

高建平,任德轩,郝建国. 基于全局 K-means 聚类算法的汽车行驶工况构建[J]. 河南理工大学学报(自然科学版) 2019, 38(1): 112-118. doi: 10.16186/j.cnki.1673-9787.2019.1.17

GAO J P, REN D X, XI J G. Construction of vehicle driving cycle based on global K-means clustering algorithm[J]. Journal of Henan Polytechnic University( Natural Science) 2019, 38(1): 112-118. doi: 10.16186/j.cnki.1673-9787.2019.1.17

## 基于全局 K-means 聚类算法的汽车行驶工况构建

高建平,任德轩,郝建国

(河南科技大学 车辆与交通工程学院,河南 洛阳 471003)

**摘要:** 为构建符合郑州市交通特征的乘用车典型行驶工况,选取 60 辆乘用车进行数据采集,将采集的有效行驶数据划分成 120 744 条运动学片段,采用主成分分析法对构建的特征参数矩阵进行降维、非线性化处理。选取戴维森-堡丁指数来确定最佳聚类数,并通过全局 K-means 聚类算法将主成分分析得到的 4 个主成分聚成 3 类,然后采用相关系数法从各类片段库中选取具有代表性的运动学片段,从而构建出郑州市乘用车典型循环工况(ZZ\_DC)。通过与传统 K-means 聚类算法构建的工况比较,采用全局 K-means 聚类构建的工况更加准确地反映了郑州市乘用车在实际道路上的行驶状态。将 ZZ\_DC 工况与国内一直采用的乘用车测试工况(NEDC 工况)比较,结果表明,ZZ\_DC 工况与 NEDC 工况差异显著,因此,有必要开发出适用于郑州市交通特征的乘用车行驶工况。

**关键词:** 乘用车; 行驶工况; 主成分分析法; 全局 K-means 聚类

中图分类号: U491.2

文献标志码: A

文章编号: 1673-9787(2019)1-112-7

### Construction of vehicle driving cycle based on global K-means clustering algorithm

GAO Jianping, REN Dexuan, XI Jianguo

(School of Vehicle and Traffic Engineering, Henan University of Science and Technology, Luoyang 471003, Henan, China)

**Abstract:** In order to construct the typical driving cycle of the passenger cars to meet the traffic characteristics of Zhengzhou city, 60 passenger cars were selected to collect the data, then the effective driving data were divided into 120 744 kinematics fragments, and principal component analysis method was used to reduce the dimensions and to execute the nonlinear processing of the constructed characteristic parameter matrix. Davies-Bouldin index(DB) was chosen to determine the optimal cluster number, and the four principal components obtained by the principal component analysis method were clustered into three categories by the global K-means clustering algorithm, then the representative kinematics fragments were selected from fragment libraries by means of the correlation coefficient method, thus the typical driving cycle of Zhengzhou passenger cars(ZZ\_DC) was constructed. Compared with the driving cycle constructed by the traditional K-means clustering algorithm, the driving cycle constructed by the global K-means clustering could more accurately reflect the driving status of the passenger cars on the actual roads in Zhengzhou. The ZZ\_DC was compared with the test driving

收稿日期: 2018-04-19; 修回日期: 2018-06-23

基金项目: 国家自然科学基金资助项目(U1604147); 河南科技大学研究生创新基金资助项目(CXJJ-2016-ZR01)

第一作者简介: 高建平(1976—),男,河南洛阳人,博士,副教授,主要从事新能源汽车整车控制方面的教学和研究工作。

E-mail: gjpcar@163.com

cycle of the domestic passenger cars( NEDC) ,and the results showed that the differences between the ZZ\_DC and the NEDC were significant. Therefore it was necessary to develop the driving cycle suitable for the traffic characteristics of Zhengzhou.

