Periodic Time Series Forecasting with Bidirectional Long Short-Term Memory

Periodic Time Series Forecasting with Bidirectional LSTM

Quoc-Dung Nguyen*
Van Lang University, 45 Nguyen Khac Nhu, Co Giang ward, District 1, Ho Chi Minh City, Vietnam
Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic

Nguyet-Minh Phan
University of Information Technology, VNU-HCM, Quarter 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam

Ivan Zelinka
Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic

ABSTRACT

Deep learning methods such as recurrent neural network and long short-term memory have recently drawn a lot of attentions in many fields such as computer vision, natural language processing and finance. Long short-term memory is a type of recurrent neural network capable of predicting future values of sequential data by learning observed data over time. Many real-world time series in business, finance, weather forecasting and engineering science have periodic property like daily, monthly, quarterly or yearly period and need efficient tools to forecast their future events and values. The forecasting study and tools in these fields are therefore essential and important. In this paper, we present a deep learning technique, called bidirectional long short-term memory, in forecasting time series data. The bidirectional long short-term memory model is evaluated based on the benchmark periodic time series dataset. The model performs well on the macro and industry categories and achieves average mean absolute percentage errors less than 9%. It is shown that the bidirectional architecture obtains the better results than the baseline models. We also test the model by tuning the time step parameter to evaluate how the time step length impacts on forecasting performance of the model.

CCS CONCEPTS

• **Computing methodologies** → Neural networks; • **Applied computing** → Forecasting.

KEYWORDS

Long short-term memory, recurrent neural network, sequence prediction, time series

*Corresponding authors. Email: dung.nguyen@vlu.edu.vn

ICMLSC'21, January 29–31, 2021, Virtual Event, Vietnam

ACM ISBN 978-1-4503-8761-3/21/01…$15.00
https://doi.org/10.1145/3453800.3453812

ACM Reference Format:
Quoc-Dung Nguyen, Nguyet-Minh Phan, and Ivan Zelinka. 2021. Periodic Time Series Forecasting with Bidirectional Long Short-Term Memory: Periodic Time Series Forecasting with Bidirectional LSTM. In *2021 The 5th International Conference on Machine Learning and Soft Computing (ICMLSC'21), January 29–31, 2021, Virtual Event, Vietnam.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3453800.3453812

1 INTRODUCTION

Time series is a sequence of data points collected sequentially in time. Predicting future values of time series is a common problem as seen in many practical fields such as finance, business planning, weather forecasting, as well as applied science and engineering. Several typical examples are forecasting the weather for the next days, daily opening and closing stock prices, electricity consumption in a household or future heart failure.

Time series introduces a dependent relationship among collected data points. Time series forecasting makes use of a prediction model to predict future values based on previous observations, as shown in Figure 1

Time series data often contains periodic variation and trend. A time series has a trend if its mean value is changing over time. Periodic variation, or seasonality, refers to the phenomenon where the data experiences predictive cycles that regularly recur over a time period often less than a year such as daily, weekly, monthly, or quarterly. For example, the number of visitors of a tourist attraction increases during the holidays and weekends, or swimsuit sales increase in the summer and decrease during the winter. Time series with trend or with seasonality are non-stationary.

Recurrent neural networks (RNN) are well-suited to supervised learning problems for the data having sequential nature such as text, voice, video and time series. The RNN networks are designed to capture the relation between the sequential values, hereby it can be applied to detecting periodic patterns in time series. Long short-term memory (LSTM) is an advanced version of RNN [1, 2]. It is capable of handling long sequence dependence among observed inputs and therefore highly suitable for sequence prediction problem. LSTM-based methods and applications have been recently deployed in a wide range of fields such as multimedia processing [3–6], abnormal detection [7–9] and engineering science [10–12]. The bidirectional LSTM network (BiLSTM) is an extension of the

Figure 1: Time series prediction model.

original LSTM architecture where the inputs are fed into the network in two ways, one in the forward direction (from past values to future ones) and one in the backward direction (from future values to past ones).

In our previous paper [13], we have experimented the different architectures of LSTM on the benchmark time series data and shown that the bidirectional LSTM achieves the best performance among the three LSTM architectures. The LSTM models have also obtained the better results than the various baseline models on the same benchmark dataset. However, the data used in the experiments are based on only the yearly dataset, not the seasonal time series. Following that result, in this paper, we verify the bidirectional LSTM model on the time series data with seasonal characteristic in order to evaluate its performance on the periodic time series. The experimental results are compared with those of the baseline models. Additionally, we examine and analyze the performance of the proposed model based on different time step settings.

