ENSO Prediction with Different LSTM Structures

Bin Mu, Cheng Peng and Shijin Yuan

School of Software Engineering, Tongji University, Shanghai 201804, China

{binmu, pengcheng}@tongji.edu.cn, yuanshijin2003@163.com

Abstract

We applied different LSTM networks to study El Niño Southern Oscillation (ENSO) in this paper, which can be formalized as a time series forecasting problem of meteorology field. We design different experiments to testify the effectiveness of prediction models with different time series steps, and explore the performance between different LSTM structures. The results show that the LSTM networks have the ability to capture the long dependencies between the SST data, then forecast the ENSO phenomenon with a high accuracy, which can be considered as a supplement model for the conventional prediction methods, especially with longer multiple steps forecast ahead.

1 Introduction

Approximately every 4 years, the sea surface temperature (SST) is higher than average in the eastern equatorial Pacific. This phenomenon is called El Niño-Southern Oscillation (ENSO) and is considered as the dominant mode of interannual climate variability observed globally (Wunsch 1990). ENSO is associated with many climate changes (Fraedrich 1994; Wilkinson et al.1999), affecting climate of much of the tropics and subtropics, then cause enormous damage worldwide, so a skillful prediction of ENSO is strongly needed.

So far, both dynamical and statistical models have been applied for predicting ENSO (Barnston 2012). As the conventional models for studying ENSO, the dynamical models use physical equations of the ocean and atmosphere to predict ENSO while the statistical models use mathematical formulations to learn from the observed data to predict ENSO. However, ENSO is still not predicted well enough up to 6 months due to the existence of predictability barrier (Goddard et al. 2001; Duan et al. 2013), and the computationally are very expensive while applying those climate models. Those reasons remain room for further study of this problem.

Recently, some researchers have attempted to apply machine learning methods for this problem, especially deep learning methods. Long Short-Term Memory (LSTM) (Hochreiter 1997) networks have been applied to predict SST (Zhang 2017) and tropical cyclones (Li 2017), which are highly related to ENSO phenomenon (Hong 2001; Catto 2012). A hybrid model with ANN and ARIMA model have also been used to predict ENSO, and the result is slightly better than an ensemble prediction model (Nooteboom 2018). Those preliminary results show great potential with LSTM networks for studying ENSO event further. However, most of the presented works focused on applying the existing model directly or constructing more complex dataset for a better result. Little work has been done to explore the influence of difference model structures to the final prediction result.

In this paper, we explore the influence of different LSTM based models to the final result. Concretely, an LSTM-based network framework is properly designed with fully-connected (FC) layers to form a regression model for ENSO prediction. LSTM layers are utilized to model the temporal relationship among time series data. FC layers are applied to map the output of previous layers to the final result. We compare the influence of different inner structure of models to the the prediction result with different lead time steps, then compare the result of optimized model with other existing methods to prove the effectiveness of proposed model.

The remainder of this paper is organized as follow: Section 2 formalizes the ENSO forecasting problem, which we transfer it as a time series problem. In Section 3, we discuss the data and applied model details. The experiment results and detailed parameters setting are reported in Section 4. The paper concludes with a summary and discussion in Section 5.

2 Problem Formulation

To quantify ENSO, there exist several indices to monitor the tropical Pacific, all of which are based on SST anomalies averaged across a given region (Figure 1). Among those different NINO indices, NINO3.4 index is the average SST anomaly in the region bounded by 5°N to 5°S, from 170°W to 120°W. This region has large variability on ENSO time scales, and is close to the region where changes in local SST are important for shifting the large region of rainfall typically located in the far western Pacific. NINO3.4 index is one of the most commonly used indices to define ENSO phenomenon. Operationally, an ENSO is said to occur when the NINO3.4 index is above the threshold of +0.5°C for at least 5 consecutive months. We use NINO3.4 index as the predictant. The prediction of ENSO can be regarded as a problem of predict NINO3.4 index with different time steps ahead.

