

Research Article

Analysis of Bus Trip Characteristics and Demand Forecasting Based on NARX Neural Network Model

Feng Sun , Wenheng Su , Weixuan Liu, Hui Cao, Dong Guo , and Ye Zhu

College of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, Shandong 255049, China

Correspondence should be addressed to Dong Guo; ghyuyu@163.com

Received 31 May 2018; Revised 10 October 2018; Accepted 18 October 2018; Published 25 November 2018

Academic Editor: Qin Xin

Copyright © 2018 Feng Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, there has been increased interest in the use of bus IC card data to analyze bus transit time characteristics, and the prediction is no longer confined to rail traffic passenger flow prediction and traditional traffic flow prediction. Research on passenger flow forecast for the bus IC card has been increasing year by year. Based on the bus IC card data of Qingdao City, this paper first analyzes the characteristics of one-day passenger flow and passenger flow during subperiods and conducts a separate study on the characteristics of the elderly. The results show that the travel of the elderly is also affected by the weekday and the weekend. Then, based on the ARIMA model and the NARX neural network model, the passenger flow forecasting (10-minute interval) is carried out using the IC card data of No. 1 bus for 5 weekdays. The prediction results show that the NARX neural network model is effective in the short-term prediction of bus passenger flow, and especially, it is more accurate in the peak hour and large-scale data prediction.

1. Introduction

Automated fare collection (AFC) system, also known as the transit smart card (SC) system, has gained more and more popularity among transit agencies worldwide. Compared with the conventional manual fare collection system, an AFC system has its inherent advantages in low labor cost and high efficiency for fare collection and transaction data archival. Although it is possible to collect highly valuable data from transit SC transactions, substantial efforts and methodologies are needed for extracting such data because most AFC systems are not initially designed for data collection [1]. People are also paying more and more attention to how to use these data to dig deeper into the characteristics of bus passengers and provide support for urban public traffic management and planning. At present, research on bus IC card data can be broadly divided into three categories: (1) analysis of characteristics of bus passenger flow, (2) evaluation of bus operating efficiency, and (3) forecast of bus travel demand.

In terms of the analysis of the bus passenger flow characteristics, Morency et al. used smart card data from Ottawa, Canada, to study the travel characteristics of

different card types in time and space [2]. Páez et al. combined AFC data with personal passenger information and used cluster analysis to obtain passengers' spatial travel characteristics [3]. Briand et al. used a Gaussian mixture generation model to cluster the travel time characteristics of passengers and revealed the travel patterns of different groups [4]. Li and Deng adopted the method based on the travel chain to obtain the passengers' drop-off point and classify the bus passengers' travel according to the passengers' travel time and space characteristics [5]. He et al. took Guangzhou residents as the research object and proposed a cyclic analysis method based on mode division [6].

On the evaluation of bus operation efficiency, Ma developed a data-driven platform for online transit performance monitoring by using the smart card and GPS data to monitor transit service quality [7].

In the aspect of bus passenger flow forecasting, Yang et al. extracted data from the IC card database of Dalian public transport, used the Fisher clustering algorithm to classify the bus peak intervals to establish a regression equation for predicting passenger flow under different peak conditions, and realized the forecast of total passenger flow in different peak intervals, but the forecast period is longer [8]. Liu et al.

used the number of passengers getting off and on at Bus No. 8 in Harbin City to present a prediction model for the number of people entering and leaving the bus station based on the improved BP neural network, but the forecasting time interval was longer [9]. Tsai et al. used multiple temporal units neural network (MTUNN) and parallel ensemble neural network (PENN) to predict the short-term subway passenger flow based on sales data from Kaohsiung to Taipei of a specific train (Train No. 1008) over four years, which had higher prediction accuracy than the traditional multilayer perceptron model (MLP) [10]. Wei and Chen used a hybrid EMD-BPN forecasting approach by combining empirical mode decomposition (EMD) and backpropagation neural networks (BPN) in the short-term passenger flow forecasting of the metro systems. The prediction results showed that this method performs well and was stable in predicting short-term subway passenger flow [11]. Jiang et al. used dynamic regression neural network (Elman) and BP neural network methods to predict the bus passenger flow in Hefei City. Experiments showed that the dynamic regression neural network (Elman) method had higher prediction accuracy [12]. Yang et al. used the time-series model and the fuzzy neural network model to predict the short-term bus passenger flow. Because of the nonstationary bus passenger flow, it failed to achieve good prediction result [13]. Zhang et al. proposed the Kalman filter as a short-term passenger flow forecasting model for public transit stations and presented the solution process of the model [14].

In summary, the current analysis of bus trip characteristics is mainly based on the long-term forecast of one-day passenger flow and one-week passenger flow. However, there are few studies on the short-term prediction of passenger flow during subperiods, and in addition, the lack of exploration of the elderly travel characteristics has led to the neglect of this important group in passenger flow forecasting. In terms of passenger flow prediction methods, with the gradual maturity of the neural network and other deep learning forecasting technologies, it has become the most mainstream traffic prediction method at present.

Therefore, in this paper, the preprocessing and cluster analysis of bus IC card data are used to explore the bus trip characteristics of passenger flow during subperiods, and the bus travel characteristics of the elderly are mostly analyzed. Then, the NARX neural network model is used to forecast the bus passenger flow during subperiods, conducting to provide basic data for the real-time scheduling and management of bus operations.

The remainder of this paper is organized as follows. Section 2 introduces the data structure of the bus IC card in Qingdao and preprocesses the original data. In Section 3, it analyzes the characteristics of passenger flow in a single day and passenger flow during subperiods and mostly analyzes the bus trip characteristics of the elderly. Section 4 introduces the construction of the bus traffic forecast model, ARIMA model, and NARX neural network model and analyzes and evaluates the example of the bus IC card data of 5 weekdays in Qingdao City. In Section 5, we summarize the research results of this paper and look forward to future research.

2. Bus IC Card Data Preprocessing

2.1. Data Structure. The research in this paper is based on the IC card data of Qingdao bus. Passengers swipe the IC card when getting on the bus in most Chinese cities. It is also true in Qingdao. The IC card data records the line name, line number, POS number, card number, card-type number, card-type name, transaction date, transaction time, transaction amount, and other information. The data structure of the Qingdao bus IC card system is shown in Table 1.

2.2. Data Preprocessing. In order to guarantee the quality of data analysis, it is necessary to preprocess the bus IC card data, extract fields that have a significant impact on the data analysis, and filter some invalid data which will adversely affect the results of data analysis. This article will clean up and filter Qingdao bus IC card data based on Microsoft SQL Server 2017, to deal with the wrong data, redundant data, and other issues in the original data and improve data mining efficiency and quality. Combining the research content of this article and the original data of Qingdao bus IC card, this data preprocessing is mainly divided into data cleaning, data transformation, and data reduction.

2.2.1. Data Cleaning. Delete the duplicate data and error data records that exist in the bus IC card data. For the same card number of the bus IC card data, when the transaction date and the transaction time are the same, these data are duplicate data, and the deletion operation is performed on such data; when there is a null value in the card number, transaction time, transaction date, etc., these data are missing data which need be deleted; through the screening we found that there exist some mistakes in bus IC card data, such as "20000"; that is, swipe card time is 2 am, which is obviously not practical. Deletion processing is required for such erroneous data.

