

## Research Article

# Anomaly Detection for Time Series with Difference Rate Sample Entropy and Generative Adversarial Networks

Keke Gao <sup>1,2</sup>, Wenbin Feng <sup>1</sup>, Xia Zhao <sup>2</sup>, Chongchong Yu <sup>2</sup>, Weijun Su <sup>2</sup>,  
Yuqing Niu <sup>2</sup> and Lu Han <sup>2</sup>

<sup>1</sup>State Key Laboratory of Coal Mine Safety Technology,

China Coal Technology & Engineering Group Shenyang Research Institute, Fushun 113122, China

<sup>2</sup>School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China

Correspondence should be addressed to Wenbin Feng; 625309069@qq.com

Received 19 June 2021; Revised 27 September 2021; Accepted 15 October 2021; Published 19 November 2021

Academic Editor: Francesco Lo Iudice

Copyright © 2021 Keke Gao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The spontaneous combustion of residual coals in the mined-out area tends to cause an explosion, which is one kind of severe thermodynamic compound disaster of coal mines and leads to serious losses to people's lives and production safety. The prediction and early warning of coal mine thermodynamic disasters are mainly determined by the changes of the index gas concentration pattern in coal mine mined-out areas collected continuously. The time series anomaly pattern detection method is mainly used to reach the state change of gas concentration pattern. The change of gas concentration follows a certain rule as time changes. A great change in the gas concentration indicates the possibility of coal spontaneous combustion and other disasters. To emphasize the features of collected maker gas and overcome the low anomaly detection accuracy caused by the inadequate learning of the normal mode, this paper adopted a method of anomaly detection for time series with difference rate sample entropy and generative adversarial networks. Because the difference rate entropy feature of abnormal data was much larger than that of normal mode, this paper improved the calculation method of the abnormal score by giving different weights to the detection points to enhance the detection rate. To verify the effectiveness of the proposed method, this paper employed simulation models of the mined-out area and adopted coal samples from Dafosi Coal Mine to carry out experiments. Preliminary testing was performed using monitoring data from a coal mine. The experiment compared the entropy results of different time series with the detection results of generative adversarial networks and automatic encoders and showed that the method proposed in this paper had relatively high detection accuracy.

## 1. Introduction

The thermodynamic compound disaster of the coal mine refers to the compound disaster of the explosion caused by the spontaneous combustion of residual coal in a mined-out area. As the main energy, coal plays an irreplaceable role in China's energy structure. The safe supply of coal is directly related to the sustainable development of the national economy and the energy security of the country. After coal mining, the remaining area is the mined-out area, which has poor ventilation and a lot of residual coal. The residual coal is oxidized continuously under the influence of air leakage, which is likely to cause gas accumulation [1–3]. As a result, it

is easy to cause coal spontaneous combustion and other coal thermodynamic disasters. The thermal power disaster in coal mines seriously affects the safety of coal mine production and workers. Therefore, it is necessary to know the factors related to the occurrence of thermal power complex disasters in coal mines and then analyze them through time series mining technology to forecast the possible thermal power disasters.

Coal mine thermal disaster prediction and early warning are very important tasks in the coalfield. The early disaster prediction methods mainly include oxygen measurement, temperature measurement, and index gas analysis [4]. The oxygen measurement method is mainly used to measure the

oxygen content in the coal mine and determine whether it reaches the threshold. The temperature measurement is to monitor the temperature of some areas and then use the data of measured temperature to infer the possible coal spontaneous combustion disasters in this area. Index gas analysis is to observe whether single or multiple quantitative indexes related to coal spontaneous combustion or gas explosion in the mined-out areas exceed the critical value, mainly including gas index, coal seam property index, or comprehensive index [5]. A large number of experimental studies showed that gases like CO will be produced in the process of coal spontaneous combustion. The amount of gas generation can change significantly with the increase of coal temperature. According to the experimental results, the appropriate gases are selected as the index gas, and the stage of the coal oxidation process can be roughly judged by analyzing the change of the formation status of the index gas [6]. With the development of gas collection and analysis technology, index gas analysis has been widely used in the early prediction of spontaneous combustion [7]. In practice, many predictions based on the gas indexes have been proposed. There are relatively simple single gas indexes, such as CO and C<sub>2</sub>H<sub>4</sub>. There are also relatively complex forms of double or compound gas indexes, such as CO/CO<sub>2</sub> and hydrocarbon ratio. The prediction index and the critical value of the index are based on the field experience and the statistical analysis of a large number of experimental data. Due to the influence of manual operation factors, uneven distribution of coal and stress and the prediction show great limitations, and the accuracy of prediction is difficult to improve [8].

With the development of information technology, more and more time series data are produced and become more complex. Anomaly detection for time series [9] has become a research hotspot in recent years. Time series anomaly usually refers to the data which is obviously different from other data in a series of data sets. This anomaly is not caused by random deviation but by differences due to different patterns. The commonly used time series anomaly detection research methods mainly contain statistical methods and machine learning methods. Canizo et al. [10] proposed a supervised multi-time series anomaly detection method based on deep learning, which combined Convolution Neural Network (CNN) and Recurrent Neural Network (RNN) in different ways, and processed each sensor separately to avoid the need for data preprocessing and greatly improve the operation speed. Since the model uses a dataset containing a fixed-length time series to verify the proposed architecture, in some real use cases, they are not. Therefore, further research must be done to analyze the performance of the proposed architecture at the time of processing time series with different frequencies. Beggel et al. [11] proposed a new unsupervised anomaly detection method based on wavelet transform of time series, which could effectively detect the abnormal time series in the test data without retraining the model by repeatedly learning the feature representation. However, further research is needed for anomaly detection of multivariate time series. Malhotra et al. [12] used a stackable LSTM network for anomaly/fault detection in time series. The network was trained on

nonabnormal data and used as a predictor at multiple time steps. The resulting prediction errors were modeled as multivariate Gaussian distributions, which were used to assess the likelihood of anomaly. Chauhan and Vig [13] adopted a deep recurrent neural network architecture with the help of Long Short Term Memory (LSTM) to develop predictive models for healthy ECG signals. The probability distribution of the prediction errors from these recurrent models was utilized to indicate normal or abnormal behaviors. But using stacked networks or deep recursive networks can reduce running speed to some extent. Rajagopalan and Ray [14] presented a wavelet-based partitioning approach for symbol generation, instead of the currently practiced method of phase-space partitioning. However, it is necessary to extend the separation to multiple time series and reduce the noise in the time series to achieve robust anomaly detection. Izakian and Pedrycz [15] considered fuzzy *c*-means (FCM) as a conceptual and algorithmic setting to deal with the problems of anomaly detection. Using a sliding window, the time series was divided into several subsequences, and the available spatiotemporal structure within each time window is discovered using the FCM method. However, with this algorithm, the number of iterations in the selection of the cluster center point is more likely to cause the algorithm model to have weak scalability and weak sensitivity and fall into a local minimum. The prediction ability of the above method is limited and the computational cost is large, which cannot be effectively detected in large data sets with large data size and dimension. The detection method based on GAN can achieve the purpose of anomaly detection without collecting a large number of abnormal data and using normal data training.

In recent years, a generative adversarial network framework [16] has been proposed to build a deep learning model through adversarial training. Li et al. [17] proposed a way of conducting multivariate anomaly detection on time series data based on the generation of adversarial network and used LSTM neural network as the basic model GAN framework to capture time-related time series distribution. An anomaly detection neural network, dual autoencoder generative adversarial network (DAGAN), was developed by Tang et al. [18] to solve the problem of sample imbalance. With skip-connection and dual autoencoder architecture, the proposed method exhibited excellent image reconstruction ability and training stability. The detection of abnormal patterns in gas timing data can provide a theoretical basis for coal spontaneous combustion or oxygenation and gas explosion.

