IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON MOBILE SERVICE COMPUTING WITH INTERNET OF THINGS

Received April 1, 2019, accepted May 5, 2019, date of publication May 16, 2019, date of current version June 10, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917277

Air Quality Forecasting Based on Gated Recurrent Long Short Term Memory Model in Internet of Things

BAOWEI WANG[1,2,3], WEIWEN KONG[1], HUI GUAN[1], AND NEAL N. XIONG[4]

[1]School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China
[2]Engineering Research Center of Electronic Information and Control, Fujian Provincial Education Department, Fuzhou 350108, China
[3]Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing 210044, China
[4]Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK 74464, USA

Corresponding author: Baowei Wang (wang@nuist.edu.cn)

This work was supported in part by the Major Program of the National Social Science Fund of China under Grant 17ZDA092, in part by the Electronic Information and Control of Fujian University Engineering Research Center Fund under Grant EIC1704, in part by the National Key R&D Program of China under Grant 2018YFB1003205, in part by the National Natural Science Foundation of China under Grant 61173136, Grant U1836208, Grant U1536206, Grant U1836110, Grant 61602253, and Grant 61672294, in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund, and in part by the Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET) Fund, China.

ABSTRACT With the continuous development of the Chinese economy and the gradual acceleration of urbanization, it has caused tremendous damage to the environment. The bad air environment seriously damages the physical and mental health of the people. The change in smog concentration will be affected by many realistic factors and exhibit nonlinear characteristics. The method proposed in this paper is to use the Internet of Things (IoT) technology to monitor the acquired data, process the data, and predict the next data using a neural network. The existing prediction models have limitations. They don't accurately capture the law between the concentration of haze and the factors affecting reality. It is difficult to accurately predict the nonlinear smog data. One algorithm proposed in this paper is a two-layer model prediction algorithm based on Long Short Term Memory Neural Network and Gated Recurrent Unit (LSTM&GRU). We set a double-layer Recurrent Neural Network to predict the PM2.5 value. This model is an improvement and enhancement of the existing prediction method Long Short Term Memory (LSTM). The experiment integrates data monitored by the IoT node and information released by the national environmental protection department. First, the data of 96 consecutive hours in four cities were selected as the experimental samples. The experimental results are close to the true value. Then, we selected daily smog data from 2014/1/1 to 2018/1/1 as a train and test dataset. It contains smog data for 74 city sites. The first 70% of the data was used for training and the rest for testing. The results of this experiment show that our model can play a better prediction.

INDEX TERMS Air pollution, Internet of Things, forecasting, LSTM, GRU.

I. INTRODUCTION

In recent years, with the rapid development of scientific information technology, IoT [1] technology has been widely used in various fields in China. With the in-depth study of relevant personnel, the IoT technology has been continuously improved and gradually matured [2]. The practical application of the IoT technology in the current stage of environmental monitoring has played an important role in the ecological environmental protection work and has a broad development prospect.

Air quality has become a topic of common concern to the whole society. Air pollution seriously affects the health of Chinese residents. Seriously, it increases the incidence of lung diseases and cancer [9]. It can cause abnormal development of infants. It is harmful to humans after inhaling people's respiratory tract. Long-term inhalation can lead to death. The bad air environment seriously damages the physical and mental

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

health of people. This has a great impact on improving the living conditions and quality of life of our people. Therefore, it is very important to predict that smog will become very important.

At present, the problem of atmospheric environmental pollution has become an important issue that restricts China s economic and social development. If this problem is not resolved in timely effective manner, it will seriously threaten the health of people and the development of China's economy and society. High-quality air pollution prediction can accurately grasp the level of the environment in the future and form complete and accurate prediction information. Therefore, it provides a scientific basis for the national prevention of air pollution and makes the flood prevention work more targeted and effective. In this context, an effective way to solve the problem of air pollution in China is to improve the accuracy of air pollution prediction.

The smog problem is serious [9]. It is still a difficult problem to accurately predict PM2.5 in the face of smog. The final result is not satisfactory. There are limitations in the monitoring and prediction of smog. The Municipal Air Quality Index is published daily by the Bureau of Meteorology. This index can only generally describe the overall air quality of a city. However, the air quality in urban and suburban areas and in different areas of the same city is different. Because of the fluidity of haze, the PM2.5 values in different regions of the same city are also different. Even indoor and outdoor values are different. During the implementation of environmental monitoring, the high cost of the base station and the difficulty of power supply affect the data acquisition.

Because of the many factors involved, the final result is unpredictable. With the demand for scientific research becomes higher and higher, more accurate monitoring quality is required. At present, China's largest IoT is the environmental pollution monitoring system. China's various provinces and cities have gradually established a pollution protection system. Relevant monitoring centers have been established in various places. China has gradually formed a smog management strategy integrating monitoring, information adoption, and data prediction. These measures have led to an overall improvement in the quality of environmental protection management.

The method proposed in this paper is based on the air quality monitoring and prediction system of the IoT [28]. It monitors air quality in real time and even gives the user a value every minute. This value reflects the air quality status in real time. The future air quality is predicted based on the monitored values. Therefore, through the IoT technology, wireless networking and other technologies to obtain node data to achieve national air quality detection, analysis and prediction. Some low-power air quality monitoring modules are installed in different areas. These monitoring modules are safe and energy efficient and can detect PM2.5 metrics. It can greatly improve the efficiency of environmental regulation and protection.

The rest of this article is organized as follows. The second section introduces the work related to IoT monitoring and data forecasting. The third section introduces the principles of IoT monitoring and data prediction. The fourth section introduces the experimental process and implementation steps and gives a predictive evaluation of the experimental results. Finally, we give concluding observations and future work.

II. RELATED WORK

Environmental monitoring technology is a monitoring and early warning technology. It promotes the development of ecological society. It has been in the development of our society for decades. Environmental monitoring technology has developed a relatively mature development in western developed countries and has been widely used. The specific application of environmental monitoring technology in China is to better promote environmental protection. The ecological environment forms a good monitoring and governance system.

During the 2008 Beijing Olympics, 11 prefecture-level cities under the jurisdiction of Beijing, Tianjin and Hebei Province established the joint control mechanism for air pollution joint defense for the first time [3], [27]. 13 urban areas implement unified standards, unified deployment, unified management, and unified decision-making. At this time, the monitoring and control of atmospheric environmental quality achieved very good results. It provides valuable experience for subsequent nationwide air pollution control. The state has increased its strength and depth in management. However, it is not mature in the technology of flood control and treatment. Excessive monitoring data makes it difficult to distinguish between spammers and regular users. To solve this problem.

