

基于改进的K-均值聚类算法构建 柴油车运行工况

李勇志

上海工程技术大学机械与汽车工程学院, 上海

收稿日期: 2023年3月16日; 录用日期: 2023年5月22日; 发布日期: 2023年5月29日

摘要

针对传统的K-均值聚类算法在聚类时容易陷入局部最优的问题, 提出一种改进的K-均值聚类算法, 并结合主成分分析法将其应用到柴油车运行工况的构建中。首先通过主成分分析完成运动学片段特征参数的降维, 再利用改进的K-均值聚类算法进行运动学片段的聚类, 最后选择具有代表性的片段合成运行工况。对自主构建的运行工况的特征参数值和总体数据的特征参数值进行比较, 各参数间的相对误差在2%~5%之间, 说明用改进后的K-均值聚类算法结合主成分分析法构建的运行工况具有很高的精度, 并且对车辆的运行状况具有很强的代表性。

关键词

改进的K-均值聚类, 主成分分析, 运行工况, 柴油车

Construction of Diesel Vehicle Driving Condition Based on Improved K-Means Clustering Algorithm

Yongzhi Li

School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai

Received: Mar. 16th, 2023; accepted: May 22nd, 2023; published: May 29th, 2023

Abstract

Aiming at the problem that the traditional K-mean clustering algorithm is prone to fall into the local optimum during clustering, an improved K-mean clustering algorithm is proposed, and combined with the principal component analysis method, it is applied to the construction of diesel ve-

hicle operating conditions. Firstly, dimensionality reduction of characteristic parameters of kinematic segments was completed through principal component analysis, and then the improved K-means clustering algorithm was used to cluster kinematic segments. Finally, representative segments were selected to synthesize operating conditions. By comparing the characteristic parameter values of the independently constructed operating conditions with those of the overall data, the relative error of each parameter is between 2% and 5%, indicating that the operating conditions constructed by the improved K-mean clustering algorithm combined with the principal component analysis method have high precision, and have a strong representation of the vehicle operating conditions.

Keywords

Improved K-Means Clustering, Principal Component Analysis, Operating Conditions, Diesel Vehicles

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

城市化和机动化的进程使机动车保有量不断增加，机动车排放对环境造成的污染问题日益严重[1]。机动车排放标准认证与车辆运行工况有十分重要的联系，车辆行驶工况表示某种车型在某种交通环境下运行时的速度 - 时间曲线，对新车型的技术开发和评估也有重要意义[2] [3]。

为了构建更为精准的运行工况，国内外学者进行大量研究。路尧等[4]调研国内多个代表性城市的实际运行工况，获得车辆实际道路运行数据，利用短行程法构建轻型汽车运行工况。Wang 等针对中国城市特点构建运行工况，并将其与欧洲运行工况和美国运行工况进行比较，然后对不同国家运行工况下的排放因子进行对比分析。彭育辉等[5]利用 K-均值聚类方法对车辆行驶数据进行分析，再用 Silhouette 函数对聚类结果进行筛选，增加了人为选择的准确度，然后根据聚类结果进行车辆运行工况的构建。曹骞等[6]用自动驾驶的方法采集车辆的行驶数据，然后用马尔科夫链算法进行行驶工况的构建，最后针对沈阳市乘用车构建出具有代表性的运行工况。郭家琛等[7]用自编构建码器对数据进行降维处理，在聚类后用 Silhouette 函数对聚类结果进行筛选，构建车辆运行工况。段宇帅等[8]用传统的 K-均值聚类算法和主成分分析法进行运行工况的构建。周溪召等[9]同样通过传统的 K-均值聚类算法对大连市汽车运行工况进行构建。上述研究在聚类时大多运用了传统的 K-均值聚类算法，在聚类结果方面存在一定的不足，比如聚类结果不稳定，聚类结果容易陷入局部最优等问题。

为此，本文通过改进 K-均值聚类算法，解决聚类时聚类结果不稳定和容易陷入局部最优的问题，并将改进的 K-均值聚类算法结合主成分分析法应用于车辆运行工况的构建。并验证所构建工况的准确性与代表性。

