# SUBWAY PASSENGER FLOW FORECASTING WITH MULTISTATION AND EXTERNAL FACTORS

1. Prof.R.Yadagiri Rao,professor,Head,H&S,Sri Indu Institute of Engineering&TechnologySIIET, Sheriguda,Ibrahimpatnam,Hydarabad,
2. 2.P.H.Swarna rekha,ASssistant Professor,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
3.K.Dheeraj,Student,CSE,SIIET, Sheriguda,Ibrahimpatnam,Hydarabad
3. 4.K.Varun, Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
4. 5.M.Prashanth Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
6.M.Baleshwar,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad

## Abstract:
With the rapid development of urban rail transit, more and more people choose to travel by subway. Therefore, accurate passenger flow forecasting is of great significance for passengers and municipal construction and contributes to smart city services. In this paper, we propose a multi-type attention-based network to forecast the subway passenger flow with multi-station and external factors. The proposed network has different types of attention mechanisms to adaptively extract relevant features, including multi-station, external factors, and historical data. In addition, the hierarchical attention mechanism is used to model the hierarchical relationship between subway lines and stations. In addition, the embedding method is applied to better combine the different kinds of data. The experiments on real subway passenger flow data in a city in China demonstrate that our method outperforms five baseline methods. Moreover, our method can visualize the impact of different stations and other factors on traffic, which plays an important role in passenger travel and subway dispatch.

## INTRODUCTION:

In recent years, rail transit has developed rapidly. As an important part of rail transit, the subway has become the main choice for people's travel with its timely and efficient advantages. Therefore, reliable and accurate subway passenger flow forecasting is significant for passengers, transit operators, and public agencies. Moreover, the study of subway passenger flow is also an important part of the smart city domain .

We define passenger flow as the number of passengers passing through the target station per unit time. For subway companies and passengers, passenger flow in the subway station is more concerned, including all inbound and outbound passengers. Metro managers can adjust the number of subway gates and the running interval of the subway according to the passenger flow in the station, and passengers can adjust their own travel routes to avoid crowding. Therefore, this paper focuses on the trend of passenger flow in subway stations. From the time dimension, the forecasting of passenger flow can be regarded as the prediction of time series data, and there have been many studies on it. These studies focus on the prediction of a single source of time series data, and try to find the interconnection in the time dimension. However, the subway passenger flow forecasting can be considered as a typical time-space problem, so it is not enough to consider only the information of the time dimension. Taking into account the characteristics of urban subway, we divide all the factors affecting passenger flow into three parts: the influence of subway stations on each other, external factors and historical data.

## A. The Influence of Subway Stations on Each Other

Urban subway can be regarded as a network system with the spatial topological relationship, nodes of which are subway stations. Passengers' travel and transfer make the subway stations interact with each other. An intuitive example is the impact among sites on the same subway line. If the passenger flow of the upstream site is large, after a period of time, the passenger flow of the downstream site will increase accordingly. In addition, due to the division of urban functional areas [5], stations that are not on the same line would also interact with each other. For example, during the morning rush hour, passengers near a residential area flood into the central business district (CBD) to work. Similarly, during the evening rush hour, passengers return to the residential area from the commercial area. We show the inbound passenger flow of a site in the residential area and the outbound flow of a site in the commercial area in Fig. 1(a). It can be seen from the graph that the trend of the two curves is roughly the same during the morning rush hour. A schematic diagram of this phenomenon is shown in Fig. 1(b). It can be seen from the above analysis that the relationship between the stations will also be affected by the subway lines. In general, for a target station, all the other stations, namely common stations, have different effect at different time. And the traffic data of the entire subway can be regarded as a hierarchical structure due to the inclusion relationship between the line and the site.

