

基于大数据与微行程的运输车辆安全驾驶行为研究

摘要

随着车联网技术的普及,从大量的驾驶行为数据中挖掘出有用信息,对驾驶员的安全驾驶行为进行实时、准确、高效的评价,对提高道路运输过程安全管理水平和运输效率有着重要的意义。

针对问题 1,首先利用 GoogleEarth 对车辆的行车轨迹进行数据可视化,观察行车轨迹分布状况。使用离群点去噪、速度异常点去噪、里程异常点去噪、路径优化四种方法对行车轨迹进行漂移去噪等处理。利用地图匹配算法,结合开源地图集项目 OpenStreetMap 对去噪后的行车轨迹进行地图匹配纠正。然后,根据同一辆运输车的时间间隔阈值以及距离间隔阈值,划分不同的路线,进而画出车辆的运输路线图。将整个路线分为高速模式和低速模式,采用经纬度求解法,运动学求解法对不同模式下的里程进行求解,得到精确到米的里程值。根据国家运输行业的现行标准,找到急加速、急减速的阈值,求出每辆车的每个时刻所走的里程值,平均行车速度,急加速,急减速等。

针对问题 2,首先确定安全驾驶行为评价相关指标,根据所给数据的离散化属性以及序列化属性,定义相关不良驾驶行为和不良驾驶行为对应的识别算法,将每一个路线分为多个微行程,深度挖掘出每辆车的每个微行程的不良驾驶行为累计次数和时长。然后,为每类指标建立相应的打分体系,利用层次分析和主成分分析相结合的权重赋值法,最后对每辆车的每个微行程进行比较、分析以及评价,及对该运输企业所给的 450 辆车进行整体评价分析。

针对问题 3,首先对附件 2 中的数据进行数据离散化处理,使用气象标准、几何关系将风向,风力,天气情况等转换为具体数值,并将附件二给定的城市道路气象数据与附件一中的行车轨迹相匹配。在问题 2 的指标基础上,增加行驶侧滑可能性,车速稳定性,行车时间利用率,日均行驶里程等指标,并且考虑天气,温度等环境因素对疲劳驾驶,超速等指标的影响,对挖掘算法相应阈值进行修改。并通过模型 2 中所采取的主观客观结合赋权值法,计算出每辆车分别在安全,效率,节能等三个方面的得分情况。本文采用改进 k-均值聚类算法,以 450 辆车的 3 个方面的得分为属性,得到驾驶行为典型样本及其所属类别。然后设计三层 BP 神经网络,将每辆车的 3 类得分为神经网络的输入,驾驶行为综合评价结果作为神经网络的输出。对 BP 神经网络进行训练,得到一个在线分类的 BP 神经网络分类器。该分类器模型可以对车联网所上传的大量实时驾驶行为数据进行快速分类,评价。

另外,经过对驾驶行为挖掘算法和评价模型研究之后,为了将本文研究的算法进行系统化的应用,本文采用 ASP.NET 技术构建了基于大数据分析的驾驶行为识别与评价平台。以此达到实施监控运输车辆不良驾驶行为的作用,更好的帮助运输车辆管理部门开展道路运输过程安全管理的数据分析,提高运输安全管理水平和运输效率。

关键词: 改进 k-means 聚类算法 BP 神经网络 不良驾驶行为挖掘 数据离散化

基于大数据与微行程的运输车辆安全驾驶行为研究.....	0
摘要.....	0
1 挖掘目标.....	2
1.1 挖掘背景.....	2
1.2 挖掘目标.....	2
2 问题分析.....	2
2.1 问题一的分析.....	2
2.2 问题二的分析.....	3
2.3 问题三的分析.....	3
3 模型假设与符号说明.....	4
3.1 模型假设.....	4
3.2 符号说明.....	4
4 数据预处理.....	4
4.1 漂移轨迹去噪.....	4
4.2 地图匹配算法.....	7
4.3 分割双向车道.....	9
4.4 划分微行程.....	10
4.5 自然气象与道路数据离散化.....	11
5 问题求解.....	14
5.1 提取分析指定车辆的行车情况.....	14
5.1.1 行车线路的划分.....	14
5.1.2 对行车各项指标的提取.....	15
5.2 不良驾驶行为挖掘与评价.....	18
5.2.1 行为数据特征评价指标提取.....	18
5.2.2 不良驾驶行为挖掘算法.....	20
5.2.3 行车安全评价模型.....	28
5.2.4 评价结果分析.....	31
5.3 改进的行车安全综合评价模型.....	33
5.3.1 改进后的评价指标体系与得分标准.....	34
5.3.2 一种基于改进 K-均值聚类和神经网络的驾驶行为分级评价模型.....	37
5.3.3 基于大数据分析的驾驶行为识别与评价平台搭建.....	42
6 总结.....	44
7 参考文献.....	44
附录.....	45

1 挖掘目标

1.1 挖掘背景

随着政府,各个行业部门对运输车辆卫星定位系统的需求的越来越迫切,车联网(车联网是指借助装载在车辆上的电子标签通过无线射频等操作识别技术了)技术的发展非常迅速。目前全国范围内不管是政府还是企业,都已经大量已建成的道路运输车辆卫星定位系统。

而运营车辆的普及为人们的出行提供极大的便利的同时,也不可避免的产生了许多交通事故,造成了难以挽回的经济损失和人员伤亡。造成交通事故的原因有多种,其中驾驶员的不良驾驶行为占据主要地位。因此,对驾驶员的行为进行评价,加强监管力度,能够在一定程度上避免道路交通事故以及其他不利于节能减排,高效运输等情况的发生。

因此题需要根据某运输企业通过车联网系统所采集的数据,分析车辆行驶过程中的驾驶行为对行车安全,运输效率与节能的影响,运用数据挖掘方法,建立有效的数学模型进行评价。

随着车联网技术的逐渐普及,如何围绕道路运输车辆卫星定位系统采集的 GPS 数据开展道路运输过程安全管理的数据分析,以提高运输企业的运输安全管理水平和运输效率,成为道路运输企业亟待解决的一部分。建立基于行车安全,运行效率与节能,并结合自然气象条件与道路状况的行车安全的综合评价模型对运输企业的管理有着重要的意义。

1.2 挖掘目标

根据附件 1 里的数据提取并分析运输车辆的运输路线以及运输过程中的速度、急加速急减速、行车里程,并画出 10 辆车每辆车每条线路的在经纬度坐标下的运输路线图。

根据附件 1 的数据,分析并定义每辆运输车辆的不良驾驶行为,确定安全驾驶行为评价相关指标,构建行车安全的评价模型;结合附件 2 的数据,综合运输车辆的安全、效率、节能,以及考虑自然气象条件与道路状况等因素,构建行车安全的综合评价指标体系和综合评价模型。

2 问题分析

2.1 问题一的分析

对附件一的数据进行分析,建立相关模型找到数据中的异常点,对异常点进行相关处理。利用道路匹配算法,对附件 1 中偏移的数据进行道路匹配,通过谷歌地图可视化相关路线。定义行车里程,平均速度,急加速急减速等相关识别算法,并计算每条路线的里程,平均速度,急加速急减速。

2.2 问题二的分析

针对问题 2，定义疲劳驾驶，急加速，急减速，怠速预热，超长怠速，熄火滑行，超速，急变道等不良驾驶行为及相应的识别算法，确定安全驾驶行为评价相关指标。将每一个路线分为多个微行程，提取出每辆车的每个微行程的不良驾驶行为累计次数和时长。为每类指标建立相应的打分体系，利用层次分析和主成分分析相结合的权重赋值法，对每辆车的每个微行程进行比较、分析以及评价。下图为题中涉及的交通数据三维多粒度数据实例图：

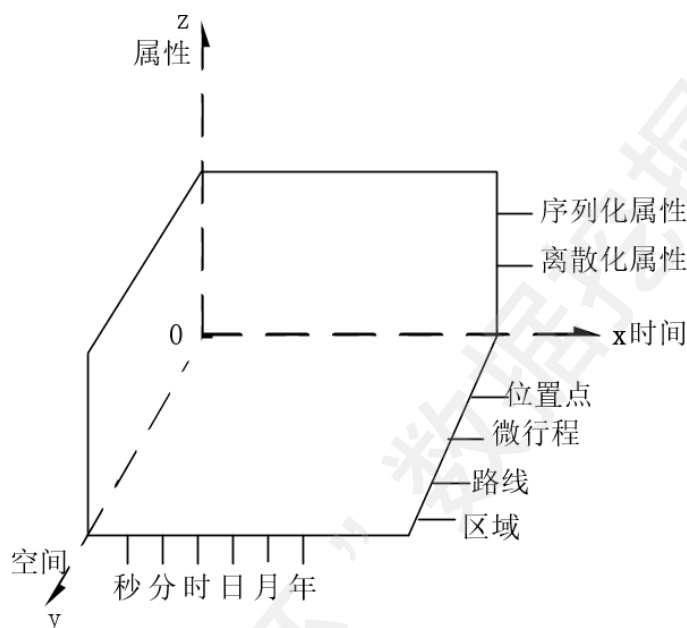


图 2-1 交通数据三维多粒度数据实例图

2.3 问题三的分析

针对问题 3，首先对附件 2 中的数据进行数据离散化处理，使用气象标准、几何关系将风向，风力，天气情况等转换为具体数值，并将附件二给定的城市道路气象数据与附件一中的行车轨迹相匹配。在问题 2 的指标基础上，增加行驶侧滑可能性，车速稳定性，行车时间利用率，日均行驶里程等指标，并且结合天气，温度等环境因素对疲劳驾驶，超速等指标的影响，对挖掘算法相应阈值进行修改。通过模型 2 中所采取的主观客观结合赋权值法，计算出每辆车分别在安全，效率，节能等三个方面的得分情况。采用改进 k-均值聚类算法，以 450 辆车的 3 个方面的得分为属性，得到驾驶行为典型样本及其所属类别。然后设计三层 BP 神经网络，将每辆车的 3 类得分作为神经网络的输入，驾驶行为综合评价结果作为神经网络的输出。对 BP 神经网络进行训练，得到一个在线分类的 BP 神经网络分类器。该分类器模型可以对车联网所上传的大量实时驾驶行为数据进行快速分类、评价。

3 模型假设与符号说明

3.1 模型假设

1. 假设所给的数据真实，可靠。
2. 假设驾驶员的驾驶行为只受自然气象条件和路况的影响，本文不考虑除二者之外的影响
3. 假设行车安全评价体系由行车安全，运行效率与节能指标来体现。

3.2 符号说明

表 3-1 符号说明图

G	行车的里程	S	急加速（减）段集合
R	地球半径	Q	超速段集合
F_f	运输车辆轮胎摩擦力	f	风力
μ	轮胎与路面的附着系数	P	疲劳段集合
M	运输车辆的载重量	V	平均行车速度
φ	急变道的转向角	α	角加速度
σ	车速标准差		

本表未涉及的符号在文中有具体介绍。

4 数据预处理

本文数据预处理主要分为 5 个部分，由于给的数据中的行车轨迹存在漂移点和噪声点，故先需要进行去噪处理，接着进一步进行道路匹配。对于一些车辆，其往往在两地之间来回行驶，故需要对其进行双向车道分割。在后续的研究中，本文引入微行程的概念，将每一辆车的总行驶轨迹划分为一个一个的微行程，将每一个微行程作为研究对象。最后针对附件二，则会对天气道路数据进行离散化处理，为后续驾驶行为研究做准备。整体流程如下：

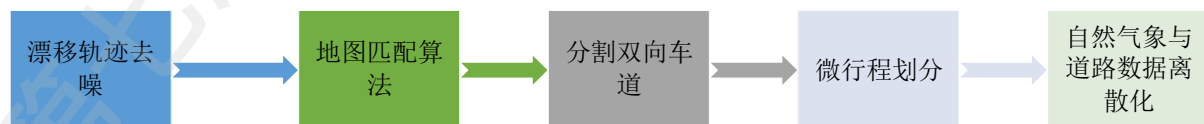


图 4-1 数据预处理流程图

4.1 漂移轨迹去噪

根据附件 1 中的原始行车轨迹，通过对其进行地图可视化处理，得到最初的行车总

路线图，这里以 AA00001 车为例：

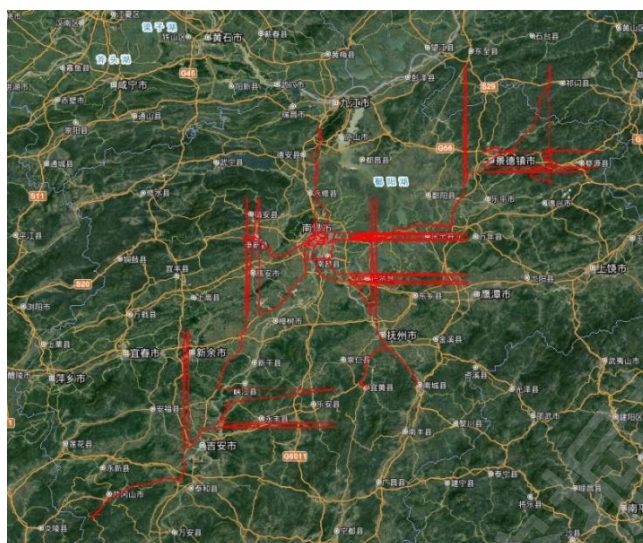


图 4-2 AA00001 最初轨迹图

由图可知，原始的 GPS 数据轨迹存在大量的偏移点和噪点，这些异常数据会对问题的分析和结果造成一定的影响，故在提取运输车辆有效轨迹数据之前需要对数据进行异常分析和预处理，将这些偏移点纠正，噪点去除。经过查阅相关文献[10-11],得到漂移数据存在以下特点：

- (1) .漂移的本质是一种噪声，不受时间和地点的约束限制。
- (2) .漂移的距离没有规律。
- (3) .部分研究表明，漂移数据一般发生在轨迹的起点与终点位置，或者建筑的边缘；
- (4) .漂移的原因可能是由信号不稳定造成的。

本文对漂移轨迹去噪，以编号为 AA00001 车为例，提出以下四种方法：



图 4-3 AA00001 车辆漂移轨迹去噪处理异常点个数

1. 离群点：

将每辆车中的轨迹点化为轨迹点群。当相邻轨迹点的距离超过一定阈值，则将其视为一个轨迹点群。部分轨迹点群只有一个轨迹点，则该点为离群点。离群点的判断条件为：

$$\begin{cases} \Delta S = R \times \arccos[\sin x_1 \times \sin x_2 + \cos x_1 \times \cos x_2 \times \cos(y_1 - y_2)] \\ \Delta S > \phi \end{cases}$$

其中 ΔS 表示两点由经纬度求出的实际距离， ϕ 为阈值，这是 ϕ 取值为 10 km 。

2. 速度异常点:

速度异常点是指车辆在某时刻的 GPS 速度与通过经纬度求出的速度之差大于某个阈值，或者该点经纬度求出的速度明显与现实生活中的速度不符，则将该点视为速度异常点。速度异常点的判断条件为:

$$\begin{cases} v = \frac{\Delta S}{\Delta t} \\ v > \varepsilon \\ |v - V| > \delta \end{cases}$$

其中 v 表示由经纬度求出的速度，其速度与 GPS 之差应小于阈值 δ ，根据运输行业经验，这里将 δ 取值为 10 km/s ， ε 取值为 200 km/s 。

3. 里程异常点:

由于车辆定位偏差，部分点往往距离实际行车点很远，故将其去除，里程异常点的判断条件为:

$$4. \begin{cases} \Delta S' = S_i - S_{i-1} \\ |\Delta S - \Delta S'| > H \end{cases}$$

$\Delta S'$ 表示两点间由车辆 GPS 行里程求出的距离。若此值这两个距离之差大于某个阈值，则将第 i 个点视为里程异常点，本文 H 点取值为 10 km 。同时，判断里程异常点的条件为车辆应在同一条行驶线路，车辆前后两点须为连续的行车轨迹。

5. 路径优化:

提取 GPS 轨迹上的连续三个采样点 A、B、C，将其视为一个三角形， h 为中间点 B 点到 AC 的垂直投影，如图所示，根据海伦公式，可以得到三角形面积 S :

$$\lambda = \sqrt{p(p-a)(p-b)(p-c)}$$

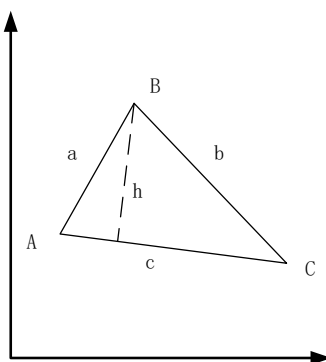


图 4-4 路径优化示意图

其中 a 、 b 、 c 分别为三角形的三边边长，即三个 GPS 轨迹点之间的水平距离， p 为半周长， S 为三角形面积。

通过比较中间点到连接线 AC 的距离与漂移数据的最大距离误差，来确定中间点 B

的取舍，即判断 h 是否满足以下公式：

$$h = 2\lambda / c \leq \Delta d_{\max}$$

其中 Δd_{\max} 为最大容许误差。若上述条件不成立，则表明该轨迹点为漂移数据点，可以对其进行删除。

删除以上 4 类异常点之后，以编号为 AA00001 车为例，得到其行车总路径如下图：

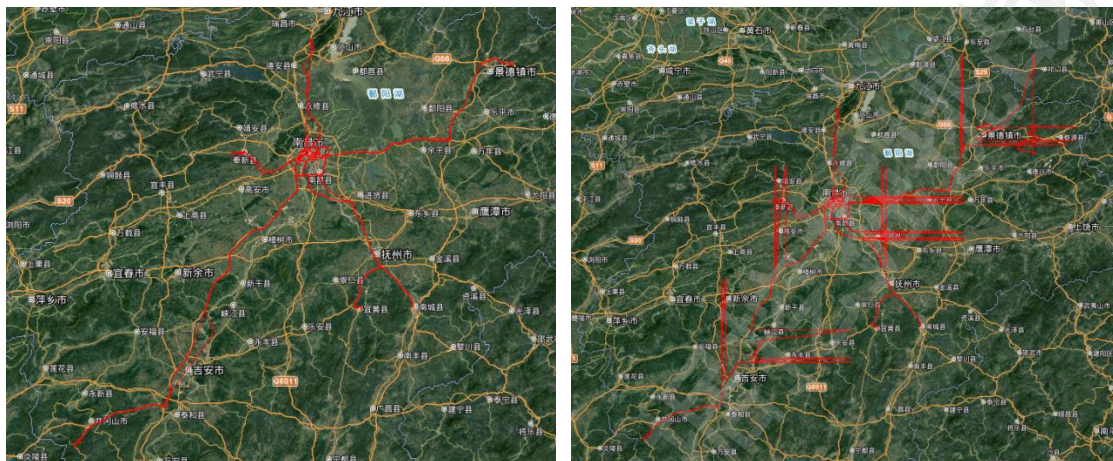


图 4-5 漂移轨迹去噪对比图

由图可以看出，在保留数据完整性的基础之上，采用离群点去噪、速度异常点去噪、里程异常点去噪、路径优化四种方法之后，漂移轨迹去噪效果较好。但是仍存在一些整体漂移的点，其路径明显与道路不符合，故采用地图匹配算法校正。

4.2 地图匹配算法

根据上述部分轨迹偏移实际交通道路的情况，本文提出了一种地图匹配的算法。将车辆的实际运行轨迹与现有的道路交通路径相结合，将偏移的路径纠正到距离最近的交通道路轨迹上，常用的地图匹配算法有：直接投影法、相关性算法、概率统计法、模糊逻辑法等等。本文采用基于几何关系的地图匹配算法，该算法共分为三种情况，分别为点到点的地图匹配算法、点到线的地图匹配算法、线到线的地图匹配算法。



图 4-6 道路匹配算法分类

a. 点到点的地图匹配算法：

点到点的地图匹配算法的原理是在合理半径的范围内识别到 GPS 观测点的最小距离节点，其算法匹配速度很快，具体的计算示例如下：

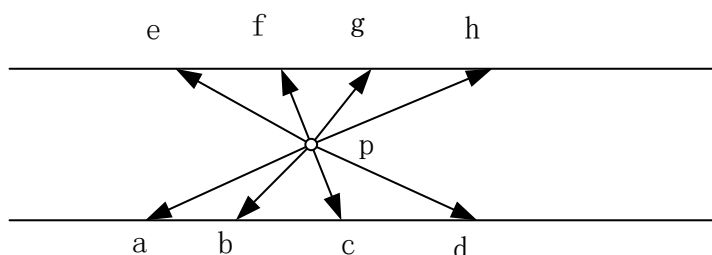


图 4-7 点到点匹配

如图所示，若 p 点为 GPS 定位点， a 、 b 、 c 、 d 、 e 、 f 、 g 、 h 分别为一条道路或者相邻道路上的节点和形状点，通过计算 p 点到这六个点的距离，取最小距离所对应的节点作为匹配结果。

b. 点到线的匹配

点到线的匹配算法原理是计算 GPS 定位点到所有待匹配路段的投影距离，最短垂直投影距离对应的路段为匹配路段。而对应的路段投影点为匹配点，具体举例如下图所示：

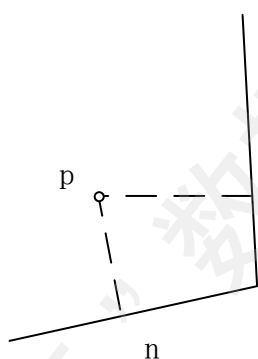


图 4-8 点到线匹配

其中 p 点为 GPS 定位点， m 为线路 1、 n 为线路 2，点到线的匹配就是寻找 p 到这两条线路谁得距离更近，更近的道路则为匹配的最终线路。

c. 线到线的匹配

其原理是线到线的匹配与其他方法相比都比较复杂，其原理是计算线与线段之间的距离，选择距离较近的道路作为匹配的道路。其中 GPS 定位点至少需要 2 个，一个作为起始点，一个作为结束点，通过比较两个定位点之间的线段和候选道路构建的曲线距离，选出距离最小的一个作为最终的匹配线路，具体示例如下：

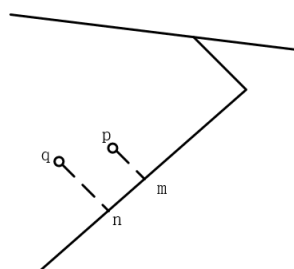


图 4-9 线到线匹配

其中若 q 点到路段的距离为 40m ， p 点到路段的距离为 30m ，则计算两条曲线之间的距离为：

$$d = |d_1 - m| + |d_2 - n|$$

进而求出 GPS 定位点到各个候选路段的距离,其中最短距离的路段则为最终匹配路段。

本文采用的地图匹配方法为 OpenStreetMap 基金会 (OSMF) 授权的开源地图项目 OpenStreetMap¹。以编号为 AA00001 车为例,其进行地图匹配算法之后的行程轨迹如下图:



图 4-10 道路匹配算法效果对比图

对比可以看出,在进行道路匹配之后,地图的部分错误轨迹得到了纠正。地图匹配算法效果较好。对于编号车辆 AA00001,共有 5 处地方得到了纠正。

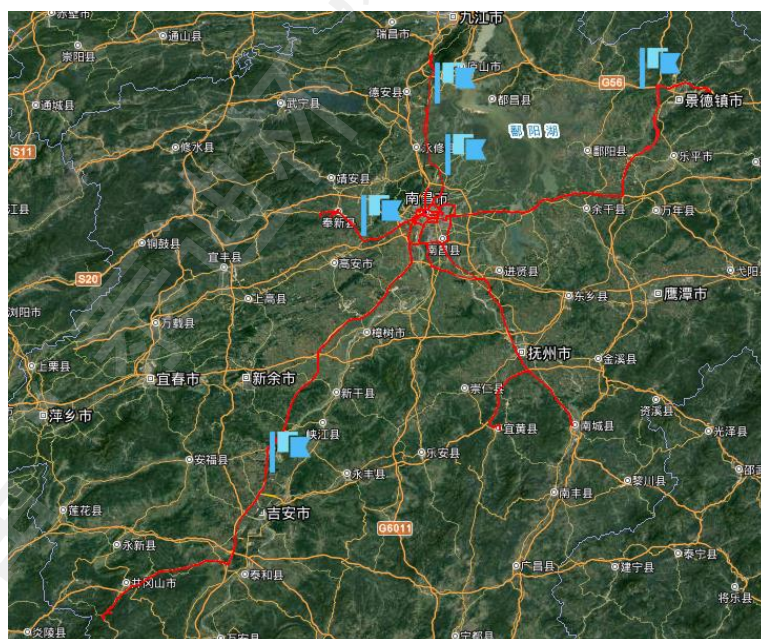


图 4-11 道路匹配算法纠正结果

4.3 分割双向车道

在车辆的实际行驶过程中,存在某一辆车在某两个地区往返行驶的情况。在实际情

¹ © OpenStreetMap 贡献者 <https://www.openstreetmap.org/>

况中双向通行路段的正反车道的交通状况存在一定的差异，故将车辆的双向行驶进行标记分割，分割算法如下：

Step1:若 L 为车辆经过的某路段，则其第一次经过该点记 r 为 1，同时标记方向为 1

Step2:若该车第二次经过该路段上的点，则先判断此时车辆的方向，若方向相同，则不进行车道分割。

Step3: 若方向不同，说明该车已经返程，此时标记该点 r 为 0，表示反向行驶。

$$r = \begin{cases} 1 & \text{正向行驶} \\ 0 & \text{反向行驶} \end{cases}$$

4.4 划分微行程

在汽车驾驶领域，汽车的行驶工况是采用统计学方法衡量某一类型车辆在某地区的实际行驶的情况。本文引入构建汽车行驶工况所使用的方法，提出采用运动学片段和微行程的概念，来体现运输车辆驾驶员的驾驶行为特征。

运动学片段的定义：

运动学片段即微行程，通常是指车辆在行驶过程中从一个怠速状态起始点到下一个怠速状态的过程，下图时间 5-25 时间段可以表示为一个微行程：



图 4-12 微行程与运动片段学定义

以编号为 AA00002 车辆为例，本文依据上述定义将其划分为 12 个微行程，下图展示该车微行程初末时间。

device_num	rection_and	lng	lat	acc_state	ht_turn	sign	turn_sign	hand_brake	foot_brake	location_time	gps_speed	mileage	isYuRei
AAA9101002	312	116.2165	28.22356	1	0	0	0	0	0	2018/8/4 2:29	93	8865	0
AAA9101002	12	115.9119	28.66967	0	0	0	0	0	0	2018/8/4 4:31	0	8963	0
AAA9101002	12	115.9119	28.66969	1	0	0	0	0	0	2018/8/5 7:25	0	8963	1
AAA9101002	136	115.783	28.58737	0	0	0	0	0	0	2018/8/5 8:17	0	8984	1
AAA9101002	136	115.783	28.5874	1	0	0	0	0	0	2018/8/5 8:18	0	8984	2
AAA9101002	37	115.9119	28.66994	0	0	0	0	0	0	2018/8/5 13:09	0	9317	2
AAA9101002	37	115.9119	28.66999	1	0	0	0	0	0	2018/8/5 23:19	0	9317	3
AAA9101002	171	115.9118	28.66949	0	0	0	0	0	0	2018/8/6 4:08	0	9546	3
AAA9101002	171	115.9118	28.6695	1	0	0	0	0	0	2018/8/6 7:09	0	9546	4
AAA9101002	350	115.9723	28.70357	0	0	0	0	0	0	2018/8/6 7:46	0	9556	4
AAA9101002	350	115.9723	28.70357	1	0	0	0	0	0	2018/8/6 7:49	0	9556	5
AAA9101002	358	115.9722	28.70396	0	0	0	0	0	0	2018/8/6 7:51	0	9557	5
AAA9101002	358	115.9722	28.70396	1	0	0	0	0	0	2018/8/6 7:51	0	9557	6
AAA9101002	347	115.9721	28.70422	0	0	0	0	0	0	2018/8/6 7:56	0	9557	6
AAA9101002	347	115.9722	28.70423	1	0	0	0	0	0	2018/8/6 8:18	0	9557	7
AAA9101002	20	115.9119	28.66973	0	0	0	0	0	0	2018/8/6 9:04	0	9567	7
AAA9101002	0	115.9119	28.66971	1	0	0	0	0	0	2018/8/7 0:00	0	9567	8
AAA9101002	6	115.9169	28.70379	0	0	0	0	0	0	2018/8/7 1:39	0	9587	8
AAA9101002	6	115.9168	28.70382	1	0	0	0	0	0	2018/8/7 1:45	0	9587	9
AAA9101002	151	115.9275	28.72082	0	0	0	0	0	0	2018/8/7 1:58	0	9589	9
AAA9101002	151	115.9275	28.72083	1	0	0	0	0	0	2018/8/7 2:01	0	9589	10
AAA9101002	169	115.9169	28.70373	0	0	0	0	0	0	2018/8/7 2:23	0	9594	10
AAA9101002	169	115.9169	28.70364	1	0	0	0	0	0	2018/8/7 3:34	0	9594	11
AAA9101002	36	115.9115	28.67008	0	0	0	0	0	0	2018/8/7 4:38	0	9614	11

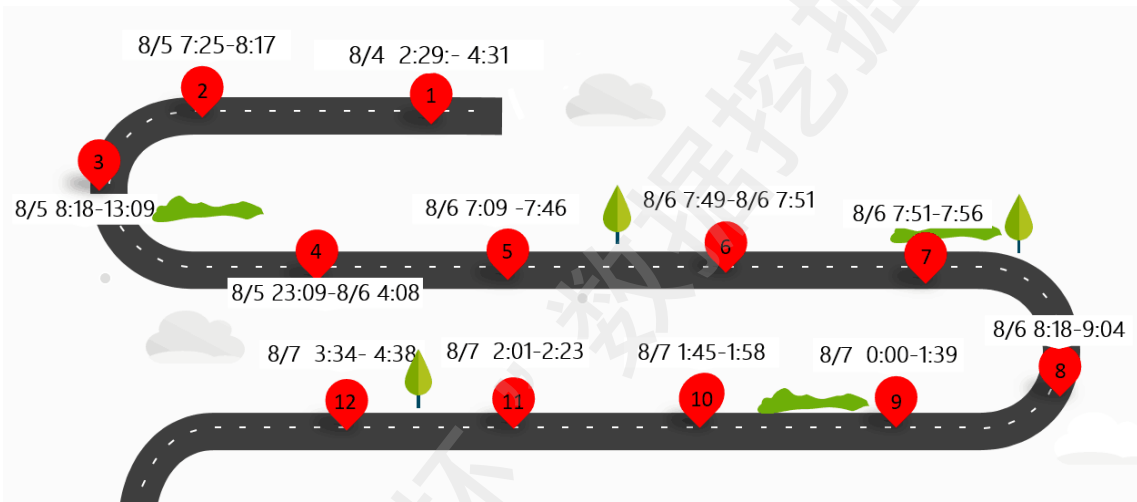


图 4-13 AA00002 车辆微行程划分结果

4.5 自然气象与道路数据离散化

附件二中的数据一共有 11 个字段，本文根据数据中的省、市、县地理位置找出其对应的经纬度坐标，然后与给定的 450 辆车辆的行车轨迹进行匹配，找出每辆车所对应的地理环境。若行车轨迹中与给定数据无法匹配的点，本文将将其视为不存在自然气象和道路信息。下图为附件二给定的地理位置图：



图 4-14 附件二地理位置

首先去除数据中的错误数据,如下表所示,其风向值为 nan 转 nan,故将其视为错误数据,删除。

province	prefecture_city	county	wind_direction
浙江	湖州	湖州	nan 转 nan

对风力和风向划分等级,将风力转化为风速,风向角表示为从定位点的正北方向起,以顺时针方向至风向的水平夹角,如南风风向为 0 度,北风风向为 180 度。风力、风向转化表如下:

wind_power	values (m/s)	wind_power	values (m/s)
1-2 级	1.8	3	4.4
3-4 级	5.65	2	2.45
4-5 级	8.1	5~6 级转 4~5 级	9.65
<3 级	3.3	<3 级转 4~5 级	5.4
3~4 级	5.65	3~4 级转 5~6 级	8.6
3~4 级转<3 级	5.4	4~5 级转 5~6 级	9.65
4~5 级转 3~4 级	7.05	3~4 级转 4~5 级	7.05
4~5 级转<3 级	5.4	5-6 级	10.9
4~5 级	8.1	5~6 级转 6~7 级	12.55
<3 级转 3~4 级	5.4	5~6 级	10.9
5~6 级转 3~4 级	8.6	7~8 级	12.3
6~7 级转 4~5 级	11.3	6-7 级	13.95
7~8 级转 5~6 级	14.35	7-8 级	12.3
6~7 级转 5~6 级	12.55	11-12 级	30.55
6~7 级	13.95	8-9 级	20.8
9-10 级	24.6	10-11 级	28.55

wind_direction	values	wind_direction	values
北风	180	西北风转东风	202.5
东风	270	西风转南风	45
东北风	225	西南风转南风	22.5
南风	0	东风转西风	180
西南风	45	东风转东北风	247.5
东南风	315	南风转西南风	22.5
西北风	135	南风转西风	45
西风	90	南风转无持续风向	0
无持续风向	NA	西南风转无持续风向	45
南风转东南风	157.5	南风转西北风	67.5
东风转东南风	292.5	西南风转东风	157.5
西北风转东北风	180	南风转北风	90
东北风转东风	247.5	东南风转西北风	22.5
...
北风转东北风	202.5	北风转南风	90