**Key words:** passenger car; driving cycle; principal component analysis method; global K-means clustering

## 0 引言

近年来,我国乘用车保有量增长迅猛,同时能源危机和汽车污染物排放问题日益凸显,而车辆的行驶工况<sup>[1]</sup>是反映汽车污染物排放的重要特征。在一个确定地区,车辆行驶工况可为汽车污染物排放量和燃油消耗量的测试以及新车型的技术开发、评估等提供参考依据。目前,国内外学者对车辆的行驶工况进行了深入研究,并拟合出了符合各地区交通特征的汽车行驶工况<sup>[2-3]</sup>。在行驶工况构建的过程中,大多数学者多采用主成分分析法与 K-means 聚类算法相结合的研究方法,但传统 K-means 聚类算法的初始聚类中心多为随机选取,易导致聚类结果陷入局部最优,从而影响工况构建的准确度。胡宸等<sup>[4]</sup>,胡志远等<sup>[5]</sup>采用主成分分析与聚类算法相结合,分别构建了哈尔滨市、上海市的典型车辆行驶工况,但在数据聚类中均没有对聚类中心进行寻优。石琴等<sup>[6]</sup>采用神经网络算法改进了 K-means 聚类算法的初始聚类中心,但神经网络算法初始权值的设置会影响学习效果,且未对最佳聚类数 K 值进行确定,不确定的 K 值也导致聚类质量的下降,从而进一步影响车辆行驶工况构建的精度。

截至 2016 年末,郑州市汽车保有量已超过 260 万辆<sup>[7]</sup>,但针对郑州市车辆行驶工况的构建及分析,尤其是对乘用车行驶工况的构建并不多见。基于上述分析,本文以在郑州市采集的 1 200 余万条有效车辆行驶数据为基础,进行运动学片段划分和全局 K-means 聚类。采用戴维森-堡丁指数( Davies-Bouldin index ,DB) 确定最优聚类数,通过全局 K-means 聚类得到各类中的最佳聚类中心,并将运动学片段聚成 3 类片段库。从各类片段库中选取最优片段,从而构建出郑州市车辆典型行驶工况。通过与采集的总样本数据进行对比分析,验证了所构建车辆行驶工况的准确性。

## 1 车辆行驶数据采集

常用的车辆行驶数据采集方法主要有车辆追踪法、平均车流量统计法、自主行驶法等<sup>[8]</sup>。其中,自主行驶法不需要提前规划好特定的试验行

驶线路,由车主按正常驾驶习惯自主驾驶。为使采集的数据更有效、更能充分表达郑州市的交通特性,选取的车辆应考虑车主所在位置、以及车主每天行驶道路的状况,因此,所选择车辆的每日行驶线路中应含有住宅区、学校、车站、公园等一些交通互动比较频繁的地方。图 1 中的粗黑线即为一条车辆经常行驶的典型线路。



图 1 车辆典型行驶线路

Fig. 1 Typical route for passenger car

采用自主行驶法在郑州市选取 60 辆乘用车进行数据采集,数据采集装置如图 2 所示,包含 CAN 通讯总线和 GPS 模块,采样频率为 1 Hz。通过 CAN 总线可读取车辆的实时油耗、发动机转速、油门踏板开度等信息,通过 GPS 模块可采集到车辆的实时位置(经纬度)、车速等信息,采集车速的分辨率为  $0.1 \text{ km} \cdot \text{h}^{-1}$ ,满足数据采集精度的要求。



图 2 数据采集装置

Fig. 2 Devices of data acquisition

## 2 主成分分析与全局 K-means 聚类

本文结合郑州市的交通特征和车辆的行驶特点,采用主成分分析与全局 K-means 聚类算法相结合进行车辆行驶工况构建,流程如图 3 所示。

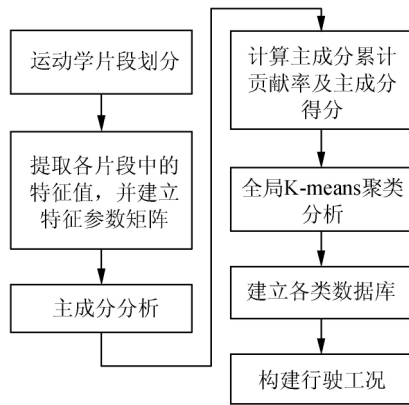


图3 车辆行驶工况构建流程

Fig. 3 Constructing process of driving cycle

### 2.1 运动学片段划分及特征参数提取

为模拟车辆在实际行驶过程中的频繁启动、加速、减速等行驶状态,把汽车从一个怠速状态开始行驶到下一个怠速状态之间的行驶过程定义为一个运动学片段,通常由一个怠速部分和一个运动部分构成,图4所示为运动学片段示意图。

