# An Analysis Model of Automobile Running State Based on Neural Network

Jie Yan[1], Xinxin Guan[2], Qingtao Zeng[1](✉), Chufeng Zhou[3], Yeli Li[1], and Fucheng You[1]

[1] Beijing Institute of Graphic Communication, Beijing, China
zengqingtao@bigc.edu.cn
[2] National Museum of China, Beijing, China
[3] China Women's News, Beijing, China

**Abstract.** A reasonable design of the operating condition curve of automobile running state is conducive to improving the credibility of the government, so it is more and more important to formulate a test condition that reflects the actual road driving conditions in China. The actual fuel consumption is very different from the regulatory certification results. In order to construct the model mainly by two-segment clustering, the initial clustering of the processed data is carried out by self-organizing mapping neural network, and the cluster number and clustering center are obtained to solve the problem of poor convergence in the K-means model in the early stage. In view of the construction of the operating condition curve of the driving characteristics of light vehicles in a city, the data pre-processing, the extraction of motion fragments and the construction of the driving conditions of a car are to be provided for the driving data set of the same vehicle in a city.

**Keywords:** Automobile running state · Self-organizing mapping neural networks · K-means model · Two-segment clustering

## 1 Introduction

### 1.1 Background

In the early years, because China's development in the automotive industry is still relatively backward, so in the domestic automotive products for energy consumption or emission certification, choose to use the European NEDC (New European Drivig Cycle) driving conditions to complete the car certification, such a practice for China's automotive energy conservation and emission reduction and technology development has provided a lot of impetus to promote the development of China's automotive industry.

Subsequently, China's economic development is more and more rapid, the national economy continues to grow, China's car security also began to rise, gradually our country has been using NEDC driving conditions began to do not conform to China's national conditions, the use of NEDC conditions as the benchmark to optimize the calibration of the car, the actual situation and NEDC operating conditions of the data deviation is


Y. Weng et al. (Eds.): TridentCom 2020, LNICST 380, pp. 83–94, 2021.
https://doi.org/10.1007/978-3-030-77428-8_7

growing. Not only that, Europe and the United States in many years of development and practice also found that NEDC conditions of many shortcomings, in order to solve the problem, Europe and the United States began to implement the use of the world light vehicle test cycle. But even WLTC conditions are not realistic for our country. The two most important characteristics of this condition: idle time ratio and average speed, and China's actual car driving conditions there is a big difference. On the other hand, China's vast geographical area, the degree of development of each city, climatic conditions and traffic conditions are not the same, so that the characteristics of the car driving conditions in each city there are obvious differences. Mixed traffic consisting of non-motor vehicles such as motor vehicles and bicycles is the main feature of urban traffic in my country. According to statistics, non-motor vehicles, mainly bicycles, account for 25% to 55% of the total traffic. Because of their different driving behavior, speed and other basic characteristics, mixed traffic has a certain impact on urban road driving conditions. In my country's mixed traffic, motor vehicles and non-motor vehicles, fast vehicles and slow vehicles may run in the same lane and interfere with each other. Therefore, it is urgent to pass in-depth research, to develop in line with the actual road driving conditions of China's cars test conditions, as the automotive industry vehicle development and evaluation of the basis.

### 1.2 The Current State of Research at Home and Abroad

Domestic and foreign scholars have conducted a lot of research on how to construct local car driving conditions. The proposed construction methods mainly include short-stroke method, clustering method and Markov method. The short-stroke method uses short-stroke as the basic unit, and randomly combines all short-strokes. The random-combined short-stroke constitutes candidate operating conditions. The method of feature parameter evaluation is used for candidate operating conditions, and the candidate with feature parameters closest to the experimental data is selected. The conditions are representative driving conditions.

