frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fenrg.2022.1054427

OPEN ACCESS

EDITED BY
Zaibin Jiao,
Xi'an Jiaotong University, China

REVIEWED BY
Juntao Chang,
Harbin Institute of Technology, China
Zhengda Yang,
China University of Petroleum, Huadong, China

*CORRESPONDENCE
Bin Wu,
bin_wu0318@163.com

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 26 September 2022
ACCEPTED 31 October 2022
PUBLISHED 13 January 2023

CITATION
Yin Z, Yang C, Yuan X, Jin F and Wu B (2023), NOx concentration prediction in coal-fired power plant based on CNN-LSTM algorithm.
Front. Energy Res. 10:1054427.
doi: 10.3389/fenrg.2022.1054427

NOx concentration prediction in coal-fired power plant based on CNN-LSTM algorithm

Zhe Yin, Chunlai Yang, Xiaolei Yuan, Fei Jin and Bin Wu*

State Grid Hebei Energy Technology Service Co., Ltd., Shijiazhuang, China

Measuring the nitrogen oxides concentration accurately at the inlet of the selective catalytic reduction denitrification system plays an important role in controlling the nitrogen oxides concentration for coal-fired power plants, and a coupling relationship exists between nitrogen oxides concentration and multiple operational variables. Here, a modeling method based on feature fusion and long short-term memory network is proposed to mine the spatial and temporal coupling relationship between input variables for improving the prediction accuracy. First, the collected data were converted to image-like sequences. Then, the high-dimensional features of image-like sequences were fused by a convolutional neural network, and the spatial coupling features among the variables were mined. Finally, the constructed fusion features were input into the long short-term memory network to further explore the time coupling characteristics among the variables and complete the prediction of nitrogen oxides concentration at the inlet of the selective catalytic reduction denitrification system. The simulation results show that the prediction error of nitrogen oxides concentration at the inlet of selective catalytic reduction denitrification system based on CNN-LSTM model is 15.15% lower than that of traditional LSTM model.

KEYWORDS

coal-fired power plants, deep learning, feature fusion, Prediction, NOx concentration

Introduction

Carbon peaking and carbon neutrality goals can be realized by continuously increasing the installed capacity of renewable energy power in China, but thermal power units need to participate in long-term flexible regulation to curb the impact of renewable energy power on the stable operation of grids (Meysam et al., 2017; Kang et al., 2020; Shahbaz et al., 2020; Zhang et al., 2021; Tan et al., 2022). Conditions change frequently during the flexible operation of thermal power units (Zeng et al., 2019; Wang et al., 2020a), resulting in large changes in the NO_x concentration at the entrance of selective catalytic reduction (SCR) denitrification systems. This scenario hinders the effective control of denitrification systems of thermal power units. Establishing an accurate prediction model of NO_x concentration at the inlet of the SCR denitrification system can provide a model basis for the optimization of boiler combustion and the optimal control of the denitrification process, which has certain engineering significance (Adams et al., 2020; Huang et al., 2022).

In recent years, with the development of machine learning algorithms, data-driven modeling methods have shown a strong ability to deal with this problem. Adams et al. (2020) proposed a deep neural network model with a modified early stopping algorithm and least square support vector machine to predict NOx emission concentration. However, the traditional neural network model still has the problem of easily falling into local convergence, which causes its prediction accuracy to be limited (Tang et al., 2022).

Boiler combustion in a power station is a typical large inertia delay system, and a certain degree of lag manifests between the change in operating parameters and the change in NO_x concentration at the inlet of the SCR denitrification system (Xie et al., 2021). Therefore, the prediction of NO_x concentration at the inlet of the SCR denitrification system is a prediction problem based on a time series (Wang et al., 2020b). Long short-term memory (LSTM) neural networks can carry out long-term memory of data and have the unique advantage of temporal data processing. Tan et al. (2019), proposed a deep learning algorithm referred to as long short-term memory to predict the dynamics of NO_x emission in a 660 MW tangentially coal-fired boiler. Han et al. (2019) proposes a production capacity analysis and energy saving model using long short-term memory based on attention mechanism. Yang et al. (2020b)) focuses on the application of LSTM neural network with principal component analysis method in modeling the relationship between operational parameters and NOx emission of a 660 MW boiler. Furthermore, the LSTM model presented certain limitations. For instance, while the temporal characteristics of the data could be mined, it was not the case for input variables with interaction coupling relationships (Xie et al., 2020). Therefore, input features should be sufficiently constructed to establish LSTM network models that can predict the NO_x concentration at the inlet of the SCR denitrification system (Rumelhart et al., 1986; You et al., 2013).

