ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20230227000384
Original scientific paper

Multi-Step Subway Passenger Flow Prediction under Large Events Using Website Data

Qun TU, Guining GENG, Qianqian ZHANG*

Abstract: An accurate and reliable forecasting method of the subway passenger flow provides the operators with more valuable reference to make decisions, especially in reducing energy consumption and controlling potential risks. However, due to the non-recurrence and inconsistency of large events (such as sports games, concerts or urban marathons), predicting passenger flow under large events has become a very challenging task. This paper proposes a method for extracting event-related information from websites and constructing a multi-step station-level passenger flow prediction model called DeepSPE (Deep Learning for Subway Passenger Flow Forecasting under Events). Experiments on the actual data set of the Beijing subway prove the superiority of the model and the effectiveness of website data in subway passenger flow forecasting under events.

Keywords: deep learning; intelligent transportation systems; multi-task learning; subway passenger flow forecasting

1 INTRODUCTION

To accurately perceive and accurately grasp the space-time distribution characteristics of passenger flow in a large-scale subway network is conducive to more scientific and reasonable intelligent decision and intelligent control [1]. Applying data mining technology and deep learning to the short-term passenger flow prediction of the subway can provide intelligent and accurate decision support for subway operations, thus improving the accuracy, timeliness, and effectiveness of subway management [2]. Passenger flow prediction is of great significance in the operation and management of subways, especially in reducing energy consumption and controlling potential risks. Accurate passenger flow prediction helps the operators to design better timetables to wisely pre-allocate resources to meet the demands. As a result, saving energy consumption from 5% to 30% [3]. Burst passenger flows for large events will quickly cause a huge traffic demand, which may pressure public security and exert a bad influence like a stampede [4]. For example, a serious stampede happened in Seoul on the Halloween Eve of 2022. The abnormal passenger flow caused by large events (such as sports games, concerts or urban marathons) often affects a few specific stations. However, due to its strong suddenness and large passenger flow, it often puts great pressure on urban rail transit operations and even leads to safety problems [5]. Predicting passenger flow under large events has become a very challenging task because of the non-recurrence and inconsistency of large events.

The challenge of subway passenger flow forecasting under large events is that capturing the event-related information from the commonly used data sources is difficult. Li et al. [6] proposed a multi-scale radial basis network to predict abnormal passenger flows at target station caused by large events by detecting irregular fluctuations in inbound passenger flow at several key stations. This method may be suitable for small urban rail transit networks. For a complex urban rail transit system (for example, the Beijing urban rail network has more than 21 lines and more than 400 stations), the large events at the target station are difficult to cause abnormal fluctuations at other stations of the whole network. Therefore, we should discover new data sources to provide event information to the model.

In the operation of the urban rail transit, only when a super-large event (such as the World Cup or the Olympic Games) occurs will the event organizer provide the urban rail operator with event information to adjust the operation plan and timetable. In recent years, the development of the Internet has provided us with a new channel for mining information related to urban traffic. The Internet, including various websites, social media, and online forums, is flooded with massive amounts of information published by users, companies, and institutions [7]. This information may include traffic conditions, accidents, weather, road construction information, news of major events, and more. Through the Internet, we can know the information about large events that have happened, are happening and even will happen.

This paper uses Internet information as a new data source to mine information related to large events and also provides a method for feature representation. Next, we construct a station-level passenger flow forecasting model, which can be used to predict the subway passenger flow for multiple future time steps under large events. Our prediction results for passenger flow during large events in the Beijing Gongti Area show that the prediction performance of this model is better than baseline models.

2 RELATED WORK

In recent years, deep learning models have achieved extraordinary performance in fields such as natural language processing, computer vision, and speech recognition, surpassing traditional machine learning models [8]-[11]. Researchers began to apply it in the field of traffic prediction, hoping to further improve the model performance with its powerful nonlinear fitting and deep feature expression abilities [12]-[14]. A Stacked Autoencoder (SAE) neural network was proposed for traffic flow prediction to obtain general traffic flow feature representations from raw data [15]. Liu et al. [16] used Stacked Autoencoders Deep Neural Networks (SAE-DNN) for traffic prediction. They visualized the output of the model's intermediate layers to demonstrate the deep network's feature learning ability. In 2014, Huang et al. [17] proposed a Deep Belief Network (DBN) consisting of a deep belief network and a multi-task regression layer. Ma et al. [18] used Long Short-Term Memory Neural Network (LSTM) for traffic prediction to improve the problem that recurrent neural networks tend to cause vanishing gradients. Liu et al. [13] combined deep learning methods with subway domain knowledge to build an end-to-end flexible subway passenger prediction model.

To combine the advantages of different models, some researchers have developed a variety of hybrid models. Zhang et al. [19] proposed a deep learning framework for subway passenger flow prediction that combines Residual Network (ResNet), Graph Convolutional Network (GCN) and LSTM. Wei et al. combined the empirical mode decomposition method with a neural network to predict subway passenger flow [20]. Based on LSTM and wavelet, Yang et al. proposed a novel Wave-LSTM traffic prediction model [21]. Jing et al. combined three models, LSTM, gradient decision boosting tree and K-nearest neighbour algorithm to achieve a better performance in passenger flow prediction [22]. Zheng et al. [23] proposed a two-phase framework that mines the spatiotemporal disturbances of multiple contextual factors for citywide traffic flow prediction. The attention mechanism performing well for image recognition and natural language processing was also used as a technology for helping neural networks [24], [25]. Li et al. applied attention mechanism to capture the different weights of features extracted from former layers of neural networks [26].

In addition to Li et al., some other scholars also have tried to build complex models to predict the passenger flow under large events. Xue et al. built a hybrid deep neural network to model three parts of passenger flow under event and add three results as final prediction [27]. Zhao et al. proposed a prediction model that combines gradient boosting decision tree and deep learning [28].

