3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019)

Research on Bus Passenger Traffic Forecasting Model based on GPS and IC Card Data

Jie Deng, Huawei Nie * and Chaojun Chen

Guizhou Communication Vocational College, Guizhou Guiyang 550000, China.

*513124857@qq.com

Abstract. In order to accurately predict the bus passenger traffic volume and optimize bus dispatching, the paper combines the prediction model of RBF neural network to propose the deep integration of GPS and IC card data to realize the forecast of bus passenger traffic. Based on the GPS data and IC card data of Chongqing bus, the paper introduces the characteristics of GPS data and IC card data, and uses GPS data and IC data fusion analysis to obtain passenger OD distribution, passenger time segment distribution and passenger flow statistics. Then, using the RBF neural network model, the obtained three sets of data are predicted. Finally, the site of the 462 line in the main business district of Chongqing is used as the data source for verification. The results show that the RBF neural network prediction model based on GPS and IC card data sources can accurately predict the bus traffic and meet the requirements of optimizing bus scheduling.

Keywords: GPS data; IC card data; RBF neural network; Bus passenger flow; Prediction.

1. Introduction

Faced with the serious imbalance between traffic demand and transportation supply in the process of urban development, if we rely solely on vigorously developing infrastructure to improve transportation, it will not be able to meet the needs of current social development. How to improve the attractiveness of public transportation, rationally deploy public transportation facilities, and meet the needs of urban transportation needs are urgently needed to solve problems. Therefore, continuously improving the bus travel environment and optimizing the bus line network structure is the key to solving the urban traffic problem.

With the continuous development and utilization of Internet of Things technology, information processing technology, positioning technology and communication technology in the field of public transportation, the use of IC card ride has become the main payment method for residents' travel, and the GPS positioning system has also been widely installed on public transport vehicles. It is possible to obtain IC card data and GPS data. How to use these data for deep mining, obtaining residents' travel needs, optimizing urban public transportation, and providing decision-making information has become a key research area of urban public transportation. Many scholars at home and abroad have studied the integration of IC card and GPS data: Wang Wei et al. proposed a method of "routing one by one, optimizing the network", aiming at the maximum direct traffic, focusing on the optimization of a single bus line. Very strong practical guidance. Xu Jianmin of South China University of Technology used GPS and bus IC card data to generate passenger OD, estimated the principle of passengers boarding the station, matched the GPS location information with the transaction record information of the IC card data, and included the location, time and other information included in the GPS. It is linked to information such as transaction time in the IC card data.

However, there is still an immature place to study the characteristics of passenger flow and the operation information of the network by integrating the bus IC card data with the GPS data. The study of bus passenger traffic forecasting model based on GPS and IC card data in this paper can be used as a basis for bus operators to optimize scheduling decisions with practical significance and application value.

2. Data Characteristics

When the bus vehicle is equipped with an IC card ticketing system and a GPS system, the information collection subsystem can obtain the IC card information and positioning data of the



passenger using the vehicle, including the ID number, the card swiping time, the vehicle number, the line, and the site name (or Number) and entry and exit time, etc. The most important data in the intelligent public transport system includes vehicle location, speed, inter-site running time and passenger flow. The vehicle time and speed can be directly obtained by the vehicle GPS positioning system, and the running time between the stations will be calculated by more GPS data, bus passenger flow and vehicle historical driving record.

Timely and reliable traffic data is the basis for scheduling optimization. The main methods for intelligent bus system to obtain passenger traffic data are APCS technology and IC card information technology. Using these technical means, automatic statistics can be realized on the passenger flow rate of the bus, the number of passengers on the bus and the corresponding time period, so that the function relationship between the passenger flow and the time change of a certain section on the bus line can be derived.

How to realize the future dispatch of the bus requires the information management system to make previous predictions on the future passenger traffic and road network conditions. Through the established real-time forecasting model of passenger traffic, the intelligent bus dispatching subsystem predicts the real-time conditions of the passenger traffic on the future bus lines, and further optimizes the line vehicles according to the analysis of the passenger traffic forecast results.

