International Journal of Control and Automation
Vol.7, No.3 (2014), pp.125-136
http://dx.doi.org/10.14257/ijca.2014.7.3.13

Research on Driving Cycle of Long-distance Passenger Vehicles Based on Principal Component Analysis and Cluster Algorithm

Yingji Liu[1], Jingyi Li[2] and Bo Shen[2]

[1]*Key Laboratory of Operation on Transport Vehicles, Ministry of Transport, Beijing, China*
[2]*Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing, China*

bshen@bjtu.edu.cn

Abstract

In order to get the vehicle' actual driving mode in a different road conditions, the driving data is divided into kinematic sequence for analysis of driving cycle. On the basis of kinematics fragments, the characteristic parameters of kinematic sequence are carried out with the principal component analysis. Then driving cycle are synthesized based on dynamic clustering algorithm analysis to get the vehicle' driving state in different road conditions.

1. Introduction

Automobile exhaust emissions have become a major factor in environmental pollution. Automobile exhaust contains a lot of dust particles and emissions of nitrogen oxides, carbon oxides and hydrocarbon gas are much harmful to the human body resulting in respiratory disease, even serious cause cancer. At present, China has realized the importance of control of vehicle exhaust emissions. Meanwhile, vehicle exhaust emissions are managed with more stringent standards. Automobile exhaust emissions and other performance measurements were done on the bench in the laboratory, to simulate the actual driving cycle of the vehicle. Therefore, researchers have developed a variety of model on different driving cycle and vehicles, which is the vehicle driving cycle. It represents the speed variation of the type of vehicle in the test area. In addition, the also has a very important reference value to understand the traffic situation in the region

The vehicle driving cycle is speed - time history of some parameters by analyzing the intrinsic link, by analyzing the internal relation of some parameters, which can be represented as two variables, the current speed and time, determine the current speed and the corresponding time for sampling.

Different countries, according to their different geographical and traffic conditions take a different driving cycle standards. United States, Europe and have a lot of research Japan for their own vehicles and traffic state, and driving cycle is also very mature. For example, European ECE15 and NEDC, American FTP-75, Japanese japan10 are all internationally recognized systems of driving cycle. China currently has not any driving cycle system universally recognized international. At present, certified passenger driving cycle in China is still using European urban steady conditions ECE-15 and NEDC that includes the suburbs driving cycle. Heavy-duty vehicles certified driving cycle is also using Europe ESC steady and transient ETC. However, China's traffic conditions, city size, vast territory and its existence are very different from other countries, so China also needs to develop according to their own inherent characteristics of the vehicle driving cycle system for testing the

ISSN: 2005-4297 IJCA

performance of the car with more scientific and emission standards and methods.

Intelligent Transportation Systems (ITS) has become the new hot spot of driving cycle, while the GPS technology as a mature and widely used in the calculation, has been the main information vehicle monitoring techniques.

Most of the current objects of study are concentrated in coverage of urban taxi or floating car, as well as vehicle speed estimation and so on. The researches of medium-sized passenger travel data analysis are less, and focus on public transport vehicle scheduling and bus arrival time prediction, but the researches of long-distance passenger car's space driving cycle analysis based on GPS data long-distance passenger space vehicle driving cycle analysis have been few mentioned. This paper is designed to verify operation of long-distance passenger vehicle driving cycle feasibility analysis, using long-distance passenger traffic GPS data from Hefei in Anhui to Zhengzhou in Henan for objects and by collecting and processing a large number of dynamic driving data, through the relevant multivariate statistical methods to analyze the results of long-distance passenger vehicle driving cycle.

Driving cycle analysis method consists of the following aspects: real-time vehicle data acquisition and storage, data processing, driving cycle analysis and driving cycle's reasonable verification. Driving cycle analysis and driving cycle's reasonable verification are the core of the study. Among them, driving cycle analysis is the most important step to build a vehicle driving cycle. What analytical methods are confirmed determines what the final driving cycle to build and whether the driving cycle to reflect actual driving characteristics of the vehicle. Many domestic and international researchers and practitioners have developed a lot of analytical methods. The author based on comprehensive consideration design of the principal component cluster analysis method and the separate cluster analysis method is used for city traffic and highway traffic analysis, in order to achieve long-distance passenger cars analytical driving cycle. Large amounts of data collected by the GPS divided into the city link data and highway link data which use different analytical methods based on different driving characteristics can make more accurate simulations of long-distance passenger cars travel status while operation is greatly reduced.

