A Multi-Channel LSTM Model for Sea Surface Temperature Prediction

Yidi Lin and Guoqiang Zhong*
Department of Computer Science and Technology, Ocean University of China, Qingdao 266100, China

Corresponding author: *gqzhong@ouc.edu.cn

Abstract: Sea Surface Temperature (SST) plays an important role in marine ecology. SST prediction raises considerable attention in ocean-related fields. Recently, deep learning models are widely used in SST prediction, but it is not easy to obtain optimal prediction results using historical observation data directly. More than temporal information, SST data also contains other features, such as trend, periodicity and disturbance. In this paper, we proposed a Multi-Channel LSTM (MC-LSTM) model to improve SST prediction. Firstly, a wavelet transform is used to decompose time-sequence data into multiple sequences representing trend, period and disturbance respectively. Secondly, we use multiple LSTM channels to train these data in parallel, and then obtain the combined prediction results. MC-LSTM can predict Sea surface temperature only by historical SST data, without the help of other spatial and climatic information, so it is easy to get the data. Compared with the direct use of LSTM for prediction, MC-LSTM can improve the prediction accuracy by 26%.

1. Introduction

Sea Surface Temperature (SST) is a critical indicator of sea temperature and a principal factor affecting ocean and atmospheric movement. SST plays an important role in aquaculture, navigation and other fields. Accurate prediction of time distribution of SST is of great significance for scientific research and application. However, due to many factors affecting sea surface temperature (e.g., solar radiation, wind fields and precipitation), accurate prediction of SST is a challenging issue till now.

There are many methods to predict SST, one of them is empirical-statistics method, which is related to data-driven models. It tries to dig out sea temperature variation rules from a large number of SST data. Recently, machine learning models have been widely used for SST prediction. These models include Support Vector Machines (SVMs)[1], neural networks[2], Long Short-Term Memory (LSTM) and so on. Among these models, LSTM models show better prediction performance.

In SST prediction using LSTM model, many new network architectures have been developed recently. Yuting et al.[3] proposed a CFCC-LSTM model, which combines the temporal and spatial information to predict future SST values. Zhang et al.[4] used a multi-layer convolutional LSTM model to predict 3-D ocean temperature. Kim et al.[5] combined a denoising AutoEncoder and convolutional LSTM to predict the ocean weather worldwide. The above methods need to use spatial data or other climate data together in SST prediction, but not all these additional historical data of the areas can be collected and saved, so that the application area of these methods is greatly limited.

In this work, we propose a Multi-Channel LSTM (MC-LSTM) model to predict SST. It combines the wavelet decomposition and LSTM to provide promising prediction results.
2. LSTM and wavelet

LSTM is a special Recurrent Neural Network (RNN) model [6]. RNN model can well process the sequential information and carry out the following prediction tasks, but it has a fatal disadvantage of predicting the long sequences. Because in the process of back propagation, RNN will multiply the parameter w many times, leading to gradient vanishing or gradient exploding. LSTM is proposed to solve this problem.

LSTM uses three gates to control the information flow: forget gate, input gate and output gate. Let \( x^t \) and \( h^{t-1} \) respectively represent the current input and the previous state. Forget gate determines how much of the information of the previous moment \( C_{t-1} \) remains to the current moment \( C_t \), where \( f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \) is the weight matrix of the forgotten gate. Input gate determines how much of the information at the current time \( x_t \) is saved to the unit state \( c_t \), where the formula is \( c_t \circ I_t \), and \( I_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \). The unit state \( C_{t+1} \) is obtained through the step: \( C_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \) multiplying the input gate result \( I_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \), then plus the forgetting gate information: \( C_{t+1} = f_t \odot C_{t-1} \oplus I_t \odot C_t \). Because of the forgetting gate control, you can save information of long sequence and make predictions. Output gate controls how much unit state \( C_t \) output, the formula is \( o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \), \( h_t = o_t \odot \tanh(C_t) \). Figure 1 shows the structure of an LSTM cell.

![Figure 1. Structure of an LSTM cell.](image1)

![Figure 2. Process of wavelet decomposition.](image2)

Wavelet decomposition can decompose one signal into a set of wavelets after displacing and scaling according to different frequencies in the image signal, and decomposes the image information layer by layer[7]. In the predicting of the seawater temperature, wavelet decomposition can decompose complex and disordered data into a series of regular data.