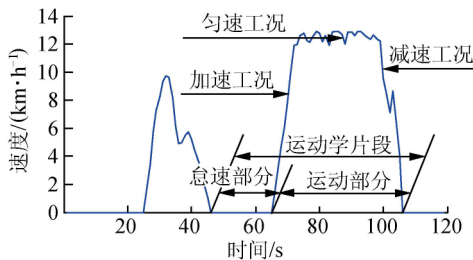


图4 运动学片段示意图

Fig. 4 Kinematics fragment sketch

为充分表达各运动学片段的特征,选取12个具有代表性的特征参数,如表1所示。在MATLAB中编程,对总样本数据进行运动学片段划

分,共获取120 744条运动学片段,提取每条运动学片段的特征参数,得到特征参数矩阵 $M_{120\,744 \times 12}$ ,如表2所示。

### 2.2 主成分分析法

在车辆行驶工况构建的过程中,需根据上述特征参数值,对运动学片段进行聚类。但选取的特征参数中,有些变量之间并不相互独立,存在一定的冗余性,如运行距离、运行时间、平均速度3个参数之间存在着一定的关联,这样使所表达的信息存在重叠;而如果仅采用一两个特征参数进行分析,又会导致信息丢失的情况。需对特征值参数矩阵 $M_{120\,744 \times 12}$ 进行主成分分析,主成分分析法主要用于大数据的降维以及通过几个互不相关的综合主成分变量完全表达原有的信息量。

表1 典型特征参数

Tab. 1 Typical characteristic parameters

名称	符号
运行距离/m	$L$
平均速度/( $\text{km} \cdot \text{h}^{-1}$ )	$V_m$
最大速度/( $\text{km} \cdot \text{h}^{-1}$ )	$V_{\max}$
运行时间/s	$T$
怠速时间/s	$T_i$
加速时间/s	$T_a$
减速时间/s	$T_d$
匀速时间/s	$T_c$
最大加速度/( $\text{m} \cdot \text{s}^{-2}$ )	$a_{\max}$
最小加速度/( $\text{m} \cdot \text{s}^{-2}$ )	$a_{\min}$
加速段的平均加速度/( $\text{m} \cdot \text{s}^{-2}$ )	$A_a$
减速段的平均减速度/( $\text{m} \cdot \text{s}^{-2}$ )	$A_d$

表2 运动片段特征参数矩阵 $M_{120\,744 \times 12}$

Tab. 2 Characteristic parameter matrix  $M_{120\,744 \times 12}$  of kinematics fragments

片段序号	$L$	$V_m$	$V_{\max}$	$T$	$T_i$	$T_a$	$T_d$	$T_c$	$a_{\max}$	$a_{\min}$	$A_a$	$A_d$
1	478.56	6.47	11.81	66	8	29	22	15	2.78	-3.50	0.64	-0.87
2	34.42	0.91	3.28	21	17	6	10	5	1.69	-0.64	0.69	-0.39
3	138.42	4.47	8.42	23	8	11	7	5	2.00	-1.69	0.76	-1.17
4	612.53	5.42	14.22	71	42	22	33	16	1.89	-1.11	0.74	-0.49
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
120 744	33.69	1.35	3.28	20	5	12	6	2	0.83	-1.47	0.35	-0.70

将表2中的特征参数矩阵用 $M$ 表示,即

$$M_{p \times q} = \begin{bmatrix} M_{11} & M_{12} & M_{13} & \cdots & M_{1q} \\ M_{21} & M_{22} & M_{23} & \cdots & M_{2q} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ M_{p1} & M_{p2} & M_{p3} & \cdots & M_{pq} \end{bmatrix}, \quad (1)$$