2. Related Technologies

Like what the article mentions at the beginning, so far China has not yet the driving cycle that is widely recognized, but our researchers has done considerable research efforts in the driving cycle, even more has made some achievements.

In 2005, Zhao Hui made a detailed study of modes of transport in Hong Kong. She obtained an actual operating conditions based on driving cycle and compare the result with the more mature driving cycle in Europe, America and Japan. Researchers in Hong Kong for a lot of the actual driving speed tests。Then they according to the number of relationships with the various criteria for determining the source in the laboratory analyzed the characteristics of the traffic situation in Hong Kong which are a long idle time, acceleration and deceleration too often, hardly moving at a constant speed, driving speed is lower than the United States and is closer to urban traffic patterns in Japan. Hong Kong Island's traffic is relatively better than in the one in the Kowloon area, mainly embodied in the average speed is higher.

Hong Kong's urban driving cycle is to obtain driving cycle. The general idea is extracting a velocity curve from a lot of actual measured driving speed, making this curve can reflect the characteristics of the measured data. First, the speed measured each time the expansion curve into a drive cycle. A preparatory driving cycle that is formed by 20 randomly selected from extensions compute its 9 numbers of criteria. If these criteria are the mean value of the number difference of 5% or less, it will be accepted as driving cycle.

In addition to driving cycle in Hong Kong analytical method, the researchers also put a lot of research into domestic urban traffic and driving cycle but also used in many other aspects of the driving cycle analysis.

Guangzhou bus driving cycle is derived using the dynamic clustering method. The basic idea of the dynamic clustering method is to classify sample (kinematic sequence) roughly according to certain principles and then correct until the classification comparison reasonable so far. By analyzing the exercise Canton bus data sample, the total sample into two categories. The first Idle speed ratio is three times of the second one. Obviously, this indicates that the first short kinematic sequence reflects the road driving characteristics in heavy transportation, and the second one reflects driving characteristics on the flow road.

In recent years, researchers of driving cycle study specific driving cycle in each city. These are basically using the fixed step size intercept method and developed based on short-stroke stroke analysis, VA (velocity and acceleration) matrix analysis, cluster analysis and principal component analysis method Markov analysis methods. These methods not only successfully establish driving cycle for different cities, different states and geographical patterns of traffic, but also provides the appropriate analytical method driving cycle for the problems studied in this paper.

3. Data Collection of Driving Cycle

There are three main methods of data collection of driving cycle: vehicle tracking method, the average traffic statistics and autonomous driving. Since the object of this study is long-distance passenger cars, the above three methods are not applicable for the object. The paper uses the "vehicle tracking method" in normal operating state, by tracking the GPS track data of long-distance buses in the normal working state to complete the original data collection. To carry out this study, work, the author collaborates with the Hefei passenger transport company in Hefei to achieve long-distance operating passenger data from Hefei to Zhengzhou. The acquisition frequency provided data is 1Hz, which fully reach the conditions of the test vehicle requirements. The data speed accuracy is 0.1m, positioning accuracy is 8m, also reach the appropriate standards of research studies and even beyond the basic requirements.

Many domestic and overseas studies show that based on data collected by GPS technology road network coverage is high, and based on data collected by GPS technology with a high degree of artificial fit the measured data. Based on the motion of the GPS based on the collected data access method is feasible and the result is reliable.

Test vehicle for the long-haul operations is in the same operating state vehicles, so the route Hefei - Zhengzhou along the road data, including the return traffic data, can capture the traffic conditions in different driving data, in order to better carry the resolution of driving cycle.

