Process Safety and Environmental Protection 137 (2020) 93–105

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

journal homepage: www.elsevier.com/locate/psep

LSTM based encoder-decoder for short-term predictions of gas concentration using multi-sensor fusion

Pingyang Lyu, Ning Chen, Shanjun Mao*, Mei Li

Institute of Remote Sensing and Geographic Information System, Peking University, Beijing 100871, China

ARTICLE INFO

Article history:
Received 5 October 2019
Received in revised form 8 January 2020
Accepted 15 February 2020
Available online 15 February 2020

Keywords:
LSTM
Encoder-Decoder
Gas concentration
Multi-Sensor information
Multi-Step prediction

ABSTRACT

Gas is one of the most dangerous byproducts of coal in mines. Before gas accidents occur, an abnormally increased gas concentration can be observed. Therefore, a prediction of the gas concentration in coal mines is of great significance to prevent the gas accident and ensure the production safety in the mines. By calculating the Pearson correlation coefficient for the gas concentration of different sensors, the spatial correlation of the gas concentration that is monitored for each mining face is verified. We present multi-step prediction results for gas concentration time series based on the ARMA model, the CHAOS model and the Encoder-Decoder model (single-sensor and multi-sensor) and compare these results. The Encoder-Decoder model provides high robustness in a multi-step prediction and can predict the gas concentration for five different time steps. Its prediction error is significantly lower than those of the ARMA and the CHAOS models. The prediction accuracy is further improved through a fusion with information of other sensors. In this way, this study provides a novel concept and method for gas accident prevention.

1. Introduction

Methane is not only the gas associated with coal, but also an important source for so-called clean energy. Due to its flammable and explosive nature, it has a critical impact on the production safety in coal mines (Karacan et al., 2011). An accurate determination of the gas concentration forms the basis for gas outburst prediction, gas explosion prevention and ventilation design (Wu et al., 2014; Dougherty and Özgen Karacan, 2011; Karacan, 2007). With the exploitation of coal resources, the mining depth of mines increases year by year. Subsequently, the gas content of coal seams increases gradually and gas control work becomes more complex (Xia et al., 2016). Gas accidents in coal mines include coal and gas outbursts, gas explosions as well as asphyxiation or poisoning of miners. An abnormal increase of the gas concentration usually occurs before such accidents (Song et al., 2019; Li et al., 2018). In order to accurately judge the gas concentration and to effectively prevent gas accidents, researchers have conducted several studies regarding emission laws, numerical simulations, machine learning approaches and obtained numerous valuable results already.

The mining face is the main production area of the mine. During the mining process in gas-rich coal mines, a large amount of gas is released from the coal into the airflow. In order to quantify and predict this gas emission from the mining face, researchers (Liang et al., 2016; Ye et al., 2006; Yang et al., 2020; Guo et al., 2018) divide gas sources into coal wall, falling coal, goaf and adjacent strata by using the idea of the *Separate Source Prediction Method* (State Administration of Work Safety, 2006). By studying the law of gas emission, the constitutive equation for the gas emission from different gas sources is established. It provides meaningful insight and aids the ventilation and gas extraction design (Noack, 1998). However, a large number of parameters of this method, such as the amount of coal falling per unit time and the amount of coal left in the goaf, are difficult to accurately quantify in practice(Zheng et al., 2019), which makes it impossible to apply these methods in the real-time prediction of the gas concentration.

Numerical simulations are important tools used to study gas in the mining industry. The transport of methane within coal satisfies Darcy's law and the diffusion law. When methane pours into the roadway, it - together with the air - forms a mixed gas and moves with the wind (Yu et al., 2000). In order to determine the law of the gas migration in the mining face, researchers have studied the gas flow equation and the gas diffusion and migration equation for the mining face. Through establishing the model of the mining face, a large number of numerical simulations have been carried out (Cao and Li, 2017; Xia et al., 2017). However, the mesh division of the numerical simulation has a significant effect on this experiment. The simulation of the gas migration in the mining face is based on

* Corresponding author.
E-mail address: sjmao_pku@163.com (S. Mao).

https://doi.org/10.1016/j.psep.2020.02.021
0957-5820/

a simplified model with ideal boundary conditions. For example, the movement of electromechanical equipment, the movement of personnel, the degree of air leakage and the internal state of the goaf are difficult to retrieve from the real scene (Wu et al. 2018). A numerical simulation demonstrates the laws of gas migration, but it is difficult to achieve a real-time prediction of the gas concentration in this way.

Within a coal mine monitoring system several sensors (methane, carbon dioxide, temperature, humidity, air flow, etc.) are used. The collected information is used to visualize original data, and to perform in-depth data analysis that can lead to significant benefits for mining (Dominik et al., 2018; Adam and Alicja, 2018). In order to realize a real-time prediction of the gas concentration, researchers have focused their studies on the prediction of the gas concentration using sensor data, and presented several methods, including chaotic time series (Gao and Yu, 2006; Li et al., 2008; Cheng et al., 2008; Zhang et al., 2007; Wang et al., 2017), ARMA model (Ma, 2018; Zhou and Yao, 2009), neural networks (Karacan, 2008), support vector machines (Yin, 2010) and LS-SVM (Qiao et al. 2011). Although, these studies achieved good results, there is still a need for improvement regarding the speed and accuracy of the algorithms. For example, ARMA as well as its improved model are both suitable for short-term and medium-term linear predictions, while support vector machines are suitable for small data samples only. Due the multi-source characteristics of gas in the working face and the migration characteristics of gas mixtures, the gas concentration is both regular and complex. This constitutes a typical non-linear prediction problem. For this reason, various combination algorithms based on neural networks, such as wavelet-ELM, EMD-MFOA-ELM (Lu et al., 2017), CAPSO-ENN (Fu et al., 2015) and IGA-DFNN (Fu et al., 2014), have been studied in the past. Neural networks and the various combination algorithms are simple, easy to use and highly applicable. They provide various possibilities for gas prediction and are mainly used as prediction algorithms at present. However, these algorithms usually rely on original data of a single gas sensor and lack the option of fusion with other sensors' data. Moreover, a manual feature extraction step to train the model is often required. There is a spatial dependence among the sensors in different areas of a coal mine that partially contain each other's information. Rational use of this partial data can support the prediction of the gas concentration (Wu et al., 2014; Zagorecki, 2015). Therefore, when building an appropriate artificial neural network model, fusing the relevant sensor information can accelerate the convergence of the model and can improve the effect of the model (Shen, 2019).

In recent years, deep learning has been widely promoted. Compared with traditional machine learning algorithms, deep learning algorithms can automatically extract features from raw data using their multi-layer networks and can thus reduce the manual work of feature engineering. They further reduce the number of shallow network parameters, and have a better fitting ability under the same order of parameters. As a typical deep learning model, the Long Short-Term Memory (LSTM) network (Hochreiter and Schmidhuber, 1997) can memorize long-term historical information and has thus been widely used in speech recognition, emotion analysis, power forecasting and other sequential learning tasks. The LSTM based Encoder-Decoder model has also been applied in machine translation (Cho et al., 2014), vehicle trajectory prediction (Park et al., 2018), air pollution prediction (Bui et al., 2018) and multi-sensor timing anomaly detection (Malhotra et al., 2016). Encoding means to compress information of the whole sequence into a fixed length vector, while decoding means to convert the fixed vector generated before into an output sequence.

In order to make full use of the characteristics of this deep learning model to realize the prediction of mine gas and to solve the shortcomings of traditional prediction models such as their experience dependence, the poor timeliness and their limitation to shallow data characteristics, the Encoder-Decoder model is used to extract the temporal and spatial characteristics of gas concentration by analyzing the historical data of gas concentration, to improve the accuracy of prediction. In the Encoder stage, the spatial and temporal characteristics of the gas concentration of multiple sensors in the working face are extracted, and the intermediate state vector C containing the characteristic distribution of historical data is generated by non-linear transformation as the input of the decoder. In the decoder stage, the multi-step prediction of the gas concentration value is carried out using the characteristics of historical data; the prediction results are then evaluated.

2. Gas sources in the mining face

The working face of gas-rich coal mines causes large amounts of gas emission during the production process. The sources of gas include mainly coal falling, coal wall, goaf and adjacent strata. Different gas sources follow different gas emission laws (Ye et al., 2006).

(1)Law of Gas Emission from Coal Falling

In the process of mining, coal is continuously mined from the coal seam and transported to the surface using the transport roadway. The coal body breaks up during the mining process, which results in changes of the conditions of gas occurrences. Therefore, a large number of gas may change from the adsorption state to a free state and may spread along with the air flow into the roadway. Gas emission from falling coal is closely related to the amount of falling coal, the fragmentation degree of the falling coal, the content of gas in the coal seam and the content of residual gas in the coal. The intensity of gas emission from falling coal is expressed as shown in Eq. (1) (Yang et al., 2009).