The NO_x concentration at the inlet of the SCR denitrification system is affected by many factors, and the correlation is high among these variables (Lv et al., 2018; SaifUl Allah et al., 2022; Lv et al., 2022). When constructing the input features of the model based on prior knowledge, the complete features of the object cannot be guaranteed. Neurons of the convolutional neural network (CNN) have learnable weights and bias parameters, and specific feature mapping rules can be established *via* supervised learning while reducing the dimension of input feature variables (Yang et al., 2020a). Song et al. (2022) proposes a SCR inlet NOx concentration prediction algorithm based on BMIFS-LSTM to filter out the auxiliary variables with maximum correlation and minimum redundancy with NOx concentration. He et al. (2020) proposed a deep learning architecture formed by integrating CNN and LSTM with CNN layers extracting features among several variables and LSTM layers learning time series dependencies for predicting NOx emissions However, the temporal characteristics of the data were destroyed in the process of using CNN or other feature processing algorithms for feature extraction, and errors appeared in the subsequent process of using the LSTM network to learn fused features. In this study, the class image structure is introduced and then the sequence folding and expansion operations are used to improve the feature extraction process. In this manner, the time sequence invariance of the data in the process of feature fusion can be ensured.

Therefore, the paper focused on an algorithm based on feature fusion and the LSTM network. In this algorithm, CNN is used as the feature fusion unit to extract the coupling interaction features among many variables related to NO_x concentration. Then, the LSTM network structure is used to learn the time series features. This method involves an explicit learning process of identifying the variable coupling relations, and more directly, the mining of variable interaction information, which is helpful in further improving the prediction accuracy of the model. The results show that the CNN–LSTM model has a higher accuracy for predicting NO_x concentration at the inlet of the SCR denitrification system compared with the traditional LSTM network prediction model.

High-dimensional feature fusion based on convolutional neural network

Basic theory of convolutional neural network

CNN is a kind of neural network that uses convolution and has a multilayer structure, which mainly includes a convolutional layer, a canonical layer, and a fully connected layer. Neurons in CNN have the characteristics of local connection and weight sharing, enabling the network to compile specific feature mapping rules into the convolution structure. Thereafter, feature dimension reduction can be achieved, and the fusion of high-dimensional input features can be completed in the modeling process of NO_x concentration at the entrance of the SCR denitrification system.

The convolutional layer network in CNN is composed of several convolution kernels $w \in R^{d\times d}$ with the same size, where d is the size of the convolution kernel. The parameters in the convolution kernel correspond to the weights of CNN neurons. The features are generated by sliding the convolution kernel on the input data $\tilde{x} \in R^{H\times H}$, where H is the size of the input data. The mathematical expression of the CNN's network input $\tilde{x}$ and convolution kernel w operation is shown in Eq. 1

TABLE 1 Arrangement of type image data.

Unit load	Burnout damper opening A	Secondary damper opening degree A	Perimeter damper opening A	Furnace temperature
Total air volume	Burnout damper opening B	Secondary damper opening degree B	Perimeter damper opening B	Flue gas temperature of air preheater
For coal	Burnout damper opening C	Secondary damper opening degree C	Perimeter damper opening C	Primary air temperature at air preheater outlet
Main feedwater temperature	Burnout damper opening D	Secondary damper opening degree D	Perimeter damper opening D	Secondary air temperature at air preheater outlet
Main steam pressure	Burnout damper opening E	Secondary damper opening degree E	Perimeter damper opening E	NOx concentration at SCR inlet (last moment)

$$P_{i,j} = f\left(\sum_{u=1}^{d}\sum_{v=1}^{d} w_{u,v} \odot \tilde{x}_{i+u-1,j+v-1}{}' + b\right) \tag{1}$$

where $P_{i,j}$ is the element in the output feature, $\tilde{x}_{i+u-1,j+v-1}$ is the input the $\tilde{x}$ element, $w_{u,v}$ is the weight parameter w in the convolution kernel, b is the bias parameter, and f is a linear rectification function (ReLu).

Input feature extraction of the NO_x prediction model

According to the operation of the unit, 25 indices (i.e., load, total air volume, and coal supply) are selected as input variables, whereas NO_x concentration at the inlet of the SCR denitrification system is selected as the output variable. Assume that the sample collected is $x \in R^p$, where p is the number of input variables ($p = 25$), and the NO_x concentration at the inlet of the SCR denitrification system corresponding to its time is y. The collected sample data x is converted to class image data $\tilde{x} \in R^{H\times H}$, where $H = 5$. The arrangement of the class image data is shown in Table 1. Compared with the data before processing, the kind of image data presented above has additional spatial features that can be easily fused by CNN.

After the class image data are input into the CNN, their respective features $\tilde{x}$ are extracted by a convolution operation, as shown in Eq. 1. The convolution layer $\tilde{x}$ contains several different convolution kernels, each representing different feature rules. Features P can be obtained after the convolution operation. Furthermore, as the CNN contains multiple convolutional layers, the obtained features P are sent as input to the next convolutional layer for convolution operation. Passing through several convolutional layers causes the data dimension to be continuously reduced and the final features $\tilde{P}$ to be fused. With $\tilde{P}$ as the input of the subsequent model, the CNN subsequently participates in the training together with the prediction model and realizes the fusion of the input features. The CNN's process of fusing the high-dimensional input features is shown in Figure 1.

