Research on Traffic Flow Prediction Based on ARIMA Model

Jiamu He[a]
School of Mathematical Sciences, Dalian University of Technology, Dalian, 116000, China

Keywords: Road Traffic, ARIMA, Traffic Flow Forecasting.

Abstract: Nowadays, road traffic has become the most mainstream mode of transportation, and its impact on people's lives is significant in terms of transportation efficiency and safety. Therefore, the accurate prediction of traffic flow is a research topic with high application value. This paper aims to establish a model for fitting and predicting the collected traffic flow data. Firstly, the ADF test will be conducted along with ACF and PACF plots to determine the approximate range of each input parameter of the model. Then, ARIMA models with different parameters will be applied to fit the data, and their mean square errors will be compared to identify the best-fitting model. The result indicates that the fitted values of this model closely align with the distribution of actual values, this proves the feasibility of ARIMA model for traffic flow fitting. Finally, the study will utilize this model to forecast data changes over a period of time in the future.

1 INTRODUCTION

The current process of urbanization is continuously accelerating, leading to an increasing demand for transportation and a continuous growth in traffic volume. Among these modes of transportation, road traffic stands out as the primary means of travel. Understanding the changing trends in road traffic flow is of significant importance for transportation planning. Therefore, predicting traffic flow has always been a focal point of research for transportation authorities. Accurately forecasting traffic flow aids relevant departments in implementing real-time traffic control measures, thereby alleviating traffic congestion and improving transportation efficiency. Although transportation authorities have introduced new technologies such as intelligent transportation systems to enhance the monitoring of traffic volume, accurately predicting changes in traffic flow remains a formidable challenge. This is due to the characteristics of road traffic flow, such as high volume, rapid fluctuations, and susceptibility to external influences.

There are various models and methods used for traffic flow prediction. Time series analysis is a typical method, whose common models include Autoregressive Integrated Moving Average (ARIMA) and its derivative models. These models attempt to identify patterns by breaking down long-term trends and extrapolating those patterns into the future. Kumar et al. (2015) obtained a suitable Seasonal ARIMA model by differentiating the data and adjusting the model parameters. Chikkakrishna et al. (2019) utilized actual data to establish the PROPHET model and SARIMA model, obtaining the optimal SARIMA model for the data. Regression analysis, which relates traffic flows to other factors, is also feasible for predicting future traffic flows. Feng et al. constructed an Adaptive Multi-kernel Support Vector Machine (AMSVM) to study the nonlinearity and randomness of traffic flow. They optimized AMSVM parameters and integrated spatial-temporal information with AMSVM, achieving accurate prediction (Feng et al., 2019). Machine learning methods can handle large amounts of data and complex features, adapting to nonlinear relationships and high-dimensional data. Mohammed et al. proposed using channel conditions for short-term prediction. The results showed slightly better predictions with the Distributed Random Forest model compared to the other methods (Mohammed and Kianfar, 2018). Deep learning models, capable of handling large-scale data, are another popular method for predicting traffic flow. Fu et al. used Long Short-Term Memory (LSTM) and Gated Recurrent Unit Neural Network methods (Fu et al., 2016). Zhang et

[a] https://orcid.org/0009-0009-1377-3313

He, J.
Research on Traffic Flow Prediction Based on ARIMA Model.
DOI: 10.5220/0012887800004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 16-21
ISBN: 978-989-758-713-9

al. converted spatio-temporal traffic flow features into two-dimensional matrix and constructed models by using Convolutional Neural Networks (Zhang et al., 2019). Lv et al. (2015) considered the spatiotemporal correlation of traffic flow, trained the hierarchical autoencoder model by greedy stratification, and then took the trained model as a building block. These different deep learning models achieved good prediction results.

Some researchers have combined various methods to predict traffic flow. Li et al. (2017) combined the ARIMA model with the Radial basis Function Artificial Neural Network model to capture different aspects of traffic flow patterns. They used two models to capture and model the linear and nonlinear components of the data, respectively. Lin (2020) et al. used Random Forest to calculate the importance of data features, eliminate redundant features, and apply the LSTM algorithm model. The above results both show that the algorithm combining the mixed models is superior to the algorithm using one of the models alone.

