2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS)

Time Series Prediction Based on Time Attention Mechanism and LSTM Neural Network

1st Jiayi Sun
College of Computer
Beijing University of Posts and Telecommunications
Beijing, China
dabaosjy@bupt.edu.cn

2nd Wenming Guo
College of Computer
Beijing University of Posts and Telecommunications
Beijing, China
guowenming@bupt.edu.cn

***Abstract*—As a collection of time observations, time series has attracted extensive attention in artificial intelligence. Time series prediction is one of the important topics to obtain future trends. Therefore, based on the discussion of time series characteristics, temporal attention mechanism and deep learning time series prediction, this paper briefly discusses the open data set, experimental environment and parameter settings, and designs an improved time series PA-LSTM prediction model based on deep learning. Finally, through specific experimental analysis. The results show that the RMSLE and MAE values of the PA-LSTM prediction method designed in this paper are 0.012 and 0.010 respectively. The error is lower than other prediction methods. Therefore, the PA-LSTM prediction method has certain advantages.**

Keywords*—*deep learning, time series, temporal attention mechanism, neural network.

I. INTRODUCTION

In the context of big data society, the development of information technology promotes data deep mining, and it is possible to explore the interaction of data. The rational use of data can help people's decision-making behavior. The method of time series has always been the focus of researchers.

Nowadays, more and more scholars have done a lot of research on time series prediction of deep learning through various technologies and system tools, and have also made some research achievements through practical research. Prabowo H proposed a general, fast and memory efficient time series prediction model. This model makes use of the time characteristics between the sequence and the structure of the network to predict the sequence. It also uses convolution digit elimination and attention mechanisms to predict future developments. The data show that the mean square error of the method is reduced by 34% and the convergence speed is reduced by 33% [1]. Nguyen T N proposed a long interval two-dimensional method to increase the prediction time of the population with high accuracy. The method consists of two parts: first, long time interval prediction, which extends the prediction length to 260-time steps; Second: input method, which uses the knowledge of different time period dimensions to improve the accuracy of long-term interval prediction. Through experiments, it is found that the proposed prediction model is suitable for inter series prediction, which proves the generalization of the method [2]. All the deep learning models used by Safat W have achieved an accuracy rate of more than 90%, and the accuracy rate of LSTM and GRU is 93%. In addition, the deep learning model always has good performance on single year data. The data shows that the accuracy of LSTM and GRU is 90% and 88% respectively. Data shows that LSTM and GRU have high precision and recall rate compared with CNN. The deep neural network method will bring better performance in the research of economic distress prediction [3]. Although the existing research on time series prediction of deep learning is very rich, there are still many problems in its practical application.

In order to predict the download speed of network traffic, this paper designed a time series prediction method based on CEEMDAN, attention mechanism, LSTM neural network and PSO optimization algorithm, and named it PA-LSTM. In this method, CEEMDAN preprocesses data so that the data becomes smooth and the variation trend becomes more obvious. The attention mechanism measures the importance of the features extracted by LSTM to the prediction results. In addition, in order to find a more appropriate attention weight, the PSO algorithm with stronger global search ability is used to adjust the attention weight, so as to make up for the SGD algorithm's deficiency in global search ability. In order to test the capability of PA-LSTM in network traffic prediction of open Abilene dataset, three other time series prediction models and the improved time series model PA-LSTM designed in this paper were selected for experiments. The data show that the PA-LSTM method is superior to the comparison method in the evaluation indexes of time series prediction and can accurately predict the change of download speed in the future.

II. RESEARCH ON TIME SERIES PREDICTION DESIGN BASED ON TEMPORAL ATTENTION MECHANISM AND LSTM NEURAL NETWORK

A. *Time Series Characteristics*

(1) Different data points in the time series will be affected by changes brought about by time, and different factors brought about by different data point values will have the same effect at different times, bringing about comprehensive results [4].

(2) From a local perspective, data points on different time nodes will change randomly over time, but there will be some correlation between these data points because of these changes, and this correlation can just give the internal operating mechanism of the predicted project [5].

