

Received January 14, 2020, accepted February 17, 2020, date of publication February 20, 2020, date of current version March 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975257

Research on a Mine Gas Concentration Forecasting Model Based on a GRU Network

PENGTAO JIA¹, HANGDUO LIU¹, SUJIAN WANG^{2,3}, AND PENG WANG⁴

¹Department of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, China

²Shaanxi Coal Industry Chemical Technology Research Institute, Xi'an 710065, China

³National and Local Joint Engineering Research Center for Green, Safe and Efficient Mining of Coal, Xi'an 710065, China

⁴Department of Safety Science and Engineering, Xi'an University of Science and Technology, Xi'an 710054, China

Corresponding author: Hangduo Liu (hangduo_liu@163.com)

This work was supported in part by the State Key Research Development Program of China under Grant 2018YFC0808303, and in part by the National Natural Science Foundation of China under Grant 51974236.

ABSTRACT To improve the level of safety in coal mine production, it is important to enhance the accuracy of coal mine gas concentration prediction. In the context of deep learning, we proposed a mine gas concentration prediction model based on gated recurrent units (GRUs). The GRU model is not only simple in structure but also offers high prediction accuracy, and it can make full use of the time-series characteristic of mine gas concentration data. First, we apply the Pauta criterion and Lagrange interpolation to preprocess mine gas concentration monitoring data. Then, a spatial reconstruction method is used to construct the training set for the prediction model. Finally, the mean square error (MSE) is used as the loss function and adaptive moment estimation (Adam) is used as the optimization algorithm to determine the learning parameters of the GRU model for predicting gas concentration values. Experimental results show that compared with models based on support vector regression (SVR), a backpropagation neural network (BPNN), a recurrent neural network (RNN) and a long short-term memory (LSTM) network, the proposed GRU-based model for gas concentration prediction achieves reduced error on the test set, and moreover, the GRU model is more efficient than the LSTM model in terms of run time. Thus, the accuracy and efficiency of gas concentration prediction are both improved, showing that the proposed model is of high practical value.


INDEX TERMS Deep learning, gated recurrent unit, gas concentration, prediction.

I. INTRODUCTION

Coal is an important pillar of China's primary energy consumption, and it is related to the economic and energy security of China. Frequent mine gas disasters have caused significant losses to China's coal industry and of miners' lives. At present, coal enterprises have installed safety monitoring systems, but the main functions provided by these systems are simply the short-term identification of and response to disasters. These monitoring systems fail to fully exploit the value of the available gas data, resulting in insufficient forecasting ability for mine gas disasters. Therefore, there is an urgent need to introduce new methods and technologies for supporting improved safety in coal mine production. Through the analysis of monitoring data, reliable and accurate forecasting of gas concentration levels can be achieved, thus improving the early warning capability for coal mine

gas disasters. Such forecasting will be of great significance for reducing the economic losses caused by mine gas disasters and protecting the lives of miners.

Mine gas concentration data are dynamic and nonlinear, and they are influenced both by many natural factors and by mining technology; thus, it is difficult to perform forecasting based on these data using traditional methods [1]. In recent years, neural networks have been widely applied for nonlinear regression modeling and forecasting. Among them, gated recurrent unit (GRU)-based networks have been theoretically proven to be able to represent nonlinear functions with arbitrary accuracy. Therefore, in this paper, we propose a GRU-based algorithm for gas concentration forecasting. The first step is to apply the Pauta criterion to process the noise in the mine gas concentration monitoring data. Then, Lagrange interpolation is applied to interpolate the missing values of the monitoring data to complete the necessary preprocessing. Subsequently, spatial reconstruction of the gas monitoring data is performed to construct the input samples to be used

The associate editor coordinating the review of this manuscript and approving it for publication was Liangxiu Han .

to train the GRU network. Then, we choose the mean square error (MSE) as the loss function and use adaptive moment estimation (Adam) as the optimization algorithm to construct a GRU learning model that is suitable for predicting the time-varying coal mine gas concentration. Finally, coal mine gas data are used for testing. An experimental evaluation shows that the error of the proposed GRU-based gas concentration prediction model is reduced by 7.9% compared with that of a model based on a long short-term memory (LSTM) network and that its run-time efficiency is simultaneously improved by 13.05%, endowing it with better practical value for application.

The remainder of this paper is organized as follows. Section 2 reviews some of the relevant literature on mine gas concentration forecasting. Section 3 describes the dataset and data processing. Section 4 introduces the theoretical background on forecasting methods and presents the proposed GRU-based method of mine gas concentration forecasting. Section 5 presents the parameter settings of the proposed model. The experimental results are analyzed in Section 6, and Section 7 concludes the paper.

II. LITERATURE STUDY

Considerable research has been conducted in the area of mine gas concentration forecasting. For example, Zhang Zhao-zhao proposed a brain-like multihierarchical modular neural network (BMNN) with applications for gas concentration forecasting [2]. The model of BMNN has a brain-like multihierarchical structure and uses a collaborative learning approach. Zhang *et al.* [3] presented research on and the application of an improved gas concentration prediction model based on gray theory and a backpropagation neural network (BPNN) for a digital mine. That paper exploited the advantages of gray prediction and the ability to revise the gray prediction model using a BPNN. Thus the researchers built an improved gas concentration prediction model and carried out specific computer simulations. Stopforth and Davrajh [4] was to develop equations for close estimations of gas concentrations for the Figaro sensors used, to allow other researchers the ability to identify the gas concentrations when using the sensors for different applications. Booth *et al.* [5] proposed a solution to the shortcomings of a traditional gas emissions forecasting method and analyzed the various characteristics and limitations of gas management design in practice. The proposed method used an improved spatial dataset for prediction while also incorporating basic physics and energy-related principles. Yu and Shi [6] proposed a gas concentration forecasting model based on a radial basis function (RBF) neural network and chaos theory, in accordance with the nonlinear and chaotic time-series characteristics of mine gas data. Liu Kun proposed a method of coal mine gas concentration analysis based on a support vector machine (SVM) model [7]. In this paper, on the one hand, the authors adopted support vector regression (SVR) to predict gas concentration values based on data from other well-performing sensors; on the other hand, they classified the gas concentration data into