The application of Website data in traffic models is not uncommon. Researchers have noticed the value of Internet data since Schweitzer et al. [29] mined public opinion information on bus routes on Twitter. For example, Gröflin et al. [30] used technical method to automatically capture and observe a large amount of information about local public transportation on the network and visualize it so as to infer whether some public transportation delays are occasional or systematic problems. Terpstra et al. [31] used the special event information, weather, and traffic data to develop a data assimilation technique using the Kalman filter, considering the difference between the current observation and the predicted state to predict future values. Pereira et al. [32] studied the public transportation system in Singapore and proposed a BP neural network prediction model using event information, weather conditions, and holidays. Ni et al. [33] discovered and verified the correlation between social media data and passenger flow and proposed a Hashtag-based event recognition method to predict the passenger flow under events. However, these models focus on the relationship between event information and passenger flow, ignoring the complex spatiotemporal correlation of passenger flow itself. These models can only predict the public transport passenger flow of arriving. They cannot predict the return passenger flow after the event. In addition, although Pereira et al. used activity information, they simply input the two pieces of information of activity location and time into the model.

To sum up, there are some researches on passenger flow forecasting under large events. However, most researches focus on the development of models, applying complex models or hybrid models to historical passenger flow data, and rarely explore new data sources. Even though some studies have developed and utilized new data sources, these studies tend to focus too much on the correlation between new data and abnormal passenger flow. In order to explain the prediction results, most of them choose relatively simple linear models or machine learning models. This paper attempts to use new data sources to capture irregular fluctuations and build a deep learning model to fit complex spatiotemporal features of passenger flow under events.

3 MODEL DEVELOPMENT

Based on LSTM, this paper proposes an end-to-end deep learning architecture--DeepSPE (Deep Learning for Subway Passenger Flow Forecasting under Events, Fig. 1) to predict abnormal passenger flow caused by large events. The model considers three features related to passenger flow under events. The first is external disturbance features that can be learned from website data, weather, and holidays. The second is spatial related features that can be learned from historical passenger flow. The third is temporal related features that can be learned from inbound or outbound passenger flows at other stations. We use three modules to extract the above three features, respectively, and then fuse them. Finally, based on multi-task learning, we use multiple output modules to output the prediction results of multi-steps passenger flow in the future.

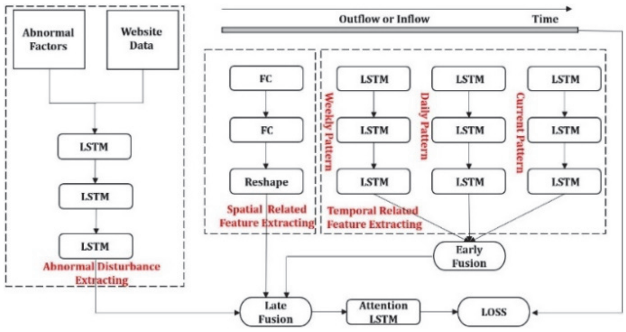

Figure1 Overall model architecture

3.1 Abnormal Disturbance Feature Extracting

Predicting passenger flow under large events is challenging because the information is often scattered among otherwise unrelated data sources. External factors such as weather and holidays often interact with large event, making it more difficult for the model to understand passenger flow. For example, two concerts of similar size, one is held on weekdays, and the other is held on weekends. If we only input large event information to the model, the model will think that the disturbance flow caused by the two concerts is the same. However, the true situation is that the difference in holding time will cause obvious differences in disturbance flow.

Therefore, to fully grasp the various types of information that disturb the passenger flow, we should consider external factors such as weather, holidays, and time in addition to the features related to activity information. Supposing we need to input the disturbance features of T_R time steps into the model to predict the future passenger flow of T time steps. Then the input of the external disturbance feature extracting module is $R_t \in \mathbb{R}^{T_R \times D_R}$.

Embedding technique is widely used in natural language processing to solve the problem of sparsity caused by one-hot encoding, by mapping sparse vectors into dense vectors. Embedding can be represented as a map as follows:

$$Y = E\left(X\right) \tag{1}$$

where the function $E\left(\cdot\right)$ is injective and has the feature of preserving structure (for example, $X_1 < X_2$ in the space where X is located; also $Y_1 < Y_2$ when mapped to the Y space). To deal with categorical features, an embedding layer is constructed with a parameter matrix $W \in R^{I \times O}$, where I represents the number of categories and O represents the dimension of the embedding space. In our model, we do not embed for a certain type of one-hot encoding but use an embedding layer to embed the input vector as a whole. In a forward propagation calculation, we conduct T_R operations on input R_t with the embedding layer and combine the T_R results to obtain the output $E_t \in \mathbb{R}^{T_R \times D_E}$.

To make predictions about passenger flow for the next T steps, we feed E_t into a many-to-many LSTM network. The network structure consists of an encoder and a decoder. Specifically, for the abnormal disturbance module, we input E_t to the model, and the model outputs $R_t \in \mathbb{R}^{T \times F}$, where T is the number of future steps to be predicted. F stands for feature dimension.

3.2 Temporal Related Feature Extracting

3.2.1 Temporal Related Feature Modeling

Existing passenger flow studies mostly use current pattern to fit future passenger flows, and the input features of current pattern are expressed as $N_t = \left[y_{t-1}, y_{t-2}, \ldots, y_{t-c}\right]$, where c represents the selected historical passenger flow time steps. In addition to current pattern, we also model daily pattern and weekly pattern. The daily pattern input feature is expressed as $D_t = \left[y_{t-d}, y_{t-2d}, \ldots, y_{t-nd}\right]$, where n represents the selected number of historical passenger flow time steps, and d represents a fixed time of one day interval, y_{t-d} represents the same moment of the previous day. The weekly pattern input feature is expressed as $L_t = \left[y_{t-w}, y_{t-2w}, \ldots, y_{t-mw}\right]$, where m represents the selected historical passenger flow time steps, and w represents a fixed one-week time interval, y_{t-w} represents the same moment in the previous week.