The concentration change of combustible gas released from floating coal follows certain patterns with time going on. When the gas concentration changes greatly, it can be considered to enter the abnormal mode, indicating that coal spontaneous combustion and other disasters may occur. Therefore, effective detection of the inflection points of monitored data in different stages can assist in the judgment of different oxidation stages and the occurrence of coal spontaneous combustion. Different coal mines have different amounts of gas accumulation. If only the amount of gas accumulation is taken as the criterion for the

determination of disasters, great errors may occur when applied to other coal mines. Therefore, the detection of abnormal patterns can improve the generalization of disaster judgment and provide a new idea for the detection of coal composite disasters.

The main contributions of this paper are as follows. (1) According to the trend characteristics of CO gas data, the difference rate entropy feature is extracted to get the processed feature sequence, in which the entropy feature value of abnormal mode is higher and that of normal mode is lower, to highlight the difference between abnormal mode and normal mode. (2) The anomaly pattern of entropy feature sequence is detected by using a generative countermeasure network, and a new calculation method for the anomaly score is proposed in the detection stage, which considers both the weighted outlier score of the generated samples and the outlier score of the discrimination results, and judges whether the data segment of the one-dimensional time series is determined by calculating the anomaly score of the sample.

The rest of this article is arranged as follows. Section 2 mainly introduces the basic concepts of difference rate sample entropy, time series, and abnormal pattern analysis. In Section 3, a one-dimension time series anomaly detection algorithm based on different rate sample entropy and generative adversarial networks is proposed. In Section 4, the validity of the algorithm and its feasibility under this background illustrated by experiments are presented. The conclusion and the future suggestions are given in the last section.

## 2. Related Work

**2.1. Gas Abnormal Patterns Analysis.** When the gas concentration increases in the early stage, the abnormal information is relatively weak and difficult to be detected in time. When the gas concentration increases obviously and reaches the threshold value, the detection will lose the significance of disaster prediction. Therefore, it is important to detect the time when the anomaly occurs as early as possible.

To detect data anomalies as early as possible, the present study defines the process that the data changes greatly as the abnormal pattern whose data changing rate is different from the normal pattern. Different from the normal mode, in the abnormal mode, the change rate of data is significantly different from that before. Taking index gas as an example, the whole process of concentration change is analyzed first, as shown in Figure 1. The whole process can be roughly divided into four stages:

In the first stage, the concentration increases at a lower rate and lasts the longest

In the second stage, the gas concentration goes up rapidly and the growth rate also increases

In the third stage, the gas concentration rises at a relatively higher rate, and the growth rate remains almost unchanged

In the fourth stage, the concentration begins to decline

Meanwhile, the general trend of the data in the first three stages is increasing, and the fourth stage begins to decline. The constant increase of CO gas concentration means that the coal spontaneous combustion oxidation process enters into a different oxidation stage. When the growth rate of gas concentration keeps increasing in the second stage, it means that the gas concentration will increase rapidly. The detection of this mode can make a judgment in advance for the identification of the oxidation stage of coal spontaneous combustion. Therefore, the stage when data starts to increase rapidly is defined as an abnormal mode.

Next, the mode of the time series is explained. The one-dimensional time series of the original gas sample is given as  $S = \{S_1, S_2, \dots, S_T\}$ ,  $S_u = (x_u, t_u)$ . It contains  $T$  time points, and each time point corresponds to a concentration value. The representation of the time series pattern can be understood as the segmentation of time series in the time dimension and the feature representation of each segment. The representation of the time series pattern can be understood as segmenting the time series on the time dimension, then, the feature representation of each sequence is carried out, and then the abnormal points and abnormal sequences can be detected by relevant algorithms.

Table 1 shows the original data sample of a monitoring point, including symbolic gases and temperature values. These gas sensors generate time-dependent multivariate responses to different gases.

Through the analysis of time series data, it can be seen that only from the data size, the abnormal changes of data are not obvious, which cannot contribute to the subsequent abnormal detection work. Therefore, according to the feature of time series data, difference rate calculation is calculated first on time series data. Then, a more comprehensive feature extraction of time series is obtained through sample entropy. The data of abnormal patterns are often complex. To determine the abnormal patterns more accurately, the Generate Adversarial Networks (GANs) are adopted. The network can generate a sample similar to the real data by the generator, and the generator and discriminator can judge the input data through the abnormal score after learning the real sample to achieve the recognition of abnormal patterns.

**2.2. Difference Rate Sample Entropy.** First, a brief introduction to differential rate calculation is given. Differential rate calculation can extract the deterministic information in the series utilizing autoregression, as shown in (1). When linear trends are included in the series, the linear features can be extracted by first-order difference. When the series contains a nonlinear trend, the second or third-order difference can be used to extract the nonlinear trend.

$$\nabla^e x_t = \sum_{i=0}^e (-1)^i C_e^i x_{t-i}. \quad (1)$$

Among them,  $\nabla^e x_t$  is the eth order difference of the series, and  $(-1)^i C_e^i$  is the numerical coefficient at the time of

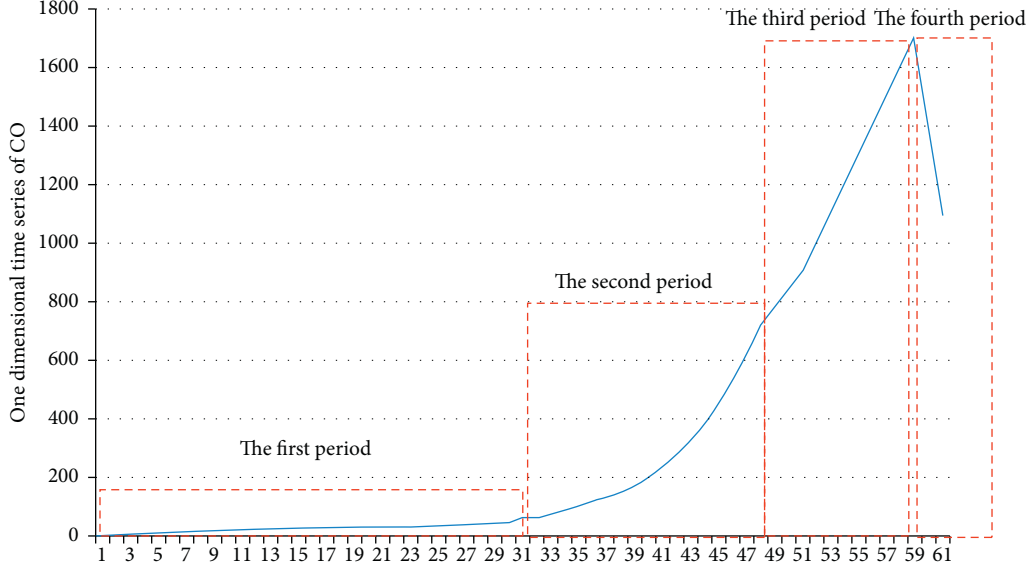


FIGURE 1: The trend of gas concentration change.

TABLE 1: Sample of original data.

Time	O <sub>2</sub>	CO	CO <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>3</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>2</sub>	T
1	9.83	0	49.08	12.85	0	0	0	23
2	9.51	0	42.16	12.69	0	0	0	36
3	9.62	0	53.17	12.41	0	0	0	49
4	9.78	1.91	62.9	12.86	0	0	0	60
5	9.5	3.96	72.16	12.3	0	0	0	72
6	9.48	31.25	482.39	18.95	0.18	0	0	152
7	9.38	73.28	629.5	21.03	0.47	0.19	0	174
8	9.04	159.7	936.26	25.38	1.93	0.84	0	202
9	8.75	401.28	1836.27	57.8	13.91	5.26	0.53	238
10	4.30	883.26	4719.39	392.17	103.85	62.71	0.79	296

$t - i$ ; For example, Figures 2(a) and 2(b), respectively, represent the characteristic diagram of CO concentration and the second-order difference characteristic diagram, where Figure 2(b) is obtained by second-order difference operation of Figure 2(a).

In view of the idea of difference, the present study proposed the concept of difference rate, which is to calculate the change of difference sequence based on the  $n$ th order difference. The definition is as follows.

