

基于纯电动汽车高频数据的驾驶风格分类方法

纪少波¹, 张珂¹, 李伦¹, 苏士斌², 何绍清³, 冯远宏², 张强³

(1. 山东大学能源与动力工程学院, 山东济南 250061; 2. 青岛海信网络科技股份有限公司, 山东青岛 256000;

3. 中国汽车技术研究中心有限公司, 天津 300300)

摘要: 基于纯电动汽车的高频运行数据选取了 15 项与驾驶安全有关的特征参数; 通过对各特征参数的统计分析, 提出了多参数组合阈值边界线进行危险驾驶行为识别, 然后对比分析了两种聚类算法和 4 种降维方法的效果, 采用 t 分布随机邻域嵌入 (t -SNE) 和高斯混合模型 (GMM) 组合算法建立了驾驶风格分类模型, 将驾驶风格分为 3 种类型。在此基础上, 研究了不同驾驶风格对应的车辆运行特性。该分类模型能够有效反映驾驶员的驾驶习惯, 为车队管理和道路安全提供可靠的评估依据。

关键词: 危险驾驶行为; 驾驶风格; 高斯混合模型; t 分布随机邻域嵌入

中图分类号: U471.15

文献标志码: A

Driving Style Classification Method Based on High-frequency Data from Pure Electric Vehicles

Ji Shaobo¹, Zhang Ke¹, Li Lun¹, Su Shibin², He Shaoqing³, Feng Yuanhong², Zhang Qiang³

(1. School of Energy and Power Engineering, Shandong University, Ji'nan 250061, China; 2. Hisense TransTech Co. Ltd., Qingdao 256000, China; 3. China Automotive Technology and Research Center Co. Ltd., Tianjin 300300, China)

Abstract: 15 characteristic parameters related to driving safety were selected based on the operation data of pure electric vehicles. According to the statistics of the parameters, threshold line with multiple parameters was proposed. Then two clustering algorithms and four dimension-reduction methods were compared. Based on the classification effect, the combination algorithm of t -Distribution stochastic neighbor embedding (t -SNE) and Gaussian Mixture Model (GMM) were used to establish the driving style classification model, which was used to classify the driving style into three classes. On this basis,

the vehicle operation characteristics of the three classes were compared. The classification model can effectively reflect the driving behavior and provide reliable evaluation basis for fleet management and road safety.

Key words: dangerous driving behavior; driving style; Gaussian Mixture Model; t -distribution stochastic neighbor embedding

根据国家统计局 2013—2020 年发布的《中国统计年鉴》, 我国道路安全事故发生情况可以看出近几年的车祸发生数量呈增长趋势, 这表明交通安全问题日渐严重。驾驶员作为交通行为的首要执行者、车辆的操控者, 是导致交通事故的主要原因。因此, 驾驶员在行驶过程中驾驶行为的安全性具有重大的研究价值。

国内外学者多采用无监督学习模型进行驾驶风格分类, 通过对分类结果的分析确定驾驶风格类别。目前研究驾驶风格用的无监督学习主要有基于划分、层次、密度及模型等不同方法的聚类方法。基于划分的无监督学习聚类方法主要有 K 均值 (K -means) 和模糊 C 均值 (FCM)。亓航等^[1]、徐婷等^[2]和余荣杰等^[3]研究学者采用 K 均值聚类算法分别对出租车、货车和分时租赁汽车的驾驶行为数据进行驾驶风格分类, 发现不同驾驶风格的驾驶员在出行特点、车辆运行参数及百公里能耗方面都具有差异性。石秀鹏等^[4]从车辆轨迹数据提取了风险指标特征, 基于这些特征利用 FCM 算法将驾驶行为聚为 4 个风险等级。基于层次的无监督学习聚类方法主要是层次聚类算法, 牛增良等^[5]基于大量重特大交通事故数据采用模糊聚类、层次聚类对危险驾驶行为划分层次进行分析。基于密度的无监督学习聚类方

收稿日期: 2021-06-28

基金项目: 国家重点研发计划 (2020YFB1600501); 山东省重大科技创新工程项目 (2019TSLH0203); 山东省自然科学基金 (ZR2020ME180)

第一作者: 纪少波 (1979—), 男, 副教授, 硕士生导师, 工学博士, 主要研究方向为新能源汽车测控技术。

E-mail: jobo@sdu.edu.cn



论文
拓展
介绍

法主要是密度聚类(DBSCAN)算法,文江辉^[6]从营运车辆监控平台获取出行高频时间段的车辆运行数据,利用DBSCAN算法对不同车辆、不同时间段及不同天气状况下驾驶行为进行了分类,用于判定车辆驾驶的稳定性程度。基于模型的无监督学习聚类方法主要是利用高斯混合模型(GMM)等模型展开研究,朱冰等^[7]基于真实驾驶数据采用GMM和KL散度定量地度量驾驶员之间的相似性,实现驾驶员驾驶风格分类。李立治等^[8]建立了驾驶员驾驶风格的数据库,使用K均值、FCM及层次聚类对驾驶风格进行分类。李卓轩等^[9]基于不良驾驶行为特征参数,使用K均值聚类和DBSCAN算法对驾驶风格进行分类。刘通等^[10]基于K均值聚类算法进行了驾驶风格的分类研究。

