ISBN: 978-981-18-7950-0
2023 the 13th International Workshop on Computer Science and Engineering (WCSE 2023)
doi: 10.18178/wcse.2023.06.051

Dual-channel Feature Fusion Model for Air Pollutants Forecast

Yujie Li, Xu Qiao, Zhiping Liu and Rui Gao +

School of Control Science and Engineering, Shandong University, Jinan, China

Abstract. Air pollution is severely injurious to human health, and hence short-term prediction for various air pollutants concentrations is beneficial for early prevention and travel planning. To reduce the contradictory information between time trend and variable correlation when predicting, the dual-channel feature fusion model is proposed. The model first obtains the time-series characteristic through the temporal feature extraction module based on TPA-LSTM, then captures the variable relevance with the structural feature extraction module aggregated GNN and LSTM, and finally integrates the two features by a fusion gate to realize forecast. Comparative experiments on air pollutants dataset from the state-controlled air station in Jinan by different models reveal that our method demonstrates superior properties.

Keywords: Air pollutants forecast, Multivariate time-series prediction, Feature fusion, LSTM, GNN

1. Introduction

Air quality is one of the important environmental factors that affect human life and health. It depends on the concentrations of various pollutants (NO, PM2.5, PM10, CO, SO_2, O_3, etc.). WHO estimates that 4.2 million people die prematurely each year from exposure to outdoor (or ambient) air pollution. Therefore, air pollutants forecast in the short term can provide reliable information on future weather pollution, which helps governments to develop preventive measures and assists residents in formulating travel plans [1-2].

Air pollutants forecast is an application of multivariate time series forecasting. Currently, the most prevalent method for it is Long Short-Term Memory network (LSTM) and Gate Recurrent Unit network (GRU). Y. Tsai et al. [3] exploited RNN and LSTM to make short-term prediction of PM2.5 concentration in the next 4 hours. J. Zhang et al. [4] proposed a CNN-LSTM model to predict future values by learning the changing laws of air quality data. Besides, the TPA-LSTM model [5] has become a favourable tool for multivariate time series forecasting. The concentrations of air pollutants not only depend on the change trend information contained in historical data, but also rely upon the comprehensive influence of variables such as meteorological environment and other pollutant concentrations. However, LSTM&GRU and its variant network neglect the correlations among variables.

To sketch the relevance between variables, the graph method is employed as a powerful mathematical tool. S. Wang et al. [6] utilized external environmental variables such as weather conditions, temperature, and wind direction to develop a graph-based model, which could explore the long-term dependence of these variables on PM2.5 and realize predictions. But the graph method was limited by the predefined graph structure, so Z. Wu et al. [7] proposed an adaptive graph learning network (MTGNN). M. Jin et al. [8] used dynamic graph god ordinary differential equation (MTGODE) to further lower algorithm complexity. However, the graph method pays more attention to the influence between different nodes, and the extracted features contain less historical time information.

In this paper, a dual-channel feature fusion model is presented to adequately obtain potential information from multivariate time series of air pollutants concentrations and reduce the contradictory information between time trend and variable correlation. The temporal feature extraction module based on TPA-LSTM is used to extract historical change features in the time direction. The structural feature extraction module integrated graph convolutional network and the long-short-term memory network is applied to conducting the structural correlations among variables. The feature fusion module is proposed for combining two

+ Corresponding author. Tel.: + 86 13505318418.
E-mail address: gaorui@sdu.edu.cn.

features to predict the concentrations of various air pollutants at the same time. Finally, the data from the state-controlled air station in Jinan is used to demonstrate the superiority of the proposed model through model comparison.

