

Article

Forecasting Short-Term Passenger Flow of Subway Stations Based on the Temporal Pattern Attention Mechanism and the Long Short-Term Memory Network

Lingxiang Wei ^{1,2} , Dongjun Guo ^{1,*}, Zhilong Chen ¹, Jincheng Yang ²  and Tianliu Feng ^{2,3}

¹ Research Center for Underground Space, Army Engineering University of PLA, Nanjing 210007, China

² School of Materials Science and Engineering, Yancheng Institute of Technology, Yancheng 224051, China

³ School of Rail Transportation, Soochow University, Suzhou 215031, China

* Correspondence: guo_dongjun@163.com

Abstract: Rational use of urban underground space (UUS) and public transportation transfer underground can solve urban traffic problems. Accurate short-term prediction of passenger flow can ensure the efficient, safe, and comfortable operation of subway stations. However, complex and nonlinear interdependencies between time steps and time series complicate such predictions. This study considered temporal patterns across multiple time steps and selected relevant information on short-term passenger flow for prediction. A hybrid model based on the temporal pattern attention (TPA) mechanism and the long short-term memory (LSTM) network was developed (i.e., TPA-LSTM) for predicting the future number of passengers in subway stations. The TPA mechanism focuses on the hidden layer output values of different time steps in history and of the current time as well as correlates these output values to improve the accuracy of the model. The card swiping data from the Hangzhou Metro automatic fare collection system in China were used for verification and analysis. This model was compared with a convolutional neural network (CNN), LSTM, and CNN-LSTM. The results showed that the TPA-LSTM outperformed the other models with good applicability and accuracy. This study provides a theoretical basis for the pre-allocation of subway resources to avoid subway station crowding and stampede accidents.

Keywords: urban underground space (UUS); subway station; short-term passenger flow forecast; temporal pattern attention (TPA); long short-term memory network (LSTM)



Citation: Wei, L.; Guo, D.; Chen, Z.; Yang, J.; Feng, T. Forecasting Short-Term Passenger Flow of Subway Stations Based on the Temporal Pattern Attention Mechanism and the Long Short-Term Memory Network. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 25. <https://doi.org/10.3390/ijgi12010025>

Academic Editor: Wolfgang Kainz

Received: 8 November 2022

Revised: 3 January 2023

Accepted: 14 January 2023

Published: 16 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The development and utilization of urban underground space (UUS) can optimize urban spatial structure, improve urban infrastructure, and effectively solve urban traffic congestion and other urbanization problems [1–5]. Therefore, recently, the interest in the utilization of UUS, due to its considerable contribution to sustainable urban development, has been growing [6–9].

With the continuous urbanization in China, underground infrastructure has gained great importance in urban space use because of limited land resources [10–14]. Underground subway systems occupy a significant position in urban underground infrastructure and have gradually become an irreplaceable public transportation tool [15–18].

According to statistics released by the China Urban Subway Association, by June 2021, 40 cities in China had opened subway lines, with a total of 6641.73 km of lines in operation. However, owing to factors such as people's travel time and geographical location of the subways, oversaturation of stations' populations often occurs [19], resulting in congestion and retention of people in stations. Consequently, this causes disharmony of resource ratio and huge potential safety problems [20,21]. For a reasonable allocation of vehicle shifts and adjustment of personnel scheduling, it is important to conduct research to forecast the passenger flow in subway stations [22,23].

Short-term patronage forecasting at subway stations is the advance estimation of patronage for a future period [24], and different methods have been used to address this problem. Ke et al. constructed a short-term passenger demand forecasting model based on long short-term memory (LSTM) [25]. The proposed predictive model is composed of multiple memory cells and has a better prediction effect than other types of prediction models. Although multiple factors are comprehensively considered for passengers, the time factor of passengers has not yet been fully analyzed and utilized. Li et al. combined the autoregressive integrated moving average model based on symbolic regression and achieved a better short-term prediction accuracy than traditional models under real conditions [26]. Chen et al. used a combination of seasonal-trend decomposition based on loess and LSTM to reduce the impact of data fluctuations [27]. The above methods include neural network, machine learning, and hybrid models [28,29], all of which aim to drive the rapid development of passenger flow forecasting.

The passenger flow data of subway stations have a chronological character and are closely linked to external time characteristics, such as workdays and weekends. Considering the characteristics of subway station passenger flow data, in this study, we integrate the time series mode attention mechanism based on the LSTM neural network, which captures the long-term dependence of the data and analyzes the short-term time series features. The contributions of this study can be listed as follows:

1. The prediction model temporal pattern attention (TPA)-LSTM, combining the time series mode attention mechanism and LSTM, was used to forecast the passenger flow of subway stations. The drawback of the disappearing LSTM gradient was avoided, and the prediction accuracy and stability was better.
2. In predicting the short-term passenger flow of TPA, instead of selecting the relevant passenger flow time steps as in the typical attention mechanism, the prediction model learns to select the relevant passenger flow time series. Considering the long- and short-term time dependences of subway passenger flow, the dependence and correlation of passenger flow data can be deeply mined and the prediction accuracy can be further improved.
3. According to the analysis of weekdays, weekends, and different types of passenger flows, the applicability of the model in subway station passenger flow prediction was verified by considering the Hangzhou Metro in China as an example.

The rest of this paper is organized as follows. Section 2 introduces the methods and research status of subway station passenger flow prediction in detail. Section 3 presents the problem of passenger flow forecasting in subway stations. Section 4 introduces the methods and process of the subway station passenger flow forecasting model in this study. Section 5 presents the model parameter setting and evaluation indexes of prediction results. Section 6 presents the forecast results and their analysis. Finally, the research content and the direction of improvement in the future are summarized in Section 7.

2. Literature Review

The short-term passenger flow forecasting methods include traditional statistical, machine learning, neural network, and other forecasting models. Traditional statistical models include the time series model [30] and Kalman filter model [31]. These models can predict the change in the number of passengers in the next period based on historical passenger flow data. Although traditional statistical models can process passenger flow data in different periods, they have high requirements in terms of data types and consistency. Furthermore, the accuracy of static and short-term predictions is considerably lower. In the past, scholars have used machine learning models to predict passenger flow. Machine learning models include the grey model (GM) [32] and the regression model [33]. Benitez et al. used the GM to build a passenger demand forecast model for the air transport industry [34], and Jiang et al. built a short-term forecast model for high-speed rail passenger flow using a grey support vector machine [35]. Owing to the increase in the amount of data and improvement in computing power, neural networks have gradually become an

important method for processing big data. For the past decade, different algorithms have continuously been developed, and many passenger flow predictions have been studied using neural networks [36]. Researchers have used deep learning methods to build subway station passenger flow prediction models, including back propagation (BP) networks [37], convolutional neural networks (CNNs) [38], and LSTM networks [39]. Chen et al. used the travel time of passenger flow as the research object and employed a BP neural network to forecast the travel time of passengers [40]. Zhang et al. [41] used an LSTM network for prediction and then employed a CNN to build a prediction model for the subway [42]. The results of these studies show that deep learning models achieve good passenger flow prediction performance.

Owing to the ability of long- and short-term neural networks to capture nonlinear interdependence, relevant research has achieved good results in recent years. The long short-term neural network was proposed by Hochreiter et al. [43]; it is a variant of a recurrent neural network (RNN) and can be used to memorize relatively long-term patterns. However, in practical applications, LSTMs cannot memorize interdependencies for a long time owing to training instability and vanishing gradient problems [44]. Based on LSTM, Jing et al. combined the LightGBM algorithm and the DRS local optimal fusion method to construct an LGB-LSTM-DRS model [45]. Compared with logistic regression, random forest, gradient boosting decision tree, and other prediction models, it has a better prediction effect. Yang et al. constructed an improved model based on an LSTM (namely, the enhanced long-term features based on LSTM prediction model), which overcomes the limitation of insufficient learning of long-term dependencies owing to the time lag [46]. Zhang et al. constructed the ResLSTM prediction model, which is a deep learning model combining a residual network, graph CNN, and LSTM [47]. Compared with other advanced prediction models, the ResLSTM has better prediction accuracy and robustness. Chen et al. built a Conv-LSTM prediction model, which can capture network spatial and temporal features [48]. Thereafter, they verified the effectiveness of the method by considering Chongqing Rail Transit as an example. The above methods can enhance the prediction accuracy of subway station passenger flow, which is significant for constructing accurate prediction models. However, these methods cannot deeply explore the long-term and short-term dependence of the subway station passenger flow prediction problem and the relationship between the correlation and prediction accuracy. If the temporal characteristics are deeply explored, the prediction accuracy can be further improved.

According to the above analysis, this study introduces the TPA-LSTM prediction model, which combines the TPA mechanism based on LSTM. Compared with traditional attention mechanisms, TPA is more conducive to the processing of subway station passenger flow data with time series characteristics [49]. Compared with the traditional LSTM, TPA-LSTM has the following advantages: (1) It avoids the problem of a vanishing gradient of LSTM and has better prediction accuracy and stability. (2) Considering the long- and short-term time dependencies of subway station passenger flow, it can deeply mine the dependence and correlation of passenger flow data. (3) Compared with LSTM, the improved TPA-LSTM model can learn longer historical data features without increasing the time complexity.

