

# 运输车辆安全驾驶行为分析

## 摘要

本文旨在基于某运输企业所采集的 450 辆运输车辆行车状态信息，结合气象条件与道路状况等信息，通过数据特征工程、数据分析与数据深入挖掘，对运输车辆驾驶行为的安全性、效率性和节能性进行综合评价。结合量化与质化数据分析，探讨运输车辆安全行驶的指导性管理政策。

针对任务一，建立了 DBSCAN 聚类模型进行异常点去噪，基于去噪后的经纬度数据绘制了待评判的十辆车的行车线路图，并且通过去噪前后的线路图对比可知，DBSCAN 聚类对该问题的去噪效果良好，得到了比较清晰的线路图。随后根据构建的指标计算了这十辆车的行车里程、平均行车速度、急加速急减速情况。

针对任务二，本文构造了（每公里）急加速次数率、（每公里）急减速次数率、速度标准差、加速度标准差、（每公里）超长怠速次数率等多个行车安全评价指标。综合比对了多个聚类算法模型的效果，最终运用谱聚类实现了 450 辆车行车状态的聚类，将 450 辆样本车辆分为了稳健型驾驶、激进型驾驶和疲劳型驾驶。并且基于指标输入与车辆的类别标签，利用 BP 神经网络模型实现了行车安全性类别的分类预测。

针对任务三，选取可能影响行车安全、节能、效率的指标，基于 SEM 模型构建驾驶员驾驶行为综合评价模型。分析还发现夜间行车的速度波动较白天更大。而根据相关分析，基于任务数据未能发现天气与驾驶行为间的显著关系。

**关键词：**行车评价；DBSCAN 聚类；谱聚类；BP 神经网络；结构方程模型

## Abstract

The purpose of this paper is to make a comprehensive evaluation of the safety, efficiency and energy saving of the driving behavior. And the guiding management policy of safe driving of transportation vehicles is discussed.

Aiming at task one, the DBSCAN clustering model is established to make abnormal point denoising, and the driving circuit diagram of 10 vehicles to be judged is drawn based on the latitude and longitude data after denoising, and the denoising effect of DBSCAN clustering on this problem is good, and a relatively clear circuit diagram is obtained by comparing the circuit diagram before and after denoising. Then according to the construction index, the driving mileage, average driving speed and rapid acceleration and deceleration of the 10 vehicles were calculated.

Aiming at task two, this paper constructs a number of driving safety evaluation indexes (per kilometer), such as the rate of acute acceleration, the rate of emergency deceleration (per kilometer), the standard deviation of speed, the standard deviation of acceleration, and the rate of extra long idling (per kilometer). The effect of multiple clustering algorithm models is compared, and the clustering of 450 vehicles is realized by using spectral clustering, and 450 sample vehicles are divided into robust driving, aggressive driving and fatigue driving. Based on the index input and Vehicle category label, the BP neural network model is used to realize the classification and prediction of driving safety category.

Aiming at task three, the indexes which may affect the safety, energy saving and efficiency of driving are selected, and the comprehensive evaluation model of driver's driving behavior is constructed by using the structural equation model. The analysis also found that the speed fluctuation of night driving was greater than during the day. According to the correlation analysis, the significant relationship between weather and driving behavior could not be found based on task data.

**Key words:** Driving evaluation; DBSCAN clustering; spectral clustering; BP neural network; structural equation model

# 目录

一、问题分析 .....	4
二、数据准备 .....	4
2.1 剔除异常样本.....	4
2.2 删除与分析无关的指标.....	5
2.3 构造分析需要的指标.....	5
2.4 标准化处理 .....	7
三、模型假设 .....	7
四、任务一 .....	7
4.1 基于 DBSCAN 对经纬度数据进行去噪.....	7
4.2 绘制运输路线图.....	10
4.3 行车状态分析.....	11
五、任务二 .....	12
5.1 聚类模型分析.....	12
5.2 三种聚类算法实证分析.....	14
5.3 基于 BP 神经网络的多分类行车安全模型.....	17
六、任务三 .....	20
6.1 探索因子分析.....	20
6.2 测量模型的构建.....	21
6.3 潜在变量之间的因果关系.....	21
七、参考文献 .....	22
八、附录 .....	23

## 一、问题分析

运输车辆具有自重大、载荷量大、行车时间长、行驶里程远、油耗高等特点，随着运输车辆与汽运交通的不断发展，道路运输行业相关部门需要基于车辆运行状态，利用车联网系统数据，对运输车辆进行监管和综合服务。对运输车辆驾驶行为的分析和基于数据分析的决策知道，有助于提高其行车安全、运输效率和节能水平。

问题给出两个附件数据，附件一包括 450 辆车的行车轨迹采集数据，包括设备号、经纬度、ACC 状态、刹车状态、GPS 速度等 13 个指标，附件二给出采集时间内全国主要城市的自然气象数据，包括地区位置信息与气候信息等 11 个指标。

问题所给的任务一要求提交给定的 10 辆车的运输路线图及行驶状态。路线追踪是行车数据观测最基本和广泛的用途。在实际判别中，要注意坐标系间的对应和离群点的处理。

任务二要求利用附件一中所给的数据指标，选取、构建相应的指标，挖掘运输车辆的驾驶行为，对其安全性进行评价。

通过阅读国内外文献我们了解到，Zoran Constantinescu 等学者 (Constantinescu Zoran, 2010) 在 2010 年选取了速度、加速度和正向加速度等参数，运用主成分分析算法和层次聚类算法研究了影响驾驶质量的各种参数因素。Eskandarian 等 (Eskandarian A, 2007) 则研究发现在疲劳状况下营运车辆司机发生危险的风险要远高于小车司机。因此我们对已有的指标进行了一定的删改增，来构建分析模型。

任务三要求结合实际气象条件，分析气象条件与上述指标对驾驶行为的影响，建立运输车辆安全、效率、节能的综合评价模型。

## 二、数据准备

### 2.1 剔除异常样本

根据后续分析中的程序运行结果发现，有两辆车的全程 GPS 速度都为 0，

本文研究的是驾驶行为特征，这些样本无法成为研究对象，故删去。

## 2.2 删除与分析无关的指标

通过问题分析明确所需要的指标，基于准确、有效、易量化的原则对给出指标进行有选择的保留。最终保留的指标有车牌号码(vehicleplatenumber)、经度(lng)、纬度(lat)、ACC 状态(acc\_state)、采集时间(location\_time)、GPS 速度(gps\_speed)、GPS 里程(mileage)。