3. Structure and Characteristics of RBF Neural Network Model

With the continuous emulation of RBF neural network technology and its advantages, it has become an effective method to make full use of RBF neural network to infinitely approximate any nonlinear mapping relationship to realize traffic research. The RBF neural network has a system-based algorithm structure. It does not need to use mathematical formulas to derive and use the input and output data to learn useful knowledge. It can be suitable for the study of random and multi-constraint problems. RBF neural network is a kind of feedforward network structure, suitable for approximation of multivariate functions, and has the only optimal approximation point. Its unique fault tolerance and generalization ability can predict well.

The general function can be expressed as a linear combination of a set of basic functions. The RBF neural network is equivalent to constructing a set of basic functions using the output of the hidden layer unit, and then using the output layer to perform linear combination for the purpose of fitting [1-3].

Assuming a given sample data $P = \{p_1, p_2 \cdots p_i \cdots p_Q\}$, $T = \{t_1, t_2 \cdots t_i \cdots t_Q\}$, the corresponding fit function is found using the RBF neural network to satisfy $t_i = F(p_i)$ $1 \le i \le Q$. Its RBF neural network topology is shown in Figure 1.

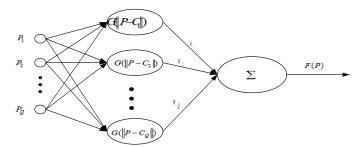


Fig.1 RBF neural network topology diagram

The fitting process is as follows:

- 1 Setting the hidden layer of the RBF neural network has Q hidden nodes;
- 2 Enter all Q samples as the center of Q hidden nodes;
- 3 each radial basis function takes the same expansion constant;
- 4 The weight can be determined by solving the linear equations. The equations are:



$$\sum_{i=1}^{Q} w_{j} G(\|p_{i} - p_{j}\|) = t_{i} (0 \le i \le Q)$$
(1)

5 Because, the output expression of the j hidden node in the i sample is:

$$\varphi_{ij} = G(\|p_i - p_j\|) \tag{2}$$

The matrix can be expressed as: $\Phi W = T$. If R is reversible, there is a formula:

$$W = \Phi^{-1}T \tag{3}$$

According to Micchelli's theorem, if the implicit node activation function uses a radial basis function and p_1, p_2, \dots, p_Q are different, the linear equations have a unique solution. That is, the RBF neural network output is expressed as:

$$F(p_i) = \sum_{j=1}^{Q} w_j \Phi(\|p_i - p_j\|)$$
 (4)

The fitting process of RBF neural network only uses input parameters and constraints as reference factors, thus reducing noise interference and predicting results more accurately.

4. Model Framework

The heterogeneous information system integration technology realizes the integration of IC ticketing system and GPS positioning system. The integrated system information processing center can easily complete the integration of bus IC card information and GPS data to form the traffic, time and bus properties [4-5]. Data sets for complete information such as lines, sites, and sites. The predictive model is to train the integrated data set through the RBF neural network, and use the RBF neural network to have the ability of autonomous learning to establish a predictive model. Using existing data, we can predict future passenger traffic and improve reliable data support for bus scheduling. The bus passenger traffic forecasting model framework is shown in Figure 2.

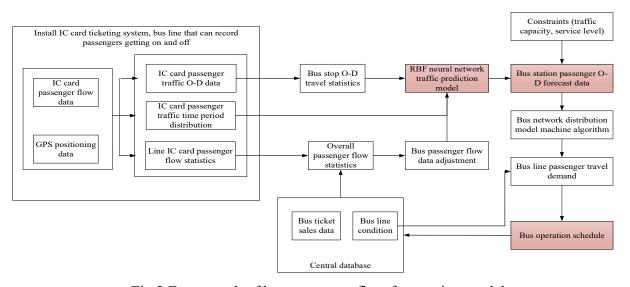


Fig.2 Framework of bus passenger flow forecasting model



The main module introduction:

- 1 IC card ticketing system and GPS positioning system combined with modules. This module mainly collects bus traffic data and bus location data, and uses the information processing technology introduced above to match the same parameters of the two data, such as bus lines, vehicle numbers, stations, and time of entry and exit, to generate IC card information and Data set of GPS data. Through the OD analysis method, the IC card passenger traffic OD data is obtained; the IC card passenger traffic time period distribution is obtained by dividing the time period; the line IC card passenger traffic flow statistics are obtained through the line statistics, thereby generating useful information for different analysis fields [6].
- 2 Line overall passenger flow statistics. The module integrates IC data and manual data to obtain actual passenger traffic on the line.
- 3 RBF neural network traffic prediction model. The module predicts the OD of a passenger in a future bus station by using historical bus station OD travel statistics, IC card passenger traffic time distribution data and IC card line passenger traffic statistics.
- 4 Bus network distribution model and algorithm. The data of the passenger OD of the bus station in the future is predicted, and the bus dispatching subsystem is used to intelligently dispatch the bus under different conditions according to the predicted data, thereby forming a scheduling model in different situations.
- 5 Bus operation schedules. According to the actual bus line conditions in the central database and the predicted line travel demand, a specific schedule is generated to realize the actual intelligent dispatch of the bus.

5. Model Prediction Process

The data is sent to the information collection subsystem processing center through the communication subsystem, and the processing center first completes the integration of the IC card information and the GPS data. Firstly, the RBF neural network is used to classify and format the integrated data according to the requirements of the RBF neural network. Then, the RBF neural network is used to process the bus traffic data in the future, so as to predict the future bus passenger volume. The specific forecasting steps are as follows:

- 1 Through manual survey data or historical data in the central database, the relationship between the number of IC card passengers and the number of cash passengers used on each bus line is statistically analyzed in the order of time segments. This ratio relationship changes with time. Update.
- 2 According to the historical statistical data of the IC card ticketing system, the data is processed and analyzed to determine the OD distribution law of the IC card passengers traveling in each time period, and the real-time data of the IC card ticketing system is continuously updated.
- 3 According to the real-time operational data of the IC ticketing system, calculate the IC card passenger travel demand status of each station and line in the past [7-8].
- 4 Using the RBF neural network algorithm to predict the demand situation of the IC bus passengers using the IC card ticketing system and the IC card passengers at the future, the specific prediction process is shown in Figure 3.

The RBF neural network processing process is as follows [1]:

- 1) Select raw data as sample training data and prediction data;
- 2) Normalize the data and transform it into an input mode of RBF neural network requirements;
- 3) Construct an RBF neural network and define the walking constants and network training goals of the network;
- 4) Adjust the number of hidden layer units and the center of the hidden layer unit, calculate the hidden layer weights by Gaussian function, and obtain the hidden layer output;
- 5) obtaining the output of the output layer by a function relationship between the hidden layer and the output layer;



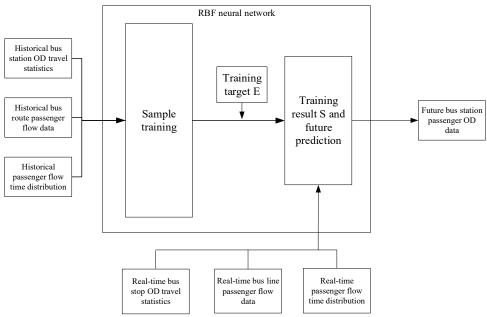


Fig.3 RBF neural network traffic forecast

- 6) The error judgment index function calculates the training error;
- 7) Whether the training error is less than the specified target, then go to step 9), otherwise go to the next step;
- 8) using the least squares method to perform weight adjustment, modify the output layer weight, and proceed to step 6);
 - 9) If all the samples are trained, go to the next step, otherwise return to step 4);
 - 10) using a neural network to predict normalized processed data for testing network performance;
 - 11) The predicted data is de-processed to obtain the true predicted value.
- 5 Calculate the travel demand of all passengers at each station and line in the future based on the predicted results and the proportional relationship between the line, station IC card passengers and all passengers.
- 6 Determine the bus network allocation model according to the passenger travel demand, and allocate the bus travel demand to each road segment and line.
 - 7 Use the line distribution model to determine the bus schedule for individual routes.

