Expert Systems With Applications 188 (2022) 116057

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Forecasting the subway passenger flow under event occurrences with multivariate disturbances

Gang Xue [a], Shifeng Liu [a], Long Ren [b], Yicao Ma [a], Daqing Gong [a,*]

[a] School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China
[b] School of Information Technology and Management, University of International Business and Economics, Beijing 100029, China

ARTICLE INFO

Keywords:
Social media
Subway passenger flow prediction
Attention
Spatiotemporal disturbances

ABSTRACT

Subway passenger flow prediction is of great significance in transportation planning and operation. Special events, as for vocal concerts and sports games, lead large-scaled passenger flow with few periodic trends. Therefore, predicting subway outbound flow during events is a challenging task. In recent years, social media has been used for socio-economic forecasting. Correlation analysis shows that the trend of social media volume can be used for passenger flow prediction under events occurrences. In this paper, besides traditional smart card data, we incorporate social media data into passenger flow prediction. The multivariate disturbance-based hybrid deep neural network (MDB-HDNN), which models the disturbances of the inbound flow from the nearby stations and the social media post trends, is proposed for subway passenger flow prediction during events. Experimental results on three real-world datasets demonstrate that the MDB-HDNN performs well under various settings and has better robustness. Our findings and results can provide decision support for schedule formulation and passenger flow guidance.

1. Introduction

Urban rail transport systems are utilized as efficient measures to alleviate the detrimental effects of fast urbanization and traffic congestion all over the world (Ma, Tao, Wang, Yu, & Wang, 2015). Urban rail transport passenger flow prediction is of great significance in transportation planning and operation for intelligent transportation systems (ITSs) (Ding, Zhao, & Jiao, 2002). Accurate passenger flow prediction can help the planning, management, and operation of the transportation system (Chen & Wei, 2011). Most of the past research focused on predicting daily time series passenger flow (Leng, Zeng, Xiong, Lv, & Wan, 2013; Sun, Leng, & Guan, 2015; Sun et al., 2015). However, in regard to special events (e.g., concerts, running races, sporting events, etc.), as a result of their irregularity, predicting subway outbound flow during events is a challenging task (Ni, He, & Gao, 2016; Pereira, Rodrigues, & Ben-Akiva, 2015; Pereira, Rodrigues, Polisciuc, & Ben-Akiva, 2015). Predicting irregular passenger flow during special events has more difficulties than regular passenger flow predicting since the events are disturbed by many factors (Li, Wang, Sun, Ma, & Lu, 2017).

Regarding the prior research, the data used mainly included historical data and spatiotemporal features. However, social media, as a new type of data, can better contribute to predicting passenger flow, especially during special events. Contents released by users in social media strengthen the contacts of users while providing vast quantities of information (Ni et al., 2016). Hence, we can obtain effective and timely events, roadworks, and traffic flow information from the social media platform (Pereira, Rodrigues, & Ben-Akiva, 2015; Pereira, Rodrigues, Polisciuc et al., 2015). Fig. 1 shows the comparison between the number of social media posts and the outbound flow at Dongsishitiao (DSST) station on October 25, 2017, which was a day with a football game at 19:35; the ordinate represents the percentage of the value in the whole day.

It is believed that there are two main limitations of the latest approaches. First, novel deep learning methods perform well in regular short-term passenger flow prediction, but they ignore the impact of social media and are unable to cope with the impact of special events. Second, studies considering social media focus on mining the relationship between the text content and passenger flow, and the impact of disturbances (all stations' inbound flow) in the spatiotemporal view has not been fully addressed; inbound flows from other stations, which are considered to have potential contributions to the sudden subway

* Corresponding author.
E-mail addresses: 19113032@bjtu.edu.cn (G. Xue), shfliu@bjtu.edu.cn (S. Liu), renl@uibe.edu.cn (L. Ren), 19120617@bjtu.edu.cn (Y. Ma), dqgong@bjtu.edu.cn (D. Gong).

https://doi.org/10.1016/j.eswa.2021.116057
Received 19 June 2020; Received in revised form 3 April 2021; Accepted 8 October 2021
Available online 18 October 2021

congestion, should be addressed.

To alleviate the limitations aforementioned, we propose multivariate disturbance-based hybrid deep neural networks (MDB-HDNNs), a three-stage framework to uncover the disturbances by utilizing the sudden increase in inbound flows from other stations and social media to predict the outbound flow during special events. We evaluate the performance of our method on three real-world datasets under different events from Dongsishitiao (DSST), Tuanjiehu (TJH), and Olympic Park Station (OPS) (Three subway stations near stadiums where large-scale sports competitions and concerts are often held) Beijing, China. The experimental results demonstrate that MDB-HDNNs outperform existing methods.

The rest of our paper is summarized as follows. Section 2 reviews the literature on passenger flow prediction. Section 3 provides the details of the proposed MDB-HDNNs method. Section 4 provides the experimental setup and compares the predictive performance between MDB-HDNNs and other approaches. Finally, we make a conclusion of our paper in Section 5.

2. Related work

There is much literature on short-term transport prediction. In general, the prediction methods can be grouped into two categories, namely, parametric and nonparametric methods. The general parametric methods included the autoregressive integrated moving average model (ARIMA) and exponential smoothing. (Hobeika & Kim, 1994; Williams, Durvasula, & Brown, 1998; Williams, 2001). Milenković, Švadlenka, Melichar, Bojović, and Avramović (2018) proposed a complex mathematical model to generate railway passenger traffic forecasts using a seasonal autoregressive integrated moving average (SARIMA). Tang, Zhao, Cabrera, Ma, and Tsui (2018) presented a three-phase data-driven short-term passenger flow forecasting framework, including a time series model with an autoregressive integrated moving average and a linear regression, Li, Yan, Zhang, and Tsung (2020) integrated tensor decomposition and Auto-Regressive Moving Average (ARMA), both indicate that the parameter methods will be more complicated to obtain better prediction power in the future.

Nonparametric nonlinear regression models (such as SVR) can capture the fluctuation features of recurrent data and are more effective for nonlinear features in automatic fare collection (AFC) data. Commonly used nonparametric techniques included support vector machine regression, neural networks, k-nearest neighbors, and gradient boosting (Guo, Krishnan, & Polak, 2013; Tsai, Lee, & Wei, 2009; Wu, Ho, & Lee, 2004; Zhang, Zhu, & Wang, 2020). Ensemble learning methods, integrating multiple nonparametric models, were used for passenger flow forecasting (Wang, Chen, Chen, Hua, & Pu, 2021; Zhao, Ren, Ma, & Jiang, 2020). With the development of artificial intelligence, many kinds of neural network-based models handling sequence problems were developed for transportation demand prediction, including time-delay neural networks (TDNNs) (Lingras & Mountford, 2001), recurrent neural networks (RNNs) (Fu, Zhang, & Li, 2016; Jia, Wu, Ben-Akiva, Seshadri, & Du, 2017; Ma et al., 2015; Wu & Tan, 2016; Zhao, Chen, Wu, Chen, & Liu, 2017). Developing improved RNNs models was an emerging trend, such as improved spatiotemporal long short time recurrent neural network model (Sp-LSTM) (Yang, Xue, Ding, Wu, & Gao, 2021). Stacked autoencoder (SAE) models were used to train general traffic flow data (Lv, Duan, Kang, Li, & Wang, 2014; Liu et al., 2017). Hybrid deep learning methods have become increasingly popular and not only capture the changing trends of passenger flow but also capture passenger traffic characteristics (Wei & Chen, 2012; Xie, Li, Zhou, & Fu, 2014; Liu et al., 2017; Tang et al., 2018; Zhang, Chen, & Shen, 2019). The attention mechanism could help machine learning models identifying the most important parts of the features and appeal to much attention of the researchers (Do, Vu, Vo, Liu, & Phung, 2019; Wu, Tan, Qin, Ran, & Jiang, 2018). Zheng et al. (2019) proposed a two-phase framework that mines the spatiotemporal disturbances of contextual variables for citywide transportation demand forecasting.

Recently, an increasing number of researchers have attempted to incorporate social media data into the traffic prediction field (Wang & Wang, 2015; Wang, 2014; Zheng et al., 2016). Research on passenger flow prediction, especially regarding special events, introduced social media as a data source (Chaniotakis & Antoniou, 2015; Pereira, Rodrigues, & Ben-Akiva, 2015; Pereira, Rodrigues, Polisciuc et al., 2015; Wang, 2014; Zheng et al., 2015). Kuppam et al. (2011) and Kuppam et al. (2013) used traditional four-step models and other methods to collect data related to special activities, thereby analyzing the needs of special events. Schulz, Ristoski, and Paulheim (2013) developed a model to identify the special events utilizing social media data, thus increasing environmental perception by extracting extra information about events. Chen, Chen, and Qian (2014) used topic models to identify traffic tie-ups applying Internet data. Ni et al. (2016) found that there was a moderate positive correlation between subway passenger flow and the volume of social media posts. Zhang et al. (2016) explored the features of a traffic tie-up near an intensity of social media posts, which was considered as the keyword clustering transportation-related incidents. Li et al. (2017) developed a novel model for predicting the unconventional fluctuations of the passenger flow. Zhao et al. (2017) used disaggregate theory to

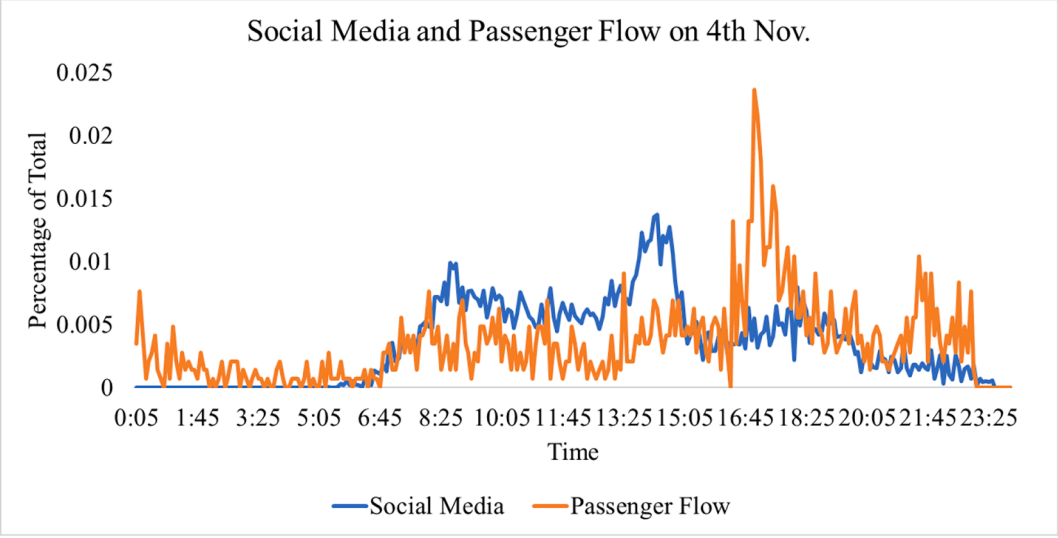

Fig. 1. Number of social media posts and the inbound traffic during an event. Notes: the unit of abscissa is hour: min, the percentage of total is dimensionless.

establish an evaluation model based on preference survey data and to predict the passenger flow of surrounding stations in actual events. Chen, Ye, Wang, and Xu (2019) focused on the dynamic volatility and nonlinear characteristics of passenger flow during special events, and a general framework was proposed.

