Prediction Study Based on TCN-BiLSTM-SA Time Series Model

He Zhang[1,*], Peng Chu[2]

[1]School of Electronic Information, Xijing University, Xi'an, China
[2]School of Electronic Information, Xijing University, Xi'an, China

*Corresponding author:Z18437379496@163.com

Abstract. To enhance the accuracy of time series prediction, this study proposes a hybrid network model called TCN-BiLSTM-SA, which combines Temporal Convolutional Network (TCN), Bidirectional Long Short-term Memory (BiLSTM), and Self Attention (SA). The TCN is employed to learn sequence features, while the BiLSTM model captures preceding and succeeding states to extract more information for prediction. The self-attention mechanism calculates weights for each time step's output, effectively utilizing the cell memory information of BiLSTM to capture global features and improve prediction accuracy. Experimental results on the Beijing PM 2.5 dataset demonstrate that the TCN-BiLSTM-SA network outperforms the BiLSTM model in terms of RMSE, MAE, and MAPE error rates, while also exhibiting greater stability. This model holds promising potential for various time series prediction applications. It has broad application prospects in time series prediction.

Keywords: Time series prediction; TCN; BiLSTM; Self-Attention

1 Introduction

The Bidirectional Long Short-Term Memory (BiLSTM) is a recurrent neural network with a bidirectional structure. By incorporating both past and future information in the network, it enables a comprehensive exploration of temporal features in the data, leading to significant advancements in time series prediction. However, the traditional BiLSTM model still faces challenges with information loss and a decrease in accuracy when dealing with longer sequences.

To address the information loss issue in the BiLSTM model, several approaches have been proposed in recent studies. Reference [1] proposes a Attention-BiLSTM model that incorporates attention mechanisms to selectively emphasize important features in the input data, leading to improved performance in short-term result prediction. Reference [2] proposes a CNN-LSTM-BiLSTM model with an attention mechanism for load forecasting in the energy system. This model combines CNN and attention mechanisms to extract effective local features, while LSTM and BiLSTM capture the temporal features of the load data. Reference [3] proposes a CNN-BiLSTM-AM model that utilizes

R. Appleby et al. (eds.), *Proceedings of the 2nd International Conference on Intelligent Design and Innovative Technology (ICIDIT 2023)*, Atlantis Highlights in Intelligent Systems 10,
https://doi.org/10.2991/978-94-6463-266-8_21

a CNN to extract relevant features and incorporates a BiLSTM with an attention mechanism for accurate outcome prediction. Reference [4] proposes an AC-BiLSTM model that employs convolutional layers to extract features and incorporates attention layers to selectively utilize information from the BiLSTM model. This approach significantly improves predictive accuracy. Reference [5] proposes the TCN-LSTM model, which combines TCN for spatial feature extraction from multivariate data and LSTM for capturing long-term correlations in temporal data. This model demonstrates superior accuracy compared to the CNN-LSTM model. Reference [6] proposes the DSA-TCN-BiGRU model, which utilizes a TCN with an attention mechanism to extract temporal features. It then employs a BiGRU model with the DSA mechanism to quantify the impact of each node on prediction results, achieving optimal performance and accuracy through Bayesian optimization of hyperparameters.

This paper proposes a prediction method based on the TCN-BiLSTM-SA time series model to improve the learning of relationships between features and enhance prediction accuracy. The method involves using TCN to learn sequence features in the time dimension, BiLSTM to capture preceding and succeeding states, and self-attention mechanism to prioritize important feature information for accurate predictions.

2 TCN-BiLSTM-SA Time Series Prediction Model

2.1 TCN

TCN, introduced by Lea et al [7], is a network architecture derived from convolutional neural networks, specifically designed for the analysis and processing of time series data. It introduces two novel operation modes, namely causal convolution and dilated convolution, to overcome the limitations of traditional one-dimensional convolution. Causal convolution enables TCN to process time series data of variable length. Dilated convolution enables TCN to effectively capture long-term time dependencies. In addition, by incorporating residual connections between network layers, TCN addresses the issue of gradient vanishing or explosion [8].

2.2 BiLSTM

BiLSTM [9] is a bidirectional recurrent structure propagation. Unlike traditional LSTM networks, BiLSTM incorporates a reverse data flow from future to past, enabling better exploration of temporal data. The hidden layers used in the past and future are independent, facilitating the analysis of temporal characteristics.

2.3 Self-attention

Self-attention [10] is a specific case of the attention mechanism where the Query, Key, and Value matrices are the same input. Self-attention not only captures global feature information of the data but also captures feature information among the same group of data vectors, facilitating the capture of important data.

2.4 Model Building

The overall structure of TCN-BiLSTM-SA proposed in this paper is shown in Fig.1.