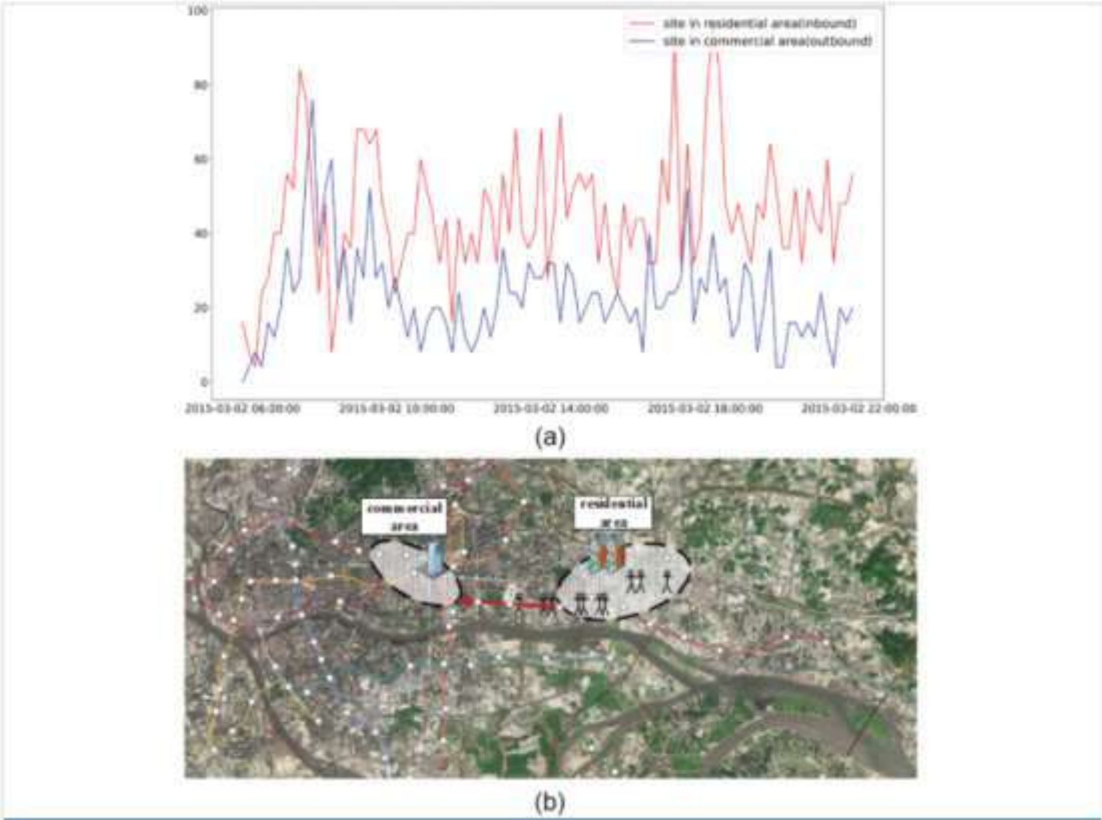


**FIGURE 1.**
(a) Inbound and outbound passenger flow in residential and commercial area during 2015-03-02 06:00 to 2015-03-02 22:00. (b) Schematic diagram of passenger flow distribution during morning rush hour.

## B. The Influence of External Factors

In addition to the subway stations, there are many factors that affect the passenger flow of the target station, including the properties of the station itself and environmental factors. For predicting time series data, the properties of the target station do not change over time, so these properties do not affect trend prediction. Environmental factors, such as weather and season, are characteristics of the time dimension and have a certain impact on passenger flow. Considering the passengers' age and occupation distributions, the traffic flow is related to the workday [6], i.e., passenger flow during the weekend and the working day are different. In addition, some studies have shown that holidays also have a certain impact on passenger flow. For example, during the Spring Festival, the passenger flow present a special form [7]. Moreover, different seasons and months also have some impact on passenger flow. All of these related factors are called external factors.

C. The Influence of Historical Data

For urban subway, the daily passenger flow is basically the same, which can be considered as a time series data with a daily cycle. And for time series data, historical data contains important information

**PROPOSED SYSTEM:.**

In this paper, we propose a Multi-Type Attention-based Network to forecast the subway passenger flow with multi-station and external factors (subMTAN). The proposed network has different types of attention mechanisms for multi-station, external factors and historical data. It consists of two parts: passenger flow representation and passenger flow forecasting. In the representation part, we use relevant factors to represent the passenger flow at a certain moment. Different attention mechanisms are used to model different data structures and dynamically adjust the weights among different factors. In the forecasting part, we use a temporal attention mechanism to select relevant states across all the timestamps. These two parts can not only adaptively select the most relevant features, but also capture the long-term temporal dependencies of the passenger flow appropriately. Specifically, the attention vector for multi-station in the first part can be used to represent the influences of common stations on the target at different time. This plays an important role in the early warning and dispatch of subway passenger flow.

## B. Problem Statement

Given the previous passenger flow of the target station, i.e. $(y_1,y_2,\ldots,y_{T-1})$ with $y_t \in R$ , as well as the current and past values of $n$ stations,

i.e., $X_{station}=(x_1,x_2,\ldots,x_T) \in R_{n\times T}$ ($x_t=(x_{1t},x_{2t},\ldots,x_{nt})_T \in R_n$ denotes a vector of $n$ stations' passenger flow at time $t$ ), and $m$ external factors, i.e. $Z=(z_1,z_2,\ldots,z_T)$ , the network aims to forecast the target station's passenger flow over $t'$ time, denoted as $y\hat{}=(y\hat{}_{T+1},y\hat{}_{T+2},\ldots,y\hat{}_{T+t'})_T \in R_{t'}$

Fig. 2 shows more vividly the problem and the features of the paper. The black circle denotes the target station, other circles of the right half denote the rest of the stations, and the plane denotes the state of these stations at a timestamp. The circles of the left half denote the external factors. As illustrated in Fig. 2, first, the passenger flow of the target station has temporal dependency on its current state and that of its previous state. Second, it is reflected by its spatial neighbors(other stations). Last, it is also reflected by some external factors. All of these factors affect the passenger flow of the target station comprehensively.