The above literature illustrates that there is a great potentiality to apply social media to explore the correct features for traffic and passenger flow-related prediction. However, few works of literature explored the applications of fusing historical data, spatiotemporal features, and social media for passenger flow prediction in subway systems.

3. Methodology

3.1. Preliminaries

In this section, we introduce the preliminaries of our methodology, which include the data description, assumptions, correlation analysis of social media volume and passenger flow, and problem statement. We first introduce the data definitions.

Definition 1. *A smart card transaction includes the following information*:

- tr_{sboard}: Passenger's inbound station
- tr_{tboard}: Passenger's inbound time
- tr_{sexit}: Passenger's outbound station
- tr_{texit}: Passenger's outbound time
- tr_{ncard}: Passenger's card number
- tr_{mcard}: Passenger's card type

Given the observations at current time t at a target station, let y_t be the observed volume (i.e., the volumes of passenger arrivals at the target station) during the t-th time interval.

Definition 2. *Suppose the subway network is* $G = (S,E)$, *where* $S = \{s_1, s_2, \ldots, s_N\}$ *is the set of stations, and N is the number of stations.* $(s_i, s_j) \in E$ *are the OD (origin–destination) path from station* s_i *to* s_j, *where* s_i *represents the origin station, and* s_j *denotes the destination station.*

To forecast the passenger flow of a target subway station, a multi-input and single-output (MISO) nonlinear system identification and parametric modeling problem is considered. In short-term passenger flow forecasting, the inbound and outbound flows of stations are of equal importance. However, under the influence of special events, the outbound passenger flow before the event needs to be accurately predicted. Braghini et al. (1990) showed that time series data can be decomposed into three parts: seasonal, trend and noise. Namely, given observed traffic flow X, there was $X = S + T + N$, where S, T, and N represent the seasonal, trend, and noise parts, respectively. Specifically, prediction methods with temporal information can capture S but are not adequate to explain T and N, leading to insufficient forecasting performances. We believe that the effects of the sudden increases in inbound flows of other stations and the intent to participate in an event collected from social media are possible explanations for T and N.

Assumption 1. *Passengers travelling from origin station o contribute to* y_t *if OD path* (o,s) *exists. Namely, if* $o,s \in S$ *and* $(o, s) \in E$, *then* $u_o(t-1), u_o(t-2)\cdots$ *is related to* y_t, *where* $u_o(t)$ *denotes the inbound flow at station o at time interval t.*

Assumption 2. *People who use social media to express their intent to attend an event contribute to* y_t.

Correlation analysis: To show that the social media posts volume can really play a role in event-related traffic prediction, we analyze the correlation between subway passenger flow and the social media posts volume. Events from April 24, 2017, to November 4, 2017, in Beijing are shown in Table 1, which are selected as the data samples. Table 2 shows the temporal and spatial information samples of social media volume,

Table 1
Events from April 24, 2017 to November 4, 2017 in Beijing.

Stadium	Event type (No.)	Time	Sample social media post
Beijing Workers Stadium	Concert (3)	2-Jul, 25-Aug, 26-Aug	18:18, August 26, 2017 (start at 19:30): Sing with you tonight, Jay Chou's 2017 concert#
	Sports game (14)	7-May, 19-May, 18-Jun, 26-Jun, 8-Jul, 15-Jul, 5-Aug, 13-Aug, 1-Sep, 3-Sep, 10-Sep, 22-Sep, 25-Oct, 4-Nov	17:15, July 8, 2017 (start at 19:35): To the workers' Stadium, suddenly a little excited
Beijing Worker's Gym	Concert (6)	30-Apr, 20-May, 17-Jun, 22-Jul, 19-Aug, 2-Sep	19:01, June 17, 2017 (start at 19:30): Ten-year escape plan. Ten years of youth
	Opera (3)	28-May, 29-May, 6-Aug	14:39, May 29, 2017 (start at 15:00): Accompany the children to the circus

which preliminarily indicates the volume of the posts of geo-tag consistent with event stadium increase while the volumes of the posts in the geo-tags inconsistent with the events stadiums have the opposite results. Then, we test the correlation of passenger flow and social media volume in different lags. Specifically, according to the events in Table 1, the subway passenger flow of Dongsishitiao (DSST) station and social media posts volume is calculated with 15 min as the statistical interval (the extraction method of social media volume is described in detail below). Then, we analyze the correlation between subway passenger flow and social media volume for different stadiums and event types. Table 3 shows the results of the correlation test, in which all the correlation coefficients are above 0.45 and significant. Among them, vocal concerts and sports games show similar correlations while operas show a relatively low correlation. We reasonably believe that there exists a moderate positive correlation between event passenger flows and social media volumes.

Problem 1. *At time interval t-1, given all subway card transactions and event information, the historical flow vector* V_h^{t-1}, *social media vector* V_s^{t-1} *and matrix* M_p^{t-1}, *which expresses a sudden increase in inbound flow of other stations, are extracted, and used to predict the target station outbound flow* y_t (*details are shown in Section 3*).

The MDB-HDNN framework is shown in Fig. 2. Specifically, in Stage 1, we first develop a feature matrix to qualify the inbound flows of other stations. Second, a 2D convolutional neural network (2D CNN) model using a spatial attention module is presented to capture the features of inbound flows of other stations. In Stage 2, a social media feature vector is extracted, and stacked autoencoder-based neural networks (SAE-DNNs) are presented to train the model. In Stage 3, we first eliminate the disturbances from inbound flows of other stations and social media from the history passenger flow. Next, the passenger flow without disturbances is fed into deep neural networks (DNNs) to learn the inherent passenger flow modes. Ultimately, the output of SAE-DNNs is combined with the disturbances of the prediction time interval to obtain the final result.

3.2. Disturbance from inbound flows of other stations

3.2.1. Spatial-temporal feature extraction

Spatiotemporal features, such as historical flow and the OD matrix, have been widely used in traffic flow prediction (Do et al., 2019; Han et al., 2019). However, due to the particular prediction challenges during events, mining the temporal and spatial features separately did not perform well. To improve the performance of the passenger flow prediction during events, mining the temporal and spatial features is widely used. Li et al. (2017) proposed that $y(t) =$

Table 2
Temporal and spatial information samples of social media volume.

Date	Event	0–1 h ahead		1–2 h ahead		2–3 h ahead		3–4 h ahead	
		BWS	BWG	BWS	BWG	BWS	BWG	BWS	BWG
Nov. 4, 2017	BWS Sports game	375	16	245	23	130	12	231	21
Aug. 26, 2017	BWS Concert	463	5	351	14	283	13	320	8
Jun. 17, 2017	BWG Concert	20	153	12	145	6	81	25	87
May 29, 2017	BWG Opera	15	148	14	121	14	93	21	85

Notes: BWS denotes Beijing Workers Stadium. BWG denotes Beijing Worker's Gym and 0–1 h ahead denotes the time interval ahead the event start time. The numerical value is the social media posts volume with corresponding geo-tag and time interval.

Table 3
Correlation of passenger flow and social media volume.

Stadium	Event type	Correlation (Lag 1)	Correlation (Lag 2)	Correlation (Lag 3)	Correlation (Lag 4)
Beijing Workers Stadium	Concert	0.652***	0.698***	0.616***	0.626***
	Sports game	0.687***	0.614***	0.632***	0.642***
Beijing Worker's Gym	Concert	0.623***	0.682***	0.653***	0.671***
	Opera	0.585***	0.458***	0.526***	0.545***
Total	Total	0.656***	0.621***	0.623***	0.636***

Notes: ***: $p < 0.01$; **: $p < 0.05$.

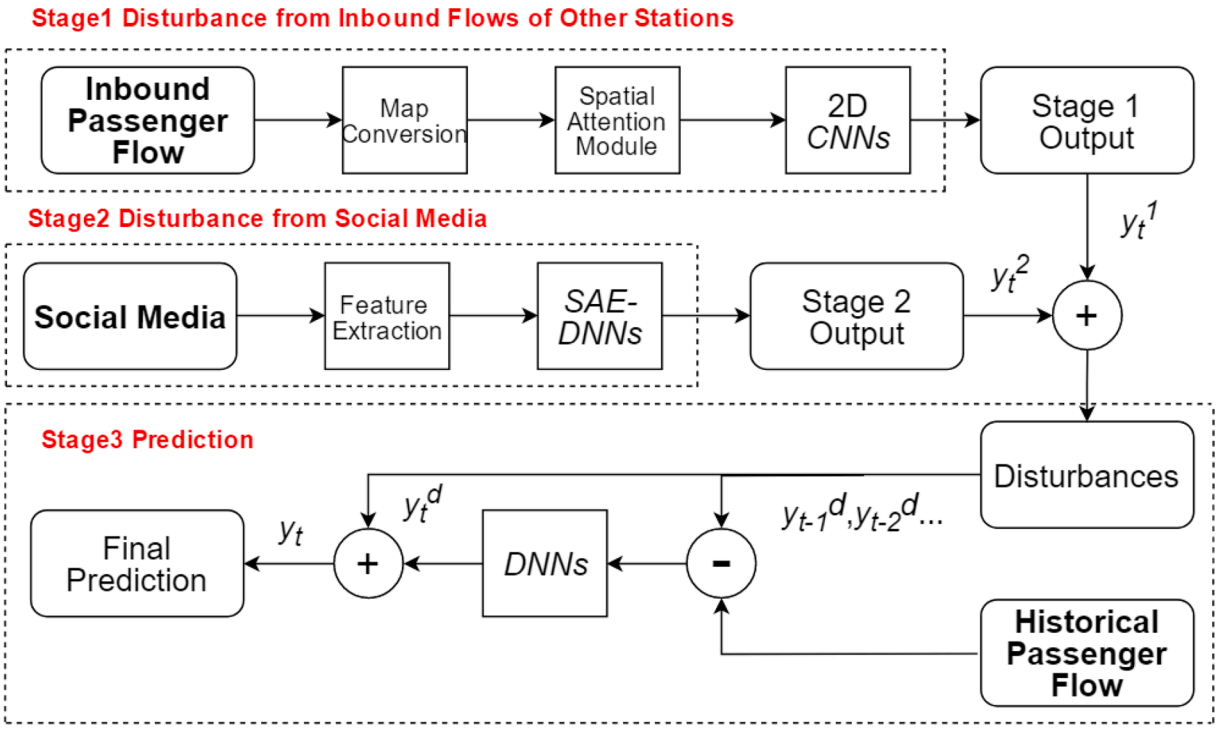

Fig. 2. Overall Framework.