2.2.2. Data Transformation. In order to meet our data analysis needs, we need to transform the original data. Specifically, it is divided into two parts: the characteristics of the passenger flow and the forecast of bus passenger flow. For the feature analysis part of the passenger flow, we will group the data after cleaning into 1-hour intervals and at the same time extract the elderly bus IC card data and group the data into 1-hour intervals in Section 3. For the bus passenger flow forecasting, we will group the data from 5:00 to 23:00 on weekdays into 10-minute intervals to serve as the data basis for the demand forecast study.

2.2.3. Data Reduction. The original data obtained from the IC card of Qingdao bus includes the fields shown in Table 1. Some of the fields are useless for the analysis of the passenger flow characteristics and passenger flow forecasts, such as the unit number, unit name, and posttransaction card balance. These data take up storage space and reduce the speed of data filtering. For this reason, we remove these fields.

TABLE 1: Data structure of Qingdao bus IC card system.

Field	Type
Unit number	INT
Unit name	NVARCHAR (20)
Line number	NVARCHAR (12)
Line name	NVARCHAR (20)
POS number	NVARCHAR (12)
Card number	NVARCHAR (16)
Card-type number	INT
Card-type name	NVARCHAR (20)
Transaction date	INT
Transaction time	INT
Transaction amount	FLOAT
Posttransaction card balance	FLOAT

3. Analysis of Bus Passenger Flow Characteristics

This section will analyze the bus IC card data from September 4th to September 10th, 2017 (Monday to Sunday), including mainly analyzing the travel time characteristics and different types of bus IC card travel characteristics.

3.1. Travel Time Distribution Characteristics

3.1.1. One-Day Passenger Flow. The total number of the bus IC card on a single day from September 4th to September 10th is shown in Table 2.

Figure 1 shows that there is a significant difference in the number of passengers on the weekend and weekdays, and the amount of bus IC cards on the weekend is significantly lower than that on weekdays. From Monday to Thursday, the number of bus IC cards remains steady at 1.54 million times. The passenger flow on Friday increases by 1.6 million times over the previous four days. The passenger flow on Saturday and Sunday decreased continuously. There are two reasons for this:

- (1) The majority of the passengers are students and commute passengers to work. The travel of these two groups has obvious periodicity and is affected by the weekend.
- (2) On Saturdays and Sundays, the travelers will choose to go out for entertainment and shopping activities. However, due to the reduction in the number of students and the number of commute passengers, the overall number of times of swiping the card is on a downward trend. Therefore, the results obtained through statistical analysis of a large number of data are consistent with the actual results.

3.1.2. Time Interval Passenger Flow. Bus passenger flow has very obvious time-varying characteristics. There are obvious morning and evening peaks. Some small cities will also have afternoon peaks. Through the analysis of bus IC card data, we can grasp the time distribution characteristics of bus passenger flow, make reasonable arrangements of the frequency of shifts, and improve the operating efficiency of the entire bus system. This section will conduct a statistical

analysis of the data from 05:00 to 23:00 at one-hour intervals, based on the bus IC card data from September 4th to September 10th.

Figures 2 and 3 are time-varying diagrams of passenger flow and passenger flow box plot of the Qingdao bus IC card in one-hour intervals. The following conclusions can be obtained by comparing the time periods and the daily passenger flow:

- (1) The passenger flow of the bus IC card in Qingdao has obvious peaks in the morning and evening. The intensity of the evening peak is weaker than that of the morning peak, but the duration is longer.
- (2) The start of the weekend morning peaks was postponed, and the duration of the evening peaks was shortened. The total number of passengers during the weekend was significantly lower than weekdays.
- (3) Figure 3 intuitively shows that there are obvious differences between the travel characteristics of the weekdays and the weekend. The number of card swiping on September 9th and September 10th is less fluctuating, and the peak amount of card swiping is an obvious decline compared to the weekdays.

The above conclusions are consistent with the actual travel situation. Travel on the weekend is relatively decentralized, and on weekdays, it is relatively concentrated. The morning peak is late on the weekend, and the evening peak end earlier.

3.2. Analysis of the Travel Characteristics of Seniors. With the ageing of population becoming increasingly prominent, the issue of seniors' travel has received extensive attention. In terms of bus travel, the city of Qingdao started to implement a free travel policy for people aged 65 and older in 2014. In this context, the seniors' bus travel demand has increased significantly, and bus travel has gradually become the primary choice for seniors to travel. This section will study the travel patterns of the elderly, based on the data from the bus IC card in Qingdao, with the purpose of avoiding the disadvantages of inaccurate and incomplete data of traditional survey and statistical methods, providing decision support for urban public transport planning, and improving the quality of travel for the seniors.

This section will study the travel patterns of the seniors from the amount of the bus IC card and time distribution characteristics.

Figure 4 shows that the travel of the seniors has the following characteristics:

- (1) The number of trips over the weekend is less than that of the weekdays. According to the statistics of the seniors' card in one week, it can be found that the amount of swiping the bus IC card on Friday is the largest, followed by Wednesday, and the least on Sunday. The traditional experience is that the travel of the seniors is affected little by the weekdays and weekend. However, actual statistics show that the amount of seniors' bus travel is also affected by the weekdays and weekend and is obvious.

TABLE 2: Total number of the bus IC card on a single day from September 4th to September 10th.

Date	09-04	09-05	09-06	09-07	09-08	09-09	09-10
Passenger flow	1543408	1542399	1549074	1549589	1600723	1467969	1405396

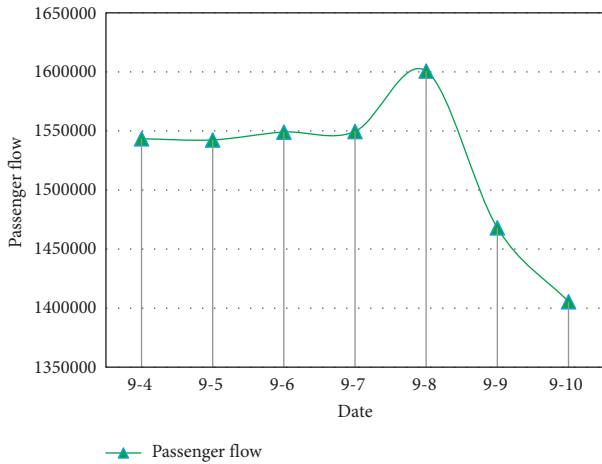


FIGURE 1: Daily change of bus IC card amount.

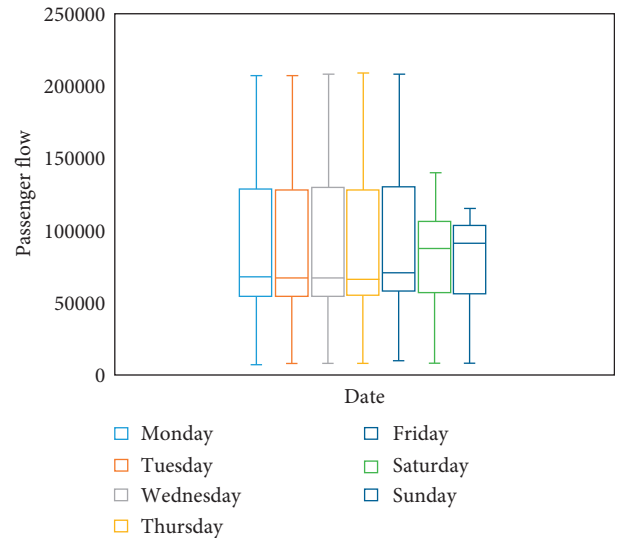


FIGURE 3: Passenger flow box plot.