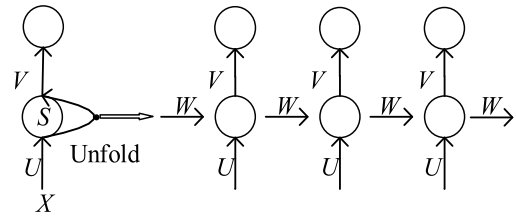


FIGURE 1. The structure of an RNN.

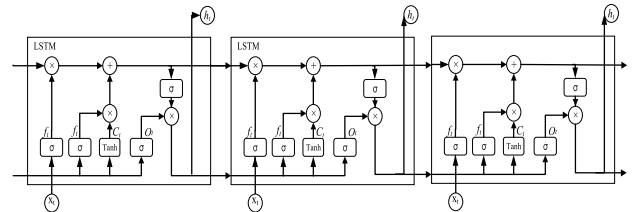


FIGURE 2. The structure of an LSTM neural network.

two classes, corresponding to either totally safe or slightly high concentrations, by applying a model constructed based on either C-support vector classification (SVC) or a one-class SVM.

However, all of the above methods focus on single-feature learning and prediction for mine gas concentration levels, and consequently, these methods are disadvantageous in terms of forecasting accuracy. To improve the accuracy of mine gas concentration forecasting, techniques based on time-series forecasting are receiving increasing interest. Since Hinton's team won the ImageNet competition using deep learning in 2012 [8], an increasing number of scholars have begun to pay attention to deep learning and to apply it in various fields. Deep learning has achieved breakthroughs in various fields, including natural language processing [9], [10], speech recognition [11], machine translation [12], and image comprehension [13]. A recurrent neural network (RNN) is a kind of deep neural network that is effective at processing sequential data. In contrast to conventional feed-forward neural networks, an RNN preserves, learns, and records the historical information contained in sequential data by means of periodically connected hidden layer nodes [14]. As shown in Figure 1, the structure of an RNN includes an input layer, hidden layers, and an output layer; U , V , and W are the weights from the input layer to the hidden layers and from the hidden layers to the output layer.

Since the parameters of an RNN are shared between layers, the number of parameters is dramatically reduced, thereby shortening the training time. However, an RNN can easily suffer from gradient vanishing or gradient explosion when processing a long sequence of data.

To solve the problems of gradient vanishing, gradient explosion and long-term dependence, Hochreiter and Schmidhuber [15] introduced the LSTM neural network architecture in 1997. LSTM is an improved design for deep neural networks based on the RNN architecture; it also has a chain-like structure, as shown in Figure 2.

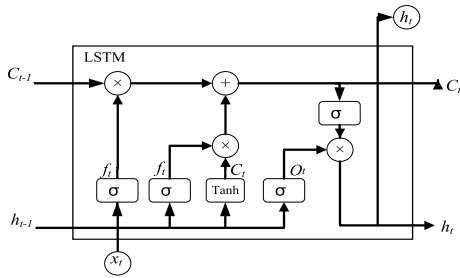


FIGURE 3. The internal structure of an LSTM neuron.

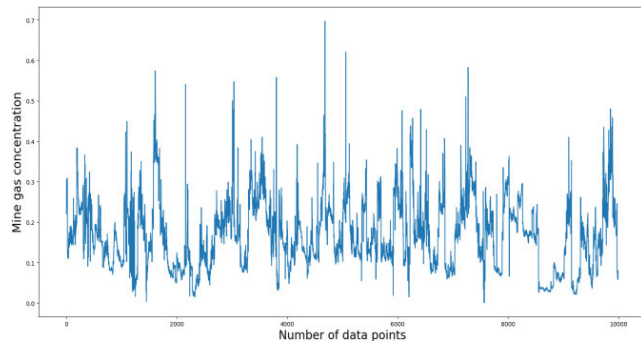


FIGURE 4. Raw data.

However, compared with the simple layers of RNN neurons, the structure of LSTM neurons is more complex, as shown in Figure 3.

The loop structure of an LSTM neuron includes three control gates: a forget gate, an external input gate, and an output gate. These gate structures allow the LSTM neuron to update, maintain, or delete information contained within the cell state. An LSTM network is trained using the Backpropagation Through Time (BPTT) algorithm to determine its parameters [16]. The BPTT algorithm calculates the error term of each LSTM neuron in the reverse direction, calculates the gradient of each weight in accordance with the corresponding error term, and then updates the weights using a gradient optimization algorithm.

LSTM networks are widely used because they can avoid the problems of gradient explosion and gradient vanishing encountered in RNNs. Li Weishan presented preliminary discussions on the application of LSTM models in coal mine gas forecasting and early warning systems [17] and confirmed their effectiveness for this purpose. However, the LSTM architecture is complex in structure and prone to overfitting. Therefore, we propose a GRU-based algorithm for gas concentration prediction. The GRU structure is simpler than the LSTM structure, with fewer parameters, a shorter training time, and lower susceptibility to overfitting.