2. 车辆运行数据采集

为了使采集的车辆运行数据能更加有效的代表当地的实际道路交通状况，使构建的运行工况更具代表性。本研究将远程监控终端设备和北斗车载定位终端设备与试验车辆内的终端接口相连接，在不干扰试验车辆原本行驶计划的情况下，让试验车辆在实际道路进行同往常一样的行驶，然后利用车载终端所具有的发动机信息解析和存储功能、卫星定位功能采集试验车辆实际道路运行状况下的车速信息和定位

信息，并利用车载终端所具有的远程通讯功能将所采集到的信息传送到远程监控数据平台，然后通过客户端登录获得试验数据，将得到的试验数据用于后续的工况构建中。数据采集装置如图 1。

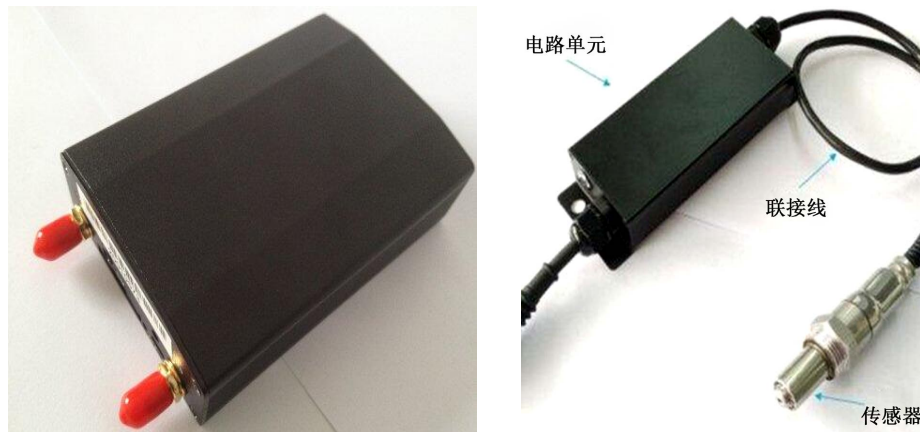


Figure 1. Data acquisition equipment
图 1. 数据采集装置图

3. 主成分分析和改进的 K-均值聚类算法

3.1. 主成分分析法

主成分分析法是把原数据中的多个变量进行处理，重新组合成新的综合变量，这些新的综合变量被称为主成分[10]。主成分能够反映原有变量的特征，并且主成分的数目相较于原来变量数目有所减少，达到数据降维的目的。经过主成分分析之后得到的主成分比原来的变量更加简洁且具有更高的独立性，为后续数据分析带来便利。

主成分分析的具体计算步骤如下：

1) 对原始数据进行向量化和标准化

采集 m 个样本，并且每个样本里包含 n 个变量，然后将其构造成一个样本矩阵，样本矩阵如式 1：

$$\mathbf{x} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{bmatrix} \quad (1)$$

矩阵中 x_{mn} 表示第 m 个样本的第 n 个变量。

由于特征参数量纲的不同，导致各变量的取值较为分散，进而导致后续数据分析的稳定性降低。所以为了避免上述情况的发生，对数据进行标准化处理。具体计算公式如式(2~4)：

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad (2)$$

$$s_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}} \quad (3)$$

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (4)$$

式中 i 为 $1, 2, 3, \dots, m$; j 为 $1, 2, 3, \dots, n$;

进行标准化后的样本矩阵为

$$\mathbf{Z} = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & Z_{23} & \cdots & Z_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Z_{m1} & Z_{m2} & Z_{m3} & \cdots & Z_{mn} \end{bmatrix}$$

矩阵中 Z_{mn} 表示标准化之后的第 m 个样本的第 n 个变量。

2) 计算 \mathbf{Z} 相关系数矩阵, 具体计算公式如式 5:

$$\mathbf{R} = \mathbf{Z}^T \cdot \mathbf{Z} \quad (5)$$

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \cdots & r_{nn} \end{bmatrix}$$

矩阵中 r_{nn} 表示变量之间的相关系数, 其中下标相同的相关系数为 1。

3) 算矩阵的特征值及特征向量

对相关系数矩阵进行计算, 根据特征方程 $|\lambda I - \mathbf{R}| = 0$, 计算出特征值, 然后将特征值从大到小按顺序排列 $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \cdots \geq \lambda_n \geq 0$; 然后计算出每个特征值 λ_i 对应的特征向量 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_n$ 。

