

基于数据的 K 均值理论及其应用

杨 金¹, 陈 林¹, 周 强², 陈建勋¹

(1. 重庆交通大学数学与统计学院, 重庆 400074)

(2. 重庆南开中学, 重庆 400074)

[摘要] 首先使用主成分分析方法对车辆行驶工况的特征参数进行处理, 然后利用 K -均值聚类原理对所有具有代表性的运动学片段进行聚类分析, 由此得出汽车行驶工况的数学原理. 为汽车节能减排的发展和建立能够正确反映我国道路工况特征的行驶工况提供理论依据.

[关键词] K -均值聚类, 主成分分析, 数据处理, 特征参数

[中图分类号] O175.13 [文献标志码] A [文章编号] 1001-4616(2021)02-0010-08

Application of K -Means Theory Based on Data

Yang Jin¹, Chen Lin¹, Zhou Qiang², Chen Jianxun¹

(1. Chongqing Jiaotong University, Chongqing 400074, China)

(2. Chongqing Nankai Middle School, Chongqing 400074, China)

Abstract: In this paper, the principal component analysis method is used to process the characteristic parameters of vehicle driving conditions, and the K -means clustering principle is used to cluster all the representative kinematics fragments, and then the mathematical principle of vehicle driving conditions is obtained. It provides a theoretical basis for the development of automobile energy saving and emission reduction and the establishment of driving conditions that can accurately reflect the characteristics of road conditions in China.

Key words: K -means cluster, principal component analysis, data processing, characteristic parameter

本世纪初,我国直接采用欧洲的 NEDC(新标欧洲循环测试)行驶工况对汽车产品能耗/排放的认证,有效促进了汽车节能减排和技术的发展^[1]. 但随着我国汽车保有量的迅速增长,交通状况的逐渐变化,继续以 NEDC 工况为基准所优化标定的汽车,其法规认证结果与使用过程中的实际油耗量相比较,差距越来越大^[2]. 因而,制定能反映我国实际道路行驶状况的行驶工况迫在眉睫. 本文主要采集了某城市同一辆轻型汽车在不同时间段共 21 d 所获得的数据,针对人工筛选的不足,首先对数据时间不连续、加、减速度异常、长期停车、长时间堵车、怠速 4 种情况的异常数据进行处理,然后对运动学片段进行划分,再通过主成分分析方法对车辆行驶工况的特征参数进行处理,根据选出的特征参数利用 K -均值聚类分析方法将所有代表性运动学片段按低速、中速和高速进行聚类,最终拟合成汽车行驶工况的全过程,由此建立了某市同一辆轻型汽车合理的行驶工况,其设计方案简单合理,具有很好的实用价值.

1 预备知识

由于 GPS 信号不良、数据采集设备故障、堵车等状况造成部分无效和异常数据,必须对无效和异常数据进行数据处理. 本文主要考虑数据时间不连续、加、减速度异常、长期停车、长时间堵车、怠速 4 种情况的异常数据,接下来我们给出上述 4 种情况的定义及其处理方式.

1.1 数据处理

为使采集的数据更加全面,本文利用可插入 GPRS 手机卡和 SD 储存卡的记录仪采集某城市同一轻型

收稿日期:2020-10-26.

基金项目:国家自然科学基金项目(11801047)、重庆市自然科学基金项目(cstc2019jcyj-msxmX0755)、重庆市研究生导师团队建设项目(JDDSTD201802)、重庆市教委科学技术研究项目(自然科学类)(KJQN201900707)、重庆市教育科学规划课题(2020-07-203).

通讯作者:杨金,博士,副教授,研究方向:生物数学. E-mail:seehom@126.com

车不同时间段在实际道路行驶的数据. 采样频率为 1 Hz, 将采集到的 GPS 车速、 X 轴加速度、 Y 轴加速度、 Z 轴加速度、发动机转速、扭矩百分比、瞬时油耗、油门踏板开度、空燃比、发动机负荷百分比、进气流量等数据从 SD 卡中导出.

