Data Preprocessing for Agricultural IoT Based on RBF Neural Network

Qihui Huang, Yajie Ma and Jin Zhang

College of Information Science and Engineering, Wuhan University of Science and Technology, Engineering Research Centre for Metallurgical Automation and Detecting Technology, Ministry of Education, Wuhan 430081, China.
Email: hqhwust@163.com

Abstract. The collected data obtained in the agricultural environment is not simply linear and stable, complex nonlinear functional relationships can be found. While RBF neural networks have the ability to approximate arbitrary nonlinear mappings and can implement the functions of nonlinear predictors with superior performance for data forecasting. Considering the strong time correlation of agricultural data, this paper proposes a time series model based on RBF neural network for preprocessing raw data. A four layer of agricultural IoT system is designed, a data processing layer is inserted into the traditional three-layer IoT system. Then the abnormal value is identified and eliminated by the "t testing criteria" test on the data preprocessing. The experimental results indicate that, comparing with auto regressive moving average model, the proposed model can achieve more competitive prediction ability.

1. Introduction

Agricultural Internet of Things (IoT) will generate a large amount of data every moment. As the agricultural environment is very complex, the collected data includes various anomalies cannot be directly analysed and mined [1]. How to deal with these data and use them are crucial issues in the step of data preprocessing. In order to improve the data quality, we consider to set up a prediction model is by using the historical data to eliminate the effect of environmental noise. And the model can be used to forecast subsequent data and handle missing values and outliers.

In our former work [1] it shows a time series model to preprocess agricultural data. The method performs well in greenhouses. However, the agricultural environment is complex, and the collected data has complex nonlinear functional relationship [2]. The traditional time series models (AR, MA, ARMA, etc.) cannot respond well to this complex relationship. While the RBF neural network has the ability to approximate arbitrary nonlinear mapping by learning, and has local approximation characteristics, which can realize nonlinear predictors and nonlinear classifiers with superior performance [2-3]. RBF neural networks have attracted great attention and become one of the most favourable methods for data modelling and prediction in the last decades [6]. G. Peter Zhang and Douglas M. Kline [7] suggest a neural network (NN) method to predict quarterly sales. Qing Ye [8] shows a prediction model based on RBF to predict the engineering valuation of Xiamen City. However, these methods do not take into account the temporal correlation of agricultural IoT data.

To address the above problems, a time series model based on RBF neural network for preprocessing raw data is proposed in this paper. Aiming at the data characteristics of the agricultural IoT, a four-layer framework is proposed. More attentions are paid on the analysis and design of the data processing layer. In view of the problems existing in the raw data, the abnormal data is identified and eliminated by the "t testing criteria" test on the data preprocessing, which improve the quality of...
data preprocessing. Then the time series model of RBF neural network is set up in the data processing layer. Compared with the ARMA model, the method achieves competitive prediction accuracy.

The remainder of this paper is organized as follows. Section II introduces the four-layer structure of agricultural IoT. Section III introduces the RBF neural network structure and modelling as well as data preprocessing. Section IV is the experiment, including neural network simulation, ARMA model and RBF neural network prediction method performance comparison. The final section is the conclusion.

2. Architecture of Agricultural IoT

Agricultural IoT is generally divided into the sensor layer, the network layer and the application layer [9]. In order to meet the data quality requirements of the agricultural IoT, a four-layer structure for agricultural IoT is designed. The data preprocessing layer is inserted between the sensor layer and the network layer. The hierarchy of agricultural IoT is shown in Figure 1.

![Figure 1. The architecture of agricultural IoT](image)

2.1. Sensor Layer

The sensing layer is mainly responsible for data collection. A large number of wireless sensor terminal nodes and sink nodes are distributed in the data acquisition subsystem. Considering the large-scale of agricultural IoT, LoRa wireless technology is applied for communication between sensor nodes. LoRa communication technology has the advantages of long distance and low power consumption. It will reduce the operating cost in agricultural IoT applications and meet the low cost and large area transmission requirements of agricultural IoT systems.

2.2. Network Layer

The network layer is a communication bridge for data transmission between the sensing layer and the application layer. The network layer sends the data collected by the sensor layer to the remote server through the GPRS network and the Internet, and also transmits the information fed back by the user in the application layer to the sensing layer.

2.3. Data Processing Layer

A time series model based on RBF neural network is put forward in data processing layer. The method will achieve competitive prediction compared with Auto Regressive ARMA mode in Agricultural IoT.
The processed data will be applied for further data mining, what’s more obtaining useful information for the user finally.

2.4. Application Layer

The application layer is the main function of data storage and applications. The user can log in to the web remote monitoring platform or Android APP to query the data according to different requirements to manage the crops in real time. It will not only achieve disaster prevention, but also take preventive measures in advance, which greatly improve the yield and quality of agricultural products [9].