4) 计算各主成分贡献率和累计贡献率

根据上面的计算结果, 计算主成分的贡献率, 贡献率的大小代表对原信息综合反映能力, 贡献率值越大代表性越强。具体计算公式如式(6~7):

$$\delta_i = \frac{\lambda_i}{\sum_{k=1}^n \lambda_k} \quad (i = 1, 2, 3, \dots, n) \quad (6)$$

$$\zeta_i = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^n \lambda_k} \quad (i = 1, 2, 3, \dots, n) \quad (7)$$

经过计算后, 通常认为累计贡献率大于 80% 时, 主成分具有较好的代表性, 可以较为准确的反映原数据特征。

3.2. 改进的 K-均值聚类算法

K-均值聚类作为一种迭代聚类算法, 其相似程度用距离来衡量, 把需要进行分类的数据按照分类标准划分成 k 类, 通过计算每类所有点的均值确定聚类中心, 用聚类中心来表示每个类[11]。K-均值聚类算法相较于其他聚类算法更加适用于数据量较大的时候。但其也存在一定的缺点, 比如聚类结果的稳定性不好、聚类时容易陷入局部最优等。所以为了解决这些问题, 对 K-均值聚类算法进行改进, 具体步骤如下:

步骤一: 对每个样本进行计算, 求其均值当作每个类的聚类中心, 计算公式如式 8:

$$z_1 = \frac{\sum_{i=1}^n f_i}{n} \quad (8)$$

步骤二: 令 $k = k + 1$ (k 初始值为 1), 当满足 $k > K$ 时, 终止聚类。

步骤三: 定义新参数 l_m , 对每个点的参数 l_m 进行计算, 当 l_m 最大时, 选其对应的点作为下一类的初

始聚类中心。 l_m 计算公式如式 9:

$$l_m = \sum_{i=1}^n \max(d_{k-1}^i - \|f_m - f_i\|, 0) \quad (9)$$

式中 d_{k-1}^i 表示 f_i 和它所在类中心的欧氏距离。

步骤四: 本研究中 f_i 和聚类中心的距离用欧式距离来表示, 寻找与 f_i 最近的中心点, 将其划分到此类。

步骤五: 对各新类的中心点重新进行计算, $z_{ji} = \frac{\sum_{i=1}^n f_{ji}}{N}$, N 代表第 i 类中数据的数量。然后计算准则函数 F 值, 计算公式如式 10:

$$F = \sum_{j=1}^k \sum_{i=1}^n \|f_{ji} - z_i\|^2 \quad (10)$$

步骤六: 判断准则函数值 F 是否收敛, 若收敛, 进行步骤二, 若不收敛进行步骤四, 在新类中心基础上进行进一步迭代。

4. 运行工况构建

4.1. 运动学片段划分

车辆在实际道路运行时, 会受到道路交通环境的影响, 导致运行过程必然经历很多次的起步和停车。所以从运动学角度出发, 将车辆的一次启停用运动学片段来描述[12]。将一个怠速开始点到下一个怠速开始点定义为一个运动学片段, 运动学片段划分示意如图 2。

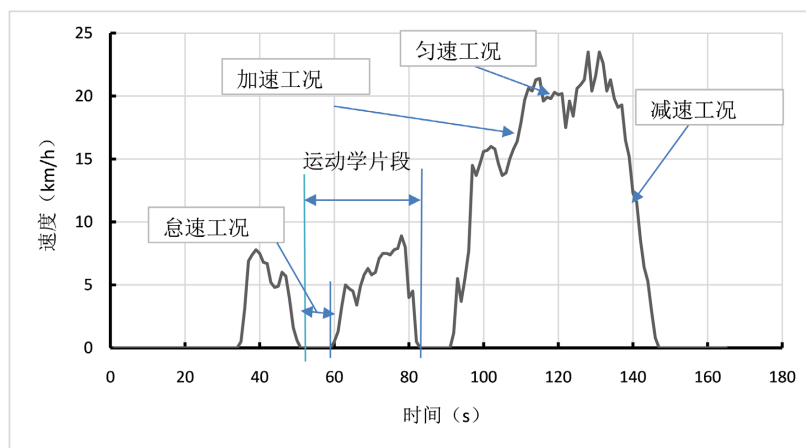


Figure 2. Schematic diagram of kinematic segment division

图 2. 运动学片段划分示意图

划分运动学片段是为了更好的分析每个运动学片的特征, 每个运动学片段包含怠速、加速、减速和匀速四种状态, 并且每种运动状态有着不同的定义方式, 具体定义方式如下:

- 1) 怠速状态: 车速为 0 并且发动机处于工作状态的车辆运行状态。
- 2) 加速状态: 车速不为 0 并且车辆加速度大于等于 0.15 m/s^2 的车辆运行状态。
- 3) 减速状态: 车速不为 0 并且车辆加速度小于等于 -0.15 m/s^2 的车辆运行状态。
- 4) 匀速状态: 车速不为 0 并且加速度的绝对值小于等于 0.15 m/s^2 的车辆运行状态。

在完成运动学状态划分后, 利用 Matlab 编写相应的程序进行运动学片段的划分, 最终将试验数据划分得到 5250 个运动学片段。完成运动学片段的划分后, 由于部分运动学片段并不适用于工况构建, 因为

部分运动学片段受到突发状况的影响，导致这些运动学片段不符合车辆正常运行情况，所以要对运动学片段进行筛选。筛选条件如表 1。

Table 1. Kinematic segment screening conditions

表 1. 运动学片段筛选条件

筛选内容	筛选条件
运动学片段时间	>10 s
速度范围(v)	5 km/h $< v < 100$ km/h
加速度持续时间	<10 s
加速度范围(a)	-5 m/s ² $< a < 5$ m/s ²
匀速持续时间	<10 s

如表 1 所示，对运动学片段时间进行限制，要求运动学片段时间大于 10 秒；加速度绝对值小于等于 5 m/s²；速度范围在 5 km/h 到 100 km/h 之间；加速度为 0 的时间不超过 10 秒；按照此筛选条件进行运动学片段的筛选，使得到的运动学片段更具代表性。在对原来的 5250 个运动学片段进行筛选后得到 4526 个运动学片段。

4.2. 特征参数选取与计算

运动学片段特征仅用速度和时间来描述不够全面，所以为了更加具体的描述运动学片段特征，需要定义一些特征参数，然后用这些特征参数定量的描述运动学片段特征。描述性特征参数如表 2，统计性特征参数如表 3。

Table 2. Descriptive feature parameter

表 2. 描述性特征参数

序号	特征参数符号	定义	单位
1	T_y	匀速时间	S
2	T_d	怠速时间	S
3	T_j	减速时间	S
4	T_s	加速时间	S
5	T	运动片段时间	S
6	V_{ym}	平均运行速度	km/h
7	V_d	最大速度	km/h
8	V_m	平均速度	km/h
9	a_x	最大减速度	m/s ²
10	a_d	最大加速度	m/s ²
11	a_m	平均加速度	m/s ²
12	a_n	平均减速度	m/s ²

Continued

13	V_{sd}	速度标准差	km/h
14	a_{sd}	加速度标准差	m/s ²

Table 3. Statistical characteristic parameters
表 3. 统计性特征参数

序号	特征参数符号	定义	单位
1	P_y	匀速状态时间比例	%
2	P_j	减速状态时间比例	%
3	P_s	加速状态时间比例	%
4	P_d	怠速状态时间比例	%
5	P_{0-10}	0 到 10 km/h 速度区间占比	%
6	P_{10-20}	20 到 30 km/h 速度区间占比	%
7	P_{20-30}	20 到 30 km/h 速度区间占比	%
8	P_{30-40}	30 到 40 km/h 速度区间占比	%
9	P_{40-50}	40 到 50 km/h 速度区间占比	%
10	P_{50-60}	50 到 60 km/h 速度区间占比	%
11	P_{60-70}	60 到 70 km/h 速度区间占比	%
12	P_{70-}	70 km/h 以上速度区间占比	%

对本研究中划分的运动学片段进行特征参数计算。得到计算结果如表 4。

Table 4. Kinematic segment characteristic parameter calculation results
表 4. 运动学片段特征参数计算结果