3. Problem Statement

Subway passenger flow forecasting is a complex and critical field of research that requires the use of high-precision and efficient forecasting methods based on historical passenger flow data. Efficient and accurate passenger flow forecasting of subway stations is significant to the organization of subway station passenger flow, which helps to rationally arrange operating shifts, improve passenger travel efficiency, and avoid station overcrowding, especially for hub sites in the subway network, as shown in Figure 1.

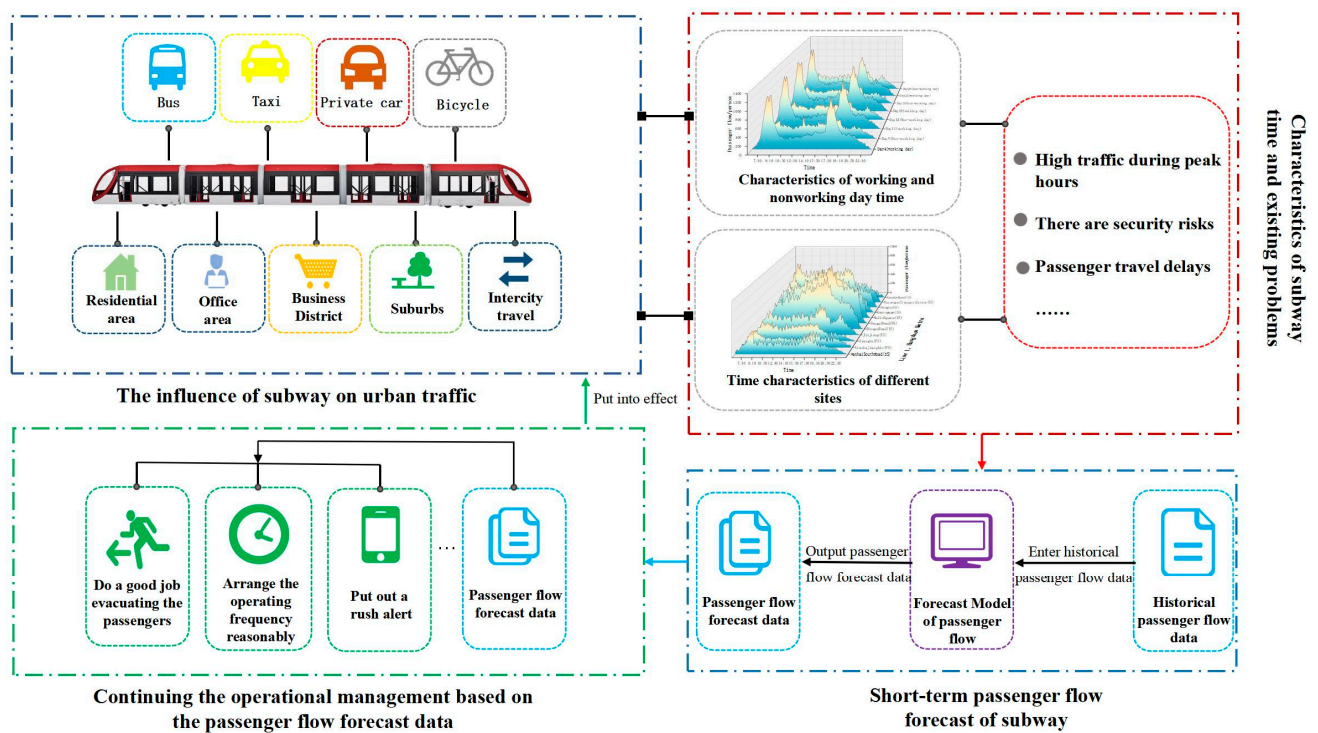


Figure 1. Frame diagram of alleviating subway passenger flow congestion.

Analyzing the temporal characteristics of subway station passenger flow is a necessary step for prediction. From the perspective of time, passenger traveling is affected by weekends and holidays. Considering Jinjiang subway station of Hangzhou City in China as an example, the number of passengers fluctuates significantly with the date. The passenger flow on weekdays shows a notable peak in the morning and evening, while this peak decreases on weekends, which is closely related to the commutes of people to work, as shown in Figure 2. There are different trends in the number of passengers on different days and at different stops. Consider Metro Line 1 in Hangzhou City, China as an example. Line 1 spans the Qiantang River and the Beijing–Hangzhou City Canal. The passenger flow is compared horizontally, as shown in Figure 3. The passenger flow of the first station, last station, and intermediate station (not first, last, or transfer station) is small, while the passenger flow at the transfer station is significantly higher than that at other types of stations. The passenger flow is strongly dependence on the change characteristics of its historical data.

This study employs the unique time characteristics of the historical data of the subway station passenger flow to predict passenger flow, alleviating the congestion problem of passenger flow in subway station. Accordingly, the input of the prediction model is the existing historical observation value of passenger flow of each site, and its mathematical expression is defined as $\{X_n | n = t - 1, \dots, t - x + 1, t - x\}$ (where n is the time step number, t is the time step length, and X is the number of passengers of the first n time steps). Y_t is the prediction model output value, and its mathematical expression is defined as $\{Y_n | n = t, \dots, t + m\}$ (where m is the predicted time step, and Y represents the predicted passenger flow of m steps).

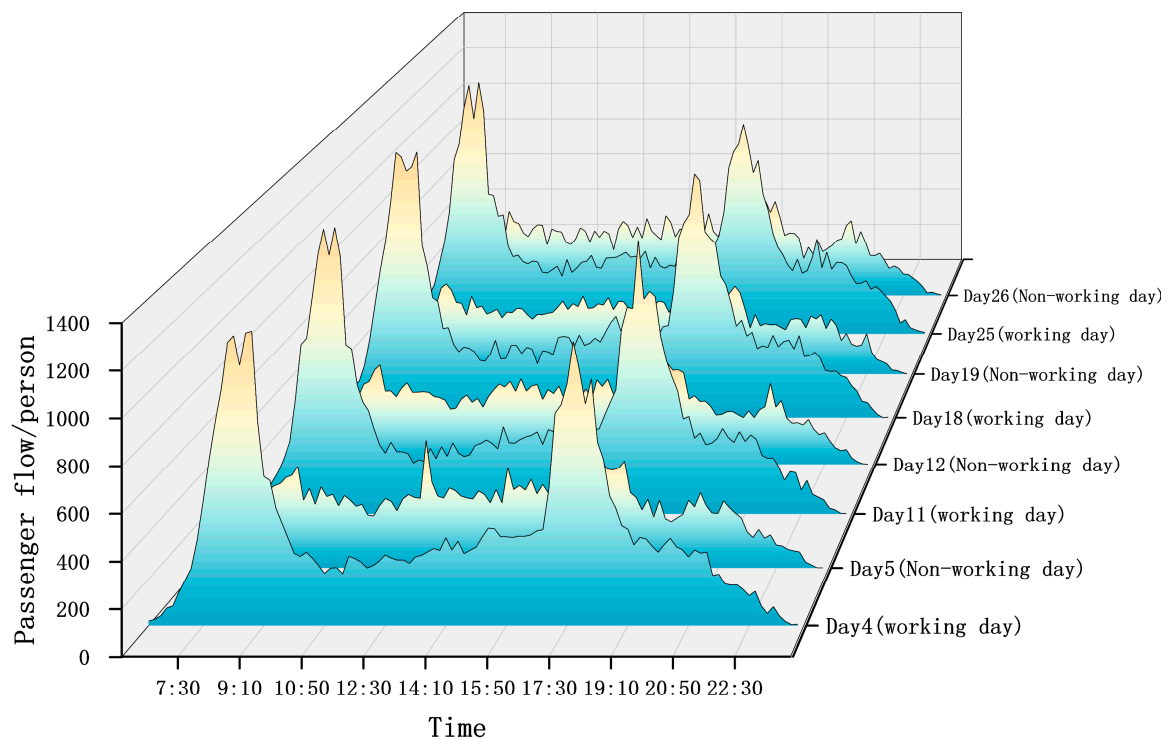


Figure 2. Passenger flow of Hangzhou Metro Jinjiang station on weekdays and weekends.