由于数据较多，部分结果如上，具体结果见附件，下图为部分时间段的风力地理数据：

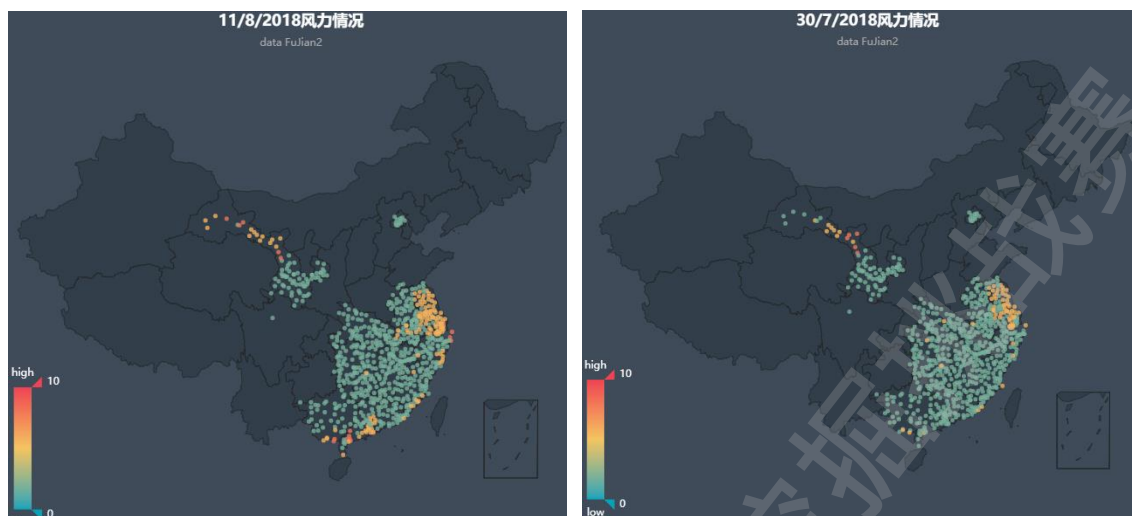


图 4-15 部分日期风力地理数据

对于气温：

气温分为最高温度和最低温度，本文对这两列气温进行线性加权处理，得到当前城市的平均气温：

$$T = \frac{T_{\max} + T_{\min}}{2}$$

对于天气环境：

将天气分为 4 个等级，不同等级代表天气的恶劣程度。以下是部分量化结果，由于天气重量较多，具体结果见附件。

conditions	程度	conditions	程度
多云	0	雷阵雨转多云	1
小雨转多云		阵雨转多云	
晴		中雨转小雨	
小雨		中雨转阴	
小雨转阴		中雨转多云	
阴		小雨转中雨	
阴转小雨		阵雨转晴	
多云转小雨		雷阵雨转晴	
晴转多云		阵雨	
……		……	
多云转小雪		小到中雨	
大雨转小雨	2	暴雨转中雨	3
大雨转多云		暴雨	
大雨转中雨		小雨转暴雨	
小雨转大雨		暴雨转阵雨	
阵雨转大雨		大雨-暴雨转阵雨	
中雨转大雨		暴雨转阴	
雷阵雨转大雨		大雨-暴雨转多云	

大雨		暴雨转多云	
.....		
阵雨转雨夹雪		大暴雨	

下图为部分日期的天气道路情况：

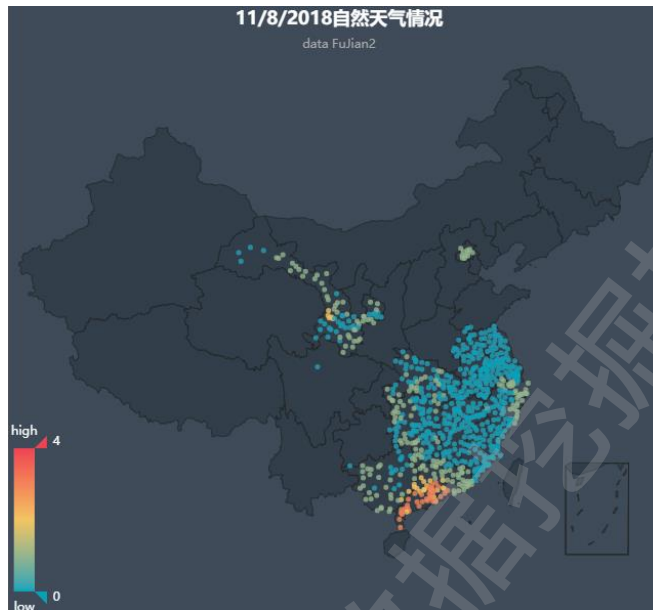


图 4-16 部分日期天气道路情况

5 问题求解

5.1 提取分析指定车辆的行车情况

5.1.1 行车线路的划分

由于附件一中给定的数据是某一车辆所有的运行轨迹，所以需要先对数据进行行车线路进行划分，确定哪一部分数据是一条线路。本文综合考虑相邻点之间的时间间隔和路程间隔，具体划分规则如下：

$$\begin{cases} T = T_i - T_{i-1} \\ S = S_i - S_{i-1} \end{cases}$$

T 取值范围/h	是否为一条线路
[0,2]	是
[2, +∞]	否

S 取值范围/km	是否为一条线路
[0,10]	是
[10, +∞]	否

其中 T 表示相邻轨迹两点时间之差， S 表示相邻轨迹车辆里程之差。

5.1.2 对行车各项指标的提取

1). 行车里程:

常见的行车里程计算通常是采用车辆的初末起始点相减得到的，其公式如下：

$$G = L_{start} - L_{end}$$

其中 G 表示该车在某一线路的行车里程， L_{start} 表示该车在某一线路的初始 GPS 里程，

L_{end} 表示该车在某一线路的最终 GPS 里程。

但是在本题中，由于所给定的行车里程不够精确，如果采用上式方法来表示运输车辆行车里程，存在着很大的偏差。结合参考文献[9], 本文提出一种基于卫星定位数据的车辆行驶里程计算方法。改进的行车里程的求解算法如下：

1. 将一条线路的所有行车数据划分为低速模式和中高速模式。
2. 遍历路线，根据阈值速度 V 以及每条数据下的 GPS 速度大小，为每辆车的每个时间戳贴上低速和中高速的标签。
3. 将连续的低速或者中高速看做一个记录集 N (N 表示该记录集中数据的条数)
4. 针对低速的记录集 N_1 ，采用**运动学求解法**。从记录集 N_1 的第一个记录开始，依

序计算到最后一条记录，计算两点相邻数据间的坐标距离 G_n ，当 $G_n < \omega \times \Delta t$ 时(其中 ω 表示小车的极限速度，用于滤除记录中的不正常值，这里取的 $\omega = 100\text{m/s}$ 。 Δt 表示这相邻两条数据之间的时间差)，利用 $G_i = G_{i-1} + G_n$ 迭代，得到该段记录集的里程数。

5. 针对高速记录集 N_2 ，采用**经纬度求解法**。从记录集 N_2 的第一个记录开始，依序计算到最后一条记录，计算两点相邻数据间的坐标距离 G_j ，GPS 速度距离 G'_j ，然后执行第 6 步，以此比较每个时间戳的相邻距离的合理性。

6. 如果 $G_j > G'_j$ ，并且同时满足 $G_j < \omega \times \Delta t$ 时， $G_n = G_j$ 。同理，若 $G_j < G'_j$ ，并且同时满足 $G'_j < \omega \times \Delta t$ 时， $G_n = G'_j$ 。找出每个时间戳下的 G_n ，采用 $G = G + G_n$ 迭代，求出该段记录集的里程数。

2). 平均行车速度:

$$\begin{cases} T' = T_{start} - T_{end} \\ V = \frac{G}{T'} \end{cases}$$

其中 T' 表示该车在该线路的总共行车时间， T_{start} 表示其初始时刻， T_{end} 表示其最终时刻， V 表示该车的平均行车速度。

根据平均行车速度结果，选取 10 辆车的第一个线路 ACC 为 1 时的平均行车速度进行比较，发现 10 辆车整体的平均行车速度均在同一量级。

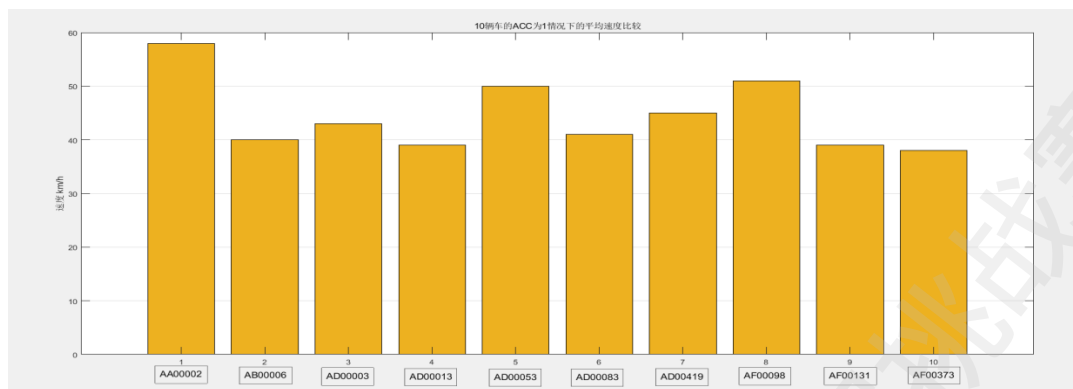


图 5-1 10 辆车平均速度对比

对于每一辆车来说，其每条线路平均行车速度是经常变化的。

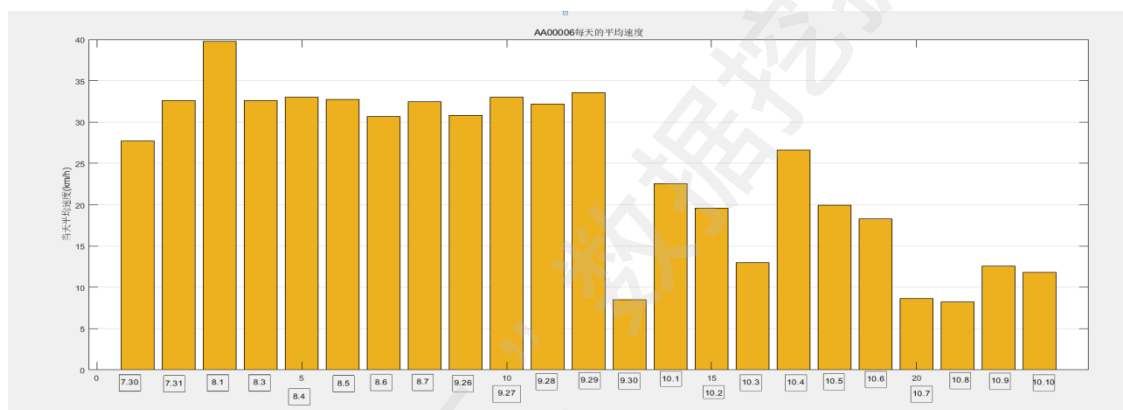


图 5-2 AB00006 部分平均行车速度

对于每一辆车某个路线，其车辆的速度也是不断变化的。

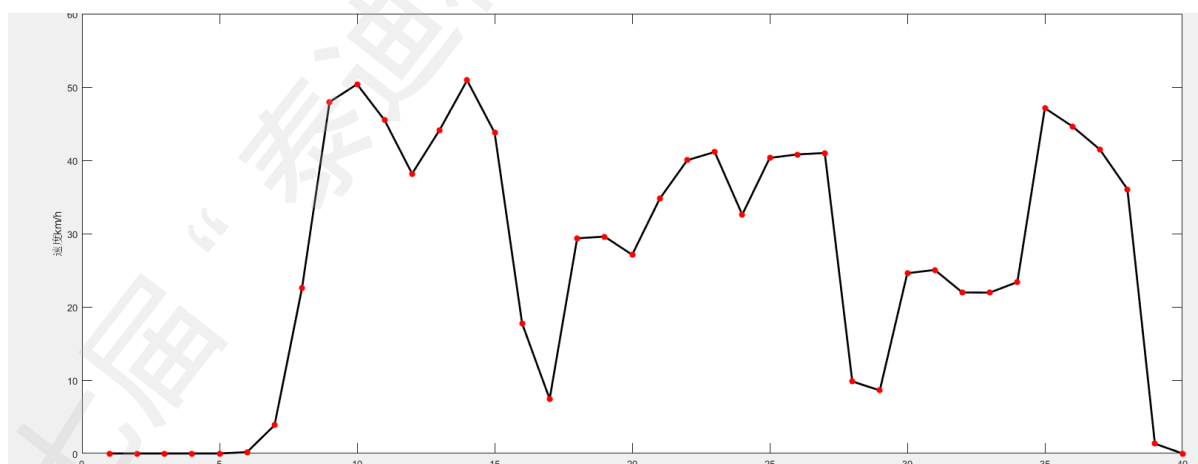


图 5-3 AB00006 某路线车速

3). 急加速急减速:

运输车辆急加速急减速情况主要由加速度决定，加速度的定义如下：

$$a = \frac{\Delta v}{\Delta t}$$

其中 Δv 表示相邻轨迹两点的速度之差， Δt 表示相邻轨迹两点的时间之差。

根据华东师范大学任君慧的研究[1],本文将急加速、急减速情况定义为:

加速度 m/s^2	行驶情况
$[-3,3]$	正常
$(3, +\infty)$	急加速
$(-\infty, -3)$	急减速

此外,当两个单一急加速或者急减速行为的时间间隔小于连续急加速行为合并时间阈值或连续急减速行为合并时间阈值时,将其合成视为一个急加速或急减速行为,并且将计算得到该急加速或急减速行为的连续单一急加速或急减速次数。

根据急加速急减速结果,以 AB00006、AB00006 为例,本文发现车辆的每条线路的加速度也是逐渐变化的,这与实际的道路情况相符合,因为车辆可能会遇到红绿灯等必要情况停车。

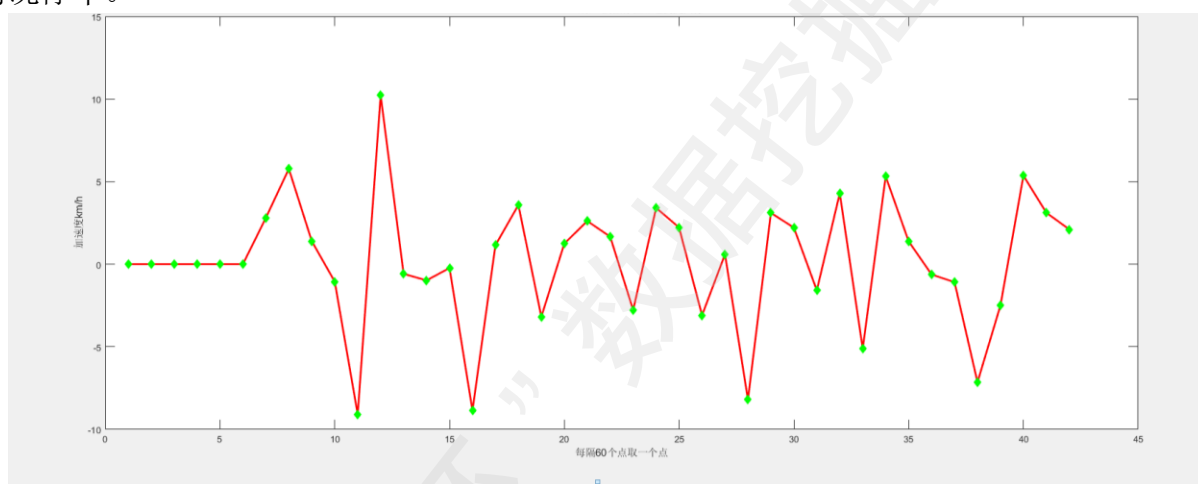


图 5-4 AB00006 线路 1 加速度变化图

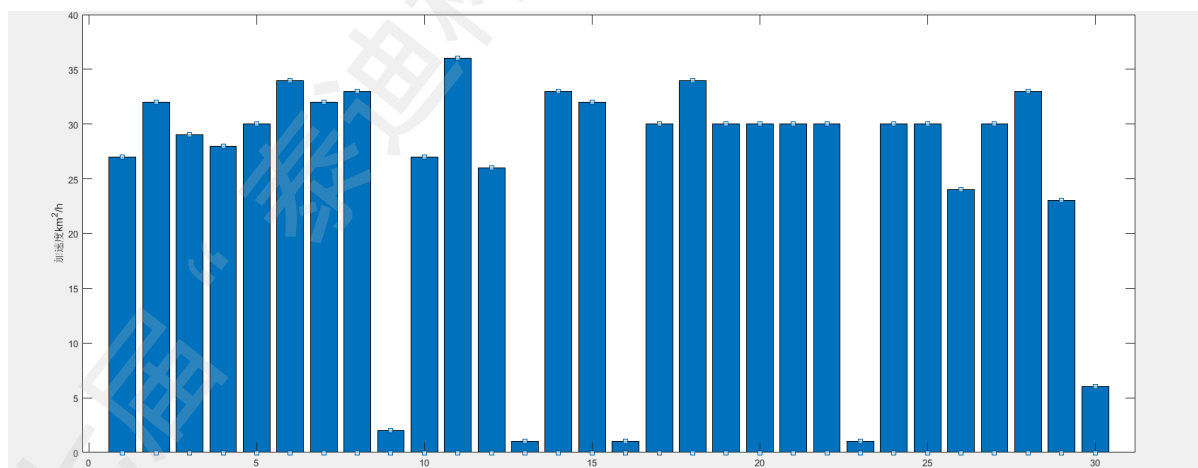
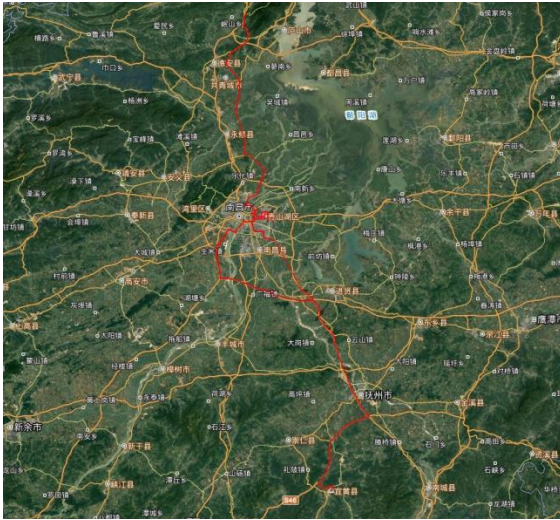