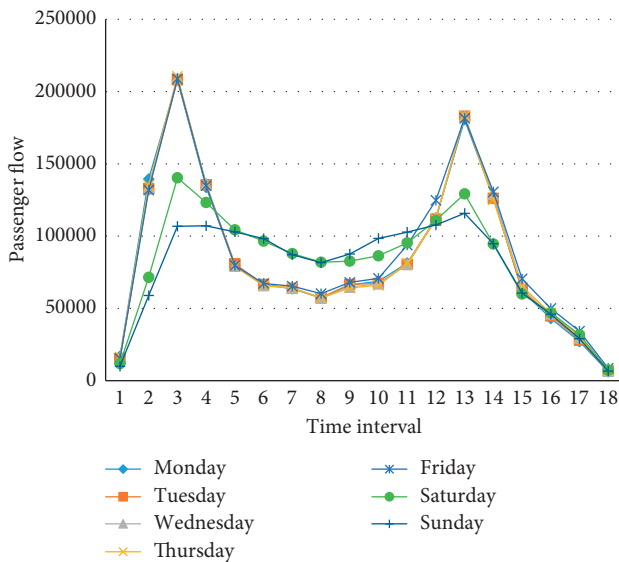


FIGURE 2: Time-varying diagram of passenger flow.

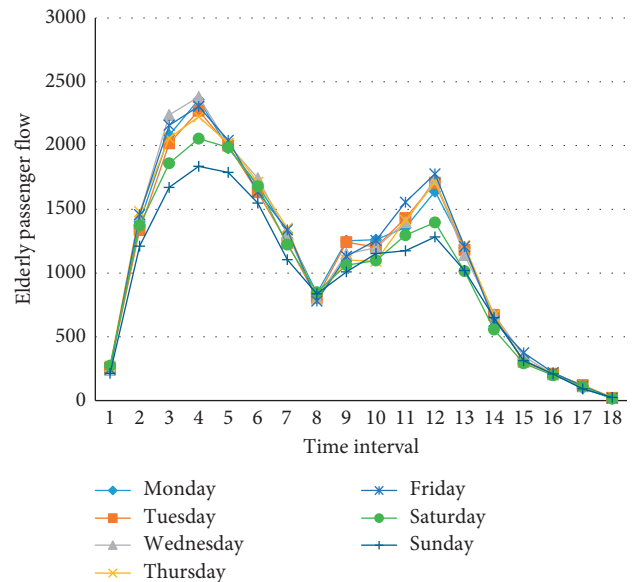


FIGURE 4: Time-varying diagram of senior passenger flow.

- (2) Travel time is more concentrated in the morning than other types of travelers. Because the seniors have more free time, and most seniors use their morning time to exercise and go out.
- (3) Seniors travel later in the morning peak and travel earlier in the evening peak. The previous analysis found that the morning peak and evening peak of other passengers except the seniors are 7:00 to 8:00 and 17:00 to 18:00, respectively. Through the analysis of seniors card data, it is found that the morning peak is between 8:00 and 9:00 and the evening peak is 16:00 to 17:00. This shows that the seniors will avoid the morning and evening peaks in the choice of

travel time. This period of time is conducive to the elderly travel, avoiding the crowded hours of work and commuting.

- (4) The change trend of seniors' travel volume is relatively flat, the gap between the peak and off-peak travel volume is small, and the slope of the passenger flow changes after the evening peak of Figures 2 and 4 shows that the number of seniors reduced faster than other types of passengers after the evening peak.

4. Bus Passenger Flow Forecasting Method

4.1. ARIMA Model. The ARIMA (Autoregressive Integrated Moving Average) model is a time-series model used for short-term forecasting. In general, the ARIMA model is given in the following equation:

$$\begin{aligned}\Phi(L)(1-L)^d R_t &= \Theta(L)e_t, \\ \Phi(L) &= 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p, \\ \Theta(L) &= 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q,\end{aligned}\quad (1)$$

where $\Phi(L)$ is the autoregressive coefficient polynomial for the stationary reversible ARMA (p, q) model, $\Theta(L)$ is the moving smoothing coefficient polynomial for the stationary reversible ARMA (p, q) model, R_t is time-series data, L is the backward shift operator, d is the differencing order, and e_t is white noise series.

The essence of the ARIMA model is the combination of the differential operation and ARMA model, with high short-term prediction accuracy.

4.1.1. Stationarity Test. The passenger flow trend of a workday is shown in Figure 5. It can be seen that passenger flow data are not stable.

Figure 6 is an autocorrelation diagram of the original data, from which it can be seen that the data have the characteristics of a typical nonstationary sequence, which is smoothed by the differential method for further prediction.

4.1.2. Differential Processing. Figures 7–9 show that the sequence has a steady trend. As can be seen from the autocorrelation diagram and the partial autocorrelation diagram, the time-series presents a trailing situation. The ADF unit root test is performed on the time series, and the p value is 0.0069, which is less than the significant level value 0.05, so the sequence after the first-order difference is a stationary time series.

4.1.3. Model Establishment. According to the AIC criterion, the ARIMA model is fixed, and the AIC value is calculated. When the AIC value reaches the minimum, the ARIMA (p, d, q) model has a p value of 4 and a q value of 5. The sequence has been subjected to first-order difference. The operation is therefore using the ARIMA (4, 1, 5) model.

4.1.4. Model Prediction. After the model is established, we use it to predict the passenger flow in the working day, and the forecast results are shown in Figure 10. It can be seen that there is a large fluctuation in some time periods, and the actual value has a large error with the predicted value, so the prediction effect of the ARIMA model on the nonlinear data needs to be improved.

4.2. NARX Neural Network Model. NARX neural network (Nonlinear Autoregressive with External (Exogenous) Input) is a kind of the dynamic neural network. Its output value can be input into the model again through feedback as an input value, so that the sensitivity of the model to

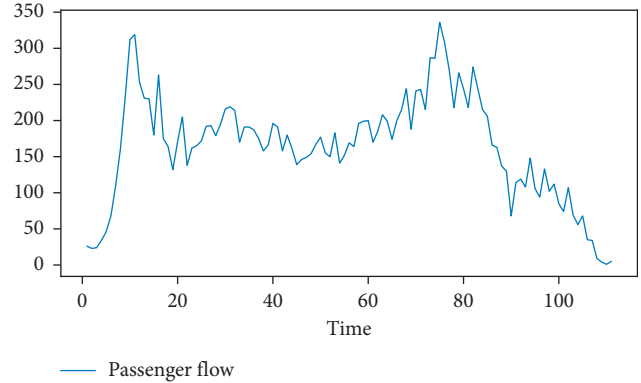


FIGURE 5: Passenger flow in different periods of a workday.

