

## 基于改进 K 均值和马尔科夫链的公交车行驶工况构建

陈宝, 谢光毅, 黄春, 付江华

重庆理工大学车辆工程学院, 重庆 400054

**摘要:**目的 为填补目前缺少符合重庆城区地形特点的公交车行驶工况空白。方法 提出一种改进的 K 均值聚类算法与马尔科夫链结合的行驶工况构建方法。通过滤波算法对采集到的重庆城区 805 路公交车有效行驶数据进行平滑处理, 从平滑后的行驶数据中提取出 1 721 条运动学片段; 采用核主成分分析法对各运动学片段的特征参数矩阵进行降维处理, 以粒子群算法选择适合的 K 均值初始聚类中心, 并通过改进后的 K 均值聚类算法对运动学片段进行分类标记; 最后通过马尔科夫链筛选出合理的运动学片段, 构建出时长为 1 310 s 的重庆城区公交车行驶工况。结果 将构建的行驶工况与实车采集数据及国内外部分典型工况进行对比, 结果表明: 所构建的工况与原始数据在速度-加速度的联合分布趋势一致, 构建的行驶工况与实车采集数据各特征值误差均小于 7%, 所构建的工况与原始数据特征高度吻合且能较好地反映出重庆城区公交车怠速时间比例高、加减速频繁等符合区域特点的实际交通状况。结论 构建的工况能为重庆市公交车行驶路线规划、排放及油耗测试等方面提供基础标准, 同时其涉及的研究方法能为新能源汽车能量管理、控制策略及其他城市工况构建和城市公交线路优化提供参考。

**关键词:** 公交车; 行驶工况; 核主成分分析; K 均值聚类算法; 马尔科夫链

**中图分类号:** U491 **文献标识码:** A **doi:** 10. 16055/j. issn. 1672-058X. 2024. 0003. 003

### Construction of Bus Driving Cycle Based on Improved K-means and Markov Chains

CHEN Bao, XIE Guangyi, HUANG Chun, FU Jianghua

School of Vehicle Engineering, Chongqing University of Technology, Chongqing 400054, China

**Abstract:** **Objective** The research in this paper aims to fill the gap of the current lack of bus driving cycles in line with the characteristics of the urban terrain in Chongqing. **Methods** An improved K-mean clustering algorithm combined with a Markov chain was proposed to construct driving cycles. The effective driving data of bus No. 805 in urban areas of Chongqing were smoothed using a filtering algorithm, from which 1, 721 kinetic segments were extracted. The feature parameter matrices of each kinetic segment were dimensionally reduced using kernel principal component analysis, suitable initial cluster centers were selected with the particle swarm algorithm for K-means, and the kinetic segments were classified and labeled using the improved K-means clustering algorithm. Finally, reasonable kinetic segments were selected through the Markov chain to construct a bus driving cycle in urban areas of Chongqing with a duration of 1, 310 seconds. **Results** The constructed driving cycle was compared with the data collected from real vehicles and some typical cycles at home and abroad. The results showed that the constructed cycle exhibited consistent trends in the joint distribution of speed and acceleration with the original data, and the errors of the characteristic values of the constructed driving cycle compared with the data collected from real vehicles were all less than 7%. The constructed cycle highly matched the characteristics of the original data and can well reflect the high idle time ratio, frequent acceleration and deceleration, and other actual traffic conditions in urban areas of Chongqing. **Conclusion** The constructed cycle can

收稿日期: 2023-02-24 修回日期: 2023-04-20 文章编号: 1672-058X(2024)03-0018-08

基金项目: 重庆市教育委员会科学技术研究计划青年项目(KJQN202101143); 重庆理工大学研究生创新项目(GZLCX20222131)。

作者简介: 陈宝(1971—), 男, 副教授, 从事车辆系统控制及其结构设计分析研究。

通讯作者: 谢光毅(1995—), 男, 硕士研究生, 从事车辆行驶数据研究。Email: 617215332@qq.com。

引用格式: 陈宝, 谢光毅, 黄春, 等. 基于改进 K 均值和马尔科夫链的公交车行驶工况构建[J]. 重庆工商大学学报(自然科学版), 2024, 41(3): 18—25.

CHEN Bao, XIE Guangyi, HUANG Chun, et al. Construction of bus driving cycle based on improved K-means and Markov chains [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(3): 18—25.

provide a basic standard for bus route planning, emission testing, and fuel consumption testing in Chongqing, and the research methods involved can serve as a reference for energy management of new energy vehicles, control strategies, as well as city driving cycle construction and optimization of city bus routes.

**Keywords:** bus; driving cycle; kernel principal component analysis; K-means clustering algorithm; Markov chain

## 1 引言

公共交通是推动城市运营的重要基础设施,在城市基础设施建设中要优化公共交通结构,提高公交效率和交通安全保障,切实减轻交通拥堵。行驶工况是一段用来描述车辆行驶速度随时间变化的曲线,主要用于确定汽车污染物排放和燃油消耗等性能,广泛应用于交通协同控制、新车型的技术开发和评估等方面<sup>[1]</sup>。全球车辆排放测试分别使用美国(FTP-75)、欧洲(ECE15+EUDC)、日本(JAPAN10-15)等行驶工况<sup>[2]</sup>。2019年10月底,中国发布能对燃油及非插电式混合动力汽车提供能耗、排放测试基础标准的汽车行驶工况<sup>[4]</sup>。我国地区特点的差异使得各个区域汽车行驶工况不尽相同<sup>[3]</sup>。

