Passenger Flow Prediction and Risk Early Warning for Metro Cross-section based on Transportation Big Data

Jiaxu Li[1], Meihan Ye[2] and Dazhi Zhao[3, *]

[1]School of Information and Computer Engineering, Northeast Forestry University, Harbin, 150040, China

[2]School of Information Science and Engineering, China University of Petroleum, Beijing, 102249, China

[3]School of Traffic and Transportation Engineering, Central South University, Changsha, 410075, China

* Corresponding Author Email: 8212200228@csu.edu.cn

Abstract. With the increasing pace of urbanization and the diversification of urban transportation modes, the demand for urban transportation in people's life is also increasing. For the purpose of relieving urban traffic, cities have made great efforts to develop public transport, especially metro transport. This paper will use the subway swipe card data of Changsha city from March 1-15, 2023, and optimize the LSTM model based on particle swarm algorithm, to predict the metro cross-sectional passenger volume. Based on the data and methods, this paper investigates the short-time cross-sectional passenger flow forecasting of the metro using the variables of initial passenger flow, inbound passenger volume and outbound passenger volume as references. It is found that the combined short-time passenger flow prediction model (PSO-LSTM) has good prediction effect. Based on the standard LSTM model, the PSO algorithm optimizes the learning rate, the number of iterations and the number of hidden neurons in the LSTM parameters to construct the PSO-LSTM model. The video recognition technology more accurately reflects the traffic congestion and makes early warning. The short-time cross-sectional passenger flow prediction of subway will help to improve people's living standard and save transportation time cost, which is of practical significance to urban rail operating companies.

Keywords: Metro, LSTM, PSO-LSTM.

1. Introduction

With the increasing pace of urbanization and the diversification of urban transportation modes, the demand for urban transportation in people's life is increasing, and the scale and complexity of urban transportation data are increasing. Efficient storage, processing, and analysis of data has become an important research direction [1]. In order to alleviate urban congestion and provide convenience to people's life, many cities are vigorously developing rail transportation. Among them, the short-time cross-sectional passenger flow forecast of metro traffic can better reflect the demand for rail in daily life. How to effectively predict the cross-sectional passenger flow of subway transportation using these data is of great significance to develop reasonable countermeasures and improve urban services [2, 3].

The operational efficiency of the metro transportation system is not only related to the degree of infrastructure construction and the level of technical equipment, but also influenced by the supervision of the metro transportation operation management. The short-time cross-sectional passenger flow of metro transportation reflects the demand generation of rail transportation. Compared with the outbound and inbound passenger flow, the control of short-time sectional capacity flow is more manipulable for rail transit operation and management [4].

When the cross-sectional passenger flow is counted over a longer period of time, the traffic forecast results are not timely, although there is also a more obvious pattern of peaks and flats, which lacks obvious timeliness. By shortening the time interval to 5~15 min, the non-linear and dynamic connection between the short-time passenger flow and the influencing factors will be more obvious

and the results will be more timely. Only on the basis of timely, accurate and high-precision forecasting of rail transit short-term passenger flow can we better induce and control metro traffic and improve the service level of metro traffic.

The field of short-term traffic forecasting has been the focus of research by experts and scholars around the world and has been significantly effective after decades of efforts [5]. Many mature models have been applied by domestic and foreign scholars, such as traditional moving average method, k-Nearest Neighbor algorithm (KNN), Recurrent Neural Network (RNN) and Convolutional Neural Networks (CNN). Subsequently, it was found that Long Short-Term Memory networks (LSTM) can overcome the "forgetting" phenomenon and has a powerful "memory" capability. In order to improve the characteristics of the LSTM model, which can only learn information unidirectionally for time series, the Bi-directional Long Short-Term Memory (BiLSTM) is introduced. Li and Wu et al. predicted short-term traffic at urban bus stops based on LSTM [6]. Mobarak used LSTM, Gate Recurrent Unit (GRU) and RNN for electricity load forecasting [7].

In short-time cross-sectional traffic forecasting, the forecasting process needs to provide accurate point forecasts and reliable analysis of impending uncertainties [8]. In the forecasting process, to address the drawbacks of a single model, research has started to combine multiple deep learning models together to form an overall model. Integrated learning methods such as weighted averaging and voting methods can be used, or the output of different models can be used as input to construct a new model for prediction by methods such as stacking [9].

This paper takes metro passenger entry and exit data in urban rail transit as an reference, and calculate the cross-sectional passenger flow of a certain train at a certain station in a certain time period by counting the time of individual's entry and exit. Subsequently, based on the historical cross-sectional passenger flow data, a particle swarm optimization algorithm combined with an LSTM model (PSO-LSTM) is used to predict future cross-sectional passenger flows [10-12]. At the same time, when making forecasts, the characteristics of metro traffic and multiple influencing factors should be taken into account to accurately grasp the changes in passenger flow and improve the scientific nature of the forecasts. Accurate forecasts can provide a reference for the scheduling of metro operating schedules and improve the operational efficiency of the metro system. Based on the forecast results, passengers can choose the right route and time to travel, thus improving the convenience and safety of their journeys. Considering that the number of people in different compartments of each train will vary to a certain extent, video recognition technology is also required. At the same time, the onboard camera is used to identify the number of people in the vehicle and to develop relevant risk indicator functions for risk assessment, issuing a risk warning when the number of people in the carriage reaches a threshold value. This helps platform staff to regulate passenger flow in a timely manner and direct passengers to queue in areas of the car where the amount of people is low. This will further improve the service level of metro transport.