已有研究采用的聚类算法多是针对低维数据集设计的,处理高维数据集通常会面临“维灾难效应”,会出现基于距离的度量函数失效、聚类中心难以确定及计算效率低等问题。本文采用先降维后聚类的研究思路,在对比不同降维及聚类方法的基础上,提出利用t-SNE和GMM组合算法建立基于驾驶安全的驾驶风格分类模型,用于对不同驾驶风格进行分类。此外,已有研究关于异常驾驶行为的判定多采用固定阈值,本文在深入分析与驾驶安全相关的特征参数变化规律基础上,提出多参数组合阈值边界线识别危险驾驶行为的方法。本文的研究可以有效

改进分类效果,提高异常驾驶识别精度。研究内容可为运输企业的安全运营提供指导,为道路安全评估提供参考。

1 数据采集与处理

1.1 数据采集

基于车辆的OBD接口,通过CAN总线获取车辆实时运行数据,并将采集的数据通过无线网络发送至服务器进行处理,数据采集流程如图1所示。驾驶员瞬态操作数据中包含丰富的驾驶行为信息,过低的采样频率将导致瞬态驾驶行为信息丢失;过高的采样频率将导致数据采集成本增加,数据处理难度加大,因此需要合理选择监测频率。心理学家认为正常人的反应时间为0.15~0.4s,最快反应时间也高于0.1s。本文数据采集频率设置为10 Hz,即采集间隔为0.1s,采样频率完全可以满足驾驶行为分析需求。

本文研究数据主要来自于北京、成都、天津、上海四座城市的200辆纯电动汽车4个月的高频运行数据,汽车用途有私家车、网约车、共享汽车。数据样本具有采样频率高、车辆用途多样、驾驶员分布广泛、出行次数频繁等优点,适用于驾驶员的驾驶行为分析。

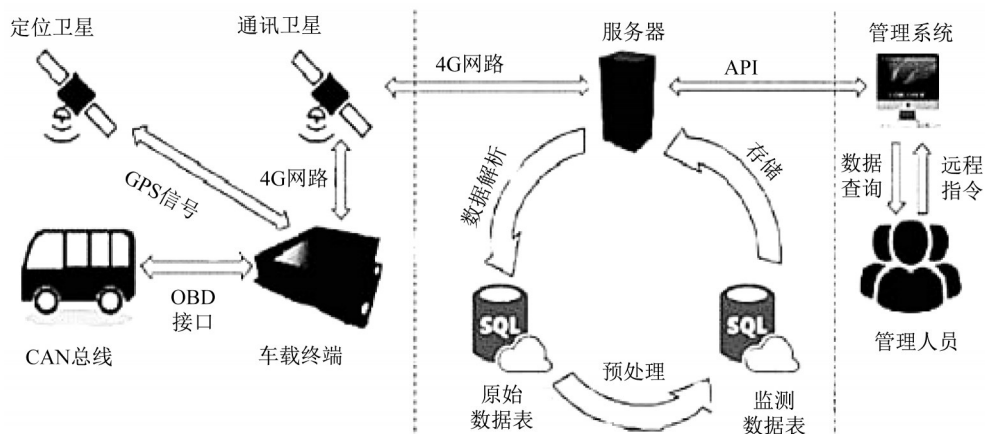


图1 车辆监测数据采集流程

Fig.1 Vehicle monitoring data collection process

1.2 数据处理

由于原始数据是按车辆VIN码分类储存的,且采集信号中存在的无效数据字段导致原始数据体量巨大,影响后续研究。因此,本文首先剔除了无效字段信号数据和充电与静置片段数据,形成每车每次

独立的运行数据文件。考虑到一次正常行车时长不会过短,且驾驶工况在较短时间内变化不会很大,所以结合车辆的上电状态及车速,剔除一次行车时长小于10min和数据缺失率超过5%的行程事件。处理后的一次行车事件数据如图2所示。

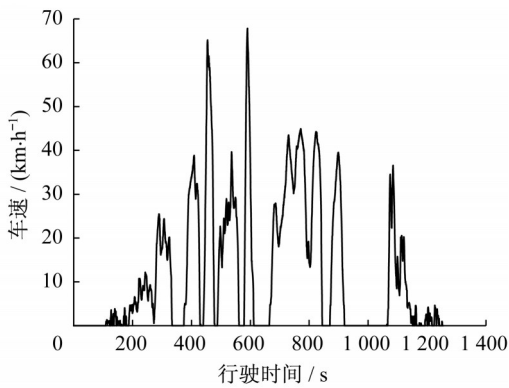


图2 一次行车事件数据

Fig.2 Data in a driving event

数据采集和传输过程受到干扰可能导致远程数据管理平台接收到的数据存在波动、异常或缺失等问题。为了消除异常数据的影响,在进行数据分析前需要对原始数据进行质量控制。首先对于轻微波动的信号数据采用滑动窗口均值滤波方法进行处理,效果如图3所示。采用统计分析和6 σ 法去除幅值异常的数据点;接收的数据中存在相邻数据时间差小于采样时间的问题,这主要是由于网络不稳导致数据重复发送造成,将这些冗余数据亦作为异常值剔除,处理后的效果如图4所示。此外,对少量的缺失数据值采用线性插值进行填补。