Figure 1. Coach urban road section

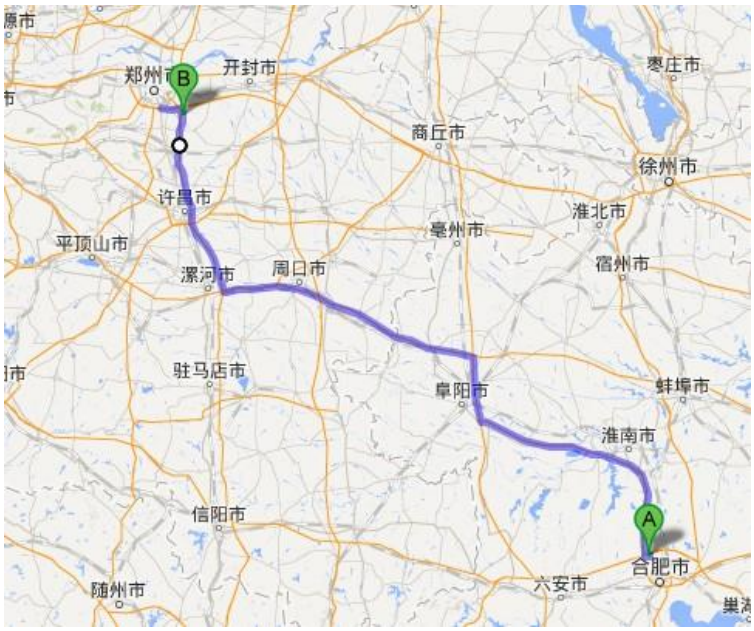


Figure 2. Coach highway sections

4. Data Processing

The original Driving data is divided according to the velocity difference in acceleration into the following four basic driving modes after collection is complete.

（1） Idle mode, the engine is in working condition and the acceleration of Ongoing process is $0\,m/s$.

（2） Acceleration mode, the acceleration of ongoing process is greater than 0.15 m/s^2.

（3） Deceleration mode, the acceleration of ongoing process is less than -0.15 m/s^2.

（4） Uniform pattern. The acceleration of ongoing process between -0.15 m/s^2 to 0.15 m/s^2.

A kinematic fragment is a movement beginning to idle mode to idle mode. Kinematic sequence is the study of driving conditions. Therefore, after getting the original data. Raw data is divided into a fragment of a kinematic, in order to carry out the next step of the analysis of driving conditions. Travel of the vehicle may be regarded as a combination of various fragments. Some of these pieces reflect the traffic situation may be the same. Different time, location and road type will appear the same fragment. Sometimes busy highway fragments may be exactly same as crowded city. These fragments were linked to the type and traffic conditions, targeted to analyze movement patterns at different speeds. And on this basis, parsing out the driving conditions is fully consistent with the objective situation.

Kinematic sequence contains a lot of kinematic characteristics. The feature is used to analyze the characteristic parameters of driving conditions. Based on the results of previous research at home and abroad as well as access to data in this study, we determine the characteristics of the nine characteristic value parameters.

Eigenvalues	Meaning / unit
t	Time /s
v_m	Average speed / m/s
v_{mr}	Average running speed / m/s
$a_{\max}$	Maximum acceleration / m/s^2
$a_{\min}$	Minimum acceleration / m/s^2
P_a	Acceleration percentage /%
P_c	Uniform speed percentage /%
P_d	Deceleration percentage /%
P_i	Idle speed percentage /%
d	Travel Distance /m

This average running speed is not included the idle mod, which is in motion.

5. Key Algorithms

Kinematics fragment has more characteristic parameters and there is a correlation between indicators, which means the information in the observed data overlap to a certain extent, this will certainly increase the difficulty of data analysis, so we need to reduce dimension of the multivariate. Principal component analysis refers that the original parameters are combined into a new set of uncorrelated comprehensive parameter to analyze instead of the original parameters, and the new parameters should contain the original information as much as possible.