6. Passenger Traffic Prediction Example

6.1 Model Data Setting

The case analysis data is derived from the public transportation special planning project of Yubei District of Chongqing Municipality [9]. The project counts the IC information and GPS data of Chongqing's main bus route in the main city. In order to study the universality and rationality, the line passenger traffic of data acquisition is relatively large and typical. Therefore, this paper selects the stations to which the 462 lines of the major commercial districts in Chongqing belong to the data source. The starting and ending points of the test sections are Xiaolongkanzhengjie Station and Jiefangbei Station, as shown in Figure 4.

A total of 22 test object sites were selected, and these 22 sites were non-holiday by group survey. A relationship data survey using IC card passengers and the entire passenger is conducted from 1 am to 8 pm on an hour-hour basis. According to the survey results, the proportion of passengers using IC cards is about 45%.





Fig.4 Schematic diagram of the line

The collected IC card information and GPS data are integrated by using heterogeneous information integration technology, and then the OD analysis method is used to obtain the OD distribution of IC card passengers traveling at each site. In order to facilitate the statistics of the database during the analysis, the specific name of the site is replaced by the number 0102-0123, and the integrated data of the same time period and the integrated data of the same site are obtained. As shown in Figure 5 and Figure 6.

When using model prediction, consider the number of passengers on board and the number of passengers on board at each site [10-12]. Through analysis, the integrated data of the passengers on the station and the passengers on the site are integrated, as shown in Figure 7 and Figure 8.

:N_STA_STARTIME 🔻	IN_STA_STOPTIME	→ IN NUMBER →	TOTALE OUT_01	02 VOUT_0103	▼ 0UT_0104	▼ OUT_0106	▼ 0UT_0107	▼ 0UT_0108	▼ OUT_0109	▼ 0UT_011
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0102	280	1	19	14	8	30	8	9
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0103	218	35	1	8	7	22	9	5
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0104	251	28	15	1	7	18	21	9
2013/3/24 8:00:00	2013/3/24 8:5	59:59 <u>0105</u>	0							
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0106	151	47	19	2	2	3	1	
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0107	406	80	36	7	5		1	7
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0108	277	41	18	8	2	2	8	3
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0109	309	54	30	7	2	7	3	6
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0110	832	186	60	29	8	27	20	5
2013/3/24 8:00:00	2013/3/24 8:5	9:59 0111	359	55	15	17	4	25	21	5
2013/3/24 8:00:00	2013/3/24 8:5	9:59 0113	437	72	31	11	3	21	26	20
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0114	864	126	55	19	17	54	50	30
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0115	271	39	11	6	2	11	14	7
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0116	384	38	10	4	6	19	12	10
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0117	315	21	9	10	3	8	24	6
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0119	612	63	13	5	3	15	5	15
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0120	66	7	1	3		1		1
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0121	0							
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0122	147	17	1			5	2	2
2013/3/24 8:00:00	2013/3/24 8:5	59:59 0123	842	61	14	10	5	16	14	15
2013/3/24 9:00:00	2013/3/24 9:5	59:59 0102	389	1	15	8	3	46	15	15
2013/3/24 9:00:00	2013/3/24 9:5	9:59 0103	235	28	5	11	13	7	7	5
2013/3/24 9:00:00	2013/3/24 9:5	9:59 0104	249	38	24	2	5	18	21	7
2013/3/24 9:00:00	2013/3/24 9:5	9:59 0105	0							
2013/3/24 9:00:00	2013/3/24 9:5	9:59 0106	151	33	23	5	1	5	1	