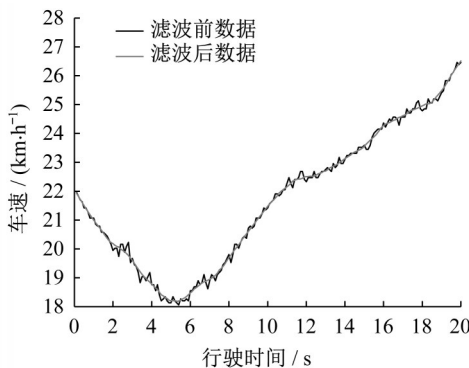


图3 滤波前后数据对比

Fig.3 Comparison before and after treatment of filter

在对原始数据进行质量控制后,参考文献[11-14]选取与驾驶安全有关的驾驶行为特征参数表1所示。通过对事故发生原因的分析,驾驶员超速、急加速、急减速和急转向等危险驾驶行为是导致事故多发的重要原因^[15]。但在我国现有的相关交通规定中未对这些危险驾驶行为进行约束,并且交通违章监控难以有效辨识,无法及时给予驾驶员危险警示以避免事故的发生。本文对急加速、急减速、超速、急转向、转向超速和疲劳驾驶等不同危险驾驶行为

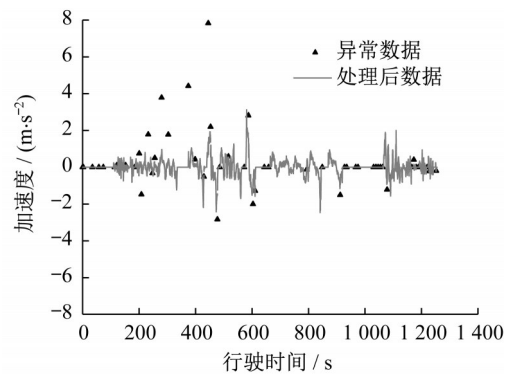


图4 异常值处理效果

Fig.4 Treatment of abnormal value

的判别方法进行了研究。

1.2.1 急加速

对车辆速度与前向加速度参数进行分析得到图5。使用式(1)对前向加速度与速度关系中的各分位值进行线性拟合,结果如表2所示。

$$f(x) = \beta_1 x + \beta_2 \quad (1)$$

式中: x 表示车辆速度; β_1 和 β_2 是拟合参数; R^2 表示拟合系数。

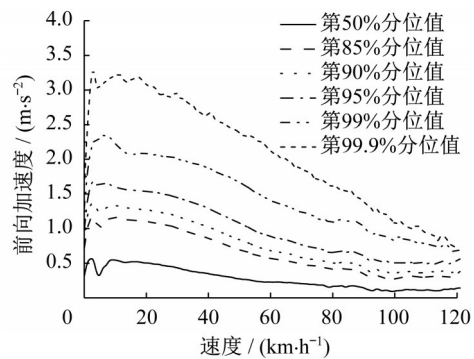


图5 前向加速度分位值随速度变化关系

Fig.5 The relationship between acceleration quantile value and velocity

文献[9]将最大加速度阈值设定为 $2.78\text{m}\cdot\text{s}^{-2}$,文献[16-19]将最大加速度阈值设定为 $3\text{m}\cdot\text{s}^{-2}$ 。由图6可知,前向加速度的峰值随速度增加呈降低的趋势,因此,如果将最大加速度阈值设为固定值,无法准确反映车辆实际运行情况。为此本文提出了多参数组合阈值边界线进行危险驾驶行为识别的思路,即随着车速的变化,最大加速度阈值亦相应调整,根据前向加速度峰值随速度的变化规律,定义了急加速判断阈值边界线为第99%分位值,如图6所示。

1.2.2 急减速

对车辆速度与制动减速度参数进行分析,各分

表1 驾驶行为特征参数

Tab.1 Driving behavior characteristic parameters

符号	含义	计算方法
$v_{\max}/(\text{km}\cdot\text{h}^{-1})$	最高车速	$v_{\max}=\text{Max}(v_1, v_2, \dots, v_n)$
$v_m/(\text{km}\cdot\text{h}^{-1})$	平均车速	$v_m = \frac{\sum_{i=1}^n v_i}{n}$
$v_{\text{std}}/(\text{km}\cdot\text{h}^{-1})$	车速标准差	$v_{\text{std}} = \sqrt{\frac{\sum_{i=1}^n (v_i - v_m)^2}{n}}$
$a_m^+/(\text{m}\cdot\text{s}^{-2})$	平均前向加速度	$a_m^+ = \frac{\sum_{i=1}^n a_i^+}{n}$
$a_{\text{std}}^+/(\text{m}\cdot\text{s}^{-2})$	前向加速度标准差	$a_{\text{std}}^+ = \sqrt{\frac{\sum_{i=1}^n (a_i^+ - a_m^+)^2}{n}}$
$a_m^-/(\text{m}\cdot\text{s}^{-2})$	平均制动减速度	$a_m^- = \frac{\sum_{i=1}^n a_i^-}{n}$
$a_{\text{std}}^-/(\text{m}\cdot\text{s}^{-2})$	制动减速度标准差	$a_{\text{std}}^- = \sqrt{\frac{\sum_{i=1}^n (a_i^- - a_m^-)^2}{n}}$
$\omega_m/((^\circ)\cdot\text{s}^{-1})$	平均方向盘角速度	$\omega_m = \frac{\sum_{i=1}^n \omega_i}{n}$
$\omega_{\text{std}}/((^\circ)\cdot\text{s}^{-1})$	方向盘角速度标准差	$\omega_{\text{std}} = \sqrt{\frac{\sum_{i=1}^n (\omega_i - \omega_m)^2}{n}}$
$F_a/(\text{次}\cdot 100\text{km}^{-1})$	100km急加次数	$F_a = \frac{N_a}{D} \times 100$
$F_d/(\text{次}\cdot 100\text{km}^{-1})$	100km急减次数	$F_d = \frac{N_d}{D} \times 100$
$F_t/(\text{次}\cdot 100\text{km}^{-1})$	100km急转向次数	$F_t = \frac{N_t}{D} \times 100$
$F_{\text{st}}/(\text{次}\cdot 100\text{km}^{-1})$	100km转向超速次数	$F_{\text{st}} = \frac{N_{\text{st}}}{D} \times 100$
$P_{\text{os1}}/\%$	车速80~120km·h ⁻¹ 的时间占比	$P_{\text{os1}} = \frac{t_{\text{os1}}}{T} \times 100$
$P_{\text{os2}}/\%$	车速超过120km·h ⁻¹ 的时间占比	$P_{\text{os2}} = \frac{t_{\text{os2}}}{T} \times 100$