Different from the network-wide prediction model, the single station prediction model only needs to predict the passenger flow of the target station. In order to avoid introducing too much interference and increasing the complexity of the model, we do not need to input the passenger flow of all stations in the network. Our input in this part is $y_t \in \mathbb{R}^{1+N}$, where 1 represents the target station, and N represents the N stations (called "related stations" in this paper) that are most similar to the target station's passenger flow pattern. Compared with using only the target station features, the benefit of adding related stations is that it can provide more information and improve the generalization ability of the model. We use the Dynamic Time Warping (DTW) distance to measure the distance between the target station's passenger flow curve and other stations' passenger flow curves, then select N stations with the closest distance to the target station.

In this paper, in order to fit the future passenger flow of T time steps, three LSTM networks are used to extract temporal related features in three patterns. The outputs of the three networks are $N_t \in \mathbb{R}^{T \times F}$, $D_t \in \mathbb{R}^{T \times F}$ and $L_t \in \mathbb{R}^{T \times F}$.

3.2.2 Dynamic Time Warping

DTW is mostly used in fields such as speech recognition and data mining. The algorithm was originally designed to calculate the similarity between two time series based on temporal heterogeneity [34]. Suppose there are two time series data X and Y. They represent the passenger flow curve I and passenger flow curve II, respectively.

$$X = x_1, x_2, \ldots, x_i, \ldots, x_I \tag{2}$$

$$Y = y_1, y_2, \ldots, y_i, \ldots, y_J \tag{3}$$

To find the best alignment between X and Y, a twisted path W is introduced, as shown below.

$$W = w_1, w_2, \ldots, w_l, \ldots w_L \tag{4}$$

where $\max\left(I, J\right) \le L < I + J$, L represents the length of W, and $w_l = \left(i, j\right)$ represents the l^{th} element of W. The twisted path W defines a mapping from X to Y that satisfies the following three conditions.

Boundary conditions: $w_1 = \left(1, 1\right)$, $w_L = \left(I, J\right)$, ensuring that every element in X and Y is taken into account.

Monotonicity condition: known $w_l = \left(i, j\right)$, $w_{l-1} = \left(i', j'\right)$, where $i - i' \ge 0$, $j - j' \ge 0$, guarantee there is no reverse path in W.

Continuity condition: known $w_l = \left(i, j\right)$, $w_{l-1} = \left(i', j'\right)$, where $i - i' \le 1$, $j - j' \le 1$, guarantee only adjacent elements are considered in calculation.

Constrained by the three conditions, the optimal path between X and Y can be found in the form of the optimal twisted path, which has the smallest twisted cost among all possible paths. In the twisted path w_l, the twist cost which represents the distance between the i^{th} element in X and the j^{th} element in Y is calculated by the following equation:

$$T_{Dist}\left(i, j\right) = \left(x_i - y_j\right)^2 \tag{5}$$

Based on $T_{Dist}\left(i, j\right)$, the DTW distance representing the shortest twisted path between X and Y can be obtained by the following loop.

$$T_D(i,j)_{\mathrm{I,\,II}} = T_{Dist}(i,j) + \\ +\min\left[T_D(i-1,j)_{\mathrm{I,\,II}},\ T_D(i,j-1)_{\mathrm{I,\,II}},\ T_D(i-1,j-1)_{\mathrm{I,\,II}}\right] \quad (6)$$

where $T_D(i,j)_{\mathrm{I,\,II}}$ represents the DTW distance between passenger flow curve I and passenger flow curve II. The DTW distance inspired us to measure the similarity of the two passenger flow curves. The shorter the DTW distance, the more similar the two passenger flow curves. In the temporal related feature fitting module, we select those stations that have the closest DTW distance with target station as related station to help the model learn the regular part of the passenger flow of the target station.

3.3 Spatial Related Feature Extracting

Regarding the spatial related subway passenger flow under large events, we consider two aspects: when the predicted target is outbound passenger flow, the spatial related passenger flow originates from the passenger flow that enters the station from other stations and will exit from the target station; when the predicted target is the inbound passenger flow, the spatial related passenger flow comes from the passenger flow that leaves the target station before the event starts and will return to the target station after the event.

3.3.1 Spatial Related Feature Extracting of Out-Flow

Assuming that the prediction target is outbound passenger flow, we adopt a method based on origin and destination (OD) passenger flow and average travel time to construct features. The steps are as follows:

In the first step, for the target station v and the inbound station k, select the K stations with the most OD passenger flow.

In the second step, calculate the average travel time from k to v:

$$T^{k,\,v} = \sum_{i=1}^{I}\left(t_i^v - t_i^k\right)/\,I \quad (7)$$

where $T^{k,\,v}$ represents the average travel time of the two stations, t_i^v represents the time for the passenger i to get on the train from the station v, t_i^k represents the time for the passenger i to get off from the station k, and I represents the total number of passengers.

The third step is to calculate on which time steps that the inbound passenger flow will directly affect the outbound passenger flow, assuming that the predicted target outbound passenger flow is at the m^{th} time step.

$$r = m - T^{k,\,v}\,/\,U \quad (8)$$

where $\lceil\cdot\rceil$ represents rounding up, and U represents the time granularity.

The fourth step is to select the inbound passenger flow of station u at time steps r, $r+1$, …, $r+e$ as related inbound passenger flow, where $0 \le e \le m - r - 1$.

We form a one-dimensional vector of all related inbound passenger flows selected from K stations and connect them with two fully connected layers, where the last fully connected layer has $T \times F$ neurons. We then connect it with a Reshape layer to get the output $S_t \in \mathbb{R}^{T\times F}$.