## 2.3 构造分析需要的指标

基于分析需求构造如下指标：

### 1) 急加速

在实际道路行驶中，急速度变化会影响后车驾驶，导致一系列不安全事件。正在同样的车速下，猛踩油门会带来更高的发动静瞬时喷油量和转速，消耗更多的燃油，造成大量的燃油浪费，加重尾气污染。用相邻两次数据的 GPS 速度差值 speed\_diff 来反映车辆的速度变化情况，变化量大于 0 即为加速。判定驾驶人是否有急加速行为产生，通过参数设置对急加速的判定阈值进行设定。在本驾驶行为评价模型中，结合学者已有研究，设定前后速度变化阈值为 4km/h，统计了每辆车的急加速次数指标 speed\_plus。

### 2) 急减速

急减速行为容易导致后车追尾，且高减速操作因利用离合下压浪费了车辆本身惯动力，造成了不必要的油耗损失。与急加速类似，我们定义前后两次数据的速度差阈值为-5km/h，并统计了每辆车的急减速次数指标 speed\_minus。

### 3) 行驶里程数

所给数据中覆盖的总时长相近，我们用最晚采集时间下的 GPS 里程减去最初采集时间下的 GPS 里程，得到每辆车的行驶里程数 mileage\_diff。

### 4) 平均速度

将 gps\_speed 为 0 的记录删除，然后对每辆车里程速度求均值 avg\_speed。

### 5) 长时间怠速

怠速是指速度为 0 且加速度为 0，但燃油仍在燃烧维持发动机转动的车辆

状态。由于短暂的怠速对于车辆预热有一定的作用，而长时怠速对油耗影响较大，且影响交通和车辆自身机械安全。因此综合考虑怠速状态与怠速时间来评判驾驶行为的安全性与节能性，根据相关技巧总结，我们定义一次怠速时长超过 60s 即为一次长时怠速时间，连续出现的长怠速事件合并为一次事件。我们统计出了每辆车长怠速次数  $dscs$ 。

#### 6) 疲劳驾驶

疲劳驾驶，是指驾驶人在长时间连续行车后，产生生理机能与心理机能的失调，而在客观上出现驾驶技能下降的现象。驾驶人睡眠质量差或者不足，会影响到驾驶人的注意、感觉、知觉、思维、判断、意志、决定与运动等方面，是严重的不安全因素。由大量文献可知，大量的高速公路的交通事故的发生都与疲劳驾驶有着密切的联系。本文中疲劳驾驶是指驾驶员连续驾驶四个小时以上，期间休息时间不到 20 分钟的情况定义为疲劳驾驶，计算了每辆车的疲劳驾驶次数  $tired$ 。

#### 7) 熄火滑行

汽车熄火滑行是将发动机熄火，变速箱置于空挡，利用汽车前进的惯性滑行，我们就判断其处在熄火滑行状态。熄火滑行对于行车安全方面有着重大影响。由于熄火滑行时，空气压缩机停止工作，会造成贮气筒内没有足够的制动空气，万一发生危险情况，制动容易失灵，从而引发交通事故。在本文中当 ACC 状态为 0，经纬度又发生了位移时，则判定车辆处于熄火滑行状态，计算了每辆车的熄火滑行次数  $slip$ 。

#### 8) 标准差指标

相对于速度或速度变化净值的大小，研究其波动性更有助于分析驾驶行为是否激进，速度、加速度多变的驾驶行为往往更不安全。本文基于上述求得的平均速度与前后两次速度差值计算每辆车的速度标准差  $std\_speed$  与速度差值标准差  $std\_speeddiff$ 。

#### 9) 比率指标

同时在分析过程中我们发现，由于行驶里程不同，直接比较相关行为发生的次数不能合理地反应司机的驾驶行为，故对上述次数指标（急加速次数、急减速次数、超长怠速次数、熄火滑行次数、疲劳驾驶次数）都除以该车的行驶里程数，得到相应的（每公里）次数率，用以反应该行为的发生频率。

对 450 辆车都求出了上述指标，保存为 `xlsx` 文件“Data.xlsx”，在附件中提交。

表 2-1 指标说明

指标名称	含义	单位	指标名称	含义	单位
speed_plus	反应车辆急加速情况	km/h	std_speeddiff	反应每辆车速度变化标准差	-----
plus_rate	每公里的急加速次数	-----	tired	疲劳驾驶次数	次
speed_minus	反应车辆急减速情况	km/h	slip	熄火滑行次数	次
minus_rate	每公里的急加速次数	-----	dscs	长时怠速次数	次
avg_speed	反应每辆车平均速度	Km/h	tired_rate	每公里的疲劳驾驶次数	次/km
mileage_diff	反应每辆车行驶里程	Km	slip_rate	每公里的熄火滑行次数	次/km
std_speed	反应每辆车速度标准差	-----	dscs_rate	每公里的长时怠速次数	次/km

2.4 标准化处理

由于各指标间的量纲相差较大，影响后续建模效果，对保留下来的变量数据进行标准化处理。

三、模型假设

为了便于问题的研究，对题目中某些条件进行简化及合理假设。

- 1) 因为车辆信息未知，假设车辆因素对问题不造成影响且车辆间无差异。
- 2) 因为收集数据的间隔时间比较均匀与数据体量相比很小，因此假定所记录的数据中一段驾驶过程中前后两个记录点间隔的时间是均匀的。
- 3) 为简化模型，认定处于点火状态但没有行驶，即 acc\_state=1 且 gps\_speed=0 的状态为怠速预热状态。

四、任务一

4.1 基于 DBSCAN 对经纬度数据进行去噪

在初始绘制路线图时，我们发现存在大量 GPS 漂移噪声点，如果仅仅将经纬度导入地图中，画出的路线多在水上漂移，基于此问题，我们最初想通过计算基于每秒的经纬度的偏移筛选，如果偏移经纬度过大，大过其汽车速度每秒能走到的范围，那么我们就将其过滤掉，但是实践过发现其效果并不好，由于偏移点大多成片，基于经纬度偏移异常不好成片识别，于是希望利用密度聚

类，将成片的异常点聚成一类，进行集体删除。

DBSCAN(Density-Based Spatial Clustering of Applications with Noise)密度聚类的算法实现如下：

for 样本 D，据邻域参数 ( $\varepsilon$ , minpoints) 确定所有 core points 后存入集合 P 中，这些 core points 的  $\varepsilon$  邻域中有大于等于 minpoints 的点。

任意从集合 P 中抽取一个 core points 作为初始点，找出密度可达的样本生成聚类簇构成第一个聚类簇 C1。

将 C1 中所有 core points 从 P 中去除，再从更新的 p\* 中任意选取一个种子样本。

Repeat 直至 core points 全被选完即 P 为空集, 其中  $\varepsilon$ -邻域定义为：对样本点  $x_j$  的  $\varepsilon$ -邻域包括样本集中与  $x_j$  的距离不大于  $\varepsilon$  的样本。core points 定义为：若  $x_j$  在  $\varepsilon$ -邻域中至少包含 Minpoints 个样本，即  $x_j$  为一个 core object。