Figure 2 shows the specific decomposition process of wavelet decomposition. The initial signal \( S \) is decomposed into high-frequency signals (detail signals) \( d \) and low-frequency signals (approximation signals) \( a \). As the decomposition goes on, the high-frequency signals gradually increase, while the low-frequency signal takes only the latest one. For example, if five decomposition is carried out, \( S = a_4 + d_1 + d_2 + d_3 + d_4 \).

3. MC-LSTM

The variation of sea water temperature is resulted from many factors, which make the change of sea water temperature have random and nonlinear characteristics. But this variation also has a certain trend and periodicity. For example, the sea water temperature is rising slowly. This rising trend shows that SST time series contains a low frequency component. Because of the season, SST shows a yearly periodic change, this indicates that the SST sequence contains a periodic component with an annual period. If the SST time sequence can be decomposed into different frequency signals to represent these
trends and periodontics, the randomness of the SST sequence will be reduced, and the learning model can achieve a higher accuracy. So we use wavelet decomposition to extract these components.

Our multi-channel LSTM network is mainly divided into two parts. The first part is wavelet decomposition, which decomposes SST time series into signals in different frequency domains. The second part is learning and prediction. The learning model is LSTM, which is composed of multiple LSTM channels, and each channel processes signals with different frequency independently. Finally, the prediction results of these channels are composed.

Figure 3. MC-LSTM network structure.

The overall structure of MC-LSTM is shown in the Figure 3. The row of decomposed SST is the real decomposed result of an SST sequence. The frequency increases from top to bottom. These decomposition results represent some trends and periodic changes. Only the bottom item has no obvious periodicity. It is a noise component, and the prediction accuracy of this channel will be relatively low. Because the decomposition can improve the accuracy of other parts, and the amplitude of this noise signal is very small, the overall prediction accuracy will be improved. The LSTM model structure of each channel is the same, which is constitute by LSTM layer (model the time series relationship) and full-connect dense layer(map the output of LSTM layer to a final prediction).

4. Experiments

4.1. Data Sets

Institute of Atmospheric Physics (IAP) ocean heat content gridded data are collected by Institute of Atmospheric Physics, Chinese Academy of Sciences, and can be downloaded from http://ddl.escience.cn/f/FiL0[8]. IAP database contains the historical global sea water temperature on a regular one-degree latitude/longitude grid and have 41 levels in the vertical at upper 2000m. We created an SST data set from the IAP database. Our data set contains monthly SST values at the point (36N, 121w) of Yellow Sea from 1940 to 2019 (960 months total).

4.2. Experimental Setup

First of all, our model needs to decompose the SST sequence by wavelet. Since prediction results will be composed, we should choose symmetric wavelet basis to avoid phase shift in the process of decomposition. In this experiment, we choose Daubechies wavelet. Wavelet decomposition also needs to consider the number of levels of decomposition. The more levels of wavelet decomposes, the more detailed the low-frequency part will be divided. However, for sequence prediction, as long as the low-frequency part is smooth enough, there is no need to select too many layers, which will increase the amount of calculation. In this experiment, we choose five levels for decomposition.

SST prediction is a sequence prediction problem. We should determine how long the historical data
is used. We define the length of this historical data as LookBack and should determine its value first. Another parameter that needs to be determined in advance is epoch. For other parameters of LSTM, we select adam optimization method, and set the initial learning rate and batch size to 0.1 and 1. In the first experiment, we put the SST sequence directly into an LSTM model for learning, so that we can find the parameters suitable for MC-LSTM by modifying the values of epoch and LookBack. At the same time, for comparison, the prediction results of the LSTM model for SST are presented.

In our experiment, data from 1940 to 2005 are used as training set, and data from 2006 to 2019 are used as test set. The performance evaluation is a fundamental issue. We use Root Mean Square Error (RMSE) and Mean Squared Error (MSE) to measure the effectiveness of different methods. Both RMSE and MSE are the smaller the better.

4.3. Determination of parameters

We first design an experiment to determine the critical values for epoch and LookBack. We put the SST sequence directly into an LSTM model, and change the epoch to search the appropriate value.

![Figure 4. RMSE against epoch for the LSTM model.](image)

We plot the RMSE at different epoch in Figure 4. It can be seen that epoch = 30 is a better choice. When epoch is less than 30, the prediction accuracy is low, but when epoch is greater than 30, the prediction accuracy is not significantly improved, and more calculation time will be consumed.

