Forecast on short-term passenger flow of urban railway stations for normal and special conditions

Guang Zhi-rui, Yang Jun, Li Jian

(Beijing Metro Network Control Center, Beijing 100101）

Abstract: Short term passenger flow forecast of urban rail transit is the key of network operation and management. Meanwhile, it is the basis of passenger flow organization and train optimal allocation. In order to accurately access the impact of various factors on the passenger flow, a predict model of passenger flow entering and departing for normal and special conditions is constructed considering the holidays, weather conditions and air quality. First, the stationarity and periodicity of the time series data were analyzed with autocorrelation and partialautocorrelation functions, and the influences of trend and periodicity were eliminated. Then the Auto Regressive Integrated Moving Average model to predict passenger flow for normal conditions was established by adding virtual indexes such as weather conditions and air quality. Second, considering the land use type, passenger flow prediction model for the circumstances of holidays, odd-and-even license plate rule, Spring Festival travel rush, summer travel peak was constructed by fuzzy *C*-means clustering method and linear regression. Finally, the model was calibrated and validated by the historical passenger flow date collected by AFC system of Beijing Metro, and the error is less than 5%. The results shows that the prediction model has highly accuracy.

key words: urban rail trasit; passenger flow forecast; Auto Regressive Integrated Moving Average; linear regression; fuzzy cluster

1 Introduction

The passenger flow forecast of urban rail transit, which can be divided into long-term forecast, medium-term forecast and short-term forecast, is the basis of line planning and operation organization[1]. Among them, short-term passenger flow forecast is of great importance in operation organization including quick response to traffic fluctuations and the effective allocation of resources, which aim to achieve the maximization of social and economic benefits. There are different characteristics of short-term passenger flow in normal condition and special situation: For normal circumstances, the passenger flow, which is vulnerable to bad weather, air quality and other factors, has three main features of seasonality, trend and randomness; and resemble samples of special circumstances are not easy to obtain because of the change of railway network structure. Thus, it is of great theoretical and practical significance to study the different characteristics of passenger flow for different conditions.

There are four main methods of passenger flow forecasting model, which are four-step method of traffic planning [2], nonparametric model [3, 4, 5] and statistical model[6,7,8,9]. The traditional four-step method ,which is mainly used in line planning , is not timely for fluctuations caused by holidays, bad weather, air quality and other factors. In order to obtain high accuracy, it is necessary to use the massive data for iterative calculation by non parametric models such as Support vector machine (SVM), fuzzy logic, and neural network model, but it is not easy to track the prediction error and to effectively guide the operation management due to its "black box" learning mode.

In this paper, in order to overcome the above problems, based on the statistical analysis of historical passenger flow, the short-term passenger flow forecasting models were set up for normal and special conditions respectively, considering the weather, the land use, and other influencing factors.

2 Prediction models for normal conditions

Based on the analysis of the passenger flow for normal circumstances, firstly linear interpolation method was used to replace the abnormal value due to holidays and other reasons; secondly, adding weather conditions, air quality as a dummy variable, two ARIMA models with/without working day attribute dummy variables were established; Finally, the method of least squares was used to estimate the weight, and a combined forecasting model for time series was constructed.

2.1 Outlier analysis

The accuracy of traffic time series prediction for normal circumstances is affected by outliers. Given that the passenger flow of same working day attribute has the characteristics of short-term stability, outliers were replaced by the linear interpolation method. In mathematics, linear interpolation is a method of curve fitting using linear polynomials to construct new data points within the range of a discrete set of known data points.

