Construction of vehicle driving cycle in Fuzhou

Sijia Zou¹, Xinru Zhu¹, Linjun Xu¹ and He Zhu¹*

¹School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, Zhejiang, 310018, China

*Corresponding author’s e-mail: 593019700@qq.com

Abstract. The current driving cycle in Europe is not consistent with China’s actual road conditions, and the traffic conditions in different cities of China also are discrepant. Therefore, it is increasingly urgent to build corresponding driving cycle based on the driving data of each city. In this paper, based on noise reduction pretreatment of vehicle driving data in three areas of Fuzhou city, K-means clustering is carried out on the first five principal components of the extracted kinematic segments obtained by principal component analysis. Then, three types of kinematic segments, which represent relatively smooth traffic, traffic congestion and smooth traffic, are obtained respectively. So the vehicle driving cycle is constructed, and the correlation coefficient between it and the sampling population for constructing is calculated. Finally, the error analysis method and correlation coefficient are used to verify the constructed driving cycle that is rational and representative.

1. Introduction

The driving speed of a certain type of vehicle changes with time under a specific time and space state, which is generally called vehicle driving cycle. The vehicle driving cycle is commonly used to measure the fuel consumption and emissions from vehicles, and is the core technology of the vehicle industry. Transient driving cycle FTP in the United States, modal driving cycle NEDC¹⁰ in Europe and JAPAN¹⁰-¹⁵ in Japan are the typical driving cycles, which are widely used in the world²-⁴. However, compared with developed countries such as Europe, America and Japan, China's research time on road driving conditions is relatively short and is still in its infancy. Moreover, the traffic conditions in different cities of China also are discrepant. Our country mostly uses European driving conditions as the test cycle⁵, but the actual road conditions in our country are quite different from those in Europe. In a word, it is of great significance to study the vehicle driving cycle in line with the traffic characteristics of various regions in China.

In recent years, domestic scholars have carried out a series of research on driving cycles and made some achievements. Shi Qin et al.⁶-⁷ used combined clustering method, FCM clustering method and principal component analysis to study the actual road conditions in Hefei City and established typical driving conditions in Hefei urban area. Hu Chen et al.⁸ adopted K-means clustering algorithm and principal component analysis to construct the driving conditions of vehicles in Harbin city. Huang Wanyou⁹ studied the driving conditions of pure electric vehicles through a large number of test data, and further constructed the driving conditions of small pure electric vehicles in Jinan. Miao Qiang et al.¹⁰ use Markov chain to synthesize and select representative working conditions in the class to form typical driving conditions. The results show that the overall characteristic parameters of bus driving conditions generated by clustering and Markov chain method have lower relative error and better effect. Wan Xia et al.¹¹ constructed the driving conditions of passenger cars in Shenzhen urban area by short-
stroke analysis. The results show that the vehicle emission conditions are better than those in other cities in China.

This paper is based on the actual road driving data of light vehicles in three regions of Fujian Province. The data comes from the data acquisition system developed by China Automotive Technology Research Center (CATC). Our study uses cross-region data on the city level for 3 regions. Among them, the data of these three regions are: the 185,725 records recorded on December 24, 2017 near Cangshan District of Fuzhou City of Fujian Province, 145,825 records of November 7, 2017 near Xiangan Avenue in Xianyou County of Putian City of Fujian Province, and the 164,914 records of December 6, 2017 near Huli Road in Mawei District of Fuzhou City of Fujian Province. Intending to construct a vehicle driving cycle, which can reflect the driving characteristics of vehicles participating in data collection, through processing abnormal data, extracting motion segments, calculating motion characteristics, and utilizing the characteristics of average speed and average acceleration of vehicle motion. The aim is to provide certain reference value for road design.

2. Data preprocessing

2.1. Time discontinuous data preprocessing

During the driving process of the car, GPS signals are often lost due to high-rise building coverage or tunnel crossing, resulting in discontinuous missing of time in the collected data.

![Cubic spline interpolation and linear interpolation comparison chart](image1)

![Comparison chart before and after glitch data elimination](image2)

Here, for all the adjacent data points whose time interval is more than 1s and the speed is greater than 0, if the time interval of the identified two adjacent data points is not more than 600s, and the absolute value of the velocity difference between the two data points is less than 20km/h, the missing relevant data are completed by cubic spline interpolation algorithm. Other time discontinuity data that do not belong to the above situation are not inserted.

