

Research Based on vehicle exhaust emission estimation

Xiusheng Xu^{*,#}, Laixin Zhang[#], Tianyi Tang[#]

¹ Department of Computer and Control Engineering, Shandong University, Yantai, China, 264005

* Corresponding Author Email: xuxiusheng2019@163.com

Abstract. With the development and progress of society, short-term traffic flow prediction has become an increasingly important topic. Accurate prediction of traffic flow can not only provide necessary guidance for the transportation department, but also bring convenience to travel, but also play an important role in the prevention and control of urban air pollution. This paper mainly studies the prediction of vehicle flow, and then estimates the vehicle exhaust through the calculated vehicle flow, trying to establish the estimation model of vehicle flow and vehicle exhaust. Through Gaode map interface and WGS-84 (GPS) coordinates of each bayonet provided, the road name corresponding to the bayonet equipment is obtained according to the distance vector, and the data is preprocessed. For the interval road motor vehicle flow, the dynamic prediction can be realized by using the interdependence between the time series observations. According to the time dimension of prediction, traffic flow prediction can generally be divided into long-term traffic flow prediction (in year), medium-term traffic flow prediction (in month and day) and short-term traffic flow prediction (in hour and minute). According to the data provided, due to the weak regularity of short-term traffic flow data of interval roads, strong random error interference, and high uncertainty, it is difficult to predict accurately, The long-term and medium-term traffic flow data have strong periodicity and weak random interference. Considering the accuracy and data provided, this paper mainly selects the prediction method based on endogenous variables represented by autoregressive moving average model ARIMA and time cycle neural network LSTM. Because the use conditions of the two models are different, it is necessary to improve the model and modify the use scenario of the model. The advantage of this method is that the data acquisition cost is low and easy to implement. Finally, the prediction effect of traffic flow is evaluated by cellular automata to improve the performance of the model.

Keywords: LSTM neural network Cellular automata Traffic flow prediction.

1. Introduction

In recent decades, with the continuous progress of society, the number of motor vehicles has increased exponentially. Especially in the rush hour and holidays, urban roads are seriously blocked, which increases the travel cost of residents, puts great pressure on road traffic, and causes a certain degree of harm to environmental problems. In the problem of urban air quality, motor vehicle exhaust emission is an important source of pollution. A large number of motor vehicles drive on urban roads every day, and the exhaust generated is retained and diffused in the air, affecting people's health, production and life. An accurate understanding of the scale, spatial distribution and temporal changes of motor vehicle exhaust emissions will help the transportation and environmental departments to carry out pollution control and provide a scientific basis for the prevention and treatment of urban air pollution. However, due to the limited coverage space of air monitoring stations in cities, it is difficult to comprehensively and pertinently monitor vehicle exhaust emissions for a long time. At present, how to obtain a more accurate scale and distribution of vehicle exhaust emissions is still a hot issue.

2. Data visualization

Nowadays, the coverage space of air monitoring stations in cities is limited and the performance of bayonet equipment is unequal. How to obtain a more accurate scale and distribution of motor vehicle exhaust emissions needs to be solved urgently. The data used in this experiment include the vehicle network monitoring data of Dezhou City, the weather data of Dezhou City in recent two years, the vehicle ownership data, and the emission inventory of previous years. As the main data set and

verification data of traffic flow prediction and exhaust gas estimation. Conduct corresponding data processing on the bayonet coordinates in the attachment. After checking and eliminating the abnormal data, import the longitude and latitude data of each bayonet equipment into Baidu map for marking. The corresponding position of the longitude and latitude of the bayonet equipment of the task is shown in Figure 1:

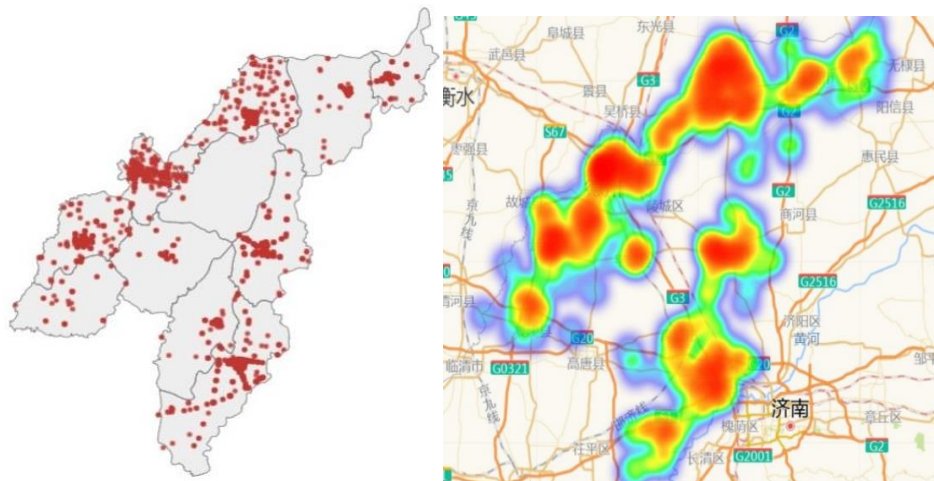


Figure 1. Bayonet equipment distribution

Figure 2. Traffic flow distribution

It can be seen from the figure that the distribution of monitoring equipment is not uniform, mainly distributed in some cities, and the monitoring distribution in some urban areas is very sparse. According to the data detected by the monitoring equipment, dynamically visualize the traffic flow change of Dezhou area in a day. Take 0:00 midnight as the starting point and gradually move forward for 24 hours. The thermal map starts to change from the state of most blue areas to the state of red areas, indicating that the traffic flow is gradually increasing. The deeper the red degree is, the greater the traffic flow here is, Figure 2 shows the traffic flow of a road at 17 o'clock.

It is observed from the figure that the main color is dark red, which indicates that the traffic flow on the road is very high at 17 o'clock. This is mainly because it is in the rush hour at 17 o'clock, and the traffic flow is relatively high. At the same time, due to the lack of bayonet equipment, the traffic flow of some sections in the figure is not displayed. As an auxiliary tool for the prediction of the whole road network segment, it is necessary to process the data of missing sections. The processing measures taken in this experiment are to drive out according to the number of branches n according to the traffic flow on the route at a certain time, and the adjacent sections also drive into the route according to the same proportion. From the calculation and Simulation of Monte Carlo, it can be concluded that the traffic flow is stable at a certain level in a short time and fluctuates gently up and down.

3. Establishment and solution of model

According to the distribution of the bayonet equipment of the task on the map, we can intuitively observe the relationship between the equipment and the road. It can be found that the bayonet equipment of the task is mainly concentrated in Decheng District, Wucheng County, Qihe County and Linyi County of Dezhou. In order to more intuitively observe the time distribution relationship of traffic flow, draw the data recorded by a bayonet equipment in the annex (equipment No.: 37142429041310000010). The corresponding relationship between traffic flow and time is shown in the figure3.

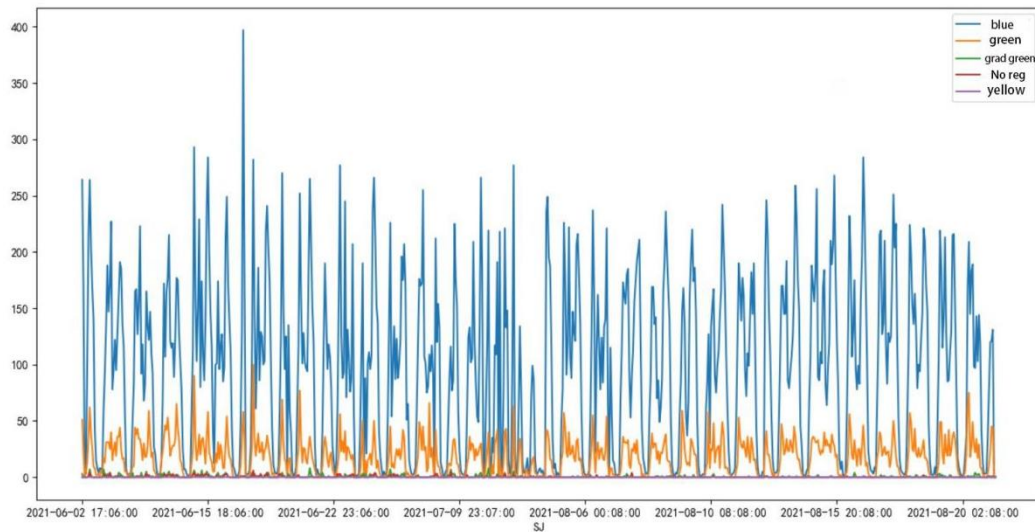


Figure 3. Equipment No.: 37142429041310000010 monitoring data

Through the data analysis of all sections, it is known that blue and green cars occupy the main body, and the number of driving in both colors is high at all times. Further clarify the above conclusions, extract the monitoring data with the equipment number of 37140000000371452055 on June 15, 2021 in the data set for visualization, as shown in the figure4:

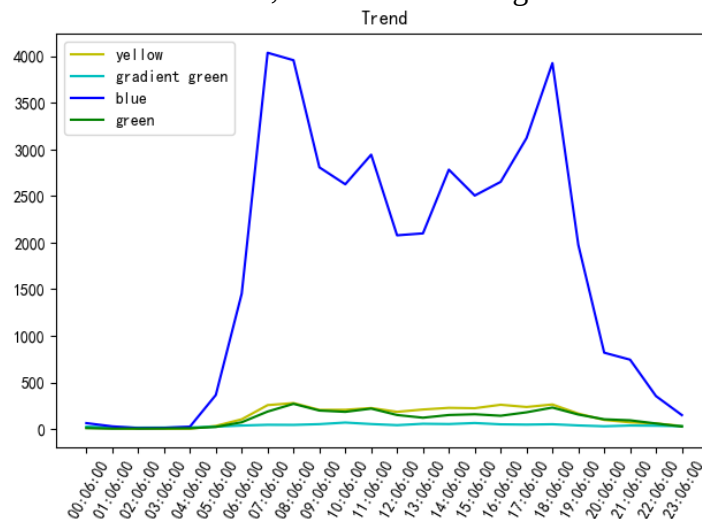


Figure 4. Traffic flow change of bayonet equipment 37140000000371452055 on June 15, 2021

3.1. data preprocessing

In order to facilitate data processing and modeling, it is necessary to process the original data. The first is the standardization of data format. Because this experiment is based on time series observation, if there are more than 12 missing values on the dates with missing values, we will use the deletion strategy to further process them, and carry out secondary classification of the data. The first classification is divided according to the given bayonet equipment number. Due to the huge amount of data, we will directly use the database to classify them, The second level of classification is divided according to the license plate color. It is mainly divided into five colors of license plates, namely blue, green, yellow, gradient green and unrecognized color license plates. The license plates with unrecognized color are further classified. According to the mathematical principle, when the sample is large enough, the mean value tends to the overall mean value. Based on this, the proportion of license plates of each color in the car ownership in Dezhou City is calculated. The roulette algorithm is used to set the priority according to the proportion of ownership, and the license plates of unrecognized color are divided into the above four known license plates.

The second is the processing of missing data. According to statistics, it is found that the data of a given data set does have a high proportion in the time dimension. Because the change of traffic flow has the characteristics of continuity in the time dimension, polynomial interpolation is used to take the data of three time dimensions before and after the missing time point dimension respectively, and polynomial interpolation fitting is carried out to finally obtain the traffic flow data that changes continuously in the time dimension.

The data style given in the data is longitude and latitude, which does not correspond to the fixed road, so the longitude and latitude are further used for road mapping. Since the longitude and latitude coordinates are GPS coordinates, Gaode API is used for positioning. Some positioning information is as follows:

Table 1. Some positioning information

ID	X	Y	Road	City
110008	117.2064	37.72785	Zhenxing West Road	Leling City
1610560	117.2186	37.71152	Fuchang East Road	Leling City
27640015	116.0321	36.98072	Victory Road	Xiajin County
3.71422160101001E+25	116.7726	37.6416	National Highway 339	Ningjin County
37140029051320000000	116.7659	36.78556	Yingbin Road	Qihe County

The bayonet equipment is further divided. According to the road location information, the equipment is divided into each road. Because the road may be located at the intersection, the traffic flow of the two roads may be detected. Therefore, the same equipment is divided into the road according to the distance from the road. If the distance is within 50m.

3.2. time series analysis

First, analyze the time series of the data and observe the change trend of the data. Whether the data changes violently or not has a great impact on the model training. If the data changes too violently with time, smooth the data to reduce the change degree.

