The 14th International Conference on
Computer Science & Education (ICCSE 2019)
August 19-21, 2019. Toronto, Canada

TuesA1.4

Prediction of short-term available parking space using LSTM model

1st Li Xiangdong
School of Information and Electronic Engineering
Zhejiang University of Science & Technology
Hangzhou, China
23809678002@qq.com

2nd Cen Yuefeng
School of Information and Electronic Engineering
Zhejiang University of Science & Technology
Hangzhou, China
cenyuefeng@126.com

3rd CEN Gang
School of Information and Electronic Engineering
Zhejiang University of Science & Technology
Hangzhou, China
gcen@163.com

4th Xu Zengwei
School of Information and Electronic Engineering
Zhejiang University of Science & Technology
Hangzhou, China
dk1092546893@163.com

***Abstract*—Real-time and accuracy prediction of available parking space plays an important role in intelligent urban traffic, which guides the driver to find a parking space efficiently. A novel prediction method of the urban short-term available parking space is proposed in this paper. Firstly, considering the time series characteristics of parking data, the Long Short-Term Memory (LSTM) model in introduced to predict the available parking space. Secondly, the genetic algorithm is introduced to adaptively adjust the parameters of LSTM, which saves lots of time for the training of this model. Experimental results show that the proposed method achieves a significant prediction accuracy comparing with back propagation (BP) neural network, while still maintaining high adaptability.**

Keywords—LSTM; parking space prediction; genetic algorithm

I. INTRODUCTION

Parking garages induction is one of the important means to solve the problem of urban parking. Information around the destination is collected such as the available parking space, road conditions, and routes to the destination. Then the information which helps the drivers reach the available parking space quickly is forwarded to the drivers. The traffic congestion and environmental pollution can also be relieved. Drivers are not only concerned about the real-time parking conditions of the parking lot, but also pay more attention to the available parking conditions when they arrive at the parking lot. Therefore, it is particularly important to make a quick and accurate prediction of urban parking space.

Nowadays, researches about the prediction of the parking space are mainly consisted of methods of influence factor analysis and methods of time series prediction. For the former methods, the influencing factors are calculated quantitatively after the analysis of the influencing factors. Since there are lots of influencing factors and the quantitative process contains many subjective factors, it is difficult to establish the model, there are few researches adopting these methods. The latter methods analyze the characteristics of the parking space data from the perspective of time series, which are widely adopted by the researchers.

Methods of the time series prediction are mainly consisted of traditional time series prediction methods and methods combined with heuristic algorithm. The traditional time series prediction method mainly include autoregressive moving average model (ARIMA) and chaotic time series prediction method [1, 2]. Methods of ensemble heuristic algorithm combined the traditional time series prediction method with some heuristic algorithm. Literature [3] proposed a combined prediction model of wavelet neural network based on wavelet transform and particle swarm optimization, and the experiment revealed a high accuracy. Reference [4] used the neural network model to predict the number of APSs. It was found that the structure of the neural network could be optimized by changing the number of neurons and learning rate, so as to effectively improve the prediction accuracy of the neural network model. Literature [5] made a comparative evaluation of the existing parking occupancy prediction methods. The ARIMA model, Kalman filtering model and BP neural network were used to predict parking occupancy in different areas. And the prediction performance of the three prediction methods was evaluated. The model of the above method is mainly based on BP neural network, which has been widely applied and achieved good results in short-term berth prediction [6-9].However, results adopting BP neural network method is prone to be the regional optimum. The selection of parameters lacks of theoretical support. Meanwhile, BP neural network can’t reflect the timing of time series.

