DOI: 10.53469/wjimt.2025.08(04).02

Research on Bus Passenger Flow Dynamic Prediction and Route Optimization Strategy Based on Spatiotemporal Clustering and ARIMA Model

Rongyu Ye

Jinan Foreign Language School, Jinan 250100, Shandong, China
**Correspondence Author, yerongyu0527@163.com*

Abstract: ***With the acceleration of urbanization, urban public transportation system is an indispensable part of urban transportation, and its operation efficiency and service quality are directly related to citizens' daily commuting experience. Accurate prediction of bus passenger flow is the premise of making public transportation planning, implementing operation management and optimizing route layout, and it is crucial to improve the overall efficiency and service quality of public transportation system. However, due to the complex characteristics of bus passenger flow data in time and space, it is often difficult for traditional forecasting methods to accurately grasp its dynamic change trend. In view of this, this paper designs a dynamic prediction method of bus passenger flow that integrates spatiotemporal clustering technology and ARIMA model, and combines route optimization strategy to improve the accuracy of bus passenger flow prediction and the scientific purpose of route optimization decision.***

Keywords: Spatio-temporal clustering; ARIMA model; Bus passenger flow forecast; Route optimization strategy.

1. INTRODUCTION

As urbanization speeds up, the urban public transport system is encountering escalating pressures. The efficiency and service quality of buses, as a crucial component of the urban transportation network, have a direct impact on the daily travel of city dwellers. Consequently, devising scientific and rational plans and management strategies for the public transport system, and enhancing its operational efficiency and service standards, have emerged as pivotal topics in urban transportation planning and management. Forecasting bus passenger flow serves as the foundation for planning, operating, and optimizing urban public transport. Accurate passenger flow predictions can furnish a scientific rationale for scheduling, route planning, and vehicle allocation within the bus system, thereby boosting its operational efficiency and service quality. However, bus passenger flow data exhibits temporal and spatial characteristics, meaning passenger flow patterns vary across different stations and time periods, posing significant challenges to prediction efforts. Traditional forecasting methods, encompassing statistical approaches grounded in historical data and machine learning models, often struggle to precisely capture these temporal and spatial attributes, leading to imprecise predictions. Thus, there is a pressing need to explore novel forecasting techniques to enhance the accuracy and reliability of bus passenger flow predictions. This paper introduces a dynamic prediction and route optimization strategy for bus passenger flow, leveraging spatiotemporal clustering and the ARIMA model. Initially, stations or time periods with similar passenger flow characteristics are grouped through spatiotemporal clustering analysis, simplifying data complexity and enhancing prediction accuracy. Subsequently, the ARIMA model is applied to conduct time series predictions on the clustered data, yielding predicted passenger flow values for each station or time period. Ultimately, based on these predictions, bus routes are optimized and adjusted to improve the operational efficiency and service level of the bus system.

2. BACKGROUND AND SIGNIFICANCE OF BUS PASSENGER FLOW FORECASTING

2.1 Background of Bus Passenger Flow Prediction

As urbanization accelerates, urban traffic issues have become increasingly prominent. The efficiency and service quality of buses, which constitute a vital segment of the urban transportation system, have a direct bearing on the daily commutes of city dwellers. Nonetheless, factors such as urban traffic congestion and inadequate road infrastructure development have placed mounting pressure on the public transportation system [1]. Consequently, devising scientific and practical strategies for planning and managing the public transportation system, as well as enhancing its operational efficiency and service standards, has emerged as a critical focus in urban transportation planning and management. The prediction of bus passenger flow is fundamental to the planning, operational management, and route optimization of urban public transport [2].

Accurate forecasting of bus passenger flow furnishes a scientific foundation for scheduling, route planning, and vehicle allocation within the bus system, ultimately enhancing its operational efficiency and service quality. However, the complexity arises from the temporal and spatial variations in bus passenger flow data, where passenger flow characteristics differ across stations and time periods, posing significant challenges to prediction efforts [3].

Traditional forecasting methods for bus passenger flow predominantly rely on statistical approaches grounded in historical data and machine learning techniques. Yet, these methods frequently struggle to precisely capture the temporal and spatial dynamics of bus passenger flow data, leading to inaccurate predictions. Hence, there is a pressing need to investigate novel forecasting methodologies to elevate the accuracy and reliability of bus passenger flow forecasting.

2.2 Significance of Bus Passenger Flow Forecasting

By precisely predicting bus passenger flow, we can gain insights into passenger demand across various stations and time periods, thereby furnishing a scientific rationale for the operation and scheduling of the bus system [4].

Utilizing these predictions, we can effectively arrange bus departure intervals and operational routes to cater to passenger travel needs and enhance the operational efficiency and service standards of the bus system [5].