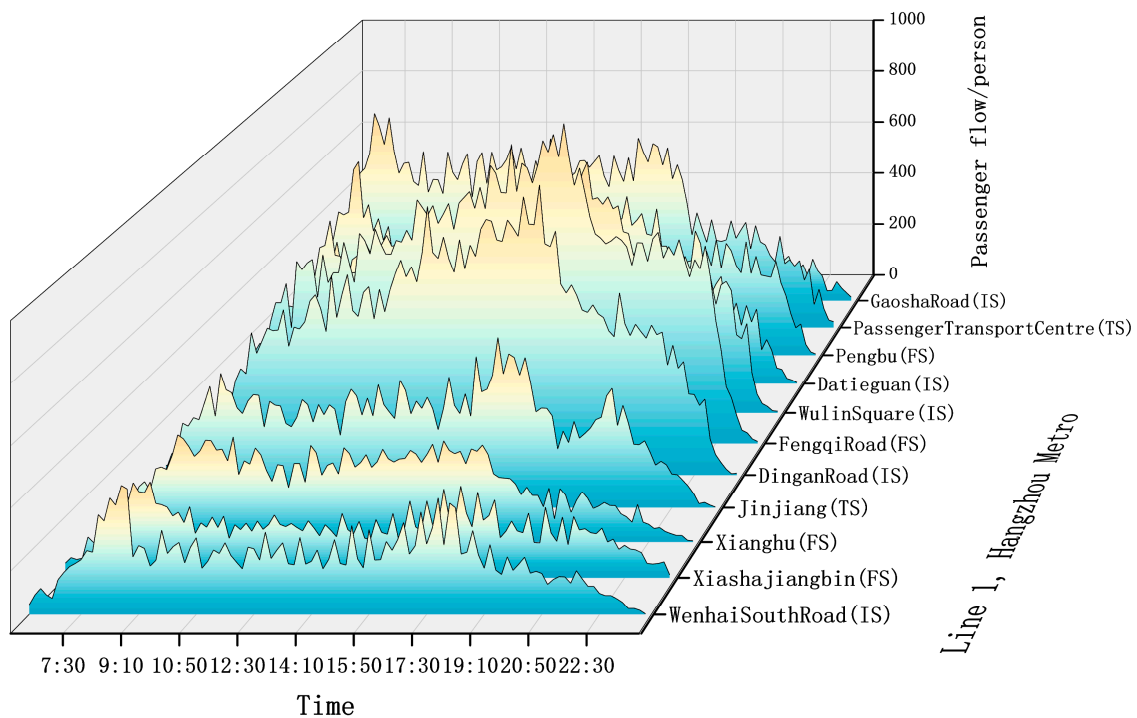


Figure 3. Passenger flow of Hangzhou Metro Line 1 at different stations. FS: first station, LS: last station, TS: transit station, IS: intermediate station (not FS, LS, or TS).

4. Methodology

In machine learning, deep learning is used to learn the inherent rules of sample data; it is widely used in word recognition, image recognition, and sound processing. Deep learning can be an effective tool for analyzing and processing big data. In this section, we

elaborate on the methods involved in subway passenger flow prediction model, including the LSTM, sequential mode attention mechanism, and TPA-LSTM prediction model.

4.1. LSTM

LSTM is an improved form of an RNN that has the function of short and long-term memory, which can deal with time series problems. The traditional RNN faces challenges in connecting the prediction results with relevant information beyond a certain distance. The RNN cannot effectively process big data with time series nature, such as subway passenger flow data [50]. To overcome the defects of the RNN, Hochreiter et al. proposed an LSTM neural network in 1997, which added an RNN-based memory module and replaced the RNN of the hidden layer with an LSTM unit, imparting long-term memory ability [43]. When new information is input, the LSTM screens out important information and saves it into long-term memory and thus avoiding the problem of gradient disappearance in the training process of the RNN. Therefore, LSTM is more suitable for the study of such problems.

LSTM adopts a gated network structure and includes the forget gate, input, and output gates (Figure 4). For the processing of passenger flow data in a subway station, the working steps are as follows:

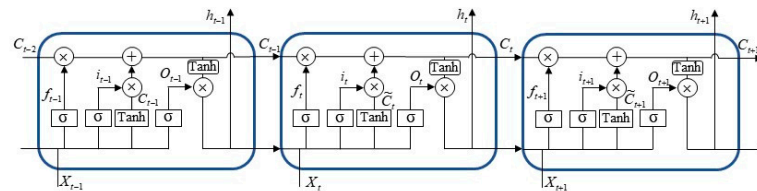


Figure 4. Long short-term memory (LSTM) structure diagram.

Step 1: When the LSTM unit receives new historical passenger flow data, the forget gate determines the information forgotten at the last moment, and it is expressed as

$$f_t = \sigma(\omega_f x_t + \alpha_f h_{t-1} + b_f), \quad (1)$$

where f_t represents the forget gate status value in the LSTM unit, σ represents the sigmoid activation function, x_t is the input value of passenger flow at the current moment, h_{t-1} is the output value of the passenger flow at the last moment, ω_f is the input weight, α_f is the cyclic weight of the forget gate, and b_f is the bias of the forget gate.

Step 2: The input gate determines whether to store the new passenger flow information in memory; that is, whether the input value and neuron state of the previous time step are updated to the neural unit of the next time step. Thus, it avoids memorizing the current unimportant information. The equation is as follows:

$$i_t = \sigma(\omega_i x_t + \alpha_i h_{t-1} + b_i), \quad (2)$$

where i_t is the state value, α_i is the cyclic weight, and b_i is the bias of the input gate.

Step 3: The sigmoid function is used to calculate the final output, and it is expressed as

$$o_t = \sigma(\omega_o x_t + \alpha_o h_{t-1} + b_o), \quad (3)$$

$$h_t = o_t * \tanh(c_t), \quad (4)$$

where o_t is the state value, ω_o is the input weight, α_o is the cyclic weight, b_o is the bias, h_t is the state of the hidden layer, and $*$ is the operation of the Hadamard product.

4.2. TPA

The attention mechanism was first used in image and natural language processing. Its function is to ensure the model focuses on important information in the historical data. In

addition, it helps generate output by reviewing the information of each time step before the current time step and focusing on relevant information. However, the traditional attention mechanism cannot capture the time pattern across multiple time steps; thus, it cannot handle the periodic subway passenger flow data well. To overcome this defect, Shih et al. proposed a model of temporal attention mechanism (i.e., temporal pattern attention mechanism (TPA)) in 2019 [49]. This model can capture data across multiple time steps and patterns according to the score function to determine the value of different time patterns. The output value is determined by the weight value. The sequential mode attentional mechanism module is shown in Figure 5.

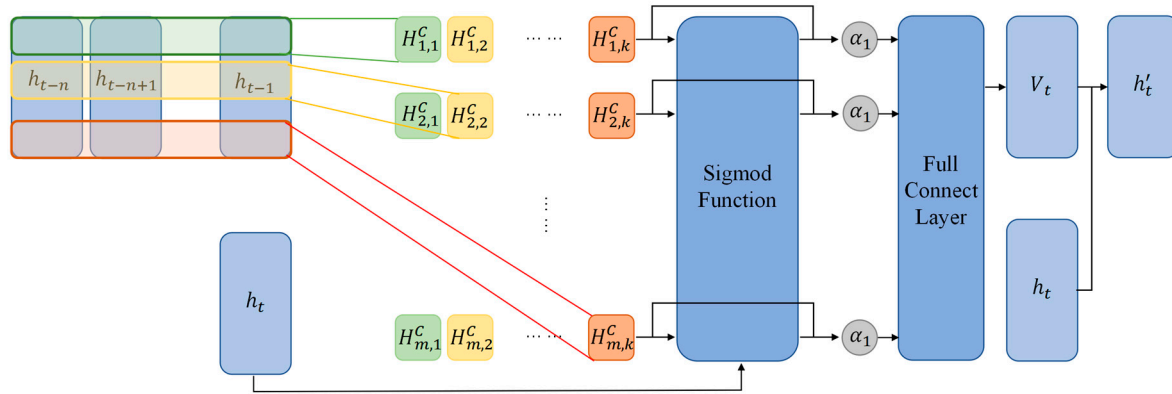


Figure 5. Attention mechanism of sequential patterns.

For the processing of subway passenger flow data, the working principle of TPA can be divided into the following three steps.

Step 1: Set $h = \{h_{m,l-w+1}, \dots, h_{m,t}\}$ as the subway passenger flow data sequence inputting TPA, where w is the length of the sequence and m is the dimension of a single data sequence. $C = \{C_1, C_2, \dots, C_T\}$ is the CNN filter, where T represents the maximum length of the attention mechanism. We set $w = T$: that is, the sequence length is consistent with the maximum length obtained by the attention mechanism. Different timing patterns can be obtained by the convolution operation of h and C , which can be expressed as

$$H_{i,j}^C = \sum_{l=1}^w h_{i,(t-w-1+l)} \times C_{j,T-w+l} \quad (5)$$

Step 2: Weight calculation

Calculate the weight of each time step data by defining a score function, which is expressed as

$$f(H_i^C, h_{m,t}) = (H_i^C) W_a h_{m,t}, \quad (6)$$

where H_i^C represents the i th row vector of H^C , $W_a \in R^{k \times m}$. Therefore, the attention weight is zero.

$$\alpha_i = \text{sigmoid}(f(H_i^C, h_{m,t})) \quad (7)$$

In Equation (7), *sigmoid* is the activation function used for weight normalization.

Step 3: TPA output

Define the context vector as shown in the following equation:

$$V_t = \sum_{i=1}^n \alpha_i H_i^C \quad (8)$$

The weight of the hidden layer at time t obtained through the attention mechanism is as follows:

$$h'_{m,t} = W_h h_{m,t} + W_V V_t \quad (9)$$

According to Equation (9), $W_h \in R^{m \times m}$, $W_V \in R^{m \times k}$.