Establishment of the NOx concentration prediction model based on the long short-term memory network

Long short-term memory modeling theory

The special structure of the LSTM network lies in the introduction of a gating mechanism to control the learning degree of the time series data, allowing the network to retain historical information for a long time. The logical unit in the LSTM neural network is shown in Figure 2.

Each logic unit has three types of gates: a forget gate, an input gate, and an output gate. The forgetting gate uses the activation function to conditionally control and select the memorized information and subsequently obtain the information needed by the logic unit. The input gate also applies this process to the new input information to retain the valid information in the input. The whole process is given by Eq. 2

$$\begin{cases} \boldsymbol{f}_t = \sigma(\boldsymbol{W}_{\mathrm{f}} \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{S}_t] + \boldsymbol{b}_{\mathrm{f}}) \\ \boldsymbol{i}_t = \sigma(\boldsymbol{W}_{\mathrm{i}} \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{S}_t] + \boldsymbol{b}_{\mathrm{i}}) \\ \boldsymbol{m}_t = \tanh(\boldsymbol{W}_{\mathrm{c}} \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{S}_t] + \boldsymbol{b}_c) \\ \boldsymbol{o}_t = \sigma(\boldsymbol{W}_{\mathrm{o}} \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{S}_t] + \boldsymbol{b}_o) \\ \boldsymbol{c}_t = \boldsymbol{f}_t \odot c_{t-1} + \boldsymbol{S}_t \odot \boldsymbol{m}_t \\ \boldsymbol{h}_t = \boldsymbol{o}_t \odot \tanh\ (\boldsymbol{c}_t) \end{cases} \tag{2}$$

where $\sigma(\cdot)$ is an activation function, W and b are the weight matrix and corresponding deviation of each gate, f_t is the forgetting gate, i_t is the input gate, o_t is the output gate, c_{t-1} is the state value at the last moment, c_t is the state value at the current moment, h_{t-1} is the input at the last moment, h_t is the output at the current moment, S_t is the input at the

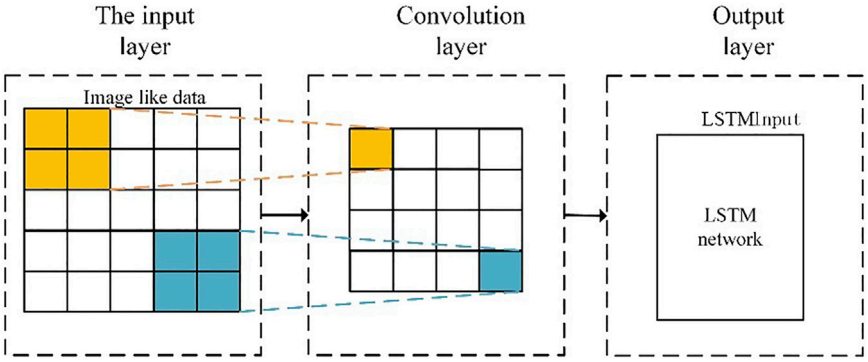

FIGURE 1
Schematic diagram of high-dimensional feature fusion based on CNN.

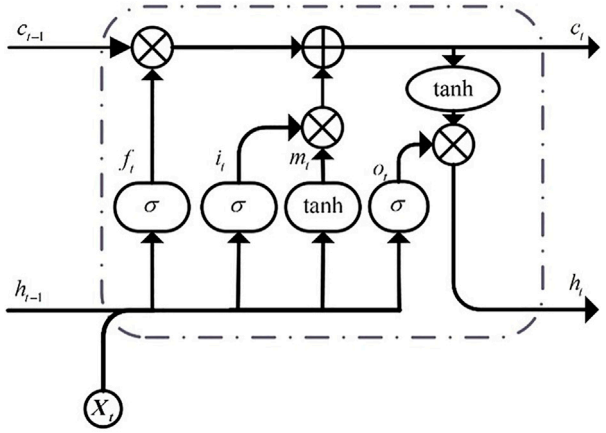

FIGURE 2
Structural diagram of the LSTM unit.

current time, K is the sequence data of length, and ⊙ is for Hadama.

Construction of the NO_x prediction model

Most variables related to NO_x concentration at the inlet of the SCR denitrification system have delay characteristics, and the LSTM network is used to introduce delay information of the input variables in the modeling process (Lv et al., 2020). The LSTM network performs effectively in obtaining the long-term correlation of variable series, but it cannot establish an effective model for discontinuous data. Therefore, a sequence folding layer and a sequence unfolding layer are added to the neural network. Then, multiple groups of continuous-time image-like data are taken as the input of the CNN. In this manner, the temporal features of the data can be preserved while fusing the high-dimensional features, further ensuring the effectiveness of the model input.