Chen et al. (2011) combined linear ARIMA models with non-linear Generalized Autoregressive Conditional Heteroskedasticity models, which could simultaneously capture more data at the same time. Compared to the predictive accuracy of the standard ARIMA model, the mixed model generally shows no improvement in performance, and sometimes even a decrease in performance. Therefore, in general, the standard ARIMA model can be used to achieve better results. Moreover, throughout the history of traffic flow prediction research, there have been many studies using ARIMA models, which indirectly reflect its good predictive performance. Therefore, this paper will also use ARIMA model to fit and forecast existing traffic flow data.

2 METHODS

2.1 Data Source

This paper uses a road traffic flow dataset from Huawei Munich Research, which contains recorded data over 3 days at an intersection in an urban area, and describes the traffic flow passing through the intersection in the form of a time series.

2.2 Indicator Selection

The paper takes intervals of every 5 minutes from the dataset as time periods and uses them as independent variables, with traffic flow as the dependent variable. This study is going to find a suitable model to fit the collected data over time and to make predictions about traffic volume for a future period. Figure 1 depicts the change of traffic flow over time.

2.3 Methodology Introduction

This paper employs ARIMA model, which is often used for handling data with trend and seasonality structures.

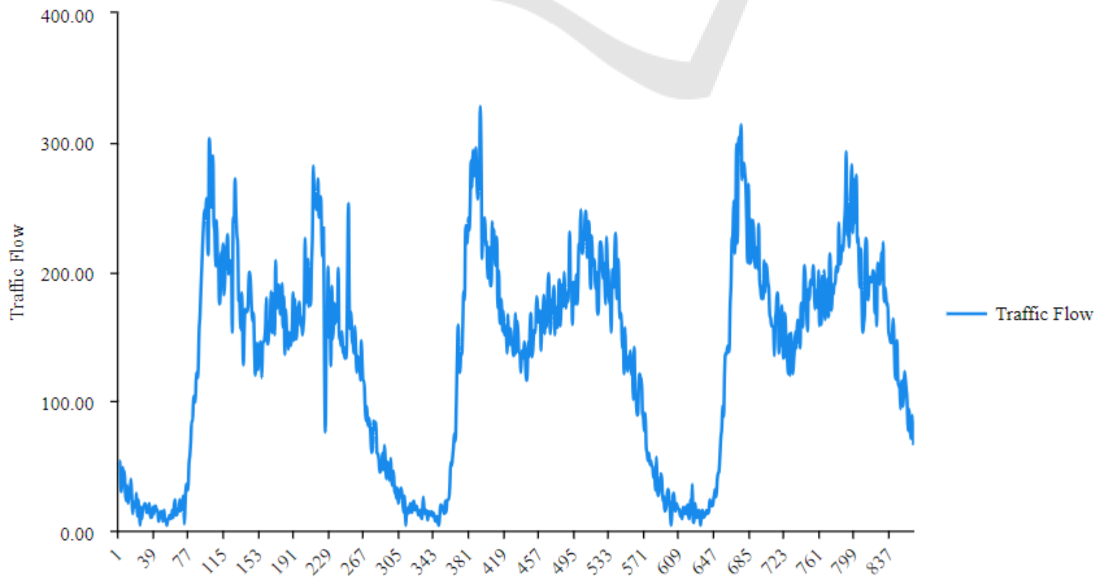

Figure 1: Time Series of Traffic Flow.

The standard representation of the ARIMA model is (p, d, q), where p, d, q are the parameters. The basic assumption of this model is that, through appropriate differencing operations, the time series can be transformed into a stationary sequence. Then, by the combination of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, the structure of the sequence can be modelled.

The ARIMA model is commonly used for short-term forecasting. For long-term predictions, more complex models may need to be considered.

3 RESULTS AND DISCUSSION

3.1 ADF Test

The Augmented Dickey-Fuller (ADF) test can verify the stationarity of time series. A P-value of less than 0.1 (sometimes 0.05) indicates that the sequence is stable at the 0.1 significance level.

Table 1: Traffic Flow - ADF Test.

d	t	P	Critical Value		
			1%	5%	10%
0	-3.677	0.004	-3.438	-2.865	-2.569

As table 1 provided, concerning the traffic flow, the P-value is 0.004 < 0.01. There is strong evidence (with over 99% confidence) that the sequence is stationary without differencing operations.