Concretely, we list the observed NINO3.4 indices as a time series sequence as input (training data) to train our model, then use the model to predict the next several monthly NINO3.4 indices with different prediction steps, and compare the result with observations and prediction result of climate models. It is worth mentioning that the length of prediction is critical here, as a prediction with lead times up to 1 year, is high desired (Nooteboom 2018).

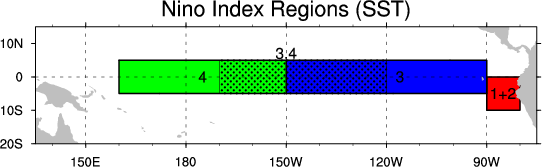


Figure 1: Different Nino index Regions.

3 Data and Model

Dataset Description

We use the National Oceanic and Atmospheric Admin- istration’s (NOAA’s) climate time series data for our experiment, which including the longest timescale monthly NINO3.4 index data (from 1870 to May 2018) to best of our knowledge, and it’s periodically updated. The data is available on this website: <https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/>.

From theories of data preprocessing, it seems wiser to investigate different data preparation techniques to rescale (normalization) the data and make the data stationary before put them into training. However, the data of NINO3.4 index seems born with those well-processed data properties – The mean is close to zero and the distribution is near Gaussian. The NINO3.4 index is both the input and output of our models.

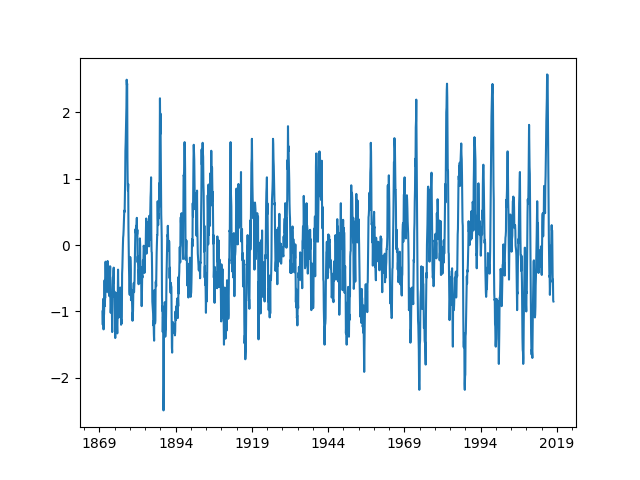


Figure 2: Line plot of monthly NINO3.4 Index from 1870 to 2017.

It should be noted that, some works have been done for the same forecasting problem. Most of them focus on figuring out the best parameters combination for a better result, and some of them use the climate network theory (Albert 2002) to involving as much data as possible. However, the potential of the prediction model itself is rarely discussed, we have only use limited size of data in this paper, and concentrate on improving the performance of model and explore how different structures influence the final result.

MIMO Time Series Forecasting

The task of multiple steps time series forecasting is to predict two or more steps ahead using single-output or multiple-output strategy (Taieb 2012), and the multiple-output strategy is called multiple-input multiple-output (MIMO) time series forecasting. It models time series data as a multiple-input multiple-output function.

More formalized, based on the previous observed sequence X = {} as the input, predict the future sequence Y = {}, *n* and *m* stand for the length of input and output respectively.

Long Short-Term Memory

LSTM is a special kind of recurrent neural network (RNN), which can capture the temporal relationship among time series data. Compare with conventional RNN, LSTM addressed the vanishing and exploding gradients problem during training process, then solving the long-range dependencies among time series data, the whole structure of a LSTM cell is showed in Figure 3, and the whole computation process can be defined as a series of equations (Graves 2013) showed in (1):

(1)

where is the activation function (here is sigmoid function), ,, and represent the recurrent weight metrics, ,,, and are the bias terms respectively. *H* represents a concatenation of the new input and the previous hidden vector .

