

Research Article

A DBN-Based Deep Neural Network Model with Multitask Learning for Online Air Quality Prediction

Jiangeng Li ^{1,2}, Xingyang Shao,^{1,2} and Rihui Sun^{1,2}

¹College of Automation, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

²Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China

Correspondence should be addressed to Jiangeng Li; lijg@bjut.edu.cn

Received 2 February 2019; Revised 17 April 2019; Accepted 20 May 2019; Published 1 July 2019

Academic Editor: Antonio Visioli

Copyright © 2019 Jiangeng Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To avoid the adverse effects of severe air pollution on human health, we need accurate real-time air quality prediction. In this paper, for the purpose of improve prediction accuracy of air pollutant concentration, a deep neural network model with multitask learning (MTL-DBN-DNN), pretrained by a deep belief network (DBN), is proposed for forecasting of nonlinear systems and tested on the forecast of air quality time series. MTL-DBN-DNN model can solve several related prediction tasks at the same time by using shared information contained in the training data of different tasks. In the model, DBN is used to learn feature representations. Each unit in the output layer is connected to only a subset of units in the last hidden layer of DBN. Such connection effectively avoids the problem that fully connected networks need to juggle the learning of each task while being trained, so that the trained networks cannot get optimal prediction accuracy for each task. The sliding window is used to take the recent data to dynamically adjust the parameters of the MTL-DBN-DNN model. The MTL-DBN-DNN model is evaluated with a dataset from Microsoft Research. Comparison with multiple baseline models shows that the proposed MTL-DBN-DNN achieve state-of-art performance on air pollutant concentration forecasting.

1. Introduction

Air pollution is becoming increasingly serious. To protect human health and the environment, accurate real-time air quality prediction is sorely needed.

There are nonlinear and complex interactions among variables of air quality prediction data. Artificial neural networks can be used as a nonlinear system to express complex nonlinear maps, so they have been frequently applied to real-time air quality forecasting (e.g., [1–5]).

Deep networks have significantly greater representational power than shallow networks [6]. To solve several difficulties of training deep networks, Hinton et al. proposed a deep belief network (DBN) in [7]. DBN is trained via greedy layer-wise training method and automatically extracts deep hierarchical abstract feature representations of the input data [8, 9]. Deep belief networks can be used for time series forecasting, (e.g., [10–15]). For these reasons, in this paper, the proposed prediction model is based on a deep neural network pretrained by a deep belief network.

Multitask learning can improve learning for one task by using the information contained in the training data of other related tasks [16]. Multitask deep neural network has already been applied successfully to solve many real problems, such as multilabel learning [17], compound selectivity prediction [18], traffic flow prediction [19], speech recognition [20], categorical emotion recognition [21], and natural language processing [22]. Collobert and Weston demonstrated that a unified neural network architecture, trained jointly on related tasks, provides more accurate prediction results than a network trained only on a single task [22].

Current air quality prediction studies mainly focus on one kind of air pollutants and perform single task forecasting. The most studied problem is the $PM_{2.5}$ concentration prediction. However, there are correlations between some air pollutants predicted by us so that there is a certain relevance between different prediction tasks. For example, SO_2 and NO_2 are related, because they may come from the same pollution sources. Studies have showed that sulfate (SO_4^{2-}) is a major PM constituent in the atmosphere [23]. And in 2016,

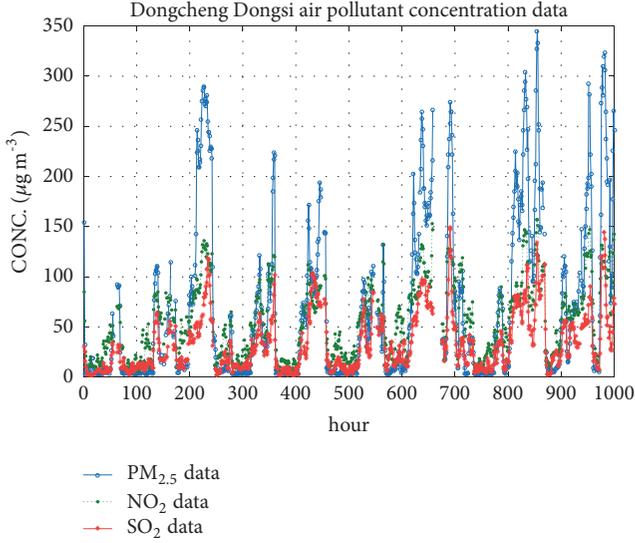


FIGURE 1: The observed data from 7 o'clock in November 30, 2014, to 22 o'clock in January 10, 2015. In the figure, time is measured along the horizontal axis and the concentrations of three kinds of air pollutants ($PM_{2.5}$, NO_2 , and SO_2) are measured along the vertical axis. There are some missing values in data sets. Dongcheng Dongsi is a target air-quality-monitor-station selected in this study.

a discovery revealed that the aqueous oxidation of SO_2 by NO_2 , under specific atmospheric conditions, is key to efficient sulfate formation, and the chemical reaction led to the 1952 London “Killer” Fog [24]. And a study published in the US journal *Science Advances* also discovered that fine water particles in the air acted as a reactor, trapping sulfur dioxide (SO_2) molecules and interacting with nitrogen dioxide (NO_2) to form sulfate [25]. Therefore, we can regard the concentration forecasting of these three kinds of pollutants ($PM_{2.5}$, SO_2 , and NO_2) as related tasks. Figure 1 shows some of the historical monitoring data for the concentrations of the three kinds of pollutants in a target station (Dongcheng Dongsi: air-quality-monitor-station) selected in this study. The three kinds of pollutants show almost the same concentration trend. Therefore, the concentration forecasting of the three kinds of pollutants can indeed be regarded as related tasks.