Foreign researchers have done a lot of research on driving conditions in some areas: KS Nesamani of the University of California and others established the driving conditions of urban buses in Chennai, India based on the data collected using GPS [1]; Nanyang, Singapore Sze-Hwee Ho and others at the Polytechnic University used the vehicle tracking method to construct vehicle driving conditions that are more in line with the actual road conditions in Singapore [2]; Matjaz Knez and others at the University of Mariol in Slovenia used the Tango GPS program to measure the importance of the actual driving of the vehicle. Parameters, and thus developed the driving conditions of the small town Celje in Slovenia [3]. There are also many achievements in the construction of actual driving conditions in China: Li Ning of Hebei Agricultural University and others used the short-stroke method to construct the road driving conditions of Tianjin through principal component analysis and cluster analysis [4]; Hefei University of Technology Qin, Ma Honglong and others used SOM network to cluster the principal components, and used the obtained weights as the initial clustering center of FCM clustering to construct the road driving conditions in Hefei [5]; Cai E, Li Yangyang of Chang'an University, etc. People constructed road driving conditions in Xi'an based on the K-means clustering algorithm [6].

## 2 Related Research

### 2.1 Self-organizing Map Neural Network

Model Overview. This modeling uses a neural network model, self-organizing map neural network (SOM), abbreviated as SOM. The SOM neural network model is mainly a method that can represent high-dimensional data in a low-dimensional space (usually one-dimensional or two-dimensional). The process of reducing the dimensionality of a vector is called vector quantisation. In addition, the SOM network can maintain the topological relationship of the original data. The self-organizing mapping neural network can map high-dimensional data to low-dimensional space through the neural network according to the characteristics of the sample and the internal rules, so as to achieve the purpose of dimensionality reduction and clustering. The structure diagram is shown in Fig. 1.

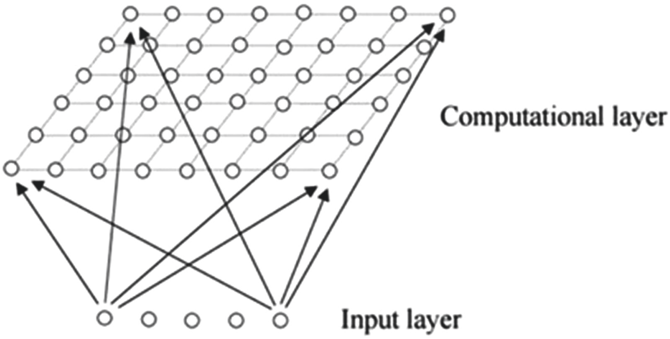


**Fig. 1.** SOM neural network structure diagram.

The link between the input layer and the output layer is primarily through the weight vector. The SOM neural network input layer corresponds to the input vector of the sample; Output Neuron nodes are widely connected to other nodes in the neighborhood, competing with each other for activation. At the same time only one neuron node is activated and the other neuron nodes are suppressed. This activated neuron node, called the winning unit, updates the weight of the winning unit and its adjacent regions, allowing the output node to maintain the topological characteristics of the input vector. However, the traditional SOM neural network still has some disadvantages, the number of neurons in the competitive layer needs to be pre-defined, the initial value of the right vector is randomly generated, this network structure restriction greatly affects the convergence speed and learning effect of the network.

### 2.2 Overview of the K-Means Model

The k-means clustering algorithm is a clustered analysis model of an iterative solution proposed by James MacQueen. The so-called clustering, that is, according to the principle of similarity, according to the degree of similarity divided into clusters, this is an unservesed process, the data object to be processed without any prior knowledge, prior knowledge is an indispensable part of the experience.

For the k-means model, there are two ways to terminate an iteration: one is to set the number of iterations T, and when the second iteration is reached, the iteration is terminated, at which point the resulting class cluster is the final clustering result [7].

The k-means model measures similarity between data objects, usually at Euclidean distance. Euclidean distance is calculated as follows:

$$dist\left(x_i, x_j\right) = \sqrt{\sum\nolimits_{d=1}^{D} \left(x_{i,d} - x_{j,d}\right)^2} \tag{1}$$

Where D represents the number of properties of the data object. $dist\left(x_i, x_j\right)$ represents Euclidean distance from $x_i$ to $x_j$. $x_{i,d}$ represents the distance from xi to the nearest cluster center. $x_{j,d}$ represents the distance from $x_j$ to the nearest cluster center.

k-means model in the process of clustering, the corresponding class cluster center because of continuous iteration re-update, corresponding to the class.

The average of all data objects in a cluster, that is, the center of the class cluster of the updated cluster. Defines the center of the class cluster for the kth class cluster is $Center_k$.

