Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/81/20241044

A study of short-term passenger flow forecasting for public transportation based on the time series method

Guiyang Meng

School of Foreign Language Middle School, Xiamen, 361000, China

menggy2008@outlook.com

Abstract. In order to solve the conflicts between the supply and demand of urban transport, improve the living standard of urban residents, raise the level of urbanization, and promote the sustainable socio-economic development of the whole city, it is necessary to increase the pace of urban transport construction, planning and management in China. And at the same time, it's necessary to take various measures to regulate the structure of urban transport and guide urban transport to develop in the direction of taking public transport as the mainstay. The experience of the development of urban transportation in China and abroad also proves that prioritizing the development of public transportation is the fundamental way to solve the problem of urban transportation. Public transport passenger flow data is the basis for optimizing public transport operations and scheduling. Accurate passenger flow forecasts on public transport routes can effectively guide public transport operation decisions, formulate operation scheduling plans and effectively improve the operational efficiency of the public transport system. The following are some of the key factors that can be taken into consideration when making passenger flow forecasts. The main contribution of this article is to propose the use of data mining technology, use bus IC card data and bus GPS data to integrate and organize them, conduct correlation analysis on the data, and on this basis, obtain the main distribution characteristics of bus passenger flow, and provide short-term Passenger flow forecast provides relevant ideas.

Keywords: Public transit passenger flow, data mining, short-term forecasting, time series method.

1. Introduction

In the exploration of passenger flow analysis and prediction, Zuo adopted the algorithm model of extreme learning machine in the prediction of urban short-time passenger flow and found the practicality and feasibility of the model in passenger flow prediction [1]. Zhou predicted the passenger flow by Autoregressive moving average model (ARMA) model, found the practicality of ARMA model prediction and predicted the advancement of ARIMA model [2]. Liu used advanced algorithmic models such as K-Nearest Neighbor (KNN) in deep learning to lay a further foundation in passenger flow prediction, and analyzed the efficiency of KNN model in depth [3]. Sun further analyzed the passenger flow prediction using machine learning models [4]. Lin predicted the local short-term passenger flow using the modeling algorithm of support vector machine [5]. Ouyang reasonably predicted the local short-time passenger flow using Long Short-Term Memory (LSTM) model [6]. Huang proposed the drawbacks of current passenger flow prediction using the traditional prediction method of urban rail

DOI: 10.54254/2755-2721/81/20241044

transit system [7]. Sun used intelligent bus system and Geographic Information Science (GIS) system to predict bus passenger flow [8]. Song proposed further thoughts and gave relevant outlooks on bus passenger flow prediction using deep learning models [9]. Jiang used ARIMA model and found its feasibility and efficiency in passenger flow prediction [10].

In today's society, the growing flow of people, urbanization process accelerated, bus passenger flow as a large part of the transportation flow, will face great challenges. In today's real bus passenger flow deployment, there are incomplete methods, poor management utility, low efficiency, the effect of the status quo is not obvious, resulting in bus deployment cannot be adjusted in time to optimize the structure is not reasonable, directly affecting the efficiency and level of the bus system operation. Therefore, it is extremely important to adopt suitable algorithmic model to be applied to the actual bus passenger flow deployment and prediction.

In the field of short-time passenger flow prediction, data mining technology, machine learning technology and deep learning technology have appeared, the former use a variety of models and algorithms have been carried out on public transportation short-time passenger flow prediction research, in the data mining, there are extreme learning machine algorithms, time series algorithms, support vector machine algorithms, and other model branches. This paper will go deeper into the data mining technology and study the time series algorithm for short-time passenger flow prediction of public transportation. By comparing the prediction results of the time series method with the actual results to analyze the characteristics of passenger flow distribution, and then make optimization measures for the bus short-time passenger flow prediction.

2. Methodology

2.1. Data sources and description

The data used in this paper were provided by a public transportation group company in a city in China, which is responsible for the major part of the city's public transportation operations. The company has 1247 vehicles in operation, more than 110 bus routes, a total network length of 1716 kilometers, 2210 bus stops, more than 170 bus drivers, and 18 large stations, with four terminals at the secondary level. The secondary units of the company include four terminals, each of which manages different routes and vehicles. The company carries more than one million passengers daily.

