Article

Prediction of Gas Concentration Based on LSTM-LightGBM Variable Weight Combination Model

Xiangqian Wang *, Ningke Xu, Xiangrui Meng and Haoqian Chang

School of Computer Science and Technology, Anhui University of Science & Technology, Huainan 232000, China; nkxu999@gmail.com (N.X.); xrmeng@aust.edu.cn (X.M.); hqchang@aust.edu.cn (H.C.)
* Correspondence: xiqwang@aust.edu.cn; Tel.: +86-15309648996

Abstract: Gas accidents threaten the safety of underground coal mining, which are always accompanied by abnormal gas concentration trend. The purpose of this paper is to improve the prediction accuracy of gas concentration so as to prevent gas accidents and improve the level of coal mine safety management. Combining the LSTM model with the LightGBM model, the LSTM-LightGBM model is proposed with variable weight combination method based on residual assignment, which considers not only the time subsequence feature of data, but also the nonlinear characteristics of data. During the data preprocessing, the optimal parameters of gas concentration prediction are determined through the analysis of the Pearson correlation coefficients of different sensor data. The experimental results demonstrate that the mean absolute errors of LSTM-LighGBM, LSTM and LightGBM are 1.94%, 2.19% and 2.77%, respectively. The accuracy of LSTM-LightGBM variable weight combination model is better than that of the two above models, respectively. In this way, this study provides a novel idea and method for gas accident prevention based on gas concentration prediction.

1. Introduction

Energy is the engine of economic development and the lifeblood of national economy [1]. Coal is crucial with respect to the energy strategy of China, which is also caused by the feature of resource distribution in China, but also it determines that the solution to energy problems should depend on coal. For a long time, safety has always been one of the important issues during the process of coal mining. Gas accidents are a particularly serious problem. Through the investigation and analysis of coal mine gas accidents, it is found that not accurately grasping the law of gas concentration changes is the main reason for gas accidents [2]. Thus, if the inner rules can be explored and the gas concentration can be predicted relatively accurately [3], it will be of great importance to reduce the occurrence of gas accidents.

So far, many domestic and foreign scholars have conducted a great amount of research on gas concentration prediction [4]. Normally, gas concentration prediction methods can be broadly divided into two categories, one of which is using gas geomathematical modeling methods, and the other of which is based on machine learning methods. However, since the change of gas concentration is not a simple static process, and there are highly complex nonlinear relationship among its the influencing factors, it is still a great challenge for the current gas concentration prediction models to predict gas concentration accurately and efficiently [5].

The prediction of gas concentration using the gas geomathematical model requires detailed measurements of multidimensional attributes of the geological environment surrounding the mine and the underground environment, such as mining depth, permeability of coal seam, stability of coal seam and thickness of the coal seam. Wang et al. [6] constructed the gas concentration prediction equation based on one-dimensional regression

analysis. Zhang et al. [7] established the multivariate prediction model of gas concentration using the actual measured parameters of gas gushing from the mined area. Lu et al. [8] combined the gas gushing characteristics and gas gushing mechanism to construct a mathematical model of gas geology. However, based on the kind of methods for gas concentration prediction, it is not easy to obtain necessary input data, and not possible to achieve real-time prediction. Furthermore, in the process of model building, the prediction equation needs to be adjusted artificially based on experience, and it lacks the consideration of gas concentration time-series correlation.

As machine learning becomes more and more widely used in many fields, machine learning algorithms have been applied to gas concentration prediction. The previous studies focusing on prediction of gas concentration are mainly based on single factor, historical gas data or conventional single machine learning models such as the recurrent neural network (RNN) [9], eXtreme gradient boosting (XGBoost) model [10], the random forest model (RF) [11], backpropagation (BP) neural network [12] and long short-term memory (LSTM) network [13]. These algorithms have been used to predict the gas concentration in the short term. A comparison between the prediction values of gas concentration in several machine learning models demonstrated that LSTM network has a better generalization ability, and it can deal with nonlinear time sequence data on the basis of solving the defect of traditional recurrent neural network [14]. The light gradient boosting machine (LightGBM) [15] operates faster and it is accurate compared with that of XGBoost in the multiple benchmarks and public data set test. To further improve the precision of gas prediction, a few researchers have attempted to predict the gas concentration by combining several single machine learning models. Xun et al. [16] constructed a CNN-LSTM model. Lin et al. [17] combined PSO-BP neural network to predict the gas content of coral beds. Wen et al. [18] developed a BP neural network model based on Gray theory. Xu et al. [19] developed a IGSA-BP combination prediction model that had a better prediction accuracy than that of the single machine learning model. Zhang et al. [20] constructed a prediction model based on a combination of wavelet noise reduction and LSTM. Han et al. [21] constructed a gas concentration residual correction model based on Markov model and Gray neural network. However, majority of the combination models place the first prediction results into another model for the secondary prediction or sum up the prediction results of the two models to utilize the average value. Combination models that adopt this strategy do not "integrate" two single-machine models; this also results in their prediction accuracy still not meeting the needs of underground coal mine safety production.

Considering the drawbacks of the abovementioned studies, in this paper, the historical data of this survey site was selected as the time sequence factor, and the historical data of other survey sites at the working face was selected as a spatial topological factor, and these were combined. An analysis of the correlation between the attribute data and gas concentration is used to define the attribute requirements of the input data. According to the data time sequence and nonlinear characteristics, the variable weight combination model [22] of the LSTM network and the LightGBM model was developed to dynamically predict the gas concentration for the next 10 h. The model conquers the difficulty in obtaining data and inability to predict in real time by traditional gas geomathematical models and improves the accuracy of gas concentration prediction using the improved variable weight combination method of residual weighting. The prediction of gas concentration change trend can be as an important reference for safety management in coal mines to take measures such as gas extraction, water misting, boosting wind speed and other methods in time to ensure a better prevention of gas accidents.

2. Date Source

Since coal is the main source of energy in China, the safety problems related to coal mining have attracted significant attention. A large volume of gas gush is generated in the working face of the gas mine during the process of the production. By referring to the pre-decessor's data collection scale when predicting the gas concentration, [23,24] in this

study, 10,000 sets of data were collected from 11 different survey sites at the working face of a coal mine in Shanxi Province from 19 March 2021 to 24 March 2021. The description of data attribute is shown in Table 1.

Table 1. Data attribute description of each measuring point at a working face.

