PAPER • OPEN ACCESS

Short-term Passenger Flow Prediction for Urban Railway Transit Based on Change-point Model

To cite this article: Enjian Yao *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **587** 012102

View the article online for updates and enhancements.

Short-term Passenger Flow Prediction for Urban Railway Transit Based on Change-point Model

Enjian Yao[1st*1,2], Chuanyu Chen[2nd,3], Muyang Lu[3nd,4], Ying Zhou[4th,3]

[1.]The School of Traffic and Transportation, Beijing Jiaotong University

[2.]Key Laboratory of Transport Industry of Big Data Appalication Technologies for Comprehensive Transport Beijing, China

[3]The School of Traffic and Transportation, Beijing Jiaotong University Beijing, China

[4]The Pennsylvania State University PA, USA

[*] Corresponding author: enjianyao@bjtu.edu.cn

Abstract-For short-term passenger flow in urban railway transit has the characteristics of nonlinear distribution, a combined forecasting model based on variable-point model and neural networks model is proposed. First, the Pettitt method, the optimal segmentation algorithm, the BG segmentation algorithm(BGSA) and the wavelet analysis method are used to detect the change point of the passenger flow sequence at the memorial hall station, identify the change point of the passenger flow curve and divide the interval，and the passenger flow is predicted and verified by using the Multilayer neural network (MLP) and the radial basis function neural network (RBF). The results show that the wavelet analysis-RBF combined model based on change point detection has the root mean square error of 24.20 and the mean absolute percentage error of 3.52%. Compared with the single neural network model, it improves forecasting accuracy.

1. INTRODUCTION

As the framework of urban public transport network, urban railway transit is of great significance for improving the operation service level of urban railway transit and providing technical support for the operation safety of urban railway transit through real-time analysis of passenger flow data and accurate prediction of passenger flow in future periods.

When the passenger flow statistics of urban railway transit are made with "hour" and "minute" time granularity, it appears more obvious morning peak, midday peak, evening peak and other phenomena. And the passenger flow is dynamic, nonlinear and uncertain, it is not easy to grasp the inherent law of passenger flow, and directly predicting the passenger flow often has certain errors. In order to ensure the consistency of the characteristics of urban railway transit passenger flow data, through the change-point detection model of the passenger flow sequence, the law of passenger flow in different periods of time is studied in detail, so as to make more accurate short-term passenger flow prediction.

The existing short-term urban railway transit passenger flow prediction methods can be roughly divided into the following two categories: 1. Linear prediction methods, mainly including time series model, Kalman filter model. [1][2]; 2. Nonlinear prediction methods, mainly including neural network model, support vector machine, gray model, non-parametric regression. [3][4][5][6]. Although the linear

prediction method is simple and fast in calculation, it is difficult to reflect the dynamic and uncertain characteristics of the urban railway passenger flow. For the nonlinear prediction methods, the neural network model has better self-learning ability and higher prediction accuracy. In this paper, through the change-point detection model[7], the whole-day passenger flow is divided into different periods of passenger flow with different characteristics, and then the neural network model is used to predict the passenger flow at each period, and finally the whole-day predicted passenger flow is processed.

2. INTERVAL DIVISION BASED ON CHANGE-POINT

Urban railway passenger flow within a day of different time periods have different data characteristics, such as the morning peak, midday peak and evening peak. The accuracy of direct passenger flow prediction is low. But change-point detection model can divide passenger flow data into intervals with similar statistical characteristics, and accurately study of passenger flow laws at different periods to effectively improve the accuracy of prediction.

The change-point theory is a classic branch of statistics. When statistical characteristics (distribution type, distribution parameter) in a sequence changes at a certain point due to systemic factors, it is called the change-point. The change-point detection model can use statistics methods to estimate the position of the change point. In this paper, there are mainly pettitt [8], optimal segmentation algorithm [9], BG segmentation algorithm [10], and wavelet analysis [11].

2.1. Pettitt Model

Pettitt model is a non-parametric test method, and its principle is as follows:
First, constructing order columns:

$$S_k = \sum_{i=1}^{k} r_i, \ k = 2, 3, \cdots, n \tag{1}$$

Where, $r_i = \begin{cases} +1, & \text{when } x_i > x_j \\ 0, & \text{when } x_i = x_j, \ j = 1, 2, 3 \ldots, n \\ -1, & \text{when } x_i < x_j \end{cases}$

Then, the time t_0 satisfy:

$$k_{t_0} = \max |s_k|, \ k = 2, 3, \cdots, n \tag{2}$$

$$P = 2 \exp \left[-6 k_{t_0}^2 / \left(n^3 + n^2 \right) \right] \tag{3}$$

When $P \le 0.5$, the point is considered to be a change point of the sequence.