2. Methods

2.1. Data Source

This paper uses Changsha metro card swipe data from March 1-15, 2023 for the study, with objective and accurate data sources.

2.2. Indicator Selection and Description

Table 1 shows the full names, data types and interpretations of the six variables used in the study. The smart card data can well record the information of passengers entering and leaving the station with very high accuracy, thus making the data of the study more authoritative. According to the characteristics of the data set and the actual situation of riding the metro shown by the data, it can be judged whether the initial station number ENTRY_STATION_ID and the final arrival station number STATION_ID are equal to determine whether they are staff.

Table 1. Name and explanation of variables

Full Name	Data type	Explanation
ENTRY_LINE_ID	INT	Initial ride route
ENTRY_STATION_ID	INT	Initial station number
LINE_ID	INT	Final arrival route
STATION_ID	INT	Terminal station number
ENTRY_DATETIME	Typedef	Entry time
DEAL_DATETIME	Typedef	Outbound time

2.3. Introduction to the Method

2.3.1 LSTM neural network

LSTM is a special recurrent neural network. LSTM avoids the problem of vanishing gradients and gradient explosions generated by traditional recurrent neural networks and is effective in learning long-term dependencies. The LSTM network structure, as shown in Figure 1 by Visio.

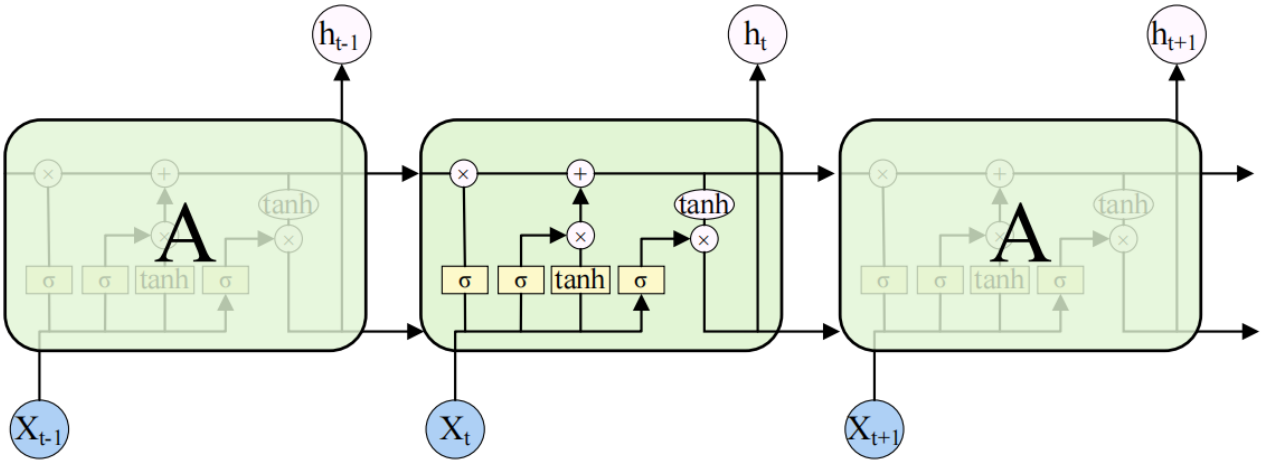


Fig. 1 LSTM network topology

A means that three cells have the same cell structure. structure. Each cell is composed of 4 main elements: input gate, forget gate, output gate, and cell state.

Through the design of the memory unit and the doorway mechanism, LSTM network can combine various input features such as historical traffic flow and weather data to better predict the passenger flow. LSTM also has strong dynamic adaptability and can adjust the network structure in real time according to real-time data. Finally, LSTM also has high flexibility and can output different continuous values (passenger flow value) or discrete values (passenger flow level classification) according to the actual demand.

In conclusion, LSTM can be used as a method to effectively predict subway passenger flow.

2.3.2 BiLSTM neural networ

In order to improve the LSTM model, which can only learn in one direction for time series, the BiLSTM was introduced. The BiLSTM is an LSTM with an additional layer of opposite orientation on top of the LSTM, and the two layers of LSTM have separate hidden layers. The internal structure of the BiLSTM model is shown in Figure 2 by Visio.

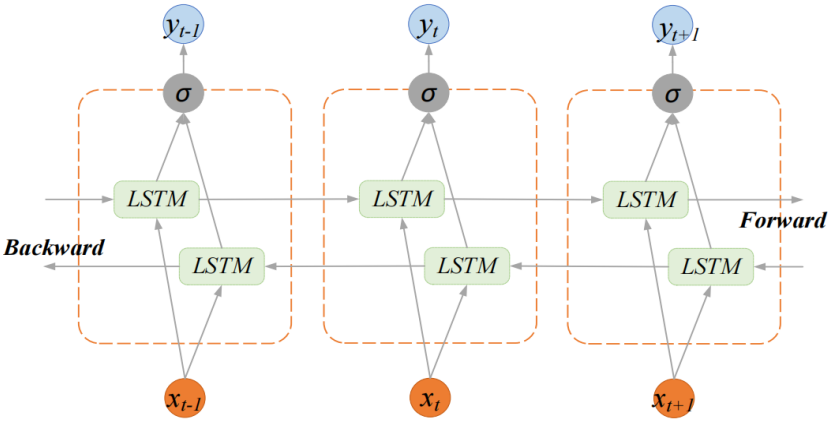


Fig. 2 BiLSTM network topology

Forward LSTM refers to the data entering the LSTM layer is in the input data direction and the reverse LSTM. It refers to the data entering the LSTM layer is in the reverse direction and the data input direction is the main difference between the two. Only the final output result can be obtained by linear integrating the LSTM of the two directions.