The elements in the initial input sample set $\{x_1, x_2, \cdots, x_t, \cdots, x_n\}$ are transformed into image-like data $\{\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_t, \cdots, \tilde{x}_n\}$ by the arrangement shown in Table 1 and then input into the CNN. After performing several convolution operations as described in Eq. 1, the sample fusion feature set $\{\tilde{P}_1, \tilde{P}_2, \cdots, \tilde{P}_t, \cdots, \tilde{P}_n\}$ can be obtained. For the sample fusion feature set, the input of the LSTM model $\{S_1, S_2, \cdots, S_t, \cdots, S_{n-K-1}\}$ is constructed by the sliding window. The method of constructing the input–output structure by the sliding window is shown in Figure 3. The relationship between S_t and sample fusion features is shown in Eq. 3.

$$\boldsymbol{S}_t = [\tilde{\boldsymbol{P}}_t, \tilde{\boldsymbol{P}}_{t+1}, \cdots, \tilde{\boldsymbol{P}}_{t+K-1}]^{\mathrm{T}} \tag{3}$$

where K is the input sequence length of the LSTM model.

After the fusion feature sample sequence S_t is input to the LSTM network model, Equation 2 model is calculated to predict the NO_x concentration of SCR denitration system entry. Then, the corresponding real output for the moment of the sample sequence of the SCR denitration system's inlet corresponding to the next-moment NO_x concentration y_{t+K} are determined to build the LSTM training model.

The mathematical expression of the NO_x concentration prediction model for the inlet of the SCR denitrification system is shown in Eq. 4.

$$y_{t+1} = F(\boldsymbol{x}_t, \boldsymbol{x}_{t-1} \ldots \boldsymbol{x}_{t-K+1}) \tag{4}$$

where F is the constructed CNN–LSTM network model.

The CNN–LSTM network's structural design and model hyperparameter setting are shown in Table 2. K is the step size in the process of convolution kernel movement in the CNN, and *Drop* is the neuron drop rate in each layer of the LSTM network (Lv et al., 2015).

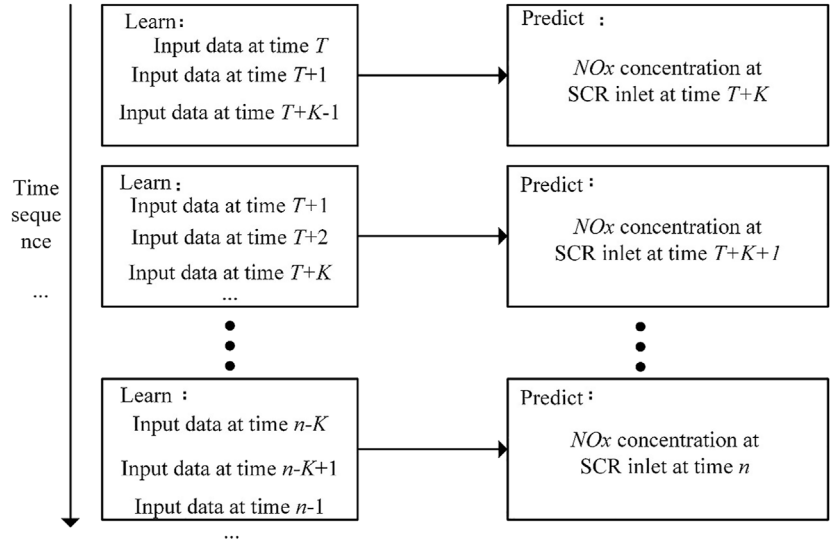

FIGURE 3
Structure of input and output in sliding window.

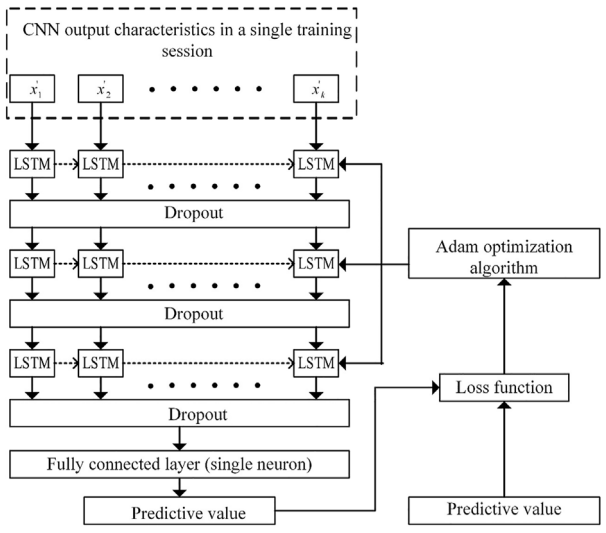

FIGURE 4
Structural diagram of the LSTM network model based on CNN.