式中, $M_{ij}$  ( $i=1, 2, 3, \dots, p$ ;  $j=1, 2, 3, \dots, q$ ) 为第*i*

个运动学片段中的第*j*个特征参数。

在所选取的特征变量中,有些变量的单位量纲不同,导致变量取值的分散程度较大,这时数据的变化幅度会影响到聚类的效果。因此,特征参数在主成分分析前需进行标准化处理,去除不同量纲可能带来的影响。本文采用正态分布标准化

法<sup>[9]</sup>,即

$$M_{ij}^* = \frac{M_{ij} - \bar{M}_{ij}}{\delta}, \quad (2)$$

$$\bar{M}_{ij} = \frac{1}{p} \sum_{i=1}^p M_{ij}, \quad (3)$$

$$\delta = \sqrt{\frac{1}{p-1} \sum_{i=1}^p (M_{ij} - \bar{M}_{ij})^2}, \quad (4)$$

式中,  $M_{p \times q}^*$  为标准化后的特征矩阵。由  $M_{p \times q}^*$  可以得到其协方差矩阵  $Q$  为

$$Q = \begin{bmatrix} \text{Cov}(1, 1) & \text{Cov}(1, 2) & \text{Cov}(1, 3) & \cdots & \text{Cov}(1, q) \\ \text{Cov}(2, 1) & \text{Cov}(2, 2) & \text{Cov}(2, 3) & \cdots & \text{Cov}(2, q) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(q, 1) & \text{Cov}(q, 2) & \text{Cov}(q, 3) & \cdots & \text{Cov}(q, q) \end{bmatrix}, \quad (5)$$

$$\text{Cov}(x, y) = \text{Cov}(y, x) = \frac{1}{n-1} \sum_{i=1}^m (x_i - x_{\text{mean}})(y_i - y_{\text{mean}}),$$

式中,  $\text{Cov}(x, y)$  为矩阵  $M_{p \times q}^*$  的第  $x$  列与第  $y$  列的协方差。

相关系数矩阵定义为  $R$ ,

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} & \cdots & R_{1q} \\ R_{21} & R_{22} & R_{23} & \cdots & R_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_{q1} & R_{q2} & R_{q3} & \cdots & R_{qq} \end{bmatrix}, \quad (6)$$

式中,  $R_{11} = R_{22} = R_{33} = \cdots = R_{qq} = 1$ ,  $R_{xy} = \frac{\text{Cov}(p, q)}{\delta_p \delta_q}$ 。

求取矩阵  $R$  的特征值  $\lambda_i$  ( $i=1, 2, 3, \cdots, 12$ ), 前  $T$  个主成分的累积贡献率为

$$\psi = \frac{\sum_{i=1}^T \lambda_i}{\sum_{i=1}^{12} \lambda_i}. \quad (7)$$

通常选取前几个累积贡献率大于 85% 的主成分。从表 3 可以得出, 前 4 个主成分的累计贡献率为 85.89%, 超过了 85%, 能够较好反映出原有信息。计算 4 个主成分与 12 个运动学片段特征参数间的相关系数, 解析出 4 个主成分分别代表哪些特征参数, 表 4 列出了部分运动学片段的得分矩阵。然后对主成分得分矩阵进行聚类。

### 2.3 全局 K-means 聚类算法

针对大数据集的聚类, 与其他聚类算法相比, K-means 聚类算法效率更高。传统 K-means 算法所得聚类结果受初始聚类中心的影响, 往往只是一个局部最优而非全局最优, 导致聚类结果不稳定, 且易受噪声点的影响<sup>[10]</sup>。而全局 K-means 聚

表 3 主成分特征值、贡献率及累积贡献率

Tab.3 Eigenvalues, contribution rates and cumulative contribution rates of principal components

序号	特征值	贡献率/%	累计贡献率/%
1	5.646	46.63	46.63
2	2.687	22.19	68.82
3	1.112	9.18	78.01
4	0.954	7.87	85.89
5	0.607	5.01	90.90
$\vdots$	$\vdots$	$\vdots$	$\vdots$
12	0.000	0.00	100.00