In this paper, based on the powerful representational ability of DBN and the advantage of multitask learning to allow knowledge transfer, a deep neural network model with multitask learning capabilities (MTL-DBN-DNN), pretrained by a deep belief network (DBN), is proposed for forecasting of nonlinear systems and tested on the forecast of air quality time series. DBN is used to learn feature representations, and several related tasks are solved simultaneously by using shared representations.

For multitask learning, a deep neural network with local connections is used in the study. Such connection effectively avoids the problem that fully connected networks need to juggle the learning of each task while being trained so that the trained networks cannot get optimal prediction accuracy for each task. The locally connected architecture can well learn the commonalities and differences of multiple tasks.

In order to get a better prediction of future concentrations, the sliding window [26, 27] is used to take the recent data to dynamically adjust the parameters of prediction model.

The rest of the paper is organized as follows. Section 2 presents the background knowledge of multitask learning, deep belief networks, and DBN-DNN and describes DBN-DNN model with multitask learning (MTL-DBN-DNN). In Section 3, the proposed model MTL-DBN-DNN is applied to the case study of the real-time forecasting of air pollutant concentration, and the results and analysis are shown. Finally, in Section 4, the conclusions on the paper are presented.

2. Methods

2.1. MultiTask Learning. Multitask learning can improve learning for one task by using the information contained in the training data of other related tasks. Multitask learning learns tasks in parallel and “what is learned for each task can help other tasks be learned better” [16].

Several related problems are solved at the same time by using a shared representation. Related learning tasks can share the information contained in their input data sets to a certain extent. Multitask learning exploits commonalities among different learning tasks. Such exploitation allows knowledge transfer among different learning tasks. The difference between the neural network with multitask learning capabilities and the simple neural network with multiple output level lies in the following: in multitask case, input feature vector is made up of the features of each task and hidden layers are shared by multiple tasks. Multitask learning is often adopted when training data is very limited for the target task domain [28].

2.2. Deep Belief Networks and DBN-DNN. Deep Belief Networks (DBNs) [29] are probabilistic generative models, and they are stacked by many layers of Restricted Boltzmann Machines (RBMs), each of which contains a layer of visible units and a layer of hidden units. DBN can be trained to extract a deep hierarchical representation of the input data using greedy layer-wise procedures. After a layer of RBM has been trained, the representations of the previous hidden layer are used as inputs for the next hidden layer. A schematic representation of a DBN is shown in Figure 2.

A DBN with l hidden layers contains l weight matrices: $\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(l)}$. It also contains $l + 1$ bias vectors: $\mathbf{b}^{(0)}, \dots, \mathbf{b}^{(l)}$ with $\mathbf{b}^{(0)}$ providing the biases for the visible layer. The probability distribution represented by the DBN is given by

$$P(\mathbf{h}^{(l)}, \mathbf{h}^{(l-1)}) \propto \exp\left(\mathbf{b}^{(l)\top} \mathbf{h}^{(l)} + \mathbf{b}^{(l-1)\top} \mathbf{h}^{(l-1)} + \mathbf{h}^{(l-1)\top} \mathbf{W}^{(l)} \mathbf{h}^{(l)}\right), \quad (1)$$

$$P(h_i^{(k)} = 1 | \mathbf{h}^{(k+1)}) = \sigma\left(b_i^{(k)} + \mathbf{W}_{:,i}^{(k+1)\top} \mathbf{h}^{(k+1)}\right) \quad (2)$$

$$\forall i, \forall k \in 1, \dots, l - 2,$$

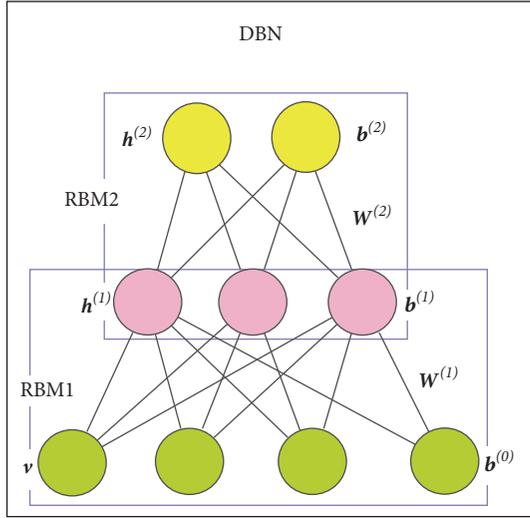


FIGURE 2: A 2-layer deep belief network that is stacked by two RBMs contains a layer of visible units and two layers of hidden units. Where $\mathbf{h}^{(1)}$ and $\mathbf{h}^{(2)}$ are the state vectors of the hidden layers, \mathbf{v} is the state vector of the visible layer, $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ are the matrices of symmetrical weights, $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(2)}$ are the bias vector of the hidden layers, and $\mathbf{b}^{(0)}$ is the bias vector of the visible layer.

$$P(v_i = 1 | \mathbf{h}^{(1)}) = \sigma\left(b_i^{(0)} + \mathbf{W}_{:,i}^{(1)\top} \mathbf{h}^{(1)}\right) \quad \forall i. \quad (3)$$

In the case of real-valued visible units, substitute

$$\mathbf{v} \sim N\left(\mathbf{b}^{(0)} + \mathbf{W}^{(1)\top} \mathbf{h}^{(1)}, \beta^{-1}\right), \quad (4)$$

with β diagonal for tractability [30]. $\sigma(x) = 1/(1 + \exp(-x))$.