2.2. BG Segmentation Algorithm

The principle of BG segmentation algorithm is as follows:
First, for a sequence $\{x_1, x_2, ..., x_j\}$, the sequence before and after each point is divided into x_{1i} and x_{2i} in turn, $i = 1, 2, 3, \cdots, n$. The joint deviation SD_i of the point i is:

$$SD_i = \left(\frac{S_{1i}^2 + S_{2i}^2}{n_1 + n_2 - 2} \right) \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \tag{4}$$

Then, constructing the t-test statistic T_i:

$$T_i = \left| \frac{\mu_{1i} - \mu_{2i}}{S_{D_i}} \right| \tag{5}$$

Where, u_{1i} and u_{2i} are the mean values of sample x_1 and x_2 respectively; S_{1i} and S_{2i} are the mean standard deviation of sample x_1 and x_2 respectively; n_1 and n_2 are the sample size of sample x_1 and x_2 respectively.

And, repeating the (4) and (5) for each x_i in turn to test the statistical significance of $T_{\max}$:

$$P\left(T_{\max}\right) = P\left(T \leq T_{\max}\right) \tag{6}$$

Where, $P\left(T_{\max}\right) \approx \left(1 - I_{u\left(v+T_{\max}^{\delta}\right)}(\delta v, \delta)\right)^{\gamma}$. From Monte-Carlo simulation: $\delta = 0.40$, $v = n - 2$, $\gamma = 4.19\ln(n) - 11.54$, $I_x(a,b)$ is the incomplete beta function.

Therefore, the point with $P\left(T_{\max}\right)$ is the change-point.

2.3. Optimal Segmentation Algorithm

The optimal segmentation algorithm is to use sum of the squares within groups(SSW) to represent the difference degree between similar samples, so as to minimize the difference between similar samples and maximize the difference between dissimilar samples.

Suppose that sequence $\{x_i, x_{i+1}, ..., x_j\}$ is the subsequence of sequence $\{x_1, x_2, ..., x_n\}$ ($i \leq j \leq n$), the mean value of the subsequence is $x(i,j) = \dfrac{1}{j-i+1}\sum_{k=i}^{j}x_k$, the mean vector of the subsequence is $x(i,j) = (x(i,j)_1, x(i,j)_2, \cdots ..., x(i,j)_m)$, the SSW of the subsequence is $D_{(i,j)} = \sum_{k=i}^{j}|x_k - x(i,j)|^2$.

Therefore, the sum of SSW is:

$$P(n,g) = \sum_{r=1}^{g} D_{(i_{r-1}+1, i_r)} \tag{7}$$

The classification corresponding to the minimum value of $P(n,g)$ becomes the optimal segmentation.

2.4. Wavelet Analysis

Based on the invariance of translation and expansion, wavelet analysis has good properties of regularity and locality. Wavelet transform method describe time sequence through frequency, time and space, so it is particularly suitable for multi-scale analysis of time sequence.

The morlet wavelet is a commonly used complex wavelet function, and its wavelet transform is:

$$\omega_{a,b} = \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{a}} \overline{\Phi}\left(\frac{x-b}{a}\right) dx \tag{8}$$

Where, $\omega_{a,b}$ is the signal square function; $a > 0$ is the distinguish scale; $b \in R$ is the translation factor; $\overline{\Phi}$ represents conjugate.

The real-part of the wavelet coefficient is the most important variable obtained by morlet wavelet transform, containing the given time and scale information, so it is usually constructed to reveal the law of time sequence.

3. EXAMPLE ANALYSIS

Based on the method proposed in this paper, the arrival volume data of Guangzhou Memorial Hall Station on Mondays from October 10, 2016 to December 5, 2016 were selected, and the original passenger flow sequence was generated using 15 minutes time granularity. The original passenger flow data was obtained for 10 days in total, and 72 arrival volume observations were obtained on a complete operating day, with the serial number set as 1-72. First, experiment fits original sequence. passenger flow through. Based on pettitt method, the optimal segmentation algorithm, BG segmentation algorithm and wavelet analysis method, it identify change-points and divided interval, and then select reasonable dividing interval method. Finally, for the division of the interval, it respectively adopt multi-layer neural network (MLP) and radial basis function neural network (RBF) to predict passenger flow, and verify it.

