

Research Article

Compound Autoregressive Network for Prediction of Multivariate Time Series

Yuting Bai ^{1,2} Xuebo Jin ^{1,2} Xiaoyi Wang ^{1,2} Tingli Su ^{1,2} Jianlei Kong ^{1,2} and Yutian Lu³

¹School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China

²Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China

³State Grid Beijing Electric Power Company, Beijing 100031, China

Correspondence should be addressed to Xuebo Jin; jinxuebo@btbu.edu.cn

Received 16 July 2019; Revised 31 August 2019; Accepted 4 September 2019; Published 22 September 2019

Guest Editor: Teddy Craciunescu

Copyright © 2019 Yuting Bai 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.

The prediction information has effects on the emergency prevention and advanced control in various complex systems. There are obvious nonlinear, nonstationary, and complicated characteristics in the time series. Moreover, multiple variables in the time-series impact on each other to make the prediction more difficult. Then, a solution of time-series prediction for the multivariate was explored in this paper. Firstly, a compound neural network framework was designed with the primary and auxiliary networks. The framework attempted to extract the change features of the time series as well as the interactive relation of multiple related variables. Secondly, the structures of the primary and auxiliary networks were studied based on the nonlinear autoregressive model. The learning method was also introduced to obtain the available models. Thirdly, the prediction algorithm was concluded for the time series with multiple variables. Finally, the experiments on environment-monitoring data were conducted to verify the methods. The results prove that the proposed method can obtain the accurate prediction value in the short term.

1. Introduction

In the information era, data play a significant role in various artificial and natural systems. Data provide the basis for machine control, industrial system running, economical market, environment management, etc. For the complex systems above, the accurate real-time data are essential for the control and operation. Moreover, the future information is also very important, which is predicted with the historical data and can guide the beforehand operation for the system adjustment, environmental adaptation, and accident avoidance. Therefore, the reliable prediction of the data in the time domain becomes an urgent issue for the complex systems. For the complicated composition and internal mechanism, the time-series data in the systems are usually nonstationary, nonlinear, and noisy. The complicated features make the prediction difficult. Besides, the variables in the time-series impact on each other to perplex the nonlinear

relation. Then, the prediction issue becomes the challenge in front of the complicated time-series characteristics and multivariate correlativity.

In the prediction issue, various explorations have been conducted to excavate the potential rules and features in the time-series data. For the practice application in some fields, the prediction methods are proposed based on mechanism models. In the methods, the inner mechanism of a system is studied deeply, in which the relations between system components are built with the approach of physics, chemistry, and biology, such as models of water environment (WASP [1] and EFDC [2]) and models of atmospheric diffusion (Gaussian puff and plume model [3]). The system change can be predicted based on the mechanism model in the view of model simulation. However, the models are difficult to build because of the complex and unknown inner structure. Moreover, the professional and interdisciplinary knowledge is also required for the mechanism analysis.

The data-driven solution has been an effective complement for the mechanism methods. Different from the mechanism methods, the data-driven methods focus on the external data characteristic instead of the inner structure relation. It develops from the statistical method to the machine learning method which can excavate more features from the mass data. Machine learning solves mainly the problem of the parametric model setting and adaption in the statistical methods such as autoregression (AR), moving average (MA), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA) models [4]. Machine learning including the traditional neural network and deep learning also face some problems in the time-series analysis. First, multiple variables usually need to be considered for the target predicted variable. In the multivariable analysis, the traditional networks mainly model the multivariable mapping relations, while neglecting the sequential features. And the deep learning methods are specialized in the sequential feature extraction of the univariate. Second, the computational efficiency should be considered in the prediction models, especially for the terminal application which cannot provide the high configuration. Third, the training methods affect the network performance largely. A suitable and extensible learning framework should be designed for the neural network. Based on the analysis of the existing research, we explore an access to the time-series prediction, in the view of multivariable modelling performance, computational efficiency, and training methods.

The rest of this paper is organized as follows: Section 2 introduces the related prediction methods, including the statistical model and machine learning method. In Section 3, the main prediction model is proposed and the compound autoregressive network is presented with the prediction algorithm. Experiments are conducted in Section 4 to test the network. The methods and results are discussed in Section 5. Finally, the paper is concluded in Section 5.

2. Related Works

The direct solution of the prediction is to figure out the change rule of the system, which is the basic thought of the mechanism-based prediction methods. Obviously, it is difficult to build the completed mechanism model to describe the system composition and change rule. Then, the data-driven method becomes a feasible solution with the external characteristic irrespective of the system inner construction and relation. The data-driven methods can be divided into two categories: statistical model and machine learning model.

2.1. Prediction Models Based on Statistics. The statistical model is based on the mathematical description and calculation of the data. The classical statistical models are built on the autocorrelation function and exponential decays of the time series. The typical models include AR, MA, and hybrid models. The AR model describes the change process of the regressor variable itself. In the model, the random

variables in the next time steps are expressed with the linear combination of the variables in the previous moments. The MA model uses the sliding window to extract the time-series features in the view of the adjacent data segment. Because the length of the sliding window impacts the feature extraction ability mainly, some exponential smoothing methods are proposed to optimize the MA model, in which the cubic exponential smoothing method is applied widely. Based on AR and MA models, the hybrid model is proposed for accurate modelling, including the ARMA and ARIMA. The ARIMA has been the typical hybrid model for the nonstationary regressive issue. It was applied in the prediction problems of environment monitoring [5], financial economy [6, 7], food safety [8], traffic system [9], etc.

The statistical model can be expressed as follows. x_t is the value of the time series at t , p is the number of autoregressive terms, q is the number of moving average terms, d is the differential order, ε_t is the white noise at t , L is the lag operator, and $\alpha_1, \dots, \alpha_p$ and β_1, \dots, β_q are the weights. Then, the AR model can be expressed as

$$x_t = \sum_{i=1}^p \alpha_i x_{t-i} + \varepsilon_t. \quad (1)$$

The MA model is

$$x_t = \sum_{i=1}^q \beta_i \varepsilon_{t-i} + \varepsilon_t. \quad (2)$$

The ARMA model is

$$x_t = \sum_{i=1}^p \alpha_i x_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i} + \varepsilon_t. \quad (3)$$

The ARIMA model is

$$\left(1 - \sum_{i=1}^p \alpha_i L^i\right) (1 - L)^d x_t = \left(1 + \sum_{i=1}^q \beta_i L^i\right) \varepsilon_t. \quad (4)$$

The statistical models rely on the assumption of stationarity in the time series. Although the models are improved and evolved, they are still limited by the transformation and process of the stationary data. Besides, it is a problem on how to select a proper model and estimate the model parameters. The practice indicates that the models perform well in the linear short-term prediction. The prediction accuracy declines markedly in the complex and long-term time series. It becomes a demand to seek new prediction solutions to the nonstationary time series.