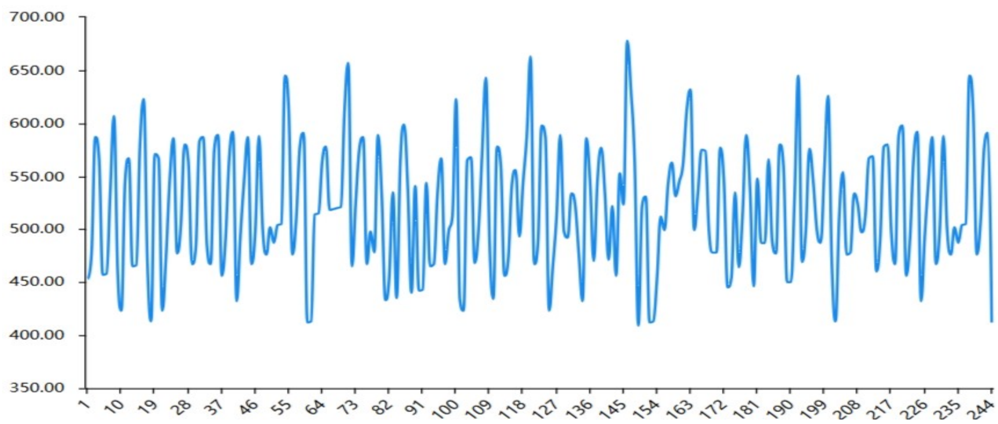

Figure 1. Passenger flow data of a bus station in Shanghai during the morning peak period.

It can be seen from Figure 1 and the data given by the public transportation group that under the general trend, the total passenger flow across the country has increased significantly, with a larger increase compared with the previous year. The growth in urban public transportation passenger flow is also very significant.

2.2. Method introduction

Temporal sequencing refers to some degree of dependence or association between sequential values. It is sometimes characterized by a definite pattern of change, usually called autocorrelation. Time series analysis is a method of analyzing and showing the pattern of change of values in the order of precedence and their statistical properties, usually referred to as mathematical modeling, which shows the inherent pattern of change of time series data and their statistical properties, referred to as time series modeling analysis.

Any time series has the model form of ARIMA(p, d, q), where P is the autoregression term, q is the number of moving average terms, and d is the number of differences made when the time series becomes stationary. The full name of the ARI MA model is the differential autoregressive moving average model. It is a famous time series forecasting method proposed by Box and Jenkins in the early 1970s, called Box-Jenkins method. According to the different components included, time series models can be divided into several types, including autoregressive (AR) models, moving average (MA) models, autoregressive moving average (ARMA) models, etc., which all belong to ARI MA (p ,d, q) special form of model.

The advantage of time series method is that all the influencing factors are attributed to the time factor, and the mathematical relationship between the time series is found by means of data fitting, etc. The method is simple and easy to realize. The disadvantage is that the time series prediction method does not take into account the influence of external factors because of the time series, so it has the shortcoming of prediction error, and when there is a big change in the external world, there will be a big deviation.

3. Results and discussion

3.1. Preliminary work

Through the smoothness test that is the ADF test on the ARIMA model to determine the number of differential order finally determined in the d value of 2, to meet the conditions of the smooth series, so d = 2, test chart as Table 1.

Table 1. ADF Test Table

Differential order	t	p	Critical values		
			1%	5%	10%
0	-0.921	0.781	-3.924	-3.068	-2.674
1	-1.38	0.592	-3.924	-3.068	-2.674
2	-2.062	0.26	-4.138	-3.155	-2.714

Subsequently, partial autocorrelation and autocorrelation function graphs are used to determine the lag order and q value in ARIMA. p=3 and q=4 can be easily obtained from the two function images in Figure 2 and Figure 3. The function images are as follows.