DBSCAN 的优点就是能够有效识别噪声点，密度聚类的三个主要要点为：

a) 密度直达：若  $x_j$  位于  $x_i$  的  $\varepsilon$ -邻域中，并且  $x_i$  为核心对象， $x_j$  由  $x_i$  密度直达；

b) 密度可达：若  $x_j$  位于  $x_i$  的  $\varepsilon$ -邻域中， $x_m$  又在  $x_j$  邻域中称  $x_m$  可由  $x_i$  密度可达；

c) 密度相连：存在样本集合中一点 o，若  $x_j$  与  $x_i$  均由 o 密度可达，那么  $x_j$  与  $x_i$  密度相连。

对于未去噪的 AA00002 数据路线图如下：



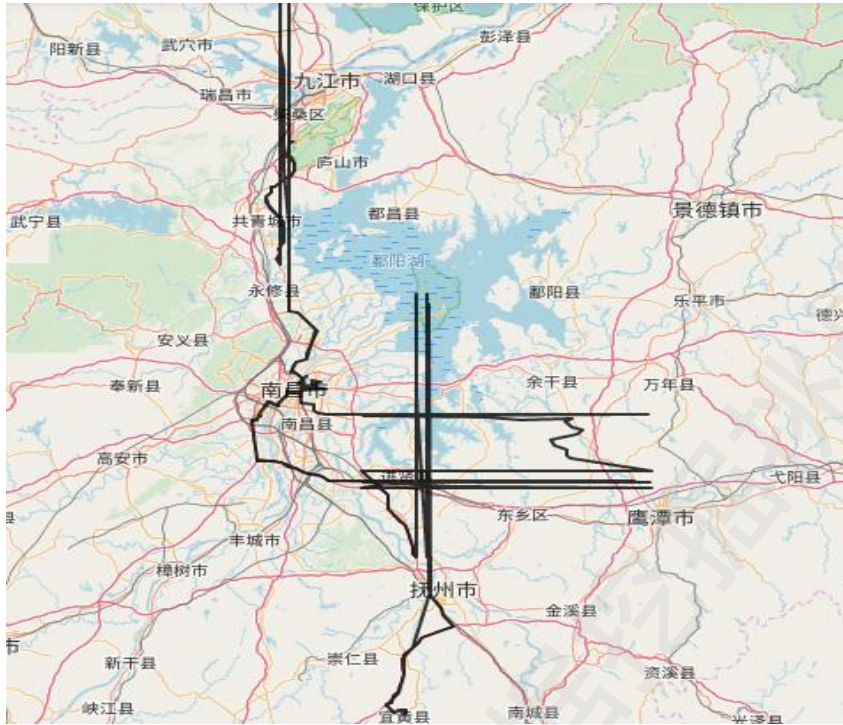


图 4-1 包含噪声点的 AA00002 行车轨迹

利用 DBSCA 去除噪声点集群后的 AA00002 路线图如下：

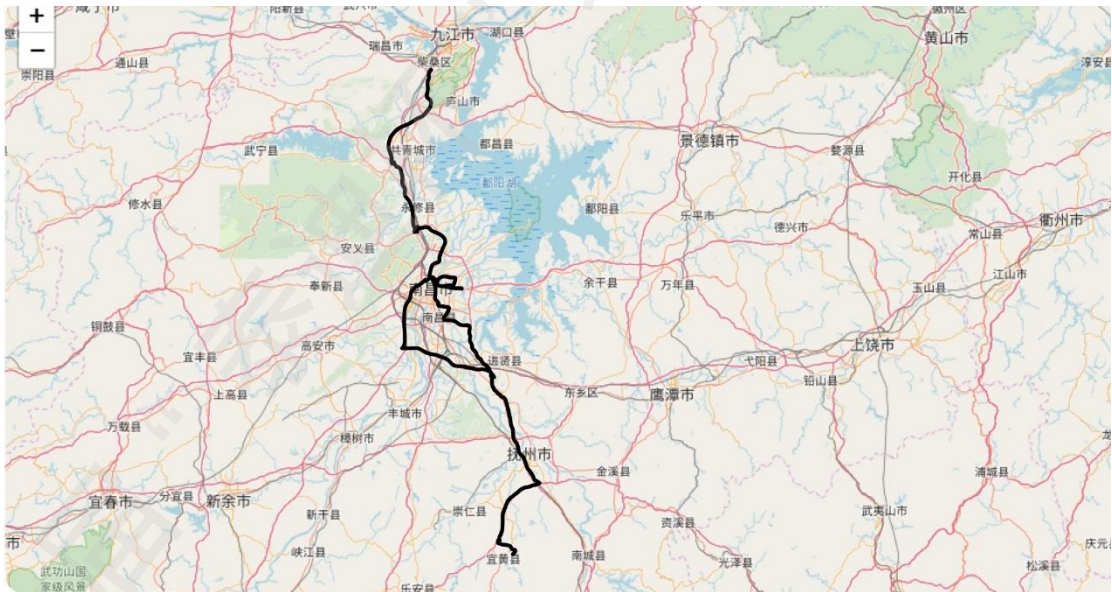


图 4-2 去除噪声点的 AA00002 行车轨迹

非常明显的能够发现噪声部分去除了东边很大一块 GPS 漂移块，路线比较规整了，路线的两端分别为江西抚州与九江，期间多次往返。

4.2 绘制运输路线图

除上文所示的 AA00002 号车外，其余九辆车的行驶线路图与大致路线说明如下：

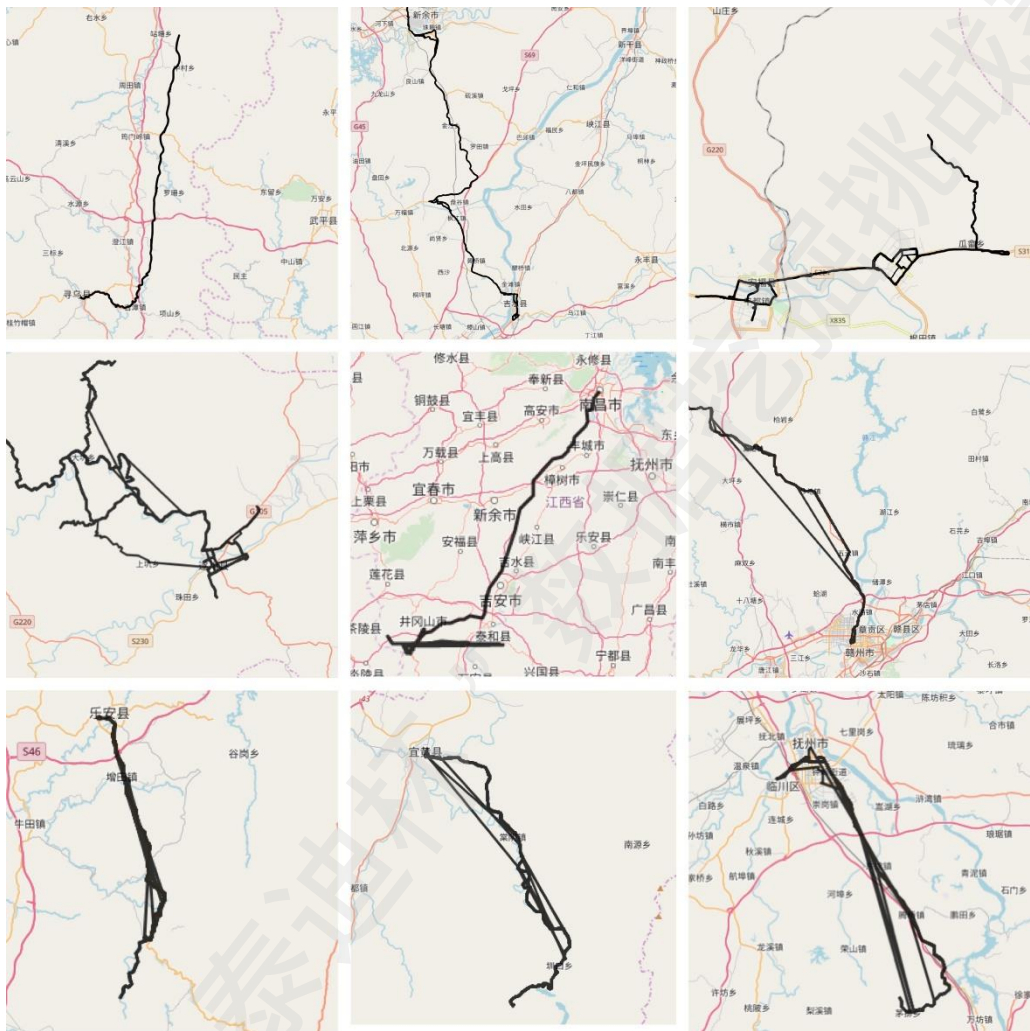


图 4-3 2-10 号车辆行车路线图

左上角为第二辆车，车牌为 AB00006，右下角为第十辆车，车牌为 AF00373，以此类推。

表 4-1 2-10 号车辆行车路线说明

编号	车辆	路线说明
2	AB00006	路线端点分别为江西省赣州市寻乌县行驶与站塘乡附近
3	AD00003	路线端点分别文江西省吉安市吉水县与江西省新余市
4	AD00013	路线途径江西省吉安市遂川县、上坑乡、大坊乡

5	AD00053	路线端点分别为江西省井冈山市和江西省南昌市
6	AD00083	路线途径江西省吉安市安福县平都镇，沿 S318 至瓜畲乡
7	AD00419	路线端点分别为江西省赣州市与江西省吉安市遂川县
8	AF00098	路线途径江西省抚州市乐安县、增田镇
9	AF00131	路线端点分别为江西省抚州市宜黄县与圳口乡南部
10	AF00373	路线端点分别为江西省抚州市临川区与茅排乡

#### 4.3 行车状态分析

通过计算我们可以得出指定十辆车的行车状态指标值，以 AA00002 为样例，该车从 8 月 4 号到 8 月 7 号期间，共行驶了 720KM (mileage\_diff)，急加速次数 (count\_plus) 为 5 次，急减速次数 (count\_minus) 为 1 次，三天内平均行驶速度 (avg\_speed) 为 61.3KM/h，疲劳驾驶次数 (tired) 1 次，熄火滑行 (slip) 1 次，超长怠速 (dscs) 19 次。

表 4-2 指定车辆行车状态

车辆编号	count_plus	count_minus	avg_speed	mileage_diff	tired	dscs	slip
