

Research on drought prediction model of LSTM with elevation of water

Yunxuan Fu*

Wuhan Britain-China School, No.10 Gutian Ce Rd., Qiaokou District, Wuhan, Hubei, P.R. China

Abstract. The Drought is one of the most widespread and damaging natural phenomena in the world and have been increasing around the world in recent years. A drought is a persistent shortage of water caused by an imbalance in water supply and demand. The water shortage can be manifested as insufficient precipitation, lack of soil moisture or low elevation of water of rivers and lakes. So, in this paper, according to the recent drought period and the elevation of water data of Lake Mead, the drought prediction model of the elevation of water used long short-term memory (LSTM) neural network was established to predict the elevation of water of Lake Mead in 2025, 2030, and 2050 and drought prediction respectively. The results show that the drought prediction model of the elevation of water used LSTM has the high accuracy in the data set.

Keywords: Elevation of water, Lake Mead, LSTM, Drought prediction.

1 Introduction

According to statistics, the average global temperature has increased by 0.85°C over the past century and global warming has led to a gradual increase in the area, intensity and frequency of droughts. Drought has become a major natural disaster in China [1]. But it is difficult for people to directly quantify the conditions, intensity and degree of drought occurrence. Therefore, relevant researchers have constructed a drought index to describe and evaluate drought conditions based on the characteristics of drought occurrence and measurable empirical data.

Drought is a complex systemic problem and there are many factors that influence its occurrence. It is very important to accurately find out the driving factors that influence drought. Currently, the methods by analysing the relationships between factors and variables include empirical methods, linear correlation methods, principal component analysis and grey correlation analysis [2-4]. But empirical methods are highly subjective and lack an objective basis, linear correlation tests cannot measure non-linear relationship, principal component analysis leads to information loss, so the principal components cannot be explained clearly and extracted, and there is lack of clear grey correlation analysis evaluation indicators.

* Corresponding author: 2962860955@qq.com

In theory, Recurrent Neural Network (RNN) can handle all the sequences, but in practice, due to design flaws such as gradient disappearance and gradient explosion, RNN can only recall the previous steps. The LSTM [5] adds an approximate "conveyor belt" cell design by improving the implicit structure, a "gate" design for selecting information, filtering out past states, filtering out information that has a large impact on the current state, and controlling it. The result of remembering the flow of information can be a good solution to the long-term information storage and memory problem of RNN [6]. Therefore, LSTM is more suitable for the predictive analysis of long-term sequences.

The definition of a drought period varies slightly from country to country and region to region according to the references. A common approach is to classify drought periods as mild, moderate or severe drought based on the number of consecutive days. Linsley et al. (1975) define a hydrological drought as "a period of time during which river runoff is unable to meet water supply needs under a given water management system". A hydrological drought is considered to have occurred if flows remain below a certain threshold over a period of time. The choice of threshold can be based on recent high or low water levels [7].

2 Method

The water level in the drought period defined in the literature is from 60-95% of the highest water level, so the elevation of water can be used to make predictions for the dry period. Firstly, the LSTM method is used to predict the elevation change; secondly, the dry season pattern is defined and predictions are made for future dry seasons. The framework of this paper is shown in Figure 1.

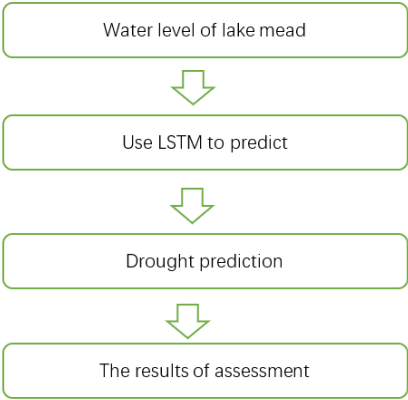


Fig. 1. The framework of this paper.

2.1 LSTM

LSTM model is the variant model of RNN neural network proposed by Hochreiter and Schmidhuber, which solves the problem of forgetting long-term sequence information in RNN network by constructing memory storage unit. The LSTM recurrent neural network mainly relies on the forgetting gate, input gate and output gate play a special role. Under the condition of fixed model parameters, the results of neural units at different times can be dynamically changed, thus avoiding the problem of gradient disappearance or gradient explosion. The schematic diagram of the neuro structure of the LSTM model is shown in Figure 2.