2. Dual-channel Feature Fusion Model

2.1. Problem statement

Consider a multivariate time series $\mathbf{Z}^{(t)} = (\mathbf{Z}_{t-T+1}, \mathbf{Z}_{t-T+2}, \cdots \mathbf{Z}_t)^T \in \mathbb{R}^{T\times N}$, where T denotes the length of the desired historical time window, and N is the feature number. The element $\mathbf{Z}_k = (\mathbf{X}_k^P, \mathbf{Y}_k^{N-P})^T = (x_k^1, x_k^2, \cdots x_k^P, y_k^1, y_k^2, \cdots y_k^{N-P})^T \in \mathbb{R}^{N\times 1} (t-T+1 \le k \le t)$ represents the multivariate variables at time step k, where $\mathbf{Y}_t^{N-P} = (y_t^1, y_t^2, \cdots y_t^{N-P})$ denotes the target variables, and $\mathbf{X}_t^P = (x_t^1, x_t^2, \cdots x_t^P)$ is other relevant variables that influence the target set. We define N as the node number in the graph, and T as the feature number of the node. This paper aims to use $\mathbf{Z}^{(t)}$ to forecast $\hat{\mathbf{Y}}_{t+\tau} = (y_{t+\tau}^1, y_{t+\tau}^2, \cdots y_{t+\tau}^{N-P})$, where τ is the time step to be predicted.

2.2. Model framework

The overall model framework of the developed dual-channel feature fusion model is shown in Fig. 1. The three modules of this model are as follows: (1) temporal feature extraction module: combined with the CNN network, it weights the temporal patterns abstracted by the LSTM network, thereby increasing the correlation between the historical time change trend characteristics and the current input. (2) structural feature extraction module: it introduces graph structure learning and graph neural network based on the MTGNN network and aggregates LSTM to make the network pay more attention to the deep information of other vectors that affect the target vector. (3) feature fusion module: it amalgamates the two characteristics with the fusion gating mechanism to improve the efficiency of information mining.

