

# 基于多层模型的“两客一危”车辆行驶状态评价系统<sup>①</sup>



曹磊, 裴莉莉, 高尧, 李伟, 户媛姣

(长安大学 信息工程学院, 西安 710064)

通讯作者: 李伟, E-mail: [grandy@chd.edu.cn](mailto:grandy@chd.edu.cn)

**摘要:** 随着我国道路运输行业的快速发展,“两客一危”车辆大幅度增长,给道路出行和乘客的生命财产安全带来了极大的考验. 本文基于海量“两客一危”车辆行驶数据提出了多层模型的“两客一危”车辆行驶状态评价系统. 首先,对行驶数据进行特征筛选、异常值清洗、归一化等处理. 然后,在宏观和微观两个层面上分别使用聚类分析模型和动态阈值模型对车辆行驶数据进行分析. 最后通过将聚类分析的结果与动态阈值的分析结果相结合即可实现对车辆行驶状态以及驾驶员的驾驶习惯的综合评价. 研究表明,本文提出的多层模型能够对车辆行程路况以及车辆驾驶员驾驶习惯进行较为准确的评估,可为“两客一危”车辆的管理监督部门以及车辆运输企业提供合理的安全生产的科学依据和数据支持.

**关键词:** 两客一危; 数据挖掘; 聚类分析; 多层模型

引用格式: 曹磊,裴莉莉,高尧,李伟,户媛姣.基于多层模型的“两客一危”车辆行驶状态评价系统.计算机系统应用,2021,30(1):94-100. <http://www.c-s-a.org.cn/1003-3254/7759.html>

## Driving Status Evaluation System of Special Transportation Vehicle Based on Multi-Layer Model

CAO Lei, PEI Li-Li, GAO Yao, LI Wei, HU Yuan-Jiao

(School of Information Engineering, Chang'an University, Xi'an 710064, China)

**Abstract:** With the rapid development of road transportation industry in China, special transportation vehicles have increased significantly, which has brought great challenges to road travel and the safety of passengers' lives and property. Based on massive special transportation vehicle driving data, this study proposes a multi-layer model of special transportation vehicle driving state evaluation system. Firstly, the data is processed for feature selection, outlier cleaning, and normalization. Then, the cluster analysis model and the dynamic threshold model are used to process vehicle driving data at the macro and micro layers, respectively. Finally, the results of cluster analysis and dynamic threshold analysis are combined to achieve a comprehensive evaluation of the vehicle's driving status. The research results show that the multi-layer model proposed in this paper can make a more accurate assessment of the vehicle's travel conditions and driving habits of vehicle drivers. It can provide reasonable scientific basis and data support for the management and supervision departments of special transportation vehicles and the vehicle transportation enterprises.

**Key words:** special transportation vehicle; data mining; clustering analysis; multi-layer model

① 基金项目: 陕西省交通运输厅 2018 年度交通科研项目 (18-31X); 长安大学中央高校基本科研业务费专项资金 (300102249102, 300102240201)

Foundation item: Transportation Scientific Research Project of Shaanxi Provincial Transportation Department in 2018 (18-31X); The Fundamental Research Funds for the Central Universities of China of Chang'an University (300102249102, 300102240201)

收稿时间: 2020-05-21; 修改时间: 2020-06-16, 2020-07-10; 采用时间: 2020-07-14; csa 在线出版时间: 2020-12-31

随着我国道路运输行业的快速发展,“两客一危”道路运输车辆数量出现大幅度的增长,在方便人们出行、促进地区经济水平发展的同时,也给道路出行和乘客的生命财产安全带来了极大的考验<sup>[1]</sup>。

然而,由于“两客一危”车辆的特殊性、高风险性,以及路网中“两客一危”车辆数据的相对独立性和地区差异<sup>[2]</sup>,各地区(省)对于活动在本省境内的“两客一危”车辆缺乏系统性研究和管理,没有能够充分挖掘大数据背景下研究分析的方法和优势,对于“两客一危”重点车辆的跟踪监测、分析、分布与风险预测没有进行深入挖掘,从而不利于政府等监管部门的监督和决策。