国内外研究学者构建了含有区域特征的行驶工况并且取得了大量成果。Brady等<sup>[5]</sup>基于电动汽车实车行驶数据构建了爱尔兰地区电动汽车的行驶工况;秦大同等<sup>[6]</sup>通过采集某市城区汽车实车行驶数据,采用K均值聚类算法对工况块进行划分和拼接,最终合成1500s的城市循环工况;Peng等<sup>[7]</sup>采集福州市区内7辆汽车的行驶数据,提出一种基于K均值聚类和马尔科夫模型相结合的驾驶周期构建法,构建了一个包含1200s速度-时间序列的驾驶循环;高建平<sup>[8]</sup>通过车载终端设备采集郑州市乘用车行驶数据,采用传统主成分分析结合优化的K均值聚类算法构建了乘用车行驶工况,通过仿真试验测试了油耗和排放性能;苗强等<sup>[9]</sup>采集济南市K139路公交车7d的行驶数据,采用聚类分析结合马尔科夫链的方法构建了该市公交车的行驶工况,并通过仿真试验测试了整车百公里油耗;Liu等<sup>[10]</sup>收集合肥市乘用车行驶数据,运用主成分分析法和K均值聚类法构建了合肥市乘用车行驶工况;Shen等<sup>[11]</sup>通过车载数据记录仪,采集上海市浦东新区993线路混合动力客车200h的行驶数据,通过主成分分析和K均值聚类方法构建了典型的行驶工况,提出了基于公交车站的行驶周期分段方法和基于全程的公交车行驶周期构造方法;Wang等<sup>[12]</sup>采用GPS设备收集北京市乘用车112辆汽车行驶数据,使用TRAFIX方法生成了1000s北京市区高峰时间行驶工况、1200s北京市区非高峰时间行驶工况、1400s北京郊区高峰时间行驶工况、1400s北京郊区非高峰时间行驶工况;Wang等<sup>[13]</sup>使用北京市40辆电动车180d的行驶数据,根据平均速度进行行驶片段划分,基于马尔科夫链

构建了北京市电动车的行驶工况;李加强等<sup>[14]</sup>采集昆明市5条代表性的公交线路,基于主成分分析和K均值聚类算法,综合考虑各类别时间比例、欧式距离和相关系数选择合适的运动学片段,构建了昆明市1200s公交车行驶工况;胡志远等<sup>[15]</sup>采集上海市10辆出租车12个月的道路行驶数据,在主成分及聚类分析的基础上利用相关系数构建了1163s出租车行驶工况;金思含等<sup>[16]</sup>采集成都市示范区内燃料电池公交车实际行驶数据,通过主成分分析和K均值聚类算法构建了示范区的公交车行驶工况。

综合分析可以发现:在进行行驶工况的构建方法中,主成分分析法、K均值聚类法较为广泛应用。主成分分析是一种线性降维方法,并不能提取出工况中运动学片段的非线性特征;K均值聚类时,结果取决于聚类数目及初始聚类中心的选取,具有随机性。为充分反映重庆市公交车的实际运行特征并合理构建行驶工况,本文在对原始数据进行预处理的基础上,采用核主成分分析法对特征参数进行降维处理,使用改进的K均值聚类算法对降维后的数据进行分类,最后,通过马尔科夫模型构建行驶工况,通过对工况进行速度-加速度联合比较和特征参数误差分析验证工况的有效性。

## 2 行驶数据采集

对重庆市公交车的运营特点进行分析,采集路线覆盖充分。它由繁华和交通量较大的商业区、居民生活区、旅游景区区、郊区及其他地段组成。最后选取以途经路线涵盖新纪元购物广场、磁器口、歌乐山、首创城等地标性站点,线路总长9.6km的805路公交车运行路线为实验路线,图1为该路公交车行驶路线图。

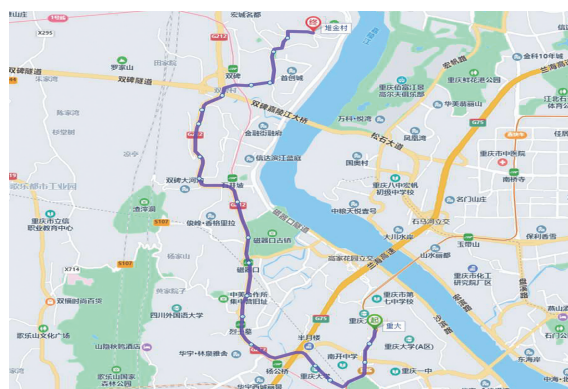


图1 重庆市805路公交车行驶路线图

**Fig.1 Route map of No. 805 bus in Chongqing**  
为保证测试数据的可靠性,车辆实验时间须包括

工作日与非工作日、上下班高峰时段以及夜晚低峰时段。由于公交车运行路线固定,采用正常运营法由驾驶员正常驾驶公交车行驶以完成实验数据采集,采用德国 VECTOR 通讯模块 VN1640A 设备采集公交车一周的行驶数据。该设备通过公交车 OBD 接口收集车辆 CAN 总线信息,通过整车通讯协议,可解析出速度、加速度、经度、纬度等信息,最终采集 108 042 条有效数据,采样频率为 1 Hz,如图 2 所示。

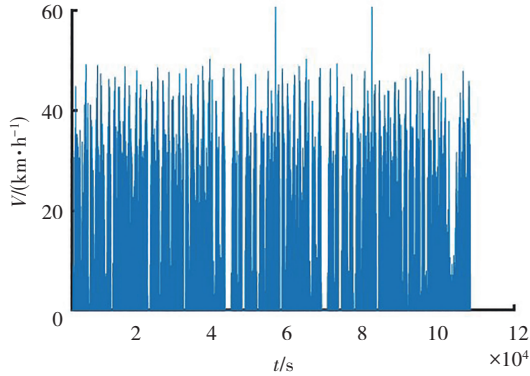


图 2 公交车行驶数据

Fig. 2 Bus travel data

### 3 行驶数据预处理

#### 3.1 速度数据滤波

在采集汽车行驶数据的过程中,因受到设备影响会导致数据采集值与真实值存在偏差(噪声),如果不对这些数据进行降噪处理,在一定程度上会影响计算结果。本文采用对称指数移动平均(Symmetric Exponential Moving Average, SEMA)算法对数据进行滤波处理,如式(1)、式(2):

$$X_i = \frac{\sum_{k=i-D}^{i+D} X_k \exp(-|i-k| \cdot d/T)}{\sum_{k=i-D}^{i+D} \exp(-|i-k| \cdot d/T)} \quad (1)$$

$$D = \min\left\{3 \frac{T}{d}, i-1, l-i\right\} \quad (2)$$

其中, $i$  为数据点对应离散时刻; $T$  为移动时间窗,取 0.5 s; $d$  为数据时间间隔,取 0.1 s; $l$  为数据总量; $X_i$  为  $i$  时刻的平滑值。

