



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 东南大学

参赛队号 19102860085

 1.陈立军

队员姓名 2.李旭涛

 3.于福忠

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

汽车行驶工况构建

摘 要：

本文研究的是汽车行驶工况构建，汽车行驶工况是描述汽车行驶的速度-时间曲线，反映了汽车道路行驶的运动学特征。由于我国国土辽阔，各个城市的交通发展、人口规模、汽车保有量、道路复杂度存在差异性，因此迫切需要基于城市自身的汽车行驶数据进行城市汽车行驶工况的构建研究。本文首先对给出的试验数据进行了预处理，去掉不良数据。紧接着进行了运动学片段的划分，引入九个特征参数，通过主成分分析、K-均值聚类得到低、中、高速三类运动学片段，再根据本文提出的“最小误差分析法”，构建出了汽车行驶工况曲线 1。同时，还利用相关性分析得到汽车行驶工况曲线 2。用综合参数值 CPV 对其进行了合理性评价，得到了最优的汽车行驶工况曲线。

对于问题一，我们首先对原始数据进行了时间变量转换，将原本的年月日时分秒的计时方式直接转变为第 n 秒 ($n=1,2,3,\dots$)。然后是对超速情况不良数据进行处理，用前 1 秒的正常数据替换掉不良数据。利用 MATLAB 进行移动平均滤波去除毛刺。接下来对堵车、断断续续的低速行驶进行状态识别，按照怠速情况处理。对于加、减速度异常的情况，我们对速度进行插值来使得加、减速度满足要求。最后我们对所有文件中的怠速情况进行了超时判断，超过 180s 的时候按照 180s 处理，获得了最终的有效试验数据。三个文件处理后的记录数依次为：178372、138909、157089。

对于问题二，我们利用 MATLAB 编程根据运动学片段的定义，将预处理后三个文件中的有效试验数据依次划分为 1186、880、857 段运动学片段。同时引入了平均速度、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、加速度标准差这九个特征参数用于后续的数据分析，建立了合理的汽车运动特征评估体系。

对于问题三，我们沿用问题二的运动学片段和特征参数。通过主成分分析，我们得到 4 个主成分，这 4 个主成分基本代表了上述 9 个特征参数的大部分信息，并且实现了对特征参数的降维处理。紧接着利用 K-均值聚类将所有的运动学片段分成了 A 类（低速）、B 类（中速）、C 类（高速）三个类别。在本文中我们采用了两种方法来构建汽车行驶工况曲线。第一种是根据我们自己提出的“最小误差分析法”来构建汽车行驶工况曲线。首先计算得到每一个类别运动学片段 9 个特征参数的特征值，然后将每一个运动学片段 9 个特征参数与所属类特征参数的特征值进行误差比较，按照误差从小到大依次排序。计算 A、B、C 类运动学片段对汽车行驶工况曲线的时长贡献，然后从各类运动学片段中依次抽取直到满足各类运动学片段的时间要求，得到了汽车行驶工况曲线 1。第二种是利用相关性分析，大体步骤与方法一类似，只是排序的依据有所不同。相关性分析是用每一个运动学片段 9 个特征参数与所属类特征参数的特征值的相关性大小替代误差进行排序的，计算得到汽车行

驶工况曲线 2。我们采用综合参数值 CPV (Comprehensive Parameter Value) 作为所建工况的合理性评价标准，对两种方法都进行了合理性评价，得到的 CPV 值分别为 0.313 和 1.3873，证明了根据最小误差分析构建的汽车行驶工况曲线 1 更加合理，加上本身就是以最小误差做为出发点考虑的，所以它更能反映实际的汽车行驶状态。

关键词:运动学片段，汽车行驶工况，主成分分析，K-均值聚类，最小误差分析法、综合参数值 CPV。

目录

1.问题重述	4
2.模型假设	6
2.1 题目对模型给出的假设	6
2.2 为简化模型的求解而追加的假设	6
2.3 符号说明	7
3. 问题 1 的建模与求解	8
4. 问题 2 的建模与求解	11
5.问题 3 的建模与求解	15
5.1 主成分分析	16
5.2 K-均值聚类	18
5.3 汽车行驶工况曲线的构建-最小误差分析	19
5.4 汽车行驶工况曲线的构建-相关性分析	21
5.5 计算结果的精度评价	21
6.总结	24
6.1 模型评价	24
6.2 模型改进	24
参考文献	25
附录	26

1. 问题重述

汽车行驶工况（Driving Cycle）又称车辆测试循环，是描述汽车行驶的速度-时间曲线，体现汽车道路行驶的运动学特征，是车辆能耗/排放测试方法和限值标准的基础。目前，欧、美、日等汽车发达国家，均采用适应于各自的汽车行驶工况标准进行车辆性能标定优化和能耗/排放认证。我国在本世纪初直接采用了欧洲的 NEDC 行驶工况（如图 1-1）对汽车产品能耗/排放的认证，取得了显著的成效。但是近年来，随着我国道路交通状况的明显变化以及路上汽车数量的大规模增加，以 NEDC 工况为基准所标定的汽车，实际油耗与法定认证结果的偏差越来越大，NEDC 行驶工况已经不再适用于我国的汽车产品认证。与此同时，欧洲也发现了 NEDC 工况存在的许多问题，改用世界轻型车测试循环（WLTC, 如图 1-2）。但是 WLTC 工况同样与我国汽车的国情不符，所以我国急需制定反映我国实际道路行驶状况的汽车行驶工况^[1]。而且，由于我国国土辽阔，各个城市在人口数量、汽车保有量、交通状况等方面存在着较大的差异，所以应该根据每个城市的实际情况量身定制各个城市自身的汽车行驶工况，使其尽可能地吻合各个城市的汽车行驶情况。

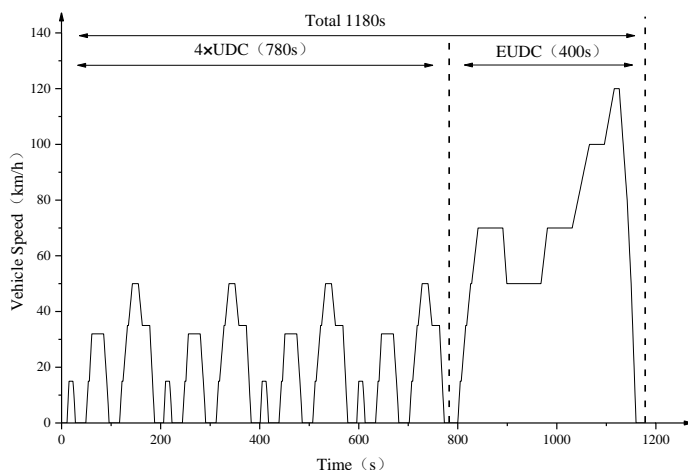


图 1-1 欧洲 NEDC 工况

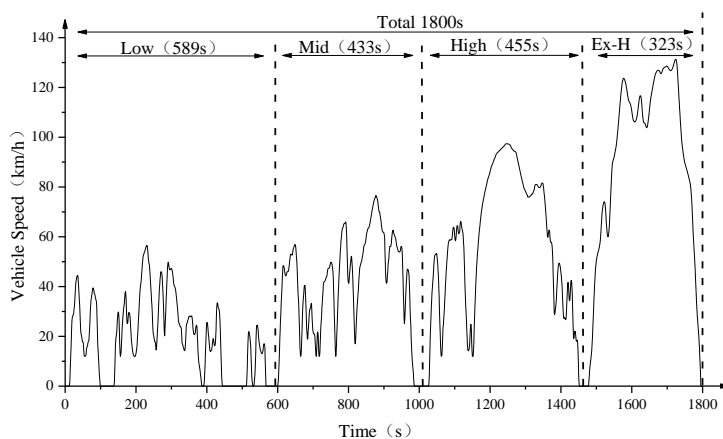


图 1-2. 世界 WLTC 工况

问题 1：数据预处理，本题的附件中给出了测试车辆的汽车行驶数据，但是由于这些原始数据是由采集设备直接记录得到的，所以存在一些不良数据值，一般包括数据缺失、尖点、毛刺等异常值，以及不符合法规与汽车性能的数值等，会对后面的结果产生严重的影响，所以需要针对不同类型的不良数据进行相应的处理，使得数据更加符合实际。

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 1-3 所示。汽车行驶工况的构建往往需要先将原始行驶数据划分成运动学片段再进行求解。所以本问题要求根据运动学片段的定义，将问题 1 中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。