III. RESEARCH DATA

A. DATASET DESCRIPTION

The dataset used in this study was constructed from gas concentration data for a mine working face. We collected the dataset, which consists of 10419 mine gas concentration data points, from January to March 2014, with a sampling

interval of 5 minutes. There are no missing raw data, as shown in Figure 4. To highlight the advantages of each model, we divided the dataset into three subsets: one containing 1041 data points, one containing 5209 data points and one containing all 10419 data points. In each subset, 70% of the data points were designated as training data, and the remaining 30% were used as test data to evaluate the prediction accuracy of each model.

B. DATA PREPROCESSING

A typical dataset for the task addressed in this paper consists of gas concentration data and other related data collected by monitoring equipment over a certain period of time (e.g., one hour, one day, or one month). However, these data will generally contain various missing values and outliers due to factors such as process alteration, equipment failure, air volume regulation or other factors related to human activity. Therefore, it is necessary to preprocess the gas concentration monitoring data that are collected in real time to improve the accuracy of mine gas concentration forecasting.

The first step of data preprocessing is to apply the Pauta criterion to process the noise in the gas concentration monitoring data. The Pauta criterion is one of the most common and simplest criteria for discriminating error. If the absolute value of a residual is larger than three times the standard deviation, meaning that inequality (2) is true when combined with formula (1), then this error is considered to be excessively large, and the corresponding measured value is considered to be an outlier that should be rejected, resulting in a missing value for the corresponding measurement time.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - x)^2}{n - 1}} \quad (1)$$

$$|x_i - x| > 3\sigma \quad (2)$$

where x is the average value of the dataset and σ is the standard deviation of the dataset.

The second step is to apply Lagrange interpolation to fill in the resulting missing values of the monitoring data sequence to obtain a complete set of processed monitoring data.

Finally, the dataset is normalized to values in the range of [0,1]. The normalization formula is as follows:

$$norm(x_i) = \frac{x_i - \min(X)}{\max(X) - \min(X)} \quad (3)$$

where $norm(x_i)$ is the normalized value of x_i , $\min(x)$ is the smallest value in the dataset, and $\max(x)$ is the largest value in the dataset.

IV. GRU-BASED FORECASTING FRAMEWORK

A. GRU MODEL

An LSTM module has a large number of parameters and a complex structure; thus, it is prone to overfitting. To overcome these deficiencies, Cho proposed the GRU structure as a variant of LSTM in 2014 [18]. The GRU architecture

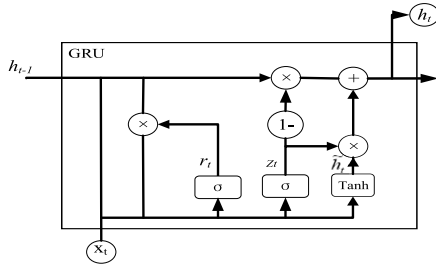


FIGURE 5. The internal structure of a GRU neuron.

maintains the characteristics of LSTM while having a simpler structure. The internal structure of a GRU neuron is shown in Figure 5.

Each of these recurrent neural network variants has a structure that consists of replicated instances of a particular module; however, the structure of the replicated module in a GRU network is slightly simpler than that in an LSTM network. A GRU neuron has only two gates, namely, an update gate and a reset gate, denoted by z_t and r_t in Figure 5. The update gate is used to control the extent to which the information of previous hidden states is carried over into the current state. The larger the value of the update gate is, the more information is preserved from previous states. Similarly, the reset gate is used to control the extent to which the information of previous states is ignored; the smaller the value of the reset gate is, the more information is ignored. Accordingly, short-term dependencies are usually captured by means of frequent activation of reset gates, while long-term dependencies are associated with the activation of update gates. Since the GRU architecture has only two types of control gates, the calculation speed of a GRU model is much faster than that of an LSTM model.

Let r_t represent the reset gate of a GRU at time t , with the following calculation formula

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (4)$$

where σ is the sigmoid function; X_t and h_{t-1} are the current input value and the last activation value, respectively; W_r is the input weight matrix; and U_r is the weight matrix of the loop connection.

Similarly, let z_t be the update gate of the GRU at time t , with the following calculation formula:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \quad (5)$$

Let h_t be the activation value of the GRU at time t , which is the median value between the last activation value h_{t-1} and the candidate activation value \tilde{h}_t :

$$h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t \quad (6)$$

\tilde{h}_t is calculated as follows:

$$\tilde{h}_t = \tanh(W_{\tilde{h}} x_t + r_t \circ U_{\tilde{h}} h_{t-1}) \quad (7)$$

where \circ represents the Hadamard product.

B. MODEL OPTIMIZATION

The magnitude of the loss function in traditional gradient descent optimization depends on the change in the parameters. Theoretically, the more iterations are performed, the less the loss function should change from one iteration to the next; however, the gradient descent algorithm still may not reach a globally optimal solution. Therefore, the Adam optimizer is chosen as the optimization algorithm in this work.

The Adam algorithm [19] was proposed by Kingma, Diederik P., and J. Ba. in 2014. The Adam algorithm is based on the adaptive estimation of a low-order matrix and is a stepwise optimization algorithm for a stochastic objective function. It can dynamically adjust the first-order and second-order matrices of the gradient of each parameter in accordance with the loss function. It has the advantages of easy implementation, high computational efficiency, and low memory consumption. In deep learning, the mini-batch technique is adopted, which causes the objective function to change with the different samples included in each batch. However, the Adam algorithm can still effectively solve the optimization problem even with this randomness in the objective function. It has been proven through experiments that the Adam algorithm is superior to random gradient descent optimization.