Measurement Point Name	Measurement Point Description	Index	Max Value	Min Value
MGas	Mixed methane concentration in air entry	% CH_4	0.7	0
EGas	Methane concentration of back air in air inlet drift	% CH_4	0.7	0
Gas1	Methane concentration in the downwind side of the tunnel	% CH_4	0.79	0.16
Gas2	Methane concentration in working face of air entry	% CH_4	0.4	0
YCO1	Concentration of carbon monoxide in the downwind side of tunnel drilling	ppm	6	0
YCO2	Concentration of carbon monoxide at the head of the belt conveyor in the air inlet lane	ppm	6	0
WS	Back air speed in air entry	m/s	1.2	0.2
FC	Dust on working face of air entry	mg/m^3	0	0
ET	Back air temperature in air entry	°C	13.3	10.8
GD	Mixed instantaneous flow in air inlet pipeline	m^3	19.29	0
SM	Smoke on the downwind side of the head of the belt driven into the air entry	mg/m^3	0	0

2.1. Missing Data Processing

Due to various force majeure factors in the data collection, transmission and storage scenarios, some data can be missing. Missing data can cause serious impediments to subsequent data correlation analysis and the construction of gas concentration prediction models. In addition to reducing the validity of the data, it can also lead to inaccuracies in the overall data analysis task and produce incorrect analysis results. Hence, this paper adopts the average method to fill in the missing data. The data filling equation is given as follows:

$$\widetilde{x} = \frac{\sum\limits_{i=1}^{n} x_i}{n} \tag{1}$$

In the above formula, $\widetilde{x}$ represents the missing data series, $\sum\limits_{i=1}^{n} x_i$ represents the total of all data in the data set and n represents the number of nonmissing data in the data set.

2.2. Normalization Process

In order to eliminate the impact of the dimensionality between the gas multiparameter time series, it is necessary to perform data normalization. Following data normalization of the raw data, the indicators are in the same order of magnitude and suitable for comprehensive comparative evaluation. Meanwhile, normalization provides a certain degree of numerical comparability of features among different dimensions. The original time series x

is normalized by applying min–max normalization. The normalization formula is given as follows:

$$x* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

where x^* is the normalized value, $x_{\max}$, $x_{\min}$ are the maximum and minimum values of the sample data respectively.

2.3. Feature Selection

After the data have been preprocessed, it is necessary to select meaningful features to input into the machine learning algorithms and models for training. Generally, feature selection is divided into the following two main steps:

2.3.1. Correlation Analysis

In order to fulfill the requirements of gas concentration prediction and to strengthen the situational awareness and extrapolation capability of the prediction model, in this paper, we use the Pearson correlation coefficient to describe the degree of correlation between gas concentration at the working face and its impact factors. The equation is given as follows:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{3}$$

In the above equation, $\rho_{X,Y}$ represent the Pearson correlation coefficient of two continuous variables X, Y, $\text{cov}(X, Y)$ represents the covariance between them, and σ_X and σ_Y represent the standard deviations of the variables X and Y.

2.3.2. Eliminate Redundant Features

Using the Pearson correlation coefficient to obtain the weights of each feature, the features with weights less than a threshold value are eliminated. Afterward, the mutual information is calculated for the features in the remaining data set two by two. Mutual information refers to the extent of information shared between two features. If the value of mutual information is greater than the threshold, the feature with the smaller weight is considered redundant and is removed. The equation for calculating mutual information is given as follows:

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{4}$$

In the above formula, $p(x,y)$ is the joint probability distribution function of X and Y, and $p(X)$ and $p(Y)$ are the marginal probability density functions of X and Y.

3. Materials and Methods

3.1. LightGBM

XGBoost should be defined before explaining about LightGBM [25], XGBoost is an improved boosting algorithm of the gradient boosting decision tree (GBDT), which is GBDT in essence, but it strives to maximize the speed and efficiency. Conventional GBDT adopts classification and regression tree (CART) as the base classifier, and XGBoost supports the multiple base classifiers to compensate for the shortcoming in the accuracy of single CART prediction. However, the disadvantages associated to XGBoost are that it stores feature sorting results, which occupy a massive amount of memory, and it severely affects cache optimization.

Compared with that of XGBoost, LightGBM [26] is a relatively new tree-based gradient boosting variant. It adopts the histogram algorithm to ensure that an algorithm utilizes less memory and has a low computational cost. Layer-by-layer growth is a conventional method used for tree based combination (including XGBoost) growth decision trees. LightGBM is different from that of XGBoost, as it does not utilize the conventional decision tree growth strategy and it introduces leaf-by-leaf growth strategy. In contrast to layer-by-layer growth,

leaf-by-leaf growth strategy converges faster and consumes lesser memory. Layer-by-layer growth strategy and leaf-by-leaf growth strategy are shown in Figure 1.
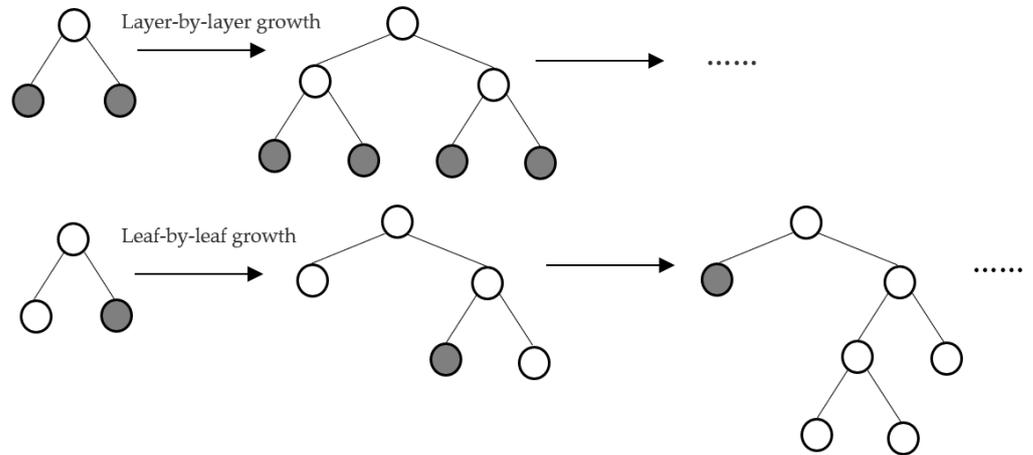

Figure 1. Layer-by-layer growth and leaf-by-leaf growth.

3.2. LSTM

LSTM [27] consists of a set of cyclic subnetworks named according to the memory blocks. Each memory block consists one or multiple self-connected memory cells and three gating units: input gate, output gate, and forget gate. Similar to that of the recurrent neural network (RNN), the hidden unit is horizontally connected back to the hidden unit. However, the hidden unit of RNN is replaced by the memory cell with gating function. The diagram of LSTM structure of a single cell is shown in Figure 2.

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\widetilde{C}_t = \tan h(w_c \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

$$C_t = f_t \times C_{t-1} + i_t * \widetilde{C}_t \tag{8}$$

$$O_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

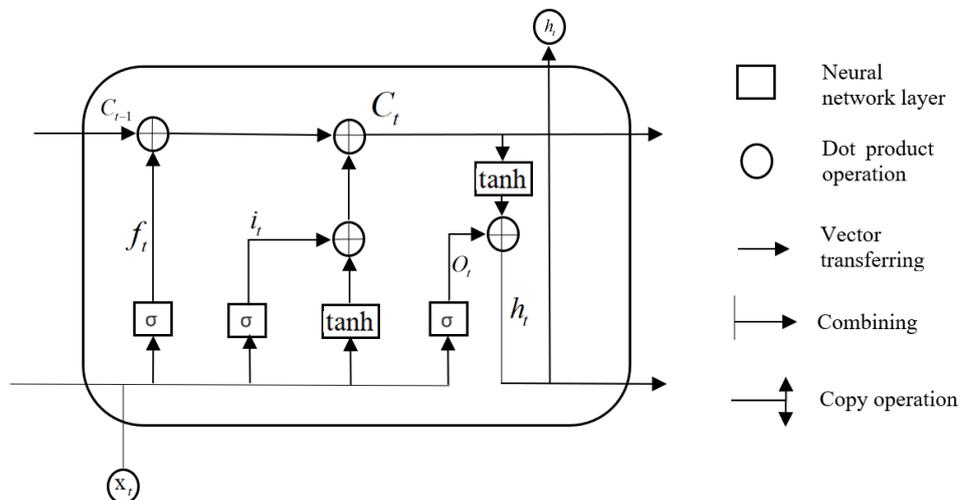$$h_t = O_t \times \tan h(C_t) \tag{10}$$

Figure 2. LSTM structure diagram.