AA00002	5	1	61.28409009	720	1	19	1
AB00006	43	26	37.18666343	1121	0	33	4
AD00003	15	17	50.41252935	415	0	12	2
AD00013	32	47	39.4037132	1056	1	42	1
AD00053	36	27	51.89650137	769	0	10	9
AD00083	33	28	41.73546366	742	0	2	5
AD00419	13	14	46.89909818	607	1	9	6
AF00098	66	67	50.27750092	957	0	33	20
AF00131	10	11	39.01510213	565	0	8	6
AF00373	13	13	38.51146771	584	1	17	4

## 五、任务二

### 5.1 聚类模型分析

#### 5.1.1 K-Means 聚类

K-means 算法是硬聚类算法，是典型的基于原型的目标函数聚类方法的代表，它是数据点到原型的某种距离作为优化的目标函数，利用函数求极值的方法得到迭代运算的调整规则。K-means 算法以欧式距离作为相似度测度，它是求对应某一初始聚类中心向量  $V$  最优分类，使得评价指标  $J$  最小。算法采用误差平方和准则函数作为聚类准则函数。

K-means 算法优缺点：

a. 优点：

- ①算法简单、快速，是解决聚类问题的一种经典算法
- ②对处理大数据集，该算法保持可伸缩性和高效性
- ③当簇接近高斯分布时，它的效果较好。

b. 缺点：

- ①在 K-means 算法中  $K$  值的选定是非常难以估计的；
- ②在 K-means 算法中，首先需要根据初始聚类中心来确定一个初始划分，然后对初始划分进行优化。这个初始聚类中心的选择对聚类结果有较大的影响；
- ③若簇中含有异常点，将导致均值偏离严重（即：对噪声和孤立点数据敏感）；
- ④不适用于发现非凸形状的簇或者大小差别很大的簇。

#### 5.1.2 高斯混合聚类

高斯混合聚类通过选择成分最大化后验概率来完成聚类。与 k-means 聚类相似，高斯混合模型也使用迭代算法计算，最终收敛到局部最优。使用高斯混合模型的聚类属于软聚类方法，各点的后验概率提示了各数据点属于各个类的可能性。

高斯混合聚类的优缺点：

a. 优点:

①高斯混合聚类应用最为广泛;

②收敛速度快;

③高斯混合聚类能扩展以用于大规模的数据集

b. 缺点:

①倾向于识别凸形分布、大小相近、密度相近的聚类;

②中心选择和噪声聚类对结果影响大

### 5.1.3 谱聚类

谱聚类的主要思想是把所有的数据看做空间中的点, 这些点之间可以用边连接起来。距离较远的两个点之间的边权重值较低, 而距离较近的两个点之间的边权重值较高, 通过对所有数据点组成的图进行切图, 让切图后不同的子图间边权重和尽可能的低, 而子图内的边权重和尽可能的高, 从而达到聚类的目的。

(1) 谱聚类算法的流程:

输入: 样本集  $D = (x_1, x_2, \dots, x_n)$ , 相似矩阵的生成方式, 降维后的维度  $k_1$ , 聚类方法, 聚类后的维度  $k_2$

输出: 簇划分  $C = (c_1, c_2, \dots, c_{k_2})$ .