Long Short-Term Memory (LSTM) is a deep Neural Network model originated from the Recurrent Neural Network (RNN). RNN abandons the full connection mode of the ordinary fully connected neural network in the hidden layer, introduces the concept of timing and adopts the mechanism of ‘recursive connection’ [10]. In this way, the time sequence characteristics of the sequence are preserved, and the previous input information is reserved in the network. There are enough hidden layer neurons in RNN therefor it can fit the predicted time series with any accuracy. However, preserving the temporal characteristics of sequences also faces a fatal challenge, with the growth of time series, the phenomenon of gradient disappearance and gradient explosion inevitably occur. Therefore, Hochreiter and Schmidhuber proposed the LSTM model, which not only effectively avoided problems such as gradient disappearance, but also improved the ability to process things with long intervals or delays in time series [11]. This paper proposes a short-term prediction algorithm for effective parking Spaces based on LSTM neural network.

II. PREDICTION OF AVAILABLE PARKING SPACES BASED ON LSTM

A. Introduction of LSTM Neural Network

Compared with RNN, a memory unit is added to its structure which is called memory cell. There are three gates with different functions in a cell, namely input gate (i_t), forgetting gate (f_t) and output gate (O_t). Input gate is used to determine the quantity of new information is added to the cell state before a new memory is created. The forgetting gate decides the utilization of the current memory according to the past memory unit. The memory unit is used to record all historical information up to the current time; the output gate is to separate the final memory from the hidden layer state. LSTM reserves historical memory and can process time series with long-term relevance. By adding these gates, the network can selectively reserve and forget historical information, thus avoiding gradient disappearance and explosion.

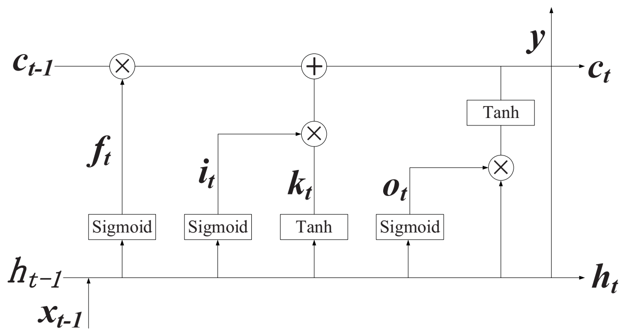

Fig.1. Illustration of LSTM model

The specific LSTM model is shown in Fig.1. There are two time-passing state chains *h* (hidden layer state) and C (cell state) in LSTM, while the relative RNN only contains the state chain *c* which transfer over time. h_{t-1} represents the values transfer to the current moment by the hidden layer at the previous moment; x_t represents the current input value. c_{t-1} represents the state value of the LSTM memory unit at the last moment; c_t represents the LSTM memory unit state value at the current moment. The information which is discarded when passes through the forgotten gate is calculated according to ht-1 and x_t.

The result of the calculation would be 0 or 1. Where 0 means that all information is forgotten, and 1 means that all information is reserved. As shown in (1).

$$f_t = \sigma(w_f \bullet [h_t - 1, x_t] + b_f) \tag{1}$$

Where *w* and *b* are the weight matrix and the bias vector in the f_t; and σ represents the activation function Sigmoid.

There are two processes for updating new information into the cell. First, the information that needs to be updated is calculated through the input gate (formula 2), and then a new value k_t is created by the *tanh* layer and added into the cell state. As shown in (3).

$$i_t = \sigma(w_i \bullet [h_{t-1}, x_t] + b_i) \tag{2}$$

$$k_t = \tanh(w_k \bullet [h_{t-1}, x_t] + b_i) \tag{3}$$

The results obtained from formula (2) and formula (3) are multiplied and added with the result obtained from the unit state value at the last time through the forgetfulness gate to obtain the unit state value at the current time, as shown in (4).

$$C_t = f_t * C_{t-1} + i_t * k_t \tag{4}$$

The final output depends on the cell state. Firstly, Sigmoid is used to classify the output results and select the data to be output. Then *tanh* function is used to process the unit state, and the state value h_t transferred by the hidden layer to the next moment was obtained. Finally, the pre-output value y of the current moment was obtained after the processing of h_t by Sigmoid function, as shown in (5) ~ (7).

$$O_t = \sigma(w_o \bullet [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = O_t * \tanh(C_t) \tag{6}$$

$$y = \sigma(w' h_t) \tag{7}$$

B. *Adaptive Optimization of Model Parameters*

When the LSTM is used to predict the number of APS, two problems should be considered, namely, the selection of input features and the super-parameter selection of model construction. Theoretically, LSTM deals well with time series with long-term correlations, and the accuracy will be improved with the increase in the number of inputs to the model. But on the other hand, it also brings huge burden to the model training and the prediction takes too much time. Therefore, the input characteristics of specific models need to be adjusted according to the actual situations.