3.2 ACF and PACF Plots

The ACF and PACF plots are used to determine p and q. If the plots do not show clear truncation, it is necessary to choose the appropriate ARIMA orders. In this case, one can select the lags from the ACF plot as q and the lags from the PACF plot as p. If both the plots exhibit truncation, indicating that the randomness of the data is large, ARIMA modelling may not be suitable in such situations.

For traffic flow, combined with Figure 2 and Figure 3, it is recommended that d is 0, p is around 1 and q is around 1.

3.3 ARIMA Prediction

Given the value of d and the approximate range of p and q, model construction can be performed. The model parameter list shows the results of the model construction, and even if the P-value exceeds 0.05, it usually does not require much attention. Predictive metrics like mean squared error (MSE), information criteria such as AIC and BIC, are employed for multiple analyses to compare models. Lower values for MSE, AIC, and BIC are considered better, with MSE having the most significant impact on model fit. The best model can be obtained by repeatedly comparing the changes in these three values.

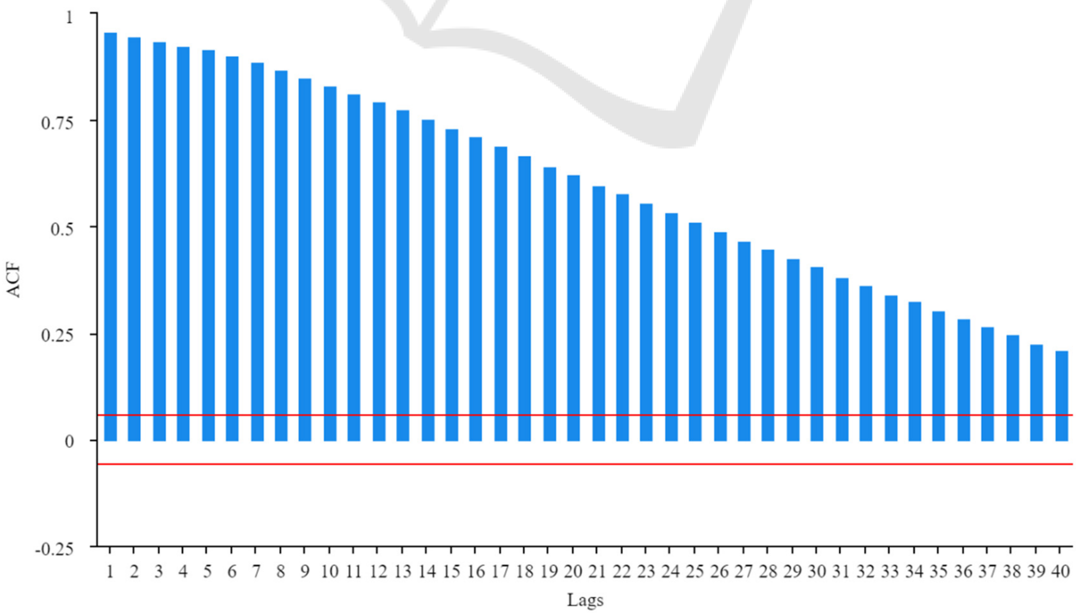

Figure 2: ACF Plot of the Traffic Flow.