The existing state-controlled environmental air monitoring points in the Beijing-Tianjin-Hebei region mainly include 80 national control city evaluation points and 11 regional points in 13 prefecture-level and above cities. The number of national control monitoring points in each city is 3 to 13 Three of the regional locations are located in Beijing, three in Tianjin and four in Hebei Province. In addition to the national control monitoring points, each city has also set up a certain number of provincial control points. They are mainly distributed in the built-up areas at the district and county levels to jointly reflect the urban ambient air quality.

There are many researches related to the Internet of Things [18], [23], [29]. Minoli *et al.* [28] introduced the concept and current status of the IoT and discussed the shortcomings of the traditional environmental online monitoring system. Internet of Things monitoring data can cause network congestion problems. To solve this problem, Qiu *et al.* [12] proposed an event scheduling scheme EABS for EIOT. Wireless sensor networks (WSNs) have been used to build intelligent meteorological observation networks. Wang *et al.* [7] proposed an air temperature error correction scheme (STCS). After preprocessing such as interpolation and time conversion, the trained neural network is applied to temperature correction. Wang *et al.* [16] propose an effective dual-chaining watermark scheme, called DCW, for data integrity protection in smart campus IoT applications. Shi *et al.* [17] propose IoT device identification via Radio Frequency Fingerprinting (RFF) extracting from radio signals which is physical-layer method for IoT security which solve security problems in this paper. Therefore, IoT technology is mature in environmental monitoring technology.

The method proposed in this paper is to use the IoT technology to monitor the acquired data and process the data. The obtained data is predicted using a neural network. One algorithm proposed in this paper is a two-layer model prediction algorithm based on LSTM&GRU of recurrent neural network (RNN). This algorithm is an improvement and enhancement of the existing prediction method LSTM [4]–[6]. This two-layer recursive neural network model replicates the first loop layer in the network. The input sequence provides input for the first layer. The reverse copy of the input sequence is passed to the second layer GRU [8]. To overcome the limitations of traditional RNN [10], [26], [27], we propose a two-layer recurrent neural network. It can be trained using past smog data. The model can predict future data more accurately based on past smoke data.

There are many methods to predict air pollution [19]–[22], [24]. Yang *et al.* [10] used discrete wavelet analysis to decompose sulfur dioxide and soot emissions into high frequency channels and low frequency channels. These two channels establish a periodic function and fit the periodic characteristics of the Fourier curve. Finally, a higher prediction effect was obtained. Fu et al. [11] updated the prediction method online by multiple linear regression models. The model is continually updated based on the test results of the day. In the absence of a large amount of forecast data, changes in smoke conditions are reflected in a timely manner.

Ordieresa *et al.* [13] used three neural network models to predict PM2.5. They are Multilayer Perceptron (MLP), Radial Basis Function (RBF), Square Multilayer Perceptron (SMLP). All of these models are based on BP neural networks. The results show that RBF networks can better predict PM2.5 results than other networks. Barai *et al.* [14] investigated some air quality prediction methods. He detected the sequential network construction model (SNCM), the RNM variable point modeling model (CPDM), the cyclic network model (RNM), and the self-organizing feature map(SOFM). It showed that the SOFM has better forecasting performance and neural networks have sufficient capabilities for air quality forecasting to handle noise and error data. Corani *et al.* [15] find that Ozone and PM10 have a great correlation with PM2.5. In some cases, lazy learning can provides better performances on samples and PNNs are superior to the other approaches in detecting of the exceedances of alarm and attention thresholds. Zhao et al. [25] established a Long short-term memory - Fully connected (LSTM-FC) neural network to predict PM2.5 of a specific air quality monitoring station over 48 hours using historical air quality data. Compared with ANN and LSTM, the result shows that our proposed model gives better predictive results based on RMSE and MAE metric.

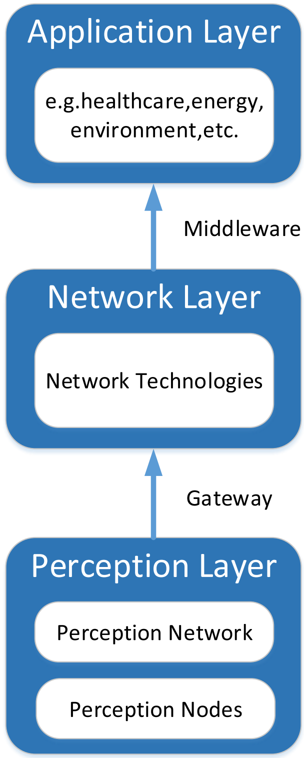

FIGURE 1. **The IoT monitoring system model.**

III. THE SMOG MONITORING AND FORECASTING SCHEME

A. OVERVIEW OF THE MONITORING SYSTEM MODEL

First, we obtain data through the IoT node. The IoT monitoring system consists of a perception layer, a network layer, and an application layer. The structure of the IoT is shown in Fig. 1.

1) PERCEPTION LAYER

The main function of the perception layer is to obtain information about environmental monitoring through sensing devices such as sensor nodes. The wireless sensor network technology forms an autonomous network and adopts a collaborative work mode. It extracts useful information and enables resource sharing and communication with other devices on the Internet through access devices. The perception layer is the base layer and core layer of the IoT technology.

2) NETWORK LAYER

The main function of the network layer is to transfer information from the sensing layer to the Internet. It is an Internet platform built on the core of IPV6/IPV4. It manages and controls the massive amount of information acquired by environmental monitoring within the network through a large central computer platform and provides a good user interface for upper-layer applications. The network layer is mainly responsible for transmitting the information collected by the sensing layer and finally outputting it to the application layer. It can also be said to be the brain of the IoT. There may be a delay in the data transfer process.

3) APPLICATION LAYER

The main function of the application layer is to summarize and convert data. It has the functions of the underlying system and builds practical applications for the environmental monitoring industry to achieve real-time monitoring and early warning. Data processing is performed by supporting sub-layers of the platform. Various environmental indicators are monitored in real time and a large amount of data is obtained.

B. MONITORING SYSTEM MODEL COMPOSITION

The IoT monitoring data is transmitted to the server through a wireless transceiver module packaged inside the processor. This achieves real-time detection of environmental pollution. Its monitoring system mainly includes three parts: monitoring node, server side, and user side. The wireless router connects them. The router connects the node to the server. The PM2.5 information can be obtained through the node to access the environmental data collected from the distributed monitoring points.

1) DISTRIBUTED MONITORING NODE

The function of the monitoring node is to read the data of the current environment from various sensors and send the data to the local server. It monitors PM2.5 through the useąŕs remote control node and automatically records PM2.5 concentration data. The monitoring side uses the sensor to process the data collected by PM2.5 to further determine the air quality. It makes corresponding decisions and estimates based on the description and assessment of integrated air quality.