The weights from the trained DBN can be used as the initialized weights of a DNN [8, 30],

$$\mathbf{h}^{(1)} = \sigma\left(\mathbf{b}^{(1)} + \mathbf{v}^\top \mathbf{W}^{(1)}\right), \quad (5)$$

$$\mathbf{h}^{(l)} = \sigma\left(\mathbf{b}_i^{(l)} + \mathbf{h}^{(l-1)\top} \mathbf{W}^{(l)}\right), \quad \forall l \in 2, \dots, m \quad (6)$$

and, then, all of the weights are fine-tuned by applying back-propagation or other discriminative algorithms to improve the performance of the whole network. When DBN is used to initialize the parameters of a DNN, the resulting network is called DBN-DNN [31].

2.3. DBN-Based Deep Neural Network Model with MultiTask Learning (MTL-DBN-DNN). In this section, a DBN-based multitask deep neural network prediction model is proposed to solve multiple related tasks simultaneously by using shared information contained in the training data of different tasks.

DBN-DNN prediction model with multitask learning is constructed by a DBN and an output layer with multiple units. Deep belief network is used to extract better feature representations, and several related tasks are solved simultaneously by using shared representations. The sigmoid function is used as the activation function of the output layer.

Each unit in the output layer is connected to only a subset of units in the last hidden layer of DBN. It is assumed that the number of related tasks to be processed is N , and it is assumed that the size of the subset (that is, the ratio of the number of nodes in the subset to the number of nodes in the entire last hidden layer) is α , then $1/(N-1) > \alpha > 1/N$. At the locally connected layer, each output node has a portion of hidden nodes that are only connected to it, and it is assumed that the number of nodes in this part is β , then $0 < \beta < 1/N$. There are common units with a specified quantity between two adjacent subsets.

The MTL-DBN-DNN model is learned with unsupervised DBN pretraining followed by backpropagation fine-tuning. The architecture of the model MTL-DBN-DNN is shown in Figure 3.

Remark. First, pretraining and fine-tuning ensure that the information in the weights comes from modeling the input data [32]. In other words, the network memorizes the information of the training data via the weights. The network needs not only to learn the commonalities of multiple tasks but also to learn the differences of multiple tasks. Locally connected network allows a subset of hidden units to be unique to one of the tasks, and unique units can better model the task-specific information. Therefore, fully connected networks do not learn the information contained in the training data of multiple tasks better than locally connected networks. Second, fully connected networks need to juggle (i.e., balance) the learning of each task while being trained, so that the trained networks cannot get optimal prediction accuracy for each task. Based on the above two reasons, the last (fully connected) layer is replaced by a locally connected layer, and each unit in the output layer is connected to only a subset of units in the previous layer. There are common units with a specified quantity between two adjacent subsets.

Input. As long as a feature is statistically relevant to one of the tasks, the feature is used as an input variable to the model.

When the MTL-DBN-DNN model is used for time series forecasting, the parameters of model can be dynamically adjusted according to the recent monitoring data taken by the sliding window to achieve online forecasting.

The Setting of the Structures and Parameters. The architecture and parameters of the MTL-DBN-DNN can be set according to the practical guide for training RBMs in technical report [33].

3. Experiments

3.1. Data Set. In this study, we used a data set that was collected in (Urban Computing Team, Microsoft Research) Urban Air project over a period of one year (from 1 May 2014 to 30 April 2015) [34]. There are missing values in the data, so the data was preprocessed in this study. We chose Dongcheng Dongsu air-quality-monitor-station, located in Beijing, as a target station. The hourly concentrations of $\text{PM}_{2.5}$, NO_2 , and SO_2 at the station were predicted 12 hours in advance.

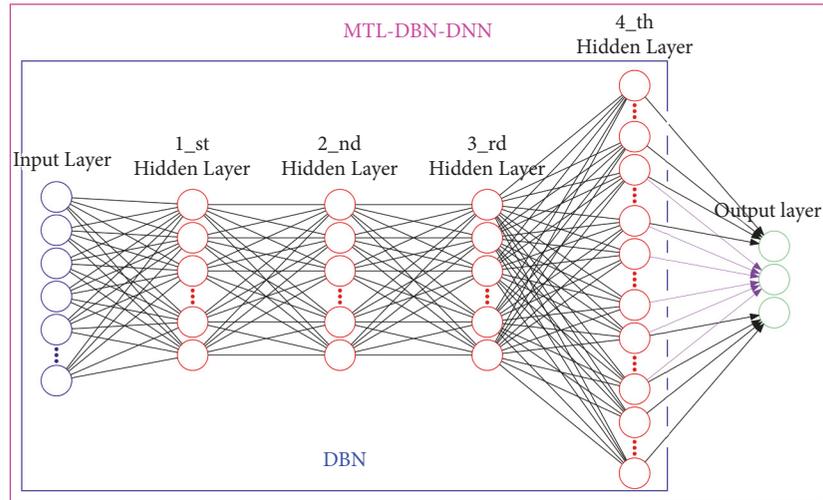


FIGURE 3: The schematic representation of the DBN-DNN model with multitask learning.

