Proceedings of the 2023 International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/9/20230077

# Short-term passenger flow prediction based on deep learning

**Zining Guo [1, 3, †], Jingchu Xu [2, †]**

[1] School of Traffic and Transportation, Central South University, Changsha, Hunan, 410075, China
[2] School of Economics and Management, Qingdao University of Science and Technology, Qingdao, Shandong, 266000, China

[3] 8212200110@csu.edu.cn
[†] These authors contributed equally.

**Abstract.** With the continuous improvement of urbanization, the problem of urban congestion has become increasingly prominent. Rail transit can greatly alleviate the congestion of ground traffic and is considered as a feasible solution to alleviate urban congestion. However, there are many factors affecting rail transit, among which the number of passengers accounts for the majority. Excessive passenger flow will sharply increase the probability of public safety events in confined space. The continuous development of deep learning has benefited from studies such as convolutional neural networks (CNN), so as to correctly predict short-term traffic. Despite continual model accuracy improvement, prior research has given little regard to the application boundaries of diverse networks in traffic prediction. In this paper, by comparing the results of previous studies, the prediction effect and deficiency of long short-term memory (LSTM), CNN, graph convolutional network (GCN) are discussed. And results reveal that prediction models relying on recurrent neural networks (RNN), such as LSTM, can capture the time dimension of short-term passengers very well, but not the spatial patterns, and model training time is lengthy. The GCN model would account for the spatial dependence of the traffic network, but it has poor performance when constructing the deep network. In general, each of these three models has its own unique aspects. Due to some problems in a single model, the multi-model combination deep learning framework is gradually emerging. In practical application, the complexity of the model, performance effect and application value should be considered comprehensively.

**Keywords:** LSTM, CNN, GCN, passenger flow forecast.

## 1. Introduction

Since the 21st century, the economic growth of various countries has accelerated, and the construction level of transportation has been improved day by day. Although rail transit brings efficiency and convenience, associated risks also come along, such as congestion and property damage. [1]. In the operation and management of rail transit, passenger flow forecasting has been a very important part.

There are already many effective models for predicting passenger numbers. There are about three stages of development. The first stage is the traditional model of mathematical statistics. In this stage, there are roughly parametric model and non-parametric model [2]. Parameter model is widespread used in passenger flow prediction, which has the characteristics of simplicity, rapidity and less input data. The main models include autoregressive integrated moving average (ARIMA) [3], seasonal autoregressive integrated moving average model (SARIMA) [4], Kalman filtering (KF) model [5], etc. ARIMA model has been regarded as a superior time series model with non-stationary properties [3]. Ye et al. combined the model and used the IC card data of Jiaozuo City to predict the passenger flow of the line, these proving that the weekend data is more accurate in the prediction [6]. They comparison mean absolute percentage error (MAPE) is smaller, so its results are relatively accurate [7]. Zhang et al. used Extended Kalman Filtering (EKF) to calibrate and improve the dynamic traffic allocation (DTA) model, which plays an important role in accurately estimating and predicting traffic conditions [8]. However, the ARIMA model refers to the stationarity of time series, with low accuracy [9]. Non-parametric machine learning algorithm can fit many different forms of function without too many assumptions. Common models are k-nearest neighbour (KNN) [10], artificial neural network (ANN) [11], etc. Lin et al. used KNN model combined with support vector regression method to predict passenger flow, and verified KNN's excellent performance in passenger flow prediction by comparing root mean squared error (RMSE) and MAPE with traditional methods [10]. G et al. used ANN model to forecast passenger numbers of a subway station in Italy, and the results proved relatively accurate.

Its main models include recurrent neural networks (RNN), convolutional neural network (CNN), graph convolutional network (GCN), etc. Ouyang et al. were the first to use the LSTM model to focus on real-time information by encoding previously recorded data and dates to provide the information needed for predictions, which proved that LSTM could obtain better accuracy compared with other methods [12]. Yang et al. with the view of achieving the purpose of improving the service level of tourism service area, a composite model of CNN is studied, that is, with the gated recurrent unit model (GRU). The experimental results obtained by them verify that, in comparison with the simple CNN model, the algorithm error of CNN combined model is significantly smaller when the total flow velocity is larger, so the results are better [13]. Since GCN can consider the spatial-temporal dependence characteristics of the transportation network, especially the topological structure relationship among stations, roads and regions, it has faster training efficiency and fewer hyperparameters than RNN and CNN. Zheng et al. discovered a new combinatorial model of GCN, using it in combination with generative adversarial networks (GAN), and used it to predict traffic flow. The model is tested by collecting traffic flow data sets at multiple intersections, they verified that their performance was more than 30.54% higher than the baseline method, showing obvious advantages in multi-step prediction [14].

