Study on Subway Passenger Flow Prediction based on Deep Learning

Yichao Zhang[1, a]

[1]Logistics Engineering College, Shanghai Maritime University, Shanghai, China.

[a]zhangyichao@stu.shmtu.edu.cn

Abstract

With the continuous advancement of urbanization, the scale of subway operation is expanding day by day, and the operation mode is also more complex. Passenger flow has become an important factor affecting the development of subway. The prediction of subway passenger flow is affected by various external factors, such as climate system, social activity system and so on. In view of the low prediction accuracy of traditional prediction methods based on mathematical statistics and machine learning, this paper, from the perspective of deep learning, forecasts the subway passenger flow by using convolution bidirectional long-short term memory network deep learning prediction model and integrating attention mechanism into the model. The simulation results show that the prediction accuracy of this method is greatly improved compared with the traditional method, and has better prediction ability.

Keywords

Passenger flow prediction; Deep learning; Convolution bidirectional long-short term memory network; Attention mechanism.

1. INTRODUCTION

In recent years, the scale of subway has been expanding. Taking Shanghai Metro as an example, as of December 30, 2018, Shanghai Metro has opened 16 lines and 413 stations with an operating mileage of 676 km, which is the longest urban rail transit system in the world. The average daily passenger traffic volume of Shanghai Metro is more than 10 million person times on weekdays and about 7 million person times on weekends. In 2017, the average daily passenger traffic volume of Shanghai Metro was about 9.693 million. In 2018, the average daily passenger traffic volume of Shanghai Metro is about 10.153 million. As of November 2018, the maximum daily passenger volume of Shanghai Metro is 12.563 million [1]. Subway almost covers people's life, more and more people choose subway travel. With the increase of population, the huge passenger flow caused subway congestion, and even once paralyzed. For example, on July 28, 2015, the power supply failure of Shanghai Metro Line 1 occurred and it was suspended for 3 hours, resulting in a large number of passengers stranded and congested along the line, resulting in traffic chaos. Similar passenger flow congestion caused by emergencies can often be seen in the news. The reason for this is that the relevant departments fail to assess the passenger flow in an emergency and do not take corresponding measures at the first time. Therefore, the prediction of urban subway passenger flow has always been the research focus of rail transit. The significance of this research as follows: (a) Provide scientific data for the transportation department, so as to effectively allocate human and material resources, and improve the safety and comfort of subway operation. (b) Provide data for relevant government departments to deal with emergency events. (c) It provides scientific data for the majority of passengers to choose the appropriate travel mode. Therefore, it is of great

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202011_6(11).0045

value to predict the subway passenger flow effectively. Because the prediction of subway passenger flow is a problem worthy of study and solution, no matter from the perspective of rail transit network construction or people's travel [2].

The prediction of passenger flow is affected by geographical location, traffic conditions and social activities, which is very complex and difficult [3]. Its complexity and difficulty are reflected in the following aspects: (a) There are many projects to be predicted, involving a lot of data, so it is difficult to collect. (b) The diversification of people's travel modes, such as buses, bike sharing and other travel tools, makes it more difficult to predict the subway passenger flow.

With the advent of the era of big data, the traditional mathematical statistics prediction method and machine learning prediction method are very difficult to deal with big data, and have great defects. Therefore, how to predict the subway passenger flow scientifically and accurately is a very difficult problem to be solved. The ability of self-learning features of deep learning model depends on a large number of training data. The continuous development of deep learning technology provides a favorable technical support for solving the problem of predicting passenger flow.

2. SUBWAY PASSENGER FLOW FORECASTING MODEL BASED ON DEEP LEARNING

2.1. Overview of Deep Learning

Deep learning is a branch of machine learning, which has experienced three waves of development [4]. Deep learning first appeared in 1940s and 1960s, which is called cybernetics, and the first deep learning model, namely perceptron, appeared. The second deep learning in the 1980s and 1990s, known as the linking mechanism, can use back propagation to train neural networks with one or two hidden layers. The third is the revival of deep learning since 2006. A model named deep belief network appears, which can effectively train by greedy layer by layer training strategy. From then on, deep learning began to appear in the papers of numerous experts and scholars, a large number of applications in the field of character recognition, face technology, intelligent monitoring and so on.