2) SERVICE-TERMINAL

The main function of the local server is to process and forward data. It is the hub for data transmission and processing throughout the system. The server processes the information collected by the sensor and saves the results to the server.

3) USER TERMINAL

The data monitored by the IoT is integrated through the processing of the server to obtain complete data. The user processes, analyzes, and stores various data and parameters collected by each monitoring node in the area. The client implements real-time detection of the air quality in the area and transmits the monitoring result to the user's client.

C. DATA PREDICTION MODEL

The algorithm used in this paper is a multi-layer RNN structure.The first layer uses LSTM and the second layer uses GRU.

In our model, we divide the architecture into two RNN layers so that the model can improve the accuracy with little increase of complexity. The air quality data is continuous and applies to the RNN model. This is also the reason for choosing LSTM&GRU.

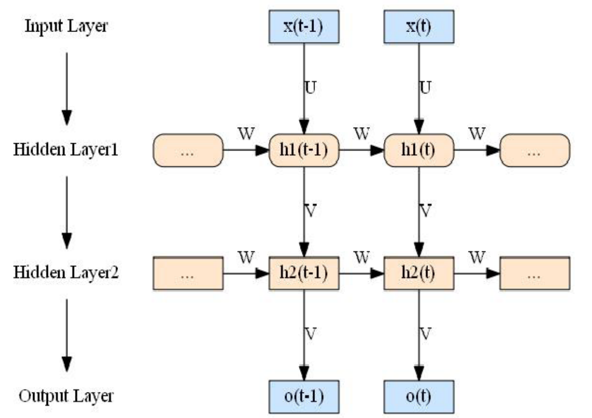

FIGURE 2. **LSTM&GRU architecture, where $h_1(t)$ means LSTM architecture and $h_2(t)$ means GRU architecture.**

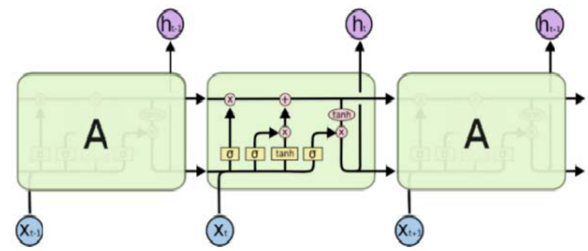

FIGURE 3. **The repeating module in an LSTM contains four interacting layers.**

1) LONG SHORT TERM MEMORY NETWORK

The first layer uses LSTM because the cells in the LSTM will judge the information as they enter the model. Information that conforms to the rules will be retained and information that does not meet the requirements will be forgotten. Using this principle, the problem of long sequence dependence in neural networks can be solved. LSTM is an excellent variant of RNN. It inherits the advantages of the RNN model and solves the problem of the disappearing gradient caused by the gradual decrease of the gradient back propagation process.

As shown in Fig. 2, there is a circular chain in all circulating neural networks. Each unit of the chain structure processes the input information and outputs it. The unit of the chain structure in the conventional recurrent neural network contains only one nonlinear function to nonlinearly transform the data, such as the tanh function. In the recurrent neural network of RNN variants such as LSTM, each unit adds different nonlinear functions at different locations to achieve different functions and solve different problems.

LSTM adds three sigmoid functions to each small unit. They are forget gate, input gate and output gate. These three functions are used to implement the gating function to control the inflow and outflow of data.

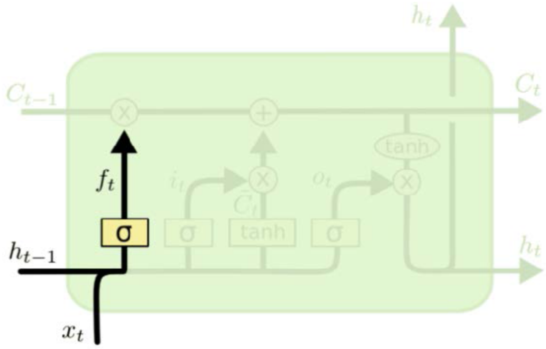

FIGURE 4. **The Forget Gate in LSTM.**

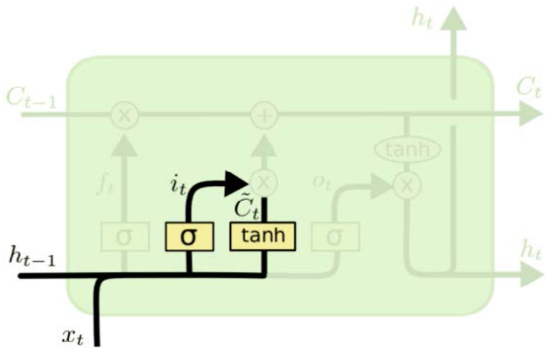

FIGURE 5. **The input gate in LSTM.**

The first gate in LSTM architectuer is the forget gate, which is shown as Fig. 4. The function of the forget gate is to determine the information to be discarded in the memory cells. It determines how much information in the memory cell C_{t-1} from the previous moment can continue to be passed to the memory cell C_t at the current moment. The mathematics of memory cells is formulated as (1).

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ &= \sigma(W_{hf} h_{t-1} + W_{xf} x_t + b_f) \end{aligned} \tag{1}$$

where σ is the logistic sigmoid function, which is often used as a threshold function for neural networks, mapping variables between 0 and 1. f_t means the forget gate at the time t, h_t means the hidden vector which is the same size as f_t, the W terms denote weight matrices(e.g. W_{xf} is the forget-hidden weight matrix), the b_f terms denote bias vectors in the forget gate, $x = [x_1, x_2, \cdots, x_t, \cdots, x_n]$ is an input sequence. (1) shows when we see a new subject through the Cell C_t, we want C_t to forget the old subject.

The second gate is the input gate, which is shown as Fig. 5. Its main function determines how much new input information can enter the memory cell C_t at the current moment. The mathematics of memory cells is formulated as (2).

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ &= \sigma(W_{hi} h_{t-1} + W_{xi} x_t + b_i) \end{aligned} \tag{2}$$

where i_t means the input gate at the time t, the b_i terms denote bias vectors in the input gate, the W terms still denote weight matrices. Then we need update the old Cell C_{t-1}, the details will be discussed below.