The rest of the paper is structured as follows. Section 2 presents in detail the architecture of bidirectional LSTM and its applications in practice. Section 3 discusses and compares the experimental results between the bidirectional LSTM model and the baseline models on the quarterly time series dataset. Finally, Section 4 gives our conclusion and future work.

2 BIDIRECTIONAL LONG SHORT-TERM MEMORY

LSTM is a deep learning model for processing sequential data. The LSTM model was introduced by Hochreiter and Schmidhuber [14] and was subsequently refined in the following works [15–17]. It has made significant advancements in various domains like natural language processing, speech recognition and computer vision. It overcomes the drawbacks in traditional RNNs by its capability of capturing long-term temporal dependencies as well as avoiding the exploding and vanishing gradient problems.

LSTM extends the memory capability of RNN by introducing three gates (input gate, output gate and forget gate) to control the flow of information inside the LSTM memory unit, also known as the memory cell. The input gate (1) controls the amount of information learned from the current input and the forget gate (2) controls the amount of information retained from the previous cell, which are together combined in the current cell state (3). The output gate (4) scales the value in the current cell used to compute the output activation or the hidden state (5) of the LSTM unit. This gated memory mechanism enables the network to remember for a long time.

$$i_t = \sigma\left(W_i\left[h_{t-1},\ X_t\right] + b_i\right) \tag{1}$$

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

$$C_t = f_t * C_{t-1} + i_t * tanh\left(W_C\left[h_{t-1},\ X_t\right] + b_C\right) \tag{3}$$

$$o_t = \sigma\left(W_o\left[h_{t-1},\ X_t\right] + b_o\right) \tag{4}$$

$$h_t = o_t * tanh\left(C_t\right) \tag{5}$$

The bidirectional LSTM model consists of two independent LSTM networks, one where the input sequence is processed from left to right and the other from right to left. This LSTM architecture allows the model to learn the input sequence in both forward and backward directions as shown in Figure 2. The interpretations at the outputs of both the forward and backward LSTM networks are combined to generate the predicted value at the next time step.

Here we apply the BiLSTM model to time series forecasting problem. The main idea is that the predicted values are produced using not only the input data, but also the previous output values. In addition, by utilizing the time series data and its reversed copy in making prediction, it can provide the supplementary context to the model and result in faster and more efficient learning on the problem. The time series components like trend or seasonality can be encoded in input sequence, the model is expected to extract these features and use them for future prediction.

3 RESULTS AND DISCUSSION

3.1 Dataset

The M-Competitions [18, 19] have been organized for empirical studies to advance the field of forecasting and make it more practical and beneficial for organizations requiring predictions in scheduling, business, production, or operation. Various methods have been proposed and compared to each other by their forecasting performance on the benchmark datasets.

In order to evaluate the performance of the bidirectional LSTM model on the seasonal sequence data, we make use of the quarterly time series dataset of the M3-Competition database [20]. This database consists of 3003 time series, mainly in business and economic domains. Each dataset is subdivided into 6 categories including micro, industry, macro, finance, demographic and other. The quarterly dataset contains 756 time series with different numbers of observations as shown in Table 1

Like the M3-Competition [18], the number of forecasts is chosen as 8 for the quarterly time series. In other words, for each quarterly time series, the last 8 observations are reserved for evaluating the forecasting performance of the LSTM model, while the preceding observations are used in developing the forecasting model. The forecasted values are subsequently compared with the actual values to measure forecasting accuracy of the models.

The forecasting accuracy is measured by the symmetric mean absolute percentage error (sMAPE) metric for the model performance