Figure 4 shows the structural diagram of the CNN–LSTM network model. CNN is used to merge the input features. Then, a three-layer LSTM neural network is used to predict the NO_x concentration at the entrance of the SCR denitrification system. At the same time, the "dropout" layer is used in each layer of the LSTM network to reduce the interaction between the hidden layer nodes during training, allowing the generalization ability of the model to be enhanced.

Model valuation criteria

The prediction accuracy of the model was evaluated more accurately by introducing the root mean square error (RMSE) and normalized root mean square error (NRMSE). The expressions of RMSE and NRMSE are shown in Eqs 5, 6:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_t - y_t)^2} \tag{5}$$

$$\mathrm{NRMSE} = \frac{1}{\bar{y}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_t - y_t)^2} \times 100\% \tag{6}$$

where y_t represents the actual measured value, $\hat{y}_t$ represents the predicted value of the model, and $\bar{y}$ represents the average value of the actual measured value

Data acquisition and model construction

In this study, a 300 MW coal-fired thermal power unit was taken as the research object. The supporting boiler of the unit is a naturally circulating single furnace with a π-type arrangement and solid slag discharge boiler adopting the four-angle cut-round combustion method and positive-pressure direct-blowing pulverizing system. The SCR denitrification system adopts the high-temperature and high-ash arrangement in the flue after the furnace operation. The system consists of two parts: a catalytic reaction zone and an ammonia zone. It is also composed of a

TABLE 2 Model structure and hyper parameter setting.

Network layer structure	Network layer name	Parameter
Sequential input layer	Input	Input data dimension (5,5,1)
Sequence folding layer	Fold	No
Two-dimensional convolutional layer	Conv1	d = 2, q = 1, $Stride$ = 1
Specification layer	Batchnorm1	No
The activation layer	ReLu1	No
Two-dimensional convolutional layer	Conv2	d = 2, q = 1, $Stride$ = 1
Specification layer	Batchnorm2	No
The activation layer	ReLu2	No
Sequence unfolding layer	Unfold	No
Flat layer	Flatten	No
LSTM layer	Hidden1	Number of hidden layers 200
School layer	Dropout1	$Drop$ = 0.6
LSTM layer	Hidden2	Number of hidden layers 240
School layer	Dropout2	$Drop$ = 0.6
LSTM layer	Hidden3	Number of hidden layers 280
School layer	Dropout3	$Drop$ = 0.6
Connection layer	Full connect	Number of hidden layers 1

2+1-layer cellular catalyst, dilution fan, ammonia spraying grid, and soot blower.

Data samples of the abovementioned variables were collected from the Supervisory Information System's operation database at the factory level over a period of 1 week, with a sampling interval of 30 s. These variables include three categories: basic operating parameters of the boiler (such as unit load, coal feed, etc.), which can represent the current operating conditions of the unit; Operating parameters and sensor measurement parameters that affect or reflect furnace flame and boiler outlet flue gas temperature (such as secondary air valve opening, air preheater flue gas temperature, etc.), these variables can characterize the combustion state in the furnace, which also affects the generation of NO_x to a considerable extent; NO_x concentration at SCR inlet at the last moment, this variable is mainly used to improve the prediction accuracy of the model by using the characteristics of LSTM network. According to field experience, the reaction process inside the boiler is usually completed within 6 min. Thus, the length of the input sequence of the LSTM model is set as K = 12. The data were preprocessed (i.e., outlier processing and normalization processing), and 41,500 sets of data samples were obtained during the continuous operation. The data timing data were kept to the first 40,000 sets of the data after processing, and they were used as the training set to establish the NO_x concentration prediction model for the inlet of the SCR denitrification system. Meanwhile, the last 1,500 sets of data were used as the test set to verify the generalization ability of the prediction model.

The mini-batch gradient descent strategy was adopted in the training of the network model. The number of samples in each mini-batch was 192, and the data were completely traversed for 10 rounds in the training. The mean square error (MSE) was used as the loss function. The expression of MSE is shown in Eq. 7:2

$$e_{mse} = \frac{1}{n}\sum_{i=1}^{n}\left(y_t - \hat{y}_t\right)^2 \tag{7}$$

where n is the number of samples, y_t is the actual measured value, and $\hat{y}_t$ is the predicted value.

Experimental results and analysis

Model training results

The generalization ability of CNN–LSTM model was determined and then compared with that of the constructed traditional LSTM model. The same input samples were used to construct the prediction model. Figure 5 shows the training results of the prediction model of NO_x concentration at the inlet of the SCR denitrification system based on CNN–LSTM network and traditional LSTM network. The predicted RMSE and NRMSE of the CNN-LSTM model used in this paper for training set samples are 1.66 mg/m^3 and 5.3%, compared with 1.92 mg/m^3 and 6.2% of the traditional LSTM model, and the fitting ability has been improved by 13.54%. The CNN-LSTM model has a higher fitting ability to the NOx concentration at the inlet of SCR denitrification system, which shows that the processing of input variables by CNN network is conducive to improving the overall expression ability of the model.