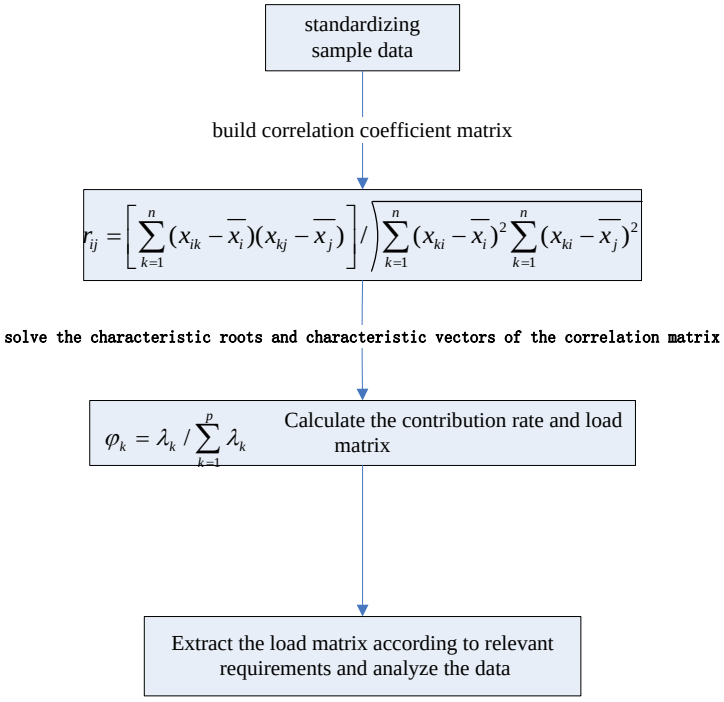


Figure 3. PCA flowchart

With n samples, each sample characteristic parameters are p，Denoted as y_1, y_2,$\cdots$, y_p, Such raw data constitutes an n × p dimensional matrix：

$$Y_{n\times p}=\begin{bmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{np} \end{bmatrix}$$

among formula， y_{ij} （i=1,2,$\cdots$,n ;j=1,2,$\cdots$,p） is the i-th sample of the j variable.

After using principal component analysis algorithm of the characteristic parameters with data，we need to use required clustering algorithm to classify put driving data. So what can be seen under different traffic conditions in different driving modules.

The basic idea of clustering method is that we roughly classify the sample (refers to

"kinematics fragments") firstly, then don't amend it in accordance with a certain principle until there are bigger difference during classification.

In this paper, we adopt a method called K-means Clustering of clustering analysis, of which the process is simple, but the effect is obvious.

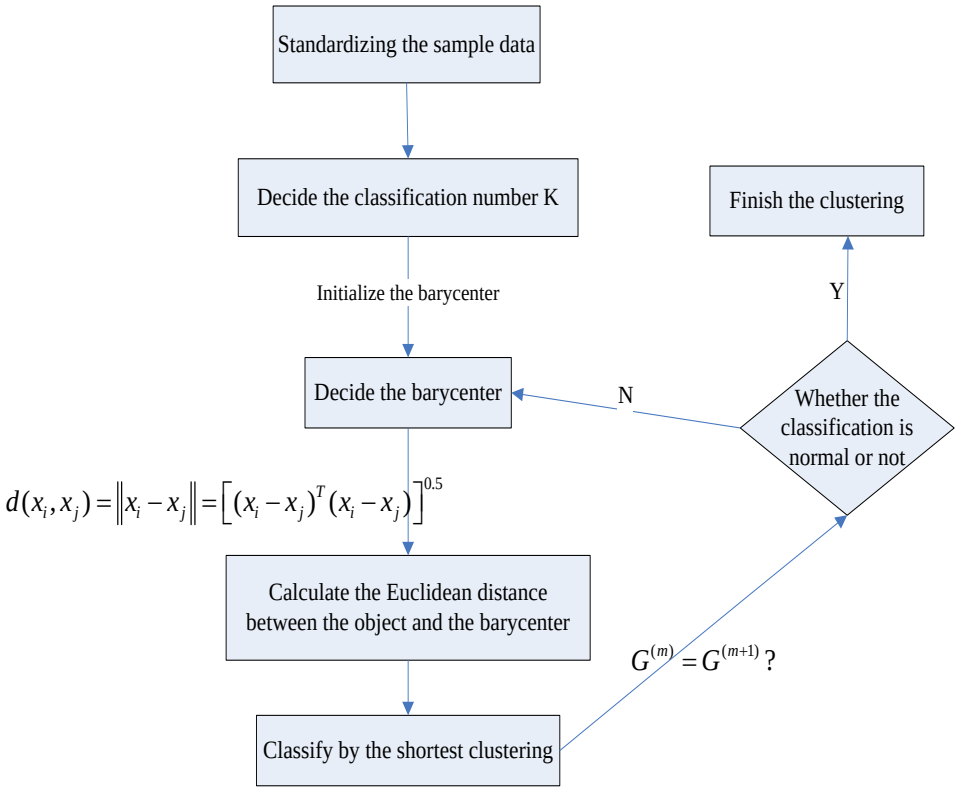


Figure 4. K-means clustering flowchart

Sampling N data into N data objects of p dimension and clustering them according to the flowchart of figure 3 before reaching a stable clustering result.