(3) To sum up, the change rules of different data points in the time series usually have certain regularity, such as long-term and integration [6]. Long term is mainly superior to the long-term impact of external environment on the project; Integration means that the change of sequence is the result of changes brought by different factors [7].

(4) The past and current data points in the time series may represent the direction of future events, and generally do not produce jumping changes, but will move forward at a small pace [8].

B. *Temporal Attention Mechanism*

At any time g, the decoder will use Temporal Attention to calculate a context vector S_g. The calculation process of S_g is shown in Equation (1), Equation (2) and Equation (3) respectively:

$$k_g^r = t_p^G \tanh(Q_p^{'}\left[f_{g-1}^{'};d_{g-1}^{'}\right] + W_p^{'} f_r + c_p^{'}) \tag{1}$$

$$\varphi_g^r = \frac{\exp(k_g^r)}{\sum_{u=1}^{G-1}\exp(k_g^r)} \tag{2}$$

$$S_g = \sum_{r=1}^{G-1} \varphi_g^r f_r \tag{3}$$

In formula (1), t_p, $Q_p^{'}$, $W_p^{'}$, $c_p^{'}$ are learning parameters, $c_p^{'}$ is bias term, $\left[f_{g-1}^{'};d_{g-1}^{'}\right]$ is the current state of decoder, and f_r is the state of encoder hidden unit at time r [9]. $\left[f_{g-1}^{'};d_{g-1}^{'}\right]$ and f_r are related by linear addition, so k_g^r represents the correlation measure between f_r and the current state. In Formula (2), k_g^r is converted into φ_g^r through Softmax function, and φ_g^r is the weight assigned to f_r; As shown in Formula (3), the weight corresponding to the encoder at time 1 to $G-1$ is weighted and summed to finally obtain the context vector S_g of the decoder at time g.

C. *Time Series Prediction of Deep Learning*

(1) The steps of deep learning process are as follows: first, learn about the data information according to the research background, so as to understand the data characteristics. The second is to obtain data sets and preprocess the data, which includes missing data processing, abnormal data standardization processing and other operations. Compared with traditional machine learning, deep learning has strong adaptability and can independently learn complex features. The process of in-depth learning includes network structure selection, model training and model evaluation [10].

(2) The time series forecasting problem can be viewed as a supervised forecasting model from a series of numbers to a series. The sequence-to-sequence model uses the characteristics of a system sequence, such as seasonality and trend in time, to realize the model, thus avoiding the traditional method of manually establishing features. Instead, the new sequence features obtained through transformation are a modeling framework suitable for modern artificial neural networks [11].

D. *Feature Extraction Based on LSTM*

LSTM is a special form of RNN. Unlike RNN, LSTM introduces a gate structure to selectively let information pass.