片段号	T_d	T_s	T_j	T_y	T	...	P_{60-70}	P_{70-}
1	11	25	17	15	68	...	0.00	0.00
2	85	103	81	72	341	...	0.10	0.00
3	128	112	95	83	418	...	0.00	0.00
4	63	72	88	132	355	...	0.21	0.00
5	95	88	79	62	324	...	0.00	0.00
6	53	125	134	207	519	...	0.15	0.00
...
4526	66	78	92	99	335	...	0.33	0.00

4.3. 主成分分析结果

根据计算得到主成分贡献率和累计贡献率如表 5, 由表可以看出主成分 1 的贡献率最高, 达到 41.17%,

说明其对原始数据的解释度最大；紧随其后的是主成分 2，其贡献率为 19.17，累计贡献率为 60.34%；主成分 3 的贡献率为 14.16%，累计贡献率为 74.50%；主成分 4 的贡献率为 9.09%，累计贡献率为 83.59%；所以前 4 个主成分的累计贡献率高于 80%，说明前四个主成分能比较全面的反映原数据特征，即将运动学片段的 14 个特征参数降维为 4 个主成分[13]。

Table 5. Principal component contribution rate and cumulative contribution rate
表 5. 主成分贡献率和累计贡献率

序号	特征值	贡献率%	累计贡献率%
1	6.25	41.17	41.17
2	2.91	19.17	60.34
3	2.15	14.16	74.50
4	1.38	9.09	83.59
5	0.77	5.07	88.66
6	0.69	4.55	93.21
7	0.31	2.04	95.25
8	0.30	1.99	97.24
9	0.15	0.99	98.23
10	0.11	0.72	98.95
11	0.09	0.59	99.54
12	0.05	0.33	99.87
13	0.02	0.13	100
14	0.00	0	100

对主成分与特征参数的相关性进行计算，得到主成分与特征参数相关性如表 6，由表 6 可以得出主成分 1 主要代表的是加速时间、运动片段时间、减速时间、平均速度、匀速时间、平均运行速度、最大速度；主成分 2 主要代表的是平均减速度、平均加速度、最大减速度、最大加速度；主成分 3 主要代表的是平均速度；主成分 4 主要代表的是怠速时间。

Table 6. Principal components are correlated with characteristic parameters
表 6. 主成分与特征参数相关性

特征值	主成分 1	主成分 2	主成分 3	主成分 4
匀速时间	0.765	0.242	-0.059	-0.305
怠速时间	0.302	0.249	0.608	0.656
减速时间	0.900	0.225	0.168	-0.121
加速时间	0.927	0.201	0.069	-0.133
运动片段时间	0.899	0.306	0.292	0.087
平均运行速度	0.759	-0.289	-0.348	0.359

Continued

最大速度	0.798	-0.359	-0.337	0.249
平均速度	0.811	-0.225	-0.495	0.019
最大减速度	-0.510	0.569	-0.289	0.169
最大加速度	0.511	-0.515	0.419	-0.260
平均加速度	-0.225	-0.568	0.489	-0.073
平均减速度	0.052	0.768	0.036	-0.152
速度标准差	0.361	0.628	0.279	0.345
加速度标准差	0.425	0.757	-0.038	-0.269

对主成分得分矩阵进行计算，结果如表 7。将其用于后续聚类分析。

Table 7. Principal component score matrix

表 7. 主成分得分矩阵

序号	主成分 1	主成分 2	主成分 3	主成分 4
1	2.6352	1.8865	-0.6948	0.5162
2	0.4256	-0.3366	0.5421	1.1562
3	1.8823	-0.3256	-0.0061	0.5053
...
4525	1.4896	0.8356	-0.4352	0.0528
4526	2.6549	-1.4201	0.2236	-3.1235

4.4. 改进的 K-均值聚类结果分析

运用改进的 K-均值聚类算法对运动学片段进行聚类，将运动学片段分成三类，聚类后各类平均特征值如表 8。由表 8 可以看出，第 1 类运动学片段的平均速度最高，怠速比例最低，代表郊区或者环城高架桥高速路况；第 2 类运动学片段平均速度和怠速比例均处于其他两类之间，代表城市道路通畅时行驶情况；第 3 类平均速度最低，怠速比例最高，代表城市道路拥堵时行驶情况。