Figure 1 is based on the data processing in Cangshan district. It can be clearly seen that the data curve interpolated by cubic spline interpolation for discontinuous time data is much smoother than the data curve interpolated by piecewise linear interpolation. Then the effect is better, and it is more in line with the actual driving condition of the car.

2.2. Abnormal data preprocessing collected by long-term parking

Under the condition of long-term parking, the data, which is also called burr data, has non-zero speed. For the processing of burr data, the data is segmented according to the time discontinuity points of the data. What’s more, the time node data with continuous movement time less than or equal to 3s in idle state is determined as idle data, and its speed is filled with 0, while other abnormal data that do not belong to the above conditions are not processed.

As can be seen from Figure 2, after the burr data is processed into idle speed data, the driving segments with the running time of the vehicle within the time interval not exceeding 3s are greatly reduced. Thus the vehicle driving data have better consistency.
2.3. Preprocessing of idle speed data for long-term traffic jams and intermittent low-speed driving

According to the processing of idle speed data of long-time traffic jam and intermittent low-speed driving, it is defined that the vehicle speed is lower than 10km/h when driving at low speed. If the driving time at low speed exceeds 60s, the long-time condition can be considered satisfied.

![Comparison chart before and after low speed driving data](image1)

According to the time break point of the data, the data is segmented. Furtherly, subintervals in which the maximum vehicle speed for more than 60s is less than 10km/h for each interval spanning more than 120s is identified, and the idling condition is determined. Furtherly, all the speeds in the subintervals are setted to 0. In each continuous time segment data, only the 180s data at the end of idle state is reserved for the part of each interval data where idle time exceeds 180s. Other long-term traffic jams and intermittent low-speed driving data that do not belong to the above situation are not processed.

The Figure 3 shows that after the data satisfying the above condition is determined as the idling condition data, the segment in which the vehicle is intermittently driven at a low speed for a long period of time is greatly reduced, so that the vehicle traveling data have better continuity.

![Contrast chart before and after cusp data processing](image2)

2.4. Preprocessing of abnormal data of vehicle acceleration and deceleration

The cusp data generally is the data with excessive instantaneous acceleration. Here, the maximum acceleration is 3.97m/s$^2$. For the processing of abnormal data of vehicle acceleration and deceleration, the acceleration is calculated firstly. Then, the cusp data is processed smoothly by the cubic spline interpolation algorithm. Finally, the data of the vehicle decelerate exceeding the deceleration threshold at a certain moment is eliminated. The acceleration of the vehicle itself based on the obtained data is calculated. And the calculation formula is as following.

$$a_{t+1,t} = \frac{(V_{t+1} - V_t) / 3.6}{T_{t+1,t} - T_t}$$

![Figure 4. Contrast chart before and after cusp data processing](image3)

According to the time discontinuity of the data, the data is segmented by Python. Here, the deceleration threshold is between 7.5 and 8 m/s$^2$, and the acceleration threshold is 3.97 m/s$^2$. If the acceleration in the data exceeds the acceleration threshold, that is the cusp data. And cubic spline interpolation algorithm is adopted to smooth it. If the deceleration in the data exceeds the deceleration threshold, the elimination process is carried out.

The Figure 4 shows that the data is more in line with the actual driving situation of the vehicle after the cusp data smoothed by cubic spline interpolation algorithm. And it also makes the kinematic fragments extracted later more reasonable.

2.5. Extraction of kinematic fragments

The kinematic segment refers to the range of vehicle speed from the start of the idle state to the start of the next idle state. It is one of the most commonly used methods to construct the driving cycle curve of vehicle based on kinematic segments.

According to the time discontinuity of the data, that is two adjacent data points with time interval greater than 1s, the data is segmented and divided into multiple continuous time interval data. Then, the kinematic segments are extracted from each interval data. Firstly, each idle state of starting point in a
certain interval data is identified, and it is used as a cutting point for dividing kinematic fragments. Then, two adjacent between two starting point in the idle state data of the speed data is divided into a piece of kinematic fragment. In this way, all the kinematic fragments in the interval data of the files, which match the definition, are extracted.