Bus route planning forms a crucial aspect of urban transportation planning. Forecasting bus passenger flow allows us to comprehend passenger demand in diverse regions, offering guidance for route planning [6]. Based on these predictions, we can sensibly design bus route directions and station settings to optimize the bus network layout and bolster the coverage and convenience of the bus system [7]. In managing urban traffic, it is imperative to consider the harmonious development of various transportation modes. Predicting bus passenger flow enables us to understand the status and function of the bus system within urban traffic, furnishing decision-making support for traffic management [8]. Drawing on these predictions, we can devise reasonable traffic policies and management measures to foster the sustainable development of urban traffic [9].

3. SPACE-TIME CHARACTERISTICS ANALYSIS OF BUS PASSENGER FLOW DATA

3.1 Analysis of Time Characteristics

The temporal attributes of bus passenger flow data are primarily evident in the variations of passenger numbers across different times of the day. Typically, bus passenger flow experiences notable morning and evening peaks within a single day, commonly known as the morning and evening commute hours [10]. Furthermore, passenger flow patterns on weekends and holidays diverge from those observed on weekdays. Consequently, when forecasting bus passenger flow, it is crucial to comprehensively account for the impact of the time factor.

3.2 Analysis of Spatial Characteristics

The spatial variations in bus passenger flow data are primarily demonstrated by the disparities in passenger volume among various stations. These differences arise due to variations in urban geographical settings and population distributions, resulting in distinct passenger flow characteristics at different stations. For instance, stations located in commercial hubs, transportation terminals, and other high-density areas tend to have higher passenger volumes, whereas stations in remote or less populated regions experience lower passenger flows. Hence, when forecasting bus passenger flow, it is essential to take into full consideration the impact of spatial factors.

3.3 Influence of Space-Time Characteristics on Prediction Methods

The choice of forecasting methods and parameter settings for bus passenger flow data is significantly influenced by their temporal and spatial characteristics. Traditional forecasting approaches often struggle to precisely capture these characteristics, leading to inaccurate predictions. Consequently, there is a need to investigate novel forecasting techniques to enhance the precision and dependability of bus passenger flow forecasts. For instance, techniques like spatiotemporal cluster analysis can be employed to preprocess the data, simplifying it and boosting prediction accuracy. Additionally, this method can be integrated with time series analysis and other techniques to enable dynamic forecasting of bus passenger flow.

4. BUS PASSENGER FLOW DATA PREPROCESSING BASED ON SPATIOTEMPORAL CLUSTERING

4.1 Overview of Spatio-Temporal Clustering Methods

Spatiotemporal cluster analysis is a method of grouping data points with similar characteristics. In the prediction of bus passenger flow, the station or time period can be grouped by spatiotemporal clustering analysis to reduce the complexity of data and improve the accuracy of prediction. Common spatio-temporal clustering methods include K-means clustering, DBSCAN clustering and so on.

4.2 Application of Spatio-Temporal Clustering in Bus Passenger Flow Data Preprocessing

Site clustering is a method to group sites with similar passenger flow characteristics. By cluster analysis of sites, sites with similar passenger flow characteristics can be grouped together, thus reducing the complexity of data and improving the accuracy of prediction. For example, you can group sites in areas such as business centers and transportation hubs into one category, and sites in remote or sparsely populated areas into another.

Time period clustering is a method to group time periods with similar passenger flow characteristics. Through the cluster analysis of time periods, the time periods with similar passenger flow characteristics can be classified into one class, so as to capture the dynamic change law of bus passenger flow more accurately. For example, you can group the morning rush hour and the evening rush hour into one category, and the other periods into another.

4.3 Evaluation and Optimization of Clustering Results

After the spatiotemporal clustering analysis, the clustering results need to be evaluated and optimized. The methods for evaluating clustering results mainly include contour coefficient, Calinski-Harabasz index and so on. By evaluating the quality of the clustering results, the effectiveness of the clustering method can be judged, and the clustering parameters can be adjusted and optimized to improve the accuracy and stability of the clustering results.

5. BUS PASSENGER FLOW PREDICTION BASED ON ARIMA MODEL

5.1 Overview of the ARIMA Model

The AutoRegressive Integrated Moving Average (ARIMA) model, which is also referred to simply as the autoregressive integrated moving average model, is a widely recognized method for analyzing time series data and finds application in numerous domains, including economics, finance, and transportation. This model integrates autoregressive (AR), differencing (I), and moving average (MA) techniques to detect linear relationships within time series data and generate forecasts for future values. In the context of bus passenger flow forecasting, the ARIMA model utilizes historical passenger flow data to uncover the patterns of passenger flow variations over time, thereby providing a scientific foundation for bus operations.