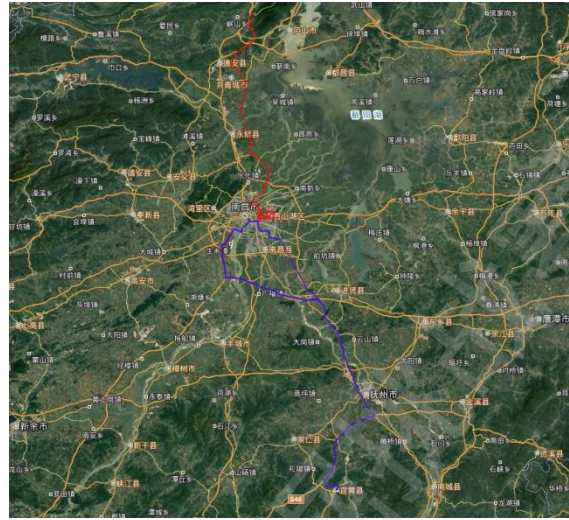
图 5-5 AB00006 线路 1 加速度比较

4.运输路线图

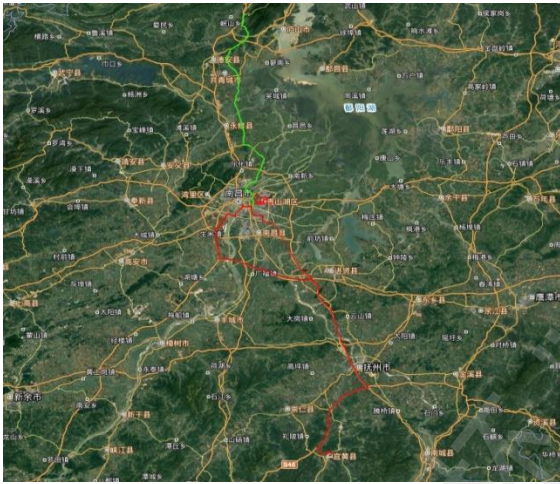
本文采用 Excel 中的三维地图和地理可视化工具 91 位图助手对车辆的运输线路进行可视化处理,并生成每辆车每条线路的运输路线图,以编号为 AA00002 的车辆运输路线图的所有线路为例,其运输路线图如下图,其余 9 辆车的运输路线图见提交的附件。



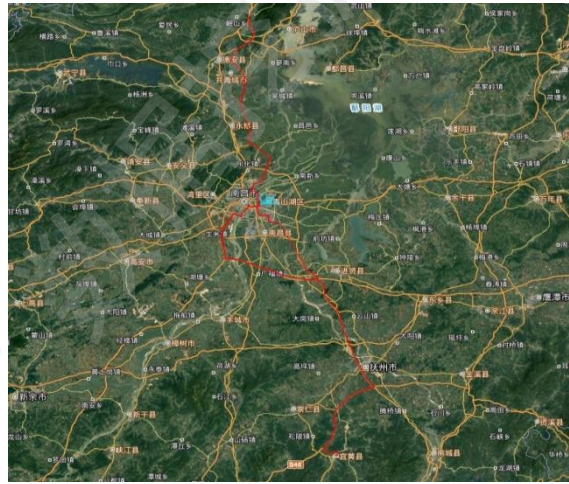
AA00002 运输总路线图



AA00002 线路 1



AA00002 线路 2



AA00002 线路 3

图 5-6 AA00002 运输路线图

由图可以看出，AA00002 的车辆主要有 3 条路线，线路 1 从南昌市城中心出发，到宜黄县往返。线路 2 从南昌市中心出发，到岷山乡往返。线路 3 在南充市青山湖区一带往返。

5.2 不良驾驶行为挖掘与评价

对于问题二，本文首先定义 9 种不良驾驶行为，将根据附件一对 450 辆车进行深度挖掘。

不良驾驶行为	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	车速稳定性	

5.2.1 行为数据特征评价指标提取

行车安全的评价模型指标提取

以运动学片段中的特征值提取[8]为借鉴，结合现有所提供的运输车辆卫星定位数据和文献[4-7]。本文主要研究和计算提取出以下行车安全评价指标：

行车安全评价指标	一级评价指标	二级评价指标
	急加速\急减速	累计时长
		加速总次数
	超速行驶	累计时长
		超速总次数
	疲劳驾驶	累计时长
		疲劳驾驶总次数
熄火滑行	累计时长	
	滑行总次数	
车速稳定性	标准差	

行车安全单项指标评价方法：

综合考虑基于卫星定位数据的驾驶行为安全与节能评价方法等公开技术文献[4-7]，根据中华人民共和国公共安全行业标准等行业规范[2-3]，以及各运营企业对驾驶员的考核制度，得到的评价指标得分标准如下表：

单项评价指标得分标准：

评价指标	评分标准		
急加速	累计时长	$y_1 = 100 - t$	$y = 0.5y_1 + 0.5y_2$
	加速总次数	$y_2 = 100 - k_1n$	
超速行驶	累计时长	$y_1 = 100 - t$	$y = 0.5y_1 + 0.5y_2$
	超速总次数	$y_2 = 100 - k_2n$	
急减速	累计时长	$y_1 = 100 - t$	$y = 0.5y_1 + 0.5y_2$
	加速总次数	$y_2 = 100 - k_3n$	
疲劳驾驶	累计时长	$y_1 = \begin{cases} 100; & t \leq 480 \\ 0; & t > 480 \end{cases}$	$y = 0.5y_1 + 0.5y_2$
	疲劳驾驶总次数	$y_2 = 100 - k_4n$	
熄火滑行	累计时长	$y_1 = \begin{cases} 100; & 0 < t \leq 1 \\ 100 - 10t; & t > 1 \end{cases}$	$y = 0.5y_1 + 0.5y_2$
	滑行总次数	$y_2 = 100 - k_5n$	
车速稳定性	标准差	$y = \begin{cases} 100; & 0 < \sigma < 20 \\ 80; & 20 < \sigma \leq 40 \\ 60; & \sigma > 40 \end{cases}$	

其中 σ 为标准差, t 表示时间, n 表示次数。 k_1 、 k_2 、 k_3 、 k_4 、 k_5 为系数。

5.2.2 不良驾驶行为挖掘算法

1. 疲劳驾驶行为挖掘算法

依据道路交通安全行为规范 2014[2], 存在以下情况者可以视为疲劳驾驶:

- 1) 驾驶人在 24 小时内累计驾驶时间超过 8 小时。
- 2) 驾驶人连续驾驶时间超过 4 小时未停车休息或每次停车休息时间少于 20 分钟。
- 3) 对于单程运行里程超过 400 公里(高速公路直达客运 600 公里)的客运车辆, 企业应当配备两名以上客运驾驶人。

本识别算法利用附件 1 的数据提取并判断驾驶员的疲劳驾驶行为, 并记录疲劳驾驶的次数和累计时长, 疲劳驾驶行为挖掘算法步骤如下:

Step1: 获取附近 1 的定位数据, 根据行业标准初始化参数;

Step2: 获取驾驶员的第 i 条数据, i 初始设置为 0;

Step3: 判断车速是否为 0, 若否, 则说明正在行车, 不做标记; 若是, 则说明车已经停止, 标记为 0, 加入到疲劳驾驶点记录 *oversleepList*;

Step4: 遍历 *oversleepList* 数据, 判断相邻两条数据间隔 T 是否小于等于 20 分钟, 若否, 则划分到不同的疲劳段, 若是, 则表明驾驶员可能已经休息时间符合要求, 划分到同一个疲劳段, 最后形成疲劳段集合 $P = \{P_1, P_2, P_3, P_4 \dots\}$;

Step5: 遍历疲劳段集合, 统计每个疲劳段的第一条以及最后一条数据的时间间隔 T_1 , 若大于 4 个小时, 则记为一次疲劳驾驶次数, 累计历史驾驶时长; 若小于 4 个小时, 累计驾驶时长, 若驾驶时长大于 8 个小时, 也记为一次疲劳驾驶次数;

Step6: 输出疲劳驾驶次数和疲劳驾驶累计时长, 算法结束。

2. 急加速急减速挖掘算法

具体急加速急减速定义见 5.1.2,

挖掘算法步骤如下:

Step1: 获取附件 1 的定位数据, 根据行业标准初始化参数;

Step2: 设驾驶车辆的卫星定位数据序号为 i , i 初始设置为 0;

Step3: 取第 i 条和第 $i+1$ 条驾驶数据, 计算两条间的加速度, 判断是否大于设定的阈值 A (如是急减速识别, 计算两条间的加速度, 判断是否小于设定的阈值 A), 若否, 不算入急加速点; 若是, 则为急加(减)速点, 加入到急加(减)速点记录里

Step4: 遍历急加速(减)点记录数据, 判断相邻两条数据间隔 T 是否小于等于合并时间阈值, 若否, 则划分到不同的急加速(减)段, 若是, 划分到同一个急加速(减)段, 最后形成急加速(减)段集合 $S = \{S_1, S_2, S_3, S_4 \dots\}$;

Step5: 遍历急加速(减)段集合, 急加速(减)段的个数为急加速(减)的次数, 统计每个急加速(减)段的第一条以及最后一条数据的时间间隔 T_1 , 累计急加速(减)时长

Step6: 输出急加速（减）次数和急加速（减）累计时长，算法结束。
急加速急减速算法流程图如下：

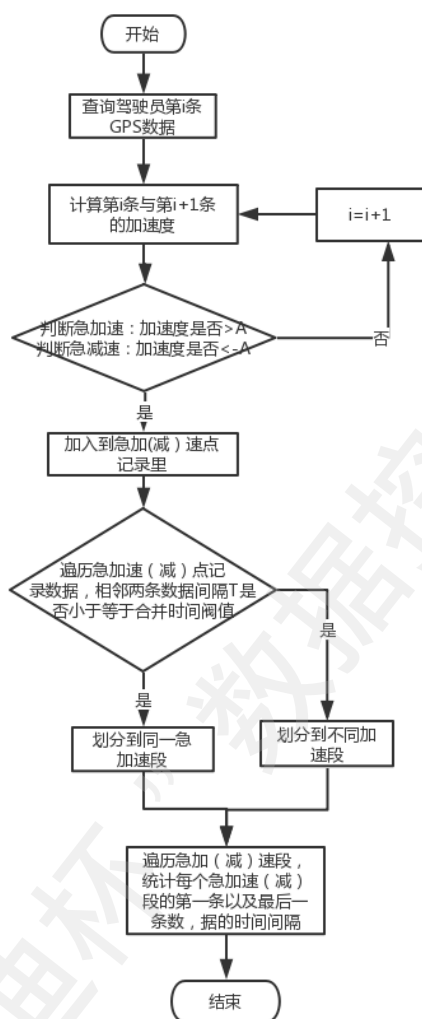


图 5-7 急加速急减速流程图

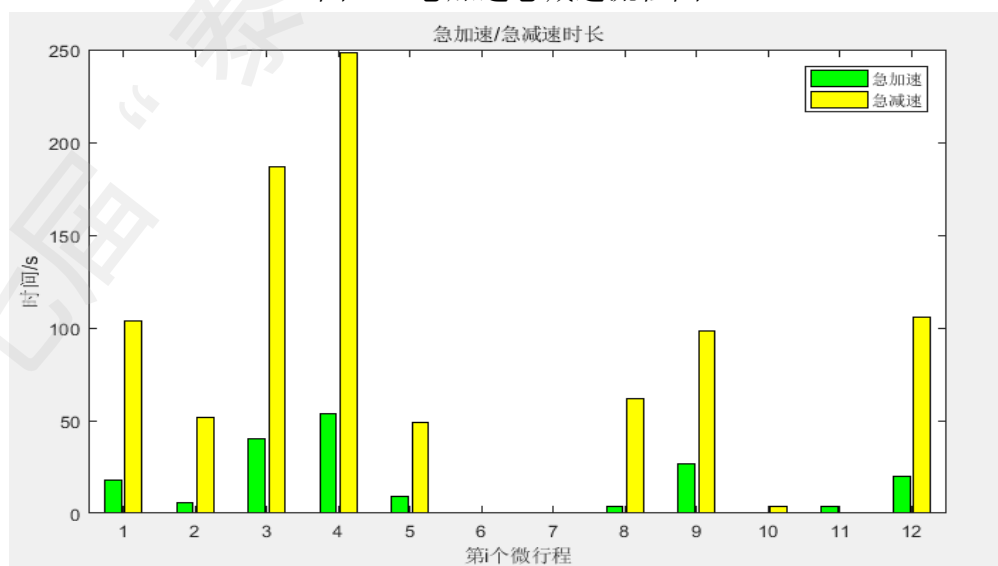


图 5-8 AA0002 急加速急减速累计时长

以 AA00002 为例，可以看出，其急减速急加速的累计时长随机分布，急减速时长大于急加速时长，这可能是由于车辆避让其他车辆造成的，比较符合现实情况。

3. 超速识别算法

设当前车载终端计算得到的车速为 V ，而驾驶车辆所对应的所在路段的车速阈值为 V_{\max} 。若 $V > V_{\max}$ ，则判断该车超速。其中车速 V ：

$$\begin{cases} V = \frac{\Delta S}{\Delta t} \\ \Delta S = R \times \arccos[\sin x_1 \times \sin x_2 + \cos x_1 \times \cos x_2 \times \cos(y_1 - y_2)] \end{cases}$$

其中 x_1, y_1 分别是纬度和经度的弧度值， R 为地球半径。根据参考文献[5]，本文规定超速的界定时长为 30 秒，即持续超速超过 30 秒认定为违规超速。

根据运输行业标准，最高车速阈值为 V_{\max} 为 100km/h，超速时间最低限值 T_a 为 3s，是为了避免卫星定位数据漂移造成不必要的误判。

超速挖掘算法步骤如下：

Step1: 获取附件 1 的定位数据，根据行业标准初始化参数；

Step2: 获取驾驶员的第 i 条数据， i 初始设置为 0；

Step3: 判断车速是否大于最高车速阈值为 V_{\max} ，若是，则说明超速，加入到超速点记录 *overspeedList*；若否，则说明正常行驶；

Step4: 遍历 *overspeedList* 数据，判断相邻两条数据间隔 T 是否小于等于 3s，若否，则划分到不同的超速段，若是，划分到同一个超速段，最后形成超速段集合 $Q = \{Q_1, Q_2, Q_3, Q_4 \dots\}$ ；

Step5: 遍历超速段集合，统计每个疲劳段的第一条以及最后一条数据的时间间隔 T_1 ，若大于 30s，则记为一次超速次数，累计历史超速时长；若小于，不记作超速行为；