The key of LSTM is the cell state, which plays the role of control the information added or removed, carefully regulated by structures called gates. There exist 3 kind of gates in (1): *i*, *f* and *o* represent *input gate*, *forget gate* and *output gate.* The input gate decides how much information enters the current cell. The forget gate decides how much information be forgotten(removed) from the previous memory vector . The output gate decides what information will be output from the current cell.

There have many works on exploring the structure of LSTM (Yao 2015, Greff 2017), but not go deep into a specified meteorology task. We need develop a model based on convention LSTM to fit our ENSO forecasting task.

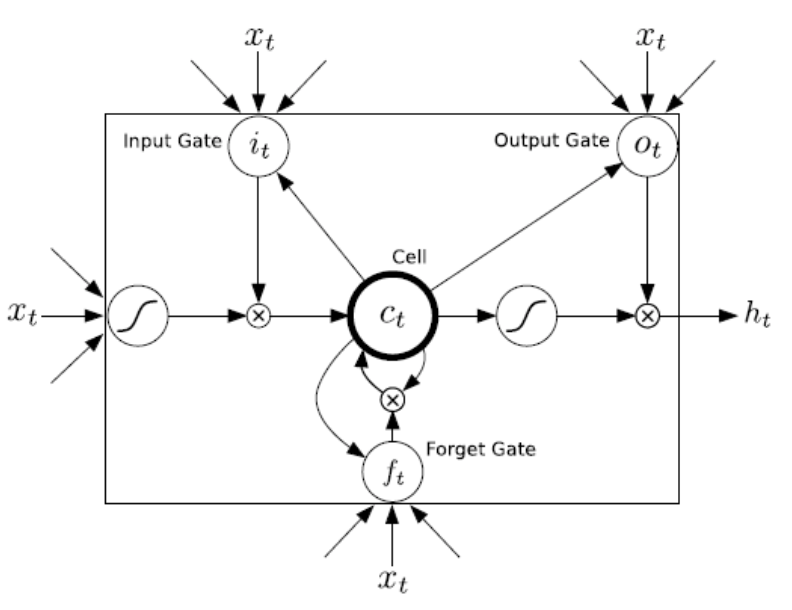


Figure 3: LSTM (Hochreiter 1997).

LSTM Based Model

From the data processing perspective, a time series data is not suitable as dataset for machine learning inborn. Before put data into the network model, we must try the prepare data process first, this process including 3 subtasks: data split, preprocessing and sequence generate (detailed parameters setting will be discussed in Section 4).

After getting the processed dataset, we put those data to LSTMs based model. In general, to construct a fully end-to-end model, we need combine the LSTM layers and Fully-Connected (FC) layers to form a LSTM block, the output of LSTM layer is a vector of the state of the last time step, then the FC layer is used for data abstraction and reduce dimensionality, then the output of network model is a series of data that stand for the next status, which means the NINO3.4 index of next several months for ENSO. The whole model structure showed in Figure 4.

