

Short-term Passenger Flow Prediction on Bus Stop Based on Hybrid Model

Zhijian Wang, Chunlei Yang* and Chao Zang

Beijing Key Lab of Urban Intelligent Traffic Control Technology, North China University of Technology, Beijing 100144, China

*Corresponding author

Abstract—Short-term passenger flow prediction on bus stop is an important base and technical support for bus dispatch strategy. In this paper, a new hybrid prediction model including two single models of BP neural network and time series model was proposed according to the periodicity and randomness properties of short-term passenger flow. By using the IC card data from buses, a three-layer BP neural network model was established to reflect the characteristics of the stability of short-term passenger flow cycle. The time series model of cash ticket passenger flow data was established. Finally, the results of the two models were fitted to get the final prediction results. The results show that the proposed hybrid prediction model can effectively predict short-term passenger flow on bus stop.

Keywords—short-term traffic flow prediction; BP neural network; time series model; hybrid model

I. INTRODUCTION

Bus is a means of transportation which makes efficient use of road resources. Grasping the changing rules of passenger flow and accurately predicting the passenger flow are the basis and key for bus companies to make scientific plans for operation [1]. Short-term passenger flow prediction on bus stop is an important base and technical support for decision-making on the intelligent bus dispatch system [2].

There are many methods for short-term passenger flow prediction: Time-series model [3], neural network model [4], nonparametric regression model [5], linear regression model [6], Kalman filter method [7], prediction model based on chaos theory [8]. Li Yuancheng et al used wavelet principle to divide the traffic flow data into many smooth vectors. Then SVM model was used to predict these vectors, and finally the data of each component were reconstructed. Tan Manchun et al. adopted ANN model and ARIMA model separately to predict the traffic flow nonlinearly and linearly. However, because of the poor adaptability of ARIMA model compared with ANN model, the parameters of the two models cannot ensure the synchronous update [9]. Fan Na proposed a new hybrid prediction model including two single models of BP neural network and non-parametric regression model, the two models more effectively combined for the short-term passenger flow prediction provides a new direction [10].

Based on the above researches, this paper proposes a new hybrid model by combining BP neural network with ARIMA model to predict short-term passenger flow, after fully considered the stability and randomness of short-term passenger flow. The hybrid model is established to give full

play to the predictive characteristics of each individual model, and at the same time has the characteristics of complementary fusion between models so as to predict the passenger flow on bus stop more effectively and improve the prediction accuracy.

II. SHORT-TERM PASSENGER FLOW CHARACTERISTICS

Short-term passenger flow on bus stop is affected by fixed factors and random factors such as vehicle type, platform design, road traffic status, etc. Under the joint action of different factors, passenger flow on bus stop has the characteristics of cyclicity, randomness and variability. The cyclical performance of the daily passenger flow similarities and the similarity of the same period passenger flow, due to the working day to comply with the law of early departure and evening, resulting in passenger bus travel time and frequency similarities, the final performance of the bus station short When the cyclical passenger flow. From the perspective of individual behaviors, it is relatively independent because each passenger's travel is affected by many factors, like weather, holidays and work characteristics. The performance of passenger demonstrates the randomness of passenger flow. Therefore, Analysis of the bus station passenger flow curve when there will be some different trends.

The characteristic of short-term passenger flow on bus stop is obviously different from that of long-term passenger flow: short-term passenger flow time series non-stationary more significant; short-term passenger flow randomness and uncertainty more obvious, this is because the latter has a large time scale for observation, impairs randomness at the expense of lose information integrity. The related factors of short-term prediction are more difficult to capture and analyze. These differences make the research of short-term passenger flow forecasting more difficult.

III. SHORT-TERM PASSENGER FLOW PREDICTION MODEL

A. BP Neural Network

BP neural network is currently the most widely used to achieve the most intuitive way, the easy way to understand the mechanism of computing, research a deep artificial neural network. The algorithm usually consists of input layer and propagation. Compared with the traditional model, the algorithm has a better fault tolerance, while both self-learning, non-linear mapping and parallel distributed processing capabilities. Generally speaking, neural network model was established after training and learning of system input-output

samples can always approach all complex nonlinear functions with arbitrary precision.

The model structure and weights of BP neural network are obtained through a learning process. The learning process are divided into two stages: Multi-layer feed-forward stage, which is from the input layer in turn calculate the actual input and output of each node of each layer, and Reverse error correction phase, that is, according to the output layer neuron output error, reverse along the road to amend the weight of each connection, so that the error decreases.