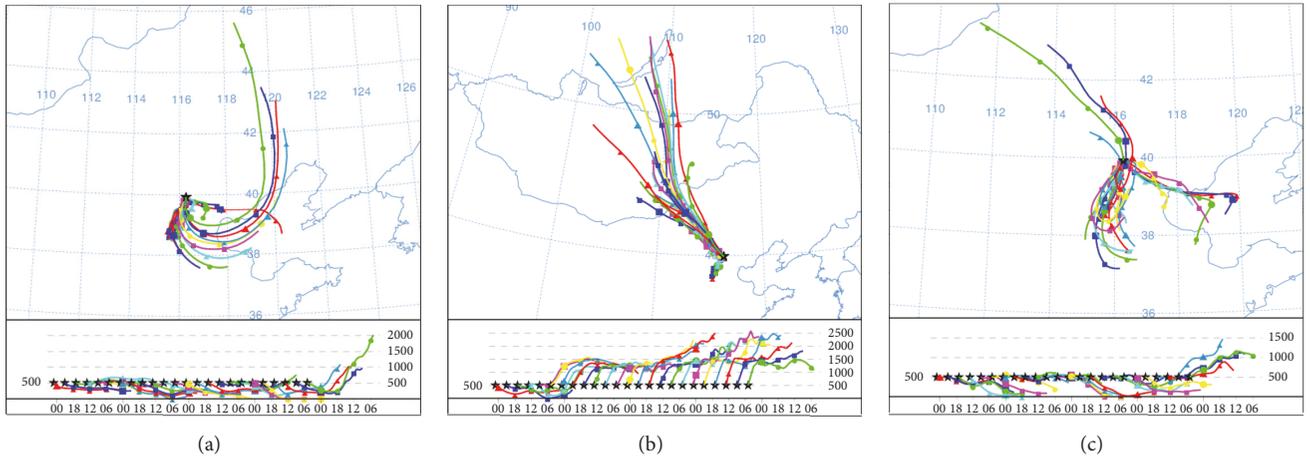


FIGURE 4: Three transport corridors, namely, southeast branch (a), northwest branch (b), and southwest branch (c), tracked by 24 h backward trajectories of air masses in Jing-Jin-Ji area.

3.2. Feature Set. According to some research results, we let the factors that may be relevant to the concentration forecasting of three kinds of air pollutants make up a set of candidate features.

Traffic emission is one of the sources of air pollutants. The traffic flow on weekdays and weekend is different. During the morning peak hours and the afternoon rush hours, traffic density is notably increased. In this paper, the hour of day and the day of week were used to represent the traffic flow data that is not easy to obtain.

Anthropogenic activities that lead to air pollution are different at different times of a year. The day of year (DAY) [3] was used as a representation of the different times of a year, and it is calculated by

$$DAY = \cos(2\pi dth/T), \quad (7)$$

where dth represents the ordinal number of the day in the year and T is the number of days in this year.

Regional transport of atmospheric pollutants may be an important factor that affects the concentrations of air pollutants. Three transport corridors are tracked by 24 h backward trajectories of air masses in Jing-Jin-Ji area [3, 35], and they are presented in Figure 4. According to the current wind direction and the transport corridors of air masses, we selected a nearby city located in the upwind direction of Beijing. Then we used the monitoring data of the concentrations of six kinds of air pollutants from a station located in the city to represent the current pollutant concentrations of the selected nearby city.

Candidate features include meteorological data from the target station whose three kinds of air pollutant concentrations will be predicted (including weather, temperature, pressure, humidity, wind speed, and wind direction) and the concentrations of six kinds of air pollutants at the present moment from the target station and the selected nearby city (including $PM_{2.5}$, PM_{10} , SO_2 , NO_2 , CO , and O_3), the hour of day, the day of week, and the day of year.

TABLE 1: The 21 elements in the candidate feature set.

#	Feature
1	The current PM _{2.5} concentration of the target station ($\mu\text{g}/\text{m}^3$)
2	The current PM ₁₀ concentration of the target station ($\mu\text{g}/\text{m}^3$)
3	The current NO ₂ concentration of the target station ($\mu\text{g}/\text{m}^3$)
4	The current CO concentration of the target station (mg/m^3)
5	The current O ₃ concentration of the target station ($\mu\text{g}/\text{m}^3$)
6	The current SO ₂ concentration of the target station ($\mu\text{g}/\text{m}^3$)
7	Weather
8	Temperature ($^{\circ}\text{C}$)
9	Atmospheric pressure (hPa)
10	Relative humidity
11	Wind speed (m/s)
12	Wind direction
13	The current PM _{2.5} concentration of the selected nearby station ($\mu\text{g}/\text{m}^3$)
14	The current PM ₁₀ concentration of the selected nearby station ($\mu\text{g}/\text{m}^3$)
15	The current NO ₂ concentration of the selected nearby station ($\mu\text{g}/\text{m}^3$)
16	The current CO concentration of the selected nearby station (mg/m^3)
17	The current O ₃ concentration of the selected nearby station ($\mu\text{g}/\text{m}^3$)
18	The current SO ₂ concentration of the selected nearby station ($\mu\text{g}/\text{m}^3$)
19	The day of year
20	The day of week
21	The hour of day

Weather has 17 different conditions, and they are sunny, cloudy, overcast, rainy, sprinkle, moderate rain, heavier rain, rain storm, thunder storm, freezing rain, snowy, light snow, moderate snow, heavy snow, foggy, sand storm, and dusty. All feature numbers are presented in the Table 1.