$f(y(t-1), \cdots y(t-n), u_i(t-1), \cdots u_i(t-m)) + \varepsilon(t)$, where y(t) denotes the outbound passenger flow of the target station at time step t. $u_i(t)$ denotes the inbound passenger flow of station i at time step t, $\varepsilon(t)$ is the estimating error at time step t, and f is a linear or nonlinear mapping function. This method considers that the inbound flow of other stations affects the outbound flow of the target station. However, when i is increasing, it is difficult to train the model when the input of the model is too large. The spatiotemporal feature is defined as a matrix that can deal with the increasing i values and the complex traffic network in large cities. The horizontal axis and the vertical axis represent the spatial and temporal dimensions (the horizontal axis and the vertical axis are interchangeable), respectively, forming a space–time characteristic map, which performed well in highway traffic speed and flow prediction (Ma et al., 2017; Yu, Wu, Wang, Wang, & Ma, 2017). However, a spatiotemporal feature map suitable for special events has not been presented. To improve the accuracy of outbound flow forecasting, this paper develops a model of constructing a spatiotemporal feature map based on the target station.

The outbound flow of the target station will increase significantly during special events; this outbound flow during events can be much higher than that during regular times. This increase is caused by the inbound flow of other stations. For prediction, y_t, the flow feature matrix of the target station at time t is defined as:

$$M_r^{t-1} = \begin{bmatrix} u_1(t-1) & u_1(t-2) & \cdots & u_1(t-q) \\ u_2(t-1) & u_2(t-2) & \cdots & u_2(t-q) \\ \vdots & \vdots & \cdots & \vdots \\ u_p(t-1) & u_p(t-2) & \cdots & u_p(t-q) \end{bmatrix}, \quad (1)$$

where $u_i(t)$ denotes the inbound flow at station i during t-1 to t under event occurrences, p denotes the number of stations in the subway system, and q denotes the maximum time lags in the previous inbound flow. M_r^{t-1} can only capture the value of the inbound flow at each time for other stations, and it cannot reflect the difference between event occurrences and no event occurrences; thus, the average flow feature matrix of the target station at time interval t is defined as:

$$M_a^{t-1} = \begin{bmatrix} {u_1}^a(t-1) & {u_1}^a(t-2) & \cdots & {u_1}^a(t-q) \\ {u_2}^a(t-1) & {u_2}^a(t-2) & \cdots & {u_2}^a(t-q) \\ \vdots & \vdots & \cdots & \vdots \\ {u_p}^a(t-1) & {u_p}^a(t-2) & \cdots & {u_p}^a(t-q) \end{bmatrix}, \tag{2}$$

where ${u_i}^a(t)$ denotes the average inbound flow at station i at time interval t under no event occurrences (the average of one month without events in the past.). Finally, the inbound intensity matrix is defined as:

$$M_p^{t-1} = M_r^{t-1} \ominus M_a^{t-1}, \tag{3}$$

where the symbol $\ominus$ indicates the subtraction of the corresponding elements in two matrices. Each element in M_p^{t-1} can illustrate the intensity of the inbound flow of each station compared with the regular time. If the element value is much higher than 1, the significant increase in inbound flow at the station is caused by the event, and the increased flow has great potential to be the outbound flow of the target station. To train the matrix by means of a machine learning model, the spatiotemporal feature matrix M_p^{t-1} can be converted to a picture (Ma et al., 2017; Yu et al., 2017). The spatiotemporal feature map is defined as:

$$S^{t-1} = convert2picture\left(M_p^{t-1}\right) \tag{4}$$

Fig. 3 shows the data flow of the generation of the spatiotemporal feature map.

3.2.2. Spatial attention module based model

Considering the sparsity of the feature map S^{t-1} (most of the values are close to 1) in stage 1, we develop a feature refinement model utilizing a spatial attention module. With the development of the Convolutional neural networks (CNNs), the performance of computer vision models has been better and better (Deng et al., 2009; Krizhevsky & Hinton, 2009; Lin et al., 2014). Nevertheless, the CNNs cannot achieve good performance in learning feature maps with no notable features. The significant effects of attention mechanism have been widely researched in state-of-the-art studies (Ba, Mnih, & Kavukcuoglu, 2014; Bahdanau, Cho, & Bengio, 2014; Gregor, Danihelka, Graves, Rezende, & Wierstra, 2015; Mnih, Heess, & Graves, 2014; Xu et al., 2015). The attention mechanism not only identifies the focus locations but also improves the performance of models. Woo, Park, Lee, and Kweon (2018) developed a Convolutional Block Attention Module (CBAM) which can be extensively used to improve the representation power of CNNs. The spatiotemporal feature map we proposed is a 2D map; the spatial attention module (the word spatial used here, which denotes the regions in one image channel, is different than the spatial of the subway system) can perform feature refinement on it.

The spatial attention module is shown as follows. Given an intermediate feature map $F \in \mathrm{R}^{C*H*W}$ as input, a 2D spatial attention map $M_s \in \mathrm{R}^{1*H*W}$ is shown in Fig. 4.

$$F^{'} = M_S(F) \otimes F, \tag{5}$$

where $\otimes$ denotes elementwise multiplication.

We generate a spatial attention map by utilizing the interspatial relationship of features, and the spatial attention focuses on the location of an informative part. To obtain the spatial attention, we first use average-pooling and max-pooling operations along the channel axis and concatenate them to obtain effective feature descriptors. Using pooling operations along the channel axis is shown to have good performance by emphasizing the most important parts (Zagoruyko & Komodakis, 2016). On the concatenated feature descriptor, we use a convolution layer to obtain a spatial attention map $M_S(F) \in R^{H \times W}$, which encodes regions to highlight or suppress. The detailed operations are described below.

The channel information of a feature map is aggregated by applying two pooling operations, obtaining two 2D maps: $F^s_{avg} \in R^{1 \times H \times W}$ and $F^s_{max} \in R^{1 \times H \times W}$. Each represents the average-pooled features and max-pooled features across the channel. These are then concatenated and convolved by a standard convolution layer, generating our 2D spatial attention map. Briefly speaking, the spatial attention is computed as:

$$M_S(F) = \sigma\left(f^{7\times7}([AvgPool(F)];[MaxPool(F)])\right) = \sigma\left(f^{7\times7}\left(F^s_{avg};F^s_{max}\right)\right), \tag{6}$$

where σ denotes the sigmoid function and $f^{7\times7}$ represents a convolution operation with a filter size of 7×7.

Due to the spatiotemporal feature map S^{t-1} is a 2D map, we apply a convolution layer to increase the channels of the feature map and then use the spatial attention module to refine the feature. Finally, another convolution layer is used to decrease the channels of the feature map, as illustrated in Fig. 5, S'^{t-1} is feature maps after feature refinement. Then, the refined feature map S'^{t-1} is used as an input for the basic deep CNNs, and finally, y_t^1 is the output of stage 1.

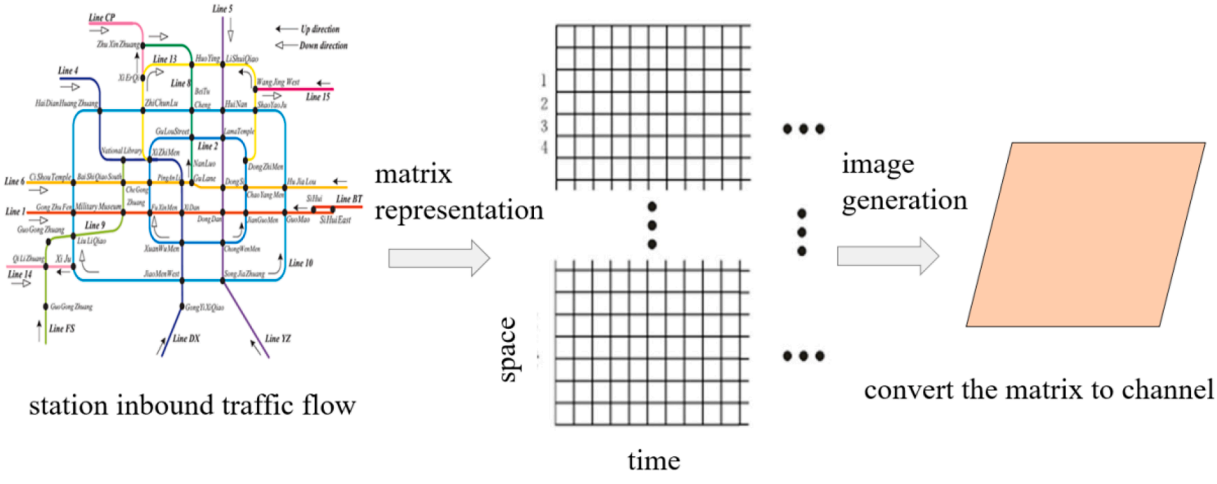

Fig. 3. Data flow of the generation of the spatiotemporal feature map.