In the development of deep learning in the past decades, it absorbed the knowledge of human brain, statistics and applied mathematics, and its popularity and applicability have been greatly developed. Deep learning also depends on the development of software infrastructure, such as theano (2010), pylearn2 (2013), torch (2011), Caffe (2013), tensorflow (2015), deep learning toolbox (2016), etc. An important sign of the rapid rise of deep learning is that convolutional neural network won the championship of Imagenet large-scale visual recognition challenge in 2012, reducing the top five error rate of the highest level from 26.1% to 15.3%. Since then, the error rate of convolutional neural network has been reduced to 3.6% in 2017 after continuous improvement [5]. While the scale and precision of deep learning are improved, the tasks they can solve are becoming more and more complex.

2.2. Convolution Neural Network

The CNN is a kind of neural network which is used to process data with similar grid structure. For example, time series data can be regarded as one-dimensional grid formed by conventional sampling on the time axis, while image data can be regarded as two-dimensional pixel grid. CNN has done very well in many areas. CNN uses convolution, which is a special linear operation. The structure of CNN prediction model includes input layer, convolution layer, pooling layer, batch normalization layer, activation layer, full connection layer, regression layer and output layer.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202011_6(11).0045

The basic work of convolution layer is to convolute the input data and convolution kernel function, and the characteristic map obtained is the output value of convolution layer. The equation as follows.

$$s(t)=(x*\omega)(t)=\sum_{a=-\infty}^{\infty}x(a)\omega(t-a) \tag{1}$$

Where the function $x(t)$ is the input and $\omega(t)$ is the kernel function. However, in the practical application of deep learning, the input is usually a high-dimensional data group, and the kernel function is also a high-dimensional parameter array generated by the algorithm. Such high-dimensional arrays are called tensors. In practice, infinite summation can be regarded as the summation of finite arrays.

For convolution of one-dimensional discrete data, each row of the matrix must be equal to that of the previous row after moving an element. This kind of matrix is called Toeplitz matrix. For the two-dimensional case, convolution corresponds to a double cyclic matrix. In convolution layer, the operation is carried out through three important ideas, namely sparse interaction, parameter sharing and equivariant representation. In addition, convolution layer also provides an input method with variable size. After convolution operations are performed in parallel by convolution layer, the feature data are normalized by normalization layer, which speeds up the convergence speed and makes CNN training more stable.

Then, the problem of gradient dispersion in the training process is solved through the activation layer to accelerate the convergence speed. The Leaky Relu function is usually used, as follows:

$$f(x)=\begin{cases}x, x>=0\\ ax, x<0\end{cases} \tag{2}$$

Then, some unimportant features are deleted by pooling layer, and the output of the convolution layer is adjusted. The pooling function uses the overall statistical characteristics of adjacent outputs at a certain location to replace the output of the network at that location.

After pooling, the data usually needs to be expanded, and then the full connection layer will sum the data processed by flatten, and the output as follows:

$$o_j = w_j p_j + b_j \tag{3}$$

Where o_j is the output of the full connection layer, p_j is an expanded one-dimensional eigenvector, w_j is the weight coefficient, p_j is the offset term. The function of full connection layer is to reduce the loss of feature data.

Finally, the prediction results are output by regression layer. For sequential regression, the loss function of regression layer is the half mean square error of prediction response. The equation as follows.

$$loss=\frac{1}{2}\sum_{i=1}^{R}(o_i-z_i)^2 \tag{4}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202011_6(11).0045

Where z_i is the actual value, o_i is the predicted value.

2.3. Bidirectional Long-Short Term Memory Network

In order to process one-dimensional time series data, the most effective sequence model in practical application is the long-term and short-term memory network model. It is based on the idea of self-circulation, and its core is to generate gradient long-term continuous flow path. It is gated and consists of input gate, forgetting gate and output gate. The weight and cumulative time scale of gated LSTM can be changed dynamically. The structure of LSTM prediction model consists of input layer, LSTM layer, full connection layer, regression layer and output layer. The LSTM can only encode forward, not reverse, when processing time series. The Bidirectional long-short term memory network can solve this problem [6]. It consists of two ordinary LSTM layers. One is the forward LSTM layer which uses the past information, and the other is the reverse LSTM layer which uses the future information. The structure of BiLSTM prediction model consists of an input layer, a bilstm layer, a full connection layer, a regression layer and an output layer. At time t, the bilstm layer can use the information of $t-1$ and $t+1$ at the same time. Therefore, the output of BiLSTM can be obtained by adding the vectors from two LSTM layers.

