Global Temperature Prediction Based on SARIMA+LSTM Model

Siyi Lyu*[a], Wenhui Wu[b], Xinmei Ma[c], Fei Hu[d]

[*a]School of Reading, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, People's Republic of China, 210044

[b]School of Science，Guilin University of Technology, Guilin, Guangxi, People's Republic of China, 541004

[c]School of Mathematics and Statistics, Sichuan University of Science and Engineering, Yibin, Sichuan, People's Republic of China, 643000

[d]School of Economics and Management, Shihezi University, Shihezi, Xinjiang, People's Republic of China, 832003

[*] Corresponding author:1184007046@qq.com

Abstract

After the Industrial Revolution, global temperatures generally began to rise. With global warming as the background, this paper builds a forecast model of future temperature, studies to determine whether the temperature rise in March 2022 leads to a larger rise, and builds more than two models to predict the temperature and evaluate the model. SARIMA(1,1,1)(1,1) model and LSTM model were established for temperature prediction. The study found that the temperature will not reach 20C in either 2050 or 2100, and the global average temperature is expected to reach 20C around 2180. Through model evaluation, the r square of SARIMA model is 0.87, and the accuracy of LSTM model is 0.783. Therefore, we think the SARIMA model is superior.

Keywords: Grey correlation analysis, global temperature warming, multiple regression analysis

1 Introduction

Since the industrial revolution, the gases such as carbon dioxide which called heat-absorbing greenhouse's amount in the atmosphere is increasing day by day. The data shows that before the industrial revolution the concentration of carbon dioxide (CO2) was around 280ppm,but it has reached 377.7ppm in 2004,The greenhouse effect of the atmosphere is also being enhanced，which have attracted the attention all over the world with series of problem it cased[1, 2].

Numerous report show that global temperatures are rising， and many countries have declared heat an emergency， Therefore，it is of great significance to study the factors affecting global climate change and make temperature prediction for controlling global warming.

2 Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Description
d	number of differential
It	Input door
Ft	Door of oblivion
Ot	Output gate
Ct	Memory cell
ht	Layer of hiding
qt	The output at time t

3 LSTM prediction model

3.1Long Short Term Network

Long Short Term Network -- commonly known as LSTM -- is a model that can keep only relevant information to make predictions and forget about irrelevant data. This applies to the long time series data in this question. There are three gates in LSTM, the forgetting gate determine the information memory and forgetting at every moment. The input gate determines how much new information is added to the cell, the oblivion gate determines whether information is forgotten at any given moment, and the output gate determines whether information is output at any given moment.

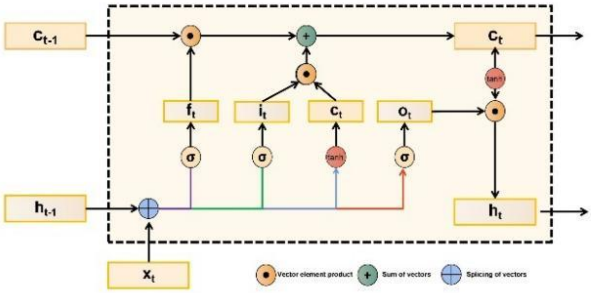

Figure 1 LSTM schematic

In the picture, i, f, o, c, h respectively represent the input gate, the output gate, the memory cell, the hidden layer and the output gate.

The first step of LSTM is to decide what information to discard for the cell state. This part of the operation is handled through a sigmoid unit called a forget gate.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

(1) The next step is to decide what new information to add to the cell state. First use h_{t-1} and x_t to determine what information to update through the operation of the input gate, then use h_{t-1} and x_t to get $\tilde{c}_t$ through a tanh layer, and then update the old c_{t-1} to c_t

(2) Finally, sigmoid test was performed to find out which states of cells needed to be outputted.

$$\begin{aligned} i_t &= \sigma(Wi \cdot [h_{t-1}, x_t] + b_t) \\ \tilde{c}_t &= \tanh(Wc \cdot [h_t, x_t] + b_c) \\ c_t &= f_t * c_{t-1} + i_t * \tilde{c}_t \end{aligned} \tag{2}$$

$$\begin{aligned} o_t &= \sigma(Wi \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(c_t) \end{aligned} \tag{3}$$

3.2 Model training and results

We put the data from 1850 to 2022 into the LSTM model for regression, and forecast the data for the next 100 years to get the results:

As can be seen from the figure, the temperature presents an overall rising trend and tends to be stable at the end.