Therefore, the output of TPA is:

$$Y_t = W_{h'} h'_{m,t} \quad (10)$$

According to Equation (10), $W_{h'} \in R^{n \times m}$, and Y_t is the output value at time t .

4.3. TPA-LSTM

LSTM is significantly affected by computing power and optimization algorithms during the operation; therefore, we introduced a TPA-LSTM prediction model that combines the time series attention mechanism and LSTM network; thus, avoiding the pitfalls of LSTM.

TPA-LSTM model uses the TPA mechanism to calculate the output value of the LSTM model. Compared with the LSTM model, this model can simultaneously focus on the output value at different time steps in history as well as at the current time [49]. In addition, the TPA-LSTM model can combine these output values to make the output values more accurate. The structure of the TPA-LSTM model is shown in Figure 6. The input of the model in the figure is the historical passenger flow data at the first x moments, and the output is the predicted value.

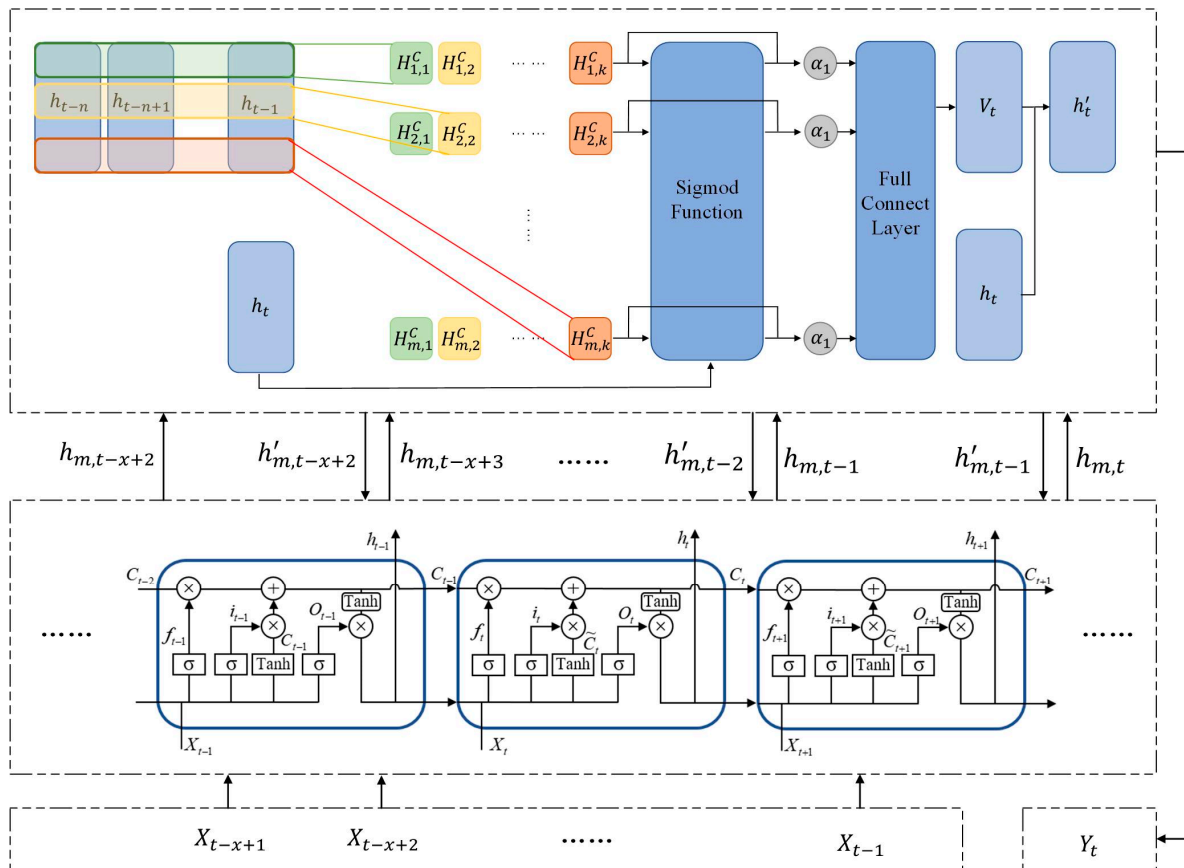


Figure 6. Temporal pattern attention (TPA)-LSTM model.

5. Case Study

5.1. Dataset Description

In this study, we considered the Hangzhou City automatic fare collection system passenger card data as the validation dataset, which included data related to the users: name, card number, credit card information, and the time and location of card use. According to privacy regulations, users' private information was hidden. In 2019, Hangzhou City only opened three subway lines, with a total of 81 stations. The basic information is as follows: Hangzhou Metro Line 1 has 34 stations, with Xianghu station and Xiashajiangbin station as

the first and last stations, respectively. Hangzhou Metro Line 2 has 33 stations, starting and ending at Liangzhu station and Chaoyang station, respectively. Hangzhou Metro Line 4 has 14 stations, with Puyan station and Pengbu station as the first and last stations, respectively. In this study, we selected the ticket checking system data from 1 January to 26 January 2019 as the experimental sample, with a total of 70 million rows of data.

Based on the interval of the historical data of 10 min, the passenger volume within a certain time interval is represented by $\{X_n | n = t - 1, \dots, t - x + 1, t - x\}$, where n represents the time step number, t represents the time step length, and x represents the passenger flow of the first n time steps. The predicted passenger flow value is expressed as $\{Y_n | n = t, \dots, t + m\}$, where m is the predicted time step, and Y is the predicted passenger flow of m time steps. The prediction process is shown in Figure 7.

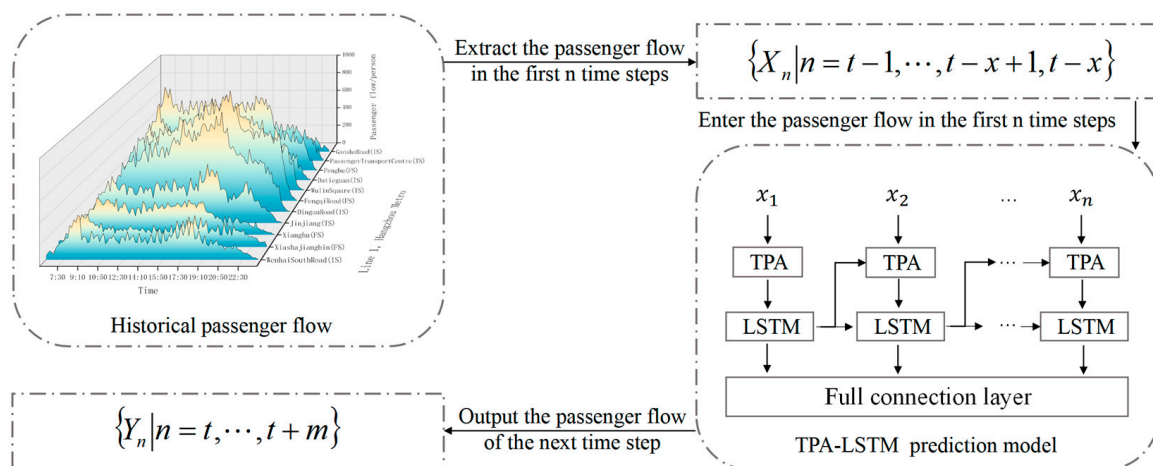


Figure 7. Data entry and forecasting.

5.2. Methods for Comparison

Experiments were mainly run on a GPU platform with an NVIDIA GeForce MX150 graphics card, whose detailed information is shown in Table 1. Python libraries, including TensorFlow, Numpy, Keras, Pandas, and Numpy, were used to build our model.

Table 1. Experimental Environment.

Items	Parameters
OS	Windows 10
Memory	16 GB
CPU	Intel(R) Core(TM) i7-8550U CPU @ 1.80 GHz
GPU	NVIDIA GeForce MX150
TensorFlow version	2.11.0
Numpy version	1.23.4
Keras version	2.11.0
Pandas version	1.5.2
Numpy version	1.24.1
Training time	1060 s

In the TPA-LSTM model established in this study, LSTM was set as three layers, and each layer had 40 LSTM neurons. The activation function was the sigmoid function; we selected mean squared error as the loss function of the model and stochastic gradient descent as the optimization algorithm. In the process of model training, the learning rate and number of iterations were 0.001 and 20, respectively. The parameter settings of the TPA-LSTM subway passenger flow prediction model are shown in Table 2.

Table 2. Parameter Settings of the TPA-LSTM model.

Parameter	Instructions	Set
Activation	The activation function	Sigmoid
Loss	Loss function	MSE
Optimizer	Optimization algorithm	Sgd
Lr	Vector	0.001
Num_epochs	The number of iterations	20

In this study, we compared the TPA-LSTM prediction model with some traditional prediction models. The three models for comparison were CNN, LSTM, and CNN-LSTM, and the parameters of these models were as follows:

CNN: The convolution kernel size of 3×3 was adopted, convolution time step was set to 5, and the pooling mode was set to maximum pooling.