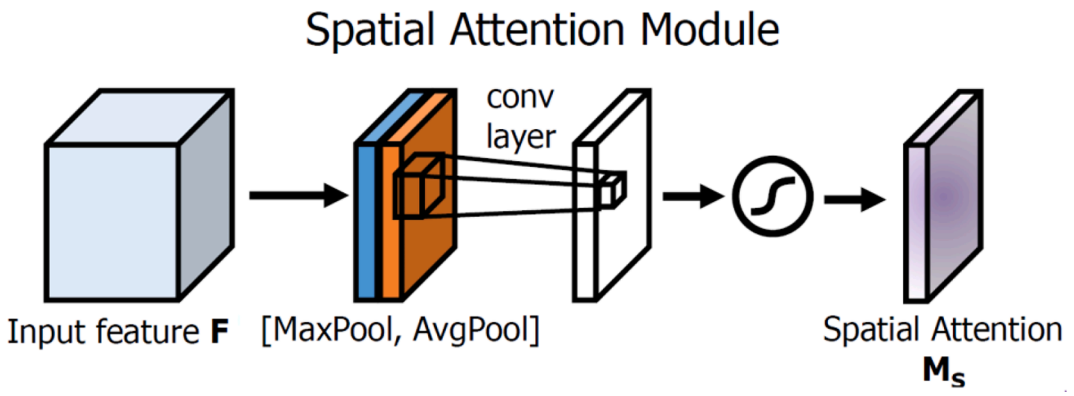

Fig. 4. An overview of the spatial attention module.

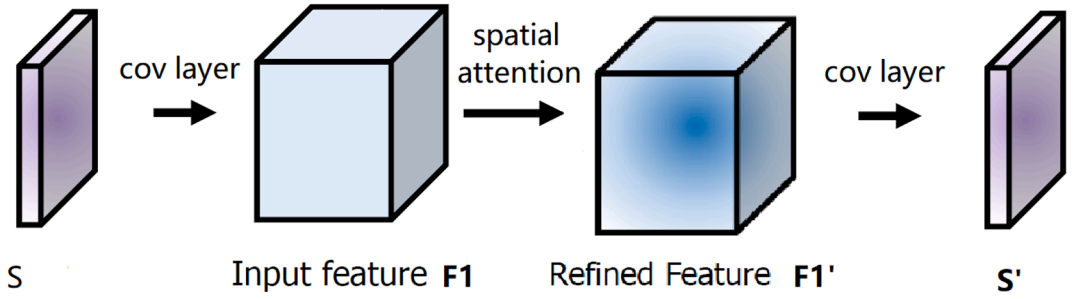

Fig. 5. Refining the Features of S. Note: the unit of ordinate is person/15 min, the unit of abscissa is hour: min.

$$y_t^1 = F\left(S^{t-1}\right), \tag{7}$$

where F denotes the basic CNN algorithm.

3.3. Disturbance from social media

3.3.1. Social media feature extraction

Social media with geotags have proven to be closely related to outbound flow event occurrences (Chen et al., 2014; Ni et al., 2016; Xue, Liu, & Gong, 2020; Zhang et al., 2016). Social media provides an economical and effective method to get traffic related data and fills the gap between short-term passenger flow and the volume of social media posts.

As shown in the preliminaries section, there is a significant positive correlation between the social media post volumes and subway passenger flow under events occurrences. Therefore, the time series of message volume in social media can reflect the trend of passenger flow. To predict y_t, the social media feature vector is defined as:

$$V_s^{t-1} = \{H_{t-1}, H_{t-2}, \cdots H_{t-m}\}, \tag{8}$$

where H_t denotes the social media volume of relevant hashtags s about the event during $t-1$ tot, and m denotes the maximum time lags in the previous social media posts. Social media data are crawled from social media (for example: Twitter and Sina microblogs) based on hashtags and geographic coordinates. Specifically, social media posts are collected in two ways. The first way is that collecting social media posts with geographic coordinates, which have the same temporal window as the events, through a geo-location filter. The spatial bounding box is set to cover only the subway station and the stadiums. These posts are all treated as relevant posts because their geotags are close to the subway station and the stadiums. The second way is that collecting social media posts that have relevant hashtags with events. Text contents and image information are semantic analyzed and topic extracted by the social media platform. According to the stadiums, we can obtain the relevant hashtags generated by social media platforms, which helps us get the events-relevant posts. Among these, irrelevant posts, as for posts not from humans, allopatric posts, are included. To address this issue, posts from users certified as media or institutions, unoriginal posts, posts with allopatric geotags are excluded. Then, the duplicate posts extracted by the two ways are removed. Finally, effective posts' time stamps, geo-locations (if any), and text content are retained, and we calculate the social media feature vector V_s using these data. The extraction algorithm is shown in Algorithm 1:

Algorithm 1 V_s extraction

Input: Extraction position, extraction time interval t
Output: Social media feature vector V_s
1. Hashtags extraction based on position
2. for $i = 1$ *to* m do
   2.1: Retrieval event in social media by hashtags and extraction time interval *t-i-1* to *t-i*
   2.2: Get text through crawler using geo-tags and hashtags
   2.3: Removing irrelevant posts from users certified as media or institutions, unoriginal posts and posts with allopatric geotags
   2.4: H_{t-i} is the value of text counting
3. $V_s = \{H_{t-1}, H_{t-2}, \cdots H_{t-m}\}$

3.3.2. Stacked autoencoder (SAE)-Based model

In stage 2, the social media feature vector V_s^{t-1} is used as an input for the basic deep neural networks (DNNs) (shown in Fig. 7); finally, y_t^2 is the output of stage 2.

$$y_t^2 = G\left(V_s^{t-1}\right) \tag{9}$$

where G denotes the stacked autoencoder-based deep neural network (SAE-DNN) algorithm. Autoencoders can be stacked (SAEs) to form a deep network to obtain an abstract representation of the input through gradual feature extraction (Duan, Lv, Liu, & Wang, 2016); SAEs have performed better than basic deep neural networks in traffic flow prediction.

3.4. Prediction

3.4.1. Historical flow

The historical data of the target station is the most widely used indicator to predict passenger flow (Tang et al., 2018; Lv et al., 2014; Liu et al., 2017), and there is a strong correlation between y_{t-1}, y_{t-2},…y_{t-n} and y_t. In addition, important features can be extracted from past events, and the arrival distributions of passenger flows are different under different events. We divide events into N classes, such as sporting events and concerts. To predict y_t, $\overline{y_t}$ also plays an important role, where $\overline{y_t}$ denotes the historical average value of y_t under the same event at time t.

The participants will not all be present near the beginning of an event, most of them will arrive in advance. Fig. 6 illustrates the outbound flow of the target station before an event (start time is 15:30); the flow increased 2 h before the event and decreased half an hour before the event. Therefore, i_t could assist the model in predicting y_t, where $i_t = t_0 - t$, i_t denotes the index of time before the event, and t_0 denotes the start time of the event.

Above all, three types of features are extracted from the historical data, and the historical feature vector is defined as:

$$V_h^{t-1} = \{y_{t-1}, y_{t-2}, \cdots, y_{t-n}, \overline{y_t}, i_t\}, \tag{10}$$

3.4.2. Predicted model

The outputs of stage 1 and stage 2 and historical vectors are used to predict the outbound flow of the target station in stage 3. Since the passenger flow includes both regular passenger modes and irregular disturbances, it is of great significance to deal with the effects of the irregular disturbances when modelling passenger flow (Zheng et al., 2019).

In order to solve this problem, in stage 3, we develop a three-step method to combine stage 1 and stage 2 with historical passenger flow to explore the multiple dependencies between inherent flow modes and the disturbances for predicting the target station outbound flow. In the first step, we eliminate the disturbances from the history passenger flow by:

$$y_t^d = y_t^1 + y_t^2, \tag{11}$$

$$\overline{V_h^{t-1}} = \{y_{t-1} - y_{t-1}^d, y_{t-2} - y_{t-2}^d, \cdots, y_{t-n} - y_{t-n}^d, \overline{y_t}, i_t\}, \tag{12}$$

where y_t^d denotes the disturbances of the inbound flow of other stations and the intent to attend an event extracted from social media, and $\overline{V_h^{t-1}}$ is the history passenger flow without disturbances. It is easier for NNs to learn the inherent passenger modes from $\overline{V_h^{t-1}}$ rather than from the raw historical passenger flow V_h^{t-1}.

In the second step, we feed $\overline{V_h^{t-1}}$ into a forecasting model to capture the underlying passenger modes for forecasting:

$$y_t^3 = H\left(\overline{V_h^{t-1}}\right), \tag{13}$$

where H denotes the basic DNNs algorithm; because of the elimination of disturbances, basic DNNs are sufficient to capture the underlying passenger patterns for prediction.

In the third step, we combine the output of the DNNs with y_t^d to consider the disturbances as follows:

$$\widehat{y_t} = y_t^3 + y_t^d. \tag{14}$$

3.5. Parameter learning

The overall MDB-HDNN model is to solve an optimization problem. The decision variables are the parameters of the whole framework, and the objective function is the mean squared error (MSE) of the forecasting result, as follows:

$$\theta = argmin_\theta \|y_t - \widehat{y_t}\|^2, \tag{15}$$

where y_t is the ground truth of the outbound flow of the target station, and $\widehat{y_t}$ is the forecasting result of the outbound flow of the target station. θ denotes the parameters of the whole framework, which can be learned using the Adam (Kingma, 2015) optimizer via back propagation.