Fig.5 Integrating data in the same time period



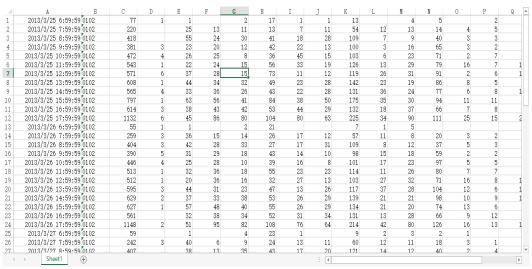


Fig.6 Integrating data with the same site

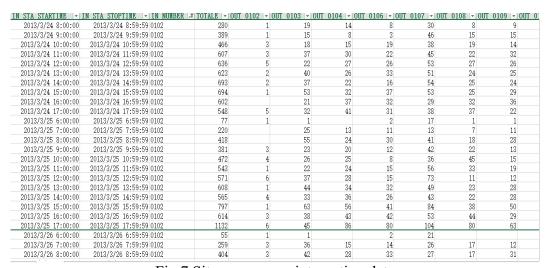


Fig.7 Site passengers integration data

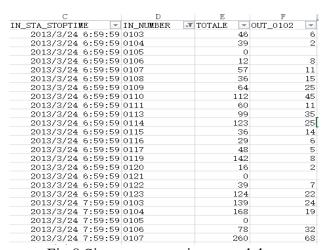


Fig.8 Site passengers integrated data

After the model is predicted, the number of passengers on board and the number of passengers on the same site at the same time can be counted, and the future passenger traffic of the station can be obtained.



6.2 Model Verification and Analysis

The model is used to predict the passenger traffic of the first station. The model input data includes the bus passenger statistics on the bus station history, the passenger time distribution statistics of the IC card used in the bus, and the passenger flow analysis data of the bus line passing through the station [13]. In order to verify the contrast between the model prediction results and the actual situation, and to improve the accuracy of the prediction results, it is necessary to maximize the capacity of the training samples. Therefore, 80% of the data is used as the model training sample, and 20% of the data is used as the unpredicted sample. The comparison of the predicted results is shown in Figures 9 and 10.

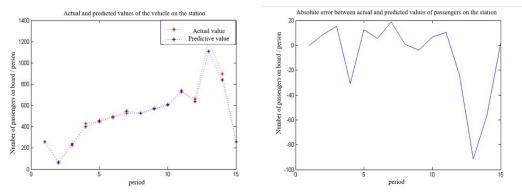


Fig.9 Curve and error curve of actual and predicted values of passenger flow on board

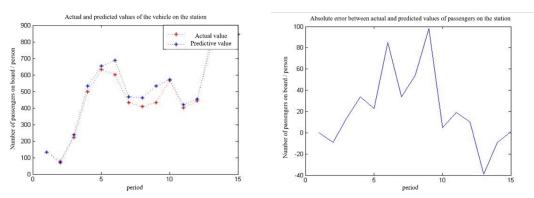


Fig. 10 Curve and error curve of actual and predicted values of passenger flow

Fig. 9 and Fig. 10 show that the change of the error curve at the late peak is relatively large. The number of passenger changes greatly at the peak of the station, and the prediction difficulty increases, resulting in a decrease in prediction accuracy. Through the error curve of the boarding and disembarking, it can be known that the bus IC card charging system adopts the boarding of the vehicle, and the vehicle is not brushed: the number of passengers who get off is the basic data of the number of passengers in the above vehicle and is calculated by using the OD model. Therefore, the predicted result of the number of passengers on board is closer to the true value than the predicted result of the number of passengers getting off.

7. Conclusion

Based on the real-time dynamic acquisition technology of bus IC card information and GPS data and the RBF neural network suitable for multivariate data processing, a bus and traffic forecast model based on GPS and IC card is proposed. Real-time dynamic passenger flow information and vehicle information are obtained through the IC card ticketing system and GPS positioning system on the bus. The data is statistically analyzed by back-calculation method to obtain the real-time passenger flow distribution of passenger traffic, and then the passenger flow in the future time period through the RBF neural network. The quantity is forecasted, and the real-time dynamic bus passenger traffic forecasting model is established, and the example is verified by Chongqing 462 bus line. The results



show that the model cannot only reduce the manpower and material resources of the survey, but also improve the prediction accuracy to a large extent, which has certain value in both theory and practical application.