2.2. Prediction Model Based on Machine Learning. Machine learning develops fast in the classification and regression research. The black-box thought of machine learning seems to provide the extensive possibility for the complex modelling problems. The backpropagation neural network (BP), radial basis function neural network (RBF), nonlinear autoregressive neural network (NAR), support vector machine (SVM), and Bayes network have been studied and applied in the prediction problems [10].

Some studies have been conducted to improve the network and prediction performance. Pradeepkumar [11] proposed a novel particle swarm optimization algorithm to train the quantile regression neural network, which was applied in the financial data prediction. Daly [12] designed the structure of the NAR to predict the video traffic in the Ethernet passive optical network. Wang [13] proposed an adaptive method based on the multiple-rate network to predict the parameters in industrial control. Liu [14] studied an improved grayscale neural network which was tested to predict the traffic stop. Some combinations of different methods are also a hotspot in the machine learning studies. Doucoure [15] predicted the wind speed with wavelet analysis and neural network. Wang [16] improved the BP with the self-adaptive differential evolution algorithm.

The machine learning methods above are mainly the shallow networks. They are suitable for multivariate modelling because of the network structure of multiple input nodes. The data in different time steps are imported independently into the network circularly, which place emphasis on the nonlinear mapping relation instead of the sequence connection in the time domain. Generally, they are limited in mass data processing and complex time-series relation modelling. Especially for the prediction issue, the sequence feature should be extracted which is difficult to realize in the traditional fully connected network. The recurrent neural network (RNN) [17] draws much attention in the sequence features. In the RNN, the nodes between the hidden layers are connected, and the input of the hidden layer includes not only the output of the input layer but also the output of the previous hidden layer. The RNN develops to the multidimensional recurrent neural network (MDRNN) [18] and to the bidirectional recurrent neural network (BiRNN) [19] for the higher performance. The long short-term memory network (LSTM) [20] is proposed for the long-term dependency problem in the traditional RNN. Some variants of the LSTM appear with the improvement and redesign of the structure or gate in the LSTM, including the bidirectional LSTM network (BiLSTM) [21] and gated recurrent unit (GRU) [22]. Although the deep networks usually perform better than the traditional networks, they are studied and applied more with the univariate instead of the multivariate. Besides, their structures are more complex, and they need more training time and computing resources.

In the prediction problem of the time series, on the one hand, we should consider the sequence feature of the time series as well as the mutual effect of the related variables. On the other hand, we should balance the network prediction accuracy with the calculating speed and resources occupied. Considering the related works mentioned above, the advantages of different networks should be utilized, including the simple structure and multivariate analysis ability in the shallow networks, as well as the sequence feature extraction in the recurrent networks. Then, the shallow recurrent neural network NAR [23] is selected as the basic network which can extract the nonlinear and sequence features in the time series. And a compound network structure and algorithm are designed to analyse multiple variables. The novel framework of the compound network can be applied in the prediction

problem of complex systems, providing an alternative solution to analyse the data change in the data-driven view.

3. Compound Autoregressive Prediction Network

For the time series in the systems, the main feature is the trend in their changing process, as well as the incidence relation among different variables. The trend means that there are potential rules in the changing data, which can be linear, periodic, or stochastic. The incidence relation means the effect on multiple variables. For example, the temperature value fluctuates in its change rule, and it is impacted by other meteorological variables such as the precipitation and humidity. Based on the two important factors in the time series, a compound neural network is built to predict the object variable. The overall network structure is introduced firstly. Then, the components and training methods are analysed. The prediction algorithm for the multivariate time series is proposed finally.

3.1. Compound Autoregressive Network. In the traditional neural networks, the NAR can realize the regression analysis of the time series itself. The network has been applied in practice and performs well in the short-term prediction. Besides, the data needed in the network training are obviously less than those of deep networks such as the LSTM and GRU. Then, the NAR can be an effective tool in the univariate prediction. Moreover, the nonlinear autoregressive network with external input (NARX) develops based on the NAR, in view of the incidence relation in the multiple variables. With the advantages of the NAR and NARX, the compound network is designed for the multivariate prediction issue, as shown in Figure 1. The compound autoregressive network proposed in this paper is abbreviated as CARN.

The CARN consists of two parts, namely, the primary network and auxiliary network. In the prediction issue, a variable is the main target to be predicted, and some variables are selected as the correlated variables according to their correlation degrees. The components in the compound network are corresponding to different types of variables. The primary network is built based on the structure of the NARX to predict the object variable. And the auxiliary network is built based on the NAR to provide the reference of the correlated variables.

For the primary network, the inputs include the object variable (Y in Figure 1) and the correlated variables (U in Figure 1). The nonlinear and complex relation in the variables is usually difficult to be analysed with mechanism modelling. But the network performs well in the black-box mapping relation mining. Then, the design of the two types of inputs can excavate the associate relation in multiple variables. Besides the two types of inputs, the other characteristic of the network is the feedback of the object variable from the output to the input. The changing trend in the object variable itself is usually more important than the multivariable relation. And the self-trend is constructed based on the feedback in the time dimension.

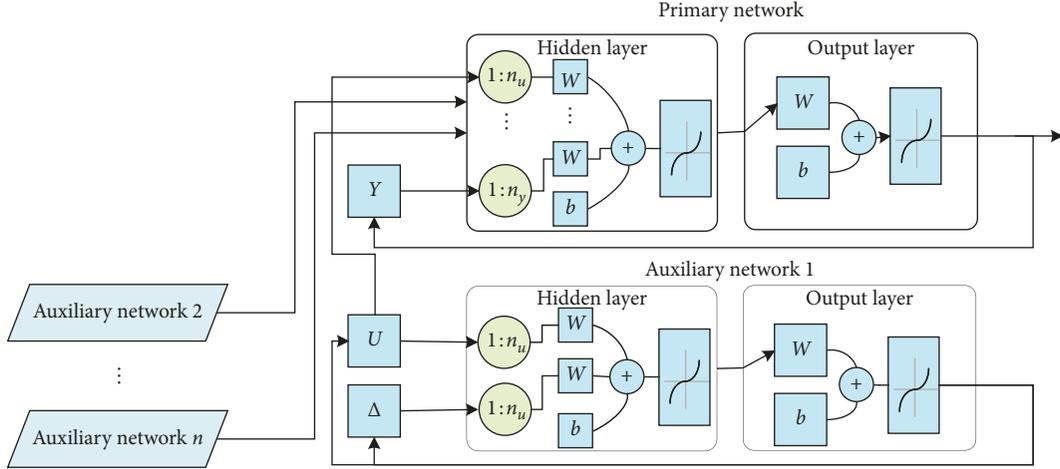


FIGURE 1: Compound autoregressive network for the multivariable time-series prediction.