注: i 为单次行驶事件中数据点; v_i 为每个采样点的车速; a_i^+ 为每个加速度数据; a_i^- 为每个减速度数据; ω_i 为每个方向盘转角数据; n 为单次行驶事件中总数据点数; N 为危险驾驶行为发生次数; D 为单次行驶总里程; t 为单次行驶事件超速累积时长; T 为单次事件行驶时长。

表2 前向加速度各分位值的拟合参数

Tab.2 The fitting parameters of each quantile value of forward acceleration

参数	第50%分位值	第85%分位值	第90%分位值	第95%分位值	第99%分位值	第99.9%分位值
β_1	-0.004 2	-0.009 0	-0.010 2	-0.011 8	-0.015 0	-0.023 8
β_2	0.531 2	1.205 6	1.400 2	1.707 6	2.500 0	3.553 4
R^2	0.912 3	0.935 1	0.938 6	0.947 7	0.970 8	0.993 0

位值结果如图7所示,由图可见,制动减速度在约25km·h⁻¹时存在拐点,这是由于研究用车辆存在制动能量回收,在25km·h⁻¹时有最大的制动回收力矩。制动减速度各分位值在速度超过100km·h⁻¹变化幅度较小,使用式(2)对制动减速度分位值进行分段线性拟合,以速度值25km·h⁻¹为线性拟合分段边界,结

参考急加速阈值线设定,本文同样定义了急减速判断阈值线为第99%分位值,如图8所示。

果如表3所示。

$$f(x) = \begin{cases} \beta_1 x + \beta_2 & x < 25\text{km}\cdot\text{h}^{-1} \\ \beta_3 x + \beta_4 & x \geq 25\text{km}\cdot\text{h}^{-1} \end{cases} \quad (2)$$

式中: x 表示车辆速度; $\beta_1, \beta_2, \beta_3, \beta_4$ 是拟合参数; R^2 表示拟合系数。

1.2.3 超速

在不同的道路环境中限速是不同的,在城市道路中,车辆的限速一般为60km·h⁻¹,高速道路中一般

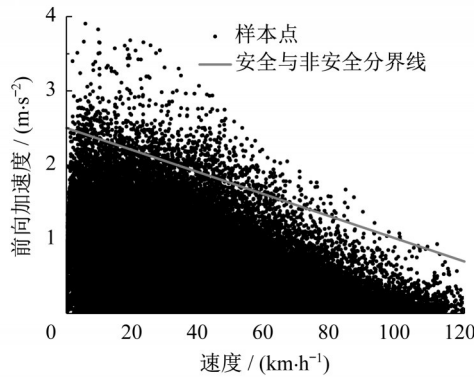


图6 急加速判断阈值线

Fig.6 The judgment threshold line of rapid acceleration

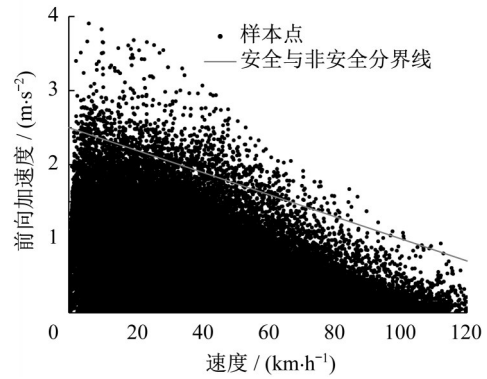


图8 急减速判断阈值线

Fig.8 The judgment threshold line of sharp deceleration

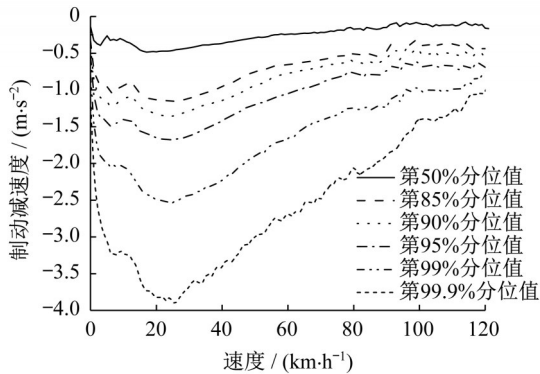


图7 制动减速度分位值随速度变化关系

Fig.7 The relationship between braking deceleration quantile value and velocity

最高限速为 $120\text{km}\cdot\text{h}^{-1}$ 。文献[20]认为车速在高于 $80\text{km}\cdot\text{h}^{-1}$ 就认为驾驶员有超速倾向或已经在超速行驶了,而车速高于 $120\text{km}\cdot\text{h}^{-1}$ 必然是超速行驶了。但准确的超速行驶行为的识别需要基于不同路段的限速情况进行,由于本文缺少道路限速信息,仅采用两个等级指标表征超速行为的严重程度:车速 $80\sim 120\text{km}\cdot\text{h}^{-1}$ 的行驶时间占总行驶时间的比例和车速高于 $120\text{km}\cdot\text{h}^{-1}$ 行驶时间占总行驶时间的比例。