如图 3 所示,采集的速度曲线存在大量噪声、突变数据,这些数据会影响加速度值的计算结果。经过滤波处理后,前后速度曲线走势基本一致,而且曲线变得更为平顺,达到了滤波的目的。

对滤波后的数据预处理,剔除怠速时间超过 180 s 的数据,线性插值由于 GPS 信号缺失而导致的间断数据和加速度( $a \geq 4 \text{ m/s}^2$  或  $a \leq -8 \text{ m/s}^2$ ) 异常数据,最终得到 89 119 组有效数据。对数据处理前后的行平均速度、平均行驶速度两个参数进行对比,数据波动在合理范围之内。

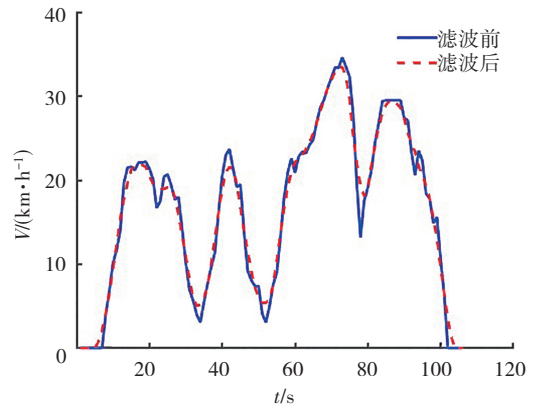


图 3 滤波前后速度数据对比

Fig. 3 Comparison of velocity data before and after filtering

#### 3.2 运动学片段划分

公交车在城市道路运行过程中,由于站点停车或道路交通状况的影响,驾驶员对车辆的加速、制动方面的操作较为频繁,为模拟车辆的实际行驶状态,将车辆介于相邻两个怠速状态车速区间定义为一个运动学片段,如图 4 所示。一个完整运动学片段由怠速、加速、匀速、减速 4 个行驶工况构成<sup>[14]</sup>。车辆的行驶过程可认为是由多个运动学片段构成。参考国标 GB/T 38146.2-2019 对重型商用车辆的片段工况划分规则,以速度  $v$ 、加速度  $a$  值为准则对车辆行驶状态进行划分:

$$\begin{cases} a \geq 0.15 \text{ m/s}^2 & \& v \neq 0 & \text{加速段} \\ a \leq -0.15 \text{ m/s}^2 & \& v \neq 0 & \text{减速段} \\ v < 0.5 \text{ km/h} & \& |a| < 0.15 \text{ m/s}^2 & \text{怠速段} \\ v \geq 0.5 \text{ km/h} & \& |a| < 0.15 \text{ m/s}^2 & \text{匀速段} \end{cases}$$

按照上述原则进行数据分割处理,得到 1 708 条运动学片段。

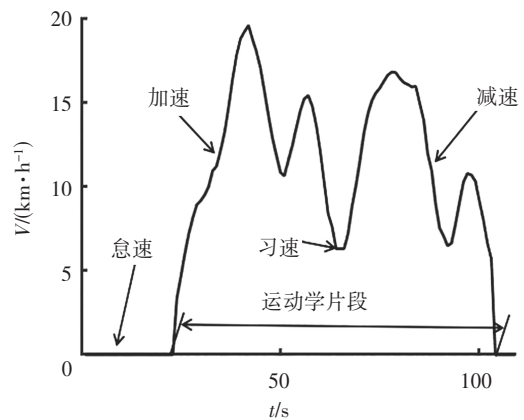


图 4 运动学片段示意图

Fig. 4 Schematic diagram of kinematic segments

#### 3.3 特征参数选取与计算

通常运动学片段的特点描述以及定量分析可用特征参数来表征。在运动学片段特征参数选择中,合适的特征参数能兼顾运算时间和准确率<sup>[6]</sup>。如表 1 所

示,选取 14 个具有代表性的特征参数并求解出所有运动学片段中的特征参数,得到 M1708×14 的片段特征参数数据库,如表 2 所示。

表 1 运动学片段的特征参数

Table 1 Characteristic parameters of kinematic segments

| 序 号 | 特征参数/单位                                       | 含 义    |
|-----|-----------------------------------------------|--------|
| 1   | $S/\text{m}$                                  | 运行距离   |
| 2   | $T/\text{s}$                                  | 运行时间   |
| 3   | $v_a/(\text{km}\cdot\text{h}^{-1})$           | 平均速度   |
| 4   | $v_r/(\text{km}\cdot\text{h}^{-1})$           | 平均行驶速度 |
| 5   | $v_{\max}/(\text{km}\cdot\text{s}^{-1})$      | 最大速度   |
| 6   | $a_a/(\text{m}\cdot\text{s}^{-2})$            | 平均加速度  |
| 7   | $a_d/(\text{m}\cdot\text{s}^{-2})$            | 平均减速度  |
| 8   | $a_{a\max}/(\text{m}\cdot\text{s}^{-2})$      | 最大加速度  |
| 9   | $a_{d\max}/(\text{m}\cdot\text{s}^{-2})$      | 最大减速度  |
| 10  | $a_{\text{std}}/(\text{m}\cdot\text{s}^{-2})$ | 加速度标准差 |
| 11  | $P_a/\%$                                      | 加速比例   |
| 12  | $P_d/\%$                                      | 减速比例   |
| 13  | $P_i/\%$                                      | 怠速比例   |
| 14  | $P_c/\%$                                      | 匀速比例   |