另一方面,交通安全是交通领域的关键问题。交通安全条件由驾驶员、车辆和驾驶环境决定。先前的研究表明,超过90%的交通事故与不安全的驾驶行为有关。驾驶行为在驾驶风险分析中起着重要作用。但是,在现实生活中很难衡量驾驶风险<sup>[3]</sup>。因此驾驶模拟器通常用于调查各种实验环境中的驾驶行为<sup>[4]</sup>。诸如自然驾驶研究(NDS)和DriveCam系统之类的一些车辆仪表技术已被广泛用于监测驾驶行为和运动学特征<sup>[5]</sup>。现有的大多数危险驾驶行为分析都依赖于碰撞数据或自我报告的问卷调查<sup>[6]</sup>。张辉等<sup>[7]</sup>通过设计分心模拟驾驶试验来采集驾驶人眼动特征数据,进行驾驶员的分心状态判别。侯海晶等<sup>[8]</sup>利用搭载了眼动仪的驾驶模拟器才采集驾驶员感知与操作的数据,利用这些数据对驾驶员的驾驶风格进行分类。薛清文等<sup>[9]</sup>通过采集高精度车辆轨迹数据评估驾驶员的整体驾驶状态,利用LGBM(Light Gradient Boosting Machine)算法对危险驾驶行为进行识别。为了充分探索交通事故中的驾驶行为,重要的是要保持真实驾驶情况下的驾驶行为习惯。

对于危险驾驶行为的研究,交通事故数据传统上是主要或唯一的数据源,但是交通事故在广义上来说是一个小概率事件,其所包含的信息很少,因此主动采取有效交通安全措施的方法已被忽略。其次,多层次或多结构的模型能够发现数据中所忽略的对交通安全的影响因素,对数据的多维度挖掘。最后,国内很少将积累的“两客一危”车辆大数据用于道路安全以及危险驾驶研究。

为了填补上述研究空白,本研究的研究目的主要分为以下两点。第一个目标是从宏观水平上分析不同类型的卡车驾驶员的驾驶习惯和危险驾驶的倾向。第

二个目标是通过在微观层次上对车辆进行动态监控,在这两个水平的基础上建立一个多级模型来对“两客一危”的车辆进行监控和管理。该研究与现有研究的不同之处在于同时考虑了以下方面:

1) 采用真实的“两客一危”车辆驾驶数据(并不是来源于模拟器产生的实验数据),并根据多个指标发现识别潜在的危险驾驶行为;

2) 使用大规模数据集,本文使用的数据集以5s的间隔记录货车和客车的行车速度,车辆位置,车辆信息相关的数据;

3) 建立反映“两客一危”车辆驾驶员危险驾驶倾向的多层次模型。

## 1 数据集构建

### 1.1 数据集概述

本文使用的数据为陕西省境内2018年9月至2019年2月共6个月的“两客一危”车辆的行驶数据,其中每天包含大约25000辆车的行驶信息,每个车辆每天的行程为一个单独的数据文件,其中包含的数据项如表1所示。

表1 车辆行驶数据字段解释

数据项	示例	单位	说明
地图经度	65470092	/	纠偏后的经度
地图纬度	20194431	/	纠偏后的纬度
GPS时间	20190224/195533	年月日/时分秒	/
GPS速度	22	km/h	/
方向	229	度	/
事件	5	/	/
报警编码	/	/	/
GPS经度	65467020	/	/
GPS纬度	20195274	/	/
海拔	80	/	/
行驶记录仪速度	28	km/h	/
里程	0	km	/
错误类型	0	/	0正常;1经度错误 2纬度错误3:时间 错误:4速度错误; 5:方向错误
系统时间	20190225/080913	年月日/时分秒	/