Updated as a formula (2).

$$Center_k = \frac{1}{|C_k|} \sum\nolimits_{x_i \in C_k} x_i \tag{2}$$

Where $C_k$ represents the kth class cluster; $C_k$ represents the number of data objects in the kth class cluster. The leveling here refers to the level of all elements in class cluster $C_k$ on each column property, so $Center_k$ is also a containing D.

Vector of properties, expressed as (3).

$$Center_k = \left(Center_{k,1}, Center_{k,2} \cdots Center_{k,D}\right) \tag{3}$$

For the k-means model, there are two ways to terminate an iteration: one is to set the number of iterations T when it comes to Up to the Tth iteration, the iteration is terminated, and the resulting class cluster is the final clustering result.

Square and criterion functions, function models such as formulas (4).

$$J = \sum\nolimits_{k=1}^{K} \sum\nolimits_{x_i \in C_k} dist(x_i, Center_k) \tag{4}$$

## 3 Improved Self-organizing Neural Mapping Network for Automotive Driving Conditions Construction

### 3.1 Basic Principle

The two-segment clustering model building K-means clustering model is a classic clustering model, but the model needs to give a exact number of clusters at the beginning, cluster center, otherwise it is easy to cause the model not convergence or local convergence problems. It was decided that the data set should be clustered before the K-means clustering model, and for this clustering, we chose to use the SOM self-organizing mapping neural network model to do so. The SOM model is an unsealed clustering method with good self-organization. During model operation, clustering is automatically based on the signs of the data, resulting in data that can be used for the number of clusters and cluster centers that the K-means model can enter. The resulting data is then used for the K-means model to solve the problem of pre-convergence of the K-means model.

### 3.2 Model Building

Here are the steps:

1. Mark the time break point in the data set and divide the data set into multiple segments according to the time break point.
2. Mark the acceleration and deceleration anomaly data in each segment, reject and further segment;
3. Long-term parking, long-term traffic jams, and long-term low-speed intermittent driving (maximum speed is less than 10 km/h) data to deal with;
4. Data that is often idled for more than 180 s is processed at a maximum value of 180 s;
5. After the above work is completed, according to the definition of the kinematic segments, the work of statistical kinematic segments is completed;
6. According to the data set after processing, a mathematical model is established, the characteristic data of the driving conditions of the car are calculated, and the driving conditions of the generation table are drawn.

### 3.3 Code Implementation

The algorithm pseudo-code is as follows Table 1.

**Table 1.** The model details the process description.

```
Input: Class cluster number K, iteration abort threshold δ.
Output: Clustered results.
For t=1,2,3…,T
    For every x_i
        Calculate dist(x_i,center_k) according to formula (1);
        Divide xi into the class clusters at the center of the closest class cluster;
    End for
    Update the center of all class clusters according to formula (5);
    According to formula (6), the interpolation ΔJ of the two iterations is calculated;
    If ΔJ<δ
        Then The result of the cluster is output;
        Break;
    End if
End for
```

# 4 Analysis of Experimental Results

## 4.1 The Experimental Process

First of all, on the data pre-processing, the first data cleaning. Wash away some dirty data to avoid affecting later results. Time is missing, acceleration and decrease speed are abnormal, and idle data is cleaned to ensure complete motion fragments. We select the vehicle from the parking idle zone of 0 to the parking idle interval of 0 for the next speed of 0 as a motion snippet. Calculate the average of the 15 indicator parameters selected in the motion fragment to form a feature parameter matrix. The initial cluster number K and cluster center Z are obtained by algorithm, and then the cluster number K and cluster center Z are used as inputs of K-means algorithm, and the cluster results are output. Finally, the driving conditions are constructed according to the established database.

Due to the different conditions in the course of car driving, the original acquisition data recorded directly by the vehicle driving data collection equipment often contains some bad data values. In these data, the bad data mainly include several types: due to high-rise buildings covered or tunneled, GPS signal loss, resulting in the data provided in the time is not continuous, car add, reduce speed abnormal data, long-term parking collected abnormal data, long-term traffic jams, intermittent low-speed driving conditions. Take the following steps to clean the data in response to the above.