表 2 运动学片段特征参数矩阵

Table 2 Matrix of kinematic segment characteristic parameters

| 片段序号  | $S$   | $T$ | $v_a$ | $v_r$ | $v_{\max}$ | ... | $P_c$ |
|-------|-------|-----|-------|-------|------------|-----|-------|
| 1     | 175.0 | 66  | 9.4   | 17.0  | 27.6       | ... | 5.9   |
| 2     | 162.6 | 42  | 13.6  | 16.3  | 21.2       | ... | 46.5  |
| 3     | 260.8 | 41  | 16.6  | 19.3  | 25.1       | ... | 42.0  |
| ⋮     | ⋮     | ⋮   | ⋮     | ⋮     | ⋮          | ⋮   | ⋮     |
| 1 708 | 32.7  | 15  | 7.4   | 9.8   | 11.9       | ... | 18.7  |

3.4 核主成分分析

在行驶工况构建过程中,直接进行聚类分析会出现计算量大、聚类效果差的问题,需要对这 14 个特征参数降维以降低计算难度。主成分分析是一种通过线性变换将原始指标量重组为一组彼此线性不相关的综合指标,以取代原指标的统计分析方法。考虑汽车行驶运动学片段的复杂性,主成分分析不能完全提取出模型中的非线性特征<sup>[17]</sup>。核主成分分析是指基于核函数的主成分分析,其实质就是把原始样本以非线性方式映射至高维特征空间上,并在此高维特征空间内再次利用主成分分析方法进行分析<sup>[18]</sup>。

本次核主成分分析采用高斯核函数,如式(3):

$$K(\boldsymbol{x}_n,\boldsymbol{y}_n)=\exp\left(-\frac{\|\boldsymbol{x}_n-\boldsymbol{y}_n\|^2}{\delta}\right)$$
 (3)

其中, $\boldsymbol{K}$  为核矩阵, $\delta$  为核参数, $\boldsymbol{x}_n,\boldsymbol{y}_n$  为特征参数矩阵中的行向量。

核主成分分析的流程:

① 主成分数据无量纲处理。计算样本均值式(4)和标准差式(5);通过式(6)进行无量纲处理。

$$\bar{x}_j=\frac{1}{m}\sum_{n=1}^mx_{nj}$$
 (4)

$$S_j=\sqrt{\frac{\sum_{n=1}^m(x_{nj}-\bar{x}_j)^2}{m-1}}$$
 (5)

$$X_{nj}=\frac{x_{nj}-\bar{x}_j}{S_n}$$
 (6)

② 主成分核矩阵的求解与中心化。通过式(3)求取核矩阵,再使用式(7)进行中心化处理。

$$\boldsymbol{K}'=\boldsymbol{K}-\boldsymbol{IK}-\boldsymbol{KI}+\boldsymbol{IKI}$$
 (7)

其中, $\boldsymbol{I}$  是与核矩阵维度相同的  $m\times m$  维矩阵。

③ 计算中心化核矩阵  $\boldsymbol{K}$  的特征值  $\lambda$ 、特征向量  $\boldsymbol{\varphi}$ ,并将特征值按照数值大小排列。

④ 通过式(8)、式(9)求解各主成分贡献率  $\omega$ 、累计贡献率  $\beta$ ;通过累计贡献率高于 85% 的指标确认主成分个数。

$$\omega_n=\lambda_n/\sum_{n=1}^l\lambda_n$$
 (8)

$$\beta_k=\sum_{n=1}^k\lambda_n/\sum_{n=1}^l\lambda_n$$
 (9)

⑤ 计算数据在特征空间上映射的特征向量  $\boldsymbol{F}$ ,表达式如下:

$$\boldsymbol{F}=\sum_{n=1}^l\boldsymbol{\varphi}_nK(\boldsymbol{x}_n,\boldsymbol{y}_n)$$
 (10)

由表 3 可知:前 5 个主成分符合累计贡献率高于 87.4% 且特征值均大于 1 的特点,足以表征整个运动学片段。

表 3 各成分特征值、贡献率及累计贡献率

Table 3 Characteristic value, contribution rate and cumulative contribution rate of each component

| 主成分 | 特征值  | 贡献率/% | 累计贡献率/% |
|-----|------|-------|---------|
| 1   | 18.6 | 39.7  | 39.7    |
| 2   | 7.3  | 15.6  | 55.2    |
| 3   | 6.5  | 13.7  | 69.0    |
| 4   | 5.9  | 12.5  | 81.5    |
| 5   | 2.8  | 6.0   | 87.4    |
| 6   | 1.9  | 4.0   | 91.4    |

3.5 改进的 K 均值聚类算法

K 均值聚类算法基本原理:选定  $k$  个聚类中心,求解所有数据点与聚类中心之间的距离最小值;根据该值将数据分配到最近的一类,迭代至各数据聚类中心不再变化<sup>[6]</sup>。

由于 K 均值聚类结果受初始聚类中心和聚类数目的影响,本文使用 Silhouette 轮廓函数来确定聚类个数,利用不同聚类个数值所对应的轮廓值<sup>[19]</sup>,结合粒子群

算法优化初始聚类中心<sup>[20]</sup>。但由于粒子群算法的寻优效果很大程度上依赖于参数(最大速度、系数等)的准确率,因此通过惯性权重调整使全局搜索和局部搜索趋于平衡,优化的聚类算法才能进一步用于 K 均值聚类中心点的初始化。

改进的 K 均值聚类算法基本过程:

① 根据式(11)的 Silhouette 轮廓系数值选择最佳聚类数目  $k$ 。

$$s_j = \begin{cases} 1 - \frac{a(j)}{b(j)} & a(j) < b(j) \\ 0 & a(j) = b(j) \\ \frac{b(j)}{a(j)} - 1 & a(j) > b(j) \end{cases} \quad (11)$$