数据时间不连续: 往往是由汽车行驶过程中 GPS 信号丢失导致的. 由于数据以秒为单位, 且数据量往往是几十万甚至上百万, 人工筛选工作量过于巨大, 因此选择编写计算机程序筛选出非连续数据. 此时需要对非连续数据进行定义及处理, 具体定义及处理方式如下:

令 t_i, t_{i+1} 为相邻两时间点且 $t_i < t_{i+1}$, v_i, v_{i+1} 分别代表 t_i, t_{i+1} 时刻的速度, 若 $(t_{i+1} - t_i) > 1$, 则定义 t_i, t_{i+1} 为时间数据不连续时刻. 此时, 包含 t_i 时刻的连续行驶片段与包含 t_{i+1} 时刻的连续行驶片段时间上不连续, 如果删除 t_i 或 t_{i+1} 时间点得到的数据集依然是不连续的, 那么需要对这个时间点附近的数据进行处理. 找到距离 t_i 时间最近且 GPS 车速为 0 的时间点 t' , 及距离 t_{i+1} 时间最近且 GPS 车速为 0 的时间点 t'' , 判断, v_i, v_{i+1} 的速度, 有以下 4 种情况:

若 $v_i = v_{i+1} = 0$, 根据运动学片段的定义, 认为 $[t', t_i], [t_{i+1}, t'']$ 时间段车辆行驶数据可以构成一个完整的运动学片段, 保留 $[t', t_i]$ 与 $[t_{i+1}, t'']$ 时间段的车辆行驶数据;

若 $v_i = 0$ 且 $v_{i+1} \neq 0$, 则 $[t', t_i]$ 时间段车辆行驶数据可以构成一个完整的运动学片段, 保留 $[t', t_i]$ 时间段的车辆行驶数据, 而 $[t_{i+1}, t'']$ 时间段车辆行驶数据无法构成一个完整的运动学片段, 删除 $[t_{i+1}, t'']$ 时间段的车辆行驶数据;

若 $v_i \neq 0$ 且 $v_{i+1} = 0$, 则 $[t', t_i]$ 时间段车辆行驶数据无法构成一个完整的运动学片段, 删除 $[t', t_i]$ 时间段的车辆行驶数据, 而 $[t_{i+1}, t'']$ 时间段车辆行驶数据可以构成一个完整的运动学片段, 保留 $[t_{i+1}, t'']$ 时间段的车辆行驶数据;

若 $v_i \neq 0$ 且 $v_{i+1} \neq 0$, 则 $[t', t_i]$ 与 $[t_{i+1}, t'']$ 时间段车辆行驶数据无法构成一个完整的运动学片段, 删除 $[t', t'']$ 时间段的车辆行驶数据, 删除 $[t', t'']$ 所有数据后, 包含 t' 时刻的连续行驶片段与包含 t'' 时刻的连续行驶片段虽然时间上不连续, 但是保证了每部分片段之间时间的连续性, 每部分片段可以认为是一个运动学片段.

加、减速度异常: 汽车行驶过程中, 对于一辆普通轿车, 一般情况下: $0 \sim 100$ km/h 的加速时间大于 7 s, 紧急刹车最大减速度在 $7.5 \text{ m/s}^2 \sim 8.0 \text{ m/s}^2$. 由于采集的数据中并不能直接得到加速度, 因此首先测算每个时刻的加速度.

加速度的定义为: 单位时间内速度改变程度的矢量. 若动点的速度 v_i 经 $(t_{i+1} - t_i)$ s 后变成 v_{i+1} , 加速度计算公式如下:

$$a_i = \frac{(v_{i+1} - v_i)}{(t_{i+1} - t_i)}. \quad (1)$$

考虑到已知数据是以秒为单位测算车辆的速度, 本文加速度测算公式如下:

$$a_i = (v_{i+1} - v_i) \times \frac{5}{18}. \quad (2)$$

将测算的加速度与普通轿车一般情况下加、减速度做比较, 根据公式 (1), 最大加速度为 4 m/s^2 , 最大减速度为 7.5 m/s^2 , 超过则为加、减速度异常情况.

按照公式 (2) 测算所有时刻的加速度, 将各个时刻的加或减速度与最大加速度为 4 m/s^2 和最大减速度比较, 判断该时刻的加或减速度是否存在问题, 具体定义及处理方式如下:

令 t_i 时刻的加速度为 a_i , v_i 代表 t_i 时刻的速度, 若 $a_i > 4 \text{ m/s}^2$ 、 $a_i < -7.5 \text{ m/s}^2$, 则定义 t_i 时刻为加或减速度异常的数据时刻, 如果删除 t_i 时刻的数据, 必然会导致原本包含 t_i 时刻的车辆连续行驶片段不连续, 因此需要对这部分数据进行处理.