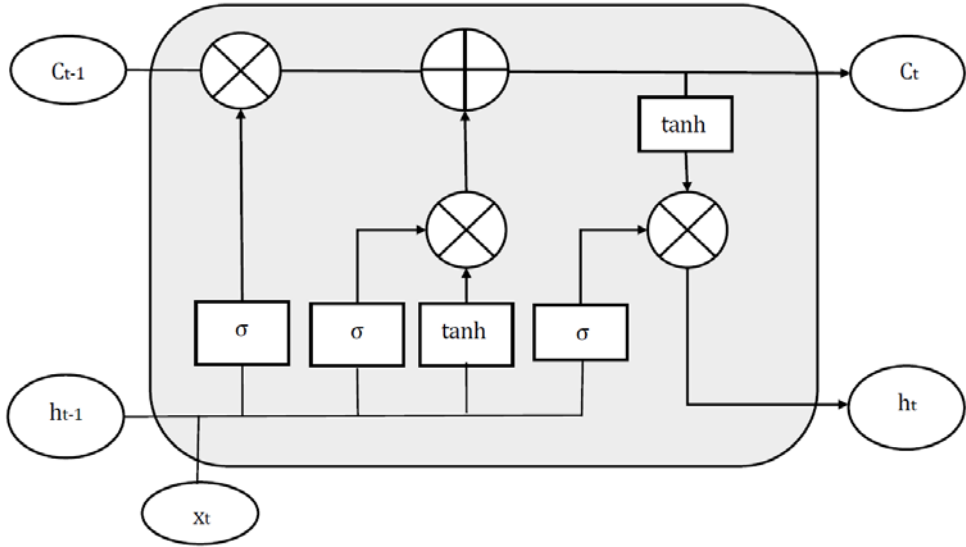


Fig. 2. LSTM model structure.

At the beginning of the work, the forgetting gate is used to control the output of the last cell state by sigmoid function combined with the input of the current state:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \tag{2}$$

Through the function of input gate, the information determines what kind of new information is added into the cell and the new candidate cell state vector is obtained, and then the cell state value at the current time t is obtained by combining the information of the old and new states:

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \tag{3}$$

$$\hat{C}_t = \tanh(W_C * [h_{t-1}, x_t] + b_C) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \tag{5}$$

The last output gate obtains the output of the current moment based on the combination of the past output results, the current moment input and the current cell state.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t * \tanh(C_t) \tag{7}$$

In which: x_t represents the input of cell state at time step t , f_t represents the activation value of forgetting door at time step t , i_t represents the activation value of input door at time step t , $\hat{C}_t$ represents the selection of cell state at time step t , o_t represents the activation value of output door at time step t , C_t and C_{t-1} represent the output cell state at time step t and $t - 1$. h_t , h_{t-1} represents the output of cell state at time step t and $t - 1$, W_f , W_i , W_C and W_o respectively represent the weight matrix of forgetting gate, input gate, cell state and output gate; b_f , b_i , b_C and b_o respectively represent the bias term of forgetting gate, input gate, cell state and output gate.

2.2 Defining the drought period

As described in the above model establishment, the measurement standard of elevation of water in dry period shall be divided into time periods. Taking the 10 years from 2011 to 2020 as the standard, the iterative threshold segmentation method is used to adaptively solve the elevation of water threshold in the drought period. The level altitude during the dry period is below 1099 feet, which was obtained during the analysis of the dry period of the past decade. The schematic diagram of the drought period in the past ten years is shown in the figure below.

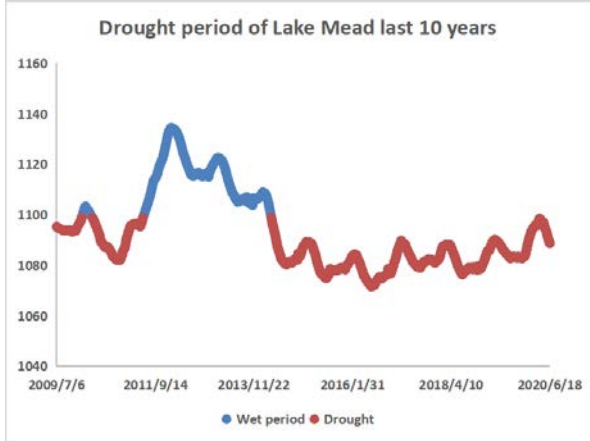


Fig. 3. Drought Period in the past 10 years.

The most recent drought lasted from the end of 2013 to the present. Given that the problem takes several years as a unit, the most recent dry period is seen as 2014-2020.

3 Results and discussion

In order to evaluate the effectiveness of this method, the elevation of water at Lake Mead (in feet above sea level) at the end of each month by year from 1971 to 2021 is used and the data found at <https://www.usbr.gov/lc/riverops.html>. In this paper, because the annual average elevation of water is used to the LSTM, the Monthly average data was processed. The annual average data are shown in Table 1.

Table 1. The elevation of water at Lake Mead from 1971 to 2021.