表 4 主成分得分矩阵

Tab.4 Score matrix of principal components

序号	1	2	3	4
1	-0.78	-1.51	0.89	-0.33
2	2.67	1.25	0.38	0.88
3	1.48	-1.18	0.89	-0.18
4	-0.37	0.35	-0.15	0.06
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
120 744	2.48	1.22	0.93	-0.73

类算法是在传统 K-means 的基础上改进的, 克服了传统 K-means 聚类算法的缺点, 比传统 K-means 算法具有更好的鲁棒性<sup>[11]</sup>, 因此, 本文选取全局 K-means 聚类算法对得分矩阵进行聚类。

全局 K-means 聚类算法是将  $K$  个簇的聚类问题转换成一系列子聚类问题, 在每一步迭代过程结束后, 用一种增量的方法选出下一簇新的最佳初始中心<sup>[13]</sup>。假定一组数据集为  $F = \{f_1, f_2, f_3, \cdots, f_n\}$ ,  $f_j \in R^D$  ( $j=1, 2, \cdots, n$ ), 将此数据集聚成  $K$  类, 每类聚类中心为  $C = \{c_1, c_2, c_3, \cdots, c_k\}$ ,  $c_i \in R^D$  ( $i=1, 2, \cdots, k$ ), 采用聚类误差平方和函数  $E$  作为聚类准则函数, 定义为

$$E = \sum_{i=1}^k \sum_{j=1}^n \|f_{ij} - c_i\|^2. \quad (8)$$

通过反复迭代聚类, 寻求最优聚类中心, 直至聚类准则函数值收敛。具体过程如下。

Step1 [初始化]: 计算数据集中所有样本的均值, 作为第一簇的聚类中心, 即  $c_1 = \frac{1}{n} \sum_{j=1}^n f_j$ , 并设  $k=1$ 。

Step2 [设置终止条件]:  $k = k+1$ , 如果  $k > K$  聚类终止。

Step3 [选择下一簇初始聚类点]: 在选择下一簇初始聚类中心时, 引入一个参数  $a_m$ , 计算数据集  $F = \{f_1, f_2, f_3, \dots, f_n\}$  中每个点处的参数  $a_m$  的值, 选择使  $a_m$  值达到最大值的那个样本点  $f_m$  作为下一簇的初始聚类中心。 $a_m$  定义为

$$a_m = \sum_{j=1}^n \max(d_{k-1}^j - \|f_m - f_j\|, 0), \quad (9)$$

式中,  $d_{k-1}^j$  为  $f_j$  与其所属类别中心的欧氏距离。

Step4 [分配数据样本集  $f_j$ ]: 本文使用欧氏距离表示  $f_j$  与聚类中心点间的距离, 找到与  $f_j$  距离最短的中心点, 并将  $f_j$  分配到该类。

Step5 [更新簇中心并计算  $E$  值]: 重新计算各簇的中心点,  $c_{ij} = \frac{1}{N} \sum_{j=1}^N f_{ij}$ ,  $N$  为第  $i$  簇中数据样本量。并计算准则函数  $E$  值。

Step6 [判断  $E$  值是否收敛]: 若准则函数  $E$  值收敛, 返回到 Step2, 否则返回到 Step4, 基于新的簇中心做进一步迭代。

在聚类之前要确定最优聚类数  $K$ , 本文选取 DB 评价指标, 该指标进行类数估计时取其最小值对应的类数作为最佳聚类数。DB 指标定义如下<sup>[12]</sup>,

$$DB(k) = \frac{1}{k} \sum_{i=1}^k \max_{j=1 \sim k, j \neq i} \left( \frac{W_i + W_j}{C_{ij}} \right), \quad (10)$$

