Short-Term Subway Passenger Flow Prediction Based on GCN-BiLSTM

Donglin Ma, Yating Guo* and Sizhou Ma
Lanzhou University of Technology, Department of computer and communication
Email: 1443071043@qq.com

Abstract. Aiming at the problem that the accuracy of passenger flow prediction is not high, this paper presents a short-term passenger flow forecasting model based on Graph Convolutional Neural Network (GCN) and Bidirectional Long-term Memory Network (BiLSTM). Firstly, the historical traffic time series is divided into three time modes: recent period, daily period and weekly period; Secondly, we construct three models based on GCN and BiLSTM to capture the spatial and temporal dependence of the three patterns; Finally, the parameter matrix is used to fuse the output of the three time modes to obtain the final prediction result. By testing the data set of subway passenger flow in a city in January 2019, the experimental results show that the root mean square error of the model is reduced by 8.515% and the average absolute error is reduced by 4.239% compared with the single BiLSTM model, it has a high fitting degree with the real passenger flow value and has certain application value for the reasonable allocation of subway capacity.

1. Introduction
With the increasingly prominent problem of urban road congestion, China has begun to vigorously develop public transport and advocate "green" travel, so as to effectively alleviate traffic congestion and reduce carbon emissions. In many public transport modes, urban rail transit system has obvious advantages in relieving ground traffic pressure due to its high speed, environmental protection, energy saving and low consumption, and large traffic volume. However, with the gradual increase of subway passenger flow, some subway stations are relatively crowded during peak hours and are prone to stampede accidents. Therefore, it is urgent to effectively and accurately predict the subway passenger flow in a short time, formulate emergency evacuation plans during the rush hour, and rationally allocate subway capacity.

At present, the short-term forecasting methods of urban rail transit passenger flow are mainly divided into traditional methods and machine learning methods. Traditional methods mainly use time series to predict short-term traffic flow, such as the Kalman filtering model [1], the Autoregressive Moving Average (ARIMA) model [2], the Multiple Linear Regression model [3] and so on. However, these traditional methods have high requirements for the smoothness of the sample time series, and the short-term passenger flow has obvious characteristics of non-linearity and volatility, resulting in a significant reduction in forecasting performance. Commonly used machine learning models continuously adjust their own parameters through adaptive learning to obtain more accurate calculation results, such as Support Vector Machine [4], decision tree model [5], K-Nearest Neighbors (KNN) [6] and so on. Although these models can solve the problem of poor performance of traditional methods, the pursuit of high precision can easily lead to large model training time costs and excessive data samples requirements [7]. With Recurrent Neural Networks (RNN), Long Short-term Memory (LSTM) and Gated Recurrent Neural Network (GRU) and other deep learning models based on RNN...
[8-9] can effectively process time series data, which is frequently used in passenger flow forecasting. Although these models can effectively learn the long-term dependence of the traffic flow sequence and capture the complex time characteristics, they do not consider the spatial structure of the traffic road network, and it is difficult to realize the problem of passenger flow prediction between road networks. In 2013, Bruna et al. [10] first proposed GCN to effectively solve the problem of network space correlation. GCN can directly use graph structure information to extract local features, and process arbitrarily connected irregular data to complete the task of traffic prediction [11].

In summary, starting from the time dimension and space dimension, this paper proposes the GCN-BiLSTM model to predict the short-term passenger flow of urban rail transit. In this model, multiple time modes of short-term, daily and weekly are considered, and GCN and BiLSTM are respectively used to capture the spatial characteristics and temporal characteristics of short-term passenger flow to produce the final prediction results.

2. Subway Passenger Flow Prediction Model

2.1. Definition of Passenger Flow Forecasting Problem

In this paper, the metro network within a certain city is defined on a graph. An undirected graph $G = (V, E, A)$ is constructed, and $V$ represents the set of subway stations. The stations in the subway network are the vertices of graph $G$. Assuming there are $N$ stations in total, the station set can be defined as $V = \{v_1, v_2, \ldots, v_N\}$; $E$ stands for edge set, and edge stands for connectivity between sites; the adjacency matrix between sites is represented by $A$, that is, the element $A_{ij}$ in $A^{N \times N}$ represents the connection relationship between vertex $v_i$ and $v_j$; Let the site feature matrix be $X \in R^{N \times P}$, where $X$ is the number of rows $N$ is the number of stations, the number of columns $P$ is the characteristic attributes of the stations, and the historical passenger flow is the characteristic attributes of each station.