C. GAS CONCENTRATION FORECASTING FRAMEWORK

The gas concentration in a coal mine varies in a time-varying and nonlinear manner; thus, it cannot be predicted by means of a simple linear relationship. Neural networks have obvious advantages in predicting complex nonlinear time-varying sequences. Therefore, a gas prediction model based on a GRU neural network is constructed to predict the trend of gas concentration data.

Let the time series of previous gas concentration values at n consecutive moments be denoted by $X(t-n+1)$, $X(t-n+2)$, \dots , $X(t-1)$, with $X(t+1)$ being the predicted value at the next moment; then, the GRU-based gas concentration prediction model can be expressed as follows:

$$X(t+1) = \text{GRU}(X(t-n+1), X(t-n+2), \dots, X(t-1)) \quad (8)$$

The three-layer network framework of the proposed GRU gas concentration prediction model, which includes an input layer, a hidden layer, and an output layer, is shown in Figure 6. The input layer is responsible for preprocessing the original time series of gas concentration data to satisfy the requirements for the input to the GRU model; in the hidden layer, GRU neurons are used to construct a 1-layer loop neural network; and the output layer is mapped to one-dimensional sequence of data through a fully connected layer to realize gas concentration prediction.

V. MODEL PARAMETER SETTINGS

TensorFlow is a GPU- and CPU-based library. It acts as a backend for the Keras library [20]. Keras is a deep learning

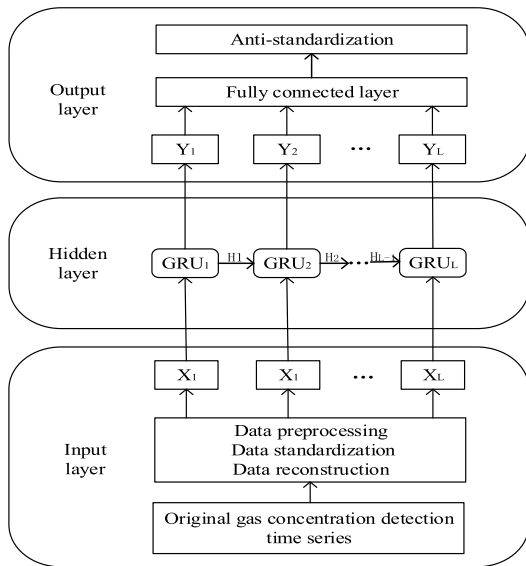


FIGURE 6. The structure of gas concentration forecasting model.

library that supports the implementation of complicated prepackaged architectures such as the RNN, LSTM and GRU architectures. TensorFlow does not provide many prepackaged architectures but rather supports the design of new architectures, whereas Keras supports new datasets for known architectures. In this work, the Keras and TensorFlow libraries were used for mine gas concentration forecasting. Additionally, Keras and TensorFlow contain many prepackaged deep learning functions, such as activation functions and loss functions. By using these functions, better results can be obtained with a given architecture. The model parameters are explained as follows.

A. PERFORMANCE INDEX

To test the effectiveness of the proposed GRU-based gas concentration prediction method, it is necessary to choose an indicator to comprehensively measure and evaluate the prediction effect. The MSE is chosen as the evaluation index in accordance with both the principles and practice of prediction effect evaluation. The MSE is defined as the expected value of the squared difference between the estimated value of the quantity being predicted and the corresponding true value. The MSE can be used to evaluate the degree of variation in the data. The smaller the value of the MSE is, the more accurately the prediction model describes the experimental data. The MSE can be calculated using the formula below:

$$MSE = \frac{1}{N} \sum_{t=1}^N (observed_t - predicted_t)^2 \quad (9)$$

where N is the number of data points.

B. ACTIVATION FUNCTION SELECTION

An activation function is used to convert an input signal into an output signal, which will then act as the input signal for the next layer. In each cell, a weight function is used for

processing in combination with the current state input and the previous state output. In this work, we use the sigmoid activation function. When information from the previous hidden state and information from the current input enter a cell, they are activated by the sigmoid function, with the values of the vectors being between 0 and 1. A value closer to 0 means that the corresponding information is more likely to be forgotten, whereas a value closer to 1 means that it is more likely to be retained. The sigmoid function is calculated using the following formula:

$$f(x) = 1 / (1 + \exp(-x)) \quad (10)$$

C. HIDDEN LAYER STRUCTURE

We have tested several representative combinations of numbers of hidden layers and hidden nodes and have found that the best choices for the values of these two parameters are 1 and 12, respectively.

D. SELECTION OF THE OPTIMIZATION ALGORITHM AND THE NUMBER OF ITERATIONS

The training algorithm for a neural network usually refers to the algorithm used to update the weight coefficients such that the value of the objective function will be gradually reduced until it converges to a global minimum [21]. In this work, we have selected the Adam as the optimization algorithm. The Adam algorithm is introduced in Section IV-B.

To address the problems of undertraining and overtraining, we apply the cross-validation method. If both the validation error and the training error are steadily decreasing, then the current model is still undertrained. If the validation error is increasing while the training error is steadily decreasing, we consider this behavior to be indicative of overtraining. Ultimately, we find that when the number of iterations is 10, the training error and verification error of the neural network model are basically stable. An SVR model has two very important parameters: C and γ . Here, C is the penalty coefficient, which represents the tolerance for error. The higher C is, the less tolerance there is for error, and the easier it is to overfit the model. The smaller C is, the less well fitted the model will be. γ is a parameter associated with the RBF kernel function and implicitly determines the distribution of the data mapped to the new feature space. The larger the value of γ is, the fewer support vectors there are, whereas a smaller γ value corresponds to more support vectors. The number of support vectors affects the speed of training and prediction. Finally, we set C equal to 1 and γ equal to 0.01.