表5 聚类后各类别的平均特征值

Tab. 5 Average eigenvalues of each category after clustering

聚类类别	$V_m$	$V_{\max}$	$T$	$T_i$	$a_{\max}$	$a_{\min}$	$A_a$	$A_d$
I 类	7.27	21.22	34.96	32.10	1.63	-1.29	0.66	-0.61
II 类	19.97	36.88	66.37	30.74	2.31	-2.19	0.71	-0.86
III 类	26.03	53.13	188.56	31.94	2.25	-2.23	0.55	-0.63

### 3 车辆行驶工况构建及对比验证

#### 3.1 工况构建

在每个聚类中选取与该类特征值相关系数最大的代表性运动学片段构建车辆的行驶工况。通常所构建的典型城市循环工况时长  $T$  约 1 200 s, 可通过各类总运行时间在总体数据中所占的时间比确定各类在最终拟合工况中所占的时间<sup>[13]</sup>, 定义为

$$T_k = \frac{T_{\text{duringcondition}}}{T_{\text{all}}} \sum_{i=1}^{N_k} T_{k,i}, \quad (11)$$

式中:  $T_k$  为第  $k$  类在最终合成车辆行驶工况中所占时间;  $T_{\text{all}}$  为整体样本所持续时间;  $T_{\text{duringcondition}}$  为所要构建车辆行驶工况持续的时间;  $N_k$  为第  $k$

式中:  $W_i$  为聚类  $C_i$  的所有样本到其聚类中心的平均距离;  $C_{ij}$  为状态  $C_i$  和状态  $C_j$  中心之间的距离。

求取在不同聚类状态数下的 DB 值并进行插值拟合处理, 如图 5 所示。由图 5 可知, 当聚成 3 类时, DB 取值最小。

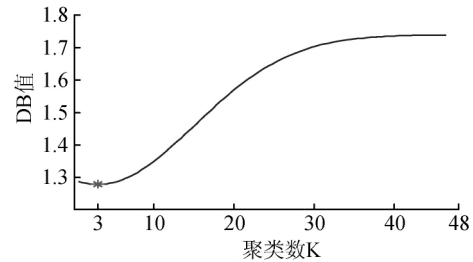


图5 类别状态数

Fig. 5 Class status number

通过上述全局 K-means 聚类, 把运动学片段聚成 I、II、III 3 类, 表 5 为聚类后各类别的平均特征值。由表 5 可知, 第 I 类运动学片段的平均速度和平均最大速度均较低, 怠速时间占总运行时间比例较高, 因此, 第 I 类为低速片段。第 II 类运动学片段的平均速度、平均最大速度及怠速比例均处于中间, 因此, 第 II 类片段为中速片段。第 III 类运动学片段的平均速度和平均最大速度均较高, 且怠速比例较低, 为高速运行片段。

类中短行程的总条数;  $T_{k,i}$  为第  $k$  类中第  $i$  条运动学片段的运行时间。

通过上述公式计算取整后, 获得 3 条 I 类低速片段 6 条 II 类中速片段 2 条 III 类高速片段, 最终构建出持续时间为 1 240 s、最高车速为  $76.5 \text{ km} \cdot \text{h}^{-1}$  的郑州市典型车辆行驶工况 (ZZ-DC), 如图 6 所示。

#### 3.2 工况验证

将采集的行驶数据作为总体样本数据, 选取平均速度  $V_m$ 、加速段的平均加速度  $A_a$ 、减速段的平均加速度  $A_d$ 、怠速比例  $P_i$ 、加速比例  $P_a$ 、减速比例  $P_d$  6 个特征参数作为判定准则, 将基于全局 K-means、传统 K-means 构建的车辆行驶工况与总体样本数据进行对比, 各工况下的特征参数值如

