

**《机器学习》**

**实验报告**

**系统名称：** 运输车辆安全驾驶行为的分析

**班 级：** 软件Q164班

**组 长：** 梁键兴

**组 员：** 李远树

雷钧光

**指导老师：** 秦胜君

**日 期：** 2019 年 5 月 24 日

**评阅成绩：**

**教师签名：**

**附件：**

**小组成员任务清单**

|  |  |  |
| --- | --- | --- |
| **组员** | **任务** | **是否组长** |
| 梁键兴 | 调用高德地图的接口，申请key，观察并了解所需数据，第三问思路分析 | 是 |
| 李远树 | 具体实现和代码编写，分析数据，构建模型，整体思路分析 | 否 |
| 雷钧光 | 查找资料，观察并了解所需数据 | 否 |
|  |  |  |
|  |  |  |
|  |  |  |

**小组成员设计成绩评价**

|  |  |  |
| --- | --- | --- |
| **组员** | **组长对组员的简单评语** | **评价** |
| 梁键兴 | 在此作业中付出的时间少，没有做到组长的职责 | 良 |
| 李远树 | 表现优秀 | 优 |
| 雷钧光 | 中规中矩 | 良 |
|  |  |  |
|  |  |  |
|  |  |  |

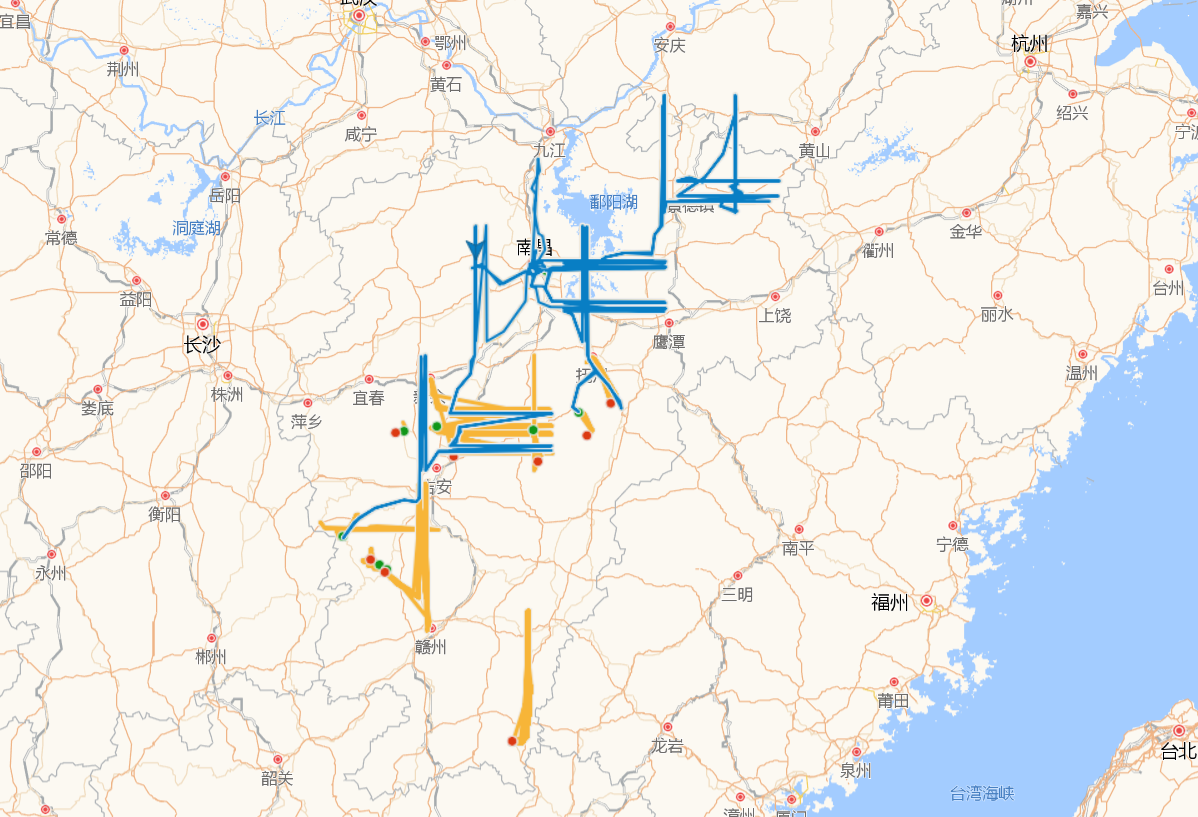
**注：**“评价”主要以“优”、“良”、“中”、“及格”、“不及格”填写。其中“优”不得超过2名。