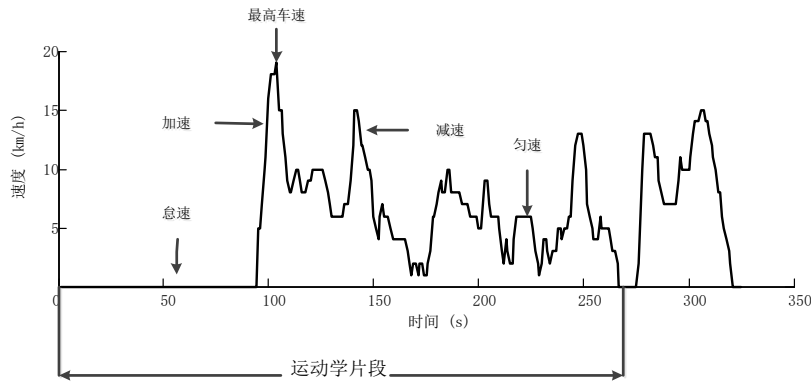


图 1-3 运动学片段的定义

问题 3：汽车行驶工况的构建，结合问题 1、2 中得到的结果，建立完备的数学模型以及合理的汽车运动特征评估体系，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）。按照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，并说明所构建的汽车行驶工况的合理性。

综合分析上述三个问题，给出本题的解题流程图如下所示：

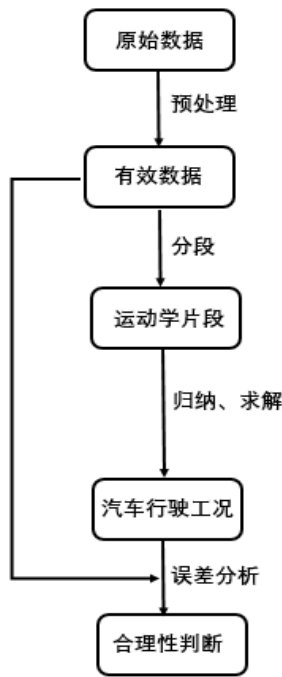


图 1-4 解题流程图

2. 模型假设

2.1 题目对模型给出的假设

1、不良数据主要包括几个类型：

第(1)类，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；

第(2)类，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 $7.5\text{--}8\text{ m/s}^2$ ）；

第(3)类，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。

第(4)类，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。

第(5)类，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。

2、题目对运动学片段的四个要素进行了明确的定义：

怠速：汽车停止运动，但发动机保持最低转速运转的连续过程。

加速：汽车加速度大于 0.1m/s^2 的连续过程。

减速：汽车加速度小于 -0.1m/s^2 的连续过程。

巡航/匀速：汽车加速度的绝对值小于 0.1m/s^2 非怠速的连续过程。

2.2 为简化模型的求解而追加的假设

1、通过分析 X、Y、Z 轴加速度的数值可知，X 轴仅有较小的波动，可认为其是竖直方向上的加速度，而且由于 X 轴加速度数值大部分时间等于 0，因此可以假设测试车辆所经过的路面都是平坦的，没有上坡和下坡，所以 GPS 速度就可以认为是测试车辆的实际行驶速度。

2、测试路面粗糙，不存在车辆轮胎打滑情况，发动机转速与车辆行驶速度成比例。

3、将长期停车状态归为汽车怠速状态。

4、因为国内交通法规定，车辆行驶速度不得超过 120km/h,所以如果速度超过 120km/h 的数据视为异常数据，用前 1 秒的正常数据替换。

2.3 符号说明

序号	符号	名称	单位
1	V	行驶速度	km/h
2	t	行驶时间	s
3	T_i	怠速时间	s
4	T_a	加速时间	s
5	T_d	减速时间	s
6	T_e	匀速时间	s
7	a_a	加速度	m/s^2
8	a_d	减速度	m/s^2
9	V_a	平均速度	km/h
10	V_r	平均行驶速度	km/h
11	a_{aa}	平均加速度	m/s^2
12	a_{ad}	平均减速度	m/s^2
13	T	运动学片段时长	s
14	P_i	怠速时间比	%
15	P_a	加速时间比	%
16	P_d	减速时间比	%
17	P_e	匀速时间比	%
18	V_{std}	速度标准差	km/h
19	a_{std}	加速度标准差	m/s^2

3. 问题 1 的建模与求解

问题 1 要求对原始数据进行预处理，获得有效的行驶数据。由假设可知，不良数据主要分成 5 大类，对于不同类型的不良数据需要进行不同的操作，而且需要考虑操作的先后顺序，所以对数据进行了如下的预处理：

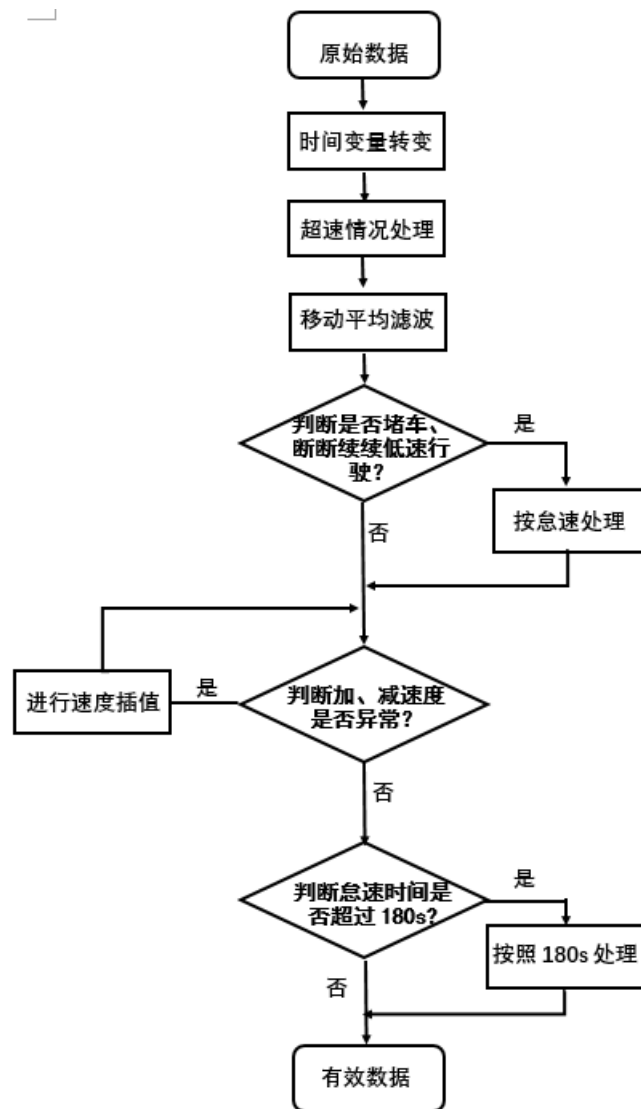


图 3-1 数据预处理流程图

第一步，时间变量转换。我们分析了附件中所给的数据，由于 Excel 中的时间变量是以文本的形式记录的，如果要将时间变量的数据直接导入到 MATLAB 中，过程比较复杂。为了简化计算，我们按照从头开始以秒为单位计时的方式来描述时间变量，例如文件 3 中有 164914 个数据，各个数据对应的时间就是从第 1 秒、第 2 秒……一直到第 164914 秒。这种计时方式使得原始数据在时间上是连续的，解决了第（1）类不良数据，当然这样的简单处理会带来诸如：加、减速度异常，毛刺等问题，但是这些问题可以通过后续处理得到有效的解决。

第二步，超速情况处理。由于我国的交通法规定车辆行驶速度不得超过 120km/h，所

以如果速度超过 120km/h 的数据视为异常数据，用前 1 秒的正常数据替换。按照时间先后，从头开始利用 MATLAB 判断数据中 GPS 速度超过 120km/h 的，用前 1 秒的参数替换。