3.1. Change-point Detection and Interval Division

The passenger flow time sequence of Guangzhou Memorial Hall station from 06:00 to 24:00 every Monday from October 10, 2016 to December 5, 2016 is obtained and fitted by 15 minutes time granularity, as shown in Fig. 1. Based on the fitting curve, the change-point is detected by Petit, optimal segmentation algorithm, BG segmentation algorithm and wavelet analysis method, and different interval partition is obtained.
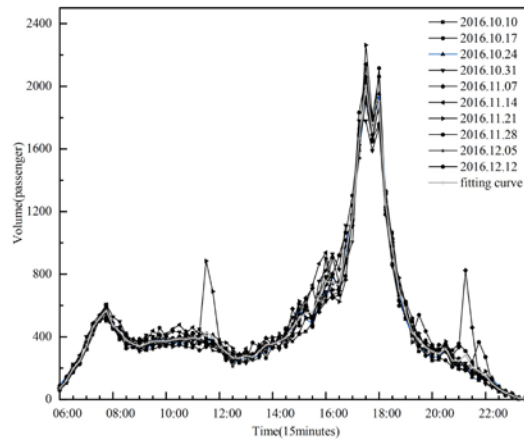

Figure 1. Time sequence of arrival passenger flow at Memorial Hall Station on different dates.
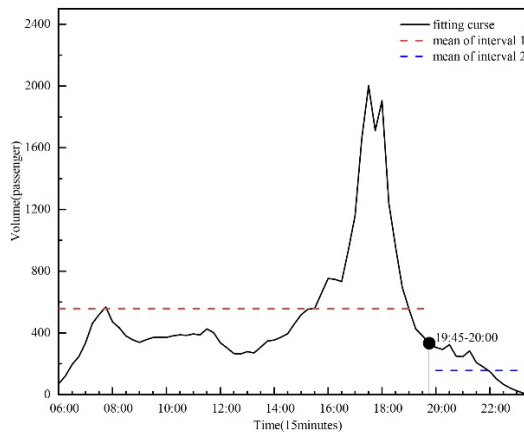

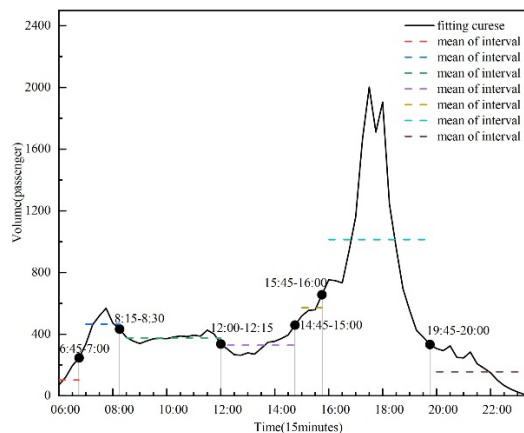Figure 2. Change-point detection and interval division based on pettitt

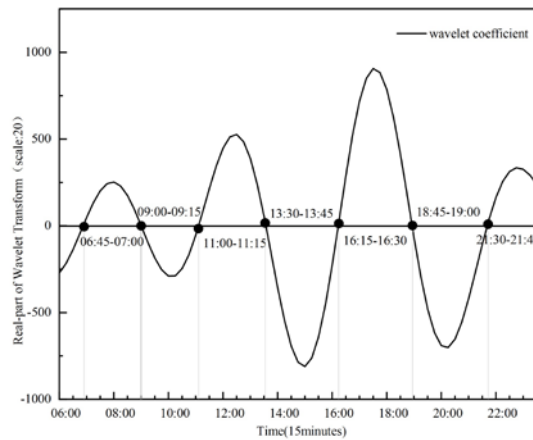Figure 3. Change-point detection and interval division based on BGSA

Figure 4. Change-point detection and interval division based on wavelet analysis

TABLE I.   CHANGE-POINT DETECTION AND INTERVAL DIVISION BASED ON OPTIMAL SEGMENTATION
ALGORITHM

No. Layer	1，2，3，4，5，6……70，71，72					
Two	[1,39]			[40,72]		
Three	[1,39]			[40,53]	[54,72]	
Four	[1,11]	[12,39]		[40,53]	[54,72]	
Five	[1,11]	[12,25]	[26,39]	[40,53]	[54,72]	
Six	[1,11]	[12,25]	[26,39]	[40,53]	[54,67]	[68,72]
Seven	[1,11]	[12,25]	[26,39]	[40,53]	[54,67]	[68,72]

As shown in Fig. 2, the interval division result of pettitt method is (06:00,20:00]、(20:00,24:00]. As shown in Fig. 3, the interval division result of BGSA method is (06:00,07:00]、(07:00,08:30]、(08:30,12:15]、(12:15,15:00]、(15:00,16:00]、(16:00,20:00]、(20:00,24:00].

