Prediction of rice yield based on LSTM long short term memory network

Haoming Mo a, Ying Zhang b,*, Yifei Liu c, Yanzi Zheng d
Guangdong Ocean University, Zhanjiang, Guangdong, China

a 1194807684@qq.com, b* Corresponding author e-mail: xzb@gdou.edu.cn, c 13543319904@163.com, d zyxdyyxiang@163.com

Abstract. The prediction of rice yield is of great significance for improving rice yield and preventing disasters. At present, there is no exact and effective solution to the problem of rice yield prediction, and there are many factors affecting the rice yield, which may lead to large errors in the prediction results. Based on the rice yield data of 81 counties in Guangxi Province of China for three years, a long short time memory network (LSTM) model was established and used to predict rice yield under the influence of meteorological factors. The results show that the performance of the recurrent neural network with LSTM structure is better than that of the standard recurrent neural network, which solves the problems of gradient dispersion and gradient explosion of RNN model, and enables the model to learn long-term laws. At the same time, the traditional machine learning model is established to compare with LSTM model. The neural network model can better mine the information hidden in the data and make full use of the rules in the data. Combined with meteorological factors, the test shows that LSTM model has higher prediction accuracy. It provides a reference for rice yield prediction.

Keywords: long short term memory network, machine learning, rice, meteorology, production forecast

1. Introduction

Rice is one of the most important food crops in the world, and also the most important food crop in southern China. Rice is the staple food of nearly half of the people in China. Moreover, the rice planting area price accounts for about 27% of the national grain planting area, and the rice yield is more than one third of the total national grain output. Guangxi is a big agricultural province in China, but due to the complex geographical environment, the workload of agricultural statistical investigation is heavy. With the development of science and technology, Guangxi has realized the comprehensive monitoring of meteorological data from ground, satellite and radar [1]. For the agricultural industry, by detecting the relationship between weather and crop growth, it provides services for agriculture, rural areas and farmers, and provides favorable conditions for improving rice yield.

At present, Chinese scholars have done a lot of research on rice yield prediction. Some of them discussed the relationship between rice production and meteorological factors, some analyzed the impact of climate change on agriculture, some studied the prediction of rice yield by establishing neural network...
rice yield prediction model, and some simulated rice yield function based on mathematical method “exponential smoothing method”. There are many factors that affect rice yield. It is not comprehensive to find the rules only through time series. Otherwise, the characteristics beyond time cannot be fully utilized, and the deeper meaning can not be mined out, which will lead to the lack of generalization ability of the prediction model and even some serious prediction errors under the interference of Meteorological factors. Therefore, based on the former research, this paper considers the time series information and combines with meteorological factors, which improves the generalization ability of the model and makes the prediction effect of the prediction model more accurate in the future.

This paper uses the Long Short Term Memory model (LSTM), which is the most famous and successful extension of the recurrent neural network. Due to the problems of gradient vanishing and gradient explosion, the learning ability of recurrent neural network is limited, and the effect in practical tasks is often not as expected [2]. LSTM can memorize valuable information for a long time [3], so as to reduce the learning difficulty of recurrent neural network, and solve the long-term dependence problem in sequence model. Therefore, LSTM is widely used in speech recognition, language modeling, machine translation text detection and recognition [4][5][6].

2. The Introduction of Data Sources and Data

2.1. Data Sources
The data used in this paper are rice yield and daily meteorological factors of 81 counties in three years in Guangxi Province. The daily sunshine duration, daily relative humidity, daily precipitation and other meteorological factors and specific dates are recorded in detail in the data, which is conducive to the application of LSTM model in this paper. Download from: https://tianchi.aliyun.com/competition/entrance/231753/information.

2.2. Evaluating Indicator
In this paper, mean square error (MSE) is used as the evaluation index to evaluate and compare the fitting effect of the model.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$  \hspace{1cm} (1)

In the formula: $y_i$ is the predicted value; $\hat{y}_i$ is the actual value; the smaller the value is, the better the prediction effect of the model on the target yield of rice; otherwise, the prediction effect is worse.

3. Neural Network

3.1. Network
Recurrent neural network (RNN) originated from Hopfield network proposed by saratha sathasivam in 1982. The cyclic neural network can effectively use the time series information in the data, and has certain memory ability [7]. Its structure is shown in Figure 1[8].