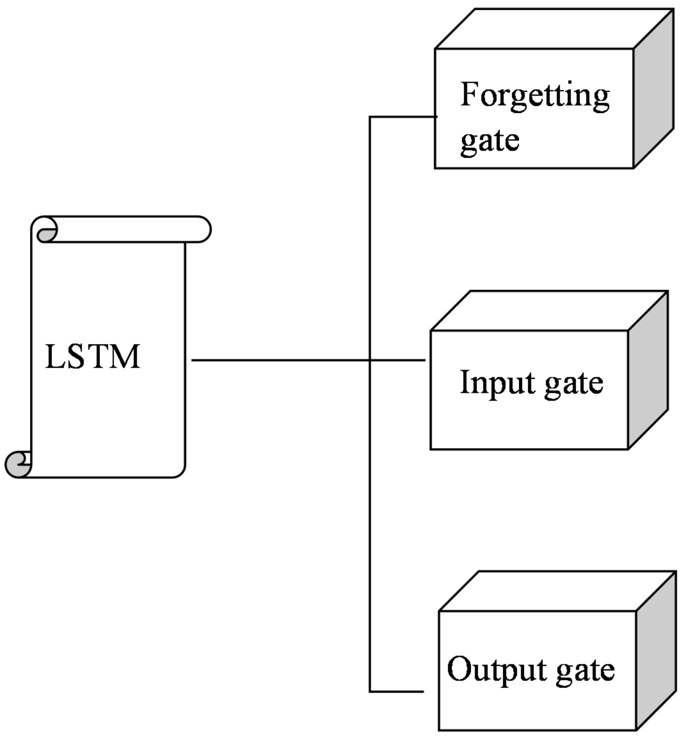

Fig. 1. LSTM structure

The gate structure achieves selective passing through sigmoid function and dot multiplication. There are three kinds of gates in LSTM: 1. Forgetting gate; 2. Input gate; 3. Output gate, as shown in Figure 1.

III. Research on Time Series Prediction Based on Temporal Attention Mechanism and LSTM Neural Network

A. Public Datasets

The public data set of Abilene is adopted in this experiment. Abilene is the American Research and Education Network, a wide area backbone network with a real topology. The real topology and historical traffic data of 12 major cities in the United States are adopted, including 12 network nodes, 144 OD flow pairs in total. The data collection time is from 0:00 on August 1, 2004 to 0:00 on August 23, 2004. The flow bandwidth values from 12 source locations to 12 destinations are collected every five minutes. There are 6336 flow matrices, 912384 data in total. The Abilene network consists of 12 peer nodes and 30 undirected links, this data set is one of the popular data sets in the field of network traffic engineering. For the prediction of the trend of time series changes, we adopt the division method of training set and test set: based on time characteristics, use the data of the first two weeks for training model parameters, and use the data of the last seven-star periods to verify the prediction performance of the model.

B. Forecast Environment and Parameter Analysis

(1) Hardware environment Surface notebook: Intel I7-8650U processor, DDR4 12G memory, NVIDIAGTX1050 video memory.

(2) The software environment operating system adopts Windows 10, 64 bit. The programming software adopts Anaconda3 and Python 3.6. Use deep learning Tesorflow-GPU1.11.0 framework to program. In order to support Tesorflow-GPU1.11.0, CudaTookit9.0 and cuDnn9.0 are also installed in this experiment.

(3) Parameter setting: The training set of PA-LSTM spatio-temporal fusion network prediction model is constructed through the open data set. The parameter settings include: the number of input and output nodes is 3 and 2, the learning step is 45, and the number of hidden layer nodes is 110.

IV. Time Series Prediction Application Research Based on Temporal Attention Mechanism and LSTM Neural Network

A. PA-LSTM Prediction Model Based on Deep Learning and Improved Time Series

In order to predict the download speed of network traffic, we need to improve the time series prediction model of deep learning. It mainly solves the following three problems: 1. Data preprocessing; 2. Build prediction model; 3. Optimization of model parameters. In this paper, the content of building prediction model and optimizing model parameters is named PA-LSTM method. The overall framework of prediction based on PA-LSTM is shown in Figure 2.