1) 根据输入的相似矩阵的生成方式构建样本的相似矩阵  $S$

2) 根据相似矩阵  $S$  构建邻接矩阵  $W$ , 构建度矩阵  $D$

3) 计算出拉普拉斯矩阵  $L$

4) 构建标准化后的拉普拉斯矩阵  $D^{-1/2} L D^{-1/2}$

5) 计算  $D^{-1/2} L D^{-1/2}$  最小的  $k_1$  个特征值所各自对应的特征向量  $f$

6) 将各自对应的特征向量  $f$  组成的矩阵按行标准化, 最终组成  $n \times k_1$  维的特征矩阵  $F$

7) 对  $F$  中的每一行作为一个  $k_1$  维的样本, 共  $n$  个样本, 用输入的聚类方法进行聚类, 聚类维数为  $k_2$ 。



8) 得到簇划分  $C = (c_1, c_2, \dots, c_{k2})$ .

(2) 谱聚类算法的优缺点:

a. 优点:

① 谱聚类只需要数据之间的相似度矩阵, 因此对于处理稀疏数据的聚类很有效。这点传统聚类算法比如 K-means 很难做到。

② 由于使用了降维, 因此在处理高维数据聚类时的复杂度比传统聚类算法好。

b. 缺点:

① 如果最终聚类的维度非常高, 则由于降维的幅度不够, 谱聚类的运行速度和最后的聚类效果均不好。

② 聚类效果依赖于相似矩阵, 不同的相似矩阵得到的最终聚类效果可能很不同。

## 5.2 三种聚类算法实证分析

利用 Python 软件对 450 辆车的不良驾驶行为进行聚类分析, 在无法判断选取何种聚类算法的情况下, 将这三种聚类算法都进行可视化分析来加以判断。

### 5.1.1 K-Means 聚类

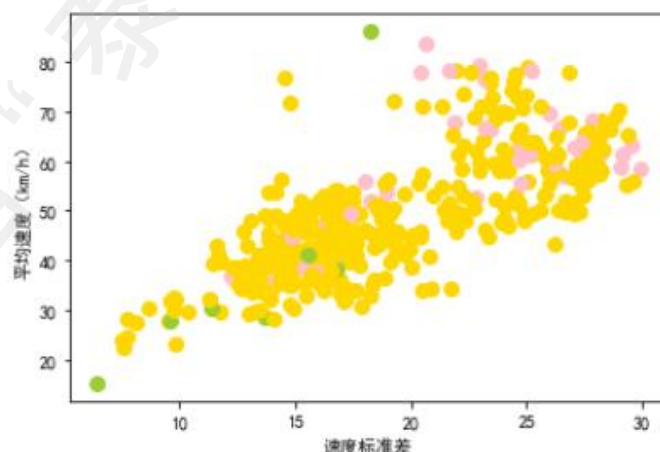


图 5-1 K-Means 聚类结果可视化

利用 k-means 聚类来对不良驾驶行为进行聚类分析, 图中虽有三个类别,

但这三个类别并不能很好地聚集为三类，黄色类别说明平均速度大的驾驶行为速度波动也越大，而粉色类别也同样说明这个现象，绿色类别所分布的较离散，不能分辨出它的状态类型，同时黄色类别与粉色类别也不能较理想地分离开，所以在这里我们不选用 k-means 聚类算法来进行研究分析。

5.1.2 高斯混合聚类

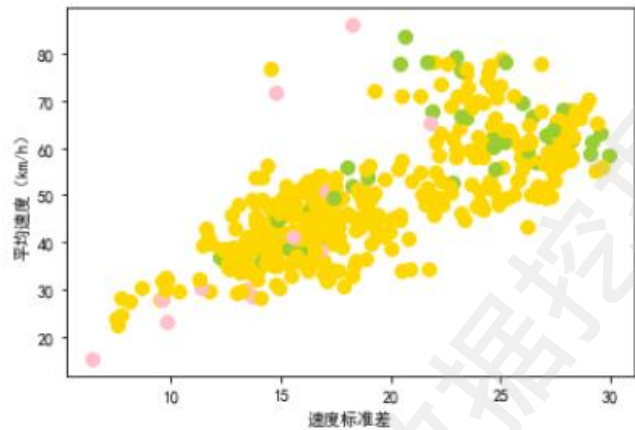


图 5-2 高斯混合聚类结果可视化

利用高斯混合聚类算法来对不良驾驶行为进行聚类分析，图中虽有三个类别，但这三个类别同样并不能很好地聚集为三类，黄色类别说明平均速度大的驾驶行为速度波动也越大，而绿色类别也同样说明这个现象，粉色类别所分布的较离散，不能分辨出它的状态类型，同时黄色类别与绿色类别也不能较理想地分离开，所以在这里我们也不选用高斯混合聚类算法来进行研究分析。

5.1.3 谱聚类

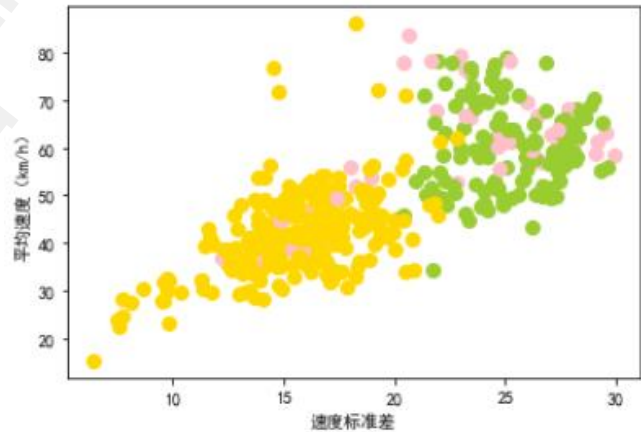


图 5-3 谱聚类结果可视化 1

在否定了前两种聚类算法之后，我们探索使用谱聚类来进行车辆驾驶行为聚类的可视化分析。从速度标准差—平均速度的关系图中可以看出驾驶行为能够较好地分成两个类别，尽管粉色代表的类别不是很能清晰地从图中直观地看出来。由黄色代表的类别在这里我们可以看到，车辆在速度标准差较小的情况下，其行驶过程中的平均速度也相对较小，这从驾驶行为类别上可以将其判断为“稳健型驾驶”；由绿色代表的类别处于速度标准差较大，同时在行驶过程中的平均速度也较大的情形，也就是说在行驶过程中车辆驾驶员倾向于行驶速度过快，而且急加速急减速情况的发生也较频繁，这类别行为我们将其判断为“激进型驾驶”；而对于粉色所代表的类别，我们根据平均速度-疲劳驾驶次数率的关系图来分析判断。