Definition of (2) difference rate: for the time series  $S = \{S_1, S_2, \dots, S_T\}$ , each series point is  $S_u = (x_u, t_u)$ .  $\nabla^e S = \{S_1^e, S_2^e, \dots, S_{T-e}^e\}$  is  $e$ th order difference series. Then, the  $e+1$  difference rate series of series  $S$  is  $S^\Delta = \{\Delta S_1, \Delta S_2, \dots, \Delta S_{T-e-1}\}$ .  $\Delta S_u$  is the element in the difference rate series. The calculation is shown in

$$\Delta S_u = \frac{x_u^e - x_{u-1}^e}{x_{u-1}^e}, \quad (2)$$

where  $x_u^e$  is the  $e$ th order difference value at the time point of  $U$ , and  $x_{u-1}^e$  is the  $e$ th order difference value at the time point of  $u - 1$ .

Definition of (2) sample entropy: it is used to measure the complexity and regularity of time series. The greater the sample

entropy of the series is, the greater the complexity of the corresponding time series will be. For time series  $X = \{x_1, x_2, \dots, x_n\}$ , it is defined as shown in (3). The length of subseries  $B_{m+1}(r)$  is the length of  $m + 1$  mean subseries similarity probability and  $r$  is the similarity threshold [19, 20].

$$\text{sample entropy} = -\ln\left(\frac{B^{m+1}(r)}{B^m(r)}\right). \quad (3)$$

Take the series in Figure 2(a) as an instance; it is a second-order difference rate sample entropy feature series as shown in Figure 3. As it is shown in the figure, the sample entropy of the segmented series fluctuates up and down and tends to increase in the later stage, which corresponds to the original sequence.

2.3. *Generating Adversarial Networks*. This paper focuses on the research of anomaly detection for time series data. Due to the complexity of industrial time series data, the traditional anomaly detection methods cannot make timely predictions, and the supervised machine learning method cannot be used due to the lack of labeled data [21]. To solve this problem, this paper proposes an unsupervised anomaly detection method based on generative adversarial networks (GANs). Previous studies have proved that a generative adversarial network is very successful in image processing tasks, such as generating high-quality images, image conversion, image repair, text generation, video generation, and enhanced photos [22]. GAN has also been proved to be effective in generating time series prediction and detection according to previous studies [23]. Different from traditional classification methods, the discriminator trained by GAN detects false data from input data in an unsupervised way, which makes it an attractive unsupervised machine learning technology [24]. The network realizes the decision of input data by learning generator and discriminator in turn and playing games with each other. Figure 4 shows the GAN network model diagram.

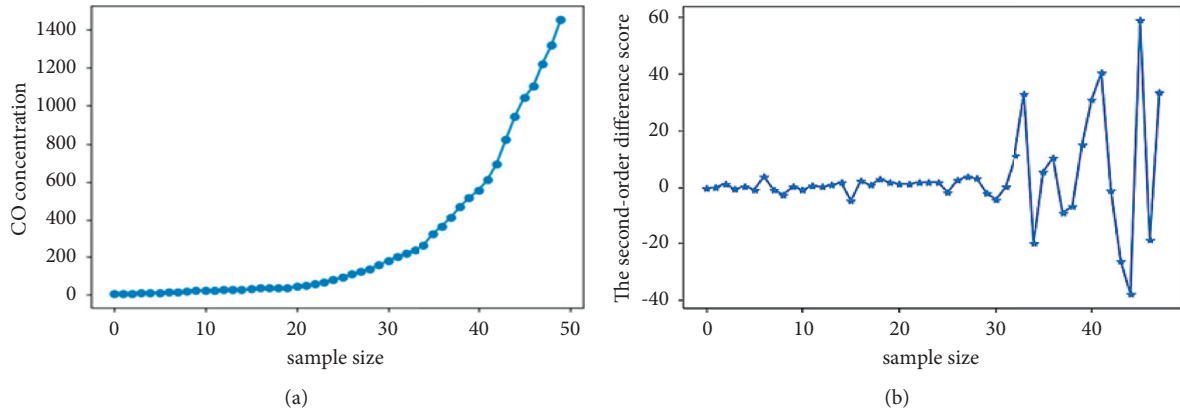


FIGURE 2: Curve series difference operation results diagram.

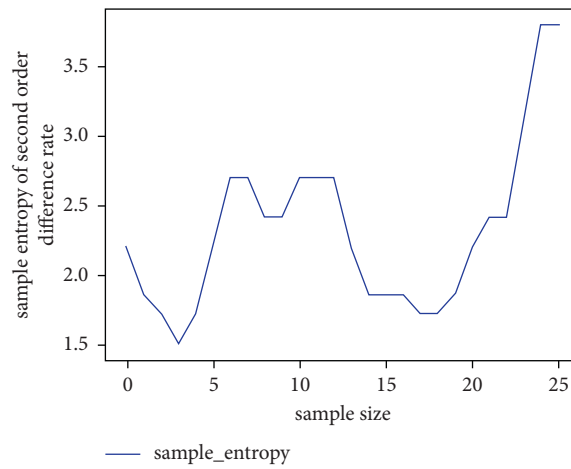


FIGURE 3: Second-order difference rate sample entropy feature series diagram.

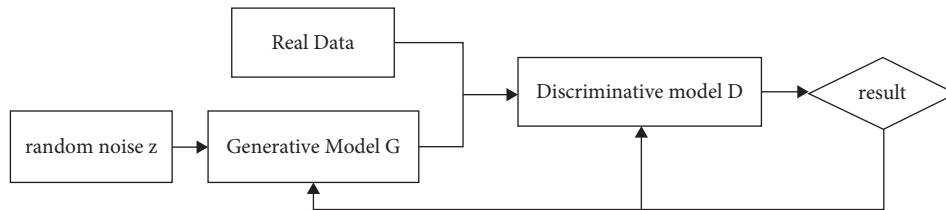


FIGURE 4: The GAN network model diagram.

In the model of Generating Adversarial Networks, the main structures are generator G and discriminator D, which are functions that can fit the corresponding generation and discrimination. In the process of network training, the generator transforms a random noise vector from a potential space to a generated sample. With the optimization of the network, the sample generated by the generator becomes more and more similar to the real samples. The discriminator accepts the generated samples from real data or generator to determine whether the input data is a real sample or a generated sample. The output of the discriminator is used to optimize the parameters of the discriminator and the generator. In this way, the generator generates more

realistic generated samples, and the discriminator can better distinguish the real samples from the generated samples.

The generative adversarial network can generate samples that are very similar to the real samples. In time series prediction, the network can be used to learn the historical data of the series and learn the pattern of the time series to generate the predicted value of the future moment.

### 3. Materials and Methods

In this section, we discuss the overall structure of our proposed model and its key elements which are presented in Figure 5. The anomaly detection framework is divided into three stages:



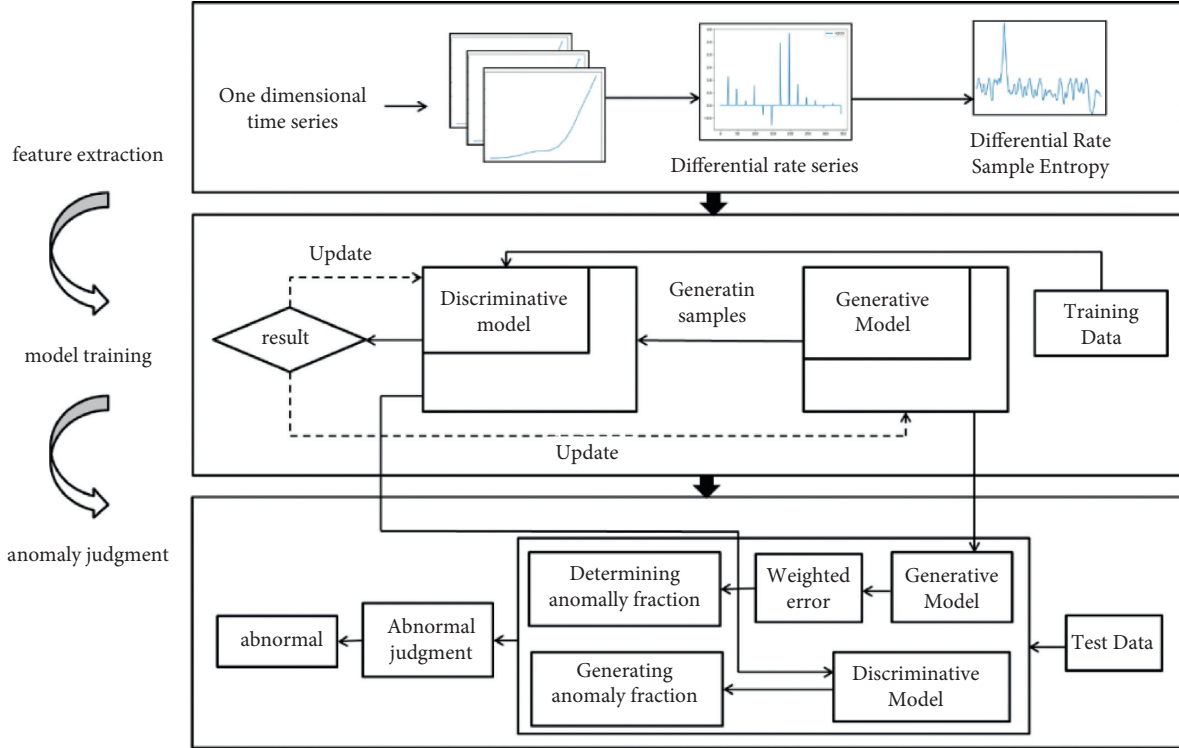


FIGURE 5: Overall flowchart of anomaly detection.

abnormal pattern feature extraction, model training, and anomaly judgment. In the stage of abnormal pattern feature extraction, the original data are extracted more thoroughly. In the model training stage, the feature series is preprocessed, and the training of the generative adversarial network model is completed. In the anomaly detection stage, the trained network model and the exponentially weighted error are used to determine the anomaly pattern.

**3.1. Entropy Feature Extraction of Difference Rate Sample.** The steps of extracting series features by using difference rate sample entropy are as follows:

Step 1: to maintain the dependence of the original time series data on the timing series, the original one-dimensional input time series data  $S = \{x_1, x_2, \dots, x_T\}$  is segmented by sliding window size  $w$  and step size  $d$ . The  $i$ th time series fragment is  $s_i = [x_{1+(i-1)d}, x_{2+(i-1)d}, \dots, x_{1+(i-1)d+w}]$ .

Step 2: carry out (2) on the series segment  $s_i$  to obtain the second-order difference rate series  $G = \{g_1, g_2, \dots, g_{w'}\}$  and its standard deviation std.

Step 3: take  $m$  time series data points as a subsegment; the second-order differential rate sequence with  $w'$  data points is divided into  $w' - m + 1$  subsegments, denoted as  $K2_i = \{q_1, q_2, \dots, q_{w'-m+1}\}$ .

Step 4: calculate the distance  $D[q_a, q_b]$  between any two subseries segments  $q_a$  and  $q_b$ , which is determined by

the maximum difference value of the corresponding position element in the two subseries segments.

Step 5: calculate the probability of similarity between the subseries  $q_a$  and other subseries, as shown in (4), which means the proportion of subseries whose distance between subseries is less than the threshold.  $r$  is the threshold of similarity,  $D[q_a, q_b]$  is the distance between any two subsequence fragments  $q_a$  and  $q_b$ ,  $m$  is  $m$  time series data, and  $w'$  is the second-order difference rate sequence with  $w'$  data points. The average probability of similarity of the second-order difference rate series is shown in (5).

$$B_a^m(r) = \frac{\text{num}[D(q_a, q_b) < r]}{w' - m}, \quad (4)$$

$$B^m(r) = \frac{\sum_{a=1}^{w'-m+1} B_a^m(r)}{w' - m + 1}. \quad (5)$$

Step 6: according to Steps 4~6, the mean similarity probability  $B^{m+1}(r)$  is recalculated with  $m + 1$  as the length of the subseries. Then, the entropy features of the second-order difference rate sample are shown in

$$\text{SE} = -\ln\left(\frac{B^{m+1}(r)}{B^m(r)}\right). \quad (6)$$

Step 7: return to Step 1 to calculate the entropy features of the second-order difference rate sample of the next

series segment  $s_{i+1}$ . Finally, a complete second-order difference rate sample entropy series is obtained.

**3.2. DRSe-GAN Anomaly Detection Model.** The anomaly detection framework in DRSe-GAN mainly includes input data, generator, discriminator, and anomaly detection. The input data is the feature sequence extracted by the differential rate sample entropy. The generator is used to capture the distribution of the data and generate new samples closer to the real data through learning. The discriminator is used to distinguish normal data from abnormal data. In the iteration process, the data generated by the generator and the real training data are used as the input of the discriminator. The parameters of the generator and discriminator are updated by the discriminated results to finally obtain better parameters. In the stage of anomaly detection, a new calculation method for anomaly scores is designed according to the features of time series data to improve the detection accuracy. If the series reconstructed by the trained generator has higher anomaly scores than the normal samples, then the current time series can be determined as an anomaly. Compared with other methods, the detection method based on DRSe-GAN can achieve the purpose of anomaly detection without collecting large amounts of abnormal data and using normal data training. The model can identify abnormal data that does not conform to the distribution of training data and improve the ability of the model by improving the loss function and the judgment score.

The training data set  $S_{T_{\text{train}}} = \{x_1, x_2, \dots, x_u, \dots, x_{T_{\text{train}}}\}^{1 \times T_{\text{train}}}$  is given to train network parameters, where  $s_u$  is the  $u$ th data point in the data set, and all data in the training data are normal. A verified dataset  $S_{T_{\text{val}}} = \{x_1, x_2, \dots, x_{T_{\text{val}}}\}^{1 \times T_{\text{val}}}$  determines the

threshold of anomaly scores. A testing dataset  $S_{T_{\text{test}}} = \{x_1, x_2, \dots, x_{T_{\text{test}}}\}^{1 \times T_{\text{test}}}$  is used for anomaly detection, validation, and verification and tests abnormal and normal data contained in the data set.

In the process of model training, both generator and discriminator use Long Short Term Memory (LSTM) to extract time information between time series data. LSTM is evolved from RNN. It adds input gate, forgetting gate, and output gate to neuron cells, which enables the network to use a longer history state than RNN to predict.

Different from the traditional neuron node, the basic unit of the hidden layer of LSTM is a special cell structure, which contains a self-connected memory cell and three gate units controlling information flow. Among them, the input gate and output gate control the flow of information into and out of neurons, respectively, and the forgetting gate controls the degree of memory cells' state before memory.

In the training process, it is necessary to customize appropriate loss functions to guide the training according to the requirements of tasks. The loss functions in this paper include two parts, discriminator loss  $G_{\text{loss}}$  and generator loss  $D_{\text{loss}}$ .  $\text{loss} = D_{\text{loss}} + G_{\text{loss}}$  means the two losses influence the change of network parameters jointly. In other words, the two parts of losses jointly affect the changes of network parameters. The generator is used to generate data similar to the real data, and the formula of the loss function is shown in (7), where  $z$  is the random input data and  $p_z$  is the distribution of the random input data. The output of the discriminator represents the probability that the output of the generated data is true. The loss function of unsupervised learning can be expressed as (8).

$$G_{\text{loss}} = E_{z \sim p_z(z)} [\log(1 - D(G(z)))], \quad (7)$$

$$D_{\text{loss}} = -E_{X \sim p_{\text{data}}(X)} [\log D(X)] - E_{z \sim p_z(z)} [\log(1 - D(G(z)))]. \quad (8)$$

The model parameters are updated according to the loss function to get the trained generator and discriminator. The following are the detailed steps of the training stage:

Step 1:  $Z = \{z_i, i = 1, 2, \dots, n\}$  is the random sampling of noise data, where  $n$  corresponds to the number of samples. The generator model is a few LSTM memory units. The number of memory units is set.  $Z$  is input into the generator model  $G$  to generate reconstructed sample series data  $G(Z)$ .