3.3. Evaluation Metrics. In this study, four performance indicators, including Mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE), and Accuracy (Acc) [34], were used to assess the performance of the models. They are defined by

$$MAE = \frac{1}{N} \sum_{i=1}^N |O_i - P_i|, \quad (8)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2}, \quad (9)$$

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{O_i - P_i}{O_i} \right|, \quad (10)$$

$$Acc = 1 - \frac{\sum_{i=1}^N |O_i - P_i|}{\sum_{i=1}^N O_i}, \quad (11)$$

where N is the number of time points and O_i and P_i represent the observed and predicted values respectively.

3.4. Experiment Setup. There is a new data element arriving each hour. Each data element together with the features that determine the element constitute a training sample

$[\mathbf{x}_t, (y_{1t}, y_{2t}, y_{3t})]$, where y_{1t} , y_{2t} , and y_{3t} represent PM_{2.5} concentration, NO₂ concentration and SO₂ concentration, respectively. \mathbf{x}_t is a set of features, and the set is made up of the factors that may be relevant to the concentration forecasting of three kinds of pollutant.

Setting the Parameters of Sliding Window (Window Size, Step Size, Horizon). In the study, the concentrations of PM_{2.5}, NO₂, and SO₂ were predicted 12 hours in advance, so, horizon was set to 12. Window size was equal to 1220; that is, the sliding window always contained 1220 elements. Step size was set to 1. After the current concentration was monitored, the sliding window moved one-step forward, the prediction model was trained with 1220 training samples corresponding to the elements contained in the sliding window, and then the well-trained model was used to predict the responses of the target instances.

Selecting Features Relevant to Each Task. The experimental procedures are as follows:

(1) After the continuous variables are discretized, for different tasks, the features were evaluated and sorted according to minimal-redundancy-maximal-relevance (mRMR) criterion.

First, the continuous variables were discretized, and the discretized response variable became a class label with numerical significance. In this paper, continuous variables were divided into 20 levels. A MI Tool box, a mutual information package of Adam Pocock, was used to evaluate the importance of the features according to the mRMR criterion.

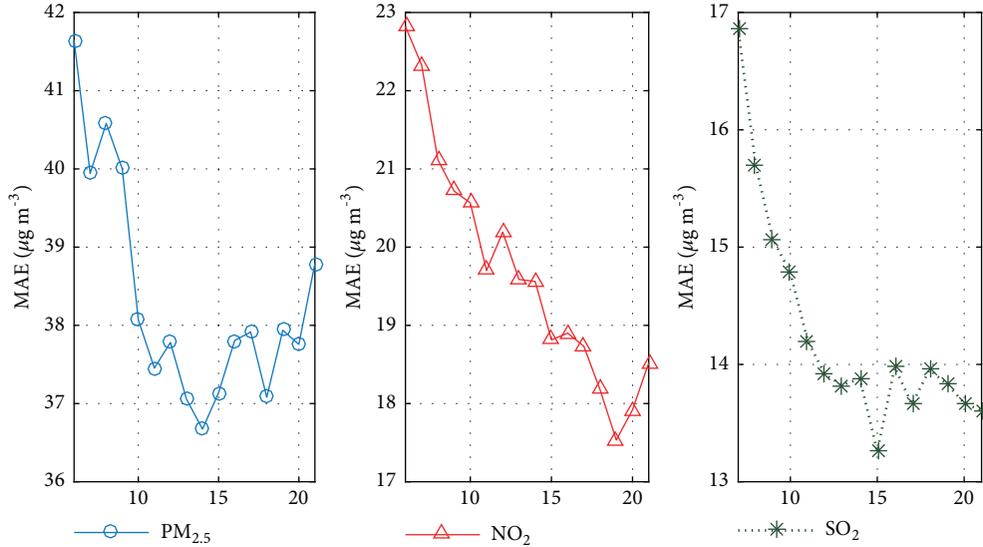


FIGURE 5: MAE vs. different numbers of selected features on three tasks.

TABLE 2: Selected features relevant to each task.

Task	Selected features	Removed features
PM _{2.5} concentration prediction	{19, 13, 1, 10, 6, 20, 3, 7, 12, 2, 11, 4, 9, 18}	{21, 8, 15, 5, 17, 16, 14}
NO ₂ concentration prediction	{19, 21, 11, 13, 10, 3, 6, 12, 20, 9, 2, 17, 8, 7, 4, 18, 5, 15, 1}	{16, 14}
SO ₂ concentration prediction	{19, 21, 6, 13, 11, 20, 18, 2, 10, 15, 7, 4, 9, 16, 5}	{17, 3, 8, 14, 1, 12}

(2) The dataset was divided into training set and test set. For each task, we used random forest to test the feature subsets from top1-topn according to the feature importance ranking, and then selected the first n features corresponding to the minimum value of the MAE as the optimal feature subset. The curves of MAE are depicted in Figure 5. Table 2 shows the selected features relevant to each task.