2.4. Convolution Bidirectional Long-Short Term Memory Network

Convolutional long-short term memory network is a comprehensive network which combines the advantages of convolutional neural network and long-short term memory network. The structure of CLSTM model consists of input layer, convolution layer, pooling layer, normalization layer, activation layer, LSTM layer, full connection layer, regression layer and output layer. The function of CNN layer is to extract feature data. The function of LSTM is to filter and update the feature data, and endow the feature with forward temporal dependence, which can better screen out the features of time series. The CLSTM model improves the classification effect of single CNN model and LSTM method.

Therefore, convolution bidirectional long-short term memory network optimizes LSTM layer on the basis of clstm. Bilstm layer is used, and a pair of bidirectional LSTM layers are used to deal with the time series before and after the network, so as to endow it with the relation between before and after.

2.5. The Model of Self Attention

The self attention model is a model that dynamically generates different weights of connections between the input and output of the same layer of network to get the output of the network. The self attention model can establish long-distance dependence within sequences. In fact, it can also be done through the full connection layer, but the number of connected sides of the fully connected network is fixed, so it is unable to process the variable length sequences. The self attention model can dynamically generate the weights of different connections, and the weights generated and the weights are variable. When a longer sequence is input, only more connected edges need to be generated. $x_{1:n}=[x_1,x_2,\cdots,x_n]$ is used to represent n input information, and the output information is $H_{1:n}=[H_1,H_2,\cdots H_n]$ of the same length. Firstly, three sets of vector sequences are obtained by linear transformation.

$$\begin{aligned} Q &= W_Q X \\ K &= W_K X \\ V &= W_V X \end{aligned} \tag{5}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202011_6(11).0045

Where K, V and Q are query vector sequence, key vector sequence and value vector sequence respectively. W_Q, W_K and W_V are the parameter matrices that can be learned. The output vector can be obtained as follows.

$$\begin{aligned} h_i &= att((Q,K,V),q_i) \\ &= soft\max(\frac{QK^T}{\sqrt{d}})V \\ &= \sum_{j=1}^{n}\alpha_{ij}v_j \end{aligned} \tag{6}$$

Where $i, j \in [1,n]$ is the position of the input and output vector sequences, and the connection weight α_{ij} is dynamically generated by the attention mechanism. The self attention model can be used as a layer of a network.

2.6. Principle of Convolution Bidirectional Long-Short Term Memory Network Combined with Attention Prediction Model

The specific workflow of CBiLSTMA prediction model used in this paper as follows.

(1) Local CNN network

CNN can effectively extract the spatial correlation in the pixel matrix through continuous convolution operation. Inspired by this, CNN is used to extract the spatial dependence of data. The input data is input into the CNN component of CBiLSTMA model as one-dimensional gray image for learning. However, if CNN is directly applied to the whole matrix, a large number of distant and weak correlation sites will be considered, which will weaken the learning ability of CNN for strong correlation between sites, and affect the overall prediction performance. After convolution, the extracted spatial association information is processed as BiLSTM network input.

(2) The BiLSTM network

The BiLSTM network recursively applies the transfer function to the input hidden state vector, which solves the gradient disappearance and gradient explosion problems that may occur when the shallow neural network processes the long-term sequence. The short-term and long-term time-dependent modeling of BiLSTM network has better performance in time-series data modeling.

$$h_t^i = BiLSTM(p_t^i, p_{t-1}^i) \tag{7}$$

Where h_t^i is the output at time t , p_t^i is the convolution output.

(3) The layers of self attention

The attention mechanism is introduced to correct the effect of time drift, and the final long-term output of bilstm network is obtained.

$$h_i = BiLSTM(att((Q,K,V),q_i)) \tag{8}$$

(4) Integration training

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202011_6(11).0045

The predicted value of o_t was obtained by activating h_i with tanh function.

$$o_t = \tanh(W_f h_t + b_f) \tag{9}$$

Where W_f and b_f are learning parameters, and the output result is within [-1,1]. The predicted value o_t was obtained by inverse normalization.