Table 8. The average eigenvalue after clustering

表 8. 聚类后各类平均特征值

聚类类别	第 1 类	第 2 类	第 3 类
V_m	55.26	31.23	17.09
V_d	69.31	52.21	30.35
T	125	238	336
V_{sd}	5.35	6.07	5.69
a_x	-3.29	-3.18	-3.02
a_d	3.25	3.16	2.92

Continued

a_m	0.92	0.81	0.73
a_n	-0.88	-0.86	-0.76

4.5. 运行工况合成

根据中国汽车行驶工况构建标准以及车辆实际运行情况，将本研究构建的运行工况的持续时间定为 1800 s，这样能更好的反映车辆的实际道路运行情况，同时也满足将其用于排放测试时的标准。

$$P_n = \frac{T_n}{T_z} \times 100\% \quad (11)$$

式中 P_n 表示第 n 类运动学片段时间占总运动学片段时间的比例， T_n 为第 n 类运动学片段总时间， T_z 是总运动学片段时间。

通过式 11 计算得到郊区或者环城高架桥高速路况所占的比例为 15.4%，城市道路通畅时运行状态所占比例为 57.8%，城市道路拥堵时运行状态所占比例为 26.8%。根据各类运行状态所占比例，在第 1 类、第 2 类和第 3 类运动学片段库选取运动学片段用于工况构建。选取运动学片段时，通常选取聚类后与本类的特征值有最大相关性的运动学片段作为代表片段，而在本研究中，依次选取相关系数较大的运动学片段组成运动学片段库，然后用于后续的运行工况合成。

按上述方法，本研究在第 1 类运动学片段中选择 107 号片段、125 号片段、212 号片段组成第 1 类运动学片段库；在第 2 类运动学片段中选择 525 号片段、561 号片段、623 号片段、651 号片段、1662 号片段、1671 号片段、1783 号片段、2485 号片段、898 号片段、995 号片段组成第 2 类运动学片段库；在第 3 类运动学片段中选择 3521 号片段、3631 号片段、3789 号片段、3965 号片段、4021 号片段、4156 号片段、4258 号片段、4321 号片段组成第 3 类运动学片段库。根据各类运动学片段在构建工况时所占的时间比例，分别选取运动学片段来构建工况。按上面所求比例在第 3 类运动学片段库中选 1 个运动学片段；在第 2 类运动学片段库中选 5 个运动学片段；在第 1 类运动学片段库中选 4 个运动学片段；将选出的运动学片段构建成运行工况曲线如图 3。

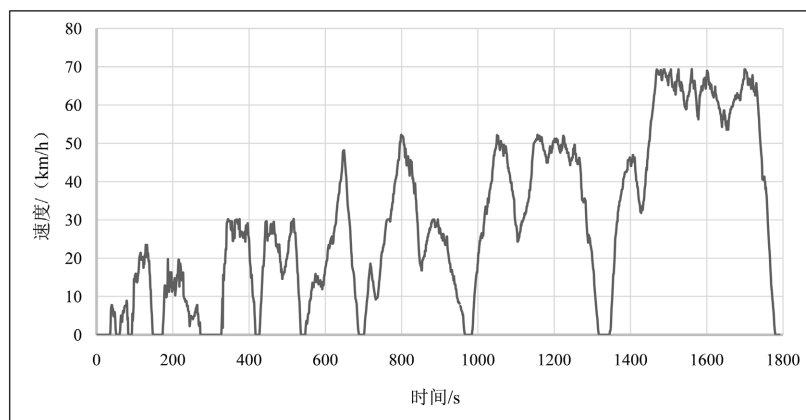


Figure 3. Vehicle driving cycle
图 3. 车辆运行工况

4.6. 工况验证

通过比较构建的车辆运行工况的特征参数值和总样本数据的特征参数值，判断构建的运行工况是否

能准确的反映车辆实际道路运行状况。选取的特征参数有平均速度 V_m 、怠速比例 P_d 、减速比例 P_j 、加速比例 P_s 、平均加速度 a_m 、平均加速度 a_n 。将这 6 个特征参数作为判定标准，把本研究构建的运行工况的特征参数值和总样本数据的特征参数值进行对比分析，结果如表 9。