Multi-layer feedforward mathematical model is:

$$\begin{cases} y_i^l = f(x_i^l) \\ x_i^l = \sum_{j=1}^{N_{l-1}} w_{ij}^l y_j^{l-1} + \theta_i^l \end{cases} \quad (i=1,2,3,\dots,L) \quad (1)$$

Where y_i^l is the output value of the i node of the l th layer; x_i^l is the l th layer the activation value of the i node; w_{ij}^l is the j node of $l-1$ layer to the i node of the l layer has the connection weight; θ_i^l is the i section of the l layer point threshold; N_l is the number of nodes in the l layer; L is the total number of layers; $f()$ is neuron activation function.

In the forward feed-forward process, according to the above formula to calculate the input and output of each layer until the error of output layer neurons cannot meet the accuracy requirements, then enter the error propagation phase. The error backpropagation stage adopts the gradient descent algorithm, which adjusts the connection weights between the neurons in each layer, so that the total error changes in a decreasing direction.

The mathematical expression is:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \quad (2)$$

The weight adjustment formula is:

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial E}{\partial w_{ij}} \quad (3)$$

B. The Time Series Model

The time series model is a series of numbers formed according to the time order of a statistical index. Its method is to compile and analyze the time series, according to the development process, the direction and the trend reflected in the time series, and so as to predict the level that will be reached after a certain period of time in the next period or later. The time series constructed in this paper also belong to the non-stationary time series, so the *ARIMA* model is chosen. The general expression of this model is:

$$A(f)(1-h)^d X_t = B(f)\varepsilon_t \quad (4)$$

Where d is the observation data for d differential processing; X_t for the time series value; f for the time t backward shift operator.

$$A(z) = 1 - \sum_{j=1}^p a_j z^j \neq 0 (|z| \leq 1) \quad (5)$$

$$B(z) = 1 + \sum_{j=1}^q b_j z^j \neq 0 (|z| \leq 1) \quad (6)$$

Where q, p is its order, short for *ARIMA*(p, d, q); ε_t is the error or impact value; a_j is the autoregressive coefficient; b_j is the moving average coefficient.

In this paper, the statistical analysis software SPSS was used in constructing the time series forecasting model by predicting and analysis the time series changes of random passenger flow. The steps are as follows:

(1) *Data processing*. If the data sequence is non-stationary, the data is smoothed, that is, the difference or the natural logarithm, and the number of differences denoted as d .

(2) *Model recognition*. After analyzing the autocorrelation function (*ACF*) and the partial autocorrelation function (*PACF*) of the time series processed in the first step, the autocorrelation order and the moving average order are determined according to the censoring of *ACF* and *PACF* in the first few steps, and the Initial model was got.

(3) *Parameter estimation and model diagnosis*. Different coefficients were used to determine the coefficients of the initial model *ARIMA* (p, d, q) accurately. The coefficients in the model were tested for significance and white noise. The values of *BIC* and *LBQ* were observed.

IV. EXPERIMENTAL STUDY

This paper selected Beijing Pingguoyuan east Station, with a total 19 bus lines travelling through, as the observation station. The bus lines are shown in Figure.1. Due to the large amount of passenger trips on bus stop at peak hours on weekdays, the passenger flow is cyclical with obvious randomness. So we chose to count the data of all-day passenger flow on the station for two continuous weeks of working day, in August 2017, and separately investigate the passenger flow with bus IC card and the cash-on-board passengers on bus stop for short-term passenger flow prediction.

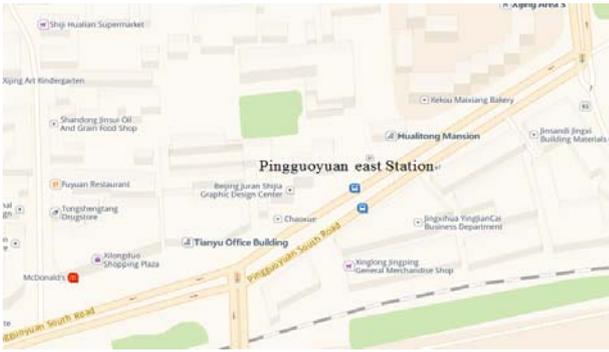


FIGURE I. THE BUS STATION ROUTE

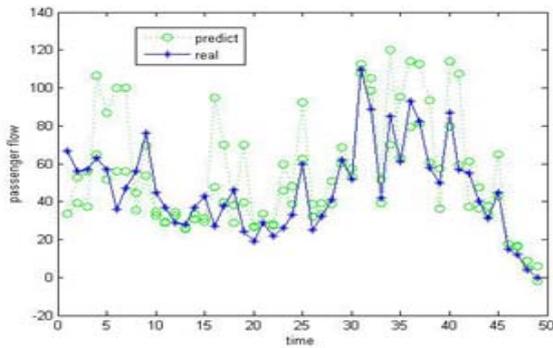


FIGURE II. NEURAL NETWORK OUTPUT PREDICTION RESULTS

A. BP Neural Network Model

The training data sources are collected from Pingguoyuan east stop of bus IC card holding passenger flow from 06: 10 to 21: 00, every 10 minutes as statistics intervals. Due to the obvious periodicity of IC card data, the program works well and reaches about 10 times of training to achieve convergence. The comparison chart between the test data and the prediction data is demonstrated in Figure.2. The error is basically within 10%, which is consistent with the tolerable range of error and the model accuracy is high.