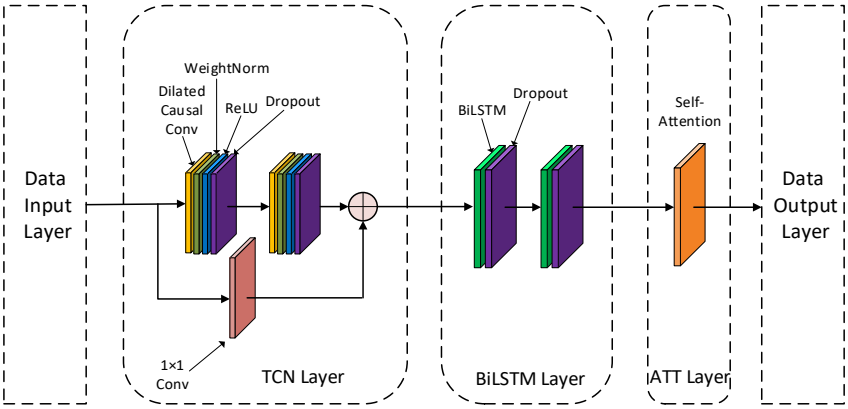

Fig. 1. Structure Diagram of TCN-BiLSTM-SA Model

In this study, we initially utilized the "series_to_supervised" function to transform the raw data into a supervised learning problem, followed by feature normalization and encoding. We then constructed a TCN-BiLSTM-SA model. First, the model extracts features through a TCN layer with 64 filters, 8 convolution sizes, and dilation rates of [1, 2, 4, 8]. The activation function is 'ReLU' and returns sequence data. Next, we used a bidirectional LSTM layer to capture time series features. This layer has 128 units and returns sequence data. We then introduced a self-attention mechanism to calculate attention weights using a sigmoid activation function so that the model can focus on more important information. Finally, we further extracted and integrated features through several fully connected layers, including a fully connected layer with 64 neurons and a 'ReLU' activation function, a Dropout layer with a probability of 0.3, and an output layer with a single neuron to generate the final output of the model. During model training, we adopted a learning rate scheduler to monitor the model training process. We trained the model for 120 epochs with a batch size of 128 and a time step of 8.

The proposed model employs TCN and bidirectional LSTM layers to effectively harness historical information and extract salient features and hidden states from time series data. The integration of a self-attention layer allows the model to selectively focus on pivotal information via attention distributions, thereby enhancing its learning capability of key features. This approach results in superior performance of our model in handling complex time series data.

3 Experimental Simulation

This study evaluated multiple models using the Beijing PM2.5 dataset. The dataset consists of 43,800 rows and 8 columns, representing hourly PM2.5 measurements in Beijing from 2010 to 2014. The PM2.5 was used as the target variable for prediction, and

the dataset was split into training and testing sets in a 7:3 ratio. Experiment employed evaluation metrics such as root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) to assess the performance. When the model's iteration count was set to 120, it underwent 20 evaluations, with the results depicted in Fig.3. Moreover, at iteration counts of 30, 60, 90, and 120, the model was evaluated 5 times each, and the average evaluation results were illustrated in Fig.4.