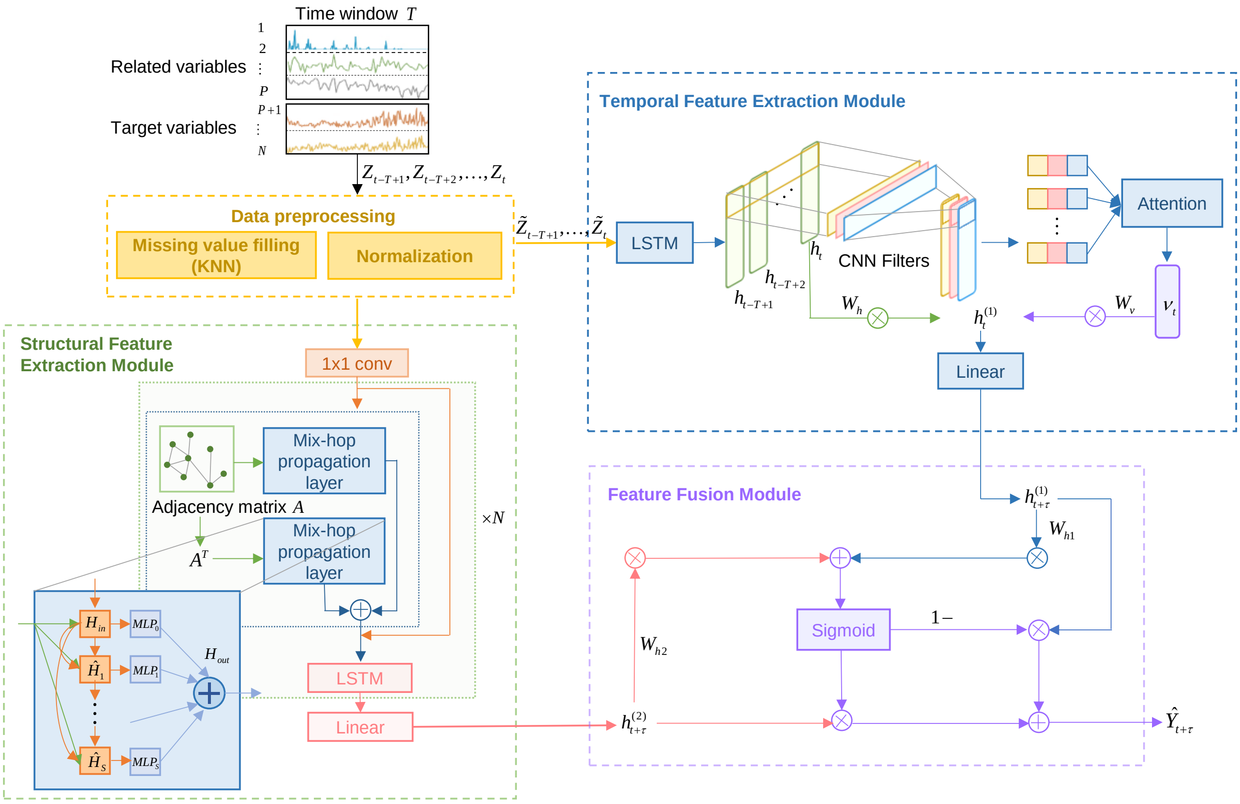

Fig. 1: The overall framework of the proposed model.

2.3. Temporal feature extraction module

The temporal feature extraction module is based on TPA-LSTM network [5]. Given a multivariate time series $\mathbf{Z}^{(t)} = (\mathbf{Z}_{t-T+1}, \mathbf{Z}_{t-T+2}, \cdots \mathbf{Z}_t)^T \in \mathbb{R}^{T\times N}$, the hidden state $\mathbf{H} = (\mathbf{h}_{t-T+1}, \mathbf{h}_{t-T+2}, \cdots, \mathbf{h}_t) \in \mathbb{R}^{M\times T}$ is obtained after the LSTM network, which is defined as

$$\mathbf{h}_k, \mathbf{c}_k = LSTM(\mathbf{h}_{k-1}, \mathbf{c}_{k-1}, \mathbf{Z}^{(k)}), \tag{1}$$

where $t-T+1 \le k \le t$. To improve the representation ability of the hidden state, $\mathbf{H}^C \in \mathbb{R}^{M\times k}$ is gained by using k CNN filters $\mathbf{C}_i \in \mathbb{R}^{1\times T}$. The attention scores for each row vector of $\mathbf{H}^C$ are calculated to obtain the content vector $\mathbf{v}_t \in \mathbb{R}^k$, which combines $\mathbf{h}_t$ to get the final timing feature. This operation is given by:

$$f\left(\mathbf{H}_i^C, \mathbf{h}_t\right) = \left(\mathbf{H}_i^C\right)^T \mathbf{W}_a \mathbf{h}_t, \tag{2}$$

$$\mathbf{v}_t = \sum_{i=1}^{M} sigmoid(f(\mathbf{H}_i^C, \mathbf{h}_t)) * \mathbf{H}_i^C, \tag{3}$$

$$\mathbf{h}_{t+\tau}^{(1)} = \mathbf{W}_y \mathbf{h}_t^{(1)} = \mathbf{W}_y(\mathbf{W}_h \mathbf{h}_t + \mathbf{W}_v \mathbf{v}_t), \tag{4}$$

where $\mathbf{W}_a$, $\mathbf{W}_h$, $\mathbf{W}_v$ and $\mathbf{W}_y$ are learnable parameters.

2.4. Structural Feature Extraction Module

The GNN module is imported from the MTGNN framework [7], which aggregates LSTM model to capture correlations between variables. Considering multivariate time series $\mathbf{Z}^{(t)} = (\mathbf{z}_1^t, \mathbf{z}_2^t, \cdots \mathbf{z}_N^t) \in \mathbb{R}^{T\times N}$, we assume $\mathbf{z}_i^t \in \mathbb{R}^{T\times 1}$ as the i-th node of the graph. The graph adjacency matrix $\mathbf{A}$ is independently learned through the graph learning layer to describe the correlation between variables illustrated as follows:

$$\mathbf{A} = \mathrm{Re}\,LU(\tanh(\lambda(\mathbf{N}_1\mathbf{N}_2^T - \mathbf{N}_2\mathbf{N}_1^T))), \tag{5}$$

where $\mathbf{N}_1 = \tanh(\lambda \mathbf{E}_1 \mathbf{\Phi}_1)$, $\mathbf{N}_2 = \tanh(\lambda \mathbf{E}_2 \mathbf{\Phi}_2)$, $\mathbf{E}_1$, $\mathbf{E}_2$ represent initialized node embedding. Then, the adjacency matrix $\mathbf{A}$ is fed into mix-hop propagation layer to effectively extract the influence of neighbor information on nodes. The output $\mathbf{H}_{out}$ is passed through the LSTM model to reduce the information conflict between the spatial structure of the graph and the time series, and obtain the final feature $\mathbf{h}_{t+\tau}^{(2)}$.