3.3.2 Spatial Related Feature Extracting of In-Flow

Wang et al. [4] discovered an interesting phenomenon of "gathering slowly and dispersing urgently" by observing the change of passenger flow at subway stations during large events. The so-called "gather slowly and disperse quickly" phenomenon refers to the fact that after the event, passengers who had previously taken subway to the venue will return to the station to leave, resulting in abnormal outbound passenger flow before the event and more concentrated abnormal inbound passenger flow after the event. This paper counts the subway passenger flow data under 11 large events in the Beijing Gongti Area from November 4, 2017 to March 24, 2018. It is found that more than 65% of the inbound flow in 30 minutes after the event is from outbound flow of the same station. In this paper, the phenomenon of "gathering slowly and dispersing urgently" is applied to the prediction of abnormal inbound passenger flow caused by large events, and the abnormal inbound passenger flow is predicted by using the abnormal outbound passenger flow information that occurred earlier.

To select the most useful outbound passenger flow to construct a feature vector, we observe and count the inbound and outbound passenger flow curves of Dongsishitiao Station after removing the average value in these 11 event days. Assuming that the time granularity U = 10, we find that the duration of all abnormal outbound curves is between 15 and 20 time steps. The abnormal outbound peaks and abnormal inbound peaks are between 20 and 30 time steps. Therefore, in the case of 10-minute time granularity, in order to predict the inbound passenger flow at time t we should select the outbound passenger flow between $t-40$ and $t-10$ to construct the input features. In order to facilitate model learning, we input the outbound passenger flow minus the historical average value, so that when there is no abnormal outbound passenger flow, each feature value is close to 0, indicating that there will be no abnormal inbound passenger flow during the forecast period. When there is non-zero values, it means that there may be abnormal inbound passenger flow in the predicted time period. y_t' represents the outbound passenger flow minus the historical average at time t. The input feature is $S_t = \left[y_{t-10}', y_{t-11}', \ldots, y_{t-40}'\right]$. We also use two fully connected layers and one reshape layer for feature extracting, and get the output $S_t \in \mathbb{R}^{T\times F}$.

3.4 Prediction

3.4.1 Feature Fusion

In the early fusion stage, we use the weight matrix to fuse the three small branches under the branch of temporal related feature extracting. The process is as follows:

$$ES_t = W_C \odot C_t + W_D \odot D_t + W_L \odot L_t \quad (9)$$

where, $\odot$ represents the Hadamard product, and W_C, W_D, W_L represent the weight matrices of different small branches, respectively.

In the later fusion stage, we use a concatenate layer to concatenate the outputs of each module. The process is as follows:

$$LS_t = ES_t \oplus S_t \oplus R_t \tag{10}$$

where $\oplus$ represents merge operation, which is performed on the feature dimension of the tensor, the merged result is $LS_t \in \mathbb{R}^{T\times 3F}$. To accomplish the multi-task learning of predicting T time steps, we operate T attention LSTMs on LS_t separately, each of which outputs a prediction value $\hat{y}_t$.

3.4.2 Attention Mechanism

Due to the successful application of the attention mechanism in the field of machine translation, many scholars have introduced it to traffic prediction [35]. In the output stage of our model, T LSTMs are required to operate on the feature LS_t and generate predictions. LS_t is a tensor that incorporates different information at different time steps (including historical passenger flow, holidays, weather, large events, etc.). Adding an attention mechanism to LSTM allows the model to assign different weights to each time step and feature according to different situations when extracting features from LS_t, helping the model to pay attention to key features. Since the output layer structure is many to one, we choose to add attention before the LSTM operation. The attention mechanism can be implemented by a fully connected layer whose activation function is SoftMax. The output of the fully connected layer is multiplied by the input of the fully connected layer to complete the allocation of attention weights. Suppose the matrix $Out \in R^{m\times n}$ represents the input before adding attention, where m represents the time step and n represents the number of features at each time step.

$$A = f\left(W \cdot Out + b\right) \tag{11}$$

$$Out' = A \cdot Out \tag{12}$$

where out' is the attention-based output, A is the weight matrix with the same shape as Out, "$\cdot$" represents the Hadamard product, f represents the fully connected layer, W is the weight matrix of f, and b is the bias.

3.4.3 Learning Process

In order to solve the optimization problem of the multi-task learning model for short-term subway passenger flow prediction under large event subways, this paper adopts the mean square error (MSE) as the loss function.

$$Loss = MSE = \frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(y_i^j - \hat{y}_i^j\right) \tag{13}$$

The objective function of parameter learning is defined as follows:

$$\theta = \mathrm{argmin}_{\theta} Loss \tag{14}$$

where y_i^j represents the true value of the i^{th} sample at the j^{th} moment, $\hat{y}_i^j$ represents the predicted value of the i^{th} sample at the j^{th} moment, M is the number of samples, N is the number of tasks, and θ is the model parameter. This paper applies the Adam [36] optimizer to the back-propagation algorithm for parameter learning.

4 EXPERIMENTS AND DISCUSSION

4.1 Data Description

Event data. This study takes Dongsishitiao Station on Line 2 as target stations. Dongsishitiao Station is adjacent to Beijing's two main large event venues, Beijing Workers' Stadium and Beijing Workers' Gym. We collect event information from ticketing websites. The time frame is from October 27, 2017, to May 31, 2018. During this period, 11 large-scale activities such as football games and concerts were held in this area.

Passenger flow data. The smart card data are provided by Beijing JingtouYiyajie Transportation Technology Co., Ltd., which is one of Beijing's urban rail transit operators. In this study, the smart card data of Beijing subway is from October 27, 2017, to May 31, 2018, a total of 216 days, and more than 1.5 billion travel records. There are 17 lines and 287 stations included in the data. We calculate the passenger flow data of each station at different times according to the records.