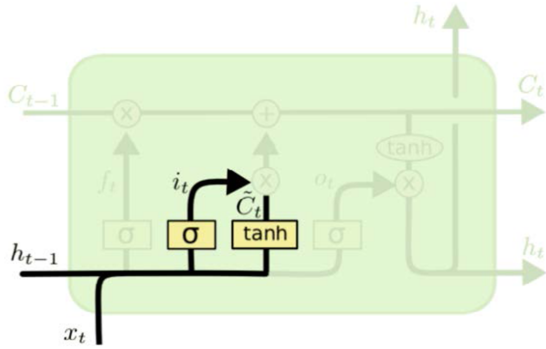

FIGURE 6. **The output gate in LSTM.**

The third gate is the output gate, which is shown as Fig. 6. Its main function is to determine whether the information in the memory cell C_t at the current moment can enter the hidden layer state h_t. The mathematics of memory cells is formulated as (3).

$$\begin{aligned} o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ &= \sigma(W_{ho} h_{t-1} + W_{xo} x_t + b_o) \end{aligned} \tag{3}$$

where o_t means the input gate at the time t, the b_o terms denote bias vectors in the output gate.

Then the LSTM architecture need to update the memory cell. Updating memory cells need use forget gates and input gates according to the structure of LSTM. The candidate memory cell C_t is generated using the tanh function as information to enter the input gate. The mathematics of memory cells is formulated as (4).

$$\begin{aligned} C_t &\approx \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ &= \tanh(W_{hC} h_{t-1} + W_{xC} x_t + b_C) \end{aligned} \tag{4}$$

where C_t means the cell state at the time t. The b_c terms denote bias vectors in the cell state.

The output of the forget gate is multiplied by the old state to determine how much old state information can enter the new memory cell. The output of the input gate is multiplied by the candidate memory cells to determine how much information is to be updated. They are linearly added to obtain memory cells that have been updated at the current time. The process of accumulating historical information is shown as (5).

$$\begin{aligned} \tilde{C}_t &= f_t C_{t-1} + i_t C_t \\ &\approx f_t C_{t-1} + i_t \tanh\left(W_{hC} h_{t-1} + W_{xC} x_t + b_C\right) \end{aligned} \tag{5}$$

where i_t means the input gate, i_t means the forget gate.

At last the model will output the information. The data is processed by the tanh function ($\tanh(C_t)$) and normalized to the interval [-1, 1]. The result of the output gate is then multiplied by the normalized data to control the amount of data output to the hidden layer. This process is formulated as (6).

$$h_t = o_t * \tanh(C_t) \tag{6}$$

Among the many models available for time series prediction, traditional prediction models such as the autoregressive model (AR) and its upgraded differential autoregressive moving average model (ARIMA) have a good theoretical basis. They are better at capturing the dynamics of systems with linear relationships. However, for nonlinear systems, deep learning-based RNN performs better than its variant LSTM. RNN is a neural network with a feedback structure whose output is related not only to the current input and the weight of the network, but also to the input of the previous network. LSTM is an excellent variant of RNN that inherits the features of most RNN models. It can train long-term dependency information and can solve the vanishing gradient problem caused by the gradual reduction of the gradient back-propagation process. LSTM is ideal for dealing with issues that are highly correlated with time series. Therefore, the data prediction in this system uses the LSTM neural network to solve the first layer of the smog concentration problem.

2) GATED RECURRENT UNIT

In this paper, the second layer of the two-layer model uses GRU. Compared to LSTM, the construction of GRU is simpler because it reduces one gate and matrix multiplication is less strained. Therefore, the GRU can save a lot of time when the training data is large. Because GRU has fewer parameters, it is faster to train and requires fewer samples. While LSTM has more parameters, it may get a better model. Therefore, LSTM is more suitable for situations with a large number of samples. As a variant of LSTM, GRU combines the forget gate and the input gate into a single update gate. The cell state and the hidden state are also mixed and some other changes are added. GRU made two major changes on the basis of LSTM.

First is the decrease of the number of gates. GRU combines the input gate and the forget gate in the LSTM into an update gate. The other is reset gate. The role of the reset gate is to control how much information is forgotten. The first gate is update gate. The main function of the update gate determines how much past information can continue to be passed to the future. The information of the previous moment and the current moment are linearly transformed by the right multiplication weight matrix. The added data is sent to the update gate and multiplied by the sigmoid function. The resulting value is between [0, 1]. This process is formulated as (7).

$$\begin{aligned} z_t &= \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \\ &= \sigma(W_{hz}h_{t-1} + W_{xz}x_t + b_z) \end{aligned} \tag{7}$$

where z_t means the update gate and the b_z terms denote bias vectors in the update gate. The update gate determines whether to update the hidden state to the new state, which has the same function as the output gate in the LSTM.

The second gate is the reset gate. The main function of the reset gate is to determine how much historical information cannot be passed to the next moment. As with the data processing of the update gate, the information of the previous moment and the current moment are linearly transformed. The values and usefulness of the weight matrix of these two data processing are different.

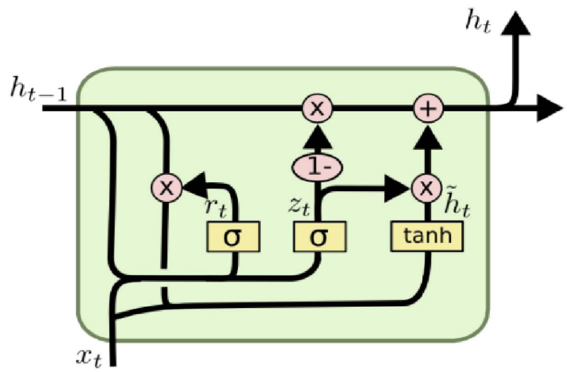

FIGURE 7. GRU architecture.

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \\ &= \sigma(W_{hr}h_{t-1} + W_{xr}x_t + b_r) \end{aligned} \tag{8}$$

where r_t means the reset gate and the b_r terms denote bias vectors in the reset gate.

The second difference between GRU and LSTM is that GRU cancels the linear self-renewing memory unit. It directly performs linear self-updating using gating directly in the hidden unit. The logic diagram of the GRU is shown as Fig. 7.

GRU no longer uses separate memory cells to store memory information. It directly records the historical state using hidden units. The amount of data of the current information and the memory information is controlled by the reset gate. Generating new memory information continue to pass forward.

$$\begin{aligned} h_t &\approx \tanh(W \cdot [r_t h_{t-1}, x_t] + x_t) \\ &= \tanh(r_t W_h h_{t-1} + W_x x_t) \end{aligned} \tag{9}$$

As shown in (9). When the reset gate is 0, the memory information is completely cleared. Conversely, when the reset gate is 1, it means that the memory information is all passed.

In summary, GRU is better at convergence time. Double-layer prediction is determined by several inputs before and after. The double-layered recurrent neural network predicts data more accurately.