1. **利用附件 1 所给数据，提取并分析车辆的运输路线以及其在运输过程中的速度、加速度等行车状态。提交附表中10辆车每辆车每条线路在经纬度坐标系下的运输线路图及对应的行车里程、平均行车速度、急加速急减速情况：**

**数据去噪条件**：连续两行或者多行中经纬度都不相同、ACC状态必须为1、GPS 速度需要大于0。只有这三个条件都符合要求，则认为汽车在行驶状态，则取数据表中的lng、lat、location\_time、gps\_speed这四列数据。

**车辆行驶路线图可视化：**

1. 调用高德地图的接口，在网上申请注册后得到一个基本的path.html文件。
2. 提取数据：使用python提取数据并写入到一个记事本文件，再把记事本中的内容手动复制到path.html文件中的经纬度参数中。(代码请参考show.py)
3. 把path.html文件内容复制到<http://www.5axxw.com/tools/web/web_run.html>中运行，便可得到附表中十辆车的运输路线图。结果如下图：



**行车里程**：获取lng、lat这两列经纬度，算出两点之间的距离，然后累计相加，最后得出行车里程。

**两个经纬度之间的平均行车速度：**利用lng、lat这两列算出距离km，location\_time列算出时间间隔h，再利用公式km/h得出每段路程的平均速度。

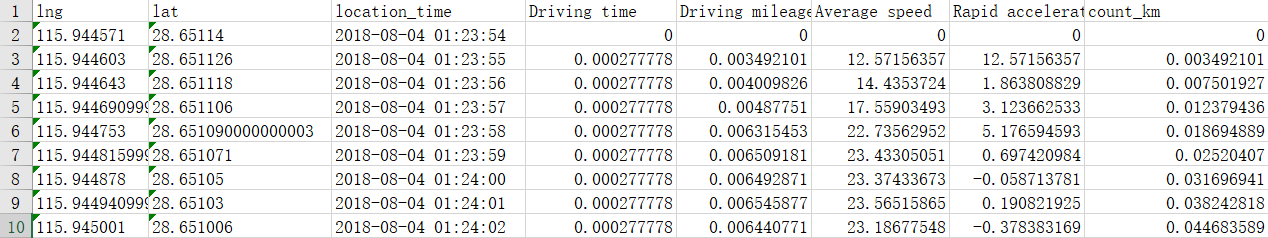
**总路程平均行车速度**：由location\_time列获取总开车时间，利用公式（行车里程/总开车时间）可得出

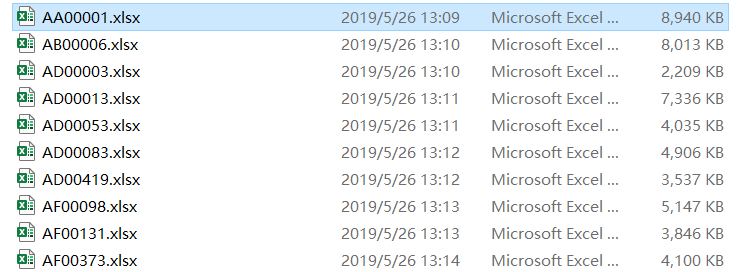
**急加速急加速**：由两个经纬度之间的平均行车速度和时间间隔算出加速度，公式为a=dv/dt;计算出来的加速度a单位为km/s^2,意义为汽车每秒钟速度变化某个公里数，其中正数部分代表加速，负数部分代表减速。

最后把10辆车的经过路线的每一个经纬度的坐标、时间与车辆的对应的行车里程、平均行车速度、急加速急减速情况数据存放在对应的excel表格中，其中每列的含义如下：

|  |  |
| --- | --- |
| lng | 经度 |
| lat | 纬度 |
| location\_time | 采集时间 |
| Driving time | 上一个经纬度到此经纬度行驶时间 单位：小时 |
| Driving mileage | 上一个经纬度到此经纬度行驶路程 单位：千米 |
| Average speed | 平均速度 单位：km/h |
| Rapid acceleration | 急加速急减速情况 单位：km/s^2 |
| count\_km | 目前行驶路程 |

