摘要：

汽车行驶工况可以用于确定车辆污染物排放量及燃油消耗量，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，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。

本文根据三组某城市汽车行驶数据，构建了一条时长1201s的汽车行驶工况曲线，较好地反映了所使用的数据源的汽车运动特征。本文首先以多种方式对脏数据进行处理，包括清除和修正等，从三个文件中分别得到了131174，108077，106737条合理的数据。进而将原始数据分别划分为1181，725，649个运动学片段，并计算每个片段的特征参数，得到了一组特征参数矩阵。之后本文对该矩阵使用Min-Max方法进行了归一化处理，并进行了主成分分析，保留了三个主成分，使用K均值聚类算法进行聚类分析，得到了四种类型的运动学片段，每种类型分别有715、616、482、742条片段。然后根据各类片段在整个数据集中所占的时间比例，合成了一条时长1201s的行驶工况。最后本文对该结果进行了验证和分析，认为其能较好地体现出该汽车的运动特征。

问题一要求对原始数据进行清洗。对于加速度异常或者时间不连续的不良数据，当其出现异常持续时间小于30s时，本文采用线性插值法进行了修正，当其出现异常时间超过30s时，本文则将其所属的整个运动学片段内的所有数据删除。对于间歇性怠速或低速行驶超过60s的数据，本文将其速度重置为0km/h，认为这是怠速情况。对于长时间怠速的数据，本文只保留180s的怠速数据。最终数据文件1剔除的异常记录数为54551条，总记录数为131174条。数据文件2剔除的异常记录数为37748条，总记录数为108077条。数据文件3剔除的异常记录数为58177条，总记录数为106737条。

问题二要求切分运动学片段。本文将自一个怠速状态开始至下一个怠速状态开始的片段视为一个运动学片段，据此进行切分，最终数据文件1得到了1181个运动学片段，数据文件2得到了725个运动学片段，数据文件3得到了649个运动学片段。

问题三要求构建一条时长1200~1300s的，能体现参与数据采集汽车行驶特征的汽车行驶工况曲线。本文对每一个片段，计算了包括平均速度、平均行驶速度、平均加速度等9个特征参数。然后使用Min-Max归一化方法对每个参数进行标准化处理，之后为避免信息重叠，进行了主成分分析，保留了三个主成分进行K均值聚类，在聚类前本文使用“手肘法”确定K值为4，最终得到了四组运动学片段，每组依次有715、616、482、742条样本。接着本文根据各类片段在整个数据集中所占的时间比例，合成了一条时长1201s的行驶工况。最后使用卡方校验等方式对所得结果进行了校验，认为其能较好地反映出该汽车的形式特征。

**关键字**：运动学片段，行驶工况，汽车行驶特征，K-Means聚类