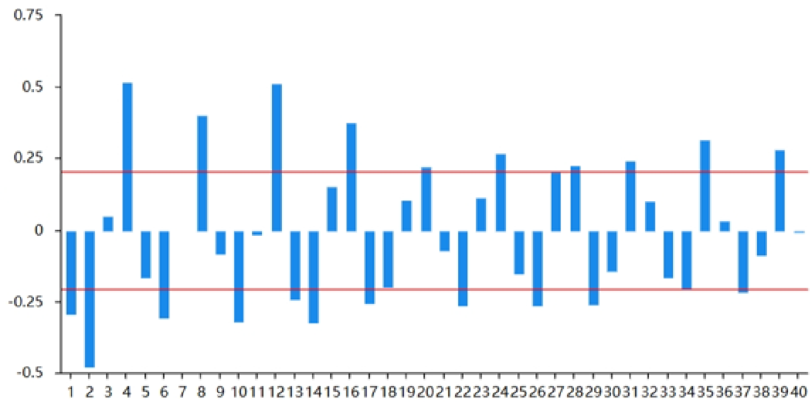

Figure 2. Plot of the autocorrelation function

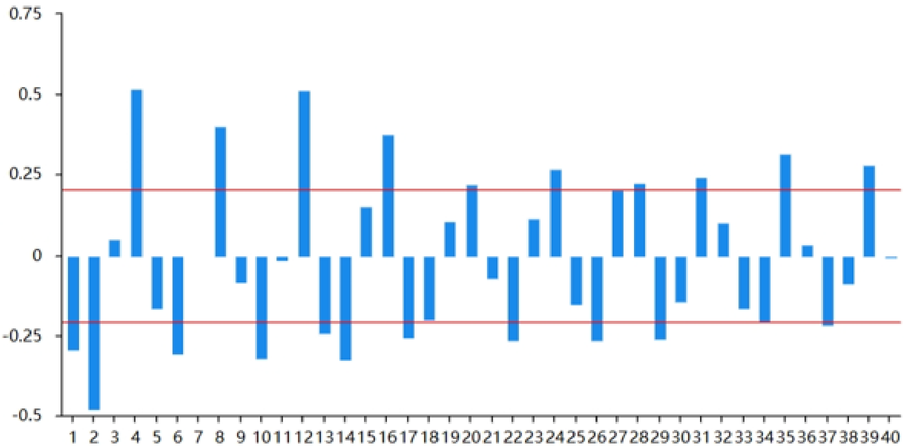

Figure 3. Image of the partial autocorrelation function

3.2. Model results

Due to meet the white noise test, and the model is a smooth series, so the model can pass the test finally ARIMA (3.2.4) as the model of the passenger flow prediction, which prediction results are shown in Figure 4.

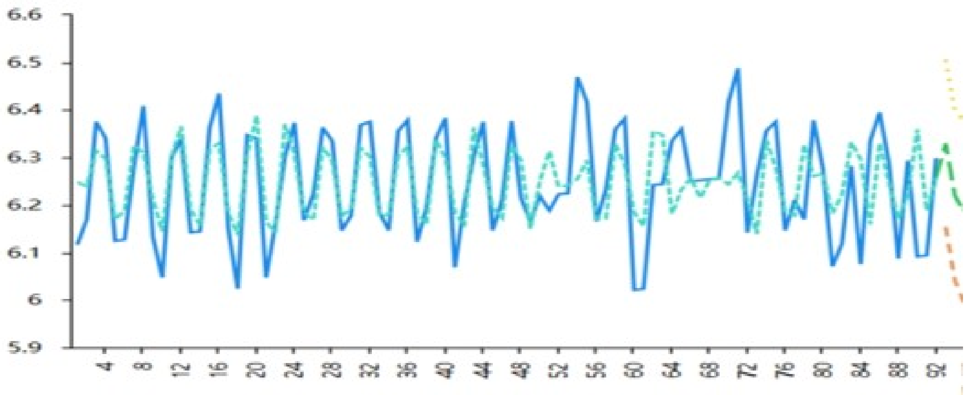

Figure 4. ARIMA (3.2.4) model for short-term passenger flow prediction results.

In this prediction result chart can be seen, ARIMA (3.2.4) for the future passenger flow prediction of the practicality and accuracy can be verified, is with a certain degree of feasibility, and the prediction of the future trend of passenger flow for the real results of the comparison is also close to the same, it can be seen that, the ARIMA model for the public transportation short-term traffic prediction of the effectiveness of the more significant, and to a certain extent to help alleviate the current acute problems of short-term passenger flow prediction management errors such as low efficiency, for the future traffic management to provide some reference. The current short-time passenger flow prediction management errors such as low efficiency and other acute problems, for the future of transportation management to provide some reference. The data in Table 2 can be concluded that the prediction error is within a reasonable range, and the relevant data are as follows.