Step 2: real sample data series  $X$  (normal mode data) and the noise generated by the sample data are input into the built discriminant model  $D$ . The generator model is a few LSTM memory units as well. The model outputs the probability that the input data is of real data, and the loss function was calculated

according to the output of the generator and the discriminator.

Step 3: update the model parameters using the gradient descent algorithm according to the value of the loss function. Update the parameters of the generator according to the noise data after the parameters of the discriminator are updated.

Step 4: save the model parameters, return to step 1 for cyclic iteration, and finally get the trained generator model  $G^*$  and discriminator model  $D^*$ .

**3.3. Anomaly Pattern Determination.** This paper utilizes the difference between the test samples and the reconstructed samples of the generator and the results of the discriminator to establish a new method to calculate the abnormal scores and detect the abnormal patterns. In this paper, the anomaly

pattern of the different rates of sample entropy is relatively larger, and the normal pattern of the sample entropy difference rate is smaller, so we consider using the relative difference between the maximum values of the reconstructed sample and the real sample to construct the abnormal scores and give different weights to the data points in the data segment to be detected. Finally, when the anomaly score of the data segment is beyond the threshold value, the sample is determined as an anomaly. The specific steps are as follows:

Step 1: first, the maximum mean difference loss function between the generated samples of random noise feature  $Z$  and the real samples is used to obtain the optimal  $Z^*$ . The maximum mean difference loss function is used to measure the distance between two different but related distributions.

Step 2: the trained discriminator  $D^*$  is used to output the probability  $P$  that the sample belongs to the real sample, and the discriminated anomaly score  $D_{\text{score}}$  is calculated as  $1 - p$ .

Step 3: the trained generator  $G^*$  is used to generate reconstructed samples based on random noise  $Z^*$ . On account of the continuous adjustment of the parameters of the generator, it can produce samples quite similar to real samples. At the same time, only samples of normal mode are included in the training samples. As a result, the distribution of the reconstructed samples is similar to the samples of normal mode. When there are anomaly pattern samples in the test sample set, the distribution of the generated sample and the real sample at the abnormal series point will produce greater errors. The generation error is used to calculate the anomaly fraction scores  $D_{\text{score}}$ .

Step 4: the anomaly fraction score  $D_{\text{score}}$  and sample generated anomaly fraction  $R_{\text{score}}$  are used to calculate anomaly fraction score  $S$ , as shown in

$$S = W_D \times D_{\text{score}} + W_G \times R_{\text{score}}, \quad (9)$$

where  $W_D$  and  $W_G$  are the weights for the discriminated anomaly fraction and the sample generated anomaly fraction, respectively. The final anomaly score can be obtained by combining the two. The calculation of determining anomaly score and sample generated anomaly score is as follows.

**3.3.1. Determining Anomaly Fraction  $D_{\text{score}}$ .** Given a test sample set  $X = \{(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)\}$ , the test sample uses the trained discriminator  $D^*$  to output the probability  $P$  that the sample is a real sample. For the sample of the normal pattern, the  $P$  value is larger when it is more consistent with the data distribution of the training set. For anomaly pattern samples, the distribution of abnormal samples is significantly different from that of normal samples because the abnormal data far deviate from the normal data, and the  $P$  value is relatively small. As a result, the discriminated anomaly score  $D_{\text{score}}$  is  $1 - P$ .

**3.3.2. Generating Anomaly Fraction  $R_{\text{score}}$ .** Assuming that the sample length is  $n$ , the generator is used to generate a sample  $G(Z^*) = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n\}$  based on random noise  $Z^*$  while the real sample is  $X_i = \{x_1, x_2, \dots, x_n\}$ . For calculating the absolute error at each time during anomaly generation, different weights are given to absolute error  $e$  considering that the biggest difference between normal and anomaly pattern is the deviation of the data from the average score in the abnormal pattern that does not exist in the normal pattern. The different weights constitute weighted absolute error. The weighted series is set as  $W_i = \{w_1, w_2, \dots, w_n\}^T$  and the weights exhibit exponential changes. The values of the weighted influence change exponentially as data that nearest to the maximum has the largest weighted influence. The anomaly score of the sample is  $R_{\text{score}} = e \cdot W_i$ . The setting of the weight series is as follows:

Step 1: sort the elements in  $e$  with an absolute error length of  $n$  from the smallest to the largest to obtain the absolute error  $E'_i = \{e'_1, e'_2, \dots, e'_n\}$ . The sizes of the element do not change but the positions of the elements change.

Step 2: calculate the average value  $M$  of the absolute error  $E'_i = \{e'_1, e'_2, \dots, e'_n\}$  after sorting. If there is abnormal data in the sample, the value of  $M$  will increase. Assume that the data element  $\{e'_k, e'_{k+1}, \dots, e'_n\}$  in  $E'_i$  is greater than the average value  $M$  with the size of  $n - k + 1$ . Initialize the weight series  $W'_i = \{w'_1, w'_2, \dots, w'_n\}^T$ ,  $w'_{1 \sim n-2} = 0$ , and set  $x'_n$  the corresponding weight  $w'_n$  as  $\lambda$ . The corresponding weight  $w'_{n-1}$  of  $x'_{n-1}$  is  $1 - \lambda$ .

Step 3: update the size of the element in the weight series  $W'_i$ . The updating of  $w'_j$  is shown in

$$w'_j = \begin{cases} 0, & j < k, \\ \lambda(1 - \lambda)^{n-j}, & j \geq k. \end{cases} \quad (10)$$

In formula (15), only the error of data elements that are greater than the mean  $M$  is calculated to reduce the number of parameters and consider the main errors between the normal pattern and the anomaly pattern. That is, the weight of data elements less than the mean  $M$  is set as 0. When  $j \geq k$ ,  $w'_j$  increases with the increase of  $j$ . The larger  $\lambda$  is, the greater the weight of the maximum value will be. When  $j = n$ ,  $w'_j = \lambda$ .

Step 4: use the updated weight series  $W'_i$  and the sorted sample  $X'_i$  to calculate the generated anomaly score  $R_{\text{score}}$  of the test sample, as shown in

$$R_{\text{score}} = E'_i \cdot W'_i. \quad (11)$$

The higher the anomaly score is, the higher the probability of anomaly is. To better distinguish between normal samples and abnormal samples, the threshold value of anomaly scores is determined by verifying the sample set  $S_{T_{\text{val}}} = \{x_1, x_2, \dots, x_{T_{\text{val}}}\}^{1 \times T_{\text{val}}}$ . That is, the maximum and minimum anomaly scores in the results of the verification



sample set are taken as the maximum and minimum boundaries and they are divided equally. The anomaly score of the  $q$ th verification is shown in

$$\text{score}_q = \min_{\text{val}} + \frac{\max_{\text{val}} - \min_{\text{val}}}{l} \times q, \quad (12)$$

where  $l$  is the set division quantity.  $\min_{\text{val}}$  is the minimum boundary, and  $\max_{\text{val}}$  is the maximum boundary. Since  $F_1$  includes both the precision and the recall of the model [25], the anomaly score corresponding to the maximum  $F_1$  score is selected as the threshold value. The calculation of  $F_1$  is shown as follows:

$$F_1 = 2 \times \frac{\text{Per} \times \text{Rec}}{\text{Per} + \text{Rec}}, \quad (13)$$

where

$$\begin{aligned} \text{Per} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Rec} &= \frac{\text{TP}}{\text{TP} + \text{FN}}. \end{aligned} \quad (14)$$

In (14),  $TP$  is the positive sample predicted to be positive by the model,  $FP$  is the negative sample predicted to be positive by the model, and  $FN$  is a positive sample predicted to be negative by the model. After determining the threshold, the testing set  $S_{T_{\text{test}}}$  is used to test the performance of the network.