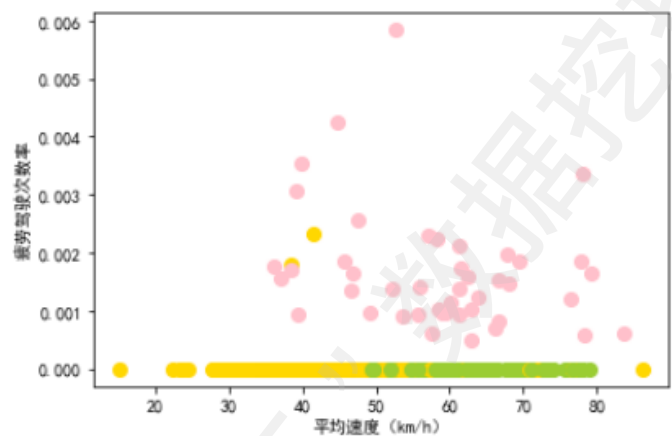


图 5-4 谱聚类结果可视化 2

在这张图中，我们发现平均速度在保持在 40-80km/h 之间的疲劳驾驶次数率较高，而在这个平均速度区间内，我们可以从图中看出粉色所代表的点集聚成一个类别，它不同于其他两类的点，能够将其准确地加以分析判别，在这里我们将其判断为“疲劳型驾驶”。

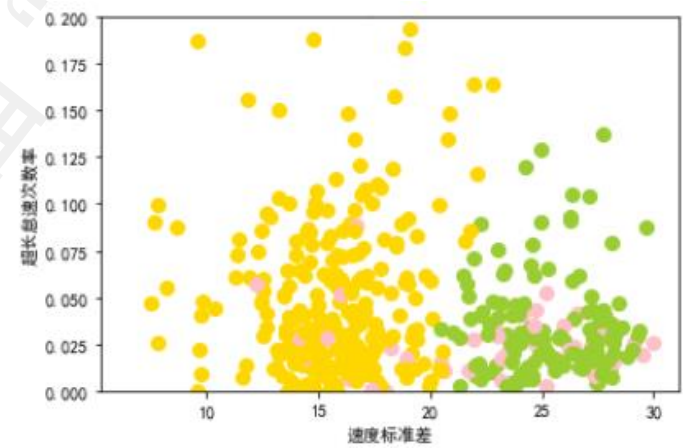


图 5-5 谱聚类结果可视化 3



进一步来判断速度标准差—超长怠速次数率的聚类关系图,“稳健型驾驶”人群的速度标准差较小,意味着其速度波动较小,但是其超长怠速次数率相对较大,这说明驾驶员在行车中遇到交通信号灯、障碍物等临时停车并保持发动机怠速运转,而超长怠速次数率大表明驾驶员在行驶过程中相对谨慎,宁可保持长一点的怠速时间,也不去超车或者急加速等以免发生交通事故;“激进型驾驶”人群的速度标准差较大,意味着其速度波动大,其超长怠速次数率相对较小,这说明这类人群会比较倾向于速度快,“走捷径”的方式去行驶,同时这也是导致事故多发的一类人群;“疲劳型驾驶”人群的速度标准差范围分布较大,即速度波动范围在 12~30 之间,从速度波动范围可以体现出这类人群行驶过程中驾驶行为的不稳定性,侧面反映了疲劳驾驶人群的显性特征,同时其超长怠速次数率更小,这表明了人在疲劳状态下反而会失去怠速的耐心,变得相对急躁而导致怠速次数率低的不良驾驶行为。速度标准差—超长怠速次数率的聚类关系图进一步验证了驾驶行为的判断类别的合理性,从而为后文展开深入地研究做好铺垫。

### 5.3 基于 BP 神经网络的多分类行车安全模型

利用谱聚类,我们将驾驶行为分为三个类别,并且给每一辆车贴上标签分别为稳健型驾驶,激进型驾驶,疲劳型驾驶。利用谱聚类客观分类出来的标签,根据 BP 神经网络多分类建立安全行车模型

BP 神经网络是一种多层的前馈神经网络,包括输入层、输出层与隐藏层。主要特点为:信号是前项传播的,误差是反向传播的。具体来说,对于本文只含一个隐层的神经网络模型,BP 神经网络过程主要分为两个阶段:第一阶段是信号的前向传播,从输入层经过隐含层,最后到达输出层;第二阶段是误差的反向传播,从输出层到隐含层,最后到输入层,依次调节隐含层到输出层的权重与偏置,输入层到隐含层的权重与偏置。本文建立的 BP 网络采用的激活函数是 Sigmoid 函数。

主要步骤如下流程图。第一步:网络初始化。给各连接权值分别赋一个区间  $(-1, 1)$  内的随机数,设定误差函数为  $e$ ,给定计算精度值  $\varepsilon$  和最大学习次数  $M$ 。第二步:随机选取第  $k$  个输入样本及对应期望输出。第三步:计算隐含层各神经元的输入与输出。第四步:利用网络期望输出与实际输出,计算误差函数对输出层的各神经元的偏导数。第五步:根据误差偏导数与隐含层的各神经元的输出来修正连接权值。第六步:计算全局误差。第七步:判断网络误差是否满足要求。当误差达到预设精度或学习次数大于设定的最大次数,则算法结束。否则,选取下一个学习样本及对应的期望输出。



图 5-6 BP 神经网络主要流程图

在本文中，我们查阅国内外研究文献，Srdjan M. Lukic 等学者 (Wang Rui, 2011) 运用模糊聚类与神经网络算法研究驾驶行为时，所选取的参数与本文所选取的类似，最后将驾驶行为分为三类。我们选取急加速率等七个指标作为输入层输入，选取不同的隐藏层与隐藏层节点数进行循环验证，验证发现选取一个隐藏层、五个隐藏层节点时，训练集与测试集的均方误差都比较小。不仅能具有较高的拟合精度，而且有效避免了过拟合。

