-ISSA-SVM</td>
<td>improved sparrow search algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm</td>
</tr>
<tr>
<td>MCMC-SSA-SVM</td>
<td>sparrow search algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm</td>
</tr>
<tr>
<td>MCMC-GA-SVM</td>
<td>genetic algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm</td>
</tr>
<tr>
<td>MCMC-PSO-SVM</td>
<td>particle swarm optimization algorithm optimized support vector machine based on Markov chain Monte Carlo filling algorithm</td>
</tr>
<tr>
<td>RF</td>
<td>random forest</td>
</tr>
<tr>
<td>KNN</td>
<td>K-nearest neighbor</td>
</tr>
<tr>
<td>RE</td>
<td>relative error</td>
</tr>
<tr>
<td>RE\text{max}</td>
<td>maximum relative error</td>
</tr>
<tr>
<td>MRE</td>
<td>average relative error</td>
</tr>
<tr>
<td>RMSE</td>
<td>root mean square error</td>
</tr>
<tr>
<td>R^2</td>
<td>coefficient of determination</td>
</tr>
<tr>
<td>GASA</td>
<td>genetic simulated annealing algorithm</td>
</tr>
<tr>
<td>BP</td>
<td>back propagation</td>
</tr>
<tr>
<td>\rho</td>
<td>resolution factor</td>
</tr>
<tr>
<td>p_{ij}</td>
<td>the weight of the characteristics of the jth evaluation object under the ith evaluation indicator</td>
</tr>
<tr>
<td>e_j</td>
<td>entropy value of the jth indicator</td>
</tr>
<tr>
<td>\omega_j</td>
<td>the entropy weight of the jth indicator</td>
</tr>
<tr>
<td>r_i</td>
<td>entropy-weighted gray correlation</td>
</tr>
<tr>
<td>ST</td>
<td>safety value</td>
</tr>
<tr>
<td>R_2</td>
<td>warning value</td>
</tr>
<tr>
<td>c</td>
<td>penalty parameters for support vector machines</td>
</tr>
<tr>
<td>g</td>
<td>kernel function parameter for support vector machine</td>
</tr>
<tr>
<td>\lambda(r)</td>
<td>variation operator</td>
</tr>
<tr>
<td>P</td>
<td>variation probability</td>
</tr>
<tr>
<td>rand</td>
<td>random number</td>
</tr>
<tr>
<td>r</td>
<td>the number of iterations</td>
</tr>
<tr>
<td>r_{\text{max}}</td>
<td>the maximum number of iterations</td>
</tr>
</tbody>
</table>
References


20. Wang, Y.H.; Sun, F.C.; Fu, H.; Xu, Y. Coal and gas prominence prediction based on optimized quantum gate node neural network. *Inf. Control. 2020*, *49*, 249–256. [CrossRef]


32. Chen, G.J.; Chen, G.F. An Improved Sparrow Algorithm Based on Small Habitats in Cooperative Communication Power Allocation. *Electronics* 2023, 12, 1153. [CrossRef]

33. Li, N.; Xue, J.K.; Shu, H.S. UAV trajectory planning based on adaptive t-distribution variational sparrow search algorithm. *J. Donghua Univ. (Nat. Sci. Ed.)* 2022, 48, 69–74. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.