If the two known points are given by the coordinates (x_0 , y_0) and (x_1 , y_1), the linear interpolant is the straight line between these points. For a value *x* in the interval [x_0 , x_1], the value *y* along the straight line is given from the equation which can be derived geometrically from the figure on the right.

$$\frac{y-y_0}{y_1-y_0}=\frac{x-x_0}{x_1-x_0} \tag{1}$$

$$y=\frac{x-x_0}{x_1-x_0}(y_1-y_0)+y_0 \tag{2}$$

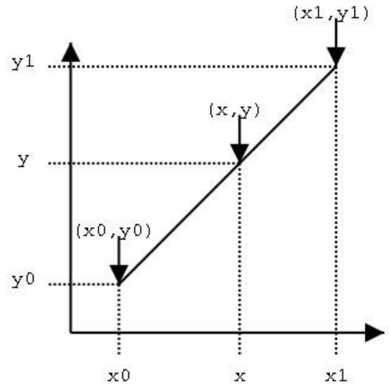


Fig. 1 Schematic diagram of the model

Linear interpolation is often used to approximate a value of some function f using two known values of that function at other points. The error of this approximation is defined as $R_T = f(x) - \rho(x)$, where ρ denotes the linear interpolation polynomial.It can be proven using Rolle's theorem that if *f* has a continuous second derivative, then the error is bounded by $|R_T| \le \frac{(x_1-x_0)^2}{8} \max_{x_0 \le x \le x_1} |f''(x)|$. That is, the approximation between two points on a given function gets worse with the second derivative of the function that is approximated. This is intuitively correct as well: the "curvier" the function is, the worse the approximations made with simple linear interpolation.

2.2 Seasonal ARIMA model

Considering the complex relationship among the seasonal effect, long-term trend effect and random fluctuations, the multiply ARIMA on the forecast of the entrance and exit passenger flow for normal circumstances was estanblished. In order to eliminate the effects of rain, snow and haze, weather conditions and air quality is added as dummy variables. And working day attributes (such as Sunday, Monday) have a direct impact on the entrance and exit passenger flow.One way is to take the seasonal differencing, which can removes the changes in the level of a time series, eliminating trend and seasonality and consequently stabilizing the mean of the time series. Another way is to add working day attribute as dummy variables, which can help accurately grasp the influence of each working day attribute on passenger flow.

Seasonal ARIMA models are usually denoted $ARIMA(p,d,q)\times(P,D,Q)_S$, where *S* refers to the number of periods in each season, and the lowercase *p,d,q* refer to the autoregressive, differencing, and moving average terms for the non-seasonal part of the ARIMA model, and the uppercase *P,D,Q* refer to the autoregressive, differencing, and moving average terms for the seasonal part.

The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The MA part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past.The seasonal ARIMA model formula is as follows[10, 11]:

$$\begin{aligned}
&ARIMA(p,d,q)\times(P,D,Q)_S \\
&\nabla^d\nabla_S^D x_t = \frac{\Theta(B)\Theta_S(B)}{\Phi(B)\Phi_S(B)}\varepsilon_t \\
&\Theta(B) = 1-\theta_1 B-\cdots-\theta_q B^q \\
&\Phi(B) = 1-\phi_1 B-\cdots-\phi_p B^p \\
&\Theta_S(B) = 1-\theta_1 B^S-\cdots-\theta_Q B^{QS} \\
&\Phi_S(B) = 1-\phi_1 B^S-\cdots-\phi_P B^{PS}
\end{aligned} \tag{3}$$

Where *B* refers to delay operator.

2.3 Combination prediction model for normal conditions

In order to improve the prediction accuracy, the linear combination forecasting model was used to combine the two ARIMA models. The combination prediction model is as follows:

$$\hat{F}(t) = \hat{\beta}_1\hat{F}_1(t)+\hat{\beta}_2\hat{F}_2(t) \tag{4}$$

Where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the weights of the two prediction methods.

The target of the forecast weight calibration is to minimize the error between the predicted value and the actual value as far as possible, expressed as the minimum sum of the square residuals. The combination forecasting parameters can be expressed as the optimization model as follows:

$$
\begin{aligned}
&\min Q = \sum (F(t) - \hat{\beta}_1 \hat{F}_1(t) - \hat{\beta}_2 \hat{F}_2(t))^2 \\
&s.t. \\
&\hat{\beta}_1 + \hat{\beta}_2 = 1 \\
&\hat{\beta}_1, \hat{\beta}_2 \in [0,1]
\end{aligned}
\tag{5}
$$

3 Prediction models for special situations

For special circumstances, there are few historical samples, and the network structure may be changed, therefore, the accuracy of time series method to predict the passenger flow entering and departing is low. In this paper, considering the land use around the railway station, the fuzzy C-means clustering method and linear regression model were used to constructe the short-term passenger flow prediction model for special circumstances.