- Time loss: For time-missing data, the car into multiple high-rise buildings, covered by high-rise buildings, or the car into the tunnel, will cause the loss of GPS positioning of the car, resulting in data loss, that is, data time is not continuous problems, therefore, the data set should be segmented from the time breakpoint to ensure complete travel fragments.
- Acceleration and deceleration anomalies: For acceleration and deceleration anomaly data, an ordinary car should typically have an acceleration time of more than 7 s from 0–100 km/h while driving, so some transient accelerations in the data set can be considered abnormal data. For deceleration anomaly data, the maximum deceleration of the emergency brake is usually 7.5–8 $m/s^2$, so data that is outside this range can also be considered abnormal.
- Idle speed: For long-term idle data, there is a long-term parking in the process of car driving, such as: parking does not stop, stop and stop, but the car's equipment is still running, as well as traffic jams for too long, intermittent low-speed driving situation. In this case, it's all handled at idle speed, but when a period of idle time exceeds 180 s, we treat it as an exception and calculate it at a maximum of 180 s.

### 4.2 Extraction of Motion Fragments

In the process of studying and constructing the driving condition curve of the automobile, constructing the motion fragment of the vehicle of this model is the most common method used to complete the study. In general, the definition of a motion fragment refers to the vehicle from the speed of 0 parking idle interval to the next stop idle interval, after the processed data set, based on the data set, according to the definition of python fragments divided into multiple psychological fragments, and statistics related data. The driving condition curve and automobile motion characteristics of the automobile by constructing the model should represent the corresponding characteristics of the collected data source (processed data), and the smaller the error between the two, the better, indicating that the representative of the vehicle driving conditions constructed is better [8]. Driving conditions refer to a period of vehicle speed change, and its main parameters such as average vehicle speed and idle driving time should be consistent or as close as possible to the actual traffic conditions in the area [9]. The driving condition is expressed as a speed-time curve. Automobile gear shifting, speed selection, acceleration and deceleration have a great influence on automobile fuel consumption [10]. Therefore, the driving condition of the vehicle on the road is expressed by some parameters that can reflect its motion characteristics, such as acceleration, deceleration, constant speed and idling speed [11]. The following picture is an example of the construction of the car condition under the complete time segment, such as Fig. 2, Fig. 3 and Fig. 4.