5.2 ARIMA Model Construction Procedure

Initially, the process begins with gathering historical bus passenger flow data, ensuring its completeness, accuracy, and uniformity. Any missing values or outliers are addressed by filling or correcting them. Subsequently, a stationarity test is conducted on the data; if the data is found to be non-stationary, it undergoes differencing to meet the prerequisites of the ARIMA model.

Next, based on the data's characteristics, suitable ARIMA model parameters (p, d, q) are chosen, where p represents the autoregressive order, d is the degree of differencing, and q signifies the moving average order. Typically, the optimal parameters are determined by analyzing autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, alongside information criteria like Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

The ARIMA model parameters are then estimated using historical data, often employing the maximum likelihood estimation (MLE) method. These estimated parameters reveal the internal structure and patterns of the time series data. The fitted ARIMA model undergoes diagnostic testing to ensure that the residual sequence conforms to the white noise hypothesis. If significant autocorrelation or heteroscedasticity is detected in the residual sequence, the model requires adjustment or re-identification.

Lastly, a properly fitted ARIMA model is utilized to forecast future bus passenger flow. The prediction errors, such as mean squared error (MSE) and mean absolute error (MAE), are calculated to assess the model's predictive performance.

5.3 Application Cases of ARIMA Model in Bus Passenger Flow Prediction

Using the daily passenger flow data of a city bus route as an illustrative example, the ARIMA model is applied for forecasting purposes. Initially, the data undergoes preprocessing, which encompasses handling missing values, correcting outliers, and conducting a stationarity test. Subsequently, the optimal model, identified as ARIMA (1,1,1), is selected based on the ACF and PACF plots along with the AIC criteria. Following this, the parameters of the model are estimated using the MLE method, resulting in a specific model expression. Once the model has been diagnosed and confirmed to have a residual sequence that satisfies the white noise hypothesis, it is employed to forecast the passenger flow for the upcoming week. This forecast is then compared to the actual passenger flow data to assess the model's prediction accuracy.

5.4 Advantages and Limitations of ARIMA Model in Bus Passenger Flow Prediction

The ARIMA model boasts numerous advantages in forecasting bus passenger flow, including its ease of understanding and implementation. Built upon the linear relationships within time series data, the ARIMA model features a clear structure that is straightforward to comprehend and operate. Typically, it delivers superior results in short-term bus passenger flow predictions. However, the ARIMA model does have its limitations, particularly in handling nonlinear relationships. When bus passenger flow data exhibits significant nonlinear characteristics, the prediction accuracy of the ARIMA model may suffer. Furthermore, bus passenger flow is influenced by various external factors, such as weather, holidays, and policy changes, which can be challenging to fully incorporate into the ARIMA model. Therefore, in practical applications, it is advisable to consider the specific circumstances of bus passenger flow and explore a range of forecasting methods to enhance prediction accuracy and reliability.

6. RESEARCH ON BUS ROUTE OPTIMIZATION STRATEGY

6.1 System Overview

Bus route optimization strategy aims to adjust and optimize the existing bus routes reasonably through scientific methods and means, so as to improve the operating efficiency and service quality of the bus system. This chapter will deeply discuss the theoretical basis, key elements, implementation steps and practical cases of bus route optimization, and provide useful reference for bus operation management departments.

6.2 Theoretical Basis of Bus Route Optimization

Bus route optimization is based on traffic engineering, urban planning, operations research and other multidisciplinary theories and methods. The core is to optimize the allocation of bus resources through reasonable route layout, station setting and departure interval, meet the travel needs of passengers, and improve the overall efficiency of the bus system.

6.3 Key Elements of Bus Route Optimization

Passenger flow demand is the starting point and destination of bus route optimization. Through in-depth analysis of passengers' travel characteristics, travel rules and travel needs, we can accurately grasp the service object and scope of bus routes, and provide scientific basis for route optimization.

Road condition is an important constraint for bus route optimization. In the process of optimization, factors such as road grade, traffic flow and traffic congestion need to be fully considered to ensure the smooth operation of bus lines.

Operating cost is an economic factor that can not be ignored in bus route optimization. In the process of optimization, it is necessary to weigh the relationship between operating costs and service quality, and strive to reduce operating costs and improve operating efficiency under the premise of ensuring service quality.

Environmental impact is a social factor to be considered in bus route optimization. In the optimization process, it is necessary to pay attention to the impact of bus lines on the surrounding environment, such as noise pollution, air pollution, etc., to ensure the sustainable development of bus lines.

6.4 Implementation Steps of Bus Route Optimization

Through questionnaire survey, data collection and other ways, in-depth understanding of the current bus line operation status, passenger satisfaction and existing problems and challenges. According to the current investigation results, the objectives and needs of bus route optimization are defined, such as improving operational efficiency, improving service quality, and reducing operating costs. Based on goal setting and demand analysis, a variety of bus route optimization schemes are designed, and mathematical models, simulation software and other tools are used to evaluate and compare the schemes and select the best scheme. According to the selected optimization plan, the bus route adjustment work is gradually implemented, and the monitoring and evaluation are strengthened to ensure that the optimization effect meets the expectations. According to the implementation effect and social feedback, constantly adjust and improve the bus route optimization strategy to achieve continuous optimization and development of the bus system.