B. Time Series Model

On the same day, the cash-on-board passenger flow of the previous model was taken as the sample data. Similarly, the data was divided into 89 time periods at intervals of 10 minutes. The previous 80 groups are the training data and the last 8 groups are the test data, as shown in Figure.3. After the model is established, the residual autocorrelation graph and partial autocorrelation graph was observed, as shown in Figure.4, and judged the residual sequence as white noise sequence, indicating that the established model is an ideal ARIMA model.

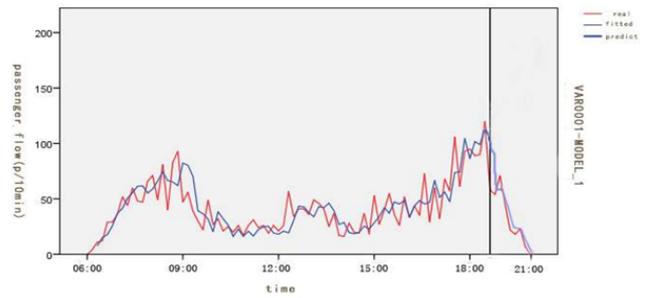


FIGURE III. TIME SERIES PREDICTION RESULTS

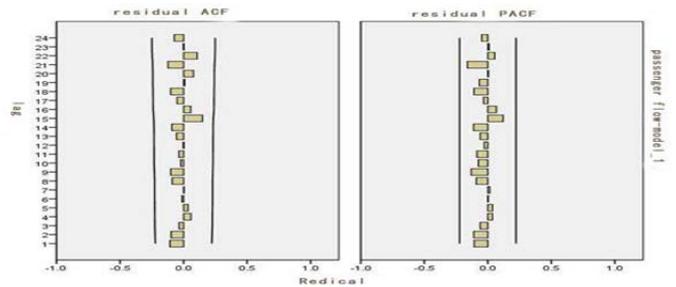


FIGURE IV. THE RESIDUAL AUTOCORRELATION AND PARTIAL AUTOCORRELATION

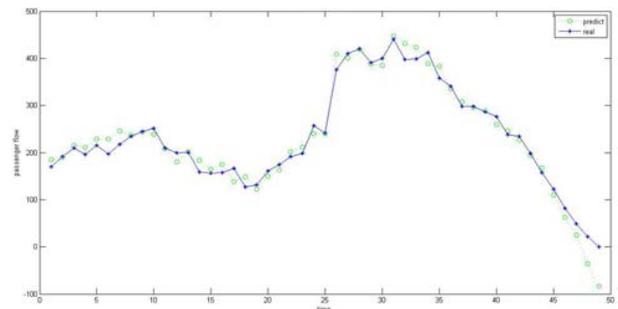


FIGURE V. THE RESIDUAL AUTOCORRELATION AND PARTIAL AUTOCORRELATION

C. Hybrid Model Test

According to the stability and randomness of passenger flow on bus stop, the data are classified. The two models are respectively used for prediction. The results are in line with the expectation. The prediction results of the two models are fitted and the errors are analyzed. As shown in Figure.5. It can be concluded that the prediction results of the hybrid model are good and the overall error of the model is within 10%, which meets the need of prediction accuracy. The performance is more outstanding, it can be applied to the urban bus stop short-term passenger flow prediction.

D. Algorithm Evaluation and Comparison

According to the above research, the passenger flow of IC bus and cash-on-boarding bus are respectively applied BP neural network model and time series model to get the passenger flow of every ten minutes on bus stop. In order to test the accuracy of the hybrid model prediction algorithm, in addition to the comparative analysis of the results of the prediction and the actual data, this paper uses Kalman filter as the comparison prediction algorithm. The prediction results with performance comparison are given figure.6.