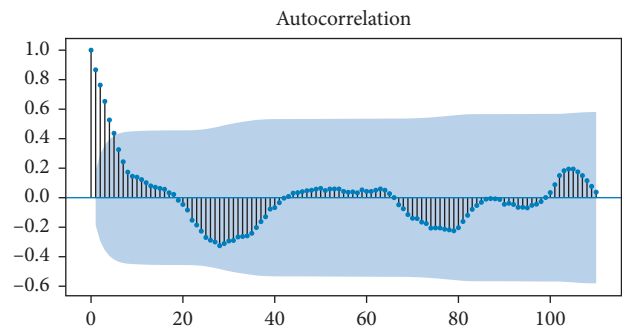


FIGURE 6: Autocorrelation diagram of the original data.

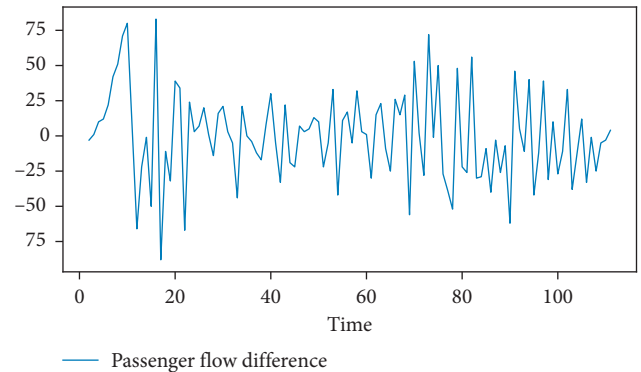


FIGURE 7: Passenger flow after the first-order differential processing.

historical data is improved and can better reflect dynamic features of the passenger flow and improve the prediction accuracy of neural networks. The NARX neural network structure is shown in Figure 11.

$$y(t) = f(x(t-1), \dots, x(t-d), y(t-1), \dots, y(t-d)), \quad (2)$$

where $x(t)$ represents the input of the neural network, $y(t)$ represents the output of the neural network, f represents the nonlinear ambiguity function, and d represents the feedback delay.

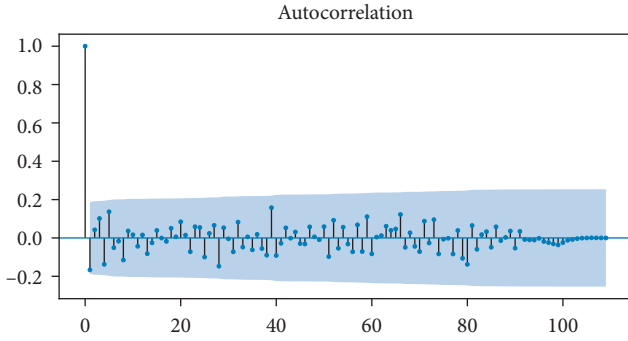


FIGURE 8: Autocorrelation diagram after the first-order differential processing.

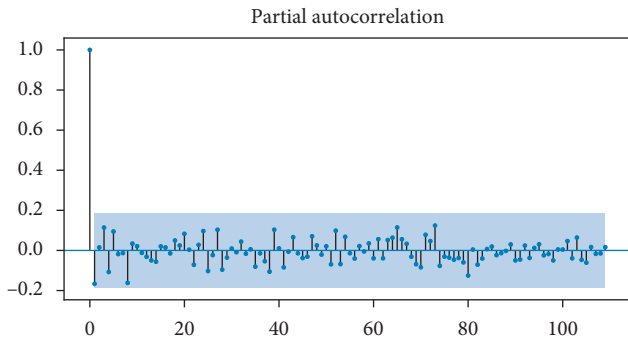


FIGURE 9: Partial autocorrelation diagram after the first-order differential processing.

4.2.1. Determination of NARX Neural Network Parameters

- (1) Determination of the number of input layer nodes: when using the neural network model for prediction research, we first need to determine the number of nodes in the input layer. In general, the number of nodes in the input layer is determined by the input of the dynamic system equations. If the dynamic system equations are not clear, step-by-step testing can be used to determine the number of input layer nodes in the network. We select the times of the swiping card as the input value and then adjust it one by one to determine the number of input layer nodes to be 4.
- (2) Determination of the number of hidden layer neurons and hidden layers: the number of neurons in the hidden layer is extremely important for the performance of the network. When the number of neurons in the hidden layer is small, the characteristics of the data cannot be well simulated. Too many neurons will increase the training time of the network, and excessive training will occur. The function of the hidden layer is to extract features from the input data. The appropriate number of hidden layers can make the neural network have better data processing capabilities. Excessive number of hidden layers will increase the training error and lengthen the network training time. We collected

data from the bus IC card for 5 weekdays and counted the times of the swiping card at intervals of 10 minutes. According to the data size of this study, we determined that the network with one hidden layer was used for training.

The final network structure is the number of input layer nodes is 4, the number of hidden layer neurons is 22, the number of hidden layers is 1, and the delay number is 2. The network structure is shown in Figure 12.

4.2.2. Training Method Selection. The number of research samples, the number of hidden layers, and the number of hidden layer nodes all play an important role in the convergence of the training algorithm. Based on the data size of this study, we compare the effects of two different training algorithms and determine the LM (Levenberg–Marquardt) training algorithm as final results.

(1) SCG (Scaled Conjugate Gradient) Algorithm. The SCG algorithm is an improved conjugate gradient algorithm. It does not need to calculate the Hessian matrix and changes the linear search mode of the conjugate gradient algorithm when calculating the search step length. It does not need a linear search to determine the optimal search path, and its convergence speed is faster. The conjugate gradient method improves the traditional gradient descent method, which can improve the network oscillation and improve the convergence speed of the network. Based on the data from this study, we used the SCG algorithm to train the network. The training effect is shown in Figure 13.

As shown in Figure 13, the errors of training, validation, and testing tend to be stable after the 45th generation, which reflects that the network convergence speed is fast and the network performance is good.

(2) LM (Levenberg–Marquardt) Algorithm. The LM algorithm does not need to calculate the Hessian matrix during the correction rate. When the error performance function has the form of the sum of squared errors, the Hessian matrix can be approximated as

$$H = J^T J, \quad (3)$$

where H is the Jacobian matrix containing the first derivative of the network error function to weights and thresholds.

The LM training algorithm is modified as follows:

$$\omega(k+1) = \omega(k) - [J^T J + \mu I]^{-1} J^T e, \quad (4)$$

where μ is the adjustment coefficient and e is the error vector. When μ is close to or equal to 0, the LM algorithm is the Newton method. When the value of the coefficient μ is large, the LM algorithm becomes the gradient descent method with smaller steps. Since the Jacobian matrix is easier to compute than the Hessian matrix, the training speed of the LM algorithm is very fast. The LM algorithm training effect is shown in Figure 14.