In terms of short-term subway passenger traffic prediction, BiLSTM can better combine the relationship of time series than LSTM. LSTM has a strong sequential dependence on the underlying data, but BiLSTM can combine the forward and reverse information flow to improve the robustness of the model. Finally, the positive and negative propagation of BiLSTM can be carried out in parallel, which can make full use of the computer capability and computing efficiency, and can efficiently train the model in the subway stations with more passenger flow data.

In conclusion, BiLSTM can also be used as a way to predict subway traffic flow.

2.3.3 PSO-LSTM model

Single-model prediction has the disadvantage of being impossible to optimize, this paper combines the particle swarm algorithm (PSO) with the LSTM and constructs the PSO-LSTM prediction model. Compared with other biological intelligent evolutionary algorithms, the biggest advantage of PSO is the simple algorithm design and fast convergence.

Individuals in a particle swarm are treated by PSO as particles in a multidimensional search space, and each particle has a fitness value determined by the objective function. PSO iterates to find the individual optimal position and the global optimal position for each particle. During this process, each particle continuously updates its position and velocity until the optimal search condition is reached. The PSO-LSTM model is able to determine the optimal hyperparameters quickly and accurately according to the characteristics of the input data, achieving an effective combination of the network structure of the LSTM and the characteristics of the data.

PSO-LSTM has better parameter optimization ability compared to LSTM. PSO-LSTM is more efficient and achieves a better parameter scheme than traditional random search. Finally, PSO-LSTM can continue to be combined with some methods to construct more complex and accurate prediction models.

In summary, good experimental results can be obtained by using PSO-LSTM to predict the short-term intermittent passenger flow. Meanwhile, it can be combined with other algorithms to construct a combined prediction model in the future, which has certain development space.

3. Results and Discussion

3.1. Descriptive Analysis

The historical cross-sectional passenger flow statistics are mainly obtained from the raw swipe data and the raw interchange data. The data of each line is gradually superimposed to obtain the cross-sectional flow of each section. Taking Changsha South Railway Station to Pingyang as an example,

the flow between stations is solved, and the flow between adjacent stations is calculated and plotted for each hour. As shown in Figure 3.

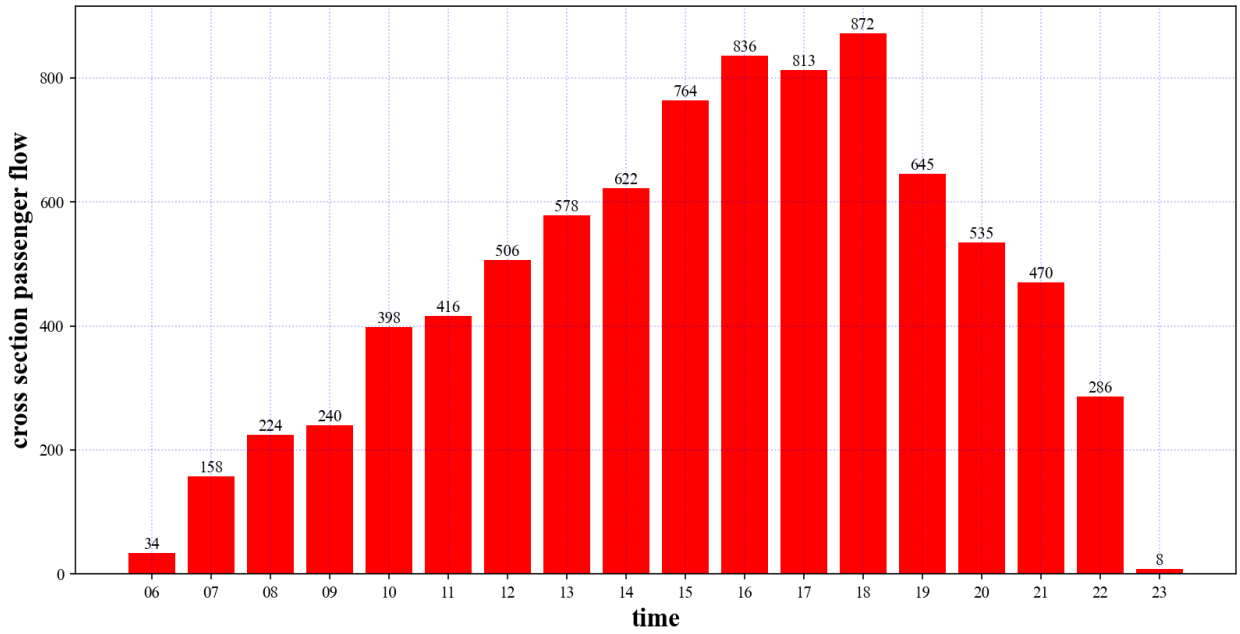


Fig. 3 Changsha South railway station to Pingyang interface flow chart

Assume that the interchange time is at least 3 min, the first train departs at 6:30, there may be more than one train on a route at the same time. Considering the train matching problem, set random_rate=0.05 and the data is processed as in Figure 4.