3.1 Influencing factors analysis

The land use along the rail transit is the "source" of the urban rail transit[12], and the passengers' travel behavior in the holidays is different from the surrounding land use. In accordance with the land use, urban railway stations can be divided into residential, office, residential + office, station hub, leisure, and so on. The traditional method to identify surrounding land use relies on field reconnaissance with a lot of manpower and material resources, and the operability and practicability of this method are not satisfactory. Therefore, it is necessary to consider alternative indicators to identify the nature of land use around railway stations[13,14].

The distribution curve of time for entrance and exit per hour passenger flow reveals that the distribution law of time for entrance and exit passenger flow in a whole day varies with the changes of the subway station type. For the residential class station (i.e. HUILONGGUAN, Figure3), there are more passengers to go to work or school than that to go back home in the morning rush hour, which is contrast to the passenger flow in the evening rush hour, so it has large entrance morning peak hour coefficient and exit evening peak hour coefficient. For the office class station, there are more passengers to go to work in the morning rush hour, and more passenger flow to go home in the evening rush hour, so it has large exit morning peak hour coefficient and entrance evening peak hour coefficient are larger. For the residential & office class station, it contains the characteristics of both residential class station and office class station, so the entrance and exit peak hour coefficient in morning and evening are large. And for the hub class station, the passenger flow is large in a whole day, and there is no significant change in morning and evening peak hour. There are a correspondence between the entrance and exit peak hour coefficient and the land use type. Thus, the entrance and exit peak hour coefficients are selected as the index of land use type around stations.

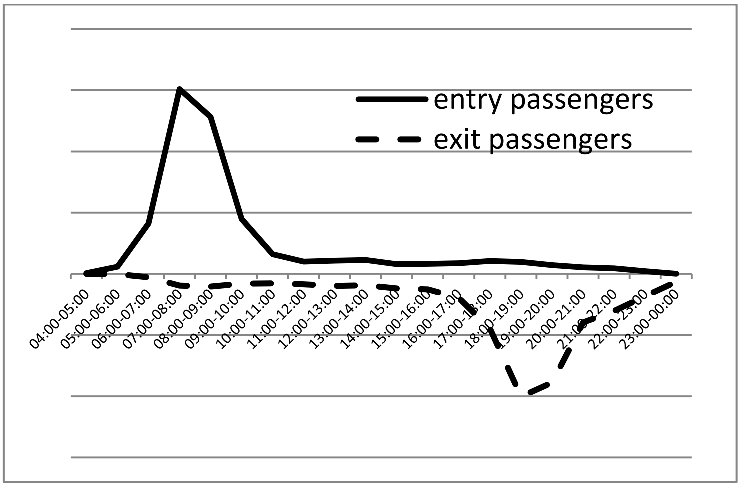
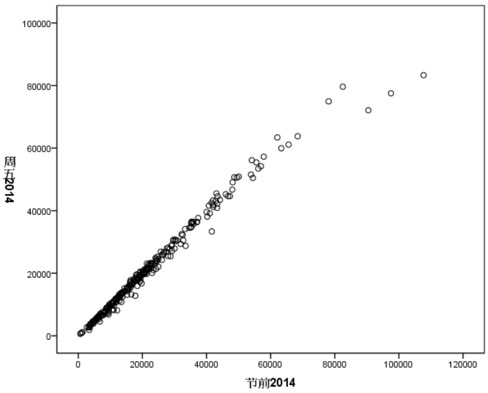


Fig. 2 The distribution curve of passenger flow entering and departing of HUILONGGUAN Station

3.2 Correlation Analysis

Take the passenger flow of the day before Qingming festival vacation in Beijing as an example, the relationship between passenger flow on the day before Qingming festival vacation (short for BQFV) and the recent weekdays was analyzed. The scatter plot and Poisson correlation coefficient shows that, there is a significant correlation between the passenger flow on the BQFV and recent weekdays, particularly on Friday. According to the different land use classification, the correlation is more significant.