其中,  $a(j)$  代表类别属性为  $j$  的样本簇内不相似度,  $b(j)$  表示类别属性为  $j$  的样本簇间不相似度,  $s(j)$  为轮廓系数。

② 初始化种群。首先将所有样本随机分配到一个类别,以此进行聚类划分;然后根据初始位置  $x_i$  和速度  $v_i$ ,确定学习因子  $b_i$ ,以此构建有效的聚类结构。

③ 根据式(12)求解适应度值。当前适应度粒子最小值为由式(12)求解出的各小类中数据与聚类中心欧式距离  $f_i$ ,全局极值为所有适应度粒子最小值的极小值。

$$d_j = \sum_{j=1}^k \sum_{i=1}^{m_j} \sqrt{((c_{ji} - c_j) \cdot (c_{ji} - c_j)^T)} \quad (12)$$

其中,  $j$  代表数据  $c_{ji}$  的类别属性,  $m_j$  表示第  $j$  类样本数据总量值,  $c_j$  代表类别属性为  $j$  的 K 均值聚类中心。

④ 采用式(13)、式(14)的惯性权重,可以有效地将粒子转变为局部准确搜索,式(15)一式(18)可以准确地定位和调整粒子的位置与速度。

$$\Delta\sigma = (\sigma_{\max} - \sigma_{\min}) \quad (13)$$

$$\sigma = \sigma_{\max} + \Delta\sigma \left( 1 - \left( \frac{d_j - d_{\text{bad}}}{d_{\text{best}} - d_{\text{bad}}} \right)^3 \right)^5 \quad (14)$$

$$\Delta P_i = b_1 \times r_1 \times (P_{bi} - x_j) \quad (15)$$

$$\Delta G_i = b_2 \times r_2 \times (G_{bd} - x_j) \quad (16)$$

$$v_i(t) = wv_i(t-1) + \Delta P_i + \Delta G_i \quad (17)$$

$$x_i(t) = x_i(t-1) + v_i(t) \quad (18)$$

其中,  $\sigma_{\max}$  表示权重最大值,  $\sigma_{\min}$  表示权重最小值;  $d_{\text{best}}$  表示粒子最佳适应度值,  $d_{\text{bad}}$  表示粒子最差适应度值; 随机数  $r_1, r_2$  均属于区间(0~1);  $P_{bi}$  表示粒子处于最佳位置时的数值;  $G_{bd}$  表示全局极值。

⑤ 由式(19)求解粒子群适应度方差  $\delta^2$ , 此处阈值设置为 0.1,  $d_a$  代表了粒子群适应度值的平均值; 最后从中选取 10 个最优粒子来初始化 K 均值聚类中心。

$$\delta^2 = - \sum_{j=1}^n (d_j - d_a)^2 \quad (19)$$

⑥ 求解各样本与此时聚类中心的欧氏距离数值,

并参考此值对数据进行聚类划分,更新下一聚类中心适应度值。

⑦ 达到最优适应度值或是预设的迭代次数,终止算法,否则继续上述步骤。

根据 Silhouette 轮廓指数的定义,当前类数下的聚类效果指数接近 1 时为最佳。通过图 5 可得出最佳聚类数目为 2,对应的 Silhouette 指标值为 0.781 8。如图 6 所示:各个类别的轮廓系数均出现了少量负值,当聚类类别数为 2 时,各类的轮廓值分布比较均匀。根据改进的 K 均值聚类算法,不同聚类数目的适应度曲线如图 7 所示:迭代次数为 100 的条件下,聚类数目为 3 和 4 时,算法在经历初期的快速下降后陷入了局部最优解,聚类数目为 2 时,搜索曲线是不断下降的,跳出了局部最优的困扰。因改进后的 K 均值聚类算法在初始聚类中心的选取仍具有随机性,本文通过重复将改进的行算法运行 5 次,以其平均值作为聚类结果来提高效果,如表 4 所示,可以发现:聚类数目为 2 时,迭代次数最多且运行时间最短。