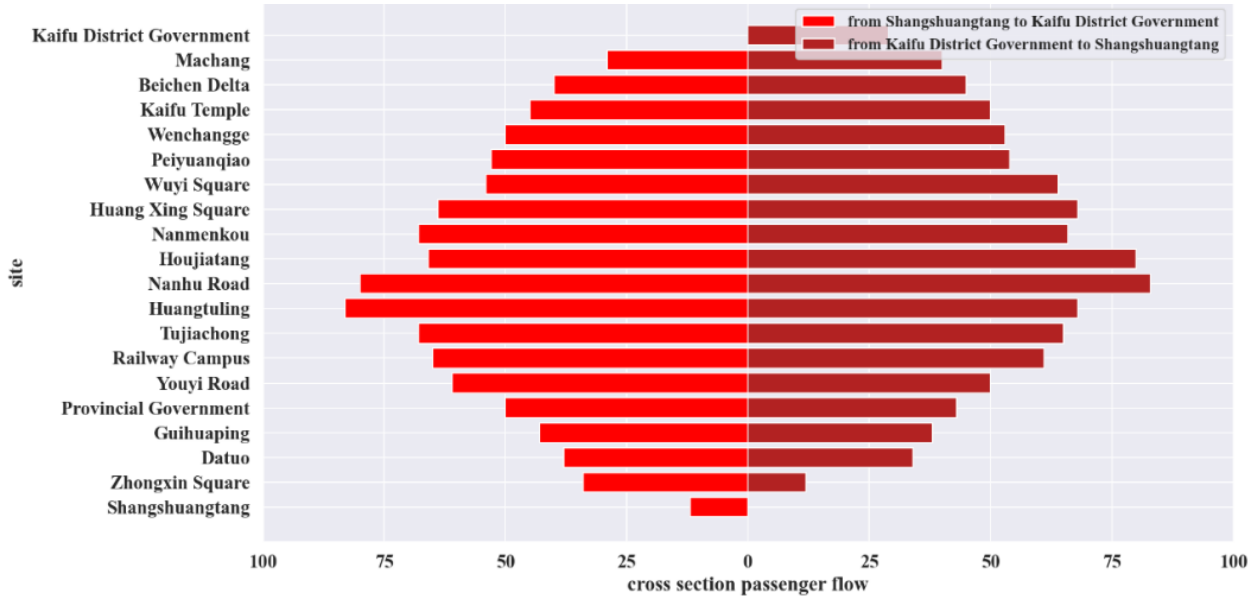


Fig. 4 Passenger flow of all trains from Central South University to Fubu River subway station

First of all, the credit card data from Central South University was used to calculate the passenger flow from Central South University to FuBuhe subway station. The results are shown in Figure 5. The shade of color represents the size of the passenger flow. This is a working day, and it can be found to increase gradually from 7 am, peak at around 8 am, and then start to decrease. During the whole day, there were three peak passenger flow, respectively 8:00,11:00 and 18:00, which was very consistent with the actual situation.

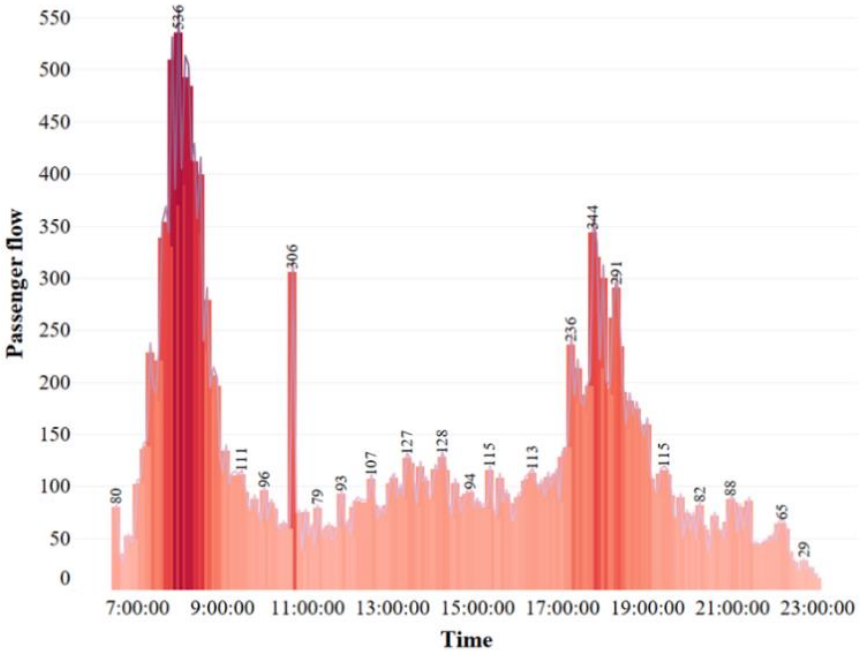


Fig. 5 Metro Line 1 6:30 train interface traffic

In addition to counting the section passenger flow of the adjacent two stations, the passenger flow of a train number in each station is also counted. The results are shown in Figure 6. The figure of the vertical coordinate in the figure represents different stations, the abscissa represents the passenger flow, and the deeper the color represents, the greater the passenger flow. It can be found that the first subway in the morning has the largest passenger flow at Tujiacong-Huangtuling station.