### 1.2 数据采样与可视化

在原始数据集中,随机选择不同日期(包含节假日、工作日)、不同天气状况、一天内不同时段、不同颜色牌照以及不同道路状况的多类车辆行驶数据对驾驶员的驾驶行为进行分析。本文共选择了9000辆

“两客一危”车辆的行程(单位:天)进行研究。

在对字段的筛选中取出与驾驶员危险驾驶行为相关的车辆位置信息,例如车辆的GPS速度,车辆时间与车辆的方向信息.其中车辆的位置信息用于判断车辆行驶道路类型,速度、时间与方向信息用于评判驾驶员的驾驶状态,在对数据字段进行筛选采样之后,对现有字段中每辆车每天行程(运行时间大于2小时)的速度变化信息与方向变化信息计算方差,形成新的字段.9000辆车的速度方差与方向方差部分数据可视化结果如图1所示。

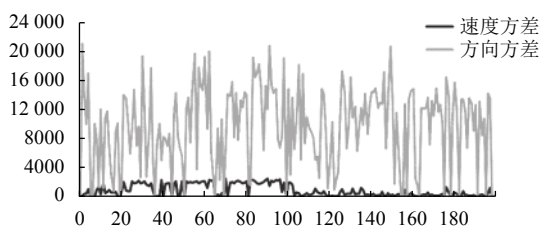


图1 方差示例数据

### 1.3 零值与异常值清洗

从数据可视化图中可以发现,数据中包含大量的零值,而这些零值表示这些车辆的没有处在行驶状态,因此需要对这些零值进行清洗,清洗后的数据如图2所示。

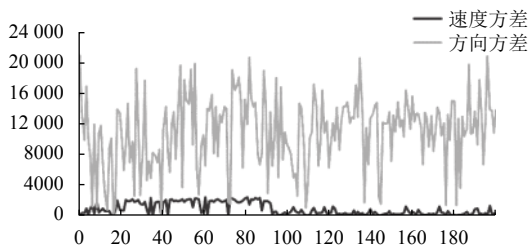


图2 零值清洗后数据

同时,在这些数据中还有一些由于传感器的误差或者其他原因产生的异常数据,使用箱型图的方法可以有效检测到这些异常值,对去除零值后的数据进行箱型图可视化如图3所示。

图3中的“+”表示数据中的异常值,将这些异常值从原数据中清洗掉,还有7895辆车的行驶状态数据,之后再对清洗后的数据示例进行可视化如图4所示。

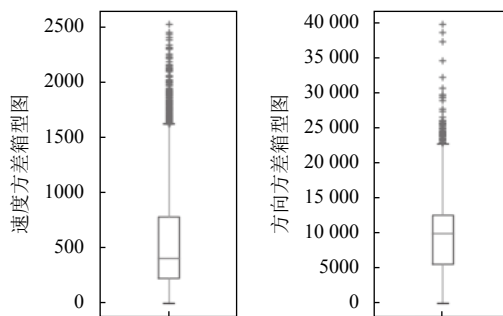


图3 箱型图检测异常值结果

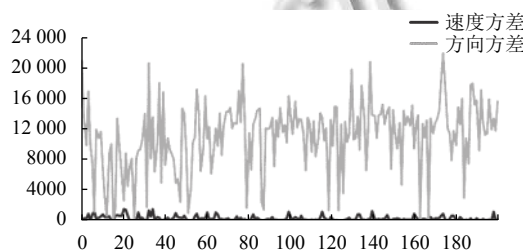


图4 数据清洗后结果

### 1.4 数据归一化

从零值与异常值清洗后的数据可视化图中可以看出,车辆行驶的速度方差和方向方差的数据分布不均衡,且波动相差较大这将在之后的距离计算中对计算结果产生影响,因此需要对数据进行归一化处理,本文使用的归一化方法为min-max归一化方法,如式(1):