While these previous efforts have improved the accuracy of short-term ridership forecasting, less attention has been paid to the exploration of the application boundary of this algorithm. The following content will focus on introducing the LSTM model based on RNN, CNN and GCN, and discuss the prediction effect and shortcomings of the three models combined with previous studies.

## 2. Method

### *2.1. LSTM*

As shown in Figure 1, RNN (cyclic neural network) has the problem of long-term dependence, a new model, namely long short-term memory (LSTM), comes into being. It is a sort of time cyclic neural network. In standard RNN, it has only one very simple structure, the tanh layer, which is the repetitive structural module in which it exists. And the chain form of the repetitive neural network module is found in all RNN.

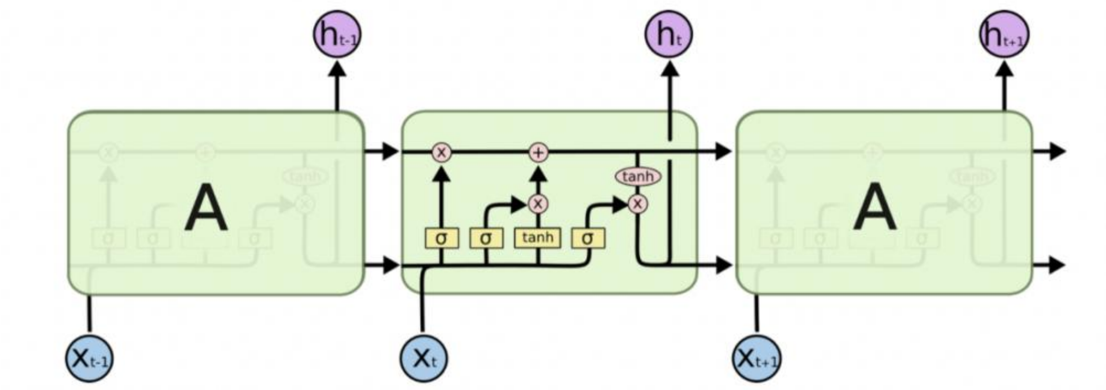


**Figure 1.** Structural diagram of LSTM model.

Yang et al. put forward an improved ELF-LSTM neural network model that can enhance long-term characteristics [15]. The diagram of the network structure is be shown in Figure 2. It is based on the LSTM model, but in order to make the model more extensible, it adds two fully connected layers for combining features and one for improving prediction accuracy. Yang et al. used the smart card data of a station in Chongqing Municipality as a data set for calculation, combined with short-term characteristics to predict the next hour outbound destination (OD) traffic, and judged. In addition to the mean absolute error (MAE), the model performance measure also includes another measure, the RMSE.