最终采用三层 BP 神经网络，输入训练向量后所得结构图如下图所示，具体各节点间的权重见附录：

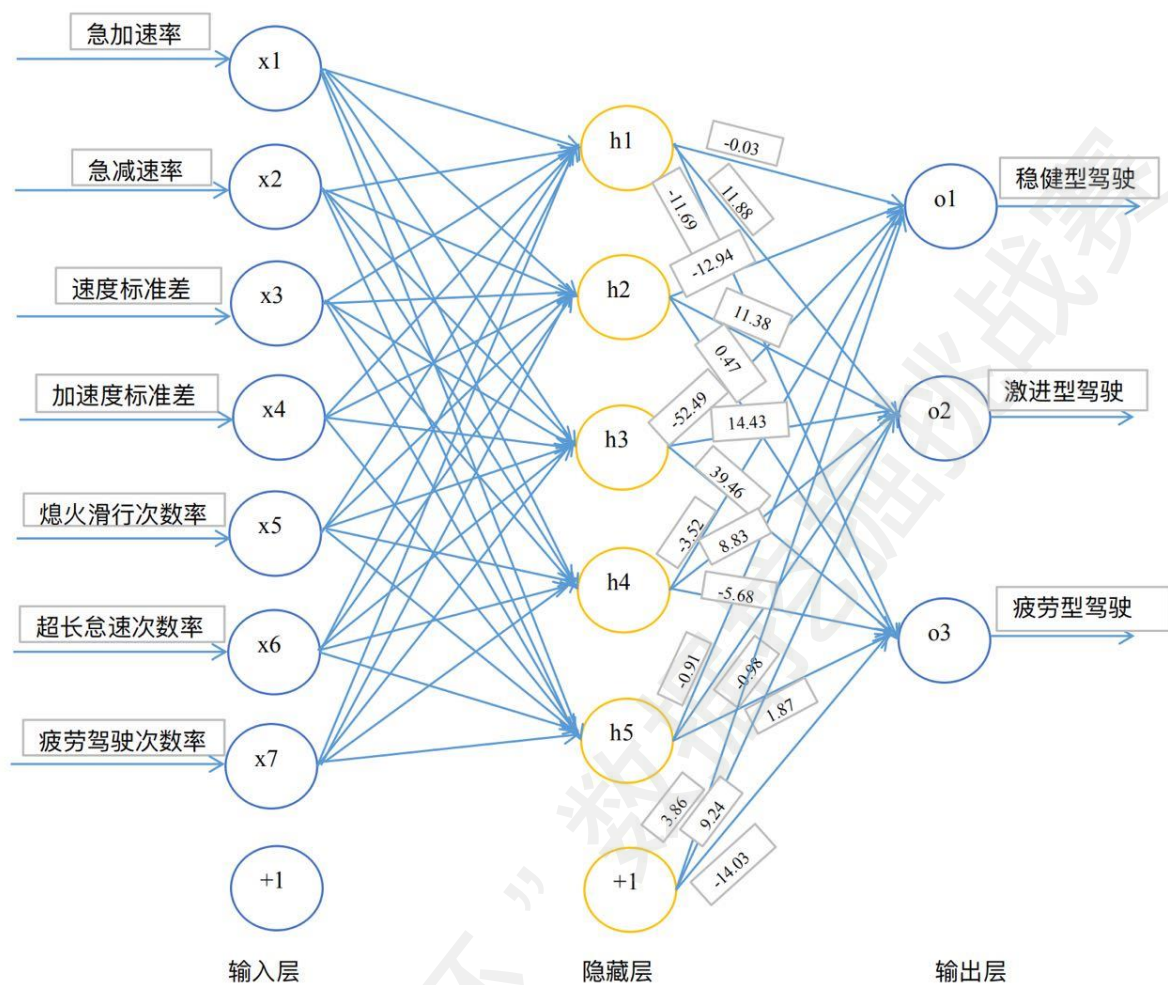


图 5-7 BP 神经网络主要流程图

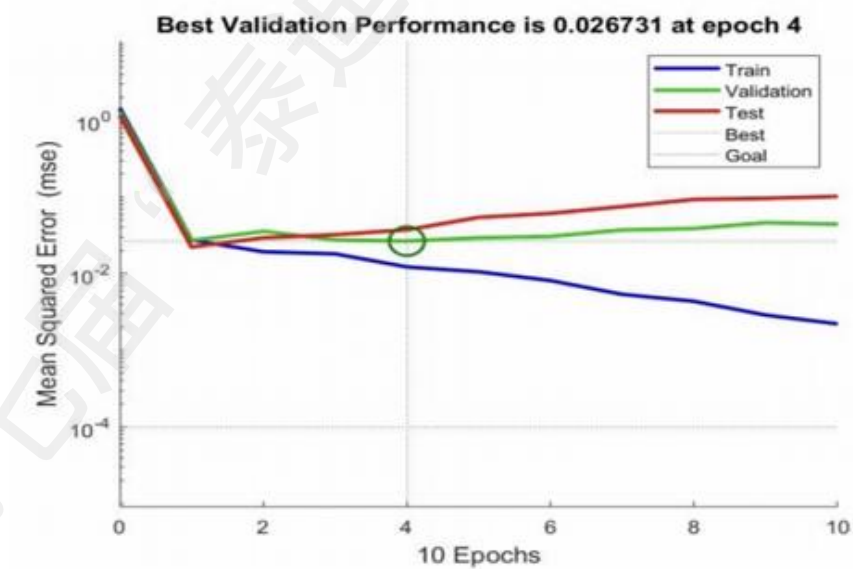


图 5-8 BP 神经网络评价图

由上图可知，训练过程中 BP 神经网络学习速度较快，很快达到目标误差

差 0.1 之下。经训练的神经网络对不良驾驶行为类别的预测值与车辆驾驶行为的实际类别值的状态识别率达到 93.2%。并且通过测试集进行结果分析得到测试误差为 6.2%，这表明此模型可以成为不良驾驶行为类别测评可行模型。

## 六、任务三

### 6.1 探索因子分析

本节选取了急加速率、急减速率、超长怠速次数率、熄火滑行次数率、加速度标准差、速度标准差、平均速度、疲劳驾驶次数率这八个观测指标，首先采用主成分法提取因子，并按照正交方差最大法进行因子旋转，得到 448 辆车的探索性因子分析的结果如下表所示。

表 6-1 探索性因子分析结果

潜在变量	观测变量	变量定义	F1	F2	F3
节能	A1	急加速率	0.788		
	A2	急减速率	0.750		
	A3	超长怠速次数率	0.734		
	A4	熄火滑行次数率	0.721		
	A5	加速度标准差	0.699		
效率	B1	平均速度		0.824	
	B2	速度标准差		0.555	
安全	C1	疲劳驾驶次数率			0.708

由上表可知，通过探索性因子分析将影响行为驾驶综合评价的 7 个指标提出 3 个公因子，由此提出 3 个潜在变量，同一潜在变量下的因子载荷基本都大于 0.6，不同潜变量的交叉载荷量均小于 0.4，表明探索因子的分析效果较好。