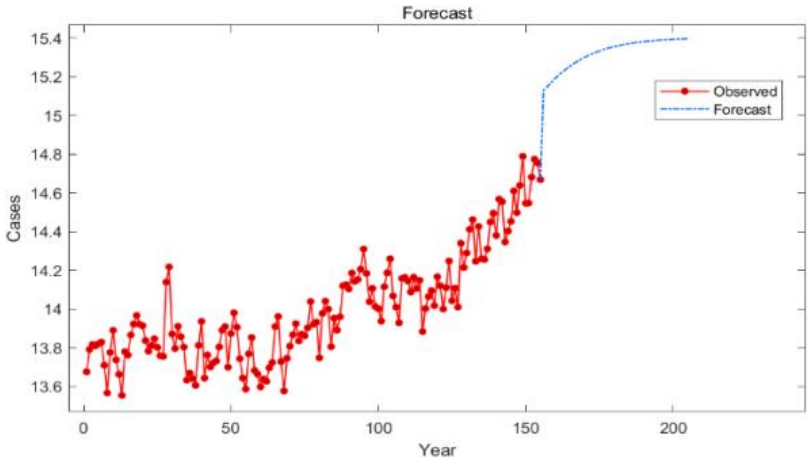

Figure 3 Global temperature forecast

Projections of global temperatures

Above, we built SARIMA(1,1,1)(1,1,1) model and ASDF model to fit the temperature, and then we will use this part of data to predict the future temperature.

SARIMA(1,1,1)(1,1,1) model predict the future temperature

Solve the SARIMA model and use it to achieve temperature predictions to 2100. The results are shown in the following table:

Table 2 Global mean temperature forecast results

year	global mean temperature(℃)	year	global mean temperature(℃)

2025	15.20	2065	16.14
2030	15.31	2070	16.27
2035	15.42	2075	16.41
2040	15.53	2080	16.55
2045	15.65	2085	16.69
2050	15.76	2090	16.83
2055	15.89	2095	16.98
2060	16.01	2100	17.13

As can be seen from the model, the total mean temperature forecast for 2050 is 15.76 and the global mean temperature forecast for 2100 is 17.13. Projections suggest that the global average temperature will not reach 20℃ in 2050 or 2100. And the temperature will reach 20 in about September 2180

3.3 LSTM model predict the future temperature

Solve the LSTM model and use it to achieve temperature predictions to 2100. The results are shown in the following table:

Table 3 Global mean temperature forecast results

year	global mean temperature(℃)	year	global mean temperature(℃)
2025	15.155233	2065	15.390944
2030	15.227476	2070	15.394888
2035	15.281361	2075	15.397607
2040	15.319721	2080	15.39949
2045	15.346284	2085	15.400801
2050	15.364437	2090	15.401716
2055	15.376789	2095	15.402358
2060	15.385200	2100	15.403351

As can be seen from the figure, the average temperature in 2050 is about 15.36℃, and the average temperature in 2100 is about 15.4℃, neither of which has reached 20℃.

3.4 Evaluation of the model

We evaluate the model by the statistics returned by the model:

Table 5 Statistics of model

		Model fit statistics			Yang Box Q (18)			
model	Number of predictive variables	Stationary R square	R square	RMSE	Statistics	DF	significance	Number of outliers
Temperature - Model 1	0	.486	.872	.150	12.121	14	.597	0

R squared represents the fitting degree of time series, the closer to 1 the better effect R squared represents the fitting degree of time series, the closer to 1 the better effect, here the R squared is 0.87, we can think the fitting effect is good. The statistical significance of Young-Box Q (18) is P=0.597, greater than 0.05. Therefore, the null hypothesis is accepted, and it is believed that the residual of this series conforms to random distribution, and no outliers appear, which reflects the acceptable data fitting effect.

Through the evaluation of the two models, the advantages and disadvantages of the two models can be inferred from the indicators such as goodness of comparison. We believe that the SARIMA model is superior.

We analyzed the results obtained above by drawing error graphs and fitting curves:

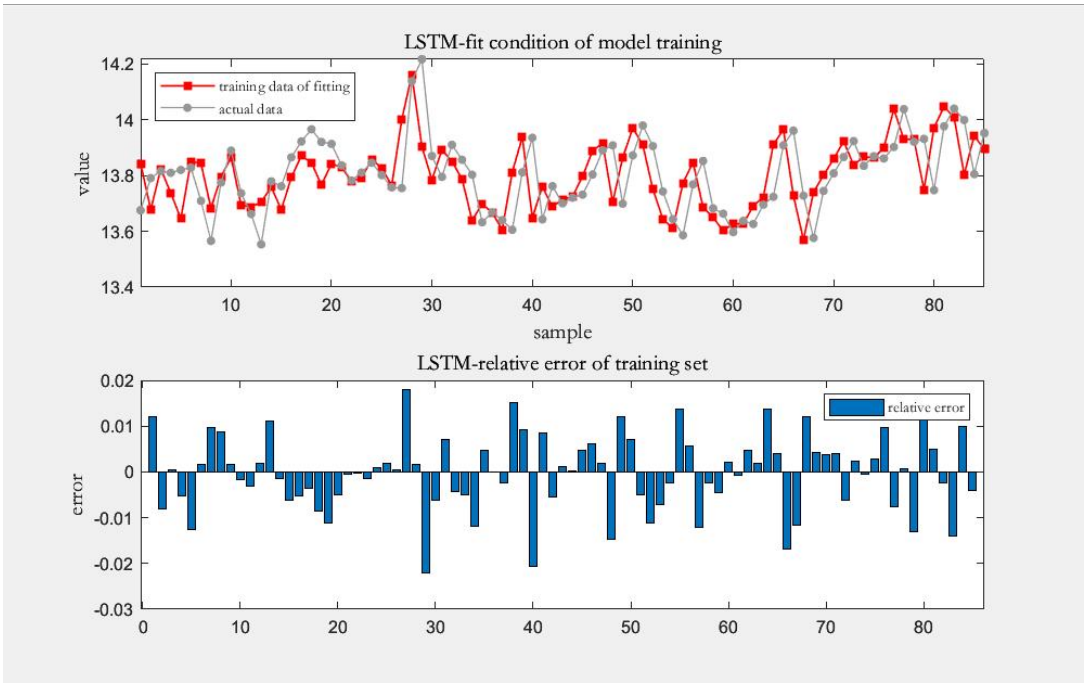

Figure 4 error graphs and fitting curves

It can be seen from the figure above that the fitting effect of the model is good, and the errors are all within 0.03. The accuracy of the returned values of the model is 0.738, indicating that the prediction effect of LSTM is good.

The fitting degree R square of SARIMA is 0.87, which is greater than the accuracy of 0.738 obtained by LSTM. Therefore, we believe that the prediction effect of SARIMA(1,1,1)(1,1) model is better. This may be because SARIMA takes into account the seasonal nature of the data, while LSTM is flattening out over the long term. As a result, we

believe that the prediction effect of SARIMA(1,1,1)(1,1) model is better.

4 Conslusion

In this paper, the future temperature prediction model is established with global warming as the background. It is found that the temperature in 2050 and 2100 will not reach 20℃, and the global average temperature is expected to reach 20℃ around 2180. Through model evaluation, the r square of SARIMA model is 0.87, and the accuracy of LSTM model is 0.783. Therefore, we think the SARIMA model is superior.

References

[1] Ding Yihui, Dai Xiaosu. Temperature changes in China during the last century [J]. Meteorological Monthly, 1994, 20(012):19-26. (in Chinese)

[2] Lang Xianmei, Sui Yue. Prediction of mean climate and extreme climate events in China under 2℃ global warming scenario [J]. Chinese Science Bulletin, 2013, 58(8):9

[3] Greenhouse temperature prediction model based on time series analysis [J]. Transactions of the Chinese Society for Agricultural Machinery, 2010(11):6.

[4] Guimei C , Tong S , Yong Z . Application of Support Vector Machine in Predicting Hot Metal Temperature of Blast Furnace [J]. Control Engineering of China, 2013.

[5] Daouda DIOUF,Awa Niang,Sylvie Thiria. Deep Learning based Multiple Regression to Predict Total Column Water Vapor (TCWV) from Physical Parameters in West Africa by using Keras Library[J]. International Journal of Data Mining & Knowledge Management Process,2019,9(6).

2023 3rd International Conference on Applied Mathematics, Modelling and Intelligent Computing (CAMMIC 2023)

03/24/2023 - 03/26/2023 Tangshan • China

Acceptance Letter

Dear Author(s):

Congratulations! Your manuscript has passed the peer review (the reviewers' comments are available in the attached file on AIS) and has been accepted by the 2023 3rd International Conference on Applied Mathematics, Modelling and Intelligent Computing (CAMMIC 2023). The conference will be held in Tangshan • China from 03/24/2023 - 03/26/2023. We are glad to invite you to attend the conference and make an oral report.

Manuscript No.: J7FNYVWDHY

Author name(s): Siyi Lyu, Wenhui Wu, Xinmei Ma, Fei Hu

Manuscript title: Global Temperature Prediction Based on SARIMA+LSTM Model

Your manuscript, after presented in the oral report or poster in the conference, will be published on Conference Proceedings, after which it will be submitted for index in EI, Scopus.

2023 3rd International Conference on Applied Mathematics, Modelling and Intelligent Computing (CAMMIC 2023)

Academic Exchange Information Center

03-[illegible]-2023

Academic Exchange Information Center AEIC

Notices:

1. Authors need to revise the manuscript as per the reviewers' comments before re-uploading the final version (in Word or PDF) to the AIS system.
2. Authors need to ensure that the submitted manuscript is an original paper with a similarity lower than 20%. Once the manuscript is submitted to AIS, the authors are not allowed to re-submit it to other journals for publication.
3. Authors need to confirm their attendance one week before the conference is held. If the authors are not able to be present on the conference after agreeing to attend the conference, the authors need to reach the conference secretary for re-arrangement.

Again, congratulations and we look forward to meeting you in Tangshan • China