展示编号为AA00001车辆的部分数据，其余具体数据请参考以下对应的文件。





**小结**：在第一个问题中实现了行车路线的可视化，但是行车过程中车辆的速度变化，路程公里数变化还无法在行车线路图上做得数据可视化，具体提取和计算数据的代码请参考shuju.py文件。

**（2）利用附件 1 所给数据，挖掘每辆运输车辆的不良驾驶行为，建立行车安全的评价模型，并给出评价结果。**

**思路：**

在车辆运输过程中，不良驾驶行为主要包括疲劳驾驶、急加速、急减速、怠速预热、 超长怠速、熄火滑行、超速、急变道等。

针对以上运输车辆的不良驾驶行为，给出不同不良驾驶行为的判别标准，行车安全评价模型如下：

**疲劳驾驶**：连续行车时间超过4小时。

提取数据思路：若某一行acc\_state列值为1并且gps\_speed列数值大于0，则认为汽车开始启动，继续扫描数据表，直到寻找到一行gps\_speed列的数值为0，则认为汽车已经处于停止状态，再根据location\_time列由两个数据获取时间间隔，判断是否属于疲劳驾驶。

**急加速、急减速**：每两个经纬度间汽车的加速度达到或者超过20km/s^2。两个经纬度间汽车的加速度具体实现思路已经在第一问中给出并且已经实现。

**怠速预热、超长怠速**：汽车的ACC状态为1，但汽车的GPS速度、经纬度无变化，这个时间间隔超过3分钟。

**熄火滑行**：汽车的ACC状态为0，但GPS速度超过0km/h。

**超速**：因为汽车在每一种道路上的限制速度不一样，而只根据附件一的数据无法知道汽车当前所在道路的限制速度，因此在汽车行驶过程中，取两个经纬度之间的平均行车速度，当速度大于120km/h,认为汽车超速。

**急变道**：

1、在汽车行驶中，当汽车方向角在10至45，速度大于100km/h；

2、当汽车方向角在45至90，速度大于80km/h；

3、当汽车方向角在90至180，速度大于60km/h；

4、当汽车方向角在180至270，速度大于40km/h；

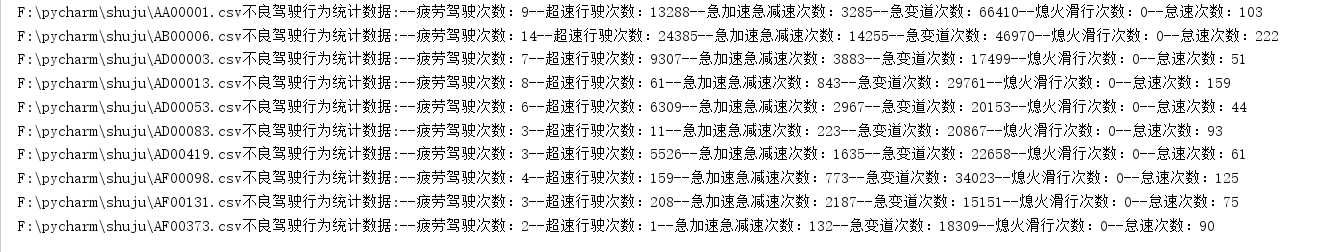
5、当汽车方向角在270至360，速度大于30km/h。

为了更详细区分汽车属于那种急变道操作，把以上五种急变道行为分别标记为1、2、3、4、5。

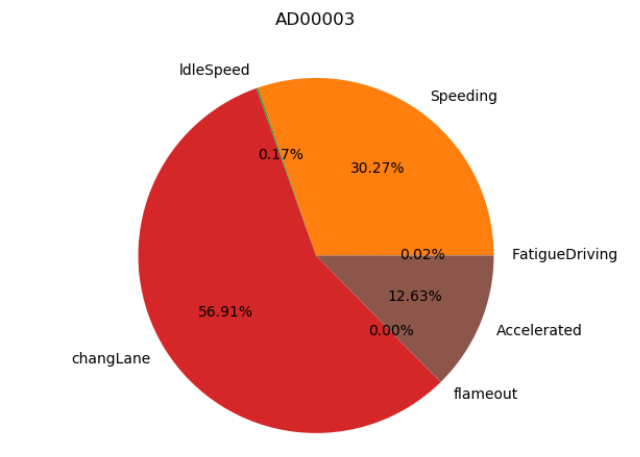