The goal of urban subway passenger flow forecasting is to predict the passenger flow for a certain period of time in the future based on historical passenger flow. As shown in Equation (1), the prediction task of this paper is to train a function $f(\cdot)$ to act on the inbound and outbound passenger flow observation value $x=(x_{t-h:t-1}, x_{t-1}, \ldots, x_{t})$ in the historical time interval of each station, and output the inbound and outbound passenger flow value $x=(x_{t+h}, x_{t+h+1}, \ldots, x_{t+h+p})$ in the future time interval.

$$f(x_{t-h:t-1}, x_{t-1}, \ldots, x_{t}) \rightarrow (x_{t+h}, x_{t+h+1}, \ldots, x_{t+h+p})$$

(1)

Where $x_t$ represents the passenger flow of the site at the time step $t$; $t_h$ represents the length of the selected historical interval; $t_p$ is the length of the interval to be predicted.

2.2. GCN Model

In the urban metro rail transit system, the metro station network has a complex non-Euclidean spatial structure, the number and location of the stations are known. and each of the upstream and downstream of the subway site is fixed, so this article use GCN to trap the subway site spatial correlation characteristics of topology structure, the characteristics of each site is regarded as the signal on the figure, directly on the diagram of the figure signal processing, to capture a meaningful patterns and characteristics in the space.

The input of the graph convolution model includes adjacency matrix $A$ and passenger flow characteristic matrix $X$. The propagation mode between the middle layer and the layer is shown in Equation (2).
\[ H^{(l+1)} = \sigma \left( \frac{1}{2} \hat{A} \frac{1}{2} \hat{D} A D \frac{1}{2} \hat{H}^{(l)} W^{(l)} \right) \]  

(2)

Where \( \hat{A} \) is the sum of the adjacency matrix and the identity matrix, that is \( \hat{A} = A + I \); \( \hat{D} \) is the degree matrix of \( \hat{A} \); \( H^{(l)}, H^{(l+1)} \) respectively represent the characteristic matrix of the \( l \) layer and the \( l + 1 \) layer. The characteristic matrix of the initial layer \( H^{(0)} \) is the passenger flow characteristic matrix \( X \). \( W^{(l)} \) represents the weight matrix of layer \( l \); \( \sigma \) is the activation function. The calculation of \( A_j \) and \( \hat{D} \) is shown in Equation (3) and Equation (4) respectively.

\[
A_j = \begin{cases} 
1, & i \neq j \\
0, & i = j 
\end{cases}
\]

(3)

\[
\hat{D}_j = \sum_{i=1}^{N} A_{ij}, \quad i = j \\
0, \quad i \neq j
\]

(4)

Where \( A_j \) represents the connection relationship from site \( i \) and site \( j \).

2.3. BiLSTM Model

Subway passenger flow prediction is a typical time-space sequence prediction problem, and the passenger flow information input before and after is interrelated. The classic RNN has a memory function for the input information and can affect the output result. However, as the input increases, the perception of information long ago decreases, resulting in long-term dependence and gradient disappearance problems. The hidden layer of the LSTM[16] network has added forget gate, input gate, output gate and storage memory unit, which effectively avoids the problems of RNN gradient disappearance and long-sequence information dependence.

BiLSTM is formed by stacking two one-way LSTM forwards and backwards, which can obtain data feature information in both positive and negative directions at the current node. It can not only capture information features of future data, but also capture information features of past data. Therefore, this paper uses BiLSTM to extract the information of the past and the future effectively, making the prediction result more accurate. The BiLSTM network structure is shown in Figure 1.

![Figure 1. Schematic diagram of BiLSTM network structure](image)

2.4. GCN-BiLSTM Network Passenger Traffic Prediction Model

In this paper, the GCN-BiLSTM model is proposed to predict the short-term passenger flow of subway. The overall framework of the model is shown in Figure 2. Due to the strong correlation between the period to be predicted and its recent segments, daily segments and weekly segments [17], we firstly...
modeled the recent, daily and weekly patterns in the historical time series. Then, each pattern uses GCN to capture the spatial correlation characteristics of the topological structure of the site to obtain the feature vector containing spatial information. BiLSTM is then used to capture the time-dependent characteristics of the positive and negative passenger flows simultaneously. Finally, the output of the three modes is further fused through the full connection layer to obtain the final prediction result.