For the auxiliary network, the main inputs are the variables associated with the object variable. The network mainly sets up the time-series trend with the structure of the feedback. In the feedback, the data change gradient is also set as the input to compensate the prediction. The NAR-based auxiliary network realizes the regression of the univariate. Moreover, there is not only one effect variable of the object variable. Therefore, there are some auxiliary networks in practice, and the number of auxiliary networks equals the variable number.

3.2. Design and Train of Discrete Networks. In the framework of the compound network, the primary and auxiliary networks are set up to predict the variables. There are two issues to be solved including the concrete network structure and the network training method. The structures of the networks are shown in Figure 2.

There are three layers in the primary network, namely, the input, hidden, and output layers. The inputs include the effect variables which are from the auxiliary networks and the object variable. In the view of the time dimension, the data of the object variable in the past are used to predict the data in the next time steps. The data at present are provided by the auxiliary networks. The nonlinear regressive function of the network can be expressed as

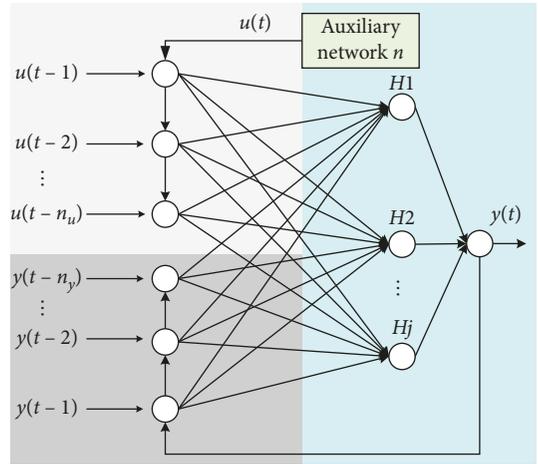
$$y(t) = f\left(u(t), u(t-1), \dots, u(t-n_u), y(t-1), y(t-2), \dots, y(t-n_y)\right), \quad (5)$$

where $y(t)$ is the prediction output, $u(t)$ is the effect variable input, $(t-i)$ means the time step, n_u is the input delay, and n_y is the output delay.

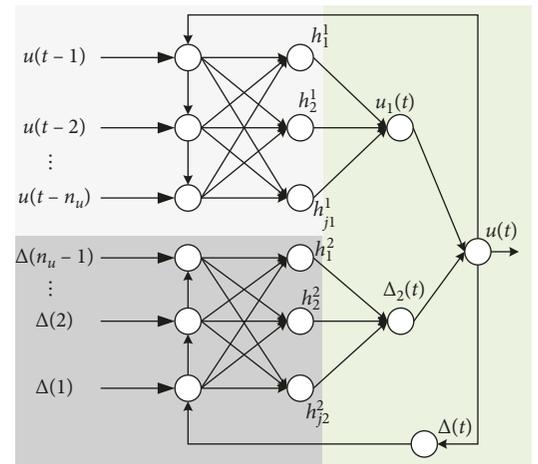
The relation between the input and hidden layers is

$$H_j = f\left(\sum_{i^1=1}^{n_u} W_{i^1 j} u_{i^1} + \sum_{i^2=1}^{n_y} W_{i^2 j} y_{i^2} + A_j\right), \quad (6)$$

where $j = 1, 2, \dots, l$, i_1 is the number of historical input data, u_{i_1} is the i_1 -th input, i_2 is the number of historical output data, y_{i_2} is the i_2 -th output, l is the number of hidden-layer



(a)



(b)

FIGURE 2: Structure design of the networks: (a) primary network; (b) auxiliary network.

neurons, f is the activation function in the hidden layer, $W_{i^1 j}$ is the connection weight between the i^1 -th input and the j -th neuron in the hidden layer, $W_{i^2 j}$ is the connection

weight between the i_1 -th linear relation weight and the j -th neuron in the hidden layer, and A_j is the threshold value of the j -th hidden neuron.

The network output O can be obtained with the hidden-layer output H_j :

$$O = f\left(\sum_{j=1}^i H_j W_j + B\right), \quad (7)$$

where W_j is the connection weight between the output neuron and the j -th neuron in the hidden layer and B is the threshold value of the output neuron.

Similar to the primary network, there are also input, hidden, and output layers in the auxiliary network. But the hidden layers are extended to two layers. The inputs include the effect variable itself and the data change gradient which can be the reference to promote the prediction accuracy. The network can be expressed as

$$u(t) = f\left(u(t-1), u(t-2), \dots, u(t-n_u), \Delta(t-1), \Delta(t-2), \dots, \Delta(t-n_u)\right), \quad (8)$$

where $u(t)$ is the effect variable input and $\Delta(t)$ is the data change gradient given by

$$\Delta(t) = \frac{(u(t-n_u) - u(t-(n_u-1)))}{t_0}, \quad (9)$$

where n_u is the input delay and t_0 is the time step interval.

The concrete model of the auxiliary network is

$$\begin{aligned} h_{j_1} &= f\left(\sum_{i_1=1}^{n_u} \omega_{i_1 j_1} u_{i_1} + a_{j_1}\right), \\ h_{j_2} &= f\left(\sum_{i_2=1}^{n_u-1} \omega_{i_2 j_2} \Delta_{i_2} + a_{j_2}\right), \end{aligned} \quad (10)$$

where $j_1 = 1, 2, \dots, l_1, j_2 = 1, 2, \dots, l_2$, i_1 is the number of historical input data, i_2 is the number of linear relation weights between $u(t)$ and $u(t-1)$, l_1 and l_2 are the number of hidden-layer neurons, n_u is the input delay, f is the activation function of the hidden layer, u_{i_1} is the i_1 -th input number, ω is the connection weight between input and hidden neurons, and a is the threshold value of the hidden neuron. The output is derived from the hidden layer:

$$o = f\left(f\left(\sum_{j_1=1}^{l_1} h_{j_1} \omega_{j_1} + b_1\right) + f\left(\sum_{j_2=1}^{l_2} h_{j_2} \omega_{j_2} + b_2\right) + c\right), \quad (11)$$

where b is the threshold of the second hidden layer and c is the threshold of the output layer.