若 $v_i \neq 0$, 找到距离 t_i 时间最近且 GPS 车速为 0 的时间点 t', t'' , 且 $t' < t_i < t''$, 删除 $[t', t'']$ 时间段的车辆行驶数据, 即完成了汽车突然加、减速度产生的异常数据的处理, 删除 $[t', t'']$ 的所有数据后, 包含 t' 时刻的车辆连续行驶片段与包含 t'' 时刻的车辆连续行驶片段虽然时间上不连续, 但是保证每部分片段之间时间的连续性, 每部分片段可以被认为是一个运动学片段;

若 $v_i = 0$, 找到距离 t_i 时间最近且 GPS 车速为 0 的时间点 t' , 且 $t' > t_i$, 删除 $[t_i, t']$ 时间段的车辆行驶数

据,即完成了汽车突然加、减速度产生的异常数据的处理,删除 $[t_i, t']$ 的所有数据后,包含 t_i 时刻的车辆连续行驶片段与包含 t' 时刻的车辆连续行驶片段虽然时间上不连续,但是保证每部分片段之间时间的连续性,每部分片段可以被认为是一个运动学片段。

长期停车:定义 $v_i=0$ 的连续时刻的累计时间超过180 s为长期停车状态,将所有 $v_i=0$ 的连续时间区间通过计算机语言筛选出来,并判断累计时间是否超过180 s,若 $[t', t'']$ 为连续时间 $v_i=0$ 区间,则累计时间计算方式如下:

$$cumulativetime = (t'' - t') + 1. \quad (3)$$

长期停车产生的异常数据具体定义及处理方式如下:令通过MATLAB编程语言筛选出 $v_i=0$ 的连续时间区间为 $[t', t'']$,根据公式(3)计算 $v_i=0$ 的连续时刻的累计时间,若 $cumulativetime > 180$ s,删除 $[t', t'']$ 时间段的车辆行驶数据,即完成了长期停车产生的异常数据处理。

堵车、怠速:由于长时间堵车、断断续续低速行驶通常按怠速情况处理,即 $v_i=0$ 的情况,因此长时间堵车、断断续续低速行驶时间超过180 s,可理解为长期停车状态,长期停车为异常情况,需要对这部分数据进行处理。

具体定义及处理方式:通过MATLAB编程语言筛选出 $v_i \neq 0$ 的连续时间范围: $\{t_i, t_{i+1}, t_{i+2}, \dots, t'\}$,其速度范: $\{v_i, v_{i+1}, v_{i+2}, \dots, v'\}$,若 $\max\{v_i, v_{i+1}, v_{i+2}, \dots, v'\} < 10$ km/h,则定义 $[t_i, t']$ 这段时间为长时间堵车、断断续续低速行驶时间,根据公式(3)计算,若 $cumulativetime > 180$ s,删除 $[t_i, t']$ 时间段的车辆行驶数据,即完成了长期停车产生的异常数据处理。

1.2 运动学片段划分

运动学片段是指汽车从怠速状态开始至下一个怠速状态开始之间的车速区间^[3]。本文划分运动学片段的标准是:

1. 从汽车速度为零的状态开始,到下个时间为零的状态结束;
2. 全过程持续时间必须是连续的,且超过50 s;
3. 过程中的最大速度大于10 km/h。

将上述3个条件用MATLAB软件对筛除异常后的数据进行运动学片段的划分,我们共得到了1 648个运动学片段。从这1 648个运动学片段中随机选出4段数据绘制其速度-时间图,如图1所示,并将第一节运动学片段的的速度-时间图提取出来。如图2所示。