VI. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETTINGS

In the experiment, the data were split into two different sets, one for training and one for testing. The training set was used to train the forecasting models, while the test set was used to evaluate the final results. The procedure applied in the

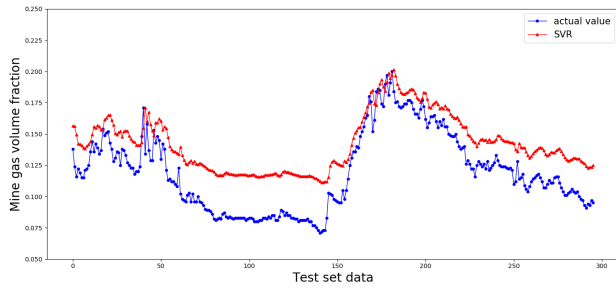


FIGURE 7. Prediction effect of the SVR model on the test set.

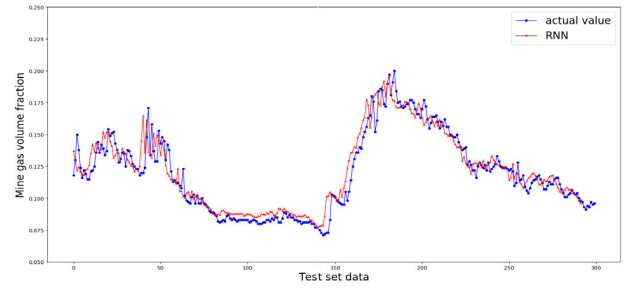


FIGURE 9. Prediction effect of the RNN model on the test set.

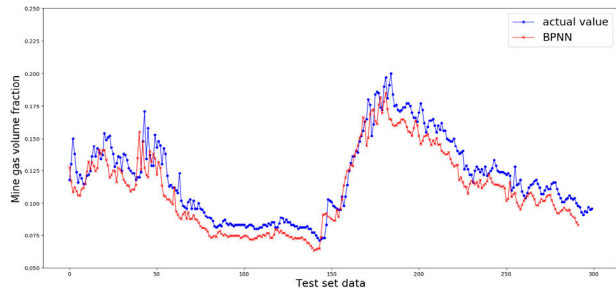


FIGURE 8. Prediction effect of the BPNN model on the test set.

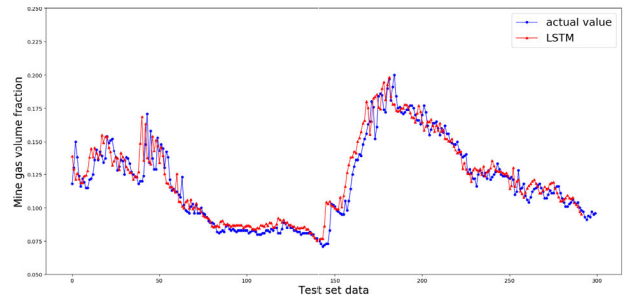


FIGURE 10. Prediction effect of the LSTM model on the test set.

mine gas concentration forecasting experiment is described as follows:

- I. Data preprocessing.
- II. Division of the dataset into a training set and a test set, with the training set comprising 70% of the data and the test set comprising the remaining 30% of the data.
- III. Construction of a 3D array from the training set for input to the GRU network via spatial reconstruction.
- IV. Determination of the parameters of the GRU model, using the MSE as the loss function and the Adam algorithm as the optimization method.
- V. Determination of the number of training iterations, namely, the number of epochs, in accordance with the loss function.
- VI. Forecasting on the test set.
- VII. Comparison with other gas concentration forecasting models of different structures and evaluation of these models in terms of the chosen performance index.
- VIII. Visualization of the prediction results after inverse normalization.

Through the above 8 steps, a mine gas concentration prediction model was established based on a GRU neural network.

B. RESULTS AND ANALYSIS

In this experiment, a neural-network-based method was used for the prediction of mine gas concentration levels. Then, we considered the MSE and the run time of the model to evaluate the modeling effect. Through a comparative experimental analysis, we found that the effect of each model is different on different subsets of data; however, as the amount of data to be processed increases, the effect of the GRU model becomes superior to that of the other models.

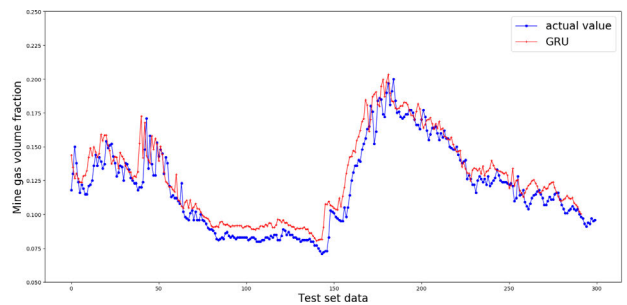


FIGURE 11. Prediction effect of the GRU model on the test set.

TABLE 1. Comparison of predictive performance.