In the above formula, f_t represents the forget gate. It is used to control whether or not to filter the hidden cellular state of the upper layer in the LSTM. i_t represents the input gate, $\widetilde{C}_t$ is the cell state at the previous moment, C_t is the cell state at the present moment, O_t represents the output gate, x_t and h_t represent the input and output at the current moment and σ and tanh represent the sigmoid function and hyperbolic tangent function, respectively. The forget gate, input gate, output gate and the weight matrix of the cell state are represented by w_f, w_i, w_o and w_c respectively. b_f, b_i, b_o and b_c represent the offset vector of the forget gate, input gate, output gate and cell state, respectively.

3.2.1. Activation Function

The sigmoid function is used as the activation function for the forgetting, input and output gates in the LSTM. The tanh function is used as the activation function when generating candidate memories. Both are saturated functions. If a nonsaturated activation function is used, the past and present memory blocks will be superimposed all the time, resulting in memory misalignment and making it difficult to achieve the gating effect [28].

Sigmoid is a commonly used activation function in gating structures. It compresses the values to between 0 and 1, which can help update and forget information. In fact, sigmoid activation function is the common choice for almost all modern neural network modules in gating.

Tanh activation function is used to generate candidate memories. This is due to the fact that tanh function has a larger gradient than the sigmoid function, which makes the model converge faster. Likewise, if a nonsaturated activation function is used to generate the candidate memory, it is likely that the output values may explode or the gradient may disappear. Hence, in this paper, we choose tanh activation function as the activation function.

3.2.2. Overfitting

High fit is a key sign of a good model. However, in the process of model fitting, if the pursuit of high *R*-squared is pursued, it is likely that some of the characteristics of the training sample itself will be taken as general properties that all potential samples will have. As a result, this can lead to a reduction in the generalization performance of the model. This phenomenon is called "overfitting" in machine learning and cannot be completely avoided in model training. All we can do is "reduce the risk", and currently, there are several ways to prevent model overfitting:

1. Data enhancement: Employing more data for model training helps to better identify signals and avoid identifying noise as signals.
2. Pretermination: Pretermination prevents overfitting by stopping the iteration of the model before it converges on the training data set.
3. Regularization: Regularization refers to the process of optimizing the objective function or cost function by adding a regular term after the objective function or cost function, typically L1 regular or L2 regular, etc.
4. Dropout: Dropout is implemented by randomly "removing" the hidden units from the neural network after the model training has started.

In order to prevent overfitting of the LSTM model in this paper, pretermination and the addition of a dropout layer are used. First, by recording the best validation accuracy so far during the training process, when after five consecutive iterations, no better validation accuracy is produced, then we can terminate the model early by default. Furthermore, we add a dropout layer to the model to reduce the complex coadaptation between neurons. Once the hidden layer neurons are randomly removed, the fully connected network is sparse, which can effectively reduce the synergistic effect of different features and enhance the generalization ability of the model. Due to the addition of the dropout layer, the model has a certain randomness in prediction, so the 10 predictions of the LSTM model are taken and averaged as the final prediction result.

3.3. Grid Search Algorithm

A reasonable set of model parameters is the basis for building a good model, and the impact of hyperparameters on the effectiveness of the model is crucial. The grid search algorithm refers to an exhaustive list of parameter values. By combining the values determined by the range of values for each parameter and the search step, a "grid" is generated by listing all possible results. Subsequently, the combinations are used to train the model, and an optimal combination of parameters is returned after all combinations have been tried.

3.4. Improved Variable Weight Combination Model

During the gas concentration prediction performed by the conventional combination model, different models are adopted to predict the gas concentration with the same working face. The appropriate weights are assigned to the prediction values, and then combined. The combined prediction model can reduce the effect of random factors of the single forecasting model and effectively improve the prediction precision.

In this study, LSTM-LightGBM equal weight combination model, LSTM-LightGBM residual weight combination model, and improved LSTM-LightGBM variable weight combination model were developed.

3.4.1. Development of Single Machine Learning Model

Ensuring the prediction accuracy and performance of single machine learning model is the basis of determining the combination model—specifically, based on previous research and parameter comparison between LSTM neural network models. Using the grid search algorithm mentioned in Section 3.3 for hyperparameter search optimization of the LSTM model, it is determined that the search range of the first layer cell count is from 20 to 200 with a search step of 20, the search range of the second layer cell count is from 10 to 100 with a search step of 20 and the number of iterations is set to 10 to 40 with a search step of 10. The layer of the network model was set to 2. The activation probability of the dropout layer was set to 0.2, the number of the unit in the first layer was set to 100, the number of units in the second layer was set to 50 and the activation function was set to Tanh. The optimization algorithm adopted the Adam algorithm, and the iteration number was set to 20 times.

Grid search algorithm [29] was used to optimize the superparameter of LightGBM model. The final parameters of the model were set as: max _depth = 6, learning _rate = 0.2, n _estimators = 180, subsample = 0.6, colsample _by tree = 0.85, silent = True.

3.4.2. Weighing of the Residual Combination Model

It is a common method to provide a single model a proper weight to develop the combination model under the condition that the accuracy of the single machine learning model remains the same. This can improve the accuracy of the model [30]. The most extensively used weighting method is equivalent weighting. In general, the method of equivalent weighting is simple, and it has a good universality and participation. However, it does not reflect the importance that the model attaches to the prediction results of different single models, and it is possible that the determined weight is considerably different from that of the actual importance of the prediction results. The residual weighting combination model is expressed as:

$$h(x_t) = \sum_{i=1}^{m} \omega_i(t-1) f_i(x_t) \tag{11}$$

$$\omega_i(t-1) = \frac{\frac{1}{\overline{\varphi_i(t-1)}}}{\sum\limits_{i=1}^{m} \frac{1}{\overline{\varphi_i(t-1)}}} \tag{12}$$

$$\sum_{i=1}^{m} \omega_i(t-1) = 1, \omega_i(t-1) \geq 0 \tag{13}$$

where $w_i(t-1)$ is the weight of the ith model at the moment of $t-1$, $f_i(x_t)$ is the prediction value of the ith model, $h(x_t)$ is the prediction value of combination model, $\overline{\varphi}_i(t-1)$ is the square sum of the predictive errors of ith model at the moment of $t-1$. The central idea of residuals weighting is to assign the weight to describe the importance of the model based on the error between the prediction value and the real value.