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

According to Eq. (1), the absolute gas emission from falling coal can be expressed as Eq. (2).

$$Q_1 = \int_0^T q_1 \theta M dt \tag{2}$$

In this equation, q_1 represents the gas emission intensity per unit mass of falling coal after $t+1$ units of time in $m^3/(\min \cdot t)$; q_{10} represents the gas emission intensity at the initial time of coal falling in $m^3/(\min \cdot t)$; t represents the exposure time of coal falling in min; α is the attenuation coefficient; Q_1 is the absolute gas emission from coal falling in the mining process in $m^3/\min$; M is the coal cutting quality per unit time in t/min; θ is the fragmentation degree of falling coal.

(2)Gas Emission Law of Coal Wall in Working Face (State Administration of Work Safety, 2006)

The gas in the coal enters the airflow through the surface of the coal wall, which conforms to Darcy's law and the diffusion law. After a variety of long-term gas extraction measures, the gas gushing from the coal wall surface is gradually decreasing and converging to a stable condition. However, in the process of continuous mining, fresh coal wall is constantly exposed, the mining pressure is constantly changing, the state of the gas pressure balance near the working face is changing, and a large amount of gas is gushing out into the roadway space along the cracks and pores of the coal body. The intensity of the gas emission from the coal wall is shown in Eq. (3) (Yang et al., 2020).

$$q_2 = \frac{q_{20}}{(1+t)^{\beta}} \tag{3}$$

According to Eq. (3), the absolute gas emission from the coal wall of the working face can be obtained as shown in Eq. (4).

$$Q_2 = \int_0^T q_2 H v dt \tag{4}$$

In this equation, q_2 represents the intensity of gas emission from coal wall after time $t+1$ in $m^3/(\min \cdot m^2)$; q_{20} represents the intensity of gas emission at the initial time of coal wall in $m^3/(\min \cdot m^2)$; t represents the exposure time of coal wall in min; β is the attenuation coefficient; Q_2 is the absolute gas emission from coal wall in the mining process in $m^3/\min$; H is the thickness of the mining seam in m; v is the cutting speed of the shearer in m/min.

(3)Gas Emission Law in Goaf

After mining, some coal blocks will be left in the goaf. In the process of mining, the residual coal in the goaf is constantly desorbing and releasing gas. The characteristics of the residual coal gas emission in the goaf are the same as that of the falling coal gas emission, which means that the intensity of the gas emission in the goaf is the same as that of the gas emission of the falling coal in Eq. (1). Therefore, the amount of gas emission from the goaf is shown in Eq. (5).

$$Q_3 = \int_0^T q_3 \theta' M' dt \tag{5}$$

In this equation, Q_3 is the absolute gas emission from the goaf in $m^3/\min$; θ' is the fragmentation degree of the coal from the goaf in % ; M' is the quality of the coal from the goaf in t.

(4)Gas Emission Law in Adjacent Coal Seam

The gas emission from adjacent coal seams is consistent with the advancing trend of the mining face, and the gas emission is positively correlated with the thickness of adjacent layers. The expression of gas emission from adjacent layers is shown in Eq. (6) (Guo et al., 2018).

$$Q_4 = \frac{Q_2 \sum h_i \eta_i}{H} \tag{6}$$

(5)Gas Emission from Mining Face

The gas emission from the mining face is mainly composed of the four parts mentioned above. Therefore, the formula for calculating the relative gas emission is shown in Eq. 7.

$$Q = Q_1 + Q_2 + Q_3 + Q_4 \tag{7}$$

After methane pours into the roadway from the above-mentioned gas sources, a mixture of gas and air with an uneven distribution concentration is formed, which is migrated mainly through concentration diffusion and convective mixing in the airflow (Wang and Fasong, 1996). In addition, due to the real-time transportation of the coal falling to the ground using the transportation system, gas diffuses to the roadway during this transportation process.

3. Temporal and spatial feature extraction of mine gas concentration sequence

The short-term gas concentration at a gas monitoring point has a certain relationship with previous gas concentrations at this measuring point and with the gas concentration of other measuring points. Within the scope of this paper, we refer to the former as the time series factor, and the latter as the spatial topological factor. Before we enter the data into the prediction model, a simple feature extraction from the original data is helpful to ensure a fast convergence of the model. The extraction of the temporal and spatial topological features is introduced in the following sections.

3.1. Temporal feature

Gas concentration data is a typical example for time series data. The short-term gas concentration value of a measuring point in the future is correlated with historical data. Extracting the time series characteristics of gas concentration data can accelerate the convergence of the model and improve the prediction accuracy of the model (Jason, 2016). This paper is aimed at time series data of gas concentrations and mainly extracts the time series characteristics including:

(1) Concentration values and first-order difference values at current time points: : x_t, z_t ;($z_t = x_t - x_{t-1}$, the same below);

(2) Concentration values and first-order difference values for the preceding period of l_0: $x_{t-1}, x_{t-2}, \ldots x_{t-l_0} z_{t-1}, z_{t-2}, \ldots z_{t-l_0}$;

(3) Sliding statistics based on concentration values and first-order difference values. These statistics include the statistical mean, the maximum, the minimum and standard deviation. The window size is l_1

(4) The residual characteristics (δ_x, δ_z) based on the concentration values and the first-order difference. The window size is l_2.

$$\delta_x = x_t - \frac{1}{l_2}\sum\nolimits_{i=1}^{l_2} x_{t-i},\ \delta_z = z_t - \frac{1}{l_2}\sum\nolimits_{i=1}^{l_2} z_{t-i} \tag{8}$$

In this paper, the gas concentration sequence data of each sensor is read and calculated according to the characteristics, forming the feature vector X_t, which is the input of the encoder.

3.2. Spatial topological characteristics

In the process of coal production, gas pours into the roadway from the falling coal, the coal wall, the goaf and other gas sources. With the diffusion due to the transportation system and the air flow, and discharges from the return air roadway to the atmosphere, there is a correlation between the values of each monitoring point (Wu, 2015).

The monitoring data of gas concentration in several coal mines are selected, as shown in Fig. 1.

As shown in Fig. 1, the peak gas concentration at each monitoring point in different mining faces follows a delay in the time dimension. In this paper, the co-correlation coefficient is used to describe this delay effect. The direction of the gas sensor topology is added according to the delay effect. The correlation of the values of each gas sensor can be calculated using the Pearson correlation coefficient, as shown in Eq. (9).

$$COR(X, Y) = \frac{\sum_1^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_1^n (X_i - \overline{X})^2 \sum_1^n (Y_i - \overline{Y})^2}} \tag{9}$$

In this equation, X_i represents the gas concentration at point X at time i, Y_i represents the gas concentration at point Y at time i, $\overline{X}$ represents the mean gas concentration at point X, $\overline{Y}$ represents the mean gas concentration at point Y.

According to Eq. 11 and taking Fig. 1 (a) as an example, the correlation coefficients of gas monitoring points with a topological adjacency relationship in the Sijiazhuang Coal Mine are calculated. Fig. 2 shows the layout of the No.15,117 working face in the Sijiazhuang Coal Mine. Fig. 3 shows the correlation coefficient curve between monitoring points with a topological adjacency relationship of the No.15,117 mining face.

Fig. 3 shows that the gas concentration sequences of the monitoring points 009A13 and 044A03 lag behind that of the monitoring point 044A02 and that of the monitoring point 044A11 lags behind that of the monitoring point 044A03. This means, the sequence 009A13 and 044A03 is dependent on 044A02while 044A11 is