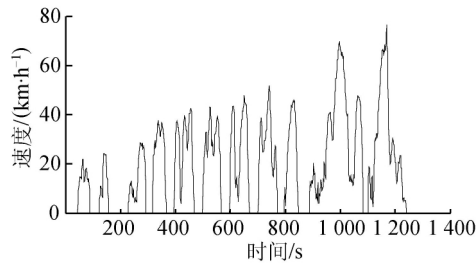


图 6 郑州市典型车辆行驶工况

Fig. 6 Typical driving cycle of passenger cars in Zhengzhou 表 6 所示。

由表 6 可知, 基于全局 K-means 聚类算法构建的车辆行驶工况与总体样本数据的特征参数值基本一致, 其中加速比例这一特征参数的相对误差最大, 但也仅为 2.6%, 各特征参数值的平均相对误差为 1.61%, 而基于传统 K-means 聚类算法构建的车辆行驶工况与总体样本数据的平均速度值相对误差为 12.46%, 各特征参数值的平均相对误差为 5.43%, 明显高于 ZZ\_DC 工况与总样本数据之间的相对误差。因此, 采用全局 K-means 聚类算法显著提高了工况构建的精度, 所构建的 ZZ\_DC 工况可以更准确地反映郑州市车辆的行驶状态。

表 6 各工况下的特征参数值

Tab. 6 Characteristic parameter values under various driving cycles

特征参数	总体样本数据	全局 K-means 构建的工况 (ZZ_DC)	传统 K-means 构建的工况
$V_m$	19.100	18.920	16.720
$A_a$	0.649	0.643	0.641
$A_d$	-0.713	-0.698	-0.654
$P_i$	0.298	0.305	0.315
$P_a$	0.269	0.276	0.261
$P_d$	0.263	0.262	0.256

将 ZZ\_DC 工况与国外标准测试工况进行比较, 结果如图 7 所示。从图 7 可以得到, ZZ\_DC 工况的平均速度比 FTP15、NEDC 标准工况分别低 44.47% 和 43.69%, 而与 Janpan10-15 工况相近, 怠速比例却远高于 3 种标准工况。ZZ\_DC 工况与 Janpan10-15 工况的加速比例、减速比例均相近, 而与 NEDC、FTP15 标准工况相比, 差异较明显。我国车辆的油耗与排放标准是采用等同欧洲 NEDC 工况的 I 型试验工况来制定的, 而郑州市车辆行驶工况的平均速度、怠速比例等特征参数与 NEDC 标准测试工况相比差异显著, 因此, NEDC 工况并不适用于郑州市车辆的测试, 有必

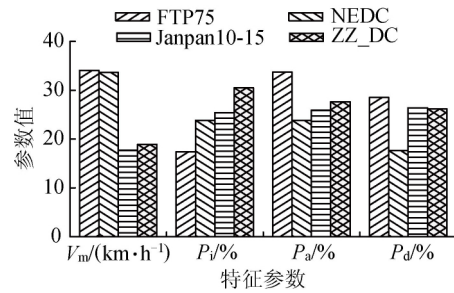


图 7 ZZ\_DC 工况与国外标准测试工况对比

Fig. 7 Comparison of ZZ\_DC with standard test driving cycle abroad

要开发出符合本地实际交通特性的车辆行驶工况。

## 4 结 语

(1) 以 60 辆车的行驶工况作为研究对象, 划分为 120 744 个运动学片段, 采用主成分分析和全局 K-means 聚类算法对特征参数矩阵进行降维和分类处理, 从而构建出了符合郑州市交通特征, 时长 1 240 s 的郑州市车辆行驶工况。

(2) 与传统 K-means 聚类算法相比, 全局 K-means 聚类算法所构建的车辆行驶工况精度更高。郑州市车辆行驶工况与欧洲 NEDC 工况相比, 存在较大差异, 尤其在平均速度、怠速比例等方面差距显著, 因此, 采用 NEDC 工况进行油耗与污染物的排放测试并不能真实反映郑州市的交通状况, 有必要开发出符合本地实际交通特性的车辆行驶工况。

(3) 混合动力汽车的动力系统匹配、控制策略的优化等严重依赖车辆的行驶工况, 不同的行驶工况所匹配、优化的结果不同, 其经济性差异也非常显著, 因此, 基于 ZZ\_DC 工况对混合动力汽车动力系统匹配、控制策略优化等方面可做进一步的研究。