Meteorological data. The meteorological data are provided by the Beijing Municipal Meteorological Bureau, including hourly climate data for 216 days from October 27, 2017, to May 31, 2018. Meteorological indicators include precipitation and weather phenomena.

4.2 Parameter Setting and Comparison Experiment

4.2.1 Parameter Setting

In this paper, we set the passenger flow time granularity as U = 10 minutes and T = 6 tasks for multi-task learning, predicting the passenger flow in the next six-time steps for 60 minutes.

The disturbance feature and the three temporal correlation feature extraction modules use the LSTM structure introduced in Section 4.2. The encoder part consists of two layers of LSTM. The two layers have 32 neurons each, and the decoder part is one layer with eight neurons. The spatial correlation feature extraction part is a two-layer fully connected layer with 64 and 48 neurons, respectively. The attention LSTM of the final output layer consists of two layers with 16 neurons and one neuron, respectively.

In order to handle the temporal related feature in the 3 modes, we set c = 6 for the current pattern (e.g. $\left[x_{t-1}^{in}, x_{t-2}^{in}, \ldots, x_{t-6}^{in}\right]$), set n = 3 for daily pattern (e.g. $\left[x_{t-1d}^{in}, x_{t-2d}^{in}, x_{t-4d}^{in}\right]$), set m = 3 for weekly pattern (e.g. $\left[x_{t-1w}^{in}, x_{t-2w}^{in}, x_{t-3w}^{in}\right]$). We set the number of related stations to eight, then the input of each time step in the temporal related feature extracting part is a nine-dimensional vector (for example, $x_{t-1}^{in} = \left[x_{t-1}^{in_t}, x_{t-1}^{in_1}, x_{t-1}^{in_2}, \ldots, x_{t-1}^{in_8} \right]$).

The evaluation metrics are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\left(y_i - \hat{y}_i\right)\right| \tag{16}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - \hat{y}_i\right|}{y_i} \tag{17}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of samples.

4.2.2 Baseline Models

We use five baseline models for comparative experiments. These models adopt multi-task learning or iterative methods to implement multi-step forecasting strategies to predict the passenger flow at multiple time steps in the future.

Integrated Moving Average Autoregression (ARIMA): A representative traditional mathematical-statistical model. An iterative multi-step forecasting strategy is employed.

Support Vector Regression (SVR): A classic supervised learning method. An iterative multi-step forecasting strategy is employed using Gaussian kernel-based general support vector regression.

Random Forest (RF): A classic ensemble learning model that averages the outputs of multiple decision trees as the result. An iterative multi-step forecasting strategy is employed.

Deep Neural Network (DNN): A neural network composed of multiple fully connected layers, which adopts a multi-step prediction strategy of multi-task learning.

Gated Recurrent Unit (GRU): A variant of recurrent neural networks that employs a multi-step prediction strategy of multi-task learning.

4.3 Prediction Results

4.3.1 Comparison with Baseline Models

In Tab. 1 and Fig. 2, we compare the prediction effect of each model for different future time steps through three indicators: RMSE, MAE and MAPE. From Tab. 1 and Fig. 2, we can see that the DeepSPF proposed in this paper is better than the other five baseline models in terms of prediction performance, whether in the comparison of different time steps or different indicators. In terms of model types, the overall performance of the neural network-based models is better than that of traditional machine learning models and statistical models, indicating that the neural network-based models have a strong ability to fit subway passenger flow data under large events. The GRU model outperforms the DNN model due to considering temporal correlations. DeepSPF outperforms GRU because it considers regular partial passenger flow and abnormal partial passenger flow, respectively, and enhances the accuracy and robustness of the prediction of these two parts.

Table 1 Comparison of prediction performance of different models at different time steps

Periods	Metrics	ARIMA	SVR	RF	DNN	GRU	DeepSPE
10 min	RMSE	54.67	47.81	56.16	41.69	40.88	**28.54**
	MAE	27.39	29.98	30.43	22.87	22.12	**15.05**
	MAPE	18.92%	22.28%	21.03%	14.07%	13.32%	**11.78%**
20 min	RMSE	56.76	58.29	62.31	48.44	44.11	**30.38**
	MAE	27.93	39.61	35.23	24.62	23.56	**16.13**
	MAPE	19.02%	23.65%	22.12%	14.65%	14.31%	**12.03%**
30 min	RMSE	57.02	76.74	79.02	50.32	46.87	**31.76**
	MAE	28.65	48.32	39.44	25.66	24.62	**17.29**
	MAPE	19.76%	28.87%	25.98%	15.12%	14.84%	**12.76%**
40 min	RMSE	57.65	103.15	90.54	55.83	52.23	**38.10**
	MAE	28.47	61.96	45.77	26.98	25.31	**19.52**
	MAPE	19.45%	32.14%	27.76%	16.65%	16.29%	**13.22%**
50 min	RMSE	57.87	142.17	105.60	61.27	60.26	**41.44**
	MAE	28.98	78.08	50.99	29.87	28.96	**21.87**
	MAPE	20.54%	35.26%	29.65%	17.07%	16.87%	**14.65%**
60 min	RMSE	58.98	176.65	122.47	68.25	63.52	**45.34**
	MAE	30.76	89.02	55.01	32.14	30.33	**23.98**
	MAPE	20.07%	37.97%	31.23%	18.68%	18.31%	**15.09%**
Mean	RMSE	57.16	100.8	86.02	54.3	51.31	**35.93**
	MAE	28.7	57.83	42.81	27.02	25.82	**18.97**
	MAPE	19.63%	30.03%	26.30%	16.03%	15.66%	**13.26%**

4.3.2 Comparison between Different Variants

To verify the effectiveness of the different modules of DeepSPE, we split the temporal related passenger flow fitting module, the abnormal disturbance passenger flow fitting module and the spatial related passenger flow fitting module to make different combinations. Tab. 2 shows the seven different DeepSPE variant models.