3.4.3. Weighting of Improved Variable Weight Combination Model

Compared with that of the conventional prediction method, there are a few improvements in data input dimension in this study. Conventional gas concentration prediction models only adopt the single dimension input model. The improved algorithm proposed in this study adopts multidimension input method based on data correlation analysis. It reveals the constraint of the single dimension input model, and it provides a theoretical basis to explore the relationship between other compounds and gas concentration.

LSTM-LightGBM variable weight combination model was developed using the improved variable weight combination method based on residual weight. The residual weighting model was improved based on weight of the moments obtained in Formula (12), and the optimal m value was calculated. The average of the weights of the first m moment was used for the initial weighting. The expression for the initial weighting is:

$$\omega_j(t) = \frac{1}{m} \sum_{k=1}^{m} \omega_i(t-k) \ (m=6), \tag{14}$$

After gaining the weight of the models from Formulas (12) and (14), the absolute value of the error between the predicted value and the true value of each combination model at the moment of t is calculated as $\delta_{i,t}$ and $\delta_{j,t}$.

$$\delta_{i,t} = \sum_{i=1}^{m} \omega_i(t) f_i(x_t) - \widehat{f(x_t)} \tag{15}$$

$$\delta_{j,t} = \sum_{i=1}^{m} \omega_j(t) f_j(x_t) - \widehat{f(x_t)} \tag{16}$$

The values of $\delta_{i,t}$ and $\delta_{j,t}$, are compared. If $\delta_{i,t} < \delta_{j,t}$, the new weight $w_j(t)$ of the combination model will replace the previous weight $w_i(t)$. Otherwise, the previous weight will remain unchanged.

3.5. Construction Flow of Prediction Model

The construction flow of the prediction model is shown in Figure 3. The main processes include data preprocessing, prediction of the single machine learning model, construction of the variable weight combination prediction model and the evaluation and analysis of the model prediction [31].

(1) Data preprocessing: Data preprocessing is an important link before data modeling, which fundamentally determines the quality of the data work and the output value. The data in this study was obtained from the working face of a coal mine in Shanxi Province. The data is relatively complete. Therefore, the data are directly normalized. The data attribute and the data correlation are considered and the suitable data from the data set is selected for the model training.

(2) Development of single machine learning model: After the data set is divided according to the scale of the training set:verification set:testing set = 7:2:1, the LSTM model and LightGBM model are trained by the data of the training set, and the data of verification set is used to adjust the parameters and monitor if the model has been fitted. The data

of the test set are placed into two models, respectively, and the prediction results of the single machine learning model are obtained.

(3) Development of improved variable weight combination model. The weight of each single machine learning model is determined by the improved weighting method shown in Section 3.4.3 to ensure that the improved prediction model can be obtained.

(4) Model evaluation analysis: According to the indexes of the model evaluation, the prediction ability of the improved model was compared and the change in the prediction effect of the model is analyzed.
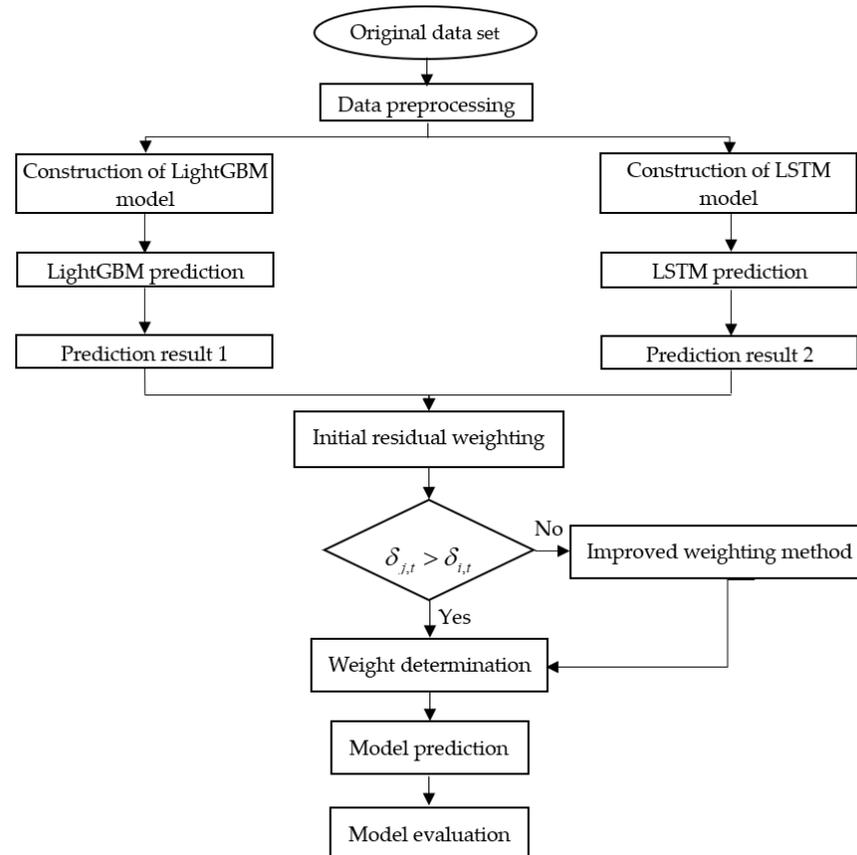

Figure 3. Prediction flow of LSTM-LightGBM variable weight combination model.

3.6. Evaluation Index

The mean absolute percentage error (*MAPE*) is not applicable because the actual value of the data used in this study includes zero. Therefore, the evaluation index used in this study is root mean square error (*RMSE*) and mean absolute error (*MAE*). The formula is as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\widehat{y_i} - y_{pre}\right)^2} \tag{17}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\left(\widehat{y_i} - y_{pre}\right)\right| \tag{18}$$

In the formula, m is the number of samples, $\widehat{y_i}$ is the true value, y_{pre} is the forecast results. The actual value will be closer to the predicted value if the value of the loss function is smaller, and this ensures a higher accuracy of the model prediction.

When there is a certain amount of error in the prediction, the value of the root mean square error will also be larger, so the root mean square error is used to characterize the degree of dispersion of the error value. As the error values of the mean absolute error are

absolutized, there is no situation where the positive and negative errors in the mean error cancel each other out. Thus, the mean absolute error can better reflect the actual situation of the prediction errors.

4. Results

4.1. Prediction Factor Analysis

There are multiple transformations and interactions between the gas mixture and other compounds at different measuring points [32]. Therefore, the correlation between the concentration of the gas mixture and other compounds is analyzed.

In statistics, the Pearson product–moment correlation coefficient (PPMMC) [33] is used to measure the correlation between variables. To avoid experimental uncertainties, data from three different coal mines were selected for correlation analysis, and the visualization of the correlation between the mixed gas concentration and the data was determined using heat diagram.