## 6.2 测量模型的构建

根据探索性因子分析得到的潜在变量结构，构建驾驶员行车综合评价的测量模型对观测变量的适应性进行检验，得到模型 M0 如下图所示：

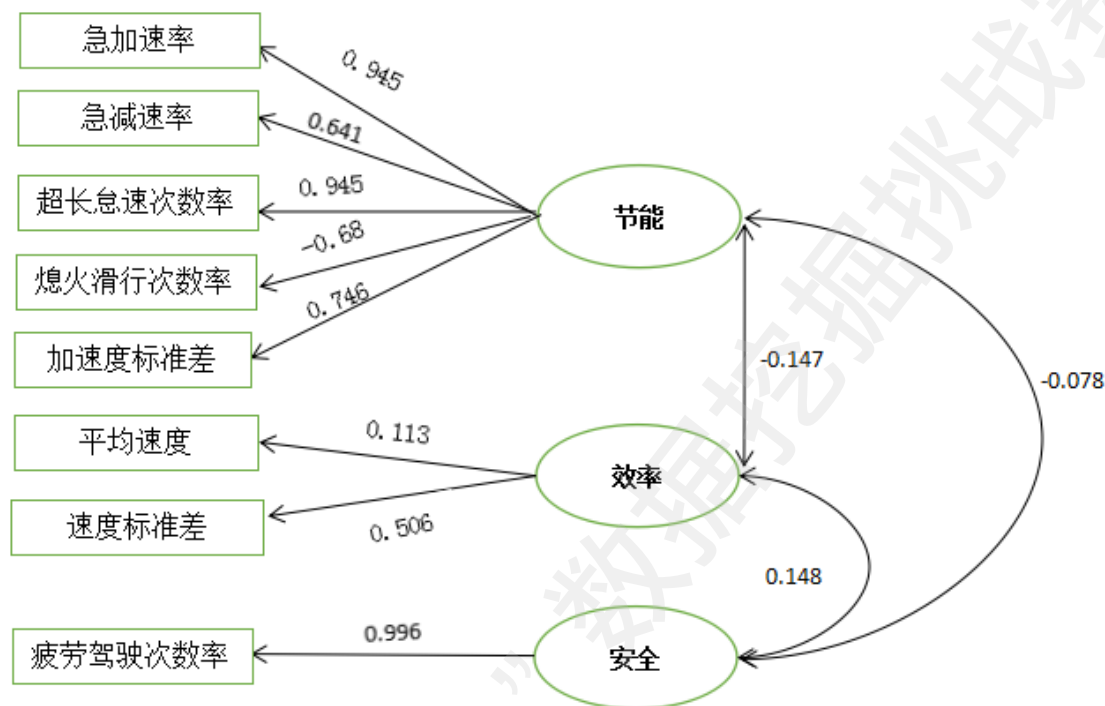


图 6-1 SEM 测量模型

## 6.3 潜在变量之间的因果关系

效率因子与安全因子之间呈正向关系，相关系数为 0.148，节能因子与效率因子同时与安全因子呈反向关系，相关系数分别为-0.147，-0.078，p 值皆小于 0.05，系数在统计学意义上显著。

在节能潜在因子的每一个指标中：急加速、急减速、超长怠速次数、加速度标准差越高，都对车辆的耗能有一个正向的影响，而熄火滑行次数越多，对油耗则有一个负向的影响。

在效率因子的指标中：平均速度与速度标准差越高，对效率有正向影响，尤其是速度标准差其系数高达 0.506。

在安全因子中：疲劳驾驶的权重非常高，这说明了在安全指标中疲劳驾驶是一个关键因素。

## 七、参考文献

- [1]刘雁飞. 驾驶行为建模研究[D]. 浙江大学, 2007.
- [2]李扬. 驾驶行为安全性多属性评价方法及应用研究[D]. 吉林大学, 2016.
- [3]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]. 北京交通大学, 2016.
- [4]张开. 基于云计算的驾驶行为分析[D]. 东南大学, 2016.
- [5]王萍, 万蔚, 张克, 王振华, 刘显青. 出租车驾驶员生态驾驶行为评价[J]. 交通工程, 2018, 18(06): p41~p44, p50.
- [6]张迪. 基于浮动车 GPS 数据分析的车辆相对异常驾驶行为研究[D]. 杭州电子科技大学, 2018.
- [7]朱晓艳. 机动车驾驶员不良驾驶行为研究[D]. 哈尔滨工业大学, 2018.
- [8]Modeling driver performance: The effects of rear-end collision warning algorithms. Brown, T. L. . 2000.
- [9]Integrated Modeling of Cognition and The Information Environment: A Closed-loop, ACT-R Approach to Modeling Approach and Landing with and without Synthetic Vision Systems (SVS) Technology (Technical Report). Byrne, M. D, Kirlik, A. . 2004.
- [10]Bridging the gap between analysis and design: improving existing driver interfaces with tools from the framework of cognitive work analysis[J] . Anders Jansson, Eva Olsson, Mikael Erlandsson. Cognition, Technology & Work . 2006 (1).
- [11]Analysis of Beijing' ' s Drivers' ' Driving Behavior. Shi J, Bai Y, Ying X. Traffic and Transportation Studies (2008) . 2008.
- [12]Driver workload and eye blink duration[J] . Simone Benedetto, Marco Pedrotti, Luca Minin, Thierry Baccino, Alessandra Re, Roberto Montanari. Transportation Research Part F: Psychology and Behaviour . 2010 (3).

八、附录

任务二 BP 神经网络各节点间权重

b->h1	i1->h1	i2->h1	i3->h1	i4->h1	i5->h1	i6->h1	i7->h1					
3.5	-11.21	8.29	-0.58	-7.43	-0.32	-0.71	-0.01					
b->h2	i1->h2	i2->h2	i3->h2	i4->h2	i5->h2	i6->h2	i7->h2					
8.12	-1.69	5.83	-3.89	7.95	-0.23	0.83	0.22					
b->h3	i1->h3	i2->h3	i3->h3	i4->h3	i5->h3	i6->h3	i7->h3					
3.68	-0.08	-0.17	-0.38	1.04	-9.83	2.93	10.64					
b->h4	i1->h4	i2->h4	i3->h4	i4->h4	i5->h4	i6->h4	i7->h4					
8.02	5.9	7.22	-3.35	28.87	0.36	-0.05	0.26					
b->h5	i1->h5	i2->h5	i3->h5	i4->h5	i5->h5	i6->h5	i7->h5					
21.52	39.38	-19.43	-3.81	1.53	-3.42	3.87	2.22					
b->o1	h1->o1	h2->o1	h3->o1	h4->o1	h5->o1	i1->o1	i2->o1	i3->o1	i4->o1	i5->o1	i6->o1	i7->o1
3.86	-0.03	-12.94	-52.49	-3.52	-0.91	0.01	-0.69	0.15	0.27	2.63	-2.07	-13.24
b->o2	h1->o2	h2->o2	h3->o2	h4->o2	h5->o2	i1->o2	i2->o2	i3->o2	i4->o2	i5->o2	i6->o2	i7->o2
9.24	11.88	11.38	14.43	8.83	-0.98	0.08	-0.6	-0.95	0.2	20.47	17.21	-13.64
b->o3	h1->o3	h2->o3	h3->o3	h4->o3	h5->o3	i1->o3	i2->o3	i3->o3	i4->o3	i5->o3	i6->o3	i7->o3
-14.03	-11.69	0.47	39.46	-5.68	1.87	-0.87	0.06	0.81	-0.04	-23.19	-15.19	28.18