2.4.1. Model input The input of the model includes the passenger flow data of the recent period, daily and weekly to be predicted. The input of the recent period refers to the historical observation value of the time interval adjacent to the period to be predicted. The input of the daily period and the weekly period refers to the historical observed values in the past few days or weeks at the same time interval as the period to be predicted. The specific input is shown in Equations (5), (6) and (7).

$$x_r = \left( x_{r, -j}, \ldots, x_{r, -2}, x_{r, -1} \right)$$ (5)

$$x_d = \left( x_{d, -d+m}, \ldots, x_{d, -2}, x_{d, -1} \right)$$ (6)

$$x_w = \left( x_{w, -w+7m}, \ldots, x_{w, -2}, x_{w, -1} \right)$$ (7)

2.4.2. Feature fusion First, the spatial characteristics of passenger flow data are extracted by the graph convolutional neural network layer. Then, the BiLSTM layer is used to extract the time characteristics of passenger flow data containing spatial information. Finally, the parameter matrix is used to fuse the output of three time modes, namely recent period, daily period and weekly period, to obtain the final prediction result, whose function expression is shown in Equation (8).

$$\hat{Y}_t = \sigma \left( W_r \odot \hat{Y}_r + W_d \odot \hat{Y}_d + W_w \odot \hat{Y}_w \right)$$ (8)

Where, $$\hat{Y}_t$$ represents the prediction target of the time interval of $$t$$; $$\odot$$ represents the product of...
Hadamar; $W_r, W_d, W_w$ represents the parameter tensors to be learned, and respectively represents the degree of influence by the recent, daily and weekly dependencies; $\hat{Y}_r, \hat{Y}_d, \hat{Y}_w$ represents the output of three time modes: recent period, daily period and weekly period; $\sigma$ represents the activation function to obtain the final prediction target.

3. Experimental Results and Analysis

3.1. Data Set Description

The smart card swiping data set of urban rail transit system records the swiping records of passengers over a period of time, reflecting the passenger flow changes of the rail transit system. In this paper, about 70 million pieces of data were collected from 81 subway stations of 3 subway lines in a certain city, and the data was collected on January 1, 2019 solstice and January 25. Since January 1st is a New Year holiday, the change of passenger flow is quite different from that of normal working days. The data of January 1st is eliminated firstly. Then, by analyzing the data field, the irrelevant property field which is not helpful to the prediction is removed. Finally, select the time and stationID of the site field, and use these two fields to calculate the passenger flow passing through a certain site every five minutes. This article uses the data from January 2th to January 22th as the training set, and the data from 23th to 25th as the test set.

3.2. The Evaluation Index

In order to better analyze the experimental results, this paper uses Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE) as the performance of the experiment indexes, index calculation formulas are shown in formula (9), (10), and (11).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$  \hspace{1cm} (9)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$  \hspace{1cm} (10)

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%$$  \hspace{1cm} (11)

Where $y_i$ represents the actual passenger flow data and $\hat{y}_i$ represents the predicted passenger flow data.

3.3. Experimental Parameter Setting

The hardware platform of the experimental environment in this paper is: graphics card GPU NVIDIA GTX 1080Ti, CPU processor Inter(R) Core(TM)i7-8750H@2.20GHz, 8GB of memory; The software environment is 64-bit Windows10 operating system, the development language is Python 3.7, the deep learning framework TensorFlow, and the IDE development tool is the most popular PyCharm to complete the building and training of subway passenger flow prediction model.

In the training process, the parameters of the GCN-BiLSTM model were set as follows: the initial learning rate was set to 0.001, the AdamOptimizer was used to optimize automatically, the batch size was set to 32, and the number of training was set to 1000. The recent period, daily period and weekly period were selected respectively. The graph convolution kernel is set to 2.

The prediction accuracy of GCN-BiLSTM model is determined by the number of convolution layers and the number of two-way memory network layers. In order to achieve the highest prediction accuracy, this paper carries out multiple comparison experiments on dynamic adjustable parameters
such as the number of graph convolution layers and the number of two-way memory network layers on the same data set, and selects the optimal parameters according to the experimental results.

As can be seen from Table 1, when the number of graph convolution layers of GCN is set to 2 and the number of BiLSTM layers is set to 4, the prediction effect of the model reaches the best. Therefore, the paper selects the combined model of the convolutional layer of two diagrams, plus four BiLSTM and one full connecting layer to make the prediction.

| GCN layers | BiLSTM layers | RMSE   | MAE   | MAPE  |
|------------|---------------|--------|-------|-------|
| 2          | 3             | 27.785 | 12.439| 13.990|
| 2          | 4             | 26.849 | 12.522| 13.447|
| 3          | 3             | 27.062 | 13.517| 14.115|
| 3          | 4             | 28.307 | 13.135| 14.672|

Table 1. Comparison of results of different model structures

3.4. Model Evaluation

BiLSTM is selected as the model for comparative analysis in this paper. BiLSTM model and GCN-BiLSTM model were used to predict passenger flow under different prediction steps, and the results are shown in Table 2. Based on the RMSE evaluation index, the GCN-BiLSTM model proposed in this paper is 8.515%, 8.926%, and 8.338% lower than the BiLSTM model at 15min, 30min, and 45min prediction steps respectively; based on the MAE evaluation index, the model proposed in this paper is compared with BiLSTM The model reduced by 4.239%, 3.489%, and 2.938% under the predicted step size of 15min, 30min, and 45min, respectively.