IV. EXPERIMENT

A. DATA PREPROCESSING

The experiment integrates data monitored by the IoT node and information released by the national environmental protection department. It contains smog data for 74 city sites. For the accuracy of the experiment, all the samples were selected as the research object in this experiment. We selected daily smog data from 2014/1/1 to 2018/1/1 as a train and test data set. The first 70% of the data was used for training and the rest for testing. We use the average to replace the missing data if some data is missing. First, the data is subjected to minimum-maximum normalization to process the data so that

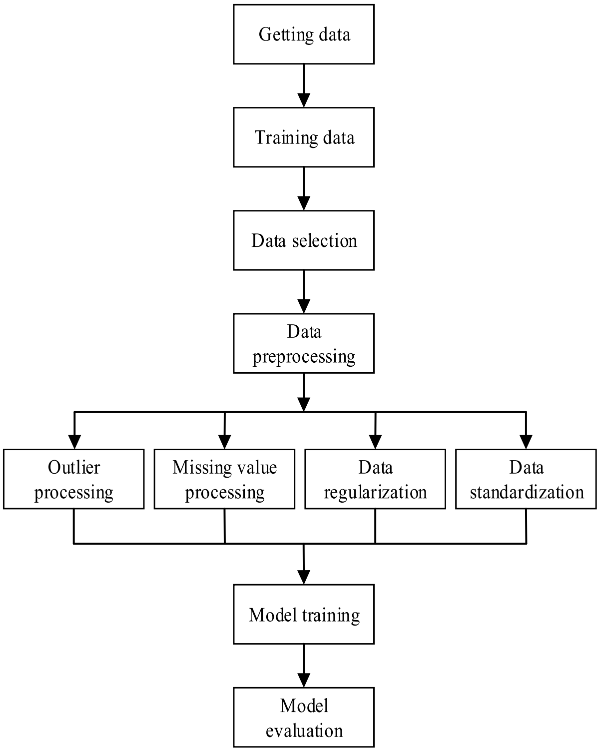

FIGURE 8. The flow chart of date preprocessing.

each data can range from 0 to 1. Then, the smog data is regularized so that the data falls within a certain interval of the statistics. In the experiment, we used the PM2.5 value of the first 5 days as a training unit to predict the PM2.5 of the next day and established a training model. After training the model, we used the same method to test the rest data. Both the training and testing indicators use RMSE to measure accuracy. Obviously the smaller the RMSE value, the more precise the model. All the experiments were run in CPU Ryzen 7 2700X, memory 8G and GPU-free environment. In order to prevent over-fitting, we set epoch as 200, learning rate as 0.15.

1) NORMALIZE

In order to eliminate the dimensional impact between indicators, data standardization is required to resolve the comparability between data indicators. After the original smog data is processed by data standardization, each indicator is in the same order of magnitude. The processed data is suitable for comprehensive comparative evaluation. The normalized method used in this experiment is min-max standardization. It also known as dispersion standardization.

It is a linear transformation of the raw data so that the resulting value is between [0, 1]. The conversion function is shown as (10).

$$x^* = \frac{x - min}{\mathrm{min} - \mathrm{min}} \tag{10}$$

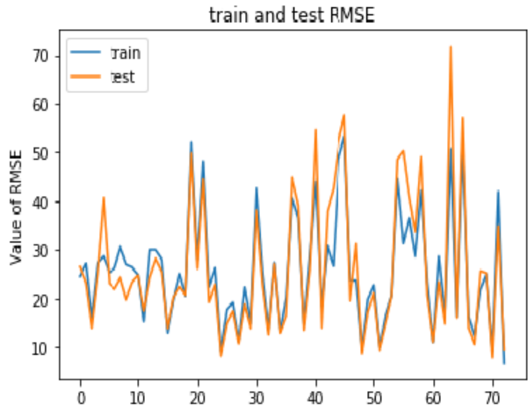

FIGURE 9. The train RMSE and the test RMSE.

where max is the maximum value of the sample data and min is the minimum value of the sample data.

2) STANDARDIZATION

The normalization of data is to scale the data down to a specific interval. Data standardization is often used in the processing of certain comparison and evaluation indicators. The unit limit of the smog data is removed and converted to a dimensionless, pure value. It is convenient for indicators of different units or magnitudes to be compared and weighted. The steps for data standardization are shown as below:

a.Find the average value x_i and standard deviation s_i of each variable.

$$x^* = \frac{x - \mu}{\sigma} \tag{11}$$

b.Standardize the data:

$$z_{ij} = \frac{x_{ij} - x_i}{s_i} \tag{12}$$

where z_{ij} is the normalized variable value, x_{ij} is the actual variable value.

c.Reverse the sign before the inverse indicator.

The normalized variable values fluctuate around zero. Greater than 0 indicates above average and less than 0 indicates below average. Normalization is the scaling of a feature rather than the feature vector of a sample.

B. EXPERIENCE RESULT

1) THE PREDICTION OF 74 CITIES

In order to predict whether the LSTM&GRU architecture is correct or not, we use the PM2.5 values of 74 cities for training and testing. The experimental indicators use the RMSE value. If the architecture is correct, the RMSE of the test dataset should be less than the one of the train dataset.

In Fig. 9 we can find it is not clearly whether the train RMSE is larger or the test RMSE, but we can easily find that most of cities has a less test RMSE than train RMSE, which can prove that the LSTM&GRU architecture is effective.