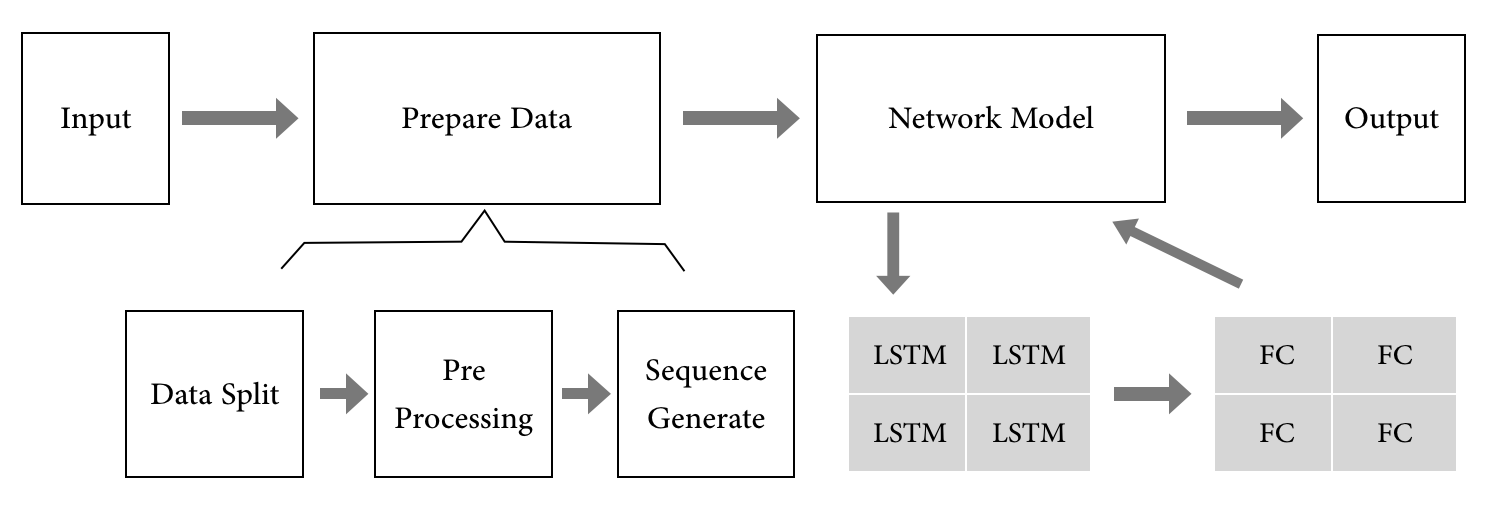


Figure 4: Model Structure.

4 Experiment

This section presents the prediction results of LSTM based model, and compared with observations and with alternative predictions from convention neural networks and climate models. Concretely, we first discuss performance between the LSTM models and the neural networks with same magnitude of parameters and same training epochs, then compares the LSTM model result with different climate models with same prediction length, which is available at: <https://iri.columbia.edu/our-expertise/climate/forecasts/enso>, and the ENSO during 2015/16 is the nearest ENSO from timescale, and it is considered as one of the most extreme ENSO up to now (Santoso 2017), We discuss the result of our model and observation during 2015/16 finally.

From the perspective of classical machine learning data partition, 80/20 is a most commonly used partition, However, since the limited size of NINO3.4 index data and the periodicity of ENSO, we use the data from 1870.01 to 2009.12 (1680 months) as the training set, and the data from 2010.01 to 2017.12 (last 96 months) as the testing set. We use rolling forecast method to construct data sequences, which is often used in long-term numerical prediction problems. We use Adam (Kingma 2014) to optimize training process, and our model is implemented based on Keras (Chollet 2015).

We use the root mean squared error (RMSE) as the performance evaluation, which is one of the most used common measurement in numerical problems, as to the ENSO during 2015/16, more details will be discussed besides purely data.

Compare with Convention Neural Networks

Before discuss the LSTM result with numerical climate models, let’s see its performance compared with convention neural networks first.

For contrast, we keep all the irrelevant variables consistent, just only difference is the structure of models. Table 1 and Table 2 show the result of LSTMs and Multilayer Perceptrons (MLPs) respectively.

The result shows that with different layers and cells, the RMSE belong to in a relatively stable interval. The LSTM network with 1 layer and 10 memory cells achieve the best prediction performance on the metric of RMSE. Furthermore, LSTM is originally designed for solving the long dependencies between data, from our result, we can see the superiority of LSTM here – with longer prediction sequence, the prediction skills decrease generally (RMSE increase), but the decrease ratio of LSTM is much slower than MLP with same structure, which means that LSTM has better performance with longer prediction steps.