As shown in Fig. 4, the interval division result of wavelet analysis method is (06:00,07:00]、(07:00,09:15]、(09:15,11:15]、(11:15,13:45]、(13:45,16:30]、(16:30,19:00]、(19:00,21:45]、(21:45,24:00].

As shown in TABLE I, the interval division result of optimal segmentation algorithm is (06:00,08:45]、(08:45,12:15]、(12:15,15:45]、(15:45,19:15]、(19:15,22:45]、(22:45,24:00].

It can be seen from Fig. 1 that there is an obvious morning peak period of passenger flow at Memorial Hall Station, corresponding to (07:00, 08:30] of BG segmentation algorithm, (06:00,08:45] of optimal segmentation algorithm, and (07:00, 09:15] of wavelet analysis. Among them, the optimal segmentation algorithm gets the earliest start time and the longest duration of the morning peak period. This method does not divide the transition period before the arrival of the early peak, so it increases the division interval of the early peak.

The arrival passenger flow of Memorial Hall station also has obvious evening peak time period, corresponding to the interval corresponding to (16:00,20:00] of BG segmentation algorithm, (15:45,19:15] of optimal segmentation algorithm, (16:15, 19:00] of wavelet analysis. As an office station, Memorial Hall Station has more obvious passenger flow in the late peak compared with the morning peak, and the volume of it is larger. It matches with the wider division interval of evening peak obtained by BG

segmentation, optimal segmentation and wavelet analysis. For the transition period between evening peak, BG segmentation algorithm and wavelet analysis also have more detailed division.

Comparing the four methods, pettitt method only divides two intervals, the optimal segmentation algorithm divides intervals of the morning peak and the evening peak time period, while the BG segmentation algorithm and wavelet analysis not only reflect the morning and evening peak time periods accurately, but also divide each peak time period carefully, which shows that the reliability of the change-point model is high. Therefore, this paper predict the passenger flow based on the divided intervals of BG segmentation algorithm and wavelet analysis.

3.2. Passenger Flow Prdiction Based on Interval Division of Change-point Detection

The combined model of change-point detection and neural network is used to predict each partition interval, and finally the complete daily passenger flow prediction curve is obtained by weighting.

$$Z_t = \sum_{i=1}^{n} \varphi_i M_i \qquad (9)$$

Where, Z_t is the forecast value of the combination model; φ_i is the weight of a partition interval prediction model in the combination model; M_i is the predicted value of a certain partition interval.

Based on the partition interval obtained by BG segmentation algorithm and wavelet analysis, MLP and RBF neural network are used to predict the passenger flow of each partition interval, and the predicted passenger flow curve of the combined model is obtained by weighting. Compared with the single neural network prediction, the prediction results of each model are shown in Fig. 5.
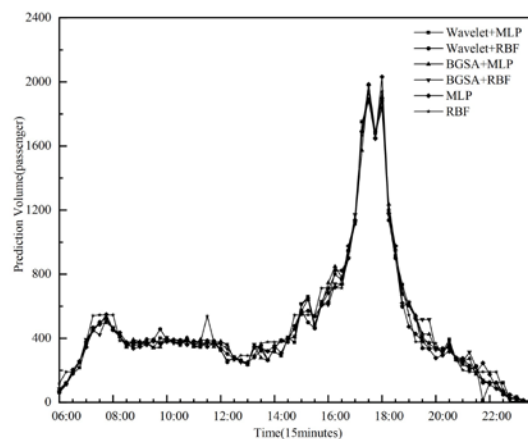

Figure 5. The passenger flow prediction results of each model

The prediction results of the combined model and the single model were analyzed to evaluate the fitting degree of each model with the real value. The model prediction error is shown in TABLE II.