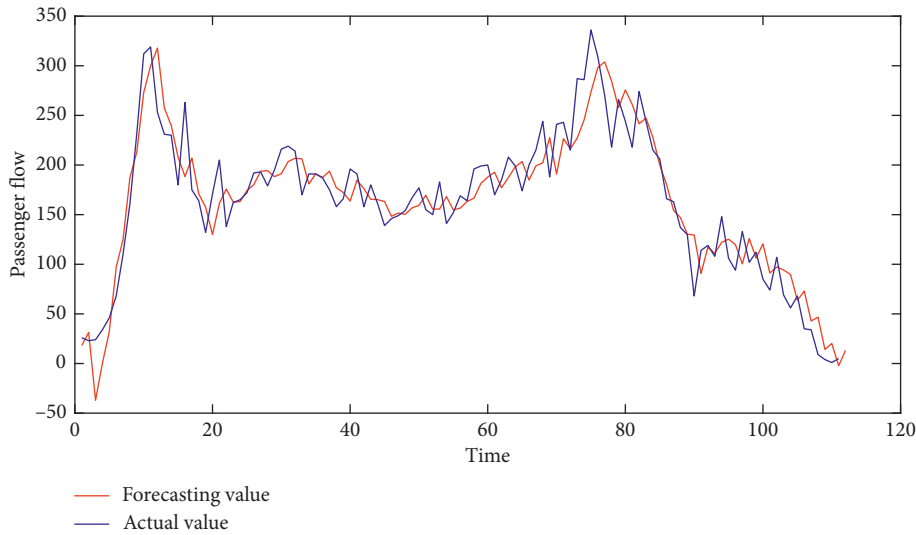


FIGURE 10: ARIMA model prediction results.

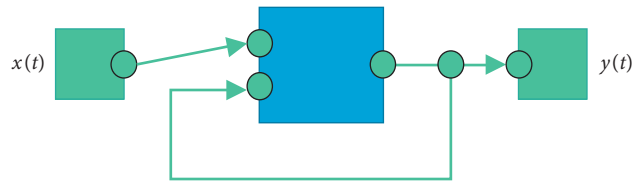


FIGURE 11: NARX neural network.

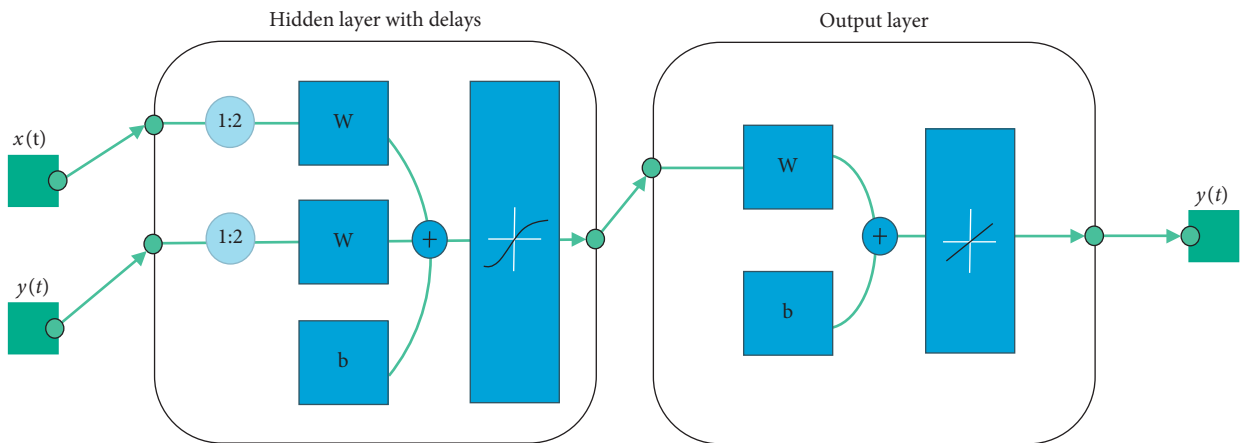


FIGURE 12: Network structure.

The error of the LM algorithm validation and test tends to be stable after the 4th generation, and the training speed is greatly improved compared with the SCG algorithm. In order to more reasonably judge the performance of the training algorithm, we compared the two algorithms by the mean squared error (MSE).

The magnitude of the mean square error reflects the discrete distribution of the error. When the mean square error is large, it indicates that the dispersion of the error distribution is high and the prediction effect is poor. The formula for calculating the mean square error is as follows:

$$MSE = \frac{1}{n} \sum_{t=1}^n (y(t) - y'(t))^2, \quad (5)$$

where n represents the number of time intervals, $y(t)$ represents the actual value, and $y'(t)$ represents the predicted value.

By comparison, we find that the mean square error of the LM algorithm is 564.32, and the mean square error of the SCG algorithm is 788.66. Therefore, the LM algorithm is finally used to train the neural network and predict with the actual data.

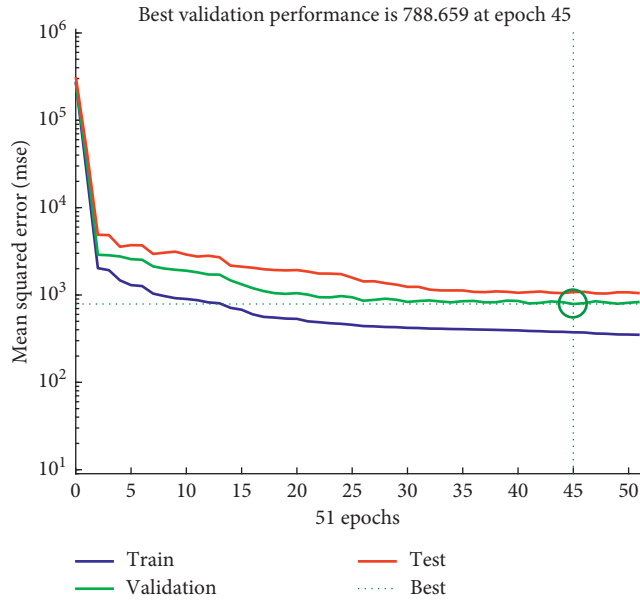


FIGURE 13: SCG algorithm training neural network characteristics.

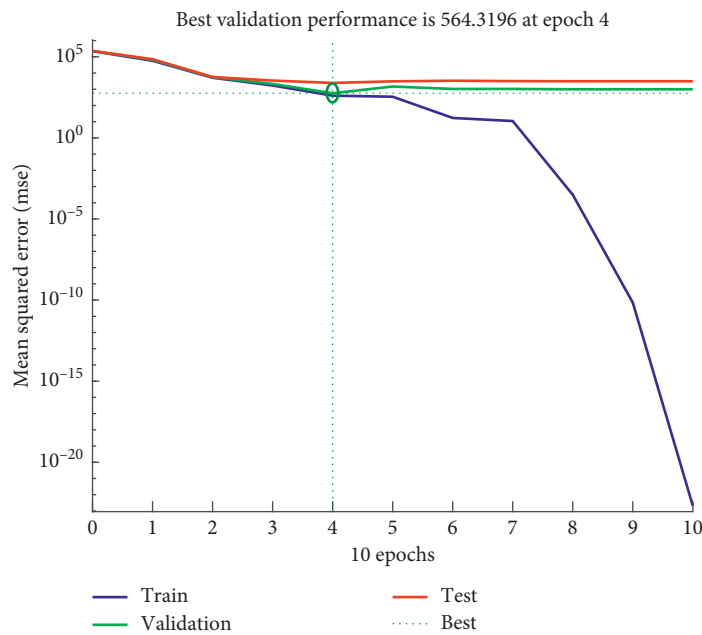


FIGURE 14: LM algorithm training neural network characteristics.

4.2.3. *Network Structure Design and Model Training.* This section uses the bus IC card data from September 4th to September 8th of the No. 1 bus in Qingdao as the database for the prediction of the NARX neural network. We have selected data from 5:00 to 23:00 and divide the data into 10-minute intervals.