Tab. 3 shows the prediction results of seven DeepSPE variant models on passenger flow in different future periods of Dongsishitiao Station. Tab. 3 shows that variant 7 (DeepSPE) is the best of all variants. Moreover, the comparison of the prediction metrics of each variant shows that the temporal related feature fitting module is the most important part. Without the temporal related feature fitting module, the accuracy dropped significantly. The abnormal disturbance fitting module and the spatial related feature fitting module cannot capture the accurate passenger flow trend individually or in combination. These two parts must be combined with the temporal related passenger flow fitting module to take full advantage.

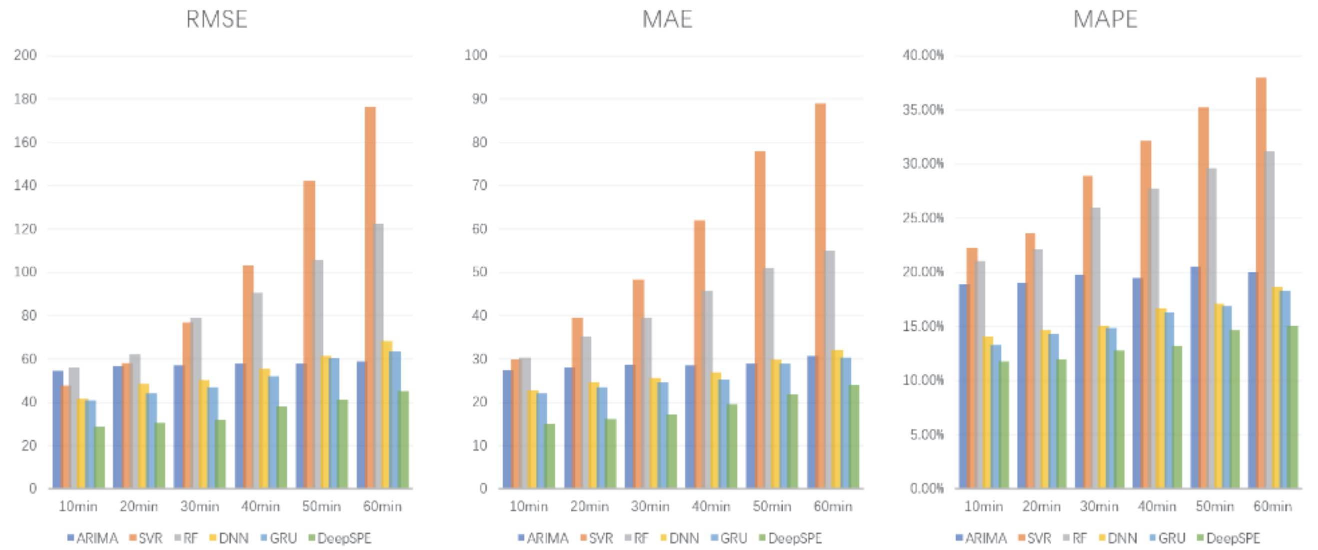

Figure 2 Comparison of prediction performance of different models at different time steps

Table 2 Deep SPF variant models

Variants	Temporal Related Feature	Abnormal Disturbance Feature	Spatial Related Feature
V1	•		
V2		•	
V3			•
V4	•	•	
V5	•		•
V6		•	•
V7	•	•	•

Table 3 Comparison of prediction effects of different variant models at different prediction time steps

Periods	Metrics	V1	V2	V3	V4	V5	V6	V7
10 min	RMSE	42.01	169.77	143.31	29.96	38.24	122.65	**28.54**
	MAPE	14.33%	51.02%	44.89%	12.05%	13.25%	46.98%	**11.78%**
20 min	RMSE	48.85	245.68	187.76	31.44	42.65	165.98	**30.38**
	MAPE	14.87%	75.23%	62.07%	12.67%	13.88%	52.89%	**12.03%**
30 min	RMSE	50.43	228.98	203.23	33.05	42.76	152.56	**31.76**
	MAPE	15.33%	68.71%	59.98%	13.24%	14.02%	49.55%	**12.76%**
40 min	RMSE	55.74%	307.72	199.08	40.98	49.17	174.38	**38.10**
	MAPE	16.09%	86.25%	59.73%	14.12%	14.89%	56.69%	**13.22%**
50 min	RMSE	62.03	316.76	278.54	43.51	54.0	198.87	**41.44**
	MAPE	17.66%	88.65%	66.23%	15.22%	16.12%	67.65%	**14.65%**
60 min	RMSE	67.54	307.54	189.97	48.98	59.77	126.76	**45.34**
	MAPE	18.72%	112.13%	78.01%	16.55%	17.36%	52.22%	**15.09%**
Mean	RMSE	45.24	262.74	200.32	37.99	47.77	156.87	**35.93**
	MAPE	16.17%	80.33%	61.82%	13.98%	14.92%	54.33%	**13.26%**

5 CONCLUSION

In this paper, we propose a station-level subway passenger flow prediction method for large events, DeepSPE, which innovatively adopts a multi-task learning strategy to predict the future multi-step passenger flow. We apply website data to subway passenger flow prediction under large events and extract information related to large events from it. In terms of model construction, we extract abnormal disturbance features, temporal related features and spatial related features based on the LSTM and fully connected layer. After merging the features of each module, we use multiple attention LSTM to output predictions at each time step. Since the prediction target is single station-level passenger flow, it allows us to do a lot of feature engineering. For example, we use DTW to select stations with similar passenger flow patterns to the target station, or use average travel time to select inbound passenger flows that are closely related to outbound traffic flow, and use de-averaged outbound passenger flow as input, etc. Although deep learning can automatically extract useful features from raw data, good feature engineering can allow models to solve problems with fewer data resources, especially when there are few abnormal samples, feature engineering can improve the informative value of data.