Based on the design of the networks above, the training method should be studied. The basic learning method is from the algorithm of backpropagation through time, in which the variable from the feedback can be regarded as a new variable.

The errors of the primary and auxiliary networks between the prediction output and the designed output are

$$\begin{cases} e_1 = \frac{1}{2}(y - o)^2, \\ e_2 = \frac{1}{2}(Y - O)^2, \end{cases} \quad (12)$$

where e_1 and e_2 are the errors, o and O are the prediction outputs, and y and Y are the designed outputs.

The connection weights $\omega_{i_1 j_1}, \omega_{i_2 j_2}, \omega_{j_1}, \omega_{j_2}, W_{i_1 j}, W_{i_2 j}, W_j, a_{j_1}, a_{j_2}, b_1, b_2, A_j$, and B are adjusted with the errors until the global error or the training iterations reach the preset value. Based on the backpropagation algorithm, the weights are obtained as

$$\begin{cases} \omega_{i_1 j_1} = \omega_{i_1 j_1} + \eta_1 h_{j_1} (1 - h_{j_1}) x_{i_1} \omega_{j_1} e_1, \\ \omega_{i_2 j_2} = \omega_{i_2 j_2} + \eta_1 h_{j_2} (1 - h_{j_2}) x_{i_2} \omega_{j_2} e_1, \\ \omega_{j_1} = \omega_{j_1} + \eta_1 h_{j_1} e_1, \\ \omega_{j_2} = \omega_{j_2} + \eta_1 h_{j_2} e_1, \\ a_{j_1} = a_{j_1} + \eta_1 h_{j_1} (1 - h_{j_1}) \omega_{j_1} e_1, \\ a_{j_2} = a_{j_2} + \eta_1 h_{j_2} (1 - h_{j_2}) \omega_{j_2} e_1, \\ b_1 = b_1 + e_1, \\ b_2 = b_2 + e_1, \\ c = c + e_1, \\ E_1 = \frac{1}{2m_1} \sum_{k_1=1}^{m_1} (y - o)^2, \\ W_{i_1 j} = W_{i_1 j} + \eta_2 H_j (1 - H_j) x_{i_1} W_j e_2, \\ W_{i_2 j} = W_{i_2 j} + \eta_2 H_j (1 - H_j) x_{i_2} W_j e_2, \\ W_j = W_j + \eta_2 H_j e_2, \\ A_j = A_j + \eta_2 H_j (1 - H_j) W_j e_2, \\ B = B + e_2, \\ E_2 = \frac{1}{2m_2} \sum_{k_2=1}^{m_2} (Y - O)^2, \end{cases} \quad (13)$$

where η_1 and η_2 are the learning rate and E_1 and E_2 are the global errors of the two networks.

3.3. Prediction Algorithm for Multivariate Time Series. Based on the CARN proposed above, the data in practice can be used to train and obtain the networks which can predict the object variable with the effect variables. The prediction algorithm for the multivariate time series is designed based on the network model. In the algorithm, the data processing and calculation process is ascertained to obtain the final prediction results. The algorithm flow is shown in Figure 3.

The inputs of the prediction algorithm include the historical data of the object variable and effect variables and the data change gradient. The output is the series of the object variable in the next time steps. The steps of the algorithm are as follows:

- (1) The effect variables are selected with the correlation degrees between the object and effect variables. The historical data of the object variable and selected effect variables are preprocessed with the normalization method. In the preprocessing, the data change gradients of the effect variables should be calculated for the auxiliary networks.
- (2) The historical data which have been processed are imported into the auxiliary networks. The networks are trained with the method in Section 3.2.
- (3) The outputs of the auxiliary networks and the historical data of the object variable are imported into the primary network to obtain the main prediction model.
- (4) The time step is set forward, and the updated data in the next time step can be obtained by repeating the steps above.

The compound network and the prediction algorithm for the multivariate time series have been proposed so far. In practice, the prediction length should be set, and the effect variables should be selected reasonably. Then, the designed prediction results of the object variable can be obtained with the historical data.

4. Experiment and Results

4.1. Experiment Data and Setting. In the experiment, we focus on the data prediction issue in the complex environment system. Two sets of the environment data are chosen to be tested. One is the atmospheric quality data from the monitoring system of an industrial park. And the other one is the meteorological forecast data.

For the atmospheric quality data, 3240 sets of data are truncated from the monitoring system in an industrial park of Hebei Province, China. The data are from different time periods which can represent different trends. The time periods include June to August in 2016 (set A), September to November in 2016 (set B), and December in 2016 to February in 2017 (set C). The monitored variables are SO_2 , NO_2 , CO, O_3 , VOC, humidity, temperature, wind speed,

atmospheric pressure, etc. And they were recorded every hour in the monitoring system. SO_2 is the main factor in the atmospheric environment management in the industrial park. Then, SO_2 is set as the object variable to be predicted, and the correlation degrees between other variables and SO_2 were calculated, as shown in Figure 4. Then, the main effect variables were selected including NO_2 , CO, O_3 , humidity, and wind speed.

For the meteorological forecast data, there are 24 sets of data in a day. And every set is about the meteorological factors, including the temperature, humidity, wind speed, precipitation, and atmospheric pressure. Similar to the atmospheric quality data, the most relevant variables are selected for the object variable temperature. The effect variables are the humidity, wind speed, and precipitation.

In the setting of the prediction models, the data were preprocessed firstly with the method of maximum and minimum. The prediction network output should be denormalized. The data were divided into the training, validation, and test sets. Their proportions are 70%, 15%, and 15%. The numbers of various sets are listed in Table 1.

In the experiments, the parameters of the network structure and training were obtained and are listed in Table 2. Then, the networks are trained to run the prediction algorithm in Section 3.3. The prediction results are presented in Section 4.2.

Some typical prediction methods are set as the contrast methods, including the ARIMA model, BP, RNN, and LSTM. The contrast methods cover the main types of the classic statistical model and machine learning methods. In the concrete experiments, the ARIMA and RNN are used to predict the object variable. The BP and LSTM are designed with multiple inputs including the object variable and effect variables.

4.2. Results of Atmospheric Quality Data. In the experiments, 162 sets of atmospheric quality data are tested for the prediction performance. The prediction results are shown in Figure 5. According to the experiment setting, the input delay means the historical data used, and the output delay means the prediction steps. For the atmospheric quality data, the historical data of the latest 6 hours are used to output the prediction, and the prediction results are the SO_2 concentration in the next 6 hours. The data are used forward circularly. In Figure 5, the reference true value and the prediction results of various methods are presented with lines in different colours, and some parts are enlarged for the obvious comparison.