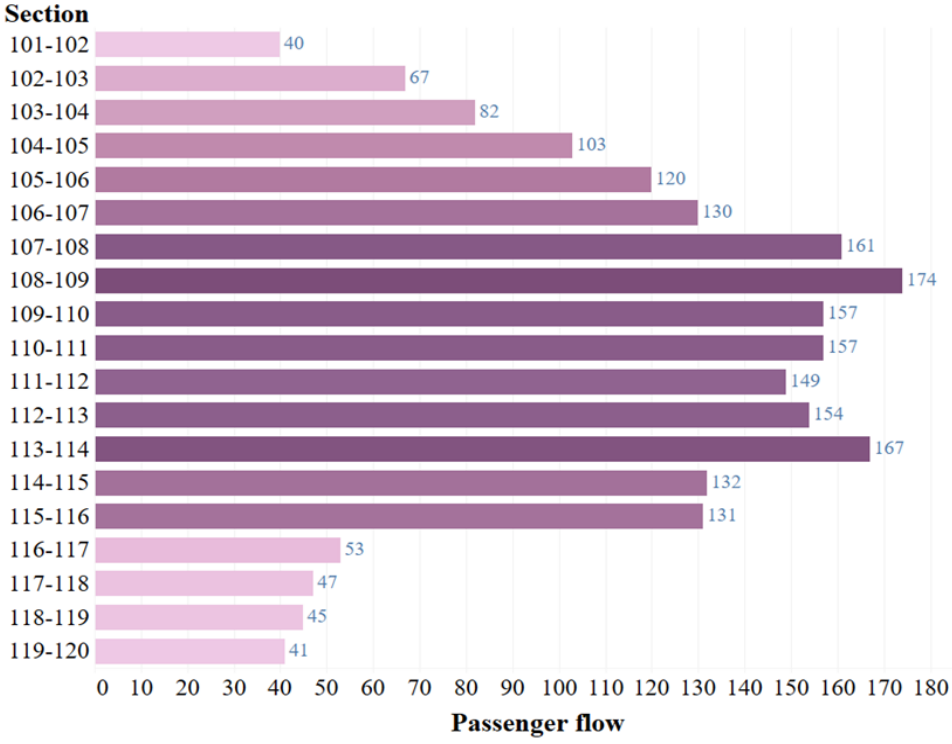


Fig. 6 Passenger flow of the first subway in the morning from Shangshuangtang to Kaifu District Government of Metro Line 1

In-depth analysis by counting the section passenger flow of two adjacent stations, the results are shown in Figure 7. Take Shangshuangtang-Zhongxin Square as an example. Passenger flow between the two stations was calculated. It has certain differences and periodicity in different times.

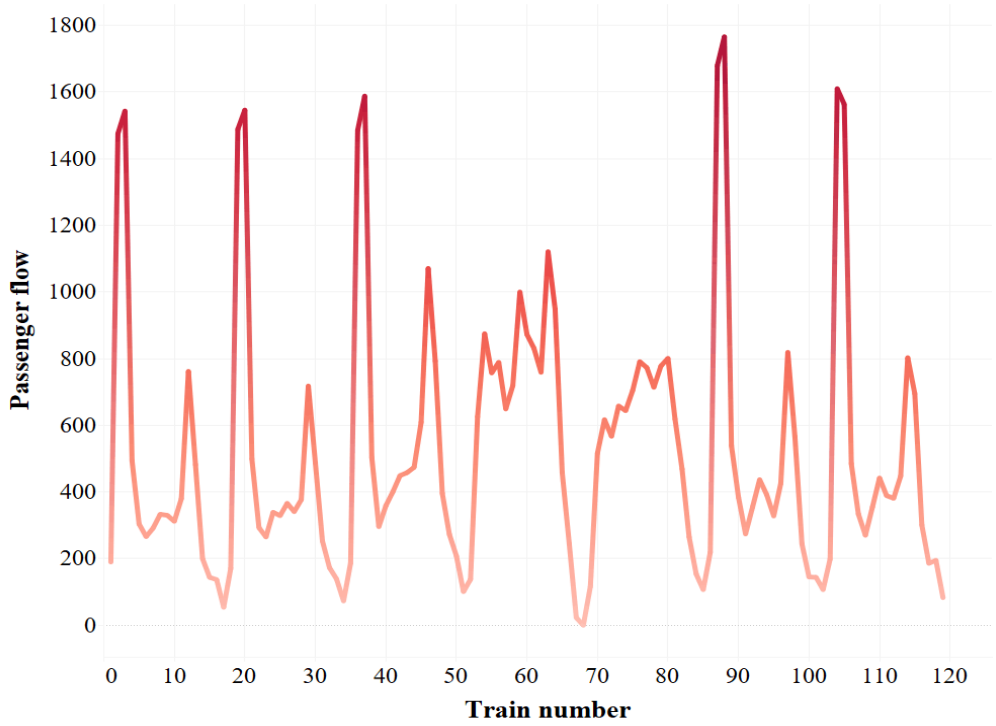


Fig. 7 Passenger flow from Shangshuangtang to Zhongxin Square in seven days is calculated by hour

After the statistical analysis of the data, the particle swarm algorithm and LSTM combined prediction model were used to forecast the passenger flow for 17 hours. The results are shown in Figure 8. By comparing the actual results, it is found that the prediction effect is good, and that the combined model can improve the prediction accuracy to a certain extent.