3. SIMULATION EXPERIMENT

3.1. Simulation Experiment of CNN Prediction

In this paper, CNN method is used to predict the passenger flow of subway station on the deep learning toolbox framework of MATLAB. The CNN training parameters are set as follows: He Kaiming initializer is used for initialization coefficient of each layer, ADAM is used for optimizer, initial learning rate is 0.1, cross entropy function is used for loss function, total number of running rounds is 100, and batch size used in model is 2000. The test data is the last 5000 data, and the rest are training data. The prediction results of CNN method are shown in Fig. 1.

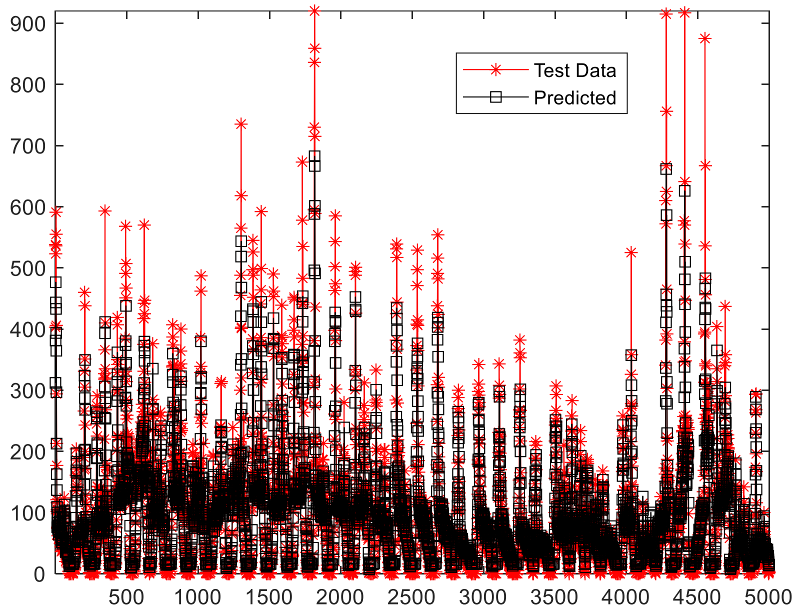

Fig 1. Graph of CNN prediction results

3.2. Simulation Experiment of Bilstm Prediction

The simulation experiment training parameters of BiLSTM are set as follows: The number of training rounds is 100, the minimum batch size is 2000, the initial learning rate is 0.005, and ADAM optimizer is used. The prediction results are shown in Fig. 2.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202011_6(11).0045

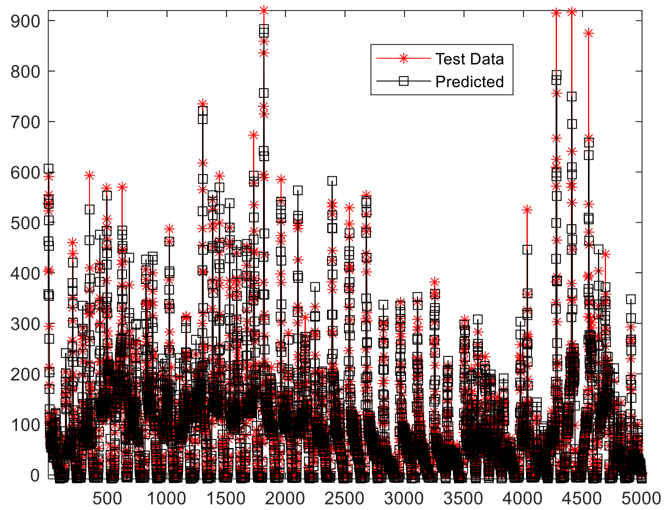

Fig 2. Graph of BiLSTM prediction results

3.3. Simulation Experiment of Cbilstm Prediction

The simulation experiment training parameters of CBiLSTM are set as follows: The number of training rounds is 100, the minimum batch size is 2000, the initial learning rate is 0.005, and Adam optimizer is used. The prediction results are shown in Fig. 3.