For the prediction results in Figure 5, all methods can trace the general trend of the SO_2 concentration data. The results of the ARIMA and RNN fluctuate more acutely than the others. The results of the CARN are closer to the true value so that the black line seems to be hidden in the figure.

For the obvious comparison of different methods, the errors are calculated and shown in Figure 6. The mean absolute error (MAE) and root-mean-squared error (RMSE) are selected as the evaluation indicators. The indicators are listed in Table 3.

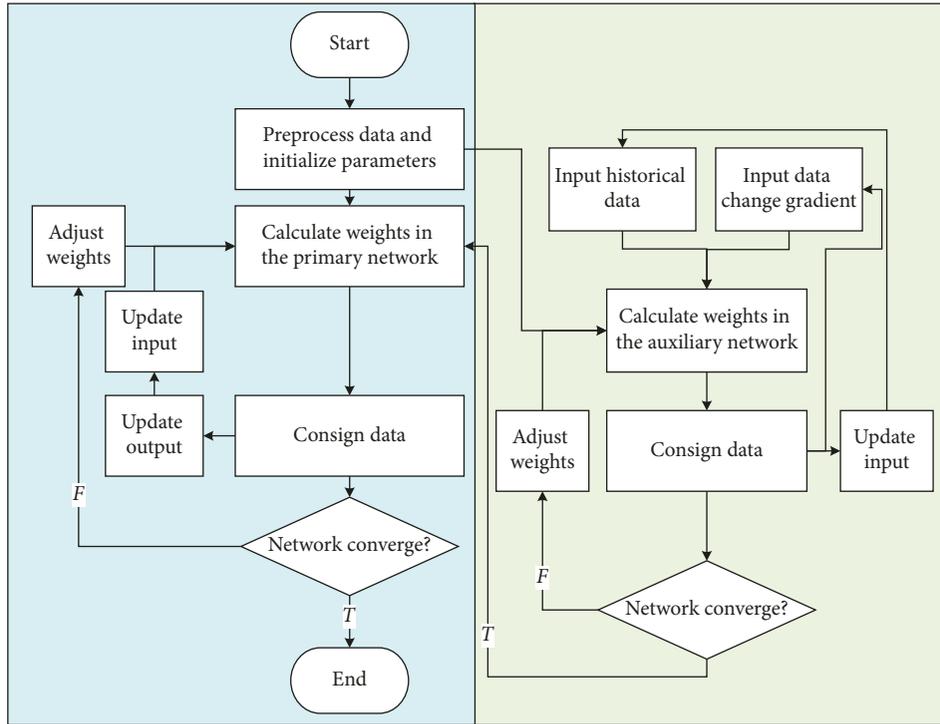


FIGURE 3: Prediction algorithm flow for the multivariate time series.

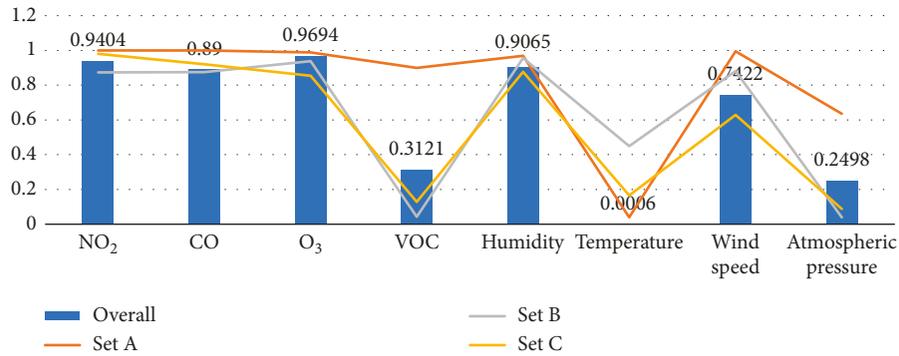
FIGURE 4: Correlation degrees between SO₂ and effect variables.

TABLE 1: Number of data in the experiments.

Datasets	Training data	Validation data	Test data
Atmospheric quality data (for each subset)	756	162	162
Meteorological forecast data	1008	216	216

The absolute errors show the similar trend of the prediction results in Figure 5. In the general view, the CARN performs more stably than other methods, in which the errors of the ARIMA and RNN change more sharply. The prediction performance can be evaluated objectively with the indicators in Table 3. The MAE is the average of all errors in their absolute value. In the indicator MAE, the CARN and LSTM perform better than the others. The MAE of results in the ARIMA is largest, while the RNN and BP show a similar MAE. The RMSE reflects the overall closeness of the results to the average value. It can indicate the stability of the

prediction methods. The sort of the RMSE in different methods is similar to the trend of the MAE, and the CARN is more stable than the others in prediction.

4.3. Results of Meteorological Forecast Data. For the prediction of meteorological forecast data, 1224 sets of data are used to train and verify the network. Then, 216 sets of data are set as the testing data. The 216 sets of the prediction results are shown in Figure 7 which also includes the reference true value and results of different methods. Different

TABLE 2: Parameters of the network structure and training.

Datasets	Atmospheric quality data		Meteorological forecast data	
	Primary network	Auxiliary network	Primary network	Auxiliary network
Number of networks	1	5	1	3
Number of training times	1000	1000	1000	1000
Learning rate	0.01	0.01	0.02	0.02
Convergence error	0.002	0.002	0.001	0.001
Input delay	1:6	1:6	1:12	1:12
Output delay	1:6	1:6	1:12	1:12
Number of inputs	6	10	4	6
Number of outputs	1	10	1	6
Number of hidden-layer neurons	10	14	7	10