4. Experiments

4.1. Data description

We focus our study on subway stations DSST on Line 2, TJH on Line 10, and Olympic Park Station (OPS) on Line 8 in Beijing. The stations were selected for three main reasons. First, “DSST” and “TJH” are adjacent to not one but two stadiums, Beijing Workers Stadium and

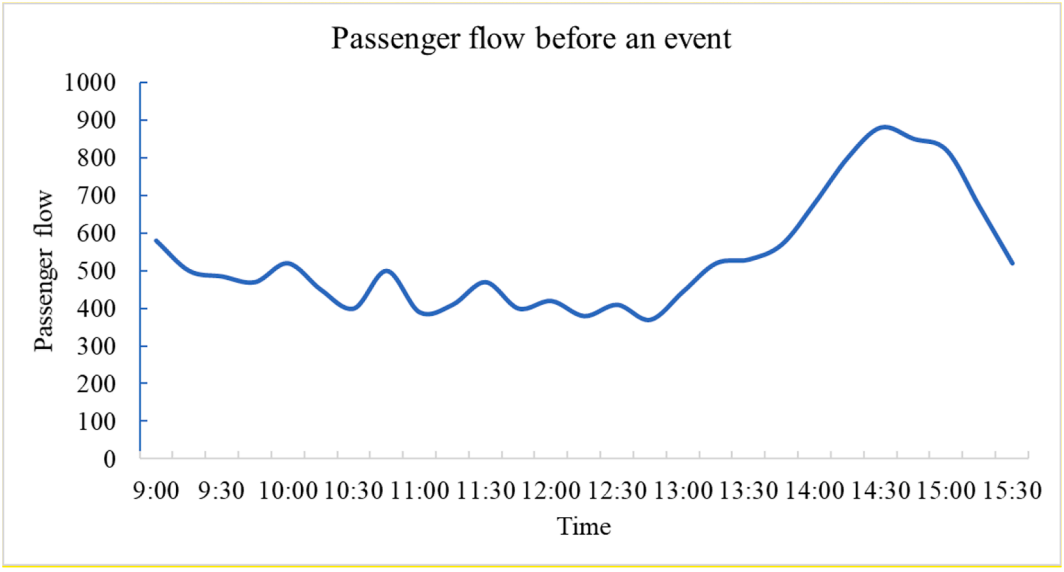

Fig. 6. Distribution of inbound passenger flow before the event. Notes: The sport game is a soccer game on November 4, 2017 (start time: 15:00, seating capacity: 80000), the concert is a vocal concert on June 7, 2017 (start time: 19:30, seating capacity: 15000), the holiday is a National Day holiday on October 2, 2017, and the weekend is Saturday on June 3, 2017, without events. The unit of ordinate is person/5min, the unit of abscissa is hour: min.

Beijing Worker's Gym. Beijing Workers Stadium is the home stadium of Beijing Sinobo Guoan F.C., and Beijing Worker's Gym hosts several concerts and sports games every year. And OPS is adjacent to Beijing National Stadium (Bird's Nest/Olympic Stadium) and Water Cube, where generates different social media topics from DSST and TJH to prove the robustness of the model. Second, sports and concert events usually appeal to public attention. As mentioned above, the volume of social media posts has a significant positive correlation with the subway outbound passenger flow. Third, since these venues have the capacity for a large number of spectators, the road traffic can be seriously congested during an event, so the outbound passenger flow of the subway station will surge. We collected event information from "Yongle ticket" (a website that sells tickets). To cover more categories of events, the time range was set from April 24, 2017, to November 4, 2017, during which, various events occurred nearby, such as football matches and concerts, were held. Fig. 7 illustrates that there are different tendencies of the outbound flows of the DSST station under different events, and the DSST and TJH stations are very suitable for testing the effect of the MDB-HDNN framework.

The Beijing Subway AFC system data from April 24, 2017, to November 4, 2017 (approximately 8 million per day) are used to extract the disturbance of inbound flows from other stations, historical flow and the ground truths of the target station predicted flow. The Sina microblog data (including text, time, geographical labels, etc.) is used to extract the disturbance indicators from social media.

4.2. Evaluation metrics and implementation

We evaluate the performance of the MDB-HDNN using the following measures:

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(\hat{y}_i - y_i)^2}, \tag{16}$$

$$MAPE = \frac{1}{T}\sum_{i=1}^{T}\frac{|y_i - \hat{y}_i|}{y_i}, \tag{17}$$

where y_i and $\hat{y}_i$ are the real value and the corresponding forecasting result, respectively, and T is the number of all real values. *RMSE* is the root mean square error ($RMSE \geq 0$, $RMSE = 0$ means a perfect model), which is used to measure the deviation between the forecasting values and the ground truths. The lower *RMSE* value denotes a better performance. *MAPE* is the mean absolute percent error ($MAPE \geq 0, MAPE = 0$ means a perfect model), which is used to measure the ratio of the absolute value of errors to ground truths. The lower *MAPE* value denotes a better performance.

We made all the experiments mainly on a workstation with 4 CPU cores (Intel Core i7-7700 CPU @ 3.60 GHz) and a GPU (NVIDIAGeForce RTX 2080). We used Python 3 with scikit-learn, TensorFlow-GPU 1.14.0, and Keras to build models.

4.3. Experimental results on the DSST

The experiments on the datasets include the following parts.

- *Comparison with baseline models.* We first compare the overall forecasting results between the MDB-HDNN and baseline models; then, we evaluate the performance under different events.
- *Performance on different training datasets.* To verify the necessity of each stage in MDB-HDNN, a portion of the features are selected and combined to form different training datasets. The removal of features from each dataset means that the corresponding stage is eliminated. All the modules can be connected through a neural network layer no matter how many neurons (number of output variables) they have.
- *Comparison with variants of the MDB-HDNN.* We first study the effect of the different approaches in each stage; second, we further study the impact of the three attention modules in stage 1.

4.3.1. Comparison with baseline methods

The proposed MDB-HDNN is compared with the competing models categorized into the following types:

Traditional models contain the autoregressive integrated moving average model (ARIMA), linear regression (LR), random forest regression (RFR), and support vector regression (SVR). Deep learning methods include deep neural networks (DNNs), recurrent neural networks (RNNs), long short-term memory (LSTM), gated recurrent units (GRUs), and stacked autoencoder-based deep neural networks (SAE-DNNs), Table 4 shows the details of the baseline models.

As shown in Table 5, MDB-HDNN has the lowest RMSE and MAPE among all the methods. MDB-HDNN achieves the best performance no matter which time intervals, indicating the robustness of the proposed model according to the time interval. More specifically, as mentioned above, only time-series history passenger flow data are used for ARIMA, and all the 3 categories of features are used in other baseline methods.

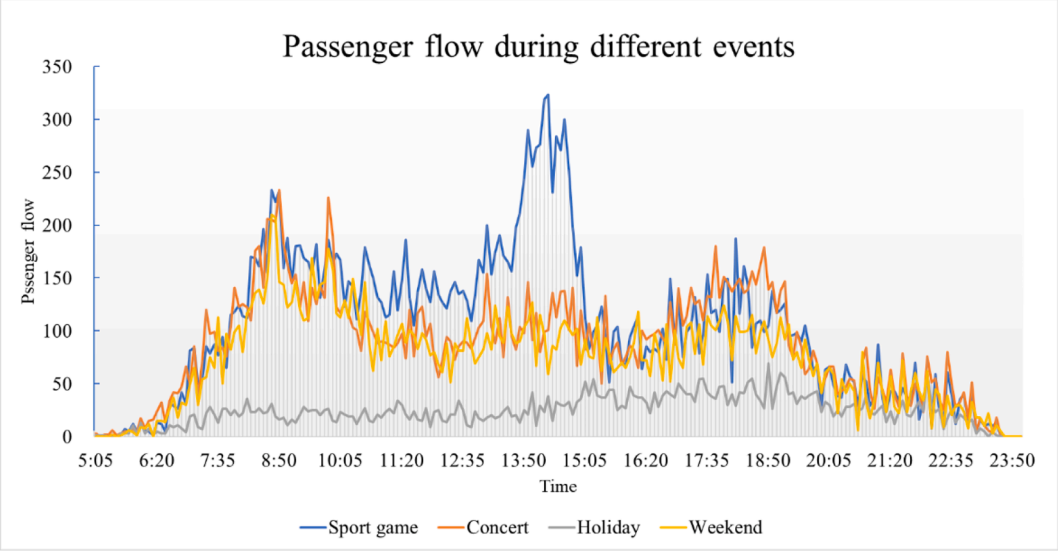

Fig. 7. Outbound flows of the DSST station under different events.

Table 4
The setting of the baseline methods.

Method	Setting
ARIMA	general ARIMA model
LR	general LR with least square method and gradient descent method
RFR	general RFR with ntree = 100 with mtry = 2, mtry = 3 and mtry = 4.
SVR	general SVR with kernel = radial function and polynomial function
DNNs	general DNNs with hidden layers = [100,100] [200,200], [200,300,200] and [100, 200, 100]
RNNs	general RNNs with RNNs Cells = 1 and 2
LSTM	general LSTM with LSTM Cells = 1 and 2
GRU	general GRU with GRU Cells = 1 and 2
SAE-DNNs	general SAE-DNNs with hidden layers = [100,100] [200,200], [200,300,200] and [100, 200, 100]

Notes: only time-series history passenger flow data are used for ARIMA, and all the 3 categories of features are used in other baseline methods. The features are flattened and connected to vectors for LR, RFR, SVR, DNNs, and SAE-DNNs, and the features are resized to vector sequences for RNNs, LSTM, and GRU. Considering that multiple sets of parameters are used to evaluate some methods, we report the best one of them to represent the performance of the competing methods thereinafter.

Table 5
Comparison of the overall predictive performance of the methods on DSST.

Method	*RMSE*	*MAPE*	*RMSE*	*MAPE*	*RMSE*	*MAPE*
	Step 1 (5 min)		Step 3 (15 min)		Step 6 (30 min)	
ARIMA	60.72	0.1369	62.53	0.1428	65.87	0.1391
LR	72.94	0.1762	70.72	0.1687	73.75	0.1661
RFR	56.08	0.1400	58.44	0.1385	58.02	0.1358
SVR	54.57	0.1321	58.78	0.1398	59.25	0.1387
DNNs	55.58	0.1270	54.29	0.1222	54.96	0.1209
RNNs	50.69	0.1004	50.50	0.1081	50.14	0.1045
LSTM	48.02	0.0885	48.74	0.0904	48.18	0.0948
GRU	49.91	0.1031	49.86	0.1054	49.32	0.0985
SAE-DNNs	52.05	0.1125	52.82	0.1158	56.21	0.1185
MDB-HDNNs	14.37	0.0268	15.42	0.0328	16.41	0.0352

Notes: We use 5 min as a time unit. In Step n (5n min), n denotes the number of time units and 5n min denotes the prediction time interval, namely predicting the passenger flow every 5n minutes.