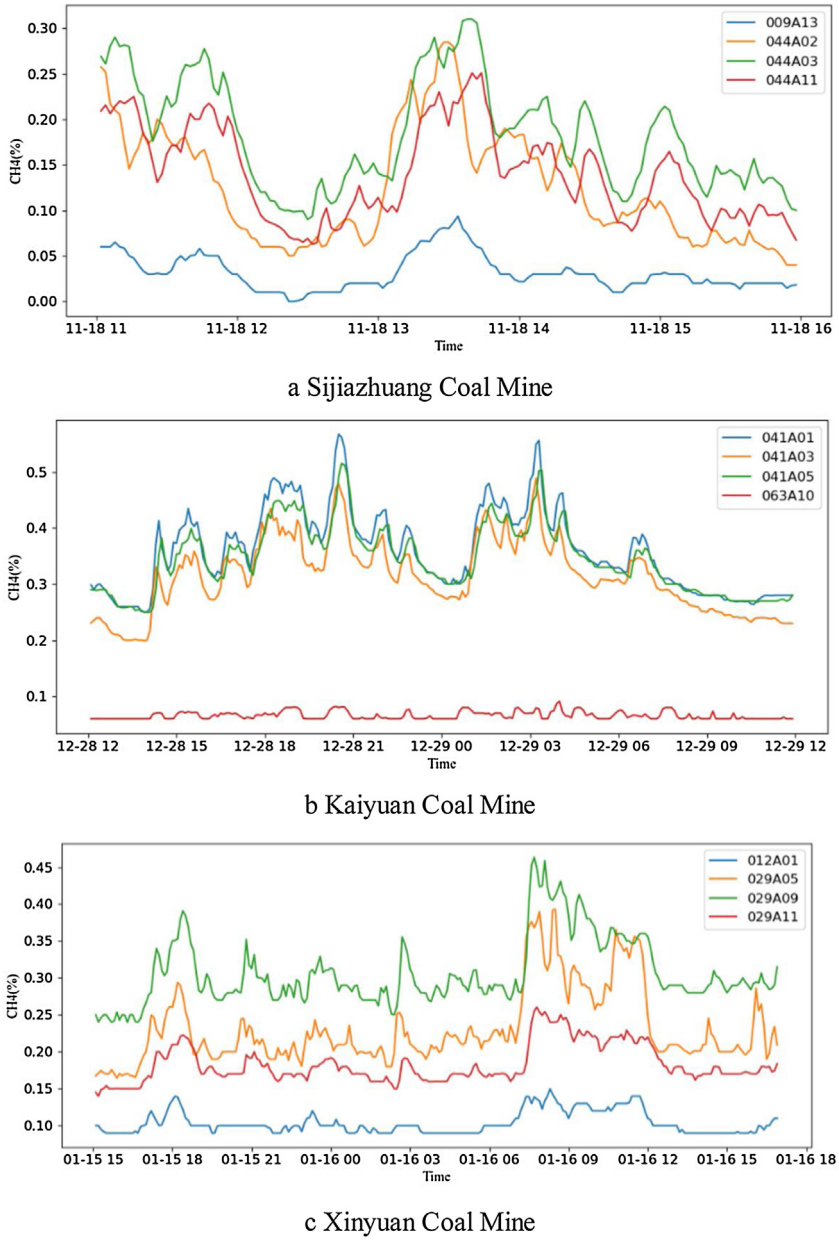


Fig. 1. Gas Concentration in Different Coal Mining Faces.

dependent on 044A03. Therefore, it is necessary to add the direction of each monitoring point in an undirected graph. According to the above correlation coefficient principle, the directed graph of gas monitoring points can be obtained as shown in Fig. 3 (d).

Influenced by the air flow, the information of the gas concentration on the inlet side will be included in the information of the gas concentration in the return air lane, forming the wind direction map of gas monitoring points, as shown in Fig. 4 (a). According to the wind direction and the delay effect, the two directed graphs of Fig. 3 (d) and Fig. 4 (a) are superimposed to form the final directed graph of the gas concentration. As shown in Fig. 4 (c), the directed graph reflects the topological relationship between different gas measuring points in the mine.

According to Fig. 4 (c), when predicting the gas concentration at a certain measuring point, other data that are related to the measuring point are extracted and entered into the model as spatial topological features. The model avoids taking all measured data as input, which reduces the training cost and the complexity of the model, while considering the spatial relationship. This is of great significance to the convergence speed and training efficiency of the model.

The gas concentration values at different monitoring points are correlated in time and space, and the accuracy of gas prediction and the efficiency of the model can be improved through fusion of the data of correlated monitoring points (Dominik et al., 2018; Wu, 2015; Lu et al., 2019).

4. Gas concentration prediction model based on Encoder-Decoder LSTM

Deep learning can train the sample data using appropriate methods, and reverse adjusting network parameters to obtain a machine learning process with a deep network structure. The Encoder-Decoder model based on LSTM was first proposed for an application in machine translation (Thang et al., 2017; Graham, 2017). In this paper, a multi-step prediction model for the gas concentration

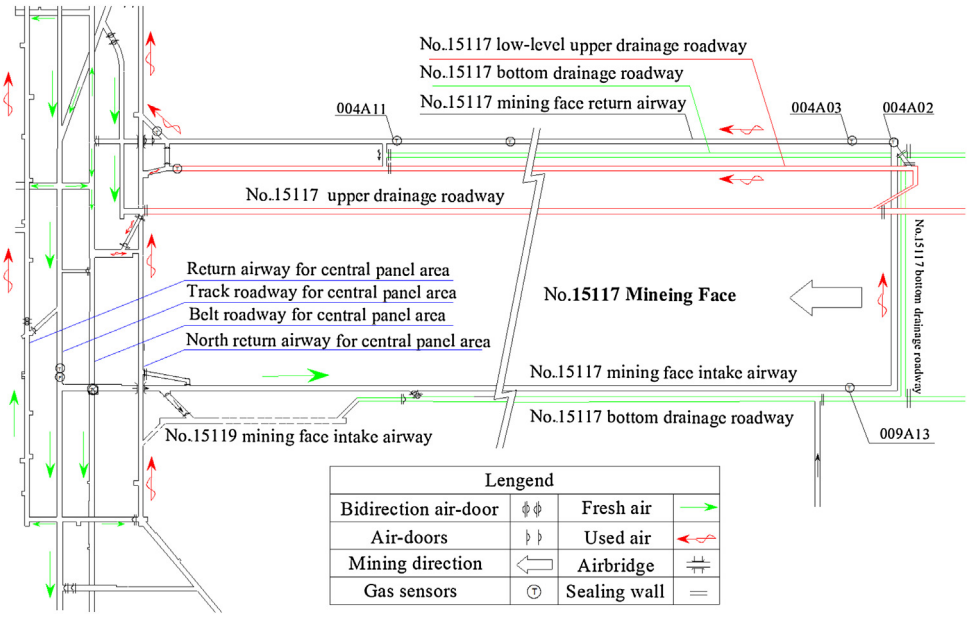


Fig. 2. Layout of No.15,117 Mining Face in Sijiazhuang Coal Mine.

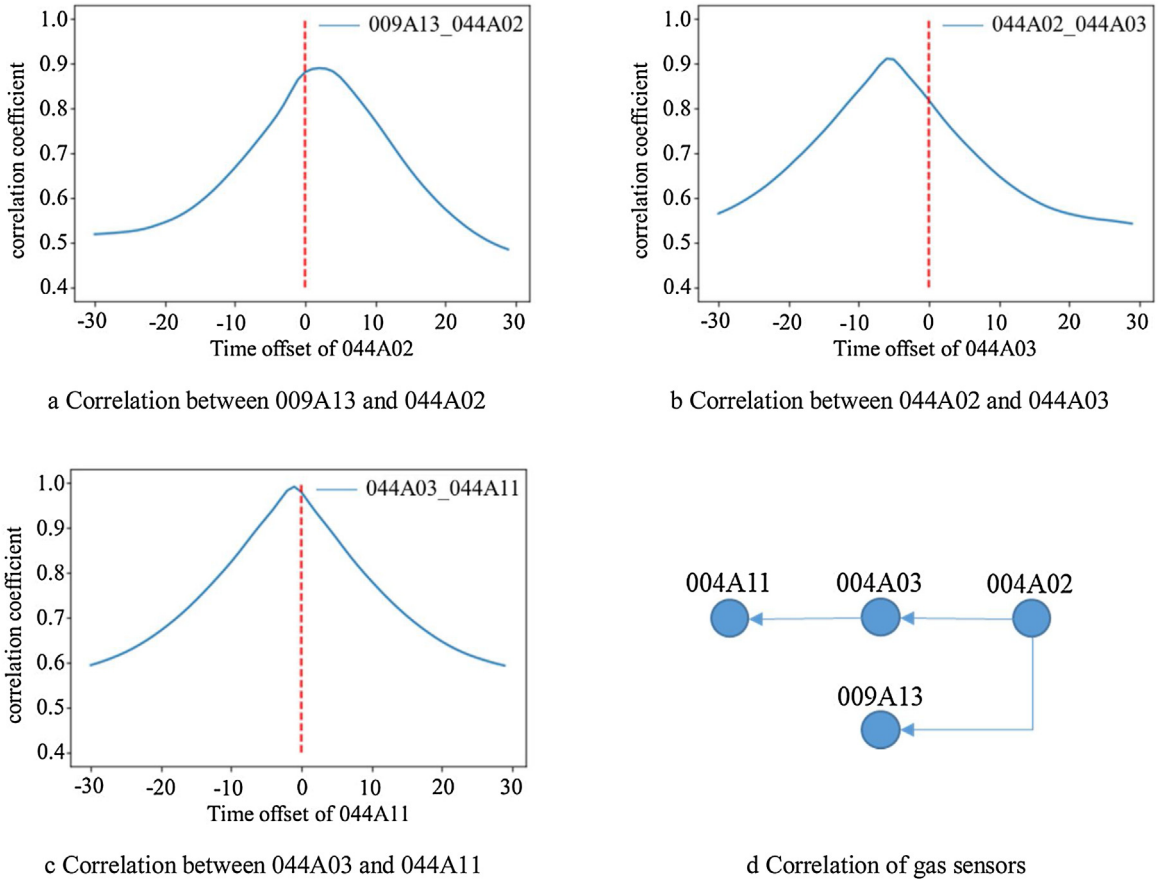


Fig. 3. Relevance of gas monitoring points.

according to a multi-sensor in a mine using the Encoder-Decoder (ED) framework in combination with LSTM is proposed. LSTM can effectively process temporal data, extract historical data features as well as related sensor data features, and inhibit the diffusion of important information.