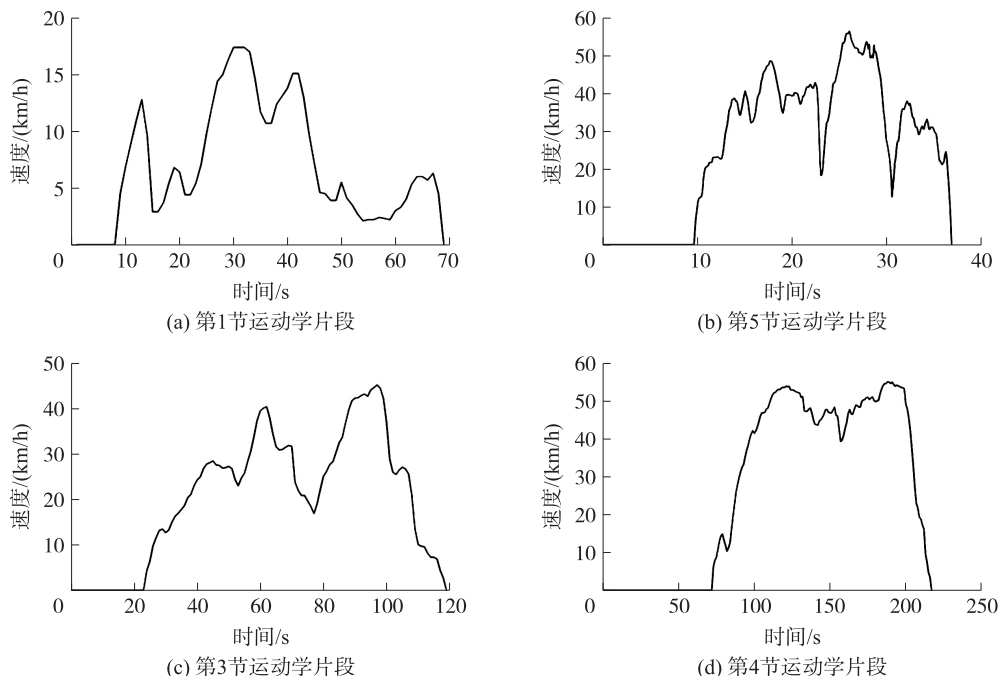


图1 4节运动学片段的的速度-时间图

Fig.1 Velocity-time diagram of four segments of kinematics

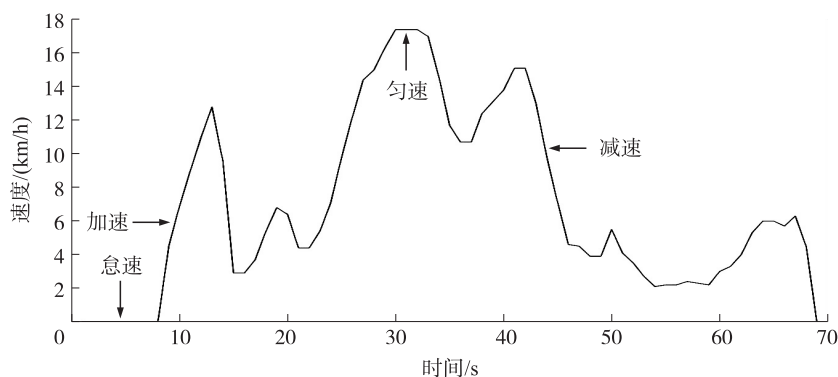


图 2 运动学片段的速度-时间图

Fig. 2 Velocity-time diagram of kinematic segments

从图 1 可以看出,这 4 节运动学片段的开头和结尾处的速度为 0,无间断区间,最大速度均大于 10 km/h,因此,这 4 节运动学片段均满足上述划分运动学片段的 3 个标准,由此可以验证编写的 MATLAB 代码的有效性. 由图 2,速度为 0 的运行状态即怠速状态,速度随着时间的增加而增加的运行状态为加速状态,速度随着时间的增加而减少的运行状态为减速状态,速度随着时间的增加不产生较大变化的状态为匀速状态.

2 算法原理及应用

前面已经介绍了数据处理方法和运动学片段的划分,下面我们将引入本文要用到的主成分分析方法和 K -均值聚类方法的算法原理.

2.1 算法原理

汽车行驶工况是描述汽车行驶速度与时间的曲线^[4]. 初步选择以下 11 个与速度相关的参数作为汽车运行特征参数:最高速度 v_{\max} ;平均速度 v_m ;平均行驶速度(不含怠速) v_{mr} ;平均加速度 a_a ;平均减速度 a_c ;速度标准差 v_{sd} ;加速度标准差 a_{sd} ;加速时间比例 P_a ;等速时间比例 P_b ;减速时间比例 P_c ;怠速时间比例 P_d . 针对选定的 11 个特征参数,我们选用主成分分析法得到对行驶工况影响最大的几个因素^[5-6],主要步骤如下:

第一步,对 11 个特征参数进行主成分分析^[7-8],得到较大影响的特征参数.