All the features are flattened and connected to vectors for LR, RFR, SVR, DNNs, and SAE-DNNs, and the features are resized to vector sequences for RNNs, LSTM, and GRU. Therefore, social media feature and inbound flow features of nearby stations can be input to LR, RFR, SVR, DNNs, RNNs, LSTM, GRU, and SAE-DNNs. However, these models learn the social media features, history passenger flow features, and the inbound flow features of nearby stations together. On the other hand, our proposed model has independent modules which are suitable for learning each kind of features. So that ARIMA, LR, RFR, SVR, DNNs, RNNs, LSTM, GRU, and SAE-DNNs perform poorly, as they poorly examine the social media disturbances and the inbound flow disturbances of nearby stations. As a result, our MDB-HDNN outperforms the other methods regardless of the time interval. For example, Fig. 8 shows the overall predicted results of 11 event days under Step 1 (5 min) (every day has 72 time intervals and there is a total of 6 h before the event), and the proposed framework shows good performance.

4.3.2. Performance on different training datasets

In this paper, the training datasets are divided into seven categories to test whether multidimensional features can improve the forecasting performance of the forecasting model. We mark the feature types with a dot in Table 6. We compare the prediction performance of different datasets in different scenarios (all of the experiments below use a 15 min time interval); scenarios include a concert, sporting event, and both events. Table 7 illustrates that training dataset 7 has the highest accuracy, and the RMSE and MAPE values of dataset 7 are the lowest in all scenarios. When forecasting separately, the performance of the history

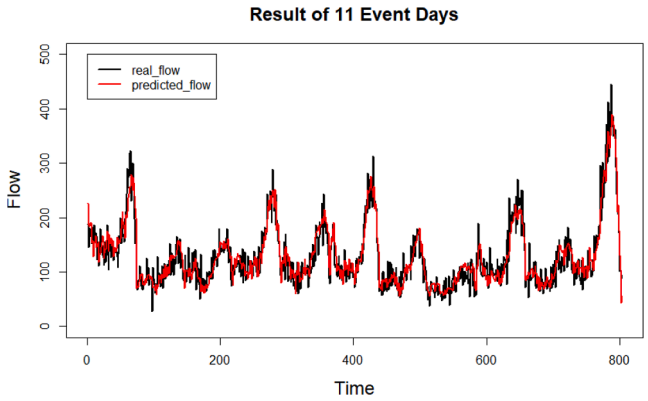

Fig. 8. Outbound flow of DSST in 11 event days at DSST. Note: the unit of ordinate is person/5min, the unit of abscissa is number of time unit. One time unit means 5 min, and every 72 units denote a total of 6 h before one event. For instance, 0 to 71 denote a total of 6 h before an event.

Table 6
Seven kinds of training datasets.

Type	History Flow Feature	Social Media Feature	Spatial-temporal Feature
1	•		
2		•	
3			•
4	•	•	
5	•		•
6		•	•
7	•	•	•

flow feature is similar to that of the spatiotemporal feature, and the social media feature does not perform well. Datasets 1, 2, and 3 have no significant differences in RMSE and MAPE under the three scenarios. However, when two random features are combined (datasets 4, 5, and 6), datasets with social media features can increase the accuracy of a single dataset slightly, and the combination of history flow features and spatial–temporal features significantly reduces the RMSE and MAPE. In addition, the performances of the three scenarios using dataset 4 are similar, but the RMSE and MAPE are higher when predicting both events than when predicting a single event using datasets 5 and 6, illustrating that the combination of history flow features and social media features can improve the stability of the prediction results in different scenarios. The use of three types of features not only has the lowest RMSE and MAPE but also performs well in the three scenarios. All of the above illustrates that history flow features and spatial–temporal features ensure the accuracy of a single scenario, and the social media features can identify different scenarios well, which ensures the accuracy of all scenarios. Fig. 9 shows the outbound flow of the DSST on 4 event days (including concerts, football games, and ice and snow sports), and the results show that the MDB-HDNN has strong performance for different events.

4.3.3. Comparison with variants of MDB-HDNN

This section verifies that the algorithm in each stage of the MDB-HDNN framework is optimal. The control variable method is used to design the experiment. We change the algorithm in one stage and keep the algorithms in other stages unchanged and then compare the RMSE and MAPE of the predicted results. Fig. 10 illustrates the experimental results on different attention arrangement modules in stage 1. According to the results, we summarize that spatial attention-based CNNs perform better than the other models; due to the single channel of the feature map in stage 1, methods including the channel attention module are not effective. Table 8 summarizes the experimental results of the three spatial–temporal feature extraction methods; ⊘ indicates the division of

Table 7
Performance comparisons of the 7 training datasets and 3 scenarios on DSST.

Events		Training datasets						
		1	2	3	4	5	6	7
concert	*RMSE*	52.08	123.94	50.63	32.05	15.88	43.34	15.15
	MAPE	0.1275	0.3132	0.1011	0.0605	0.0374	0.0784	0.0343
Sporting event	*RMSE*	52.00	130.00	52.97	31.94	18.222	39.60	14.71
	MAPE	0.1270	0.3309	0.1066	0.0661	0.0355	0.0726	0.0310
both events	*RMSE*	54.80	146.19	64.28	33.64	24.6951	40.08667	15.42
	MAPE	0.1137	0.3043	0.1572	0.0783	0.0521	0.0741	0.0328

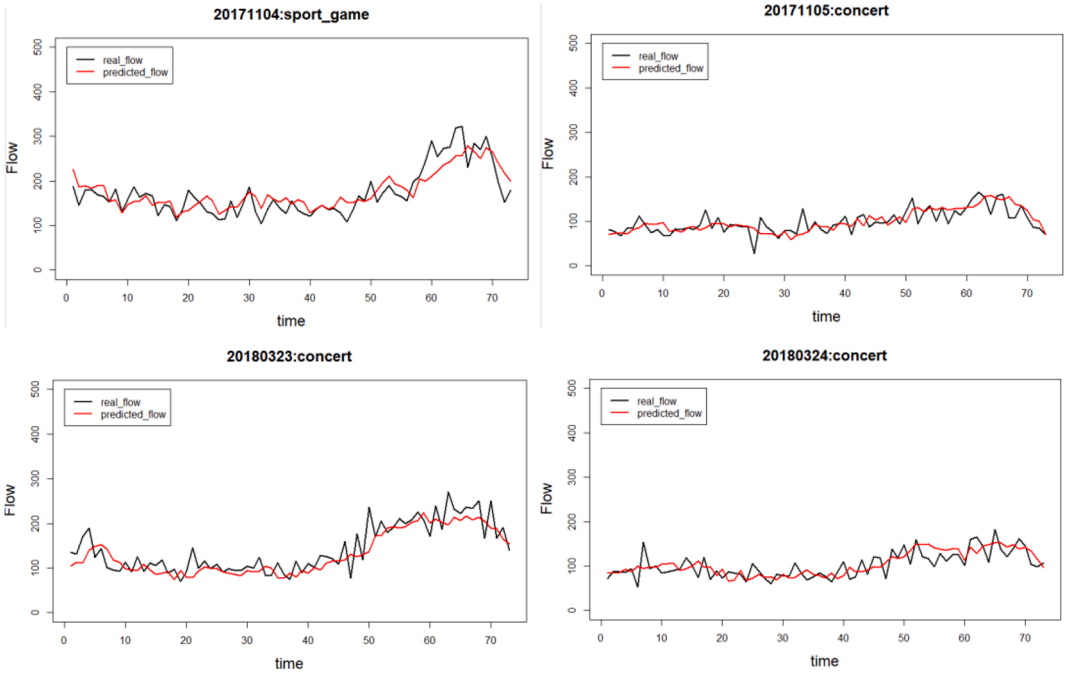

Fig. 9. Outbound flow of DSST on 4 event days at DSST. Note: the unit of ordinate is person/5min, the unit of abscissa is number of time unit. One time unit means 5 min, and every 72 units denote a total of 6 h before one event.

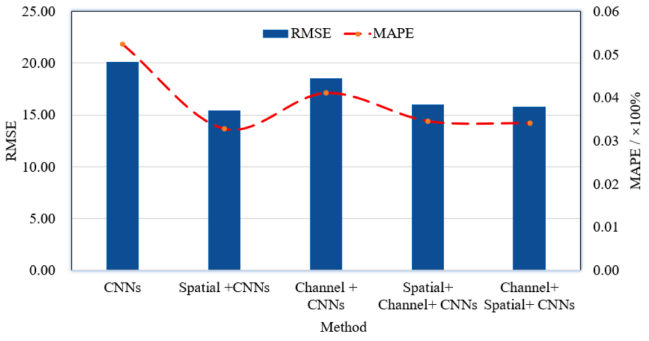

Fig. 10. Comparison of attention modules in stage 1. Note: the unit of RMSE is person/5min, the MAPE is dimensionless.

Table 8
Comparison of matrices in stage 1.

Method	M_r	$M_r \oslash M_a$	$M_r \ominus M_a$
RMSE	17.53	16.72	15.42
MAPE	0.0422	0.0396	0.0328

the corresponding elements in the two matrices. The $M_r \ominus M_a$ model achieves better performance; compared with $\oslash$, $\ominus$ more accurately reflects the change in inbound passenger flow.

Fig. 11 shows the experimental results of the different algorithms in stage 2. DNN-based algorithms perform better than traditional algorithms in stage 2, and those results demonstrate that using SAE to pretrain the features is effective to improve the forecasting performance.

In stage 3, we eliminate disturbances from the history flow and combine the disturbances at the forecasting time interval. To test the

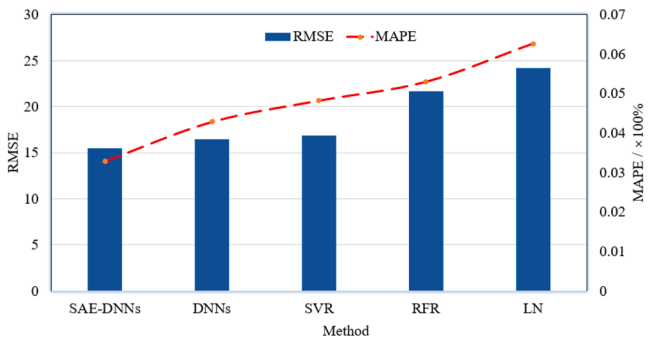

Fig. 11. Comparison of stage 2 methods. Note: the unit of RMSE is person/5min, the MAPE is dimensionless.

benefit of the approach, we conduct experiments with and without disturbance elimination. Fig. 12 illustrates that algorithms with disturbance elimination perform better than those without disturbance elimination, and DNN-based algorithms also perform better than traditional algorithms, but there is no significant difference between the prediction accuracies of the DNNs and SAE-DNNs. Considering the extra computing resources and training time of SAE, the basic DNNs are sufficient.