3. Modeling

After the design and implementation of the Agricultural IoT, 640 atmospheric temperature data were collected as a raw data set $X$ this paper presents the time series analysis method of RBF neural network to preprocess these data. While the collected atmospheric temperature data has a strong time correlation, only the trend reflected by the historical data will be model and predict the Subsequence.

Assuming the 640 temperature data as: The values of the last $M$ moments are predicted by the values of the $N$ moments before the sequence. The data is divided into $k$ data segments of length $M+N$. Each data segment can be regarded as a sample. The first $N$ values of each sample are regard as input to the RBF neural network, and the last $M$ values as target outputs. The method Mappings by learning to achieve the purpose of time series prediction [10].

The main processing of RBF neural network is divided into the following steps: (1) data preprocessing; (2) RBF neural network training; (3) RBF neural network simulation.

3.1. Data Preprocessing

The data preprocessing in this paper includes singular value culling, data normalization processing, as well as data division.

3.1.1. Abnormal value culling. Pre-treatment of the raw data is required before neural network training. There are some singular values in the agricultural IoT data. These singular values may cause bad effect in subsequent modelling. Hence it’s significant to eliminate these singular value data. The procedure as follow: (1) the logic test is used to roughly identify the singular value, while the observation value has a reasonable range. When it exceeds this range, it is judged as a singular one. (2) Based on the logic test, the “$t$ test criteria” are used to further identify the singular values. The Algorithm is as follows.

Step1: Assuming $x_{i}, \{x_{i} \mid x_{i} \in X, i = 0,1,2,...,639\}$, is abnormal, then calculate the average $\bar{x}_{n-1}$ and Standard variance $S_{n-1}$ of the remaining data;

Step2: if $|x_{i} - \bar{x}_{n-1}| > k(n,a)$

$x_{i}$ is abnormal value ;

end

($k(n, a)$ is identified in the “$3\sigma$criterion”)

3.1.2. Normalized processing. In order to prevent super saturation of neurons, these data should be normalized between [0, 1]. The normalization formula is shown in formula (1):

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X}$$

(1)

3.1.3. Data division. $R$ can be divided into 80 data segments with length 7+1. Here, each of the seven data is regarded as an input for RBF neural network to predict the latter data. Therefore, the number of input neurons is seven. The output neurons are one, and the output value is the latter data. Hence, the training dataset is 560 and the test dataset is 80.
3.2. RBF Neural Network training

3.2.1. The RBF neural network. The RBF network is a three-layer feed-forward network; the structure is shown in Figure 2, which includes an input layer, a hidden layer and an output layer [11]. The input vector of RBF network is the time series modelled. Hidden layer composed of radial basis functions. The output layer is a linear combination of the output of the hidden layer neurons. Let the input be a d-dimensional vector and the output be a real value, then the RBF network can be defined as equation 2.

\[ \varphi(x) = \sum_{i=1}^{q} w_i \exp(-\beta \| x_i - c \|^2) \]

\[ \text{Figure 2. The structure of RBF neural network} \]

where \( q \) represents the node number in the hidden layer, \( x_i \) represents \( d \)-dimensional input vector, \( c_i \) is the centre vector, \( w_i \) is the weight.

3.2.2 Parameter setting. The formula of RBF network contains three unknown parameters, the centre of the neuron \( c_i \), the weight \( w_i \) and the variance \( \beta_i \). The key point of network construction is the selection of the centre of the neuron. Considering that the self-organizing learning algorithm is simple in calculation, small amount of calculation and high in efficiency, the self-organizing learning is applied to select the centre of neurons in this paper. This method makes the selection of the centre of the neural network more precise by adjusting the cluster centre. Self-organizing learning method will be divided into neuron centre selection, calculate variance and calculate mean.

1) The selection of the centre of the neuron. The centre of the neuron is obtained by the K-means clustering algorithm. The Euclidean distance formula is defined in formula (3).

\[ \text{dist}(X, Y) = \left( \sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}} \]

2) Calculate the variance: The Gaussian radial basis function variance is shown in equation (4). Where is the maximum distance of the selected centre.

\[ \beta_i = \frac{c_{\text{max}}}{(2K)^{\frac{1}{2}}} \quad i = 1, 2, 3...h \]

3) Calculate the weight: The connection weight of the hidden layer to the output layer can be directly calculated by the least squares method, as shown in equation (5):

\[ w = \exp\left( K c_{\text{max}} \| x_p - c_i \|^2 \right) \quad p = 1, 2,...P; i = 1, 2,...K \]
4. Experiment

4.1. RBF Neural Network Simulation

After modelling the time series model of RBF neural network, we use the Matlab toolbox to create the neural network. The net defined as: 
\[ \text{net} = \text{newrb} (P, T, \text{GOAL}, \text{SPREAD}, MN, DF) \]
where \(P\) and \(T\) are the input vector and the target vector of the training sample. \(\text{GOAL}\) is the root mean square error of the network, the default is 0. \(\text{SPREAD}\) is the expansion speed of RBF, the default is 1. \(MN\) is the maximum number of neurons; the default is the number of samples. And \(DF\) is the number of neurons added between the two displays, the default is 25. The built neural network is simulated by the simulation function \(\text{sim} (\text{net}, p)\). The fit of the RBF neural network to the raw data is shown in Figure 3. It suggests from the figure that the RBF fits well with the raw data.