Different time series show different properties, so the choice of parameters in the LSTM also needs conclude the specific situations. Hyperparameters are parameters that are set before the start of the learning process, rather than the parameter data obtained through training. Usually, the hyperparameters need to be adjusted to select a set of optimal parameters for the model. On the one hand, it can improve the accuracy of prediction. On the other hand, it can prevent overtraining. These parameters include learning rate, number of hidden neurons, number of iterations, activation function, optimizer and Dropout probability, etc. In order to solve the problems of model construction in the above two aspects, the model setting is usually repeatedly debugged after the establishment and training of the model, so as to finally obtain the best model setting. This process is greatly influenced by subjective factors, so genetic algorithm is adopted in this paper to optimize it and select the optimal parameters of the network intelligently, so as to achieve the purpose of adaptive adjustment of model parameters. In the proposed algorithm, the number of hidden layer neurons, learning rate, Dropout probability and input quantity in network parameters are selected to form individuals in the population.

The specific process is as follows:

- Step 1: According to the method of real number coding, each individual is a real number channeling, which is composed of the number of hidden neurons, learning rate, Dropout probability and input quantity.
- Step 2: The initial weights of the neural network and the values of other unselected hyperparameters are solid default values. According to the three parameters of the individual neural network and the input characteristics of the model, the LSTM model after training is used to predict the output. The absolute value of the error between the predicted output and the actual value is taken as the calculation formula of the individual fitness function *F*.

$$F = k\left(\sum_{i=1}^{n} asb(y_i - O_i)\right) \tag{8}$$

Where n represents the number of output nodes, y_i represents the predicted output of output nodes in LSTM, and O_i represents the actual output.

- Step 3: The roulette wheel is used to select individuals, and the selection probability p_i of each individual is:

$$p_i = J_i /(\sum_{j=1}^{N} f_j), \quad f_i = k / F_i \qquad (10)$$

In the formula, f_i represents the fitness value of individual I. Since the smaller the fitness value is, the better it is, the fitness value is reciprocal before individual selection. k represents the coefficient; N is the number of individuals in the population.

- Step 4: As individuals adopt real number coding, the crossover operation method adopts real number crossover method. The specific crossover operation of a_k and a_i of chromosome k and chromosome i is as follows:

$$a_{kj} = a_{kj}(1-b) + a_{ij}b \qquad (10)$$

$$ai_j = ai_j(1-b) + a_{kj}b \qquad (11)$$

Where b is a random number in [0, 1].

- Step 5: The specific method to select the jth gene a_{ij} of the ith individual for mutation operation is as follows:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{\max}) * f(g), r > 0.5 \\ a_{ij} + (a_{ij} - a_{\min}) * f(g), r \le 0.5 \end{cases} \qquad (12)$$

Where a_{max} represents the upper bound of gene a_{ij}; Amin represents the lower boundary of genes; r_2 is a random number; g is the number of current iterations; G_{max} is the maximum number of evolutions; r is a random number between [0,1].

- Step 6: Repeat steps 3, 4, and 5 until a new population is created.

III. EXPERIMENTAL ANALYSIS

This paper chooses a roadside parking lot as the research object in Shenzhen, China. The daily real-time APS data was collected in the parking lot during the 14 days period from s November 04, 2018 to November 21, 2018. The time series of APS data is extracted at an interval of 3 minutes. The total amount of intervals is 6720. 5760 intervals are selected as training samples, and the remaining 960 intervals were used as test sets to test the LSTM model. The experimental fundamentals are as following: CPU is core i7-7700; the running memory is 8G; Anaconda is selected as the Integrated Development Environment; the programming language is python3.6.