LSTM: 40 LSTM neurons with three layers were set. In the process of model training, the learning rate and iteration times were consistent with the TPA-LSTM model.

CNN-LSTM: We set parameters consistent with those of the CNN and LSTM models.

5.3. Evaluation of Forecast Results

In this study, the following four evaluation indicators, including the mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2), were used to analyze the accuracy of the model.

$$MAE = \frac{1}{n} \times \sum_{t=1}^n |A_t - Y_t| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{t=1}^n (\log(A_t + 1) - \log(Y_t + 1))^2} \quad (12)$$

$$MAPE = \frac{1}{n} \times \sum_{t=1}^n \left| \frac{A_t - Y_t}{A_t} \right| \times 100\% \quad (13)$$

$$R^2 = \frac{\sum_{i=1}^n (A_t - Y_t)^2}{\sum_{i=1}^n (A_t - Y_t)^2 + \sum_{i=1}^n (A_t - A_{ta})^2} \quad (14)$$

Here, A_t is the actual value of subway passenger volume, A_{ta} is the average value of the actual passenger volume, Y_t is the predicted value of subway passenger volume, and n is the sample number of subway passenger volume.

6. Results Analysis

6.1. Prediction Results

Considering Linping station of Hangzhou Metro Line 1 as an example, the prediction results of 25 and 26 January can be obtained by considering the historical passenger flow data from 1 January to 24 January as the input value. To test the prediction accuracy of each model, we specifically analyzed the prediction results on 25 January, as shown in Figure 8. The figure contains the prediction results of the four prediction models: CNN, LSTM, CNN-LSTM, and TPA-LSTM.

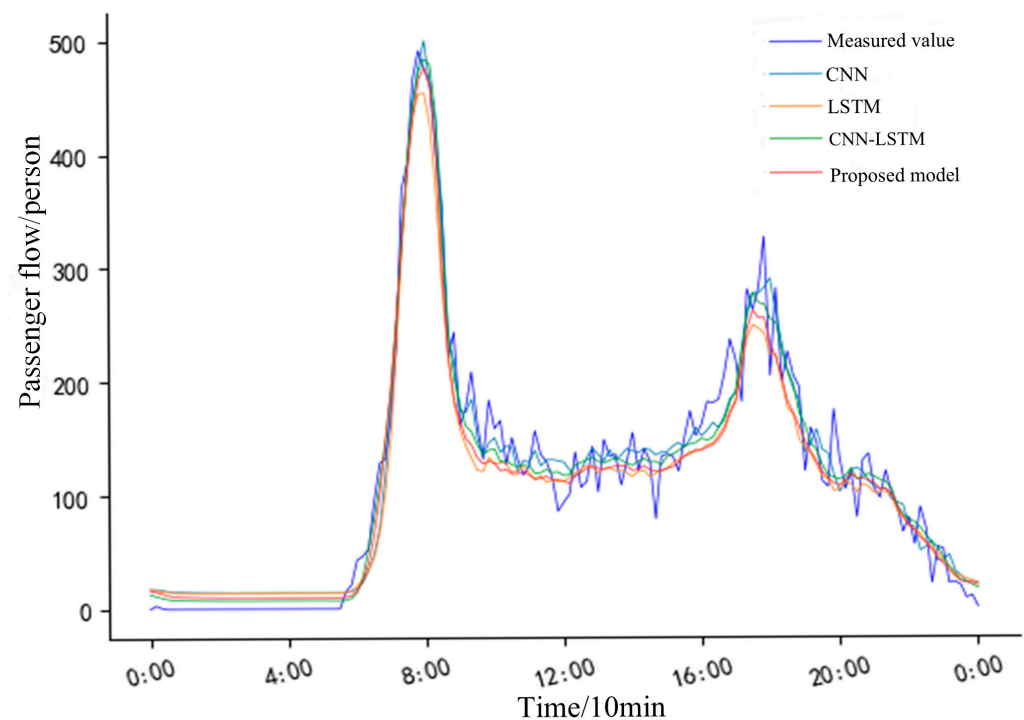


Figure 8. Comparison of passenger flow forecast of various models (Linping station, January 25).

It can be observed from Figure 8 that (1) all prediction models exhibit good capture of morning and evening peaks in the passenger flow prediction of Linping station of Hangzhou Metro Line 1. Both the CNN and LSTM models can predict the fluctuation trend of passenger flow. In a detailed analysis, the TPA-LSTM prediction model has good applicability to Hangzhou Metro passenger flow prediction and can accurately predict the peak of passenger flow. Therefore, the TPA-LSTM prediction model has high prediction accuracy and strong learning ability. (2) Although the CNN and LSTM models in the figure can accurately predict the general trend of Hangzhou City subway passenger flow change on 25 January, compared with CNN-LSTM and TPA-LSTM prediction models, their fitting degree is slightly worse. This result also verifies the limitation of the single structure of the two prediction models, while the hybrid prediction model has higher accuracy for subway passenger flow prediction. (3) In most cases, the prediction effect of CNN-LSTM is close to that of the TPA-LSTM model; however, TPA-LSTM is closer to the measured value in peak and peak phases. This is because the TPA mechanism plays an important role in capturing time patterns across multiple time steps. Moreover, it determines the weights of different time patterns according to the scoring function to generate the output value, making the TPA-LSTM model a better fitting model.

Table 3 contains various evaluation indicators of the prediction model. The indices of the mixed model are superior to those of the single model, and the TPA-LSTM prediction model adopted in this study is superior to other models. In terms of MAPE, the value of the TPA-LSTM was 6.8090%; that is, 1.1865% lower than that of the CNN-LSTM. Furthermore, the TPA-LSTM outperformed the other models in MAE, RMSE, and R^2 . Therefore, it proves the superiority of the prediction model established in this study in the short-term prediction of subway passenger flow.

Table 3. Error analysis of prediction results of various models at Linping station.

Model	The Evaluation Index			
	MAPE (%)	MAE	RMSE	R ²
CNN	15.8189	18.8550	24.1583	0.9263
LSTM	8.6346	21.4699	27.5674	0.9041
CNN-LSTM	7.9955	17.4831	22.9633	0.9334
TPA-LSTM	6.8090	17.0147	22.3449	0.9370

6.2. Performance during Workdays and Weekends

In the problem statement section, we indicated that there is a difference in the number of passengers on weekdays and weekends; thus, this section will forecast and analyze the passenger flow on workdays and weekends at the Linping station in Hangzhou Metro Line 1 in China. The historical passenger flow data from 1 January to 24 January was the input of the model; the forecast results of 25 and 26 January were analyzed, in which 25 (Friday) is a workday and 26 (Saturday) is a rest day.

Figure 9 shows the prediction results of each model. Several commuters take the subway in the peak period on weekdays, and the number of passengers in the morning peak is significantly higher than in the evening peak; however, this phenomenon is not obvious on weekends. Each model can reflect the change trend of the number of passengers at different times, but there are some differences in the accuracy of each model. The prediction curve of the CNN and LSTM models for Linping station on 26 January (Saturday) is relatively smooth. Although the LSTM can predict the change trend of passenger flow, it cannot capture the randomness of the data. Consequently, the prediction curve of the subway passenger flow on weekends is relatively smooth, and the rapid fluctuation within a small range cannot be accurately predicted. Compared with the LSTM prediction model, the TPA-LSTM prediction model constructed in this study is more accurate in predicting small fluctuations of passenger flow and can accurately predict multiple local passenger flow peaks on weekends. The peak hours on weekdays and weekends are similar. This is because some passengers need to commute during the weekends during peak hours. However, the number of passengers during peak hours on weekends will not be more than that during weekdays.

Therefore, we believe that the prediction model constructed in this study has a good ability to capture local peak values for the passenger flow data during weekends, which are scattered and have certain random fluctuations.

Table 4 contains the evaluation indicators of the prediction results. The error comparison results show that the TPA-LSTM model adopted in this paper is superior to the CNN, LSTM, and CNN-LSTM models. Compared with the other models, it can better capture the time characteristics of subway passenger flow and accurately predict the change trend between peak and nonpeak hours. Furthermore, this indicates that the TPA-LSTM and CNN-LSTM hybrid models are more suitable for solving the problem of passenger flow prediction of public transport than single models, such as CNN and LSTM.

Table 4. Error analysis of prediction results of each model on weekdays and weekends.