## 4. Experiment and Analysis

**4.1. Experimental Data.** The data used in this experiment is the experimental data of the mined-out area model and the real data of Dafosi Coal Mines. The prototype of the model experimental platform is Dafosi 40118 fully mechanized caving face. The thickness of residual coal in two lanes of Dafosi working face is 12 meters, and the thickness of residual coal in the middle of the mined-out area is 0.92 meters. The size of the mined-out area model is  $1.2 \times 1.2 \times 0.6$  (m), the geometric similarity ratio is 150 : 1, the thickness of residual coal in two lanes of the mined-out area model is 8 cm, and the thickness of residual coal in the central mined-out area is 0.6 cm. The model is divided into three layers: upper, middle, and lower. Each layer is divided into nine square areas. Each area has four monitoring positions, and each location has gas sensors, temperature sensors, and pressure sensors. Therefore, there are 108 monitoring points. The experimental configuration of the platform is shown in Table 2.

In the experiment, the series values of CO gas variation with time in the oxidation process of residual coal were collected, including the variation of CO gas concentration in different oxidation stages. Training data is  $S_{T_{\text{train}}} = \{x_1, x_2, \dots, x_{T_{\text{train}}}\}$ . Validation data is  $S_{T_{\text{val}}} = \{x_1, x_2, \dots, x_{T_{\text{val}}}\}$ . Testing data is  $S_{T_{\text{test}}} = \{x_1, x_2, \dots, x_{T_{\text{test}}}\}$ .

First, the difference rate entropy feature series of the original data is extracted, and then, the feature series is

TABLE 2: Experimental configuration.

Category	Type of gas mixture
Coal sample selection	Dafosi Coal
Gas type	O <sub>2</sub> , CO <sub>2</sub> , CO, C <sub>2</sub> H <sub>4</sub> , C <sub>3</sub> H <sub>6</sub> , C <sub>2</sub> H <sub>2</sub> , CH <sub>4</sub>
Temperature range	30°C–530°C
Gas flow	50 ml/min
Coal particle size	10–20, 20–40, 40–80 mesh

normalized, as shown in formula (15).  $\min(S)$  is the minimum value in the series, and  $\max(S)$  is the maximum value in the series.

$$x_i = \frac{x_i - \min(S)}{\max(S) - \min(S)}. \quad (15)$$

Secondly, the normalized series is segmented and averaged in the way described in formula (5) to obtain the real input data of the network model. Experimental data description is shown in Table 3. The experimental object is Dafosi Coal, and the dimension of the time series is one-dimensional, that is, one-dimensional series. The data size is as follows: training data  $S_{T_{\text{train}}} = \{x_1, x_2, \dots, x_{8360}\}$ , validation data  $S_{T_{\text{val}}} = \{x_1, x_2, \dots, x_{5248}\}$ , and testing data  $S_{T_{\text{test}}} = \{x_1, x_2, \dots, x_{3574}\}$ . There are only normal samples in the training data and normal and abnormal samples in the validation set.

Finally, the sliding window is used to segment the data in order to effectively extract the data patterns contained in different stages of coal spontaneous combustion. In this way, the data in the window can be detected and analyzed in time. Set the window size as  $w$  and step size as  $b$ ; then, each window data segment is a sample, corresponding to a label  $\in (0, 1)$ . 0 means normal and 1 means abnormal and the total number of samples is  $n = (T_{\text{train}} - w) / (b)$ . When the window size and step size are different, the number of samples will be different, and the experimental results will be different.

**4.2. Experimental Results.** Figure 6(a) shows the original series segment containing the normal pattern and the anomaly patterns that the 100~250 data segments are the abnormal pattern data segment with the concentration of gas changing greatly. Figure 6(b) shows the differential rate series curve of the original series segment in response. The window size of the series segment is set as 10 and the step size is 1. It can be seen that the maximum value of the difference rate corresponding to the data segment of abnormal mode is 0.6, and the data change is negative. The feature series of sample entropy is shown in Figure 6(c), the length of subseries is 2, and the threshold of similarity comparison is the standard deviation  $\times 2$ . The larger part of the eigenvalue corresponds to the anomaly pattern of the original series. The characteristic value of the normal model fluctuates roughly between 0.25 and 0.75, while the fluctuation range of the abnormal model fluctuates far beyond the range of other parts, reaching 1.79 at the highest. The series after series aggregation averaging process segment by

TABLE 3: Description of experimental data.

Categories	Value	Number of periods	Anomaly pattern determination	Anomaly rate (%)
Select coal sample	Dafosi Coal	—	—	—
Variable dimension	1	—	—	—
Training set	8360	16	Normal	0
Validation set	5248	10	Abnormal	20.84
Testing set	3574	7	Abnormal	19.23

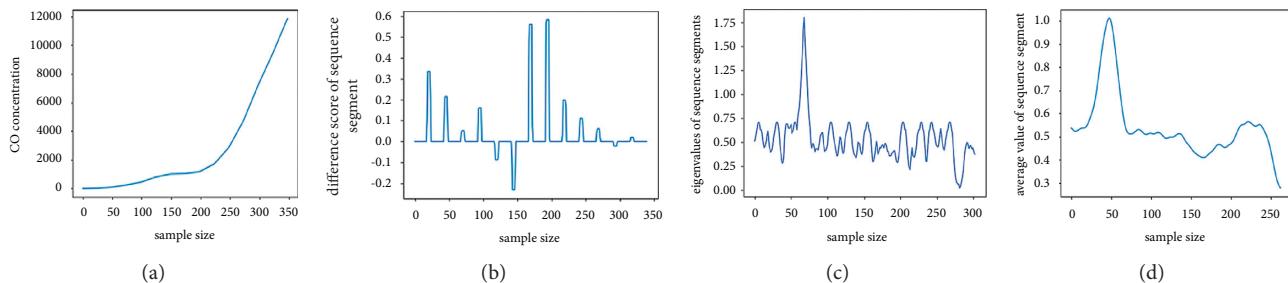


FIGURE 6: Series difference result figure.

segment is shown in Figure 6(d). It can be seen that the data is smoother. It can reduce the fluctuation of normal mode data and bring large errors for abnormal mode detection without changing the eigenvalue difference between abnormal mode and normal mode.

Figure 7 shows the characteristics of different entropy series of the same difference rate sequence in which Figures 7(a)–7(c) are, respectively, Shannon entropy, sample entropy, and permutation entropy. These entropies measure the complexity of time series on a single scale. The corresponding statistical features are shown in Table 3, including the minimum value, maximum value, and average value of data segments of normal mode and abnormal mode, as well as the threshold in the abnormal mode. Table 4 shows the percentage difference between the statistical features of the anomaly pattern and that of the normal patterns.

As can be seen from Table 4, the minimum, maximum, and average values of the anomaly pattern are lower than those of the normal pattern for Shannon entropy and permutation entropy, while the sample entropy is the opposite. In addition, before the occurrence of the anomaly pattern, the entropy series has found the state where the anomaly pattern begins to appear in advance, that is, the threshold where the anomaly pattern appears, which helps with prediction and warning. The minimum threshold of the sample entropy is 62, and the maximum threshold of the Shannon entropy is 111, indicating that the sample entropy can manifest the anomaly pattern first. As can be seen from Table 5, as for the percentage error of the statistical characteristics of each entropy series of the anomaly pattern, the sample entropy has obtained large results, which indicates that, in the sample entropy series, the anomaly pattern is significantly different from the normal pattern. It is more conducive to distinguishing the two patterns.

The data processed by the differential rate entropy feature is preprocessed as the input signal for generative adversarial networks to train the network. The number of

network training iterations is 1000, the learning rate is 0.1, and the number of generator and discriminator training is set as 100. Under this setting, the variation curves of discriminator loss and generator loss function are shown in Figure 8.

It can be seen from Figure 8 that the generator and discriminator have trained alternately in turn during the training process and the corresponding generator loss function and discriminator loss function change in opposite trends. The generator loss first decreases, then increases, then decreases, and then increases, while the discriminator loss first increases, then decreases, then increases, and then decreases. The final discriminator loss function converges to about 0.1311, and the generator loss function converges to about 13.99.