After the data of Cangshan District, Xiangang Avenue and Mawei District are extracted and processed, the number of kinematic fragments finally are respectively 1,077, 899 and 741.

3. Construction of the vehicle driving cycle

3.1. Construction of the vehicle driving cycle curve

3.1.1. Calculation of eigenvalues

Based on the business characteristics, this paper selects some characteristic parameters to describe the kinematic segments, including car driving distance (m), car driving time (s), acceleration time (s), deceleration time (s), constant speed time (s), idle time (s), average speed (km/h), average driving speed (km/h), maximum speed (km/h), maximum acceleration (m/s²), minimum acceleration (m/s²), average acceleration (m/s²), average deceleration (m/s²), speed standard deviation (km/h) and acceleration standard deviation (m/s²).

At the same time, the characteristic parameters describing the overall distribution are selected, including the idle time ratio (%), the acceleration time ratio (%), the deceleration time ratio (%), the uniform time ratio (%), \( P_{0-10} \), \( P_{10-20} \), \( P_{20-30} \), \( P_{30-40} \), \( P_{40-50} \), \( P_{50-60} \), \( P_{60-70} \) and \( P_{70} \). The meaning of \( P_{0-10} \) indicates the ratio of speed between 0 and 10km/h, and \( P_{70} \) indicates the ratio of speed greater than 70km/h. Then Python is used to calculate the eigenvalues of each kinematic segment in the data file.

3.1.2. Principal component analysis

Since there is still a large amount of data after data preprocessing and each characteristic parameter is not completely independent, the principal component analysis method is adopted here to conduct the dimensionality reduction processing for all the characteristic parameters.

According to the gravel graph drawn by Python and the cumulative contribution rate of variance, the contribution rate of the first five principal components is as high as 89.06%. Hence, the majority of the information of the initial characteristic parameter variable set can be reflected by retaining only five principal components.

3.1.3. Cluster analysis

When the amount of data is very large, K-means clustering method can efficiently classify and process a large amounts of data sets. Therefore, it is adopted to classify the kinematic segments extracted.

Firstly, according to the actual vehicle driving experience, the initial condensation points are set to 3. Then, based on the selected characteristic parameter variables, Euclidean Distance is used to compare kinematic segments with three clustering center distance, and all the kinematic segments are divided into three categories.

![Figure 5. Kinematic fragment clustering result](image)

The first category, which represents a relatively smooth traffic condition, and includes 1,321
kinematic segments. Its average speed is 24.21km/h and its idle time ratio is 23.40%. The second category, which represents traffic congestion condition, and includes 1,284 kinematic segments. Its average speed is 5.53km/h and its idle time ratio is 46.36%. The third category, which represents smooth traffic condition, and includes 112 kinematic segments. Its average speed is 43.94km/h, and its idle time ratio is 4.64%.

3.2. Vehicle driving cycle

3.2.1. Construction steps

Firstly, the number of kinematic segments needed to fit the driving cycle of the vehicle is determined. Then, the correlation coefficients between the sample parameter values in each type of sample and the characteristic parameter values used to represent the kinematic segments of each sample are calculated respectively, and the threshold value of the correlation coefficient selected here is 0.9. If its absolute value does not exceed 0.9, it is eliminated. And the formula is as following.

\[ \varphi_j = \frac{\sum_{i=1}^{N} t_j \times 1200}{\sum_{i=1}^{N} t_j} \]  

The number of kinematic segments required for fitting the real driving cycle of the vehicle by using the remaining kinematic segments is calculated according to the formula, the new segments are randomly combined. The characteristic parameter values of the long segment and its correlation coefficient with the eigenvalues representing the kinematic segments are calculated. Thereby the maximum combination of correlation coefficients is selected, and the final driving cycle of the vehicle is constructed based on it. Finally, the most representative combinations of the selected kinematic segments are extracted together to construct the driving cycle of the vehicle which are consistent with the actual driving cycle of the vehicle.

3.2.2. The first type of vehicle driving cycle

As shown in Figure 6, in such cycle, the continuous driving time to maintain the vehicle in the medium speed driving state is 1,300s, with a total of 5 kinematic segments, and the average duration of the kinematic segments is 260s. The number of vehicle stops is less, and the continuous driving time of the vehicle is longer. Therefore, it has obvious traffic smoother characteristics.