2.5. Feature Fusion Module

The $\mathbf{h}_{t+\tau}^{(1)}$ represents the timing characteristic, and $\mathbf{h}_{t+\tau}^{(2)}$ is the variable structure feature. In order to better fuse the two features and improve the accuracy of prediction, the calculation method of the feature fusion module [9] is:

$$\mathbf{G}_{t+\tau} = \sigma(\mathbf{W}_{h1}\mathbf{h}_{t+\tau}^{(1)} + \mathbf{W}_{h2}\mathbf{h}_{t+\tau}^{(2)}), \tag{6}$$

$$\mathbf{R}_{t+\tau} = (r_{t+\tau}^1, r_{t+\tau}^2, \cdots, r_{t+\tau}^N) = \mathbf{G}_{t+\tau} \odot \mathbf{h}_{t+\tau}^{(1)} + (\mathbf{1} - \mathbf{G}_{t+\tau}) \odot \mathbf{h}_{t+\tau}^{(2)}, \tag{7}$$

$$\hat{\mathbf{Y}}_{t+\tau} = (y_{t+\tau}^1, y_{t+\tau}^2, \cdots y_{t+\tau}^{N-P}) = (r_{t+\tau}^{P+1}, r_{t+\tau}^{P+2}, \cdots, r_{t+\tau}^N), \tag{8}$$

where $\mathbf{W}_{h1}$ and $\mathbf{W}_{h2}$ are learnable parameters, and σ is the sigmoid function.

3. Experiments

3.1. Data sources

In this paper, the proposed model is tested and evaluated using the daily data from the state-controlled air station in Jinan from January 1, 2013 to December 20, 2021. The dataset consists of 7 environmental impact variables, including cumulative precipitation from 20 to 20 o'clock (mm), atmospheric temperature (℃), relative humidity (%), sunshine hours (h), wind speed (m/s), barometric pressure (hPa), surface temperature (℃) and 4 air pollutant concentration variables, including NO_2, PM10, PM2.5, CO (mg/m^3).

3.2. Data preprocessing

Due to missing records and uncollected information in the real air station daily data, we first use the KNN proximity algorithm to fill in the missing values. The KNN proximity algorithm measures the similarity of samples in the dataset space by sample distance, and calculates the average value of the most similar k samples to estimate the value of missing data points. In this paper, k = 5.

3.3. Comparative models

To illustrate the superiority of the proposed model, we compare our model with the following 5 models:

LSTM: Long Short-Term Memory network for multidimensional input and output time series forecasting.

CNN-LSTM: a hybrid model of CNN and LSTM [4].

MTGNN: a multi-dimensional time series forecasting model that aggregates the graph convolution module and the time series convolution module, which can automatically learn the graph structure [7].

MTGODE: a forecasting model using dynamic graph neural ordinary differential equations [8].

3.4. Evaluation criteria

To quantify model performance, we choose two evaluation metrics, the root relative squared error (RRSE) and the relative absolute error (RAE). Since this article only considers single-step forecasting, the calculation methods of the two indicators are as follows:

$$RRSE = \frac{\sqrt{\sum_{i=1}^{N}(z_{t+\tau,i} - \hat{z}_{t+\tau,i})^2}}{\sqrt{\sum_{i=1}^{N}(z_{t+\tau,i} - \overline{z}_{t+\tau})^2}}, \tag{9}$$

$$RAE = \frac{\sum_{i=1}^{N}\left|z_{t+\tau,i} - \hat{z}_{t+\tau,i}\right|}{\sum_{i=1}^{N}\left|z_{t+\tau,i} - \overline{z}_{t+\tau}\right|}, \tag{10}$$

where $z_{t+\tau,i}$ and $\hat{z}_{t+\tau,i}$ respectively denote the real value and predicted value of the i -th variable at time $t+\tau$, $\overline{z}_{t+\tau}$ represents the mean of $\mathbf{z}$. Both indicators are the lower the better.