Step6: 输出超速次数和超速累计时长，算法结束。

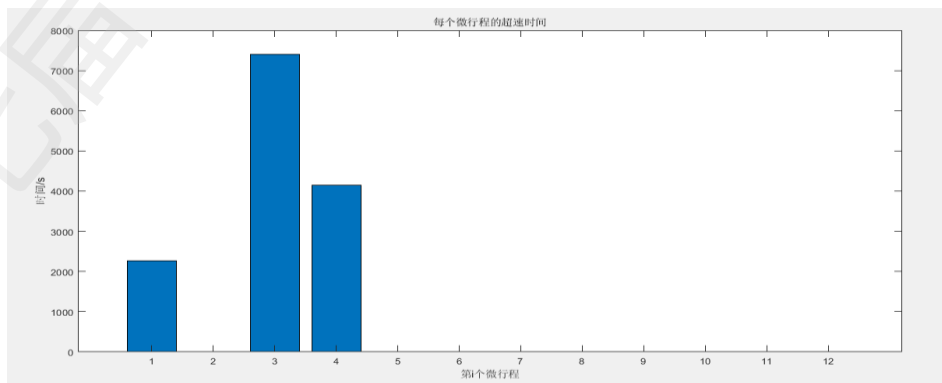


图 5-9 AA00002 每个微行程超速累计时间

超速算法流程图如下：

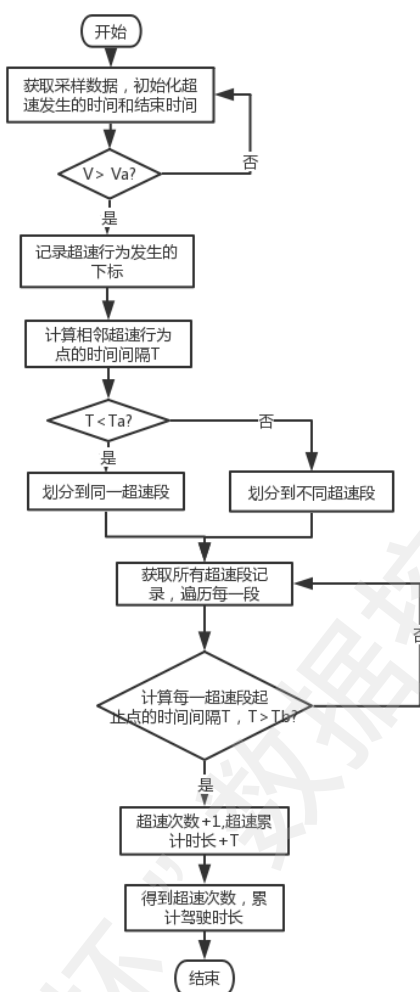


图 5-10 超速算法流程图

以 AA00002 为例，可以看出，其不是每个微行程都存在超速过程，只有部分时间才有超速不良驾驶行为发生。

4. 怠速预热挖掘算法

怠速是指汽车发动机发生空转现象，怠速预热是指车辆从发动机启动到车辆自身起步的一个过程，车辆怠速预热满足下列条件：

- 1). 当前车辆速度为 0；
- 2). 发动机 ACC 状态为 1；
- 3). ACC 状态为 1 时是该车辆当天首次点火或该数据上一条 ACC 状态数据为 0；

怠速预热终止条件是车辆速度 $V > 0$ ，表示该车辆起步终止。同时，车辆的怠速预热时长问题，怠速时间越长，表明驾驶员的不良驾驶行为越明显。当 $T > T_{\max}$ 时，记录该车辆驾驶员不良驾驶行为次数。怠速预热状态表如下表所示：

ACC 状态	车速	备注
1	$V = 0$	当天首次点火或上条 ACC 状态为 0

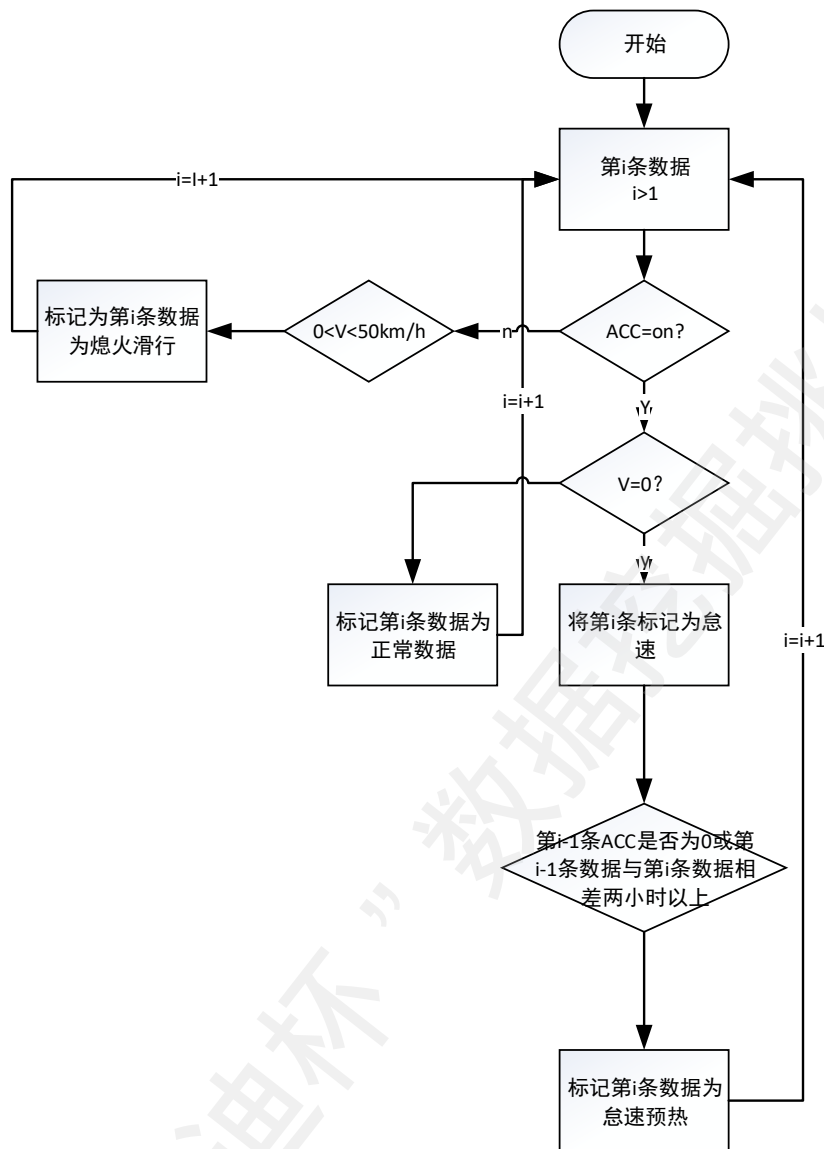


图 5-11 每条数据的标记算法

怠速预热的挖掘算法如下：

Step1: 判断当前数据是否 ACC 状态为 on，且车速为 0。若是，则标记该条数据为怠速状态。

Step2: 判断当前数据是否与上条数据相差 2 个小时或上条 ACC 的数据为 off，则定义该条数据为怠速预热的起始点。定义 flag=1

Step3: 如果上一条数据为怠速预热状态，且该条数据 acc 为 on 且车速为 0，则也将该条数据标记未怠速预热

Step4: 若上条语句为怠速预热状态，且本条数据不满足怠速的条件，则证明本条数据的开始时，该段怠速预热已结束。若下一条还有数据，转到 step1

Step5: 遍历完后，对连续的，且上一段怠速预热接收和下一段怠速预热开始相差只有 3 秒以下的进行合并

Step6: 设置怠速预热的最低时间门限值，只有大于这个门限值的时间才能将该段定为一次怠速预热，统计并记录段数（也就是怠速预热的次数）以及时间，并返回。

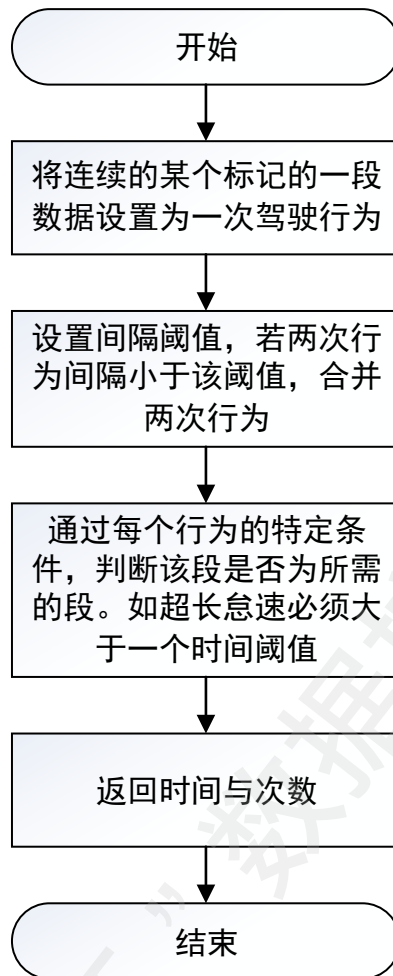


图 5-12 对于标记后的数据分段算法

5. 超长怠速挖掘算法

依据中华人民共和国交通运输部《汽车驾驶节能操作规范》2011 年版[2], 车辆停车超过 60 秒, 应将发动机熄火, 当发动机转速不为 0 但是车辆本身速度为 0 时, 表明该车超长怠速。超长怠速状态表如下表所示:

ACC 状态	车速	停车时间
1	$V = 0$	$T > 60s$

挖掘算法如下:

Step1: 遍历这个微型程, 判断车辆的每一秒是否处在怠速阶段。

Step2: 定义最短时间差阈值, 若小于这个阈值, 则将两端合成一段。

Step3: 根据超长怠速的阈值, 判断每段怠速是否为超长怠速。

Step4: 统计超长怠速的次数和时间。

6. 熄火滑行挖掘算法

熄火滑行是指汽车的发动机已经熄火, 但是汽车刹车置于空挡, 所以汽车会按照惯性继续向前行驶, 熄火滑行从节能的角度出发确实可以节约少量能源, 但是对于汽车本身的伤害确实很大, 因为当发生紧急情况需要立刻制动停车的时候, 熄火滑行可能使刹车失灵, 造成交通事故。熄火滑行状态表如下所示:

ACC 状态	车速
--------	----

挖掘算法步骤如下：

Step1: 判断当前数据是否 ACC 状态为 off，且车速是否不为 0 且小于一个 50km/h。若是，则标记该条数据为熄火滑行。

Step2: 转到 step1，直到遍历完所有数据

Step3: 将所有连续的被标记为熄火滑行的一段数据，标记为一次熄火滑行。设置最低时间限制 3s，当该次熄火滑行的时长大于这个限制才被标记为有效的熄火滑行。

Step4: 累加并记录熄火滑行的总时间与次数。

7. 车速稳定性挖掘算法

车速稳定性是指车辆再行驶过程中车辆的稳定程度，在车辆行驶过程中，如果车速波动过大，可能造成油耗量瞬时剧增，增加车辆的稳定性和不安全性，结合查阅的参考文献本文对车速提取均值、标准差来衡量车辆的运行状态，标准差来表示车速的离散程度，其公式具体如下：

$$\begin{cases} \bar{v} = \frac{1}{N} \sum_{i=1}^N v \\ \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (v_i - \bar{v})^2} \end{cases}$$

当标准差越大的时候，表示车速的离散度越大，车速的变化频率越大，造成交通事故的可能性越大。

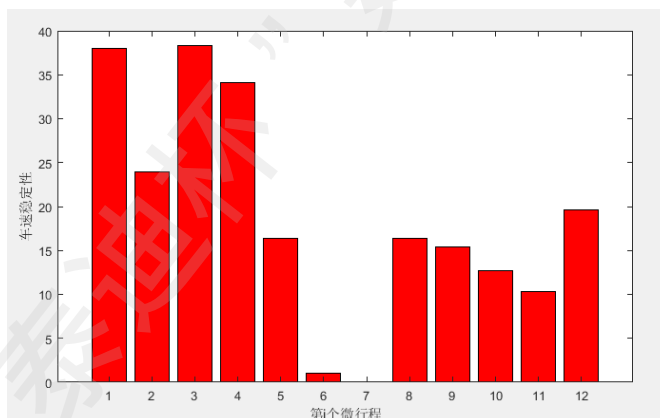


图 5-13 AA00002 车速稳定性变化

以 AA00002 为例，可以看出，其每个微行程下的车速稳定性大致保持在 20 左右，说明该辆车驾驶员开车时的车速稳定性较为一般。

8. 急变道挖掘算法

急变道指当汽车按一定速度行驶在道路上时，驾驶员迅速改变方向来改变车道，这种行为容易造成交通事故。本文将急变道定义为当驾驶员处于一定速度区间时车辆的角加速度超出阈值即为急变道行为。

依照国家现行的道路行业交通标准等文献，得到道路宽度一般为 3.5m，将汽车每一次急变道视为从一条道变换到邻近一条道路，若汽车变道变向相隔道路，则将其视为两次或多次连续相邻变道。

同时本文认定车辆变道完成一般为 1.5 秒，根据车辆所在的速度区间，求出其规定变道时间所需要的角加速度区间，公式如下：

$$\begin{cases} L_1 = V \times T \\ \cos \varphi = \frac{L_1}{L_2} \\ \omega = \frac{\varphi}{t} \\ \alpha = \frac{\Delta \omega}{\Delta t} \end{cases}$$

其中 L_1 表示急变道时车辆的行驶长度， L_2 为车辆急变道的横向宽度， φ 为急变道的转向角， α 为角加速度。

急变道状态表如下所示：

角加速度	车速	是否超速
$\alpha > 1.9rad / s^2$	$V > 120km / h$	是
$\alpha > 2.55rad / s^2$	$120km / h > V > 90km / h$	是
$\alpha > 3.83rad / s^2$	$90km / h > V > 60km / h$	是
$\alpha > 7.76rad / s^2$	$60km / h > V > 30km / h$	是
$\alpha > 11.88rad / s^2$	$30km / h > V > 20km / h$	是
$\alpha > 28.5rad / s^2$	$20km / h > V > 10km / h$	是
α 为任意角	$10km / h > V$	否

急变道挖掘算法如下：

Step1: 若两条相邻数据时间未超过一个阈值，则说明这两条数据是连续记录的点。通过相邻两条数据的方向角计算，计算出相应的角加速度

Step2: 通过该时刻的角加速度以及该时刻速度，利用 4.5 的定义，对每条数据标记上是否出现变道

Step3: 通过时间阈值判断出改变到是否为急变道

Step4: 记录急变道的时间与急变道次数。

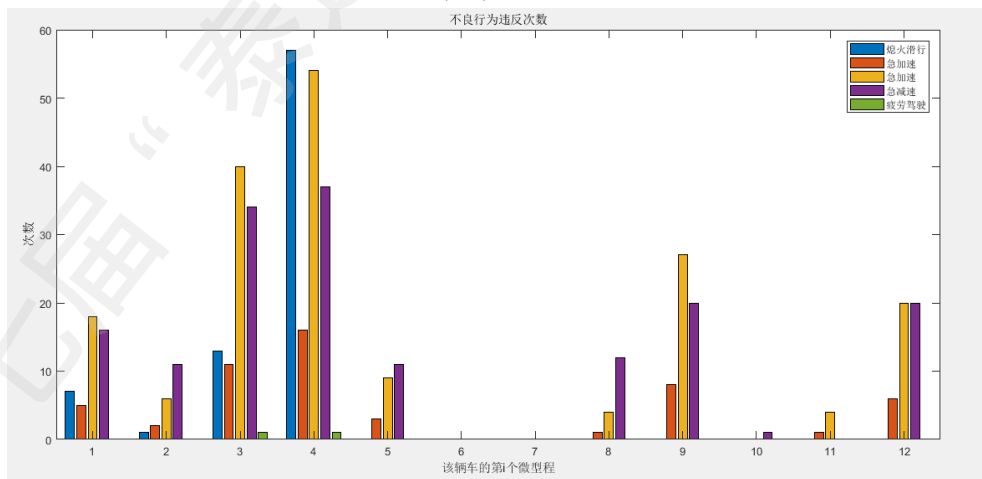


图 5-14 AA00002 微行程不良驾驶行为次数

最后以 AA00002 每一个微行程进行整体分析，可以看出该车辆在第 3、4 个微行程违反的设计安全的不良驾驶行为次数较多，在 6、7 个微行程中不存在不良驾驶行为。

5.2.3 行车安全评价模型

在驾驶行为安全的评价模型中，主要常见的方法有层次分析法、模糊评价法，熵权法等。根据参考文献[6-7]，根据交通运输行业研究部门、企业专家等的研究，若只是单一的只用一种方法建立行车安全评价模型，则会显得不够客观、科学。故本文将主客观的评价体系相结合，提出一种主成分分析与层次分析法相结合的评价模型：

$$\bar{P} = p_1^\alpha \times p_2^{1-\alpha}$$

其中 P_1 代表层次分析法评价模型得到的权重， P_2 代表主成分分析模型得到的权重， α 为系数。

层次分析法：

本文采用层次分析法解决问题的具体步骤为：

1. 驾驶员安全评价指标建立。

建立的评价指标表见 5.2.3。

2. 构造比较判断矩阵并求权重

现采用标度法构造重要程度层次：

数字	重要程度含义
1	两元素比较，其同等重要
3	两元素比较，前者比后者稍重要
5	两元素比较，前者比后者明显重要
7	两元素比较，前者比后者强烈重要
9	两元素比较，前者比后者极端重要
2,4,6,8	上述判断的中间值
倒数	若元素 i 与元素 j 的重要性之比为 a_{ij} ，则 元素 j 与元素 i 的重要性之比为 $a_{ji} = 1/a_{ij}$