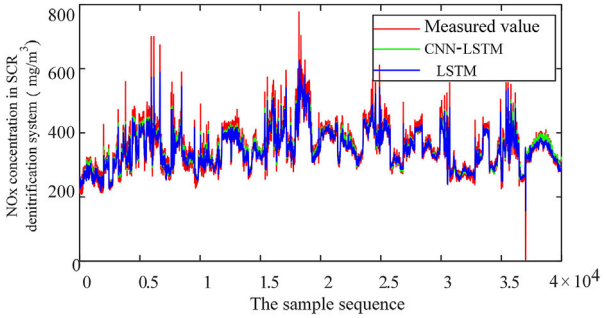

FIGURE 5
Comparison of training results between the CNN–LSTM model and traditional LSTM model.

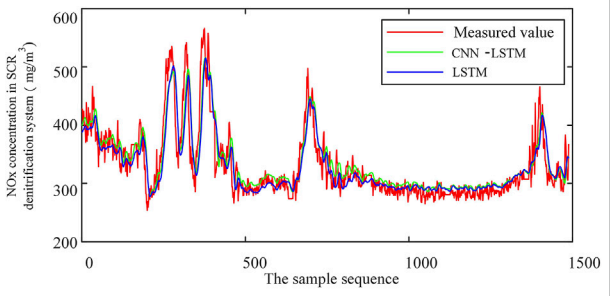

FIGURE 6
Comparison of prediction results between the CNN–LSTM model and traditional LSTM model.

TABLE 3 Comparison of prediction errors between the CNN–LSTM model and traditional LSTM model.

Model	Training set		Test set	
	RMSE (mg/m^3)	NRMSE (%)	RMSE (mg/m^3)	NRMSE (%)
LSTM	1.92	6.2	2.46	7.9
CNN–LSTM	1.66	5.3	2.09	6.7

Figure 6 shows the prediction results of the CNN–LSTM model and traditional LSTM model on the test set, and Table 3 presents the prediction errors of the model. The predicted value of the CNN-LSTM model is highly consistent with the actual measured value, and the maximum deviation does not exceed 3.82 mg/m^3.

The RMSE and NRMSE of the CNN–LSTM model on the training and test set samples are smaller than those of the traditional LSTM model. For the test set samples, the RMSE and NRMSE predicted by the CNN–LSTM model are 2.09 mg/m^3 and 6.7%, respectively; by contrast, the corresponding values of the traditional LSTM model are 2.46 mg/m^3 and 7.9%, indicating a reduction in prediction error by 15.15%. The traditional LSTM model and CNN–LSTM model both have high approximation ability, but the generalization ability of the latter is better than that of the former. The difference can be attributed to the eliminated redundant information after the CNN has integrated the high-dimensional features, allowing the generalization ability of the model to be improved.

Comparative test

The necessity of integrating high-dimensional features based on CNN was further verified by comparing the established

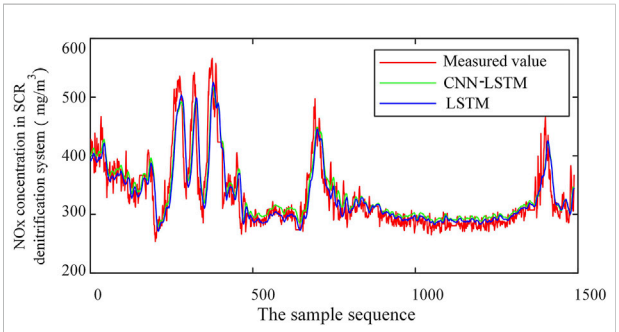

FIGURE 7
Comparison of prediction results between the CNN–LSTM and LSTM model.

CNN–LSTM model with the LSTM model that was established by empirically selecting ten features as the input for the same dataset (Han et al., 2019). The characteristics include unit load, coal feed, total air volume, four-layer secondary damper opening, two-layer burnout damper opening, and inlet NO_x concentration of the SCR denitrification system over historical time.

According to the analysis of the prediction accuracy of the model shown in Figure 7, the RMSE and NRMSE of the aforementioned LSTM model are 2.92 mg/m^3 and 9.0%,

respectively. Comparing this result with the error data in Table 3, it can be found that the LSTM model established after manually filtering variables has obviously insufficient prediction accuracy due to filtering out too much information. From the setting of CNN network parameters in the CNN-LSTM model structure provided in Table 2 and the convolution calculation process of Eq. 1, it can be seen that the CNN-LSTM model constructed in this paper only has 9-dimensional data input into the LSTM network. However, compared with the LSTM model with 10 variables selected by experience as input, the prediction accuracy has been improved by 28.42%, which shows that the input variable processing method provided in this paper is of reference value for LSTM network model.