$$V'_i = \frac{V_i - V_{\min}}{V_{\max} - V_{\min}} \quad (1)$$

其中,  $V_i$  是真实值,  $V'_i$  是规范化之后的值。

对归一化后的数据进行可视化如图5所示。

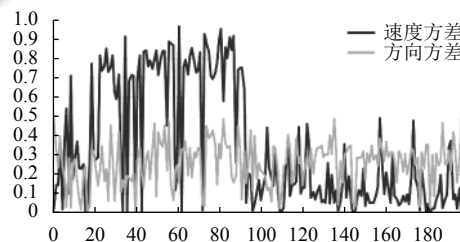


图5 归一化数据可视化

从图5中可以看出数据的分布已经比较均衡,之后以此数据作为输入样本对驾驶情况进行聚类分析。

## 2 多层模型构建

本文采用多层次的模型来对“两客一危”车辆驾驶员的驾驶行为进行分析评价,多层次主要体现在宏观

与微观两个层次. 宏观层面上对驾驶员的一次行程信息进行分析以评判其驾驶平稳性, 微观层面对驾驶员驾驶车辆在不同速度下的危险驾驶行为进行识别. 多层模型的结构如图6所示.

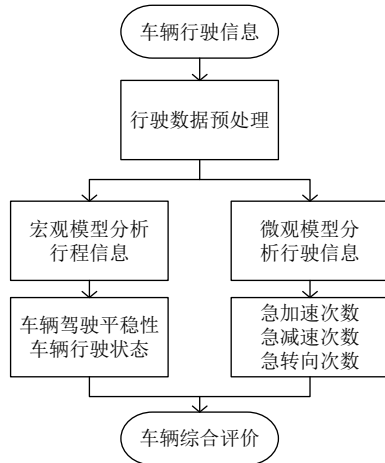


图6 “两客一危”车辆行驶状态多层评价模型

## 2.1 宏观层次聚类分析模型

对车辆行驶数据的宏观层次分析可以使用聚类的方法. 聚类分析是一种无监督的学习技术, 可将一组物理或抽象对象划分为几个相似的聚类以获得全局数据图或对特定聚类进行进一步分析. 通过聚类生成的类是一组数据对象, 与原始组中的其他对象(基于相似性进行聚类)相比, 它们具有更大的相似性. 相似性由研究对象的属性值确定, 相对距离是一种常用的措施.

本文选择基于相对距离的聚类算法 K-means 对数据进行聚类, 该方法能够将数据划分为预定数量的聚类(假设有足够多的不同情况).

基于距离的算法依靠距离度量(函数)来度量数据点之间的相似度. 距离度量的标准是欧氏距离、余弦或快速余弦距离. 根据所使用的距离度量将数据点分配给最近的群集, 该算法认为两个数据对象的距离越近, 相似度就越大; 距离越远, 相似度就越小. 它基于样本空间中最有代表性的点, 迭代地将所有数据样本划分为不同的类别, 使聚类出来的每个簇的聚合度最高, 簇间的分离度最高. 对于距离度量本文采用欧式距离计算方法计算, 公式如式(2)所示:

$$d(x_i - x_j) = \sqrt{\sum_{n=1}^m (x_{in} - x_{jn})^2} \quad (2)$$

其中,  $x_i$ 和 $x_j$ 为计算距离的两个点,  $m$ 为样本维度数,  $n$ 为当前维度.

K-means 算法因其算法框架清晰简单易懂, 处理大数据集的算法相对可扩展且高效的优点其才得到大量的应用; 当数据集的类密集且类与类之间的差异明显时, 该算法处理的效果最好. 使用该算法对数据进行聚类分析的首要任务就是给出要生成的类的数目  $k$ ,  $k$  值是否合适可以通过计算  $SSE$ (簇内误差平方和)来评价. SSE-Kmeans 聚类算法中的核心思想是:

1) 在聚类分析中随着数据簇数  $k$  的增加, 样本拆分变得更加复杂并且精细, 而且每个类别的聚合强度逐渐增加, 因此平方误差和  $SSE$  自然降低.

