



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 三峡大学

19110750006

参赛队号

1.杨蕴鹏

队员姓名 2.成孟腾

3.刘莉莉

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：

本文针对汽车行驶工况构建问题，首先通过数据挖掘，采用 LocaSpace Viewer 4 软件平台在谷歌地图中绘制出了每一辆车的行驶路径轨迹图，分析了 3 个数据文件的一致性并采用时间对齐的方式按天对于数据进行分割，通过对于 7 种不同的异常数据进行分析并给出了各自的处理方式，制定了数据剔除、数值插补和数据替换这三种用于数据预处理操作的方式，给出了每一个文件中每一种异常数据的处理方式与处理数量；然后在预处理之后的数据中划分出每一个文件中所包含的运动学片段，划分了汽车的四种行驶模式，构建了 15 个用于描述运动学片段的特征参数以及 12 个用于描述整体分布的特征参数，计算出每一个运动学片段的所有特征参数值；接着通过主成分分析法和聚类分析法，分别构建了低、中、高速的行驶工况并通过组合的方式构建出了整体道路的汽车行驶工况；最后选取了 12 个特征参数对于每一个汽车行驶工况进行了误差分析，结合整个建模过程对于本文所构建的汽车行驶工况做了一个评价分析。

针对问题一，分为三步对于原始数据完成了预处理操作。首先明确数据采集方式并挖掘出数据包含的隐藏内容，得知采集数据的方式是不选定区域及时间段，使用多辆车长时间进行数据采集的方式，数据采集使用的设备为 GPS 导航仪，采集地点为福建省福州市和莆田市，共计有 496464 组原始数据；然后在对数据按照时间对齐的方式进行分割后，制定了 7 种不同类型的异常数据并选择了各自的数据预处理方式；最后采取数据剔除、数值插补和数据替换三种方式对于每一个文件中每一种异常数据进行了预处理，得到的数据预处理结果如下表：

处理类型	原始数据	GPS 设备异常缺失	经纬度异常剔除	长期停车异常剔除	堵车低速片段	堵车低速异常剔除	GPS 短时丢失插值	GPS 长期丢失剔除
文件 1	185725	332412	3	3174	12	6003	110	332302
文件 2	145825	372454	300	402	35	10529	1722	370732

文件 3	164914	266870	0	1765	40	11277	341	266529
	加速度异常剔除	减速度异常剔除	速度异常剔除	毛刺数据替换	数据剔除	数值插补	数据替换	预处理后数据
文件 1	27	2	0	425	9209	110	425	176516
文件 2	332	82	0	1077	11645	1722	1077	134180
文件 3	115	39	313	664	13509	341	664	151405

针对问题二，首先对问题一预处理之后的数据进行了运动学片段的提取，得出的结果如下表所示；之后将汽车的行驶模式划分为怠速模式、加速模式、减速模式和匀速模式；然后构建了运行时间等 **15 个用于描述运动学片段的特征参数**以及怠速时间比等 **12 个用于描述整体分布的特征参数**；紧接着计算出每一个运动学片段的所有特征参数值；最后采用统计学分析方法对于这些特征参数进行了分布分析。

	文件 1	文件 2	文件 3	
运动学片段	830	613	547	1990

针对问题三，首先采用主成分分析法对于问题二中的 15 个运动段特征参数进行降维处理得到 5 个主成分；接着使用这 5 个主成分为依据，采用 k-means 聚类分析法将 1990 个运动学片段划分成 3 大类；然后对于每一类的道路构建出各自的行驶工况；之后将这三个工况提取出部分合称为一个整体道路的汽车行驶工况,其中各个路段行驶工况参见图 6-7 至图 6-9；最后分别使用特征指标和速度—加速度联合概率分布来对于所构建的工况进行评价，整体路段的汽车行驶工况参见图 6-10，分析出本题所构建的工况对于中、低速路段的效果比较好。

关键词：行驶工况 数据挖掘 主成分分析 k-means 聚类分析 速度—加速度联合分布

一、问题重述

1.1 问题背景

汽车工况曲线不仅是测定或者评价机动车污染物排放量和燃油消耗量的重要依据^[1]，也还对于制定汽车的节能减排方案具有指导性的作用，所以其重要性不言而喻。

在本世纪初，我国是直接采用欧洲的 NEDC 实行工况对汽车产品能耗/排放的认证，有效促进了汽车节能减排和技术的发展。但是近些年来，国人发现 NEDC 标准存在着诸多的问题，以 NEDC 工况为基准所优化标定的汽车，实际油耗与法规认证结果偏差越来越大。此外由于我国地域辽广，各个城市的发展程度、气候条件及交通状况的不同，使得各个城市的汽车行驶工况特征存在明显的不同。

因此，制定出能够反映我国实际道路行驶状况的测试工况，显得越来越重要。

1.2 需要解决的问题

围绕着汽车工况构建问题，需要通过建立数学模型解决以下几个问题：

问题一：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，请设计出合理有效的方法对原始数据个预处理，并给出各个文件数据经过数据预处理之后的记录数。

问题二：请设计合理的方法，将问题一中预处理之后的数据划分为多个运动学片段，并给出个数据文件最终得到的运动学片段数量。

问题三：请根据上述经处理后的数据，选择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），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。此外，按照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，并说明所构建的汽车行驶工况的合理性。

二、模型假设

为了简化问题，可以设定一系列的理想条件，如下：

假设 1： 文件中的数据真实有效 ；

假设 2： 连续 3 秒中的第 2 秒数据缺失，是由 GPS 瞬时异常造成的，可以修复 ；

假设 3： 长期停车持续时间超过 180s 视为异常数据 ；

假设 4： 速度连续不等于 0 的时长小于 4s 为环境干扰造成的波动数据，可以抹零；

假设 5： 堵车、长期低速运行（速度小于 10km/h），持续 180s 以上做怠速处理 ；

三、符号说明

为了便于阅读方便，所有符号在正文中首次出现时均做了阐述说明，在此仅列出部分符号说明以供参考。

符号	符号说明
N	原始数据的总计个数
a_i	第 i 个原始数据的加速度，单位为
T	运行时间
T_i	怠速时间
T_a	加速时间
T_d	减速时间
T_c	匀速时间
v_m	平均速度
v	平均行驶速度
v_{\max}	最大速度
v_{sd}	速度标准差
a_a	平均加速度
$a_a(\max)$	最大加速度
$a_a(sd)$	加速度标准差
a_d	平均减速度
$a_d(\min)$	最小减速度
$a_d(sd)$	减速度标准差

四、问题一：数据预处理

4.1 问题分析

对于问题一而言，题目所给的原始数据中存在着诸多的不良数据值，如果直接对所给的数据进行使用的话会严重影响到后续的数据挖掘部分的执行效率，降低整体模型的可信度，甚至可能会导致最后得出错误的结论，所以在进行数据挖掘与建模之前对于原始数据进行预处理操作就显得尤为重要。

通过查阅工况构建研究流程相关资料，得知工况构建的目的是以某地域大量的行驶数据作为依据，构建能够反应当地车辆运行工况的“速度-时间”曲线。通常来说，工况曲线一般都需要划分为几个不同的工况运动段，这些运动段采集并整理自原始的车辆工况曲线。通常工况曲线构建的技术路线如下图所示：

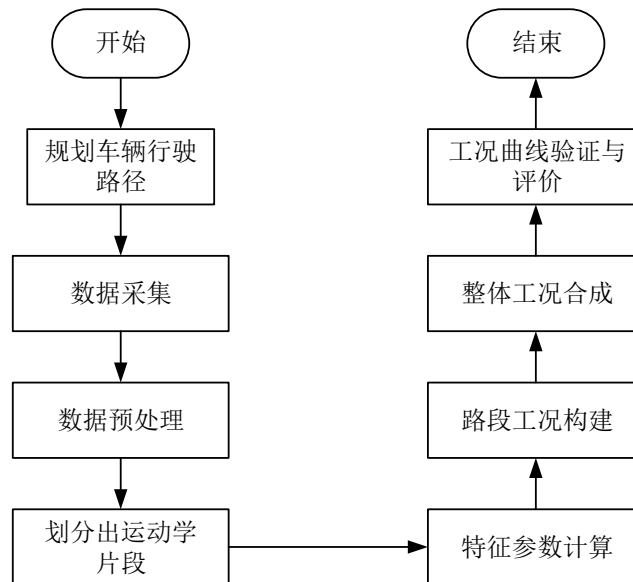


图 4-1 汽车工况构建的流程图

从上图中可以看出，构建汽车工况基本步骤为：根据区域内的交通流量规划车辆行驶路径－数据采集－数据预处理－短行程划分－特征值参数计算－路段工况构建－整体工况合成－工况曲线验证和评价。对于本题目来说，路径规划、数据采集工作已经完成，后续工作需要我们来处理，因此我们在第一问中进行数据的预处理流程。

题目中给出了不良数据主要的 5 种类型，所以我们基于此并对其进行拓展，制定了一套较为完善的数据预处理流程，如下图所示。对于数据的预处理，我们首先明确了本次数据处理的目标以及预期的效果，接着通过附件所给出的数据建立起数据集，之后对于每一个原始数据表（共计 3 个）建立样本数据子集并通过经纬度在地图中绘制出车辆行驶的路径；然后对于数据的质量做进一步的分析并对于缺失值与异常值数据进行清洗，之后对于数据进行属性规约，最后对于处理过后的数据进行特征分析。

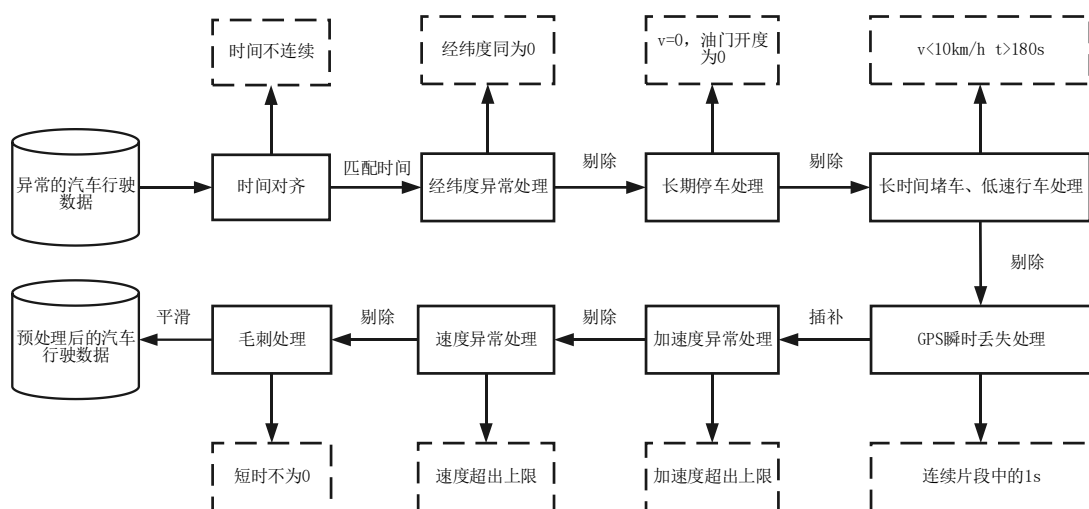


图 4-2 数据预处理流程图

4.2 明确数据采集方式

根据题目所给的三个数据文件，由于每一个原始数据表中均有经纬度的数据，因此我们可以地图软件平台定位出车辆在各个时间节点所处的位置，并将这些时间节点的位置连接起来，由此便可以得到每一辆车（共三辆）原始数据的行驶轨迹。

我们采用的地图绘制软件为 **LocaSpace Viewer 4**，该软件采用了谷歌地图作为基础。此外该软件提供了“**Excel 转 Kml**”功能，能够将包含有经纬度数据的表格区域转换成地图中的经纬度坐标，并添加到地图上。



图 4-3 图新地图 excel 转 kml 操作界面

将原始数据的文件 1 导入到 LocaSpace Viewer 4 中，并提取经纬度数据，将其转换为 kml 格式的路径文件。通过该软件解析经纬度，并将每一个点在谷歌地形图中进行标注，之后将每一个点连接起来便绘制出车辆采用原始数据的整个轨迹图，可以直观感受到数采集的区域，以及行车路径变化情况，如下所示图中很长的红色线条即为异常经纬度数据。



图 4-4 数据预处理前的文件 1 行车轨迹

从上图中我们可以明显看出，行驶轨迹在地球表面横跨大陆海洋，这说明在采集到的经纬度数据中，存在异常数据。继续放大局部后，可以看到，主要的行车区域为福建省马尾区附近。

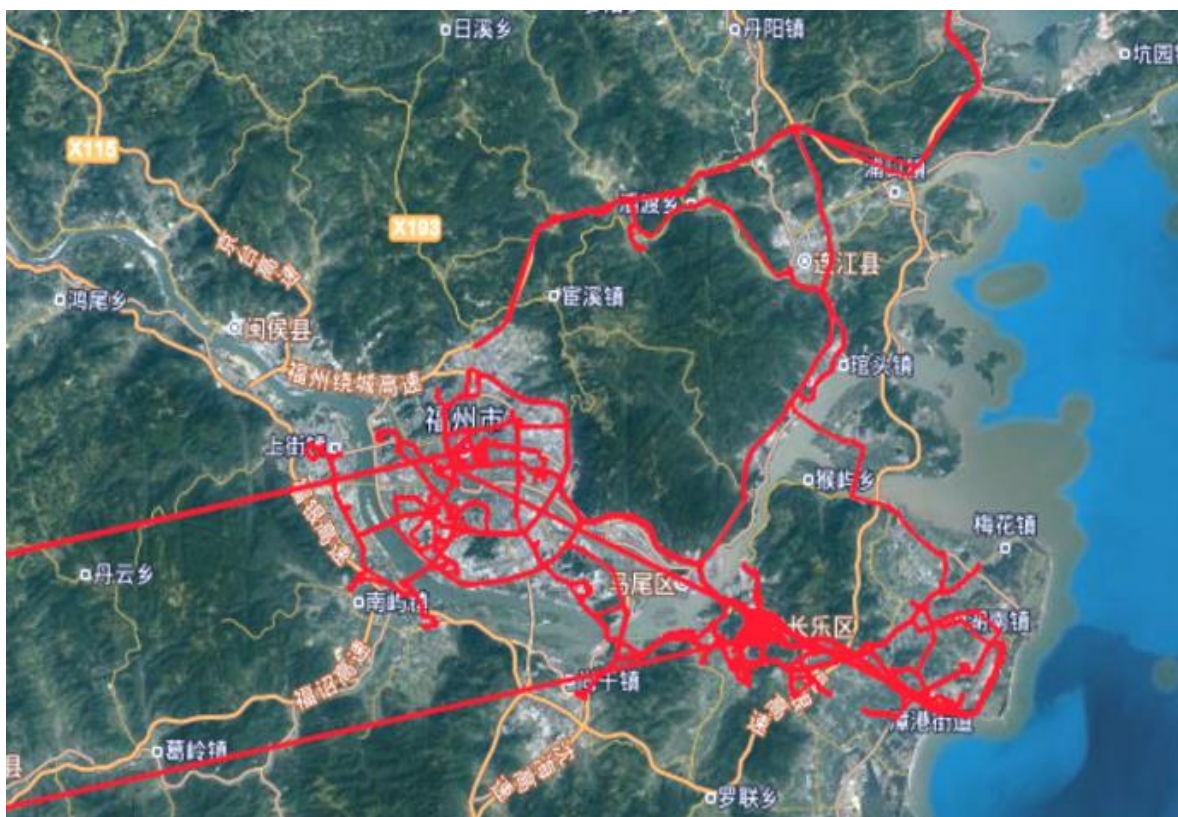


图 4-5 文件 1 行车轨迹放大图

同理，可以得到文件 2 和文件 3 在数据预处理之前的行车轨迹图，将所得的所有行车轨迹放到一起对比可以发现，文件 1、2 中存在的经纬度为 0 的情况，所以我们推断这可能是由于设备出现异常或者设备停止工作导致的异常数据，在后续的数据预处理中，应当

将这些异常的坏数据全部剔除掉。

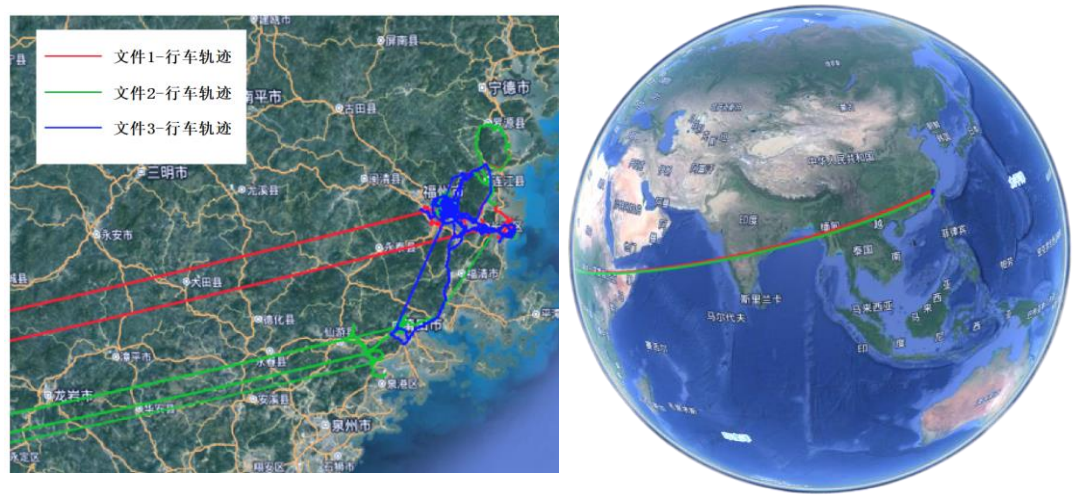


图 4-6 数据预处理前汽车行车轨迹

通过三辆汽车在数据预处理之前的行车轨迹图，我们发现数据采集的地点为福建省，并且测试的地点多为福州市与莆田市。初步确定样本数据的质量较差，样本中存在大量的不良数据，所以应当对其进行后续的预处理。

4.3 明确数据采集内容

每个文件均包含了时间、GPS 车速、X 轴加速度、Y 轴加速度、Z 轴加速度、经度、纬度、发动机转速、扭矩百分比、瞬时油耗、油门踏板开度、空燃比、发动机负荷百分比、进气流量等 14 个数据字段。其中 GPS 车速、经纬度应该由 GPS 采集设备等信号类型的采集设备采集得到；XYZ 三轴加速度比较小，并且没有具体单位，因此暂时先不考虑；剩下的字段全部是关于车辆行驶过程中的各种运行参数，这些数据相互之间的组合能够反应出一定的工况特征，需要仔细研究。对于时间这个字段最为特殊，题目明确，当 GPS 数据丢失时，将会造成数据时间不连续，通过比对数据采集起止时间和数据长度，我们也发现了存在大量时间不连续的数据丢失的情况，因此在做数据预处理时需要仔细考察。

时间	GPS车速	X轴加速度	Y轴加速度	Z轴加速度	经度	纬度	发动机转速	扭矩百分比	瞬时油耗	油门踏板开度	空燃比	发动机负荷百分比	进气流量
2017/12/18 13:42:13.000	0	0	-0.396	-0.9	119.367781	25.99242	775	18	58.02	0	0.1465	22	2.3
2017/12/18 13:42:14.000	0	0	-0.378	-0.882	119.367783	25.992418	775	17	60.3	0	0.1465	21	2.39
2017/12/18 13:42:15.000	0	0	-0.396	-0.882	119.367785	25.992416	775	17	55.24	0	0.1464	22	2.19
2017/12/18 13:42:16.000	0	0	-0.378	-0.9	119.367785	25.992416	762	17	55.75	0	0.1471	21	2.21
2017/12/18 13:42:17.000	0	0	-0.396	-0.882	119.367785	25.992416	762	16	56	0	0.1471	21	2.22

图 4-7 文件数据字段内容

4.4 数据质量分析

经过上述步骤，我们初步发现在原始数据中存在着诸多的不良数据，数据的整体质量较差，所以我们在进行数据的清洗之前应当对原始数据的质量进行进一步的分析，从而确保预处理之后的数据具有较高的价值，数据质量分析的具体指标如下图 5-10 所示。

通过对于数据与题目的分析，我们得知采集数据的方式是不选定区域及时间段，使用多辆车长时间进行数据采集的方式；本题中使用的采集设备为 GPS 导航仪；本题中所给数

据的采集地点为福州市与莆田市；数据的采集频率为 1HZ，即每间隔 1 秒就会采集一组数据；最终得到共计有 496464 组数据。

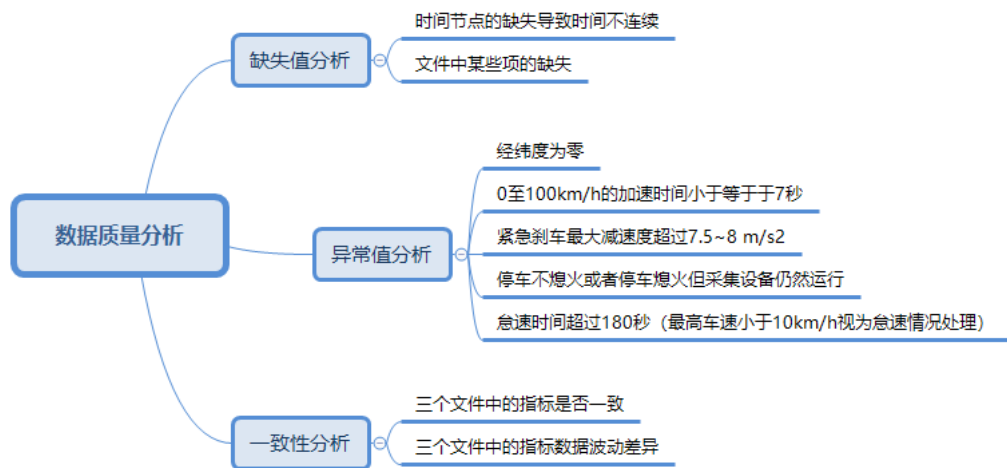


图 4-8 数据质量分析

4. 4. 1. 对于数据文件的一致性分析

通过对比三个文件的数据情况，可以发现每一个文件中的指标项目是完全相同的，所以不存在指标不一致的情况。但是各个数据文件中各项指标是否一致需要通过对于数据进行分析才能下定论，在本问中采用指标的算数平均值的差异性来判断三个文件中的数据指标是否一致。

如下表所示，可以看出三个文件差异性较大的是 GPS 车速、X 轴加速度、Z 轴加速度这三项指标，而其他指标差异性较小可以视为具有一致性。对于 GPS 车速，结合 5.3 的车辆行驶路径，三辆车存在差异可能是因为它们进行采集数据时各自侧重的道路类型不一样导致的。而对于 X 轴加速度、Z 轴加速度这两项，文件 1 和 2 的平均值均为负，而文件 3 的平均值确均为正，初步推测可能是由于三轴加速度传感器的测量方向与 1、2 不一致，在使用者两类数据的时候应当慎重考虑。在本文中，由于已知时间与速度，所以通过属性规约直接采用加速度的定义式来构造加速度，并将其用于后续的分析中，并未直接使用 X、Y、Z 轴加速度的数据。

表 4-1 原始数据指标平均值比较表

文件（平 均值）	GPS 车 速	X 轴加 速度	Y 轴加 速度	Z 轴加 速度	经度	纬度	发动机转 速
文件 1	26.6514	-0.0023	-0.3873	-0.8914	119.4749	25.9883	1196.5486
文件 2	23.7699	-0.0035	-0.4620	-0.8531	118.9657	25.8061	1146.8841
文件 3	30.1536	0.2661	-0.9509	0.3351	119.3810	25.9932	1233.4259
文件（平 均值）	扭矩百 分比	瞬时油 耗	油门踏 板开度	空燃比	发动机负 荷百分比	进气流 量	
文件 1	25.4762	13.8157	0.0466	0.1446	31.1415	5.4899	
文件 2	25.4477	12.7151	0.0420	0.1443	31.2386	5.2539	

其中加速度的计算公式如下：

$$a_i = \frac{v_i - v_{i-1}}{3.6} m/s^2, i=1, 2, \dots, N$$

式中： N ——原始数据的总计个数；

a_i ——第 i 个原始数据的加速度，单位为 m/s^2 ；

v_i ——第 i 个原始数据的 GPS 车速，规定 $v_0=0km/h$ ；

4. 4. 2. 对于时间不连续问题

由前文分析可以得知，本题中的 GPS 设备的采集频率为 1Hz，那么在设备正常工作的情况下，采集到的数据必然是在时间轴上连续且递增的时间序列数据。而由于设备运行过程中因为内部或者外部的因素导致的设备异常或者信号屏蔽干扰，都将会造成时间上的不连续，即 GPS 信号丢失。但是 GPS 信号丢失造成的时间不连续问题需要分情况讨论，因为作为 GPS 信号采集设备，首先需要供电。若驾驶员在行车时，有停车休息、吃饭、休息区加油等操作时，必然会导致 GPS 信号采集设备断电，致使信号丢失，这种情况一般持续较长时间；此外，由于高层建筑覆盖或过隧道等造成信号屏蔽，也会丢失信号，但这种情况一般持续时间不长。

因此在数据预处理之前应当对于数据进行分割，即按照时间对齐的方式对于整个文件划分出多个片段，其中时间对齐是指将数据按照时间递增方向，匹配到连续的时间轴上，使之成为以一个在时间上完整的数据。以文件 1 作为示例，将文件的数据经过时间对齐操作后，画出其“速度-时间”曲线如下图所示。

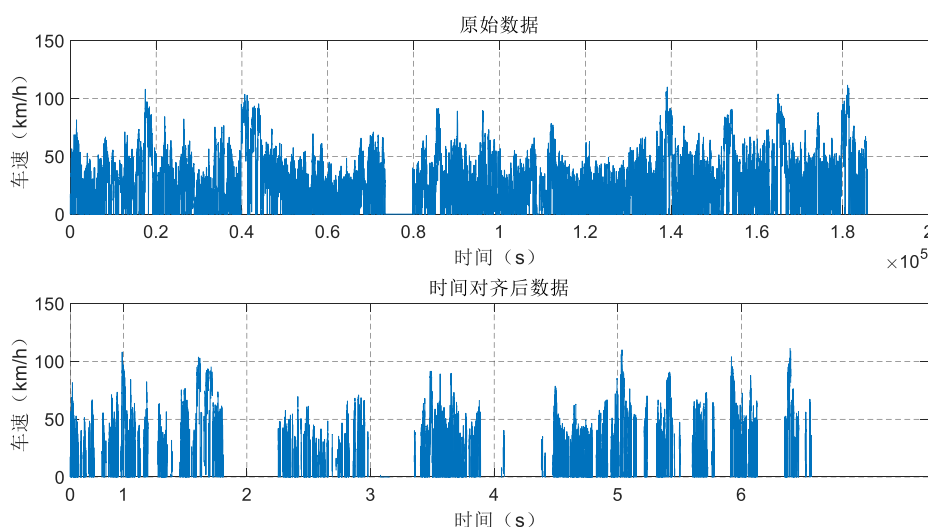


图 4-9 时间对齐前后对比

从上图中可以看出，当时间对齐后，数据的可视化效果变强。在时间对齐数据中以每天的分割节点划分了网格，通过图像可以很明显看出，驾驶员主要的行车时段。可以看出，

文件 1 中所采集车辆的驾驶员在第 1、5、6 天有通宵加班的情况出现，而在其他天里，正常作息，于是出现了长时间的 GPS 数据丢失情况(图中未出现曲线的部分)。此外，在一天的行程中，也存在短时间的 GPS 数据丢失情况，可以推测在该时间段内，驾驶员已经停车。

4.4.3. 对于经纬度丢失问题

在本题中，通过前文绘制出的汽车行驶路径图，我们得知在原始数据中，部分数据的经纬度数值存在着异常的情况（经纬度为 0），而当经纬度为 0 的时候其时间是连续的，并且此时速度一直保持固定数值不变，这极有可能是由于 GPS 的拆机设备发生了故障而引起的，所以我们将经纬度中存在 0 的数据全部视作异常数据处理，需要对其进行剔除处理。

4.4.4. 对于 GPS 信号丢失问题

由 4.4.2 的分析可以得知，GPS 信号丢失会造成时间节点的不连续，因此需要对其进行处理。而对于 GPS 信号丢失的情形需要分情况讨论，分为 GPS 信号短期丢失与长期丢失问题，对于这两种问题所采取的处理方式是不同的，具体如下：

（1）GPS 在采集信号时，可能会由于受到天气、建筑遮蔽物或信号接收设备情况等影响因素的作用，亦或者 GPS 电路异常等原因造成设备短时异常，出现短暂的信号丢失，随后立马恢复正常。这种情况属于可以掌控的异常数据，这里可以将 GPS 丢失的那几秒（直接取 1s）做插值处理；

（2）GPS 采集设备处理自身原因外，还有可能因为驾驶员的停车熄火操作，导致供电失败，从而丢失一段时间内的数据。这种情况可直接视为 GPS 信号丢失异常数据，做剔除处理。

4.4.5. 对于长期停车异常问题

由于汽车在采集数据的之后会出现一些较为特殊的情况，比如说停车不熄火等人、停车熄火了但是拆机设备仍处于运行状态，在这些情况下所采集到的数据均为异常的数据，需要对其进行处理。

我们判定当出现速度为 0，油门开度也为 0 且发动机的转速小于平均转速的时间持续 120 秒时，就将其定义为长期停车异常数据，对于这种类型的数据采取剔除的方式。

4.4.6. 对于加速度异常问题

由题目可以得知，普通轿车在一般情况下 0 至 100km/h 的加速时间大于 7 秒(约 4m/s^2)，紧急刹车最大减速度在 $7.5\sim 8\text{m/s}^2$ ，所以我们将加速度大于 4m/s^2 或者减速度大于 8m/s^2 的数据全部视为加速度异常数据，对于这种类型的异常数据进行剔除的方式。

4.4.7. 对于速度异常问题

依据我国目前对于轿车行驶速度的限制可以将速度大于 120km/h 的数据视为异常数据，经过在 Excel 中对于数据文件 3 进行分析，我们发现原始数据中速度大于 120km/h 的数据有 298 条，最高速度达到了 261.4，显然文件 3 中存在着速度异常的数据。而在文件 1 与

文件 2 中，速度均未超过 120km/h。据此，将速度大于 120km/h 的数据视为速度异常数据，对于这种类型的数据采取剔除的方式。

4.4.8. 对于怠速异常问题

由题目可以得知，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，所以将怠速时间超过 180 秒的数据视为怠速异常数据，其中将车速小于 10km/h 的数据均视为怠速处理。对于这种长时间怠速导致的怠速异常数据，采取剔除的方式。

4.4.9. 对于毛刺数据问题

参考文献资料^[2]，毛刺数据指的是汽车在行驶过程中，由于 GPS 信号或者其他的一些原因所导致的采集得到的速度数据较长时间的怠速数据中，出现了个别速度并不为 0 但是接近 0 的数据，这种情况是不符合实际情况的，而这一类的数据会对于后续问题二中的汽车运动学片段的划分有着不小的干扰作用，因此需要对于这类的异常数据进行分析处理。

依照实际中汽车的运行情况，当车辆发动后处于运行时一般会持续一段时间，我们将这个时间定为 4 秒，即当汽车如果出现连续运行时间小于 4 秒的情况时，就将这些数据视为毛刺问题数据，对于这种异常数据，我们对其做归零处理（在后文中将这种操作记为数据替换），即将所有的毛刺数据直接修改为 0，并将之视为怠速点。

4.5 数据预处理

按照 4.4 中数据质量分析的原则，对时间对齐后的每一个数据分块进行数据剔除、数值插补和数据替换这三种操作。

4.5.1. 数据剔除

由前文可以得知，进行异常数据剔除操作的类型共有 6 种类型，分别为：经纬度丢失问题数据、GPS 信号长期丢失问题数据、长期停车异常数据、加速度异常数据、速度异常数据、怠速异常问题数据。对于 3 个文件中的原始数据进行剔除操作可以得到具体的操作结果如下表，其中 GPS 长期丢失数据的剔除是在将所有时间不连续节点补齐之后再行剔除的（此时该节点并无数据），而原始数据中并不包含这些时间不连续节点的数据，所以并未将这一类剔除数据计入到最终的数据剔除总数中，具体的数据剔除结果如下表所示：

表 4-2 数据剔除

处理类型	原始数据	经纬度异常剔除	长期停车异常剔除	堵车低速异常剔除	GPS 长期丢失剔除	加速度异常剔除	减速度异常剔除	速度异常剔除	数据剔除
文件 1	185725	3	3174	6003	332302	27	2	0	9209
文件 2	145825	300	402	10529	370732	332	82	0	11645
文件 3	164914	0	1765	11277	266529	115	39	313	13509

4.5.2. 数值插补

由前文可以得知，进行异常数据剔除操作的类型共有 1 种类型，即 GPS 信号短期丢失

问题数据。在本文中采用均值插补方式，即使用平均值对于这些短期丢失的数据进行补完，具体的数值插补如下表所示：

表 4-3 数值插补

处理类型	原始数据	GPS 短时丢失插值	数值插补
文件 1	185725	110	110
文件 2	145825	1722	1722
文件 3	164914	341	341

4.5.3. 数据替换

由前文可以得知，进行异常数据剔除操作的类型共有 1 种类型，即毛刺问题数据。在本文中将这此毛刺数据全部用 0 替换掉。具体的数据替换结果如下表所示：

表 4-4 数值替换

处理类型	原始数据	毛刺数据片段	毛刺数据替换	数据替换
文件 1	185725	243	425	425
文件 2	145825	659	1077	1077
文件 3	164914	376	664	664

4.6 结果分析

使用 matlab 编写程序对原始数据进行处理，程序运行结果如下图所示：

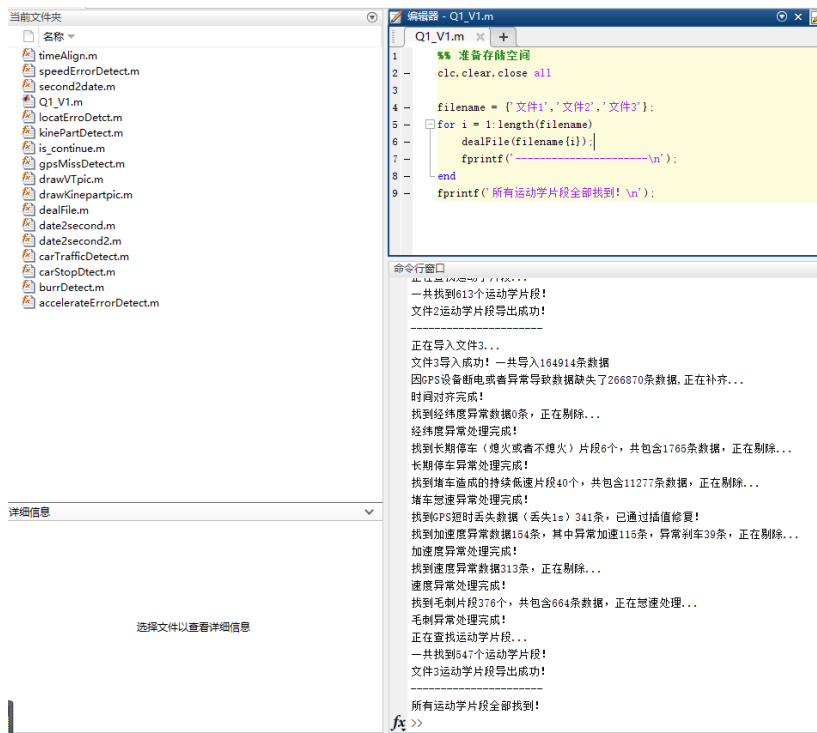


图 4-10 数据预处理程序运行结果

从前文中可以得知问题一数据预处理的整个结果如下表所示：

表 4-5 数据预处理结果

处理类型	原始数据	GPS 设备异常缺失	经纬度异常剔除	长期停车异常剔除	堵车低速片段	堵车低速异常剔除	GPS 短时丢失插值	GPS 长期丢失剔除
文件 1	185725	332412	3	3174	12	6003	110	332302
文件 2	145825	372454	300	402	35	10529	1722	370732
文件 3	164914	266870	0	1765	40	11277	341	266529
	加速度异常剔除	减速度异常剔除	速度异常剔除	毛刺数据替换	数据剔除	数值插补	数据替换	预处理后数据
文件 1	27	2	0	425	9209	110	425	176516
文件 2	332	82	0	1077	11645	1722	1077	134180
文件 3	115	39	313	664	13509	341	664	151405

以文件 1 为例,其原始数据有 185725 条,经过数据预处理还剩下 176516 条有效数据,数据剔除的比例约为 4.96%,数值插补的比例为 0.059%,数据替换比例为 0.229%,总体数据预处理比例为 5.246%,所以进行预处理比例较小能够很好保留住原始数据的信息。各文件数据处理比例具体如下表所示。

表 4-6 数据处理比例

	数据剔除比例	数值插补比例	数据替换比例	总计
文件 1	4.958%	0.059%	0.229%	5.246%
文件 2	7.986%	1.181%	0.739%	9.905%
文件 3	8.192%	0.207%	0.403%	8.801%
平均值	7.045%	0.482%	0.457%	7.984%

在经过数据的预处理之后,可以对处理之后的数据在进行经纬度数据对齐处理之后还原出某一些连续片段的汽车行驶路径轨迹图。下图是在文件 1 中选取了 3 个不同的连续时间段组合绘制出的路径轨迹图,根据时间对齐原则,原本因为 GPS 信号丢失造成的时间不连续的异常数据被时间插补对齐而离散开。这样非常有利于短行程的拆分。如下图所示,将经纬度从时间对齐后的数据中提取出来,将会得到一系列的短行程数据。并且在轨迹上可以明显看出, GPS 信号丢失, (图中蓝色圆圈部分,本应该是连续的路径,中间却没了经纬度数据)。

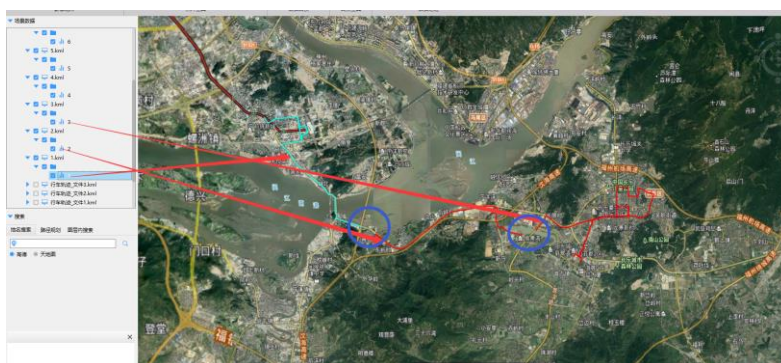


图 4-11 连续时间片段的运动轨迹图 (示例)

五、问题二：运动学片段的提取

5.1 问题分析

对于问题二，题目要求将问题一中预处理过后的数据划分为多个运动学片段，但是运动学片段又是目前构建汽车行驶工况曲线的最常用方法，所以问题二和问题三存在着诸多的联系与交集。基于篇幅考虑，将问题三中的运动学特征指标放入到问题二中进行探讨，因为在提取出运动学片段之后对这些指标进行分析计算会有利于后续的步骤。

在问题一中，我们得到了按照时间的连续性分割的片段，在本问中我们统计出每个片段中完整的运动学片段数量和不完整的运动学片段（由于 GPS 车速数据丢失过多而导致在其连续时间节点上开始或者结束的时候 GPS 车速并不为 0），由于不完整运动学片段缺失的数据过多，难以对其进行有效的数值插补，所以为了保证数据的质量将其全部剔除掉，只留下完整的运动学片段，并统计出每一个数据文件中完整的运动学片段的数目。

因此，首先对于汽车的行驶模式进行划分，之后从问题一预处理之后的数据中提取出完整的运动学片段并给出各数据文件分别得到的运动学片段数量，然后构建出能够体现运动学特征的一系列指标并计算出每一个运动学片段中的每一项特征指标的数值，最后对这些指标进行统计量分析，便于更为直观的理解车辆行驶中各运行模式特点。

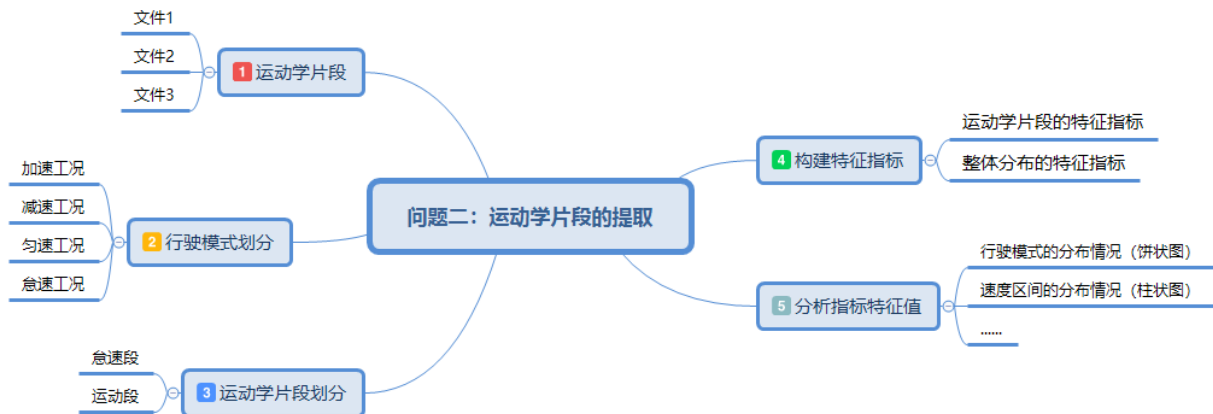


图 5-1 问题二思维导图

5.2 运动学片段提取

经过预处理后，得到的将会是一些已经剔除部分数据，断断续续的时间序列数据。而作为运动学片段，具有非常明显的特征，凭借这些特征可以实现运动学片段的提取。

对于一个运动学片段来说，主要包括怠速段和运动段。在怠速段，速度始终保持为 0，并且持续时间不会超过 180s（由前面的预处理可以知道）。在最后一个速度为 0 的时间节点处，汽车开始加速，运动段随之开始，中间可能经历各种加速减速，最后汽车速度为 0，运动段到此结束。同时运动段与前面的怠速段合成一个完整的运动学片段。基于上述运动学片的特征，可以画出选取运动学片段的流程图如下：

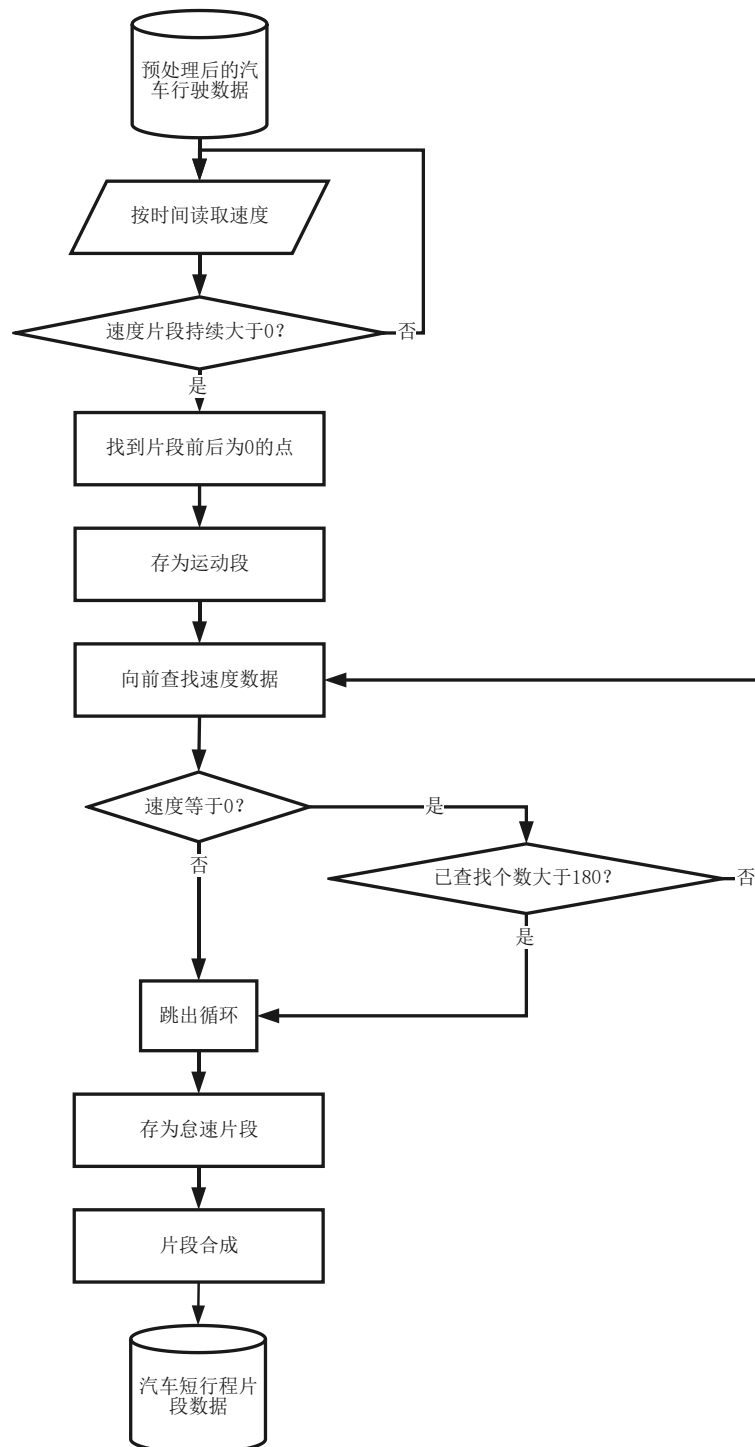


图 5-2 提取运动学片段的流程图

首先检测出所有的数据中，连续大于 0 的片段，再检测其前一个和后一个数据是否为 0。确认后，即可保证该片段为一个运动段。在运动段前面，是怠速段，数据全为 0，但是不可能超过 180 个。所以只需要找出运动段前面的时刻中连续为 0 且连续个数小于的 180 的那部分片段，存为怠速段。最后将两个片段整合到一起组成一个完整的运动学片段。最后我们从三个文件中划分得到 1190 个运动学，其中文件 1 有 830 个运动学片段，文件 2 有 613 个运动学片段，文件 3 有 547 个运动学片段，具体如下表所示：

表 5-1 运动学片段数量

	文件 1	文件 2	文件 3	
运动学片段	830	613	547	1990

我们以时间节点为 2017/12/18 13:42:14—2017/12/18 13:55:07 这一个连续的数据块作为示范，绘制出其时间段内的时间-速度曲线如下图所示，从图中我们可以看出，这是一个较为完美的连续时间片段，因此这从侧面上反映出本文问题一中数据预处理的方法较为合理。

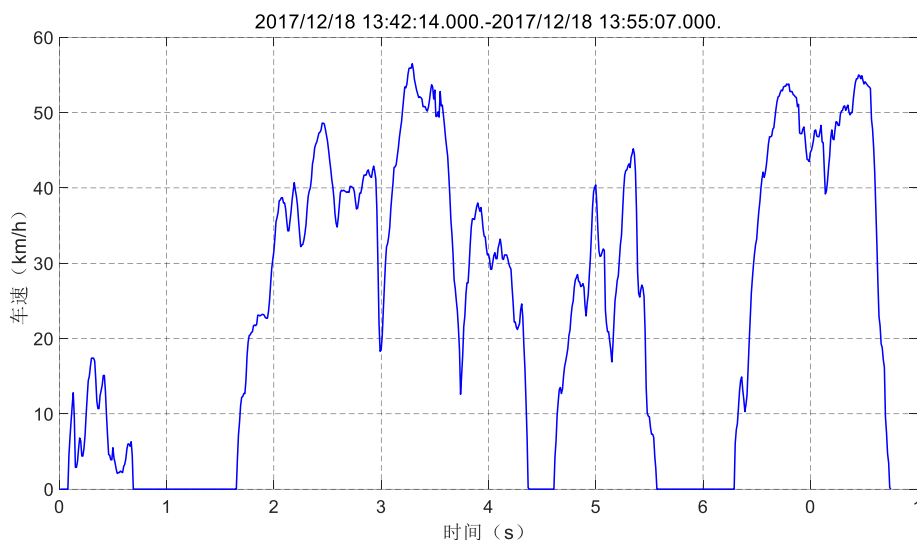


图 5-3 文件 1 第一个连续时间片段

从上图的连续时间片段中提取出完整的运动学片段，采用上述的提取流程与方法可以从文件 1 的第一个连续时间片段中提取出 4 个完整的运动学片段，分别绘制出它们的速度—时间曲线便可以得到下图，将图 5-3 与图 5-4 进行对比分析可以得知运动学片段的提取十分准确，而且每一个运动学片段都十分明显，效果较好。

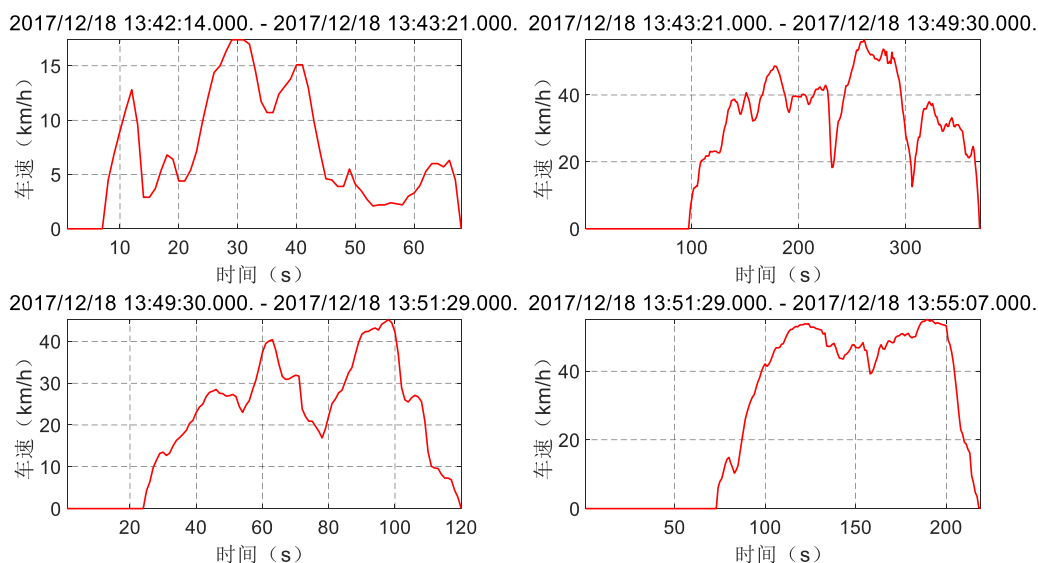


图 5-4 文件 1 第一个连续片段的 4 个运动学片段

如图，经过运动学片段拆分后，我们为了方便后续研究和查看，采取元胞数组来存储片段信息，其中存储形式如下所示。作为一个运动学片段，具有和原始数据一模一样的字段属性，并且在片段内，不存在时间不连续问题，详细可以参照下表。此得到运动学片段后，我们可以在此基础上做进一步的研究。

表 5-2 运动学片段的数据信息

时间	GPS 车速	...	发动机 负荷百分比	进气流量
2017/12/18 13:42:14.000.	0	...	21	2.39
2017/12/18 13:42:15.000.	0	...	22	2.19
2017/12/18 13:42:16.000.	0	...	21	2.21
...
2017/12/18 13:43:19.000.	6.3	...	22	2.18
2017/12/18 13:43:20.000.	4.5	...	21	2.13
2017/12/18 13:43:21.000.	0	...	21	2.25

5.3 行驶模式划分

由预处理之后的数据可以得知，汽车在道路上行驶的时候车速的变化较为复杂。而在构建行驶工况时，可以将汽车的行驶状态划分为四种：怠速模式、加速模式、匀速模式以及减速模式。通过参考文献^[3]得知目前存在三种不同的划分标准，我们结合题目所给的名词解释里对于加速和减速的定义，采用第一种标准，具体如下：

- 1) 怠速模式：汽车停止运动（即车速为 0m/s ），但发动机保持最低转速运转的连续过程的模式。此外，当长时间堵车、断断续续低速行驶情况（最高速度小于 10km/h ），按照怠速情况处理；
- 2) 加速模式：车辆在行驶过程中处于加速状态的模式，其中汽车加速度大于 0.1 m/s^2 ；
- 3) 减速模式：车辆在行驶过程中处于减速状态的模式，其中汽车加速度小于 -0.1 m/s^2 ；
- 4) 匀速模式：汽车加速度的绝对值小于 0.1m/s^2 ，并且处于非怠速的连续运行状态；

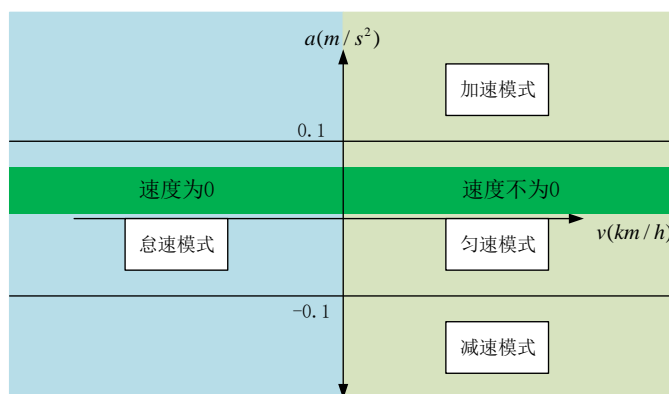


图 5-5 行驶模式的划分细则

5.4 构建特征指标

为了能够对于车辆的行驶工况进行合理的描述，需要引入汽车的运动特征评估体系。本文在题目要求的 9 种必要指标基础上，结合文献^[4]和文献^[5]，引入了运行时间、怠速时间、加速时间、减速时间、匀速时间等 5 个时间指标，平均速度、平均行驶速度、最大速度、速度标准差等 4 个速度指标，平均加速度、最大加速度、加速度标准差等 3 个加速度指标，平均减速度、最小减速度、减速度标准差等 3 个减速度指标，故而总共引入 15 个用于描述运动学片的特征参数，具体的运动学片的特征参数名称与单位如下表所示：

表 5-3 用于描述运动学片的特征参数

序号	特征参数	代表符号	单位
1	运行时间	T	s
2	怠速时间	T_i	s
3	加速时间	T_a	s
4	减速时间	T_d	s
5	匀速时间	T_c	s
6	平均速度	v_m	km/h
7	平均行驶速度	v	km/h
8	最大速度	v_{\max}	km/h
9	速度标准差	v_{sd}	km/h
10	平均加速度	a_a	m/s^2
11	最大加速度	$a_a(\max)$	m/s^2
12	加速度标准差	$a_a(sd)$	m/s^2
13	平均减速度	a_d	m/s^2
14	最小减速度	$a_d(\min)$	m/s^2
15	减速度标准差	$a_d(sd)$	m/s^2

结合文献[太原市轻型车实际行驶工况构建_田宇]和文献[城市道路轻型汽车运行工况

构建_姚延钢], 引入了怠速时间比、加速时间比、减速时间比、匀速时间比等 4 个时间比例指标, 0~10km/h 速度段比例、10~20km/h 速度段比例, ..., 60~70km/h 速度段比例、大于 70km/h 速度段比例等 8 个速度比例指标, 故而总共引入 12 个用于描述整体分布的特征参数, 具体的运动学片段的特征参数名称与单位如下表所示:

表 5-4 用于描述整体分布的特征参数

序号	特征参数	代表符号	单位
1	怠速时间比	P_i	%
2	加速时间比	P_a	%
3	减速时间比	P_d	%
4	匀速时间比	P_c	%
5	0~10km/h 速度段比例	P_{0-10}	%
6	10~20km/h 速度段比例	P_{10-20}	%
7	20~30km/h 速度段比例	P_{20-30}	%
8	30~40km/h 速度段比例	P_{30-40}	%
9	40~50km/h 速度段比例	P_{40-50}	%
10	50~60km/h 速度段比例	P_{50-60}	%
11	60~70km/h 速度段比例	P_{60-70}	%
12	大于 70km/h 速度段比例	P_{70}	%

5.4.1. 运动学片段特征值计算

(1) 运行时间 T

运行时间是指运动学片段的长度, 由于本题中数据采样的频率为 1HZ, 所以运行时间等于该运动学片段中采集到的行驶车速个数, 故由:

$$T = N, N \text{ 为该运动学片段中采集到的行驶车速个数}$$

(2) 怠速时间 T_i 、加速时间 T_a 、减速时间 T_d 、匀速时间 T_c

怠速时间是指在运动学片段中汽车处于怠速状态的时间, 依据怠速模式的定义则有:

$$T_i = \text{运动学片段中速度为0的个数}$$

同理可得： T_a = 运动学片段中加速度大于 0.1m/s^2 的个数；

T_d = 运动学片段中加速度小于 -0.1m/s^2 的个数， $T_c = T - T_i - T_a - T_d$ ；

(3) 平均速度 v_m 、平均行驶速度 v 、最大速度 v_{\max} 、速度标准差 v_{sd}

平均速度为一段时间周期内汽车速度的算术平均值；平均行驶速度为汽车在行驶状态下汽车速度的算术平均值，即不包含汽车怠速状态的平均速度；最大速度为一个运动学片段中最大的速度；速度标准差为一段时间周期内（包括怠速状态）汽车速度的标准差。所以依据各指标的含义有：

$$v_m = \frac{\sum_{i=1}^N v_i}{N}, \text{ 式中 } v_i \text{ 为第 } i \text{ 个时间点的速度；}$$

$$v = \frac{\text{运动段内速度大于0的速度总和}}{\text{运动段内速度大于0的个数}}; \quad v_{\max} = \max(v_1, v_2, \dots, v_N); \quad v_{sd} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (v_i - v_m)^2};$$

(4) 平均加速度 a_a 、最大加速度 $a_a(\max)$ 、加速度标准差 $a_a(sd)$

平均加速度为汽车在加速状态下各单位时间（秒）加速度的算术平均值；最大加速度为在运动段内的数值最大的加速度；加速度标准差为一段时间周期内处在加速状态的汽车加速度的标准差。则有：

$$a_a = \frac{\sum_{i=1}^{N_a} a_i}{N_a}, \text{ 式中 } N_a \text{ 为加速段的行驶车速个数；}$$

$$a_a(\max) = \max(a_1, a_2, \dots, a_N); \quad a_a(sd) = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N_a} (a_i - a_a)^2}$$

减速度的 4 个指标计算方法与加速度的指标完全一致，故而在本文限于篇幅不再赘述。

5.4.2. 整体分布特征值计算

(1) 怠速时间比 P_i 、加速时间比 P_a 、减速时间比 P_d 、匀速时间比 P_c

怠速时间比为一段时间周期内，怠速状态的累计时间长度占该时间周期总时间长度的百分比，因此有：

$$P_i = \frac{T_i}{T}$$

同理可得其它几个状态的时间比：

$$P_a = \frac{T_a}{T}, P_d = \frac{T_d}{T}, P_c = \frac{T_c}{T}$$

(2) 各速度段的比例

$$P_{0-10} = \frac{\text{运动段内 } v_i < 10\text{km/h} \text{ 的数据个数}}{N}, \dots, P_{70} = \frac{\text{运动段内 } v_i > 70\text{km/h} \text{ 的数据个数}}{N}$$

5.4.3. 计算特各个运动学片段的各项特征值

按照上述的计算方法，以本文的文件 1 为例，计算其 830 条运动学片段的特征参数进行计算，采用 MATLAB 进行编程计算，可以得到下表所示的结果：

表 5-5 描述运动学片段的特征参数值

片段	1	2	...	829	830
运行时间	67	369	...	285	486
怠速时间	15	120	...	30	73
加速时间	26	111	...	78	143
减速时间	19	86	...	99	106
匀速时间	7	52	...	78	164
平均速度	7.020588	26.11378	...	33.42273	51.55421
平均行驶速度	7.956667	35.52243	...	34.50866	53.19258
最大速度	17.4	56.5	...	56.1	67.2
速度标准差	5.244373	18.49912	...	17.91624	14.93277
平均加速度	0.425214	0.382132	...	0.372507	0.273893
最大加速度	1.25	1.083333	...	2	1.861111
加速度标准差	0.242514	0.227593	...	0.320847	0.234862
平均减速度	-0.58187	-0.49419	...	-0.29433	-0.37448
最小减速度	-0.11111	-0.11111	...	-0.11111	-0.11111
减速度标准差	0.446991	0.419243	...	0.221723	0.322401

同理以本文的文件 1 为例，计算其 830 条运动段的整体分布特征参数进行计算，采用 MATLAB 进行编程计算，可以得到下表所示的结果：

表 5-6 描述整体分布的特征参数值

片段	1	2	...	829	830
怠速时间	0.223881	0.325203	...	0.105263	0.150206
加速时间	0.38806	0.300813	...	0.273684	0.294239
减速时间	0.283582	0.233062	...	0.347368	0.218107
匀速时间	0.104478	0.140921	...	0.273684	0.337449
v0-10	0.666667	0.018382	...	0.111913	0.006356
v10-20	0.333333	0.0625	...	0.187726	0.010593
v20-30	0	0.202206	...	0.090253	0.042373
v30-40	0	0.382353	...	0.104693	0.091102
v40-50	0	0.205882	...	0.245487	0.129237
v50-60	0	0.128676	...	0.259928	0.358051
v60-70	0	0	...	0	0.362288
v70-inf	0	0	...	0	0

5.5 特征指标分析

以文件 1 位示例，绘制出其各类行驶工况分布图和总体速度区间分布图如下所示：

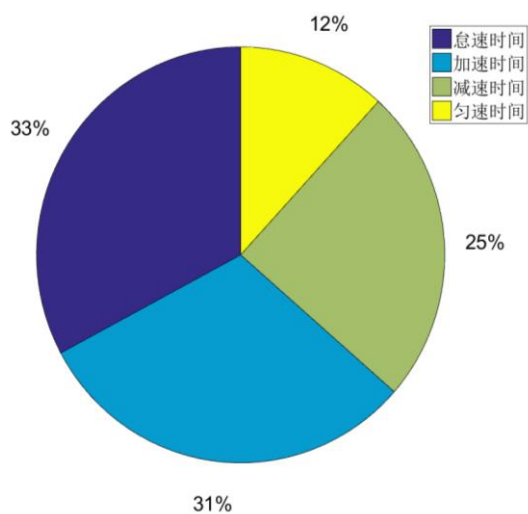


图 5-6 文件 1 各类行驶工况分布图

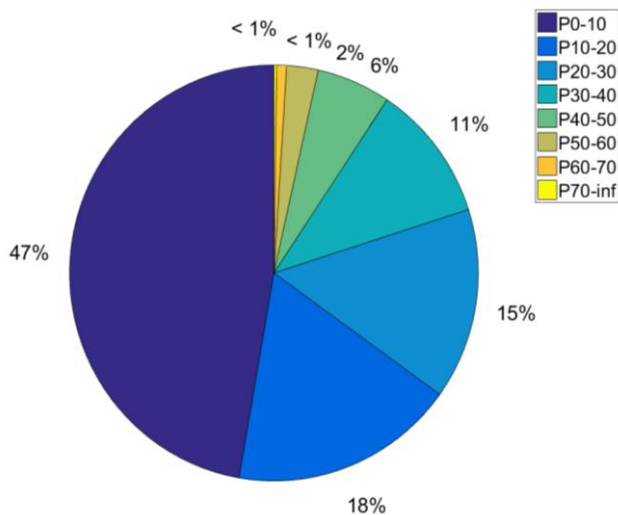


图 5-7 文件 1 各速度区间分布图

从上图中可以看出，文件 1 中怠速时间与加速时间占比较大，速度集中分布在 0-30 的段中，这个区间的速度占比高达 80%，初步推断在文件 1 中存在着众多的低速路段的运动学片段。

六、问题三：汽车行驶工况的构建

6.1 问题分析

对于问题三，题目要求构建出一条汽车行驶工况曲线。由参考文献^[1]可以得知，目前我国已有的较为成熟的工况曲线构建有短行程法、小波变换法、马尔科夫法、主成分分析法、聚类分析和马尔科夫结合法等诸多方法，从该文献得出的结论中我们发现运用主成分分析与聚类分析法，把行驶的路段划分成为不同的车速路段可以较好反映出汽车的行驶情况。

基于上述思考，本文采用主成分分析与聚类分析结合的办法，由于问题二中定义的运动学指标数量较多，并且这些指标之间存在着重叠的信息，所以首先通过采用主成分分析法对于问题二中是计算出的用于描述运动学片段的各项指标进行降维处理；在主成分分析得出主成分之后，尽管此时已经对数据进行了降维处理，但是整体的数据量依然十分庞大，所以还需进行进一步的降维处理。故而再次借用这些主成分来对于运动学片段进行聚类处理，将问题二中划分出来的运动学片段按照汽车行驶道路的差别进行聚类，最后得到 3 个大类的运动学片段（即低速路段、中速路段和高速路段），再分别对于每一个大类的运动学片段进行处理，为最后的道路实行工况的拟合做好准备。

在得到三大类的汽车运行道路路段之后，对于每一类道路，结合问题二中定义好的用于描述整体分别特征的 12 个参数，计算出该类型道路每一项整体分布指标的平均值。然后采用这 12 个指标，通过皮尔逊相关系数法计算出该类型道路中每一个运动学片段与的相关度并进行排序。之后，从中按照相关度大小从高到低依次取出运动段，直到取出的运动段累计时间处于 1200-1300 秒。最后将这些取出的运动段组合在一起便构建出该类型路段的汽车行驶工况。

在构建出三种类型道路的汽车行驶工况后，结合参考文献^[4]中对于低、中、高速的速度节点的定义，从聚类之后的数据中统计出三种速度各自的占比，并将这个比例定为最后整个道路汽车行驶工况曲线中各自的大概时间占比。之后从三个工况中按照占比取出相应的运动学片段按照低-中-高的方式组合成为整个道路的汽车行驶工况。

构建汽车行驶工况的具体流程如下图所示：

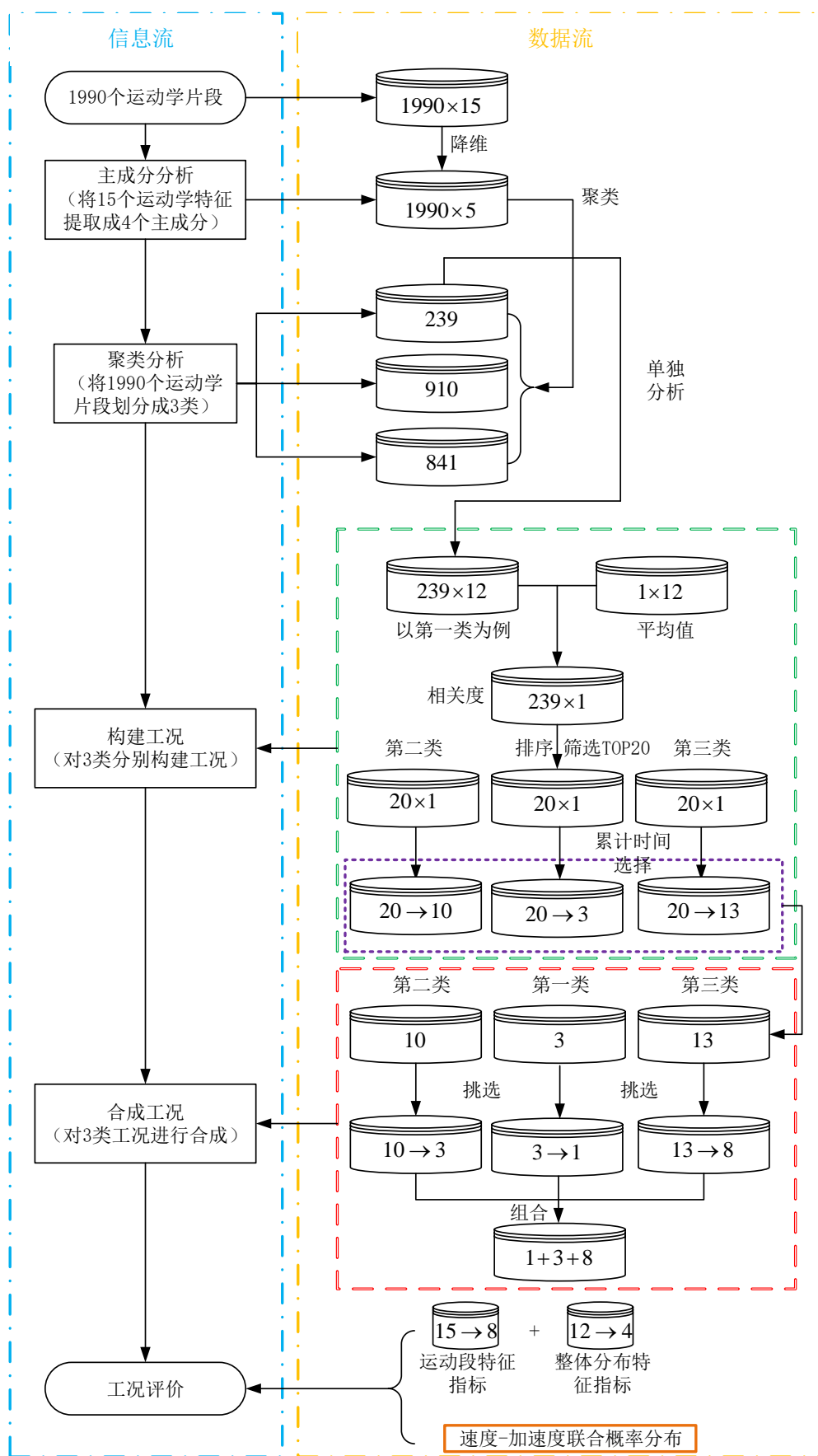


图 6-1 问题三流程图

6.2 主成分分析

由问题二得到 15 个用于描述运动学片的特征参数，如若直接采用这些参数来联合表示运动学片的信息会使得问题的复杂度骤然上升，此外这 15 个参数之间并不是相互独立的，相反这些参数往往是相互影响、相互关联的，它们之间存在着许多的重叠部分。所以需要首先采用主成分分析的方法对于这些指标进行降维处理，从而在减小工作量的同时又保留这些指标所代表的绝大部分信息。

主成分分析的步骤流程如下图所示，具体步骤如下：

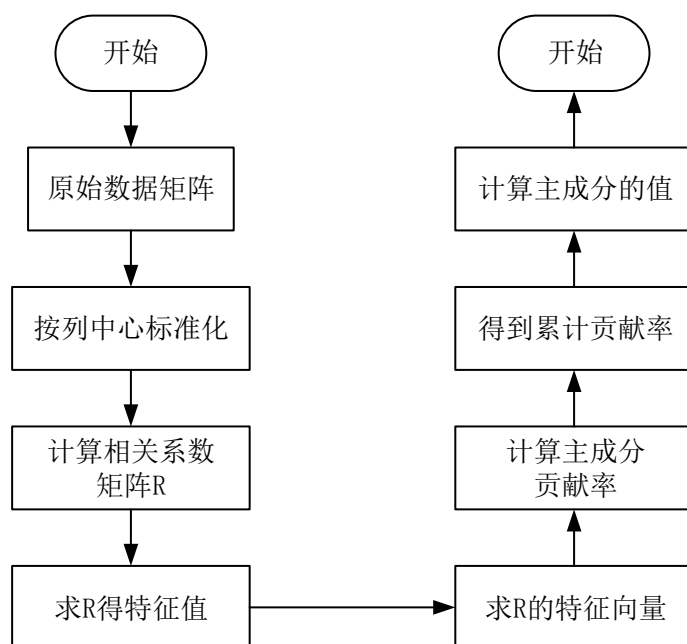


图 6-2 主成分分析法流程

Step1: 构建原始数据矩阵。以运动学片作为行，15 项特征指标作为列，构建出 1990×15 的原始数据矩阵；

Step2: 数据标准化。将原始数据矩阵按列进行中心标准化处理，得到标准化后的数据矩阵，为方便描述将其称为标准矩阵，记为 X ，其中 $m=15$ ， $n=1990$ ：

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} = [X_1 \quad X_2 \quad \dots \quad X_m]$$

Step3: 计算相关系数矩阵。相关系数矩阵 $R = (r_{ij})_{m \times m}$ ，有：

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i) \cdot (x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \cdot \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}$$

式中： $r_{ii}=1, r_{ij}=r_{ji}$ ， r_{ij} 表示第 i 个指标与第 j 个指标的相关系数。

Step4: 计算 R 的特征值和对应的特征向量。计算得 R 的 p 个特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ，对应的特征向量组成的矩阵为 $u=[u_1, u_2, \dots, u_p]$ 。

Step5: 计算主成分贡献率。记主成分贡献率为 b_j ，则有：

$$b_j = \frac{\lambda_j}{\sum_{k=1}^p \lambda_k}, j=1, 2, \dots, p$$

Step6: 计算累积贡献率。记累积贡献率为 α_t ，则有：

$$\alpha_t = \frac{\sum_{k=1}^t \lambda_k}{\sum_{k=1}^p \lambda_k}, t \leq p$$

由参考资料^[5]我们认为当 $\alpha_t \geq 85\%$ 时，这部分的主成分就包含了代表原指标的绝大多数信息。

Step7: 计算主成分的值。

$$Z = (Xu_1, Xu_2, \dots, Xu_t) = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_t)$$

则有 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_t$ 为 1 至 t 个主成分，它们之间线性不相关。

将问题二得到的数据导入到 MATLAB 中，按照上述流程编写代码对所有的运动学片段的 15 个特征指标进行主成分分析，得到的结果如下表所示：

表 6-1 各主成分贡献率/累计贡献率

主成分	贡献率%	累计贡献率%
M1	42.84%	42.84%
M2	19.87%	62.71%
M3	12.02%	74.74%
M4	7.85%	82.59%
M5	6.35%	88.94%
M6	4.74%	93.68%
M7	2.74%	96.42%
M8	1.23%	97.65%
M9	0.78%	98.43%
M10	0.71%	99.14%
M11	0.49%	99.63%
M12	0.15%	99.78%

M13	0.14%	99.92%
M14	0.08%	100.00%
M15	0.00%	100.00%

依据每一个主成分的贡献率，将其绘制成如下图所示，从图中可以明显看出各个主成分贡献率的整体变化趋势，前五个主成分的贡献率变化趋势较大，存在着明显的拐点，而从第 6 个主成分开始，贡献率的变化趋势逐步较小，变化较为平缓。

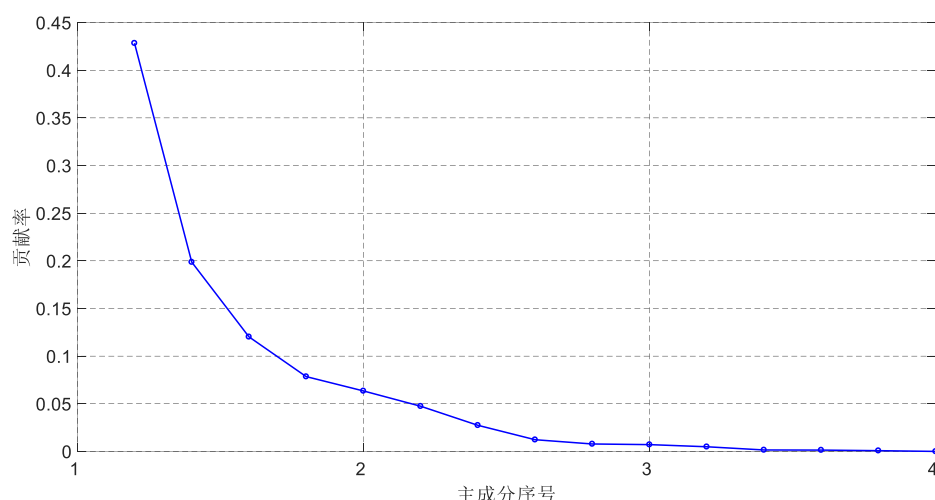


图 6-3 主成分贡献率图

从表 6-1 和上图 6-3 中可以看出第一个主成分的贡献率为 42.84%，第二个主成分的贡献率为 19.87%，第三个主成分的贡献率为 12.02%，第四个主成分的贡献率为 7.85%，第五个主成分的贡献率为 6.35%，前五个主成分的累计贡献率达到了 88.94%，大于预定的 85%，所以前五个主成分中就已经包含了问题二中设定的 15 个用于描述运动学片段的特征指标的绝大部分信息，所以可以采用这五个主成分来替代原来的指标。每一个运动学片段的前五个主成分的值如下表所示：

表 6-2 运动片段主成分值

运动片段 序号	第一 主成分	第二 主成分	第三 主成分	第四 主成分	第五 主成分
1	-1.6852	1.2479	-0.7131	-0.1399	0.0673
2	4.8746	1.7223	0.8440	-1.5923	0.9890
3	0.9748	0.2632	-1.3319	0.1548	0.2660
...
1988	-3.6985	1.1376	0.7689	0.6491	-0.5002
1989	-2.8130	1.4845	-1.0425	-0.1405	-0.1935
1990	0.7906	0.6820	-1.5963	0.0806	0.4829

6.3 聚类分析

依据文件 1、2、3 的行车路径，可以将本题中的汽车行驶到路划分为三大类，即行车比较拥堵速度最慢的市区中心路段，行车较为通畅速度较快的郊区路段，行车十分通畅且

速度最快的环城高速路段。因此，在对所有运动学片段进行聚类分析的时候采用 k-means 聚类分析法，将其分为三类。在 6.2 得到的表 6-2 的基础上，依据每一个运动学片段的各个主成分（5 个）的数值将其划分成为三大类。其中 k-means 聚类分析法的流程如下图所示：

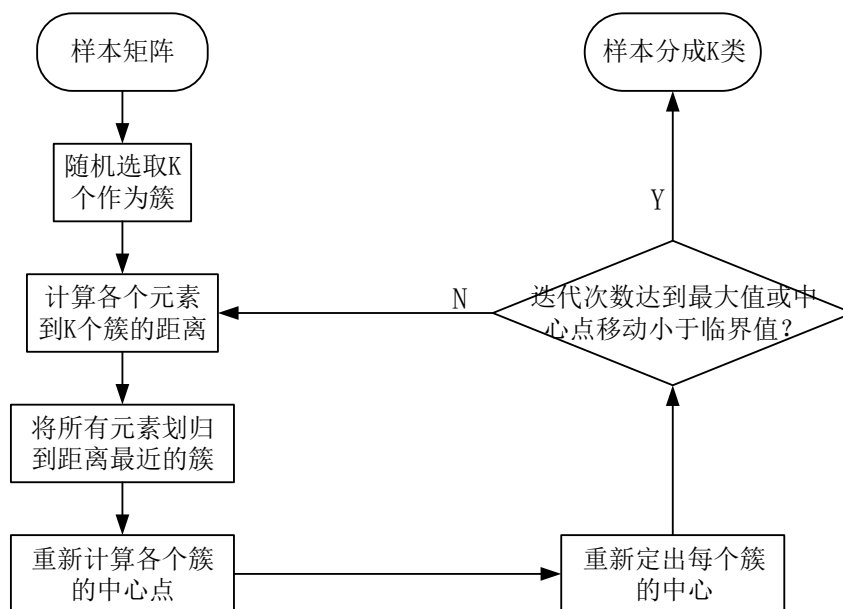


图 6-4 k-means 聚类流程图

通过 k-means 聚类分析，将 1990 个运动学片段分为三类，其中第一类中包含 239 个运动学片段，第二类中包含 841 个运动学片段，第三类中包含了 910 个运动学片段，结合问题二中定义的 12 个用于描述整体分布的特征指标，计算出每一类所有运动学片段的 12 个指标的平均值，结果如下表所示，其中整体均值为所有运动学片段的 12 个指标的算数平均值。

表 6-3 各类别的整体特征指标值

	第一类片段	第二类片段	第三类片段	整体
怠速比例	18.22%	33.49%	45.87%	32.53%
加速比例	37.24%	33.09%	22.73%	31.02%
减速比例	29.94%	23.96%	21.42%	25.11%
匀速比例	14.61%	9.46%	9.98%	11.35%
P ₀ -P ₁₀	2.98%	14.25%	65.37%	27.54%
P ₁₀ -P ₂₀	12.58%	23.75%	32.90%	23.08%
P ₂₀ -P ₃₀	16.50%	29.43%	1.72%	15.88%
P ₃₀ -P ₄₀	23.86%	26.07%	0	16.64%
P ₄₀ -P ₅₀	18.55%	6.49%	0	8.35%
P ₅₀ -P ₆₀	13.42%	0	0	4.47%

$P_{60}-P_{70}$	4.57%	0	0	1.52%
$P_{70}-P_{inf}$	0.19%	0	0	0.06%
运动段数量	239	910	841	1190
平均速度	34.0527	16.1677	4.6942	18.3049

利用上表中的数据，绘制出每一类运动段的行驶模式比例分布情况如下图所示：

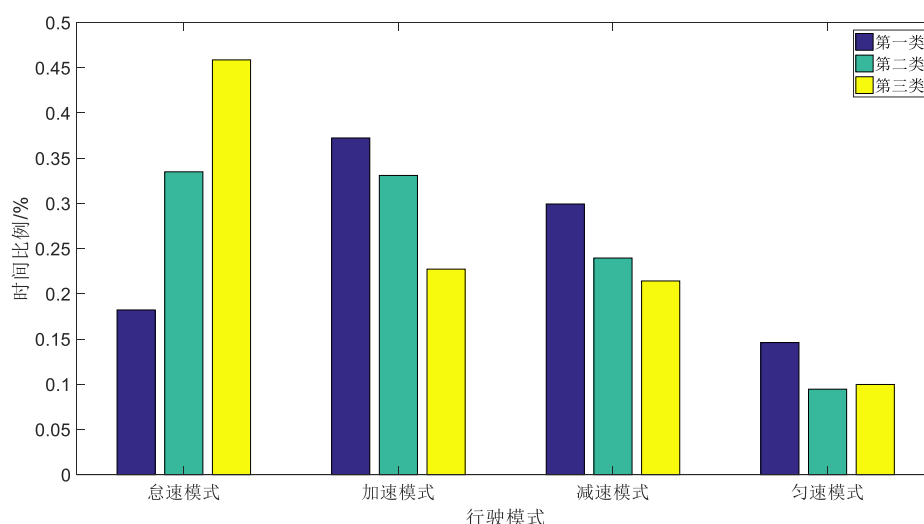


图 6-5 各类运动段的行驶模式比例图

结合表 6-3 和图 6-5，可以分析得知第一类运动学片段的平均速度为 34.0527km/h，汽车的行驶方式主要分布于加速度模式与减速模式；第二类运动学片段的平均速度为 16.1677km/h，汽车的行驶方式主要分布于加速度模式与怠速模式；第三类运动学片段的平均速度为 4.6942km/h，汽车的行驶方式主要分布于怠速模式：得出如下结论：

（1）第一类片段代表着高速行驶路段，即汽车高速行驶工况。平均速度相对较高，而且汽车加减速频繁，车速高于 50km/h 以上行驶的比例有 18.18%，说明汽车在高速路段行驶时道路十分通畅，路况良好，这些路段可能时郊区路段或者高速路段，而且时间上也错开了上下班的高峰期；

（2）第二类片段代表着中速行驶路段，即汽车中速行驶工况。平均速度相对适中，而且汽车处于加速模式和怠速模式的占比较大，所有车速均位于 50km/h 以下，说明汽车在中速路段行驶时道路中交叉路口和红绿灯较多，道路较为拥挤，这些路段可能位于市中心，存在大量红绿灯，但是错开了交通高峰期，较少出现堵车情形；

（3）第三类片段代表着低速行驶路段，即汽车低速行驶工况。平均速度非常小，将近一半的时间处于怠速模式，所有车速均位于 30km/h 以下，说明汽车在低速路段行驶时道路十分拥挤，汽车经常处于怠速状态，可能是在这些时段中汽车行驶于市中心地点且行驶时间与交通的高峰期相重合，出现了大量堵车现象。

6.4 工况构建

在将运动学片段划分为 3 大类之后，每个类别中均有大量的运动学片段，需要从个类别的众多运动学片段中提取出一定的片段来进行拟合得出该类型路段的汽车型工况曲线。最后再分别从三个不同路段的汽车工况曲线中提取出一定数量的运动学片段，将他们组合在一起即可得到本文中汽车的行驶工况曲线。

依据参考文献^[6]和^[7]，使用皮尔逊相关系数法从每一类的路段中选出相关度较大的几个运动学片段来构建各自的工况曲线，并采用类似 WLTP 的行驶工况合成步骤来合成出本题的汽车行驶工况曲线，具体步骤如下所示，流程如图所示：

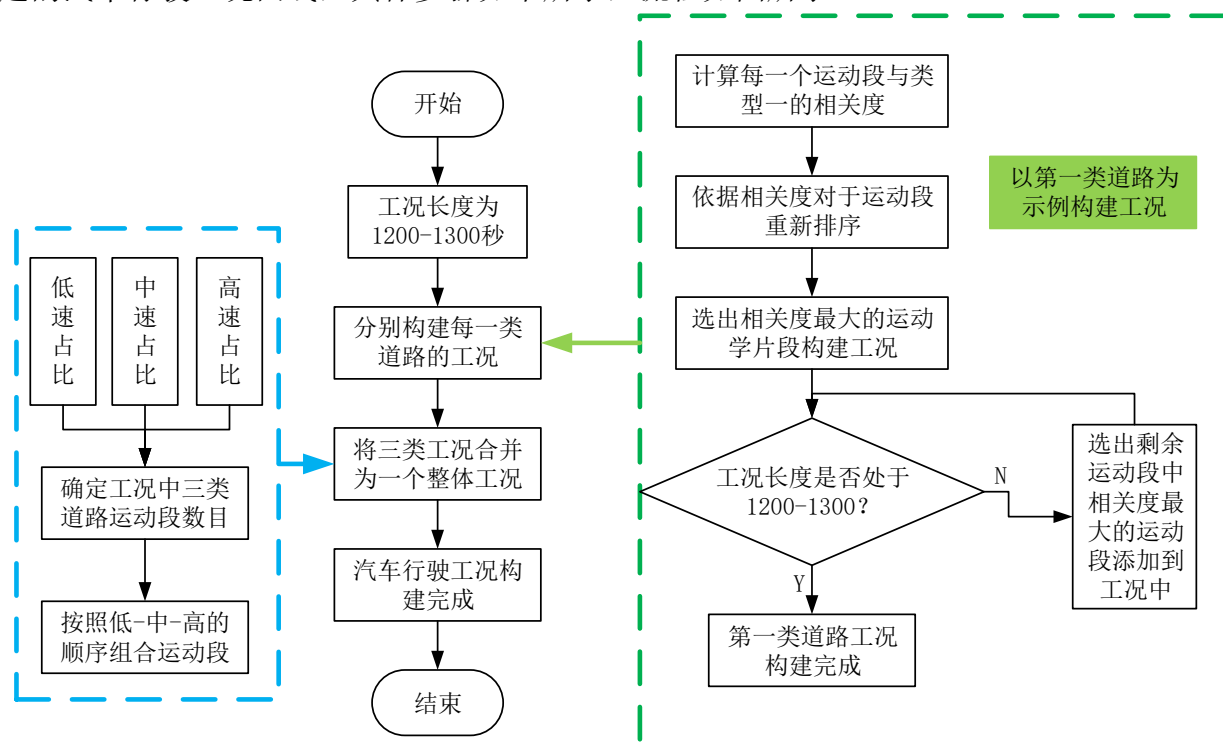


图 6-6 工况构建流程图

(1) 首先明确构建工况的长度。题目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，而该曲线的时间长度为 1200-1300 秒，所以本文将车辆道路行驶工况的长度定为 1200 秒左右。因此在后续步骤中构建出来的每一类道路的工况均需要符合该条件；

(2) 在得到三大类的汽车运行道路路段之后，对于每一类道路，结合问题二中定义好的用于描述整体分别特征的 12 个参数 X ，计算出该类型道路每一项整体分布指标的平均值 Y 。然后采用这 12 个指标，通过皮尔逊相关系数法计算出该类型道路中每一个运动学片段与的相关度并进行排序。之后，从中按照相关度大小从高到低依次取出运动段，直到取出的运动段累计时间处于 1200-1300 秒。最后将这些取出的运动段组合在一起便构建出该类型路段的汽车行驶工况。对于每一种类型的道路，分别计算出该道路类型中每一个运动段用于描述整体分布的特征指标 X 与该类型道路所有运动段的平均特征指标 Y 之间的相

关系数。采用皮尔逊相关系数法来计算两者之间的相关度，公式如下：

$$\rho(X_i,Y)=\frac{Cov(X_i,Y)}{\sqrt{D(X_i)}\cdot\sqrt{D(Y)}}$$

式中：Cov(X,Y)为X_i与Y的协方差，D(X_i)和D(Y)分别为X_i与Y的方差，X_i为第i个运动学片段的整体分布特征值组成的序列。

基于该步骤，采用matlab编程，得到按照皮尔逊相关系数从大到小的顺序给出每一种道路的运动学片段具体信息如下表所示，从表中我们可以得知高速路段中第1001个运动学片段的相关度最大，为0.984，而该运动段的持续时长为649秒；中速路段中第1178个运动学片段的相关度最大，为0.966，而该运动段的持续时长为139秒；低速路段中第44个运动学片段的相关度最大，为0.991，而该运动段的持续时长为59秒。

表 6-4 三类路段运动学片段的相关系数

第一类 路段	相关 系数	片段 时长	第二类 路段	相关 系数	片段 时长	第三类 路段	相关 系数	片段 时长
1001	0.984	649	1178	0.966	139	44	0.991	59
856	0.961	396	111	0.965	124	1841	0.990	59
397	0.949	279	652	0.963	202	387	0.990	33
884	0.944	578	1086	0.962	75	568	0.989	78
863	0.940	380	468	0.959	77	1879	0.988	105
922	0.939	330	775	0.959	48	318	0.987	41
116	0.939	539	1223	0.956	241	1298	0.986	90
1830	0.938	199	866	0.949	189	1242	0.985	53
919	0.935	194	419	0.942	80	18	0.984	102
6	0.934	220	1261	0.941	128	31	0.984	37
...

（3）最后从每一类路段中选出代表性的运动学片段组合成为一个汽车行驶工况。首先确定出每一种类型的道路分别取几个运动学片段，由聚类之后每一类的速度段分布情况确定出中速速度节点为30km/h，高速速度节点为50km/h，统计出低速（0-30）、中速（30-50）、高速（50-120）的个数占比，将这个比例定为最后整个道路汽车行驶工况曲线中各自的大概时间占比。其中低速比例为0.4021，中速比例为0.2217，高速比例为0.3762。再从每种道路的运动学片段中选出相关度较高的片段（并进行修正）即可得到用于构建整体道路汽车行驶工况的运动学片段（必须保证工况为1200-1300秒）。在选出片段之后，按照低速-中速-高速的顺序将片段进行排列，在同一类路段中按照相关度的大小依次进行排列。

依照上述的步骤与流程，可以得到每一类道路的行驶工况如下图 6-7 至图 6-9 所示，

其中从高速路段选用了三个时长分别为 649 秒、396 秒和 199 秒的三个运动学片段，最后构建出来的高速路段汽车行驶工况为 1244 秒，这三个运动段的相关系数均高于 93%，所以最后构建出来的这条工况基本上能够体现出真实的情况。从中速路段按照相关系数从大到小的顺序依次选用了 10 个运动学片段，最后构建出来的中速路段汽车行驶工况为 1249 秒，这 10 个运动段的相关系数均高于 94%，所以最后构建出来的这条工况基本上能够体现出真实的情况。从低速路段中按照相关系数从大到小的顺序依次选用了 13 个运动学片段，最后构建出来的高速路段汽车行驶工况为 1224 秒，这三个运动段的相关系数均高于 95%，所以最后构建出来的这条工况基本上也能够体现出真实的情况。

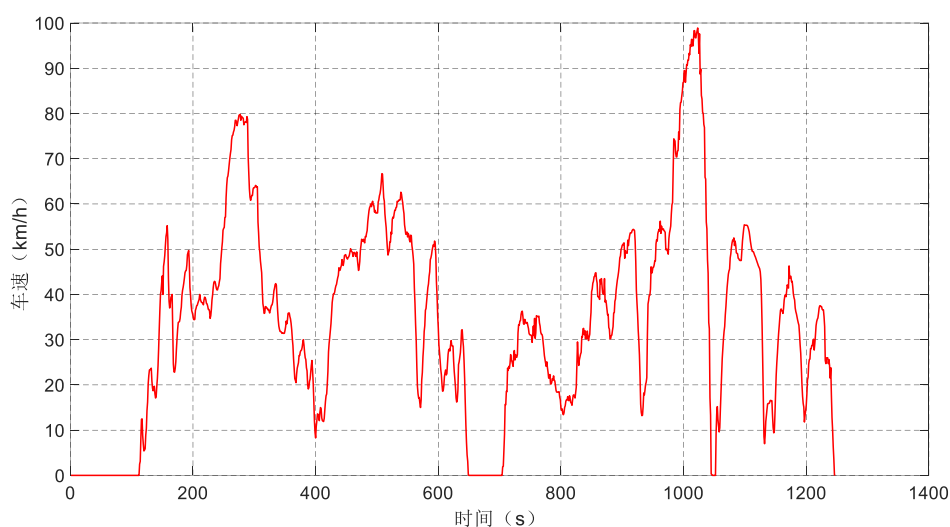


图 6-7 第一类（高速路段）汽车行驶工况

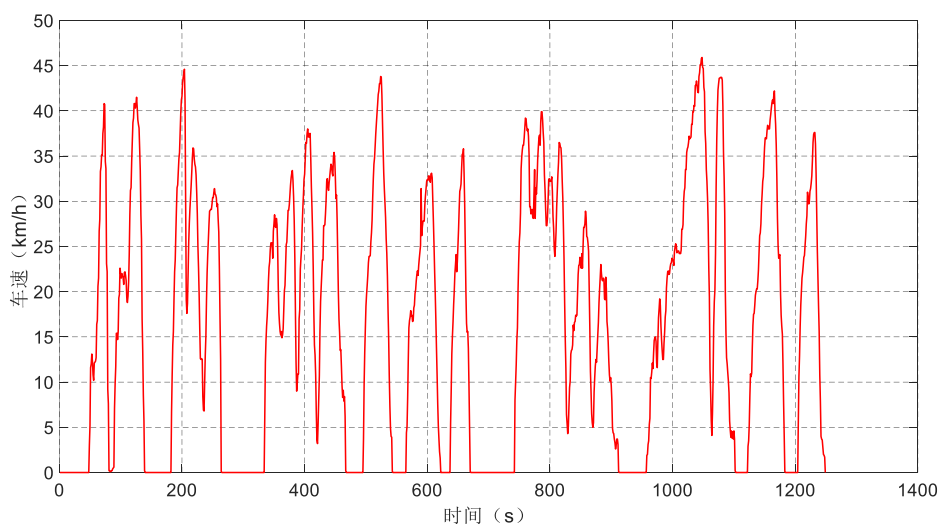


图 6-8 第二类（中速路段）汽车行驶工况

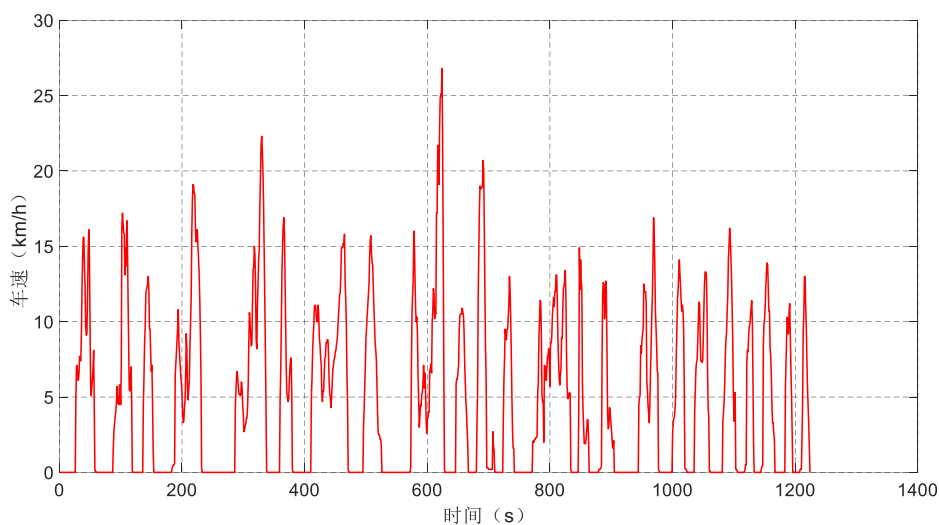


图 6-9 第三类（低速路段）汽车行驶工况

按照相关度从小到大的顺序，从高速路段中选出 1 个时长为 396 秒的运动学片段、从中速路段中选出 3 个运动学片段，从低速路段中选出 8 个运动学片段，之后按照低-中-高的排列方式将它们组合在一起得到的 1208 秒之间的整体道路工况，如下图 6-10 所示。

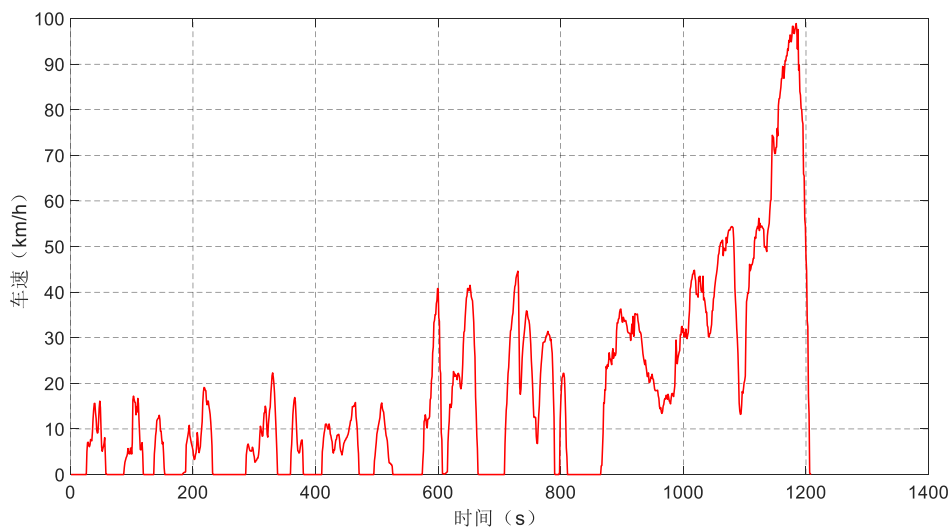


图 6-10 整体路段的汽车行驶工况

6.5 工况分析与评价

6.5.1. 特征参数误差分析

从问题二的 15 个运动学片段参数指标中选择平均速度、速度标准差、平均加速度、加速度标准差、平均减速度、减速度标准差等 8 个指标，并将这 8 个指标与问题二中的怠速时间比、加速时间比、减速时间比和匀速时间比 4 个整体分布参数指标组合在一起，将这 12 个特征参数统称为工况评价指标体系。对于每一个指标，分别计算所构建出的该类型的道路行驶工况的特征值指标与该道路类型的所有运动段这 12 个指标的平均值，然后计算

两者之间的相对误差和绝对误差，误差的计算公式如下所示，得到的结果如表 6-5 所示

$$\delta_i(j)=\frac{|\sigma_0(i)-\sigma_j(i)|}{|\sigma_j(i)|}\times100\%,i=1,2,...,9,j=1,2,3;$$
$$\Delta_i(j)=|\sigma_0(i)-\sigma_j(i)|,i=1,2,...,9,j=1,2,3;$$

式中：i ——评价指标的序号； j ——道路类型；

$\delta_i(j)$ ——第 j 类型道路中，第 i 个评价指标的相对误差；

$\Delta_i(j)$ ——第 j 类型道路中，第 i 个评价指标的绝对误差；

$\sigma_0(i)$ ——构建出的工况的第 i 类评价指标的值；

$\sigma_j(i)$ ——样本中第 j 类道路的所有运动段第 i 类评价指标的平均值值；

依据选出的 12 个评价指标，分别计算出三类道路工况以及整体道路工况与各自样本平均参数之间的相对误差值，结果如下表所示：

表 6-5 道路工况的相对误差

工况曲线	第一类 工况曲线	第二类 工况曲线	第三类 工况曲线	构建的 工况曲线
平均速度	130.35%	9.37%	68.25%	18.34%
平均行驶速度	94.24%	14.84%	59.48%	26.03%
最大速度	203.91%	41.05%	17.65%	203.91%
速度标准差	107.76%	33.43%	48.63%	99.78%
平均加速度	11.11%	4.16%	4.73%	0.79%
加速度标准差	13.57%	12.48%	1.81%	4.56%
平均减速度	6.23%	12.68%	13.37%	2.73%
减速度标准差	23.08%	19.87%	15.11%	43.77%
怠速时间	47.15%	2.83%	33.07%	0.23%
加速时间	28.17%	13.90%	21.76%	0.94%
减速时间	21.59%	2.69%	12.99%	0.73%
匀速时间	23.22%	20.24%	15.85%	1.45%
平均值	59.20%	15.63%	26.06%	33.61%

从上表中可以看出最大速度这项评价指标的误差非常大，这是因为最大速度这个指标所反映的是总体中样本的最大值，并不能够体现出我们构建所得到的道路行驶工况与样本总体之间的相关性，所以不应该将其作为评价指标考虑。从整体上来看，第一类工况的误差较大，而第二类、第三类以及整体的工况误差均较小。这就反映出了本文中所构建的工况更适用于中低速行驶的路况，不太适合用于高速的工况中，从原因上来说，这可能是由于样本数据中高速路段的数据占比较少导致的。下面仅以中速路段的工况作为示例，绘制出第二类工况与实际样本的指标相对误差如下图所示，从图中可以看出，与样本数据相比，第二类工况所有指标的相对误差均小于 40%，绝大部分指标误差位于 20% 以下。所示中速

路段的工况较为合理。同理也可得到第三类工况和整体道路工况也都较为合理，可信度较高，但是第一类（高速路段）工况的效果较差，可以考虑对其进行修正。

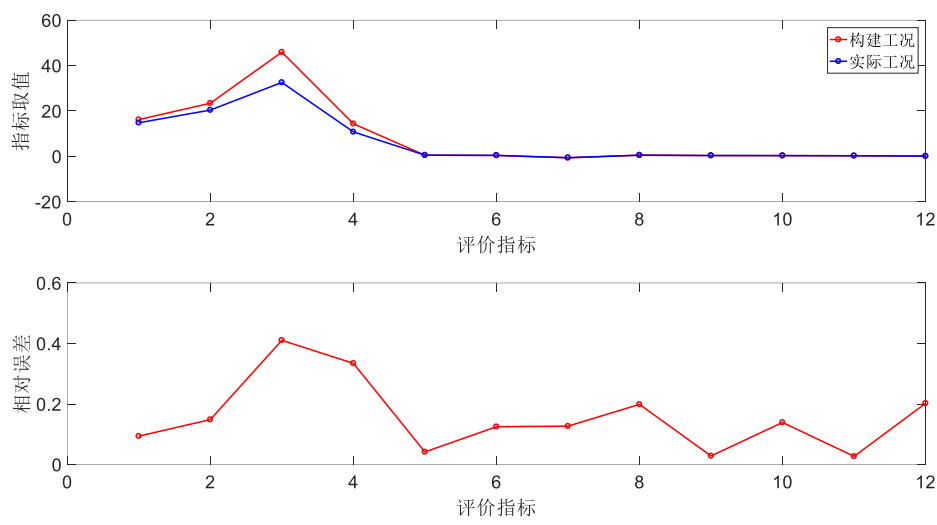


图 6-11 第二类工况和样本的 2 指标误差

6. 5. 2. 速度-加速度联合概率分布

在对道路工况的效果进行评价的之后，常用的方法为用速度—加速度联合概率分布来对构建出的工况进项检验，通过绘制速度—加速度的概率分布图可以很好地反映出所构建的道路工况与采样总体之间的相似程度，并且从分布图中也能够看出两者之间存在差异的地方，从前文中可以初步确定中速路段（第二类）行驶工况效果较好，现在采用速度—加速度联合概率分布图来对其进行验证。

从中速路段工况的所有运动学片段中统计出所有数据的速度与加速度的联合概率分布，如下表 6-6 所示，再从第二类路段的所有运动学片段中统计出其所有数据的速度与加速度的联合概率分布，如下表 6-7 所示，表中横向代表速度，纵向代表加速度。分别将其绘制成图 6-12 和图 6-13，对比分析这两个速度—加速度联合分布图可以看出，中速路段所构建的工况与其采样总体速度—加速度联合分布概率最大的区间一致均在 $0km/h \leq v \leq 10km/h$ 且 $0m/s^2 \leq a < 1m/s^2$ 的区间，其概率分别为 0.6114 和 0.5384。概率值很接近，从概率最大区间来分析可以得出中速路段的工况效果较好，从而验证了上一节说得出的结论。此外，中速路段工况的速度—加速度联合概率分布图与第二类样本的速度—加速度联合概率分布图的分布在形状基本保持一致，差别之处非常少。综上所述，中速路段的行驶工况是合理的，具有一定的可靠性。

表 6-6 中速路段工况的速度—加速度联合概率分布表

	[0,10)	[10,20)	[20,30)	[30,40)	[40,50)	[50,60)	[60,70)	>=70
<-3	0.0024	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[-3,-2]	0.0048	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[-2,-1]	0.0393	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

[-1,0]	0.2043	0.0337	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[0,1]	0.6114	0.0721	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[1,2]	0.0288	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[2,3]	0.0016	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
>3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

表 6-7 第二类样本的速度—加速度联合概率分布表

	[0,10)	[10,20)	[20,30)	[30,40)	[40,50)	[50,60)	[60,70)	>=70
<-3	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[-3,-2]	0.0036	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[-2,-1]	0.0280	0.0037	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
[-1,0]	0.1858	0.0777	0.0035	0.0000	0.0000	0.0000	0.0000	0.0000
[0,1]	0.5384	0.1306	0.0059	0.0000	0.0000	0.0000	0.0000	0.0000
[1,2]	0.0165	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[2,3]	0.0028	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
>3	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

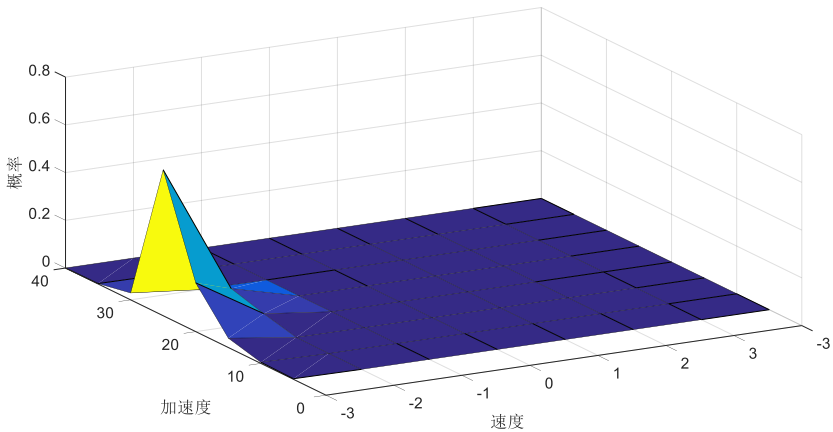


图 6-12 中速路段工况的速度—加速度联合概率分布图

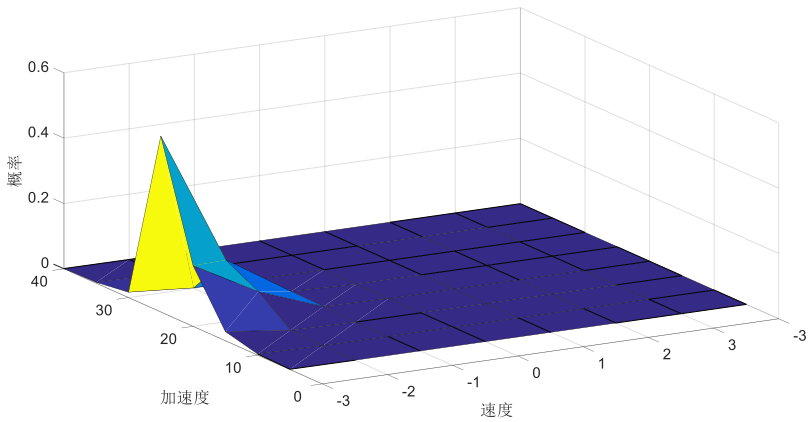


图 6-13 第二类样本的速度—加速度联合概率分布图

同理可以得出高速路段工况的速度—加速度联合概率分布图和第一类样本的速度—

加速度联合概率分布图，从者两个图中我们发现这两者差距较大，其效果不佳。这也从侧面验证了上一小节对于高速路段工况的分析。

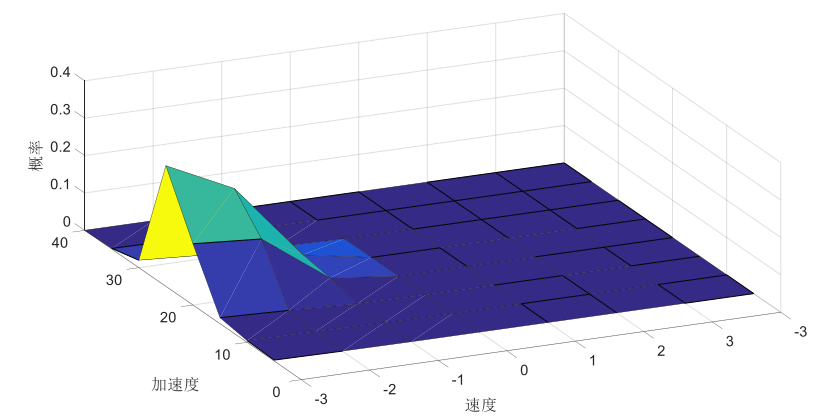


图 6-14 高速路段工况的速度—加速度联合概率分布图

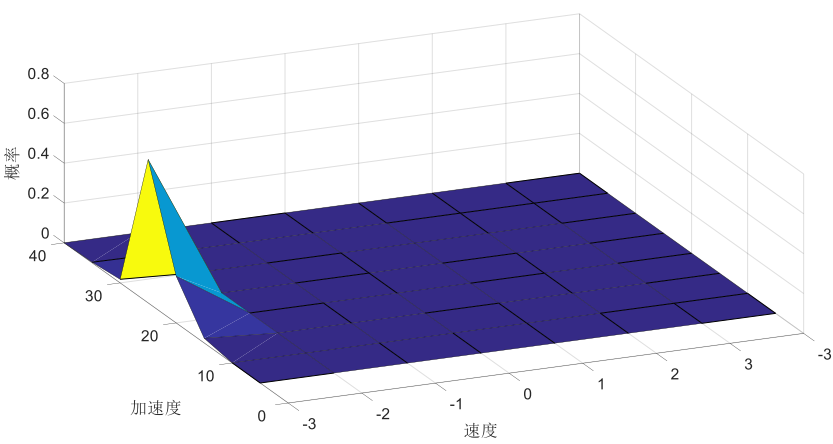


图 6-15 第一类样本的速度—加速度联合概率分布图

七、模型的评价与推广

7.1 模型优缺点

7.1.1. 模型的优点

(1) 问题一中，对于数据进行了时间对齐的方式对于整个文件的数据进行了分割，这样既保证了在进行数据预处理时的效率，又增强了数据的可视化效果；

(2) 问题一中，依据数据的经纬度绘制出了行车的路径轨迹，能够直观的反映出数据采集的地点与路段，有助于直观理解异常数据的种类；

(3) 问题三中，采用主成分分析对于特征指标进行了降维处理，有效较少了数据的运算量，采用聚类分析的方式聚为三类正好对应低中高速；

(4) 采用数据挖掘的方式对于数据进行了预处理，通过对于异常数据的处理而得到质量较高的处理后数据，有利于进行后续的运动学片段划分和行驶工况构建。

7.1.2. 模型的缺点

(1) 问题一中，虽然对于数据进行了大量的预处理操作，但是由于采集到的数据存在有一部分的噪声数据，而这一部分噪声数据会对于后续的运动学片段的提取产生一定的影响，可以通过降噪处理来进一步提高数据的质量；

(2) 在构建道路行驶工况的过程中是采用低、中、高速度路段工况合成的方式，由于低、中、高速并未有一个明确的固定区间，所以使得采用这种方式构建出来的整体道路行驶工况存在着一定的随机性与偶然性；

7.2 模型改进与推广

在问题一中，可以考虑采用指数平滑法对数据做进一步的平滑处理，以此来提高数据的质量，此外也可以对于数据进行去噪和滤波处理。在问题三中构建每一种道路类型的工况时，可以采用从样本运动学片段中随机挑选一定的片段组合出一个工况，再依据评价指标来评判工况的效果，如若效果不好则改变随机组合，直达最后能够得到一个满意的效果为止，即可以联合现代智能算法（例如模拟退火算法）来从样本中搜索出一个满意的局部最优解。

在本文中，问题一对于数据进行的预处理操作使用于绝大多数的数据处理部分，能够将该问题的模型推广到其它需要进行数据挖掘的问题中去，所以从整体上来看，本文所建立的模型具有较好的推广性。

八、参考文献

- [1]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.
- [2]李阿午. 太原市轻型车行驶工况构建与排放特性的研究[D].太原理工大学,2018.
- [3]田宇,朱建军,周博雅.太原市轻型车实际行驶工况构建[J].汽车技术,2018(03):51-55.
- [4]路尧. 中国轻型汽车行驶工况开发[D].北京理工大学,2016.
- [5]司守奎, 孙玺菁. 数学建模算法与应用[M]. 国防工业出版社, 2011.
- [6]石敏. 轻型汽车行驶工况构建的研究[D].天津理工大学,2014.
- [7]黄永青,訾琨,涂先库,杨仁法.宁波市道路汽车行驶工况调查研究[J].武汉理工大学学报(交通科学与工程版),2006(04):720-723.

九、附 录

附录一 问题一代码

● 主程序

```
%% 准备存储空间
clc,clear,close all

filename = {'文件 1','文件 2','文件 3'};
for i = 1:3
    dealFile(filename{i});
    fprintf('-----\n');
end
```

● 文件处理

```
function dealFile(filename)
% 载入数据
filename = '文件 1';
fprintf('正在导入%s...\n',filename);
[~,~,data] = xlsread(filename);
fprintf('导入成功! 一共导入%d 条数据\n',length(data)-1);

% 时间对齐处理
data1 = timeAlign(data);

% 经纬度异常处理
data2 = locatErrorDetct(data1);

% 长期停车不熄火处理
data3 = carStopDtect(data2);

% 堵车怠速处理
data4 = carTrafficDetect(data3);

% GPS 短时丢失处理
data5 = gpsMissDetect(data4);

% 加速度异常处理
data6 = accelerateErrorDetect(data5);

% 速度异常处理
data7 = speedErrorDetect(data6);

% 毛刺异常处理
data8 = burrDetect(data7);

% 选取运动学片段
kinepart = kinePartDetect(data8);
```

```

% 保存运动学片段
save([filename,'运动学片段'],'kinepart');

end

● 时间对齐
function datanew = timeAlign(data)
% 计算有记录 GPS 的时间
record_time = date2second2(data(2:end,1))';
% 计算起止点
time_start = date2second(data{2,1});
time_end = date2second(data{end,1});
time = time_start:time_end;
% 时间对齐, 查找对齐序号
[~,col] = ismember(record_time,time);
errodataNum = length(time) - length(col);
text1 = sprintf('因 GPS 设备断电或者异常导致数据缺失了%d 条数据, 正在补
齐...\n',errodataNum);
fprintf(text1);
% 构造修复后元胞
data_deal = cell(length(time)+1,size(data,2));
data_deal(:) = {nan};
data_deal([1,col+1],:) = data;
datanew = data_deal;
% 时间还原
datanew(2:end,1) = second2date(data{2,1},time);
fprintf('时间对齐完成! \n');
end

● 经纬度异常处理
function datanew = locatErroDetct(data)
datanew = data;
% 提取经纬度数据
data_temp =sum( cell2mat(datanew(2:end,6:7)),2);
% 找到全为 0 的数据, 将其剔除 (置为 nan)
erro_index = find(data_temp==0);
text1 = sprintf('找到经纬度异常数据%d 条, 正在剔除...\n',length(erro_index));
fprintf(text1);
datanew(erro_index+1,2:end) = {nan};
fprintf('经纬度异常处理完成! \n');
end

● Gps 瞬时异常处理
function datanew = gpsMissDetect(data)
datanew = data;
% 提取速度
data_temp = cell2mat(datanew(2:end,2));
% 短时 GPS 丢失修复 (2s 均值插补法)
fix_num = 0; % 记录处理的数据个数
for i = 2:length(data_temp)-1
    if ~isnan(data_temp(i-1)) && isnan(data_temp(i)) &&
~isnan(data_temp(i+1))
        data_temp(i) = mean([data_temp(i-1),data_temp(i+1)]);
    end
end

```

```

        fix_num = fix_num + 1;
    end
end

for i = 2:length(data_temp)+1
    datanew{i,2} = data_temp(i-1);
end
fprintf('找到 GPS 短时丢失数据（丢失 1s）%d 条，已通过插值修复！\n', fix_num);
end

```

● 毛刺处理

```

function datanew = burrDetect(data)
datanew = data;
% 提取速度
data_temp = cell2mat(datanew(2:end,2));
% 找出速度大于 0 的数据
v_index = find(data_temp(:,1)>0);
% 找出连续的片段
deal_index1 = is_continue(v_index);
% 统计连续片段长度
for i = 1:length(deal_index1)
    index1_length(i) = length(deal_index1{i});
end
% 找出长度小于 4s 的数据段
deal_index2 = find(index1_length < 4);
% 异常数据记录
erro_part = length(deal_index2);
erro_num = sum(index1_length(deal_index2));
text1 = sprintf('找到毛刺片段%d 个，共包含%d 条数据，正在怠速处  
理...\n',erro_part,erro_num);
fprintf(text1);
% 对应连续数据段进行怠速处理（数据抹成 0）
for i = 1:length(deal_index2)
    datanew(cell2mat(deal_index1(deal_index2(i)))+1,2) = {0};
end
fprintf('毛刺异常处理完成！\n');
end

```

● 堵车异常处理

```

function datanew = carTrafficDetect(data)
datanew = data;
% 提取速度
data_temp = cell2mat(datanew(2:end,2));
% 找出速度小于 10km/h 的数据
v_index = find(data_temp(:,1)< 10);
% 找出连续的片段
deal_index1 = is_continue(v_index);
% 统计连续片段长度
for i = 1:length(deal_index1)
    index1_length(i) = length(deal_index1{i});
end

```

```

% 找出超过 180s 的数据段
deal_index2 = find(index1_length > 180);
% 异常数据记录
erro_part = length(deal_index2);
erro_num = sum(index1_length(deal_index2));
text1 = sprintf('找到堵车造成的持续低速片段%d 个，共包含%d 条数据，正在剔除...\n',erro_part,erro_num);
fprintf(text1);
% 对应连续数据段进行怠速处理（数据抹成 0）
for i = 1:length(deal_index2)
    datanew(cell2mat(deal_index1(deal_index2(i))+1,2) = {0};
end
fprintf('堵车怠速异常处理完成！\n');
end

```

● 片段提取

```

function kinepart = kinePartDetect(data)
datanew = data;
% 提取速度
data_temp = cell2mat(datanew(2:end,2));
fprintf('正在查找运动学片段...\n')
% 找出速度大于 0 的数据
v_index = find(data_temp(:,1)>0);
% 找出连续的片段
deal_index1 = is_continue(v_index);
% 统计连续片段长度
for i = 1:length(deal_index1)
    index1_length(i) = length(deal_index1{i});
end
% 找出超过 10s 的数据段
deal_index2 = deal_index1(index1_length > 10,1);
% 找出前后都为 0 的，符合要求的运动片段
is_part = zeros(length(deal_index2),1);
for i = 1:length(deal_index2)
    part_start = deal_index2{i,1}(1)-1;
    part_end = deal_index2{i,1}(end)+1;
    if (data_temp(part_start)==0) && (data_temp(part_end)==0)
        is_part(i) = 1;
    end
end
deal_index3 = deal_index2(is_part==1,1);
kinepart = cell(length(deal_index3),1);
% 找出前面 180s 范围内为 0 的那一段
for j = 1:length(deal_index3)
    for k = 1:180
        if deal_index3{j,1}(1)-k > 1 %不可以超出索引
            if data_temp(deal_index3{j,1}(1)-k) ~= 0
                break;
            end
        else

```

```

        break;
    end
end
now_index = (deal_index3{j,1}(1)-k+1:deal_index3{j,1}(end)+1);
kinepart{j,1} = datanew([1,now_index+1,:]);
end
fprintf('一共找到%d个运动学片段! \n',length(deal_index3));
end

```

附录二 问题二代码

● 主程序

%% 准备存储空间

```
clc,clear,close all
```

```
filename = {'文件 1','文件 2','文件 3'};
```

```
for i = 1:length(filename)
```

```
    dealFile(filename{i});
```

```
    fprintf('-----\n');
```

```
end
```

```
fprintf('所有运动学片段特征参数、总体分布特征全部分析完毕! \n');
```

● 特征计算

```
function [Kinefeature,Distfeature] = calFeature(kinepart)
```

% 定义存储特征值的结构体

```
Kinefeature = struct(...
```

```

    'run_time',{},...           % 运行时间
    'neglect_time',{},...       % 怠速时间
    'accelerate_time',{},...    % 加速时间
    'slowdown_time',{},...      % 减速时间
    'unispeed_time',{},...      % 匀速时间
    'av_speed',{},...           % 平均速度
    'av_runspeed',{},...        % 平均行驶速度
    'max_speed',{},...          % 最大速度
    'speed_std',{},...          % 速度标准差
    'av_accelerate',{},...      % 平均加速度
    'max_accelerate',{},...     % 最大加速度
    'accelerate_std',{},...     % 加速度标准差
    'av_slowdown',{},...        % 平均减速度
    'min_slowdown',{},...       % 最小减速度
    'slowdown_std',{},...       % 减速度标准差

```

% 定义分存储分布特征的结构体

```
Distfeature = struct(...
```

```

    'neglect_time',{},...       % 怠速时间比
    'accelerate_time',{},...     % 加速时间比
    'slowdown_time',{},...      % 减速时间比
    'unispeed_time',{},...      % 匀速时间比
    'v0_10_time',{},...         % 0-10km 速度段比例
    'v10_20_time',{},...        % 10-20km 速度段比例

```

```

'v20_30_time',{},{},...      % 20-30km 速度段比例
'v30_40_time',{},{},...      % 30-40km 速度段比例
'v40_50_time',{},{},...      % 40-50km 速度段比例
'v50_60_time',{},{},...      % 50-60km 速度段比例
'v60_70_time',{},{},...      % 60-70km 速度段比例
'v70_inf_time',{},{},...      % 70-infkm 速度段比例

m = length(kinepart);

for i = 1:m
    % 获取运动学片段
    data = kinepart{i};
    % 提取速度
    data_temp = cell2mat(data(2:end,2));
    v_delta = (data_temp(2:end) - data_temp(1:end-1)); % 计算速度差

    % 计算运动段特征值
    % 时间特征值 单位: s
    Kinefeature(i).run_time = length( data_temp)-1; % 计算运行时间
    Kinefeature(i).neglect_time = sum(v_delta == 0); % 计算怠速时间
    Kinefeature(i).accelerate_time = sum( v_delta > 0.1*3600/1000); % 计算加速时间
    Kinefeature(i).slowdown_time = sum( v_delta < -0.1*3600/1000); % 计算减速时间
    Kinefeature(i).unispeed_time = sum( v_delta >= -0.1*3600/1000 &
v_delta <= 0.1*3600/1000 & v_delta ~= 0); % 计算匀速时间

    % 速度特征值 单位: km/h
    Kinefeature(i).av_speed = mean(data_temp); % 计算平均速度
    Kinefeature(i).av_runspeed = mean(data_temp(data_temp~=0)); % 计算平均行驶速度
    Kinefeature(i).max_speed = max(data_temp); % 计算最大速度
    Kinefeature(i).speed_std = std(data_temp); % 计算速度标准差

    % 加速度特征值 单位: m/s^2
    v_delta2 = v_delta*1000/3600; %单位换算
    Kinefeature(i).av_accelerate = mean(v_delta2(v_delta2 > 0.1)); % 计算平均加速度
    Kinefeature(i).max_accelerate = max(v_delta2(v_delta2 > 0.1)); % 计算最大加速度
    Kinefeature(i).accelerate_std = std(v_delta2(v_delta2 > 0.1)); % 计算加速度标准差

    Kinefeature(i).av_slowdown = mean(v_delta2(v_delta2 < -0.1)); % 计算平均减速度
    Kinefeature(i).min_slowdown = max(v_delta2(v_delta2 < -0.1)); % 计算最小减速度
    Kinefeature(i).slowdown_std = std(v_delta2(v_delta2 < -0.1)); % 计算减速度标准差

```



```

% 计算运动段总体分布特征
Distfeature(i).neglect_time = sum(v_delta == 0)/(length( data_temp)-1); % 计算怠速时间比
Distfeature(i).accelerate_time = sum( v_delta > 0.1*3600/1000)/(length( data_temp)-1); % 计算加速时间比
Distfeature(i).slowdown_time = sum( v_delta < - 0.1*3600/1000)/(length( data_temp)-1); % 计算减速时间比
Distfeature(i).unispeed_time = sum( v_delta >= -0.1*3600/1000 & v_delta <= 0.1*3600/1000 & v_delta ~= 0)...
/(length( data_temp)-1); % 计算减速时间比
Distfeature(i).v0_10_time = sum(data_temp>0 & data_temp<=10)/(sum( data_temp>0)); % 0-10km 速度段比例
Distfeature(i).v10_20_time = sum(data_temp>10 & data_temp<=20)/(sum( data_temp>0)); % 10-20km 速度段比例
Distfeature(i).v20_30_time = sum(data_temp>20 & data_temp<=30)/(sum( data_temp>0)); % 20-30km 速度段比例
Distfeature(i).v30_40_time = sum(data_temp>30 & data_temp<=40)/(sum( data_temp>0)); % 30-40km 速度段比例
Distfeature(i).v40_50_time = sum(data_temp>40 & data_temp<=50)/(sum( data_temp>0)); % 40-50km 速度段比例
Distfeature(i).v50_60_time = sum(data_temp>50 & data_temp<=60)/(sum( data_temp>0)); % 50-60km 速度段比例
Distfeature(i).v60_70_time = sum(data_temp>60 & data_temp<=70)/(sum( data_temp>0)); % 60-70km 速度段比例
Distfeature(i).v70_inf_time = sum(data_temp>70)/(sum( data_temp)-1); % 70-infkm 速度段比例
end

end

```

附录三 问题三代码

```

● 主程序
% 准备存储空间
clc,clear,close all

% 载入所有运动学片段、运动学特征、整体分布特征
[kine_part_all,kine_feature_all,dist_feature_all] = loadAll();

% 运动学特征 主成分分析降维
[kine_feature_dim,mylatent] = mypca(kine_feature_all,0.85);

% 运动学特征聚类分析
%group = my_k_means(kine_feature_dim); % 得到聚类到一起的序号
%save('group','group');
load group
% 画碎石图
plot(mylatent(:,1),'bo-','linewidth',2)

```

```

xlabel('主成分序号'),ylabel('方差')
grid on
set(gca,'xticklabel',{1:15},'GridLineStyle','--',
'','GridColor','k','GridAlpha',1,'fontsize',24);

%% 工况分析

% 按聚类小组的均值相关性(按相关性降序排列片段索引号) 构建小组的工况曲线
[corr_index1,choose_num1,kine1]=
calWorkCurve(group{1},dist_feature_all,kine_feature_all,kine_part_all);
[corr_index2,choose_num2,kine2]=
calWorkCurve(group{2},dist_feature_all,kine_feature_all,kine_part_all);
[corr_index3,choose_num3,kine3]=
calWorkCurve(group{3},dist_feature_all,kine_feature_all,kine_part_all);

plot(kine3,'r-','linewidth',2)
xlabel('时间 (s)'),ylabel('车速 (km/h)')
grid on
set(gca,'GridLineStyle','--','GridColor','k','GridAlpha',1,'fontsize',24);

% 合并运动学特征和总体分布特征构成评价指标
judge_feature_all =
[kine_feature_all(:,[6:10,12,13,15]),dist_feature_all];

% 计算整体分布特征
dist_kineof3 = {kine1,kine2,kine3};
[Kinefeature,Distfeature] = calFeature2(dist_kineof3);

data = [...
    Distfeature.neglect_time;...
    Distfeature.accelerate_time;...
    Distfeature.slowdown_time;...
    Distfeature.unispeed_time
];
% 画运行模式比例柱状图
bar(data)
set(gca,'xticklabel',{'怠速模式','加速模式','减速模式','匀速模式'},
'','fontsize',24);
xlabel('行驶模式'),ylabel('时间比例/%')
legend({'第一类','第二类','第三类'})

% 计算高中低速度工况下的比例, 确定挑选长度
speed_0_30 = 0;
speed_30_50 = 0;
speed_50_inf = 0;
for i = 1:length(kine_part_all)
    for j = 2:length(kine_part_all{i})
        if kine_part_all{i}{j,2} > 0 && kine_part_all{i}{j,2} <= 30

```

```

        speed_0_30 = speed_0_30 +1;
    elseif kine_part_all{i}{j,2} > 30 && kine_part_all{i}{j,2} <= 50
        speed_30_50 = speed_30_50 +1;
    else
        speed_50_inf = speed_50_inf +1;
    end
end
end

% 得到挑选长度区间
time_low = (speed_0_30/(speed_0_30+speed_30_50+speed_50_inf))*[1200,1300];
time_middle =
(speed_30_50/(speed_0_30+speed_30_50+speed_50_inf))*[1200,1300];
time_high =
(speed_50_inf/(speed_0_30+speed_30_50+speed_50_inf))*[1200,1300];

% 获取片段
is_choose = choosePart2(corr_index1(:,3),time_high(1),time_high(2));
high_part = kine_part_all(corr_index1(is_choose==1,1));

is_choose = choosePart2(corr_index2(:,3),time_middle(1),time_middle(2));
middle_part = kine_part_all(corr_index2(is_choose==1,1));

is_choose = choosePart2(corr_index3(:,3),time_low(1),time_low(2));
low_part = kine_part_all(corr_index3(is_choose==1,1));

% 片段整合
choose_part = {low_part,middle_part,high_part};
newkine = [];
for i = 1:3
    for j = 1:length(choose_part{i})
        newkine_temp = [choose_part{i}{j,1}{2:end,2}];
        newkine = [newkine,newkine_temp];
    end
end

plot(newkine,'r-','linewidth',2)
xlabel('时间 (s)'),ylabel('车速 (km/h)')
grid on
set(gca,'GridLineStyle','--','GridColor','k','GridAlpha',1,'fontSize',24);

% 计算新的联合特征参数
% 计算构建工况的所有特征
[Kinefeaturenew,Distfeaturenew] = calFeature2({kine2});
Kinefeaturenew = cell2mat(struct2cell(Kinefeaturenew));
Distfeaturenew=cell2mat(struct2cell(Distfeaturenew));

build_feature = [Kinefeaturenew(:,[6:10,12,13,15]),Distfeaturenew(:,1:4)];
all_feature =
mean([kine_feature_all(:,[6:10,12,13,15]),dist_feature_all(:,1:4)]);

```

```
erro_absolute = abs(all_feature - build_feature);
erro_relative = abs(erro_absolute./all_feature);
subplot(211)
plot(1:12,build_feature,'ro-',1:12,all_feature,'bo-','linewidth',2);
xlabel('评价指标'),ylabel('指标取值'),set(gca,'fontsize',24)
legend({'构建工况','实际工况'})
subplot(212)
plot(1:12,erro_relative,'ro-','linewidth',2);
xlabel('评价指标'),ylabel('相对误差'),set(gca,'fontsize',24)
```