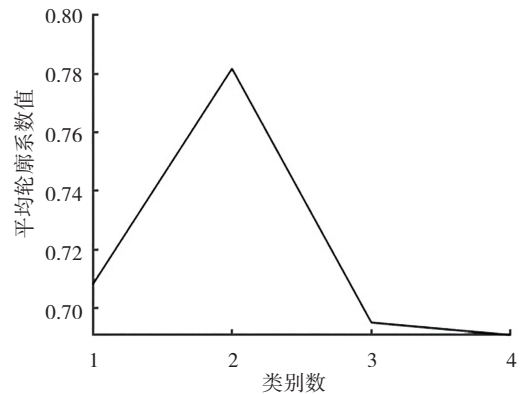


图 5 轮廓系数图

Fig. 5 Profile coefficient diagram

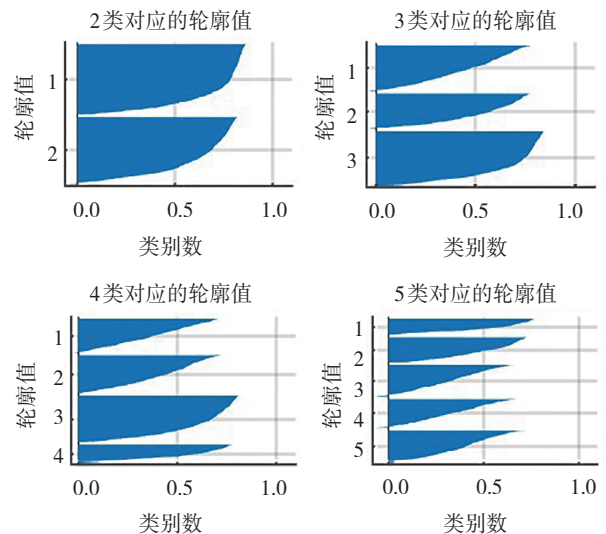


图 6 轮廓系数曲线图

Fig. 6 Diagram of profile coefficient curves

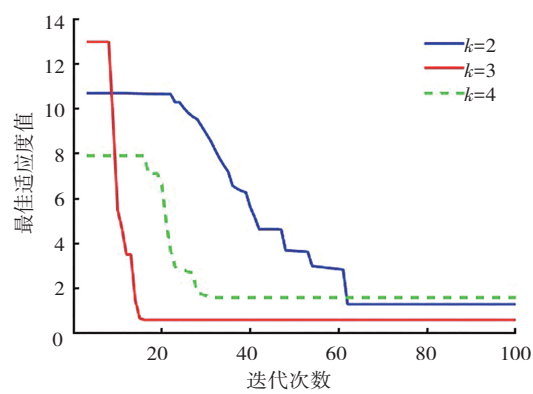


图 7 不同聚类数目的适应度曲线

Fig. 7 Fitness curves for different numbers of clusters

表 4 不同聚类数目下性能对比

Table 4 Performance comparison with different numbers of clusters

| 聚类数目 | 迭代次数 | 时间/s |
|------|------|------|
| 2    | 60   | 42.6 |
| 3    | 13   | 55.5 |
| 4    | 30   | 71.5 |

从表 5 的各类特征参数可看出:类 1 中的高怠速比例以及平均速度较小,符合城区道路拥堵段特征;类 2 中加速、减速时间占比以及平均速度较大的特点正符合城郊道路畅通的交通特征。

表 5 两类部分特征参数

Table 5 Two types of partial characteristic parameters

| 特征参数                        | 类 1   | 类 2    |
|-----------------------------|-------|--------|
| $S/m$                       | 49.04 | 54.49  |
| $T/s$                       | 69.93 | 288.73 |
| $v_a/(km \cdot h^{-1})$     | 5.01  | 18.75  |
| $v_r/(km \cdot h^{-1})$     | 11.71 | 21.86  |
| $v_{max}/(km \cdot s^{-1})$ | 32.68 | 46.88  |
| $a_a/(m \cdot s^{-2})$      | 0.51  | 0.53   |
| $a_d/(m \cdot s^{-2})$      | -0.52 | -0.62  |
| $a_{amax}/(m \cdot s^{-2})$ | 1.69  | 1.75   |
| $a_{dmax}/(m \cdot s^{-2})$ | -1.55 | -1.83  |
| $v_{std}/(km \cdot h^{-1})$ | 5.28  | 10.62  |
| $P_a/\%$                    | 23.01 | 38.58  |
| $P_d/\%$                    | 23.17 | 32.97  |
| $P_i/\%$                    | 35.68 | 12.37  |
| $P_c/\%$                    | 18.15 | 16.09  |

4 行驶工况构建

4.1 马尔科夫状态转移概率矩阵获取

已知  $t_0$  时刻的随机过程  $X(t)$  且未来时刻所处的状态只与  $t_0$  时刻的状态有关,该随机过程  $X(t)$  称为马尔科夫过程<sup>[21]</sup>。

根据 K 均值聚类后的结果,进行状态划分。由于每一个运动学片段都由怠速段、加速段、匀速段、减速

段 4 个状态组成,通过马尔可夫模型对数据进行划分,计算不同状态的事件数量。

在马尔科夫模型中,可通过式 (20) 来计算状态转移概率矩阵。 $N_{mn}$  表示当前时刻从状态  $m$  向状态  $n$  转移的次数, $P_{mn}$  为状态转移过程中的概率。

$$P_{mn} = \frac{N_{mn}}{\sum_n N_{mn}}$$

(20)

根据上述原理,计算两类中的状态转移概率矩阵,如表 6 所示。

表 6 各状态间转移概率矩阵

Table 6 Transition probability matrix between states

| 类 别 | 当前状态 | 下一状态  |       |       |         |
|-----|------|-------|-------|-------|---------|
|     |      | 加速    | 减速    | 匀速    | 怠速      |
| 1   | 加速   | 0.907 | 0.001 | 0     | 0.009 3 |
|     | 减速   | 0.013 | 0.896 | 0.047 | 0.044   |
|     | 匀速   | 0.164 | 0     | 0.836 | 0       |
|     | 怠速   | 0.074 | 0.166 | 0     | 0.760   |
| 2   | 加速   | 0.844 | 0.001 | 0     | 0.155   |
|     | 减速   | 0.010 | 0.849 | 0.114 | 0.027   |
|     | 匀速   | 0.034 | 0     | 0.966 | 0       |
|     | 怠速   | 0.052 | 0.208 | 0     | 0.740   |

4.2 行驶工况的合成

本次行驶工况的总时长选取 1 310 s。在马尔可夫基础上,通过计算各类中时间对工况的时长贡献,根据状态转移矩阵,以怠速状态为初始状态,由当前状态转移到下一状态的转移概率大于零,筛选出符合要求的运动学片段,直到达到理想的长度位置,最后拼接成最终工况,如图 8 所示。

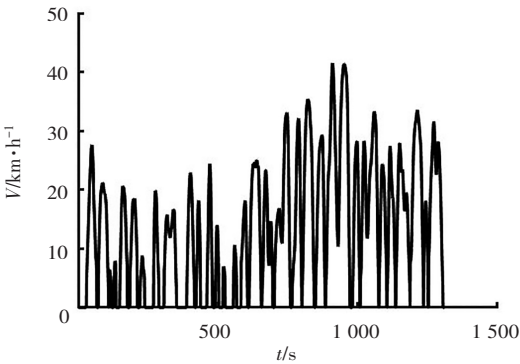


图 8 公交车行驶工况图

Fig. 8 Bus driving conditions

5 工况验证与分析

5.1 运动学片段特征参数误差分析

在构建的行驶工况基础上,计算工况曲线运动学片段的特征参数,并与经过预处理的行驶数据特征值进行对比分析,相对误差分析如式 (21) 所示:

$$\Delta = \frac{\sum_{j=1}^n |M_j - C_j|}{M_j}$$

(21)