Model	The Evaluation Index			
	MAPE (%)	MAE	RMSE	R ²
CNN	16.0155	18.2653	24.1652	0.9193
LSTM	8.8547	20.3674	26.5563	0.9102
CNN-LSTM	7.8976	17.5467	22.5839	0.9375
TPA-LSTM	6.9623	16.3247	21.8946	0.9402

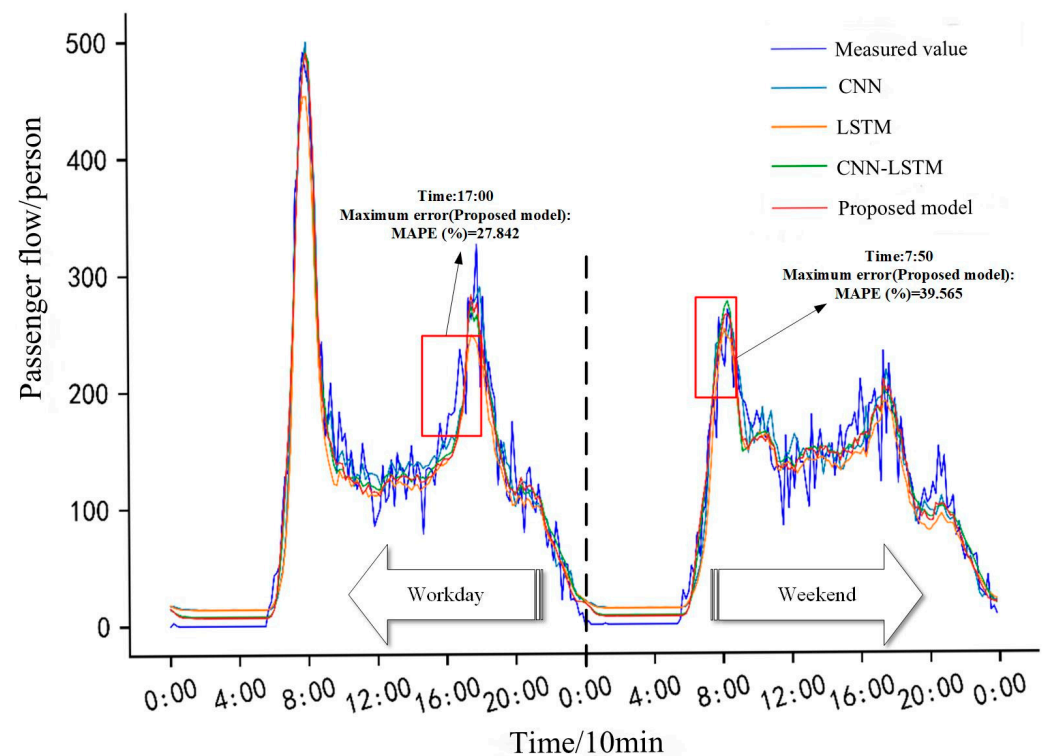


Figure 9. Comparison of passenger flow prediction between models (weekday and rest day at Linping station).

6.3. Performance at Different Station Types

Owing to the differences and distribution in passenger flow among different stations, the first and last stations, as well as the intermediate and transfer stations, were selected for passenger flow prediction. To analyze the difference of passenger flow between different types of sites, this section compares and analyzes passenger flow between different types of sites on weekdays and weekends.

(1) Passenger flow prediction at the first and last station

Linping station is the first and last station of Line 1. Passenger flow on 25 January and 26 January has been predicted in the previous section, in which 25 is a weekday and 26 is a weekend. The forecast results show that the predicted value of passenger flow is typically consistent with the measured value. Passenger flow on weekdays shows significant morning and evening peaks, while weekend peaks are significantly lower than those on weekdays. The fluctuation of the predicted value is slightly larger in the flat peak period of weekends owing to the strong randomness of the variation of the passenger flow on weekends, and the fluctuation range of the predicted value also increases.

Compared with other models, the TPA-LSTM model has higher prediction accuracy and minimal error, displaying better prediction.

(2) Passenger flow prediction at an intermediate station

This study predicts the passenger flow of Xinfeng station on 25 (weekday) and 26 January (weekend), and Figure 10 shows the prediction results. Each model can well predict the data characteristics of intermediate stations on weekdays and weekends, which is typically consistent with the measured values. Compared with the weekday passenger flow, the fluctuation range of the predicted value increases owing to the stronger randomness of the variation in the weekend passenger flow; however, the peak value of the non-weekday passenger flow decreases.

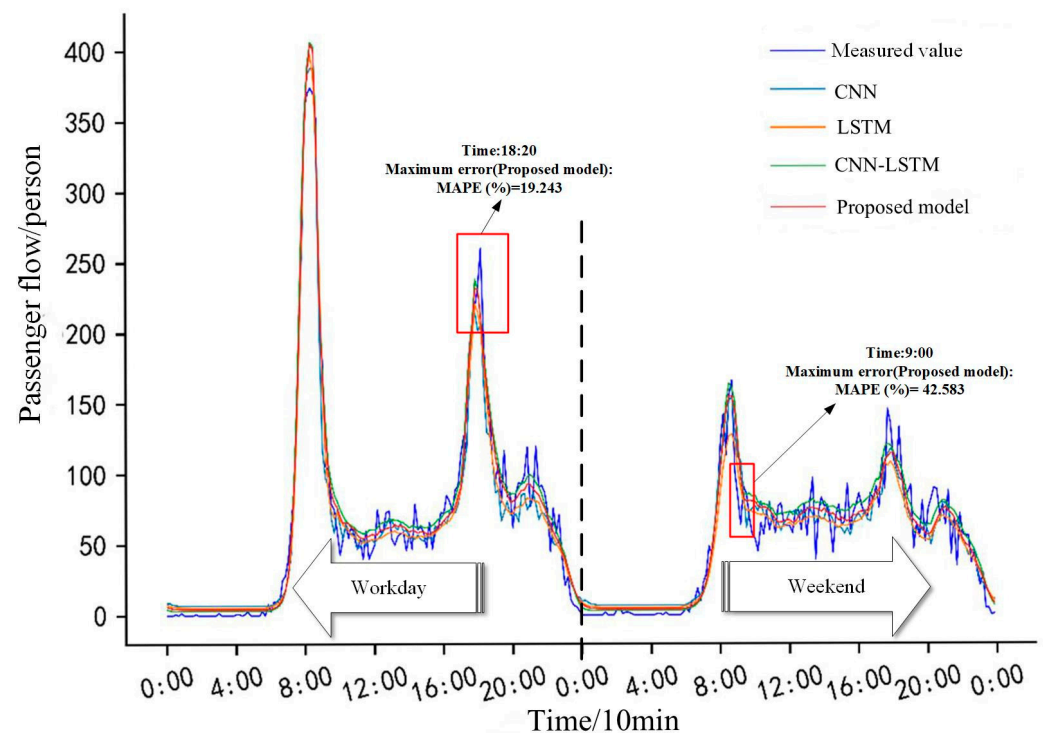


Figure 10. Comparison of passenger flow prediction among models of Xinfeng station.

Table 5 shows the error comparison results of each model. Both the single model CNN and hybrid model TPA-LSTM can predict the basic characteristics of passenger flow of Xinfeng station on weekdays and weekends; however, there is a large gap in accuracy between the models. The comparison results show that the MAPE, MAE, RMSE, and R^2 of TPA-LSTM are 20.6221%, 9.6947, 13.1371, and 0.9539, respectively. Compared with the other models, the TPA-LSTM model has higher prediction accuracy and better fitting.

Table 5. Error analysis of prediction results of each model in Xinfeng station.

Model	The Evaluation Index			
	MAPE (%)	MAE	RMSE	R^2
CNN	28.2893	11.3994	14.8537	0.9411
LSTM	27.3533	10.8343	14.2638	0.9456
CNN-LSTM	21.0220	9.7496	13.6759	0.9500
TPA-LSTM	20.6221	9.6947	13.1371	0.9539

(3) Passenger flow prediction at an intermediate station

In this study, we used passenger data from Jinjiang station for transfer station prediction. The prediction days were 25 and 26 January. Compared with Linping and Xinfeng stations, the data of Jinjiang station are larger and the situation is more complicated. The passenger flow prediction results of each model at Jinjiang station are shown in Figure 11. Each model can well predict the characteristics of passenger flow at the Jinjiang station, which is typically consistent with the measured values. For weekends, the prediction results of Linping and Xinfeng stations have notable fluctuations.

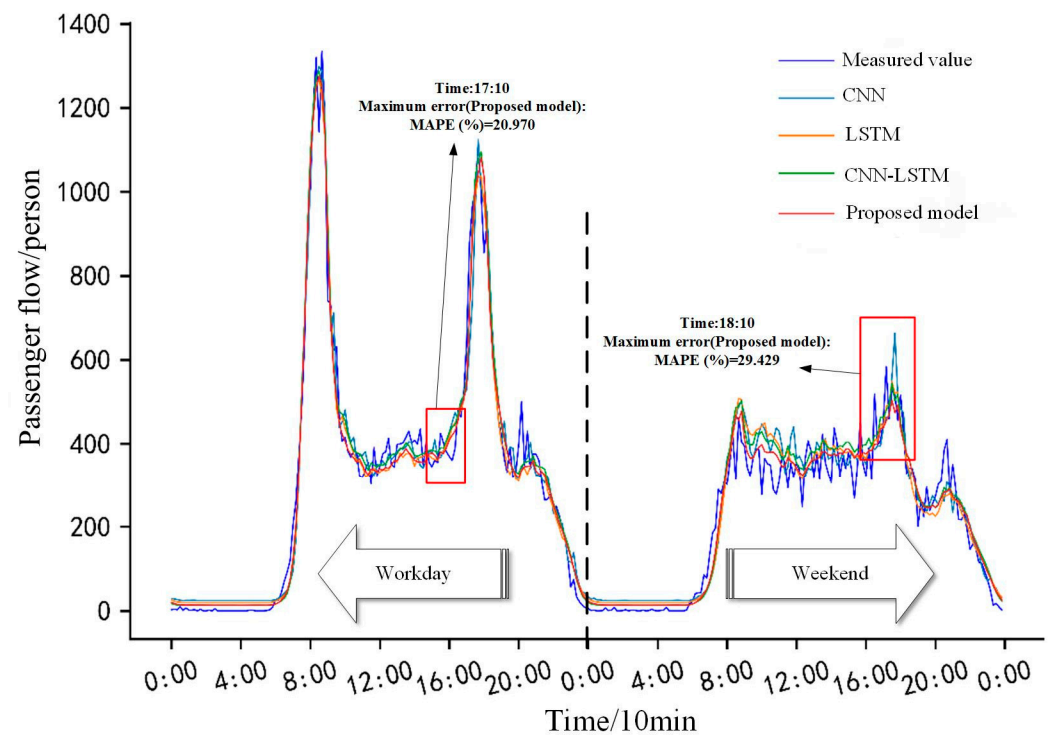


Figure 11. Comparison of passenger flow prediction of different models at Jinjiang Station.