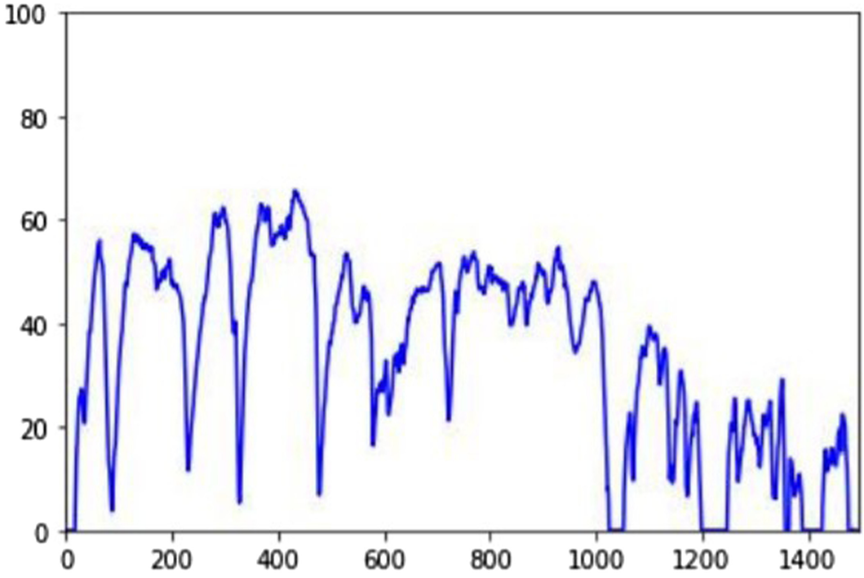


**Fig. 2.** Motion fragment 1

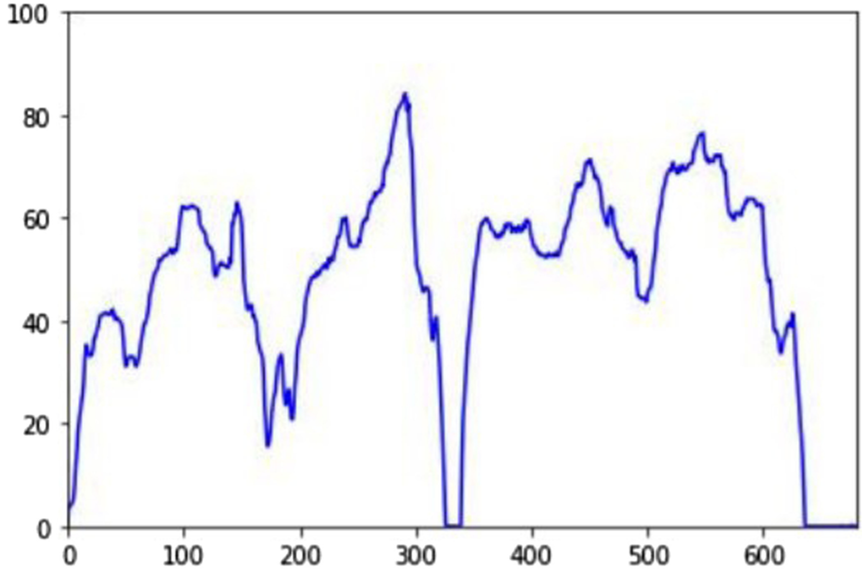


**Fig. 3.** Motion fragment 2.

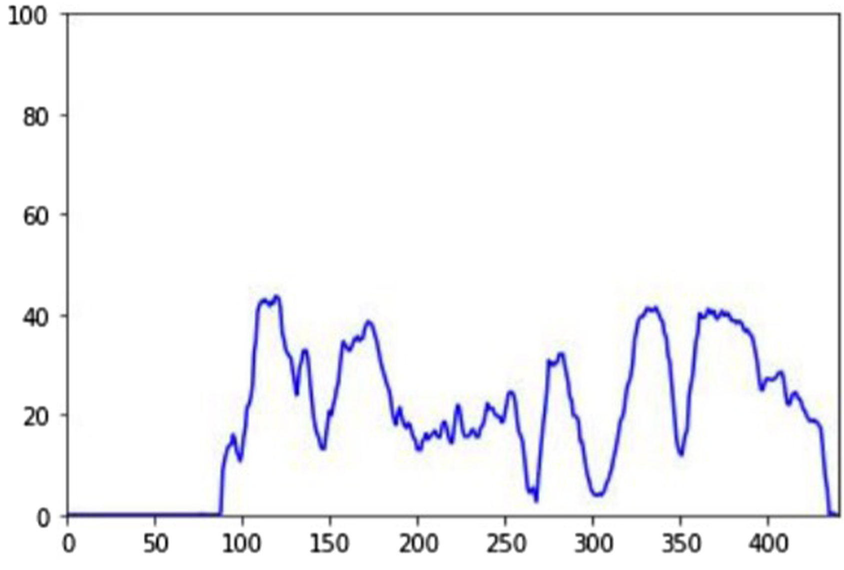


**Fig. 4.** Motion fragment 3.

### 4.3 Selection and Calculation of Feature Parameters

In order to construct the automobile motion characteristic evaluation system, 15 automobile motion characteristic evaluation indicators have been established. As shown in Table 2.

**Table 2.** Feature parameter indicator.

| Serial number | Feature parameters | Meanings |
|---|---|---|
| 1 | T(s) | Running time |
| 2 | S(m) | Driving distance |
| 3 | $S_{ni}$(m) | Non-idle travel distance |
| 4 | V(km/h) | Average speed |
| 5 | $\bar{V}_{ni}$(km/h) | Average travel speed |
| 6 | $V_{max}$(km/h) | The maximum speed |
| 7 | $\bar{A}_a\left(m/s^2\right)$ | Average acceleration |
| 8 | $\bar{A}_d\left(m/s^2\right)$ | Average reduction |
| 9 | $F_i$(%) | Idle time ratio |
| 10 | $F_a$(%) | Acceleration time ratio |
| 11 | $F_d$(%) | Deceleration time ratio |
| 12 | $V_s$(km/h) | The speed standard is poor |
| 13 | $V_{sa}$(km/h) | Increase the speed standard deviation |
| 14 | $V_{sd}$(km/h) | Reduce the speed standard deviation |
| 15 | $A_{sa}(m/s^2)$ | Acceleration standard deviation |