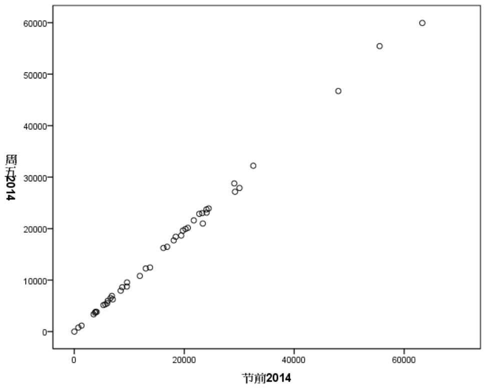


Fig. 3 Scatter plot of the passenger flow on BQFV and Friday

Fig. 4 Scatter plot of the passenger flow on BQFV and Friday of residential class

Tab. 1 the Correlations between each factor and the day before vacation

	recent weekdays	Friday	Friday (residential class)
Pearson correlation	0.986	0.991	0.999

3.3 Prediction model for special situtionr

Starting with the classification of station type by FCM method; then the linear regression model to predict passenger flow for special circumstance was constructed according with the station type; finally, passenger data from the same history holidays was carried out to calibrate parameters.

According to clustering method, data is divided into several clusters by the rule that data in same cluster have high similarities and data differs largely in different clusters. In this paper, Fuzzy C-means (FCM), which is more suitable to data that is more or less evenly distributed around the cluster centers, is applied to classify the subway stations. In order to classify the land use around subway stations, the orginal data of the entrance and exit peak hour coefficients were fuzzy processed. The fuzzy similarity matrix was constructed by using entrance and exit peak hour coefficients as the clustering variables. The subway station was divided into *n* classes according to the land use type by Fuzzy c-means, and *n* is the optimal classification number determined *F* statistic.

The fuzzy C-means clustering algorithm is based on the optimization of the objective function, and the objective function is as follows[15]:

$$J_m(U,V)=\sum_{k=1}^{N}\sum_{i=1}^{C}(u_{ik})^{m}\cdot(d_{ik})^2$$

$$\begin{cases} \sum_{k=1}^{N} u_{ik} \rangle 0 \\ \sum_{k=1}^{N} u_{ik} = 1, i = 1,2,\ldots,\mathrm{C};\ \mathrm{k} = 1,2,\ldots,\mathrm{N} \\ 0 \le u_{ik} \le 1 \end{cases} \tag{6}$$

Where $(d_{ik})^2$ refers to Euclidean distance; *m* refers to the fuzzy exponent, $m \in (1, \infty)$.

The determination of the optimal threshold λ: With different threshold λ, the classification of the fuzzy clustering analysis differs. In general, there are two methods to calculate the threshold.

NO1. According to actual demand, adjust the value of λ to get the appropriate classification in the dynamic clustering diagram, and without the need to advance accurate estimates of the value of categories.

NO2. Calculate the optimal value of λ by F statistic.

$$F > F_{\alpha}(r-1,n-r) \tag{7}$$

Where *n* refers to the total number of samples, *r* refers to the number of categories.

Set the number of categories corresponding to the λ value is r, the number of samples of the j-th class is n_j, the *J*-th class samples are recorded as: $X_1^{(j)}, X_2^{(j)}, \ldots, X_{nj}^{(j)}$ $\mathrm{X}_1^{(j)}, \mathrm{X}_2^{(j)}, \ldots, \mathrm{X}_{nj}^{(j)}$, the *J*-th class cluster center vector is $\overline{X} = (X_1^{(j)}, X_2^{(j)}, \ldots, X_m^{(j)})$ $\overline{\mathrm{X}}^{(j)} = (\overline{\mathrm{X}}_1^{(j)}, \overline{\mathrm{X}}_2^{(j)}, \ldots, \overline{\mathrm{X}}_m^{(j)})$, which $X_k^{(j)}$ $\overline{\mathrm{X}}_k^{(j)}$ is average of the k-th characteristic, that is $\overline{X}^{(j)} = \frac{1}{n}\sum_{k=1}^{n_j} X_k^{(j)} (k = 1,2,\ldots,m)$ $\overline{\mathrm{X}}_k^{(j)} = \frac{1}{n_j}\sum_{k=1}^{n_j} \mathrm{X}_k^{(j)}$