In order to verify whether the application of multitask learning and online forecasting can improve the DBN-DNN forecasting accuracy, respectively, and assess the capability of the proposed MTL-DBN-DNN to predict air pollutant concentration, we compared the proposed MTL-DBN-DNN model with four baseline models (2-5):

- (1) DBN-DNN model with multitask learning using online forecasting method (OL-MTL-DBN-DNN).
- (2) DBN-DNN model using online forecasting method (OL-DBN-DNN).
- (3) DBN-DNN model.
- (4) Air-Quality-Prediction-Hackathon-Winning-Model (Winning-Model) [36].
- (5) A hybrid predictive model (FFA) proposed by Yu Zheng, etc. [34].

For the single task prediction model, the input of the model is the selected features relevant to single task. For the multitask prediction model, as long as a feature is relevant to one of the tasks, the feature is used as an input variable to the model.

Remark. For the first two models (MTL-DBN-DNN and DBN-DNN), we used the online forecasting method. To

be distinguished from static forecasting models, the models using online forecasting method were denoted by OL-MTL-DBN-DNN and OL-DBN-DNN, respectively.

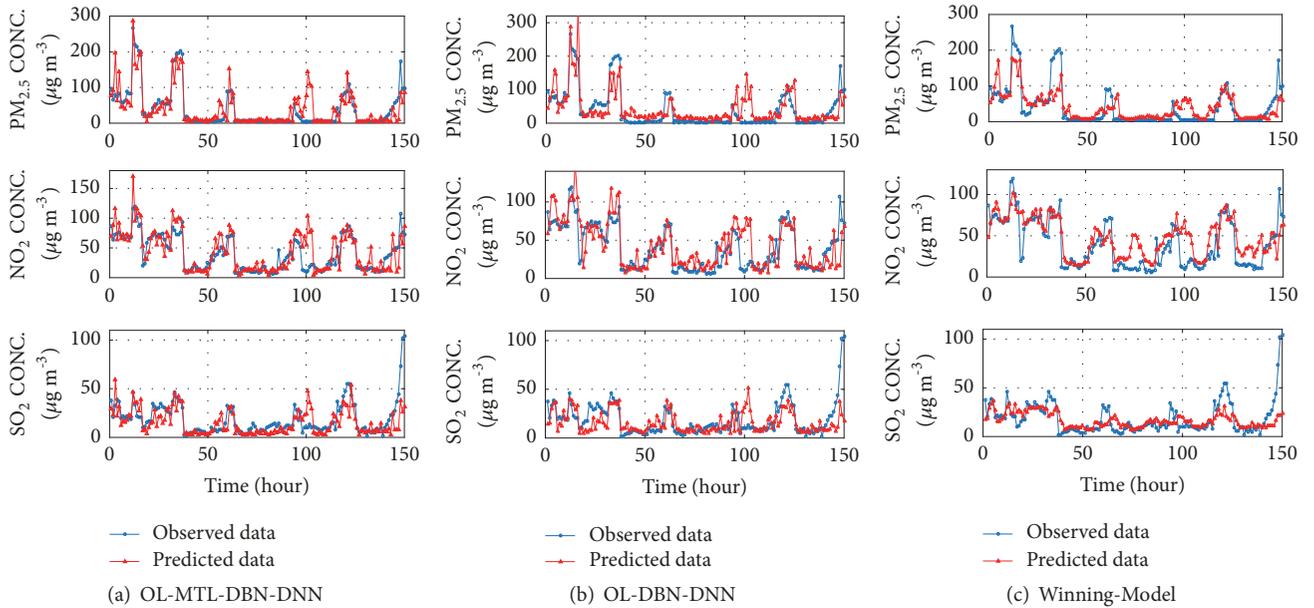
For the first three models above, we used the same DBN architecture and parameters. According to the practical guide for training RBMs in technical report [33] and the dataset used in the study, we set the architecture and parameters of the deep neural network as follows. In this study, deep neural network consisted of a DBN with layers of size G-100-100-100-90 and a top output layer, and G is the number of input variables. The DBN was constructed by stacking four RBMs, and a Gaussian-Bernoulli RBM was used as the first layer. In the pretraining stage, the learning rate was set to 0.00001, and the number of training epochs was set to 50. In the fine-tuning stage, we used 10 iterations, and grid search was used to find a suitable learning rate. For the OL-MTL-DBN-DNN model, the output layer contained three units and simultaneously output the predicted concentrations of three kinds of pollutants. Each unit at output layer was connected to only a subset of units at the last hidden layer of DBN.

For Winning-Model, time back was set to 4. Since the dataset used in this study was released by the authors of [34], the experimental results given in the original paper for the FFA model were quoted for comparison.

Because the first two models above are the models that use online forecasting method, the training set changes over time. For the sake of fair comparison, we selected original 1220 elements contained in the window before sliding window begins to slide forward, and used samples corresponding to

TABLE 3: Comparison among different models.

Models	PM _{2.5}				NO ₂				SO ₂			
	MAE	RMSE	MAPE	Acc	MAE	RMSE	MAPE	Acc	MAE	RMSE	MAPE	Acc
Winning-Model [36]	23.63	35.33	243.63	0.40	14.34	19.55	80.50	0.64	8.56	14.24	70.01	0.54
OL-MTL-DBN-DNN	<i>18.52</i>	<i>30.99</i>	<i>200.36</i>	<i>0.53</i>	<i>13.82</i>	<i>20.04</i>	<i>52.62</i>	<i>0.65</i>	<i>8.37</i>	<i>13.64</i>	<i>52.13</i>	<i>0.55</i>
OL-DBN-DNN	25.09	36.66	305.78	0.37	15.06	21.36	65.40	0.62	8.96	14.49	61.52	0.52
DBN-DNN	26.49	36.85	390.55	0.33	17.43	25.90	63.52	0.56	10.11	14.72	80.58	0.46

FIGURE 6: The prediction performances of different models for a 12-h horizon. In the pictures, time is measured along the horizontal axis and the concentrations of three kinds of air pollutants (PM_{2.5}, NO₂, SO₂) are measured along the vertical axis.

these elements as the training samples of the static prediction models (DBN-DNN and Winning-Model). The four models were used to predict the concentrations of three kinds of pollutants in the same period. The experimental results of hourly concentration forecasting for a 12h horizon are shown in Table 3, where the best results are marked with italic.