We can see that with more layers and more neural nets, the result is not always better, both LSTM and MLP does. Actually, experience in previous study show that neural layer is not the more the better. During experiment, we found that more layers and more cells is likely to get unstable result and cost more computing resources. We have tried apply more complex models and involving more training parameters, but the improvement is not obvious.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | (1-1) | (1-10) | (1-20) | (2-1) | (2-10) | (2-20) |
| 6-month lead | 0.588 | **0.550** | 0.582 | 0.567 | 0.586 | 0.558 |
| 9-month lead | 0.699 | **0.661** | 0.704 | 0.674 | 0.720 | 0.677 |
| 12-month lead | 0.771 | **0.727** | 0.777 | 0.748 | 0.807 | 0.754 |

Table 1: LSTM result with different month lead (6-, 9- ,12-).

(x-y means x layers with y cells for each layer, all networks end with a FC layer.)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | (1-1) | (1-10) | (1-20) | (2-1) | (2-10) | (2-20) |
| 6-month lead | 0.614 | 0.581 | 0.577 | 0.610 | **0.572** | 0.579 |
| 9-month lead | 0.742 | 0.711 | 0.708 | 0.735 | **0.704** | 0.712 |
| 12-month lead | 0.825 | 0.795 | 0.795 | 0.817 | **0.795** | 0.799 |

Table 2: MLP result with different month lead (6-, 9- ,12-).

Compare with Climate Models

A 9-month ahead prediction of different climate models of ENSO is available on International Research Institute (IRI) ENSO forecast. We take the average prediction result of all models as the benchmark of climate models here, then make comparison with LSTM result.