3.5. Parameter settings

We choose the time window size as 48, that is, we use the data of the past 48 days to predict the concentrations of air pollutants in the future. The prediction horizon is set to {1,3,7,14}, which means forecasting the concentrations of air pollutants after one day, three days, seven days, and fourteen days. During the training process, the trainer is selected as Adam, the initial learning rate is set to 0.0005, the training epoch is 1000, and the batch size is 32.

3.6. Experimental results

Table 1 lists the evaluation results of our proposed model and other models based on the daily data from the state-controlled air station in Jinan to predict air pollutants concentrations. The performance of CNN-LSTM is superior to that of LSTM, indicating that convolutional networks effectively improve the ability of LSTM to capture temporal information. In addition, prediction errors of LSTM and CNN-LSTM change relatively slowly with prediction-step increasing, because historical trend exerts a profound influence on the predicted values. The MTGNN and MTGODE further reduce deviations compared to the former two. Other variables do affect the targets, so adding structural information can increase the prediction accuracy. But as the prediction-step increases, the inaccuracies of the graph method change relatively significantly, which is related to the complexity of graph structure training.

Table 1: Model comparison results

		Horizon			
Methods	Criteria	1	3	7	14
LSTM	RRSE	0.0247	0.0256	0.0248	0.0256
	RAE	0.0214	0.0228	0.0217	0.0229
CNN-LSTM	RRSE	0.0234	0.0237	0.0242	0.0247
	RAE	0.0187	0.0192	0.0208	0.0214
MTGNN	RRSE	0.0211	0.0237	0.0240	0.0245
	RAE	0.0161	<u>0.0183</u>	**0.0182**	0.0193
MTGODE	RRSE	0.0204	0.0235	0.0242	0.0243
	RAE	<u>0.0146</u>	0.0184	0.0186	0.0192
Present model	RRSE	**0.0203**	**0.0230**	**0.0237**	**0.0235**
	RAE	**0.0146**	**0.0183**	0.0187	**0.0186**

The model presented has optimum performances in RRSE, which changes gently from 0.0203 to 0.0237. It outperforms other models in RAE when horizon = {1,3,14}, just slightly inferior when horizon = 7 in contrast with MTGNN and MTGODE. Fig. 2 shows the partial prediction results of the model on the test set when horizon = 1, which illustrates that our proposed method can effectively simulate the trend of time series changes. Overall, the model sufficiently combines the advantages of the above models and further improves the forecast accuracy.