Figure 9 is the schematic diagram of the generated sample graph and the real sample at different iteration numbers. Figure 9(a) is the comparison diagram of the generated sample and the real sample at the fifth generation. It can be seen that the generated sample cannot learn the curve pattern of the real data well. At this time,  $D_{loss}$  is 0.1825, and  $G_{loss}$  is 5.701. Figure 9(b) shows the comparison between the generated samples and the real samples in the 78th generation. It can be seen that the generated samples can better learn the curve pattern of the real data. At this time,  $D_{loss}$  is 4.4680 and  $G_{loss}$  is 0.0509.

The series segment length of the sample can affect detected results to a certain extent. It can be either too short to include whole pattern distribution or too long to handle the difficulty of network learning, which results in a higher error rate. Therefore, it is necessary to find proper sample series segment length to optimize detection precision. Table 6 shows detection results for different sample lengths.

As shown in Table 5, for all sample lengths, the precision is lower than the recall, indicating that some normal samples are detected as abnormal samples, while abnormal samples are almost detected. When the series segment is 40, the F1

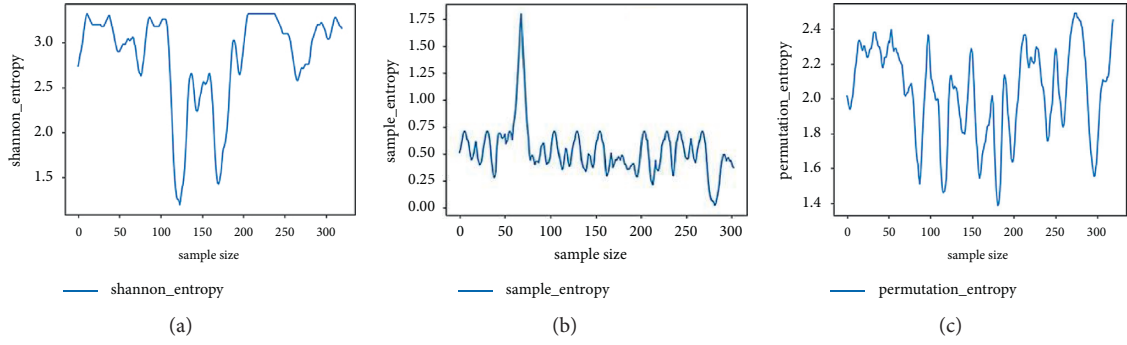


FIGURE 7: Comparison of entropy of different time series.

TABLE 4: Statistical characteristics of different entropy series.

Entropy category	Normal pattern			Anomaly pattern		Threshold
	Max	Min	Mean	Min	Max	
Shannon entropy	2.5792	3.3219	3.0829	1.1934	2.6592	111
Sample entropy	0.0206	0.7082	0.4911	0.7255	1.7987	62
Permutation entropy	1.5551	2.4905	2.1563	1.3868	2.3655	80

TABLE 5: Percentage difference.

Entropy category	Min (%)	Max (%)	Mean (%)
Shannon entropy	-53.72	-19.94	-33.83
Sample entropy	3421	153.9	139.2
Permutation entropy	-10.8	-5.02	-13.73

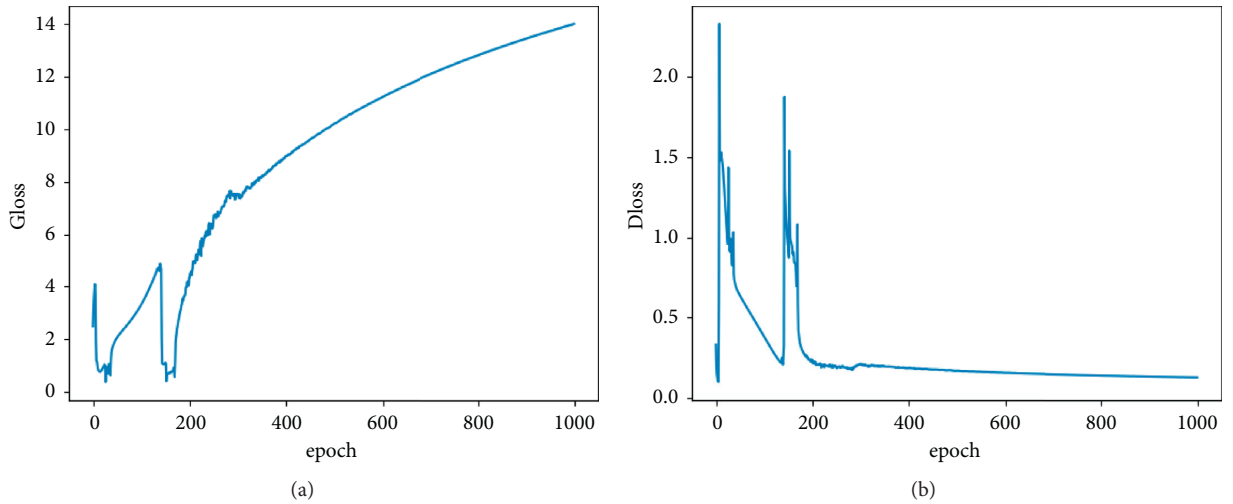


FIGURE 8: Variation curves of generator loss and discriminator loss functions.

score reaches the maximum value of 0.8916, the corresponding precision is 83.87%, and the recall is 96.3%. When the series segment is larger than 40, the detection ability significantly decreases and the precision is relatively low. It indicates that when the series length is too long, many normal patterns will also be determined as abnormal patterns. When the series segment is less than 40, the detection rate also decreases. It indicates that the distribution of the

normal pattern has not been well learned, which affects the determination of the anomaly pattern. Table 7 shows the comparison of the proposed method, some common unsupervised methods, and results without being processed by difference rate sample entropy.

As shown in Table 6, the test results of the DRSe-GAN model are the best, and the prediction results of the KNN network model are the worst. The prediction results of the GAN

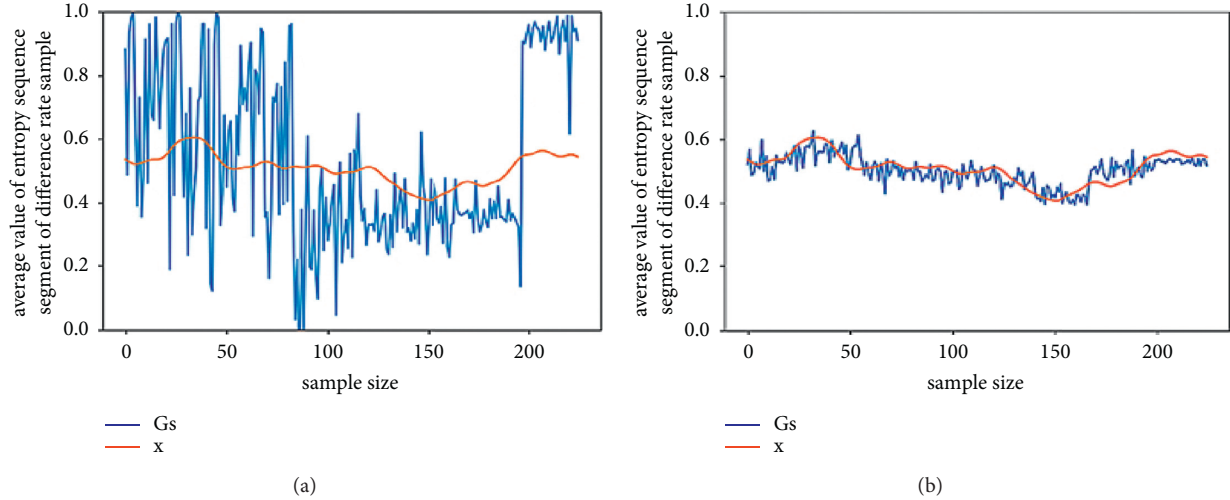


FIGURE 9: Schematic diagram of generating samples with different iteration numbers.