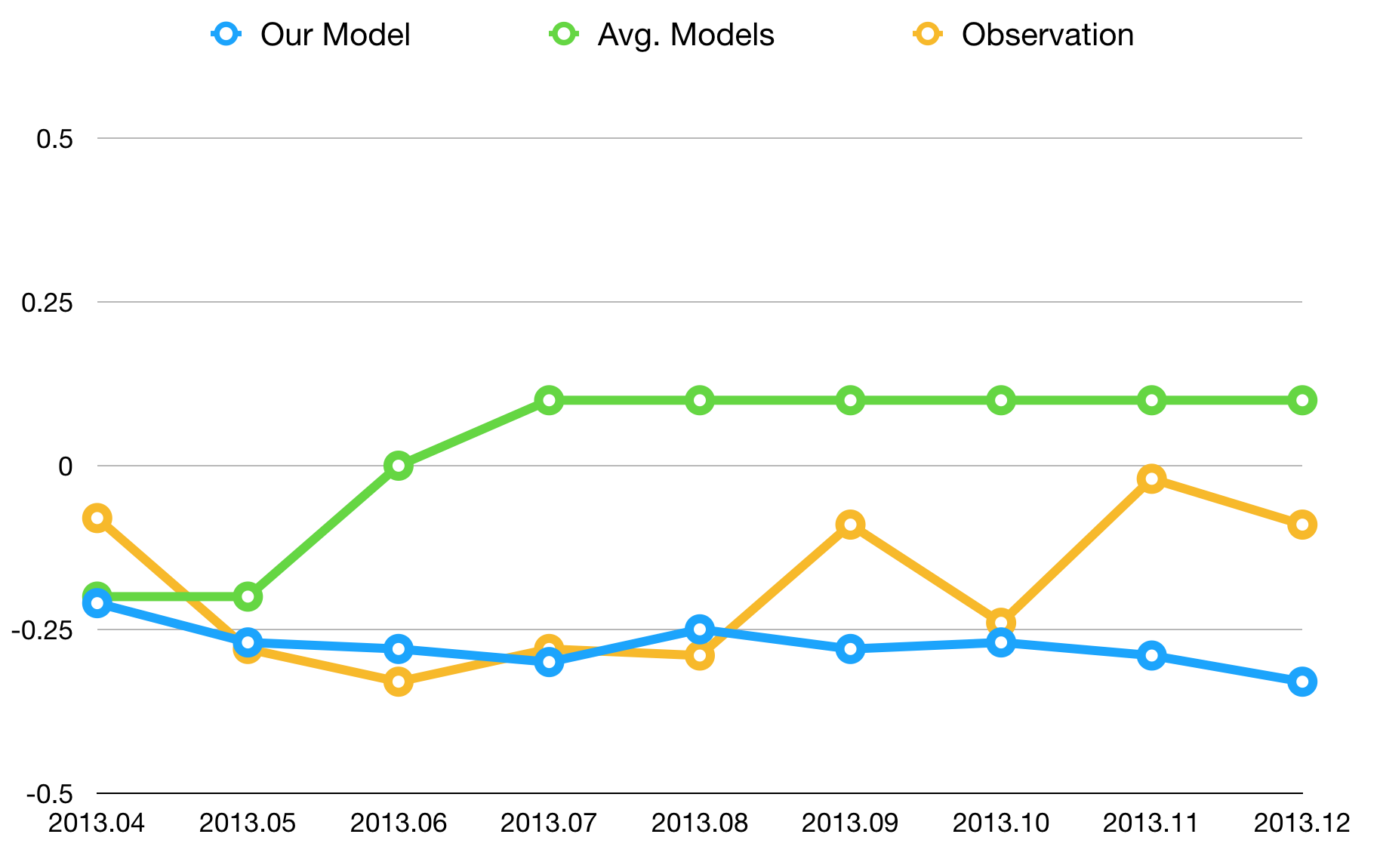


Figure 5: The 9-month ahead prediction of 2013.04 between our model, climate models and observation. (Best viewed in color)

The change of NINO3.4 index is relatively stable during 2013/14, especially during year of 2013 (after a moderate ENSO of 2009/10). We take the 2013.04 as an example here, predicting 9 months ahead. From the result of climate models, most of them predict an increasing trend of NINO3.4 index, which show great potential of an ENSO event in 2014. However, the final observation indicates that the prediction of most of the models forecast the next ENSO too early. In fact, the next ENSO comes during mid-2015. The prediction result of our model fits the observation better than climate models obviously here, both the RMSE metric and the actual fitting curve result (Figure 5). Our model shows better performance during the moderate stage of NINO3.4 index.

ENSO Event During 2015/16

The ENSO of 2015/16 is considered as the first extreme ENSO of the 21st century. From the perspective of NINO3.4 index, the peak of the NINO3.4 index reached the maximum nearly 2015.11, then decreased nearly begin of the year 2016.

For comparison, we list the development of NINO3.4 index between our model, climate models (both statistical model and dynamical model) and the observation from 2015.08 (Figure 6). As this is the strongest ENSO event observed since 1950 and took place almost two decades after the previous major event in 1997/98(L’Heureux 2017), all the prediction models underestimated the peak value of this event. However, all the 3 predictions result shows that the peak value will arrived at 2015.11~2015.12, which is coherent with observation.

After the ENSO arrived the peak value at Dec 2015, it keeps about 3 months, while the prediction results of both the statistical models and dynamical models decrease with a very high speed after reach the peak, our proposed model keep about 3 months from line point, which is more coincide with reality.