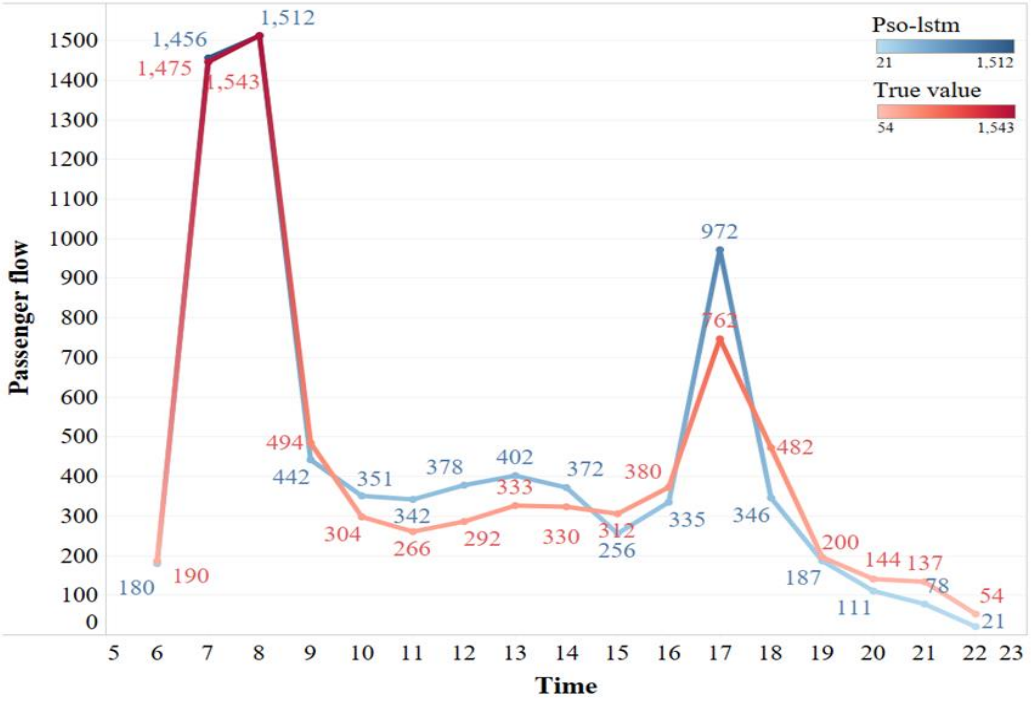


Fig. 8 17-hour passenger flow forecast using particle swarm optimization algorithm and LSTM

The risk matrix method combined with video recognition of the number of people can be used to determine whether the interior of an underground carriage is loaded. The following is a simple implementation idea.

First, passenger flow data are collected and recorded for different time periods and carriages, including peak and off-peak hours. These data can be used to build a risk matrix. Second, the passenger flow is divided into different levels or intervals based on the historical data, such as low risk, medium risk and high risk. The range of passenger traffic for each level is determined based on actual conditions. Then, the amount of people inside the subway cars is detected in real time using video recognition techniques, such as deep learning-based target detection algorithms.

Next, the number of people detected in real time is compared with the risk matrix. Based on the level of the number of people, the current passenger flow risk level of the carriage is determined. For example, if the real-time headcount is in the high risk range, it means that the current passenger flow has reached or is close to the load and measures need to be taken to relieve congestion. Finally, based on the risk level of passenger flow, an early warning message can be sent to passengers or metro managers to alert them of the current crowding situation in the carriage. In addition, corresponding measures such as increasing the frequency of trains and directing passengers to other cars can be taken to relieve the passenger flow pressure.

3.2. Inferential Analysis

Through the processing of passenger credit card data from March 1 to March 7, the passenger flow of Shangshuangtang-CITIC Square in Metro Line 1 was obtained. LSTM, BILSTM and PSO-LSTM were selected to predict the passenger flow of Shangshuangtang-CITIC Square from 7 to 8 PM on March 7, and compared with the actual value in Figure 9. The forecast results of PSO-LSTM are closer to the true value, followed by BILSTM, and the LSTM has the worst prediction effect.