TABLE 6: Detection results for different sample length.

Sample length	Precision (%)	Recall (%)	F1 score
20	75.01	93.75	0.8284
30	83.03	92.6	0.8753
40	83.87	96.3	0.8916
50	54.84	94.44	0.6893
60	36.67	91.67	0.5198

TABLE 7: Test results of different methods.

Methods	F1 score	Computation time (s)
KNN	0.4214	208.6
AutoEncoder	0.6039	230.5
GAN	0.6348	346.2
DRSe-GAN	0.8916	457.8

network are similar to those of AutoEncoder and are even better than them. Compared with the general unsupervised network, the GAN network can learn the distribution of the normal mode better and is relatively more sensitive to anomaly patterns. At the same time, carrying out difference rate sample entropy feature processing can highlight the differences between abnormal and normal samples to improve the detection precision of the model. At the same time, the running time of the DRSe-GAN model is slightly longer compared with other models due to the adding of the different rate entropy sample processing of the data. Nevertheless, the running time is shorter than the collection interval of gas sensor data. To sum up, time series anomaly detection based on difference rate entropy feature and the generative adversarial network is more suitable for processing the coal mine index gas concentration data used in the present study, with the highest detection accuracy and relatively appropriate running time.

## 5. Conclusion

Based on the features of mined-out area coal compound disasters, this paper analyzes the data of the index gas CO

that can manifest the process of the disaster and propose a method utilizing time series entropy feature and generative adversarial network to detect the anomaly. The method detects whether the series contains abnormal patterns to provide references for the judgment of the occurrence of thermodynamic disasters in coal mines.

This method has the following characteristics:

- (1) The method consists of two modules: anomaly pattern extraction module and anomaly pattern detection module. The anomaly pattern extraction module processes the original data and finally produces the original difference rate sample entropy series as the input of the detection network.
- (2) In the abnormal pattern detection module, according to its data characteristics, a new calculation method of anomaly score is proposed. The difference between the generated sample and the real sample is added based on the discriminant output produced by the generative adversarial network. Different weights are given to the errors at different time points to improve the detection rate.

- (3) The experiment compared the detection accuracy of this method with other methods, which proved the effectiveness of the proposed method and the highest detection accuracy.

This method is aimed at the iconic gas data in a specific mined-out area. Although the change rule of this type of data is similar, for different coal mines, the boundary of the data and the types of the iconic gas will be different. Therefore, in future research, we hope to further study the idea of data feature extraction, further improve the model algorithm, increase the generalization of the model, and strive to apply the model to more coal mines.

## Data Availability

Data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This study was funded by the National Key Research and Development Project (2018YFC0807903), the Natural Science Foundation of Liaoning Province of China (2020-KF-23-06) in 2020, and the Central Guidelines for Local Science and Technology Development Special Fund Subsidies (2021JH6/10500134).

## References

- [1] M. Onifade and B. Genc, "A review of research on spontaneous combustion of coal," *International Journal of Mining Science and Technology*, vol. 30, no. 3, pp. 303–311, 2020.
- [2] H. Sechman, M. J. Kotarba, S. Kędzior et al., "Fluctuations in methane and carbon dioxide concentrations in the near-surface zone and their genetic characterization in abandoned and active coal mines in the SW part of the Upper Silesian Coal Basin, Poland," *International Journal of Coal Geology*, vol. 227, Article ID 103529, 2020.
- [3] S. Song, S. Wang, S. Jiang et al., "Multifield coupled dynamic simulation of coal oxidation and self-heating in longwall coal mine gob," vol. 2020, Article ID 9075657, 16 pages, 2020.
- [4] R. K. Mishra, P. N. S. Roy, V. K. Singh et al., "Detection and delineation of coal mine fire in Jharia coal field, India using geophysical approach: a case study," *Journal of Earth System Science*, vol. 127, no. 8, pp. 1–10, 2018.
- [5] X. Hu, S. Yang, X. Zhou, Z. Yu, and C. Hu, "Coal spontaneous combustion prediction in gob using chaos analysis on gas indicators from upper tunnel," *Journal of Natural Gas Science and Engineering*, vol. 26, pp. 461–469, 2015.
- [6] D. Jun, L. I. Bei, L. I. Zhen-Bao et al., "Experiment study on gas indexes optimization for coal spontaneous combustion prediction," *Coal ence and Technology*, vol. 4, no. 21, pp. 2055–2056, 2014.
- [7] X. Y. Wan and H. Yang, "Research on prediction system of spontaneous combustion in coal mine goaf based on multi-sensor fusion technology," *Advanced Materials Research*, vol. 550–553, pp. 2887–2890, 2012.
- [8] H.-y. Niu, X.-l. Deng, S.-l. Li et al., "Experiment study of optimization on prediction index gases of coal spontaneous combustion," *Journal of Central South University*, vol. 23, no. 9, pp. 2321–2328, 2016.
- [9] J. Li, W. Pedrycz, and I. Jamal, "Multivariate time series anomaly detection: a framework of Hidden Markov Models," *Applied Soft Computing*, vol. 60, pp. 229–240, 2017.
- [10] M. Canizo, I. Triguero, A. Conde, and E. Onieva, "Multi-head CNN-RNN for multi-time series anomaly detection: an industrial case study," *Neurocomputing*, vol. 363, pp. 246–260, 2019.
- [11] L. Beggel, B. X. Kausler, M. Schiegg, M. Pfeiffer, and B. Bischl, "Time series anomaly detection based on shapelet learning," *Computational Statistics*, vol. 34, no. 3, pp. 945–976, 2019.
- [12] P. Malhotra, L. Vig, G. Shroff et al., "Long short term memory networks for anomaly detection in time series," *Proceedings, Presses Universitaires de Louvain*, vol. 89, 2015.
- [13] S. Chauhan and L. Vig, "Anomaly detection in ECG time signals via deep long short-term memory networks," in *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1–7, IEEE, Paris, France, October 2015.
- [14] V. Rajagopalan and A. Ray, "Symbolic time series analysis via wavelet-based partitioning," *Signal Processing*, vol. 86, no. 11, pp. 3309–3320, 2006.
- [15] H. Izakian and W. Pedrycz, "Anomaly detection and characterization in spatial time series data: a cluster-centric approach," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 6, pp. 1612–1624, 2014.
- [16] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.
- [17] D. Li, D. Chen, L. Shi et al., "MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks," 2019.
- [18] T. W. Tang, W. H. Kuo, J. H. Lan, C. F. Ding, H. Hsu, and H. T. Young, "Anomaly detection neural network with dual auto-encoders GAN and its industrial inspection applications," *Sensors*, vol. 20, no. 12, 2020.
- [19] Y. Jia and H. Gu, "Sample entropy combined with the K-means clustering algorithm reveals six functional networks of the brain," vol. 21, no. 12, 2019.
- [20] D. Looney, T. Adjei, and P. Danilo, "Mandic. A novel multivariate sample entropy," *Algorithm for Modeling Time Series Synchronization*, vol. 20, no. 2, 2018.
- [21] T. Amarbayasgalan, V. H. Pham, N. Theera-Umpon, and K. H. Ryu, "Unsupervised anomaly detection approach for time-series in multi-domains using deep reconstruction error," *Symmetry*, vol. 12, no. 8, p. 1251, 2020.
- [22] T. Fujioka, K. Kubota, M. Mori et al., "Efficient anomaly detection with generative adversarial network for breast ultrasound imaging," *Diagnostics*, vol. 10, no. 7, 2020.
- [23] M. N. Fekri, A. M. Ghosh, and K. Grolinger, "Generating energy data for machine learning with recurrent generative," *Adversarial Networks*, vol. 13, no. 1, 2019.
- [24] Z. Zhai, "Auto-encoder generative adversarial networks," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 3, pp. 1–7, 2018.
- [25] A. Borji, "Pros and cons of GAN evaluation measures," *Computer Vision and Image Understanding*, vol. 179, pp. 41–65, 2019.