![Fig.1 Structure diagram of recurrent neural network](image_url)

The main purpose of the recurrent neural network is to process and predict the sequence data. Compared with the fully connected network, it connects the nodes of each layer separately, so that the previous information can be used. In Fig. 1, the sequence of length t can be modeled as a feedforward network.
neural network of T layer after expansion. It can be calculated by the formula of full connection layer, but another parameter $h_t$ is introduced, is the hidden variable of the time step, $f, g$ is the activation function.

$$h_t = f\left(Ux_t + Ws_{t-1}\right),$$

$$s_t = f\left(h_t\right), o_t = g\left(Vs_t\right)$$

(2)

Although the recurrent neural network has a certain memory ability, it cannot capture the long-distance dependence when using the BPTT back-propagation algorithm, which involves the gradient vanishing problem in the neural network. The following is explained by derivation:

At the beginning, $W$ is randomly initialized, $W$ is the weight of input, the loss value is calculated by forward propagation, $E$ is the total loss value, and then $W$ is corrected by back propagation.

$$\frac{\partial E_t}{\partial W} = \sum_{k=0}^{t} \frac{\partial E_t}{\partial O_t} \frac{\partial O_t}{\partial S_t} \left( \prod_{j=k+1}^{t} \frac{\partial S_j}{\partial S_{j-1}} \right) \frac{\partial S_k}{\partial W}$$

(3)

Because the activation function used in this paper is a function, and the maximum value of the derivative of the function is 1, when the length of the sequence is too long, the gradient will disappear due to the reason of multiplication. Because of the large amount of data and long time series, it is easy to appear the phenomenon of gradient disappearance [10]. Therefore, this paper adopts LSTM model, which is an improved RNN neural network.

3.2. Long Short Time Neural Network

Long- and short-time neural network is a special network structure with "gate" structure added on the basis of cyclic neural network. There are three kinds of gate structures: forgetting gate, input gate and output gate. "Forgetting gate" controls whether the information of the state before and the information of the current state is forgotten, which is conducive to discard the useless information and keep important information. The input gate is to update the transition state of this stage by combining input with the previous stage. The output gate updates the output $h_t$ through $x_t$ input, hidden state $h_{t-1}$ and the latest state. Its structure is shown in Figure 2.

![Fig.2 Structure of long short time neural network](image)

Where $C_t$ is the state at time $t$, $h_t$ is the hidden state at time $t$, and $x$ is the input information. With the help of "gate" structure, the neural network is easier to learn the long-term dependence between sequences.
3.3. Using Neural Network Model to Predict

3.3.1. The data normalization. Firstly, the data is cleaned and the character features are encoded by one hot. Because there are a small number of missing values, the method used in this paper is to fill in with zeros to retain the data information. The maximum and minimum values of the data are normalized, and the calculation method is shown in formula (4). All the data are transformed linearly to map the result to the range of [0,1], so that the original data can be scaled equivalently to eliminate the influence of dimension. The normalization formula is as follows:

$$x_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

(4)

3.3.2. The building of models. The activation function uses the default activation function of the model, which is generally sigmoid function, because the sigmoid function is continuous and differentiable everywhere, which is conducive to the calculation of back propagation; Generally, the setting range of learning rate is between 0.01 and 0.2. In this paper, the setting of learning rate is 0.02. The setting of learning rate is related to the speed of model training or whether the final convergence can be achieved. If the setting is too small, it will lead to that the model training speed is too slow, and too large may not converge. The general learning rate is selected by final adjustment[11].

3.3.3. Forecast. The data used includes the early and late rice yield and meteorological factors of 81 counties in Guangxi from 2015 to 2017. In this paper, the data from 2015 to 2016 are used for the training model, and then the trained model is used to predict the rice late rice in 2017. The predicted data are compared with the actual rice yield, and the calculation error is calculated. The MSE on the training set is reduced to 0.1304, while the MSE on the test set is 0.00645.

4. Machine Learning

In order to further verify the effect of LSTM model, the support vector regression model, which has better performance in rice yield prediction, was used to compare the prediction effect with LSTM model.