The purpose that PCA analyzes the original travelling data is to reduce the dimension, which can reduce the calculation amount of clustering analysis that makes the result easier to be analyzed. Different from other papers which use some principal components whose accumulation contribution rate reaches 80 percent, this paper uses the parameters which have great influence on the principal component among more than 80 percent of principal component. Thus in the case of calculation accuracy basic remaining unchanged, the calculation will be more simple, and it is effective for the large amount of engineering analysis.

6. Working Condition Building Process and Result

This article selected long-distance bus routes along the Hefei-Zhengzhou as test line. The long-distance passenger car vehicle type is Yutong Bus. Using GPS data acquisition system, we collect a large number of real-time traffic data as samples of working condition analysis.

Sample data contains the urban road and highway outside the city. So this article needs to complete the construction of two roads. According to the directions of track playback, we identify urban road and expressway intersects latitude and longitude. Construction of driving working condition analysis as a statistical process must bring some relative error. Therefore, the latitude and longitude you have found do not need to be very precise. Two sections of the vast majority of traffic data of two sections should be contained in subparagraph.

GPS data are divided into two kinds of sample data of different roads. Sample data needs to be corresponding pre-processed. The most important thing is to remove unwanted information from the sample data, for example, to remove non-operating state data, status of garbage in your GPS data as well as data whose speed is 0. In addition, you also need to add the information required to build a working condition For example, we divide the data into accelerate condition, slow conditions, constant speed and idle condition.

Already mentioned in the above vehicle, driving conditions are very complex. Due to vehicles' frequent starting, accelerating, decelerating, for ease of analysis, we define the kinematics fragment as driving between 2 adjacent parking spots. That means the vehicle from an idle state to the next idle state is a unit of data. The test procedure consists of kinematics fragments. The data is made up of many units.

The data is divided into a number of kinematic data fragments using Visual Basic programming. Then we sort out the kinematics characteristic parameters of every segment using Excel, forming a characteristic parameter matrix. The below table is a part of the characteristic parameters of matrix.

Number	Length of time（unit s）	Driving distance（unit m）	Idle ratio（%）	Constant ratio（%）	Accelerating ratio（%）	…
1.5	1020	18723.6	0.00098	0.47843	0.281373	…
1.6	225	28970186	0.29333	0.275556	0.262222	…
1.7	170	1749.154	0.370588	0.258824	0.276471	…
1.8	38	140.1172	0.026316	0.263258	0.289474	…
1.9	61	350.2198	0.098361	0.131148	0.42623	…
…	…	…	…	…	…	…

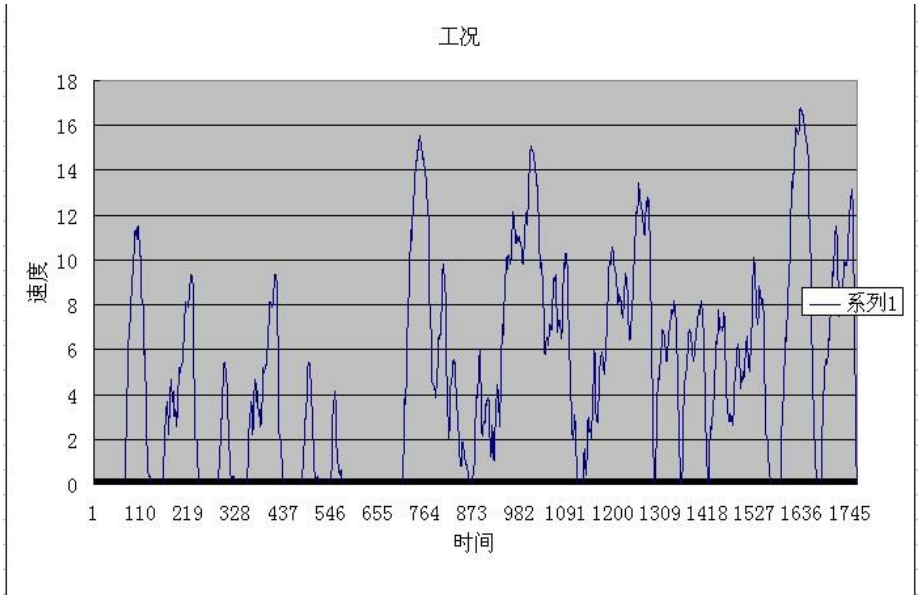
We conduct a parameter principal component analysis of this characteristic parameter matrix using SPSS software. So we get 10 principal components. We select cumulative contribution rate principal components which are above 80%. The below table is the results of principal components analysis matrix of the urban road in Hefei.