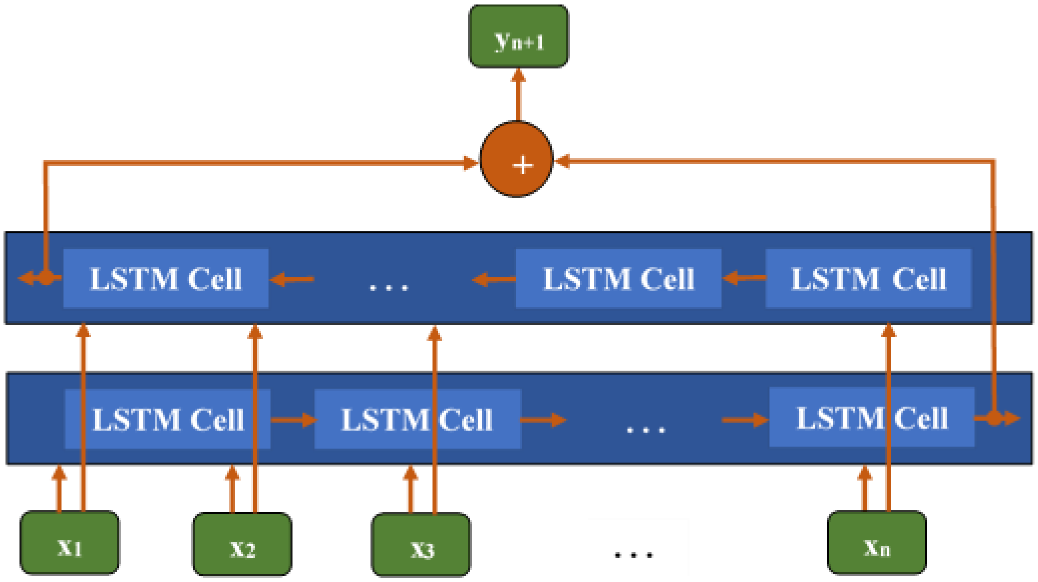

Figure 2: Structure of the bidirectional LSTM.

Table 1: The categories of the quarterly time series in the M3-Competition

Types of time series	Number of time series	Minimum observations	Maximum observations
Micro	204	36	46
Industry	83	32	72
Macro	336	24	53
Finance	76	35	72
Demographic	57	35	72
Other	0	NA*	NA

* Not available.

evaluation, defined as:

$$\frac{100}{N}\sum_{i=1}^{N}\frac{2*|a_i - f_i|}{(a_i + f_i)} \tag{6}$$

where a_i is the actual value, f_i is the forecasted value and N is the number of forecasts. The sMAPE metric is averaged across the horizon of all the forecasts. This metric is often used as an accuracy measure in forecasting competitions because it avoids the problem of large errors when the actual values a_i are close to zero, and the asymmetry in other absolute percentage error metrics when the values a_i and f_i are different.

3.2 Experimental Results and Discussion

The experiments have been run on the system Intel(R) 2-core Xeon CPU 2.20GHz, 13GB RAM. The system is installed with the library packages including Tensorflow version 1.15 and Keras version 2.3 for training and evaluating the bidirectional LSTM model.

Our model consists of one BiLSTM layer of 128 units, followed by a dense layer of a single unit with a linear activation (see Table 2). A dropout of 0.2 is applied between the BiLSTM layer and the dense layer. The learning rate is set to 0.001. The mean squared error loss function was minimized using the Adam optimizer [21]. The gradient clipping in the Adam optimizer is also used to avoid the exploding gradient problem.

Table 3 shows the sMAPE values of the BiLSTM model on the different categories of the quarterly time series dataset. It can be seen that the BiLSTM model obtains the good results on the macro and industry categories with the average sMAPE around 5.04% and 8.60% respectively, while its performances on the other three types of the time series data are worse with the average sMAPE more than 11%. Besides, the overall average sMAPE of the model is less than 9%.

Table 4 presents the sMAPE values of the BiLSTM model on the different forecasting horizons for the quarterly dataset. The BiLSTM model achieves lower absolute percentage errors at the first horizons. The errors become higher for the next time steps due to the error accumulation at each forecasting step. This table also combines the sMAPE results of the baseline models on the same dataset in the M3-Competition. It is seen that the BiLSTM model shows the better results than the baseline models regarding the average sMAPE on the next four, six and eight forecasts. In particular, the average sMAPE of the BiLSTM is 6.98% lower than

Table 2: The layers in the proposed BiLSTM model

Layer	Output Shape	Number of Params
Bidirectional LSTM	(None, 256)	133120
Dropout	(None, 256)	0
Dense	(None, 1)	257

Total params: 133,377

Trainable params: 133,377

Non-trainable params: 0

Table 3: The sMAPE values of the BiLSTM model on the different categories

Category of time series						Average
Micro	Industry	Macro	Finance	Demographic	Other	
11.50*	8.60	5.04	14.25	11.04	NA**	8.55

* The sMAPE values are rounded to two decimal places.