Here's our formula for calculating feature parameters.

$$S(m) = V_1T_1 + V_2T_2 + \cdots + VnTn \tag{5}$$

$$S_{ni}(m) = V_{1ni}T_{1ni} + V_{2ni}T_{2ni} + \cdots + V_{nni}T_{nni} \tag{6}$$

$$\bar{V}\left(\frac{km}{h}\right) = \frac{s}{T} \times 3.6 \tag{7}$$

$$\bar{V}_{ni}\left(\frac{km}{h}\right) = \frac{S - S_i}{T - T_i} \times 3.6 \tag{8}$$

$$\bar{A}_a\left(\frac{m}{s^2}\right) = \frac{\frac{\Delta V_{a1}}{\Delta t_1} + \frac{\Delta V_{a2}}{\Delta t_2} + \cdots + \frac{\Delta V_{an}}{\Delta t_n}}{n} \tag{9}$$

$$F_d(\%) = \frac{T_d}{T} \tag{10}$$

$$\bar{V}\left(\frac{km}{h}\right) = \frac{V_1 + V_2 + \cdots + V_n}{n} \tag{11}$$

$$V_s\left(\frac{km}{h}\right) = \sqrt{\frac{\left(V_1 - \bar{V}\right)^2 + \left(V_2 - \bar{V}\right)^2 + \cdots + \left(V_n - \bar{V}\right)^2}{n}} \tag{12}$$

$$\bar{A}_{sa}\left(\frac{m}{s^2}\right) = \frac{A_{1sa} + A_{2sa} + \cdots + A_{nsa}}{n} \tag{13}$$

$$A_{sa}\left(\frac{m}{s^2}\right) = \sqrt{\frac{\left(A_{1sa} - \bar{A}_{sa}\right)^2 + \left(A_{2sa} - \bar{A}_{sa}\right)^2 + \cdots + \left(A_{nsa} - \bar{A}_{sa}\right)^2}{n}} \tag{14}$$

$$\bar{V}_a\left(\frac{km}{h}\right) = \frac{V_{1a} + V_{2a} + \cdots + V_{na}}{n} \tag{15}$$

$$V_{sa}\left(\frac{m}{s^2}\right) = \sqrt{\frac{\left(V_{1sa} - \bar{V}_a\right)^2 + \left(V_{2sa} - \bar{V}_a\right)^2 + \cdots + \left(V_{nsa} - \bar{V}_a\right)^2}{n}} \tag{16}$$

$$\bar{V}_d\left(\frac{km}{h}\right) = \frac{V_{1d} + V_{2d} + \cdots + V_{nd}}{n} \tag{17}$$

$$V_{sd}\left(\frac{m}{s^2}\right) = \sqrt{\frac{\left(V_{1sd} - \bar{V}_d\right)^2 + \left(V_{2sd} - \bar{V}_d\right)\bar{V}_d^2 + \cdots + \left(V_{nsd} - \bar{V}_d\right)^2}{n}} \tag{18}$$

## 4.4 Results and Analysis

The average of 15 indicator parameters is calculated, and the traditionally calculated driving conditions are constructed from the two-segment clustering method.

The operating conditions are compared with the driving conditions constructed separately using the k-means method, and the results are shown in Table 3.