4.2. Analysis of Results

In our former work [1], we have implemented data preprocess of agricultural IoT based on ARMA model, in addition confirmed the ARMA (2, 10) model. In order to compare the performance of ARMA (2, 10) model and RBF neural network model in data preprocessing of agricultural IoT. In this paper, a comparative experiment was conducted to train the same 560 atmospheric temperature data with ARMA (2, 10) to predict the last 80 data. The fit of the atmospheric temperature and the raw data between the RBF neural network and the ARMA model is shown in Fig. 3. The results demonstrate that RBF model fits better compared with ARMA model. Further comparison from the error distribution, As shown in Figure 4, it suggests from the histogram that 57% of the ARMA model has a relative error of less than 5%, 43% between 5% and 10%. While a relative error in the error distribution of the RBF model, which is less than 5% accounted for 85%, and only 15% of the data ranged from 5% to 10%. The performance of the RBF model is better than the ARMA (2, 10) model in error distribution. In order to further compare the prediction performance of the two models, this paper applies the maximum absolute error (MAE), the average relative error (ARE), and the root mean square error (RMSE as Equation 6) as the predictive model performance index. As shown in Table 1. It demonstrates from the table that the indicators predicted by the RBF model are competitive to the ARMA (2, 10) model, indicating that the RBF neural network prediction is superior to the ARMA model in agricultural IoT data processing.

\[
\text{RMSE} = \left( \frac{1}{N} \sum_{i=1}^{N} (X_{\text{obs},i} - X_{\text{predict},i})^2 \right)^{\frac{1}{2}}
\]  

(6)

| Table 1. ARMA and RBF performance comparison |
|---------------------------------------------|
|                | MAE (%) | ARE (%) | RMSE (%) |
| ARMA           | 2.34    | 4.0458  | 1.3763   |
| RBF            | 2.0254  | 1.8064  | 0.0301   |
5. Conclusion
In this paper, a time series model based on RBF neural network is designed to model and predict the agricultural data. This method is compared with the AMRA model. The results show that the fitting accuracy of the time series method using RBF neural network achieves more competitive than ARMA model, and the performance is superior to ARMA in terms of Maximum Absolute Error (MAE), average relative error (are) as well as root Mean Square Error (RMSE).

6. Acknowledgement
This work is supported by the National Natural Science Foundation General Program of Hubei Province (NO.2016CFB463). Key Projects of Science Research Program of Hubei Provincial Ministry of Education (NO.D20151106).

7. References
[1] Ma Y, Jin J, Huang Q, et al 2018 Data Preprocessing of Agricultural IoT Based on Time Series Analysis[C] International Conference on Intelligent Computing. Springer, Cham
[2] Yi J, Yong-Peng L 2007 The Application of RBF Neural Networks to Stock Price Forecasting [J]. Mind & Computation
[3] Li L, Wen Y E 2010 Precipitation prediction of time series model based on BP artificial neural network [J]. Journal of Water Resources & Water Engineering, 21(5) 156-159
[4] Jia L, Fan L U, Yun-Zhong J, et al 2011 Application of RBF Neural Network in Runoff Time Series Prediction[J]. Yellow River
[5] Zhang G, Patuwo B E, Hu M Y 1998 Forecasting With Artificial Neural Networks: The State of the Art [J]. International Journal of Forecasting, 14(1) 35-62
[6] Zhang G P, Kline D M.2007 Quarterly time-series forecasting with neural networks [J]. IEEE Transactions on Neural Networks, 18(6)1800-14
[7] Qing Y E 2012 Prediction Model of the Project Cost Estimation Based on RBF Neural Network [J]. Journal of Huaqiao University
[8] Zhuang L Q, Zhang J B, Zhang J B, et al 2005 Data management for wireless sensor networks: research issues and challenges[C] International Conference on Control & Automation. IEEE
[9] Li S, Cui J, Li Z 2011 Wireless Sensor Network for Precise Agriculture Monitoring[C] Fourth International Conference on Intelligent Computation Technology & Automation. IEEE Computer Society
[10] Yurui Z 2005 Time Series Prediction Based on RBF Neural Network [J]. Computer Engineering & Applications, 11(74) 76
[11] Mitchell T, Buchanan B, Dejong G, et al 2003 Machine Learning [M]. China Machine Press