3.5. Results and Discussions. Table 3 shows that the best results are obtained by using OL-MTL-DBN-DNN method for concentration forecasting. Three error evaluation criteria (MAE, RMSE, and MAPE) of the OL-MTL-DBN-DNN are lower than that of the baseline models, and its accuracy is significantly higher than that of the baseline models. The prediction performance of OL-DBN-DNN is better than DBN-DNN, which shows that the use of online forecasting method can improve the prediction performance. The performance of OL-MTL-DBN-DNN surpasses the performance of OL-DBN-DNN, which shows that multitask learning is an effective approach to improve the forecasting accuracy of air pollutant concentration and demonstrates that it is necessary to share the information contained in the training data of three prediction tasks. It is worth mentioning

that learning tasks in parallel to get the forecast results is more efficient than training a model separately for each task.

The experimental results show that the OL-MTL-DBN-DNN model proposed in this paper achieves better prediction performances than the Air-Quality-Prediction-Hackathon-Winning-Model and FFA model, and the prediction accuracy is greatly improved. For example, when we predict PM_{2.5} concentrations, compared with Winning-Model, MAE and RMSE of OL-MTL-DBN-DNN are reduced by about 5.11 and 4.34, respectively, and accuracy of OL-MTL-DBN-DNN is improved by about 13%. These positive results demonstrate that our model MTL-DBN-DNN is promising in real-time air pollutant concentration forecasting.

When the prediction time interval in advance is set to 12 hours, some prediction results of three models are presented in Figure 6.

Figure 6 shows that predicted concentrations and observed concentrations can match very well when the OL-MTL-DBN-DNN is used. The advantage of the OL-MTL-DBN-DNN is more obvious when OL-MTL-DBN-DNN is used to predict the sudden changes of concentrations and the high peaks of concentrations.

4. Conclusion

In this paper, a deep neural network model with multitask learning (MTL-DBN-DNN), pretrained by a deep belief network (DBN), is proposed for forecasting of nonlinear systems and tested on the forecast of air quality time series.

The MTL-DBN-DNN model can fulfill prediction tasks at the same time by using shared information. In the model, each unit in the output layer is connected to only a subset of units in the last hidden layer of DBN. There are common units with a specified quantity between two adjacent subsets. Such connection effectively avoids the problem that fully connected networks need to juggle the learning of each task while being trained, so that the trained networks cannot get optimal prediction accuracy for each task. The locally connected architecture can well learn the commonalities and differences of multiple tasks.

$PM_{2.5}$, SO_2 , and NO_2 have chemical reaction and almost the same concentration trend, so we apply the proposed model to the case study on the concentration forecasting of three kinds of air pollutants 12 hours in advance. Comparison with multiple baseline models shows our model MTL-DBN-DNN has a stronger capability of predicting air pollutant concentration. Therefore, by combining the advantages of deep learning, multitask learning and online forecasting, the MTL-DBN-DNN model is able to provide accurate real-time concentration predictions of air pollutants.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Additional Points

Section 3.2 of this paper (feature set) cites the author's conference paper [37].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by National Natural Science Foundation of China (61873008) and Beijing Municipal Natural Science Foundation (4182008).