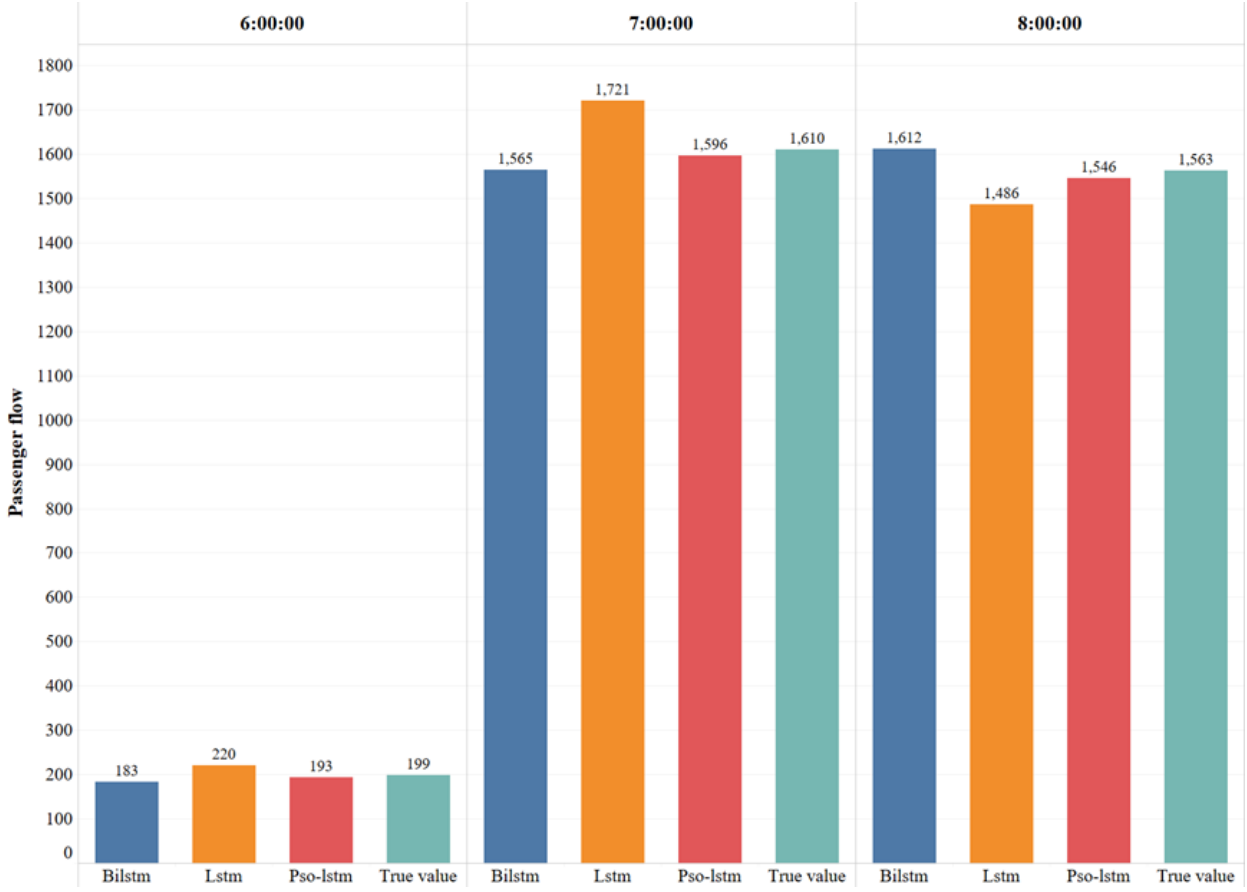


Fig. 9 Comparison chart of forecast and reality

Subsequently, the PSO-LSTM prediction model was repeatedly learned for 200 times, and the mean absolute error (MAE) results are shown in Figure 10.

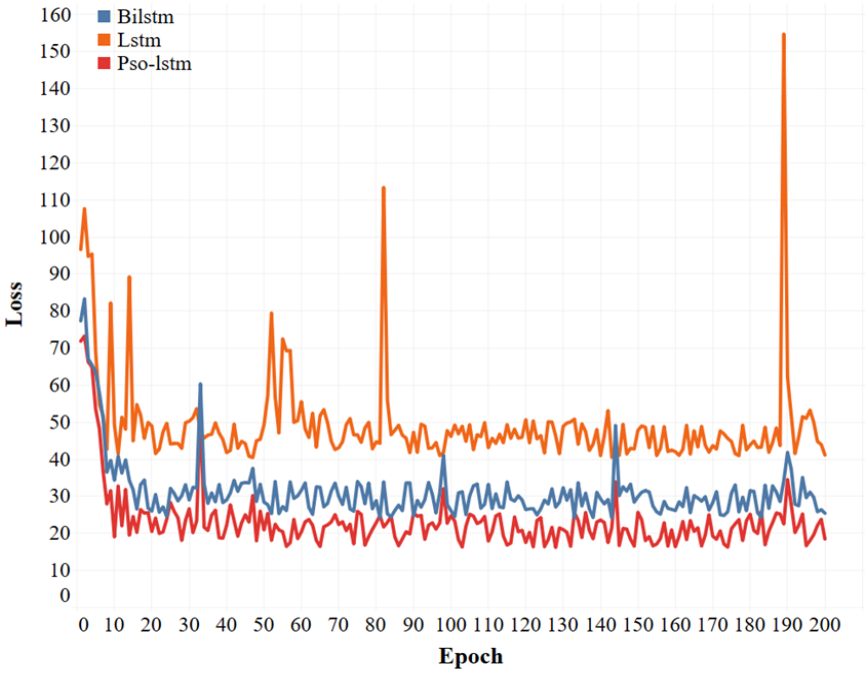


Fig. 10 Error decline chart

It can be found that the curve drops to the lowest point when overlapping the number of learning times to 150 times, and then basically remains unchanged, with the minimum MAE of 0.00006, so the combined prediction model can achieve good results, as figure 11 shows.