4.1. Long-Short term memory (LSTM)

Long-Short Term Memory (LSTM) is an improved model of Recurrent Neural Networks (RNN). By introducing gates (input gates, forgetting gates, output gates) to control and maintain the

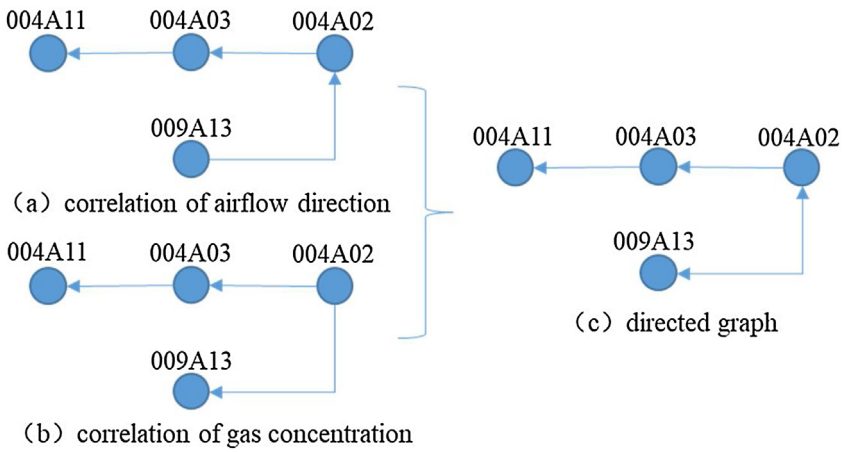


Fig. 4. Directed Map of Gas Monitoring Points.

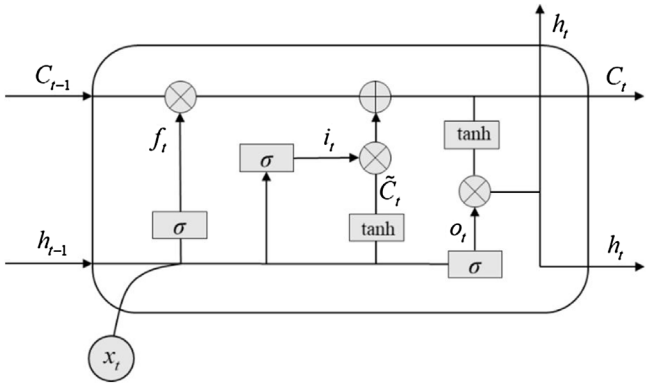


Fig. 5. Structural sketch of LSTM cells.

cell state of LSTM, long-term dependences and the retention of information in the memory are realized. Compared with RNN, LSTM uses memory cells and gate functions to control the flow of information, which can prevent gradients from disappearing too quickly. The cell structure of LSTM is shown in Fig. 5.

In Fig. 5 C_{t-1} is the state of LSTM cells at the previous time point, h_{t-1} is the hidden state at the previous time point and x_t is the input at the current time point. After passing through LSTM cells, both the cell state C and the hidden layer state h (also the output state) are updated to form a new cell state C_t and a new hidden layer state h_t. In LSTM cells, F_t is the forgetting gate, I_t is the input gate, $\tilde{C}_t$ is the temporary cell state, o_t is the output gate, t is the time, W_f, W_i, W_c, W_o are the linear transformation matrices, and b is the bias term. The corresponding formulas are given in Eq. (10) to Eq. (15).

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{10}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{11}$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{12}$$

$$C_t = f_t * C_{t-1} + I_t * \tilde{C}_t \tag{13}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{14}$$

$$h_t = o_t * \tanh(C_t) \tag{15}$$

The core of LSTM is represented by the update process of the cell state. This means, cell state C_{t-1} discards part of the information using the forgetting gate at the previous time step, and obtains a new state C_t by adding part of the information through the input gate, while a new hidden state h_t is controlled and updated using the output gate. The LSTM cyclic network has an internal LSTM cellcycle (self-cycle) besides the external RNN cycle. In this way, LSTM does not simply impose an element-by-element nonlinearity on the input unit and the cyclic unit after an affine transformation (Ian et al., 2016).

4.2. LSTM encoder-decoder architecture

The problem of mapping an input sequence to an output sequence in a deep learning context is called seq2seq (sequence to sequence) problem (Sutskever et al., 2014). The ED model is a way to realize the seq2seq problem, which includes two parts of a network: the encoder and the decoder. The encoder and the decoder can be freely combined using CNN, RNN, LSTM, GRU, etc., as shown in Fig. 6.

The main purpose of the encoder stage is to extract features from the input time series data. A variable-length sequence data $x = \{x_1, x_2, \ldots, x_m\}$ is used as input and the encoder encodes the sequence into a fixed-length state vector C, which is used as the input of the decoder. In the decoder stage, the decoder decodes the state vector C and predicts the next time sequence Y by combining the input data of the current time.

In the ED model, the hidden layer state h is updated every time the input data is read. When the end of sequence X is read, the hidden layer state vector C can be regarded as a summary of the whole input sequence. This means, the information in the sequence has been extracted and mapped into vector C.

4.3. LSTM Based Encoder-Decoder Model for gas prediction using multi-sensor information fusion

Based on the ED framework, this paper establishes a multi-step predication model for gas concentrations using multi-sensor information fusion. By adjusting network parameters and the loss function and by training with a large number of data, the temporal predication of a gas concentration is realized. Its structure is shown in Fig. 7. The encoder part of the model is a single-layer LSTM network, which maps the characteristic historical data of the gas concentration of multiple sensors into state vectors. The decoder part is composed of LSTM and fully connected layers, which are used to decode the state vectors into future gas concentration sequences. The two parts are connected by the state vector C.

As shown in Fig. 7, in the encoder stage, at a certain time point t, the model takes the historical gas concentration and characteristic data $\{x_{t-m}, \ldots, x_{t-1}, x_t\}$ as input (length m). The LSTM network maps

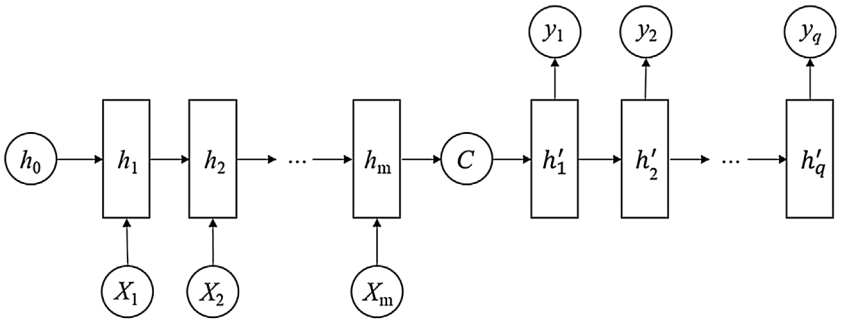


Fig. 6. Encoder-Decoder Model.

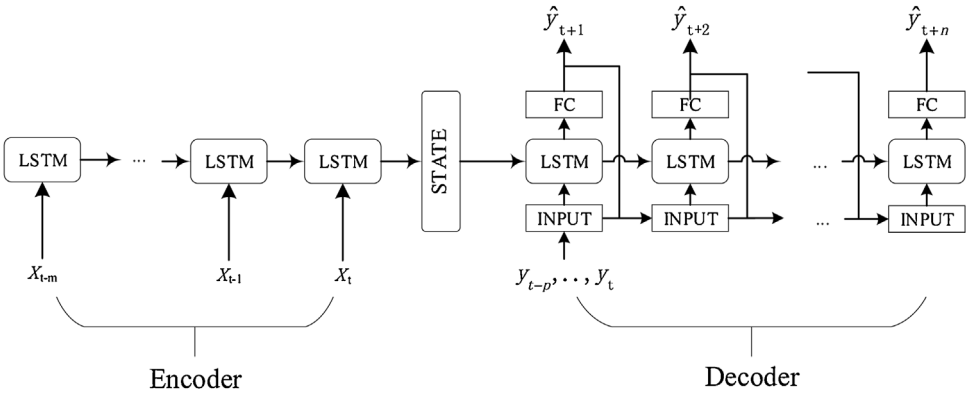


Fig. 7. Gas concentration prediction model based on Encoder-Decoder.

the input data to the state vector C and the hidden layer state h, and thus provides the output of the encoder of the model. According to the feature extraction method in Section 3, the feature vector X_t is calculated using temporal data of the gas concentration of each sensor.