4.1. Support vector regression

4.1.1. Algorithm principle. Machine learning refers to computer learning, self-renewal and progress through observing environment and interacting with environment. Because the data has more than ten features and high dimension, this paper uses the classical algorithm Support Vector Regression (SVR) algorithm in machine learning. Support Vector Machine (SVM) model is used to fit the function, and the parameters are determined by solving the optimal solution of convex quadratic programming. In this paper, when using SVR to solve the regression problem, we need to determine the absolute value of error e, the loss value can be calculated. It is equivalent to an acceptable range. When the predicted value is within this range, such error can be ignored here. If the actual function is $F(x)$ and the fitting function is $f(x)$, then the error is E. This paper takes linear regression as an example, and figure 3 is the schematic diagram of support vector regression. The red color shows the e -septum. When the predicted value is between the spacer bands, the loss is not calculated. For the selection of e-septum, the default parameters of the model are used in this paper.
The SVR problem can then be formalized as \( \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} l_{ei} (f(x_i) - y_i) \), where \( C \) is the regularization constant and the function \( l_{ei} \) becomes the \( e \)-insensitive loss function, in form of:

\[
l_{ei} = \begin{cases} 0, & |e| \leq e \\ |e| - e, & \text{other} \end{cases}
\]  \hspace{1cm} (5)

4.1.2. Dual problem. By introducing the relaxation variables \( \xi_i \), \( \hat{\xi}_i \) and Lagrange multiplier, the Lagrange function of SVR problem can be obtained by Lagrange multiplier method.

\[
L(w, b, \alpha, \hat{\alpha}, \xi_i, \hat{\xi}_i, \mu, \hat{\mu}) = \frac{1}{2} \|w\|^2 + 
C \sum_{i=1}^{m} (\xi_i + \hat{\xi}_i) - \sum_{i=1}^{m} \mu_i \xi_i - \sum_{i=1}^{m} \hat{\mu}_i \hat{\xi}_i + 
\sum_{i=1}^{m} \alpha_i (f(x_i) - y_i - e - \hat{\xi}_i) + 
\sum_{i=1}^{m} \hat{\alpha}_i (y_i - f(x_i) - e - \hat{\xi}_i)
\]  \hspace{1cm} (6)

Then, the dual problems of \( w = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) x_i \) and SVR are obtained by calculating the partial derivatives of the parameters \( w, b, \xi_i, \hat{\xi}_i \) and making the partial derivatives zero.

\[
\max_{\alpha_i, \hat{\alpha}_i} \sum_{i=1}^{m} y_i \left( \alpha_i - \alpha_i \right) - e \left( \hat{\alpha}_i + \alpha_i \right) - 
\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} (\alpha_i - \alpha_i) (\hat{\alpha}_j - \alpha_j) x_i^T x_j
\]  \hspace{1cm} (7)

\[
S.t. \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) = 0, 0 \leq \alpha_i, \hat{\alpha}_i \leq C
\]  \hspace{1cm} (8)

From \( w = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) x_i \) we can get the formal solution of SVR: \( f(x) = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) x_i^T x + b \).

4.1.3. Prediction of rice yield using SVR model. Similarly, this paper preprocessed the data, including the same steps of pre-processors as coding useful character data, filling missing values, standardizing and normalizing the data. Parameters are selected by grid search. Moreover, the triple cross validation method is used to reduce the risk of overfitting, so as to improve the generalization ability of the model.
The model is built with the help of python third-party library sklearn[13]. In this paper, the data of 2015 and 2016 is used to train the model, and then the trained model is used to predict the yield of late rice in 2017. In the training process, the average of three cross validation errors is taken as the final training error, and the MSE of the test set is 1.395792.

5. Result Analysis

The model construction process proposed in this paper was applied to preprocess the rice yield data. The data were divided into training set and test set, and the model parameters were preliminarily determined to construct LSTM and SVR models. After the model was trained by the training set, the test set was used for prediction, and the prediction effect of the model was compared.

| Table 1. Comparison of Model Prediction Results |
|-----------------------------------------------|
| Model          | Model parameter | MSE         |
|----------------|-----------------|-------------|
| LSTM           | Batch=1         | Epoch=1000  | 0.006450    |
| SVR            | C=1.0           | Ε=0.3       | 1.395792    |

It can be seen from table 1 that the mean square error of LSTM model fitting test set is 0.006450, while that of SVR model is 1.395792. It can be concluded that the prediction effect of LSTM model on test set is obviously better than that of SVR model.