在以上的汽车安全评价模型中，疲劳驾驶、急加速、急减速、超速、急变道属于汽车点火之后的行驶动作，可根据第一个问题中得到的相关数据进行再次筛选判别，便可得出其在汽车点火之后的不良驾驶行为。怠速预热、 超长怠速、熄火滑行这三个不良驾驶行为可根据判别条件再次从源数据表中获取。最后再根据每辆车的不同不良驾驶行为进行统计次数，画出对应的饼状图，根据数据得出司机偏好于那种不良驾驶行为，给出评价结果，要求司机注意改正。

**数据可视化：**

画出相对应车辆每种不良驾驶行为所占总不良驾驶行为的比例，让人直观地看到司机的不良驾驶行为偏好。同时也会有车辆对应的每种不良行为违规数据汇总，展示图如下（图一为十辆车的每种违规数据统计，图二为编号为AD00419车辆的每种不良行为所占的百分比），因为篇幅问题，其余车辆的安全评价模型请到提交的电子版数据中查看：



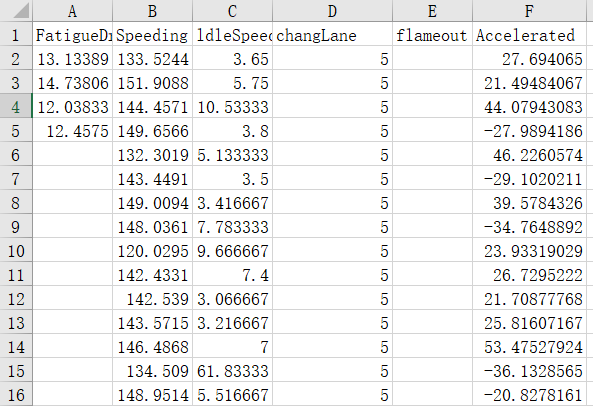
（图一）



（图二）

不良驾驶行为的具体数据存放在“不良驾驶行为具体数据【车辆编号】.xlsx”文件中，其中每一列内容代表一种不良驾驶行为的具体数据，展示编号为AF00098车辆的部分数据截图（图三）：

|  |  |
| --- | --- |
| FatigueDriving | 疲劳驾驶，每个数据代表一次疲劳驾驶的总时长，单位为小时 |
| Speeding | 超速，记录两个经纬度间的平均速度（location\_time的时间间隔一般为一秒，可认为是汽车的瞬时速度），单位为km/h |
| ldleSpeed | 怠速预热、超长怠速。数据为每次怠速的时长，单位为分钟 |
| changLane | 急变道。具体属于哪类急变道不良驾驶行为请参考急变道的判别准则 |
| flameout | 熄火滑行。 |
| Accelerated | 急加速、急减速。数据代表两个经纬度间的加速度，单位为km/h^2。整数代表急加速，负数代表急减速。 |



（图三）

数据的筛选和计算的代码均使用python完成，详细代码请参考car\_safety.py文件。鉴于车辆样本过多和电脑配置过低导致对数据的处理过慢，本题中只为十辆汽车进行了不良驾驶行为的挖掘，提交的数据中也只包含这十辆汽车的不良驾驶行为数据。但代码存在可扩展性，行车安全评价模型已经在代码中实现，只需要在主函数中加入其它车辆的采集数据，便可得出汽车的不良驾驶行为数据，如可在ff变量中加入其它车辆编号即可（图四）：