TABLE II. THE ERROR OF PREDICTION MODEL

Prediction Model	RMSE	MAPE(%)
Wavelet-MLP	30.57	6.43
Wavelet -RBF	24.20	3.52
BGSA-MLP	30.04	14.09
BGSA-RBF	37.86	6.10
MLP	48.53	19.30
RBF	59.33	27.15

As shown in TABLE II, the root mean square error (RMSE) and mean absolute percentage error (MAPE) of the combined model based on interval partition are decreased compared with the single neural

network model. Among them, the RMSE of wavelet-RBF model is 24.20, the MAPE of wavelet-RBF model is 3.52%. And the prediction error of it is the lowest, so the prediction accuracy is highest.

It shows that the volatility of passenger flow data has a great impact on the prediction results, and wavelet analysis is more suitable to deal with the volatility of daily passenger flow time series. And the radial basis function neural network (RBF) has better characteristics of capturing the periodic variation of passenger flow than the multi-layer neural network (MLP). It can capture the law of each partition interval with time and accurately predict the future passenger flow.

The wavelet analysis is used to identify the change-point of passenger flow time sequence and divide the interval. The radial basis function neural network (RBF) is used to predict the passenger flow. And the prediction accuracy of wavelet-RBF model is higher than the single neural network model. Therefore, the wavelet-RBF combination model based on change-point detection proposed in this paper has good prediction effect on urban rail passenger flow.

4. CONCLUSION

According to the characteristics of urban railway passenger flow data with different passenger flow rules in different time periods, the passenger flow rules in different periods are studied in detail, so as to make more accurate short-term passenger flow prediction. In this paper, the change-point detection method which identifies the change-points of passenger flow sequence and divides different passenger flow intervals is applied to the short-term passenger flow field of urban railway transit. Through the change-point detection of passenger flow sequence in Memorial Hall Station, the change-point of passenger flow is identified and the interval is divided. The multi-layer neural network (MLP) and radial basis function neural network (RBF) models are used to predict and verify the passenger flow. The results show that the wavelet-RBF combined model based on change-point detection has better prediction effect than single neural network model.

REFERENCES

[1] Zhang Zhiyong, Zhang Dandan, JIA Janlin, Liang Tianwen. Short-term Passenger Flow Forecasting of Rail Transit Platform Based on Improved Kalman Filter[J]. Journal of Wuhan University of Technology (Transportation Science Engineering) , 2017,41(06):974-977. （in Chinese）

[2] Meng Pinchao, Li Xueyuan, Jia Hongfei, Li Yanzhong. Short-time rail transit passenger flow real-time prediction based on moving average. Journal of Jilin University(Engineering and Technology Edition),2018,48(02):448-453. （in Chinese）

[3] Gao Meng-qi, Chang Ximing Wang Huan. Short-term forecasting of the subway passenger flow based on wavelet decomposition and long short-term memory network[J].Shandong Science,2019,32(04):56-63. （in Chinese）

[4] Chen Q, Song Y, Zhao J. Short-term traffic flow prediction based on improved wavelet neural network[J]. Neural Computing and Applications,2020(2).

[5] Yang Jing, Zhu Jingwei, Liu Bo, Feng Cheng, Zhang Hongliang. Short-term Passenger Flow Prediction for Urban Railway Transit Based on Combined Model[J]. Journal of Transportation Systems Engineering and Information Technology,2019,19(03):119-125. （in Chinese）

[6] Yoshiyuki, Ninomiya. Information criterion for Gaussian change-point model[J]. Statistics and Probability Letters,2004,72(3).

[7] Wang Huimin，He Xingshi，Zhao Wenzhi. The statistical analysis and case study of change points[J]. Electronic Test,2013(18):38-40. （in Chinese）

[8] Conte,Bayer,Bayer. Bootstrap Pettitt test for detecting change points in hydroclimatological data: case study of Itaipu Hydroelectric Plant, Brazil[J]. Hydrological Sciences Journal,2019,64(11).

[9] Xu Yumu Niu Xiuling. Study on the Influence of Different Domains on Flood Season Division Based on Improved Optimal Dissection Method[J]. Journal of Hydroelectric Engineering,2019,45(11):19-22+116. （in Chinese）

[10] Li Jing.Analysis of Characteristics and Influencing Factors for River Runoff[D]. Taiyuan University of Technology,2015. （in Chinese）

[11] Qiu Sen.Prediction of Urban Rail Transit Passenger Flow Rime Series Based on Wavelet Analysis and Neural Network[D]. Beijing Jiaotong University,2017.(in Chinese)