We further analyze the impact of the hyperparameters in the MDB-HDNN, including the number of inbound feature time steps, number of historical feature time steps, number of social media feature time steps, structure of the CNNs layer in stage 1, structure of the SAE-DNNs in stage 2, and structure of the DNNs layer in stage 3. In this paper, we used the proposed model to predict the 5 min outbound flow, 15 min outbound flow, and 30 min outbound flow. We choose feature time steps of 1–30, hidden layers sizes within 1 to 5, and the number of hidden units within [100, 200, 300, 400, 500]. After conducting the experiments, we obtain the best structure for each forecasting task, which is illustrated in Table 9.

4.4. Experimental results on TJH and OPS

The baseline models on the TJH dataset are the same as those on the DSST dataset. We report the *RMSE* and *MAPE* of the different methods in Table 10. The results according to *RMSE* and *MAPE* are the same. The results indicate that the MDB-HDNN has better performance comparing all baseline models in Step 1, Step 2, and Step 3.

More specifically, as shown in Fig. 13, we observe that although the *RMSE* and MAPE decrease when considering inbound flows from other stations, the ratio of promotion is reduced. Comparing the history flow feature + spatial–temporal feature (dataset 5) and history flow feature (dataset 1), the RMSE decreased by 54.94% on DSST; comparing the history flow feature + social media feature + spatial–temporal feature (dataset 7) and history flow feature + social media feature (dataset 4), the RMSE decreased by 54.16% on DSST, while these figures in regard to the TJH are 20.32% and 46.55%, respectively. The DSST station is on Line 2, and the TJH station is on Line 10; Line 2 has more transfer stations than Line 10, attracting more passengers from other lines. This indicates that the MDB-HDNN performs better when the line of the target station has more transfer stations, and with the expansion of the rail transit system, the performance of the MDB-HDNN will increase.

OPS dataset, which includes different stadiums and social media topics from DSST and TJH dataset, is used to evaluate the robustness of our model. The *RMSE* and *MAPE* of the competing methods are shown in Table 11, which indicates the MDB-HDNN outperforms all baseline methods. As shown in Fig. 14, we observe that the performance of dataset 7 is the best which indicates introducing more data sources can achieve better prediction results. The prediction results of these three stations indicate that our model is effective and robust. Utilizing social

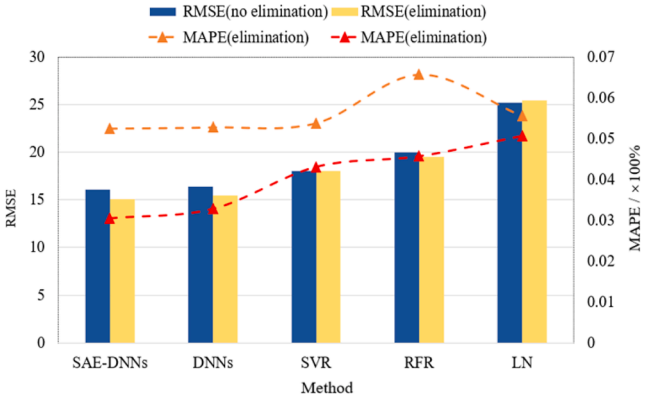

Fig. 12. Comparisons of algorithms in stage 2 and stage 3. Note: the unit of RMSE is person/5min, the MAPE is dimensionless.

Table 9
Structure of MDB-HDNNs for outbound flow prediction.

task	Stage	Feature time step	Hidden layers	Hidden units
5-min	1	24	3	[100,200,100]
	2	6	3	[400,400,400]
	3	24	4	[100,200,200,100]
15-min	1	8	3	[100,200,100]
	2	4	2	[500,500]
	3	16	3	[200,400,200]
30-min	1	4	3	[100,200,100]
	2	3	2	[400,400]
	3	8	2	[100,200,100]

Table 10
Overall predictive performance comparison on TJH.

Method	*RMSE*	*MAPE*	*RMSE*	*MAPE*	*RMSE*	*MAPE*
	Step 1 (5 min)		Step 3 (15 min)		Step 6 (30 min)	
ARIMA	74.39	0.1799	72.86	0.1751	77.90	0.1698
LR	62.14	0.1433	64.36	0.1438	70.81	0.1414
RFR	57.92	0.1433	62.21	0.1455	58.04	0.1379
SVR	57.98	0.1391	63.52	0.1463	62.72	0.1464
DNNs	51.72	0.1019	53.20	0.1177	53.96	0.1098
RNNs	59.30	0.1284	58.27	0.1229	55.96	0.1209
LSTM	49.05	0.0957	49.39	0.0923	49.70	0.0996
GRU	53.98	0.1078	54.45	0.1101	52.96	0.1035
SAE-DNNs	55.88	0.1162	54.56	0.1198	56.28	0.1195
MDB-HDNNs	26.53	0.0584	28.70	0.0616	28.85	0.0631

Notes: We use 5 min as a time unit. In Step n (5n min), n denotes the number of time units and 5n min denotes the prediction time interval, namely predicting the passenger flow every 5n minutes.

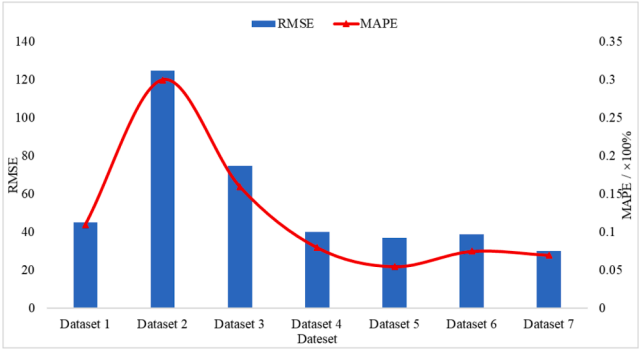

Fig. 13. Performance comparisons of the 7 training datasets on TJH. Note: the unit of RMSE is person/5min, the MAPE is dimensionless.

Table 11
Overall predictive performance comparison on OPS.

Method	*RMSE*	*MAPE*	*RMSE*	*MAPE*	*RMSE*	*MAPE*
	Step 1 (5 min)		Step 3 (15 min)		Step 6 (30 min)	
ARIMA	80.45	0.2114	79.80	0.2030	85.45	0.2010
LR	69.13	0.1658	70.52	0.1696	77.21	0.1704
RFR	65.62	0.1809	68.75	0.1732	64.84	0.1728
SVR	64.72	0.1689	69.96	0.1810	68.86	0.1825
DNNs	59.56	0.1330	59.96	0.1496	61.63	0.1399
RNNs	65.59	0.1516	64.98	0.1479	63.59	0.1503
LSTM	56.36	0.1246	56.52	0.1193	56.82	0.1324
GRU	60.41	0.1371	60.79	0.1361	59.95	0.1291
SAE-DNNs	63.45	0.1478	60.65	0.1554	63.40	0.1520
MDB-HDNNs	34.26	0.0865	36.51	0.0891	36.37	0.0886

Notes: We use 5 min as a time unit. In Step n (5n min), n denotes the number of time units and 5n min denotes the prediction time interval, namely predicting the passenger flow every 5n minutes.

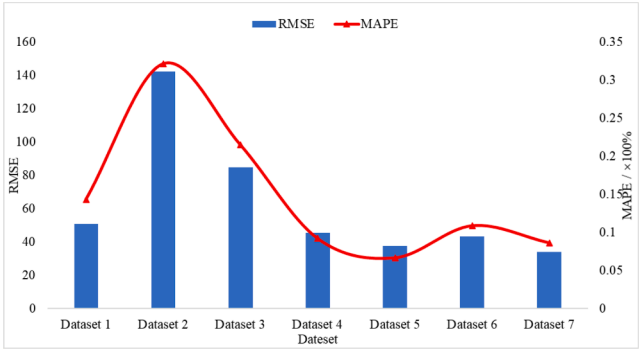

Fig. 14. Performance comparisons of the 7 training datasets on OPS. Note: the unit of RMSE is person/5min, the MAPE is dimensionless.

media volume, the inbound passenger flow of nearby stations, and historical passenger flow, our proposed model can well predict the outbound subway passenger flow when the event occurs.

5. Conclusion

Predicting subway outbound flow during events is a major challenging task for traffic management because it is of great difficulty to estimate nonregular passenger flow utilizing traditional models. In our work, we develop an MDB-HDNN framework to model the effects of the inbound flow of other stations and of social media to determine inherent flow modes and the multivariate disturbances for passenger flow forecasting during events. The proposed MDB-HDNN contains three stages. In stage 1, we develop a novel method to qualify the inbound passenger flow of other stations by a matrix and train the spatiotemporal feature using attention module-based CNNs. In stage 2, we construct a social media vector and train it using SAE-DNNs, which pretrains the feature and improves the predicted performance. In stage 3, we eliminate disturbances from the history traffic flow to overcome multivariate disturbances from capturing the inherent passenger flow modes and combine disturbances at the forecasting time interval. Experimental results on three real-world datasets indicate that MDB-HDNN performs well under various settings.

The MDB-HDNN performs well during concerts and sporting events. Whether it can be used to predict the traffic flow at stations near scenic spots and locations holding large-scale gatherings has not been verified. In the future, we will conduct further research using an emotional analysis of social media, which could perform well in predicting passenger flow under large-scaled gatherings. In addition, we only consider one channel when building the spatiotemporal feature map. If we introduce multiple channels, extracting more dimensional spatiotemporal features may also improve the accuracy of the model. Furthermore, the event scale-related indicators, include the number of tickets sold, the seating capacity of the stadium, are influential factors for the passenger flow. However, the accessibility of the data restrains the development of the model. Combing our model with the event information identification model is valuable a future work.

CRediT authorship contribution statement

Gang Xue: Data curation, Writing – original draft, Conceptualization, Methodology, Software. **Shifeng Liu:** Supervision, Investigation, Validation. **Long Ren:** Validation. **Yicao Ma:** Visualization, Data curation. **Daqing Gong:** Conceptualization, Methodology, Software, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Beijing Social Science Foundation [grant numbers 19JDGLA002, 18JDGLA018], the MOE (Ministry of Education in China) Project of Humanities and Social Sciences [grant number 19YJC630043], the National Natural Science Foundation of China [grant number J1824031] and was partially supported by Beijing Logistics Informatics Research Base. We appreciate their support very much.