Table 9. Comparison of character parameter values

表 9. 特征参数值对比

特征参数	总样本数据	构建的运行工况
平均速度 V_m	26.532	27.073
怠速比例 P_d	0.281	0.288
减速比例 P_j	0.279	0.273
加速比例 P_s	0.265	0.259
平均加速度 a_m	0.769	0.786
平均加速度 a_n	-0.756	-0.792

由表 9 可得，自主构建的运行工况和总样本数据的平均速度相对误差为 2.0%，怠速比例相对误差为 2.5%，减速比例相对误差为 2.2%，加速比例相对误差为 2.3%，平均加速度相对误差为 2.2%，平均减速度相对误差为 4.8%。由此可见本研究构建的运行工况和总体数据的特征参数值相对误差很小，可以较好地反映车辆实际道路运行情况。

5. 结论

1) 对数据预处理后进行运动学片段的划分与筛选得到 4526 个运动学片段，用主成分分析法将数据降维成 4 个主成分，再用改进的 K-均值聚类算法聚类为 3 类，最后选取每类的代表片段构建车辆运行工况。

2) 与传统的 K-均值聚类算法构建的运行工况相比，用改进的 K-均值聚类算法构建的运行工况精度更高。用改进的 K-均值聚类算法构建的运行工况和总样本数据的平均速度相对误差为 2.0%，怠速比例相对误差为 2.5%，减速比例相对误差为 2.2%，加速比例相对误差为 2.3%，平均加速度相对误差为 2.2%，平均减速度相对误差为 4.8%。说明该工况能很好的反映当地的车辆实际道路交通状况，该工况具有很高的可信度。

3) 构建的工况中郊区或者环城高架桥高速路况所占的比例为 15.4%，城市道路通畅时运行状态所占比例为 57.8%，城市道路拥堵时运行状态所占比例为 26.8%，运行工况的路况构成存在一定的地域差异，本工况的构建为后续其他城市工况构建时的地域加权提供数据依据。

参考文献

- [1] 中华人民共和国生态环境部. 中国移动源环境管理年报(2022) [R]. 北京: 中华人民共和国生态环境部, 2022.
- [2] Mora, J., Willems, F., Seykens, X., et al. (2018) An OBD Strategy to Estimate SCR Ageing and Detect Urea Injection Faults. *IFAC Papers on Line*, **51**, 369-376. <https://doi.org/10.1016/j.ifacol.2018.10.076>
- [3] Fu, L., Hao, J., He, D., et al. (2001) Assessment of Vehicular Pollution in China. *Journal of the Air & Waste Management Association*, **51**, 58-668. <https://doi.org/10.1080/10473289.2001.10464300>
- [4] 路尧. 中国轻型汽车行驶工况开发[D]: [硕士学位论文]. 北京: 北京理工大学, 2016.
- [5] 彭育辉, 杨辉宝, 李孟良, 等. 基于 K-均值聚类分析的城市道路汽车行驶工况构建方法研究[J]. *汽车技术*, 2017, 506(11): 13-18.

- [6] 曹骞, 李君, 刘宇, 等. 基于马尔科夫链的长春市乘用车行驶工况构建[J]. 吉林大学学报(工学版), 2018, 48(5): 1366-1373.
- [7] 郭家琛, 姜衡, 雷世英, 等. 城市道路汽车行驶工况构建方法[J]. 交通运输工程学报, 2020, 20(6): 1671-1673.
- [8] 段宇帅. 基于主成分分析与 K-Means 聚类的汽车行驶工况构建[J]. 软件导刊, 2022, 21(5): 175-180.
- [9] 周溪召, 刘启超. 基于 K-Means 聚类分析的汽车行驶工况构建[J]. 物流科技, 2020, 43(11): 93-96.
- [10] 宋怡帆. 基于聚类和 Python 语言的深圳市道路车辆行驶工况构建[D]: [硕士学位论文]. 西安: 长安大学, 2018.
- [11] 丁一峰. 长春市轻型车城市道路行驶工况研究[D]: [硕士学位论文]. 长春: 吉林大学, 2018.
- [12] 赵慧, 张镇顺, 熊永达. 香港城区汽车行驶工况的研究[J]. 环境科学学报, 2000, 20(3): 312-315.
- [13] 胡志远, 秦艳, 谭丕强, 等. 基于大样本的上海市乘用车行驶工况构建[J]. 同济大学学报(自然科学版), 2015, 43(10): 1523-1527.