The input layer node of the NARX neural network is set to be 4 because we use the September 4th to September 7th bus IC card data as input data, it is a matrix with 4 columns, and the number of hidden layers is 1. In the process of neural network training, the training effect is judged by observing the error autocorrelation function and the input-output

correlation function and R value after each training, the number of hidden layer neurons in the network is adjusted, and the number of hidden layer neurons is determined 22 and the delay number is 2 by using the LM (Levenberg–Marquardt) algorithm that trains the network. 70% of all sample data are selected as training data, 15% as validation data, and finally 15% as test data, until network training is effective.

After each training, we need to judge the training effect based on the error autocorrelation function and the input-output correlation function. The error autocorrelation function reflects the correlation between each data. In the

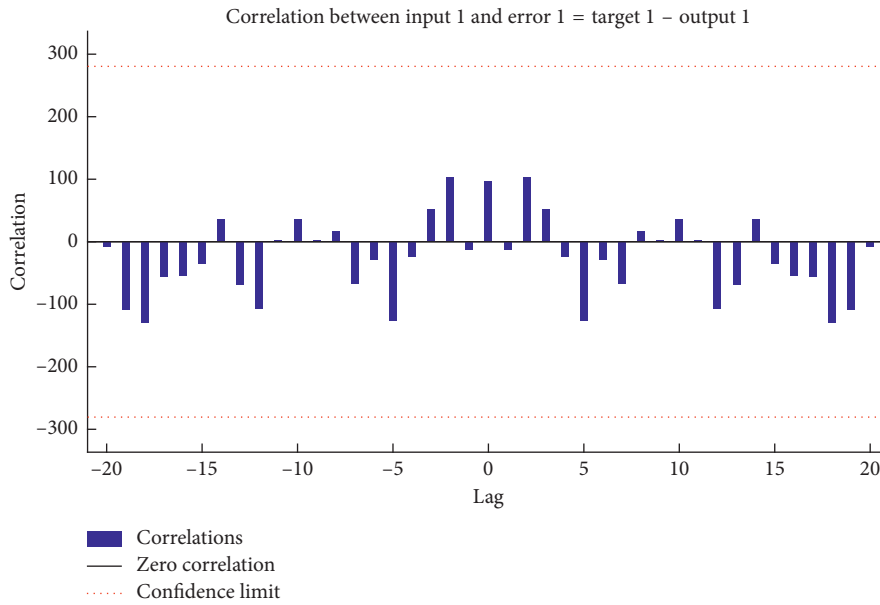


FIGURE 15: Input-output error correlation function value.

ideal state, the model has zero delay only when it is in the zero state; that is, it is completely irrelevant. In general, the error autocorrelation function falls within the confidence interval, indicating that the network training effect is good and can be predicted. The input-output correlation function in Figure 15 and the error autocorrelation function in Figure 16 show error within the confidence interval (in the red line range), indicating that the training results are good and can be predicted.

4.2.4. Prediction Process and Accuracy Evaluation. After the NARX neural network model was trained, the passenger flow on a single weekday (in 10-minute intervals) was predicted. We used Matlab to output the prediction effect error (Figure 17) and the fitting effect (Figure 18) and generated a comparison chart of the actual value and the predicted value, and the relative error between the actual value and the predicted value was calculated.

Through the distribution of the error line (solid yellow line) in Figure 17, we can see that there is a large error in only a few time periods, indicating that the model has a good effect on passenger flow forecasting. Figure 18 shows that the training data have an R value of 0.99803, the validation data have an R value of 0.91471, the test data have an R value of 0.90578, and the overall data have an R value of 0.9691. This also shows that the NARX neural network has a good effect on the time-division forecasting of bus passenger flow. Figure 19 more intuitively shows that the model's performance in bus passenger flow forecasting meets the prediction requirements.

4.2.5. Error Analysis. After the forecasting process is over, the relative error E_r between the predicted value and the actual value is used to evaluate the forecasting effect. Its expression is as follows:

$$E_r = \left| \frac{y(t) - y'(t)}{y(t)} \right|, \quad (6)$$

where $y(t)$ represents the actual value and $y'(t)$ represents the predicted value.

The error distribution curve is shown in Figure 20. We can see that the error during the peak hour is below 0.2, and there is a large error during the off-peak hours. The maximum relative error reaches 0.35. The reason for the analysis is that due to the large and concentrated passenger flow during peak hours, less passenger flow during off-peak hours, and relatively dispersed passenger flow, the NARX neural network model performs well in the extraction of data features during peak hours, achieving high accuracy. Due to the poor regularity of passenger flow during the off-peak hours, the forecasting model's ability to extract data features is lower than that during peak hours, and the prediction accuracy during off-peak hours remains to be improved.

4.2.6. Nonparametric Tests. We use the Wilcoxon signed-rank test to test the difference between the actual and prediction values. Table 3 shows that there is no significant difference between the actual value and the prediction value. The probability is 0.941 (the significance value in the table). Compared with the significance level of 0.05, 0.034 is large enough to be a high probability event; that is, the probability that there is no significant difference between the true value and prediction value is 0.941, preserving the null hypothesis. There is no significant difference between the real value and the predicted value, which is statistically significant.

Figure 21 shows the more detailed statistical results of this analysis, and the conclusion is unchanged.

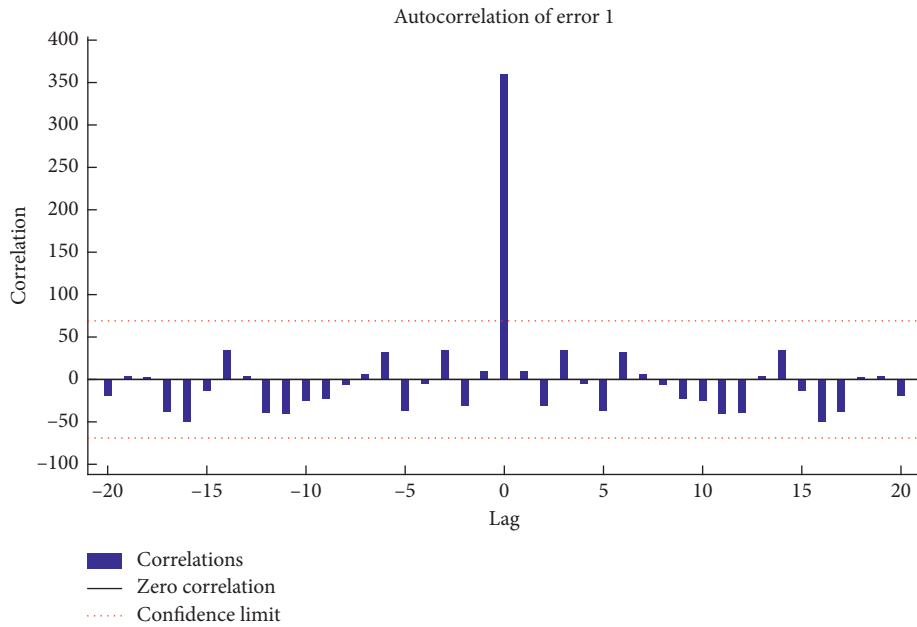


FIGURE 16: Error autocorrelation function value.