As shown in Figure 4, the "FC" data in this working face are zero, and a correlation with the mixed gas concentration was absent. There is a strong correlation between "EGas", "Gas1", "Gas2" and the mixed gas. However, by calculating the values of mutual information between "EGas", "Gas1" and "Gas2", we found that "EGas" has the largest mutual information value and is greater than the threshold value, so it can be considered that "Gas1" and "Gas2" are redundant features; thus,"Gas1 "and "Gas2" are not used as input data.
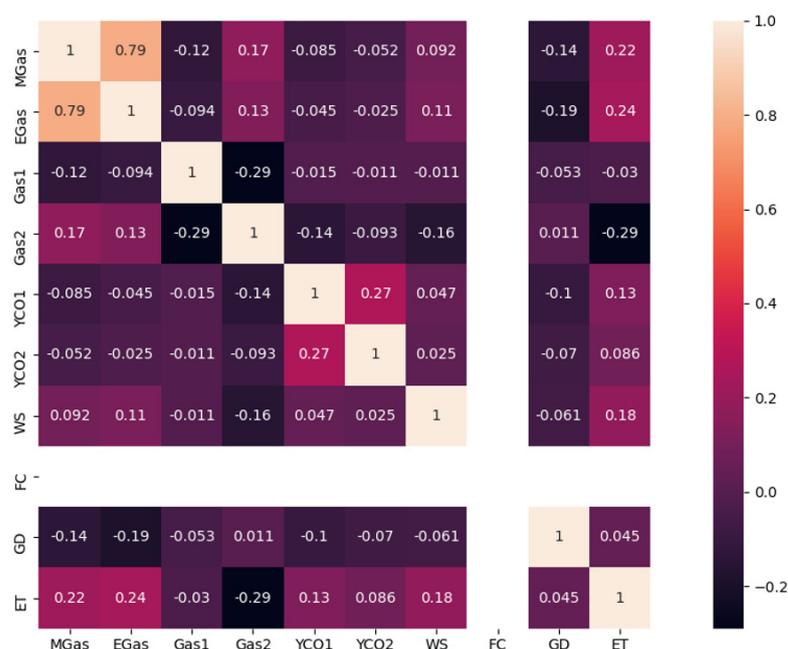

Figure 4. Correlation analysis of data.

The four variables "EGas, WS, ET and GD" were selected as the input of the prediction model, and the correlation analysis between the input variables and the mixed gas concentration is shown in Figure 5. According to previous experiments conducted on methane adsorption, an increase in temperature can reduce the gas adsorption capacity and it can effectively promote the rapid desorption and diffusion. Meanwhile, the activity of the methane molecule increases, which promotes the pore expansion of coal bodies, particularly of the small gaps. This significantly improves the methane diffusion of coal bodies. The diffusion coefficient dynamically changes with an increase in the temperature. In this study, the least squares method was used for fitting, as shown in Figure 5a. A positive correlation

between the concentration of mixed gas and the ambient temperature was observed, which revealed the mechanism of the dynamic process of gas diffusion proposed by Liu [34] et al.
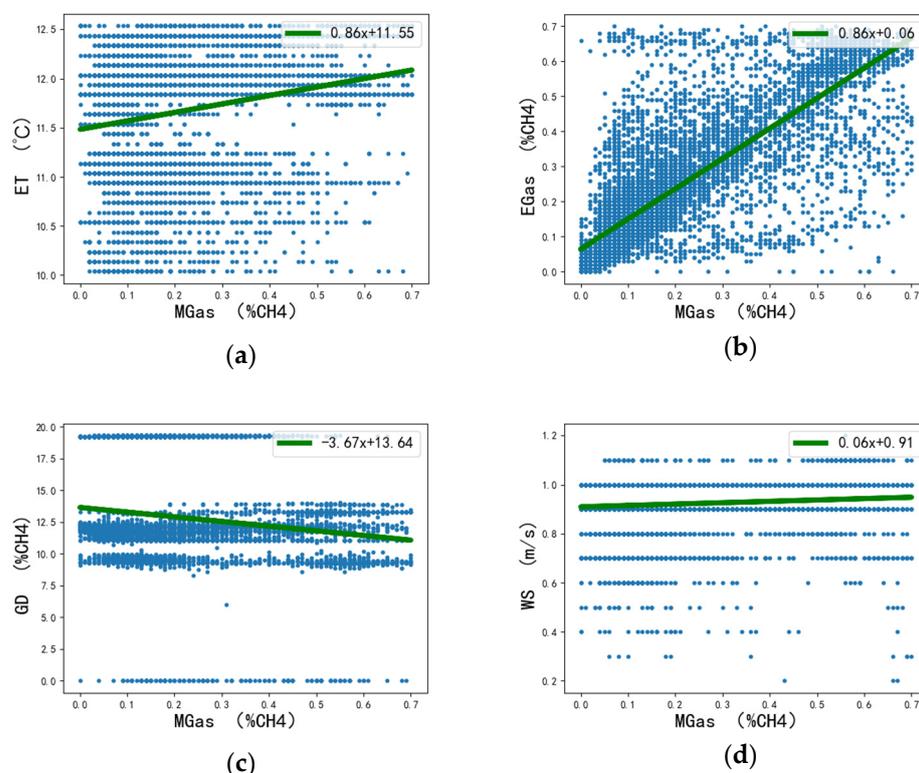

(a)

(b)

(c)

(d)

Figure 5. Correlation analysis between gas concentration and input data: (**a**) Scatter plot of correlation between mixed methane concentration and ambient temperature; (**b**) Scatter plot of correlation between mixed methane concentration and back air methane concentration; (**c**) Scatter plot of correlation between the concentration of mixed methane and the instantaneous flow of pipeline; (**d**) Scatter plot of correlation between mixed methane concentration and working velocity and back air.

In this study, the back air methane concentration and mixed methane concentration exhibited a stronger correlation. The back air pipe is mainly used to receive the air flow after cleaning the working face, and a large volume of gas will be produced during the process of production at the mine working face. At the working face, the main gas sources are the falling coal gas emission and coal wall gas. Different gas sources follow different rules of gas emission [35].

4.1.1. Law of Falling Coal Gas Emission

The coal body will crack during the process of mining, causing a change in gas occurrence conditions. A large volume of gas changes into a free state from the adsorption state, and it might enter into the tunnel with the air flow. The volume of falling coal gas emission is closely related to falling coal, the falling coal fragmentation, the content of coal seam gas and residual gas. The intensity of coal falling gas emission is shown in Formulae (19) and (20).

$$q_1 = \frac{q_{10}}{(1+t)^{\alpha}} \tag{19}$$

$$Q_1 = \int_0^T q_1 \theta M dt \tag{20}$$

In the function, q_1 represents the emission intensity per weight of falling coal gas at unit time of $t + 1$, unit is $m^3/(min.t)$. q_{10} represents the intensity of gas emission at initial moment of falling coals with the unit of $m^3/(min.t)$. t represents the exposure time of

falling coals with the unit of min. α is the attenuation coefficient, Q_1 is the absolute gas emission from falling coals in the process of mining with the unit of $m^3/$min. M represents the mining weight per unit time with the unit of $t/$min. θ is the degree of fragmentation.