Table 6 shows the error analysis of prediction results of various models of Jinjiang station. The MAE, RMSE, and determination coefficient of TPA-LSTM are 34.3704, 48.4728, and 0.9636, respectively. Compared with other models, TPA-LSTM model has the better prediction effect. However, the prediction accuracy of TPA-LSTM model for transfer stations decreased owing to the greater volume of passengers at the interchange stations, complicating the situation.

Table 6. Error analysis of prediction results of various models at Jinjiang Station.

Model	The Evaluation Index			
	MAPE (%)	MAE	RMSE	R^2
CNN	65.4696	40.9921	56.7982	0.9500
LSTM	64.4587	38.4220	52.2657	0.9576
CNN-LSTM	63.7833	36.2154	49.8794	0.9614
TPA-LSTM	61.7694	34.3704	48.4728	0.9636

Based on the comprehensive analysis, the TPA-LSTM model adopted in this study is also suitable for passenger flow prediction of transfer stations. Compared with other prediction models, it has better training accuracy and a better prediction effect.

7. Conclusions and Future Work

In this study, the TPA-LSTM method, developed as a combination of a TPA mechanism and LSTM network, was used for short-term prediction of passenger flows at subway stations. The TPA mechanism can capture time patterns across multiple time steps, which cannot be achieved by traditional attention mechanisms. Compared with the traditional LSTM model, this model can simultaneously focus on the hidden layer output value of different time steps in history and the hidden layer output value of the current time. In addition, it can calculate the correlation between these output values and further improve the prediction accuracy. The following conclusions were drawn:

1. Based on the card swiping data of passengers in Hangzhou Metro, the validation results show that TPA-LSTM has good applicability and high precision for subway station passenger flow prediction. It can assist subway stations in formulating subway network planning and operation schemes and in reducing traffic accidents caused by mobility, such as stampedes.
2. In this study, the TPA-LSTM model was compared with three classical short-term prediction models of passenger flow at subway stations: CNN, LSTM, and CNN-LSTM. The short-term passenger flow prediction results for weekdays, weekends, and three types of stations show that the TPA-LSTM model had the best accuracy and stability, in terms of MAPE, MAE, RMSE, and R^2 , among the four prediction models. Consequently, the TPA mechanism can further mine the temporal characteristics of the historical data of subway station passenger flow.

Some aspects of this paper can be improved upon. In this study, the time dependence and correlation of subway passenger flow data were explored. Limitations and tolerances of this study include: (1) the influence of external factors, such as weather and holidays, were not considered; (2) the time span of the datasets was only a few months, which may have resulted in not investigating some temporal external factors such as seasons. In future studies, the effects of external features, such as weather and holidays, on short-term passenger flow prediction of subway stations can be accommodated to improve the anti-interference of the proposed model. We will further study the application of the TPA-LSTM model in longer datasets.

Author Contributions: Lingxiang Wei: formal analysis, investigation, visualization, methodology, writing—original draft. Dongjun Guo: conceptualization, funding acquisition, project administration, methodology, writing—review & editing. Zhilong Chen: investigation, validation, supervision, writing—review & editing. Jincheng Yang: methodology, visualization. Tianliu Feng: data curation, validation. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to thank the National Natural Science Foundation of China (NNSFC) for the funding that enabled this research to be carried out under Grant No. 51878660 and No. 52078481.

Data Availability Statement: Data sharing is not applicable to this article.

Acknowledgments: The authors would like to acknowledge the editors and reviewers of *ISPRS International Journal of Geo-Information* for their constructive comments and suggestions. The authors are very grateful to Tianchi Data Sets for providing datasets (<https://tianchi.aliyun.com/dataset/dataDetail?dataId=21904>, accessed on 1 November 2022) for research use.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Han, Y.; Peng, T.; Wang, C.; Zhang, Z.; Chen, G. A Hybrid GLM Model for Predicting Citywide Spatio-Temporal subway Passenger Flow. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 222. [\[CrossRef\]](#)
2. Bobylev, N. Underground space as an urban indicator: Measuring use of subsurface. *Tunn. Undergr. Sp. Tech.* **2016**, *55*, 40–51. [\[CrossRef\]](#)
3. Bai, J.; Zhu, J.; Song, Y.; Zhao, L.; Hou, Z.; Du, R.; Li, H. A3T-GCN: Attention temporal graph convolutional network for traffic forecasting. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 485. [\[CrossRef\]](#)
4. Qiao, Y.K.; Peng, F.L.; Wu, X.L.; Luan, Y.P. Visualization and spatial analysis of socio-environmental externalities of urban underground space use: Part 1 positive externalities. *Tunn. Undergr. Sp. Tech.* **2022**, *121*, 104325. [\[CrossRef\]](#)
5. Shao, J.; Liu, G.; Yuan, H.; Song, Q.; Yang, M.; Luo, D.; Zhang, X.; Tan, Y.; Zhang, Y. Evaluation and scale forecast of underground space resources of historical and cultural cities in China. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 31. [\[CrossRef\]](#)
6. Bobylev, N. Mainstreaming sustainable development into a city's Master plan: A case of Urban Underground Space use. *Land Use Policy* **2009**, *26*, 1128–1137. [\[CrossRef\]](#)
7. Wang, W.; Wang, S.; Chen, H.; Liu, L.; Fu, T.; Yang, Y. Analysis of the Characteristics and Spatial Pattern of the Catering Industry in the Four Central Cities of the Yangtze River Delta. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 321. [\[CrossRef\]](#)
8. Cui, J. Building three-dimensional pedestrian networks in cities. *Undergr. Space* **2021**, *6*, 217–224. [\[CrossRef\]](#)
9. Lin, D.; Broere, W.; Cui, J. Underground space utilisation and new town development: Experiences, lessons and implications. *Tunn. Undergr. Sp. Tech.* **2022**, *119*, 104204. [\[CrossRef\]](#)