![Figure 6. The first type vehicle driving cycle](image)

3.2.3. The second type of vehicle driving cycle

As shown in Figure 7, in such cycle, the time to maintain the vehicle running continuously at low speed is 1,263s, with a total of 26 kinematic segments. The average duration of the kinematic segments is 48 seconds, and the driving time of the vehicle is relatively short. Moreover, the maximum driving speed of the vehicle in this kind of driving cycle is only 19.2km/h, and the vehicle runs relatively slowly. Therefore, it has obvious traffic congestion characteristics.

![Figure 7. The second type vehicle driving cycle](image)
3.2.4. The third type of vehicle driving cycle

As shown in Figure 8, in this kind of cycle, the time to maintain the car running continuously at low speed is 1,265s, there are two kinematic segments. The average duration of the kinematic segments is 632s. The driving time of the car is longer, the highest driving speed in this kind of cycle is 69km/h, and the average speed is 45.38 km/h. Therefore, compared with the first and second types of cycle, it shows higher average driving speed and driving time, which has obvious traffic flow characteristics.

The correlation coefficients between the three types of vehicle driving cycles constructed and the sampling population constructed for relevant type of cycles are respectively 0.9986, 0.9921 and 0.9880. Therefore, it is reasonable to construct the three types of vehicle driving cycles to describe the real vehicle driving cycles.

![Figure 8. The third type vehicle driving cycle](image)

3.2.5. Overall vehicle driving cycle

As can be seen from the foregoing, four kinematic segments with lengths of 68s, 218s, 136s and 152s are extracted from the first type of kinematic segments, four kinematic segments with lengths of 69s, 72s, 66s and 60s are extracted from the second type of kinematic segments. One segment with lengths of 448s is extracted from the third type of kinematic segments. And all of them are put together to construct a vehicle driving cycle with a total length of 1,289s.

As can be seen from Figure 9, when the car is running in this cycle, the idling time ratio is 22.42%, the acceleration time ratio is 25.89%, the deceleration time ratio is 21.56%, and the uniform speed time ratio is 30.14%. The correlation coefficient is 0.9979 between the constructed vehicle driving cycle and the sampled population of the constructed vehicle driving cycle. Therefore, it is reasonable to describe the real vehicle driving cycle by the constructed vehicle driving cycle.

![Figure 9. Overall vehicle driving cycle](image)

3.3. Construction of evaluation system of vehicle motion characteristics

The constructed driving cycle of the vehicle is constructed by extracting a small number of kinematic segments from a large number of kinematic segments, and fitting construction of the driving cycle of the vehicle by using the extracted small number of kinematic segments is realized.

| Characteristic Parameters | Characteristic Parameters | Characteristic Parameters | Characteristic Parameters |
|---------------------------|---------------------------|---------------------------|---------------------------|
| driving cycle             | driving cycle             | driving cycle             | driving cycle             |
| Vms                       | Vm                        | a_p                       | a_q                       |
| 33.991                    | 26.766                    | 0.443                     | -0.522                    |
| 35.690                    | 27.690                    | 0.924                     | -0.532                    |
| 1.699                     | 0.924                     | 0.020                     | 0.011                     |
| 4.759                     | 3.337                     | 4.703                     | 2.037                     |
| Tw                        | P_10                      | P_10-20                   | P_20                      |
| 31.420                    | 0.372                     | 0.063                     | 0.078                     |
| 30.140                    | 0.350                     | 0.108                     | 0.119                     |
| 1.279                     | 0.022                     | 0.045                     | 0.041                     |
| 4.245                     | 6.195                     | 41.591                    | 34.648                    |
| Vm                        | P_10                      | P_10-30                   | P_30                      |
| 22.850                    | 0.154                     | 0.078                     | 0.090                     |
| 26.232                    | 0.123                     | 0.119                     | 0.069                     |
| 3.382                     | 0.031                     | 0.057                     | 0.021                     |
| 12.894                    | 25.605                    | 51.080                    | 29.595                    |
| a_p                       | P_30                      | P_30-40                   | P_40                      |
| 0.425                     | 0.168                     | 0.016                     | 0.069                     |
| 0.464                     | 0.111                     | 0.069                     | 0.069                     |
| 0.039                     | 0.057                     | 0.021                     | 0.021                     |
| 8.448                     | 51.080                    | 29.595                    | 6.195                     |
| T_h                       | P_50                      | P_50-60                   | P_60                      |
| 21.257                    | 0.090                     | 0.090                     | 0.090                     |
| 22.415                    | 0.069                     | 0.069                     | 0.069                     |
| 1.158                     | 0.021                     | 0.021                     | 0.021                     |
| 5.167                     | 29.595                    | 6.195                     | 4.245                     |