![Figure 5. RMSE against LookBacks for the LSTM model.](image)

LookBack parameter defines the longest memory input of LSTM, which will affect the accuracy of training and prediction. Figure 5 shows the training and testing effects of different LookBack. It can be seen that better prediction results can be obtained when LookBack is multiple of 12. This is because the data used in the experiment is multi-year observation data in the unit of months, and SST has a cyclical law by year. Meanwhile, when the LookBack is in the range of 12-24, RMSE will maintain a low level. Considering the amount of calculation and prediction accuracy, the LookBack parameter of LSTM model is 16.

4.4. Wavelet decomposition

The data used in the experiment were the monthly average SST within a 1 *1 grid of latitude and longitude area around the position point (36N, 121w) of the Yellow Sea. Data length is 960 (1940-2019). The time series was decomposed into six sequences by DB6 wavelet. Among them, a5 shows an obvious upward trend of fluctuations, and the values are big (14-16), d2-4 show periodicity, and d1 is a noise signal with small amplitudes. s is the sum of d1-d5. The MSE of s and the original SST sequence is 4.25E-6, which indicates the decomposition result can represent the original sequence.
4.5. Prediction result

We trained LSTM model and MC-LSTM model with the same set of SST data. Epoch, LookBack and other parameters are set in the same, with epoch = 30 and LookBack = 16. Figure 6 shows prediction results of decomposed sequence, LSTM and MC-LSTM. The RSME and MSE of these results are listed in Table 1. It can be seen that the decomposed sequence can obtain lower RSME, which is between 0.02 and 0.31. Compared with the LSTM model without decomposition, the RME of MC-LSTM is reduced by 26%, the MSE of MC-LSTM is reduced by 48%. From Figure 6, we can also find that low and high frequency series can obtain higher accuracy while middle frequency obtain lower accuracy.

|                        | a5 | d5 | d4 | d3 | d2 | d1 | MC-LSTM | LSTM |
|------------------------|----|----|----|----|----|----|--------|------|
| RSME of Training data  | 0.02 | 0.03 | 0.15 | 0.31 | 0.20 | 0.18 | 0.70   |
| RSME of validation     | 0.03 | 0.05 | 0.23 | 0.44 | 0.30 | 0.24 | **0.6064** | 0.82 |
| MSE of validation      | 0.00067 | 0.0022 | 0.0509 | 0.1947 | 0.0881 | 0.0558 | **0.3678** | 0.7072 |

Figure 6. Prediction results of decomposed sequence, LSTM and MC-LSTM.

5. Conclusion

In this paper, the MC-LSTM model that combines wavelet decomposing and LSTM is proposed to predict sea surface temperature. It can be predicted only by historical SST data, without the help of other spatial and climatic information. Experimental results show MC-LSTM model offers better prediction accuracy than LSTM model. Because we do not use spatial and climatic information other than SST, the performance of our model is lower than that of other models. In the future work, we will try to improve the performance of our model in order to apply it in the practical application.

References

[1] Lins I, Moura M, Silva M, Droguett E, Veleda D, Araujo M and Jacinto C, Sea Surface Temperature prediction via Support Vector Machines combined with Particle Swarm Optimization. 2010 Int. Probabilistic Safety Assessment & Management Conf.
[2] Tangang F, Hsieh W and Tang BJCD. Forecasting the equatorial Pacific sea surface temperatures by neural network models. 1997 Clim. Dyn. 13 135-47.
[3] Yang Y, Dong J, Sun X, Lima E and Wang X, A CFCC-LSTM Model for Sea Surface Temperature Prediction. 2017 IEEE Geosci. Remote. Sens. Lett. 15 207-11
[4] Zhang K, Geng X, Yan X, Prediction of 3-D Ocean Temperature by Multilayer Convolutional LSTM. 2020 IEEE Geosci. Remote. Sens. Lett. 17 1303-7
[5] Kim K, June-Beom L, Roh M, Han, K and Lee G, Prediction of Ocean Weather Based on Denoising AutoEncoder and Convolutional LSTM. 2020 J. Mar. Sci. Eng. 8 805
[6] Hochreiter S and Schmidhuber J, Long Short-Term Memory. 1997 Neural Comput. 9 1735-80
[7] Mallat S, A theory for multiresolution signal decomposition: the wavelet representation. 1989 IEEE Trans. Pattern Anal. Mach. Intell. 11 674-693
[8] Cheng L, Trenberth K, Fasullo J, Boyer T, Abraham J and Zhu J, Improved estimates of ocean heat content from 1960 to 2015. 2017 Sci. Adv. 3 e1601545