In the decoder stage, the model takes the output of the encoder stage (state vector C and hidden layer state h) as input, as well as the historical concentration data $\{y_{t-p}, \ldots, y_{t-1}, y_t\}$ (length p) as input of each time step of the LSTM network. Through fully connected layers, the prediction result $\{\hat{y}_{t+1}, \hat{y}_{t+2}, .., \hat{y}_{t+n}\}$ of the gas concentration of the next time steps (length n) is obtained. In the multi-step predication process, the current time point is taken as the benchmark, the previous gas concentration sequence (length p) is obtained and is used as the input of the current time point. However, in a multi-step prediction process, for some previous time steps p real gas concentration data may not exist, so it is necessary to fill these missing values using predicted values. For example, when predicting $\hat{y}_{t+2}$, the INPUT vector is $\langle y_{t-p+1}, \ldots, y_t, \hat{y}_{t+1} \rangle$, as shown in Fig. 8.

Gas concentration prediction is a classic regression problem; MSE is used as loss function of the model in this paper, as shown in Eq. (16). In the Eq. (16), b is the batch size, n is the prediction step, t is the current time, y is the real gas concentration value, and $\hat{y}$ is the predicated gas concentration value.

$$loss = \frac{1}{b}\sum_{i=0}^{b}\frac{1}{n}\sum_{j=0}^{n}\left(y_{i,t+j} - \hat{y}_{i,t+j}\right)^2 \tag{16}$$

In order to prevent over-fitting, the L1 regularization term is added to the loss function, which can effectively improve the generalization ability of the model training. Eq. (17) shows the loss function after adding the L1 regularization term. W denotes the whole parameter matrix of the model. $\left||W|\right|_1$ is the1-norm of W.

$$loss = \frac{1}{b}\sum_{i=0}^{b}\frac{1}{n}\sum_{j=0}^{n}\left(y_{i,t+j} - \hat{y}_{i,t+j}\right)^2 + \left||W|\right|_1 \tag{17}$$

5. Result

In this paper, the monitoring data of gas sensor in the 15,117 working face of the Sijiazhuang Coal Mine are taken as samples the sampling interval is 2 min). 19 days of observation data from 2017-10-30 to 2017-11-17 are selected as the training set, 10 h of data from 2017-11-18 02:12:00 to 2017-11-18 12:12:00 are taken as the test set. The *Auto Regressive Average Moving* (ARMA) model, the *Chaos Time Series* (CHAOS) model, the ED model (single-sensor) and the ED model (multi-sensor) were used to predict gas concentration changes for up to five sampling intervals (2 min, 4 min, 6 min, 8 min, 10 min) from the current time point. By comparing the MAE values of the prediction results of each model, the gas concentration prediction accuracy of the ED model (multi sensor) with fusion of multi-sensor information is verified.

5.1. Data preprocessing

According to the laws of gas emission from different sources as described in Section 2, gas diffuses in the roadway network along with the air flow and the coal flow after gushing out from each gas source in the mining face. In order to monitor the gas concentration in real time and to ensure safety in the production, sensors are arranged at different positions of the mining face in accordance with the relevant provisions of the *"Coal Mine Safety Regulations"*(State Administration of Work Safety, 2016). Taking the No. 15,117 mining face in the Sijiazhuang Coal Mine as an example,

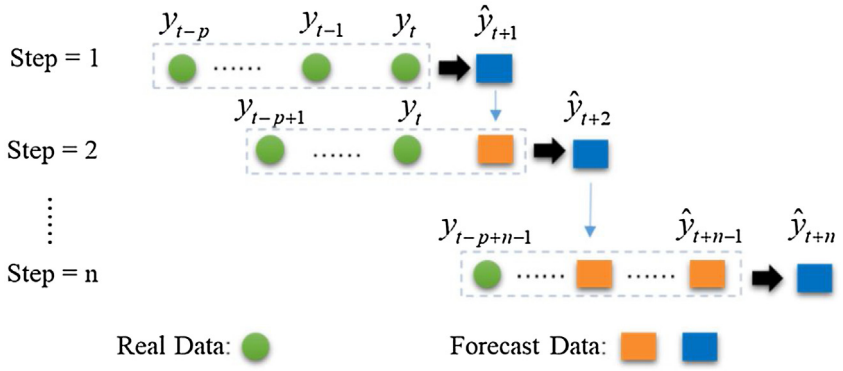


Fig. 8. Multi-step prediction of gas concentration.

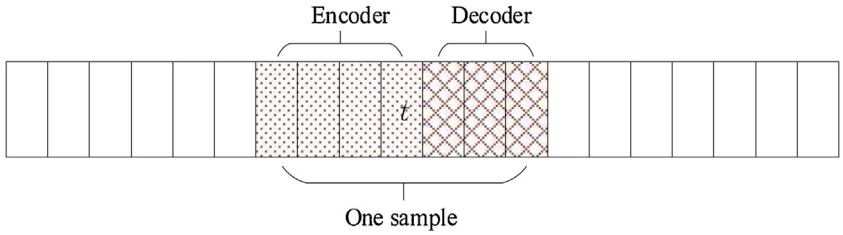


Fig. 9. Single sample random sampling.

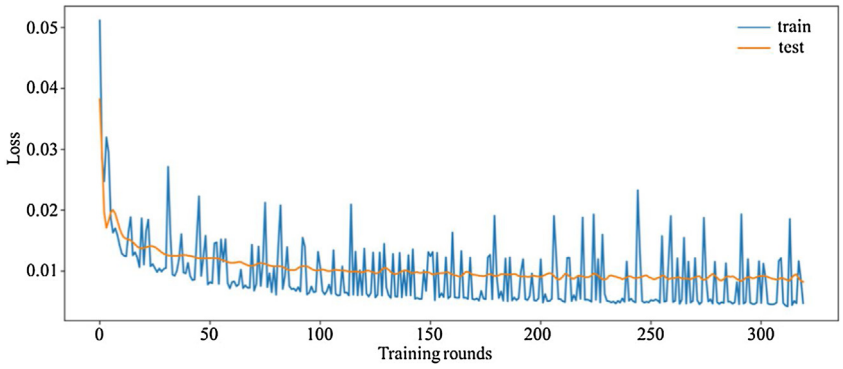


Fig. 10. Loss of ED model.

the gas sensor arrangement is shown in Fig. 2. The Sijiazhuang Coal Mine is a coal and gas outburst mine. At present, the gas content in 15 # coal seam, where the No. 15,117 working face is located, is 24.62 m^3/t. The gas emission from adjacent seams accounts for more than 80 % of the total gas emission in the mine (Guo, 2018; Zhang, 2018).

The No. 15,117 working face includes four gas concentration monitoring points as shown in Fig. 2: An intake air monitoring point (No.009A13), an upper corner monitoring point (No.044A02), a return air monitoring point (No.044A03) and a mixed return air monitoring point (No.044A11), as well as a gas flow sensor in the high-drainage roadway. The initial values of gas concentration at each monitoring point are shown in Fig. 1 (a). Because the peak value of the gas concentration at No.044A03 is much higher than that of other measuring points, we select No.044A03 as a prediction point, and create a gas prediction model for it.

In order to derive a stable training and to prevent excessive weight deviation of the model, the first-order difference of the gas concentration is adopted in the establishment of the prediction model based on deep learning. The first-order difference value is centered on 0, which can effectively avoid excessive weight deviation in the training and can quickly restore the true concentration of the model according to the first-order difference value.

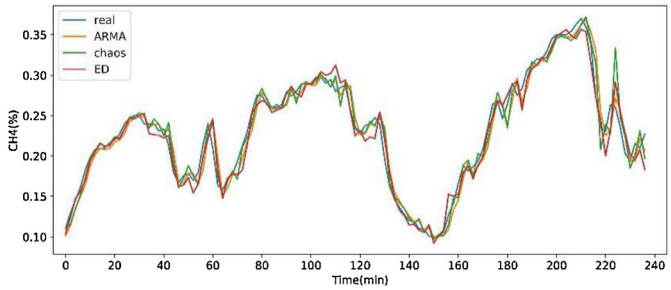
5.2. Data sampling

In this paper, the batch training method is used to train the model. For continuous temporal data, batch sampling is needed before the training. We first define the single sample random sampling method, before defining the batch sampling according to the single sample sampling, and finally we define each epoch data according to the batch sampling.

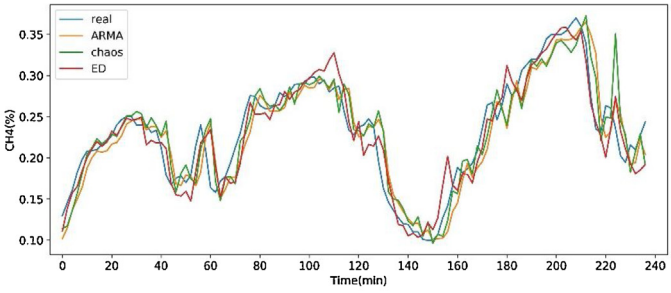
(1) Single sample random sampling: Fig. 9 shows a single sample scheme. For the given sequence data, a time t is randomly selected. According to the time t, the sequence of [t-p, t + n] is sampled, which is the corresponding sample at the time t.