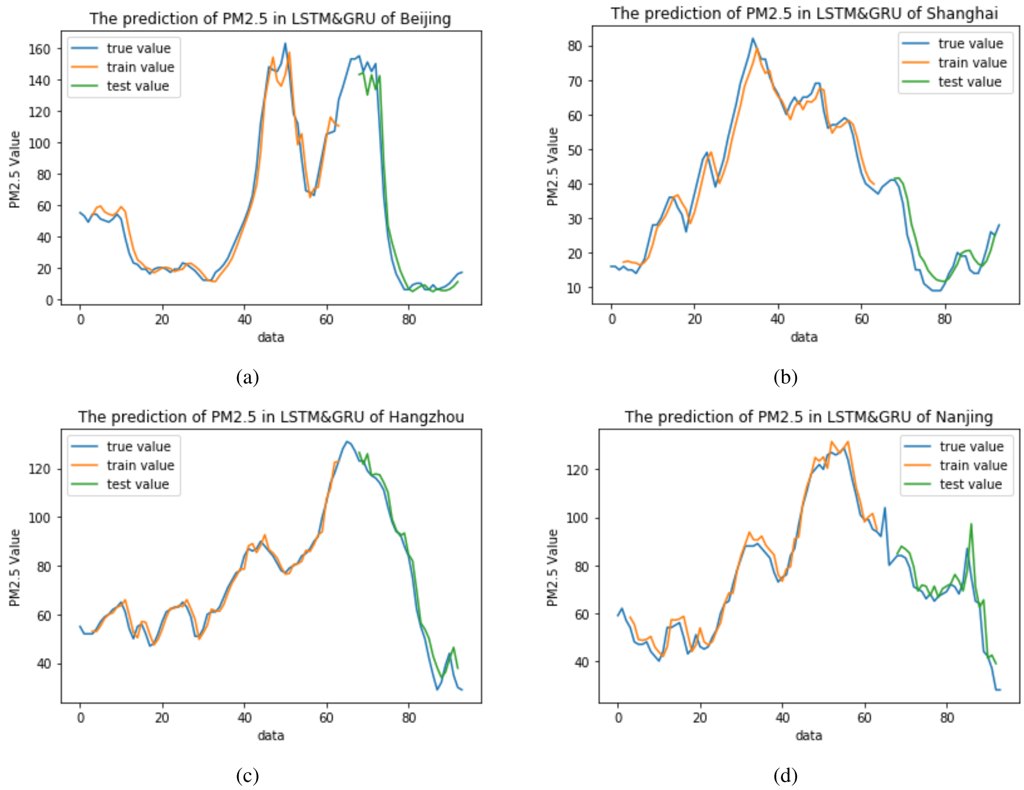

FIGURE 10. The prediction of PM2.5 in four cities. (a) The prediction of PM2.5 in Beijing. (b) The prediction of PM2.5 in Shanghai. (c) The prediction of PM2.5 in Hangzhou. (d) The prediction of PM2.5 in Nanjing.

2) THE PREDICTION OF FOUR CITIES IN CONTINUOUS 96 HOURS

We selected Beijing, Shanghai, Hangzhou and Nanjing four cities, arranged nodes in these cities, and successfully collected 4days, 96 hours of data in the period of 2019/1/1-2019/1/4. Then we use our model to predict the value of PM2.5 in each cities and compare the deviation of the prediction value of each city. In the experiment, we divide the first 70% data as our training dataset and the remaining 30% as our test dataset. The data every 5 hours is used as a training unit to get the PM2.5 predicted value of the next hour and compare with the real value. The method of measuring the contrast error is still measured by RMSE. The experimental results are shown in Fig.8. We can find the three lines basically match which means our model has a better performance to predict future data. In the four cities above, the models obtained good prediction results.So we can almost assume that our model can accurately predict the change of PM2.5.

3) COMPARING WITH LSTM

In this paper, we use the PM2.5 values of 74 cities for training and testing. The experimental indicators use the RMSE value. Since the LSTM predictions of some cities are too close to the RMSE values predicted by the methods in this paper, we use a completely new method to judge the pros and cons of the two models. In the original experiment, we find there

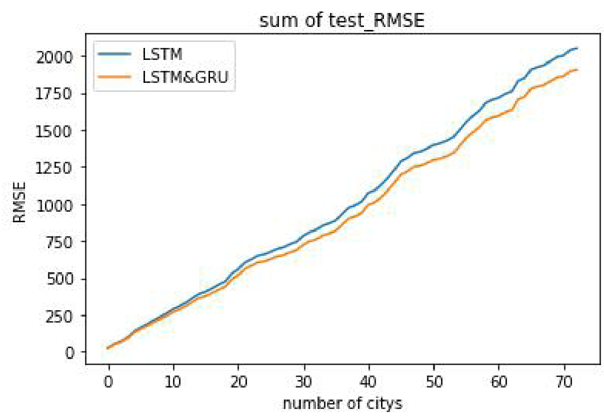

FIGURE 11. Sum of test RMSE, the blue line indicates LSTM and the orange one indicates ours.

is little difference of the two model in each city. To describe the difference between the two model rapidly, we accumulate the RMSE values trained in 74 different cities in two different models. Through continuous accumulation, we can find that the sum of the RMSE values of the method is gradually smaller than the sum of the RMSE values of LSTM, so that the method is superior to the overall method.

Fig. 11 and Fig. 12 show that with the increase of the number of cities, the gap between the two becomes bigger and

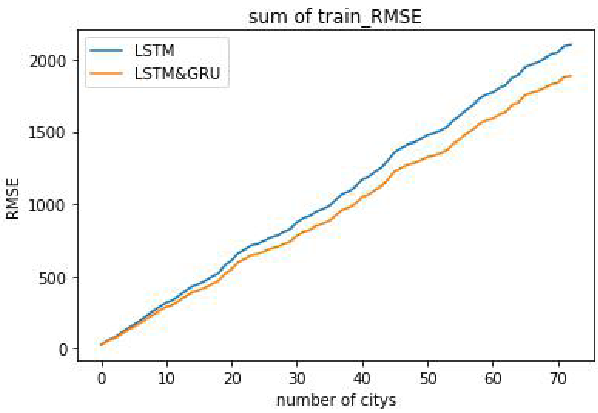

FIGURE 12. Sum of train RMSE, the blue line indicates LSTM and the orange one indicates ours.

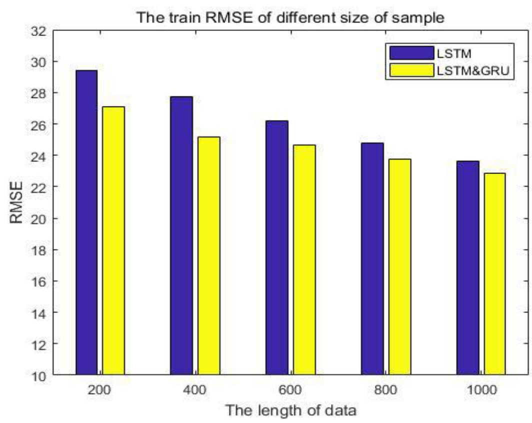

FIGURE 13. The train RMSE of different size of samples.

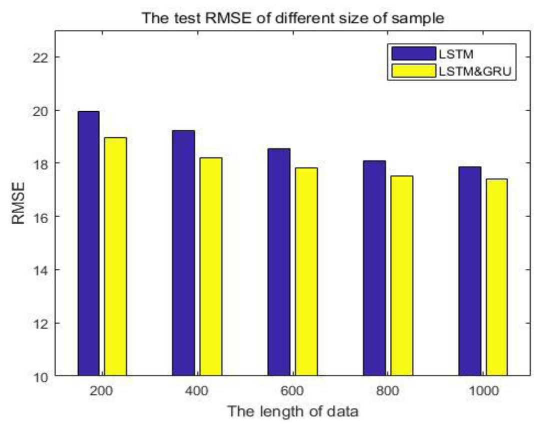

FIGURE 14. The test RMSE of different size of samples.

bigger, especially in the sum of train RMSE. There is a small difference in each city between LSTM&GRU architecture and LSTM architecture and a significant differences between the two model, therefore we choose the sum of RMSE instead single city RMSE. RMSE will have a large value due to the large change in PM2.5. The reason why the difference of the test is smaller than the train probably is that the training is locally optimal, leading to more stable on test results. Fig. 11 and Fig. 12 proof that LSTM&GRU architecture is widely better than LSTM architecture, with the improve of the model we can get a more accuracy result.