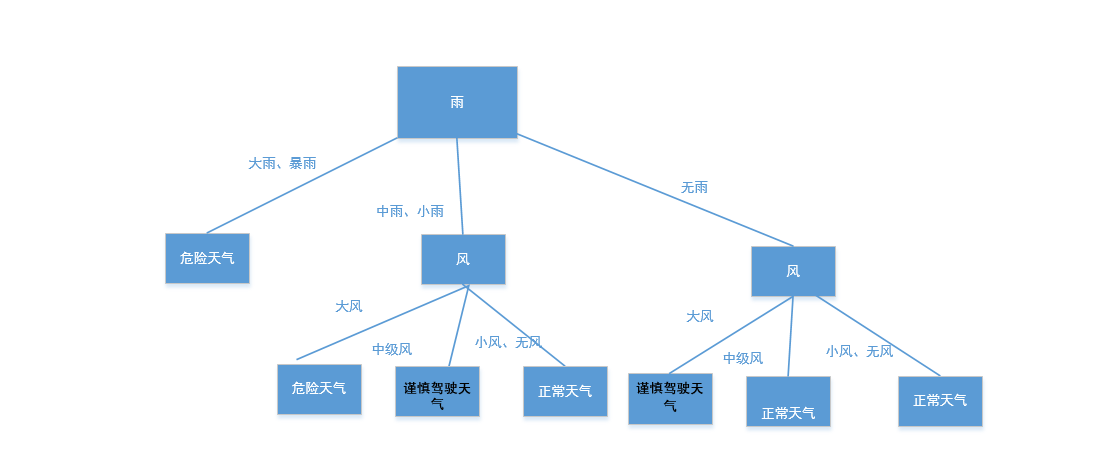
（图四）

**小结**：根据自己所建立的行车安全评价模型，挖掘出十辆车辆的不良驾驶行为中，通过数据可视化的饼图展示中，可直观地感受到驾驶汽车主要存在的不良行为主要包括在急变道和超速行驶这两个方面。在数据汇总中可得出在急变道和超速行驶这两个方面，这两个不良行为出现的次数高达上万甚至有些司机高达好几万，虽然两个数据的时间间隔一般只有1秒，但是总计起来这两个不良驾驶行为的时间也已经很长。所以该运输企业应该加强对着两方面的管理，防止司机因不良驾驶行为而发生意外。

**(3) 综合考虑运输车辆的安全、效率和节能，并结合自然气象条件与道路状况等情况，为运输车辆管理部门建立行车安全的综合评价指标体系与综合评价模型。**

行车安全的综合评价模型建立应该以车辆行驶所在位置为基础，实时获取当前位置的路况和气象条件，然后结合当前车辆的驾驶状况给出当前车辆行驶行为是否属于危险行为，若是危险行为，给出改正措施。

本题应该使用决策树来解决相关问题。下面给出一个判断天气为危险天气、谨慎驾驶天气、正常天气的用例：



**思路：**

**气象：**通过高德地图api，用gps坐标获取每段车程当日日期下的所在地，与气象数据进行比对。

1. **雨天**
2. 小雨天气下有三次及以上急加速或急减速或急变道，则标记为危险驾驶行为。
3. 中雨天气下两次及以上急加速或急减速或急变道，则判断为危险驾驶行为。
4. 大雨到暴雨天气下有一次及以上急加速或急减速或急变道，则判断为危险驾驶行为。
5. **雾天**

由于数据集中并没有给出具体的能见度情况，所以在此处定义车速时速超过40km/h为危险驾驶行为。

**3、温度**

1. 当最高温度达到35℃及以上时，驾驶时间不宜过长，否则驾驶员可能会心情受影响，车辆的状况也会出现下滑。定义驾驶时间超过3小时为疲劳驾驶。
2. 当最低温度低至0℃及以下时，当车辆存在急转弯或急减速时，标记为危险驾驶行为。

**4、大风天气**

1. 6级及以上大风时，定义车速时速超过40km/h为危险驾驶行为。
2. 8级及以上大风时，只要车辆在行驶状态时定义为危险驾驶行为。

**道路状况：**

需要获取到题目所给时间段的道路状况，目前还没有能力爬取到这些数据，所以这个小问暂时无法完成。

**小结**：此问并没有做到具体的代码实现。在扬沙和浮尘天气条件下，我并不知道如何处理此时的驾驶行为判断。

**总结**

本次作业还是存在遗憾，因为本次的题目难度比较高，对我们的技术能力提出了比较高的要求。有的组员并没有把足够多的时间放在这次作业上，导致这次并没有把问题全部解决，不过本次题目给我们以后的数据挖掘学习指明了方向。爬取到想要的数据和分析数据同样重要。