	Component 1	Component 2	Component 3
Length of time	0.403	0.749	0.230
Driving distance	0.691	0.500	0.357
Idle ratio	-0.753	0.596	-0.252
Constant ratio	0.321	-0.466	0.769
Accelerating ratio	0.771	-0.463	-0.080
Decelerating ratio	0.706	-0.410	-0.348
Average speed	0.927	0.228	0.067
Average driving speed	0.880	0.295	-0.039
Max acceleration	0.675	-0.029	-0.449
Min acceleration	-0.849	-0.172	0.227

We can conclude that the cumulative contribution rate of first three principal components has reached 87.74%. And we could know the role of each parameter in different ingredients. By finding out the characteristic parameters in these principal components which makes larger role, we can deduce the dimension of the characteristic parameters, without losing too much driving information at the same time.

After deduce the dimension, we can use the SPSS software to cluster the characteristic parameter matrix and perform the 2-means, 3-means and 4-means clustering analysis on it by using the characteristics of the K-means clustering. Through the analysis, we could conclude that the difference between each state's time scales is not so obvious in 3-means clustering and 4-means clustering, especially for the idle speed ratio. However, the relative difference is rather significant in 2-means analysis. It's more reasonable for us to choose it.

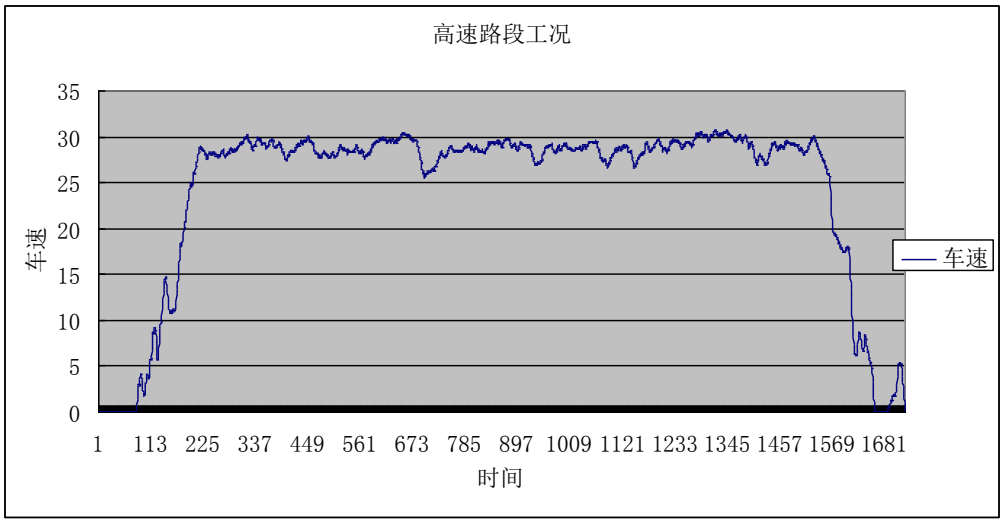
According to the results of the principal components and the cluster analysis, we can extract the typical fragment proportionally from the two kinds of kinematics fragments above to fit the Hefei city's road. Part of the typical driving cycle is shown in the graph below.



According to the study on Hefei city's driving cycle of other documents', the idle state accounts for larger proportion. This is due to the departure time is 4:20 PM, the driving

process encounters the peak after work, which make the percentage of idle speed higher. Meanwhile, each parameter's relative error between the driving cycle and the overall data is 17.2% .So it's in accordance with traffic data information

Because of the singleness of the high speed part's data model, it does not need large amount of information to complete the driving cycle's building work. Therefore, we can perform the clustering analysis directly on the kinematics fragments, which come from the high-speed road traffic information of Hefei to Zhengzhou. According to above steps, the relative difference in 2-means clustering analysis is very obvious. Driving cycle on high-speed road is shown in the diagram below.