References

- [1] P. S. G. De Mattos Neto, F. Madeiro, T. A. E. Ferreira, and G. D. C. Cavalcanti, "Hybrid intelligent system for air quality forecasting using phase adjustment," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 185–191, 2014.
- [2] K. Siwek and S. Osowski, "Improving the accuracy of prediction of PM_{10} pollution by the wavelet transformation and an ensemble of neural predictors," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 6, pp. 1246–1258, 2012.
- [3] X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, and J. Wang, "Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation," *Atmospheric Environment*, vol. 107, pp. 118–128, 2015.
- [4] W. Tamas, G. Notton, C. Paoli, M.-L. Nivet, and C. Voyant, "Hybridization of air quality forecasting models using machine learning and clustering: An original approach to detect pollutant peaks," *Aerosol and Air Quality Research*, vol. 16, no. 2, pp. 405–416, 2016.
- [5] A. Kurt and A. B. Oktay, "Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7986–7992, 2010.
- [6] A. Y. Ng, J. Ngiam, C. Y. Foo, Y. Mai, and C. Suen, *Deep Networks: Overview*, 2013, <http://deeplearning.stanford.edu/wiki/index.php/Deep-Networks:Overview>.
- [7] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [9] S. Azizi, F. Imani, B. Zhuang et al., "Ultrasound-based detection of prostate cancer using automatic feature selection with deep belief networks," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, Eds., vol. 9350 of *Lecture Notes in Computer Science*, pp. 70–77, Springer, Munich, Germany, 2015.
- [10] M. Qin, Z. Li, and Z. Du, "Red tide time series forecasting by combining ARIMA and deep belief network," *Knowledge-Based Systems*, vol. 125, pp. 39–52, 2017.
- [11] X. Sun, T. Li, Q. Li, Y. Huang, and Y. Li, "Deep belief echo-state network and its application to time series prediction," *Knowledge-Based Systems*, vol. 130, pp. 17–29, 2017.
- [12] T. Kuremoto, S. Kimura, K. Kobayashi, and M. Obayashi, "Time series forecasting using a deep belief network with restricted Boltzmann machines," *Neurocomputing*, vol. 137, pp. 47–56, 2014.
- [13] F. Shen, J. Chao, and J. Zhao, "Forecasting exchange rate using deep belief networks and conjugate gradient method," *Neurocomputing*, vol. 167, pp. 243–253, 2015.
- [14] A. Dedinec, S. Filiposka, A. Dedinec, and L. Kocarev, "Deep belief network based electricity load forecasting: An analysis of Macedonian case," *Energy*, vol. 115, pp. 1688–1700, 2016.
- [15] H. Z. Wang, G. B. Wang, G. Q. Li, J. C. Peng, and Y. T. Liu, "Deep belief network based deterministic and probabilistic wind speed forecasting approach," *Applied Energy*, vol. 182, pp. 80–93, 2016.
- [16] R. Caruana, "Multitask learning," *Machine Learning*, vol. 28, no. 1, pp. 41–75, 1997.
- [17] Y. Huang, W. Wang, L. Wang, and T. Tan, "Multi-task deep neural network for multi-label learning," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 2897–2900, Melbourne, Australia, 2013.
- [18] R. Zhang, J. Li, J. Lu, R. Hu, Y. Yuan, and Z. Zhao, "Using deep learning for compound selectivity prediction," *Current Computer-Aided Drug Design*, vol. 12, no. 1, pp. 5–14, 2016.
- [19] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.
- [20] D. Chen and B. Mak, "Multi-task learning of deep neural networks for low-resource speech recognition," *IEEE Transactions on Audio, Speech and Language*, vol. 23, no. 7, pp. 1172–1183, 2015.

- [21] R. Xia and Y. Liu, "Leveraging valence and activation information via multi-task learning for categorical emotion recognition," in *Proceedings of the 40th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2015*, pp. 5301–5305, Brisbane, Australia, April 2014.
- [22] R. Collobert and J. Weston, "A unified architecture for natural language processing: deep neural networks with multitask learning," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 160–167, Helsinki, Finland, July 2008.
- [23] R. M. Harrison, A. M. Jones, and R. G. Lawrence, "Major component composition of PM10 and PM2.5 from roadside and urban background sites," *Atmospheric Environment*, vol. 38, no. 27, pp. 4531–4538, 2004.
- [24] G. Wang, R. Zhang, M. E. Gomez et al., "Persistent sulfate formation from London Fog to Chinese haze," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 48, pp. 13630–13635, 2016.
- [25] Y. Cheng, G. Zheng, C. Wei et al., "Reactive nitrogen chemistry in aerosol water as a source of sulfate during haze events in China," *Science Advances*, vol. 2, Article ID e1601530, 2016.
- [26] D. Agrawal and A. E. Abbadi, "Supporting sliding window queries for continuous data streams," in *IEEE International Conference on Scientific and Statistical Database Management*, pp. 85–94, Cambridge, Massachusetts, USA, 2003.
- [27] K. B. Shaban, A. Kadri, and E. Rezk, "Urban air pollution monitoring system with forecasting models," *IEEE Sensors Journal*, vol. 16, no. 8, pp. 2598–2606, 2016.
- [28] L. Deng and D. Yu, "Deep learning: methods and applications," in *Foundations and Trends® in Signal Processing*, vol. 7, pp. 197–391, Now Publishers Inc, Hanover, MA, USA, 2014.
- [29] G. E. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, article no. 5947, 2009.
- [30] Y. Bengio, I. Goodfellow, and A. Courville, *Deep Generative Models, Deep Learning*, MIT Press, Cambridge, Mass, USA, 2017.
- [31] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [32] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *The American Association for the Advancement of Science: Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [33] G. Hinton, "A practical guide to training restricted Boltzmann machines," in *Neural Networks: Tricks of the Trade*, G. Montavon, G. B. Orr, and K.-R. Müller, Eds., vol. 7700 of *Lecture Notes in Computer Science*, pp. 599–619, Springer, Berlin, Germany, 2nd edition, 2012.
- [34] Y. Zheng, X. Yi, M. Li et al., "Forecasting fine-grained air quality based on big data," in *Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '15)*, pp. 2267–2276, Sydney, Australia, August 2015.
- [35] X. Feng, Q. Li, Y. Zhu, J. Wang, H. Liang, and R. Xu, "Formation and dominant factors of haze pollution over Beijing and its peripheral areas in winter," *Atmospheric Pollution Research*, vol. 5, no. 3, pp. 528–538, 2014.
- [36] "Winning Code for the EMC Data Science Global Hackathon (Air Quality Prediction), 2012," <https://github.com/benhamner/Air-Quality-Prediction-Hackathon-Winning-Model>.
- [37] J. Li, X. Shao, and H. Zhao, "An online method based on random forest for air pollutant concentration forecasting," in *Proceedings of the 2018 37th Chinese Control Conference (CCC)*, Wuhan, China, 2018.