1.2.4 急转向

对车辆速度与方向盘角速度参数进行分析,各分位值结果见图9,随车速的增加,方向盘角速度持

续降低,为此使用式(3)对方向盘角速度的各分位值进行曲线拟合,结果如表4所示。

$$f(x) = a + b \times x^c \quad (3)$$

式中: x 表示车辆速度; a 、 b 、 c 是拟合参数; R^2 表示拟合系数。

同理,本文定义了急转向判断阈值线为第99%分位值,阈值线如图10所示。

1.2.5 转向超速

对车辆速度与方向盘转角参数进行对比分析,各分位值结果见图11,使用式4对不同方向盘转角下的车速各分位值曲线进行拟合,结果如表5所示。

$$f(x) = a + b \times x^c \quad (4)$$

式中: x 表示方向盘转角; a 、 b 、 c 是拟合参数; R^2 表示拟合系数。

同理,本文定义了转向安全车速分界线为第99%分位值,转向超速行为通过该分界线进行识别,如图12所示。

1.2.6 疲劳驾驶

统计结果表明疲劳驾驶造成的交通事故占事故总数的20%左右,占特大交通事故总数的40%以上^[21]。行车时长是影响疲劳驾驶的直观因素,本文将单次行车累计时长超过4h记为疲劳驾驶。

表3 制动减速度各分位值的拟合参数

Tab.3 The fitting parameters of each quantile value of braking deceleration

参数	第50%分位值	第85%分位值	第90%分位值	第95%分位值	第99%分位值	第99.9%分位值
β_1	-0.008 7	-0.012 8	-0.014 6	-0.018 2	-0.030	-0.056 5
β_2	-0.290 3	-0.852 9	-1.014 7	-1.264 8	-1.850 0	-2.643 3
β_3	0.004 7	0.010 5	0.012	0.014 4	0.022 0	0.030 1
β_4	-0.545 6	-1.354 4	-1.580 1	-1.974 1	-3.150 0	-4.570 0
R^2	0.826 6	0.826 3	0.840 2	0.883 8	0.949 4	0.922 0

表4 方向盘角速度各分位值的拟合参数

Tab.4 The fitting parameters of each quantile value of the steering wheel angular velocity

参数	第50%分位值	第85%分位值	第90%分位值	第95%分位值	第99%分位值	第99.9%分位值
a	0.97	5.82	7.64	9.83	19.56	3.80
b	-53.27	-209.30	-270.77	-373.06	-680.00	-692.02
c	0.948	0.940	0.936	0.936	0.935	0.949
R^2	0.993 2	0.998 2	0.998 5	0.998 5	0.994 2	0.993 5

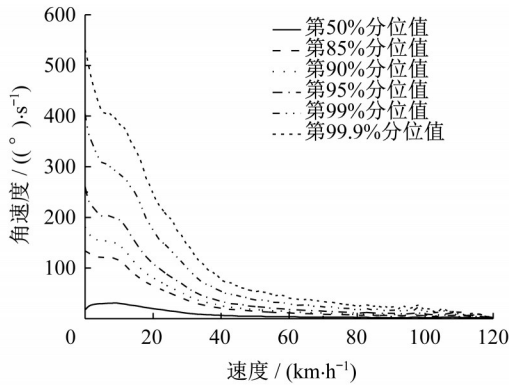


图9 方向盘角速度分位值随速度变化关系

Fig.9 The relationship between steering wheel angular quantile value and velocity

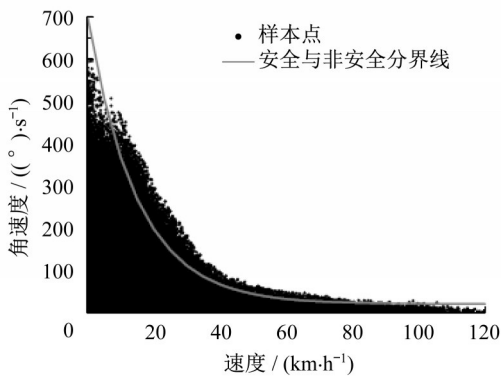


图10 急转向判断阈值线

Fig.10 The judgment threshold line of sharp turning

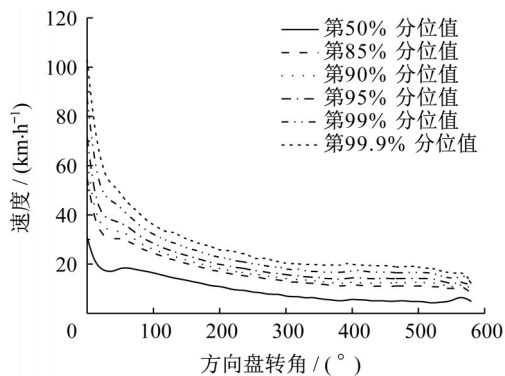


图11 车速各分位值随方向盘转角变化关系

Fig.11 The relationship between speed quantile value and the steering wheel angle

2 驾驶风格分析

2.1 高斯混合模型

在聚类算法中 K 均值聚类算法因原理易懂、收敛速度快的优点得到广泛应用。但 K 均值聚类算法结果受初始聚类中心影响非常大,而且其要求簇的形状必须是圆形的,若实际样本点是椭圆分布的,其聚类结果可能会出现多个圆形的簇混在一起,聚类效果差。高斯混合模型(GMM)可以弥补 K 均值聚类算法的不足,GMM未将每个样本点置于明确簇中,而给出了该样本点在各簇中的概率或可能性,因此能够有效避免硬分配,并且簇的形状可以是任意椭圆状,而不仅局限于圆形。