From the randomly selected data, it can be concluded that the data changes are too sharp at some times, which may be accidental factors leading to sudden changes in traffic flow, such as weather, traffic disasters, road maintenance and other special circumstances. For this case, we do not take it into account in the model, but use the smoothing strategy to process it as smooth data. The specific methods are as follows:

The weather data of Dezhou City in the past two years are crawled and saved through the crawler framework sweep. In this data, the date when the weather is rainy is filtered out and saved as abnormal date data. The traffic flow data in abnormal date is filtered out from the traffic flow data. After comparing with other data, it is found that there are too few vehicles on some roads in the data of rainy days, but the traffic flow direction of some roads becomes abnormally large, The conjecture is that different changes are caused by different road accessibility. For road accessibility, we use TOPSIS comprehensive evaluation method to quantify it according to the road type. The evaluation goal is road accessibility. The evaluation system indicators are road type, number of branches and traffic flow. The road type refers to the types classified according to the grade of national roads and provincial roads, and the number of branches is screened and queried according to the previously climbed data of Dezhou transportation official website, Traffic flow refers to the traffic flow given in the data as the index. Generally speaking, the higher the traffic flow, the higher the road accessibility. After comprehensive evaluation, the partial accessibility results of the road are as follows:

Table 2. The partial accessibility results of the road

road	Road accessibility
Wuzhouzhong Avenue	0.274

Zhenxing Road	West	0.401
Wenchang Road	West	0.49
Fu'an Road		0.13
National Highway 513		0.87

The above indicators are all positive indicators, so they do not need to be positive, so the positive ideal solution and negative ideal solution are the maximum and minimum values of each indicator. In order to eliminate the impact of different dimensions on the evaluation results, make the evaluation indicators compare under the same dimensional system, and normalize the original data. The processing method is as follows:

$$z_{ij} = \frac{x_{ij} - \min_{1 \leq i \leq n} x_{ij}}{\max_{1 \leq i \leq n} x_{ij} - \min_{1 \leq i \leq n} x_{ij}} \quad (1)$$

The Euclidean distance between each road and positive ideal solution and negative ideal solution is calculated as follows:

$$d^+ = \sqrt{\sum_{j=1}^7 w_{ij} \times (z_{ij} - z_j^+)^2} \quad (2)$$

The relative proximity formula is as follows:

$$C_{ij} = \frac{d^-}{d^+ + d^-} \quad (3)$$

3.3. Smoothing

The above filtered data will not be processed completely, because the traffic flow is relatively small in normal weather due to the remoteness of some roads, so this part of the data will not be processed. In the experiment, the roads with road accessibility greater than 0.5 are selected for smoothing. After screening the abnormal data, the first-order difference is carried out on the data, and the difference data are counted. The calculation formula of the smoothing interval is as follows:

$$high = td[75\%] + 1.5 \times (td[75\%] - td[25\%]) \quad (4)$$

$$low = td[75\%] - 1.5 \times (td[75\%] - td[25\%]) \quad (5)$$

Where low and high are the left and right boundaries of the interval, and TD is the difference data sequence.

Calculate the index of abnormal data according to this interval, smooth these data, iterate each abnormal index data, find reasonable data consistent with the change trend of abnormal data, and replace this data as the abnormal data until all abnormal data are eliminated.

3.4. multiplication factor decomposition

The traffic flow data has obvious periodicity in the statistics of days, showing the trend of convex in the middle and concave on both sides. Therefore, the multiplication factor decomposition is used to extract the hidden trend components, periodic components and residual components. Some separation results are as follows: ^[4] after instantiating the model, divide the data set, divide the training set and test set according to the ratio of 7:3, and use the training set to train the model, Then, the test

set is used for prediction, and compared with the true value of the test set to calculate the error., The model effect is shown in the figure:

3.5. LSTM model

For the time series data, the traffic flow of a place is not only related to its adjacent nodes, but also related to the traffic flow of its adjacent time points. In this experiment, LSTM model is used, and convolution and the features extracted by LSTM are added on the basis of it for fusion to predict. It is compared with ARIMA model to observe the effect of the model. The improved network structure is shown in the figure 5: [7]