YEAR	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
WL	1152.3	1158.9	1180.6	1174.1	1177.6	1181.6	1182.2	1186.6	1197.6	1202.8
YEAR	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
WL	1197.0	1200.3	1214.9	1209.2	1209.6	1206.8	1208.9	1204.2	1193.5	1183.2
YEAR	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
WL	1175.3	1176.5	1189.0	1181.8	1183.0	1192.1	1203.7	1213.6	1210.5	1203.5
YEAR	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
WL	1185.1	1162.2	1145.3	1131.4	1140.2	1131.2	1117.2	1109.4	1099.1	1091.1
YEAR	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
WL	1108.0	1121.5	1110.5	1090.1	1080.5	1077.2	1083.2	1081.5	1085.86	1088.6

In order to measure the performance of the model in the past 50 years, mean square error (MSE) is used to quantitatively analysis the error. At the same time, the ablation experiments are provided through multiple regression analysis, ARIMA time series analysis, LSTM to compare the effectiveness of the fitting and prediction of the elevation of water.

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

(2)

The fitting result of the elevation of water for Lake Mead is shown if Fig.4.

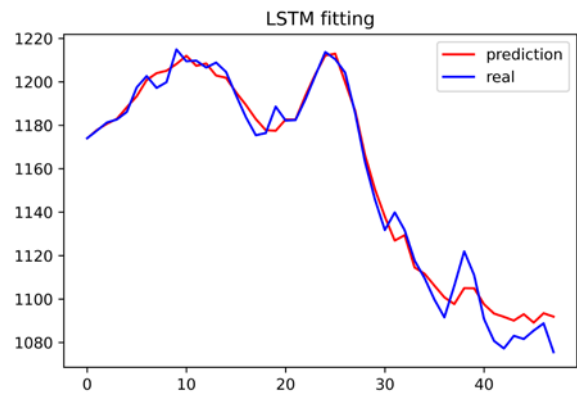


Fig. 4. Drought Period in the past 10 years.

The mean-square error obtained of the is shown in the following table.

Table 2. Formatting sections, subsections and subsubsections.

Method	Lasso	ARIMA	LSTM
MSE	127.35	73.88	44.27

For different data, the performance of these models is often different. Therefore, this article calculates the most suitable model based on historical data, and then uses the model to make predictions. However, LSTM has the best performance in predicting the elevation of water of Lake Mead. Then, uses the elevation of water data from 2005 to 2020 for training, and the predicting effect (from 2021) is as follows.

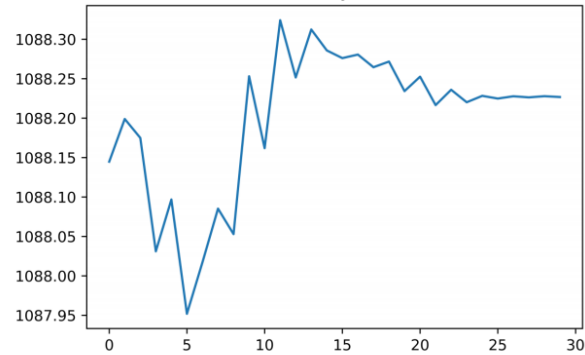


Fig. 5. Prediction of the annual average elevation of water for the next thirty years.

References

1. Du Jianbin. Effects of drought on grain yield in major grain-producing provinces of China and drought countermeasures [D]. Chinese Academy of Agricultural Sciences, 2020.

2. Guo Yafei, Fan Chao, YAN Hongtao. Grain yield prediction based on principal component analysis and particle swarm optimization neural network [J]. Jiangsu Agricultural Sciences, 2019, 47 (19): 241-245.
3. Sun Ying, Zhou Jinlong, Qi Zizhi, Sun Zhenhai, Ji Yanzhen. Variation characteristics and genesis of phreatic evaporation at daily and hourly scales in arid region [J]. Irrigation
4. Journal of drainage, 2020, 39 (S2): 14-19.
5. Wei Fuwei, Huang Rongjuan, Shi Pengfei. Research on coordinated Development of Regional Tourism in Guangxi Based on Grey Relational Analysis [J]. Practice and Understanding of Mathematics, 2020, 50 (16): 103-114.
6. Hochreiter S, Schmidhuber J R. Long short-term memory[J]. Neural Computation, 1997,9(8);1735-1780.
7. Wu Dasuo, Zhang Chuanlei, Chen Jia. Application Research of Computers,2020, 37(S1):86-87+107.
8. L. Drever, W. Roa, A. Mcewan, and D. Robinson. Iterative threshold segmentation for pet target volume delineation. Medical Physics, 2007, 34(4):1253–1265.