	MSE on training set	MSE on test set	Run time on CPU
SVR	0.0346	0.02544	4.447 s
BPNN	0.02497	0.01154	56.209 s
RNN	0.02079	0.00742	91.841 s
LSTM	0.02077	0.00725	134.726 s
GRU	0.02083	0.00758	123.929 s

First, we tested the models on the subset consisting of 1042 data points. The prediction effects of the SVR, BPNN, RNN, LSTM, and GRU models on the corresponding test set are shown in Figures 7, 8, 9, 10, and 11, respectively. A further performance comparison of these models is presented in Table 1.

From Table 1, we find that the RNN and LSTM models achieve higher prediction accuracy than the proposed GRU model does on the subset with a length of 1042 data points; therefore, for small data sets, RNN and LSTM models are still preferred in terms of prediction accuracy.

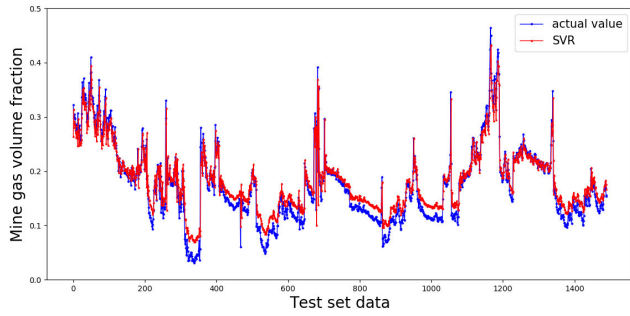


FIGURE 12. Prediction effect of the SVR model on the test set.

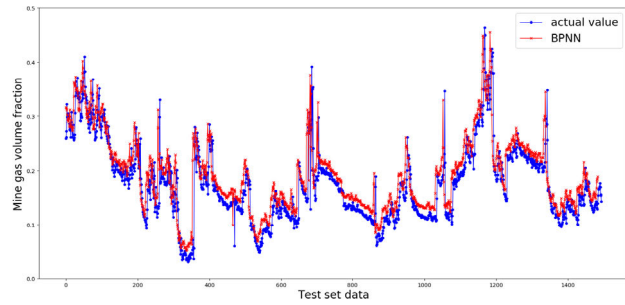


FIGURE 13. Prediction effect of the BPNN model on the test set.

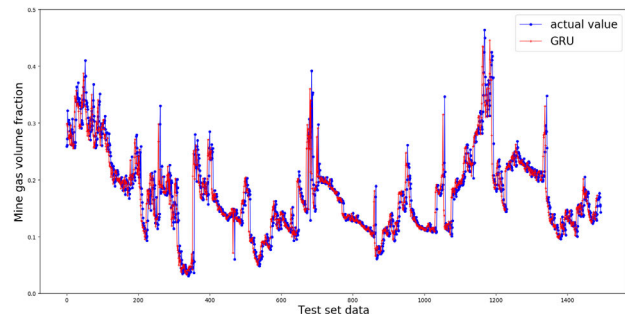


FIGURE 14. Prediction effect of the RNN model on the test set.

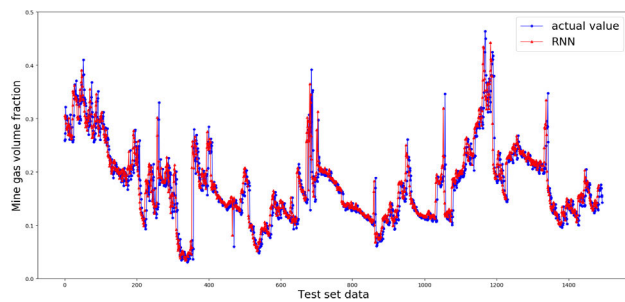


FIGURE 15. Prediction effect of the LSTM model on the test set.

Then, we conducted the same experiment on the subset consisting of 5209 data points. The prediction effects of the SVR, BPNN, RNN, LSTM, and GRU models on the corresponding test set are shown in Figures 12, 13, 14, 15, and 16, respectively, and the MSE and run time of each model are listed in Table 2.

From Table 2, it can be concluded that the error of the GRU model is less than that of the other four models, indicating that the prediction accuracy of the GRU model increases with an increasing number of data.

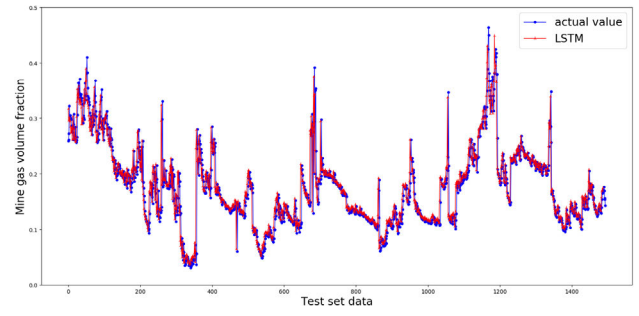


FIGURE 16. Prediction effect of the GRU model on the test set.

TABLE 2. Comparison of predictive performance.

	MSE on training set	MSE on test set	Run time on CPU
SVR	0.04024	0.04165	26.998 s
BPNN	0.03202	0.03781	275.35 s
RNN	0.02703	0.03345	456.58 s
LSTM	0.02699	0.03336	685.85 s
GRU	0.02625	0.03284	618.36 s

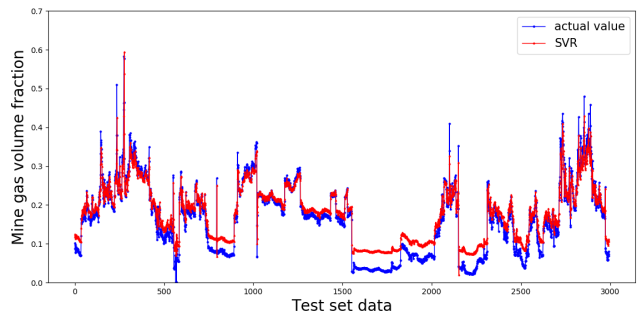


FIGURE 17. Prediction effect of the SVR model on the test set.