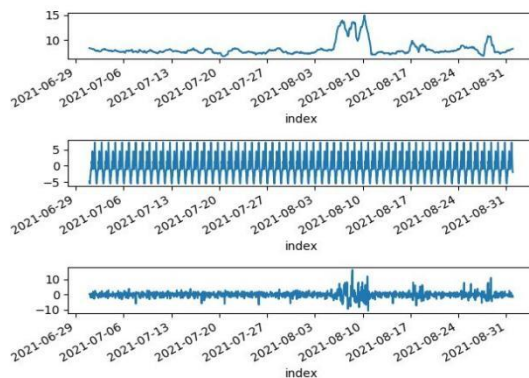


Figure 5. Decomposition results

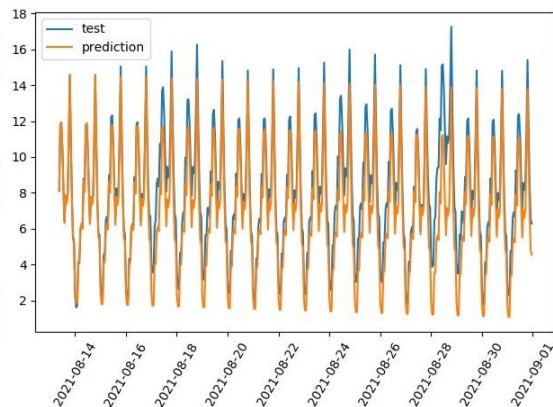


Figure 6. Model training

The observation window of this experiment is set as 2, that is, the data of the first two hours are used to predict the current data. The middle layer defines 48 nodes. The change of training error with training times is shown in the figure 7: predict the traffic flow and compare the difference between the predicted value and the real value, as shown in the figure 8: [5]

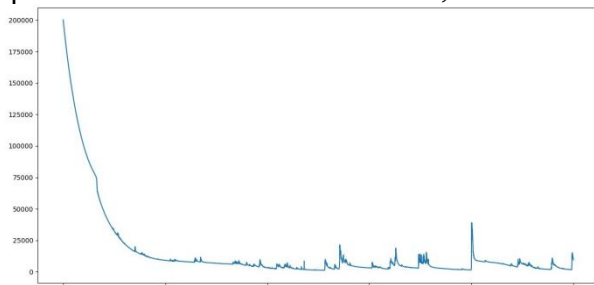


Figure 7. Training error

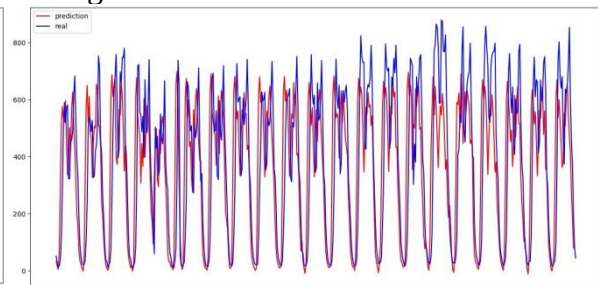


Figure 8. Comparison of real and predicted values

3.6. Cellular automata model evaluation

The evaluation of the model is simulated by cellular automata, and the following assumptions are made [2]:

Hypothesis 1: each road has n branches, and the probability of each vehicle changing at each intersection is $1/n$

Hypothesis 2: the Monte Carlo method is used to simulate the weather and road disasters. The probability of rain is the rainfall frequency of Dezhou in recent two years, which is 0.07, and the probability of road disasters is 0.01. When it rains, the vehicle chooses the road with the nearest road accessibility higher than 0.5 to drive out. In case of road disasters, the vehicle chooses the nearest intersection to drive out.

Hypothesis 3: all vehicles drive at the speed limit of the road. In case of accidents in hypothesis 2, the speed will be reduced to 0.4 times the speed limit, and the speed reduction will not be considered when changing the road. Once a change of road occurs, the speed shall be adjusted to the speed limit of the route immediately.

Hypothesis 4: a two-way road is divided into two one-way roads, and only one-way lanes are considered, that is, it is assumed that the lanes joining and leaving the route are one-way lanes.

Hypothesis 5: the initial traffic flow adopts the traffic flow predicted at 12 a.m. and gradually generates vehicles randomly into the road according to the data predicted by the model to adjust the traffic flow as time goes on.

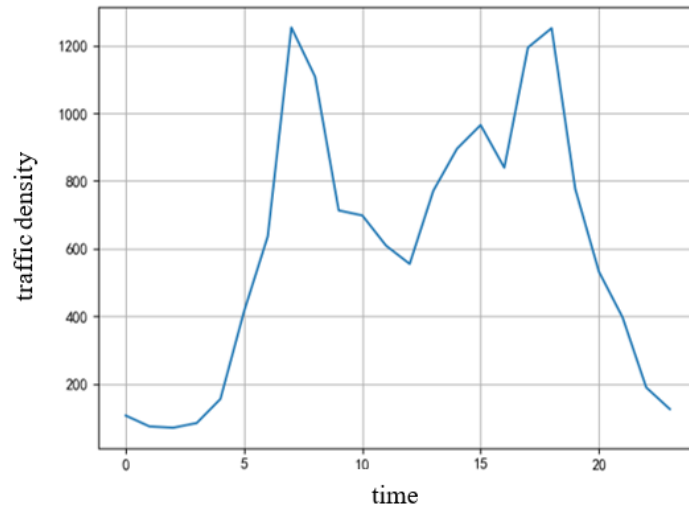


Figure 11. Traffic density change

The simulation results of cellular automata are shown in the figure11 above. It can be seen that the simulation results are more consistent with the actual route traffic flow, that is, the traffic flow predicted by the model is more accurate and has good robustness.