4) THE IMPACT OF SAMPLE SIZE ON THE MODEL

By enlarging the size of the sample, we find that an increase in the sample size will reduce the RMSE of both training and testing. Fig. 13 and Fig. 14 show the change in the RMSE value of the model after the sample size change. Fig. 13 and Fig. 14 show that with the increase of length of data, the RMSE will decrease rapidly. And no matter how the length of data, the accuracy of ours is better than LSTM. It also shows that the size of sample can significantly affect the accuracy of the model. However, enlarging the size of sample is not a wise action because it will increase the training time of the model so that we cannot get the result timely.

V. CONCLUSION

This paper analyzes and predicts the smog model of Internet of Things monitoring. This model gives full play to the IoT monitoring information. It realizes smog pollution prediction and early warning and decision analysis. Through the Internet of Things monitoring and data prediction, the technical architecture of smog dynamic monitoring integrating information collection, transmission, processing and prediction is constructed. Then we use LSTM&GRU double-recursive neural networks to use historical data to predict future air pollution data. The results of this experiment show that our model can play a better prediction.

The advantages of IoT technology in environmental monitoring are fully utilized. This model can achieve a comprehensive perception of environmental pollution. It can provide new technical means for the prevention and control of sudden smog pollution and improve the intelligence level of early warning and emergency prevention. In the experiment, we found that not all the prediction results of our model were better than those of LSTM. This may be that the data characteristics were not obvious, or the data fluctuated so much that there is no suitable model could be trained.

In the future work, the advantages of IoT integrated sensing technology, communication technology and information technology will be used to solve the uncertainty problems in the smog monitoring work. More accurate haze prediction data is obtained by gradually reducing the error. We will extract more functional values for our models and improved models for training. At the same time, we will also compare the advantages and disadvantages of different neural networks. The smog data prediction model looks for a better model to improve prediction accuracy. As a typical information technology, the IoT technology fully integrates environmental protection work and can better promote the efficiency of environmental monitoring.

REFERENCES

[1] H. Schaffers, N. Komninos, M. Pallot, B. Trousse, M. Nilsson, and A. Oliveira, ''Smart cities and the future Internet: Towards cooperation frameworks for open innovation,'' in *The Future Internet—FIA* (Lecture Notes in Computer Science), vol. 6656, J. Domingue *et al.*, Eds. Berlin, Germany: Springer, 2011, pp. 431–446.

[2] M. Jia, A. Komeily, Y. Wang, and R. S. Srinivasan, ''Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications,'' *Automat. Construct.*, vol. 101, pp. 111–126, May 2019.

[3] R. Montella, D. Di Luccio, L. Marcellino, A. Galletti, S. Kosta, G. Giunta, and I. Foster, ''Workflow-based automatic processing for Internet of floating things crowdsourced data,'' *Future Gener. Comput. Syst.*, vol. 94, pp. 103–119, May 2019.

[4] T. Lin, B. G. Horne, P. Tino, and C. L. Giles, ''Learning long-term dependencies in NARX recurrent neural networks,'' *IEEE Trans. Neural Netw.*, vol. 7, no. 6, pp. 1329–1338, Nov. 1996.

[5] F. A. Gers, J. Schmidhuber, and F. Cummins, ''Learning to forget: Continual prediction with LSTM,'' in *Proc. Int. Conf. Artif. Neural Netw.*, Oct. 2000, pp. 850–855.

[6] S. Hochreiter and J. Schmidhuber, ''Long short-term memory,'' *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] B. Wang, X. Gu, and S. Yan, ''STCS: A practical solar radiation based temperature correction scheme in meteorological WSN,'' *Int. J. Sensor Netw.*, vol. 28, no. 1, pp. 22–33, 2018. doi: 10.1504/IJSNET.2018.094709.

[8] K. Cho, B. Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, ''Learning phrase representations using RNN Encoder–Decoder for statistical machine transalation,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2014, pp. 1724–1734.

[9] F. Lu, D. Xu, Y. Cheng, S. Dong, C. Guo, X. Jiang, and X. Zheng, ''Systematic review and meta-analysis of the adverse health effects of ambient $PM_{2.5}$ and PM_{10} pollution in the Chinese population,'' *Environ. Res.*, vol. 136, pp. 196–204, Jan. 2015.

[10] W. Yang, L. Lin, and L. Tian, ''The haze weather analysis and prediction based on the wavelet analysis,'' *J. Shaanxi Univ. Sci. Technol. (Natural Sci. Ed.)*, vol. 34, no. 3, pp. 166–170, 2016.

[11] Q. Fu, ''Research on haze prediction based on multivariate linear regression,'' *Comput. Sci.*, vol. 43, no. S1, pp. 526–528, 2016.

[12] T. Qiu, R. Qiao, and D. Wu, ''EABS: An event-aware backpressure scheduling scheme for emergency Internet of Things,'' *IEEE Trans. Mobile Comput.*, vol. 17, no. 1, pp. 72–84, Jan. 2018.

[13] J. B. Ordieresa, E. P. Vergara, and R. S. Capuz, ''Neural network prediction model for fine particulate matter ($PM_{2.5}$) on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua),'' *Environ. Model. Softw.*, vol. 20, no. 5, pp. 547–559, 2005.

[14] S. V. Barai, A. K. Dikshit, and S. Sharma, ''Neural network models for air quality prediction: A comparative study,'' in *Soft Computing in Industrial Applications* (Advances in Soft Computing), vol. 39, A. Saad, K. Dahal, M. Sarfraz, and R. Roy, Eds. Berlin, Germany: Springer, 2007, pp. 290–305.

[15] G. Corani, ''Air quality prediction in Milan: Feed-forward neural networks, pruned neural networks and lazy learning,'' *Ecolog. Model.*, vol. 185, pp. 513–529, Jul. 2005.

[16] B. Wang, W. Kong, W. Li, and N. N. Xiong, ''A dual-chaining watermark scheme for data integrity protection in Internet of Things,'' *Comput., Mater. Continua*, vol. 58, no. 3, pp. 679–695, 2019.