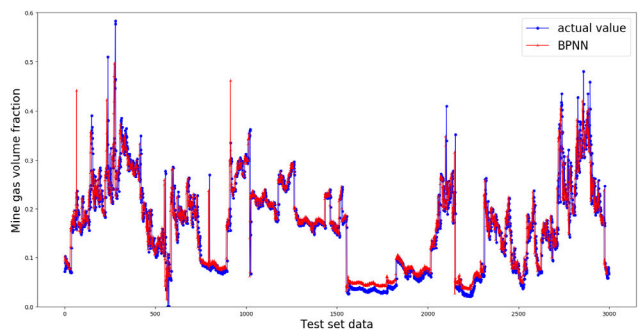


FIGURE 18. Prediction effect of the BPNN model on the test set.

Finally, we conducted the same experiment again on the whole collected dataset, with a total of 10419 data points. The prediction effects of the SVR, BPNN, RNN, LSTM, and GRU models on the corresponding test set are shown in Figures 17, 18, 19, 20, and 21, respectively, and the MSE and run time of each model are listed in Table 3.

From Table 3, we can see that the GRU model achieves smaller errors on both the training set and the test set compared with the other models. Compared with the SVR model,

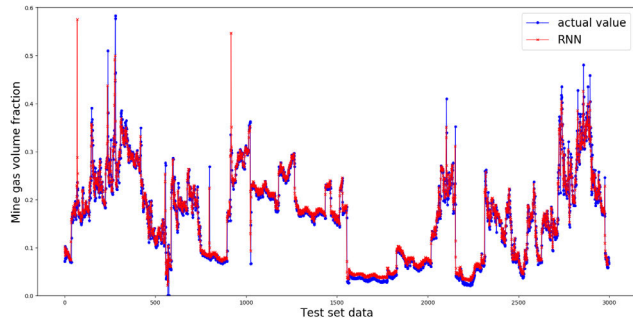


FIGURE 19. Prediction effect of the RNN model on the test set.

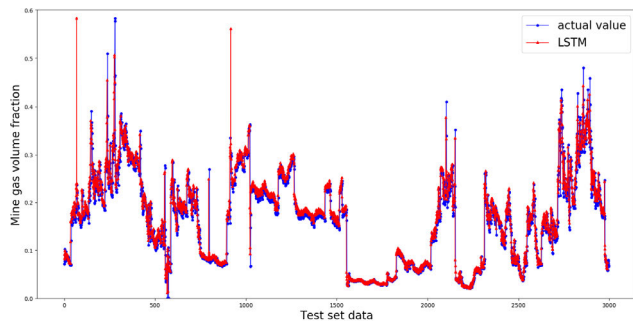


FIGURE 20. Prediction effect of the LSTM model on the test set.

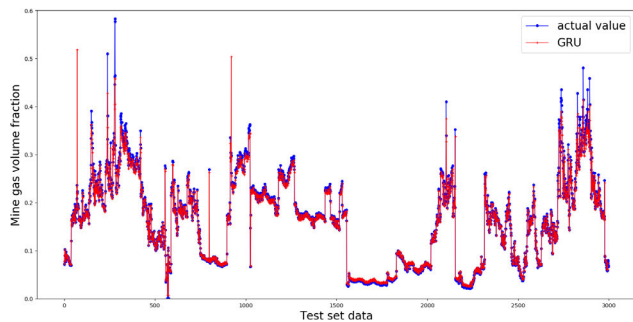


FIGURE 21. Prediction effect of the GRU model on the test set.

TABLE 3. Comparison of predictive performance.

	MSE on training set	MSE on test set	Run time on CPU
SVR	0.4635	0.04326	38.185 s
BPNN	0.03361	0.03064	537.59 s
RNN	0.03051	0.02612	875.34 s
LSTM	0.02968	0.02582	1348.97 s
GRU	0.02836	0.02405	1172.87 s

the error of the GRU model on the test set is reduced by 44.41%. Compared with the BPNN and RNN models, the error of the GRU model on the test set is reduced by 21.05% and 7.9%, respectively. Compared with the LSTM model, the error of the GRU model on the test set is reduced by 6.8%, and the GRU model also shows a 13.05% efficiency improvement over the LSTM model in terms of run time.

VII. CONCLUSION

To reduce the economic losses caused by mine gas disasters and protect the lives of miners, it is important to develop an effective method of mine gas concentration forecasting.

Therefore, in this paper, we have proposed a gas concentration forecasting algorithm to enable the prediction of gas concentration time series. The optimal number of training rounds to avoid overfitting was determined by considering the loss function on the training set. Case studies showed that with increasing data volume, compared with SVR, BPNN, RNN, and LSTM models, the proposed GRU model achieves both a better prediction effect and lower time complexity, endowing it with better practical application value.

In future work, we will consider the impact of various external factors, such as the temperature, humidity, and pressure in the coal mine, on mine gas forecasting and the incorporation of prior knowledge to achieve better performance.