Conclusion

A prediction method for determining nitrogen oxides concentration at the inlet of the selective catalytic reduction denitrification system based on feature fusion and deep learning has been proposed. On the basis of the collected data, the deep features in the high-dimensional data were extracted by the convolutional neural network, then the deep features were modeled by the long short-term memory network to predict the nitrogen oxides concentration at the inlet of the selective catalytic reduction denitrification system, and the CNN–LSTM model was established. The CNN–LSTM model used in this study can further improve the prediction accuracy of the model with respect to the traditional long short-term memory network model. The root mean square error and normalized root mean square error of the CNN–LSTM model on the test set were 2.09 mg/m^3 and 6.7%, respectively. These results indicate the shortcomings of the prediction model of nitrogen oxides concentration at the inlet of the selective catalytic reduction denitrification system based on the manual selection of auxiliary variables. The feature fusion method based on the convolutional neural network enables the long short-term memory network model to be more generalized.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The data set in this paper is collected from the SIS system of a 300 MW coal fired thermal generator set in China. Requests to access these datasets should be directed to BW, bin_wu0318@163.com.

Author contributions

The author's personal contributions are as follows: manuscript writing, XL; data collection, CY; and model construction, ZY and XY; Visualization, FJ. All authors have read and agreed to the published version of the manuscript.

Funding

This study was funding from the Science and technology project of Hebei Electric Power Co., Ltd. (TSS2020-11). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

ZY, CY, XY, FJ, and BW were employed by State Grid Hebei Energy Technology Service Co., Ltd.

Publisher's note

References

Adams, D., Oh, D. H., Kim, D. W., Lee, C. H., and Oh, M. (2020). Prediction of SOx-NOx emission from a coal-fired CFB power plant with machine learning: Plant data learned by deep neural network and least square support vector machine. *J. Clean. Prod.* 270, 122310. doi:10.1016/j.jclepro.2020.122310

Han, Y., Fan, C., Xu, M., Geng, Z., and Zhong, Y. (2019). Production capacity analysis and energy saving of complex chemical processes using LSTM based on attention mechanism. *Appl. Therm. Eng.* 160, 114072. doi:10.1016/j.applthermaleng.2019.114072

He, Wei, Li, Jufeng, Tang, Zhihe, Wu, Beng, Hui, Luan, Chen, Chong, et al. (2020). A novel hybrid CNN-LSTM scheme for nitrogen oxide emission prediction in FCC unit. *Math. Problems Eng.* 12, 1–12. doi:10.1155/2020/8071810

Huang, D., Tang, S., Zhou, D., and Hao, J. (2022). NOx emission estimation in gas turbines via interpretable neural network observer with adjustable intermediate layer considering ambient and boundary conditions. *Measurement* 189, 110429. doi:10.1016/j.measurement.2021.110429

Ioffe, S., and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proc. 32nd Int. Conf. Inter- Natl. Conf. Mach. Learn.* 37, 448–456.

Kang, J.-N., Wei, Y.-M., Liu, L.-C., Han, R., Yu, B.-Y., and Wang, J.-W. (2020). Energy systems for climate change mitigation: A systematic review. *Appl. Energy* 263, 114602. doi:10.1016/j.apenergy.2020.114602

Lv, M. L., Zhao, J., Cao, S., and Shen, T. (2022). Prediction of the 3D distribution of NOx in a furnace via CFD data based on ELM. *Front. Energy Res.* 10, 848209. doi:10.3389/fenrg.2022.848209

Lv, Y., Fang, F., Yang, T., and Romero, C. E. (2020). An early fault detection method for induced draft fans based on MSET with informative memory matrix selection. *ISA Trans.* 102, 325–334. doi:10.1016/j.isatra.2020.02.018

Lv, Y., Romero, C. E., Yang, T., Fang, F., and Liu, J. (2018). Typical condition library construction for the development of data-driven models in power plants. *Appl. Therm. Eng.* 143, 160–171. doi:10.1016/j.applthermaleng.2018.07.083

Lv, Y., Yang, T., and Liu, J. (2015). An adaptive least squares support vector machine model with a novel update for NOx emission prediction. *Chemom. Intelligent Laboratory Syst.* 145, 103–113. doi:10.1016/j.chemolab.2015.04.006

Meysam, Q., Hossein, A., Goran, S., and Nicholas, J. (2017). Efficacy of options to address balancing challenges: Integrated gas and electricity perspectives. *Appl. Energy* 190, 181–190. doi:10.1016/j.apenergy.2016.11.119

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323, 533–536. doi:10.1038/323533a0