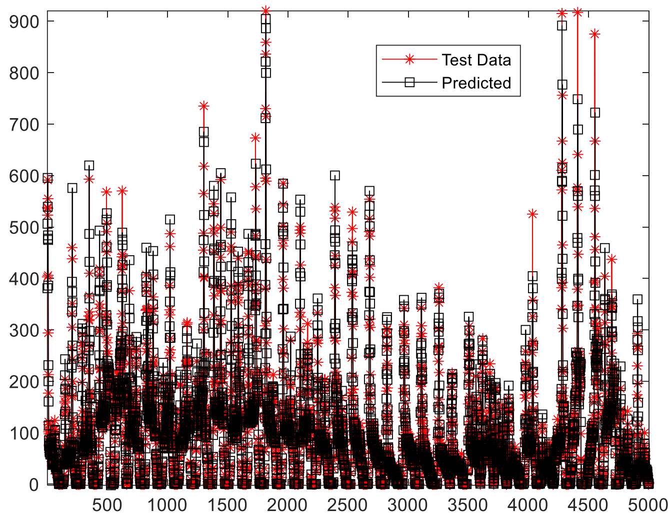

Fig 3. Graph of CBiLSTM prediction results

3.4. Simulation Experiment based on Cbilstma Model

The simulation experiment training parameters of CBiLSTMA are set as follows: Using Adam optimizer, the number of training rounds is 200, the minimum batch size is 3000, and the initial learning rate is 0.01. The prediction results are shown in Fig. 4.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202011_6(11).0045

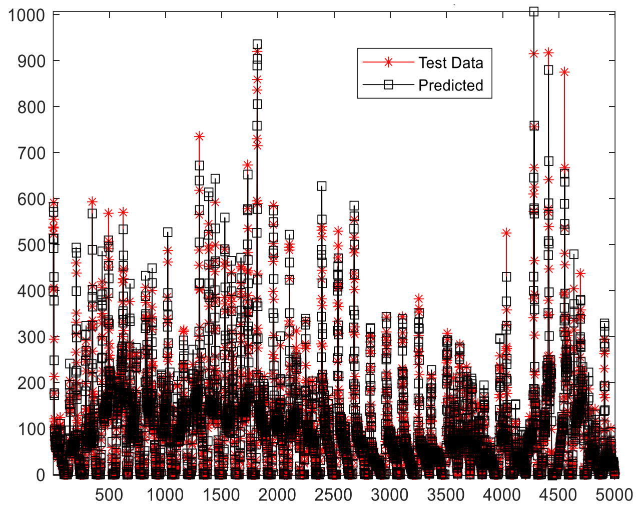

Fig 4. Graph of CBiLSTMA prediction results

3.5. Results Analysis and Comparison

In this paper, the prediction error RMSE and correlation coefficient R are used to quantitatively analyze and compare the experimental results. The results are shown in Table 1.

Table 1. Comparison of experimental results of different deep learning prediction models

Prediction model	RMSE	R
CNN	33.9819	0.9694
BiLSTM	25.8269	0.9701
CBiLSTM	25.1125	0.9713
CBiLSTMA	24.7188	0.97217

4. CONCLUSION

With the expansion of subway scale, the arrival of artificial intelligence and big data era, people have higher and higher requirements for subway passenger flow prediction. The traditional method is very difficult to solve the problem of big data passenger flow prediction, and the prediction error is relatively large, which is limited in practical application. Therefore, it is urgent to find new methods to improve the prediction effect. In order to improve the prediction accuracy of subway passenger flow, this paper applies the deep learning prediction and attention mechanism to the prediction of subway passenger flow. A prediction method based on convolution bidirectional long-short term memory network and attention mechanism is used in this paper. The experimental results show that the deep learning method can effectively improve the accuracy of subway passenger flow prediction.

REFERENCES

[1] Statistical analysis of passenger flow data of Shanghai Metro, [EB/OL]. https://www.ditietu.com/p/1824?from=singlemessage, 2018-12-29.

[2] M. Yuan. Current situation and development trend of passenger flow forecast in Urban Rail Transit [J]. Science and Technology in China, Vol.13 (2018), p. 209-210.

[3] Q.M. Zou, X.P. He. Summary of short-term prediction methods for subway passenger flow [J]. Journal of Chongqing Technology and Business University, Vol.37 (2020), p. 25-32.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202011_6(11).0045

[4] D. Bahdanau, K. Cho, Y. Bengio. Neural Machine Translation by Jointly Learning to Align and Translate [J]. Computer ence, Vol.3 (2014), p. 13-17.

[5] X.P. Qiu. Neural network and deep learning [M]. Machinery Industry Press, China, 2020, p.212-215.

[6] Y. Liu, C. Sun, L. Lin. Learning Natural Language Inference using Bidirectional LSTM model and Inner-Attention [J]. International Core Journal of Engineering, Vol.6 (2016), p. 26-29.