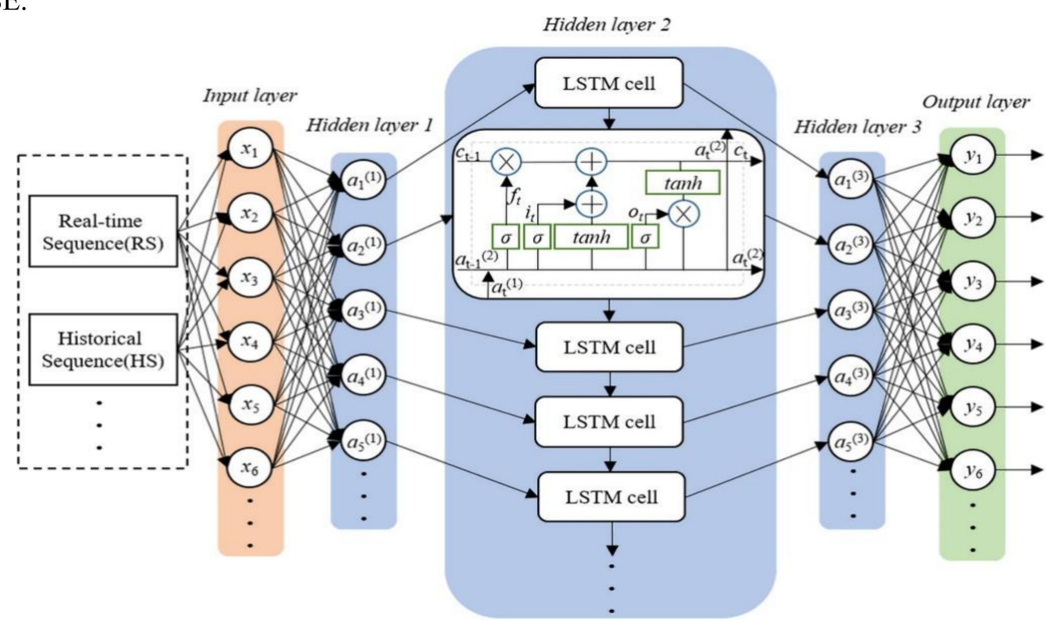


**Figure 2.** ELF-LSTM model structure diagram.

### 2.2. CNN

There is a very representative algorithm, belongs to a feedforward neural network. It is a deep structure of a deep learning algorithm and convolutional calculation, that is, convolutional neural network (CNN). Convolutional neural networks have another name besides their own, which is called "SIANN". because of its presentation learning ability, it can classify input information shift invariant according to its hierarchy. As shown in Figure 3, given a training image, a CNN will extract its features by combining the activation layer, full connection layer, pooling layer and convolution layer. The learned feature will be fed into a classification layer or other function layer to predict the final results.

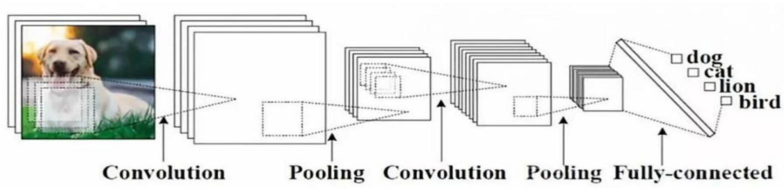


**Figure 3.** Structural diagram of CNN model.

### 2.3. GCN

There are many methods of graph convolutional neural network, including spectral domain based on method and spatial domain based on method. Here another Angle can be found to process the graph signal and define the graph convolution by introducing filters. This approach has a separate technical name, also known as spectral domain based on approach. Removing noise from graph signal is one interpretation of graph convolution operation. Graph convolution is represented by aggregation of feature information from neighbours, which is called spatial domain based on method.

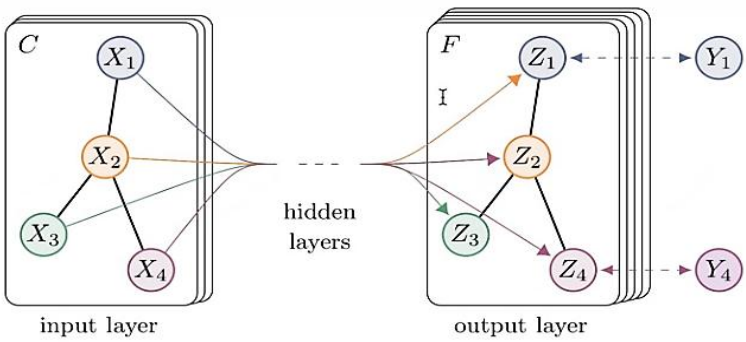


**Figure 4.** Structural diagram of GCN model.

## 3. Experiment

### 3.1. Evaluation index

The RMSE represents the average size of the deviation and it will be used to evaluate model performance, which can be used to measure stability. In addition, MAE is also selected as an evaluation index. In addition to being useful in measuring the risk index of a model, it is not surprising that this indicator can also be used to measure model accuracy. The RMSE and MAE can be calculated by the formulation (1) and (2), as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{I=1}^{n}\left(Y_i - f(x_i)\right)^2} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \tilde{y}_t| \tag{2}$$

### 3.2. Performance comparison of different models

*3.2.1. Performance of LSTM.* Table 1 shows when the time lag is one week, the experimental conclusion proves that the ELF-LSTM model has better predictive performance. Compared with the LSTM model, it could accurately catch the long-term dependence of the data, thus obtaining more accurate results.

DOI: 10.54254/2755-2721/9/20230077

**Table 1.** Comparison of LSTM and ELF-LSTM.

| Model | RMSE | MAE | Time, min |
|---|---|---|---|
| LSTM | 0.297 | 0.209 | 56 |
| ELF-LSTM | 0.269 | 0.192 | 7 |

*3.2.2. Performance of CNN.* In the CNN model, the following parameters are used as initial defaults for training 15 epochs. The default training results can be used as a benchmark, and then the network can be fine-tuned to observe the change of the results. The learning rate was 0.001. The batch size is 128. Re LU was selected for activation function. The loss function is defined as the cross entropy function. SGD mode optimization. The experimental input is a graph data composed of validation data, and the nodes are validation events and event-related attribute nodes. Such as IP, Device ID, UA and other nodes. The output of the experiment is to classify the event nodes as normal or abnormal.