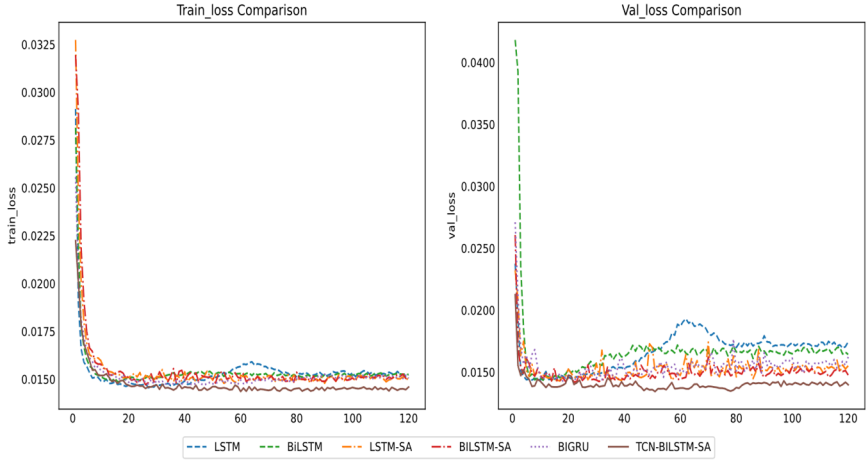

Fig. 2. Loss Comparison Graph

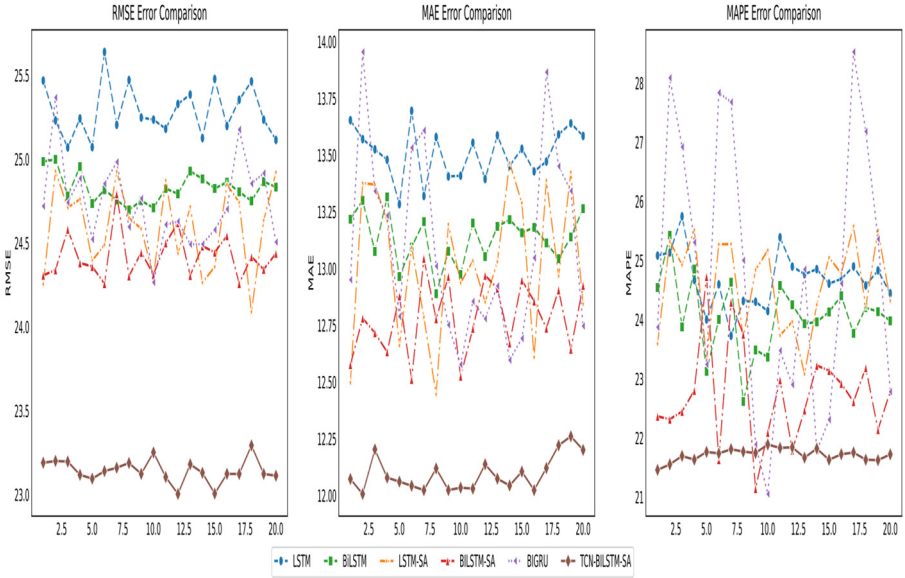

Fig. 3. Error Comparison Graph

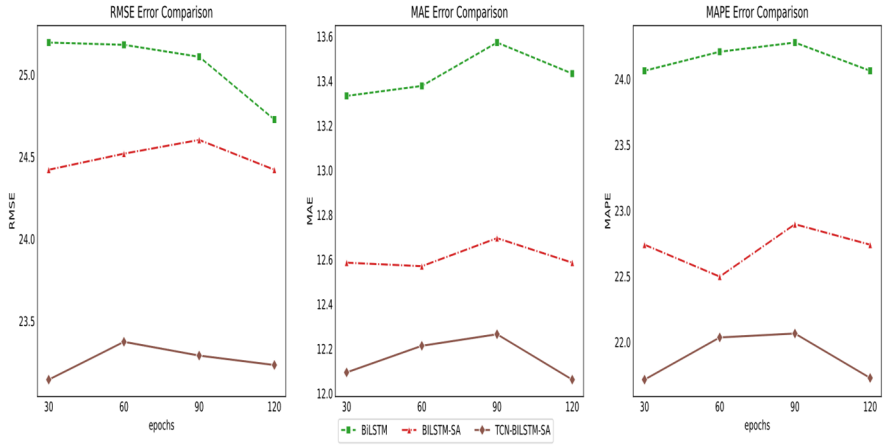

Fig. 4. Average Error Comparison Graph

As illustrated in Fig.2 to Fig.4, the BiLSTM model, due to the issue of information loss, exhibits low precision and instability in both loss values and evaluation outcomes. The incorporation of a self-attention mechanism enhances accuracy to a certain degree, yet the model remains unstable with significant discrepancies in evaluation results. To further enhance the precision and stability of the model, this study introduces a TCN network on top of the BiLSTM-SA model. Experimental evidence suggests that the TCN-BiLSTM-SA model surpasses the BiLSTM model in terms of loss values, evaluation results, and stability.

4 Conclusions

In summary, this study introduces a TCN-BiLSTM-SA time series prediction model to improve the accuracy and stability of BiLSTM in time series prediction. The model utilizes the TCN network to capture sequence features, the BiLSTM network to capture long-term dependencies, and the self-attention mechanism to prioritize salient feature information. Experimental results show that the proposed model outperforms traditional BiLSTM and holds promising potential for applications in time series prediction tasks.

References

1. Du J, Cheng Y, Zhou Q, et al. Power load forecasting using BiLSTM-attention[C]//IOP Conference Series: Earth and Environmental Science. IOP Publishing, 2020, 440(3): 032115. https://doi.org/10.1088/1755-1315/440/3/032115
2. Wu K, Wu J, Feng L, et al. An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system[J]. International Transactions on Electrical Energy Systems, 2021, 31(1): e12637. https://doi.org/10.1002/2050-7038.12637

3. Lu W, Li J, Wang J, et al. A CNN-BiLSTM-AM method for stock price prediction[J]. Neural Computing and Applications, 2021, 33: 4741-4753. https://doi.org/10.1007/s00521-020-05532-z
4. Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification[J]. Neurocomputing, 2019, 337: 325-338. https://doi.org/10.1016/j.neucom.2019.01.078
5. Hu C, Cheng F, Ma L, et al. State of charge estimation for lithium-ion batteries based on TCN-LSTM neural networks[J]. Journal of the Electrochemical Society, 2022, 169(3): 030544. https://doi.org/10.1149/1945-7111/ac5cf2
6. Huang J, Ding W. Aircraft Trajectory Prediction Based on Bayesian Optimised Temporal Convolutional Network–Bidirectional Gated Recurrent Unit Hybrid Neural Network[J]. International Journal of Aerospace Engineering, 2022, 2022. https://doi.org/10.1155/2022/2086904
7. Lea C, Vidal R, Reiter A, et al. Temporal convolutional networks: A unified approach to action segmentation[C]//Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14. Springer International Publishing, 2016: 47-54. https://doi.org/10.1007/978-3-319-49409-8_7
8. Zhang R, Sun F, Song Z, et al. Short-term traffic flow forecasting model based on GA-TCN[J]. Journal of Advanced Transportation, 2021, 2021: 1-13. https://doi.org/10.1155/2021/1338607
9. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural networks, 2005, 18(5-6): 602-610. https://doi.org/10.1016/j.neunet.2005.06.042
10. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30. https://doi.org/10.48550/arXiv.1706.03762