A. The Influence of Input Step Size on Prediction Results

Single hidden layer network structure is adopted. Number of hidden layer nodes is set as 50. The number of output layer nodes is set as 1.Small batch gradient descent method is selected to train the network. Training subset size is set as 256. RMSProp is selected as the optimization algorithm. The number of epoch is set as 30. In order to value the influence of the length of steps, values are set as 4, 6, 10, 20 and 30 respectively. The prediction results are shown in Table 1.

Table 1 Influence of different length of input steps

The length of input steps	MSE	MAE	MAPE	Training time(s)
4	4.76	1.34	20.22	99.14
6	2.99	0.99	20.14	101.42
10	4.53	1.28	20.91	172.69

As can be seen from table 1, the prediction performance of the model is different under different input time steps. Mean square error (MSE) between 3 and 5; MAE fluctuates around 1. The mean absolute percentage error (MAPE) is about 20%. As the input time step increases, the model training time also increases. But the mean square error (MSE) and mean absolute error (MAE) both decrease first and then increase. The more inputs that do not meet the expected assumptions, the higher the model accuracy. The reason is related to the characteristics of the sequence itself. The number of APS at the time interval of 6 steps before the prediction point has a large relationship with the number of APSs at the prediction point, so when the input is small, the error is large. The relationship between the data 10 steps or more before the prediction point and the prediction point becomes smaller and smaller, and there is noise information in the data. The improved noise algorithm in the preprocessing stage can change this result to a certain extent, but considering that the accuracy of the prediction results with the input length of 30 and the input step size of 6 are not very different, an optimal input feature can be selected to make up for this result. Overall, the optimal selection of input step size is between 4 and 10. Further calculation is required to obtain the optimal solution. This experiment further proves that adaptive adjustment of the initial parameters of the model can improve the prediction accuracy of the model.

B. Model Prediction and Contrast Experiments

BP neural network and algorithm proposed in this paper were used to predict the available parking space respectively. The default parameter of input steps length in the input layer is set as 6. Number of the feature equals 1.The node number of the output layer is set as 1.And the node number of the hidden layer is set as 200. Small-batch gradient descent method is adopted to train the network. The training subset size is chosen to be 256. The optimization algorithm is RMSProp. The number of epoch is set as 40. The predicted time intervals are set as 1, 5 and 10 minutes, namely 3 minutes, 15 minutes and 30 minutes.

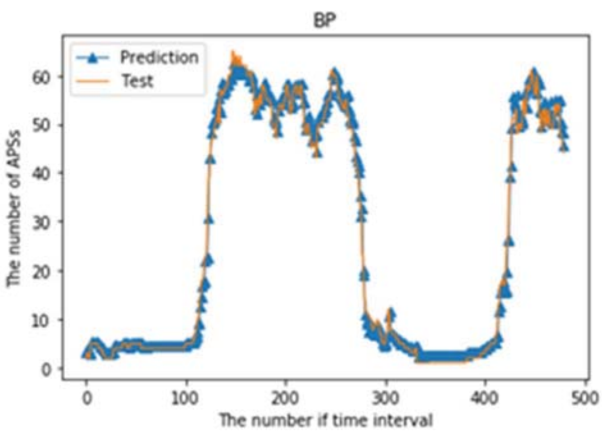

Fig. 2. The prediction results of BP neural network with a prediction interval of 1.