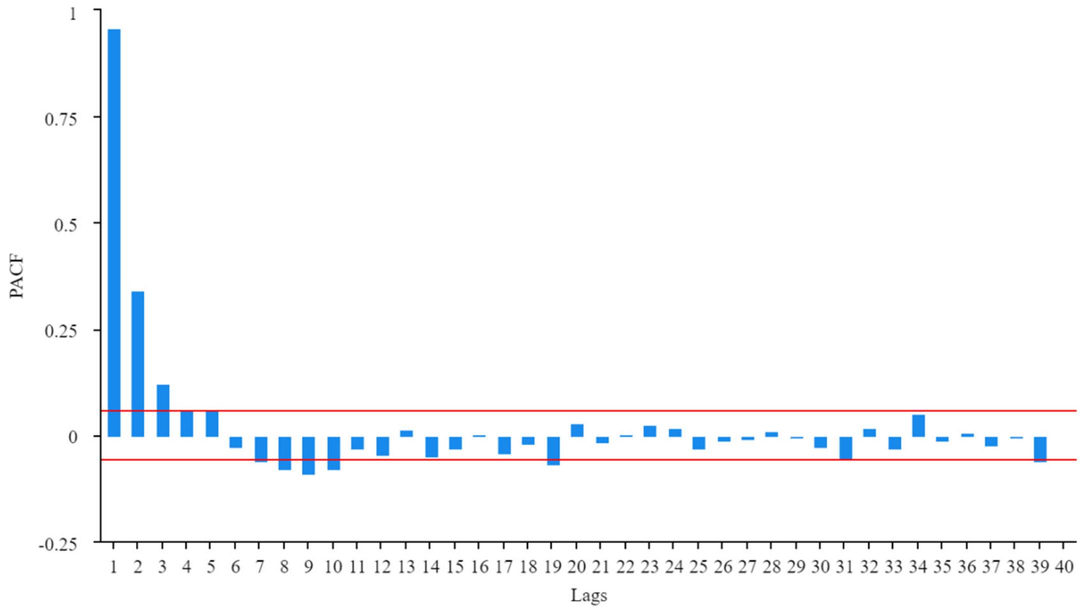

Figure 3: PACF Plot of the Traffic Flow.

Table 2: Model Evaluation.

Model	MSE	AIC	BIC
A(1,0,1)	390.660	7598.992	7618.034
A(2,0,1)	390.417	7600.323	7624.125
A(1,0,2)	390.393	7600.245	7624.048
A(1,0,0)	452.195	7722.305	7736.586
A(0,0,1)	2501.802	9207.272	9221.554

Compared with several models in Table 2, the mean square error, AIC and BIC values of A(1,0,1), A(2,0,1) and A(1,0,2) are all small with no significant difference. The mean square error of A(1,0,2) is the smallest, so it can be considered that the model has the best fitting effect. At this point, the model formula can be obtained as (Table 3):

$$y_t = 1.452 + 0.988 * y_{t-1} - 0.409 * \varepsilon_{t-1} + 0.028 * \varepsilon_{t-2} \quad (1)$$

Table 3: A(1,0,2) Model Parameter List.

Item	Sign	Coe.	S.E.	Z	P	95% CI
Constant	c	1.452	1.328	1.093	0.274	-1.151 ~ 4.054
AR	α1	0.988	0.007	137.375	0.000	0.974 ~ 1.002
MA	β1	-0.409	0.023	-18.048	0.000	-0.453 ~ -0.364
	β2	0.028	0.027	1.041	0.298	-0.025 ~ 0.080

According to Figure 4, the fitting degree of the numerical model is high and close to the distribution state of the true value. This model can meet the requirement of data fitting and prediction. Therefore, it is reliable to use A(1,0,2) to predict the traffic flow in the future period based on the existing data.

Table 4: Predicted Value.

Prediction	Value	Prediction	Value
5	75.190	80	82.481
10	75.188	85	82.957
15	75.750	90	83.427
20	76.306	95	83.892
25	76.855	100	84.351
30	77.397	105	84.805
35	77.933	110	85.253
40	78.463	115	85.696
45	78.986	120	86.134
50	79.504	125	86.567
55	80.015	130	86.994
60	80.520	135	87.417
65	81.019	140	87.834
70	81.512	145	88.247
75	81.999	150	88.654

According to Table 4, the prediction for traffic flow within the next 150 minutes indicates a gradual decline within the first 5 minutes, followed by an upward trend. By the 150th minute, the traffic flow is projected to reach approximately 88 vehicles every 5 minutes.