2) 如果  $k$  小于真实簇的数量, 则  $k$  的增加将大大增加每个簇的内聚性, 因此  $SSE$  的下降程度将会很大. 并且当  $k$  达到真实簇的数量时, 再通过增加  $k$  的值得到的聚合程度的增加将会迅速变小, 因此随着  $k$  值的持续增加,  $SSE$  的下降率迅速下降并逐渐趋于平稳. 也就是说,  $SSE$  和聚类类别数  $k$  之间的关系呈肘形, 其中肘形图中肘部对应的  $k$  值就是数据中真实簇的数量.

对于一个特定的  $d$  维数据集  $D=(x_1, x_2, \dots, x_n)$ , SSE-Kmeans 算法的步骤如图7所示.

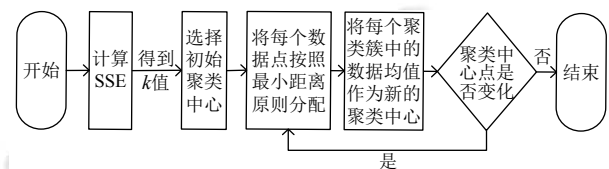


图7 SSE-Kmeans 算法步骤

## 2.2 微观层次动态阈值评判

对车辆行驶数据的微观层次分析中使用动态阈值的分析方法. 在车辆危险行驶状态的评判及等级划分的研究中, 当前学者多采用固定阈值的方法, 但是车辆在不同速度的情况下危险驾驶的评判应当也不相同, 如速度越快急转向的评判阈值应该越小, 因此基于速度的车辆危险行驶状态的动态阈值评判更符合实际情况.

Han 等<sup>[10]</sup> 利用车辆黑匣子收集了速度、加速度及横摆角速度数据, 识别了急加速、急减速、急转弯、突然换道 4 种车辆危险行驶状态, 并提出了基于不同速度区间的阈值划分方法, 如表 2 和图 8 所示.

表2 微观模型动态阈值

速度(km/h)	急加速(g)	急减速(g)	急转向(°)	突然换道(°)
0-9	0.22	0.61	/	13
10-19	0.22	0.61	/	12
20-29	0.21	0.61	/	11
30-39	0.2	0.61	/	9
40-49	0.19	0.58	12.7	9
50-59	0.15	0.58	10.7	8
60-69	0.15	0.58	10.7	7.5
70-79	0.14	0.55	10.7	7
80-89	0.13	0.55	10.5	6.5
90-	0.12	0.54	10.5	6.5

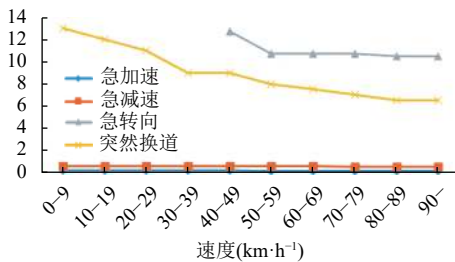


图8 动态阈值分布图

本文使用上述阈值对“两客一危”车辆每条记录的行驶状态进行评判并记录,并结合宏观模型对车辆行驶平稳状态的评估得到车辆的总体评价<sup>[11-13]</sup>。接下来主要介绍宏观层次模型的应用。

### 3 数据聚类分析

对数据使用 SSE-Kmeans 聚类算法进行聚类首先需要确定簇的个数(即  $k$ ),  $k$  值可以通过簇内误差平方和 (within-cluster SSE) 确定,  $SSE$  的计算方法如式(3)所示:

$$SSE = \sum_{i=1}^n \sum_{j=1}^m w(i, j) = \|x^{(i)} - \mu^{(j)}\|_2^2 \quad (3)$$

其中,  $x^{(i)}$  表示第  $i$  个数据点,  $\mu^{(j)}$  表示  $j$  簇的中心,  $n$  和  $m$  表示样本的维度。