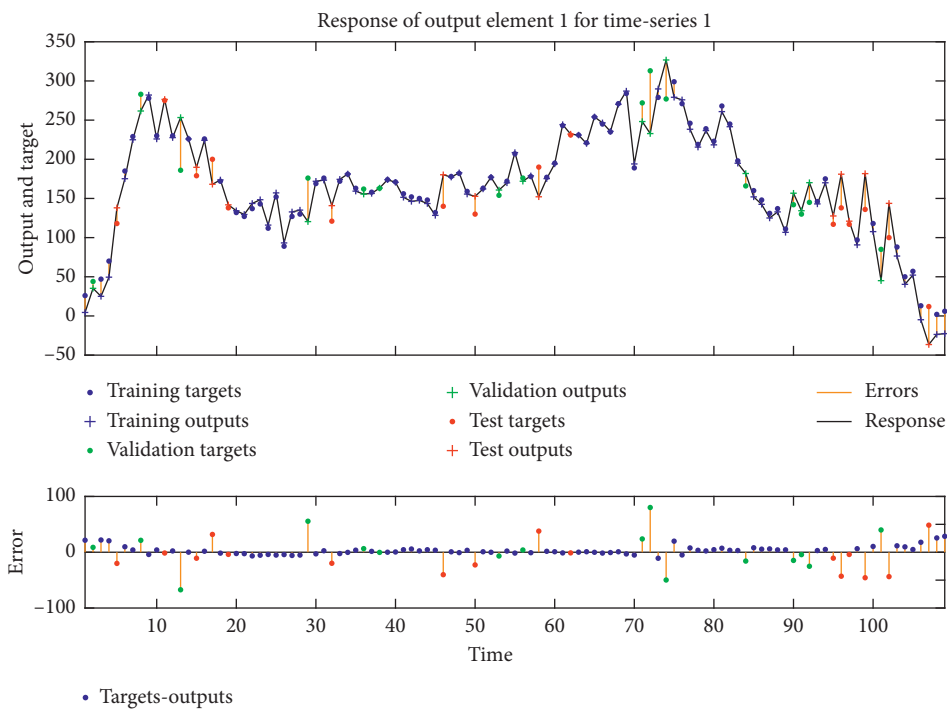


FIGURE 17: Prediction effect error diagram.

5. Conclusions

Based on the IC card data of Qingdao bus, this paper first studies the distribution characteristics and regularity of bus passenger flow in time, analyzes it from a single day passenger flow and subperiod passenger flow, and compares the different weekdays and weekend features. In addition, seniors are individually analyzed and their characteristics in bus travel were obtained. In the aspect of bus passenger flow

forecasting, the ARIMA model and NARX neural network prediction model were used according to the characteristics of the IC card data and the forecasting demand. The prediction results were compared with the actual values and found that the NARX neural network prediction model had good prediction accuracy and achieved expected results. In the future research, we should also study the characteristics of time-division trips for multiple groups. At the same time, we noticed that Qingdao implemented different fare policies

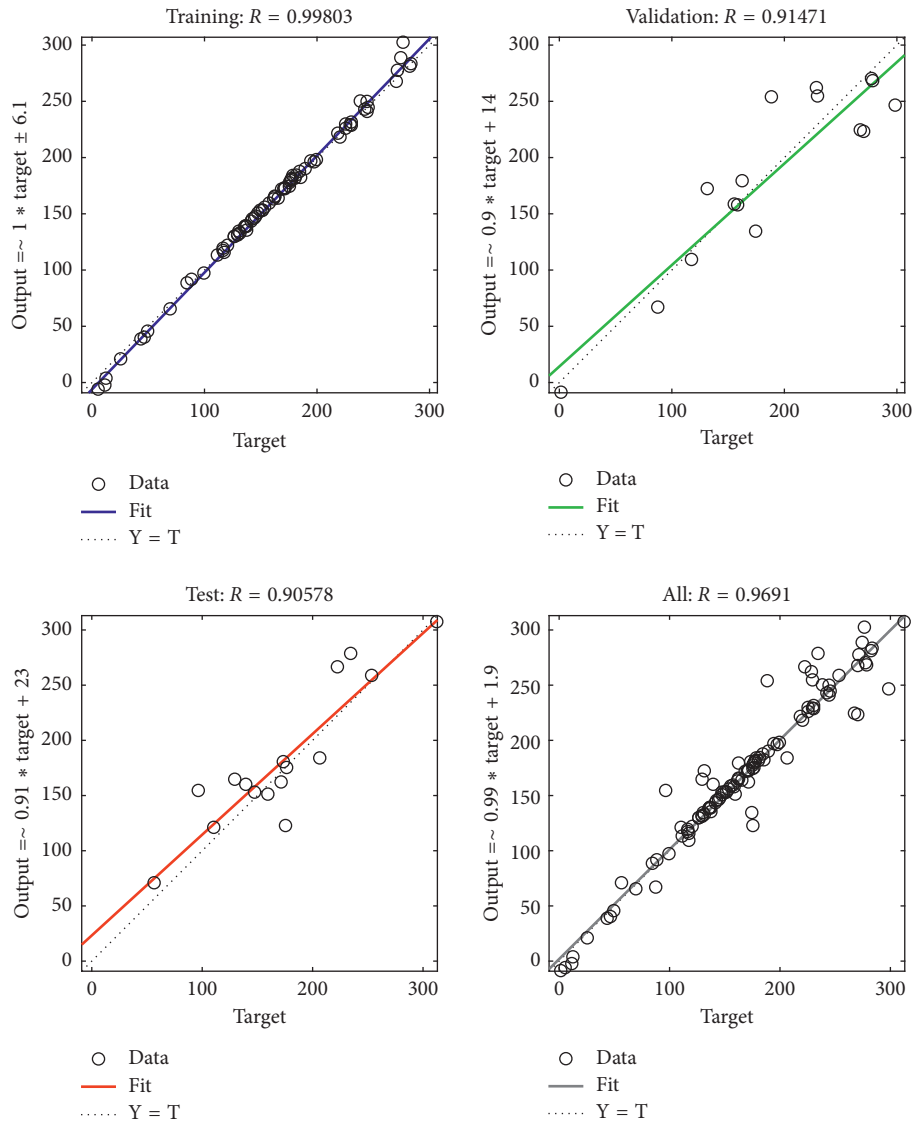


FIGURE 18: NARX neural network prediction goodness of fit R.

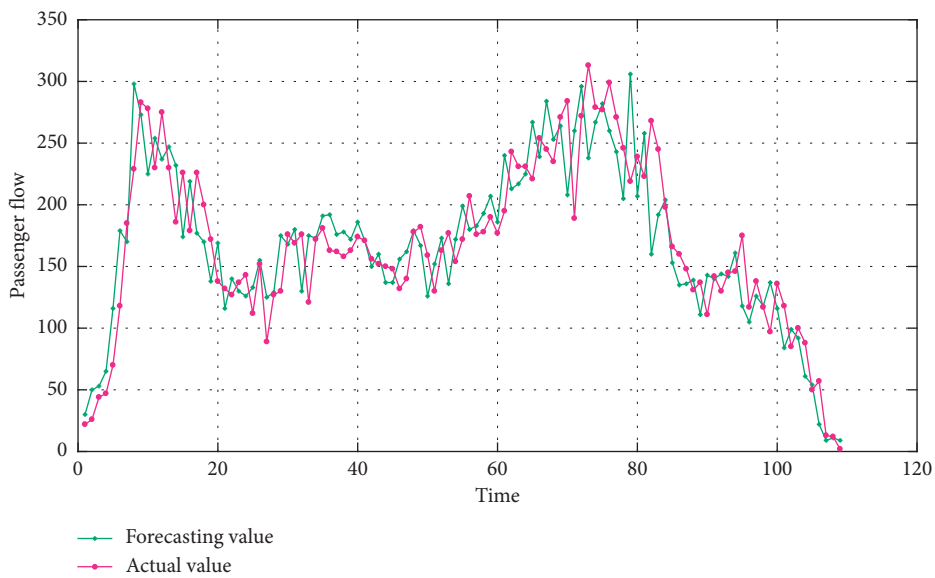


FIGURE 19: Comparison of predicted values of real values.