REFERENCES

- [1] Z. Shining and L. Boquan, *The Theory of Gas Flow and Storage in Coal Seams*. Beijing, China Bureau of Coal Industry Press, 1999, p. 25.
- [2] Z. Zhao-zhao and Q. Jun-fei, "A brain-like multi-hierarchical modular neural network with applications to gas concentration forecasting," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 398–403.
- [3] S. Zhang, B. Wang, X. Li, and H. Chen, "Research and application of improved gas concentration prediction model based on grey theory and BP neural network in digital mine," *Procedia CIRP*, vol. 56, pp. 471–475, 2016.
- [4] R. Stopforth and S. Davrajh, "Gas concentration and equation correlation: Of the figaro sensors, used for dangerous environments," in *Proc. IEEE AFRICON*, Sep. 2017, pp. 1420–1425.
- [5] P. Booth, H. Brown, J. Nemcik, and R. Ting, "Spatial context in the calculation of gas emissions for underground coal mines," *Int. J. Mining Sci. Technol.*, vol. 27, no. 5, pp. 787–794, Sep. 2017.
- [6] G. Yu and L. Shi, "The gas concentration forecast based on RBF neural network and chaotic sequence," in *Proc. 2nd Int. Symp. Instrum. Meas., Sensor Netw. Autom. (IMSNA)*, Dec. 2013, pp. 745–748.
- [7] L. Kun, Y. Ling-Kai, Z. Mei-Ling, and C. Jian, "Coalmine gas concentration analysis based on support vector machine," in *Proc. 3rd Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, Jul. 2016, pp. 257–261.
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [9] H. A. F. Sak, "Long short term memory based recurrent neural network architectures for large vocabulary speech recognition," *Comput. Sci.*, vol. 13, pp. 338–342, Jan. 2014.
- [10] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM neural networks for Language modeling," in *Proc. Interspeech*, 2012, vol. 31, no. 43, pp. 601–608.
- [11] A.-R. Mohamed, T. N. Sainath, G. Dahl, B. Ramabhadran, G. E. Hinton, and M. A. Picheny, "Deep belief networks using discriminative features for phone recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 5060–5063.
- [12] R. Zbib, E. Malchiodi, J. Devlin, D. Stallard, S. Matsoukas, R. Schwartz, J. Makhoul, O. F. Zaidan, and C. Callison-Burch, "Machine translation of Arabic dialects," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2012, pp. 49–59.
- [13] C. Liu, C. Wang, F. Sun, and Y. Rui, "Image2Text: A multimodal image captioner," in *Proc. ACM Multimedia Conf. (MM)*, 2016, pp. 746–748.
- [14] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," *J. Mach. Learn. Res.*, vol. 37, pp. 2342–2350, Jul. 2015.
- [15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [16] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.

[17] L. Weishan, W. Lin, and W. Chen, "Application and design of LSTM in coal mine gas prediction and early warning system," *J. Xi'an Univ. Sci. Technol.*, vol. 38, no. 6, pp. 1027–1035, 2018.

[18] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1–15.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <https://arxiv.org/abs/1412.6980>

[20] W. Yichao and W. Jianwen, "Application exploration and practice of TensorFlow based on high performance computing platform," *Lab. Res. Explor.*, vol. 36, no. 12, pp. 125–128, 2017.

[21] X. Wang, R. Jiang, L. Li, Y. Lin, X. Zheng, and F.-Y. Wang, "Capturing car-following behaviors by deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 910–920, Mar. 2018.



SUJIAN WANG was born in Hancheng, Shaanxi, China. He received the B.S. degree in coal field geology and exploration from the Xi'an University of Mining and Technology, in 1984, and the M.B.A. degree from the Xi'an University of Technology, in 2003.

Since 2013, he has been the Executive Deputy Director of the National Local Joint Engineering Research Center for Green, Safe and Efficient Mining of Coal, the Director of the National Coal Industry's Green Coal Mining Engineering Research Center, and the Postmaster of the Shaanxi Coal Chemical Group Postdoctoral Workstation. He has published more than 30 scientific research articles and holds 15 utility model patents. His research interests include mine disaster prevention technology and management.



PENGTAO JIA was born in Pucheng, Shaanxi, China, in 1977. She received the B.S. and M.S. degrees in computer application technology from the Xi'an University of Science and Technology, China, in June 2002, and the Ph.D. degree in computer science and technology from Northwestern Polytechnical University, China, in June 2008.

Since 2015, she has been a Professor with the School of Computer Science and Technology, Xi'an University of Science and Technology. She

has published more than 20 scientific research articles and one monograph. She holds more than ten software copyrights and utility model patents. Her research interests include data mining, applications of artificial intelligence theory, and visualizations of coal mine safety data.

Dr. Jia has received the Shaanxi Science and Technology Award for Excellence.



HANGDUO LIU was born in Qianxian, Shaanxi, China, in 1994. He received the B.S. degree in electronic information engineering from Fuzhou University, in 2016. He is currently pursuing the M.S. degree in computer science and technology with the Xi'an University of Science and Technology.

His research interests include machine learning, recurrent neural networks, artificial intelligence, and the manufacture of intelligent robots. He has

received awards and honors from the Electronics Design Contest, artificial intelligence competitions, and the Innovation and Entrepreneurship Competition.



PENG WANG was born in Hanzhong, Shaanxi, China. He received the M.S. degree in safety technology and engineering from the Xi'an University of Science and Technology, in 2010.

From 2013 to 2019, he has served as the Director of the General Laboratory of Safe and Efficient Local Coal Mining of the National and Local Joint Engineering Research Center. He has been engaged in product research and development related to mine safety technology for many years.

He has applied for more than ten patents and published more than ten academic articles. His research interests include the development of coal mine disaster prevention and control equipment and the design of intelligent early warning algorithms.

...