** Not available since there are no time series of the Other category in the quarterly dataset.

Table 4: The sMAPE values of the BiLSTM and baseline models on the different forecasting horizons

Model	Forecasting Horizon							Average		
	1	2	3	4	5	6	8	1 to 4	1 to 6	1 to 8
Holt[a]	5	6.9	8.3	10.4	11.5	13.1	15.6	7.67	9.21	10.67
Winter[b]	5	7.1	8.3	10.2	11.4	13.2	15.3	7.65	9.21	10.61
Dampen[c]	5.1	6.8	7.7	9.1	9.7	11.3	12.8	7.18	8.29	9.33
B–J automatic[d]	5.5	7.4	8.4	9.9	10.9	12.5	14.2	7.79	9.1	10.26
Autobox1[e]	5.4	7.3	8.7	10.4	11.6	13.7	15.7	7.95	9.52	10.96
Autobox2[e]	5.7	7.5	8.1	9.6	10.4	12.1	13.4	7.73	8.89	9.9
Autobox3[e]	5.5	7.5	8.8	10.7	11.8	13.4	15.4	8.1	9.6	10.93
ARARMA[f]	5.7	7.7	8.6	9.8	10.6	12.2	13.5	7.96	9.09	10.12
Theta[g]	5	6.7	7.4	8.8	9.4	10.9	12	7	8.04	8.96
Automat ANN[h]	5.5	7.6	8.3	9.8	10.9	12.5	14.1	7.8	9.1	10.2
BiLSTM	**5.96**	**6.62**	**7.43**	**7.92**	**8.90**	**9.90**	**11.02**	**6.98**	**7.79**	**8.55**

[a] Automatic Holt's Linear Exponential Smoothing (two parameter model).

[b] Holt–Winter's linear and seasonal exponential smoothing (two or three parameter model).

[c] Dampen Trend Exponential Smoothing.

[d] Box–Jenkins methodology of 'Business Forecast System'.

[e] Robust ARIMA univariate Box–Jenkins with/without Intervention Detection.

[f] Automated Parzen's methodology with Auto regressive filter.

[g] Specific decomposition technique, projection and combination of the individual components.

[h] Automated Artificial Neural Networks for forecasting purposes.

that of the Theta model (7%) on the next four forecasts, 7.79% and 8.55% lower than 8.04% and 8.96% of the Theta on the next six and eight forecasts respectively, wherein the Theta model is the best performer on the quarterly dataset with the sMAPE accuracy measure in the M3-Competition. Moreover, although the BiLSTM model has the higher sMAPE than those of the baseline models on the first forecasting horizon, it outperforms these models on the remaining horizons from two to eight.

The forecasting performance of the BiLSTM model regarding the length of input sequence (the number of time steps) is depicted in Figure 3. The time series with ID N 1334 is selected as an example for this experiment. It is shown that the model performance varies relying on the time step setting. It achieves the lowest average sMAPE value about 6.4 at the input time steps 4 and 12 but gets the highest average sMAPE value over 50.0 at the time step length 57. Moreover, many minimum points of the plot are the multiples

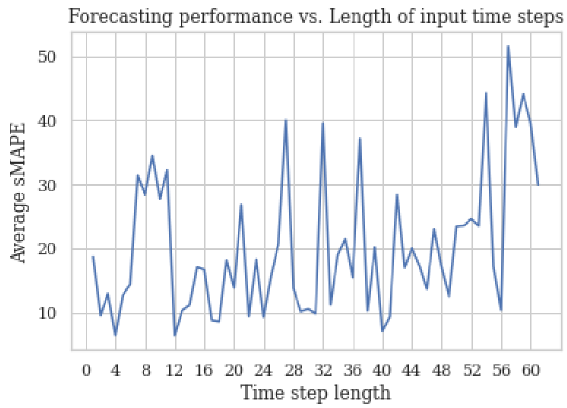

Figure 3: The BiLSTM forecasting performance regarding the length of input time steps.

of 4, which is corresponding to the quarterly period of the time series. It can be derived that the BiLSTM model is able to detect seasonality in the time series, and therefore helps reduce the overall forecast error in the forecasting problem. Besides, we observe that the optimal time step configuration is specific to each time series. Therefore, it is worth computing some descriptive statistics of the time series in order to get a better view of its features for making a reasonable setting of the model parameters.