(2) Batch sampling: Using the single sample sampling method mentioned above, without playback sampling b times, b samples can be obtained to form a batch.

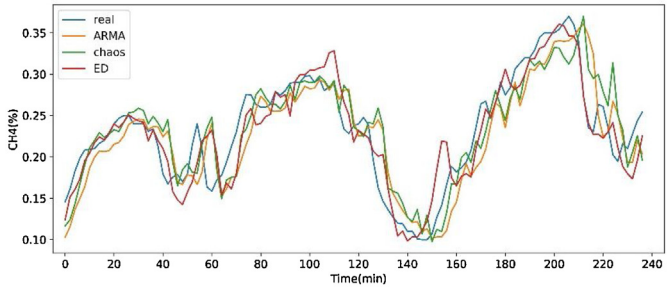
(3) Epoch sampling: Repeated batch sampling (without playback sampling). When the set of sampled time points covers the predefined set of valid time points, an epoch sampling has been completed.



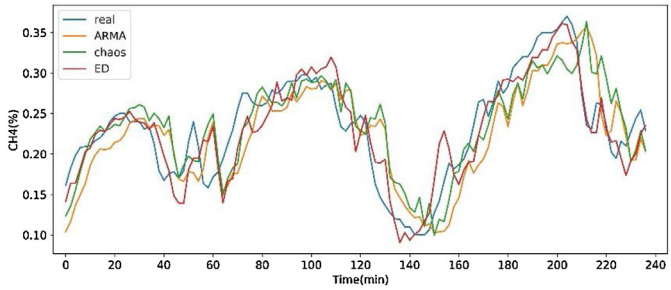
a Comparisons of the results of the first time step (2 minutes later)



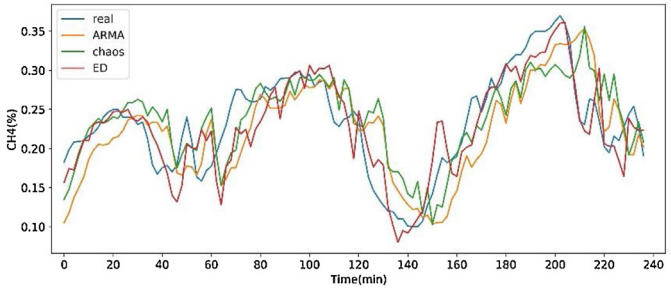
b Comparisons of the results of the second time step (4 minutes later)



c Comparisons of the results of the third time step (6 minutes later)



d Comparisons of the results of the fourth time step (8 minutes later)



e Comparisons of the results of the fifth time step (10 minutes later)

Fig. 11. Comparison of Multi-step Prediction Results of Gas Concentration(ARMA, CHAOS, ED).

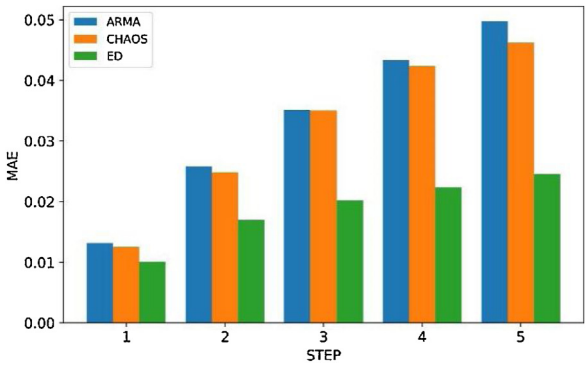


Fig. 12. Comparison of MAE(ARMA, CHAOS, ED).

5.3. Prediction experiment

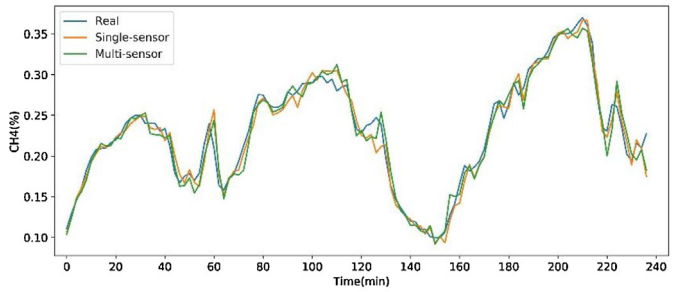
In this section, the ARMA model and the CHAOS model are used as baseline models to verify the prediction effect of the ED model. The ARMA model is a widely used time series short-term prediction method which provides high accuracy. Different time series data of gas concentrations can be well applied to this model (Tang, 2018). The CHAOS model is one of the most frequently used methods in gas concentration prediction. In this paper, the phase space of the gas concentration is reconstructed, and the CHAOS gas concentration prediction model is determined through combining *Radial Basis Function* (RBF) neural networks (Wei, 2015).

Compared with the baseline models (ARMA and CHAOS), the ED model can add various factors to improve the prediction accuracy. In order to ensure that different models input equal information, this paper compares and analyses the prediction results of the three models without adding the information of other sensors first. For the test set, data from a ten hours timeframe (from 2017-11-18 02:12:00 to 2017-11-18 12:12:00) is selected. During this period, the working face is in the production state, and the data fluctuates greatly. The gas concentration for five successive time steps (2 min, 4 min, 6 min, 8 min, 10 min) is predicted. Fig. 10 shows the changes of training loss and test loss during the training of the ED model.

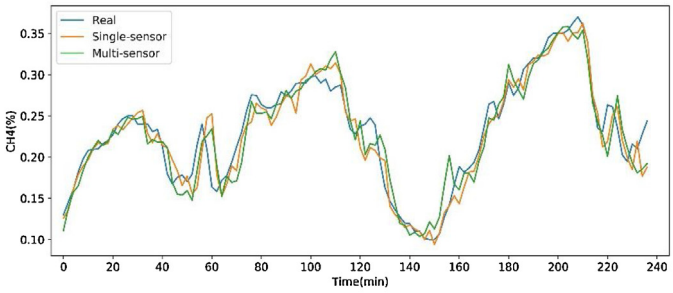
Fig. 11 shows the prediction results of five different time steps of the ARMA model, the chaotic time series model and the ED model without fusing other sensor information on the test set respectively. Because of the large volume of the test data set, only 4 h of test data predication results are selected and shown in Fig. 11.

Fig. 11 shows that the gas concentration prediction results of the first time step (2 min later) of the three models are close to the true values, with the ED model performing slightly better than the ARMA and the CHAOS models. With an increase of the time steps, the advantages of the ED model become more and more apparent. For the prediction results of step five (10 min later), the prediction effect of the ED model is significantly better than that of other models.

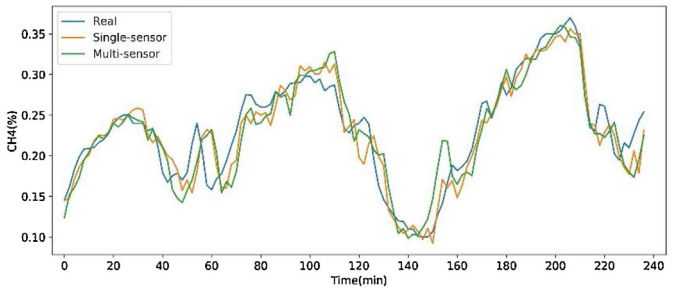
To quantify the prediction effect, Fig. 12 compares the MAE (mean absolute error) of the ED model, the ARMA model and the CHAOS model for different time steps. Fig. 12, shows that the ED model is superior to the ARMA model and CHAOS model regarding the multi-time step prediction of gas concentrations. Although the prediction errors of the three models increase with the increase of time steps, the increase of the prediction errors of the ED model is significantly smaller than that of the ARMA model and the CHAOS model. Thus, the ED model performs better than other models in the multi-step gas concentration prediction. The results show that the ED model based on LSTM can extract more information from long-term historical data, and thus obtains better gas concentration prediction results.



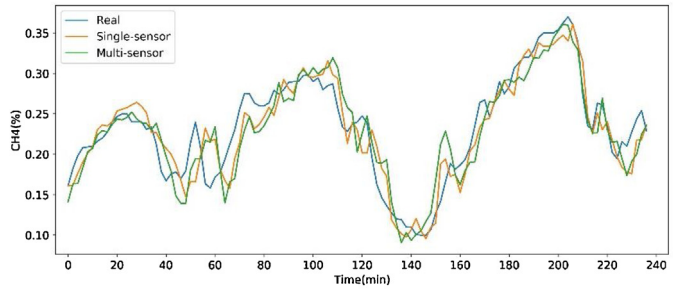
a Comparisons of the results of the first time step (2 minutes later)



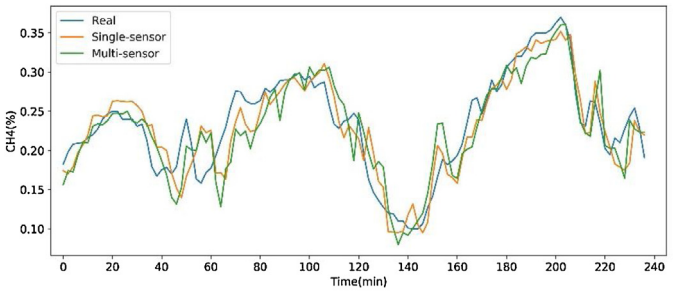
b Comparisons of the results of the second time step (4 minutes later)



c Comparisons of the results of the third time step (6 minutes later)



d Comparisons of the results of the fourth time step (8 minutes later)



e Comparisons of the results of the fifth time step (10 minutes later)

Fig. 13. Comparison of Multi-step Prediction Results of Gas Concentration(Single-sensor, Multi-sensor).