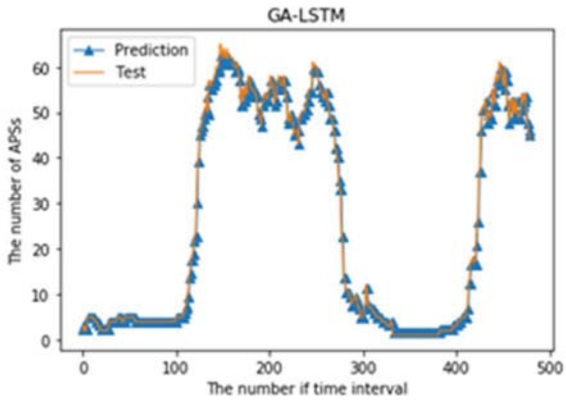

Fig. 3. The prediction results of GA-LSTM with a prediction interval of 10.

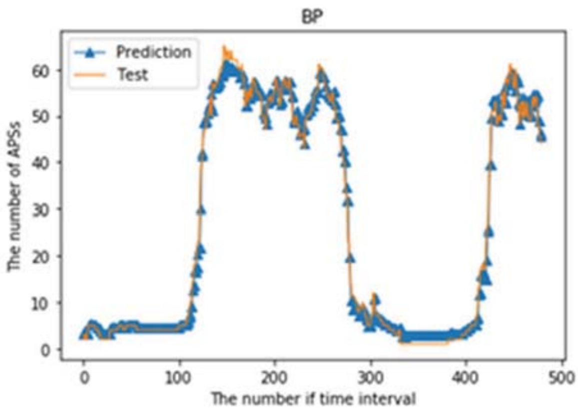

Fig. 4. The prediction results of BP neural network with a prediction interval of 5.

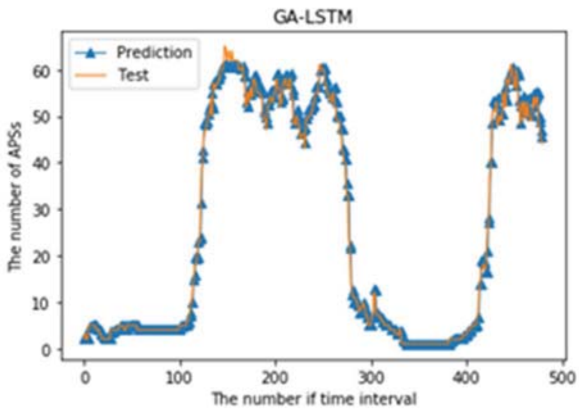

Fig. 5. The prediction results of GA-LSTM with a prediction interval of 5.

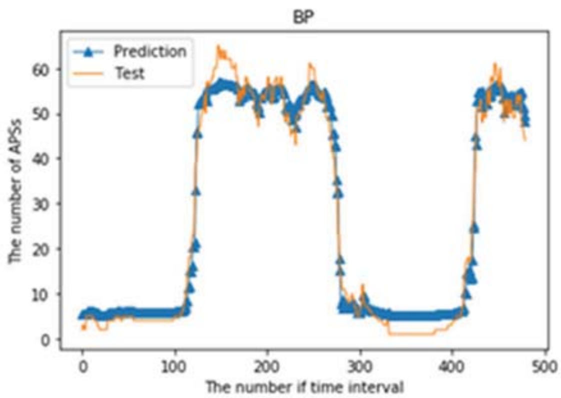

Fig. 6.The prediction results of BP neural network with a prediction interval of 10.

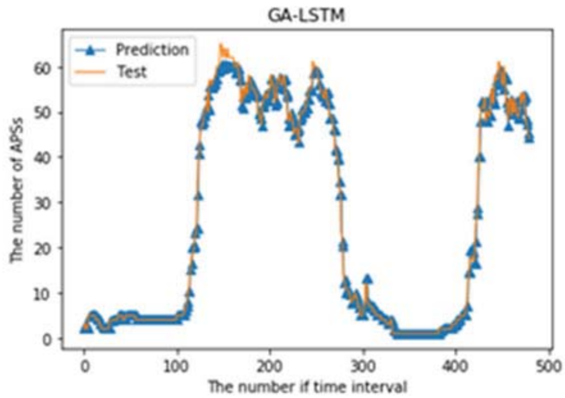

Fig. 7. The prediction results of GA-LSTM with a prediction interval of 10

Fig. 2~7 show the results of the two models with the prediction time interval of 3min, 15min and 30min respectively. The abscissa represents the predicted time period of the test set, and a total of 480 time points are selected. It can be seen that the curve fitting degree of both models is very high when the prediction time interval is relatively short. When the prediction time interval increases gradually, the fitting degree of BP neural network begins to decline, while the LSTM model still maintains a high fitting degree. Therefore, LSTM reveals a better adaptability than BP neural network in the prediction of available parking space.