上述 11 个特征参数和所有区间片段的参数矩阵如下:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1s} \\ x_{21} & x_{22} & \cdots & x_{2s} \\ \cdots & \cdots & \cdots & \cdots \\ x_{r1} & x_{r2} & \cdots & x_{rs} \end{bmatrix}, \quad (4)$$

式中, x_{ij} ($i=1,2,\cdots,r, j=1,2,\cdots,s$) 是第 i 个区间片段的第 j 个参数; r 是所有区间片段数; s 是所有特征参数个数.

对参数矩阵 X 进行标准化,得到矩阵 Y 如下:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1s} \\ y_{21} & y_{22} & \cdots & y_{2s} \\ \cdots & \cdots & \cdots & \cdots \\ y_{r1} & y_{r2} & \cdots & y_{rs} \end{bmatrix}, \quad (5)$$

式中,

$$y_{ij} = (x_{ij} - \bar{x}_j) / s_j, \quad (6)$$

$$\bar{x}_j = \frac{1}{r} \sum_{i=1}^r x_{ij}, \quad (7)$$

$$s_j^2 = \frac{1}{r-1} \sum_{i=1}^r (x_{ij} - \bar{x}_j)^2, \quad (8)$$

由矩阵 \mathbf{Y} 计算协方差矩阵 \mathbf{Z} , 其定义如下:

$$\mathbf{Z} = \begin{bmatrix} s_1^2 & \text{cov}(1,2) & \cdots & \text{cov}(1,s) \\ \text{cov}(2,1) & s_2^2 & \cdots & \text{cov}(2,s) \\ \cdots & \cdots & \cdots & \cdots \\ \text{cov}(s,1) & \text{cov}(s,2) & \cdots & s_s^2 \end{bmatrix}, \quad (9)$$

式中,

$$s_k^2 = \text{cov}(k,k), \quad (10)$$

$$\text{cov}(x,y) = \text{cov}(y,x) = \frac{1}{r-1} \sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y}). \quad (11)$$

由矩阵 \mathbf{X} 得到相关矩阵 \mathbf{G} , 定义如下:

$$\mathbf{G} = \frac{1}{r-1} \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1s} \\ g_{21} & g_{22} & \cdots & g_{2s} \\ \cdots & \cdots & \cdots & \cdots \\ g_{r1} & g_{r2} & \cdots & g_{rs} \end{bmatrix}, \quad (12)$$

式中,

$$g_{xy} = \frac{\text{cov}(x,y)}{s_x s_y}. \quad (13)$$

用 λ_i 表示矩阵 \mathbf{G} 的特征值, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, 相应正交化特征向量如下:

$$[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \cdots, \mathbf{e}_s] = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1s} \\ e_{21} & e_{22} & \cdots & e_{2s} \\ \cdots & \cdots & \cdots & \cdots \\ e_{s1} & e_{s2} & \cdots & e_{ss} \end{bmatrix}. \quad (14)$$

用 R_i 表示第 i 个主成分的贡献率, 定义如下:

$$R_i = \lambda_i / \sum_{j=1}^s \lambda_j, \quad (15)$$

式中, 贡献率 R_i 越大, 表示该主成分 i 所表达的信息越多.

用 LR_θ 表示前 θ 个特征值的贡献率之和, 定义如下:

$$LR_\theta = \frac{\sum_{i=1}^{\theta} \lambda_i}{\sum_{j=1}^s \lambda_j}, \quad (16)$$

式中, 如果 $LR_i \geq 80\%$, 前 θ 个特征值就能满足应用需要, 即每个区间片段有 θ 个主成分. 由此可以得到 θ 个对车辆行驶工况有较大影响的特征参数.

第二步, 将主成分分析确定的 θ 个特征参数用于聚类分析中, 用于评价片段之间的相关程度^[9-10].