For the *F* statistic,

$$F = \frac{\sum_{j=1}^{r} n_j \left\| \overline{X}^{j} - \overline{X} \right\| /(r-1)}{\sum_{j=1}^{r}\sum_{i=1}^{n_j} \left\| \overline{X}_i^{(j)} - \overline{X}^{(j)} \right\| /(n-r)} \tag{8}$$

F-statistic is *F* distribution which complies with (r-1, n-r), the molecular describes the distance between different classes; the denominator describes the distance between different samples in one class.

Compared with the analysis of the passenger flow law of each station, the advantage of analyzing the law of passenger flow of each type station is that it can adapt to the change of the railway network structure. In statistics, linear regression is an approach for modeling the relationship between a scalar dependent variable *y* and one or more explanatory variables (or independent variables) denoted *x*. Passenger flow prediction formula in holidays is as follows:

$$\hat{y}_{ci} = a_c + b_c x_{ci} \tag{9}$$

$$y_{ci} = \hat{y}_{ci} + \triangle_c \tag{10}$$

$$\triangle_c = t_c s_{yxc} \tag{11}$$

Where $\hat{y}_{ci}$ refers to the passenger flow predictions for special circumstance of station *i*, which belong to cluster *c*; a_c 、 b_c refers to parameters to be determined; x_{ci} refers to passenger flow on recent weekdays or weekend depending on the forecast day attribute. t_c is critical value of t distribution; s_{yxc} means standard error.

4 Evaluation

In this paper, the passenger flow data collected by the AFC of Beijing urban rail transit were used to calibrate the parameters of the model, and the short-term passenger flow forecasting models were constructed for normal conditions and special circumstances, and the error analysis was carried out.

4.1 Accuracy criterion

In order to judge the accuracy of the model, the mean absolute error (MAE) and the mean absolute percentage error (MAPE) were used to analyze the prediction results.

$$\begin{aligned} MAE &= \frac{1}{n}\sum\left|\hat{F}(t) - F(t)\right| \\ MAE &= \frac{1}{n}\sum\left|\frac{\hat{F}(t) - F(t)}{F(t)}\right| \times 100\% \end{aligned} \tag{12}$$

4.2 Prediction example for normal conditions

Taking Tiantongyuan as an example, the model parameters were calibrated by the time series from March to July in 2016. Then the calibrated model was used to dynamic predict the passenger flow in the year of August and September. Finally, the errors of the two ARIMA models and the combined forecasting model were compared and analyzed.

First, outliers has been replaced by the linear interpolation method. The series has a strong and consistent seasonal pattern, an order of seasonal differencing should be used. After that, the series also appear to be non-stationary, so an additional first difference was taken.

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial Autocor]
1	0.2899	0.2903	18.242	0.0000		
2	-0.3564	-0.4813	45.931	0.0000		
3	-0.3439	-0.0756	71.833	0.0000		
4	-0.3461	-0.5086	98.205	0.0000		
5	-0.3549	-0.5534	126.06	0.0000		
6	0.2776	0.1413	143.19	0.0000		
7	0.9586	0.9690	348.38	0.0000		
8	0.2785	-0.2168	365.78	0.0000		
9	-0.3461	0.0034	392.79	0.0000		
10	-0.3353	-0.1778	418.26	0.0000		
11	-0.3364	-0.0248	444.02	0.0000		
12	-0.3456	-0.2060	471.35	0.0000		
13	0.2656	0.0073	487.57	0.0000		
14	0.9246	0.3822	685.16	0.0000		
15	0.2680	-0.1016	701.84	0.0000		
16	-0.3377	-0.0653	728.46	0.0000		
17	-0.3262	-0.1061	753.43	0.0000		
18	-0.3243	0.0205	778.23	0.0000		
19	-0.3358	-0.1217	804.95	0.0000		
20	0.2540	-0.0312	820.33	0.0000		

Fig. 5 Figure of the ACF and PACF

Method 1:

To find an appropriate ARIMA model based on the ACF and PACF was the aim. The significant spike at lag 1 in the ACF suggests a non-seasonal MA(1) component, and the significant spike at lag 7 in the ACF suggests a seasonal MA(1) component. Consequently, we begin with an $ARIMA(0,1,1)\times(0,1,1)_7$ model, indicating a first and seasonal difference, and non-seasonal and seasonal MA(1) components. All the spikes were within the significance limits, so the residuals appear to be white noise. A Ljung-Box test also shows that the residuals have no remaining autocorrelations.