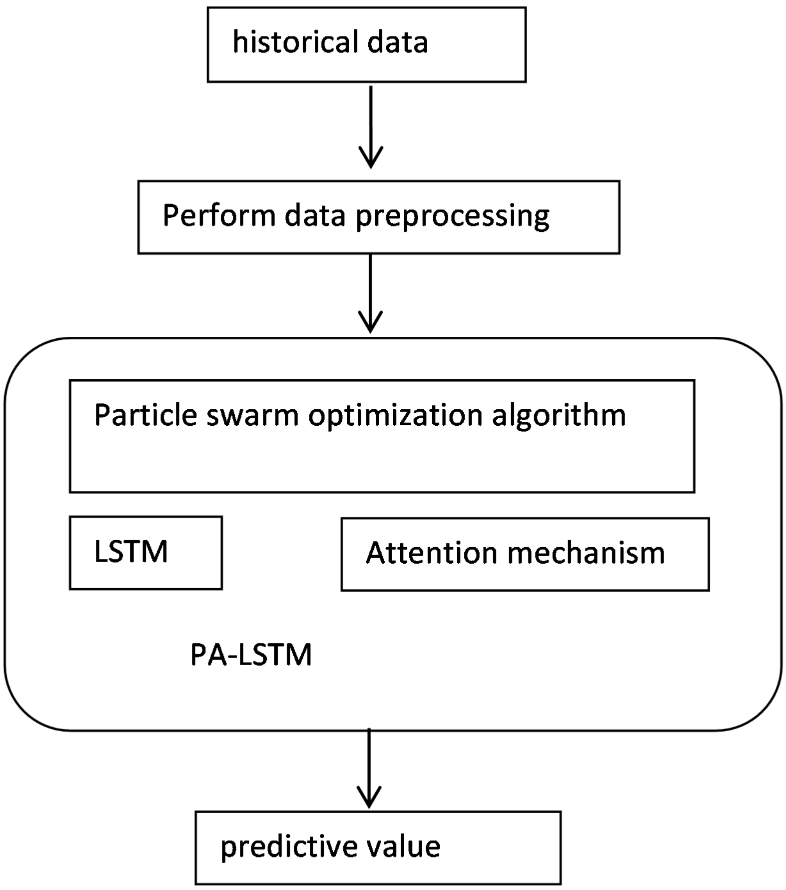

Fig. 2. Performance Comparison of PA-LSTM Model and Other Models on Abilene Dataset

(1) The network traffic data obtained from Abilene dataset constitutes the original sequence data. Then, in order to remove the high frequency components in the original sequence data, so that the data becomes

smoother and easier to predict, CEEMDAN is introduced to preprocess the original sequence data. Complete Empirical Mode Decomposition (CEEMDAN) CEEMDAN effectively solves the problem of insufficient classification of EEMD method by adding white noise with opposite phase to the original data sample, then processing it, and obtaining the classification result by calculating the average value.

(2) A time series prediction model combining LSTM and attention mechanism was constructed. In addition, the PSO algorithm is proposed to optimize the attention weight, and a PA-LSTM time series prediction model is constructed. The particle swarm optimization algorithm was used to optimize the attention weight, the size of the sliding window (input data dimension) and the number of hidden layer neurons. Suppose that the total number of parameters needs to be optimized, then the solution space of PSO is dimensional, each particle will be randomly initialized as a point in the solution space, and then the optimal solution will be found through iteration. During each iteration, each particle updates itself by tracking two extremes: one is the optimal solution of the particle itself; The other extreme is the optimal solution for the whole population, the global optimal solution.

For the model structure, since the number of input neurons and output neurons of the network have been determined by the problem itself, the selection of the depth neural network structure is equivalent to the selection of the number of hidden layer nodes. At present, there is no consensus on the selection of the number of neurons in the hidden layer. Therefore, this paper compares and analyzes the models with different number of neurons in the hidden layer and selects the optimal parameters. Set the number of neurons in the hidden layer as 16, 32, 64, 128, 256, corresponding to T1, T2, T3, T4, T5, and then train the model. The experimental results are shown in Table I.

TABLE I. INFLUENCE OF MODEL STRUCTURE ON MODEL ACCURACY

Model	Estimate rate	Actual accuracy
T1	36.6	53.2
T2	61.7	52.7
T3	62.1	55.3
T4	49.7	50.1
T5	68.2	64.1

It can be seen from Table I that reducing the number of neurons in the hidden layer reduces the accuracy of the model, while increasing the number of neurons in the hidden layer does not significantly improve the accuracy of the model. The specific process of parameter optimization using PSO algorithm in the PA-LSTM method is as follows: Step 1: Calculate the dimension of solution space. Calculate the solution space dimension D according to the number of parameters to be optimized. Assuming the length of attention weight is Wn, the number of hidden layer neurons is Hn, and the size of sliding window is the parameter to be optimized, step 2: initialize particle swarm. Step 3: Set optimization objectives. In order to obtain the optimal parameters of the model established by this method. The goal of particle swarm optimization is set to be the same as the goal of model training, that is, the minimum loss value. Step 4: Update the particle's position and speed. During the whole process, the optimal solution of each particle and the global optimal solution are updated in real time until the stop condition is reached (the loss value of the model is no longer reduced, or the specified number of iterations is reached).