References

Ba, J., Mnih, V., & Kavukcuoglu, K. (2014). *Multiple Object Recognition with Visual Attention*. Computer ence.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). *Neural machine translation by jointly learning to align and translate*. Computer Science.

Chaniotakis, E., & Antoniou, C. (2015). Use of Geotagged Social Media in Urban Settings: Empirical Evidence on Its Potential from Twitter. *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on. IEEE*.

Chen, P. T., Chen, F., & Qian, Z. (2014). Road traffic congestion monitoring in social media with hinge-loss Markov random fields. In *2014 IEEE international conference on data mining* (pp. 80–89). IEEE.

Chen, M. C., & Wei, Y. (2011). Exploring time variants for short-term passenger flow. *Journal of Transport Geography, 19*(4), 488–498.

Chen, E., Ye, Z., Wang, C., & Xu, M. (2019). Subway passenger flow prediction for special events using smart card data. *IEEE Transactions on Intelligent Transportation Systems*.

Deng, J., Wei, Dong, W., Socher, R., Li, L. J., Kai, L., & Li, F. F. (2009). Imagenet: A large-scale hierarchical image database. *Proceedings of IEEE Computer Vision & Pattern Recognition*, 248–255.

Ding, A. L., Zhao, X. M., & Jiao, L. C. (2002). Traffic flow time series prediction based on statistics learning theory. *IEEE International Conference on Intelligent Transportation Systems. IEEE*.

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.

Duan, Y., Lv, Y., Liu, Y. L., & Wang, F. Y. (2016). An efficient realization of deep learning for traffic data imputation. *Transportation Research Part C, 72*(nov.), 168–181.

Fu, R., Zhang, Z., & Li, L. (2016). Using LSTM and GRU neural network methods for traffic flow prediction. *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE*.

Gregor, K., Danihelka, I., Graves, A., Rezende, D. J., & Wierstra, D. (2015). Draw: A recurrent neural network for image generation. *Computer Science*, 1462–1471.

Guo, F., Krishnan, R., & Polak, J. (2013). A computationally efficient two-stage method for short-term traffic prediction on urban roads. *Transportation Planning and Technology, 36*(1), 62–75.

Han, Y., Wang, S., Ren, Y., Wang, C., Gao, P., & Chen, G. (2019). Predicting station-level short-term passenger flow in a citywide metro network using spatiotemporal graph convolutional neural networks. *ISPRS International Journal of Geo-Information, 8*(6), 243.

Hobeika, A. G., & Kim, C. K. (1994). Traffic-flow-prediction systems based on upstream traffic. In *Proceedings of VNIS'94-1994 Vehicle Navigation and Information Systems Conference* (pp. 345–350). IEEE.

Jia, Y., Wu, J., Ben-Akiva, M., Seshadri, R., & Du, Y. (2017). Rainfall-integrated traffic speed prediction using deep learning method. *IET Intelligent Transport Systems, 11*(9), 531–536.

Krizhevsky, A., & Hinton, G. (2009). *Learning multiple layers of features from tiny images*. Technical Report TR-2009. Toronto: University of Toronto.

Kuppam, A., Copperman, R., Lemp, J., Rossi, T., Livshits, V., Vallabhaneni, L., Jeon, K., & Brown, E. (2013). Special events travel surveys and model development. *Transportation Letters, 5*(2), 67–82.

Kuppam, A., Copperman, R., Rossi, T., Livshits, V., Vallabhaneni, L., Brown, T., & DeBoer, K. (2011). Innovative methods for collecting data and for modeling travel related to special events. *Transportation Research Record, 2246*(1), 24–31.

Leng, B., Zeng, J., Xiong, Z., Lv, W., & Wan, Y. (2013). Probability tree based passenger flow prediction and its application to the Beijing subway system. *Frontiers of Computer Science, 7*(2), 195–203.

Li Z., Yan H., Zhang C., Tsung F. (2020), Long-Short Term Spatiotemporal Tensor Prediction for Passenger Flow Profile, in: IEEE Robotics and Automation Letters, 5 (4), pp. 5010-5017, Oct. 2020, doi: 10.1109/LRA.2020.3004785.

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.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L. (2014). Microsoft coco: common objects in context. European Conference on Computer Vision.

Lingras, P., & Mountford, P. (2001). Time Delay Neural Networks Designed Using Genetic Algorithms for Short Term Inter-City Traffic Forecasting. Engineering of Intelligent Systems. *14th International Conference on Industrial and Engineering*

Applications of Artificial Intelligence and Expert Systems, IEA/AIE 2001, Budapest, Hungary, June 4-7, 2001, Proceedings. Springer-Verlag.
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.
Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors, 17*(4), 818.
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.
Milenković, M., Švadlenka, L., Melichar, V., Bojović, N., & Avramović, Z. (2018). SARIMA modelling approach for railway passenger flow forecasting. *Transport, 33*(5), 1113–1120.
Mnih, V., Heess, N., & Graves, A. (2014). Recurrent models of visual attention. *Advances in Neural Information Processing Systems*, 2204–2212.
Ni, M., He, Q., & Gao, J. (2016). Forecasting the subway passenger flow under event occurrences with social media. *IEEE Transactions on Intelligent Transportation Systems*, 1–10.
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.
Pereira, F. C., Rodrigues, F., Polisciuc, E., & Ben-Akiva, M. (2015). Why so many people? Explaining nonhabitual transport overcrowding with internet data. *IEEE Transactions on Intelligent Transportation Systems, 16*(3), 1370–1379.
Schulz, A., Ristoski, P., & Paulheim, H. (2013). I see a car crash: Real-time detection of small scale incidents in microblogs. In *Extended semantic web conference* (pp. 22–33). Berlin, Heidelberg: Springer.
Sun, Y., Leng, B., & Guan, W. (2015). A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing, 166*, 109–121.
Tang, L., Zhao, Y., Cabrera, J., Ma, J., & Tsui, K. L. (2018). Forecasting short-term passenger flow: An empirical study on shenzhen metro. *IEEE Transactions on Intelligent Transportation Systems, 20*(10), 3613–3622.
Tsai, T. H., Lee, C. K., & Wei, C. H. (2009). Neural network based temporal feature models for short-term railway passenger demand forecasting. *Expert Systems with Applications, 36*(2), 3728–3736.
Wang, F. Y. (2014). Scanning the issue and beyond: Real-time social transportation with online social signals. *IEEE Transactions on Intelligent Transportation Systems, 15*(3), 909–914.
Wang, P., Chen, X., Chen, J., Hua, M., & Pu, Z. (2021). A two-stage method for bus passenger load prediction using automatic passenger counting data. *IET Intelligent Transport Systems, 15*, 248–260. https://doi.org/10.1049/itr2.12018
Wang, F. Y., & Wang, F. (2015). Scanning the issue and beyond: Transportation games for social transportation. *IEEE Transactions on Intelligent Transportation Systems, 16*(3), 1061–1069.
Wei, Y., & Chen, M. C. (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research, 21C*(1), 148–162.
Williams, B. M. (2001). Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling. *Transportation Research Record, 1776*(1), 194–200.
Williams, B. M., Durvasula, P. K., & Brown, D. E. (1998). Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record, 1644*(1), 132–141.
Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: convolutional block attention module.
Wu, Y., & Tan, H. (2016). Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. arXiv preprint arXiv:1612.01022.
Wu, C. H., Ho, J. M., & Lee, D. T. (2004). Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems, 5*(4), 276–281.
Wu, Y., Tan, H., Qin, L., Ran, B., & Jiang, Z. (2018). A hybrid deep learning based traffic flow prediction method and its understanding. *Transportation Research Part C: Emerging Technologies, 90*, 166–180.
Xie, M. Q., Li, X. M., Zhou, W. L., & Fu, Y. B. (2014). Forecasting the short-term passenger flow on high-speed railway with neural networks. *Computational Intelligence and Neuroscience*.
Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., & Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. *Computer Science*, 2048–2057.
Xue, G., Liu, S., & Gong, D. (2020). Identifying abnormal riding behavior in urban rail transit: a survey on ''in-out'' in the same subway station. IEEE Transactions on Intelligent Transportation Systems, PP (99), 1-13.
Yang, X., Xue, Q. C., Ding, M. L., Wu, J. J., & Gao, Z. Y. (2021). Short-term prediction of passenger volume for urban rail systems: A deep learning approach based on smart-card data. *International Journal of Production Economics, 231*, Article 107920.
Yu, H., Wu, Z., Wang, S., Wang, Y., & Ma, X. (2017). Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors, 17*(7), 1501.
Zagoruyko, S., & Komodakis, N. (2016). Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer.
Zhang, J., Chen, F., & Shen, Q. (2019). Cluster-Based LSTM Network for Short-Term Passenger Flow Forecasting in Urban Rail Transit. *IEEE Access, 7*, 147653–147671.
Zhang, Z., Ni, M., He, Q., Gao, J., Gou, J., & Li, X. (2016). Exploratory study on correlation between Twitter concentration and traffic surges. *Transportation Research Record, 2553*(1), 90–98.
Zhang, Y., Zhu, C., & Wang, Q. (2020). LightGBM-based model for metro passenger volume forecasting. *IET Intelligent Transport Systems, 14*, 1815–1823. https://doi.org/10.1049/iet-its.2020.0396
Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). Lstm network: A deep learning approach for short-term traffic forecast. *Iet Intelligent Transport Systems, 11*(2), 68–75.
Zhao, Y. Y.; Ren, L.; Ma, Z.L.; Jiang, X.G. (2020). Novel Three-Stage Framework for Prioritizing and Selecting Feature Variables for Short-Term Metro Passenger Flow Prediction. *Transportation Research Record: Journal of the Transportation Research Board*, 036119812092650–. doi:10.1177/0361198120926504.
Zheng, X., Chen, W., Wang, P., Shen, D., Chen, S., Wang, X., … Yang, L. (2015). Big data for social transportation. *IEEE Transactions on Intelligent Transportation Systems, 17*(3), 620–630.
Zheng, X., Chen, W., Wang, P., Shen, D., Chen, S., Wang, X., … Yang, L. (2016). Big Data for Social Transportation. *IEEE Transactions on Intelligent Transportation Systems, 17*(3), 620–630. https://doi.org/10.1109/tits.2015.2480157
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*.