The prediction equation is as follows:

$$F_1(t)=F_1(t-7)+F_1(t-1)-F_1(t-8)+-4.114-623.261x_{wea}-0.724\varepsilon_{t-1}-0.812\varepsilon_{t-7}+0.724\times0.812\varepsilon_{t-8}$$

Method 2:

A seasonal ARIMA model was constructed to add weather conditions and workday attributes as dummy variables. The fitting results shows that except for coefficients of seasonal AR (1) and MA (1) were not statistically significant, other factors were significantly correlated. But the seasonal AR (1) and MA (1) are still useful, the residuals would not be tested by the white noise if the two items were removed.

The prediction equation is as follows:

$$F_2(t) = 32058.24 - 624.71x_{wea} + 21186.55x_{day1} + 21355.49x_{day2} + 21426.77x_{day3} + 21098.23x_{day4} + 21998.37x_{day5} + 4315.71x_{day6} + 0.969(F_2(t-1) - \hat{F}_2(t-1)) + 0.187(F_2(t-7) - \hat{F}_2(t-7)) - 0.969 \times 0.187(F_2(t-8) - \hat{F}_2(t-8)) + \varepsilon_t - 0.706\varepsilon_{t-1} - 0.053\varepsilon_{t-7} + 0.706 \times 0.053\varepsilon_{t-8}$$

Method 3:

Then the combination forecasting model was calibrated by the method of least squares. The combination prediction equation is as follows:

$$\hat{F}(t) = 0.2 \times \hat{F}_1(t) + 0.8 \times \hat{F}_2(t)$$

The table of two kinds of ARIMA model and combined forecasting model prediction error shows that the mean absolute errors of the three models are less than 1700 people, the mean absolute percentage error is less than 4%. Among them, the error of the combined forecasting model is the smallest, the mean absolute error is 935 people, and the mean absolute percentage error is 2.20%

Tab. 2 Prediction performance of different models

Variable	Mean	Median	Std. Dev.
AE	**935**	**796**	**876**
AE1	1643	1180	1410
AE2	1101	918	882
APE	**2.20%**	**1.67%**	**2.25%**
APE1	3.77%	2.41%	3.59%
APE2	2.51%	1.81%	2.31%

The error statistics figure (Fig.6) shows that prediction errors of 89% days are less than 4%, indicating that the combined forecasting model has high accuracy and good applicability.