假设数据集 $X = \{x_1, x_2, \dots, x_n\}$ 服从由 K 个多元高斯分布组成的GMM,即可分为 K 类。对于每个样本 x_i ,高斯混合分布的概率密度函数为

$$P(x_i|\Theta) = \sum_{k=1}^K \pi_k N_k(x_i|\mu_k, \sigma_k) \quad (5)$$

式中: π_k 是混合系数; μ_k, σ_k 为每个高斯成分的均值和方差。 $N_k(x_i|\mu_k, \sigma_k)$ 为第 k 个高斯成分的概率密度函数,即 x_i 属于第 k 类的概率,可表示为

$$N_k(x_i|\mu_k, \sigma_k) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \quad (6)$$

使用期望最大化(EM)算法对GMM进行非线性概率函数的优化。首先根据类别数目设定模型参数初值 (μ_{0k}, σ_{0k}) ,迭代多次求出参数 (μ_k, σ_k) ,即每一个样本属于第 k 类的概率,由最大后验概率准则可知,后验概率最大的那项类别为该样本所属聚类结果^[22]。

2.2 降维方法选择

为了避免高维数据集导致的“维灾难效应”,在利用聚类算法进行驾驶风格分类时需要提前对驾驶行为数据集进行降维处理。常见的降维方法有多维尺度分析(MDS)、主成分分析(PCA)、等距特征映射(ISOMAP)和 t 分布随机邻域嵌入(t -SNE)等^[23-24]。文献[24]指出在处理不同的高维数据集时,需要针对数据集特点选择适合的方法。因此本文采

表5 车速各分位值拟合参数

Tab.5 The fitting parameters of each speed quantile value

参数	第50%分位值	第85%分位值	第90%分位值	第95%分位值	第99%分位值	第99.9%分位值
a	26.00	76.00	113.0	341.0	3 395	6 166
b	-0.303 2	-24.00	-53.00	-266.0	-3 300	-6 053
c	0.690 9	0.169 2	0.110 8	0.035 1	0.004 1	0.002 7
R^2	0.952 6	0.981 6	0.985 5	0.987 7	0.991 6	0.989 2

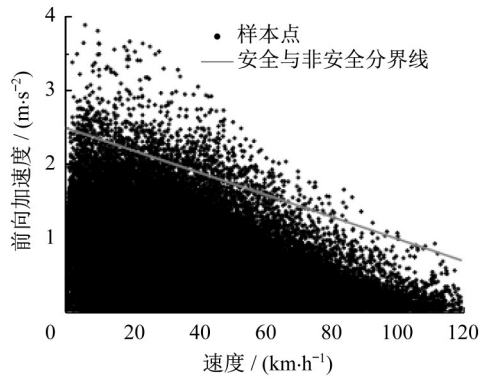


图12 转向超速判断阈值线

Fig.12 The judgment threshold line of overspeed while turning

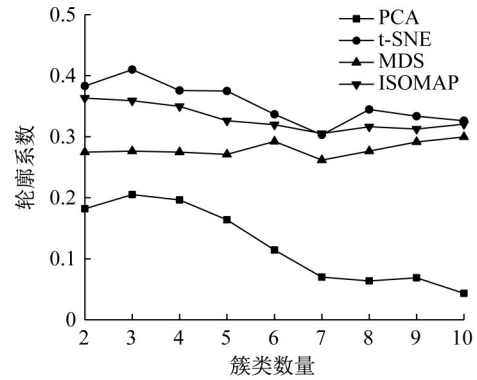


图13 不同K取值的轮廓系数

Fig.13 Silhouette coefficient with different values of K

用PCA、MDS、 t -SNE和ISOMAP 4种降维方法分别对驾驶风格特征参数进行降维,再将降维结果分别输入GMM算法中进行驾驶风格聚类。不同降维方法的效果采用轮廓系数S进行对比。

$$S(i) = \frac{1}{K} \sum_{k=1}^K \left\{ \frac{1}{n_k} \sum_{i \in I_k} \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \right\} \quad (5)$$

式中: K 表示分类数量; n_k 表示第 k 类的观测值总数; $a(i)$ 是观测值 x_i 与所在第 k 类其他值的平均距离; $b(i)$ 是观测值 x_i 与其他集群中所有值的平均距离中最小值; I_k 表示第 k 类集群观测值的索引集。

本文基于驾驶风格特征参数数据集比较了不同 K 取值的轮廓系数,确定最佳聚类 K 值,由图13可知将驾驶风格特征参数为3类效果最好。

将上述降维方法结合 K 均值聚类算法和GMM两种聚类算法进行聚类效果的对比,结果如表6所示。结果表明对于本文所用驾驶风格数据集而言,相对于 K 均值聚类算法,GMM与每种降维方法组合后普遍表现出更优的聚类效果,且 t -SNE和GMM组合聚类方法的轮廓系数最高。因此,本文最终选择 t -SNE和GMM算法把驾驶风格特征参数数据集

根据表7中各类特征参数平均值结果,发现第1类样本的百公里急加/急减/急转向/转向超速次数、加(减)速度平均值与标准差都是3类中最高的,而这些参数都与驾驶粗暴相关,故推断第1类为激进

聚为3类驾驶风格。

表6 不同降维与聚类算法聚类效果

Tab.6 Clustering effect of different dimension reduction and clustering algorithms

