



中国研究生创新实践系列大赛  
“华为杯”第十六届中国研究生  
数学建模竞赛

学 校

福州大学

---

参赛队号

19103860074

---

队员姓名

1.林露馨

2.黄文婷

3.李政阳

---

# 中国研究生创新实践系列大赛

## “华为杯”第十六届中国研究生

### 数学建模竞赛

题目 基于改进 K-means 聚类 and 隐马尔可夫  
链的汽车行驶工况构建

#### 摘 要：

汽车行驶工况的构建对于车辆能耗、排放测试，以及汽车性能指标优化有着重要意义。而欧洲 NEDC 工况、世界 WLTC 工况等为基准的行驶工况构建并不完全适用我国汽车行驶工况的构建，基于此，制定反映我国实际道路行驶状况的工况显得越来越重要。

**数据预处理（问题一）**首先，利用高德地图的 API 接口，批量导入数据文件中的经纬度信息，绘制出一辆车在连续一周内的**行驶轨迹图**。对行驶轨迹图进行分析可得：本文测试车辆行驶主要路段在福州境内。其次，由于 GPS 信号在车辆过隧道或进入高层建筑时存在漂移和信号不稳定的情况，故删除  $x$ 、 $y$ 、 $z$  三轴的加速度，并通过计算相邻数据的速度间隔和时间间隔之比，得到该车辆的加速度。在完成数据初步处理后，对问题一要求中的不良数据类型进行分情况处理：（1）对于时间不连续的数据，做删除和修正处理，其操作分别为，将时间间隔大于 3 秒的数据进行删除，和对时间间隔为 2 秒的数据进行**插值补充**；（2）对于加、减速度异常数据，设定**加、减速度的阈值**对其进行删除处理；（3）对于长时间（时间持续大于 180 秒）停车、堵车等怠速行驶状态进行删除处理。最后，得到三个数据文件的**剩余记录数**，分别为 **174917**，**131141**，**145319**。经数据预处理后，保留了约 **85%** 的原始采集数据，有助于构建准确的汽车行驶工况。

**运动学片段的提取（问题二）**根据查阅文献，依据现有研究对怠速状态的判定标准为汽车的速度满足  $v < 1\text{km/h}$ ，并且加速度满足  $|a(i)| < 0.1\text{m}\cdot\text{s}^{-2}$ ，本文采取该标准来判定怠速状态。运动学片段的提取步骤为：（1）对怠速状态的第一个初始时刻进行标记，以此为约束条件，用 MATLAB 分别标记出怠速状态的起始点位和终止点位；（2）对时间间隔不为 1 秒的数据进行删除操作，保证了运动学片段的连续性；（3）根据标记点位，对时间连续的运动学片段进行提取。得到三个数据文件的**运动学片段数量**，分别为 **674** 个，**413** 个，**374** 个。

**汽车行驶工况的构建（问题三）**首先，通过查阅文献对福州的发展程度、交通状况有了初步了解，为问题三的模式构建奠定了理论基础；其次，为了构建一条时间间隔为 1200-1300 秒的行驶工况曲线，本文提出了**两种汽车行驶工况模型**，为：（1）基于改进的 k-means 聚类和隐马尔可夫链的汽车行驶工况模型，（2）多目标 0-1 线性规划的汽车行驶工况模型；然后，确定汽车运动特征评估体系，综合各方考量，本文提出了 **16 个特征指标**，即：**平均速度、平均行驶速度、怠速时间比、片段持续时间、加速时间比、减速时间比、匀速时间比、最高速度、最大加速度、最大减速度、平均加速度、平均减速度、速度**

标准差、加速度标准差、减速度标准差、行驶距离；最后，对特征指标计算其与实际数据的相对误差进行模型检验，通过实验发现：本文提出的两种模型，其相对误差均小于8.032%，基本满足工业背景需要。但是，多目标0-1线性规划的汽车行驶工况模型相较于改进的k-means聚类和隐马尔科夫链的汽车行驶工况模型性能较弱，与真实值间的相对误差较大，故本文采用基于改进的k-means聚类和隐马尔科夫链的汽车行驶工况模型作为本文的汽车行驶工况模型。

最后，关于构建汽车行驶工况模型，本文提出了几点想法：（1）可以从曲线拟合的角度，考虑求解高阶拟线性方程组；（2）可以根据汽车行驶工况曲线和道路路径分类（由GPS根据经纬度解析地址，由解析出的地址作为道路路径分类标准）对瞬时油耗进行分析。

**关键词：改进 k-means；隐马尔科夫链；转移矩阵；多目标线性规划；汽车工况；**

# 目录

一、问题重述	1
1.1 研究背景	1
1.2 问题重述	1
二、模型假设	2
三、模型符号说明	2
四、问题一的模型建立与求解	4
4.1 问题分析	4
4.2 数据预处理	5
4.2.1 GPS 速度异常数据	6
4.2.2 时间不连续数据	6
4.2.3 缺失数据补充	7
4.2.4 删除不良数据	8
4.2.4 涉及的计算公式	11
4.2.5 预处理后的统计分析	11
五、问题二的模型建立与求解	12
5.1 问题分析	12
5.2 数据处理	12
5.2.1 工况状态划分标准	12
5.2.2 时间不连续处理	12
5.3 提取运动学片段	12
六、问题三的模型建立与求解	14
6.1 问题分析	14
6.2 运动特征提取	14
6.3 数据处理	14
6.3.1 标记工况状态	14
6.3.2 计算特征指标值	15
6.3.3 确定候选片的特征矩阵	17
6.4 汽车行驶工况模型的建立与求解	21
6.4.1 基于改进 k-means 聚类与隐马尔科夫链的行驶工况模型——模型 1	21
6.4.2 模型 1 的求解	22
6.4.3 基于多目标 0-1 线性规划的行驶工况模型——模型 2	23
6.4.4 模型 2 的求解	24
6.5 模型求解与检验	26
6.5.1 模型求解	26
6.5.2 模型检验	28
七、模型评价	30
7.1 模型的优点	30
7.2 模型的不足	30
7.3 模型的改进	30
八、参考文献	31
九、附录	32

# 一、问题重述

## 1.1 研究背景

汽车行驶工况，是描述汽车行驶过程的速度-时间曲线。该曲线是体现汽车道路行驶的运动学特征之一。通过对汽车行驶工况曲线进行分析，可以得到以下特征指标，其分别为平均速度、平均行驶速度、平均加速度、平均减速度、怠速时间比、加速时间比、减速时间比、速度标准差、加速度标准差等。通过这些特征指标，可以进一步为车辆性能标定优化和能耗/排放认证提供扎实的理论依据。

目前，一些发达国家和组织制定了适合自己的汽车行驶工况的标准，如，欧洲 NEDC 工况、世界 WLTC 工况等行驶工况构建。2000 年初，我国开始直接采用欧洲的 NEDC 行驶工况对汽车产品能耗/排放的认证。然而经过多年实践发现：欧洲的 NEDC 行驶工况实际油耗与我国法规认证结果偏差越来越大。因而我国开始转变采用世界轻型车测试循环（WLTC）作为我国汽车行驶工况的标准。但该工况中的怠速时间比和平均速度这两个主要工况特征，与我国实际汽车行驶工况的差异较大。因此，进行基于城市自身数据的汽车行驶数据的城市汽车行驶工况的构建研究至关重要。

## 1.2 问题重述

基于上述的研究背景，本文需要研究和完成以下问题：

### 问题 1：数据预处理

由汽车行驶数据的采集设备直接记录的原始采集数据往往会包含一些不良数据值，不良数据主要包括几个类型：

（1）由于高层建筑覆盖或过隧道等，GPS 信号丢失，造成所提供数据中的时间不连续；

（2）汽车加、减速度异常的数据（普通轿车一般情况下：0 至 100km/h 的加速时间大于 7 秒，紧急刹车最大减速度在 7.5-8 m/s<sup>2</sup>）；

（3）长期停车（如停车不熄火等候人、停车熄火了但采集设备仍在运行等）所采集的异常数据。

（4）长时间堵车、断断续续低速行驶情况（最高车速小于 10km/h），通常可按怠速情况处理。

（5）一般认为怠速时间超过 180 秒为异常情况，怠速最长时间可按 180 秒处理。

请设计合理的方法将上述不良数据进行预处理，并给出各文件数据经处理后的记录数。

### 问题 2：运动学片段的提取

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，如图 1 所示（基于运动学片段构建汽车行驶工况曲线是日前最常用的方法之一，但并不是必须的步骤，有些构建汽车行驶工况曲线的方法并不需要进行运动学片段划分和提取）。请设计合理的方法，将上述经处理后的数据划分为多个运动学片段，并给出各数据文件最终得到的运动学片段数量。

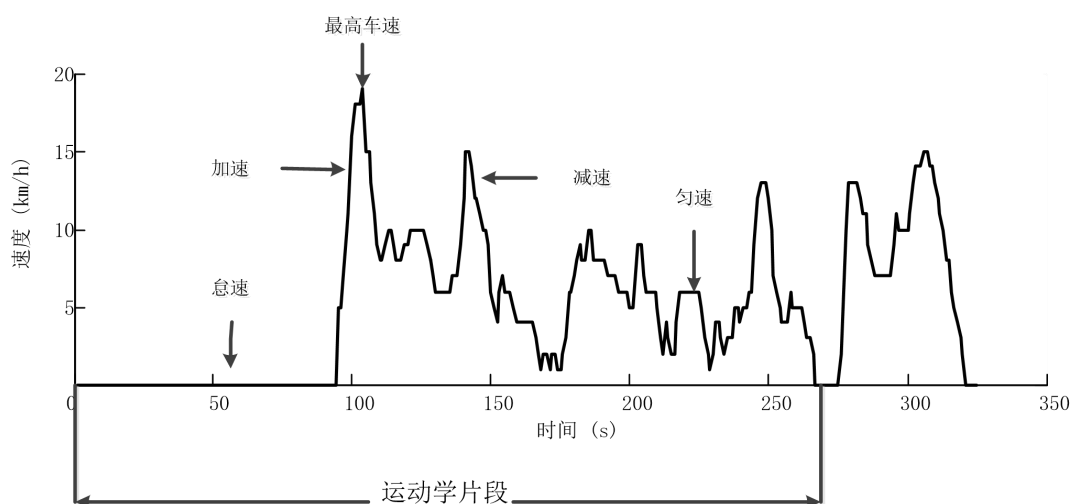


图 1 运动学片段

### 问题 3：汽车行驶工况的构建

请根据上述经处理后的数据，构建一条能体现参与数据采集汽车行驶特征的汽车行驶工况曲线（1200-1300 秒），该曲线的汽车运动特征能代表所采集数据源（经处理后的数据）的相应特征，两者间的误差越小，说明所构建的汽车行驶工况的代表性越好。要求：

- (1) 科学、有效的构建方法；
- (2) 合理的汽车运动特征评估体系；
- (3) 按照你们所构建的汽车行驶工况及汽车运动特征评估体系，分别计算出汽车行驶工况与该城市所采集数据源（经处理后的数据）的各指标（运动特征）值，并说明你们所构建的汽车行驶工况的合理性。

## 二、模型假设

- (1) 假设暂不考虑气候条件、发展状况等对汽车行驶工况的影响；
- (2) 除速度外的其余单位均采用国际单位，速度单位采用 km/h；
- (3) 假设汽车的加、减速度等于相邻速度差与时间间隔之比，暂不考虑  $x, y, z$  轴加速度的矢量合成。
- (4) 对于单个数据缺失，假设汽车在较短的时间间隔内是匀速行驶的，对数据进行补充。
- (5) 假设长期停车、长时间堵车、断断续续低速行驶均可按怠速情况处理，时间间隔超过 180s 即认定为长时间停车、堵车。
- (6) 假设汽车在高速公路上行驶速度受限制，最高车速被限制在 120km/h，超过这个速度为异常数据，做删除处理。

### 三、模型符号说明

符号	符号说明
$V_a / km \cdot h^{-1}$	平均速度
$V_x / km \cdot h^{-1}$	平均行驶速度
$T_i / \%$	怠速时间比
$a_v / m \cdot s^{-2}$	平均加速度
$a_d / m \cdot s^{-2}$	平均减速度
$T_a / \%$	加速时间比
$T_d / \%$	减速时间比
$V_{std} / km \cdot h^{-1}$	速度标准差
$a_{std} / m \cdot s^{-2}$	加速度标准差
$\rho_{OA}(i)$	瞬时能耗
$L_1(i)$	某时刻的经度
$L_2(i)$	某时刻的纬度
$r$	地球平均半径
$D$	经纬度换算距离
$\bar{v}_1 / km \cdot h^{-1}$	平均速度
$\bar{v}_2 / km \cdot h^{-1}$	平均行驶速度
$t_1 / \%$	怠速时间比
$t_3 / \%$	减速时间比
$t_4 / \%$	匀速时间比
$t_2 / \%$	加速时间比
$V_{max} / km \cdot h^{-1}$	最高速度
$a_{max} / m \cdot s^{-2}$	最大加速度
$a_{min} / m \cdot s^{-2}$	最大减速度
$\bar{a}_1 / m \cdot s^{-2}$	平均加速度
$\sigma_1 / km \cdot h^{-1}$	速度标准差
$\sigma_2 / m \cdot s^{-2}$	加速度标准差

## 四、问题一的模型建立与求解

### 4.1 问题分析

根据问题一要求，由于汽车行驶的采集设备直接记录的原始采集数据会包含一些不良的数据，主要包括：（1）高层建筑覆盖或隧道路段出现 GPS 信号丢失；（2）汽车加减速速度；（3）长期停车；（4）长时间堵车、断断续续低速行驶；（5）怠速超时等五种类型。所涉及到的异常数据处理较复杂，需要考虑路段交通状况等更多因素。

题目已知的条件是某城市轻型汽车实际道路行驶采集的数据，3 个数据文件分别为同一辆车在不同时段内采集的数据。首先，对数据进行初步分析，发现存在不良数据（数据缺失、数据异常、经纬度出现零值）等问题，通过制定具体、有效的汽车工况数据的处理流程，实现对三个数据文件的数据预处理工作。

其次，需要对五类不良数据产生的情况进行分析，主要通过 GPS 速度、时间差、平均加速度、平均减速度以及发动机转速等指标参数，对不良数据按汽车工况类型进行数据处理，这样可以较为精确地处理异常数据值，提高汽车行驶工况模型的建立和求解的效率和精度。数据处理的流程图如下图 2 所示：

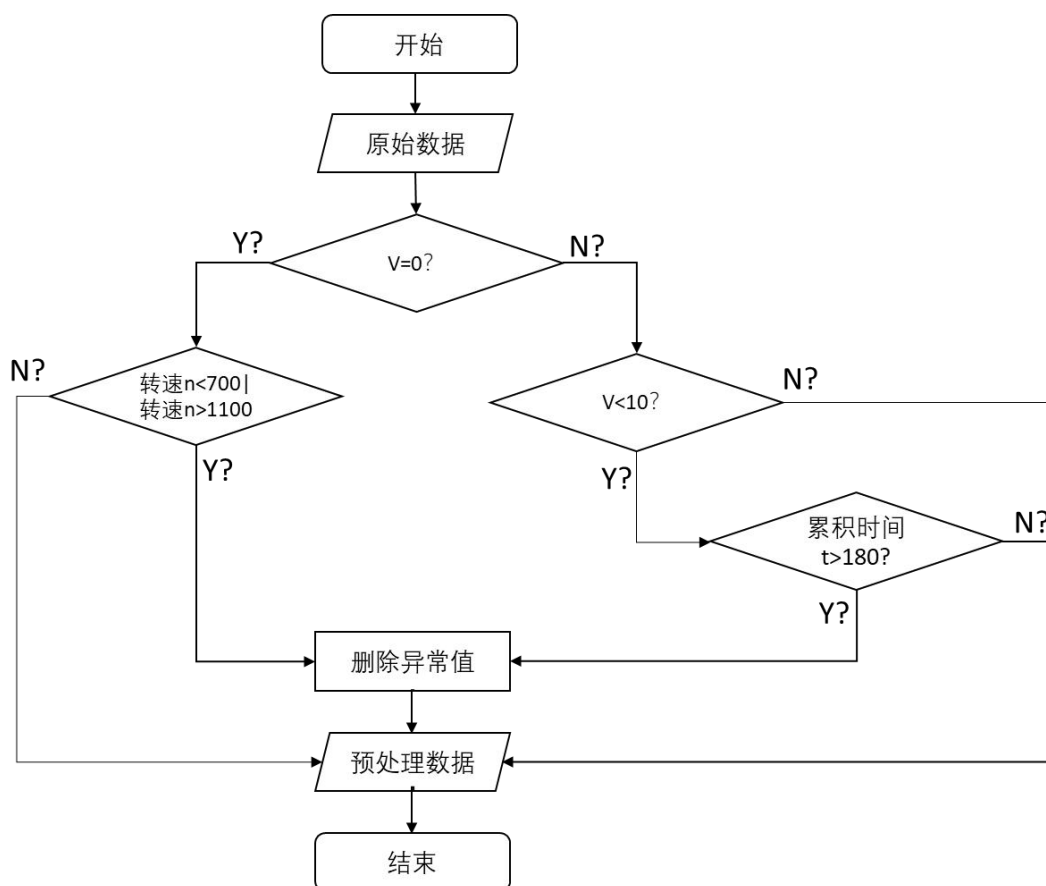


图 2 数据预处理流程



## 4.2 数据预处理

首先，借助高德地图的搜索地区经纬度功能，导入每个数据条的经纬度信息，可以分别得到同一辆车在不同时段下的行驶路径。如下图3、图4、图5的汽车行驶轨迹所示：

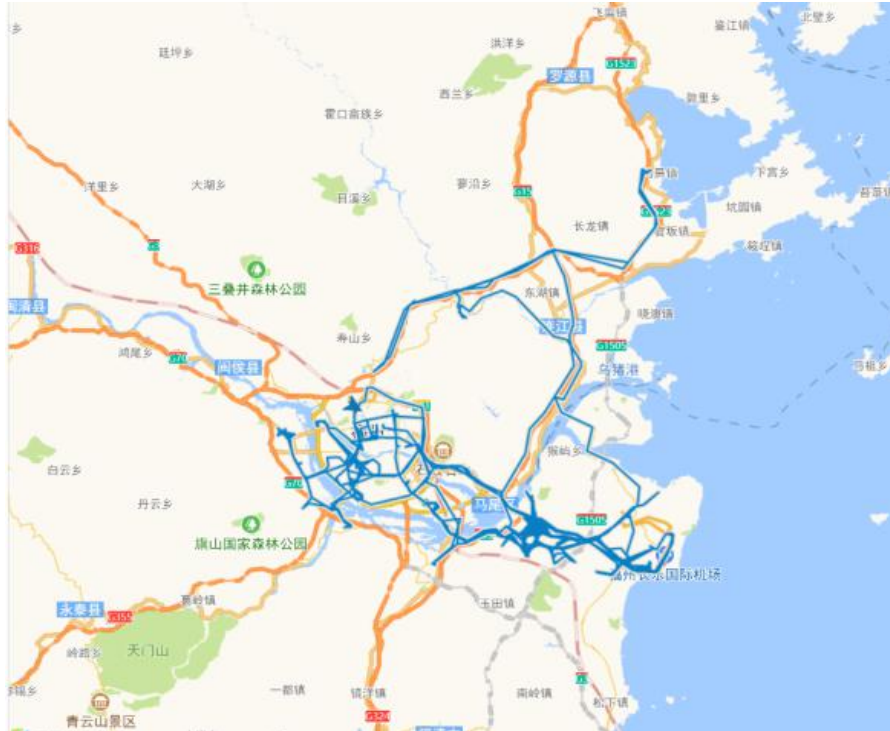


图3 汽车行驶轨迹（文件1）

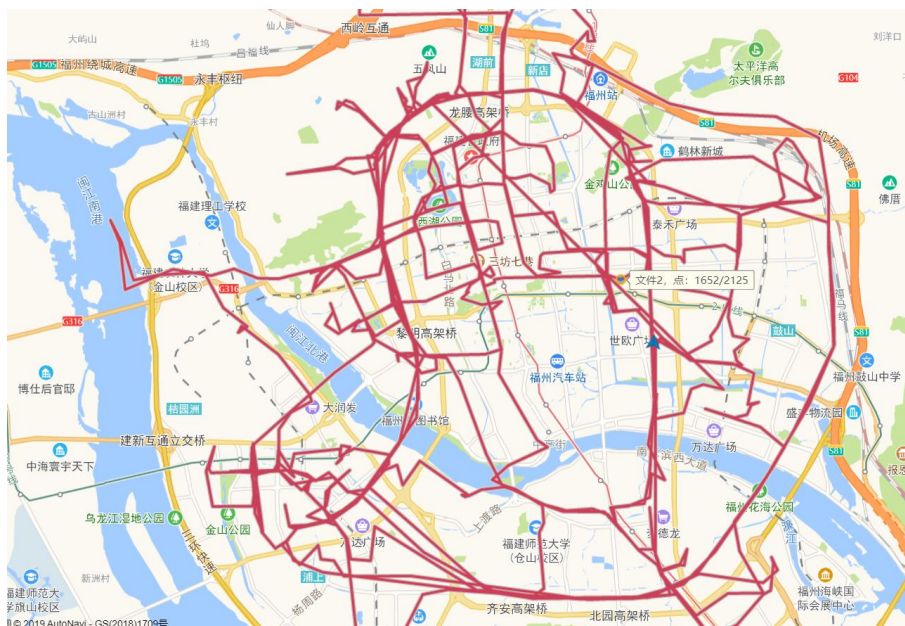


图4 汽车行驶轨迹（文件2）



图 5 汽车行驶轨迹（文件 3）

根据图 3、4、5 的汽车行驶轨迹，可以确定该问题研究的是轻型汽车福州市的行驶工况以及福州市的经纬度信息，经度为 119.30，纬度为 26.08。以下 4.2.1 节、4.2.2 节、4.2.3 节详细地列出来具体的数据处理过程。

#### 4.2.1 GPS 速度异常数据

首先，利用 Excel 对文件 1，2，3 的 GPS 速度进行筛选，发现文件 1，2 的 GPS 车速均为 120km/h 以内，而文件 3 有 298 个的 GPS 速度远超 120km/h，需要对这部分的超速数据进行处理。

汽车在高速公路上行驶的最高车速限制为 120km/h，所以文件 3 中出现的 GPS 车速大于 120km/h 的数据都要做相应的删除处理。

#### 4.2.2 时间不连续数据

处理第（1）种情况的异常，即：高层建筑覆盖或过隧道导致 GPS 信号丢失，数据存在时间不连续，对于这类不良数据，其相邻两条数据的时间间隔为  $\Delta t \neq 1$ ，考虑到 GPS 信号丢失的时间长度也不短，故删除时间间隔大于 3s 的数据，对于时间间隔为 2s 的数据，可以用插值补充方式修正数据。

对于 3 个数据文件，进行 Excel 分析各个工况下的经纬度数据，通过筛选得到“文件 3”的经纬度数据异常值，筛选出 300 条经纬度为 0 的异常记录条，如下表 1 截取了文件 3 的经纬度异常部分数据所示：

表 1 文件 3 的经纬度异常数据

时间	经度	纬度
2017/11/02 11:14:03	0	0
2017/11/04 10:37:59	0	0
2017/11/04 10:38:00	0	0
2017/11/04 10:38:01	0	0
.....	.....	.....
2017/11/04 10:38:30	0	0
2017/11/04 10:38:31	0	0
2017/11/04 10:38:32	0	0

由于文件 2 的总记录数为 145825 条，300 条异常数据对后续研究影响较小，可考虑先剔除该 300 条数据。在进行模型优化的时候再通过取经纬度平均值加入具体研究中。

### 4.2.3 缺失数据补充

处理第 (1) 种情况的异常，即：高层建筑覆盖或过隧道导致 GPS 信号丢失，数据存在时间不连续，对于这类不良数据，可以先对数据缺失个数进行分类，分为单个数据缺失和多组数据缺失。下面分别就两种情况的数据处理进行讨论。

#### (1) 单个数据缺失

由于汽车经过高层建筑或隧道的情况下，出现数据通道的不顺畅而导致数据缺失的情况，在这种较短采样时段内的 GPS 数据缺失或者漂移问题，假设汽车在较短的时间间隔内是匀速行驶的，故可利用数据缺失前后的 GPS 速度、加减速度、经纬度等数据进行取均值的补充，具体计算公式如式 (4-2-1) 所示：

$$\begin{cases} v(i) = \frac{v(i-1) + v(i+1)}{2} \\ L_1(i) = \frac{L_1(i-1) + L_2(i+1)}{2} \\ L_2(i) = \frac{L_2(i-1) + L_2(i+1)}{2} \end{cases} \quad (4-2-1)$$

其中， $v(i-1)$ ， $v(i+1)$  表示第  $i$  个时刻和第  $(i+1)$  的 GPS 速度， $(L_1(i-1), L_2(i-1))$ ， $(L_1(i+1), L_2(i+1))$  分别表示为缺失的时刻位置点前后两点的经纬度数据， $(L_1(i), L_2(i))$  为取均值后的当前缺失时刻的经纬度数据。由于汽车行驶的采集设备采集时间间隔为 1 秒，通过 Excel 筛选，得到相邻时间间隔为 2（缺失单个数据）的记录数为 111 条，下表 2 为（以文件 1 为例）截取的缺失单一数据的部分数据。

表 2 缺失单一数据的处理

时间	GPS 速度	经度	纬度
2017/12/18 14:17:18	37.85	119.4775	25.9580
2017/12/18 14:26:22	39.5	119.5041	25.9624
2017/12/18 15:44:34	21.1	119.5041	25.9622
.....	.....	.....	.....
2017/12/18 21:33:34	5.45	119.5181	25.9643
2017/12/18 00:01:44	87.6	119.4509	26.0070

## (2) 多组数据缺失

汽车在实际的行驶过程中，会因为气候、地理位置或路过高架桥等原因下造成长时间的数据通路不顺畅，从而导致数据集出现大片数据缺失的情况。在该情况下，数据缺失的数量多达十几个到几千个。其缺失的数据量信息巨大，仅通过取均值或者线性插值的方式都不能较好还原原始的汽车行驶数据，故对于这种情况，当数据的缺失量巨大而对整个分析结果造成影响，应该剔除汽车在该时段内的所有记录数据，减小该段数据对后续运动学片段提取带来干扰，增加分析的难度。[1]

### 4.2.4 删除不良数据

在上面的 4.2.2 中实现了对第一类不良数据的预处理。该节主要从删除不良数据的角度对问题一提出的第（2）、（3）、（4）、（5）种类型（以下记作 1,2,3,4 情况）进行数据处理操作。

不良数据的删除主要包括以下四个方面：

1、汽车加、减速度异常的数据（普通轿车一般情况下：0 至 100km/h 的加速时间大于 7 秒，紧急刹车最大减速度在  $7.5\sim 8\text{ m/s}^2$ ）；

2、长期停车（如停车不熄火等候人、停车熄火了但采集设备仍在运行等）所采集的异常数据。

3、长时间堵车、断断续续低速行驶情况（最高车速小于 10km/h），通常可按怠速情况处理。

4、一般认为怠速时间超过 180 秒为异常情况，怠速最长时间可按 180 秒处理。

对于这四种不良数据可以分为加减速度异常和怠速异常两种情况的分部处理。其中，1 是加减速度工况下的不良数据处理，2、3、4 是怠速工况下的不良数据处理。

#### (1) 加、减速度异常

由于普通轿车在一般情况下，从 0 加速到 100km/h 的加速时间要满足大于 7s，紧急刹车最大减速度在  $7.5\sim 8\text{ m/s}^2$ 。由于不同车型在紧急刹车的最大减速度存在差异，可以取最大减速度为  $7.75\text{ m/s}^2$ （减速度为 7.5 和 8 的均值）。这里假设加、减速度是根据速度差与时间间隔之比进行计算的，可忽略  $x, y, z$  轴的加速度数据，为了后续研究考虑，利用 Excel 对  $x, y, z$  轴加速度的数据列采取删除操作后，进行加、减速度的不良数据处理。

时间间隔如下式 (4-2-1) 所示:

$$\Delta t_i = t_i - t_{i-1} \quad (4-2-2)$$

加速度异常的表达式, 见下式 (4-2-3) 所示:

$$\begin{cases} v_i = 0 \text{ km/h} \\ v(i+7) \geq 100 \text{ km/h} \\ 0 < a_i < 10 \text{ m} \cdot \text{s}^{-2} \end{cases} \quad (4-2-3)$$

对加速度**设定阈值范围**, 则加速度的最大值不能超过  $10 \text{ m} \cdot \text{s}^{-2}$ 。

减速度异常的表达式, 见下式 (4-2-4) 所示:

$$\begin{cases} a(i) = \frac{v(i) - v(i-1)}{\Delta t} \\ a(i) < 0, a(i) \geq -7.75 \text{ m/s}^2 \end{cases} \quad (4-2-4)$$

其中,  $\Delta t_i$  表示从第  $(i-1)$  个时刻到第  $i$  个时刻的时间间隔,  $a(i)$  表示第  $i$  个时刻的瞬时加速度。下表 3 是利用 MATLAB 分别得到文件 1、文件 2、文件 3 的删除记录数据量统计情况。

表 3 加、减速度删除记录数

数据文件编号	加速度异常记录数	减速度异常记录数
1	85	354
2	159	972
3	39	671

在约束条件下, 得到不满足该约束的加减速速度对应的记录行, 对加减速速度异常的数据进行删除处理, 下表 4 为加、减速度异常删除的部分数据, 以文件 1 为例。

表 4 加、减速度异常数据

时间	GPS 车速	时间差	减速度
2017/12/20 08:01:56	26.2	1	-23.5
2017/12/20 08:10:10	45.6	1	-10.1
2017/12/20 12:00:41	43.8	1	-12.3
2017/12/20 12:10:49	20.6	1	-17.2
2017/12/21 11:50:36	77.5	1	-23.5
2017/12/21 13:20:24	51.1	1	-10.1
2017/12/21 19:13:14	0	1	-24
2017/12/22 18:13:55	7.4	1	-31.6
2017/12/23 09:43:58	3	1	-74.1
2017/12/23 14:57:53	19.6	1	-17.4
2017/12/24 00:42:09	33.5	1	-10.1

## (2) 怠速异常

假设长期停车、长时间堵车、断断续续低速行驶均可按怠速情况处理, 时间间隔超过 180s 即认定为长时间停车、堵车。因此不良数据类型 (3)、(4)、(5) 统一归为怠速情况处理。

### 类型 (3) 数据处理

首先分析类型 (3) 的异常数据处理, 长期停车主要包括: 1、停车不熄火; 2、停车

熄火但采集设备仍在运行。满足  $\Delta t_i \geq 180s$  的情况为长期停车。对于 1、停车不熄火的行驶指标状态进行分析，汽车停车不熄火状态下，速度  $v = 0$ ，不熄火则汽车发动机仍在运行，加速度  $a_i \neq 0$ ，以此为约束条件，进行删除处理。

对于 2、停车熄火但设备仍在运行情况进行分析，汽车停车熄火下，速度  $v_i = 0$ ，熄火后汽车发动机不运行，汽车不会发生位置移动，有： $a_i = 0, s_i = 0$ 。

其中， $s_i$  表示汽车在  $i$  时刻的位移数据，这里的位移是根据平均速度与时间间隔的乘积公式计算得到的。

#### 类型（4）数据处理

其次，对长时间堵车、断断续续低速行驶的状态进行分析，根据题目所给的条件，断断续续低速行驶下的最大车速  $\max v_i < 10km/h$ ，且通常可按照怠速状态处理，以此为约束，当断断续续低速行驶下的车速大于等于  $10km/h$  的数据记录进行删除操作，具体约束条件如下：

$$\begin{cases} \max v_i < 10km/h \\ \Delta t_i \geq 180s \end{cases} \quad (4-2-5)$$

根据上式（4-2-6）的约束，可以得到长时间堵车的部分异常值处理。

#### 类型（5）数据处理

怠速状态的判断方法：据题目所给的关于怠速的名词解释可知，怠速状态下，汽车是停止运动的，即满足  $v_i$  为 0；其次，怠速状态下汽车发动机未熄火，故其加速度不为 0；同时，根据一般下的怠速时间超过 180 为异常情况，可以得到下式（4-2-6）的约束条件：

$$\begin{cases} v_i < 1km/h, \\ |a_i| < 0.1m \cdot s^{-2}, \\ \Delta t_i \geq 180s \end{cases} \quad (4-2-6)$$

根据上式，可以得到怠速超时异常的数据。

### 4.2.4 涉及的计算公式

其次，对数据进行经纬度换算距离，归一化处理，列出了相关的计算公式；最后利用 MATLAB 得出最终处理后的数据记录数。

涉及的计算公式如下：

（1）经纬度换算距离公式：

$$D = \arccos((\sin L_2(i) \cdot \sin L_2(i+1)) + (\cos L_2(i) \cdot \cos L_2(i+1) \cdot \cos(L_1(i) - L_1(i+1)))) \cdot r \quad (4-2-7)$$

其中， $L_1(i)$  表示  $i$  时刻的经度， $L_2(i)$  表示  $i$  时刻的纬度， $r$  为地球平均半径。

(2) 扭矩与发动机转速关系:

$$P(i) = \pi \cdot M(i) \cdot n(i) \quad (4-2-8)$$

其中,  $P(i)$ 表示第*i*个时刻的汽车行驶功率,  $M(i)$ 表示第*i*个时刻的汽车扭矩,  $n(i)$ 表示第*i*个时刻的汽车发动机转速。

(3) 空燃比公式:

$$\delta(i) = \frac{M_1(i)}{M_2(i)} \times 100\% \quad (4-2-9)$$

其中,  $M_1(i)$ 表示第*i*个时刻的汽车燃料燃烧所需的空气量,  $M_2(i)$ 表示第*i*个时刻的汽车行驶燃烧的汽油量,  $\delta(i)$ 表示第*i*个时刻的汽车空燃比。

(4) 扭矩百分比公式:

$$\gamma(i) = \frac{M(i)}{M_{\max}} \times 100\% \quad (4-2-10)$$

其中,  $\gamma(i)$ 表示第*i*个时刻的汽车扭矩百分比,  $M_{\max}$ 表示第*i*个时刻的汽车在该转速下所能达到的最大扭矩。

#### 4.2.5 预处理后的统计分析

由此完成了整个数据预处理过程, 经过问题一的数据预处理后, 文件 1, 2, 3 的剩余数据记录数见下表 5 所示:

表 5 经处理后的记录数

数据文件编号	原始采集数据量	删除数据量	数据补充数据量	数据的剩余记录数
1	185725	10676	132	174917
2	145825	16328	164	131141
3	164914	19728	133	145319

## 五、问题二的模型建立与求解

### 5.1 问题分析

根据问题二的要求, 基于问题一得到的预处理后的数据集进行运动学片段的提取。由于汽车在行驶过程中, 受到路面交通情况的影响, 会出现多种怠速、加速、匀速和减速的状态。故定义了运动学片段为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。

由问题一得到经处理后的数据, 会存在某段时间上的不连续, 但是根据运动学片段的定义, 说明了每个运动学片段都是汽车在某段连续时间上行驶发生的速度变化, 反映了汽车在该时段的行驶工况。因此, 在进行运动学片段划分前, 需要对不连续的数据进行进一步的数据处理工作。在此基础上, 根据运动学片段的定义, 利用怠速状态进行片段的划分

和提取。

## 5.2 数据处理

### 5.2.1 工况状态划分标准

为了明确地区分汽车不同的工况，需要对不同工况下汽车行驶的速度和加速度设定一个标准【3】，如下所示：

- (1) 怠速工况：汽车的速度满足  $v < 1km/h$ ，并且加速度满足  $|a(i)| < 0.1m \cdot s^{-2}$  的工作状态；
- (2) 加速工况：汽车的速度满足  $v \neq 0km/h$ ，并且加速度满足  $a \geq 0.1m \cdot s^{-2}$  的工作状态；
- (3) 减速工况：汽车的速度满足  $v \neq 0km/h$ ，并且加速度满足  $a \leq -0.1m \cdot s^{-2}$  的工作状态；
- (4) 匀速工况：汽车的速度满足  $v \geq 1km/h$ ，并且加速度满足  $a > 0.1m \cdot s^{-2}$  的工作状态。

### 5.2.2 时间不连续处理

由于问题一预处理后的数据在时间上不一定是连续的，需要利用相邻数据条的时间间隔是否为 1 秒，作为判断数据是否连续的指标。利用 MATLAB 进行时间间隔的处理，仅保留时间间隔  $\Delta t = 1s$  的数据集，对于时间间隔  $\Delta t > 1s$  的数据，进行相应的删除。

## 5.3 提取运动学片段

首先根据 5.1.1 工况状态划分标准，确定了怠速情况下的汽车速度满足  $v < 1km/h$ ，并且加速度满足  $|a(i)| < 0.1m \cdot s^{-2}$ 。其次，对一个完整的运动学片段过程进行分析，可知运动学片段的第一个初始时刻和运动学片段的结束时刻，其速度都满足  $v = 0km/h$ ，且  $a \neq 0$ 。以此为约束条件，用 MATLAB 分别标记出运动学片段的起始点位和终止点位。

下图 6, 7, 8 分别绘制了文件 1,2,3 任意一个运动学片段的图像。

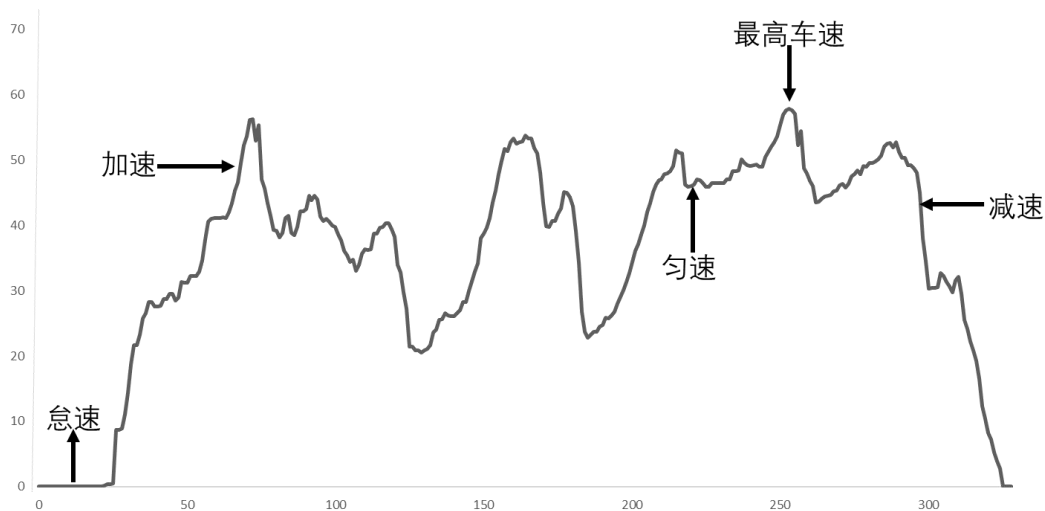


图 6 文件 1 的某个运动学片段



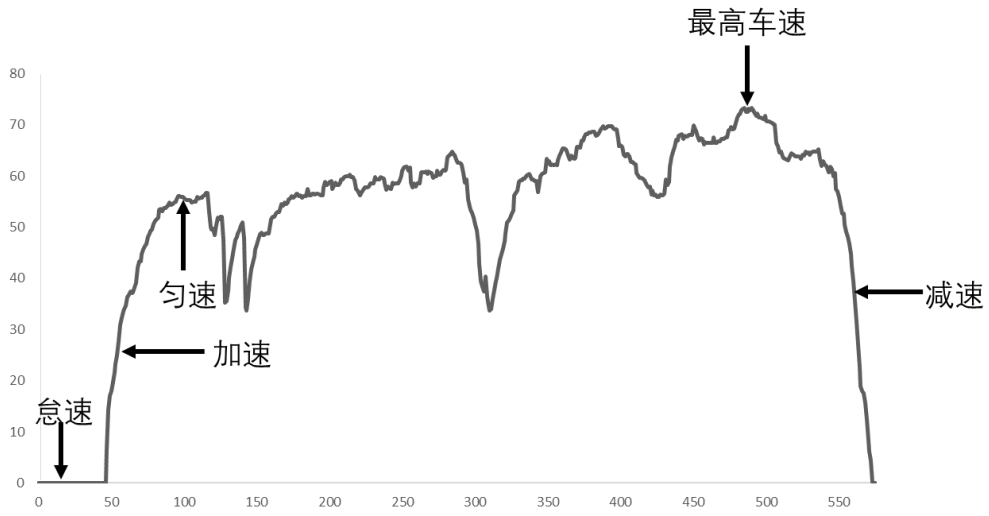


图 7 文件 2 的某个运动学片段

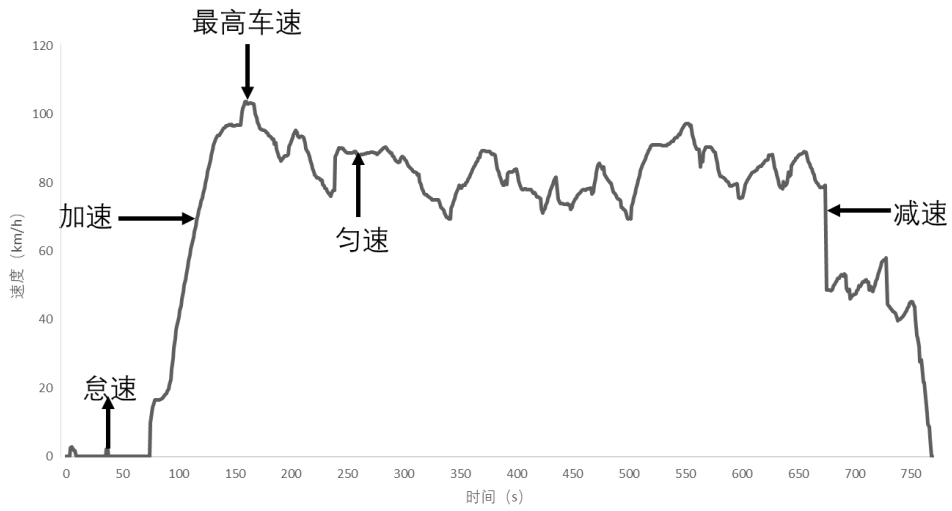


图 8 文件 3 的某个运动学片段

利用 MATLAB 软件，根据标记点位，对时间连续的运动学片段进行提取，得到三个数据文件的运动学片段数量，分别为 **674 个**，**413 个**，**374 个**。

## 六、问题三的模型建立与求解

### 6.1 问题分析

根据问题 3 的要求,构建一条能体现参与数据采集汽车行驶特征的汽车行驶工况曲线,且约束了时间条件,其时间的横坐标满足 1200-1300s。这里,我们取 1200s 作为要绘制的行驶工况曲线的横坐标时间范围。

首先,需要提取影响速度和时间的运动特征,如平均速度、平均行驶速度、怠速时间比、加速时间比等指标参数,对各个指标的数据进行处理,得到每个数据文件对应的一个候选工况分布特征矩阵;其次,通过建立合理的汽车运动特征评估体系,构建出汽车行驶工况曲线;最后,建立综合评价模型,对模型得到的拼接运动学片段与该城市(经处理后的数据)的实际行驶工况曲线进行误差对比,检验模型的效果。

### 6.2 运动特征提取

由于运动学片段为某一怠速开始至下一个怠速开始前的车速区间,而汽车实际的行驶过程可看作是大量的运动学片段的拼接,依照这一原则,利用问题二得出的三个数据文件对应的运动学片段数目,选取合理的运动特征指标,使得每一个运动学片段都有与之对应的运动特征指标,并且由这些运动特征指标确定对应的每一个运动学片段。

下表 6 展现的是我们用于描述运动学片段所选取的 16 个特征指标。

表 6 用于描述运动学片段的 16 个特征指标

序号	特征参数	定义	序号	特征参数	定义
1	$\bar{v}_1$	平均速度	9	$a_{\max}$	最大加速度
2	$\bar{v}_2$	平均行驶速度	10	$a_{\min}$	最大减速度
3	$t_1$	怠速时间比	11	$\bar{a}_1$	平均加速度
4	$n$	片段持续时间	12	$\bar{a}_2$	平均减速度
5	$t_2$	加速时间比	13	$\sigma_1$	速度标准差
6	$t_3$	减速时间比	14	$\sigma_2$	加速度标准差
7	$t_4$	匀速时间比	15	$\sigma_3$	减速度标准差
8	$v_{\max}$	最高速度	16	$s_1$ 或 $s_2$	行驶距离

### 6.3 数据处理

#### 6.3.1 标记工况状态

由于汽车的工况状态,主要分为怠速工况状态,加速工况状态,匀速工况状态,减速工况状态四种状态。为了后续数据处理的方便,首先对每个运动学片段额四种工况状态出

现个数进行标签化，利用 MATLAB 将怠速状态标记为 1，加速状态标记为 2，减速状态标记为 3，匀速状态标记为 4，统计得到第  $i$  个运动学片段的怠速状态个数为  $l_i$ ，加速状态个数为  $k_i$ ，匀速状态个数为  $h_i$ ，减速状态个数为  $g_i$ 。现截取文件 1 预处理后的任意一个运动学片段，下表 7 是该部分运动学片段的标记情况和各个文件工况状态的个数。

表 7 工况状态标记和个数统计（部分）

具体时刻	$v$	$a$	标签编号
2017/11/1 19:11:50	0	0	1
2017/11/1 19:15:09	11.2	0.056281	4
2017/11/1 19:15:10	10.3	-0.9	3
2017/11/1 19:15:11	8.6	-1.7	3
2017/11/1 19:15:12	7.5	-1.1	3
2017/11/1 19:15:13	9.2	1.7	2
2017/11/1 19:15:14	12.2	3	2

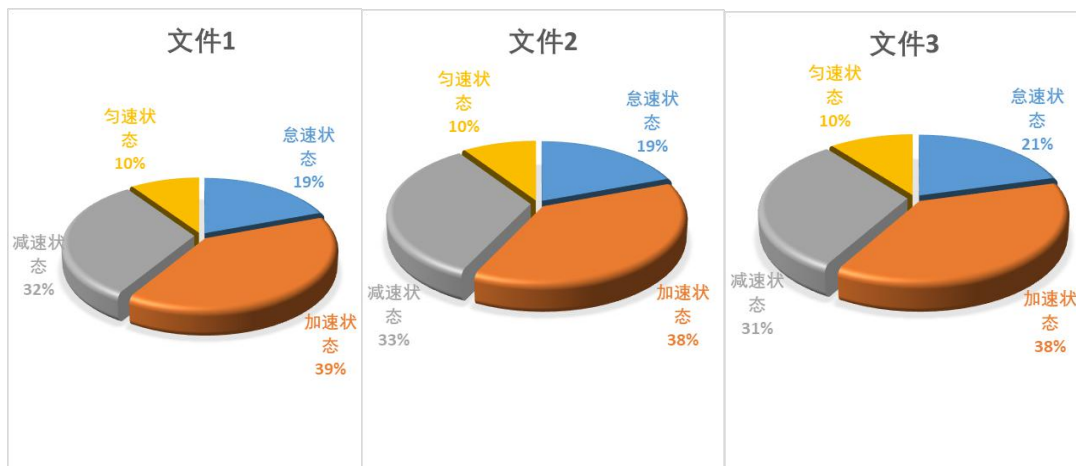


图 9 汽车工况状态频率

### 6.3.2 计算特征指标值

在确定了 16 个特征指标后，需要得到每个运动学片段对应的 16 个特征指标的具体值。下面是这 16 个特征指标的计算公式：

#### (1) 片段持续时间

$$n_i = (\Delta t)_i \quad (6-3-1)$$

其中， $n$  表示第  $i$  个运动学片段的持续时间； $(\Delta t)_i$  表示第  $i$  个运动学片段的结束时刻与初始时刻之差。

#### (2) 平均速度

$$\bar{v}_i = \frac{\sum_{j=1}^n v_j}{n} \quad (6-3-2)$$

其中， $\bar{v}_1$  表示第  $i$  个运动学片的平均速度； $v_j$  表示第  $j$  个时刻的 GPS 速度。

### (3) 平均行驶速度

$$\bar{v}_2 = \frac{\sum v_j}{m} \quad (6-3-3)$$

其中， $\bar{v}_2$  表示第  $i$  个运动学片的平均行驶速度； $l_i$  表示第  $i$  个运动学片中怠速状态的标签个数。

### (4) 怠速时间比

$$t_1 = \frac{l_i}{n_i} \times 100\%$$

(6-3-4)

其中， $t_1$  表示第  $i$  个运动学片的怠速时间比， $l_i$  为第  $i$  个运动学片的怠速状态个数。

### (5) 加速时间比

$$t_2 = \frac{k_i}{n_i} \times 100\% \quad (6-3-5)$$

其中， $t_2$  表示第  $i$  个运动学片的加速时间比， $k_i$  为第  $i$  个运动学片的加速状态个数。

### (6) 减速时间比

$$t_3 = \frac{g_i}{n_i} \times 100\% \quad (6-3-6)$$

其中， $t_3$  表示第  $i$  个运动学片的减速时间比， $g_i$  为第  $i$  个运动学片的减速状态个数。

### (7) 匀速时间比

$$t_4 = \frac{h_i}{n_i} \times 100\% \quad (6-3-7)$$

其中， $t_4$  表示第  $i$  个运动学片的匀速时间比， $h_i$  为第  $i$  个运动学片的匀速状态个数。

### (8) 最高速度

$$v_{\max} = \max v_j \quad (6-3-8)$$

其中， $v_{\max}$  表示第  $i$  个运动学片中的速度最大值。

### (9) 最大加速度

$$a_{\max} = \max a_j \quad (6-3-9)$$

其中， $a_{\max}$  表示第  $i$  个运动学片中的加速度最大值。

### (10) 最大减速度

$$a_{\min} = \min a_j \quad (6-3-10)$$

其中， $a_{\min}$  表示第  $i$  个运动学片中的减速度最大值。

### (11) 平均加速度

$$\bar{a}_1 = \frac{\sum a_{k_i}}{k_i} \quad (6-3-11)$$

其中， $\bar{a}_1$  表示第  $i$  个运动学片中的平均加速度， $k_i$  为第  $i$  个运动学片的加速状态个数。

(12) 平均减速度

$$\bar{a}_2 = \frac{\sum a_{g_i}}{g_i}$$

(6-3-12)

其中， $\bar{a}_2$ 表示第*i*个运动学片段中的平均减速度， $g_i$ 为第*i*个运动学片段的减速状态个数。

(13) 速度标准差

$$\sigma_1 = \sqrt{\frac{1}{n_i} \cdot \sum_{n_i} (v_j - \bar{v}_1)^2}$$
 (6-3-13)

其中， $\sigma_1$ 表示第*i*个运动学片段中的速度标准差（包括怠速状态）。

(14) 加速度标准差

$$\sigma_2 = \sqrt{\frac{1}{k_i} \cdot \sum_{k_i} (a_j - \bar{a}_1)^2}$$
 (6-3-14)

其中， $\sigma_2$ 表示第*i*个运动学片段中的加速度标准差。

(15) 减速度标准差

$$\sigma_3 = \sqrt{\frac{1}{g_i} \cdot \sum_{g_i} (a_j - \bar{a}_2)^2}$$
 (6-3-15)

其中， $\sigma_3$ 表示第*i*个运动学片段中的减速度标准差。

(16) 行驶距离

$$s_1 = D_i \text{ 或 } s_2 = \bar{v}_1 \cdot \Delta t$$
 (6-3-16)

其中， $s_1$ 是通过经纬度换算距离的方式； $s_2$ 是通过平均速度乘以时间差计算的距离。由于经纬度换算距离计算出来的 $s_1$ 是直线距离， $s_2$ 计算出来的是矢量位移距离，是符合汽车行驶轨迹的，所以在计算行驶距离时，应当使用 $s_2$ 的公式求解。

6.3.3 确定候选片段的特征矩阵

根据 6.3.2 的 16 个特征指标计算公式，可以分别得到 3 个数据文件的每个运动学片段的特征矩阵，如下表 8 所示。

表 8 文件 1 的运动学片段特征矩阵

片段	$\bar{v}_1$	$\bar{v}_2$	$\bar{a}_1$	$\bar{a}_2$	$t_1$	$t_2$ $t_4$	$t_3$	
1	6.9188	7.9733	1.3931	-1.9143	0.1304	0.4203	0.3043	0.1739
2	26.1249	35.4452	1.1848	-1.4009	0.2629	0.3577	0.3008	0.0840
3	20.1551	25.3117	1.4018	-1.8972	0.2034	0.4746	0.3051	0.0339
4	27.1863	40.7863	1.1493	-1.5561	0.3333	0.3425	0.2603	0.0731
.....	.....	.....	.....	.....	.....	.....	.....	.....
672	27.2545	35.8821	1.3071	-1.4708	0.2403	0.3636	0.3117	0.0974
673	33.4227	34.6373	0.9865	-0.8736	0.0350	0.3881	0.4510	0.1329
674	51.5542	53.1947	0.7110	-0.9064	0.0308	0.4497	0.3511	0.1725

续上表

片段	$n$	$s_1$	$\sigma_1$	$\sigma_2$	$\sigma_3$	$v_{\max}$	$a_{\max}$	$a_{\min}$
1	69	477.4	5.2738	0.9227	1.6298	17.4	4.5	-6.7
2	369	9640.1	18.5222	0.8711	1.4349	56.5	3.9	-7.2
3	118	2378.3	14.3473	1.0310	1.7826	45.2	4.3	-7.5
4	219	5953.8	22.8238	0.9850	1.5805	55	6.1	-6.6
5	169	6610.9	25.8528	1.3985	1.9898	68.7	8	-7.5
6	220	6560.3	20.0338	1.0176	1.9256	65.5	5.3	-7.7
.....	.....	.....	.....	.....	.....	.....	.....	.....
672	154	4197.2	18.7255	0.9294	1.5079	47.1	5.1	-7.2
673	286	9558.9	17.9162	1.1116	0.8211	56.1	7.2	-4.7
674	487	25106.9	14.9327	0.7819	1.0747	67.2	6.7	-5.3

表 9 文件 2 的运动学片段特征矩阵

片段	$\bar{v}_1$	$\bar{v}_2$	$\bar{a}_1$	$\bar{a}_2$	$t_1$	$t_2$	$t_3$	$t_4$
1	1.9426	7.6947	1.438	-1.3578	0.7466	0.12	0.1266	0.02
2	3.7613	5.7413	1.6153	-1.612	0.3409	0.2954	0.3636	0.0454
3	2.6128	6.86	1.3777	-2.75	0.6153	0.2307	0.1538	0.0512
4	36.724	37.514	1.7615	-1.8070	0.0209	0.4764	0.4450	0.0680
5	51.563	52.558	1.0433	-1.5485	0.0188	0.5660	0.2547	0.1666
6	40.033	43.117	1.2935	-1.5321	0.0714	0.4537	0.3529	0.1302
.....	.....	.....	.....	.....	.....	.....	.....	.....
411	4.8690	5.2060	1.0792	-0.8619	0.0634	0.3380	0.4437	0.1690
412	7.1089	8.9139	1.4857	-1.3813	0.2000	0.3111	0.3556	0.1778
413	27.0116	30.5921	2.4167	-2.4278	0.1163	0.4186	0.4186	0.0930

续上表

片段	$n$	$s_1$	$\sigma_1$	$\sigma_2$	$\sigma_3$	$v_{\max}$	$a_{\max}$	$a_{\min}$
1	150	291.4	3.5341	1.0505	1.4268	11.8	4.5	-4.9
2	44	165.5	4.6277	1.2239	1.2690	13.2	4	-4.9
3	39	101.9	4.1648	0.8540	1.7201	12.2	2.7	-4.2
4	191	7014.3	15.1232	1.5843	1.5522	61.7	12.2	-6.9
5	318	16397.3	20.9777	0.9912	1.7450	73.6	6.7	-6.8
6	238	9528	14.1329	1.3309	2.2028	58.3	7.7	-7.7
.....	.....	.....	.....	.....	.....	.....	.....	.....
411	45	319.9	5.4440	1.2817	1.1255	17.2	4.2	-4.3
412	43	1161.5	15.7867	3.4908	2.1243	42.4	14.6	-6.2
413	181	3030.8	11.9877	0.9311	1.5213	42.4	4.2	-6.1

表 10 文件 3 的运动学片段特征矩阵

片段	$\bar{v}_1$	$\bar{v}_2$	$\bar{a}_1$	$\bar{a}_2$	$t_1$	$t_2$	$t_3$	$t_4$
1	24.9981	30.6909	1.6977	-2.0621	0.1851	0.4166	0.3425	0.0740
2	11.7366	21.3590	1.8469	-1.802	0.4502	0.2565	0.2617	0.0418
3	11.2980	19.9322	2.37096	-3.0461	0.4326	0.2980	0.25	0.0384
4	2.77	9.8612	2.4	-1.6421	0.7181	0.1181	0.1727	0.0090
5	17.053	18.823	1.9222	-1.6281	0.0933	0.36	0.42666	0.1466
6	8.7075	19.3685	2.0636	-1.765	0.55	0.1833	0.25	0.0333
.....	.....	.....	.....	.....	.....	.....	.....	.....
372	5.1296	14.6315	3.125	-2.1	0.6481	0.1481	0.2222	0.0185
373	14.3621	24.8254	1.6227	-1.2379	0.4210	0.2315	0.3052	0.0631
374	31.3570	34.0186	0.95	-0.8013	0.0781	0.375	0.4492	0.1054

续上表

片段	$n$	$s_1$	$\sigma_1$	$\sigma_2$	$\sigma_3$	$v_{\max}$	$a_{\max}$	$a_{\min}$
1	2699.8	2699.8	17.3269	1.2108	1.76057	56.5	5.2	-6.2
2	2241.7	2241.7	12.6886	1.7806	1.6683	42	11.5	-7.2
3	1175	1175	13.2243	1.6737	2.3505	43	7.4	-7.6
4	304.7	304.7	5.5638	2.0932	1.6981	17	6.5	-6.4
5	1279	1279	9.0640	2.2288	1.2131	30.1	11.7	-5.1
6	1044.9	1044.9	11.0212	2.1669	1.9042	30.5	8.3	-7.45
.....	.....	.....	.....	.....	.....	.....	.....	.....
372	54	277	8.4128	3.2521	1.5857	22.4	10.4	-5.2
373	95	1364.4	14.0476	1.5313	1.8370	35.3	6.5	-6.9
374	256	8027.4	13.2246	0.7666	1.1252	45.3	4.5	-7.5

表 11 文件 1、2、3 的总体特征指标

文件	$\bar{v}_1$	$\bar{v}_2$	$\bar{a}_1$	$\bar{a}_2$	$t_1$	$t_2$	$t_3$	$t_4$
1	27.9579	34.6812	1.2163	-1.4743	0.1939	0.3949	0.3149	0.0964
2	26.0086	32.2425	1.4625	-1.5654	0.1933	0.3786	0.3335	0.0945
3	32.8194	41.3960	1.3235	-1.4627	0.2072	0.3757	0.3120	0.1051

续上表

文件	$n$	$s_1$	$\sigma_1$	$\sigma_2$	$\sigma_3$	$v_{\max}$	$a_{\max}$	$a_{\min}$
1	174507	$4.8788 \times 10^{-6}$	23.6214	1.2365	1.5210	109.9	55.7	-7.7
2	127457	$3.3150 \times 10^{-6}$	22.3900	1.8591	1.6244	116.6	48.3	-7.2
3	141304	$4.6375 \times 10^{-6}$	28.4093	1.3672	1.5739	116	19.7	-7.6

片段的特征指标选取的越多，汽车在运动行驶的过程中所传达的信息就越清楚。每一个特征指标都表征着汽车运动工况的某些信息，但是有些特征指标间存在着一定的相关

性，传达的信息存在一定的重叠，这就给分析和解释特征指标上带来麻烦。因此在利用这些特征指标进行汽车行驶工况模型的构建之前，需要对这 16 个特征指标进行因子分析，研究其内部的相关性程度。

## 6.4 汽车行驶工况模型的建立与求解

### 6.4.1 基于改进 k-means 聚类与隐马尔科夫链的行驶工况模型——模型 1

#### (一) 改进 k-means 聚类

改进k-means聚类的过程是把不同的运动学片段的特征集  $J_i = \{J_{i,1}, J_{i,2}, J_{i,3}, \dots, J_{i,15}, J_{i,16}\}$

分类到不同的聚类群中。改进的  $k$  均值聚类是一种非监督而且快速聚类的方法，能够使特征集分类后达到相同组内相似度较高，不同类间的相似度较低的目的。改进的k-means聚类模型的具体实施过程为：

(1) 针对汽车的行驶工况数据特点确定  $k$  个初始聚类中心。[5]

(2) 计算第  $i$  个运动学片段  $J_{ij}$  中的第  $k$  个特征指标到最近聚类中心  $z_j$  的相对距离，确定出各个运动学片段的归属类别。相对距离的计算公式如下式 (6-4-1) 所示：

$$D_n = 1 - \frac{\left( \sum_{k=1}^{16} J_{ik}^2 \right)^2}{\sum_{k=1}^{16} (J_{ik}^2 - J_{jk}^2)} \quad (6-4-1)$$

(3) 重新计算求出同一聚类的特征指标均值，并将这一均值作为新聚类中心点。

(4) 重复迭代步骤 (2) ~ (3)，直至聚类中心收敛为一定值。

按上述步骤，对主要的16个特征指标进行改进的k-means聚类，把汽车行驶特征指标相似的运动学片段聚为同一类区间，可以将所有的运动学片段分为三大类，三类片段样本集的特征参数差异明显，分别代表三类不同的行驶特征。

#### (二) 基于隐马尔科夫链的运动学片段合成

由于汽车在行驶过程中，车速是随城市的交通情况而随机的变化。而马尔科夫链过程也是一个随机过程，可以用来描述一段时间内的汽车工况状态的转移概率情况，是在假设每一个汽车工况状态已经发生的前提下，接下来发生哪个工况状态的概率分布是已知的。

因此可以把汽车的行驶工况过程看作是随机变化的离散马尔科夫过程。

##### (1) 工况状态划分

在 6.3 的数据处理中，完成了对汽车实际行驶过程的 4 种工况状态的标记工作，同时得到了每个运动学片段中，怠速状态、加速状态、匀速状态、减速状态出现的次数。

将每个运动学片段确定为模型事件，记为  $Y_\gamma = \{Y_1, Y_2, \dots, Y_\gamma \mid (\gamma = 1, 2, \dots, \gamma)\}$ ，把运动学特征接近的片段整合为模型事件集，构成马尔科夫链的状态，记作  $K = \{K_1, K_2, \dots, K_\gamma\}$ ， $(\gamma = 1, 2, \dots, k)$ 。

##### (2) 构建状态转移矩阵

汽车当前的工况状态  $K_\gamma$  的概率只与前一个状态  $K_{\gamma-1}$  有关，即：离散的马尔科夫过程在固定的时间间隔下会保持稳定性，并且假设未来时刻的取值只与当前时刻的取值有关，而与过去的时刻无关。[10]



即对于从任意状态  $\gamma-1$  到状态  $\gamma$  满足条件概率:

$$P\{Y_\gamma = k | Y_{\gamma-1} = p\} = P\{Y_{\gamma+\gamma'} = k | Y_{\gamma+\gamma'-1} = p\} \quad (6-4-2)$$

条件概率  $P\{Y_\gamma = k | Y_{\gamma-1} = p\}$  定义为状态  $\gamma-1$  到状态  $\gamma$  的转移概率, 在马尔科夫链过程中, 转移概率的物理意义是当前时间为  $\gamma-1$  时状态  $p$  的模型事件, 转移到下个时间为  $\gamma$  时状态  $k$  的模型事件的概率。转移概率矩阵  $P'$  为

$$P' = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{bmatrix} \quad (6-4-3)$$

对于一个固定的马尔科夫过程, 根据最大似然函数, 得到状态转移概率方程

$$P_{pk} = \frac{N_{pk}}{\sum_k N_{pk}} \quad (6-4-4)$$

其中,  $N_{pk}$  表示时间从  $\gamma-1$  到  $\gamma$ , 状态从  $p$  转移至  $k$  的事件数。

#### 6.4.2 模型 1 的求解

利用 MATLAB 的隐马尔科夫工具箱, 可以得到工况状态的转移概率矩阵, 见表 12 所示:

表 12 工况状态的转移概率矩阵

聚类类别	当前状态	下一状态			
		加速工况	减速工况	匀速工况	怠速工况
聚类 1	加速工况	0	0.456	0.544	0
	减速工况	0.421	0	0.430	0.149
	匀速工况	0.429	0.537	0	0.034
	怠速工况	0.764	0.015	0.221	0
聚类 2	加速工况	0	0.306	0.684	0
	减速工况	0.370	0	0.630	0
	匀速工况	0.541	0.459	0	0
	怠速工况	1	0	0	0

#### 6.4.3 基于多目标 0-1 线性规划的行驶工况模型——模型 2

**模型 1** 是从概率的角度构建的行驶工况模型。现在从**数据**本身的角度构建模型 2——**基于多目标线性规划的行驶工况模型**。由于问题三要求我们构建的汽车行驶工况曲线, 要尽可能地贴近汽车实际行驶工况曲线。而汽车行驶工况模型是基于 16 个特征指标构建的, 从特征指标的角度就是每一个候选运动学片段的 16 个特征要越贴近汽车实际行驶的 16 个特征指标值。基于这一想法, 建立了基于多目标 0-1 线性规划的行驶工况模型, 以下是具体的模型构建过程:

多目标 0-1 线性规划是有两个及两个以上的目标函数, 同时满足目标函数和约束条件均为线性的情况。对于多个候选的运动学片段选取其中的若干个片段, 拼接成汽车行驶工况曲线, 所以对于每一个候选的运动学片段的决策变量只能取 0 或者 1, 而决策变量取值为 1 的这些运动学片段就是最后用来拼接的片段。在建立多目标 0-1 线性规划的行驶工况模型时, 目标函数应当满足实际行驶的特征指标与选取的若干个运动学片段的每个特征指标尽可能的靠近, 16 个特征指标的误差都尽可能趋近 0。

首先，定义汽车实际行驶的特征指标集为  $W = \{W_1, W_2, W_3, \dots, W_{15}, W_{16}\}$ ；每一个运动学片段  $J_i$  的特征指标集为  $J_i = \{J_{i,1}, J_{i,2}, J_{i,3}, \dots, J_{i,15}, J_{i,16}\}$ ；每一个运动学片段  $J_i$  的决策变量为  $\delta_i = \{\delta_{i,1}, \delta_{i,2}, \delta_{i,3}, \dots, \delta_{i,15}, \delta_{i,16}\}$ 。

其次，目标函数表达式满足下式 (6-4-1) 所示：

$$\min \begin{cases} W_1 - \delta_{i,1} \cdot J_{i,1} \\ W_2 - \delta_{i,2} \cdot J_{i,2} \\ W_3 - \delta_{i,3} \cdot J_{i,3} \\ \dots \\ W_{15} - \delta_{i,15} \cdot J_{i,15} \\ W_{16} - \delta_{i,16} \cdot J_{i,16} \end{cases} \quad (6-4-1)$$

约束条件满足下式 (6-4-2) 所示：

$$\begin{cases} W_1^- \leq J_{i,1} \leq W_1^+ \\ W_2^- \leq J_{i,2} \leq W_2^+ \\ \dots \\ W_{16}^- \leq J_{i,16} \leq W_{16}^+ \\ \sum_{i=1}^{16} \delta_{i,j} J_{i,j} \leq W_j \\ \delta_{i,j} = 0 \text{ or } 1 \end{cases} \quad (6-4-2)$$

上述的约束条件必须满足选取的运动学片段的怠速时间之和，小于汽车实际行驶中的怠速时间比，运动学片段的加速时间之和，小于汽车实际行驶中的加速时间比，运动学片段的匀速时间之和，小于汽车实际行驶中的匀速时间比，运动学片段的减速时间之和，小于汽车实际行驶中的减速时间比，决策变量  $\delta_{i,j}$  的取值为 0 或者 1。

#### 6.4.4 模型 2 的求解

对于模型 2 提出的多目标 0-1 线性规划模型的求解，可以采用模糊数学模型，考虑将多目标规划转化为单目标规划进行求解的思路。

对每一个目标  $W_i - \delta_{i,1} \cdot J_{i,1}$ ， $i = 1, 2, \dots, n_i$  给出了每个目标的模糊伸缩指标  $\varepsilon_i (\varepsilon_i > 0)$ 。伸缩指标  $\varepsilon_i$  的选取要根据各个子目标的重要性进行选择，具体的取值原则为：越是重要的目标，其伸缩的指标应越小，根据这一想法，可以实现对每一个子目标进行模糊化处理。[13]

##### (1) 构建模糊目标

对子目标  $W_i - \delta_{i,1} \cdot J_{i,1}$  构造一个模糊目标  $U_i$ ，定义其隶属函数为：

$$\begin{aligned}
U_i(\delta_{i,j}) &= h_i \left( \sum_{j=1}^{n_i} J_{i,j} \cdot \delta_{i,j} \right) \\
&= \begin{cases} 0, & \sum_{j=1}^{n_i} J_{i,j} \cdot \delta_{i,j} < W_i - \delta_{i,j} \cdot J_{i,j} - \varepsilon_i \\ 1 - \frac{1}{\varepsilon_i} \left( W_i - \delta_{i,j} \cdot J_{i,j} - \sum_{j=1}^{n_i} J_{i,j} \cdot \delta_{i,j} \right), & W_i - \delta_{i,j} \cdot J_{i,j} - \varepsilon_i \leq \sum_{j=1}^{n_i} J_{i,j} \cdot \delta_{i,j} < W_i - \delta_{i,j} \cdot J_{i,j} \\ 1, & W_i - \delta_{i,j} \cdot J_{i,j} \leq \sum_{j=1}^{n_i} J_{i,j} \cdot \delta_{i,j} \end{cases}
\end{aligned} \tag{6-4-3}$$

记： $U = \bigcap_{i=1}^{n_i} U_i$ ，则称  $U$  为多目标线性规划问题的模糊目标。  
 $R = \{\delta_{i,j} \mid \delta_{i,j} \cdot J_{i,j} \leq W_j, \delta_{i,j} = 0 \text{ or } 1\}$  是满足约束条件的全体可能解的集合。

## (2) 模糊最优解

取模糊判决为： $M_i = M \cap U$ ，则称满足  $M_i(\delta_{i,j}) = \max_{\delta \geq 0} (M(\delta_{i,j}) \wedge U(\delta_{i,j})) = \max_{\delta_{i,j} \in M} U(\delta_{i,j})$  的  $\delta_{i,j}$  为模糊最优解，同时也是原目标函数的最优解。

所以，求解多目标线性规划的模糊最优解问题可转化为如下形式：

$$\begin{cases} \max W_i - \delta_{i,j} \cdot J_{i,j} = U \\ \sum_{j=1}^{n_i} \delta_{i,j} \cdot J_{i,j} - \varepsilon_i \cdot U \geq W_i - \delta_{i,j} \cdot J_{i,j} - \varepsilon_i, i = 1, 2, \dots, n_i \\ \sum_{j=1}^{n_i} \delta_{i,j} \cdot J_{i,j} \leq W_j, \\ U \geq 0, \delta_{i,j} \geq 0 \end{cases} \tag{6-4-4}$$

则上式 (6-4-4) 被转化为了一个普通的单目标 0-1 线性规划问题，通过 MATLAB 的函数 bintprog 求解 0-1 线性规划问题。

## 6.5 模型求解与检验

### 6.5.1 模型求解

利用 SPSS 软件，设定聚类个数为 3 类，依据每个运动学片段的 16 个特征指标，得到对文件 1 的 674 个片段的分类情况，如下表 13 所示：

表 13 聚类与聚点距离

个案号	聚类	距离	个案号	聚类	距离
1	1	670.151	541	2	1827.72
2	2	3721.51	542	2	1152.651
3	1	1277.569	543	1	699.845
4	2	25.05	544	1	766.761
5	2	732.103	545	2	349.149
6	2	656	546	3	377.03
.....	.....	.....	.....	.....	.....
301	1	549.111	667	1	1209.278
302	1	202.987	668	2	2482.709
303	1	554.004	669	1	1692.217
304	3	1701.145	670	1	1296.809
305	1	996.114	671	2	1555.632
306	1	75.328	672	2	1747.732
307	1	906.982	673	2	3615.657
308	1	78.292	674	3	8592.003

表 14 聚类 1,2,3 的特征指标值

特征指标	1	2	3
$\bar{v}_1$	11.8562	29.2041	41.8439
$\bar{v}_2$	17.7607	34.5528	44.9693
$\bar{a}_1$	1.6897	1.2332	0.9965
$\bar{a}_2$	-1.9198	-1.5738	-1.2532
$t_1$	0.3572	0.1504	0.0743
$t_2$	0.3193	0.4398	0.445
$t_3$	0.2792	0.3323	0.36
$t_4$	0.0769	0.088	0.1259
$n$	90.3119	207.596	400.5
$s_1$	1147.0021	5944.099	16515.3429
$\sigma_1$	9.8988	16.3779	16.9312
$\sigma_2$	1.5725	1.1617	0.9431
$\sigma_3$	1.5344	1.5568	1.3985
$v_{\max}$	28.9337	51.6434	65.4357
$a_{\max}$	6.5882	7.3429	7.3571
$a_{\min}$	-5.4489	-6.6634	-6.9214

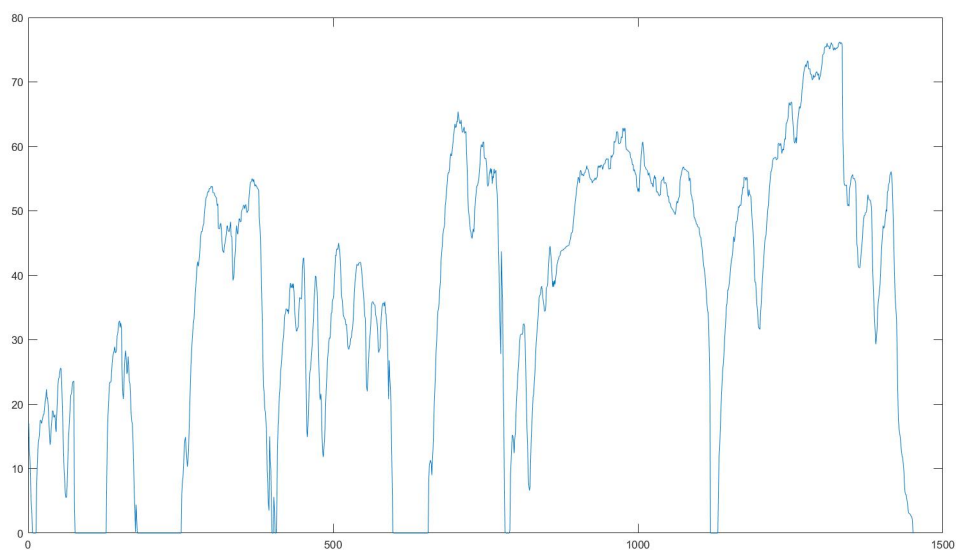
然后，以聚类点为依据，选择出距离每个聚点最近的三个运动学片段，展示如下：

表 15 距离聚点最近的片段

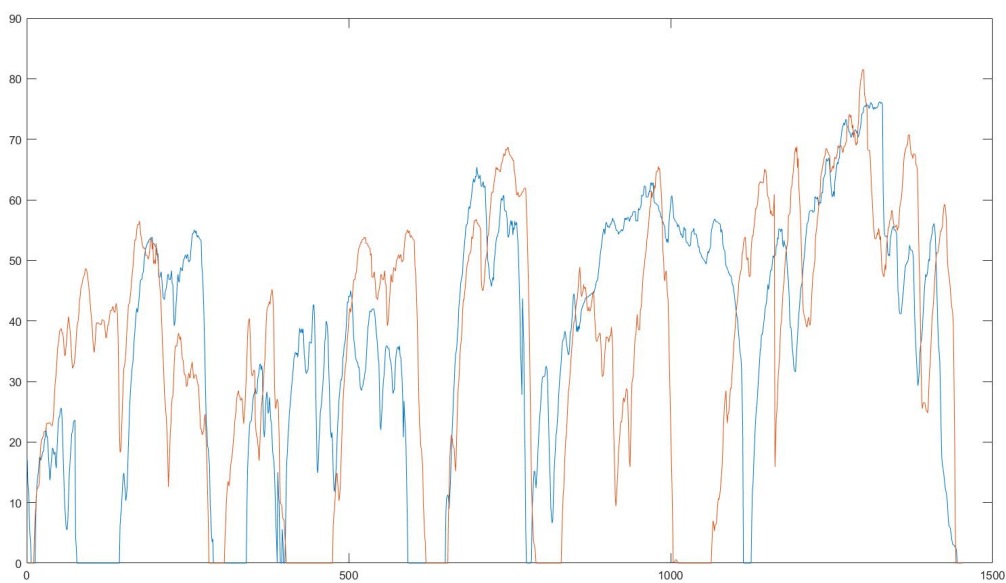
个案号	聚类	距离
211	1	20.064
439	1	22.145
597	1	7.866
4	2	25.05
134	2	58.568
294	2	15.347
72	3	823.172
546	3	377.03
592	3	343.355

这里，我们设定1200秒作为汽车行驶工况曲线的时间区间，按照聚点的片段持续时间这一指标，分配1200秒，聚类1、聚类2、聚类3的样本时间长度比例分别为12.9310%，29.7237%和57.3454%。

然后，根据最大似然估计法，不断进行下一状态筛选，直到合成候选工况达到预定的时间区间长度 1200 秒，下图 10 是得到的汽车行驶工况曲线。



(a) 合成汽车行驶工况曲线



(b) 合成汽车行驶工况曲线与原始数据的对比

图 10 汽车行驶工况曲线

根据隐马尔可夫链模型，选取了编号为 4, 72, 134, 211, 294, 439, 546 的候选片段拼接成图 10 (a)，随机截取一段相同长度的原始信号，两个信号（橙色为）对比可以发现：本文合成的信号除时间延迟滞后或提前以外，能较为完整拟合原始数据片段。

## 6.5.2 模型检验

### (1) 基于特征指标的相对误差检验

下表 17 是选取的候选工况与实际行驶数据的特征指标的相对误差对比，以文件 1 为例，同理，文件 2 和文件 3 也可以由该方法得到。

**表 17 特征的相对误差对比**

特征指标	实际行驶数据	代表工况	相对误差/%
$\bar{v}_1$	27.9579	26.79	4.177%
$\bar{v}_2$	34.6812	35.201	-1.499%
$\bar{a}_1$	1.2163	1.32	-8.526%
$\bar{a}_2$	-1.4743	-1.52	-3.100%
$t_1$	0.1939	0.181	6.653%
$t_2$	0.3949	0.3966	2.892%
$t_3$	0.3149	0.3144	1.915%
$t_4$	0.0964	0.082	5.432%
$n$	174507	174200	0.176%
$s_1$	$4.8788 \times 10^{-6}$	0.00005	-2.484%
$\sigma_1$	23.6214	22.67	4.028%
$\sigma_2$	1.2365	1.31	-8.032%
$\sigma_3$	1.5210	1.49	3.530%
$v_{\max}$	109.9	110	-0.091%
$a_{\max}$	55.7	54.7	1.795%
$a_{\min}$	-7.7	-8.3	-7.792%

根据上表的相对误差数据，可知：汽车实际行驶特征指标与代表的工况特征指标的相对误差绝对值控制在 8.032%。同时，说明了汽车实际行驶状况与片段运动学的 16 个特征指标基本一致，该模型的效果较好。

### (3) 工况模型对比分析

问题三分别建立了基于改进的 k-means 聚类 and 隐马尔科夫链的汽车行驶工况模型和基于多目标 0-1 线性规划的汽车行驶工况模型，得到下表 18 为两种模型的相对误差效果。

表 18 相对误差效果

特征指标	模型 1	模型 2
$\bar{v}_2$	35.201	39.421
$n$	174200	180130
$s_1$	0.00005	0.00028
$\sigma_1$	22.67	26.71
$\sigma_2$	1.31	2.48
$\sigma_3$	1.49	2.04
$v_{\max}$	110	143
$a_{\max}$	54.7	60.54
$a_{\min}$	-8.3	-9.46
$\bar{v}_1$	26.79	29.26
$\bar{a}_1$	1.32	5.49
$\bar{a}_2$	-1.52	-3.87
$t_1$	0.181	0.341
$t_2$	0.3966	0.912
$t_3$	0.3144	0.645
$t_4$	0.082	0.312

根据上表，可以明显地看出模型 2 的 16 个特征指标值较模型 1 的差距较大。故利用改进的 k-means 聚类和隐马尔科夫链的汽车行驶工况模型效果较好；然而，**利用多目标 0-1 线性规划求解汽车行驶工况模型使用较少，是本文提出的一大创新点。**后期可通过先进行主成分分析，减少特征指标个数，再进行多目标 0-1 线性规划模型求解可能效果会有所提高。



## 七、模型评价

### 7.1 模型的优点

(1) 本模型采用了较为创新的方法，从**数据本身和概率论**的角度，分别建立了基于**多目标 0-1 线性规划**求解汽车行驶工况的模型，以及**改进的 k-means 聚类**和**隐马尔科夫链模型**，得到的检验效果较好，并且基于多目标 0-1 线性规划模型，目前使用该模型进行汽车行驶工况研究的实践甚少，是本文提出的一大创新点。

(2) 较为准确地构建了一条误差较小的汽车工况曲线，通过该曲线可判断出汽车在 4 种工况状态下的平均速度，平均加速度等 16 个特征指标，实现了从隐信息（工况分析）表征显信息（汽车车速）的目标。

(3) 在数据预处理上，处理较为完整，将速度高于 120km/h 的记录数进行删除，并且将时间间隔大于 1 秒的记录也进行全部删除，保留了 85% 的原始采集数据，对后续汽车行驶工况曲线的构建起到较好的作用。

### 7.2 模型的不足

(1) 由于时间有限，在建立多目标 0-1 线性规划模型的时候，没有先进行 PCA 分析，对特征指标进行降维，导致目标函数过多，可能会引起求解的误差较大。

(2) 由于问题一数据预处理的过程较为复杂，可能出现数据处理不彻底，导致运动学片段划分出现偏差。

(3) 模型采取了 16 个特征指标进行分析，未考虑到降维和未降维对模型求解产生的影响。

### 7.3 模型的改进

(1) 后期可通过先进行主成分分析，通过降维减少特征指标的个数，再进行多目标 0-1 线性规划模型求解可能效果会有所提高。

(2) 可以从小波分析、SOM 神经网络模型、三次样条插值，**高阶拟线性方程组**（从曲线拟合角度出发，不需要进行运动学片段的划分，通过加速度  $a$  以及车辆工况状态等数据直接对汽车行驶工况曲线进行时间序列的拟合【1】）。

(3) 可以进一步地扩展，根据汽车行驶工况曲线和道路路径划分（由 GPS 根据经纬度解析地址，通过地址划分道路类型）对瞬时油耗进行分析。

## 八、参考文献

- [1]Lin J, Niemeier D A. Exploratory analysis comparing a stochastic driving cycle to California's regulatory cycle[J]. Atmospheric Environment, 2002, 36(38):5759-5770.
- [2]Karande, S., Olson, M., and Saha, B. Development of Representative Vehicle Drive Cycles for Hybrid Applications[J]. SAE Technical Paper 2014-01-1900, 2014, doi:10.4271/2014-01-1900.
- [3]Knez M, Muneer T, Jereb B, et al. The estimation of a driving cycle for Celje and a comparison to other European cities[J]. Sustainable Cities and Society, 2014, 11:56-60.
- [4]Ho, Sze-Hwee, Wong, Yiik-Diew, Chang, Victor Wei-Chung. Developing Singapore Driving Cycle for passenger cars to estimate fuel consumption and vehicular emissions [J]. Atmospheric Environment,2014,97:353-362.
- [5]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 3(2): 164-169.
- [6]郑殿宇, 吴晓刚, 陈汉, 等. 哈尔滨城区乘用车行驶工况的构建[J]. 公路交通科技, 2017, 34(4): 101-107.
- [6]石琴, 郑与波, 姜平. 基于运动学片段的城市道路行驶工况的研究[J]. 汽车工程, 2011 9(3): 256-261.
- [7]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.+
- [8]余曼, 赵轩, 魏朗, 等. 基于 FCM 聚类算法的电动车城市循环工况构建[J]. 公路交通科技, 2018, 35(10): 140-149.
- [9]姜平, 石琴, 陈无畏, 等. 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 18(1): 70-73.
- [10]杜爱民, 步曦, 陈礼璠, 等. 上海市公交车行驶工况的调查和研究[J]. 同濟大學學報(自然科學版), 2006, 34(7): 943-946.
- [11]张璇, 刘鹏. 一种汽车工况构建中的数据预处理方法[J]. 汽车实用技术, 2017 (10): 258-260.
- [12]姜平. 城市混合道路行驶工况的构建研究[D]. 合肥: 合肥工业大学, 2011 24(11):101-108.
- [13]张锐. 城市道路汽车行驶工况的构建与研究[D]. 合肥: 合肥工业大学, 2009 34(4): 1101-1107
- [14]姜平, 石琴, 陈无畏. 聚类和马尔科夫方法结合的城市汽车行驶工况构建 [J]. 中国机械工程, 2010 21(23): 2893-2897.
- [15]郑与波, 石琴, 王世龄. 合肥市汽车行驶工况的研究[J]. 汽车技术, 2010 35(10): 34-39.
- [16]陈弘, 刘海, 乔胜华, 王亚飞. 基于三次样条插值的车辆行驶数据分析. 汽车技术. 2013(8):54-7.弘, 刘海, 乔胜华, 王亚飞. 基于三次样条插值的车辆行驶数据分析. 汽车技术. 2013(8):54-7.
- [17]高英, 郭彦平, 葛渭高. 二阶拟线性微分方程组边值问题的三个对称正解. 系统科学与数学. 2004;24(4):513-519.
- [18]姜平, 石琴, 陈无畏. 基于马尔科夫的城市道路行驶工况构建方法[J]. 农业机械学报, 2009 24(11): 26-30.

## 九、附录

程序 1：读取表格文件

```
1.
2. dataname=strcat(num2str(i),'.xlsx')
3. savename=strcat(num2str(i),'_handle.mat');%设置保存路径
4. data=readtable(dataname);
5. %% 对 xlsx 表格进行预处理
6. data.Properties.VariableNames{1} = 'date';
7. data.Properties.VariableNames{2} = 'v';
8. data.Properties.VariableNames{3} = 'x';
9. data.Properties.VariableNames{4} = 'y';
10. data.Properties.VariableNames{5} = 'z';
11. data.Properties.VariableNames{6} = 'longitude';
12. data.Properties.VariableNames{7} = 'latitude';
13. data.Properties.VariableNames{8} = 'mottoV';
14. data.Properties.VariableNames{9} = 'niuju';
15. data.Properties.VariableNames{10} = 'youhao';
16. data.Properties.VariableNames{11} = 'youmen';
17. data.Properties.VariableNames{12} = 'kongran';
18. data.Properties.VariableNames{13} = 'fuzai';
19. data.Properties.VariableNames{14} = 'jinqi';
20. data.Properties.VariableNames{15} = 's';
21. data = removevars(data, 'x');
22. data = removevars(data, 'y');
23. data = removevars(data, 'z');
```

程序二：数据预处理（处理得到加速度  $a$  和相邻两段时间间隔  $\Delta t$  和标志位  $flag$ ）

```
24.
25. %% 计算时间间隔
26. format long; %日期型时间变量数据相差较小，需要修改数据类型
27. datetime=datenum(data.date(:,1));%将日期转换成数值
28. data.t(1,1)=1;%假设第一段时间间隔为 1
29. data.a(1,1)=0;%假设第一段加速度为 0
30. n=size(datetime,1);%
31. unit_t=1/60/60/24;%计算单位时间的数值差
32.
33. for j=2:n
34.     if data.v(j,1)<10
35.         data.flag(j,1)=1;
36.     else
37.         data.flag(j,1)=0;
38.     end
```

```

39. data.t(j,1)=round((datetime(j,1)-datetime(j-1,1))/unit_t);
40. %计算上下数据前后时间间隔
41. data.a(j,1)=(data.v(j,1)-data.v(j-1,1))/ data.t(j,1);%计算加速度 a=v/s
42. end

```

程序三：异常数据的标记

```

43. n=size(data,1);
44.
45. errNumTaisu=1;
46. errNumJiansu=1;
47. errNumTime=1;
48. errNumTingche=1;
49. errNumV=1;
50.
51. for j = 1:n-180
52.     flag = 0;
53.     %% 超速判断
54.     if data.v(j,1)>=120
55.         del.v(errNumV,1) = j;
56.         errNumV = errNumV+1;
57.         flag = 1;
58.     end
59.     %% 时间缺失判断
60.     if data.t(j,1) ~= 1
61.         del.time(errNumTime,1) = j;
62.         errNumTime = errNumTime+1;
63.         flag = 1;
64.     end
65.     %% 加速度异常
66.     if data.a(j,1)>20 & flag ==0
67.         del.jiansu(errNumJiasu,1) = j;
68.         errNumJiasu = errNumJiasu+1;
69.         flag = 1;
70.     end
71.     end
72.
73.     %% 减速度异常
74.     if data.a(j,1) < a_warning & flag ==0
75.         del.jiansu(errNumJiansu,1) = j;
76.         errNumJiansu = errNumJiansu+1;
77.         flag = 1;
78.     end
79.     %% 长期停车、堵车等长期怠速时间处理

```

```

80.     if data.flag(j,1) == 1 & flag ==0
81.         if data.v(j,1) == 0 &
82.             ( data.mottoV<700 | data.mottoV>1100)
83.             del.tingche(errNumTingche,1) = j;
84.             errNumTingche = errNumTingche +1;
85.             flag = 1;
86.         end
87.     if sum(data.flag(j:j+179,1)==1)==180 & flag ==0
88.         del.taisu(errNumTaisu,1) = j;
89.         errNumTaisu = errNumTaisu+1;
90.     elseif (sum(data.flag(j:j+179,1)==1) == 179
91.             & data.flag(j+179)== 0)& flag ==0
92.         for o=0:178
93.             del.taisu(errNumTaisu,1) = j+o;
94.             errNumTaisu = errNumTaisu+1;
95.         end
96.     end
97. end
98. end
99. end

```

程序四：异常数据的删除，再次计算

```

100.
101. dell=[del.taisu;del.jiansu;del.time];
102. if i==3
103.     dell=[del.taisu;del.jiansu;del.time;del.v];
104. end
105.
106. data(dell,:)=[];%删除数据
107.
108. % data(del.jiasu,:)=[];%删除数据
109. %% 更新
110. format long;
111. datetime=datenum(data.date(:,1));%将日期转换成数值
112. data.t(1,1)=1;%假设第一段时间间隔为 1
113. data.a(1,1)=0;%假设第一段加速度为 0
114. n=size(datetime,1);%
115. unit_t=1/60/60/24;%计算单位时间的数值差
116.
117. for j=2:n
118.     data.t(j,1)=round((datetime(j,1)-datetime(j-1,1))/unit_t);%计算上下数据前后时间间隔
119.     data.a(j,1)=(data.v(j,1)-data.v(j-1,1))/ data.t(j,1);%计算加速度 a=v/s
120.     data.s(j,1)= distance(data.latitude(j,1),data.longitude(j,1),data.latitude(j-1,1),data.longitude(j-1,1))

```

```
)/180*pi*6370*1000;  
121. end
```

#### 程序五：车辆工况状态标记

```
1.  
2. for j=1:n  
3.     if abs(data.a(j,1)) < 0.1 |abs(data.a(j,1)) ==0.1  
4.         if data.v(j,1)>1  
5.             data.flag(j,1)=4;%|a|<=0.1 &v>1  
6.         else  
7.             data.flag(j,1) =1;%|a|<=0.1 &v<1  
8.  
9.         end  
10.    elseif data.a(j,1)>0.1  
11.        data.flag(j,1) =2;%a>0.1&v>1  
12.    elseif data.a(j,1)<-0.1  
13.        data.flag(j,1) =3;%a<-0.1&v>1  
14.  
15.    end  
16.  
17. end
```

#### 程序六：Gps 信号绘制

```
18. <!doctype html>  
19. <html lang="zh-CN">  
20.  
21. <head>  
22.     <!-- 原始地址： //webapi.amap.com/ui/1.0/ui/misc/PathSimplifier/examples/simple-demo.html -->  
23.     <base href="//webapi.amap.com/ui/1.0/ui/misc/PathSimplifier/examples/" />  
24.     <meta charset="utf-8">  
25.     <meta name="viewport" content="initial-scale=1.0, user-scalable=no, width=device-width">  
26.     <title>简单路径</title>  
27.     <style>  
28.     html,  
29.     body,  
30.     #container {  
31.         width: 100%;  
32.         height: 100%;
```

```

33.     margin: 0px;
34.   }
35. </style>
36. </head>
37.
38. <body>
39.   <div id="container"></div>
40.   <script type="text/javascript" src="//webapi.amap.com/maps?v=1.4.15&key=您申请的 key 值
    '></script>
41.   <!-- UI 组件库 1.0 -->
42.   <script src="//webapi.amap.com/ui/1.0/main.js?v=1.0.11"></script>
43.   <script type="text/javascript">
44.     //创建地图（基于高德地图 API）
45.     var map = new AMap.Map('container', {
46.       zoom: 4
47.     });
48.
49.     AMapUI.load(['ui/misc/PathSimplifier', 'lib/$'], function(PathSimplifier, $) {
50.
51.       if (!PathSimplifier.supportCanvas) {
52.         alert('当前环境不支持 Canvas! ');
53.         return;
54.       }
55.
56.       var pathSimplifierIns = new PathSimplifier({
57.         zIndex: 100,
58.         //autoSetFitView:false,
59.         map: map, //所属的地图实例
60.
61.         getPath: function(pathData, pathIndex) {
62.
63.           return pathData.path;
64.         },
65.         getHoverTitle: function(pathData, pathIndex, pointIndex) {
66.
67.           if (pointIndex >= 0) {
68.             //point
69.             return pathData.name + ', 点' + pointIndex + '/' + pathData.path.length;
70.           }
71.
72.           return pathData.name + ', 点数量' + pathData.path.length;
73.         },
74.         renderOptions: {

```

```

75.
76.         renderAllPointsIfNumberBelow: 100 //绘制路线节点,当前节点数量为 100
77.     }
78. });
79.
80.     window.pathSimplifierIns = pathSimplifierIns;
81.
82.     //设置数据
83.     pathSimplifierIns.setData([ {
84.         name: '路线 0',
85.         path: [
86. //绘制路径图所需的坐标
87. //
88. //
89. //     ]
90.     }]);
91.
92.
93.     //对第一条线路（即索引 0）创建一个巡航器
94.     var navg1 = pathSimplifierIns.createPathNavigator(0, {
95.         loop: true, //循环播放
96.         speed: 1000000 //巡航速度，单位千米/小时
97.     });
98.
99.     navg1.start();
100.     });
101. </script>
102. </body>
103.
104. </html>
105.

```

程序七：建立汽车运动特征评估体系

```

106. n=size(split.place,1);
107. b=1;
108.
109. for j=1:n-1
110.     start =split.place(j,1);
111.     finish =split.place(j+1,1);
112.     m=finish-start+1;
113.     v1=0; v2=0;v3=0; numTai=1;numJiasu=1;numYunsu=1;tolv1=1;numJiansu=1;
114.     v=1;a2=1;v2=1;a3=1;v3=1;
115.     if split.flag(j,1)==0

```



```

116.     for o=1:m
117.         tolv1=data.v(o+start-1,1)+tolv1;
118.         v(o)=data.v(o+start-1,1);
119.         if data.flag(o+start-1,1)==1%怠速
120.             numTai=numTai+1;
121.         elseif data.flag(o+start-1,1)==2%加速
122.             a2(numJiasu)=data.a(o+start-1,1);
123.             v2(numJiasu)=data.v(o+start-1,1);
124.             numJiasu=numJiasu+1;
125.         elseif data.flag(o+start-1,1)==3%减速
126.             a3(numJiansu)=data.a(o+start-1,1);
127.             v3(numJiansu)=data.v(o+start-1,1);
128.             numJiansu=numJiansu+1;
129.         elseif data.flag(o+start-1,1)==4%匀速
130.             numYunsu= numYunsu+1;
131.         end
132.
133.     end
134.     chuli.advV1(b,1)=mean(v);
135.     chuli.advV2(b,1)=tolv1/(m-numTai);
136.     chuli.advA2(b,1)=mean( a2);
137.     chuli.advA3(b,1)=mean( a3);
138.     chuli.t1(b,1)=numTai/m
139.     chuli.t2(b,1)=size(a2,2)/m;
140.     chuli.t3(b,1)=size(a3,2)/m;
141.     chuli.t4(b,1)=numYunsu/m;
142.     chuli.t(b,1)=m;
143.     chuli.s1(b,1)=m*chuli.advV1(b,1);
144.     % chuli.s2(b,1)=;
145.     chuli.sd1(b,1)=std(v,0);
146.     chuli.sd2(b,1)=std(a2,0);
147.     chuli.sd3(b,1)=std(a3,0);
148.     chuli.maxv(b,1)=max(v);
149.     chuli.maxa1(b,1)=max(a2);
150.     chuli.maxa2(b,1)=min(a3);
151.     chuli.start(b,1)=start;
152.     chuli.finish(b,1)=finish;
153.
154.     b=b+1;
155.
156.     end
157. end

```

程序八：分割运动片段

```
158. load (dataname);%读取文件
159.
160. n=size(data,1);
161. split.date=zeros(100,1);
162.
163. split.date=datestr(data.date(1,1));%假设分割的第一段的起点为处理后的数据的第一个点
164. split.place=1;
165. flag=0;
166. flag2=0;
167. num=1;
168. split.numSelect=1
169. for j=2:n
170.
171.     if data.v(j,1) == 0 & flag==1
172.         split.date=[split.date;datestr(data.date(j,1))];
173.         split.place=[split.place;j];
174.         flag=0;
175.         num=num+1;
176.         flag2=0;
177.     elseif data.v(j,1) >10
178.         flag =1;
179.     end
180.
181.     if (data.t(j,1)>2 )&flag2==0
182.         split.flag(num,1)=1;
183.         split.numSelect=split.numSelect+1;
184.         flag2 = 1;
185.
186.     elseif (data.t(j,1)==1 )& flag2==0
187.         split.flag(num,1)=0;
188.     end
189.
190.
191. end
192. split.numSelect=sum(split.flag);
```

程序九：隐马尔科夫链处理汽车工况数据

```
1.
2. for ii=1:4
3.     %%%lloyds 和量化公式对特征样本进行量化得到离散码本
```

```

4. for q=20*(ii-1)+1:20*ii
5.   a=y(q,:);
6.   [partition,codebook] = lloyds(a,2^4);
7.   N=length(partition)+1;
8.   for i=1: length(a)
9.     if a(i)>partition(N-1)
10.      c(i)=N;
11.    else if a(i)<=partition
12.      c(i)=1;
13.    else
14.      for j=1:N-2
15.        if partition(j)<a(i)&a(i)<=partition(j+1)
16.          c(i)=j+1;
17.        end
18.      end
19.    end
20.  end
21. end
22. d(q,:)=c;
23. end
24. % initial guess of parameters 初始化概率矩阵
25. prior1 = normalise(rand(Q,1));
26. transmat1 = mk_stochastic(rand(Q,Q));
27. obsmat1 = mk_stochastic(rand(Q,O));
28.
29. % improve guess of parameters using EM12*(j-1)+1):12*j,:; j 用码本进行训练
30. [LL, prior2(:,ii), transmat2(:,ii), obsmat2(:,ii)] = dhmm_em(d((20*(ii-1)+1:20*ii,:),), prior1, trans
    mat1, obsmat1, 'max_iter', 100);
31. figure(1);
32. plot(LL);%最大似然估计值， 即训练曲线
33. hold on
34. end
35.
36. % use model to compute log likelihood
193.
194. y1=[%%%%%%%%%%%%%%输入测试样本 ]
195.
37.
38. for ii=1:28
39.   %%%%%%%%%%%%%%%对测试样本量化
40.   a=y1(ii,:);

```

```

41. [partition,codebook] = lloyds(a,2^4);
42. N=length(partition)+1;
43. for i=1: length(a)
44.     if a(i)>partition(N-1)
45.         c(i)=N;
46.     else if a(i)<=partition
47.         c(i)=1;
48.     else
49.         for j=1:N-2
50.             if partition(j)<a(i)&a(i)<=partition(j+1)
51.                 c(i)=j+1;
52.             end
53.         end
54.     end
55. end
56. end
57. C1(ii,:)=c;%选用测试数据样本
58. end

```