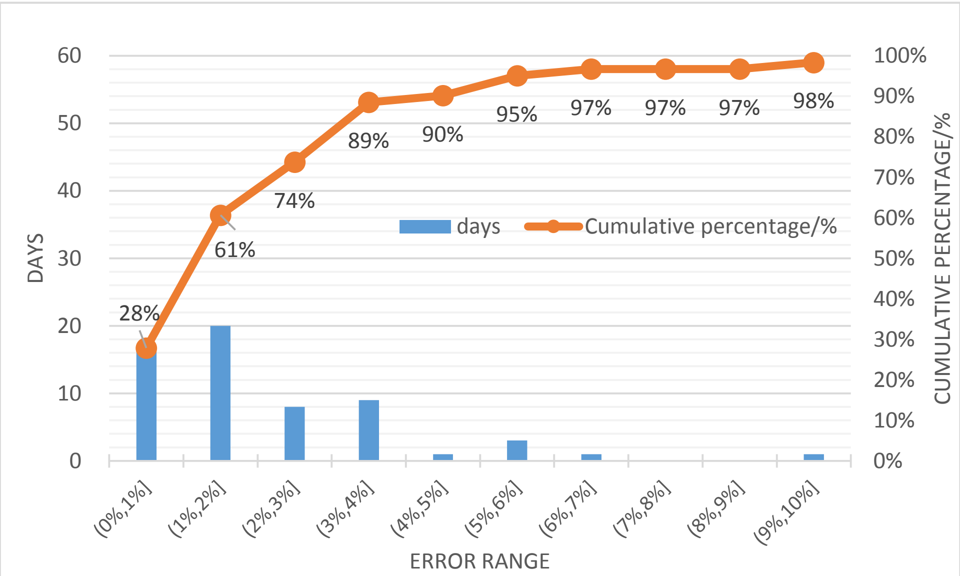


Fig. 6 Error analyses for combined forecasting models

4.3 Prediction example for special sitution

The passenger flow on the day before Ching Ming Festival in 2014 was carried out to calibrate the model, and the model was used to predict the passenger flow in 2015. Finally, the error was analyized compared with neural network and multiple linear regression model.

The station types were divided into 6 classes combining the optimal threshold λ with the specific situation, and Table 3 shows Fuzzy clustering kernel matrix. (*RC* refers to residential class, *RSC* refers to residential superiority class, *OSC* refers to office superiority class, *HC* refers to transportation hub class and *OC* refers to other class). There are a correspondence between the entrance and exit peak hour coefficient and the land use type. (*ENM* refers to entrance morning peak hour coefficient, *EXM* refers to exit morning peak hour coefficient, *ENE* refers to entrance evening peak hour coefficient, *EXE* refers to exit evening peak hour coefficient)

Tab. 3 Fuzzy clustering kernel matrix

Type	*ENM*	*EXM*	*ENE*	*EXE*
RC	0.485	0.066	0.071	0.300
RSC	0.374	0.113	0.117	0.264
OC	0.081	0.403	0.383	0.090
OSC	0.149	0.329	0.311	0.140
HC	0.176	0.252	0.246	0.173
OC	0.268	0.206	0.200	0.211

The residuals has been randomly scattered without showing any systematic patterns. R^2 of each station type, which can be calculated as the square of the correlation between the observed y values and the predicted $\hat{y}$ values is close to 1, thus the predictions are close to the actual values. There were significant differences in regression coefficient significance tests. Above all, the simple linear model is appropriate for these data.

Tab. 4 Regression results of passenger flow entering

Type	N	Adj R-squared	Prob>F	_b_X1	_b_cons	_t_X1
1	32	0.9976	0.0000	1.03	87.28	136.86
2	48	0.9951	0.0000	1.02	182.05	119.49
3	22	0.9980	0.0000	0.98	51.20	100.61
4	41	0.9948	0.0000	0.99	403.70	105.93
5	12	0.9581	0.0000	1.02	1807.62	35.03
6	10	0.9805	0.0000	1.18	306.13	17.97
7	50	0.9922	0.0000	1.00	250.78	125.61

According to the error statistic table of three models, the MAPE of proposed prediction method is 3.98%, which is smallest compared with the neural network model and multivariate linear regression model. The kernel density estimation chart of error analyses for models shows that the error of most stations are less than 4% by proposed prediction method, and they are less than 10% by multivariate linear regression model, and they are less than 20% by neural network model.

Tab. 5 Prediction performance of different models

Statistical indicators	The proposed method (%)	The neural network model (%)	Multivariate linear regression model (%)
Mean	3.98	19.10	8.57
Median	2.40	9.40	6.86
Std. Dev.	5.10	59.47	7.65