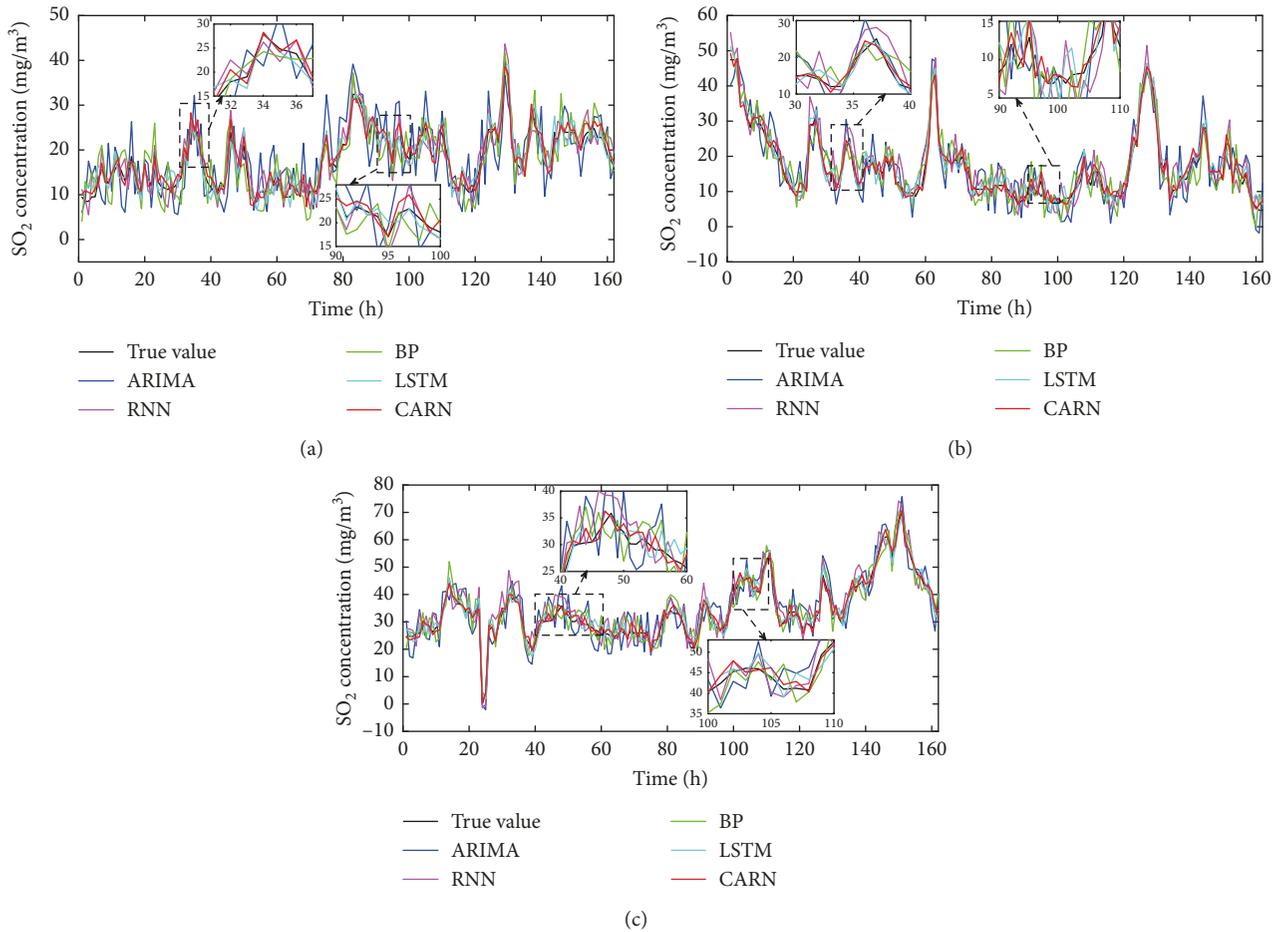


FIGURE 5: Prediction results of atmospheric quality data in three time periods: subset data (a) from June to August in 2016, (b) from September to November in 2016, and (c) from December in 2016 to February in 2017.

from the experiment of atmospheric quality data, the input and output delays are set to 12. The latest 12 sets of data are used to predict the temperature in the next 12 hours.

The data shown in Figure 7 present an obvious periodic trend. In fact, 216 sets of data are the meteorological data in 9 days. The temperature changes circularly in the period of one day. Then, the data change rule is more distinct. The prediction results of the CARN are closer to the true value than the others, in which the ARIMA and RNN fluctuate because they are predicted only with the

object variable and other methods use the object variable with effect variables.

The errors are calculated and presented in Figure 8 and Table 4. Figure 8 shows the errors of different prediction results. Table 4 lists the error evaluation indicators MAE and RMSE.

From the prediction results in Figure 7 and errors in Figure 8, it can be seen that all methods can trace the data change rule closely because of the periodicity in the meteorological data. The errors mainly occur in the fluctuation. The maximal MAE reaches 4.43°C in the ARIMA which is near to