其中,  $M_j$  表示原数据中序号为  $j$  的特征值,  $C_j$  表示构建工况中的序号为  $j$  的特征值,  $n$  表示特征参数总数。

表 7 可知: 行驶工况曲线与原始数据特征参数相对误差较小, 各个特征参数的相对误差均在 7% 以下。说明构造的行驶工况可以体现总体特征, 可作为代表工况用于后续分析。

表 7 部分特征参数相对误差对比  
Table 7 Comparison of relative errors of some characteristic parameters

| 特征参数/单位                               | 实验数据 | 工况数据 | 相对误差 / % |
|---------------------------------------|------|------|----------|
| $v_e/(\text{km} \cdot \text{h}^{-1})$ | 16.9 | 16.6 | 1.8      |
| $v_a/(\text{km} \cdot \text{h}^{-1})$ | 12.1 | 11.7 | 3.3      |
| $P_a/\%$                              | 27.1 | 28.1 | 3.6      |
| $P_d/\%$                              | 24.3 | 24.7 | 1.6      |
| $P_i/\%$                              | 32.2 | 30.1 | 6.5      |
| $P_c/\%$                              | 17.4 | 17.1 | 1.7      |

5.2 速度-加速度联合分布图

通过绘制构建的汽车行驶工况与预处理后的行驶数据速度-加速度联合分布图, 以验证构建行驶工况与原始数据之间的差异(图 9、图 10)。

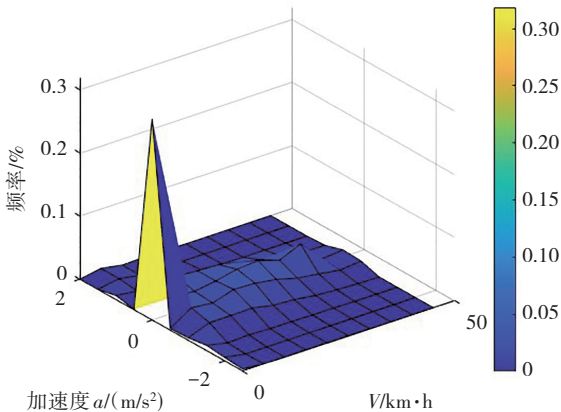


图 9 原始数据速度-加速度联合分布图  
Fig. 9 Joint velocity-acceleration distribution of original data

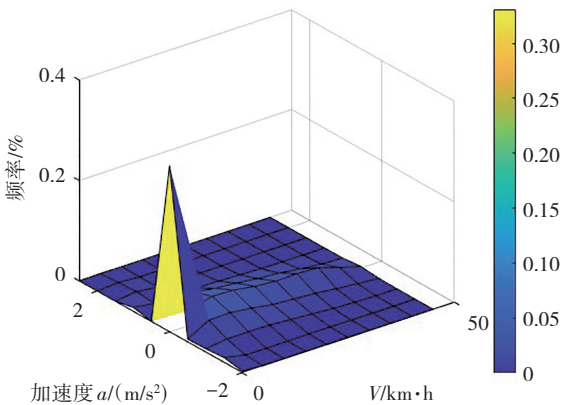


图 10 行驶工况速度-加速度联合分布图  
Fig. 10 Joint speed-acceleration distribution under driving conditions

由联合分布图可知: 所构建的工况与原始数据分布及整体趋势保持一致且主要分布在速度和加速度较低的区域, 与重庆公交车的行驶特点相符。

5.3 典型工况对比

由表 8 可看出: 所构建的重庆市公交车行驶工况与国内典型工况相比较, 平均速度比其他城市低, 怠速比例较高, 加速、减速、匀速比例存在一定差异; 该城市工况中怠速时间占比相对较大, 表明重庆市公交车怠速频繁, 符合重庆市公交车实际情况; 与欧洲 ECE15、日本 Japan10-15 工况对比, 平均速度、匀速比例存在显著差异。因此, 采用 ECE15 工况进行油耗测试实验并不能反映出真实的交通状况, 有必要建立该地区道路行驶工况。

表 8 循环工况特征参数对比  
Table 8 Comparison of characteristic parameters under cycle conditions

| 特征参数                                  | 上海   | 北京   | 重庆   | ECE15 | Japan10-15 |
|---------------------------------------|------|------|------|-------|------------|
| $v_a/(\text{km} \cdot \text{h}^{-1})$ | 14.2 | 15.6 | 11.7 | 18.7  | 17.7       |
| $P_a/\%$                              | 26.7 | 29.1 | 28.1 | 21.5  | 25.9       |
| $P_d/\%$                              | 28.6 | 25.6 | 24.7 | 18.5  | 26.4       |
| $P_i/\%$                              | 22.0 | 30.4 | 30.1 | 30.7  | 22.2       |
| $P_c/\%$                              | 22.7 | 14.9 | 17.1 | 29.3  | 25.4       |

6 结 论

采集 108 042 条有效数据, 提取 1 708 条运动学片段, 根据核主成分分析对数据进行特征参数矩阵降维, 将 Silhouette 轮廓函数所得的聚类数目运用到 K 均值聚类算法中, 采用粒子群算法优化聚类中心, 基于马尔可夫链筛选出符合要求的运动学片段, 最终构建出时长 1 310 s 的行驶工况。

对所构建的行驶工况与原始数据特征参数对比, 总体误差小于 7%; 对比速度-加速度联合分布图, 所构建的工况与原始数据特征高度吻合; 通过比较国内外工况, 构建的该城市公交工况符合当地实际交通特征, 提升了工况的真实性。

本文所构建的行驶工况可在新能源汽车控制策略的搭建与优化方面做进一步研究, 同时文中所涉及的研究方法也可其他城市行驶工况的构建及城市公交网络优化提供参考。

参考文献(References):

[1] FOTOUHI A, MONTAZERIGH M. Tehran driving cycle development using the k-means clustering method[J]. Scientia Iranica, 2013, 20(2): 286—293.  
[2] 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, 158.  
SHI Qin, QIU Duo-yang, ZHOU Jie-yu. Driving cycle

- construction and precision analysis based on combination clustering[J]. *Automotive Engineering*, 2012, 34(2): 164-169, 158.
- [3] 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.
- CAO Qian, LI Jun, LIU Yu, et al. Construction of driving cycle of passenger cars in Changchun based on Markov chain[J]. *Journal of Jilin University (Engineering Edition)*, 2018, 48(5): 1366—1373.
- [4] 杜莎. 中国汽车行驶工况正式发布[J]. *汽车与配件*, 2019(21): 6—7.
- DU Sha. China's automobile driving cycle officially released[J]. *Automobile and Accessories*, 2019(21): 6—7.
- [5] BRADY J, O' MAHONY M. Development of a driving cycle to evaluate the energy economy of electric vehicles in urban areas[J]. *Applied Energy*, 2016, 177(1): 165—178.
- [6] 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 Da-tong, JASON, QI Zheng-gang, et al. Construction method of driving cycle based on K-means clustering algorithm[J]. *Journal of Jilin University (Engineering Edition)*, 2016, 46(2): 383—389.
- [7] PENG Y, ZHUANG Y, YANG Y. A driving cycle construction methodology combining K-means clustering and Markov model for urban mixed roads[J]. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 2020, 234(2-3): 714—724.
- [8] 高建平,孙中博,丁伟,等. 车辆行驶工况的开发和精度研究[J]. *浙江大学学报(工学版)*, 2017, 51(10): 2046—2054.
- GAO Jian-ping, SUN Zhong-bo, DING Wei, et al. Development and precision research of vehicle driving conditions[J]. *Journal of Zhejiang University (Engineering Edition)*, 2017, 51(10): 2046—2054.
- [9] 苗强,孙强,白书战,等. 基于聚类和马尔可夫链的公交车典型行驶工况构建[J]. *中国公路学报*, 2016, 29(11): 161—169.
- MIAO Qiang, SUN Qiang, BAI Shu-zhan, et al. Construction of typical bus driving cycle based on clustering and Markov chain[J]. *Journal of China Highway*, 2016, 29(11): 161—169.
- [10] LIU B, SHI Q, HE L, et al. A study on the construction of Hefei urban driving cycle for passenger vehicle[J]. *IFAC-PapersOnLine*, 2018, 51(31): 854—858.
- [11] SHEN P, ZHAO Z, LI J, et al. Development of a typical driving cycle for an intra-city hybrid electric bus with a fixed route[J]. *Transportation Research Part D: Transport and Environment*, 2018, 59: 346—360.
- [12] WANG H, ZHANG X, OUYANG M. Energy consumption of electric vehicles based on real-world driving patterns: A case study of Beijing[J]. *Applied Energy*, 2015, 157: 710—719.
- [13] WANG Z, ZHANG J, LIU P, et al. Driving cycle construction for electric vehicles based on Markov chain and monte carlo method: A case study in Beijing[J]. *Energy Procedia*, 2019, 158: 2494—2499.
- [14] 李加强,王洪荣,周建文,等. 基于聚类分析法的公交车行驶工况构建研究[J]. *汽车工程学报*, 2017, 7(6): 400—406.
- LI Jia-qiang, WANG Hong-rong, ZHOU Jian-wen, et al. Research on bus driving cycle construction based on cluster analysis[J]. *Journal of Automotive Engineering*, 2017, 7(6): 400—406.
- [15] 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 Zhi-yuan, QIN Yan, TAN Pi-qiang, et al. Large-sample-based car-driving cycle in Shanghai city[J]. *Journal of Tongji University(Natural Science Edition)*, 2015, 43(10): 1523—1527.
- [16] 金思含,彭忆强,武小花,等. 成都市燃料电池公交车示范运行行驶工况的构建[J]. *汽车安全与节能学报*, 2022, 13(1): 202—208.
- JIN Si-han, PENG Yi-qiang, WU Xiao-hua, et al. Construction of the driving cycle for fuel cellobus running Chengdu demonstration area[J]. *Journal of Automotive Safety and Energy Conservation*, 2022, 13(1): 202—208.
- [17] 胡嘉蕊,吕震宙. 基于核主成分分析的多输出模型确认方法[J]. *北京航空航天大学学报*, 2017, 43(7): 1470—1480.
- HU Jia-rui, LYU Zhen-zhou. Multi-output model validation method based on nuclear principal component analysis[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2017, 43(7): 1470—1480.
- [18] 胡伟鹏,胡海峰,顾建权,等. 核主成分分析网络的人脸识别方法[J]. *中山大学学报(自然科学版)*, 2016, 55(5): 48-51, 56.
- HU Wei-peng, HU Hai-feng, GU Jian-quan, et al. Face recognition method based on kernel principal component analysis network[J]. *Journal of Sun Yat-sen University (Natural Science Edition)*, 2016, 55(5): 48-51, 56.
- [19] 周世兵. 聚类分析中的最佳聚类数确定方法研究及应用[D]. 江苏 无锡: 江南大学, 2011.
- ZHOU Shi-bing. Research and application of the method for determining the best cluster number in cluster analysis[D]. Wuxi, Jiangsu: Jiangnan University, 2011.
- [20] 汤深伟,贾瑞玉. 基于改进粒子群算法的K均值聚类算法[J]. *计算机工程与应用*, 2019, 55(18): 140—145.
- TANG Shen-wei, JIA Rui-yu. K-means clustering algorithm based on improved particle swarm optimization[J]. *Computer Engineering and Applications*, 2019, 55(18): 140—145.
- [21] GONG Q, MIDLAM-MOHLER S, MARANO V, et al. An iterative Markov chain approach for generating vehicle driving cycles[J]. *SAE International Journal of Engines*, 2011(1): 1035—1045.