对归一化数据进行  $SSE$  计算结果如图9所示。

图9中可以看出当  $k$  值为4时,正好是手肘的位置,即为最佳聚类簇数。使用 SSE-Kmeans 算法对数据进行聚类,得到结果如图10所示。

图10中每一个颜色代表一个数据簇,“X”符号代表每类数据点的中心。

同时采用基于密度的聚类方法 DBSCAN ( $Eps=0.5$ ,  $nPts=10$ ) 对数据进行聚类可以得到的结果如图11所示。

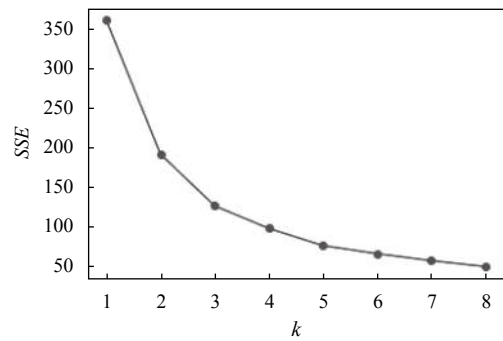


图9 SSE与k值的关系

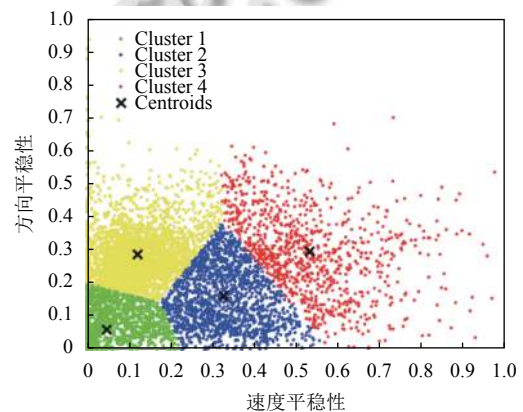


图10 SSE-Kmeans 算法聚类结果图

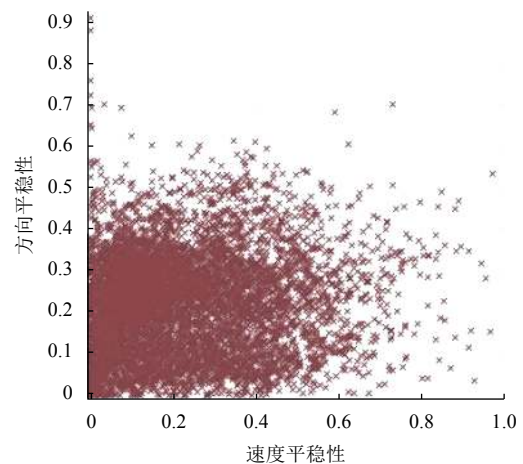


图11 DBSCAN 算法聚类结果图

从图11可以看出, DBSCAN 算法将数据聚类为一类。与图10对比可知,当数据量的类密集时,基于密度的聚类算法 DBSCAN 对行驶平稳性数据的聚类效果并没有基于相对距离的算法 SSE-Kmeans 对行驶平稳性数据的聚类效果好<sup>[14,15]</sup>。因此本文将对 SSE-Kmeans 算法的聚类结果进行分析与讨论。

## 4 结果分析与讨论

图10中每类的数据点数及每类数据点占总数据点的比例如图12所示。

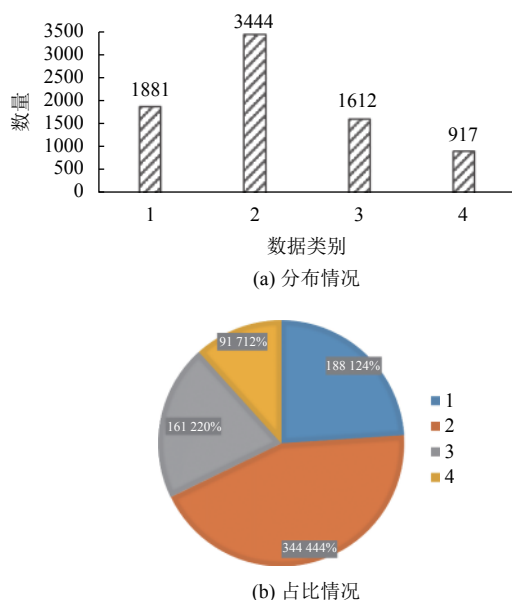


图12 每类数据的分布情况与所占比例

对聚类结果从每类数据点的特点进行分析可以得出以下结论:

1) 第1类数据点(图10中绿色点)代表了这些车辆中行驶最为平稳的一些个体,这些个体在一天的行程中速度和方向的变化都较为稳定,因此处在这一类的车辆驾驶员潜在危险驾驶的倾向性特别低。

2) 第2类数据(图10中黄色点)则代表了车辆速度平稳性较好但方向平稳性较差的个体,说明这些车辆在这一天的行程中有可能行驶在弯道较多路况较差的道路上,虽然其方向平稳性较差,但是速度变化稳定,因此这一类的车辆驾驶员的危险驾驶倾向比较低。

3) 第3类数据(图10中蓝色点)代表了车辆方向平稳性较好但速度平稳性较差,说明这些车辆在这一天的行程中有可能行驶在弯道较少路况较好的道路上,但由于其速度平稳性较差即速度变化较大,好在其方向的变化性较小,因此处在这一类的车辆驾驶员的危险驾驶倾向也比较低。

4) 第4类数据(图10中红色点)代表了车辆方向平稳性较差且速度平稳性也较差或者方向平稳性较好但速度平稳性差的车辆,说明这些车辆在这一天的行程中有可能以很差的速度平稳性行驶在路况不好的道

路上,或者以很差的速度平稳性行驶在路况较好的道路上,但是由于路况较好时车速也更快其危险驾驶行为造成的后果也更严重,因此这种特点都说明处于这一类的车辆驾驶员的危险驾驶倾向比较高。

结合上面的分析结果,可以发现88%的车辆都处于低,或者较低的危险驾驶倾向区域,其中有24%的车辆在这一天的行程中速度和方向的平稳性均比较低,剩余的12%的车辆在这一天的行程中含有较高的危险驾驶行为的倾向。聚类结果分布如图13。

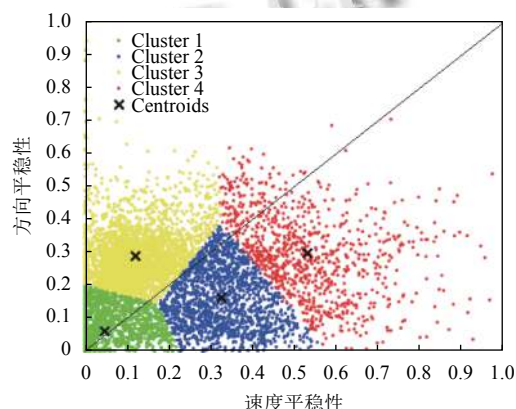


图13 聚类结果分布图

此外,据图13的分布情况可以看出,车辆数据点中的大部分都集中在数据分布图的右下方,陕西省内的“两客一危”车辆主要行驶在弯道较少路况较好的道路上,因此应该主要关注这些车辆的速度以及加速度的变化即可。

当需要对某车辆的行驶平稳性进行评价时,在宏观层面上首先计算其行程方向与速度数据方差,之后判断其属于哪一类数据簇,那么其行驶状态就具有那一类数据的特点。同时在微观层面上对其行驶过程中的急加速、急减速、急转向、突然换道次数进行计算。结合两个层面的分析结果对车辆的形式状态以及驾驶员的驾驶习惯进行全面准确的评估。