Table 1
MAE Comparison of Gas Concentration Prediction Results.

MAE	1st step	2ed step	3rd step	4th step	5th step
ARMA	0.0115	0.0214	0.287	0.0358	0.0409
CHAOS	0.0106	0.0188	0.0264	0.0331	0.0370
ED(single-sensor)	0.0100	0.0169	0.020	0.0223	0.0246
ED(multi-sensor)	0.007	0.0123	0.0141	0.0150	0.0165

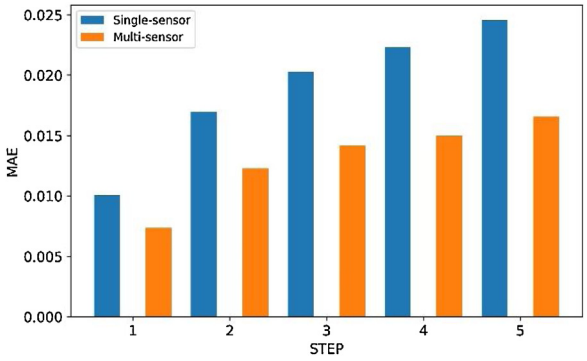


Fig. 14. Comparison of MAE(Single-sensor, Multi-sensor).

In order to further improve the prediction effect of gas concentration, the ED model is fused with relevant gas sensor information. The prediction effect for the five time steps in the ED model before and after fusion of multi-sensor information is shown in Fig. 13, and the prediction error MAE is shown in Fig. 14.

Fig. 13 and Fig. 14 show that the ED model provides high accuracy in the single-step gas concentration prediction without the fusion of multi-sensor information. However, the prediction error increases with an increase of the prediction time steps. The fusion of multi-sensor information can effectively improve the prediction accuracy of the ED model. The prediction error MAE of the fifth step (10 min later) of the ED (multi-sensor) model is 0.0165 and thus even lower than that of the second step (0.0169) of the ED (single-sensor) model. The results show that the data of other sensors contains information that affects the gas concentration prediction. The fusion of these information can help to improve the prediction accuracy of the ED model.

The MAE of the time series prediction results of gas concentrations based on the ARMA model, the CHAOS model, the ED model (single-sensor) and the ED model (multi-sensor) are presented in Table 1, with the result of the ED model being significantly better than those of other models.

6. Discussion

The mining face is an important production place in the mine. Safe and efficient mining processes are the basic assurance for a good operation of the coal mine enterprises. Before serious accidents such as coal and gas outburst or gas explosion occur, the gas concentration usually increases. In order to ensure safety during the coal mine production, to prevent gas accidents or to alert in case of impending accident, China's ¨*Coal Mine Safety Regulation*stipulates that the maximum gas concentration in mining face should not exceed 1%. Following this rule effectively reduces the number of malignant gas accidents in China. However, the mining depth in China is increasing year by year, and the gas content in coal seams increases correspondingly. As the gas concentration exceeds this 1% limit, the mining operation faces interruptions which restricts the production efficiency and cause panic among the miners. Therefore, a real-time prediction of gas concentration and measures such as reducing the shearer speed and increasing the air volume in advance can effectively ensure the continuous production of the working face.

After gas is released from the gas sources, it diffuses into the air flow of the roadway. Usually, the gas emission from coal wall and adjacent strata is relatively stable, while the gas emission from falling coal and goaf is affected by the mining speed, the coal fragmentation degree and the wind speed. The existing monitoring system lacks real-time records of the production of the mining face. Moreover, the degree of coal fragmentation is difficult to characterize. The amount of gas emission from coal falling is related to the amount of coal falling and the degree of coal fragmentation. This part of information is reflected in the value measured by the gas sensors in transport roadways to a certain extent. Therefore, this paper proposes a multi-sensor fusion, which can compensate for the lack of monitoring information and effectively improve the prediction of the gas concentration.

In addition, the current research mainly focuses on single-step predictions of the gas concentration. This means, they only predict the gas concentration for a single sampling interval after the current time point. On the one hand, if this temporal sampling interval is short, the advance prediction time is too short to meet the requirements for a timely warning. On the other hand, if the temporal sampling interval is long, a lot of information may be omitted. Taking the sampling interval of two minutes as an example, this paper predicts the gas concentration for five time steps, thus after 2 min, 4 min, 6 min, 8 min and 10 min. This approach realizes a multi-scale prediction regarding the time dimension, which allows for the accuracy to reach higher levels.

In future work, the real-time coal production, the shearer speed and the air volume should be taken into consideration as well. Moreover, we plan to use image processing techniques to calculate the degree of coal fragmentation, and to analyze the relationship between the gas concentration at monitoring points and the above parameters, to allow for gas concentration data to prevent the gas accident and guide the coal production.

7. Conclusion

An accurate prediction of the gas concentration is closely related to mine safety and production. Several gas sources exist in a mine. Gas migrates and diffuses into the mine roadway after mixing with the air flow. The gas concentration at different monitoring points is correlated. Based on these observations, this paper presents a deep learning prediction model for gas concentrations based on the Encoder-Decoder framework, which integrates the gas concentration of several monitoring points in the underground mining face to build a sequence model. Taking the data from 19 days (from 2017-10-30 to 2017-11-17) as the training set and ten hours (from 2017-11-18 02:12:00 to 2017-11-18 12:12:00) as the test set, the gas concentration after five sampling points from the current time point was predicted. The results show that:

(1) The sources of gas in the working face mainly includes coal falling, goaf, coal wall and adjacent strata. The gas emission

is closely related to the mining speed, coal falling fragmentation and coal seam thickness. With the diffusion into the air flow and through the transportation system into the roadway network, the gas concentration at each monitoring point is correlated.

(2) The multi-step gas concentration prediction model based on the Encoder-Decoder framework provides better prediction accuracy than the ARMA model and the CHAOS model, as well as higher robustness. It can predict the gas concentration well at five time steps in the test set.

(3) The accuracy of the gas prediction and the stability of the multi-step prediction results of ED can be improved through automatically extracting spatial, topological characteristics of the gas concentration at different monitoring points and by fusing multi-sensor information in the prediction model.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgement

This work was supported by the Ministry of Science and Technology of the People's Republic of China. (grant number 2016YFC0801805).

References

Adam, D., Alicja, K., 2018. Forecast of methane emission from closed underground coal mines exploited by longwall mining – a case study of Anna coal mine. J Sustain Min 17, 184–194, http://dx.doi.org/10.1016/j.jsm.2018.06.004.

Bui, T.C., Le, V.D., Cha, S.K., 2018. A Deep Learning Approach for Forecasting Air Pollution in South Korea Using LSTM.

Cao, J., Li, W.P., 2017. Numerical simulation of gas migration into mining-induced fracture network in the goaf. Int. J. Min. Sci. Technol. 27 (4), 681–685, http://dx.doi.org/10.1016/j.ijmst.2017.05.015.

Cheng, J., Bai, J.Y., Qian, J.S., et al., 2008. Short-term forecasting method of coalmine gas concentration based on chaotic time series. J of China U Min Techno 2, 231–235, http://dx.doi.org/10.1016/S1872-5813(08)60033-X.

Cho, K., Merrienboer, B., Gulcehre, C., et al., 2014. Learning Phrase Representations Using RNN Encoder-decoder for Statistical Machine Translation. Computer Science., http://dx.doi.org/10.3115/v1/D14-1179.

Dominik, Ś., Marek, G., Andrzej, J., et al., 2018. A framework for learning and embedding multi-sensor forecasting models into a decision support system: a case study of methane concentration in coal mines. Inform Sciences, 451–452, http://dx.doi.org/10.1016/j.ins.2018.04.026, 112-133.

Dougherty, H.N., Özgen Karacan, C., 2011. A new methane control and prediction software suite for longwall mines. Comput. Geosci. 37 (9), 1490–1500, http://dx.doi.org/10.1016/j.cageo.2010.09.003.

Fu, H., Li, W.J., Meng, X.Y., et al., 2014. Application of IGA-DFNN for prediction coal mine gas concentration. Chin. J. Sens. Actuators 27, 262–266, http://dx.doi.org/10.3969/j.issn.1004-1699.2014.02.022.