The model construction process proposed in this paper was applied to preprocess rice yield data. The rice yield data were divided into training set and test set, and the model parameters were preliminarily determined to construct TTS and ASN models. After the model was trained by the training set, the test set will be used for predictions, which will be compared on their effect later.

6. Conclusion

In this paper, a rice yield prediction model based on LSTM long-term neural network is established. The data from 2015 to 2017 are used to train and predict the LSTM model, and the prediction results are obtained. The experimental results show that:

In the years of 2015-2017, a rice yield prediction model based on LSTM long-term neural network was established on the rice yield and meteorological data sets of 81 counties. It is verified that LSTM long-term neural network has good fitting effect on rice yield prediction. The mean square error of the model on the training set is small, which can guide the rice yield in the next few years.

The training set is trained by observing the LSTM long-term and short-term neural network model. The mean square error of the test set decreases with each epoch. The maximum mean square error is 214.4780, and the minimum mean square error is 0.1304. The training process can show that the LSTM long-term and short-term neural network has good stability, and each training can obtain good fitting effect, which has high reliability for the prediction results of rice yield.

LSTM long-term neural network model has better accuracy than SVR model in rice yield prediction under the influence of meteorological factors, but because of LSTM long-term Neural network has complex network structure, modeling is more complex than SVR, not suitable for real-time prediction.

Acknowledgments

The National Natural Science Foundation of China (11974084); Characteristic innovation projects of colleges and universities in Guangdong Province(2018KTSCX091); College Students’ Innovative Entrepreneurial Training Plan Program (202010566021).

References

[1] Guoping Yu.” Analysis on the strategic position of rice in China's food security”. New West(the latter half of the month) 11(2009):31-33. doi:CNKI:SUN:XXBL.0.2009-11-026.
[2] Sepp Hochreiter, Jürgen Schmidhuber. Long Short-term Memory[J]. Neural Computation.
9(9):1735–1780,1997.

[3] Gustavsson A, Magnuson A, Bolmberg B, et al. On the difficulty of training Recurrent Neural Networks[J]. Computer Science, 2013

[4] ZHAO Shufang and DONG Xiaoyu. “Research on Speech Recognition Based on Improved LSTM Deep Neural Network.” Journal of Zhengzhou University (Engineering Science). 39.05 (2018):63-67. doi:10.13705/j.issn.1671-6833.2018.02.004.

[5] Wang Yi, Xie Juan and Cheng Ying. "Deep Neural Networks Language Model Based on CNN and LSTM Hybrid Architecture." Journal of The China Society for Scientific and Technical Information | J Chin Soc Sci Tech Inf 37.02(2018):194-205. doi:CNKI:SUN:QBXB.0.2018-02-008.

[6] Pengfei Zhou. "Research on text detection and recognition technology natural scene image". 2019. Xi’an University of Technology, MA thesis.

[7] Sathasivam S. Logic Learning in Hopfiled Networks[J]. Modern Applied Science, 2009.

[8] Liu P, Qiu X, Huang X. Recurrent neural network for text classification with multitask learning[J]. 2016:2873-2879

[9] Jiang Ping. "Research on Prediction of Landslide Displacement Time Series based on Recurrent Neural Network". 2017. Huazhong University of Science and Technology, PhD dissertation.

[10] Rao Zhang. "Action Function Awareness of RNN Algorithm Optimization". 2017. Zhejiang University, MA thesis.

[11] Zhaodong Pan, et al. "RESEARCH ON SLIDING MODE DECENTRALIZED CONTROL BASED ON ADAPTIVE LEARNING RATE RBF NEURAL NETWORK FOR LARGE-SCALE ENGINEERING STRUCTURES." ENGINEERING MECHANICS 36.09(2019):120-127. doi:CNKI:SUN:GCLX.0.2019-09-015.

[12] Jiao Liu, et al. "A combined tide prediction model based on harmonic analysis and ARIMA-SVR." Journal of Shanghai Maritime University 40.03(2019):93-99. doi:10.13340/j.jsmu.2019.03.016.

[13] Zeng Shaohua. "The Theory Research of Algorithm on Support Vector Regression and Application". 2006. Chongqing University, PhD dissertation.