(3) Finally, the pre-processed data is used to train the prediction model, and the trained model is used to predict the download speed of network traffic. PA-LSTM prediction model uses LSTM neural network as the core structure. In order to overcome the distraction problem of LSTM neural network when extracting sliding window features. The attention layer is also introduced to build the prediction model together with LSTM.

B. *Time Series Prediction Application Based on Deep Learning*

For traffic time series data, this paper uses the public Abilene dataset to conduct experimental analysis on the PA-LSTM model. In order to verify the effectiveness and robustness of the models proposed in this paper, based on the model evaluation indicators of RMSLE and MAE values, this paper compares and evaluates the performance of these time series prediction models. In the experiment, the first 80% of the data is used as the training set, and the last 20% is used as the test set. Finally, the RMSLE (root mean square error) and MAE (average absolute error) values of different prediction models can be obtained. The specific experimental results are shown in Table II.

TABLE II. PERFORMANCE COMPARISON BETWEEN PA-LSTM MODEL AND OTHER MODELS ON ABILENE DATASET

Model	RMSLE	MAE
LSTM	0.038	0.018
PA-LSTM	0.012	0.010
ConvLSTM	0.021	0.013
Attention+LSTM	0.036	0.025

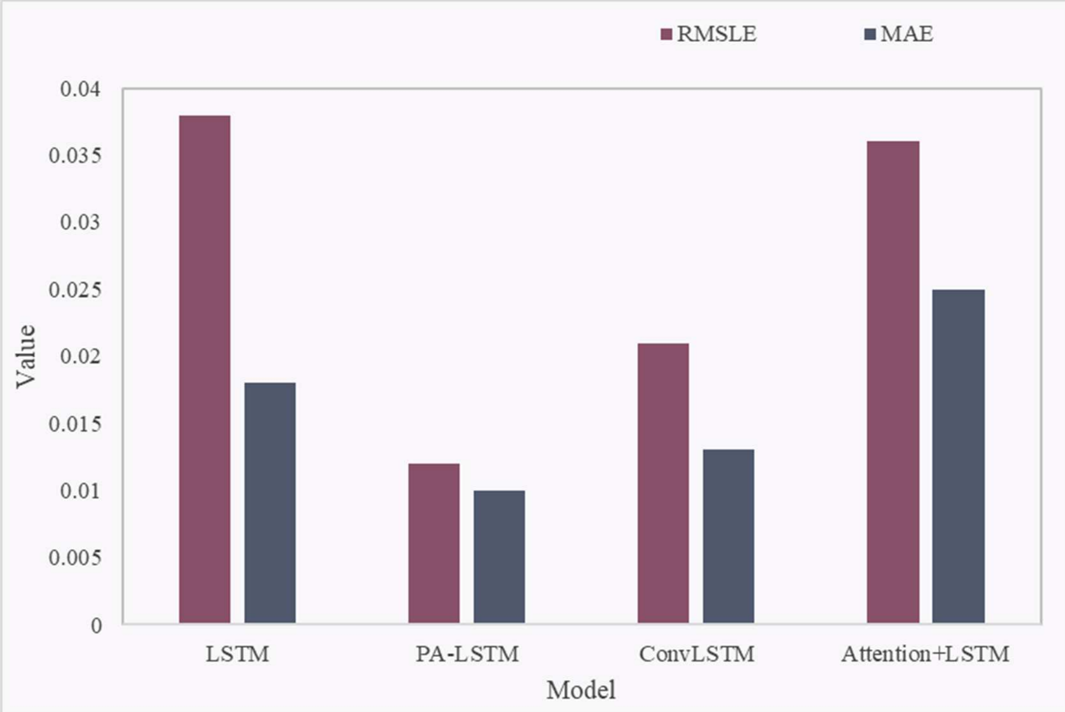

Fig. 3. Performance Comparison of PA-LSTM Model and Other Models on Abilene Dataset