## 参考文献:

- [1] 万霞, 黄文伟, 强明明. 深圳市乘用车道路行驶工况构建[J]. 深圳大学学报(理工版), 2016, 33(3): 281-287.  
WAN X, HUANG W W, QIANG M M. Construction of driving cycle for passenger vehicles in Shenzhen[J]. Journal of Shenzhen University(Science and Engineering), 2016, 33(3): 281-287.
- [2] FOTOUHI A, MONTAZERIGH M. Tehran driving cycle development using the K-means clustering method[J]. Scientia Iranica A, 2013(20): 286-293.
- [3] 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(11):13-48.
- PENG Y H ,YANG H B ,LI M L ,et al. Research on the construction method of driving cycle for the city car based on K-means cluster analysis [J]. Automobile Technology 2017(11) 13-48.
- [4] 胡宸,吴晓刚,李晓军,等. 哈尔滨城市公交工况的构建[J]. 哈尔滨理工大学学报 2014,19(1):85-89.
- HU C ,WU X G ,LI X J ,et al. Construction of Harbin city driving cycle [J]. Journal of Harbin University of Science and Technology 2014,19(1):85-89.
- [5] 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.
- HU Z Y ,QIN Y ,TAN P Q ,et al. Large-sample-based car-driving cycle in Shanghai city [J]. Journal of Tongji University( Natural Science) ,2015,43(10):1523-1527.
- [6] 石琴,仇多洋,周洁瑜. 基于组合聚类法的行驶工况构建与精度分析[J]. 汽车工程 2012,34(2):164-169.
- SHI Q ,QIU D Y ,ZHOU J Y. Driving cycle construction and accuracy analysis based on combined clustering technique [J]. Automotive Engineering 2012,34(2):164-169.
- [7] 公安部交通管理局. 2016年全国机动车和驾驶人保持快速增长. 2017-01-10. <http://www.mps.gov.cn/n2255040/n4908728/c5595634/content.html>.
- TRAFFIC MANAGEMENT BUREAU OF THE PUBLIC SECURITY MINISTRY. The nation's motor vehicles and drivers maintained rapid growth in 2016. 2017-01-10. <http://www.mps.gov.cn/n2255040/n4908728/c5595634/content.html>.
- [8] 马志雄,李孟良,朱西产,等. 乘用车实际行驶工况开发方法的研究[J]. 武汉理工大学学报(信息与管理工程版) 2004,26(3):182-184.
- MA Z X ,LI M L ,ZHU X C ,et al. Study of the methodology for car driving cycle development [J]. Journal of Wuhan University of Technology( Information & Management Engineering) 2004,26(3):182-184.
- [9] 黄万友,程勇,李闯. 基于车辆能耗状态的济南市道路行驶工况构建[J]. 西南交通大学学报 2012,47(6):989-995.
- HUANG W Y ,CHENG Y ,LI C. Driving cycle construction of city road based on vehicles energy consumption in Jinan [J]. Journal of Southwest Jiaotong University, 2012,47(6):989-995.
- [10] 赵丽. 全局K-均值聚类算法研究与改进[D]. 西安:西安电子科技大学 2013.
- ZHAO L. Research and improvement of the global K-means clustering algorithm [D]. Xi'an: Xidian University 2013.
- [11] ARISTIDIS L ,NIKOS V ,JAKOB. The global K-means clustering algorithm [J]. Pattern Recognition, 2003,36:451-461.
- [12] 周士兵. 聚类分析中的最佳聚类数确定方法研究应用[D]. 无锡:江南大学 2011.
- ZHOU S B ,Research and application on determining optimal number of clusters in cluster analysis [D]. Wuxi: Jiangnan University 2011.
- [13] 秦大同,詹森,漆正刚,等. 基于K-均值聚类算法的行驶工况构建方法[J]. 吉林大学学报(工学版), 2016,46(2):383-389.
- QIN D T ,ZHAN S ,QI Z G ,et al. Driving cycle construction using K-means clustering method [J]. Journal of Jilin University( Engineering and Technology Edition) 2016,46(2):383-389.

(责任编辑 袁兴起)