4. Exhaust emission visualization

Vehicle emissions are not only related to exhaust emission standards, but also related to running speed, driving distance and driving year.

First, calculate the length of the driving route. When calculating the length of a route, use ArcGIS to analyze, find the two bayonet devices located on the same road with the farthest distance between longitude and latitude coordinates, and convert them into geographical distance as a rough estimate of the length of the route. The driving speed is simulated according to the road speed limit, and the driving year is determined. Here, it is assumed that the vehicle age follows the normal distribution in the car ownership of Dezhou City, The following formula is obeyed:

$$p = \frac{1}{\sqrt{2\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (6)$$

4.1. Tail gas estimation

According to the above assumptions, the results can be calculated based on the number of vehicle age segments in driving.

According to Hypothesis 5, the estimated amount of road exhaust emission at a certain time can be obtained, and the calculation formula is as follows: ^[3]

$$Z = N \times \frac{S}{V} \times \sum_{i=1}^4 p_i \times \int_0^{15} \frac{1}{\sqrt{2\sigma}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} (k_i t + e_i) dt \quad (7)$$

In the above formula, Z represents the estimated amount of exhaust emission, n represents the predicted traffic flow, s represents the road length estimated according to the bayonet equipment, V represents the driving speed, P represents the proportion of ownership corresponding to each vehicle type, K represents the linear growth coefficient between the limited displacement and the age of the vehicle, e represents the initial emission standard, T represents the age of the vehicle, and only the

exhaust emission of vehicles aged less than one year to 15 years is calculated in this experiment. Some calculation results are as follows:

Table 3. Some calculation results

road	time	Emissions
National Highway 105	00:00	18.56
National Highway 309	11:00	105
Wenchang West Road	13:00	24.81
Fuyuezhuang Street	19:00	15.74
Xiao Li Lu	21:00	0.26

5. Conclusion

The exhaust emission model comprehensively considers the indicators that have an obvious impact on vehicle exhaust emission. It has strong generalization performance and is suitable for long-term prediction of exhaust.

Using the monitoring data of simple mechanical equipment to process, convert and predict, the accurate estimation of tail gas emission at low cost is realized, and the financial and material costs are saved

The model is completely based on the endogenous variables that affect the change of traffic flow. It has low requirements for the universality of the data, realizes the full use of the data, and can accurately understand the scale, spatial distribution and time change of vehicle exhaust emission, which is helpful for the transportation and environment departments to carry out pollution control and provide a scientific basis for the prevention and control of urban air pollution

6. Reference

- [1] Yu Yongbo and Cheng Xiaoming Study on exhaust emission prediction of hybrid oil electric bus In 2018 China Annual Conference on urban transport planning, pages 1859 – 1868
- [2] Liu Tianyang Research on simulation modeling of left turn traffic flow on main road of unsignalized intersection based on cellular automata Master, 2021
- [3] Zuo Fushan, Li Zhengyuan, LV Xiao, and Zhang Ying Prediction of gasoline engine exhaust emission based on BP neural network Journal of Jiangsu University (NATURAL SCIENCE EDITION), 41 (03): 307 – 313, 2020
- [4] Yang Donglong Research on short-term traffic flow prediction of interval roads based on ARIMA model Electronic design engineering, 29 (13): 38 – 42, 2021
- [5] Guo Pengyu Research and application of trajectory and traffic flow prediction Master, 2021
- [6] Lu Fei Research on vehicle detection and traffic flow analysis in complex environment based on deep learning Master, 2021
- [7] Wei Jian, Zhao Hongtao, Jia Heping Short term traffic flow prediction based on improved LSTM model Technological innovation and application, 11 (12): 25 – 27, 2021
- [8] Guo Pengyu Research and application of trajectory and traffic flow prediction Master, 2021
- [9] Lu Fei Research on vehicle detection and traffic flow analysis in complex environment based on deep learning Master, 2021
- [10] Wei Jian, Zhao Hongtao, Jia Heping Short term traffic flow prediction based on improved LSTM model Technological innovation and application, 11 (12): 25 – 27, 2021