It can be seen from the data in Figure 3 that the predicted RMSLE and MAE values of PA-LSTM model are 0.012 and 0.01 respectively. The RMSLE and MAE values of LSTM model are 0.038 and 0.018 respectively. The error of PA-LSTM model is significantly reduced by 0.026 and 0.08 compared with LSTM. The RMSLE and MAE values of ConvLSTM model are 0.021 and 0.013 respectively. The error of PA-LSTM model is significantly reduced by 0.009 and 0.003 than that of ConvLSTM. The RMSLE and MAE values of Attention+LSTM model are 0.036 and 0.025 respectively. The error of PA-LSTM model is obviously 0.024 and 0.015 less than Attention+LSTM model. To sum up, we can see the advantages of PA-LSTM model in dealing with the problem of network traffic download speed prediction. Therefore, the model is applicable to the download speed of network traffic.

V. CONCLUSION

In order to obtain the global optimal attention weight, PSO algorithm is introduced in the time series prediction of depth learning in this paper. Inspired by the predation of birds, PSO algorithm uses swarm intelligence to constantly search in the solution space to obtain the optimal solution, which has the advantages of high accuracy and fast convergence. In the method, PSO algorithm is used to adaptively search for the globally optimal attention weight and some parameters of the neural network. The attention mechanism and particle swarm optimization algorithm are introduced to improve the traditional LSTM neural network, and a time series prediction method named PA-LSTM is proposed. The network structure of PA-LSTM is composed of LSTM and attention layer. Then, PSO is used to optimize the parameters: attention weight, hidden layer node number and window size, so as to obtain the final prediction model.

REFERENCES

[1] Prabowo H, Hidayat A A, Cenggoro T W, et al. Aggregating Time Series and Tabular Data in Deep Learning Model for University Students' GPA Prediction. IEEE Access, 2021, PP(99):1-1.

[2] Nguyen T N, Lee S, Nguyen-Xuan H, et al. A novel analysis-prediction approach for geometrically nonlinear problems using group method of data handling. Computer Methods in Applied Mechanics and Engineering, 2019, 354(SEP.1):506-526.

[3] Jagannathan, Preetha, Sasikumar Gurumoorthy, Andrzej Stateczny, Parameshachari Bidare Divakarachar, and Jewel Sengupta. "Collision-aware routing using multi-objective seagull optimization algorithm for WSN-based IoT." *Sensors* 21, no. 24 (2021): 8496.

[4] Harper R, Southern J. A Bayesian Deep Learning Framework for End-To-End Prediction of Emotion from Heartbeat. IEEE Transactions on Affective Computing, 2020, PP(99):1-1.

[5] Moews B, Herrmann J M, Ibikunle G. Lagged correlation-based deep learning for directional trend change prediction in financial time series. Expert Systems with Applications, 2018, 120(APR.):197-206.

[6] Pramanik P, Sinhababu N, Kwak K S, et al. Deep Learning Based Resource Availability Prediction for Local Mobile Crowd Computing. IEEE Access, 2021, PP(99):1-1.

[7] Al-Tahmeesschi A, Umebayashi K, Iwata H, et al. Feature-Based Deep Neural Networks for Short-Term Prediction of WiFi Channel Occupancy Rate. IEEE Access, 2021, PP(99):1-1.

[8] Ardeshiri R R, Ma C. Multivariate gated recurrent unit for battery remaining useful life prediction: A deep learning approach. International Journal of Energy Research, 2021, 45(11):16633-16648.

[9] Gomez-Gonzalez S, Prokudin S, Scholkopf B, et al. Real Time Trajectory Prediction Using Deep Conditional Generative Models. IEEE Robotics and Automation Letters, 2020, 5(2):970-976.

[10] Wilson, Shyju, and C. Krishna Mohan. "Coherent and noncoherent dictionaries for action recognition." IEEE Signal Processing Letters 24.5 (2017): 698-702.

[11] Srivastava P R, Zhang J, Eachempati P. Deep Neural Network and Time Series Approach for Finance Systems: Predicting the Movement of the Indian Stock Market. Journal of Organizational and End User Computing, 2021, 33(5):204-226.