As shown in Table 2, the effect of multi-layer convolutional networks is more optimized than that of single-layer convolutional networks, and the final accuracy of the former is better than that of the latter. This is because the depth of the network deepens, so that the fitting ability of the model is strengthened, and the feature extraction and compression are also better. This is the advantage of multi-layer network.

**Table 2.** Loss comparison between single-layer and multi-layer networks.

| *Model* | train loss | test loss | Accuracy |
|---|---|---|---|
| Single-layer network | 0.1444 | 0.1347 | 96.09 % |
| Multi-layer network | 0.0847 | 0.0748 | 97.57 % |

*3.2.3. Performance of GCN.* For the GCN, the Adam optimizer is used to train the model, and the learning *rate* was set to 0.001, whose results can be found in Figure 5. Compared with the GBDT, which is the most popular shallow classifier at present and can only learn feature information as the benchmark, grid Search searches for super parameters. It is obvious that GCN model has relatively small attenuation accuracy, while the attenuation of GBDT is very serious. It is obviously that the GCN model has good man-machine discrimination effect and robustness. To sum up, the structural information learned by GCN is not only effective in man-machine discrimination, but also has better robustness.

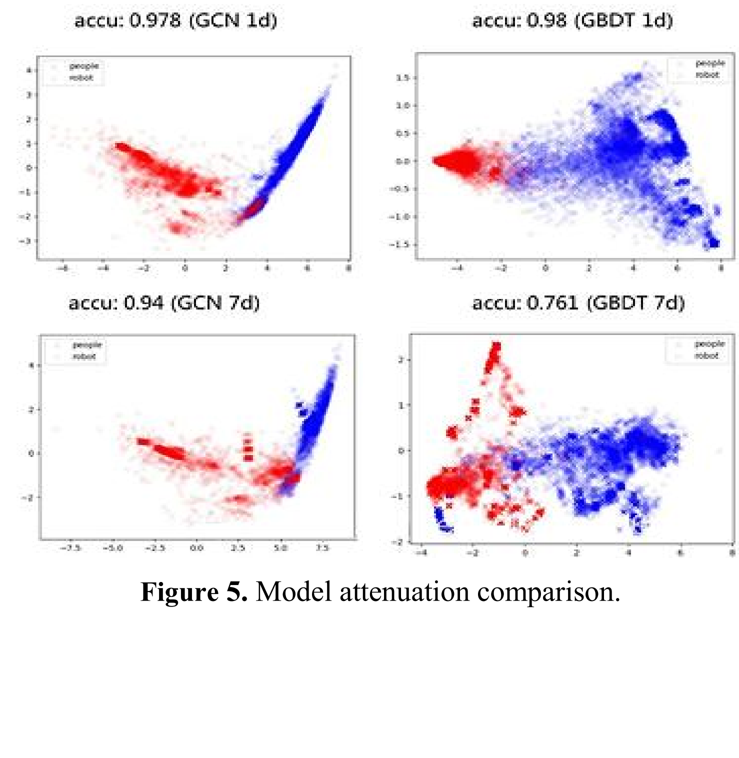


**Figure 5.** Model attenuation comparison.

## 4. Discussion

This part analyzes and compares the advantages and disadvantages of three different models, LSTM, CNN, and GCN. LSTM model can effectively improve the long-term dependence problem in RNN. In addition, it can preserve important features through various gate functions, which can be very appreciably reduced in long sequence problems where gradient disappearance or gradient explosion may occur. It can't handle 1000 orders of magnitude or longer, but it can handle at least 100 orders of magnitude. The training efficiency of RNN is much higher than that of RNN, because it is not better than RNN under the same computational power due to its relatively complex internal structure. There are four fully connected layers (MLPS) per existence, which means that LSTM units exist. If the created LSTM have a large time span and a large network depth, the calculation will be time-consuming.

The CNN model can process high dimensional data without pressure while sharing the convolutional core. Its convolution layer features automatic feature extraction and training with specific tasks, which can be used to learn the global optimal parameters. The indispensable convolution kernel (filter) plays an important role in extracting the required features by convolution. When the network layer is too deep, the BP propagation parameter modification method will cause the parameter change near the input layer is not obvious. In the case of gradient descent algorithm, the training result can be changed to a possible result with a high probability, that is, convergence to the local minimum rather than the global minimum. A module called the convergence layer can cause a lot of valuable information to be lost, and can lead to a very small loss of correlation between the local and the whole.