Acknowledgments

This work was supported by fusion algorithm and simulation research based on vehicle GPS data (project number: 2015-121-007).

References

- [1]. Tang Wei. Public exchange information extraction based on IC card and vehicle GPS data. China Urban Planning Society Urban Transport Planning Academic Committee. Innovation Drive and Wisdom Development Proceedings of the 2018 China Urban Transport Planning Annual Conference China Urban Planning Society Urban Transportation Planning Academic Committee: China Urban Planning and Design Institute Urban Transportation Research Institute, Vol. 9 (2016) No.28, p. 103-105.
- [2]. Li Wenfeng, Lin Yanling, Cheng Yuan, Lin Xiaowei, Wang Huaquan. Visualization Analysis of Bus Passenger Flow Based on Bus IC Card and GPS Data. Transportation Technology and Economy, Vol. 5 (2018) No.20, p. 55-59+80.
- [3]. Qin Zheng. Study on passengers getting on and off the site based on bus IC card and GPS data. Western Transportation Science & Technology, Vol. 8 (2017) No.18, p. 115-119.
- [4]. Yin Changyong. Estimation Method of Public Transport OD Matrix Based on GPS and IC Card Data. Shanxi Architecture, Vol. 19(2017) No.43, p. 11-13.
- [5]. Li Xingyue. Data Fusion and Bus Modeling Analysis Based on IC Card and Vehicle GPS. China Urban Planning Society, Shenyang Municipal People's Government. Planning 60 Years: Achievements and Challenges Proceedings of 2016 China Urban Planning Annual Conference (04 Cities Planning new technology applications). China Urban Planning Association, Shenyang Municipal People's Government: China Urban Planning Association, Vol. 8 (2016) No.4, p. 125-128.
- [6]. Yang Wanbo, Wang Wei, Ye Xiaofei, Xu Wei, Jiang Dongxue. The OD Estimation Method of Public Transport Based on GPS and IC Card Data. Journal of Chongqing Jiao tong University: Natural Science Edition, Vol. 3 (2015) No.34, p. 117-121.
- [7]. Huang Jie. Inferring the type of travel activity based on bus IC card data and GPS data. Southwest Jiao tong University, Vol. 112016) No.21 p. 148-151.
- [8]. Li Ying, Weng Xiaoxiong. Transition Recognition Method Based on Bus IC Card and GPS Data. Journal of Guangxi University (Natural Science Edition), Vol. 2(2017) No.42, p. 579-586.
- [9]. Luo Yunhui. Research on Bus Passenger Flow and Bus Service Level Analysis and Evaluation Method Based on IC Card and GPS Information. China Urban Planning Society Urban Transport Planning Academic Committee. 2016 China Urban Transport Planning Annual Conference Proceedings. Proceedings of the 2016 China Urban Transport Planning Annual Conference: China Urban Planning and Design Institute Urban Transport Professional Research Institute, Vol. 10 (2016) No.31, p. 55-56.
- [10]. Su Yan, Liang Wu. Predicting the Mechanical Properties of Casting and Rolling 7050 Aluminum Alloy Based on RBF Neural Network. Hot Working Technology, Vol. 21 (2018) No.47, p. 145-147+151.



- [11]. Han Xiangxi, Zhang Heng, Feng Zhiqiang, Qiu Ang, Yu Jianxing, Fu Wei. Two-objective Optimization of Fishing Boat Tail Profile Based on RBF Neural Network. Ship Engineering, Vol.10 (2018) No.40, p. 35-39+58.
- [12]. Liu Dongyang, Li Yuelin, Ge Jin, Liu Bofu, Zhang Yin. Prediction of ignition advance angle of gasoline engine based on chaotic RBF neural network. Journal of Automotive Engineering, Vol. 5(2018) No.8, p. 352-357.
- [13]. Lu Baichuan, Deng Jie, Ma Qinglu, Liu Quafu, Zhang Kai. Short-term bus passenger traffic forecast based on IC card and RBF neural network. Journal of Chongqing Jiao tong University: Natural Science Edition, Vol. 6 (2015) No.34, p. 106-110.