根据标度法最终得到的比较矩阵如下所示：

$$M = \begin{bmatrix} 1 & 1/2 & 1/3 & 1 & 1/2 & 1/3 \\ 2 & 1 & 1 & 2 & 1 & 1/2 \\ 3 & 1 & 1 & 3 & 1 & 1 \\ 1 & 1/2 & 1/3 & 1 & 1/2 & 1/3 \\ 2 & 1 & 1 & 2 & 1 & 1/2 \\ 3 & 2 & 1 & 3 & 1/2 & 1 \end{bmatrix}$$

根据比较矩阵，对列向量归一化得到对应矩阵 Q 和得其特征向量 ω ：

$$Q = \begin{bmatrix} 0.0625 & 0.0625 & 0.05 & 0.0625 & 0.076 & 0.058 \\ 0.125 & 0.125 & 0.15 & 0.125 & 0.153 & 0.088 \\ 0.1875 & 0.125 & 0.15 & 0.1875 & 0.153 & 0.176 \\ 0.0625 & 0.0625 & 0.05 & 0.0625 & 0.076 & 0.058 \\ 0.125 & 0.125 & 0.15 & 0.125 & 0.153 & 0.088 \\ 0.25 & 0.25 & 0.15 & 0.1875 & 0.076 & 0.176 \end{bmatrix}$$

$$\omega = [0.063 \quad 0.13 \quad 0.16 \quad 0.063 \quad 0.13 \quad 0.167]$$

3. 进行一致性检验

对构造的判断矩阵进行一致性检验，并以此来确定权重是否合理，公式如下：

$$CR = \frac{CI}{RI}$$

$$CI = \frac{\lambda - n}{n - 1}$$

其中， CR 为一致性比例， CI 为一致性指标， λ 为最大特征根， n 为判断矩阵阶数。当 $CR < 0.1$ 时，即通过了一致性检验，当 $CR > 0.1$ 时，则判断矩阵不符合一致性要求，需要对该矩阵进行重新修正。

矩阵阶数	1	2	3	4	5	6	7	8
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41

经计算，一致性比例均小于 0.1，权重较为合理。

4. 指标评价权重

最终得到的安全驾驶行为权重如下：

安全驾驶行为	权重/ w
急加速	0.635
急减速	0.635
超速行驶	0.16
疲劳驾驶	0.16
熄火滑行	0.13
车速稳定性	0.13

最后根据综合评价公式算出每一辆车对应的每一个微行程的综合安全驾驶评分，其公式如下：

$$P_1 = \sum_{k=1}^n yw$$

其中 P 为最终综合评价得分， y 为其单项指标得分， w 为其单项指标权重。

主成分分析法：

由于在对驾驶行为进行评价的时候，层次分析法存在一定的主观性，故本文采用主成分分析法继续对驾驶行为进行分析。主成分分析的主要步骤如下：

Step1: 采集 p 维向量 $X = (x_1, x_2, x_3, \dots, x_p)'$ 的 n 个样品 $x_j = (x_{j1}, x_{j2}, \dots, x_{jp})'$ 列出样本矩

阵 X 。

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Step2: 对样本矩阵 X 进行变换得到 $Y = [y_{ij}]_{n \times p}$ ，其中：

$$y_{ij} = \begin{cases} x_{ij}, & \text{对正指标} \\ -x_{ij}, & \text{对逆指标} \end{cases}$$

Step3: 采用如下公式得到新矩阵 X_{ij}^* ：

$$X_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{var}(x_j)}} \quad i=1,2,\dots,n; \quad j=1,2,\dots,p$$

其中 \bar{x}_j 和 $\sqrt{\text{var}(x_j)}$ 分别是第 j 个变量的平均值和标准差。最终将得到的数据进行标准化处理，得到标准化矩阵 Z ：

$$Z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix}$$

Step4: 计算矩阵的样本相关系数阵 R ：

$$R = [r_{ij}]_{p \times p} = \frac{Z^T Z}{n-1}$$

Step5: 求出特征值：

$$|R - \theta I_p| = 0$$

最后可以解得 p 个特征值 $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_p \geq 0$ 。

Step6: 确定 m 值，使信息的利用率达到 80% 以上，确定方法如下：

$$\frac{\sum_{j=1}^m \theta_j}{\sum_{j=1}^p \theta_j} \geq 0.8$$

对每个 θ_j ， $j=1,2,\dots,m$ 。求解方程组 $Rb = \theta_j b$ ，得单位向量 b_j^0 ：

$$b_j^0 = \frac{b_j}{\|b_j\|}$$

Step7: 最终求出 $z_i = (z_{i1}, z_{i2}, \dots, z_{ip})'$ 的 m 个主成分分量

$$u_{ij} = z_i b_j^0, j = 1, 2, \dots, m$$

得到最终决策矩阵，以编号车辆为 AA00002 为例，本文最终得到其第一个微行程下的驾驶行为评价模型权重为：

安全驾驶行为	权重/ w
急加速	0.950
急减速	0.950
超速行驶	0.915
疲劳驾驶	0.927
熄火滑行	0.001
车速稳定性	0.762

最后依据主成分分析得到的评价公式算出每一辆车对应的每一个微行程的综合安全驾驶评分，其公式如下：

$$P_2 = \sum_{k=1}^n yw$$

其中 P 为最终综合评价得分， y 为其单项指标得分， w 为其单项指标权重。

5.2.4 评价结果分析

以 AA00002 车辆每一个微行程为例：

采用上述模型将两种评价方法结合在一起，就得到了最终评价结果，最终得到其指标权重雷达图如下图所示：

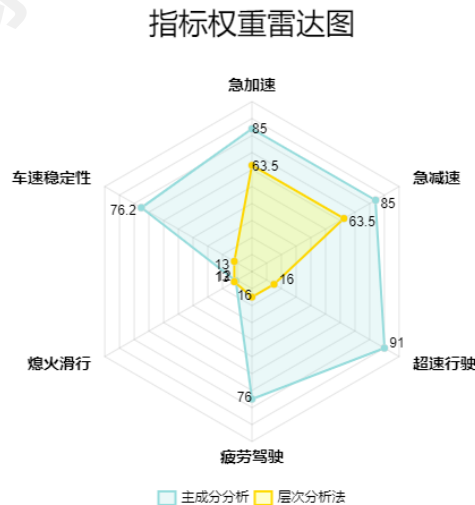


图 5-15 指标权重雷达图

此时本文对其每一个微行程进行驾驶行为评价：

微行程	xhxcnt		overspeed		jiAddspeed		jijiansunum		pilaonum		chesustables	ZH_Score
	熄火滑行		超速		急加速		急减速		疲劳驾驶			
	次数	时长	次数	时长	次数	时长	次数	时长	次数	时长		
1	0	0	7	2261	5	18	16	104	0	0	38.0190823	68.17315
2	0	0	0	0	2	6	11	52	0	0	23.94434735	81.55462
3	0	0	13	7399	11	40	34	187	1	12732	38.28739618	25.16974
4	0	0	57	4144	16	54	37	248	1	16641	34.10123228	11.85477
5	0	0	0	0	3	9	11	49	0	0	16.40027064	83.06709
6	0	0	0	0	0	0	0	0	0	0	1	96.25
7	0	0	0	0	0	0	0	0	0	0	0	96.25
8	0	0	0	0	1	4	12	62	0	0	16.43030248	85.37036
9	0	0	0	0	8	27	20	98	0	0	15.3547224	68.3015
10	0	0	0	0	0	0	1	4	0	0	12.65393513	95.47584
11	0	0	0	0	1	4	0	0	0	0	10.32511862	94.69597
12	0	0	0	0	6	20	20	106	0	0	19.65078881	71.38662

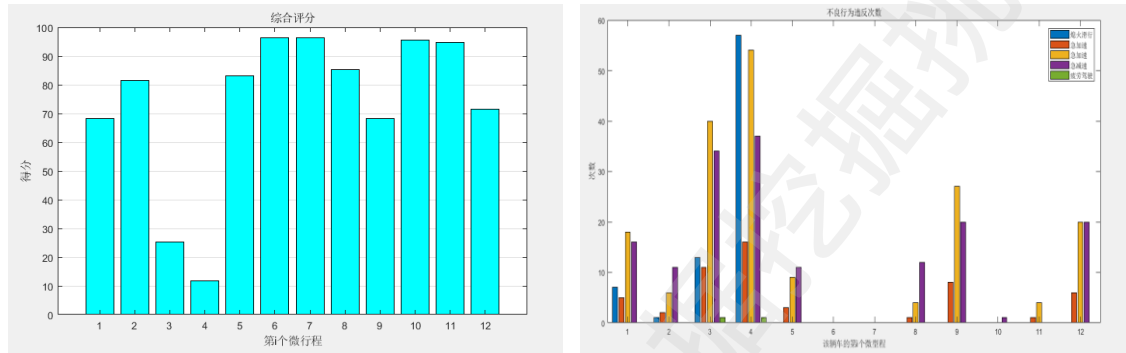


图 5-16 AA00002 微行程评价得分

由评价结果得知，因为 AA00002 的 3、4 个微行程不良驾驶行为次数和累计时间都较多，故其 3、4 段微行程评价得分较低，其 6、7 个行程不良驾驶行为很少，故分最高，但是依旧没有到达满分，可能是由于其车速稳定性不够造成的。

以 450 辆车整体为例：

根据总共 450 辆车，我们最终得到这 450 辆车的驾驶行为评价结果，具体结果见附件，评价结果分析如下：



图 5-17 450 辆车整体评价得分

由图可以看出,层次分析法与主成分相结合的 450 辆车综合评价结果与其各自单独评价结果相比,更加科学、客观。同时,大部分最终综合评分在 60 分以上,60 分以下的评分较少,而 60 分以下的车辆驾驶行为确实存在较多不良驾驶行为次数和不良驾驶时间。

5.3 改进的行车安全综合评价模型

针对问题三,需要综合考虑运输车辆的安全,效率和节能,充分结合自然气象条件和道路状况为运输车辆管理等部门建立行车安全的综合评价指标体系与模型。

首先对附件二的自然气象与道路数据进行离散化处理(见 4.5),根据这些数据中的省、市、县地理位置找出其对应的经纬度坐标,然后与给定的 450 辆车辆的行车轨迹进行匹配,找出每辆车所在的地理环境。综合考虑道路环境因素,重新定义不良驾驶行为的挖掘算法阈值。同时,为了从运输车辆的效率方面评价,结合自然气象条件和道路状况,引入了行驶侧滑可能性、行车时间利用率、日均行驶里程这 3 个指标。

行驶侧滑可能性:

在风力较大的条件下,由于路面冰雪湿滑等状况可能引起车辆轮胎侧向摩擦力小于侧风力,导致车辆失控或者被侧向风吹翻。

在理想情况下,假设风是以柱状前进的,并假定其吹入一个底端封闭的管中,当风到达底端后其速度很快减为零,散到管底周围的力忽略不计。在管中取一段长为 Δh 的风柱,空气密度为 ρ ,风速为 v ,管底面积为 s ,则在 Δt 时间内会有

$$m = \rho * s * \Delta h = \rho * s * (v * \Delta t)$$

的空气冲向管子的底部。若设底面积受到风力为 f ,由动量定理知

$$f * \Delta t = m * v - m * 0$$

即:

$$f * \Delta t = \rho * s * (v * \Delta t) * v$$

得风力为,

$$f = \rho * s * v^2$$

根据运动学的相对性,此时的速度应该为风速和运输车的行驶速度的合速度,由平行四边形法则可知:

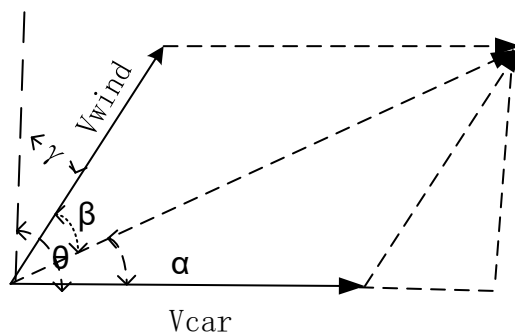


图 5-18 行驶侧滑分析图

通过计算可得,合加速度为

$$v = \sqrt{v_{wind}^2 + v_{car}^2 + 2 * v_{wind} * v_{car} * \cos(\alpha + \beta)}$$

而 β 的大小为:

$$\beta = \begin{cases} \tan^{-1} \left(\frac{v_{wind} * \sin(\alpha + \beta)}{v_{wind} * \cos(\alpha + \beta) + v_{car}} \right), & \frac{v_{wind} * \sin(\alpha + \beta)}{v_{wind} * \cos(\alpha + \beta) + v_{car}} \geq 0 \\ 180 + \tan^{-1} \left(\frac{v_{wind} * \sin(\alpha + \beta)}{v_{wind} * \cos(\alpha + \beta) + v_{car}} \right), & \frac{v_{wind} * \sin(\alpha + \beta)}{v_{wind} * \cos(\alpha + \beta) + v_{car}} < 0 \end{cases}$$

运输车辆轮胎与路面的附着力可以表示为:

$$F_f = \mu * M * g$$

其中, μ 表示为轮胎与路面的附着系数, 不同天气条件下, 取值不同, 一般情况下, 雨雪天气下的附着系数约为晴朗天气下的 $1/8 \sim 1/4$, M 表示运输车的载重量, g 为重
力加速度, 取值为 $9.8 m/s^2$ 。

运输车发生侧滑的条件为:

$$F_f \leq f * \sin(\beta)$$

日均行驶里程:

$$L_{ed} = \frac{L_{end} - L_{start}}{T}$$

其中 L_{ed} 表示日均行驶里程, L_{start} 表示运输车辆开始的里程数, L_{end} 表示运输车辆
结束时的里程数, T 表示总天数。

行车时间利用率:

$$\kappa = \frac{T_{move}}{T_{move} + T_{stop}}$$

其中 κ 表示行车利用率, T_{move} 表示一个微行程内运输车辆的运行的时长, T_{stop} 表示
一个微行程内运输车辆的休息时长。

5.3.1 改进后的评价指标体系与得分标准

所以在 5.2.1 中提出的行为数据特征评价指标就需要作出进一步改进。改进后的驾
驶行为安全与节能评价指标体系如下:

行车安全评价指标	一级评价指标	二级评价指标
	急加速\急减速	累计时长
		加速总次数
超速行驶	累计时长	

		超速总次数
	疲劳驾驶	累计时长
		疲劳驾驶总次数
	熄火滑行	累计时长
		滑行总次数
	车速稳定性	标准差
	急变道	累计时长
		滑行总次数
侧滑	累计时长	
	滑行总次数	
节能驾驶评价指标	一级评价指标	二级评价指标
	急加速\急减速	累计时长
		加速总次数
	超速行驶	累计时长
		超速总次数
	疲劳驾驶	累计时长
		疲劳驾驶总次数
	超长怠速	累计时长
超长怠速总次数		
怠速预热	累计时长	
	怠速预热总次数	
车速稳定性	标准差	
高效驾驶评价指标	一级评价指标	二级评价指标
	超长怠速	累计时长
		超长怠速总次数
	怠速预热	累计时长
		怠速预热总次数
	车速稳定性	标准差
	行车时间利用率	行车时间百分比
日均行驶里程	里程百分比	

改进之后的单项评价指标得分标准如下表：

评价指标	评分标准		
急加速	累计时长	$y_1 = 100 - t$	$y = 0.5y_1 + 0.5y_2$
	加速总次数	$y_2 = 100 - k_1n$	
急减速	累计时长	$y_1 = 100 - t$	$y = 0.5y_1 + 0.5y_2$
	加速总次数	$y_2 = 100 - k_3n$	