相关程度通过相关系数衡量, 相关系数的定义如下:

$$c_{ij} = \frac{\sum_{k=1}^r (c_{ki} - \bar{c}_i)(c_{kj} - \bar{c}_j)}{\left[\left(\sum_{k=1}^r (c_{ki} - \bar{c}_i)^2 \right) \left(\sum_{k=1}^r (c_{kj} - \bar{c}_j)^2 \right) \right]^{1/2}}, \quad (17)$$

式中, c_{ii} 表示片段 i 和片段 j 的相关系数, c_{ki} 表示第 i 个片段的第 k 个参数的特征值, \bar{c}_i 和 \bar{c}_j 分别表示片段 i 和片段 j 的所有特征值的均值. 片段之间的相关系数 c_{ii} 也可以转换成片段之间的距离 DIS_{ij} , 转化公式如下:

$$DIS_{ij}=\sqrt{1-c_{ii}^2}. \tag{18}$$

通过公式(17)–(18)可对所有代表性运动学片段进行聚类^[11–12],按照低速、中速、高速 3 类运行状态将所有代表性运动学片段归为 3 类.

第三步,对于 3 类运行状态提取的不同数量运动学片段条数进行数据拟合.

本文构建的车辆行驶工况时长在 1 200 s~1 300 s 之间,可通过各类总运行时间在总体数据时间中所占的时间比确定各类最终拟合工况所占的时间,定义如下:

$$T_k=\frac{T_{\text{duringtime}}}{T_{\text{all}}}\sum_{i=1}^{N_k}T_{ki}, \tag{19}$$

式中, T_k 为第 k 类在最终拟合的车辆行驶工况图所占时间; T_{all} 为所有带变形运动学片段的持续时间; $T_{\text{duringtime}}$ 为要构建的车辆行驶工况的持续时间; N_k 为第 k 类运动学片段的总条数; T_{ki} 为第 k 类中第 i 条运动学片段的运行时间. 由公式(19)可分别从 3 类运行状态下提取不同数量的运动学片段条数,最后拟合成车辆行驶工况图.

2.2 算法案例

从预处理后的数据中,按时间长度在 1 100 s~1 400 s 之间的连续行驶区间将所有连续行驶区间提取出来,一共提取了 11 个区间,并对这 11 个区间按照运动学片段的定义划出了 97 个代表性运动学片段.

根据选出的 11 个特征参数计算 97 个运动学片段的部分特征参数值(表 1).

表 1 97 个运动学片段的特征参数

Table 1 Characteristic parameters of 97 kinematic segments

片段序号	最高速度	平均速度	平均行驶速度	…	减速时间比	速度标准差	加速度标准差
1	12.4	4.972	6.119	…	0.344	3.852	1.615
2	10.8	2.129	29.246	…	0.200	3.638	1.030
3	36.7	19.838	29.246	…	0.287	14.892	1.629
…	…	…	…	…	…	…	…
95	30.7	6.118	21.958	…	0.095	10.609	0.807
96	51.6	32.009	33.600	…	0.343	11.959	1.562
97	45.3	31.358	34.016	…	0.465	13.222	1.222

根据表 1 的数据进行主成分分析,得到 11 个特征参数即 11 个主成分的特征值、贡献率及累计贡献率数据表,以及主成分载荷矩阵表,如表 2、3 所示.

表 2 各个主成分的特征值、贡献率及累计贡献率

Table 2 Eigenvalues, contribution rate and cumulative contribution rate of each principal component

成分	1	2	3	4	5	6	7	8	9	10	11
特征值	5.285	2.476	1.795	0.593	0.511	0.179	0.076	0.045	0.035	0.005	–3.26E–16
贡献率/%	48.043	22.507	16.314	5.388	4.649	1.63	0.689	0.412	0.320	0.048	–2.96E–15
累计率/%	48.043	70.550	86.864	92.253	96.902	98.532	99.221	99.632	99.952	100	100

表 3 主成分载荷矩阵表

Table 3 Principal component load matrix

特征参数	平均速度	最高速度	平均行驶速度	怠速时间比	加速时间比	速度标准差	等速时间比	加速度标准差	平均减速度	平均加速度	减速时间比
成分 1	0.951	0.893	0.856	–0.830	0.807	0.710	0.550	0.216	0.088	–0.479	0.648
成分 2	0.007	0.042	0.002	–0.093	0.324	0.154	–0.439	0.925	–0.904	0.682	–0.069
成分 3	0.132	0.391	0.469	0.539	–0.337	0.611	–0.282	–0.210	–0.178	–0.130	–0.674

从表 2 看出,前 3 个主成分累计贡献率超过 80%,因此选取前 3 个主成分进行分析,通过 SPSS 软件分析,可得到前 3 个主成分的载荷矩阵.