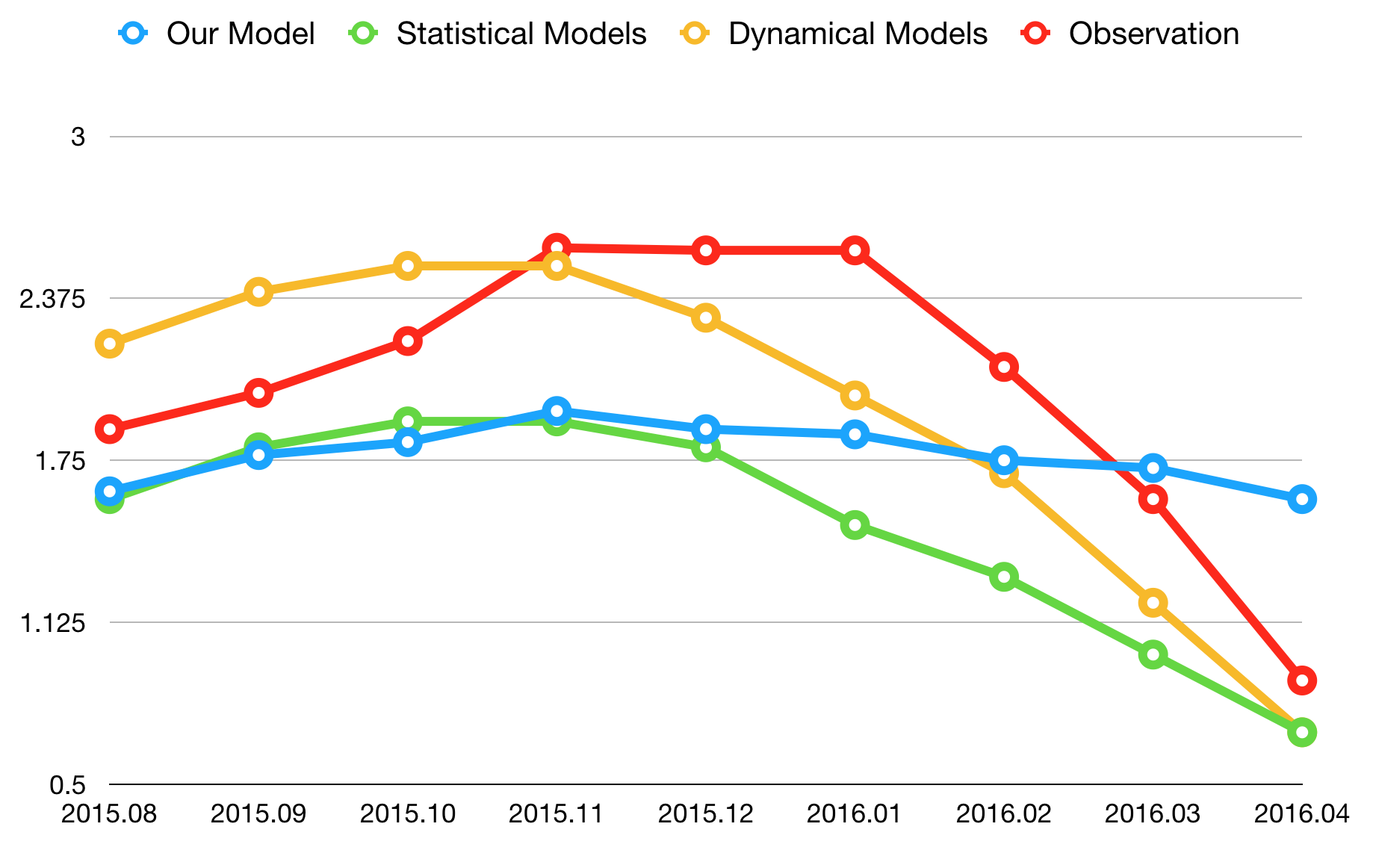


Figure 6: The 9-month ahead prediction ENSO during 2015/16, comparsion between our model, climate (statistical and dynamical) models and observation. (Best viewed in color)

5 Conclusion and Future Works

In this paper, we have made a successful attempt of applying the LSTM based models to study ENSO event and discuss the influence between different LSTM structures to the final prediction result. The result shows that even with limited input data, the performance of well-structured neural network model can still be very effective, even performance better compares with the mainstream approaches in some time.

Even though the currently applied network measures showed inspiring result, there still got some questions to be discussed: the training problem with limited dataset, the interpretability of how network work in this case and how to improve the model.

Over-fitting problem was occurred during our training process, even we have try to extend the dataset with moving average method, the result is not improved obviously. Traditional data enhancement methods seem not very helpful to this topic.

Interpretability among different network models is still a stumbling block to put them into real applications. Still in our ENSO prediction, how the model works is still ambiguous, reasoning the result is a valuable research issue in this case.

For future work, we will investigate how to involve more related data to construct a highly interpretable model for the same question. Besides, single NINO Index prediction is not suitable to solve all the ENSO-related problems, so it’s still a big challenge to consider the spatiotemporal information among the NINO regions, which can simulate the development of SST and offer more information to call for more research works.