Fu, H., Liu, Y.Z., Li, H.X., et al., 2015. Short term forecasting model of gas concentration in coal mine using the CAPSO-ENN. Chin. J. Sens. Actuators 28, 717–722, http://dx.doi.org/10.3969/j.issn.1004-1699.2015.05.018.

Gao, L., Yu, H.Z., 2006. Prediction of gas emission based on information fusion and chaotic time series. J of China U Min Techno 16, 94–96, http://dx.doi.org/10.1016/j.ahj.2015.07.001.

Graham, N., 2017. Neural Machine Translation and Sequence-to-sequence Models: A Tutorial.

Guo, Y.H., 2018. Exploration on comprehensive prevention and control of coal and gas outburst model and technical in Sijiazhuang Mine. Coal Sci. Technol 46, 139–142.

Guo, J.H., Cheng, Z.H., Kong, W.Y., 2018. Establishment and application of mathematical prediction model of gas emission rate in fully mechanized coal face. J. Coal Sci. Eng. 50, 109–113.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780, http://dx.doi.org/10.1162/neco.1997.9.8.1735.

Ian, G., Yoshua, B., Aaron, C., 2016. In: Ian, G., Yoshua, B., Aaron, C. (Eds.), Deep Learning. , 1rd edn, pp. 404–407, Boston, Massachusetts.

Jason, B., 2016. Basic Feature Engineering With Time Series Data in Python. https://machinelearningmastery.com/basic-feature-engineering-time-series-data-python.2018.11.13.

Karacan, C., 2007. Development and application of reservoir models and artificial neural networks for optimizing ventilation air requirements in development mining of coal seams. Int. J. Coal Geol. 72 (3), 221–239, http://dx.doi.org/10.1016/j.coal.2007.02.003.

Karacan, C., 2008. Modeling and prediction of ventilation methane emissions of U.S. Longwall mines using supervised artificial neural networks. Int. J. Coal Geol. 73 (3), 371–387, http://dx.doi.org/10.1016/j.coal.2007.09.003, 2008.

Karacan, C., Ruiz, F.A., Cotè, M., Phipps, S., 2011. Coal mine methane: a review of capture and utilization practices with benefits to mining safety and to greenhouse gas reduction. Int. J. Coal Geol. 86 (2), 121–156, http://dx.doi.org/10.1016/j.coal.2011.02.009.

Li, G., Hu, Y.J., Yu, H.Z., 2008. Prediction of gas emission time series based on W-RBF. J China Coal Soc 1, 67–70, http://dx.doi.org/10.1016/S1872-5791(08)60057-3.

Li, L., Qin, B.T., Ma, D., Zhuo, H., Liang, H.J., Gao, A., 2018. Unique spatial methane distribution caused by spontaneous coal combustion in coal mine goafs: an experimental study. Process. Saf. Environ. Prot. 116, 199–207, http://dx.doi.org/10.1016/j.psep.2018.01.014.

Liang, C., Wang, E.Y., Feng, J.J., Kong, X.G., Li, X.L., Zhang, Z.B., 2016. A dynamic gas emission prediction model at the heading face and its engineering application. J. Nat. Gas Sci. Eng. 30, 228–236, http://dx.doi.org/10.1016/j.jngse.2016.02.004.

Lu, G.B., Li, X.Y., Zu, B.H., et al., 2017. Research on time-varying series forecasting of gas emission quantity based on EMD-MFOA-ELM. J Safety Sci Techno 13, 109–114.

Lu, J.X., Zhang, X.P., Yang, Z.H., et al., 2019. Short-term load forecasting method based on CNN-LSTM hybrid neural network model. Autom. Electr. Power Systems 43, 131–137.

Ma, Y.L., 2018. Study on Prediction for Gas Concentration in Fully-mechanized Coal Mine Based on Time Series Analysis. Xi'an University of Science and Technology.

Malhotra, P., Ramakrishnan, A., Anand, G., et al., 2016. LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection.

Noack, K., 1998. Control of gas emissions in underground coal mines. Int. J. Coal Geol. 35 (1), 57–82, http://dx.doi.org/10.1016/S0166-5162(97)00008-6.

Park, S.H., Kim, B.D., Kang, C.M., et al., 2018. Sequence-to-Sequence Prediction of Vehicle Trajectory Via LSTM Encoder-decoder Architecture.

Qiao, M.Y., Ma, X.P., Lan, J.Y., 2011. Time series short-term gas prediction based on weighted LS-SVM. J Min Safety Eng, 310–314.

Shen, X.D., 2019. A survey of time series algorithms based on deep learning. Inf. Technol. Inf 1, 71–76.

Song, Y.W., Yang, S.Q., Hu, X.C., et al., 2019. Prediction of gas and coal spontaneous combustion coexisting disaster through the chaotic characteristic analysis of gas indexes in goaf gas extraction. Process. Saf. Environ. Prot. 129, 8–16, http://dx.doi.org/10.1016/j.psep.2019.06.013.

State Administration of Work Safety, 2006. AQ 1018–2006 Prediction Method of Mine Gas Emission, 2018.12.10 http://www.mkaq.org/Article/jishugf/201005/Article_24543.html.

State Administration of Work Safety, 2016. Coal Mine Safety Regulations. China Coal Industry Publishing House, Beijing, 2019.02.25 http://english.www.gov.cn/state_council/2014/09/09/content_281474986284037.htm.

Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to Sequence Learning With Neural Networks. NIPS.

Tang, J., 2018. Big Data Prediction Based on Methane Concentration Time Series. China University of Mining and Technology.

Thang, L., Eugene, B., Zh, Rui, 2017. Neural Machine Translation (seq2seq) Tutorial. https://github.com/tensorflow/nmt.2018.10.19.

Wang, E.Y., Fasong, B., 1996. Study on the mechanism and process of the methane movement in the tunnel. Shanxi Mining Inst. Learned J. 2, 24–29.

Wang, F.Y., Wang, Q.F., Zhang, X.Q., 2017. Establishment of prediction model of abnormal gas emission based on chaotic time series. China Coal 43, 138–143.

Wei, Y.Q., 2015. Interpolation and Prediction of Coal Mine Gas Monitoring Data. China University of Mining and Technology.

Wu, J.J., 2015. The Research of Gas Concentration Prediction Based on Space-time Neural Network Model , China University of Mining and Technology.

Wu, X., Qian, J.S., Huang, C.H., et al., 2014. Short-term coalmine gas concentration prediction based on wavelet transform and extreme learning machine. Math Probl Eng 2014, 1–8.

Wu, H.S., Yuan, S.Q., Zhang, C., et al., 2018. Numerical estimation of gas release and dispersion in coal mine using Ensemble Kalman Filter. J. Loss Prev. Process Ind. 56, 57–67.

Xia, T.Q., Zhou, F.B., Wang, X.X., et al., 2016. Controlling factors of symbiotic disaster between coal gas and spontaneous combustion in longwall mining gobs. Fuel 182, 886–896, http://dx.doi.org/10.1016/j.fuel.2016.05.090.

Xia, T.Q., Zhou, F.B., Wang, X.X., et al., 2017. Safety evaluation of combustion-prone longwall mining gobs induced by gas extraction: a simulation study. Process. Saf. Environ. Prot. 109, 677–687, http://dx.doi.org/10.1016/j.psep.2017.04.008.

Yang, M.L., Xue, Y.X., Jiang, Y.D., et al., 2009. Study on pattern of gas emission at fully-mechanized coal face in Liliu mining area. J China Coal Soc 34, 1349–1353.

Ye, Q., Lin, B.Q., Jiang, W.Z., 2006. The study of methane law in coal mining face. China Mining Magazine 05, 38–41.

Yin, H.S., 2010. Gas Time Series Analytical Method and Its Early-warning Application in Coalmine. China University of Mining and Technology.

Yu, Q.X., Wang, K., Yang, S.Q., 2000. Study on pattern and control of gas emission at coal face in China. J of China U Min Techno 29, 9–14, http://dx.doi.org/10.1103/PhysRevC.50.1713.

Zagorecki, A., 2015. Prediction of Methane Outbreaks in Coal Mines From Multivariate Time Series Using Random Forest , Berlin, Germany.

Zhang, L., 2018. Analysis of coal and gas outburst mechanism in Sijiazhuang Coal Mine. J. Coal Sci. Eng. 50, 84–86.
Zhang, J.Y., Cheng, J., Hou, Y.H., et al., 2007. Forecasting coalmine gas concentration based on adaptive neuro-fuzzy inference system. J of China U Min Techno 4, 494–498, http://dx.doi.org/10.3321/j.issn:1000-1964.2007.04.015.
Zheng, C.S., Jiang, B.Y., Xue, S., Chen, Z.W., Li, H., 2019. Coalbed methane emissions and drainage methods in underground mining for mining safety and environmental benefits: a review. Process. Saf. Environ. Prot. 127, 103–124, http://dx.doi.org/10.1016/j.psep.2019.05.010.
Zhou, Y.G., Yao, E.Y., 2009. Modeling and forecast of time series based wavelets. Microcomputer Information 25, 29–30.