## 5 结语

本文采用陕西省内的“两客一危”车辆的行驶GPS数据,提出了用以评价“两客一危”车辆行驶状态的多层次模型,其中多层次模型包含宏观评价模型与微观评价模型,宏观模型基于SSE-Kmeans算法对车辆行程的速度与方向方差进行聚类分析从而评价车辆行程行驶状态稳定性,微观模型主要通过动态阈值的方法

评价车辆行驶中急加速、急减速、急转向的次数。且这些评价均与实际情况相符,能够较为准确地对车辆行驶状态进行评价。对于管理部门而言,能够根据该结果及时对有危险驾驶倾向的驾驶员做出提醒并重点监测,提高车辆的安全驾驶程度,降低“两客一危”车辆的事故发生率,从而保障人民的生命财产安全,提高运输效率、应急处置和政策决策能力。

### 参考文献

- 1 陈小妮,郭晓炜.建设交通运输安全生产“两客一危”车辆智能监管平台的探析.公路,2019,64(8):255-259.
- 2 唐亮.信息化条件下营运车辆安全监管关键技术研究[博士学位论文].重庆:重庆大学,2012.
- 3 Eboli L, Guido G, Mazzulla G, *et al.* Investigating car users' driving behaviour through speed analysis. *Promet-Traffic-Traffico*, 2017, 29(2): 193-202. [doi: 10.7307/ptt.v29i2.2117]
- 4 Pankok C Jr, Kaber D. The effect of navigation display clutter on performance and attention allocation in presentation- and simulator-based driving experiments. *Applied Ergonomics*, 2018, 69: 136-145. [doi: 10.1016/j.apergo.2018.01.008]
- 5 彭海维,王站立.汽车危险驾驶行为检测系统的实现.江苏科技信息,2019,36(35):33-36,46. [doi: 10.3969/j.issn.1004-7530.2019.35.010]
- 6 Lord D, Mannering F. The Statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 2010, 44(5): 291-305. [doi: 10.1016/j.tra.2010.02.001]
- 7 张辉,钱大琳,邵春福,等.模拟驾驶环境下驾驶人分心状态判别.中国公路学报,2018,31(4):43-51. [doi: 10.3969/j.issn.1001-7372.2018.04.006]
- 8 侯海晶,金立生,关志伟,等.驾驶风格对驾驶行为的影响.中国公路学报,2018,31(4):18-27. [doi: 10.3969/j.issn.1001-7372.2018.04.003]
- 9 薛清文,蒋愚明,陆键.基于轨迹数据的危险驾驶行为识别方法.中国公路学报,2020,33(6):84-94. [doi: 10.3969/j.issn.1001-7372.2020.06.008]
- 10 Han I, Yang KS. Characteristic analysis for cognition of dangerous driving using automobile black boxes. *International Journal of Automotive Technology*, 2009, 10(5): 597-605. [doi: 10.1007/s12239-009-0070-9]
- 11 高悦尔,阚小溪,胥川,等.基于“两客一危”数据的高速公路服务区路段车辆行驶模式研究.福州大学学报(自然科学版),2017,45(3):398-403.
- 12 刘永涛,乔洁,魏朗,等.危险驾驶行为辨识算法研究.计算机工程与设计,2014,35(4):1322-1326. [doi: 10.3969/j.issn.1000-7024.2014.04.039]
- 13 Ahmed MM, Ghasemzadeh A. The impacts of heavy rain on speed and headway behaviors: An investigation using the SHRP2 naturalistic driving study data. *Transportation Research Part C: Emerging Technologies*, 2018, 91: 371-384. [doi: 10.1016/j.trc.2018.04.012]
- 14 Pantangi SS, Fountas G, Sarwar T, *et al.* A preliminary investigation of the effectiveness of high visibility enforcement programs using naturalistic driving study data: A grouped random parameters approach. *Analytic Methods in Accident Research*, 2019, 21: 1-12. [doi: 10.1016/j.amar.2018.10.003]
- 15 张雅楠,唐阳山,刘昊,等.基于行车数据的驾驶员危险驾驶行为的判定.汽车实用技术,2019,(15):247-248.