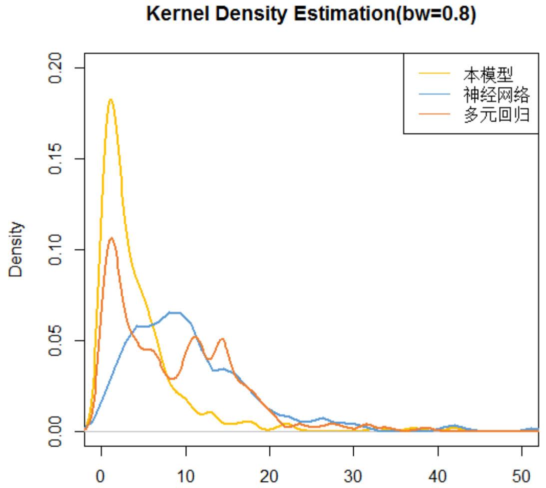


Fig. 7 The kernel density estimation chart of error analyses for models

5 Conclusions

The study on predicting the short-term passenger flow is of great importance in practical. Compared with related studies on the prediction of short-term passenger flow entering and departing, the paper, based on the seasonal ARIMA and linear regression, puts up an practical method to predict the short-term passenger flow for the normal and special conditions.Finally, the model is evaluated with extensive tests carried out in Beijing rail transit network. The results show that the proposed method is easier, reliable, and helpful to urban rail transit operation and management decision-making.

Due to the complex factors of urban rail transit passenger flow, it is necessary to carry on the post evaluation of the passenger flow forecast results, and to analyze the specific reasons for the error. On the basis of distinguishing controllable and uncontrollable factors, the short-term passenger flow forecasting model established in this paper needs to be improved.

ACKNOWLEDGEMENT

This research was supported by the State Key Laboratory of Rail Traffic Control and Safety（Contract No. I11K00050,Beijing Jiaotong University）.

References

[1] VLAHOGIANNI E I, KARLAFTIS M G,GOLIAS J C.Short-term traffic forecasting:where we are and where we're going[J].Transportation research part C:emerging technologies,2014,43:3-19.

[2] SHAO Chunfu. Traffic planning [M]. Beijing: China Railway Publishing House.2004. (in Chinese)

[3] ZHOU Jiazhong, ZHANU Dianye. Direct ridership forecast model of urban rail transit stations based on spatial weighted LS-SVM [J].Journal of the China Railway Society, 2014,36(1):1-7.

[4] LI Xiaojun,Lv Xiaoyan. Forecast of railway short-term passenger flow based on rbf neural network [J].Railway Transport and Economy, 2011,33(6):86-89. (in Chinese)

[5] KHAN,A M.Bayesian predictive travel time methodology for advanced traveler information system[J].Journal of advanced transportation,2012,46(1):67-79.

[6] MU T,JIANG J,WANG Y.Heterogeneous delay embedding for travel time and energy cost prediction via regression analysis[J].Intelligent transportation systems,IEEE transactions on intelligent transportation systems,2013,14(1):214-224.

[7] WANG Lina.Traffic forecasting and scheduling of the urban passenger rail network based on historical data [D]. Beijing: Beijing Jiaotong University, 2011. (in Chinese)

[8] LI Dewei,YAN Yixing, ZENG Xianfeng.Combined Short-term Prediction Model of Station Entry Flow in Urban Rail Transit [J].Urban Rapid Rail Transit, 2017,30(1):54-58. (in Chinese)

[9] Prediction of urban railway station's passenger entrance and exit flow based on multiplys ARIMA model[J]. Journal of Beijing Jiaotong Uneversity, 2014, 38(2):135-140.

[10] Geoge E, Box P.Time series analysis: foreasting and control [M].GU Lan, FAN Jineheng, trans. Beijing: China Statistic Press,1997:101-135. (in Chinese)

[11] WANG Yan.Applied time series analysis[M].Beijing:China Renmin University Press,2005.(in Chinese)

[12] YANG liya, Interactive relationship between urban transportation and land use: models and methodology [M]. Beijing: China Architecture&Building Press, 2012. (in Chinese)

[13] GUANG Zhirui. Passenger flow prediction for urban railway station's entrance and exit based on land-use and accessibility [D]. Beijing: Beijing Jiaotong University, 2011. (in Chinese)

[14] GUANG Zhirui,YAO Enjian,ZHANG Yongsheng. Prediction of urban railway station's entrance and exit passenger flow based on fuzzy clustering analysis [C] //Inter-national Conference on Railway Engineering,2012:70-75.

[15] GAO Xinbo. Fuzzy cluster analysis and its applications [M]. Xi'an: Xidian University Press, 2004. (in Chinese)