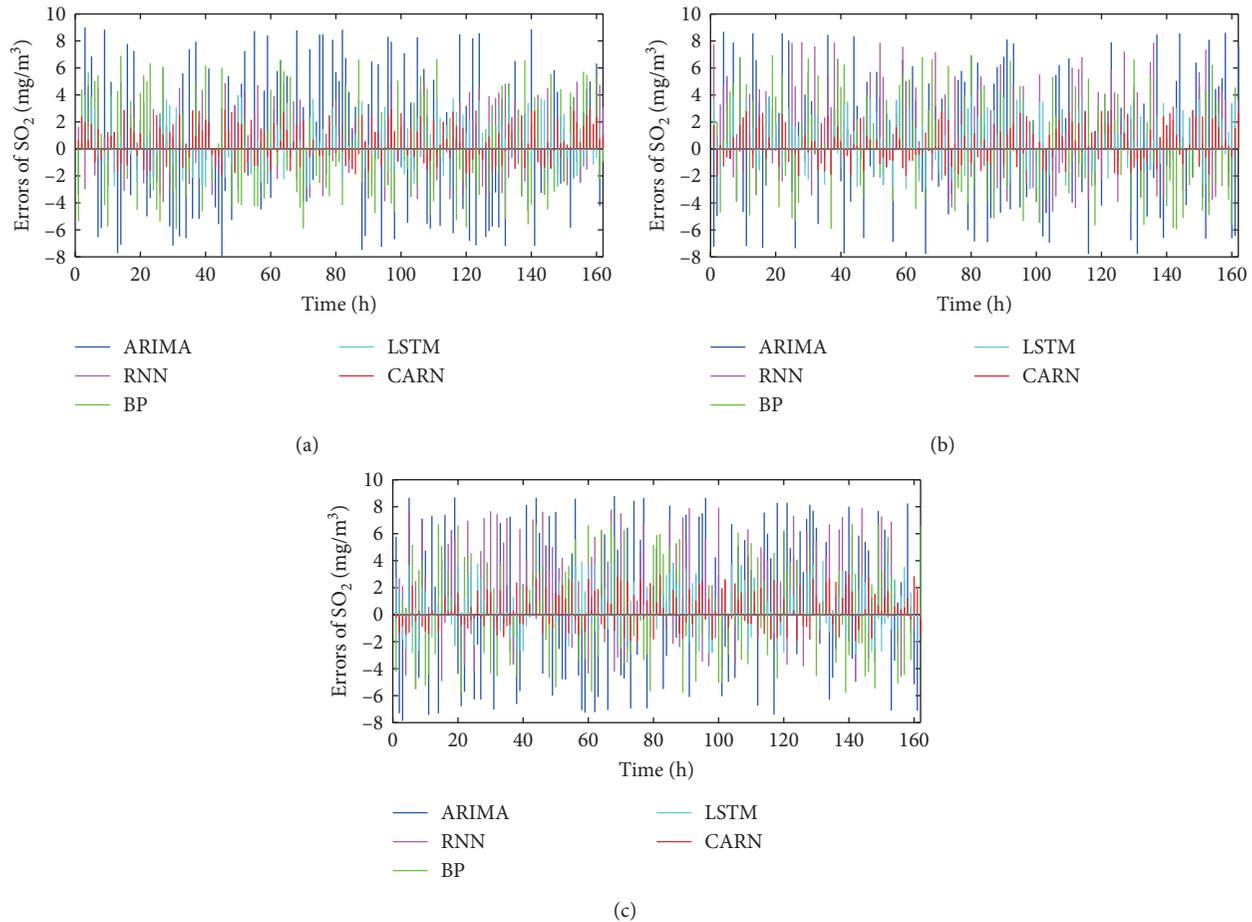


FIGURE 6: Errors of the prediction results of atmospheric quality data in different methods: subset data (a) from June to August in 2016, (b) from September to November in 2016, and (c) from December in 2016 to February in 2017.

TABLE 3: Error evaluation indicators of the prediction results of atmospheric quality data.

Data subsets		ARIMA	RNN	BP	LSTM	CARN
A	MAE	4.3617	2.3389	3.0917	1.6595	1.3275
	RMSE	5.0748	2.7257	3.6026	1.9791	1.5546
B	MAE	4.1136	2.9919	3.1235	1.8687	1.2421
	RMSE	4.7969	3.6575	3.6624	2.1634	1.4779
C	MAE	4.6690	3.3906	3.1084	1.7715	1.2660
	RMSE	5.3136	4.0724	3.6529	2.0841	1.4966

20% of the original measurement. The errors of the ARIMA, RNN, and BP exceed the usual expectation, while the errors of the CARN and LSTM (lower than 2°C) are acceptable.

5. Discussion

For the prediction issue of the multivariate time series, a compound network framework is introduced in which the structure of the nonlinear autoregressive network and the prediction algorithm are designed. The experiments are conducted within the environment data, including the atmospheric quality data and meteorological forecast data. The

prediction methods and results will be discussed in this section.

Firstly, the method shows the favourable short-term tracking performance in the data change rule. Generally, the prediction methods cannot avoid the divergency in the long term. It seems that there is not divergency in our prediction results. It is not that our approach is perfect, while the good regressive results derive from the setting of prediction time. The prediction time steps of the experiment are 6 and 12, which belong to the short-term prediction. The practical true values are imported into the model circularly to output the data in the future. Therefore, the prediction results show the good regressive effects. The results indicate that the proposed method can meet the short-term prediction need.

Secondly, the proposed method focuses on the prediction problem with multiple variables. For the accurate prediction, the related variables should be considered based on the target variable to be predicted. In the experiments, SO_2 and temperature are set as the object variable, and the related variables are selected as the effect variables. In the comparison methods, the ARIMA and RNN only use the object variable to predict the data themselves. The BP, LSTM, and CARN use multiple variables to obtain more accurate results. It is indicted that the effect variables help

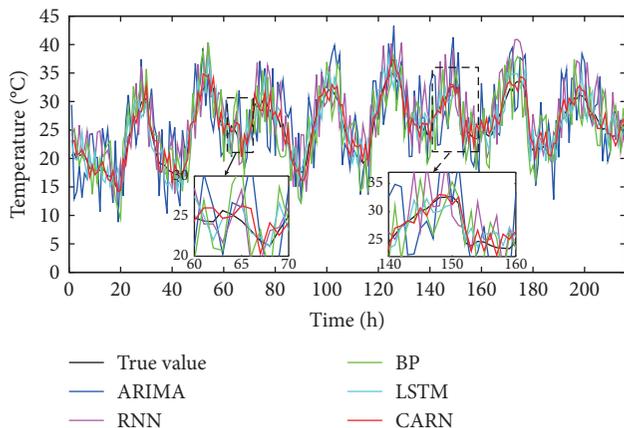


FIGURE 7: Prediction results of meteorological forecast data.

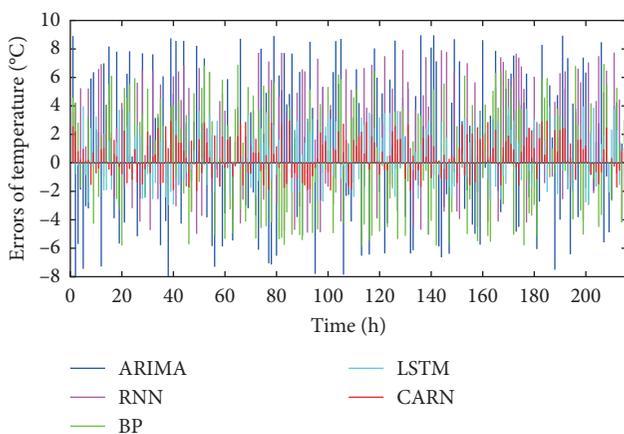


FIGURE 8: Errors of the prediction results of meteorological forecast data in different methods.

TABLE 4: Error evaluation indicators of the prediction results of meteorological forecast data.

	ARIMA	RNN	BP	LSTM	CARN
MAE	4.4331	3.3896	3.4048	1.7582	1.2564
RMSE	5.0999	4.0747	3.8864	2.0608	1.4829

improve the prediction performance. In the proposed method, the design of the auxiliary network meets the need of multivariate analysis.

Thirdly, the proposed method seeks the balance of precision performance and calculation resource occupancy. As mentioned in the introduction of related works, deep learning shows the excellent performance in prediction. It can be proved in the experiment where the result of the LSTM is similar to that of the CARN. However, the structure of the deep network is more complex than that of the network NAR, which may lead to the large consumption of the calculation resources. In the proposed method, networks based on the NAR are combined to obtain the expected prediction accuracy. Meanwhile, the simple structure of the NAR can reduce the calculation resource demand. The balance of accuracy and calculation resource in our method is beneficial to the application in practice.

The proposed CARN reaches the expectant effect in the time-series prediction. The effect is guaranteed with the compound structure of the primary and auxiliary networks to model the multivariable relation. Meanwhile, the training method in the CARN is also tested with experimental results based on the adjustment of the network parameters.