4.1.2. Law of Coal Gas Emission of in Working Face

The gas released from the coal enters the air stream through the surface of the coal wall according to Duthie's law and the law of diffusion. During the process of continuous mining, fresh coal wall is constantly exposed, mining pressure constantly changes, and the gas pressure balance state near the working face changes. A large volume of gas flow out along the coal cracks and pores gushing lane, the gushing intensity of the coal wall gas is shown in Formulas (21) and (22).

$$q_2 = \frac{q_{20}}{(1+t)^\beta} \tag{21}$$

$$Q_2 = \int_0^T q_2 Hvdt \tag{22}$$

In this function, q_2 represents gas emission intensity of back coal wall at the time of $t + 1$ with the unit of $m^3/(\text{min}.m^2)$. The q_{20} is gas emission intensity at the initial moment of coal wall with the unit of $m^3/(\text{min}.m^2)$. t is the exposure time of coal wall, with the unit of min. β is the attenuation coefficient, Q_2 is absolute emissions of coal wall gas in the process of mining with the unit of $m^3/$min. H is the thickness of coal mining layer with the unit of m. v is the cutting speed of coal mining machine with the unit of $m/$min.

After entering the lane from the above gas source, methane will form a mixture of gas and air with uneven concentration, and the mixture will migrate by concentration diffusion and convection mixing in the airflow. After fresh air flow passes through the working face of mines, partial methane gas in the mining face is diluted and carried. Therefore, the methane concentration in the back air can accurately reflect the change in the methane concentration in the mining face.

4.2. Model Prediction Analysis and Comparison

To verify the accuracy of the improved LSTM-LightGBM, the LSTM, LightGBM, XG-Boost, LSTM-LightGBM (Equivalent weighting) and LSTM-LightGBM (Residual weight) were selected for comparative experiments. The errors of the different models were compared as shown in Figure 6.

From the figure above, it can be observed that the prediction accuracy of the variable weight combination model is higher than that of the single machine learning model and the conventional combination weighting model. The comparison between the values of *MAE* and *RMSE* of the models is shown in Table 2.

Table 2. Comparison between evaluation indexes of each model.

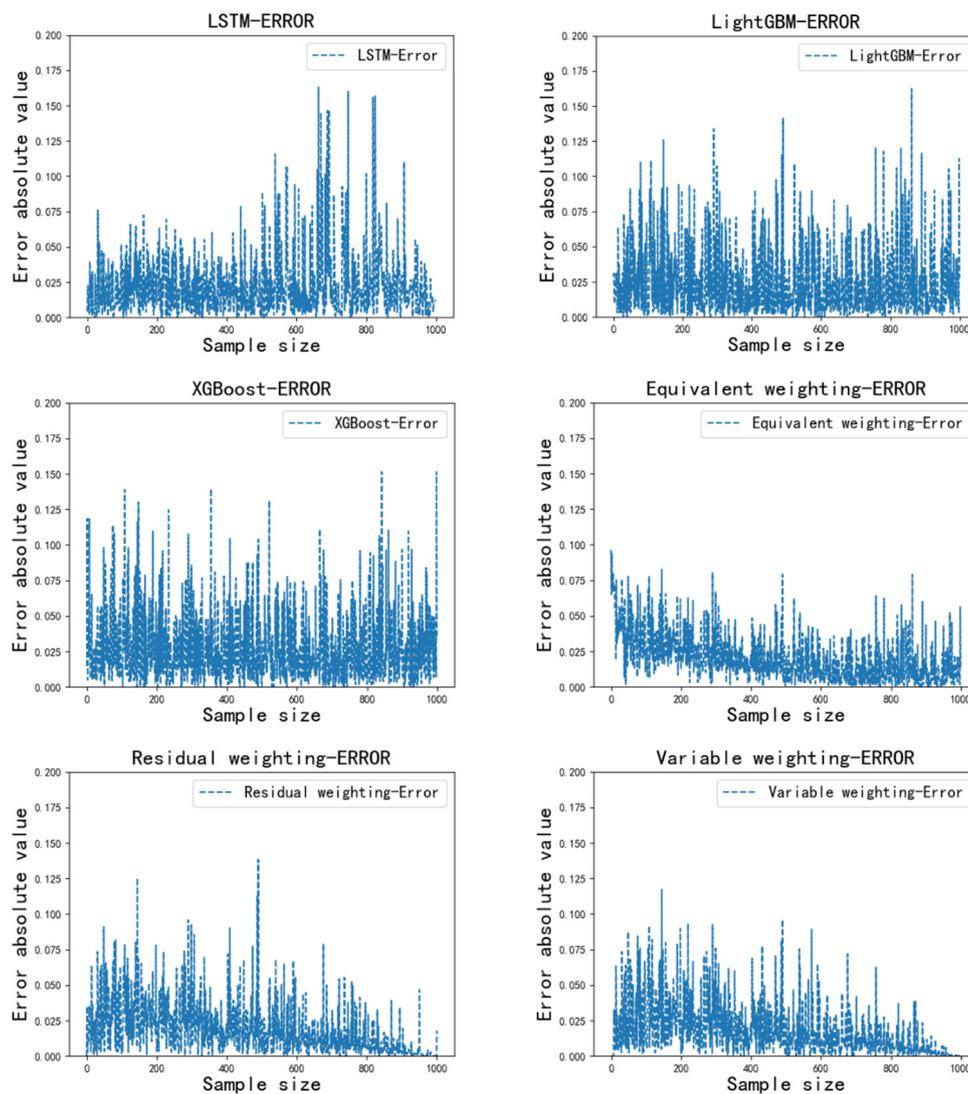| Model	*MAE*	*RMSE*
LSTM	0.0219	0.0306
LightGBM	0.0277	0.0377
XGBoost	0.0253	0.0352
LSTM-LightGBM (Equivalent weighting)	0.0214	0.0276
LSTM-LightGBM (Residual weighting)	0.0201	0.0279
LSTM-LightGBM (Variable weight combination)	0.0194	0.0261

Figure 6. Prediction result and real results of each model.

The *MAE* and *RMSE* values of LSTM model were the average value of the LSTM model which were trained ten times. After the analysis, the *MAE* and *RMSE* values of the improved LSTM-LightGBM variable weight combination model were increased by 3.5% and 6.5%, respectively, compared with that of the LSTM-LightGBM residual weight combination model, and by 11.4% and 14.7%, respectively compared with that of the LSTM-LightGBM single machine learning model. The improved variable weight combination method has a higher prediction accuracy.

4.3. Model Universality Analysis

During the selection of study area, strong local features were observed at different working faces of the coal mine at different locations. To verify the universality of the algorithm, the prediction and analysis of gas concentration were performed in different coal mines. The coal mines selected were Mine A in Shanxi, Mine B in Guizhou, and Mine C in Anhui.