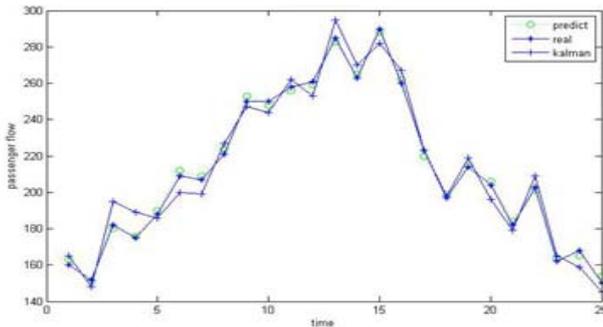


FIGURE VI. PASSENGER FLOW COMPARISON ON BUS STOP

The mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE) and mean square percentage error (MSP) are utilized to measure the accuracy of above model for short-term passenger flow prediction, which are defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \tag{7}$$

$$MSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \tag{8}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{9}$$

$$MSP = \sqrt{\frac{\sum_{i=1}^n \left(\frac{\hat{y}_i - y_i}{y_i} \right)^2}{n}} \tag{10}$$

where y_i and \hat{y}_i are respectively the actual and predicted short-term passenger flow.

Using two prediction algorithm get the forecast data of passenger flow on bus stop, compared with the actual value, the result is shown in Fig.6. According to the prediction performance index, the performance indexes of the two prediction algorithm are calculated as shown in Table I.

TABLE I. PERFORMACE FLOW COMPARISON ON BUS STOP

Algorithms	MAE	MSE	MAPE	MSP
Hybrid Model	4.2817	0.8852	0.4621	0.0765
Kalman filter	9.4762	1.6328	0.8520	0.2631

It can be seen from TABLE I that the prediction results of short-term passenger flow using Hybrid Model are obviously better than the results of Kalman filter prediction. The value of MAE of the Hybrid prediction method is 4.2817, the value of MSE is 0.8852, the prediction accuracy is higher, and the prediction error index of the Hybrid prediction is smaller than the Kalman filter prediction model. The Hybrid Model prediction is fast and the fitting value with the actual value is high.

V. SUMMARY AND OUTLOOK

Bus intelligent dispatching system needs real-time data to make decision, and bus stop short-term prediction data is an important part of it. This paper analysis and summarizes the time characteristics of passenger flow in urban bus station. According to the random difference between bus IC card and cash passenger flow, the hybrid prediction model is used to prediction Short-term passenger flow of stops and verified with the actual survey data of bus stop in Beijing. According to the bus IC card passenger flow is cyclical than the obvious cash passenger flow, BP neural network model is established to reflect the characteristics of the short-term passenger flow cycle stability, and the time series model of cash ticket passenger flow data is established and finally the two predict the results of fitting, with the overall model error less than 10%, which is in line with the needs of prediction accuracy.

ACKNOWLEDGMENT

This research is supported by Project of National Natural Science Foundation (Project No.:61503006) and Beijing Municipal Education Commission science and technology Program (Project No.: KM201510009002), the People’s Republic of China.

REFERENCES

- [1] Yang Zhaosheng. Urban Intelligent Public Transport System Theory and Methods [M].beijing: China Railway Press, 2002.
- [2] Zhang Chunhui, Songrui, SUN Yang. Kalman Filter-Based Short-Term Passenger Flow Forecasting on Bus Stop [J]. Journal of Transportation Systems Engineering and Information Technology, 2011, 11(4): 154-159.
- [3] Williams BM Hoella ,Modeling and vehicular traffic flow as a seasonal ARIMA process : theoretical basis and empirical result s[J]. Journal Transportation Engineering, 2003, 129(6):664-672.
- [4] Vlahogianni E , Karlaftis MG, Goliasjc.Optimized and meta-optimized neural networks for short-term traffic flow prediction: agenetic approach [J].Transportation RESEARCH Part C :Emerging Technologies, 2005, 13(3):211-234
- [5] Davis GA, Nihan NL , Nonparametric regression and short -term forecasting [J] . Journal of Transportation Engineering, 1991,117(2):178-188.
- [6] Yang Xinfeng Liu Lanfen. Short term Passenger Flow Forecasting on Bus Station Based on Affinity Propagation and Support Vector

- Machine. Journal of Wuhan University of Technology, 2016, 40(1): 36-40
- [7] Dai Shihua, Zhou Xinrong. Application of Kalman filtering theory on prediction of short-term traffic volume [J] Journal of Harbin University of Commerce (Natural Sciences Edition), 2005, 21(6): 728-730, 735.
- [8] Li Chunxiao, Li Haiying, Jiang Xi ,Xu Xinyue Zhao Aqun. Short-term entrance passenger flow prediction at urban rail transit station based on generalized dynamic fuzzy neural networks . Metro Rapid Transit, 2015, 8 28(4)
- [9] Tan Manchun, Feng Luobin, Xu Jianmin, Traffic Flow Prediction Based on Hybrid ARIMA and ANN Model [J]. China Journal of Highway and Transport 2007
- [10] Fan Na, Zhao Xiangmo, Dai Ming, An Yisheng. Short term traffic flow prediction model[J]. Journal of Traffic and Transportation Engineering 2012