10. Feng, J.R.; Gai, W.M.; Yan, Y.B. Emergency evacuation risk assessment and mitigation strategy for a toxic gas leak in an underground space: The case of a subway station in Guangzhou, China. *Safety Sci.* **2021**, *134*, 105039. [\[CrossRef\]](#)
11. Guo, D.; Chen, Y.; Yang, J.; Tan, Y.H.; Zhang, C.; Chen, Z. Planning and application of underground logistics systems in new cities and districts in China. *Tunn. Undergr. Sp. Tech.* **2021**, *113*, 103947. [\[CrossRef\]](#)
12. Zhang, Z.; Han, Y.; Peng, T.; Li, Z.; Chen, G. A Comprehensive Spatio-Temporal Model for Subway Passenger Flow Prediction. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 341. [\[CrossRef\]](#)
13. Yin, L.; Zhu, J.; Li, W.; Wang, J. Vulnerability Analysis of Geographical Railway Network under Geological Hazard in China. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 342. [\[CrossRef\]](#)
14. Wang, Q.; Geng, P.; Guo, X.; Wang, X.; Li, P.; Zhao, B. Case study on the seismic response of a subway station combined with a flyover. *Undergr. Space* **2021**, *6*, 665–677. [\[CrossRef\]](#)
15. Jin, M.; Wang, L.; Ge, F.; Xie, B. Understanding the Dynamic Mechanism of Urban Land Use and Population Distribution Evolution from a Microscopic Perspective. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 536. [\[CrossRef\]](#)
16. Liu, A.H.; Ellul, C.; Swiderska, M. Decision making in the 4th dimension—Exploring use cases and technical options for the integration of 4D BIM and GIS during construction. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 203. [\[CrossRef\]](#)
17. Lin, D.; Broere, W.; Cui, J. Metro systems and urban development: Impacts and implications. *Tunn. Undergr. Sp. Tech.* **2022**, *125*, 104509. [\[CrossRef\]](#)
18. Pang, R.; Chen, K.; Fan, Q.; Xu, B. Stochastic ground motion simulation and seismic damage performance assessment of a 3-D subway station structure based on stochastic dynamic and probabilistic analysis. *Tunn. Undergr. Sp. Tech.* **2022**, *126*, 104568. [\[CrossRef\]](#)
19. Zhou, J.; Wang, H.; Sun, D.; Xu, S.; Lv, M.; Yu, F. Optimization Scheme of Large Passenger Flow in Huoying Station, Line 13 of Beijing Subway System. *Cmc-Comput. Mater. Con.* **2020**, *63*, 1387–1398. [\[CrossRef\]](#)
20. Yang, X.; Xue, Q.; Ding, M.; Wu, J.; Gao, Z. Short-term Prediction of Passenger Volume for Urban Rail Systems: A Deep Learning Approach Based on Smart-card Data. *Int. J. Prod. Econ.* **2021**, *231*, 107920. [\[CrossRef\]](#)
21. Lee, S.J.; Shin, S.I. A Study on Improving Subway Crowding Based on Smart Card Data: A Focus on Early Bird Policy Alternative. *J. Inf. Technol. Serv.* **2020**, *19*, 125–138. [\[CrossRef\]](#)
22. Sun, J.; Yao, J.J.; Wang, M.X. Subway Passenger Flow Analysis and Management Optimization Model Based on AFC Data. *J. Intell. Fuzzy. Syst.* **2021**, *41*, 4773–4783. [\[CrossRef\]](#)
23. Chen, E.H.; Ye, Z.R.; Wang, C.; Xu, M.T. Subway Passenger Flow Prediction for Special Events Using Smart Card Data. *IEEE T. Intell. Transp.* **2020**, *21*, 1109–1120. [\[CrossRef\]](#)
24. Yuan, Y.; Shao, C.F.; Cao, Z.C.; Chen, W.X.; Yin, A.T.; Yue, H.; Xie, B.L. Urban Rail Transit Passenger Flow Forecasting Method Based on the Coupling of Artificial Fish Swarm and Improved Particle Swarm Optimization Algorithms. *Sustainability* **2020**, *11*, 7230. [\[CrossRef\]](#)
25. Ke, J.T.; Zheng, H.Y.; Yang, H.; Chen, X.Q. Short-term Forecasting of Passenger Demand Under on-demand Ride Services: A Spatio-temporal Deep Learning Approach. *Transp. Res. C-Emer.* **2016**, *85*, 591–608. [\[CrossRef\]](#)
26. Li, L.C.; Wang, Y.G.; Zhong, G.; Zhang, J.; Ran, B. Short-to-medium Term Passenger Flow Forecasting for Metro Stations Using a Hybrid Model. *KSCE J. Civ. Eng.* **2018**, *22*, 1937–1945. [\[CrossRef\]](#)
27. Chen, D.W.; Zhang, J.H.; Jiang, S.X. Forecasting the Short-Term Metro Ridership With Seasonal and Trend Decomposition Using Loess and LSTM Neural Networks. *IEEE Access* **2020**, *8*, 91181–91187. [\[CrossRef\]](#)
28. He, Y.X.; Zhao, Y.; Tsui, K.L. Short-term Forecasting of Origin-destination Matrix in Transit System via A Deep Learning Approach. *Transp. A Transp. Sci.* **2022**, *1*–28. [\[CrossRef\]](#)
29. Han, Y.; Wang, C.; Ren, Y.; Wang, S.; Zheng, H.; Chen, G. Short-Term Prediction of Bus Passenger Flow Based on a Hybrid Optimized LSTM Network. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 366. [\[CrossRef\]](#)
30. Li, W.; Sui, L.Y.; Zhou, M.; Dong, H.R. Short-term Passenger Flow Forecast for Urban Rail Transit Based on Multi-source Data. *EURASIP J. Wirel. Comm.* **2021**, *2021*, 9. [\[CrossRef\]](#)
31. Shekhar, S.; Williams, B.M. Adaptive Seasonal Time Series Models for Fore Casting Short-term Traffic Flow. *Transp. Res. Rec.* **2007**, *2024*, 116–125. [\[CrossRef\]](#)
32. Xie, N.M. Explanations about Grey Information and Framework of Grey System Modeling. *Grey Syst. -Theory Appl.* **2017**, *7*, 179–193. [\[CrossRef\]](#)
33. Kasza, J.; Wolfe, R. Interpretation of Commonly Used Statistical Regression Models. *Respirology* **2014**, *19*, 14–21. [\[CrossRef\]](#)
34. Benitez, R.B.C.; Paredes, R.B.C.; Lodewijks, G.; Nabais, J.L. Damp Trend Grey Model Forecasting Method for Airline Industry. *Expert Syst. Appl.* **2013**, *40*, 4915–4921. [\[CrossRef\]](#)
35. Jiang, X.S.; Zhang, L.; Chen, X.Q. Short-term Forecasting of High-speed Rail Demand: A Hybrid Approach Combining Ensemble Empirical Mode Decomposition and Gray Support Vector Machine with Real-world Applications in China. *Transp. Res. C-Emer.* **2014**, *44*, 110–127. [\[CrossRef\]](#)
36. Liu, Y.; Liu, Z.Y.; Jia, R. DeepPF: A Deep Learning Based Architecture for Metro Passenger Flow Prediction. *Transp. Res. C-Emer.* **2019**, *101*, 18–34. [\[CrossRef\]](#)
37. Wang, L.; Zeng, Y.; Chen, T. Back Propagation Neural Network with Adaptive Differential Evolution Algorithm for Time Series Forecasting. *Expert Syst. Appl.* **2015**, *42*, 855–863. [\[CrossRef\]](#)

38. Han, Y.; Wang, S.K.; Ren, Y.B.; Wang, C.; Gao, P.; Chen, G. Predicting Station-Level Short-Term Passenger Flow in A Citywide Metro Network Using Spatiotemporal Graph Convolutional Neural Networks. *Isprs Int. J. Geo-Inf.* **2019**, *8*, 243. [[CrossRef](#)]
39. Wang, J.L.; Zhang, J.; Wang, X.X. Bilateral LSTM: A Two-dimensional Long Short-term Memory Model with Multiply Memory Units for Short-term Cycle Time Forecasting in Re-entrant Manufacturing Systems. *IEEE T. Ind. Inform.* **2018**, *14*, 748–758. [[CrossRef](#)]
40. Chen, C.; Wang, H.; Yuan, F.; Jia, H.Z.; Yao, B.Z. Bus travel time prediction based on deep belief network with back-propagation. *Neural Comput. Appl.* **2019**, *32*, 10435–10449. [[CrossRef](#)]
41. Zhang, J.L.; Chen, F.; Guo, Y.A.; Li, X.H. Multi-graph Convolutional Network for Short-term Passenger Flow Forecasting in Urban Rail Transit. *IET Intell. Transp. Sy.* **2020**, *14*, 1210–1217. [[CrossRef](#)]
42. Zhang, J.L.; Chen, F.; Shen, Q. Cluster-based LSTM Network for Short-term Passenger Flow Forecasting in Urban Rail Transit. *IEEE Access* **2019**, *7*, 147653–147671. [[CrossRef](#)]
43. Hochritter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
44. Alkheder, S.; Taamneh, M.; Taamneh, S. Severity prediction of traffic accident using an artificial neural network. *J. Forecast.* **2017**, *36*, 100–108. [[CrossRef](#)]
45. Jing, Y.; Hu, H.T.; Guo, S.Y.; Wang, X.; Chen, F.Q. Short-Term Prediction of Urban Rail Transit Passenger Flow in External Passenger Transport Hub Based on LSTM-LGB-DRS. *IEEE T. Intell. Transp.* **2021**, *22*, 4611–4621. [[CrossRef](#)]
46. Yang, D.; Chen, K.R.; Yang, M.N.; Zhao, X.C. Urban rail transit passenger flow forecast based on LSTM with enhanced long-term features. *IET Intell. Transp. Sy.* **2019**, *13*, 1475–1482. [[CrossRef](#)]
47. Zhang, J.L.; Chen, F.; Cui, Z.Y.; Guo, Y.A.; Zhu, Y.D. Deep Learning Architecture for Short-Term Passenger Flow Forecasting in Urban Rail Transit. *IEEE T. Intell. Transp.* **2021**, *22*, 7004–7014. [[CrossRef](#)]
48. Chen, W.; Li, Z.P.; Liu, C.; Ai, Y. A Deep Learning Model with Conv-LSTM Networks for Subway Passenger Congestion Delay Prediction. *J. Adv. Transp.* **2021**, *2021*, 6645214. [[CrossRef](#)]
49. Shih, S.Y.; Sun, F.K.; Lee, H.Y. Temporal pattern attention for multivariate time series forecasting. *Mach. Learn* **2019**, *108*, 1421–1441. [[CrossRef](#)]
50. He, Y.X.; Li, L.S.; Zhu, X.T.; Tsui, K.L. Multi-Graph Convolutional-Recurrent Neural Network (MGC-RNN) for Short-Term Forecasting of Transit Passenger Flow. *IEEE T. Intell. Transp.* **2022**, 1–20. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.