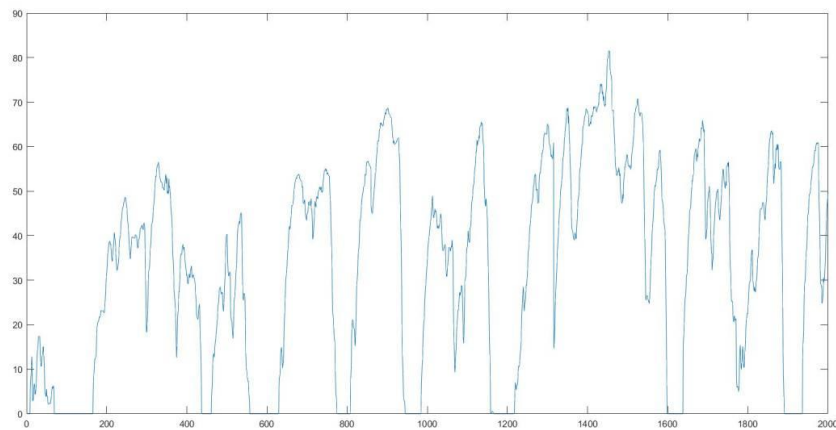
第三步，移动平均滤波。由于从信息采集器获得的数据存在一定的噪声，因此我们使用移动平均滤波器对试验数据进行过滤处理，移动平均滤波器具有原理简单、处理性能优良等优点。移动平均滤波是基于统计规律，将连续的采样数据看成一个长度固定为 N 的队列，在新的一次测量后，上述队列的首数据去掉，其余 $N-1$ 个数据依次前移，并将新的采样数据插入，作为新队列的尾；然后对这个队列进行算术运算，并将其结果做为本次测量的结果^[2]。假设输入为 x ，输出为 y ，则移动平均滤波器的计算公式如下所示：

$$y(n) = \frac{x(n) + x(n-1) + x(n-2) + \dots + x(n-N+1)}{N} \quad (3-1)$$

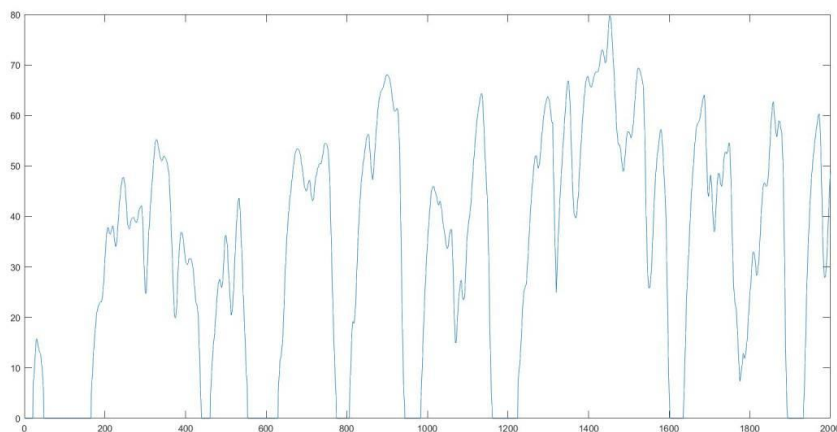
以 $N=4$ 为例说明，

$$y(n) = \frac{x(n) + x(n-1) + x(n-2) + x(n-3)}{4} \quad (3-2)$$

我们将试验数据导入 MATLAB 中，然后利用 MATLAB 实现对试验数据的移动平均滤波，下图是一小段时间内滤波前后的对比图。



(a)



(b)

图 3-2 部分试验数据图 (a)移动平均滤波前 (b)移动平均滤波后

第四步，将堵车、断断续续低速行驶情况按照怠速处理。接下来，我们需要处理第（4）类不良数据，即判断汽车长时间堵车、断断续续低速行驶情况（最高车速小于 10km/h），我们采用了邻域判定的方法：取一个点，以及与它相邻的 n 个点进行速度大小判断，如果车速都小于 10km/h，则认为是长时间堵车或者断断续续低速行驶情况，可按照怠速处理。

第五步，对加、减速异常情况进行速度插值。由于普通轿车一般情况下：0 至 100km/h 的加速时间大于 7 秒，即最大加速度大约为 4m/s^2 ，紧急刹车最大减速度在 $7.5\text{--}8\text{ m/s}^2$ 。第一步对时间变量的处理方式，可能会引入加、减速度异常的情况。通过我们的假设，GPS 速度就等于汽车的行驶速度，汽车的加、减速度可由下式求得（需要注意单位的换算），

$$a_j = \frac{V_{j+1} - V_j}{t_{j+1} - t_j} \times \frac{1000}{3600} = \frac{V_{j+1} - V_j}{3.6} \quad (3-3)$$

式中的 j 表示第 j 秒，如果 $a_j \geq 4\text{m/s}^2$ 或者 $a_j \leq -7.5\text{m/s}^2$ ，则为加、减速度异常的情况，此时我们对速度进行多次插值，直到加、减速度满足实际情况。

第六步，怠速时间处理。我们对怠速时间长度进行判断，如果超过 180s，则将超出部分去掉，即总的怠速时间不超过 180s。

经过以上六步处理，我们获得了有效的试验数据，每个文件中的记录数见表 3-1。

表 3-1 处理后各文件的记录数

文件名	处理后的记录数/个
文件 1	178372
文件 2	138909
文件 3	157089

4. 问题 2 的建模与求解

本题需要将问题 1 中得到的有效试验数据进行运动学片的分类，以及运动学特征参数的求解。首先由题目可知，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。所以，我们利用 MATLAB 编程将有效试验数据划分为多个片段，又因为一个典型的运动学片段应该包括：怠速状态、加速状态、匀速状态、减速状态四个运动状态^[3]。于是我们对划分的片段进一步筛选，将有效试验数据划分为多个运动学片段，给出各数据文件最终得到的运动学片段数量，并计算每个运动学片段的特征参数，建立问题 3 中所需合理的汽车运动特征评估体系。整个提取运动学片段的过程如下：

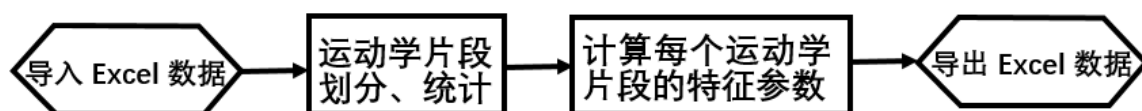


图 4-1 运动学片段的提取流程图

本题在最后给出了运动学片段各个特征参数的名词解释，表 4-1 是我们所考虑的各个特征参数的表达符号以及单位。

表 4-1 运动学片段的特征参数

序号	运动学片的特征参数		
	符号	名称	单位
1	V_a	平均速度	km/h
2	V_r	平均行驶速度	km/h
3	T	运动学片段时长	s
4	T_i	怠速时间	s
5	T_a	加速时间	s
6	T_d	减速时间	s
7	T_e	匀速时间	s
8	a_{aa}	平均加速度	m/s^2
9	a_{ad}	平均减速度	m/s^2
10	P_i	怠速时间比	%
11	P_a	加速时间比	%
12	P_d	减速时间比	%

13	P_e	匀速时间比	%
14	V_{std}	速度标准差	km/h
15	a_{std}	加速度标准差	m/s ²

以下是几个主要特征参数的计算公式^[4]， k 表示该运动学片段所有数据点的个数。

(1) 怠速时间比 P_i

$$P_i = \frac{\text{怠速状态的点数和}}{k} \times 100 \quad (4-1)$$

(2) 加速时间比 P_a

$$P_a = \frac{\text{加速状态的点数和}}{k} \times 100 \quad (4-2)$$

(3) 减速时间比 P_d

$$P_d = \frac{\text{减速状态的点数和}}{k} \times 100 \quad (4-3)$$

(4) 匀速时间比 P_e

$$P_e = 100 - P_i - P_a - P_d \quad (4-4)$$

(5) 平均速度 V_a

$$V_a = \frac{\sum_{j=1}^k V_j}{T}, j=1, 2, 3 \dots k \quad (4-5)$$

(6) 平均行驶速度 V_r

$$V_r = \frac{\sum_{j=1}^k V_j}{T - T_i}, j=1, 2, 3 \dots k \quad (4-6)$$

式中 T_i 表示该运动学片段中的怠速时间。

(7) 速度标准差 V_{std} 的计算公式：

$$V_{std} = \sqrt{\frac{1}{k-1} \sum_{j=1}^k (V_j - V_a)^2}, j=1, 2, \dots k \quad (4-7)$$

(8) 加速度标准差 a_{std} 的计算公式：

$$a_{std} = \sqrt{\frac{1}{k-1} \sum_{j=1}^k a_j^2} \quad (4-8)$$

结合以上公式，我们利用 MATLAB 编程实现了对有效试验数据的运动学片段划分，各文件中的运动学片段数量见表 4-2，具体的运动学片段划分结果见表 4-3 至表 4-5。