 

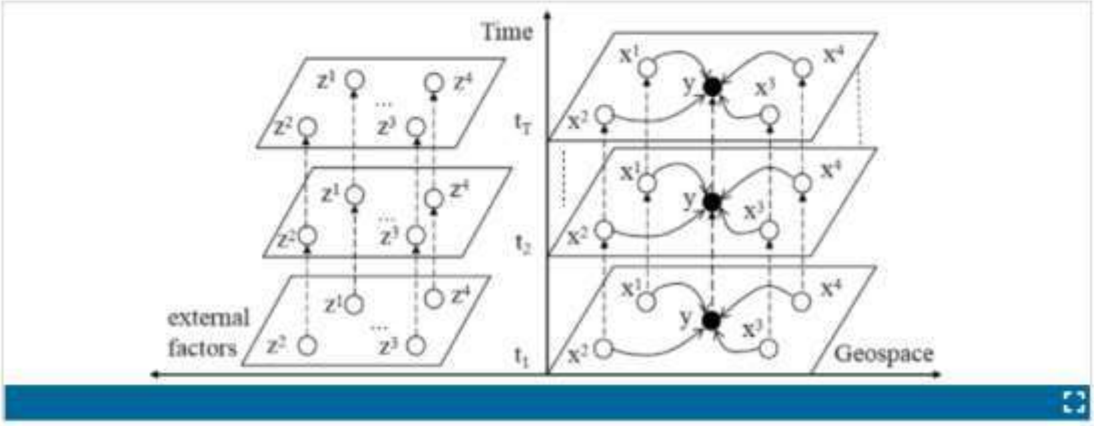


FIGURE 2.
The philosophy of the proposed model.

## Model Comparison and Prediction Results

To demonstrate the effectiveness of the proposed model, we compare our model with five baselines as follows:

**ARIMA** : A well-known model for forecasting future values in a time series. The basic idea of the ARIMA model is to treat the data sequence formed by the predicted object over time as a random sequence, and use a mathematical model to approximate the sequence. In this paper, we do stationary processing and test the order of the model, and finally use ARIMA(0,1,1).

**GBRT** : An ensemble method for the regression tasks and widely used in practice. Gradient Boosting Regression Tree (GBRT) models a data set based on relevant features and predicts the time series using the tree model. In this paper, we use common stations and external factors as relevant features. With grid search, we set the max depth to 10 and the estimator (number of regression trees) to 200.

**DNN** : A deep neural network(DNN)-based prediction model for spatio-temporal data. Same as GBRT, the DNN use all relevant features as input. In this paper, we set two dense layers, and each layer has 256 hidden units.

**LSTM** : A classic recurrent neural network for time series data prediction. Using a unique cellular structure, the LSTM cell can capture the long-term temporal dependencies. In this paper, we use two layers of LSTM with 128 hidden units.

**DA-RNN** : A dual-staged attention model for time series prediction, which shows the state-of-the-art performance in time series prediction. In the first stage, the model use an input attention mechanism to adaptively extract relevant driving series at each timestamp. In the second stage, a temporal attention mechanism is used to select relevant encoder hidden states across all the timestamps. In this paper, we use the same hyperparameters of the DA-RNN and use other stations' passenger flow as driving series.

## Conclusion:

In this paper, we propose a multi-type attention-based network for forecasting the subway passenger flow with the multi-station and external factors. The model contains three different attention mechanisms to adaptively select the relevant spatial and temporal features for the target passenger flow. With this weighted representation, we use encoder-decoder architecture to predict the passenger flow. In order to include more different kinds of data, we bucket the numerical data and add an embedding layer to unify categorical and numerical data. Besides, we use hierarchical attention to model hierarchical data structures. The experiments show that our proposed model achieves the best performance against five baselines in terms of three metrics (RMSE, MAE and MAPE) simultaneously. Moreover, we visualize the attention weights to show the interpretation of all stations in urban subway.

## Futurework:

In the future, we will expand more relevant features, including some text or image information. We believe that more information will help with the prediction. Moreover, we will explore more efficient encoder unit and better model structures. Accurate passenger flow forecasting is important for both cities and passengers, and have a positive impact on the construction of smart city.

## REFERENCES:

S. Hasan, C. Schneider, S. V. Ukkusuri and M. C. González, "Spatiotemporal patterns of urban human mobility", J. Statist. Phys., vol. 151, no. 2, pp. 304-318, Apr. 2013.

Y. Zheng, L. Capra, O. Wolfson and H. Yang, "Urban computing: Concepts methodologies and applications", *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 38, Oct. 2014.

M. D. Lytras and A. Visvizi, "Who uses smart city services and what to make of it: Toward interdisciplinary smart cities research", *Sustainability*, vol. 10, no. 6, pp. 1998, 2018.

J. Yuan, Y. Zheng and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs", *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 186-194, Aug. 2012.

 