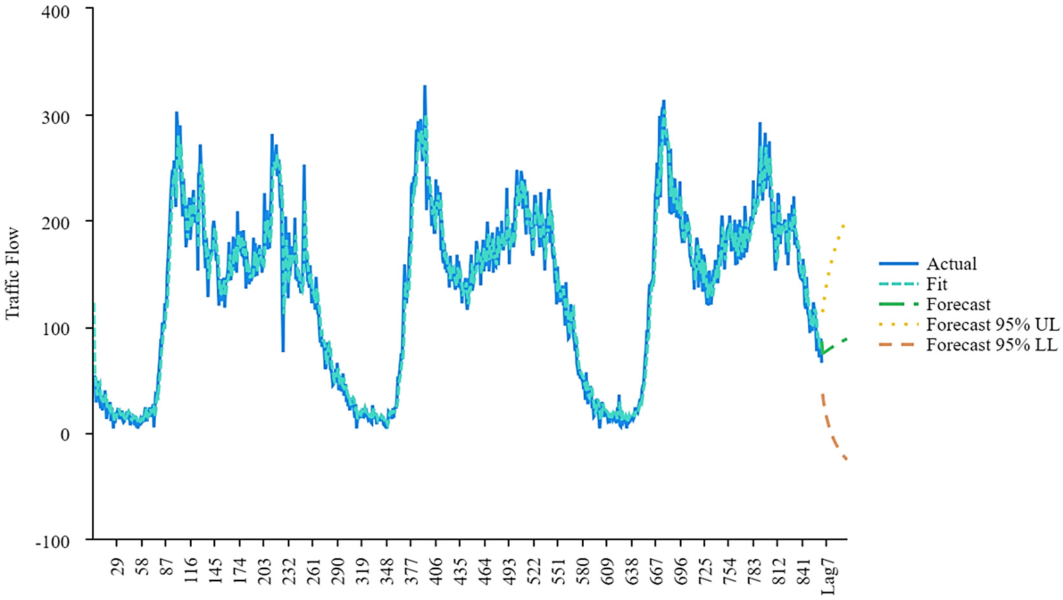

Figure 4: Fitting and Prediction of the Traffic Flow.

4 CONCLUSION

This article utilizes existing data to construct an ARIMA model for fitting and forecasting the traffic flow at a certain intersection. By selecting and comparing various parameters, the fitting results under different scenarios were examined to find the most accurate model. This demonstrates that ARIMA model is feasible in traffic flow prediction.

However, the ARIMA model also has some limitations. First and foremost, the prerequisite for the successful establishment of the model is that the data is stationary. Therefore, it is necessary to perform tests for stationarity and transform the collected data before applying the model. What's more, outliers in the data can lead to deviations in the fitting results and predicted values of the model. Therefore, it is advisable to use other models for data preprocessing to reduce the impact of outliers before employing the ARIMA model for modelling.

In conclusion, the ARIMA model possesses certain advantages and potential applications in road traffic flow prediction whereas it also has some limitations. The traffic flow predicted by this model holds promise in assisting traffic dispatching or accident early warning systems, thereby enhancing traffic efficiency or reducing accident rates. Future research could explore alternative models such as regression models or deep reinforcement learning models to enhance traffic flow prediction.

REFERENCES

Chen, C., Hu, J., Meng, Q. and Zhang, Y., 2011. Short-time traffic flow prediction with ARIMA-GARCH model, *2011 IEEE Intelligent Vehicles Symposium (IV)*, *Baden-Baden, Germany*, 607-612.

Chikkakrishna, N.K., Hardik, C., Deepika, K. and Sparsha, N., 2019. Short-term traffic prediction using SARIMA and FBPROPHET, *2019 IEEE 16th India Council International Conference (INDICON)*, Rajkot, India, 1-4.

Feng, X., Ling, X., Zheng, H., Chen, Z. and Xu, Y., 2019. Adaptive multi-kernel SVM with spatial–temporal correlation for short-term traffic flow prediction, in *IEEE Transactions on Intelligent Transportation Systems*.

Fu, R., Zhang, Z. and Li, L., 2016. Using LSTM and GRU neural network methods for traffic flow prediction, *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, *Wuhan, China,* 324-328.

Kumar, S.V., Vanajakshi, L., 2015. Short-term traffic flow prediction using seasonal ARIMA model with limited input data, *Eur. Transp. Res. Rev.* 7, 21.

Li, K. L., Zhai, C. J. and Xu, J. M., 2017. Short-term traffic flow prediction using a methodology based on ARIMA and RBF-ANN, *2017 Chinese Automation Congress (CAC), Jinan, China,* 2804-2807.

Lin, S. and Tian, H., 2020. Short-term metro passenger flow prediction based on random forest and LSTM, *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China,* 2520-2526.

Lv, Y., Duan, Y., Kang, W., Li, N. and Wang, F.Y., 2015. Traffic flow prediction with big data: a deep learning approach, in *IEEE Transactions on Intelligent Transportation Systems*, 865-873.
Mohammed, O. and Kianfar, J., 2018. A machine learning approach to short-term traffic flow prediction: a case study of interstate 64 in Missouri, *2018 IEEE International Smart Cities Conference (ISC2), Kansas City, MO, USA,* 1-7.
Zhang, W., Yu, Y., Qi, Y., Shu, F. and Wang, Y., 2019. Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning, *Transportmetrica A: Transport Science*, 1688-1711.