方法	PCA	MDS	t -SNE	ISOMAP
$S_{K-means}$	0.154	0.147	0.136	0.153
S_{GMM}	0.194	0.288	0.408	0.348

2.3 驾驶风格分类结果分析

驾驶行为是指驾驶员在某一次驾驶过程具体的操作行为,驾驶风格是指驾驶员在驾驶时所表现的综合行为特征,也是驾驶员养成的基本固定的驾驶习惯。也就是说驾驶行为能够反映驾驶风格,驾驶风格从一定程度上决定驾驶行为。为了揭示3种不同驾驶风格对应的车辆运行特征,本文从监测数据服务器接收的私家车与网约车数据中,提取同一款纯电动车型30名驾驶员一个月的驾驶事件(共2 227次)对应的驾驶数据。运用 t -SNE算法对驾驶数据进行降维,采用GMM算法对降维后数据进行驾驶风格分类,聚类效果如图14所示,分类结果中各特征参数的平均值如表7所示。

型风格;相对而言,第3类样本的车速、加(减)速度平均值与标准差都是3类中最低的,故推断第3类为保守型风格,推断第2类为普通型风格。保守型的平均速度比激进型低约 $15 \text{ km}\cdot\text{h}^{-1}$,这是因为保守型

表 7 各类特征参数平均值

Tab.7 Average value of various characteristic parameters

类别	v_{max}	v_m	v_{std}	a_m^+	a_{std}^+	a_m^-	a_{std}^-	ω_m	ω_{std}
1	90.11	36.02	24.51	0.54	0.5	0.38	0.51	12.59	39.18
2	84.63	27.01	21.56	0.42	0.38	0.31	0.42	9.19	32.04
3	67.05	22.18	17.76	0.40	0.37	0.29	0.39	15.21	43.54
	F_a	F_d	F_l	F_{st}	P_{os1}	P_{os2}			
1	32.41	5.32	21.42	8.38	7.05	0.03			
2	10.99	2.65	9.17	3.19	1.25	0			
3	9.84	1.87	7.34	2.70	0.11	0			

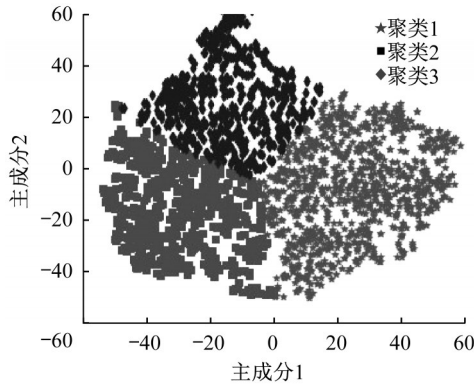


图 14 驾驶风格聚类效果

Fig.14 Clustering effect of driving style

驾驶风格为确保行车安全,将车速稳定在安全车速以内。普通型驾驶风格也是熟练驾驶人群众常用的驾

驶风格,保持速度稳定的能力较强,激进型驾驶风格对车速控制能力最差。图 15 为不同驾驶风格对应的危险驾驶行为发生次数对比结果,由图可见,激进型驾驶风格的危险驾驶行为在高发区间的占比最高。说明激进型对车辆控制能力最差,危险驾驶行为发生次数最多,发生交通事故的风险最高,对于这一类驾驶员群体,可针对性进行安全教育,提高驾驶员安全意识,减少交通事故伤亡和损失。相对而言,对于保守型驾驶风格而言,危险驾驶行为发生次数主要集中在低发区。普通型风格对应的危险驾驶行为发生次数介于保守型和激进型之间。不同风格驾驶员的车辆运行特性分类结果也验证了本文提出的分类算法是有效的。