表 4-2 各文件中的运动学片段数量

文件名	运动学片段数量/段
文件 1	1186
文件 2	880
文件 3	857

表 4-3 文件 1 运动学片段划分结果

序号	T/s	$V_a/\text{km h}^{-1}$	$V_r/\text{km h}^{-1}$...	$P_i/\%$	$P_a/\%$	$P_d/\%$	$V_{\text{std}}/\text{m s}^{-1}$	$a_{\text{std}}/\text{m s}^{-2}$
1	47	6.4	11.9	...	45.8	18.8	25.0	6.325	0.540
2	391	24.6	35.4	...	30.4	23.2	19.9	18.771	0.166
3	115	20.7	26.2	...	20.7	39.7	30.2	13.650	0.196
4	220	26.9	41.2	...	34.8	27.1	16.7	22.734	0.216
5	170	39.0	48.7	...	19.9	35.7	21.6	25.360	0.196
.									
.									
.									
483	176	23.2	25.2	...	7.9	34.5	32.2	10.972	0.310
484	345	41.2	44.4	...	7.2	33.5	23.4	18.270	0.158
485	277	38.9	39.4	...	1.4	21.9	24.5	15.286	0.247
.									
.									
.									
1183	260	36.0	37.2		3.1	15.3	23.8	16.291	0.277
1184	84	14.5	24.7		41.2	29.4	16.5	14.567	0.242
1185	487	51.4	52.8		2.7	22.1	17.2	14.805	0.167
1186	132	23.1	27.5		15.8	33.8	17.3	12.870	0.219

表 4-4 文件 2 运动学片段划分结果

序号	T/s	$V_a/\text{km h}^{-1}$	$V_r/\text{km h}^{-1}$...	$P_i/\%$	$P_a/\%$	$P_d/\%$	$V_{\text{std}}/\text{m s}^{-1}$	$a_{\text{std}}/\text{m s}^{-2}$
1	636	22.5	31.3	...	28.3	17.7	16.6	15.919	0.234
2	129	0.7	8.7	...	91.5	3.1	2.3	2.470	0.751
3	65	1.8	9.3	...	80.3	10.6	7.6	3.847	0.585
4	38	2.1	8.4	...	74.4	12.8	7.7	3.782	0.672
5	262	30.9	31.6	...	2.3	27.4	20.2	9.504	0.232
.									
.									
.									
470	190	0.5	8.4	...	94.2	2.6	2.1	2.031	0.496
471	219	33.3	34.6	...	3.6	41.8	36.4	15.060	0.202
472	320	40.0	41.9	...	4.7	41.1	30.2	16.687	0.207
.									
.									
.									
877	39	4.7	10.5		55.0	25.0	17.5	5.643	0.475
878	203	18.6	24.5		24.0	27.0	25.0	12.426	0.249
879	23	8.4	10.1		16.7	16.7	37.5	4.214	0.920
880	16	7.1	9.3		23.5	29.4	29.4	3.925	0.785

表 4-5 文件 3 运动学片段划分结果

序号	T/s	$V_a/\text{km h}^{-1}$	$V_r/\text{km h}^{-1}$...	$P_i/\%$	$P_a/\%$	$P_d/\%$	$V_{\text{std}}/\text{m s}^{-1}$	$a_{\text{std}}/\text{m s}^{-2}$
1	780	68.4	75.8	...	9.7	26.9	28.3	29.257	0.156
2	108	24.6	30.2	...	18.3	46.8	22.9	16.983	0.205
3	184	12.0	20.8	...	42.2	24.3	23.2	12.299	0.197
4	185	7.9	27.8	...	71.5	13.4	7.0	14.116	0.186
5	98	11.8	20.4	...	42.4	26.3	26.3	12.476	0.222
.									
.									
.									
389	234	14.4	14.9	...	3.0	12.3	8.1	4.033	0.373
390	134	11.8	16.1	...	26.7	13.3	6.7	7.620	0.335
391	264	4.5	14.1	...	67.9	4.5	3.4	6.774	0.441
.									
.									
.									
854	337	13.6	29.1		53.3	11.8	12.4	15.625	0.208
855	81	21.0	23.6		11.0	34.1	32.9	10.668	0.257
856	193	11.4	25.6		55.7	20.6	10.8	14.558	0.191
857	735	53.0	56.4		6.1	27.6	21.7	24.632	0.136

5. 问题 3 的建模与求解

经过问题 1 和问题 2 的处理后，我们获得了一系列的运动学片段，而且已经建立起了合理的汽车运动特征评估体系。接下来，我们根据这些运动学片段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）。

主要流程见下图：

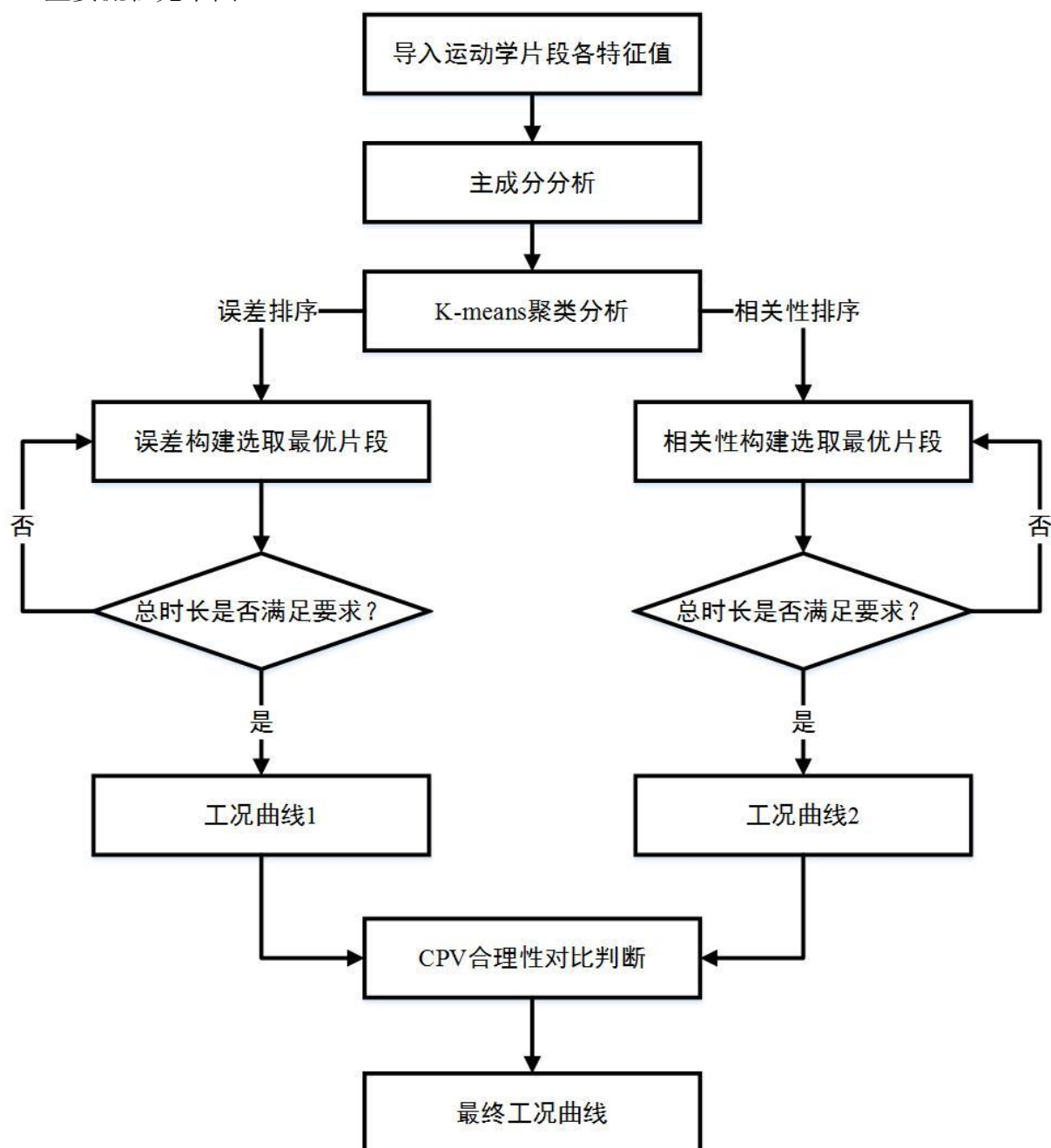


图 5-1 问题 3 的解题流程图

5.1 主成分分析

我们的评价体系包括了平均速度 V_a 、平均行驶速度 V_r 、平均加速度 a_a 、平均减速度 a_{ad} 、怠速时间比 P_i 、加速时间比 P_a 、减速时间比 P_d 、速度标准差 V_{std} 、加速度标准差 a_{std} 这九个特征参数。为了构建出真实反映汽车道路行驶情况的工况曲线，同时降低运算复杂度，我们对带有 9 个特征值的总计 2923 个运动学片段进行了主成分分析，既保留了 9 个特征参数的大部分信息，又起到了降维的作用^[5]。本文使用应用数理统计软件 SPSS 对特征参数进行主成分分析。

将数据导入到 SPSS 中，首先对对各变量进行相关性检验，得到图 5-2 所示的相关矩阵，可知 9 个特征参数之间具有相关性，可以进行主成分分析。

		相关矩阵								
相关		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	标准速度差	加速度标准差
	平均速度	1.000	.939	-.309	-.148	-.724	.511	.445	.747	-.403
	平均行驶速度	.939	1.000	-.284	-.274	-.517	.367	.264	.890	-.453
	平均加速度	-.309	-.284	1.000	-.055	.310	-.333	-.094	-.215	.349
	平均减速度	-.148	-.274	-.055	1.000	.040	-.365	.041	-.453	.299
	怠速时间比	-.724	-.517	.310	.040	1.000	-.787	-.788	-.391	.349
	加速时间比	.511	.367	-.333	-.365	-.787	1.000	.724	.388	-.368
	减速时间比	.445	.264	-.094	.041	-.788	.724	1.000	.239	-.260
	标准速度差	.747	.890	-.215	-.453	-.391	.388	.239	1.000	-.484
	加速度标准差	-.403	-.453	.349	.299	.349	-.368	-.260	-.484	1.000

图 5-2 相关矩阵

运用 SPSS 软件经过迭代后计算出原本特征参数的主成分，得到 9 个影响特征参数的主成分，表 5-1 是 SPSS 软件运算得出的各成分特征值、贡献率以及累计贡献率，一般来说，选择“特征值”大于 1 的成分作为主成分，由于成分 4 的特征值接近 1，为了结果的准确性，所以我们选择前四个成分作为本题的主成分，合计的贡献率为 88.997%，也就是说反应了特征参数 88.997% 的信息，已超过限定的 85%。

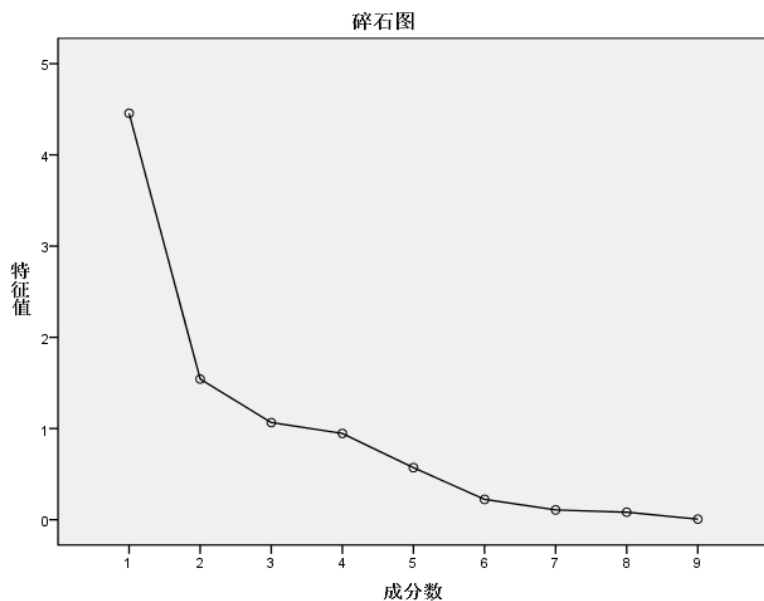


图 5-3 各主成分特征值的碎石图

表 5-1 各成分特征值贡献率以及累积贡献率

成分	初始特征值		
	特征值	贡献率 %	累积贡献率%
1	4.456	49.517	49.517
2	1.542	17.133	66.649
3	1.065	11.837	78.486
4	.946	10.511	88.997
5	.571	6.343	95.339
6	.223	2.480	97.819
7	.108	1.197	99.017
8	.082	.913	99.930
9	.006	.070	100.000

图 5-3 是各个主成分特征值的碎石图，从图中可以看出各主成分特征值的变化趋势。碎石图特征值的大小代表了主成分的方差贡献率大小和重要性程度^[6]。结合这个碎石图可以直观地看出我们所选四个主成分包含了大部分特征参数信息，因此可以将包含信息少的其他成分略去，实现对试验数据的降维。

表 5-2 主成分载荷矩阵

特征参数	成份			
	1	2	3	4
平均速度 V_a	.890	-.105	-.219	.319
平均行驶速度 V_r	.840	-.389	-.215	.294
平均加速度 a_{aa}	-.419	-.068	.591	.604
平均减速度 a_{ad}	-.337	.532	-.675	.276
怠速时间比 P_i	-.825	-.479	-.022	-.085
加速时间比 P_a	.771	.377	.339	-.232
减速时间比 P_d	.636	.650	.222	.111
速度标准差 V_{std}	.788	-.498	-.005	.172
加速度标准差 a_{std}	-.609	.222	.035	.462

表 5-2 给出的是主成分载荷矩阵，主要反映了各主成分与 9 个特征参数的相关性关系。从表中可以明显地看出，第一主成分包括：平均速度 V_a 、平均行驶速度 V_r 、怠速时间比 P_i 、加速时间比 P_a 、减速时间比 P_d 、速度标准差 V_{std} 、加速度标准差 a_{std} ；第二主成分包括：平均速度 V_a 、平均加速度 a_{aa} 、减速时间比 P_d ；第三主成分包括：平均加速度 a_{aa} 、平均减速度 a_{ad} ；第四主成分包括：平均加速度 a_{aa} 。至此已经完成了对运动学片段特征参数的主成分分析，得到结果如下：

表 5-3 各片段主成分

片段序号	主成分 1	主成分 2	主成分 3	主成分 4
1	5.19714560	-2.35020977	-0.55895889	2.27337359
2	18.38084961	-7.23267754	-3.10532504	6.07594680
3	14.41610274	-5.28553113	-2.13749931	4.73919285
4	21.16528975	-8.55515061	-3.53207944	6.91252909
5	26.72018502	-10.00416920	-4.57691647	8.71155036
.
.
.
2921	13.21647893	-4.46531801	-2.12746460	4.48343542
2922	11.67669056	-5.49973717	-1.61232440	3.83793596
2923	31.87382408	-11.00187728	-6.11689422	10.57069260

5.2 K-均值聚类

聚类是将众多的对象按照一定的规则将其聚合为若干类的过程，同一类中的对象相似性较高，不同类之间的相似性较低，通过聚类，可以将众多的运动学片段分为低速 A 类、中速 B 类和高速 C 类分别进行处理，然后组合而成最终的汽车工况曲线。

本文采用的是 K-均值聚类，该方法具有计算速度快、占用内存少等优点，适合于本题对大样本进行聚类的情况^[7]。本文采用 SPSS 软件自带的 K-均值聚类功能对主成分分析得到的结果进行聚类。首先要定好聚类中心，按照上文所述，本次聚类为 3 类，初始的聚类中心如下所示：

表 5-4 初始聚类中心

	聚类		
	A	B	C
主成分1	.43708188	25.13330737	49.94211557
主成分2	-1.49160005	-9.61122888	-15.22797051
主成分3	.30705659	-4.38817056	-11.33648693
主成分 4	2.13666061	8.70868682	17.35272018

最终数据在迭代23次之后达到收敛，初始中心间的最小距离为23.469，最终的聚类中心和每个类的案例数如下所示：

表5-5 最终聚类中心

	聚类		
	A	B	C
主成分1	5.89064754	14.72811154	26.55983217
主成分2	-2.76568969	-5.64857687	-9.27982648
主成分3	-.58436696	-2.17568424	-4.79912571
主成分4	2.37209905	4.93699258	8.90525631

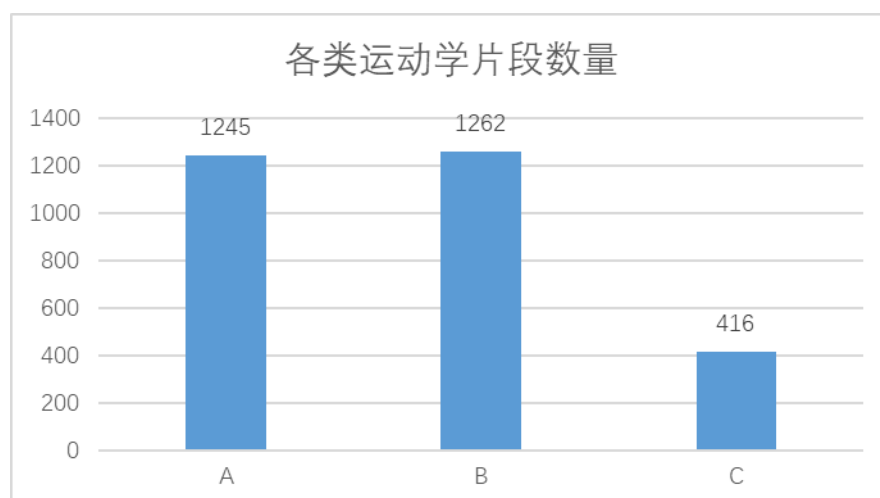


图 5-4 各类运动学片段数量

可以看出，通过 K-均值聚类，2923 个运动学片段按照 4 个主成分特征被分为 3 类，根据之前的主成分分析，我们知道主成分 1 对应的是运动学片段中和速度相关的特征，因此整个运动学片段被合理的分成了低速、中速和高速三个阶段，每个类分别含有 1245、1262、416 个运动学片段。

5.3 汽车行驶工况曲线的构建-最小误差分析

为了使得最终构建的汽车行驶工况曲线尽可能地反映该城市所采集的数据源（处理后的数据），即误差尽可能的小，本文提出了“最小误差分析”的办法来实现。具体处理如下：

- (1) K-均值聚类已经将运动学片段分为 A（低速）、B（中速）和 C（高速）三类。首先，我们利用 MATLAB 编程计算得到每类运动学片段各特征参数的整体特征值 $Y_{mn}(m=A,B,C; n=1,2,3\dots 9, \text{表示 } m \text{ 类运动学片段的第 } n \text{ 个特征参数的整体特征值})$ 。

- (2) 计算各类片段中每一个运动学片段的各特征参数 y_{mn} ($m=A,B,C$; $n=1,2,\dots,9$, 表示 m 类中的某一个片段的第 n 个特征参数值) 与该类片段特征参数整体特征值 Y_{mn} 的误差大小, 在类内按照误差大小将运动学片段以升序的顺序排列排序, 得到三组误差递增的运动学片段集合 A', B', C' 。
- (3) 可根据下式求得 A' 、 B' 、 C' 类运动学片段对汽车工况曲线的时长贡献

$$t_k = \frac{1200 * T_k}{T_{all}} \quad (5-1)$$

式中:

t_k --- k 类 ($k=A', B', C'$) 所运动学片段对汽车工况曲线的时长贡献;

T_k --- k 类运动学片段的总时间;

T_{all} --- 所有运动学片段的总时间。

注: 汽车工况曲线一般为 1200s 至 1300s, 在此处假设为 1200s。

通过此计算公式, 我们可以得到每类数据在总的工况曲线中贡献时间所占的比例如下:

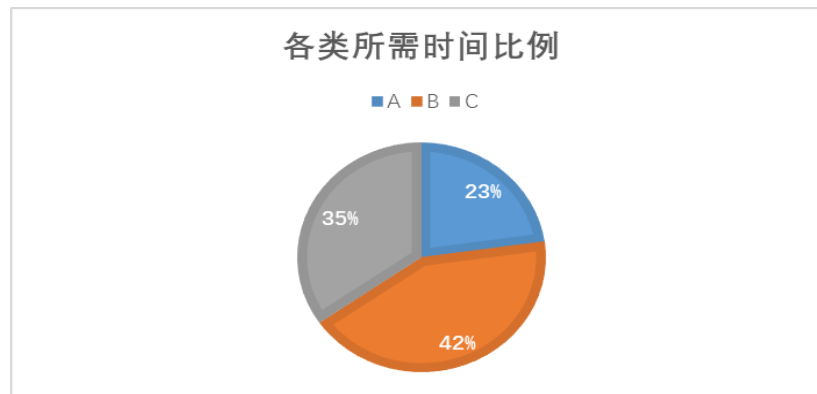


图 5-5 汽车行驶工况曲线中各类运动学片段所贡献的时间比

- (4) 从运动学片段集合 A', B', C' 中依次抽取各类运动学片段直到满足各类运动学片段的时间要求, 得到了与有效试验数据误差最小的汽车行驶工况曲线 1 (1279s), 如图 5-6 所示。

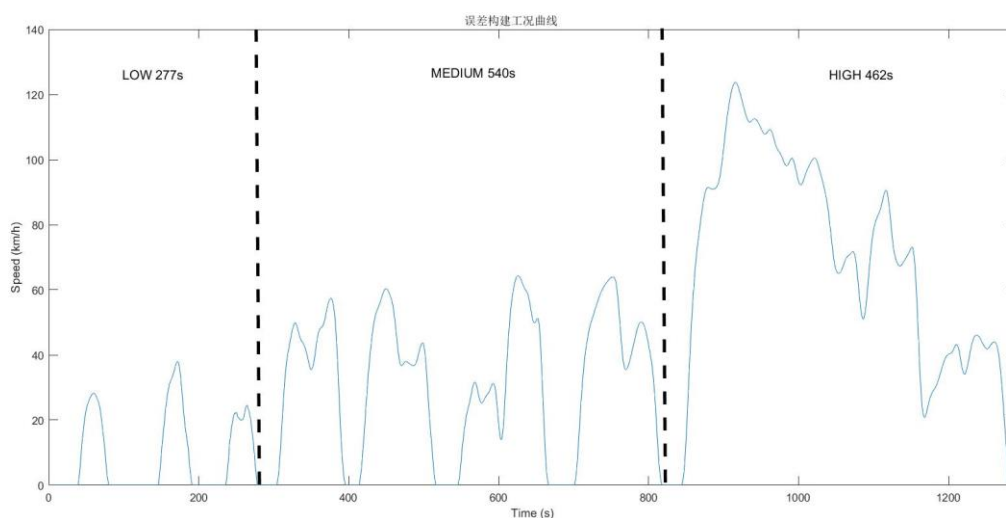


图 5-6 最小误差分析构建的汽车行驶工况曲线 1

5.4 汽车行驶工况曲线的构建-相关性分析

相关性分析是把特征参数的相关性作为依据来构建工况曲线，大体步骤与最小误差分析的方法差不多，只不过在第（2）步的时候，是按照每个片段的特征值与该片段所属类总体的特征值的相关系数大小，将三类运动学片段在各自类内进行降序排序，得到新的三类运动学片段集合 A",B",C"。相关系数的计算公式如下：

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \quad (5-2)$$

式中 $Var(X)$ 和 $Var(Y)$ 分别为向量 X 和向量 Y 的方差， ρ 为向量 X 和向量 Y 的相关系数。

我们通过相关性分析得到的汽车行驶工况曲线 2（1255s），如下所示：

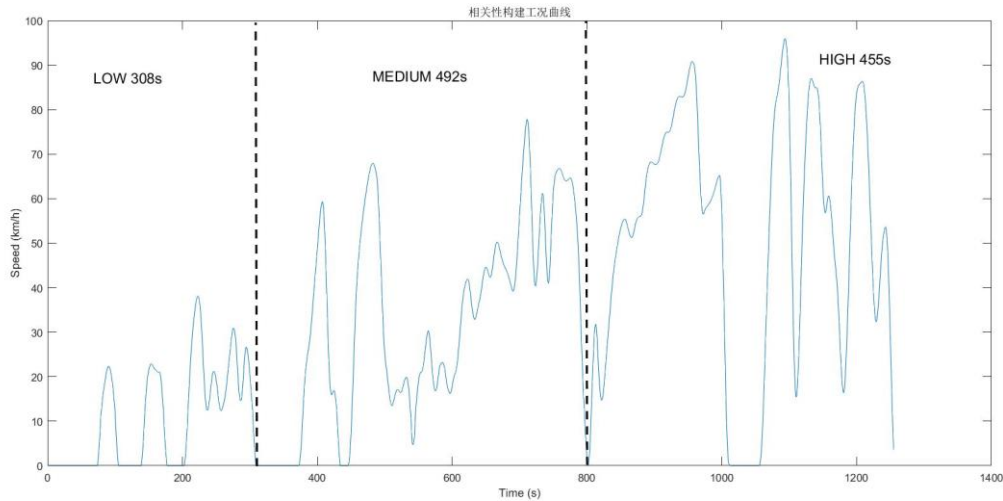


图 5-7 相关性分析构建的汽车行驶工况曲线 2

5.5 计算结果的精度评价

在得到两条工况曲线后，我们需要对其进行合理性判断，选出更符合试验数据的一条工况曲线。为此，我们采用了综合参数值 CPV (Comprehensive Parameter Value)来作为工况曲线合理性评价标准^{[8][9][10]}。综合参数值的计算公式如下：

$$CPV = \frac{|\Delta V_a|}{V_A} + \frac{|\Delta V_r|}{V_R} + \frac{|\Delta a_{aa}|}{A_a} + \frac{|\Delta a_{ad}|}{A_d} + \frac{|\Delta P_i|}{Z_i} + \frac{|\Delta P_a|}{Z_a} + \frac{|\Delta P_d|}{Z_d} + \frac{|\Delta V_{std}|}{V_{STD}} + \frac{|\Delta a_{std}|}{A_{std}} \quad (5-3)$$

式中： V_A 表示原始数据平均速度； V_R 表示原始数据平均行驶速度； A_a 表示原始数据平均加速度； A_d 表示原始数据平均减速度； Z_i 表示原始数据怠速时间比； Z_a 表示原始数据加速时间比； Z_d 表示原始数据减速时间比； V_{STD} 表示原始数据速度标准差； A_{std} 表示原始数据加速度标准差； Δ 表示合成工况与原始数据对应参数的差异值。

由式（5-3）可知， CPV 值是工况与原始数据各代表参数差异率的代数和，所以 CPV 越小，工况与实际行驶状态相似性越高^[1]。我们计算了总数据特征值、根据最小误差分析构建的特征值以及根据相关性构建的特征值，同时计算了两种构建方法的 CPV 值，结果如

下:

表 5-6 特征参数及 CPV 值对比

特征参数	总数据特征值	误差构建特征值	误差构建 CPV	相关性构建特征值	相关性构建 CPV
$V_a/\text{km} \cdot \text{h}^{-1}$	27.99606	28.50035	0.3130	24.06654	1.3873
$V_r/\text{km} \cdot \text{h}^{-1}$	37.91920	38.05002		31.23424	
$a_{aa}/\text{m} \cdot \text{s}^{-2}$	0.346909	0.341189		0.384778	
$a_{ad}/\text{m} \cdot \text{s}^{-2}$	-0.40666	-0.37533		-0.503358	
$P_i/\%$	26.1691	25.0977		22.9482	
$P_a/\%$	27.2766	27.5215		33.6255	
$P_d/\%$	21.6235	24.0031		24.4621	
$V_{\text{std}}/\text{km} \cdot \text{h}^{-1}$	25.40788	24.63628		19.89706	
$a_{\text{std}}/\text{m} \cdot \text{s}^{-2}$	0.268574	0.266518		0.273861	

由表 5-6 可知，使用最小误差分析构建的汽车工况曲线的 CPV 值远小于用相关性构建的，而且很小，证明采用误差分析构建的汽车工况曲线与实际行驶状态相似性很高，该方法具有极高的合理性，本文最终选择的汽车行驶工况曲线是根据最小误差分析构建的汽车行驶工况曲线 1。最终的汽车行驶工况曲线、工况曲线与该城市所采集数据源（经处理后的数据）特征参数值（见表 5-7）如下：

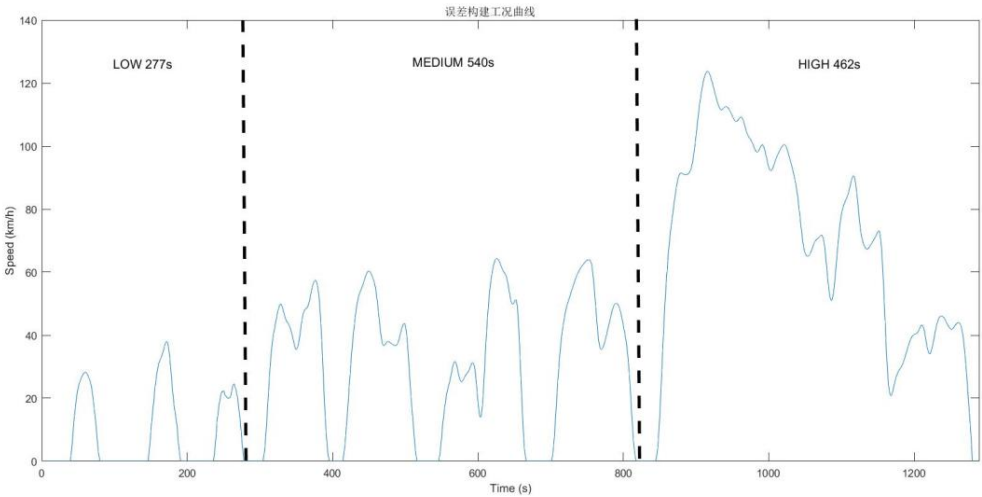


图 5-8 最终的工况曲线

表 5-7 试验数据、汽车行驶工况的特征值

特征参数	总数据特征值	误差构建特征值
$V_a/\text{km} \cdot \text{h}^{-1}$	27.99606	28.50035
$V_r/\text{km} \cdot \text{h}^{-1}$	37.91920	38.05002
$a_{aa}/\text{m} \cdot \text{s}^{-2}$	0.346909	0.341189
$a_{ad}/\text{m} \cdot \text{s}^{-2}$	-0.40666	-0.37533
$P_i/\%$	26.1691	25.0977
$P_a/\%$	27.2766	27.5215
$P_d/\%$	21.6235	24.0031
$V_{std}/\text{km} \cdot \text{h}^{-1}$	25.40788	24.63628
$a_{std}/\text{m} \cdot \text{s}^{-2}$	0.268574	0.266518

6. 总结

6.1 模型评价

本文研究了汽车行驶工况的构建，针对问题一，本文通过条件假设、约束，建立了汽车行驶状况模型。利用 MATLAB 进行数据的预处理，得到的试验数据具有很好的合理性。

针对问题二，本文根据运动学片段的定义，利用 MATLAB 对问题一得到的有效试验数据进行了运动学片段划分，同时引入了 9 个特征参数，建立了合理的汽车运动特征评估体系。

针对问题三，利用问题二中得到的运动学片段以及特征参数，通过主成分分析、K-均值聚类得到了低速、中速、高速三大类的运动学片段。本文提出了“最小误差分析”法来构建汽车行驶工况曲线，同时还通过相关性分析来构建工况曲线，对两个结果进行对比分析，利用综合参数值 CPV 判断两条汽车行驶工况曲线与实际行驶情况的相似度，大大提高了模型的可靠度。

6.2 模型改进

问题一中为了方便计算，时间变量处理得不够精确，对后面得计算产生一定的误差。问题二中还可以考虑更多特征参数，更加准确地描述汽车的行驶状态。问题三中还可以引入别的模型进行对比，使得结果更加精确，可信度更高。

参考文献

- [1] 杨辉宝, 福州市区典型道路汽车行驶工况的研究[D],福州大学, 2017.6.
- [2] 学之之博未若之之要-知之之要未若行之之实, 移动平均滤波的原理---matlab 函数的实现 smooth, <https://blog.csdn.net/datase/article/details/77036571>,2019.9.20.
- [3] 李洋, 基于聚类算法的汽车行驶工况研究[D],北京理工大学, 2016.6.
- [4] 石敏, 轻型汽车行驶工况构建的研究[D],天津理工大学, 2013.12.
- [5] 田宇, 太原市轻型车道路行驶工况构建与分析[D],太原理工大学, 2018.5.
- [6] 常胜, 秦浪, 汤弟伟, 重庆市各区县综合发展水平差异分析——基于 S P S S 主成分分析法和聚类分析法[J], 湖北民族学院学报, 第 35 卷第 1 期, 2017.3.
- [7] 赵丽, 全局 K-均值聚类算法研究与改进[D].西安电子科技大学, 2013.
- [8] Sze-Hwee Ho, Yiik-Diew Wong, Victor Wei-Chung Chang, Developing Singapore Driving Cycle for passenger cars to estimate fuel consumption and vehicular emissions, Atmospheric Environment, Volume 97, November 2014,Pages 353-362.
- [9] Matijaz Knez, Tariq Muneer, Borut Jereb, Kevin Cullinane, The estimation of a driving cycle for Celje and a comparison to other European cities,Sustainable Cities and Society,Volume 11,February 2014,Pages 56-60.
- [10]K.S. Nesamani, K.P. Subramanian, Development of a driving cycle for intra-city buses in Chennai, India, Atmospheric Environment, Volume 45,Issue 31, October 2011,Pages 5469-5476.

附录

%读取数据

```
a1=xlsread('d:\MATLAB\myprojects\shumo\shuju1.xlsx');  
a2=xlsread('d:\MATLAB\myprojects\shumo\shuju2.xlsx');  
a3=xlsread('d:\MATLAB\myprojects\shumo\shuju3.xlsx');  
a=[a1;a2;a3];  
v_initial=a(:,1);
```

%移动均值滤波

```
v_initial=smooth(v_initial,10);  
vp=v_initial;
```

%去除拥堵路段

```
Total=size(v_initial);  
for i=4:Total-3  
  
if(v_initial(i-3)<9&&v_initial(i-2)<9&&v_initial(i-1)<9&&v_initial(i)<9&&v_initial(i+1)<9&&  
&v_initial(i+2)<9&&v_initial(i+3)<9)  
    vp(i)=0;  
end  
end
```

%去除停车时的毛刺

```
for i=1:Total-5  
    if vp(i)==0&&vp(i+1)~=0&&vp(i+2)==0  
        vp(i+1)=0;  
    end  
    if vp(i)==0&&vp(i+1)~=0&&vp(i+2)~=0&&vp(i+3)==0  
        vp(i+1)=0;  
        vp(i+2)=0;  
    end  
    if vp(i)==0&&vp(i+1)~=0&&vp(i+2)~=0&&vp(i+3)~=0&&vp(i+4)==0  
        vp(i+1)=0;  
        vp(i+2)=0;
```

```

        vp(i+3)=0;
    end
    if vp(i)==0&&vp(i+1)~=0&&vp(i+2)==0&&vp(i+3)~=0&&vp(i+4)~=0&&vp(i+5)==0
        vp(i+1)=0;
        vp(i+2)=0;
        vp(i+3)=0;
        vp(i+4)=0;
    end
end
end

```

%对加速度异常时段插值处理

```

a=0;
for j=1:10
    for i=2:Total-1
        if abs(vp(i)-vp(i-1))/3.6>=3||abs(vp(i+1)-vp(i))/3.6>=3
            vp(i)=(vp(i-1)+vp(i+1))/2;
            a=a+1;
        end
    end
end
end

```

%使怠速时间不大于 180s

```

shan=[];
count=0;j=0;
for i=1:Total
    %     if vp(i)>120
    %         vp(i)=vp(i-1);
    %     end
    if(vp(i)==0) count=count+1;
    else count=0;
    end
    if(count<180)
        j=j+1;
    end

    vpp(j)=vp(i);
end

```

%提取运动学片段，并获得片段特征值

```

[t,Total]=size(vpp);
m=0;n=1;

```

```

for i=1:Total-1
    if m==0&&vpp(i)~=0
        break
    elseif m==0&&vpp(i)==0
        m=m+1;n=1;
    end
    duan(m,n)=vpp(i)/3.6;
    pianduan(n)=vpp(i)/3.6;
    n=n+1;

    if(vpp(i+1)==0&&vpp(i)~=0)
        %平均速度
        averagev=sum(pianduan)/n;
        %平均行驶速度
        I=find(pianduan(:)~=0);
        averagev_running=sum(pianduan)/length(I);
        %平均加/减速度 百分比 标准差
        count1=0;count2=0;
        for j=1:n-2
            if pianduan(j+1)-pianduan(j)>0.1
                count1=count1+1;
                aplus(count1)=pianduan(j+1)-pianduan(j);
            elseif pianduan(j+1)-pianduan(j)<-0.1
                count2=count2+1;
                aminus(count2)=pianduan(j+1)-pianduan(j);
            end
        end
        average_aplus=sum(aplus)/count1;
        aplus_percent=count1/n;
        average_aminus=sum(aminus)/count2;
        aminus_percent=count2/n;
        std_aplus=std(aplus);
        %怠速百分比
        daisu=1-length(I)/n;
        %速度标准差
        std_v=std(pianduan);
        %清空
        if aminus_percent==0
            m=m-1;

        else

        duan_te(m,:)=[averagev,averagev_running,average_aplus,average_aminus,daisu,aplus_percent,a

```

```

minus_percent,std_v,std_aplus,n-1];
    shan=[shan,pianduan];
end
    aplus=[];aminus=[];
    n=1;m=m+1;pianduan=[];
end
end

```

%计算各大类特征值

```

lei=xlsread('d:\MATLAB\myprojects\shumo\julei.xls');
leiA=lei(:,1);
leiB=lei(:,2);
leiC=lei(:,3);
ss=max(find(leiA~=0));
leiA=leiA(1:ss);
ss=max(find(leiB~=0));
leiB=leiB(1:ss);
ss=max(find(leiC~=0));
leiC=leiC(1:ss);

```

```

A=[];
for i=1:size(leiA)
    temp=leiA(i);
    ss=max(find(duan(temp,:)~=0));
    A=[A,duan(temp,1:ss)];
end
[t,lenA]=size(A);
%平均速度
averagevA=sum(A)/lenA;
%平均行驶速度
I=find(A~=0);
averagevA_running=sum(A)/length(I);
%平均加/减速度 百分比 标准差
count1=0;count2=0;
for j=1:lenA-1
    if A(j+1)-A(j)>0.1
        count1=count1+1;
        aplusA(count1)=A(j+1)-A(j);
    elseif A(j+1)-A(j)<-0.1&&A(j+1)~=0
        count2=count2+1;
        aminusA(count2)=A(j+1)-A(j);
    end
end

```

```

end
average_aplusA=sum(aplusA)/count1;
aplus_percentA=count1/lenA;
average_aminusA=sum(aminusA)/count2;
aminus_percentA=count2/lenA;
std_aplusA=std(aplusA);
    %怠速百分比
daisuA=1-length(I)/lenA;
    %速度标准差
std_vA=std(A);
lei_te(1,:)=[averagevA,averagevA_running,average_aplusA,average_aminusA,daisuA,aplus_per
centA,aminus_percentA,std_vA,std_aplusA];

```

%绘制工况曲线，并计算其特征值

```

number=[1786,1355,2237,706,2241,1299,2217,1333];    vg=[];
for i=1:8
    temp=number(i);
    ss=max(find(duan(temp,:)~=0));
    vg=[vg,duan(temp,1:ss)];
end
[t,Len]=size(vg);

```

```

% AA=[];
% clear vg aplusvg aminusvg
% AA=xlsread('d:\MATLAB\myprojects\shumo\gongkuang4.xlsx');
% vg=AA(:,2);
% [Len,t]=size(vg);

```

```

%平均速度
averagevg=sum(vg)/Len;
%平均行驶速度
I=find(vg~=0);
averagevg_running=sum(vg)/length(I);
%平均加/减速度 百分比 标准差
count1=0;count2=0;
for j=1:Len-1
    if vg(j+1)-vg(j)>0.1
        count1=count1+1;
        aplusvg(count1)=vg(j+1)-vg(j);
    elseif vg(j+1)-vg(j)<-0.1&&vg(j+1)~=0
        count2=count2+1;
    end
end

```

```
        aminusvg(count2)=vg(j+1)-vg(j);
    end
end
average_aplusvg=sum(aplusvg)/count1;
aplus_percentvg=count1/Len;
average_aminusvg=sum(aminusvg)/count2;
aminus_percentvg=count2/Len;
std_aplusvg=std(aplusvg);
    %怠速百分比
daisuvg=1-length(I)/Len;
    %速度标准差
std_vg=std(vg);
lei_te(5,:)= [averagevg,averagevg_running,average_aplusvg,average_aminusvg,daisuvg,aplus_percentvg,aminus_percentvg,std_vg,std_aplusvg];

wucha=[];
for i=1:9
    wucha(i)=abs((lei_te(5,i)-lei_te(4,i))/lei_te(4,i))*100;
end
sum(wucha)

xlswrite('D:\lei_te.xlsx',lei_te);

vg=[vg,0];
vg=vg*5;
vg=smooth(vg,5);
plot(vg)
xlabel('Time (s)');
ylabel('Speed (km/h)')
```