By conducting experiments on real data sets, this paper draws the following main conclusions:

(1) The prediction performance of the DeepSPE model at multiple time steps is better than that of traditional statistical models, machine learning models and general deep learning models. DeepSPE maintains a sound forecasting performance even if it encounters abnormal passenger flow caused by large events.

(2) Abnormal disturbance feature extracting module, temporal related feature extracting module, and spatial related extracting module all play important roles in passenger flow prediction under large events. The lack of any module will greatly reduce the prediction effect of the model.

(3) The selection of related stations adds more information to the model and improves the generalization ability, but the number of related stations is not as good as possible. Too many related stations will introduce redundant information and affect the prediction precision.

For operators, the short-term passenger flow forecast results can be used to balance the demand and the traffic volume, change the timetable, allocate more staff to avoid congestion accidents, save operating costs, and ensure the service level of the rail system. For the general public, the short-term passenger flow forecast results can be used to grasp the real-time spatial and temporal distribution characteristics of passenger flow, reasonably choose the travel route and travel time and avoid congested routes.

Acknowledgments

This research was supported by the Fundamental Research Funds for the Central Universities (JD2330), and Funds for First-class Discipline Construction in Beijing University of Chemical Technology (XK1802-5).

6 REFERENCES

[1] Li, H., Jin, K., Sun, S. et al. (2022). Metro passenger flow forecasting though multi-source time-series fusion: An ensemble deep learning approach. *Applied Soft Computing*, *120*, 108644. https://doi.org/10.1016/j.asoc.2022.108644

[2] Tu, Q., Zhang, Q., Zhang, Z., Gong, D., & Tang, M. (2022). A Deep Spatiotemporal Fuzzy Neural Network for Subway Passenger Flow Prediction with COVID-19 Search Engine Data. *IEEE Transactions on Fuzzy Systems*, *31*(2), 394-406. https://doi.org/10.1109/TFUZZ.2022.3179779

[3] Su, S., Li, X., Tang, T., & Gao, Z. (2013). A subway train timetable optimization approach based on energy-efficient operation strategy. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), 883-893. https://doi.org/ 10.1109/TITS.2013.2244885

[4] Wang, H., Li, L., Pan, P., Wang, Y., & Jin, Y. (2019). Early warning of burst passenger flow in public transportation system. *Transportation Research Part C: Emerging Technologies*, *105*, 580-598. https://doi.org/10.1016/j.trc.2019.05.022

[5] Ni, M., He, Q., & Gao, J. (2016). Forecasting the subway passenger flow under event occurrences with social media. *IEEE Transactions on Intelligent Transportation Systems*, *18*(6), 1623-1632. https://doi.org/10.1109/TITS.2016.2611644

[6] Li, Y., Wang, X., Sun, S., Ma, X., & Lu, G. (2017). Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. *Transportation Research Part C: Emerging Technologies*, *77*, 306-328. https://doi.org/10.1016/j.trc.2017.02.005

[7] Salminen, J., Hopf, M., Chowdhury, S. A., Jung, S. G., Almerekhi, H., & Jansen, B. J. (2020). Developing an online hate classifier for multiple social media platforms. *Human-centric Computing and Information Sciences*, *10*, 1-34. https://doi.org/10.1186/s13673-019-0205-6

[8] Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., & Yin, B. (2021). Deep learning on traffic prediction: Methods, analysis, and future directions. *IEEE Transactions on Intelligent Transportation Systems*, *23*(6), 4927-4943. https://doi.org/10.1109/TITS.2021.3054840

[9] Xiwen, Q. I. N., Dingxin, X. U., & Jiajing, G. U. O. (2021). Correlation analysis and prediction of stock based on vmd-lasso model. *Economic Computation & Economic Cybernetics Studies & Research*, *55*(3). https://doi.org/10.24818/18423264/55.3.21.16

[10] Sarkar, T., Salauddin, M., Hazra, S. K., Choudhury, T., & Chakraborty, R. (2021). Comparative approach of artificial neural network and thin layer modelling for drying kinetics and optimization of rehydration ratio for bael (Aegle marmelos (L) correa) powder production. *Economic Computation & Economic Cybernetics Studies & Research*, *55*(1). https://doi.org/10.24818/18423264/55.1.21.11

[11] Yun, S. (2022). Predictive Performance of Building Construction Estimation: An Analysis based on ANN Model. *Journal of System and Management Sciences*, *12*(2), 325-335. https://doi.org/10.33168/JSMS.2022.0216

[12] Yang, J. M., Peng, Z. R., & Lin, L. (2021). Real-time spatiotemporal prediction and imputation of traffic status based on LSTM and Graph Laplacian regularized matrix factorization. *Transportation Research Part C: Emerging Technologies*, *129*, 103228. https://doi.org/10.1016/j.trc.2021.103228

[13] Bau, Y. T., Sasidaran, T., & Goh, C. L. (2022). Improving Machine Learning Algorithms for Breast Cancer Prediction. *Journal of System and Management Sciences*, *12*(4), 251-266. https://doi.org/10.33168/JSMS.2022.0416

[14] Teoh, C. W., Ho, S. B., Dollmat, K. S., & Tan, C. H. (2022). Predicting Student Performance from Video-Based Learning System: A Case Study. *Journal of Logistics, Informatics and Service Science*, *9*(3), 64-77. https://doi.org/10.33168/LISS.2022.0306

[15] Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2014). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, *16*(2), 865-873. https://doi.org/10.1109/TITS.2014.2345663

[16] Liu, L. & Chen, R. C. (2017). A novel passenger flow prediction model using deep learning methods. *Transportation Research Part C: Emerging Technologies*, *84*, 74-91. https://doi.org/10.1016/j.trc.2017.08.001