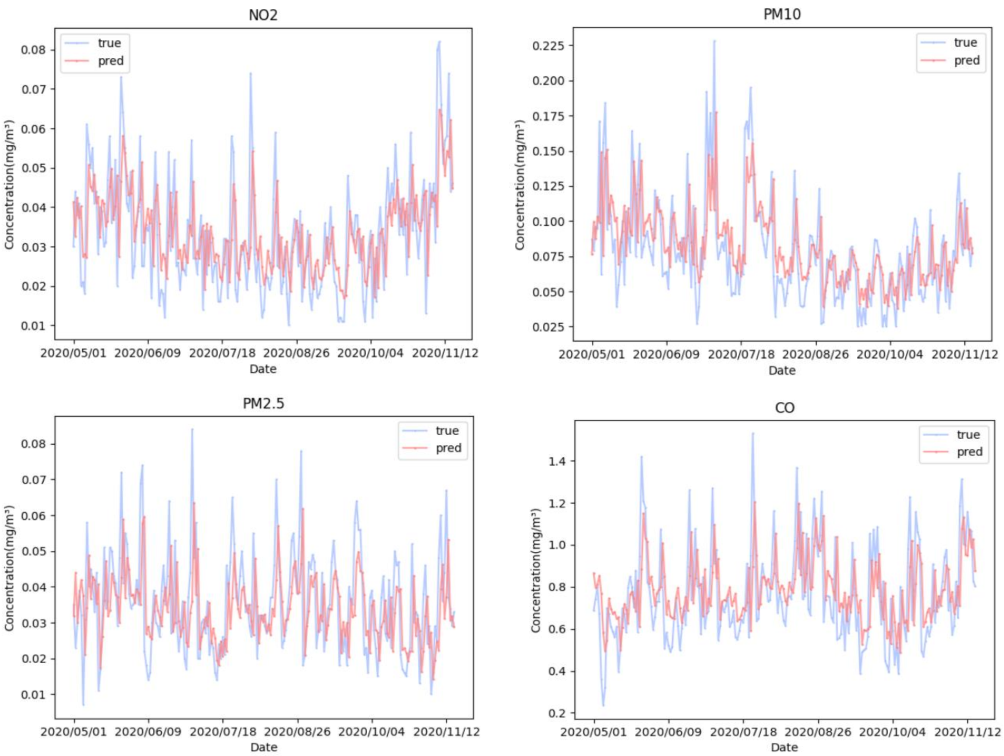

Fig. 2: The partial prediction results.

To verify that our proposed model effectively fuses the feature information extracted by the two modules, we test each branch separately. The specific operations and experimental results are as follows:

w/o T: the proposed model without temporal feature extraction module

w/o S: the proposed model without structural feature extraction module

Table 2: Ablation study results

	Horizon (**w/o T**)				Horizon (**w/o S**)			
	1	3	7	14	1	3	7	14
RRSE	0.0225	0.0237	0.0243	0.0245	0.0231	0.0238	0.0239	0.0241
RAE	0.0182	0.0189	0.0194	0.0199	0.0181	0.0188	0.0191	0.0196

The results of the proposed model are superior to those of any two branches running separately. These tests highlight that the feature fusion module effectively melds the information from different features and reduces information conflicts.

4. Conclusions

In this paper, a novel framework for simultaneously predicting various air pollutants concentrations considering time trend and variable correlation fusion is developed. The method is applied to air pollutants dataset from the state-controlled air station in Jinan, and experimental results show it outperforms other models. In the future, simplifying the computational complexity and further applying the model to other multivariate time series forecasting scenarios is the research direction we focus on.

5. Acknowledgements

The work was funded by the NSFC-Shandong Province Joint Grant No. U1806202.

6. References

[1] Y. Chang, H. Chiao, S. Abimannan, Y. Huang, Y. Tsai, and K. Lin. An LSTM-based aggregated model for air pollution forecasting. *Atmospheric Pollution Research*. 2020, **11**(8): 1451-1463.

[2] R. Rakholia, Q. Le, B. Ho, K. Vu, and R. Carbajo. Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam. *Environment International*. 2023, **173**: 107848.

[3] Y. Tsai, Y. Zeng, and Y. Chang. Air pollution forecasting using RNN with LSTM. *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. Athens, Greece. 2018, pp. 1074-1079.

[4] J. Zhang and S. Li. Air quality index forecast in Beijing based on CNN-LSTM multi-model. *Chemosphere*. 2022, **308**: 136180.

[5] S. Shih, F. Sun, and H. Lee. Temporal pattern attention for multivariate time series forecasting. *Machine Learning*. 2019, **108**: 1421-1441.

[6] S. Wang, Y. Li, J. Zhang, Q. Meng, L. Meng, and F. Gao. Pm2. 5-gnn: A domain knowledge enhanced graph neural network for pm2. 5 forecasting. In: C. Lu, et al (eds.) *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*. Seattle, WA, USA. 2020, pp. 163-166.

[7] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Virtual Event, CA, USA. 2020, pp. 753-763.

[8] M. Jin, Y. Zheng, Y. Li, S. Chen, B. Yang, and S. Pan. Multivariate time series forecasting with dynamic graph neural ODEs. *IEEE Transactions on Knowledge and Data Engineering*. 2022, doi: 10.1109/TKDE.2022.3221989.

[9] X. Geng, X. He, L. Xu, and J. Yu. Graph correlated attention recurrent neural network for multivariate time series forecasting. *Information Sciences*. 2022, **606**: 126-142.