Observing the experimental results of the GCN-BiLSTM model at 15min, 30min, and 45min respectively, it is found that as the prediction step size increases, the prediction accuracy of the GCN-BiLSTM model gradually decreases, mainly due to the medium-term (30min) and long-term (45min) There are significant disturbance factors (such as weather factors, holiday factors) and other problems in passenger flow prediction. Therefore, the reference value of medium and long-term passenger flow prediction results is limited.

Table 2. Comparison results of prediction performance of different models

| Model       | Evaluation Index | Prediction Step /min |
|-------------|------------------|----------------------|
|             |                  | 15       | 30       | 45       |
| BiLSTM      | RMSE             | 15.886   | 17.615   | 18.147   |
|             | MAE              | 10.280   | 11.086   | 12.275   |
|             | MAPE/%           | 0.1468   | 0.1508   | 0.1597   |
| GCN-BiLSTM  | RMSE             | 13.837   | 14.689   | 16.809   |
|             | MAE              | 8.928    | 9.597    | 11.337   |
|             | MAPE/%           | 0.1021   | 0.1163   | 0.1318   |

Figure 3. Comparison of prediction results of each model in a station on January 23
Figures 3 to 5 visually show the comparison of the actual passenger flow value of a certain station on January 23, 24, and 25 with the prediction results of the BiLSTM model and the GCN-BiLSTM model. The results show that the prediction results of the GCN-BiLSTM model proposed in this paper have a higher degree of fit with real passenger flow data values.

4. Conclusion

In this paper, a combined model based on GCN-BiLSTM is proposed to predict the passenger flow of metro rail transit. By establishing the station adjacency matrix, the model makes use of GCN and BiLSTM to fully capture the spatial and temporal characteristics of subway rail transit passenger flow data. The experimental results show that the GCN-BiLSTM combination model proposed in this paper has a high prediction accuracy, and the predicted passenger flow data can keep a high degree of consistency with the real passenger flow data, so it is a practical passenger flow prediction model. In future work, the influence of external factors such as weather and events on passenger flow will be considered to further improve the prediction accuracy of the model.

5. References

[1] Jie Xiong, Wei Guan, Yuxing Sun, Metro transfer passenger forecasting based on Kalman filter[J]. Journal of Beijing Jiaotong University, 2013, 37(03):112-116+121.
[2] Changjun Cai, Enjian Yao, et al. Prediction of urban railway station’s entrance and exit passenger flow based on multiply ARIMA model[J]. Journal of Beijing Jiaotong University, 2014, 38(02):135-140.
[3] GUO J, HUANG W, WILLIAMS B, Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification[J]. Transportation Research Part C Emerging Technologies, 2014, 43:502014.
[4] Sun Y, Leng B, Guan W. A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system[J]. Neurocomputing, 2015, 166(Oct.20):109-121.
[5] Hongjun Xue, Guangjiao Chen, Xinmin Li, et al. A Short-term Forecasting of Traffic Flow Parameters Based on Decision Tree Theory[J]. Traffic information and safety, 2016, 34(3):64-71.

[6] Peiqun Lin, Litian Chen, Yongwei Lei, Short-Term Forecasting of Subway Traffic Based on K-Nearest Neighbour Pattern Matching[J]. Journal of South China University of Technology (Natural Science Edition), 2018, 46(1):50-57.

[7] Xiangxue Wang, Lunhui Xu, Short-term Traffic Flow Prediction Based on Deep Learning[J]. Transportation system engineering and information, 2018, 18(01):81-88.

[8] HUANG W, SONG G, HONG H, et al. Deep Architecture for Traffic Flow Prediction: Deep Belief Networks with Multitask Learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(5):21914-2201.

[9] WEN H, LIU Y, CHE W, et al. Sequence-to-Sequence Learning for Task-Oriented Dialogue with Dialogue State Representation[J]. Transportation Research Part C: Emerging Technologies, 2018, 107:287-300.

[10] Bruna J , Zaremba W , Szlam A , et al. Spectral Networks and Locally Connected Networks on Graphs[J]. Computer Science, 2013.

[11] Han, Wang, Ren, et al. Predicting Station-Level Short-Term Passenger Flow in a Citywide Metro Network Using Spatiotemporal Graph Convolutional Neural Networks[J]. International Journal of Geo-Information, 2019, 8(6):243.