Table 2 Results comparison of these models

The number of Prediction interval	BP	GA-LSTM
	MAPE（%）	MAPE（%）
1	8.97	6.82
5	18.21	9.72
10	25.64	12.33

As shown in Table 2, at the early stage, both of two models can achieve good prediction performance, and the prediction accuracy is above 90%. As the number of interval increases, the prediction accuracy of BP neural network decreases greatly. When the prediction interval is 10, the prediction accuracy is less than 80%. This phenomenon indicates that the performance of BP neural network model begin to decline and the adaptability of the model is poor as the prediction interval becomes longer. However, the algorithm proposed in this paper still maintains a good prediction accuracy. LSTM optimized by genetic algorithm can achieve higher accuracy and better prediction effect on APS prediction.

IV. CONCLUSION

In this paper, we propose a novel algorithm to predict the urban available parking space. LSTM model is introduced to this paper because its excellent performance in deal with the data of time series. After that, the genetic algorithm is introduced to realize the parameters adjustment in LSTM. This proposed algorithm is validated by comparing the results of BP neural network.

ACKNOWLEDGMENT

This research has been funded by the Science and Technology Public Welfare Research Project of Zhejiang Province, China (Project No. 2017C31038) and General

Scientific Research Project of Zhejiang Education Department, China (Project No.Y201839557).

REFERENCES

[1] A. E. Dunning, "Method and system for projecting dynamic parking availability based on an ongoing survey for remote lots with high demand," M. US, 2006.

[2] S. Liu, H. Guan, and H. Yan, "Unoccupied Parking Space Prediction of Chaotic Time Series," Proceedings of the Tenth International Conference of Chinese Transportation Professionals, 2010.

[3] J. Y. Jie, C. X. Shi, and W. Wei, "Short-term prediction of effective parking berth based on wavelet transform and particle swarm wavelet neural network combination model," J. Journal of Jilin university, 2016, pp. 399-405.

[4] S. Yong, L. C. Ping, and W. Y. Huai, "A forecasting model for parking guidance system," IEEE C. Computer Science and Information Engineering. [2009 WRI World Congress on. Los Anyeles, p. 607-611, 2009].

[5] T. K. Shuang, H. Z. Kang, and Y. X. B. Wen, "Evaluation of parking occupancy prediction method," J. Journal of Tongji university, 2017, pp. 533-543.

[6] E. I. VLAHOGIANNI, K. KEPAPTSOGLOU, and V. TSETSOS, "A real-time parking prediction system for smart cities," J. Journal of Intelligent Transportation Systems, 2016, pp. 192.

[7] J. Y. Jie, W. Wei, and D. Wei. "Available parking space occupancy change characteristics and short-term forecasting model," J. Journal of Southeast University, 2007, pp. 604-8.

[8] Y. Ji, D. Tang, "Forecasting available parking space with largest Lyapunov exponents method," J. Journal of Central South University, 2014, pp. 1624-32.

[9] Y. Ji, D. Tang, and P. Blythe, "Short-term forecasting of available parking space using wavelet neural network model," J. Intelligent Transport Systems, 2015, pp. 202.

[10] X. Ma, Z. Tao, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," J. Transportation Research Part C, 2015, pp. 187-197.

[11] S. Hochreiter, J. Schmidhuber, "Long short-term memory," J. Neural Computation, 1997, pp. 1735-1780.

.