4 CONCLUSIONS

In this paper, we present the BiLSTM model in time series forecasting. The model is evaluated based on the benchmark quarterly dataset with seasonal property. The model has the better performance on the macro and industry categories and achieves the average sMAPE error less than 9%. It is shown that the BiLSTM architecture obtains the better results than the baseline models in the M3-Competition. By tuning the time step parameter, the model performance changes depending on the time step length setting and is specific to each time series itself. In future work, ensemble learning models combined with LSTM will be used in forecasting time series data, as well as these models will be evaluated on various model parameters such as the number of time steps, the number of layers, the number of neurons, and their combinations.

REFERENCES

[1] Rumelhart, D., Hinton, G., and Williams, R. 1986. Learning representations by back-propagating errors. Nature 323, 533-536. https://doi.org/10.1038/323533a0

[2] Karpathy, A., Johnson, J., and Li, F.-F. 2015. Visualizing and understanding recurrent networks. arXiv preprint. https://arxiv.org/abs/1506.02078

[3] Sutskever, I., Vinyals, O., and Le, Q.V. 2014. Sequence to Sequence Learning with Neural Networks. Advances in Neural Information Processing Systems 27, 3104-3112.

[4] Li, X., and Wu, X. 2015. Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD, 4520-4524.

[5] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. 2015. Show and Tell: A Neural Image Caption Generator. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3156-3164.

[6] Ullah, A., Ahmad, J., Muhammad, K., Sajjad, M., and Baik, S.W. 2017. Action Recognition in Video Sequences using Deep Bi-Directional LSTM with CNN Features. IEEE Access 6, 1155-1166.

[7] Kim, T.-Y., and Cho, S.-B. 2018. Web traffic anomaly detection using C-LSTM neural networks. Expert Systems with Applications 106, 66-76. https://doi.org/10.1016/j.eswa.2018.04.004

[8] Malhotra, P., Vig, L., Shroff, G., and Agarwal, P. 2015. Long short-term memory networks for anomaly detection in time series. In: ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 89-94.

[9] Chauhan, S., and Vig, L. 2015. Anomaly detection in ECG time signals via deep long short-term memory networks. In: *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Paris, 1-7. https://doi.org/10.1109/DSAA.2015.7344872

[10] Wu, Y., Yuan, M., Dong, S., Lin, L., and Liu, Y. 2018. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. Neurocomputing 275, 167-179. https://doi.org/10.1016/j.neucom.2017.05.063

[11] Zhao, R., Wang, J., Yan, R., and Mao, K. 2016. Machine health monitoring with LSTM networks. In: *2016 10th International Conference on Sensing Technology (ICST)*, Nanjing, 1-6. https://doi.org/10.1109/ICSensT.2016.7796266

[12] Le, D., Thi, D., Lee, J., Rabczuk, T., and Nguyen-Xuan, H. 2019. Forecasting Damage Mechanics by Deep Learning. CMC-Computers, Materials & Continua 61, 3, 951-977. https://doi.org/10.32604/cmc.2019.08001

[13] Nguyen, D.Q., Phan, M.N., and Zelinka, I. 2020. Forecasting Time Series with Long Short-Term Memory Networks. Can Tho University Journal of Science 12, 2, 53-59. https://doi.org/10.22144/ctu.jen.2020.016

[14] Hochreiter, S., and Schmidhuber, J. 1997. Long shortterm memory. Neural Computation 9, 8, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[15] Gers, F.A., Schmidhuber, J., and Cummins, F. 1999. Learning to forget: continual prediction with LSTM. In: 9th International Conference on Artificial Neural Networks: ICANN '99, Edinburgh, UK, 850-855. https://doi.org/10.1049/cp:19991218

[16] Gers, F.A., Schmidhuber, J., and Cummins, F. 2000. Learning to Forget: Continual Prediction with LSTM. Neural Computation 12, 10, 2451-2471. https://doi.org/10.1162/089976600300015015

[17] Cho, K., Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 1724-1734. https://doi.org/10.3115/v1/D14-1179

[18] Makridakis, S., and Hibon, M. 2000. The M3-Competition: results, conclusions and implications. International Journal of Forecasting 16, 4, 451-476. https://doi.org/10.1016/S0169-2070(00)00057-1

[19] Makridakis, S., Spiliotis, E., and Assimakopoulos, V. 2018. The M4 Competition: Results, findings, conclusion and way forward. International Journal of Forecasting 34, 4, 802-808. https://doi.org/10.1016/j.ijforecast.2018.06.001

[20] The M3-Competition Database. The 3003 Time Series of The M3-Competition, accessed on 01 August 2020. Available from https://forecasters.org/resources/time-series-data/m3-competition/

[21] Kingma, D.P., and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA.