**摘要**

最后再写（约2页）

[问题背景与问题重述 1](#_Toc1225478369)

[1.1 问题背景 1](#_Toc99289406)

[1.2 问题重述 2](#_Toc162252923)

[基本假设与符号说明 2](#_Toc1828128818)

[问题一的分析与建模 2](#_Toc1312506497)

[问题二的分析建模 2](#_Toc344673095)

[问题三的分析建模 3](#_Toc1157311706)

[3.1 特征参数的选取 3](#_Toc1178451863)

[3.2 运动学片段的参数计算 3](#_Toc2146268807)

[3.3 聚类分析 3](#_Toc1057020590)

[3.4 工况的合成 4](#_Toc1360328146)

[模型评估 4](#_Toc924243860)

[4.1 指标计算 4](#_Toc1017336269)

[4.2 合理性分析 4](#_Toc105875669)

[参考文献 4](#_Toc1335909167)

[附录 4](#_Toc683840384)

# 1 问题背景与问题重述

## 1.1 问题背景

汽车行驶工况又称为车辆测试循环，是描述典型车辆行驶的速度-时间曲线。它可以用于确定车辆污染物排放量及燃油消耗量，体现汽车在道路行驶中的燃料经济性、排放等运动学特征，是汽车工业一项核心技术。目前，我国乘用车行驶工况采用的是欧洲的NEDC，但由于我国的交通状况、行驶状况与道路状况等因素都与欧盟国家有很大不同，因此NEDC工况与我国城市汽车的实际行驶数据偏差较大。另一方面，我国地域辽阔，不同城市之间的交通状况、发展程度与气候条件等都存在着差异，这也导致各个城市的汽车行驶工况也不尽相同。因此，基于各个城市自身的汽车行驶数据制定反映我国实际道路行驶状况的汽车行驶工况，显得越来越重要。

## 1.2 问题重述

在上述背景下，题目给出了某城市同一辆测试汽车在不同时间段采集的实际行驶数据，要求利用该数据解决以下几个问题：

问题一：由于数据采集过程中存在着不可避免的实际问题，原始数据中会包含一些不良值。要求通过合理的方法对异常数据进行预处理，并给出处理后的数据记录数。

问题二：我们将车辆从一个怠速开始到下一个怠速开始之间的运动定义为运动学片段。题目要求设计合理的方法，将处理后的数据划分为多个运动学片段，并给出每个数据文件得到的运动学片段数。

问题三：基于上面处理好的数据，采用科学有效的方法构建一条能体现采集数据源行驶特征的汽车行驶工况曲线。同时使用合理的评估体系对行驶工况曲线进行评估，说明其合理性。

# 2 基本假设与符号说明

假设普通轿车在一般情况下最大的加速度为4m/s2，紧急刹车时的最大减速度为8m/s2。

假设当汽车车速小于10km/h时，就认为是怠速状态。当怠速状态时间超过180秒就被认为是异常情况，怠速时间最长可按180秒处理。

我们定义汽车加速度大于0.1m/s2的连续过程为加速阶段，汽车加速度小于-0.1m/s2的连续过程为减速阶段。

假设采样频率为1Hz

# 3 问题一的分析与建模

题目给出了某城市同一辆测试汽车在不同时间段采集的实际行驶数据，覆盖了工作日与节假日、高峰期与非高峰期等多时段的数据，保证了测试数据的可靠性。但由于多种主客观因素影响，原始测试数据中可能包含多种不良数据。问题一就是要求通过多种合理的处理方法对这些异常数据进行识别与预处理。

原始的测试数据包含3个数据文件，其中数据文件1的原始数据记录数为xx条，数据文件2的原始数据记录数为xx条，数据文件3的原始数据记录数为xx条。

## 3.1 加减速度异常的数据处理

从上文的基本假设中我们可知，在一般情况下普通轿车从0加速至100km/h的时间应大于7秒，即最大加速度约为4m/s2，而轿车紧急刹车的最大减速度约为8m/s2。在实际道路行驶过程中可能会出现车辆急加速与急刹车的情况，这种情况下往往会导致车辆加速度过大。而超出最大加减速度阈值的不良数据显然会对我们构建行驶工况曲线造成影响，因此需要对这些不良数据进行处理。

在题目给出的原始采集数据中也存在这种加减速度异常的数据值。如下图所示，（放一张折线图说明异常情况）

对于加减速度异常的原始数据，我们首先对异常数据记录进行剔除，消除离群点数据对于构建行驶工况曲线带来的影响。删除后造成的数据缺失问题在后文3.2中进行进一步的处理。

经处理后，数据文件1剔除的异常记录数为xxx条，总记录数为xxx条。数据文件3剔除的异常记录数为xxx条，总记录数为xxx条。数据文件3剔除的异常记录数为xxx条，总记录数为xxx条。

## 3.2 时间不连续的缺失数据处理

汽车行驶数据通过车载的终端采集设备直接采集，采集过程中可能会出现采集设备异常等问题，导致数据丢失。此外，现实路况也可能给数据采集带来一定的影响，例如在汽车过隧道时，会导致GPS信号丢失，无法进行正常的数据采集。这些问题造成了提供的数据中时间不连续，部分数据丢失的问题。

在对缺失数据段进行观察后，我们设置允许最长时间不连续的阈值为60秒。连续缺失数据未超过阈值的数据我们通过插值进行填充处理，而缺失数据超过60秒的数据我们则进行删除。

填充的方法待完善。

在构建行驶工况曲线的过程中，我们选择基于运动学片段进行构建。考虑到缺失数据过多会对整个运动学片段带来影响，因此对于超过阈值需要删除的异常数据，（如下图所示）我们选择删除该段缺失数据所在的**整个运动学片段**的数据记录。

经处理后，数据文件1剔除的异常记录数为xxx条，总记录数为xxx条。数据文件3剔除的异常记录数为xxx条，总记录数为xxx条。数据文件3剔除的异常记录数为xxx条，总记录数为xxx条。

## 3.3 怠速异常的数据处理

由于汽车行驶工况受交通状况、道路状况、行驶状况等多种主客观因素影响，在数据采集的过程中往往会出现多种类型的行驶数据。我们的原始数据覆盖了工作日与节假日、高峰期与非高峰期等多时段的数据，这其中也包含了几种怠速异常的不良数据。

一种情况如下图所示，数据中存在着一段长时间内行驶速度均为0km/h的怠速异常情况，在现实中长期停车熄火等人，或者停车熄火了但仍在采集数据等都会造成这种异常情况的发生。从上文的基本假设中我们可知，怠速时间超过180秒被认为是异常情况，怠速时间最长可按180秒进行处理。因此对于这类异常我们的处理方法是识别出整段怠速片段，删除多余的异常数据，只保留180秒（即最长怠速时间）内的数据。

另外一种异常情况与实际较为相关。在现实生活中，高峰期或节假日时期的路况一般较为拥堵，极易出现长时间堵车的情况。这时汽车一般处于断断续续行驶且只能低速行驶的情况。这对应于数据中存在着一段长时间行驶速度不为0但均小于10km/h的异常情况。对于这类不良数据，我们的处理方法是将这类数据按照怠速情况进行处理。若连续低速的数据记录超过180秒（说明是**长时间**堵车）则将其速度记录全部置为0km/h，认为这是怠速状况，并按照上述怠速异常的方法进行处理。若连续低速的数据记录未超过180秒，则认为是正常情况，保留原始的速度记录。

经处理后，数据文件1剔除的异常记录数为xxx条，总记录数为xxx条。数据文件3剔除的异常记录数为xxx条，总记录数为xxx条。数据文件3剔除的异常记录数为xxx条，总记录数为xxx条。

# 4 问题二的分析建模

在现有的构建行驶工况曲线的方法中，基于运动学片段进行构建是最常用的方法之一。在车辆行驶的过程中，会出现频繁的启动、加速、减速等过程，运动学片段就是指车辆从一个怠速状态开始至下一个怠速状态开始之间的运动过程。如下图所示便是一个运动学片段的定义说明。



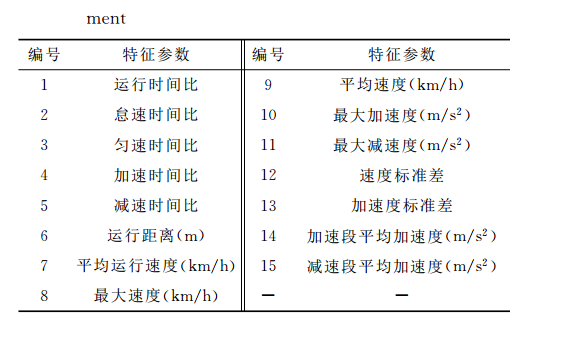
在经过问题一的预处理后我们得到了一份处理后的数据源，下图展示了处理后的部分数据，图中可以清晰地看出在这段数据中存在着多个运动学片段。我们基于上述运动学片段定义对数据进行划分。自一个怠速状态开始至下一个怠速状态开始被认为是一个运动学片段，图中xx段即是一个片段。在问题一中我们也提到，对于缺失数据超过阈值的异常数据（如图x中xx段所示），我们选择删除这段异常数据所在的整个运动学片段。

按上述方法划分后，最终数据文件1我们得到了xx个运动学片段，数据文件2我们得到了xx个运动学片段，数据文件3我们得到了xx个运动学片段。

# 5 问题三的分析建模

## 5.1 特征参数的选取

需要提前设计好特征参数



## 5.2 运动学片段的参数计算

划分出每个片段时，需要对每个片段计算相应特征参数

对特征参数进行标准化

主成分分析，选取关键参数

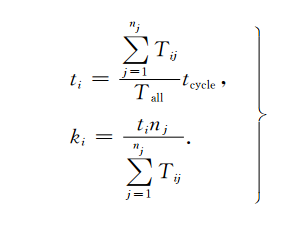
## 5.3 聚类分析

依据关键特征参数进行聚类分析

可考虑采用多种聚类方法

## 3.4 工况的合成

根据聚类结果，合成工况



# 6 模型评估

## 6.1 指标计算

对于我们合成的工况，计算各个运动特征值，如平均速度等

对于处理后的数据集全体进行计算各个特征值

较为简单

## 6.2 合理性分析

看4.1的结果是否接近了。

吹一下不合理的原因

与其他标准比较

改进的思路等

# 参考文献

# 附录