For the objective appraisal, the performance and application of the proposed network can be extended in the future. For the network performance, the training method is derived from the framework of backpropagation through time, which is an effective and simple solution in network learning. The related works on the backpropagation learning method are abundant. The improvement methods can be imitated based on the compound network structure. For the application, the proposed network can solve the direct prediction problems, such as the forecast of the weather, environment, economic market, and health management. It can also solve the data prediction in other complex systems indirectly. For example, the network may help the prediction of the control parameter in the nonlinear time-delay system [24]. The prediction result will be the important information for the control and management issues.

6. Conclusion

For the intelligent and advance management in the information era, the data-driven prediction method is studied in this paper. Considering the characteristic of the non-stationary and multivariate effect in the nonlinear time series, a compound prediction framework is designed based on the autoregressive neural network. The experiments on the environment data are conducted to verify the performance of the method. The method shows the favourable accuracy and appropriate calculation scale. The proposed network realizes the prediction of the multivariate. Besides, it takes the computational efficiency into account as well as the prediction performance. Furthermore, the principle of the network training in this paper is practical. It provides a feasible solution to the nonlinear multivariate time series with the shallow neural network. In the future work, the training method can be improved based on the advanced research, and the long-term prediction performance should be promoted. Moreover, the compound autoregressive network can be applied in other fields, including the direct forecasting of the time series and indirect prediction of the parameters and components in the complex systems.

Data Availability

The CSV data used to support the findings of this study are currently under embargo, while the research findings are commercialized. Requests for data, 6 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China under No. 2017YFC1600605, National Natural Science Foundation of China under Nos. 61673002 and 61903009, and Beijing Municipal Education Commission under Nos. KM201910011010 and KM201810011005.

References

- [1] H. Yao, W. Zhuang, Y. Qian et al., "Estimating and predicting metal concentration using online turbidity values and water quality models in two rivers of the Taihu Basin, Eastern China," *PLoS One*, vol. 11, no. 3, Article ID e0152491, 2016.
- [2] X. Liu, H. Liu, L. Wang et al., "The EFDC model integration and application in the Three Gorges reservoir," *Research of Environmental Sciences*, vol. 31, no. 2, pp. 283–294, 2018.
- [3] K. M. Mok, A. I. Miranda, K. U. Leong, and C. Borrego, "A Gaussian puff model with optimal interpolation for air pollution modelling assessment," *International Journal of Environment and Pollution*, vol. 35, no. 1, pp. 111–137, 2008.
- [4] G. E. P. Borrego and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco, CA, USA, 1976.
- [5] Y. Wang, C. Wang, C. Shi, and B. Borrego, "Short-term cloud coverage prediction using the ARIMA time series model," *Remote Sensing Letters*, vol. 9, no. 3, pp. 274–283, 2018.
- [6] Z. H. Xiao and M. Khashei, "A least squares-based parallel hybridization of statistical and intelligent models for time series forecasting," *Computers & Industrial Engineering*, vol. 118, pp. 44–53, 2018.
- [7] M. Khashei and Z. Hajirahimi, "A comparative study of series ARIMA/MLP hybrid models for stock price forecasting," *Communication in Statistics-Simulation and Computation*, vol. 48, no. 9, pp. 2625–2640, 2018.
- [8] L. Chen and H. Xu, "Autoregressive integrated moving average model in food poisoning prediction in Hunan province," *Journal of Central South University*, vol. 37, no. 2, pp. 142–146, 2012.
- [9] D.-w. Xu, Y.-d. Wang, L.-m. Jia, Y. Qin, and H.-h. Dong, "Real-time road traffic state prediction based on ARIMA and Kalman filter," *Frontiers of Information Technology & Electronic Engineering*, vol. 18, no. 2, pp. 287–302, 2017.
- [10] H. Yang, Z. Pan, and W. Bai, "Review of time series prediction methods," *Computer Science*, vol. 46, no. 1, pp. 21–28, 2019.
- [11] D. Pradeepkumar and V. Ravi, "Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network," *Applied Soft Computing*, vol. 58, pp. 35–52, 2017.
- [12] C. Daly, D. L. Moore, and R. J. Haddad, "Nonlinear autoregressive neural network model for forecasting Hi-Def H.265 video traffic over Ethernet passive optical networks," in *Proceedings of the IEEE SoutheastCon 2017*, pp. 1–7, IEEE, Charlotte, NC, USA, March 2017.
- [13] T. Wang, H. Gao, and J. Qiu, "A combined adaptive neural network and nonlinear model predictive control for multirate networked industrial process control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 2, pp. 416–425, 2015.
- [14] C. Liu, T. Shu, S. Chen, S. Wang, K. K. Lai, and L. Gan, "An improved grey neural network model for predicting transportation disruptions," *Expert Systems with Applications*, vol. 45, pp. 331–340, 2016.
- [15] B. Doucoure, K. Agbossou, and A. Cardenas, "Time series prediction using artificial wavelet neural network and multi-resolution analysis: application to wind speed data," *Renewable Energy*, vol. 92, pp. 202–211, 2016.
- [16] L. Wang, Y. Zeng, and T. Chen, "Back propagation neural network with adaptive differential evolution algorithm for time series forecasting," *Expert Systems with Applications*, vol. 42, no. 2, pp. 855–863, 2015.
- [17] P. Rodriguez, J. Wiles, and J. L. Elman, "A recurrent neural network that learns to count," *Connection Science*, vol. 11, no. 1, pp. 5–40, 1999.
- [18] A. Graves, S. Fernández, and J. Schmidhuber, "Multi-dimensional recurrent neural networks," *Lecture Notes in Computer Science in Proceedings of the International Conference on Artificial Neural Networks*, vol. 4668, pp. 549–558, 2007.
- [19] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [21] A. Graves, S. Fernández, and J. Schmidhuber, "Bidirectional LSTM networks for improved phoneme classification and recognition," in *Proceedings of the International Conference on Artificial Neural Networks*, pp. 799–804, Munich, Germany, September 2005.
- [22] K. Cho, B. V. Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1–14, Lisbon, Portugal, September 2014.
- [23] R. Matkovskyy and T. Bouraoui, "Application of neural networks to short time series composite indexes: evidence from the nonlinear autoregressive with exogenous inputs (NARX) model," *Journal of Quantitative Economics*, vol. 17, no. 2, pp. 433–446, 2019.
- [24] T. Wang, J. Qiu, and H. Gao, "Adaptive neural control of stochastic nonlinear time-delay systems with multiple constraints," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 1875–1883, 2017.



Hindawi

Submit your manuscripts at
www.hindawi.com