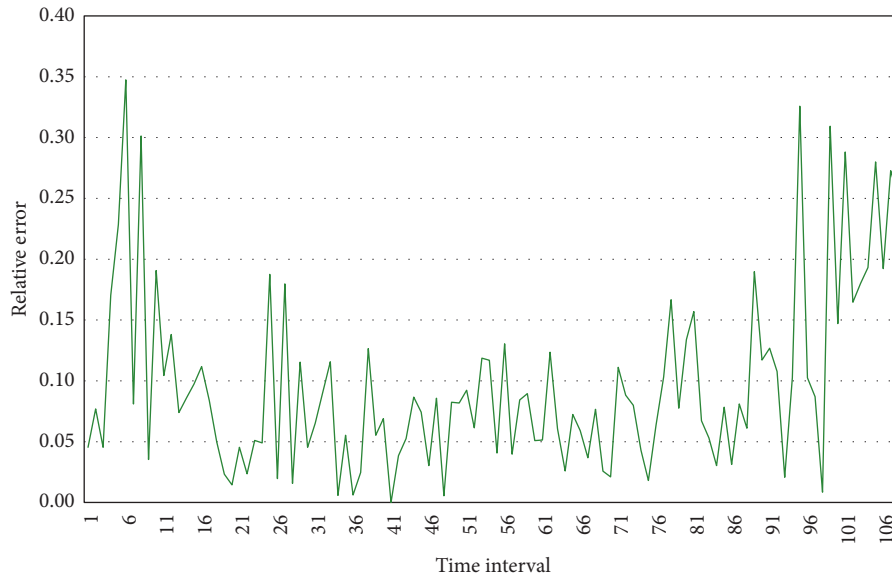
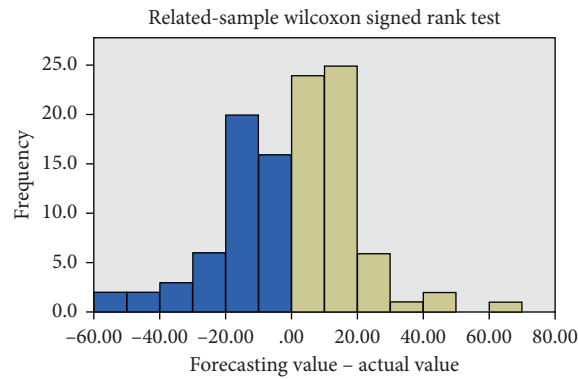


FIGURE 20: Error distribution curve.

TABLE 3: Hypothesis test summary.

Hypothesis test summary			
Null hypothesis	Test	Sig.	Decision
The median of differences between actual value and forecasting value equals 0. Asymptotic significances are displayed. The significance level is 0.05.	Related-samples Wilcoxon signed-rank test	0.941	Retain the null hypothesis.



■ Positive differences (N = 58)
■ Negative differences (N = 49) (number of ties = 1)

Total N	108
Test static	2,865.000
Standard error	321.683
Standardized test statistic	-.075
Asymptotic sig. (2-sided test)	.941

FIGURE 21: Detailed statistical results.

for elderly people of different ages. Its impact on the characteristics of seniors' trips is also a research direction. In the aspect of bus passenger flow forecasting, the short-term prediction of passenger flow for the bus IC card on multiple lines and multiple areas is a focus of attention in the future.

Data Availability

The bus IC card data used to support the findings of this study have not been made available because these data relate to personal privacy.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (Grant nos. 61573009 and 71501137) and the Science and Technology Planning Project of the Shandong Province (Grant no. 2016GGB01539).

References

- [1] X.-L. Ma, Y.-H. Wang, F. Chen, and J.-F. Liu, "Transit smart card data mining for passenger origin information extraction," *Frontiers of Information Technology and Electronic Engineering*, vol. 13, no. 10, pp. 750–760, 2012.
- [2] C. Morency, M. Trépanier, and B. Agard, "Measuring transit use variability with smart-card data," *Transport Policy*, vol. 14, no. 3, pp. 193–203, 2007.
- [3] A. Pérez, M. Trépanier, and C. Morency, "Geodemographic analysis and the identification of potential business partnerships enabled by transit smart cards," *Transportation Research Part A Policy and Practice*, vol. 45, no. 7, pp. 640–652, 2011.
- [4] A. S. Briand, E. Côme, M. Trépanier, and L. Oukhellou, "Analyzing year-to-year changes in public transport passenger behaviour using smart card data," *Transportation Research Part C Emerging Technologies*, vol. 79, pp. 274–289, 2017.
- [5] J. Li and H.-P. Deng, "Classification of passengers' travel behavior based on IC card data," *Journal of Chongqing Jiaotong University(Natural Science)*, vol. 35, no. 6, pp. 109–114, 2016.
- [6] Z.-C. He, C. Yu, and M.-X. Xu, "Analyzing methods of residents' travel characteristics considering travel patterns and periodicity," *Journal of Transportation Systems Engineering and Information Technology*, vol. 16, no. 6, pp. 135–141, 2016.
- [7] X. Ma, "Development of a data-driven platform for transit performance measures using smart card data and GPS data," *Journal of Transportation Engineering*, vol. 140, no. 12, 2014.
- [8] Z.-W. Yang, Q. Zhao, S.-C. Zhao, L. Jin, and Y. Mao, "Passenger flow volume forecasting method based on public transit intelligent card(IC) survey data," *Transport Research*, no. 9, pp. 115–119, 2009.
- [9] C. Liu, Y.-Q. Zhang, and H.-R. Chen, "Transit station's temporal getting on/off flow forecasting model based on BP neural network," *Transport Research*, no. 5, pp. 186–189, 2008.
- [10] T. H. Tsai, C. K. Lee, and C. H. Wei, "Neural network based temporal feature models for short-term railway passenger demand forecasting," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3728–3736, 2009.
- [11] Y. Wei and M. C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C Emerging Technologies*, vol. 21, no. 1, pp. 148–162, 2012.
- [12] P. Jiang, Q. Shi, W.-W. Chen, and W.-H. Zhang, "Prediction of passenger volume based on Elman type recurrent neural network," *Journal of Hefei University of Technology (Natural Science)*, vol. 31, no. 3, pp. 340–342, 2008.
- [13] X.-M. Yang, W. Wang, W.-P. Gu, and M.-B. Zhou, "Applying fuzzy neural network to predict bus line passenger flow," *Journal of Highway and Transportation Research and Development*, vol. 17, no. 4, pp. 38–40, 2000.
- [14] C.-H. Zhang, R. Song, and Y. Sun, "Kalman filter-based short-term passenger flow forecasting on bus stop," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 4, pp. 154–159, 2011.



Hindawi

Submit your manuscripts at
www.hindawi.com