Therefore, the characteristic parameters, which contain average speed, standard deviation of speed, standard deviation of acceleration, idle speed time ratio, acceleration time ratio, deceleration time ratio,
uniform speed time ratio, \(P_{0-10}, P_{10-20}, P_{20-30}, P_{30-40}, P_{40-50}, P_{50-60}, P_{60-70}\) and \(P_{70}\). And they are selected to calculate the relative error and absolute error of the vehicle driving cycle and the whole sampling data source respectively, so as to evaluate the representativeness of the constructed cycle. It is assumed here that the relative error is less than 15%, which is within an acceptable and reasonable range.

From Table 1, it can be seen that the characteristic parameter variable with the largest absolute error value is the velocity standard deviation, its specific value is 12.89%, and its value is less than 15%. So the reliability is high. Considering all the relative errors, the average relative error is 4.905% after removing the relative errors of variables other than driving speed greater than 10km/h. Therefore, the constructed driving cycle is reasonable.

4. Conclusions
The kinematics segments extracted from the noise reduction preprocessed data are more reasonable and representative. To this end, based on the three kinematics segments of K-means clustering, the driving cycle of vehicle in Fuzhou is constructed, and its rationality is verified furtherly. The innovations in this paper include that, the time discontinuity data and cusp data are smoothed by cubic spline interpolation algorithm, and the abnormal deceleration data is eliminated. In addition, the long time idle data is processed with 0. The results show that the constructed driving cycle of vehicle is reasonable, and can truly reflect the driving characteristics of vehicles participating in data collection in Fuzhou.

References
[1] Andre, M. (1996) Driving cycles development: characterization of the methods. SAE Technical Paper, 06: 103-108.
[2] Yang, Y.X., Cai, X.L., Du, Q., et al. (2002) Research on the Driving Condition of Road Cars in Tianjin. Automotive Engineering, 03: 200-204.
[3] Bata, R., Yacoub, Y., Wang, W., et al. (1994) Heavy duty testing cycles: survey and comparison. SAE transactions, 07: 450-462.
[4] Andre, M., Hickman, AJ, Hassel, D., et al. (1995) Driving cycles for emission measurements under European conditions. SAE transactions, 06: 562-574.
[5] Huang, W.Y. (2012) Research on matching technology of pure electric vehicle powertrain system. Shandong University.
[6] Shi, Q., Qiu, D.Y., Wu, J. (2012) Study on Driving Cycles based on Principal Component Analysis and Fuzzy C-M clustering. Research of Environmental Sciences, 25: 70-76.
[7] Shi, Q., Qiu, D.Y., Zhou, J.Y. (2012) Driving Cycle Construction and Accuracy Analysis Based on Combined Clustering Technique. Automotive Engineering, 34: 164-169.
[8] Hu, C., Wu, X.G., Li, X.J. (2014) Construction of Harbin city bus cycles. Harbin Polytechnic University, 19: 85-89.
[9] Huang, W.Y., Cheng, Y., Li, W. (2012) Driving Cycle of City Road Based on Vehicles Energy Consumption in Jinan. Journal of Southwest Jiaotong University, 47: 989-995.
[10] Miao, Q., Sun, Q., Bai, S.Z., et al. (2016) Construction of Typical Driving Cycle of Bus Based on Clustering and Markov Chain. China Journal of Highway and Transport, 29: 161-169.
[11] Wan, X., Huang, W.W., Qiang, Mi.M. (2016) Construction of Driving Cycle for passenger vehicles in Shenzhen. Journal of Shenzhen University Science and Engineering, 33: 281-287.