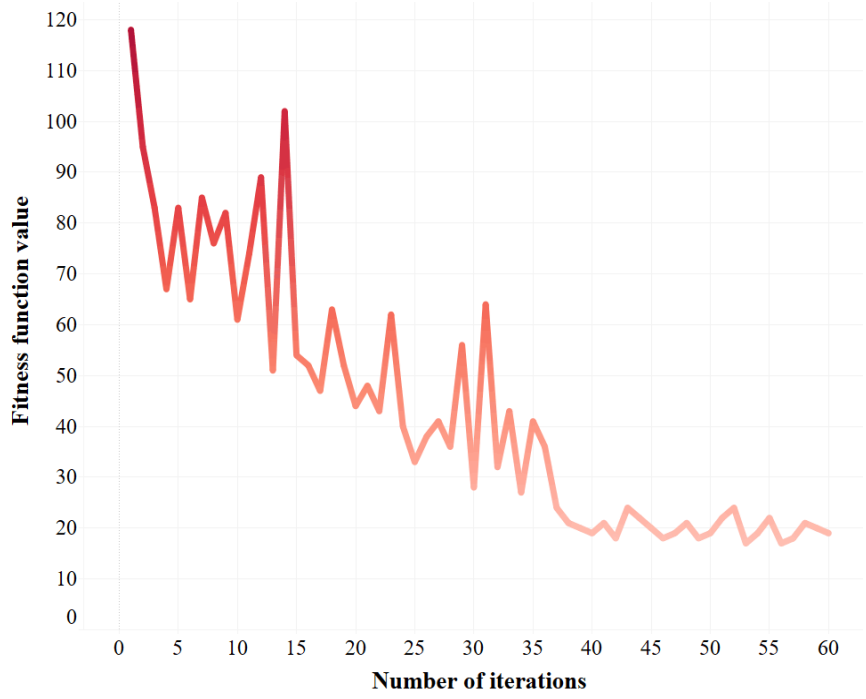


Fig. 11 Fitness function curve of particle swarm optimization

The horizontal coordinate is the number of iterations, and the vertical coordinate is the fitness function value. The fitness function uses the average error, and the function is:

$$f = \frac{\sum_{i=1}^{n} |\hat{y} - y|}{n} \tag{1}$$

Through the above exploration of the passenger volume statistics and prediction of the card data, the passenger flow over a period of time is intuitively shown. First, the results are in line with the changing trend of daily passenger flow, with more passenger flow on working days than on non-

working days. Meanwhile, the change trend of passenger volume in different periods of the same day is consistent with the travel situation of residents. Finally, the prediction effect is good, which further verifies the advantages of the combined model over the single model, and paves the way for the subsequent in-depth research.

4. Conclusion

The short-time passenger flow prediction combination model (PSO-LSTM) has good effectiveness. The PSO-LSTM model is constructed by selecting the standard LSTM model and optimizing the number of hidden neurons, learning rate and number of iterations in the LSTM parameters using particle swarm arithmetic. The results show that compared with LSTM and BiLSTM, the combined model has higher prediction accuracy.

However, at the same time, due to the few days of the sample set, more accurate model results cannot be obtained, so that the long-term prediction results do not conform to the future actual situation. At the same time, in the future model prediction, some contingency factors should be considered, such as holidays, ticket prices, other transportation modes of influence, which will make the results more accurate and has certain practical significance for urban rail operation companies.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Li Taizeng, Cheng Xinyue, Li Aidi, Wei Boan. Design and implementation of a multi-source traffic data visualization system based on distributed architecture. Urban Construction Theory Research (Electronic Edition), 2023.

[2] Liu Yaohong. Research on short-time inbound passenger volume forecast of track transportation based on combined model. Chongqing Jiaotong University, 2022.

[3] Zhang Kaiwen, He Yong. Passenger volume prediction of urban orbital transportation stations based on EMD-SSA-LSTM model. Journal of Wuhan University of Technology (Transportation Science and Engineering Edition), 2023, 1-9.

[4] Guo H F, Fang L J, Yu L. Fuzzy Kalman filter-based short-time traffic flow prediction method. Journal of Zhejiang University of Technology, 2013, 41: 218-221.

[5] Cui L. Cheng. Research on traffic flow prediction method of urban road network based on multi-section information. Dalian Maritime University, 2012.

[6] Gao-sheng Li, Ling Peng, Xiang Li, Tong Wu. Short-term traffic forecast of urban bus stations based on long short-term memory. Journal of Highway and Transportation Research and Development (English Edition), 2019, 13(2).

[7] Abumohsen Mobarak, Owda Amani Yousef, Owda Majdi. Electrical load forecasting using LSTM, GRU, and RNN algorithms. Energies, 2023, 16(5).

[8] Stefano Cabras, Sun He. A Bayesian spatial-temporal model for predicting passengers occupancy at Beijing Metro. Spatial Statistics, 2023.

[9] Kadari R, Zhang Y, Zhang W, et al. CCG super-tagging via Bidirectional LSTM-CRF neural architecture. Neurocomputing, 2018, 283: 31-37.

[10] Gajamannage Kelum, Park Yonggi, Jayathilake Dilhani, et al. Real-time forecasting of time series in financial markets using sequentially trained dual-LSTMs. Expert Systems With Applications, 2023, 223.

[11] Huo Fengcai, Chen Yi, Ren Weijian, et al. Prediction of reservoir key parameters in 'sweet spot' on the basis of particle swarm optimization to TCN-LSTM network. Journal of Petroleum Science and Engineering, 2022, 214.

[12] Kewen Li, Lu Liu, Jiannan Zhai, et al. The improved grey model based on particle swarm optimization algorithm for time series prediction. Engineering Applications of Artificial Intelligence, 2016, 55: 285-291.