Saif Ul Allah, M. W., Khan, J., Ahmed, F., Salman, C. A., Gillani, Z., Hussain, A., et al. (2022). Computationally inexpensive 1D-CNN for the prediction of noisy data of NOx emissions from 500 MW coal-fired power plant. *Front. Energy Res.* 10, 945769. doi:10.3389/fenrg.2022.945769

Shahbaz, M., Raghutla, C., Chittedi, K. R., Jiao, Z., and Vo, X. V. (2020). The effect of renewable energy consumption on economic growth: Evidence from the renewable energy country attractive index. *Energy* 207, 118162. doi:10.1016/j.energy.2020.118162

Song, M., Xue, J., Gao, S., Cheng, G., Chen, J., Lu, H., et al. (2022). Prediction of NOx concentration at SCR inlet based on BMIFS-LSTM. *Atmosphere* 13, 686. doi:10.3390/atmos13050686

Tan, P., He, B., Zhang, C., Rao, D., Li, S., Fang, Q., et al. (2019). Dynamic modeling of NOx emission in a 660 MW coal-fired boiler with long short-term memory. *Energy* 176, 429–436. doi:10.1016/j.energy.2019.04.020

Tan, P., Zhu, H., He, Z., Jin, Z., Zhang, C., Fang, Q., et al. (2022). Multi-step ahead prediction of reheat steam temperature of a 660 MW coal-fired utility boiler using long short-term memory. *Front. Energy Res.* 10, 845328. doi:10.3389/fenrg.2022.845328

Tang, Z., Wang, S., Chai, X., Cao, S., Ouyang, T., and Li, Y. (2022). Auto-encoder-extreme learning machine model for boiler NOx emission concentration prediction. *Energy* 256, 124552. doi:10.1016/j.energy.2022.124552

Wang, C., Qiao, Y., Liu, M., Zhao, Y., and Yan, J. (2020a). Enhancing peak shaving capability by optimizing reheat-steam temperature control of a double-reheat boiler. *Appl. Energy* 260, 114341. doi:10.1016/j.apenergy.2019.114341

Wang, G., Awad, O., Liu, S., Shuai, S., and Wang, Z. (2020b). NOx emissions prediction based on mutual information and back propagation neural network using correlation quantitative analysis. *Energy* 198, 117286. doi:10.1016/j.energy.2020.117286

Xie, P., Gao, M., Zhang, H., Niu, Y., and Wang, X. (2020). Dynamic modeling for NOx emission sequence prediction of SCR system outlet based on sequence to sequence long short-term memory network. *Energy* 190, 116482. doi:10.1016/j.energy.2019.116482

Xie, P. R., Zhang, G., Niu, Y., and Sun, T. (2021). Selective catalytic reduction system Ammonia injection control based on deep deterministic policy reinforcement learning. *Front. Energy Res.* 9. doi:10.3389/fenrg.2021.725353

Yang, G. T., Wang, Y. N., and Li, X. L. (2020b). Prediction of the NO emissions from thermal power plant using long-short term memory neural network. *Energy* 192, 116597. doi:10.1016/j.energy.2019.116597

Yang, T., Ma, K., Lv, Y., and Bai, Y. (2020a). Real-time dynamic prediction model of NOx emission of coal-fired boilers under variable load conditions. *Fuel* 274, 117811. doi:10.1016/j.fuel.2020.117811

You, L., Jizhen, L., Tingting, Y., and Weiyi, S. (2013). NOx emission characteristic modeling based on feature extraction using PLS and LS-SVM. *Chin. J. Sci. Instrum.* 34, 2418–2424.

Zeng, D., Gao, Y., Hu, Y., and Liu, J. (2019). Optimization control for the coordinated system of an ultra-supercritical unit based on stair-like predictive control algorithm. *Control Eng. Pract.* 82, 185–200. doi:10.1016/j.conengprac.2018.10.001

Zhang, G. M., Xie, P., Huang, S., Chen, Z., Du, M., Tang, N., et al. (2021). Modeling and optimization of integrated energy system for renewable power penetration considering carbon and pollutant reduction systems. *Front. Energy Res.* 9, 767277. doi:10.3389/fenrg.2021.767277

Glossary

SCR selective-catalytic-reduction
RBF radial basis function
LSTM long short term memory
CNN convolutional-neural-networks
SIS supervisory information system
RMSE root mean square error
NRMSE normalized root mean square error
Symbols
W several convolution kernels
D size of the convolution kernel
H size of the input data
$P_{i,j}$ element in the output feature
$\tilde{x}_{i+u-1,j+v-1}$ input element
$w_{u,v}$ the weight parameter w
b bias parameter
f linear rectification function
P Features
$\tilde{P}$ final features
$\sigma(\cdot)$ activation function
W weight matrix
b corresponding deviation
f_t forgetting gate
i_t input gate
o_t output gate
c_t state value at the current moment
h_t output at the current moment
S_t input at the current time
K sequence data of length
y_t actual measured value
$\hat{y}_t$ predicted value