急变道	累计时长	$y_1 = 100 - t$	$y = 0.5y_1 + 0.5y_2$
	加速总次数	$y_2 = 100 - k_6 n$	
超速行驶	累计时长	$y_1 = 100 - t$	$y = 0.5y_1 + 0.5y_2$
	超速总次数	$y_2 = 100 - k_2 n$	
疲劳驾驶	累计时长	$y_1 = \begin{cases} 100; & t \leq 480 \\ 0; & t > 480 \end{cases}$	$y = 0.5y_1 + 0.5y_2$
	疲劳驾驶总次数	$y_2 = 100 - k_4 n$	
熄火滑行	累计时长	$y_1 = \begin{cases} 100; & 0 < t \leq 1 \\ 100 - 10t; & t > 1 \end{cases}$	$y = 0.5y_1 + 0.5y_2$
	滑行总次数	$y_2 = 100 - k_5 n$	
超长怠速	累计时长	$y_1 = \begin{cases} 100; & 0 < t \leq 1 \\ 100 - 0.5t; & t > 1 \end{cases}$	$y = 0.5y_1 + 0.5y_2$
	超长怠速总次数	$y_2 = 100 - k_7 n$	
怠速预热	累计时长	$y_1 = \begin{cases} 100; & 0 < t \leq 2 \\ 100 - 2t; & t > 2 \end{cases}$	$y = 0.5y_1 + 0.5y_2$
	怠速预热次数	$y_2 = 100 - k_8 n$	
车速稳定性	标准差	$y = \begin{cases} 100; & 0 < \sigma < 20 \\ 80; & 20 < \sigma \leq 40 \\ 60; & \sigma > 40 \end{cases}$	
平均每日里程	里程百分比	$y = \begin{cases} 100; & \vartheta \geq 200 \\ 80; & 50 \leq \vartheta < 200 \\ 60; & \vartheta < 50 \end{cases}$	
行车利用率	行车时间百分比	$y = \begin{cases} 100; & 0.6 < \psi < 1 \\ 80; & 0.3 < \psi \leq 0.6 \\ 60; & \psi \leq 0.3 \end{cases}$	

其中 σ 为标准差, t 表示时间, n 表示次数。 k_1 、 k_2 、 k_3 、 k_4 、 k_5 、 k_6 、 k_7 、 k_8 为

系数。

5.3.2 一种基于改进 K-均值聚类 and 神经网络的驾驶行为分级评价模型

在之前的评价方法中，都是采用车辆行车安全的历史数据、根据评价模型、依据评级分数计算得到的评价结果。这种评价过程很少从数据挖掘的角度研究驾驶行为与道路安全的潜在关系，故依据国内最近两年的研究情况[14-15]，提出一种基于 k-means 聚类与神经网络相结合的评价模型。其流程图如下：

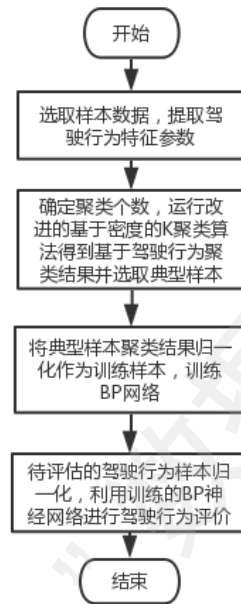


图 5-19 基于 k-means 聚类与神经网络相结合的评价模型流程图

得到安全、节能、效率各自的评价指标之后，本文采用 K-均值聚类的综合评价模型。首先确定将驾驶行为的评价结果分为 4 个等级：

顺序	等级
1	优秀
2	良好
3	合格
4	不合格

k-means 聚类的步骤如下：

Step1: 确定聚类类数 K ，由于是将评价结果分为 5 个等级，故 $K=5$ ，初始化当前确定的初始聚类中心数 $k=0$ ，当前迭代次数 $t=1$ ；并计算样本间距离 d_{ij} ，得到距离矩阵 D

Step2: 计算样本集平均差异度 M ，确定领域半径值 $\varepsilon = \frac{M}{4}$ ，计算每个样本点 X_i 的 ε 邻域内样本数量 N_i ，形成集合 N ；

Step3: 将 ε 邻域内样本数量数量 N_i 最大的样本点 X_i 作为第一个聚类中心 C_i ， $k=k+1$ ，并将矩阵 N 中相应 N_i 归 0，继续寻找矩阵 N 中最大的样本点 X_j ，计算与已有聚类中心

(C_1, C_2, \dots, C_k) 的距离, 并将矩阵 N 中相应 N_j 归 0

Step4: 若 X_j 与 (C_1, C_2, \dots, C_k) 的距离无不小于 M , 则 $k=k+1$, $C_{k+1} = X_j$, 否则返回

Step3;

Step5: 判断当前 k 的取值是否等于 K , 若是, 则进行 K 个初始聚类中心输出, (C_1, C_2, \dots, C_k) ; 若否, 则返回 Step4;

Step6: 实现聚类中心与样本对象距离的计算, 根据距离最小的对象所属类别划分, 对每个聚类中心的样本平均值计算, 并对聚类中心更新, $t=t+1$; 直至满足收敛条件;

Step7: 得到聚类集合 (C_1, C_2, \dots, C_k) 。若 C_i 包含 n 个样本, $C_{k+1} = (p_1, p_2, p_3, \dots, p_n)$, 统计该类中样本距离的最小值和最大值, $\min(D_i), \max(D_i)$;

Step8: 将差值 $\min(D_i) - \max(D_i)$ 进行划分, 划分为 m 个区间, 并统计每个区间的样本数; 找出样本数最多的区间 j , 则计算 C_i 对应的 $Eps_i = \min(D_i) + j\Delta m - \frac{\Delta m}{2}$, 其中 $\Delta m = \frac{\min(D_i) - \max(D_i)}{m}$, 将 C_i 中所有样本在 Eps_i 邻域内包含的样本数量最小值设定为

$MinPts_i$;

Step9: 重复执行上述步骤, 获得不同类中心对应的 Eps 和 $MinPts$, 选取最大的 $MinPts_i$ 作为所有类的 $MinPts$;

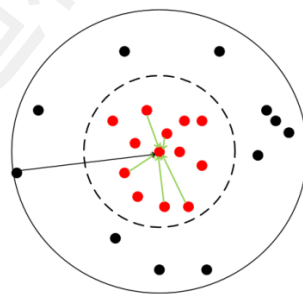


图 5-20 典型簇对象

Step10: 遍历每个类的样本点, 在 $(Eps_i, MinPts)$ 条件下, 若该样本点不是非核心点, 则标记为离群点并从该类中删除, 将去除离群点的簇集合为最终的簇集合。类中心如下表:

类序号	类别	行驶安全 score	节能 score	运输效率 score
1	合格	67.2521394847946	65.0958729777397	45.9063419466781
2	良好	79.9839926973786	78.9830919627184	73.5897373662136
3	优秀	86.650234664186	84.6852994523256	42.0241936574419
4	不合格	46.8784006209722	44.7809295775278	58.1366554990278

得到的最终结果如下图：

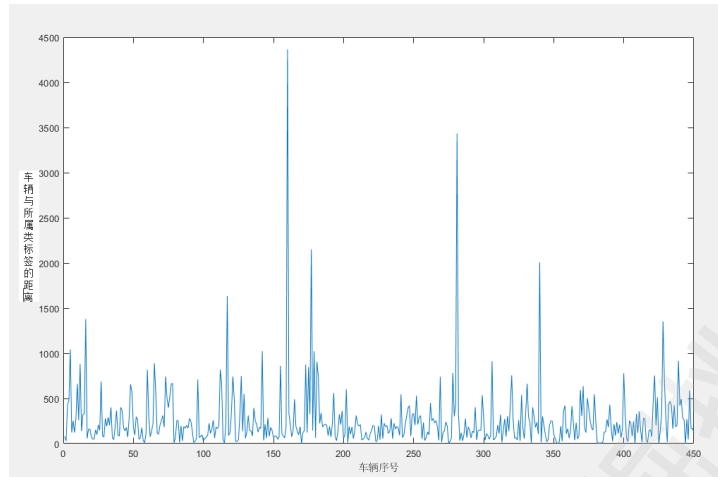


图 5-21 改进 k-means 聚类结果图

通过上图可以看出，存在距所属类中心的距离偏大的车辆，将 500 作为阈值，超过该阈值的样本点作为离群点，将其去除。

将车辆所属类信息可视化，如下图：

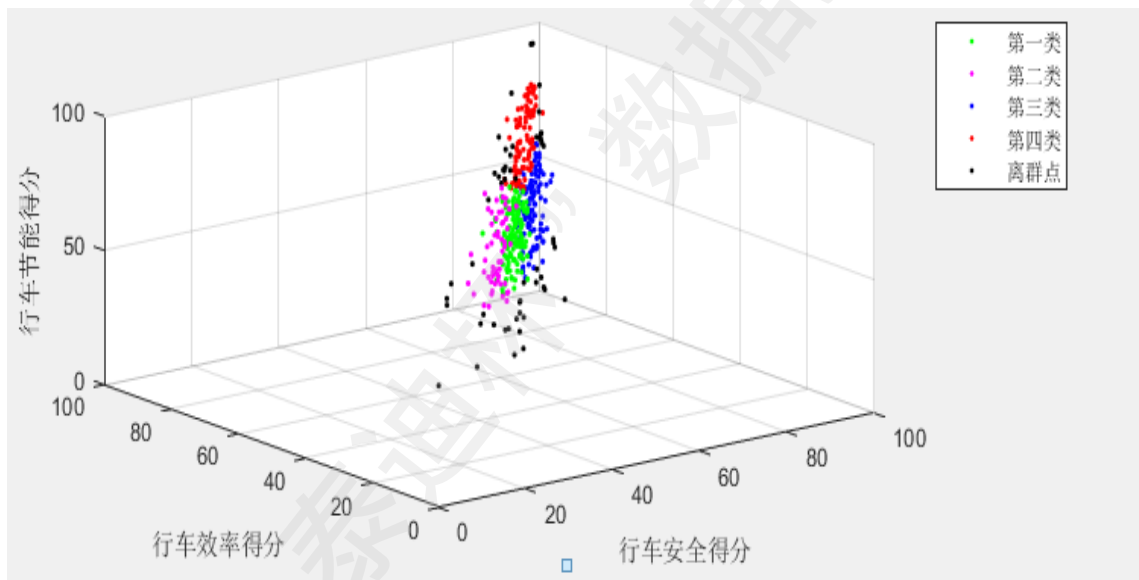


图 5-22 聚类结果

图中黑色的点为离群点。

BP 神经网络算法设计:

本文进行驾驶行为分类评价模型为输入层、输出层和隐含层所组成的一个三层 BP 神经网络。本文中驾驶行为特征参数为，即对应 个 BP 神经网络的输入节点；隐含层通过 10 个神经元组成，输出层所对应的是分类的结果，本文输出层节点数为 1，数值为驾驶行为对应的类别。将样本结果归一化作为训练样本，训练 BP 网路，利用训练完毕的 BP 神经网络进行驾驶行为分类评价。