As shown in the figure, the idle state part of the driving cycle on high-speed road counts for fewer proportion. It's in high speed most of the time, with low speed and idle state sometimes the same as the time zone, which conform to the objective facts.

7. Prospects

In this paper, the principal component analysis and dynamic clustering combination of analytical methods to get working coach driving conditions get a satisfied results. Although relatively simple lines of this study, the result is too thin, but for the future multiline research work provides an effective method and experience.

Since previous studies during driving conditions need to add additional vehicle terminal equipment to carry out data collection, and this article is based on research data operational state of the coach car GPS-based data collection. It will not only reduce the working conditions for the subsequent study of the cost of research to provide a good experience, but also increase the real degree of the data. It avoids driving conditions of the previous research in the process which needs to specify the line to bring inconsistent with the objective situation.

References

[1] M. Zhixiong, L. Mengliag and Z. Xichan, "Principal component analysis (pca) in the application of the development of vehicles' real riding condition", Journal of wuhan university of technology (management and information engineering), vol. 26, no. 4, **(2004)**, pp. 32- 34.

[2] L. Bo, "Probability Theory and Mathematical Statistics", Chongqing: Chongqing university press, **(1995)**.
[3] X. Ci and L. Xilin, "Related factors to explore the influence on vehicle emission factor", environmental protection, no. 4, **(2004)**, pp. 4547.
[4] Z. Xichan and L. Mengliang, "Working condition of vehicle development methods", Journal of Su Daxue (Natural science brand), no. 26, **(2005)** February, pp. 110-111.
[5] W. Junfang, D. Yan and W. Aijuan, "The motor vehicle driving cycle research of Beijing", Journal of environmental engineering technology, vol. 2, no. 3, **(2012)**, pp. 240-246.
[6] J. ping and C. Wuwei, "Clustering and markov method combined with the city building cars driving cycle", China Mechanical Engineering, no. 23, **(2010)**, pp. 2893-2897.
[7] J. ping, S. Qin and C. Wuwei, "Based on wavelet analysis of working condition of city road building research", Automotive Engineering, no. 1, **(2011)**, pp. 70-73, 51.
[8] C. Chenghe, Q. Huiyuan and Z. Zhijian, "The electric car fixed regional condition of development studies", Agricultural Equipment and Vehicle Engineering, no. 6, **(2011)**, pp. 7-10, 17.
[9] S. Qin, Z. Yubo and J. Ping, "Based on the research of kinematics section of urban road conditions", Automotive Engineering, no. 3, (2011), pp. 256-261.
[10] P. Meiyue and Z. Yafu, "The city bus driving cycle research", Machinery Design and Manufacture, no. 2, **(2010)**, pp. 263-265.
[11] S. Qin, C. Duoyang and W. Jing, "Based on principal component analysis and research of running condition of FCM clustering", Research of Environmental Sciences, vol. 25, **(2012)**, pp. 70-76.
[12] M. Andre, "Driving Cycles Development: Characterization of the Methods", SAE Special Publications, vol. 1201, no. 12, **(1996)**, pp. 312~ 322.
[13] M. Andre, A. J. Hickman, D. Hassel, *et al.*, "Driving Cycles for Emissions Measurements Under European Conditions", SAE Special Publications, vol. 1197, no. 3, **(1995)**, pp. 116-124.
[14] N. Demaski, Y. Guezennec and A. Soliman, "Analysis and experimental refinement of real-world driving cycles", Washington DC; SEA Technical Paper, **(2002)**.
[15] N. N. Clark, M. Gautam, W. S. Wayne, *et al.*, "Creation and test cycle evaluation of a medium heavy-duty truck", SAE Transactions, vol. 112, no. 4, **(2003)**, pp. 2654-2668.
[16] M. Montazeri and M. Naghizadeh, "Development of the Tehran car driving cycle", International Journal of Environment and Pollution, vol. 30, no. 1, (2007), pp. 106-118.