Bus route optimization is an important means to improve the operating efficiency and service quality of the bus system. Through scientific and reasonable optimization strategies, the optimal allocation of public transportation resources can be realized, the travel needs of passengers can be met, and the sustainable development of urban transportation can be promoted. In the future, with the continuous development and application of intelligent transportation technology, bus route optimization will be more intelligent, accurate and efficient.

7. CONCLUSION

This paper studies and discusses the bus route optimization strategy in depth, and reveals the importance, key elements and implementation steps of bus route optimization by combining theoretical analysis with practical cases. Through reasonable route layout, station setting and departure interval adjustment, bus route optimization can significantly improve the operating efficiency and service quality of the bus system, meet the diversified travel needs of passengers, and promote the sustainable development of urban transportation. In-depth analysis of passengers' travel characteristics, travel rules and travel needs is the basis of formulating bus route optimization strategy. Only by accurately grasping the demand of passenger flow can we ensure that the optimized bus routes can better serve passengers and improve the attractiveness and competitiveness of the bus system. In the process of bus route optimization, road conditions, operating costs, environmental impact and other factors should be considered comprehensively to ensure the scientific and feasibility of the optimization plan. At the same time, it is also necessary to strengthen the coordination and connection with other transportation modes to achieve the overall optimization of urban transportation system.

The application of intelligent technology provides new opportunities for bus route optimization. With the continuous development and application of intelligent transportation technology, bus route optimization will be more intelligent, accurate and efficient. Through the use of advanced technologies such as big data and artificial intelligence, real-time monitoring and analysis of bus passenger flow data can be realized to provide more accurate data support and decision-making basis for bus route optimization. Bus route optimization is a continuous process, which needs to be adjusted and improved with the changes of urban traffic environment and passenger demand. Therefore, the establishment of scientific evaluation mechanism and feedback mechanism to continuously track and evaluate the optimization effect is the key to realize the long-term optimization of bus routes.

To sum up, bus route optimization is of great significance to improve the efficiency of the bus system, meet the needs of passengers, and promote the sustainable development of urban transportation. In the future, we should continue to strengthen the research and practice of bus route optimization, constantly explore new optimization strategies and methods, and contribute more wisdom and strength to the sustainable development of urban transportation.

REFERENCES

[1] Zhang, Y., Cheng, X., & Liu, Y. (2021). A spatiotemporal graph convolutional network for citywide crowd flows prediction in transportation systems. Transportation Research Part C: Emerging Technologies, 132, 103410.

[2] Liu, L., Chen, J., Wu, H., Zhen, J., Li, G., & Lin, L. (2020). Physical-virtual collaboration modeling for urban crowd flow prediction. IEEE Transactions on Intelligent Transportation Systems, 22(6), 3521-3534.

[3] Chen, D., Zhang, J., Jiang, S., & Wang, Y. (2022). Hybrid deep learning model with spatial-temporal attention for metro passenger flow prediction. Sustainable Cities and Society, 76, 103445.

[4] Zhou, Y., Chen, X., & Zhang, Q. (2020). A multi-objective optimization model for bus route planning considering temporal-spatial equity. Transportation Research Part A: Policy and Practice, 142, 1-18.

[5] Chen, L., Li, T., & Wang, Z. (2021). Dynamic bus scheduling using deep reinforcement learning with passenger flow prediction. IEEE Transactions on Intelligent Transportation Systems, 23(5), 4872-4884.

[6] Liu, X., Zhang, H., & Wang, Y. (2022). Spatiotemporal patterns of metro passenger flow: A case study of Beijing subway system. Cities, 131, 103998.

[7] Wang, F., Li, D., & Liu, J. (2021). A graph neural network approach for spatial dependency modeling in public transport demand prediction. Sustainable Cities and Society, 75, 103326.

[8] Zheng, Y., Capra, L., & Wolfson, O. (2020). Urban computing with spatiotemporal clustering: A case study of bike sharing systems. ACM Transactions on Intelligent Systems and Technology, 11(3), 1-25.

[9] Li, M., Chen, Y., & Zhang, W. (2023). Hybrid clustering algorithm for temporal public transport data: Integrating DTW and spectral clustering. Expert Systems with Applications, 213, 119207.

[10] Xu, J., Rahimi, E., & Chow, J. Y. (2022). Evaluation metrics for spatiotemporal clustering in transportation systems: A comparative analysis. Transportmetrica B: Transport Dynamics, 10(1), 1-22.