[17] Huang, W., Song, G., Hong, H., & Xie, K. (2014). Deep architecture for traffic flow prediction: Deep belief networks with multitask learning. *IEEE Transactions on Intelligent Transportation Systems*, *15*(5), 2191-2201. https://doi.org/10.1109/TITS.2014.2311123

[18] Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies, 54*, 187-197. https://doi.org/10.1016/j.trc.2015.03.014

[19] Zhang, J., Chen, F., Cui, Z., Guo, Y., & Zhu, Y. (2020). Deep learning architecture for short-term passenger flow forecasting in urban rail transit. *IEEE Transactions on Intelligent Transportation Systems*, *22*(11), 7004-7014. https://doi.org/10.1109/TITS.2020.3000761

[20] Tang, T., Fonzone, A., Liu, R., & Choudhury, C. (2021). Multi-stage deep learning approaches to predict boarding behaviour of bus passengers. *Sustainable Cities and Society*, *73*, 103111. https://doi.org/10.1016/j.scs.2021.103111

[21] Yang, X., Xue, Q., Yang, X., Yin, H., Qu, Y., Li, X., & Wu, J. (2021). A novel prediction model for the inbound

passenger flow of urban rail transit. *Information Sciences*, *566*, 347-363. https://doi.org/10.1016/j.ins.2021.02.036

[22] Jing, Y., Hu, H., Guo, S., Wang, X., & Chen, F. (2020). Short-term prediction of urban rail transit passenger flow in external passenger transport hub based on LSTM-LGB-DRS. IEEE Transactions on Intelligent Transportation Systems, *22*(7), 4611-4621. https://doi.org/10.1109/TITS.2020.3017109

[23] Zheng, C., Fan, X., Wen, C., Chen, L., Wang, C., & Li, J. (2019). DeepSTD: Mining spatio-temporal disturbances of multiple context factors for citywide traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, *21*(9), 3744-3755. https://doi.org/10.1109/TITS.2019.2932785

[24] Do, L. N., Vu, H. L., Vo, B. Q., Liu, Z., & Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation research part C: emerging technologies*, *108*, 12-28. https://doi.org/10.1016/j.trc.2019.09.008

[25] Liu, Y., Bao, Z., Zhang, Z., Tang, D., & Xiong, F. (2020). Information cascades prediction with attention neural network. *Human-centric Computing and Information Sciences*, *10*(1), 1-16. https://doi.org/10.1016/j.trc.2019.09.008

[26] Li, H., Zhang, J., Yang, L., Qi, J., & Gao, Z. (2022). Graph-GAN: A spatial-temporal neural network for short-term passenger flow prediction in urban rail transit systems. arXiv preprint arXiv:2202.06727. https://doi.org/ 10.48550/arXiv.2202.06727

[27] Xue, G., Liu, S., Ren, L., Ma, Y., & Gong, D. (2022). Forecasting the subway passenger flow under event occurrences with multivariate disturbances. *Expert Systems with Applications*, *188*, 116057. https://doi.org/10.1016/j.eswa.2021.116057

[28] Zhao, Y. & Ma, Z. (2022). Naïve Bayes-Based Transition Model for Short-Term Metro Passenger Flow Prediction under Planned Events. *Transportation Research Record*, *2676*(9), 309-324. https://doi.org/10.1177/03611981221086645

[29] Schweitzer, L. (2012). How are we doing? Opinion mining customer sentiment in us transit agencies and airlines via twitter. (No. 12-2659).

[30] Gröflin, A., Weber, M., Guggisberg, M., & Burkhart, H. (2017). Traffic flow measurement of a public transport system through automated Web observation. *2017 11th international conference on research challenges in information science (RCIS)*, 156-161. https://doi.org/10.1109/RCIS.2017.7956532

[31] Terpstra, F. P., Meijer, G. R., & Visser, A. (2004, August). Intelligent adaptive traffic forecasting system using data assimilation for use in traveler information systems. *The Symposium on Professional Practice in AI a stream within the First IFIP Conference on Artificial Intelligence Applications and Innovations AIAI-2004.*

[32] Pereira, F. C., Rodrigues, F., & Ben-Akiva, M. (2015). Using data from the web to predict public transport arrivals under special events scenarios. *Journal of Intelligent Transportation Systems*, *19*(3), 273-288. https://doi.org/10.1080/15472450.2013.868284

[33] Ni, M., He, Q., & Gao, J. (2016). Forecasting the subway passenger flow under event occurrences with social media. IEEE *Transactions on Intelligent Transportation Systems*, *18*(6), 1623-1632. https://doi.org/10.1109/TITS.2016.2611644

[34] Berndt, D. J. & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *KDD workshop*, *10*(16), 359-370.

[35] Zhang, Z., Li, M., Lin, X., Wang, Y., & He, F. (2019). Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. *Transportation research part C: emerging technologies*, *105*, 297-322. https://doi.org/10.1016/j.trc.2019.05.039

[36] Kingma, D. P. & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv: 1412.6980. https://doi.org/10.48550/arXiv.1412.6980

Contact information:

Qun TU
School of Economics and Management,
Beijing University of Chemical Technology,
Beijing 100029, China
No.15, Beisanhuandong Road, Chaoyang District, Beijing, China
E-mail: tuqun@buct.edu.cn

Guining GENG
360 Digital Security Technology Group Co., Ltd.,
Beijing 100015, China
No. 6, Jiuxianqiao Road, Chaoyang District, Beijing, China
E-mail: gengguining@360.cn

Qianqian ZHANG
(Corresponding author)
School of Information, Beijing Wuzi University,
Beijing 101149, China
No. 1 Fuhe Street, Tongzhou District, Beijing, China
E-mail: zhangqianqian@bwu.edu.cn