从表 3 可以看出:

- (1)第一类主成分主要反映平均速度、最高速度、平均行驶速度;
- (2)第二类主要反映加速度标准差;
- (3)第三类主要反映速度标准差.

因此根据主成分分析选出的特征参数:平均速度、最高速度、平均行驶速度、加速度标准差、平均加速这 5 个特征参数进行聚类分析,将所有代表性运动学片段按照低速、中速、高速进行聚类,最终聚类中心对比如图 3 所示。

从图 3 中可以看出,第 I 类片段的加速比例较高,反映了车辆在畅通无阻的道路上的行驶情况,可以定为高速片段;第 II 类片段的怠速比例较大,反映了车辆在拥挤的道路上行驶,可以定为低速片段;第 III 类片段的匀速比例较大,反映了车辆稳定形式的状态,可以定为中速片段。

通过 SPSS 软件分析,可得到 K 均值聚类分析的聚类成员列表,如表 4 所示。

表 4 K 均值聚类分析的聚类成员列表

Table 4 List of cluster members of K -means cluster analysis

片段序号	类别	距离	片段序号	类别	距离
1	2	12.732	46	3	11.353
2	2	14.515	47	3	5.498
3	3	7.407	48	3	10.427
...
44	2	15.234	96	3	14.576
45	3	4.362	97	3	12.079

从表 4 可以看出,97 个运动学片段被分成 3 类,第 I 类包括了 16 个运动学片段,第 II 类包括了 32 个运动学片段,第 III 类包含了 49 个运动学片段。公式 (19) 计算取整后,可获得 2 条第 I 类高速片段,4 条第 II 类低速片段,5 条第 III 类中速片段,最后拟合成持续时间为 1 207 s 的车辆行驶工况图,如图 4 所示。在 0~272 s 之间为车辆低速行驶状态,且低速状态的车速低于 24.6 km/h;在 272 s~930 s 之间为车辆中速行驶状态,且中速行驶状态的车速普遍处于 24 km/h~43.7 km/h;在 930 s~1 207 s 之间为车辆高速行驶状态,且高速行驶状态车速最高达到了 64.4 km/h。

3 结论

本文通过采集某城市同一辆轻型汽车在不同时间段共 21 s 所获得的数据,主要介绍了如何通过处理大量的汽车行驶数据来得到汽车运动学原理。首先对数据时间不连续、加减速异常、长期停车、长时间堵车和怠速 4 种情况的异常数据进行处理,然后运用 MATLAB 软件对筛选异常后的数据进行运动学片段的划分,并从中取出具有代表性的运动学片段,再利用主成分分析方法求出对车辆行驶工况有较大影响的特征参数,用得到的特征参数求出用于评价片段之间的相关程度的片段间的相关系数和片段之间的距离,再使用 K -均值聚类方法对所有代表性运动片段进行聚类分析,将所有运动学片段按照低速、中速、高速的运动状态分为 3 类,并得到了聚类中心对比图。最后根据 K -均值聚类分析的聚类成员对 3 类运行状态提取的不同数量运动学片段条数进行数据拟合,最终拟合成包含低速、中速和高速 3 种运动状态的汽车行驶速度与时间的曲线,即持续时间为 1 207 s 的汽车行驶工况图,由此得到汽车运动学的数学原理。本文通过对汽车运动学数学原理的探索,为相关部门制定出反应我国实际情况的汽车行驶工况提供理论依据,为车辆设计的动力匹配、提高发动机的燃油消耗率、有效降低尾气排放提供参考,对我国汽车工业的发展具有重要意义。