References

Chollet, F. 2015. Keras.

Wunsch, C. 1990. El Nino, La Nina, and the Southern Oscillation. *Science* 248(4957), 904-906.

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., & DeWitt, D. G. 2012. Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing?. *Bulletin of the American Meteorological Society* 93(5), 631-651.

Zhang, Q., Wang, H., Dong, J., Zhong, G., & Sun, X. 2017. Prediction of sea surface temperature using long short-term memory. *IEEE Geoscience and Remote Sensing Letters* 14(10), 1745-1749.

Nooteboom, P. D., Feng, Q. Y., López, C., Hernández-García, E., & Dijkstra, H. A. 2018. Using Network Theory and Machine Learning to predict El Niño. *arXiv preprint arXiv*:1803.10076.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., ... & Kaplan, A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres* 108(D14).

Goddard, L., Mason, S. J., Zebiak, S. E., Ropelewski, C. F., Basher, R., & Cane, M. A. 2001. Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology* 21(9), 1111-1152.

Li, Y., Yang, R., Yang, C., Yu, M., Hu, F., & Jiang, Y. 2017. Leveraging LSTM for rapid intensifications prediction of tropical cyclones*. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4.

Hong, C. H., Cho, K. D., & Kim, H. J. 2001. The relationship between ENSO events and sea surface temperature in the East (Japan) Sea. *Progress in oceanography* 49(1-4), 21-40.

Catto, J. L., Nicholls, N., & Jakob, C. 2012. North Australian sea surface temperatures and the El Niño–Southern Oscillation in observations and models*. Journal of Climate* 25(14), 5011-5029.

Dijkstra, H. A. 2006. The ENSO phenomenon: theory and mechanisms. *Advances in Geosciences* 6, 3-15.

Trenberth, Kevin & National Center for Atmospheric Research Staff (Eds). Last modified 02 Feb 2016. "*The Climate Data Guide: Nino SST Indices (Nino 1+2, 3, 3.4, 4; ONI and TNI)*." Retrieved from <https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni>.

Albert, R., & Barabási, A. L. 2002. *Statistical mechanics of complex networks. Reviews of modern physics* 74(1), 47.

Taieb, S. B., Bontempi, G., Atiya, A. F., & Sorjamaa, A. 2012. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert systems with applications* 39(8), 7067-7083.

Hochreiter, S., & Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8), 1735-1780.

Graves, A. 2013. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:* 1308.0850.

Yao, K., Cohn, T., Vylomova, K., Duh, K., & Dyer, C. 2015. Depth-gated recurrent neural networks. *arXiv preprint. arXiv preprint arXiv:*1508.03790, 9.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. 2017. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* 28(10), 2222-2232.

Santoso, A., Mcphaden, M. J., & Cai, W. 2017. The defining characteristics of ENSO extremes and the strong 2015/2016 El Niño. *Reviews of Geophysics* 55(4), 1079-1129.

Kingma, D. P., & Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:* 1412.6980.

Broni-Bedaiko, C., Katsriku, F. A., Unemi, T., Shinomiya, N., Abdulai, J. D., & Atsumi, M. 2018. El niño-southern oscillation forecasting using complex networks analysis of LSTM neural networks. *In ISBC 3rd, January 18-20, 2018, Beppu, Japan.*

Feng, Qing Yi, et al. 2016.ClimateLearn: A machine-learning approach for climate prediction using network measures. *Geoscientific Model Development.*

Duan, W., & Wei, C. 2013. The ‘spring predictability barrier’ for ENSO predictions and its possible mechanism: results from a fully coupled model. *International Journal of Climatology* 33(5), 1280-1292.

L’Heureux, Michelle L., et al. 2017. Observing and predicting the 2015/16 El Niño. *Bulletin of the American Meteorological Society* 98.7 1363-1382.