DOI: 10.54254/2755-2721/81/20241044

Table 2. ARIMA Error Plot ARMA(3,2,4) model parameter table

item	Symbol	coefficients	SE	z-value	p-value	95% CI
constant	c	10.907	120.258	0.091	0.928	-224.793 ~ 246.608
AR	α1	0.022	10.332	0.002	0.998	-20.228 ~ 20.272
	α2	-0.824	0.463	-1.778	0.075	-1.732 ~ 0.084
	α3	0.056	8.479	0.007	0.995	-16.563 ~ 16.675
MA	β1	0.135	10.34	0.013	0.99	-20.131 ~ 20.401
	β2	0.4	2.074	0.193	0.847	-3.665 ~ 4.464
	β3	0.014	4.246	0.003	0.997	-8.307 ~ 8.335
	β4	-0.017	0.493	-0.035	0.972	-0.983 ~ 0.948
AIC: -163.356						
BIC: -140.660						

4. Conclusion

In the course of this thesis, an in-depth analysis of public transportation data, specifically from the 1C card and GPS systems, has been conducted utilizing sophisticated data mining techniques. The incorporation of a time series algorithm for the purpose of forecasting short-term fluctuations in public transportation passenger flow has been a pivotal aspect of this research. The results have conclusively demonstrated the superior efficiency and remarkable accuracy of the time series approach in anticipating the dynamics of passenger traffic over brief periods. This method has not only provided valuable insights into the temporal patterns of public transit usage but has also offered a robust predictive tool for urban transportation planning and management.

Utilizing the existing data sources like IC cards and GPS, this study employs data mining to explore the intricate patterns of public transport passenger flow across time and space. It delves into the temporal and spatial distribution traits, with a particular focus on the short-term dynamics of these flows. A comparative analysis of the variations in passenger traffic during weekdays and weekends highlights the nuanced characteristics of short-term fluctuations, offering a comprehensive understanding of public transport usage patterns.

Implementing the time series algorithm for the precise short-term prediction of public transportation passenger flow, this thesis presents a comparative analysis that underscores the superiority of the E LM algorithm. It excels in managing large datasets for short-term forecasts, demonstrating swift convergence and striking accuracy. The E LM algorithm's capacity to handle complex data patterns and its ability to provide reliable predictions even with high-volume inputs highlight its potential as a cornerstone for advanced transportation analytics.

The research conducted within the confines of this thesis is a modest contribution, representing a singular perspective amidst a vast field of inquiry. Recognizing the limitations and the preliminary nature of these findings, the author calls for a collective effort to delve deeper into the intricacies of short-term passenger flow prediction. Future endeavors in this domain could be enriched by considering a multifaceted approach, as suggested by the subsequent steps.

The integration of the Intelligent Public Transportation Scheduling System (IPTS) with the Geographic Information System (GIS) is proposed as a strategic enhancement to bolster short-term forecasting capabilities. This synergy not only fortifies the predictive models with spatial intelligence but also enables the dynamic visualization of forecast outcomes. Such real-time forecasting and deployment strategies are envisioned to revolutionize the way public transportation systems are managed, offering a more adaptive and responsive service to commuters. This proactive stance towards transportation planning, underpinned by robust data analytics, is poised to elevate the efficiency and effectiveness of urban mobility solutions.

DOI: 10.54254/2755-2721/81/20241044

At present, there are a lot of research results on short-time traffic prediction in China and abroad. This paper can establish a combined prediction model of short-time public transportation passenger flow by taking the methods of short-time vehicle prediction and combining the advantages of different models to improve the prediction accuracy.

References

[1] Zuo Q K 2017 Research on Public Transportation Passenger Flow Analysis and Short-Term Prediction Based on Data Mining. Southwest Jiaotong University, 1-16.
[2] Zhou W P 2023 Research on passenger flow characteristics of urban rail transit by data mining. Nanjing University of Science and Technology.
[3] Peng L 2023 Research on bus passenger flow prediction and departure schedule optimization method based on multi-source data. Dalian Maritime University, 1-22.
[4] Sun K and Zheng C J 2021 Station passenger flow estimation based on bus IC card swipe data. Journal of Guizhou University.
[5] Lin X W 2020 Optimization Research on Bus Combination Scheduling for Large Stops Based on Multi-source Data. Fuzhou University, 1-23.
[6] Ouyang Q 2019 Research on Bus Passenger Flow Prediction and Traffic Optimization Model Based on Big Data and Machine Learning. Beijing Jiaotong University.
[7] Huang E 2018 Research on Short-Term Passenger Flow Prediction Algorithm for Public Transportation Based on Data Mining. Hunan University,1-25.
[8] Sun W H 2017 Research on Public Transportation Passenger Flow Travel Characteristics Analysis and Forecasting Methods. Shandong University of Technology, 1-18.
[9] Song M L 2017 Research on optimization model and algorithm of urban bus network based on passenger flow driving. Shandong University of Technology.
[10] Jiang D H 2015 Research on Conventional Bus Scheduling Optimization Based on Public Transportation Data. Chongqing University of Posts and Telecommunications, 8-12.