It can be observed from Figure 7 that the prediction error of the modified variable weight combination model is smaller than that of the conventional model, and the increase in Mine A is the most obvious. *MAE* value increased by 18.5% and 29.2%, respectively, compared with that of the LSTM model and the LightGBM model. *RMSE* increased by 22.9% and 30.4%, respectively, compared with the LSTM model and the LightGBM

model. Therefore, the prediction results of the improved variable weight combination model with three different coal mine gas concentrations demonstrated that the prediction accuracy was improved. This demonstrates the universality of the improved variable weight combination model.
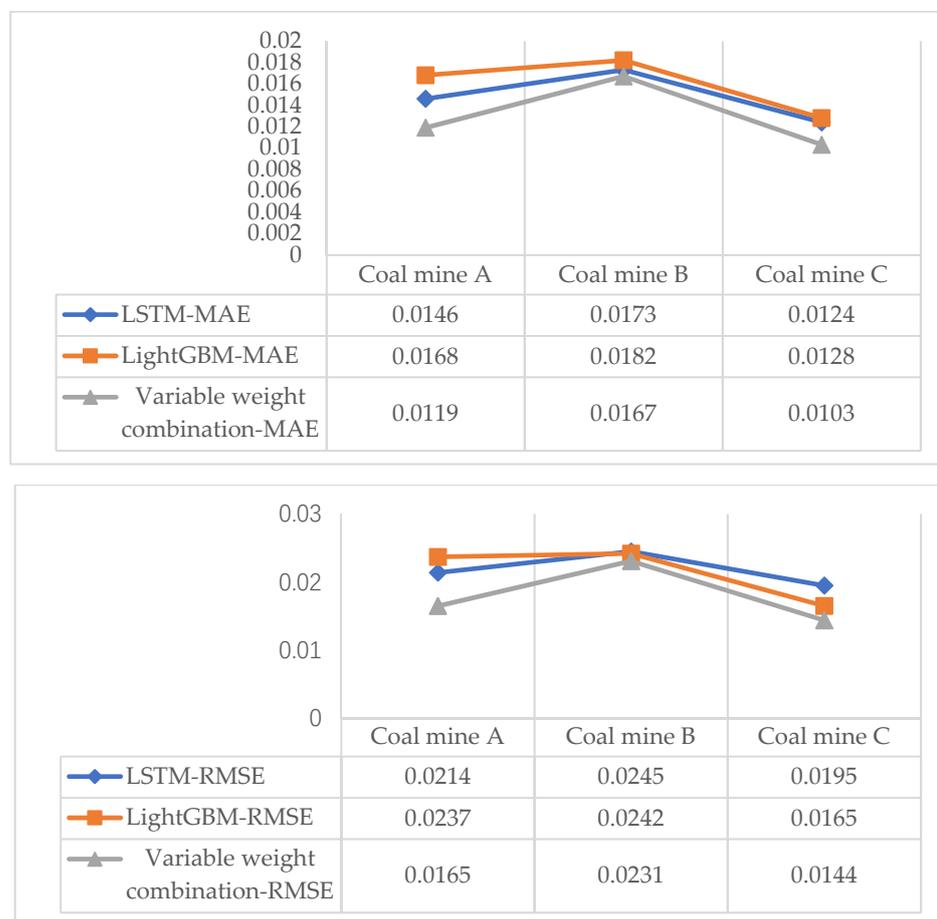
	Coal mine A	Coal mine B	Coal mine C
LSTM-MAE	0.0146	0.0173	0.0124
LightGBM-MAE	0.0168	0.0182	0.0128
Variable weight combination-MAE	0.0119	0.0167	0.0103

	Coal mine A	Coal mine B	Coal mine C
LSTM-RMSE	0.0214	0.0245	0.0195
LightGBM-RMSE	0.0237	0.0242	0.0165
Variable weight combination-RMSE	0.0165	0.0231	0.0144

Figure 7. Analysis of evaluation index applied to different coal mines.

5. Discussion

In this study, a variable weight combination model was developed by adopting the methane concentration, wind speed, ambient temperature, gas drainage, and the historical data of mixed gas. Working faces of different mines were selected to predict the gas concentration in the future 10 h. In the improved LSTM-LightGBM variable weight combination model, the *MAE* value and *RMSE* value were 0.0194 and 0.0261, respectively. These values were smaller than that of the prediction values of 0.0224 and 0.0317 obtained in the ARIMA model proposed by Zhang et al. [36] and the 0.0207 and 0.0303 of S-GRU model proposed by Chang et al. [37]. This was because an LSTM neural network with better time sequence prediction and the LightGBM model with better performance in the nonlinear model were predicted in the form of variable weight combination. It considered the time sequence feature of the data and the nonlinear feature of data. For analysis and comparison result of the gas concentration, the improved LSTM-LightGBM variable weight combination model was better than that of the conventional LSTM-LightGBM equivalent weight assignment model and LSTM-LightGBM residual weight assignment model. Considering the difference in prediction error between the LSTM network and the LightGBM model at different moments, the combination model adopted different weights for the prediction values at different moments to combine the advantages of both the models.

In this study, data from coal mine at several locations were selected to explore the performance of the regional model. Additionally, downhole temperature, wind speed and methane gas were selected as prediction factors to determine the effect of factors for gas concentration prediction [38]. To improve the prediction accuracy of gas concentration, suitable factors such as weather and ground surface temperature, depth of coal seam, inclination of coal bed, top and bottom lithology of coal bed should be considered in the future.

6. Conclusions

Based on LSTM and LightGBM model with the variable weight combination model, the prediction method of gas concentration was improved. In this model, the time sequence feature and the nonlinear relationship between the input feature and gas concentration were considered. By the data pre-processing and feature selection, it makes the model converge faster and avoids the degradation of prediction accuracy due to redundant features. Sigmoid function is selected for the activation function of the gate structure of the LSTM model. Tanh activation function is selected to generate candidate memories. These gates increase the convergence speed of the model. Moreover, they guarantee that the model does not suffer from the problem of exploding output values and vanishing gradients. In comparison to traditional single machine learning gas concentration prediction models, LSTM models have a higher prediction accuracy.

Compared with that of single machine learning model and other conventional combination weighting models, the prediction result of the variable weight combination model was closer to that of the real value with a small error. It provides better prediction accuracy, and high reliability. It can give a reference for gas accidents prevention and promote the safety of coal mines.

This study focused on the prediction and analysis of gas concentration using the underground attribute information only including temperature, wind speed, methane gas. Nevertheless, the change of gas concentration is affected by complex factors and conditions [39]. In future research, it is important for us to consider more comprehensive factors of gas concentration, such as roof pressure, minging depth, inclination angle of coal seam and ground weather information.

Author Contributions: Conceptualization, X.W. and N.X.; methodology, N.X.; software, N.X.; validation, X.M. and X.W.; formal analysis, H.C.; investigation, X.W.; resources, X.W.; data curation, N.X.; writing—original draft preparation, N.X.; writing—review and editing, X.W.; visualization, H.C.; supervision, X.W.; project administration, X.M.; funding acquisition, X.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository. The partial data presented in this study are openly available in [Gas concentration prediction data set, Mendeley Data], doi:10.17632/p3n7k6hxgw.1.