**Table 3.** The results of the experiment were compared.

| Feature parameters | Experimental data | K-means | Two-segment clustering |
|---|---|---|---|
| T(s) | 14.49 | 14.49 | 14.49 |
| S(m) | 676.82 | 676.82 | 676.82 |
| $S_{ni}$(m) | 671.61 | 671.61 | 671.61 |
| V(km/h) | 14.52 | 14.01 | 14.63 |

(*continued*)

**Table 3.** (*continued*)

| Feature parameters | Experimental data | K-means | Two-segment clustering |
|---|---|---|---|
| $\bar{V}_{ni}$(km/h) | 21.81 | 20.83 | 22.34 |
| $V_{max}$(km/h) | 35.81 | 35.81 | 35.81 |
| $\bar{A}_a\left(m/s^2\right)$ | 1.86 | 1.75 | 1.92 |
| $\bar{A}_d\left(m/s^2\right)$ | −2.11 | −1.93 | 2.15 |
| $F_i$(%) | 43.01 | 41.25 | 43.93 |
| $F_a$(%) | 28.95 | 30.48 | 28.07 |
| $F_d$(%) | 25.12 | 22.81 | 24.40 |
| $V_s$(km/h) | 11.8 | 11.8 | 11.8 |
| $V_{sa}$(km/h) | 1.59 | 1.72 | 1.51 |
| $V_{sd}$(km/h) | 1.76 | 1.59 | 1.81 |
| $A_{sa}$($m/s^2$) | 2.59 | 2.57 | 2.57 |

The results show that the average relative error obtained by the two-segment clustering method is small, only 3%, and the error of K-means clustering method is greater than 5%. Therefore, the two-segment clustering method perfectly avoids the K-means clustering because the data set is too large, resulting in the initial convergence difficulties of the model run, there may be local optimal solution of the problem, and the model built by K-means represents the driving conditions closer to the experimental data to better reflect the actual road traffic conditions of the city in which the car is located.

## 5 Conclusion

Through the test, get a large number of actual driving speed data, experiment repeatability is better, through the obtained motion fragment analysis, the construction of driving conditions and Europe's NEDC and WLTC operating conditions are different, which proves that each city has different driving conditions characteristics, can not have some standard driving conditions to fully reflect.

**Acknowledgments.** 1. Major special project of science and technology of Guangdong Province, No: 190826175545233.

2. Beijing science and technology innovation service capability construction project (PXM2016_014223_000025).

3. BIGC Project(Ec202007).

## References

1. Nesamani, K.S., Subramanian, K.P.: Development of a drivingcycle for intra-city buses in Chennai, India. Atmos. Environ. **45**(31), 5469–5476 (2011)

2. Ho, S.H., Wong, Y.D., Chang, W.C.: Developing Singapore Driving Cycle for passenger cars to estimate fuel consumption and vehicular emissions. Atmos. Environ. **97**, 353–362 (2014)
3. Knez, M., Muneer, T., Jereb, B., et al.: The estimation of a driving cycle for Celje and a comparison to other European cities. Sustain. Cities Soc. **11**(2–3), 56–60 (2014)
4. Li, N.: Construction and Research of Urban Road Vehicle Driving Conditions. Hebei Agricultural University, Baoding (2013)
5. Shiqin, M.H., Jianxun, D., et al.: Improved FCM clustering method and its application in the construction of driving conditions. China Mech. Eng. **25**(10), 1381–1387 (2014)
6. Wei, C., Li, Y., Li, C., et al.: Based on the K-average clustering algorithm, Xi'an automobile driving condition synthesis technology research. Autom. Technol. **8**, 33–36 (2015)
7. Jianping, G., Dexuan, R., Jianguo, X.: Construction of automobile driving conditions based on global K-means clustering algorithm. Henan Li J. Univ. Technol. (Nat. Sci. Edn.) **38**(01), 112–118 (2019)
8. Wang, Y., Zhang, N., Sun, Z., Wu, Y., Liu, B.: Vehicle driving condition recognition algorithm based on LVQ model. Agricul. Equip. Veh. Eng. **57**(05), 1 − 4+8 (2019)
9. Liu, X., Yan, D.: Investigation and research on driving conditions of urban vehicles in my Country. Environ. Sci. Res. **13**(1), 23–28 (2000)
10. Yang, Y., Cai, X., Du, Q., et al.: Research on driving conditions of road vehicles in Tianjin. Autom. Eng. **24**(3), 200–204 (2002)
11. Ma, Z., Zhu, X., Li, M., et al.: Dynamic clustering method in actual driving conditions of vehicles application in development. J. Wuhan Univ. Technol. **27**(11), 69–71 (2005)