[17] C. Shi, ''A novel ensemble learning algorithm based on D-S evidence theory for IoT security,'' *Comput., Mater. Continua*, vol. 57, no. 3, pp. 635–652, 2018.

[18] J. Wang, Y. Gao, W. Liu, W. Wu, and S.-J. Lim, ''An asynchronous clustering and mobile data gathering schema based on timer mechanism in wireless sensor networks,'' *Comput., Mater. Continua*, vol. 58, no. 3, pp. 711–725, 2019.

[19] A. Russoa, F. Raischel, and P. G. Lind, ''Air quality prediction using optimal neural networks with stochastic variables,'' *Atmos. Environ.*, vol. 79, pp. 822–830, Nov. 2013.

[20] X. Hu, L. A. Waller, A. Lyapustin, Y. Wang, M. Z. Al-Hamdan, W. L. Crosson, M. G. Estes, Jr., S. M. Estes, D. A. Quattrochi, S. J. Puttaswamy, and Y. Liu, ''Estimating ground-level $PM_{2.5}$ concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model,'' *Remote Sens. Environ.*, vol. 140, pp. 220–232, Jan. 2014.

[21] L. Yan, M. Zhou, Y. Wu, and L. Yan, ''Long short term memory model for analysis and forecast of PM2.5,'' in *Cloud Computing and Security—ICCCS* (Lecture Notes in Computer Science), vol. 11063, X. Sun, Z. Pan, and E. Bertino, Eds. Cham, Switzerland: Springer, 2018, pp. 623–634.

[22] F. Karim, S. Majumdar, H. Darabi, and S. Chen, ''LSTM fully convolutional networks for time series classification,'' *IEEE Access*, vol. 6, pp. 1662–1669, 2017. doi: 10.1109/ACCESS.2017.2779939.

[23] T. Qiu, X. Wang, C. Chen, M. Atiquzzaman, and L. Liu, ''TMED: A spider-Web-like transmission mechanism for emergency data in vehicular ad hoc networks,'' *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8682–8694, Sep. 2018.

[24] Y. T. Tsai, Y. R. Zeng, and Y. S. Chang, ''Air pollution forecasting using RNN with LSTM,'' in *Proc. IEEE 16th Int. Conf. Dependable, Autonomic Secure Comput., 16th Int. Conf. Pervasive Intell. Comput., 4th Int. Conf. Big Data Intell. Comput. Cyber Sci. Technol. Congr.*, Aug. 2018, pp. 1074–1079.

[25] J. Zhao, F. Deng, Y. Cai, and J. Chen, ''Long short-term memory—Fully connected (LSTM-FC) neural network for $PM_{2.5}$ concentration prediction,'' *Chemosphere*, vol. 220, pp. 486–492, Apr. 2019.

[26] G. Somani, X. Zhao, S. N. Srirama, and R. Buyya, ''Integration of cloud, Internet of Things, and big data analytics,'' *Softw., Pract. Exper.*, vol. 49, no. 4, pp. 561–564, 2019.

[27] X. Deng, P. Jiang, X. Peng, and C. Mi, ''An intelligent outlier detection method with one class support tucker machine and genetic algorithm toward big sensor data in Internet of Things,'' *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4672–4683, Jun. 2019.

[28] D. Minoli, K. Sohraby, and B. Occhiogrosso, ''IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems,'' *IEEE Internet Things J.*, vol. 4, no. 1, pp. 269–283, Feb. 2017.

[29] T. Qiu, H. Wang, K. Li, H. Ning, A. K. Sangaiah, and B. Chen, ''SIGMM: A novel machine learning algorithm for spammer identification in industrial mobile cloud computing,'' *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2349–2359, Apr. 2018. doi: 10.1109/TII.2018.2799907.

BAOWEI WANG was born in Liaocheng, Shandong, China, in 1982. He received the B.S. degree in computer science from Hunan University, Changsha, Hunan, China, in 2005, and the Ph.D. degree in computer science from Hunan University, China, in 2011. He is currently an Associate Professor with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China. He has published over 50 international journal and conference papers. Some of his works were published in the JIT, IJSNet, INFOCOM, ICPADS, and ICCCS. His research interests include the Internet of Things and sensor networks, data security, and mobile computing. He is a member of the ACM and CCF. He has been the TPC Chair, the Workshop Chair, and a TPC member of many international conferences, including the TURC-AIS 2019, ACM TURC 2018 and 2019, ICCCS 2018, and ICAIS 2019, and as a Reviewer of about 10 international journals, including IEEE ACCESS and the *Computer Communication*.

WEIWEN KONG was born in Xuzhou, Jiangsu, China, in 1994. She received the B.S. degree in software engineering from the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China, in 2018, where she is currently pursuing the M.S. degree. Her research interests include wireless sensor networks, machine learning, and data security. She is a member of the ACM.

HUI GUAN was born in Shenyang, Liaoning, China, in 1973. She received the B.S. degree in applied meteorology from the Nanjing Meteorological College, and the M.S. degree in education from Nanjing University. She is currently pursuing the Ph.D. degree with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China. Her research interests include the field of smog and health, machine learning, and data security.

NEAL N. XIONG received the Ph.D. degree from Wuhan University (about sensor system engineering), and the Japan Advanced Institute of Science and Technology (about dependable sensor networks). He is currently an Associate Professor (3rd year) with the Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK, USA. Before he attended Northeastern State University, he was with Georgia State University, Wentworth Technology Institution, and Colorado Technical University (Full Professor for about 5 years) about 10 years. His research interests include cloud computing, security and dependability, parallel and distributed computing, networks, and optimization theory. He has published over 200 international journal papers and over 100 international conference papers. Some of his works were published in the IEEE JSAC, the IEEE or ACM transactions, ACM Sigcomm workshop, the IEEE INFOCOM, ICDCS, and IPDPS. He has received the Best Paper Award in the 10th IEEE International Conference on High-Performance Computing and Communications (HPCC-08) and the Best Student Paper Award. He has been the General Chair, the Program Chair, the Publicity Chair, a PC member, and a OC member of over 100 international conferences, and as a Reviewer for about 100 international journals, including the IEEE JSAC, the IEEE SMC (Park: A/B/C), the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON MOBILE COMPUTING, and the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS. He is serving as an Editor-in-Chief, Associate Editor or Editor member for over 10 international journals, including an Associate Editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS: SYSTEMS, an Associate Editor for *Information Science*, the Editor-in-Chief for the *Journal of Internet Technology* (JIT), and the Editor-in-Chief for the *Journal of Parallel and Cloud Computing* (PCC)), and a Guest Editor for over 10 international journals, including the *Sensor Journal*, WINET, and MONET.