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

References

1. Teng, J.; Qiao, Y. Analysis of coal demand, exploration potential and efficient utilization in China. *Chin. J. Geophys.* **2016**, *59*, 4633–4653.
2. Cheng, J.; Bai, J.Y. Short-term prediction of mine gas concentration based on chaotic time series. *J. China Univ. Mine. Technol.* **2008**, *02*, 231–235.
3. Deng, G. Current status and prospects of coal and gas outburst prediction and prevention technology. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *651*, 32096. [CrossRef]
4. Fu, H.; Liu, Y.Z. Prediction of gas concentration based on multi-sensor-deep long and short time memory network fusion. *J. Sens. Technol.* **2021**, *34*, 784–790.
5. Lai, X.W.; Xia, Y.N. Improved grey gas concentration series prediction based on ensemble learning. *China Work. Saf. Sci. Technol.* **2021**, *17*, 16–21.
6. Wang, Y.S. Mathematical models in the study of gas gush prediction in mines. *Coal Technol.* **2015**, *263*, 185–187.
7. Zhang, Z.X.; Yuan, C.F. Research on prediction of gas gush in mines by gas geological mathematical model method. *J. China Coal Soc.* **1999**, *4*, 34–38.
8. Lu, X.L.; Fu, X.M. Study on the geological pattern of gas in Kongzhuang coal mine. *Coal Sci. Technol.* **1997**, *2*, 73–76.
9. Song, S.; Li, S.G. Research on a multi-parameter fusion prediction model of pressure relief gas concentration based on RNN. *Energies* **2021**, *14*, 1384. [CrossRef]
10. Zhang, D.Y.; Gong, Y. The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure. *IEEE Access* **2020**, *8*, 220990–221003. [CrossRef]
11. Wen, T.X.; Zhang, B. A random forest model for coal and gas protrusion prediction. *Comput. Eng. Appl.* **2014**, *50*, 233–237.
12. Yin, G.Z.; Li, M.H. An improved BP neural network-based model for predicting gas permeability in coal bodies. *J. Coal* **2013**, *38*, 1179–1184.
13. Zhang, T.J.; Song, S. Research on gas concentration prediction models based on LSTM multidimensional time series. *Energies* **2019**, *12*, 161. [CrossRef]
14. Li, W.S.; Wang, L. Application and design of LSTM in coal mine gas prediction and warning system. *J. Xi'an Univ. Sci. Technol.* **2018**, *38*, 1027–1035.
15. Sun, Q.M.; Qu, Z.J. Situation-aware multimodal transport recommendation based on particle swarm optimization and LightGBM. *J. Electron.* **2021**, *49*, 894–903.
16. Xun, X.X.; Su, C. CNN-LSTM based coal mine gas concentration prediction. *Mod. Inf. Technol.* **2020**, *4*, 149–150.
17. Lin, H.F.; Gao, F. PSO-BP neural network prediction model for coal seam gas content and its application. *Chin. J. Saf. Sci.* **2020**, *30*, 80–87.
18. Wen, J.Q.; Zhang, Y. Prediction of gas content based on gray theory-BP neural network. *Energy Technol. Manag.* **2020**, *45*, 44–45, 55.
19. Xu, Y.S.; Qi, C.Y. Prediction model of gas gushing based on IGSA-BP network. *J. Electron. Meas. Instrum.* **2019**, *33*, 111–117.
20. Zhang, X.J.; Liu, F. Gas concentration prediction in coal mines based on wavelet noise reduction and recurrent neural networks. *Coal Technol.* **2020**, *321*, 145–148.
21. Han, T.T.; Wu, S.Y. Gas concentration prediction based on Markov residual correction. *Ind. Min. Autom.* **2014**, *216*, 28–31.
22. Kang, J.F.; Tan, J.L. Short-term PM2.5 concentration prediction with the support of XGBoost-LSTM variable weight combination model—Shanghai as an example. *China Environ. Sci.* **2021**, *7*, 1–16.
23. Wang, Y.H.; Wang, S.Y. Research on multi-parameter gas concentration prediction model based on improved locust algorithm optimized long-short-term memory neural network. *J. Sens. Technol.* **2021**, *34*, 1196–1203.
24. Li, D.; Sun, Z.M. Research on AWLSSVM gas prediction based on chaos particle swarm. *Saf. Coal Mines* **2020**, *51*, 193–198.
25. Qiu, Y.G.; Zhou, J. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. *Eng. Comput.* **2021**, *21*, 1393–1399. [CrossRef]
26. Zhang, X. Ion channel prediction using Lightgbm Model. In Proceedings of the 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Fuzhou, China, 10–12 April 2020.
27. Shu, X.; Zhang, L. Host–parasite: Graph LSTM-in-LSTM for group activity recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 663–674. [CrossRef] [PubMed]
28. Zhang, H.; Zhang, Q. The development and properties of activation function are reviewed. *J. Xihua Univ.* **2021**, *1*, 10.
29. Hossain, M.Z.; Sohel, F.; Shiratuddin, M.F.; Laga, H. A comprehensive survey of deep learning for image captioning. *ACM Comput. Surv.* **2019**, *51*, 118. [CrossRef]
30. Cheng, T.J.; Wang, M. Trend prediction of online public opinion on unexpected events based on variable weight combination. *Comput. Sci.* **2021**, *48*, 190–195.
31. Wang, H.Q.; Liang, W. Scene mover: Automatic move planing for scene arrangement by deep reinforcement learing. *ACM Trans. Graph.* **2020**, *39*, 233. [CrossRef]
32. Peng, S.P.; Gao, Y.F. Theoretical discussion and preliminary practice of AVO detection of gas enrichment in coal seams—A case study of Huainan coalfield. *J. Geophys.* **2005**, *6*, 262–273.
33. Kong, L.; Nian, H. Fault detection and location method for mesh-type DC microgrid using pearson correlation coefficient. *IEEE Trans. Power Deliv.* **2021**, *36*, 1428–1439. [CrossRef]

34. Liu, Y.W.; Wei, J.P. The law and mechanism of temperature influence on the dynamic process of coal particle gas diffusion. *J. Coal* **2013**, *38*, 100–105.

35. Lyu, P.Y.; Chen, N. LSTM based encoder-decoder for short-term predictions of gas concentration using multi-sensor fusion. *Process Saf. Environ. Prot.* **2020**, *137*, 93–105. [CrossRef]

36. Zhang, Z.; Zhu, Q.J. Construction of ARIMA prediction model for gas concentration based on Python and its application. *J. North China Inst. Sci. Technol.* **2020**, *17*, 23–28.

37. Chang, L.; Zhang, H. An improved GRU gas concentration prediction model. *J. Heilongjiang Univ. Sci. Technol.* **2020**, *30*, 532–535.

38. Yu, G.F.; Fei, W. A compromise-typed variable weight decision method for hybrid multiattribute decision making. *IEEE Trans. Fuzzy Syst.* **2019**, *27*, 861–872. [CrossRef]

39. Wang, L.L.; Cao, Q.G. Research on the influencing factors in coal mine production safety based on the combination of DEMATEL and ISM. *Saf. Sci.* **2018**, *103*, 51–61. [CrossRef]