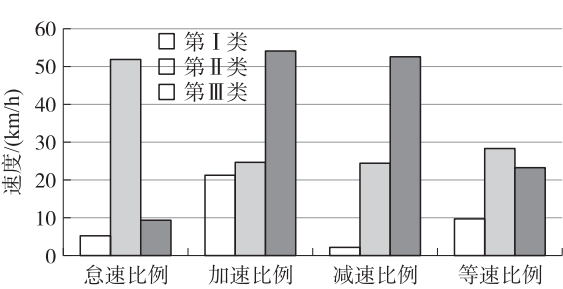


图 3 聚类中心对比图

Fig 3 Comparison of cluster centers

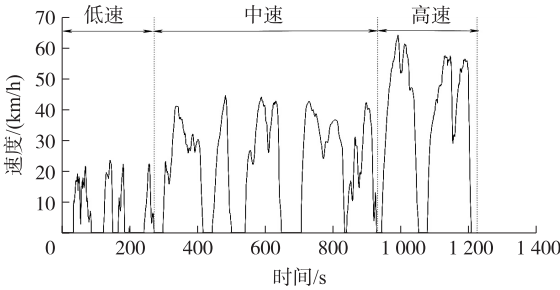


图 4 构建的汽车工况速度-时间图

Fig. 4 Constructed speed-time diagram of vehicle working condition

[参考文献]

- [1] HO S H, WONG Y D, CHANG W C. Developing singapore driving cycle for passenger cars to estimate fuel consumption and vehicular emissions[J]. Atmospheric environment, 2014, 97: 353–362.
- [2] LIN J, NIEMEIER D A. An exploratory analysis comparing a stochastic driving cycle to California's regulatory cycle[J]. Atmospheric environment, 2002, 36(38): 5759–5770.
- [3] 姜平, 石琴, 陈无畏. 基于小波分析的城市道路行驶工况构建的研究[J]. 汽车工程, 2011, 33(1): 70–73.
- [4] 石琴, 郑与波, 姜平. 基于运动学片段的道路行驶工况的研究[J]. 汽车工程, 2011, 33(3): 256–261.
- [5] ANDERSON T W. Asymptotic theory for principal component analysis[J]. Annals of mathematical statistics, 1963, 34(1): 122–148.
- [6] 余平. 基于 FPCA 的部分函数型线性模型的复合分位数回归估计[J]. 山西师范大学学报(自然科学版), 2019, 33(3): 5–12.
- [7] YEUNG K Y, RUZZO W L. Principal component analysis for clustering gene expression data[J]. Bioinformatics, 2019, 17(9): 763–74.
- [8] KORHONEN P, SILJAMKI A. Ordinal principal component analysis theory and an application[J]. Computational statistics & data analysis, 1998, 26(4): 411–424.
- [9] 白奕. 多指标综合评价的主成分分析模型及原理[J]. 陕西师范大学学报(自然科学版), 1998, 26(2): 105–106.
- [10] 刘靖明, 韩丽川, 侯立文. 基于粒子群的 K 均值聚类算法[J]. 系统工程理论与实践, 2005, 25(6): 54–58.
- [11] 温瑞英, 王红勇. 基于因子分析和 K -means 聚类的空中交通复杂性评价[J]. 太原理工大学学报, 2016, 47(3): 384–388, 404.
- [12] DING C, HE X. K -means clustering via principal component analysis[J]. Applied and computational mathematics, 2004, 1: 1–8.

[责任编辑:陆炳新]

(上接第 9 页)

- [10] LIDL R, MULLEN G L. When does a polynomial over a finite field permute the elements of the field? [J]. AM math mon, 1993, 100(1): 71–74.
- [11] BALL S, ZIEVE M. Symplectic spreads and permutation polynomials[M]. Berlin: Springer, 2004.
- [12] ZENG X, ZHU X, HU L. Two new permutation polynomials with the form $(x^{2^k} + x + \delta)^s + x$ over F_{2^m} [J]. Applicable algebra in engineering communication and computing, 2010, 21(2): 145–150.
- [13] HOU X. Permutation polynomials over finite fields—a survey of recent advances[J]. Finite fields and their applications, 2015, 32(1): 82–119.
- [14] ZHA Z, HU L. Two classes of permutation polynomials over finite fields[J]. Finite fields and their applications, 2012, 18(4): 781–790.
- [15] TU Z, ZENG X, JIANG Y. Two classes of permutation polynomials having the form $(x^{2^m} + x + \delta)^s + x$ [J]. Finite fields and their applications, 2015, 31(1): 12–24.

[责任编辑:陆炳新]