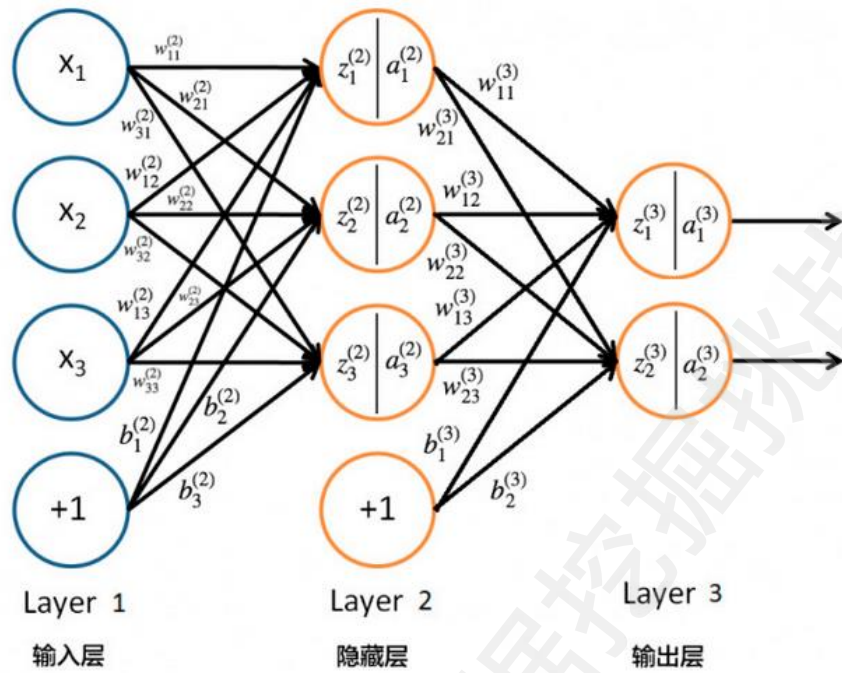


图 5-23 三层神经网络

在训练 BP 评价模型时，将典型样本的 70%数据作为训练集，15%的数据作为验证集，15%的数据作为测试集。

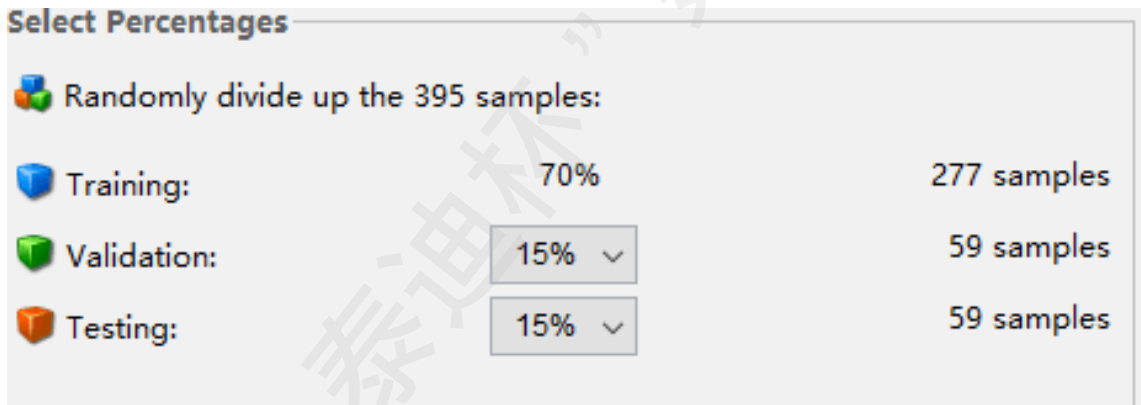


图 5-24 神经网络训练集与测试集

本模型神经网络的结构图如下，由图可知，输入变量一共是 3 个，输出变量是运输车辆类标签，隐藏层神经元的个数设置为 10，输出层神经元个数 1 个。

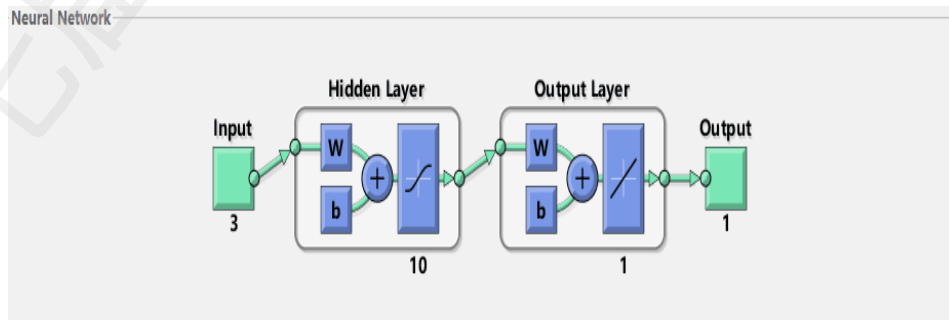


图 5-25 神经网络结构图

本文 BP 神经网络算法实现步骤如下：

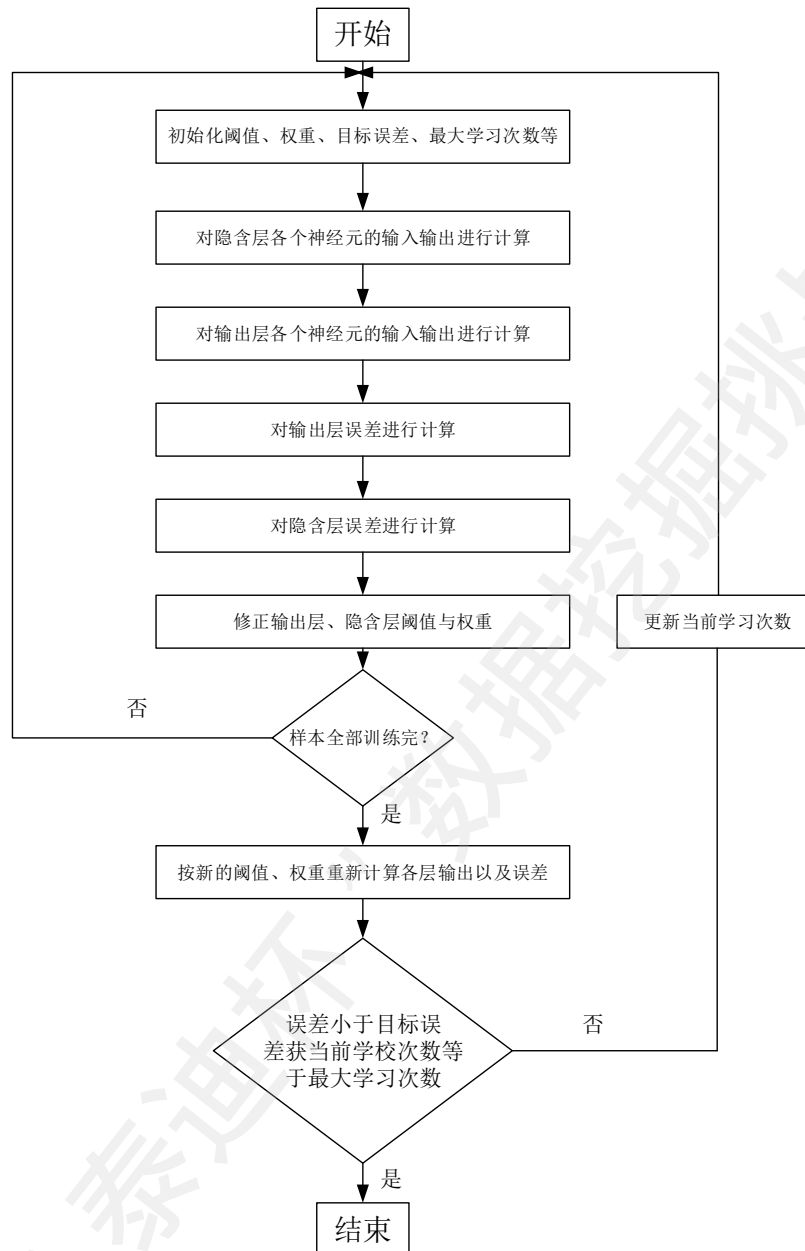


图 5-26 BP 神经网络算法流程图

通过结果分析，训练 21 次误差达到预设值，MSE 的值较小，其拟合图 R 接近 1，模型拟合的较好。使用神经网络计算得出的最终评价结果如下图所示：

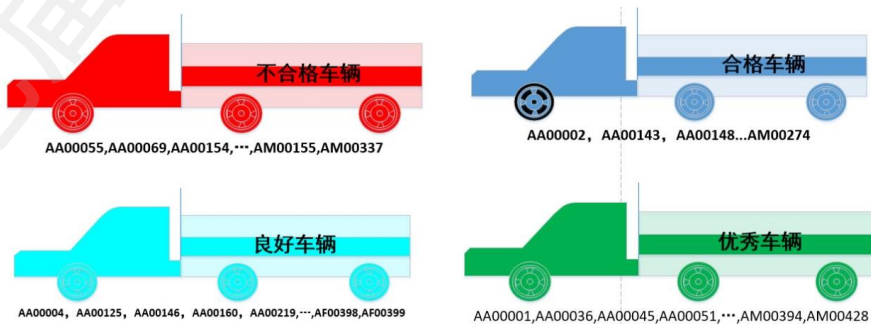


图 5-27 450 车辆评价结果图

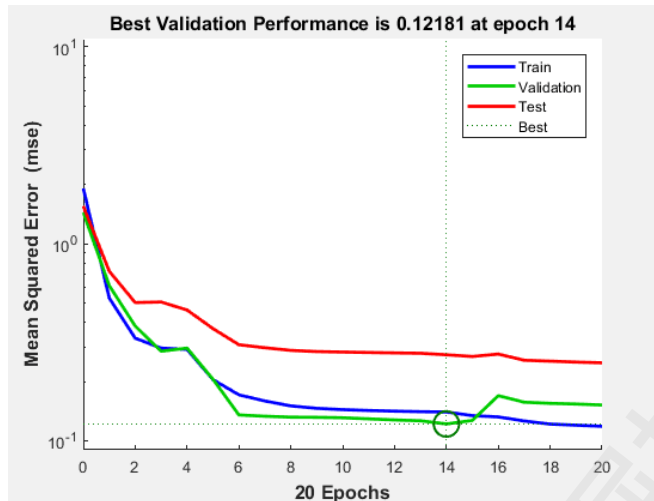


图 5-28 神经网络训练效果图

5.3.3 基于大数据分析的驾驶行为识别与评价平台搭建

经过对驾驶行为挖掘算法和评价模型研究之后，通过评价结果可以看到，本文提出的算法有效可行。为了将本文研究的算法进行系统化的应用，为帮助运输车辆管理部门开展道路运输过程安全管理的数据分析，提高运输安全管理水平和运输效率。本文构建了一个基于大数据分析的驾驶行为识别与评价平台。

平台总体框架路线：

基于大数据分析的驾驶行为识别与评价平台基于 B/S 模式的数据分析平台。系统采用 ASP.NET 技术，主要采用 MVC 前后端分离思想，后端开发语言为 C#，分别构建表现层、业务逻辑层、数据访问层，以此降低系统耦合性，达到“高内聚、低耦合”的思想。前端开发主要采用 HTML、CSS、JavaScript、Jquery 等技术，主要采用的前端框架为 Easy-UI。本平台采用的开发工具为 Microsoft Visual Studio2015 版本，使用的数据库为 SQL Server 2008。

系统界面展示：

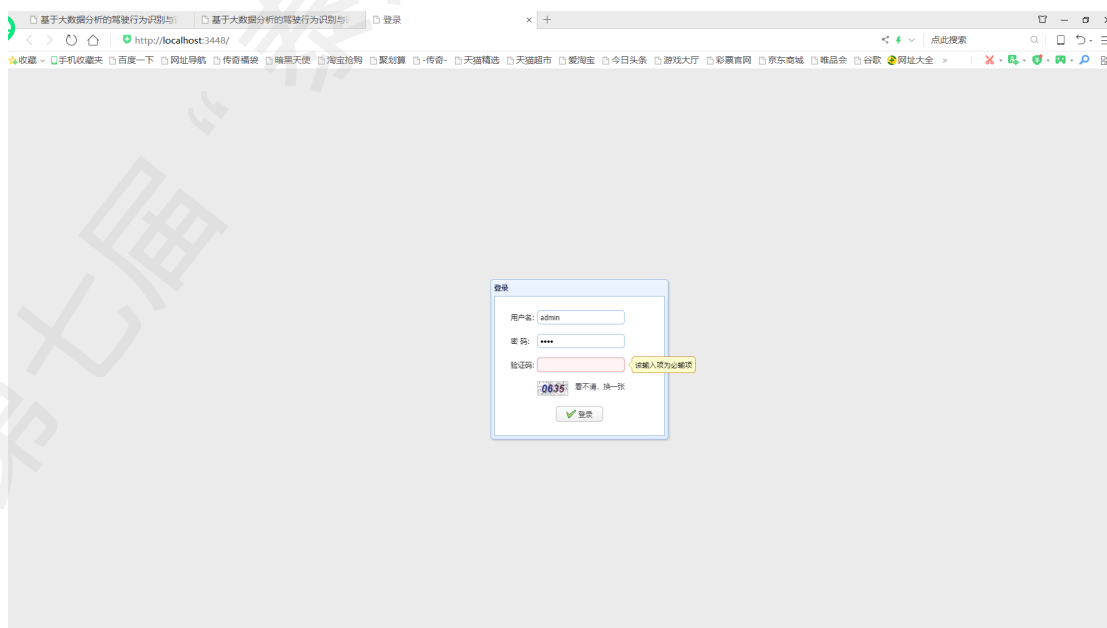


图 5-29 平台登录界面：
用户使用账号、密码、验证码即可登录。



图 5-30 平台主界面
主界面主要包括 4 个大功能点，分别为卫星定位数据、行车安全驾驶、行车节能驾驶、行车效率驾驶。



图 5-31 数据预处理界面
数据预处理界面可以对导入的行车轨迹数据按照本文的数据预处理算法(见 4.1-4.4)进行数据一键预处理。

超速车编号	超速地址经度	超速地址纬度	超速开始时间	超速结束时间
1 AB00006	115.8263	24.7014	2018-07-30T18:31:46	2018-07-30T18:32:23
2 AB00006	115.7429	24.9523	2018-07-30T18:59:21	2018-07-30T18:10:34
3 AB00006	115.4529	24.7254	2018-07-30T14:45:23	2018-07-30T14:45:59
4 AB00006	115.5482	24.9453	2018-07-30T12:49:11	2018-07-30T12:50:21
5 AB00006	115.2469	24.9543	2018-07-30T12:12:45	2018-07-30T12:13:21
6 AB00006	115.7329	24.5754	2018-07-30T11:09:55	2018-07-30T11:10:52
7 AB00006	115.4429	24.5644	2018-07-30T02:07:53	2018-07-30T02:06:25
8 AB00006	115.3229	24.3258	2018-07-31T02:08:21	2018-07-31T02:09:31
9 AB00006	115.1429	25.2574	2018-07-31T03:09:33	2018-07-30T03:10:23

图 5-32 超速驾驶识别界面

超速驾驶识别界面，系统可以根据数据预处理后的数据，进行超速挖掘算法识别。其余不良驾驶行为类似，本文不再一一展开。

6 总结

本文的主要目的利用数据挖掘与数学建模技术建立运输车辆安全驾驶行为评价模型。首先，通过对所给 GPS 轨迹数据等进行可视化，分析并提取不良驾驶行为指标，为安全评价模型的建立做准备。

首先定义疲劳驾驶，急加速，急减速，怠速预热，超长怠速，熄火滑行，超速，急变道等不良驾驶行为及相应的识别算法。将每一个路线分为多个微行程，提取出每辆车的每个微行程的不良驾驶行为累计次数和时长。为每类指标建立相应的打分体系，利用层次分析和主成分分析相结合的权重赋值法，对每辆车的每个微行程进行比较、分析及评价。然后结合天气，温度等环境因素对疲劳驾驶，超速等指标的影响，对挖掘算法相应阈值进行修改，计算出每辆车分别在安全，效率，节能等三个方面的得分情况。采用改进 k-均值聚类算法，以 450 辆车的 3 个方面的得分为属性，得到驾驶行为典型样本及其所属类别。最后设计三层 BP 神经网络，将每辆车的 3 类得分作为神经网络的输入，驾驶行为综合评价结果作为神经网络的输出。对 BP 神经网络进行训练，得到一个在线分类的 BP 神经网络分类器。该分类器模型可以对车联网所上传的大量实时驾驶行为数据进行快速分类、评价。

7 参考文献

- [1].任慧君, 许涛, 李响. 利用车载 GPS 轨迹数据实现公交车驾驶安全性分析[J]. 武汉大学学报·信息科学版, 2014, 39(6): 739-744.
- [2].GA/T 1148-2014, 道路交通安全管理规划编制指南[S].

-
- [3]. JT/T 807-2011, 汽车驾驶节能操作规范[S].
 - [4].刘应吉, 曾诚, 王书举, et al. 基于卫星定位数据的驾驶行为安全与节能评价方法[J]. 公路交通科技, 2018(1):121-128.
 - [5].刘应吉, 赵侃, 李强, et al. 基于卫星定位数据的违规驾驶行为辨识方法[J]. 公路交通科技, 2017(11):130-139.
 - [6].许书权. 基于车辆运行监控系统的驾驶行为安全与节能评价方法研究[D]. 2015.
 - [7].夏杰. 基于道路运输企业安全生产管理数据的驾驶行为安全与节能评价方法[D]. 2016.
 - [8].李敬怡. 营运车辆运行工况数据处理分析平台相关技术的研究[D]. 北京交通大学, 2014.
 - [9].刘应吉,蔡凤田, 赵侃, et al. 一种使用卫星定位数据进行车辆行驶里程计算的方法.
 - [10].薛雪莉.基于个人 GPS 轨迹数据的地图匹配算法研究[D].
 - [11].郑少波. 基于 GPS 轨迹数据的地图匹配方法[D]. 2016.
 - [12].张文. 大数据平台下地图匹配算法的研究与实现[D]. 2017.
 - [13].张鹏. 基于主成分分析的综合评价研究[D]. 南京理工大学, 2004.
 - [14].吴紫恒. 基于数据挖掘的驾驶行为与道路拥堵分析[D].中国科学技术大学,2018.
 - [15]罗正明. 公共汽车行驶参数建模及优秀驾驶行为挖掘方法研究[D].重庆大学,2017.

附录

1.建模所使用软件:

写作: Word2016 Excel2016

编程软件: MATLABb2017、python2.7、python3.6、Microsoft Visual Studio2015、 SQL Server 2008、Eclipse4.7

2.BP 神经网络训练出来的模型(其余代码见附件):

```
1. function [Y,Xf,Af] = myNeuralNetworkFunction(X,~,~)
2. %MYNEURALNETWORKFUNCTION neural network simulation function.
3. %
4. % Generated by Neural Network Toolbox function genFunction, 26-Apr-2019 00
:45:22.
5. %
6. % [Y] = myNeuralNetworkFunction(X,~,~) takes these arguments:
7. %
8. % X = 1xTS cell, 1 inputs over TS timesteps
9. % Each X{1,ts} = Qx3 matrix, input #1 at timestep ts.
10. %
11. % and returns:
12. % Y = 1xTS cell of 1 outputs over TS timesteps.
```



```

13. % Each Y{1,ts} = Qx1 matrix, output #1 at timestep ts.
14. %
15. % where Q is number of samples (or series) and TS is the number of timeste
ps.
16.
17. %#ok<*RPMT0>
18.
19. % ===== NEURAL NETWORK CONSTANTS =====
20.
21. % Input 1
22. x1_step1.xoffset = [35.10350942;32.17199784;20.66645099];
23. x1_step1.gain = [0.0308183074635528;0.0294863468819586;0.0308136548058833]
;
24. x1_step1.ymin = -1;
25.
26. % Layer 1
27. b1 = [-2.9406122800882084078;3.4693437431562306195;-4.4881600674894537306;
4.362208431962644184;2.8046094643102477129;-2.3296915663178463518;-3.21068074382875
19041;4.9782305104931685591;4.6912972713467784303;3.2208550816289007734];
28. IW1_1 = [1.7016296752312241036 2.84691111108556912626 8.8604360369978927281
;0.11714835680180338273 -5.2094980684190366205 3.0600149126849829351;1.419857148044
3125115 -0.40804106667026318611 0.77131968344628321077;-4.1884080663570601999 -0.43
355823536386323136 0.50149028672103701521;-3.6473701695130822387 -9.427459799537318
8297 -5.6061766532923877904;-2.7664516415023587825 -3.967870390245350265 0.95624422
031340428063;-3.4891704056287098368 -1.1798378349718987579 -0.18301337319191518516;
2.0967993549038945744 3.4621303293844012394 5.3915286581642423158;-0.30821278593733
125728 -7.0611549061131366756 3.0083496818564401387;1.8813272518119872334 -2.285382
5703033701089 -0.32838535638794397187];
29.
30. % Layer 2
31. b2 = -0.096104692272601668912;
32. LW2_1 = [0.57672130421682277213 -2.4667056966759011161 -0.3887490392942881
6785 0.38937260671402312839 -0.50670386064977723795 0.41262600446684338795 -0.02253
2933469148575673 -0.0071840894168166960151 2.0264063223616552989 0.1134137682735252
5614];
33.
34. % Output 1
35. y1_step1.ymin = -1;
36. y1_step1.gain = 0.6666666666666667;
37. y1_step1.xoffset = 1;
38.
39. % ===== SIMULATION =====

```

```

40.
41. % Format Input Arguments
42. isCellX = iscell(X);
43. if ~isCellX
44.     X = {X};
45. end
46.
47. % Dimensions
48. TS = size(X,2); % timesteps
49. if ~isempty(X)
50.     Q = size(X{1},1); % samples/series
51. else
52.     Q = 0;
53. end
54.
55. % Allocate Outputs
56. Y = cell(1,TS);
57.
58. % Time loop
59. for ts=1:TS
60.
61.     % Input 1
62.     X{1,ts} = X{1,ts}';
63.     Xp1 = mapminmax_apply(X{1,ts},x1_step1);
64.
65.     % Layer 1
66.     a1 = tansig_apply(repmat(b1,1,Q) + IW1_1*Xp1);
67.
68.     % Layer 2
69.     a2 = repmat(b2,1,Q) + LW2_1*a1;
70.
71.     % Output 1
72.     Y{1,ts} = mapminmax_reverse(a2,y1_step1);
73.     Y{1,ts} = Y{1,ts}';
74. end
75.
76. % Final Delay States
77. Xf = cell(1,0);
78. Af = cell(2,0);
79.
80. % Format Output Arguments
81. if ~isCellX

```

```
82.     Y = cell2mat(Y);
83. end
84. end
85.
86. % ===== MODULE FUNCTIONS =====
87.
88. % Map Minimum and Maximum Input Processing Function
89. function y = mapminmax_apply(x,settings)
90. y = bsxfun(@minus,x,settings.xoffset);
91. y = bsxfun(@times,y,settings.gain);
92. y = bsxfun(@plus,y,settings.ymin);
93. end
94.
95. % Sigmoid Symmetric Transfer Function
96. function a = tansig_apply(n,~)
97. a = 2 ./ (1 + exp(-2*n)) - 1;
98. end
99.
100. % Map Minimum and Maximum Output Reverse-Processing Function
101. function x = mapminmax_reverse(y,settings)
102. x = bsxfun(@minus,y,settings.ymin);
103. x = bsxfun(@rdivide,x,settings.gain);
104. x = bsxfun(@plus,x,settings.xoffset);
105.
```