If researchers want to describe the learning goal of the GCN model, they can use a noun here called aggregator. It is different from the representation of a single node. This way of thinking has considerable advantages, it can improve the flexibility and malleability of the model. In addition, mass training of this model can improve its convergence speed. GCN itself has many drawbacks, such as poor flexibility, transitivity, and scalability. What’s more, validation sets were used here early on to help improve the accuracy of the model, which contradicted the original intention, namely semi-supervision.

## 5. Conclusion

By combing and comparing the results of previous studies, this paper introduces three deep learning models short - term passenger flow forecast field is introduced in detail, including LSTM, CNN and GCN. The main contribution of this paper is a comprehensive discussion of the three main methods of deep learning, which will help to better learn and apply this field in the future. It is also found that no matter the single model such as LSTM, CNN or GCN, the model training time is long and the model performance is poor. Therefore, the deep learning framework of composite models has become more popular in recent years. For example, the Spatial-temporal Graph Convolutional Network (ST-GCN) model is considered as a deep learning framework with good performance, but it still has some problems, for example, the model is relatively complex, which needs a lot of time and computing resources. With the continuous development of short-time passenger flow prediction algorithms, algorithms represented by deep learning have entered a new stage of development. Therefore, in the present situation, the passenger flow characteristics, model performance, practical value and model complexity of rail transit network should be fully considered and weighed when building the model, so as to achieve a better prediction effect. Finally, in the field of passenger flow forecasting, further research is needed to propose a new model with better effect, so as to facilitate the in-depth study of this subject and the construction of intelligent transportation system.

## References

[1] Xiong Z, Zheng J, Song D, Zhong S and Huang Q. (2019). Passenger Flow Prediction of Urban Rail Transit Based on Deep Learning Methods. Smart Cities,2(3):371-387

[2] Bai Y, Sun ZZ, Zeng B, Deng J and Li C. (2017). A multi-pattern deep fusion model for short-term bus passenger flow forecasting. Applied Soft Computing, Volume 58, Pages 669-680, ISSN 1568-4946
[3] Feng S and Cai G. (2016). Passenger flow forecast of metro station based on the ARIMA model. Proc. of the 2015 Int. Conf. on Electrical and Information Technologies for Rail Transportation, Berlin, Heidelberg, vol. 378, pp. 463– 470
[4] Kumar SV and Vanajakshi L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. Eur. Transp. Res. Rev. 7, 21.
[5] Jiao P, Li R, Sun, T, Hou, Z and Ibrahim A. (2016). Three revised Kalman filtering models for short-term rail transit passenger flow prediction. Math. Probl. Eng,9717582.
[6] Ye Y, Chen L and Xue F. (2019). Passenger Flow Prediction in Bus Transportation System using ARIMA Models with Big Data. 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Guilin, China, pp. 436-443
[7] Giraka O and Selvaraj VK. (2019). Short-term prediction of intersection turning volume using seasonal ARIMA model. Transportation Letters,12:7, 483-490, DOI: 10.1080/19427867.
[8] Zhang H, Seshadri R, Prakash AA, Pereira F, Antoniou C and Ben-Akiva ME. (2019). Improved Calibration Method for Dynamic Traffic Assignment Models: Constrained Extended Kalman Filter. Transportation Research Record, 2667(1), 142–153.
[9] Tian C and Chan WK(V). (2021). Spatial-temporal attention wavenet: A deep learning framework for traffic prediction considering spatial-temporal dpendencies. IET Intell Transp Syst,15: 549– 561.
[10] Guancen L, Aijing L and Danlei G. (2022). Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. Volume 608, Pages 517-531, I SSN 0020-0255
[11] Gallo M, De Luca G, D'Acierno L and Botte M. (2019). Artificial Neural Networks for Forecasting Passenger Flows on Metro Lines. Sensors.19(15):3424
[12] Ouyang Q, Lv Y, Ma J and Li J. (2020). An LSTM-Based Method Considering History and Real-Time Data for Passenger Flow Prediction. Appl. Sci. 10, 3788.
[13] Yang YQ, Lin J and Zheng YB. (2022). Short-Time Traffic Forecasting in Tourist Service Areas Based on a CNN and GRU Neural Network. Appl. Sci.12, 9114.
[14] Zheng H, Li X, Li Y, Yan Z and Li T. (2022). GCN-GAN: Integrating Graph Convolutional Network and Generative Adversarial Network for Traffic Flow Prediction. In IEEE Access, vol. 10, pp. 94051-94062, doi: 10.1109/ACCESS.2022.3204036.
[15] Yang D, Chen K, Yang M and Zhao X. (2019). Urban rail transit passenger flow forecast based on LSTM with enhanced long-term features. IET Intell. Transp. Syst., 13: 1475-1482.