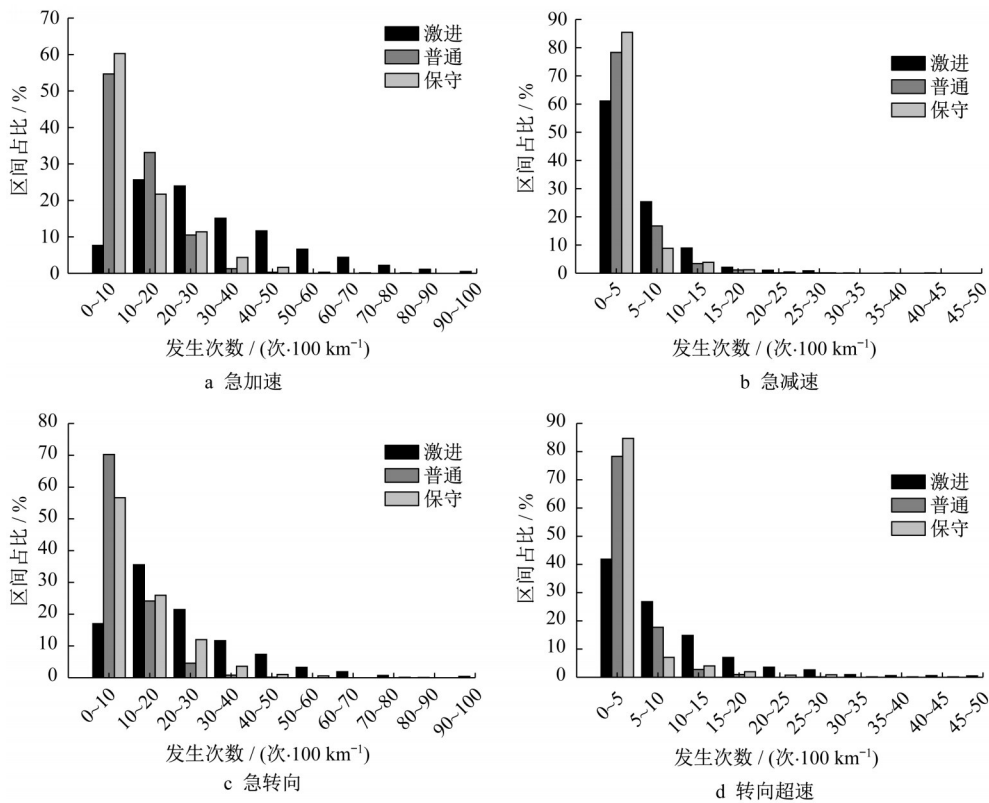


图 15 危险驾驶行为发生次数分布情况

Fig.15 Distribution of the occurrence times of dangerous driving behavior

3 结论

为提高驾驶员安全认知,探究驾驶风格划分方法,本文利用纯电动汽车高频运行数据开展研究工作,主要结论如下:

(1)对驾驶行为数据进行统计分析,根据分析结果提出了采用多参数组合阈值边界线进行异常驾驶行为的识别,并得到了不同驾驶行为特征参数的危险驾驶阈值边界线。

(2)选取了15个与驾驶安全有关的驾驶风格特征参数,对多参数降维及聚类算法进行研究,通过轮廓系数对2种聚类算法和4种降维方法的性能进行评价,结果表明 t -SNE和GMM组合算法分类效果最好。

(3)以 t -SNE和GMM组合算法建立了基于驾驶安全的驾驶风格分类模型,将驾驶风格分为3种类型,不同驾驶风格对应的车辆运行特性分析结果验证了分类算法的有效性。

作者贡献声明:

纪少波:论文研究思路提出及语言组织。

张珂:车辆运行特性统计分析。

李伦:驾驶风格分类方法对比。

苏士斌:试验样本方案制定。

何绍清:车辆运行数据采集。

冯远宏:数据接收方法研究。

张强:数据预处理方法研究。

参考文献:

- [1] 亓航,刘畅,伍毅平,等.基于图谱的驾驶行为特征安全性描述方法研究[J].交通工程,2019,19(6):1.
QI Hang, LIU Chang, WU Yiping, *et al.* Research on safety description method of driving behavior characteristics based on graph [J]. Traffic Engineering, 2019, 19(6): 1.
- [2] 徐婷,张香,张亚坤,等.基于AdaBoost算法的货车驾驶人安全倾向性分类[J].安全与环境学报,2019,19(4):1273.
XU Ting, ZHANG Xiang, ZHANG Yakun, *et al.* Truck driver safety tendency classification based on adaboost algorithm [J]. Journal of Safety and Environment, 2019, 19(4): 1273.
- [3] 余荣杰,龙晓捷,涂颖菲,等.基于低频轨迹数据的分时租赁驾驶人驾驶风格分析[J].同济大学学报(自然科学版),2019,47(10):1463.
YU Rongjie, LONG Xiaojie, TU Yingfei, *et al.* Driving style analysis of time-sharing rental drivers based on low-frequency trajectory data [J]. Journal of Tongji University (Natural Science Edition), 2019, 47(10): 1463.
- [4] SHI Xiupeng, WONG Yiik Diew, LI Michael Zhifeng, *et al.* A feature learning approach based on XGBoost for driving assessment and risk prediction [J]. Accident Analysis & Prevention, 2019, 129: 170.
- [5] 牛增良,李海斌,王文峰,等.基于聚类分析的营运驾驶人危险驾驶行为研究[J].山东交通学院学报,2014,22(1):19.
NIU Zengliang, LI Haibin, WANG Wenfeng, *et al.* Study on dangerous driving behavior of operating drivers based on cluster analysis [J]. Journal of Shandong Jiaotong University, 2014, 22(1): 19.
- [6] 文江辉,杨先敏,吴超仲.大数据环境下营运车辆驾驶行为特征分析[J].交通信息与安全,2018,36(4):42.
WEN Jianghui, YANG Xianmin, WU Chaozhong. Analysis on driving behavior characteristics of operating vehicles in big data environment [J]. Traffic Information and Safety, 2018, 36(4): 42.
- [7] ZHU Bing, JIANG Yuande, ZHAO Jian, *et al.* Typical-driving-style-oriented personalized adaptive cruise control design based on human driving data [J]. Transportation Research Part C: Emerging Technologies, 2019, 100: 274.
- [8] 李立治,杨建军,刘双喜,等.国内人群的驾驶风格分类及识别方法研究[J].重庆理工大学学报(自然科学),2019,33(11):33.
LI Lizhi, YANG Jianjun, LIU Shuangxi, *et al.* Study on the classification and recognition of driving style of domestic crowds [J]. Journal of Chongqing University of Technology (Natural Science), 2019, 33(11): 33.
- [9] 李卓轩,林凯迪,郭建华,等.基于车联网数据的运输车辆安全评价模型[J].南通大学学报(自然科学版),2020,19(1):26.
LI Zhuoxuan, LIN Kaidi, GUO Jianhua, *et al.* Transportation vehicle safety evaluation model based on internet of vehicles data [J]. Journal of Nantong University (Natural Science Edition), 2020, 19(1): 26.
- [10] 刘通,付锐,张名芳,等.融合K-means与高斯混合模型的驾驶风格聚类研究[J].中国安全科学学报,2019,29(12):40.
LIU Tong, FU Rui, ZHANG Mingfang, *et al.* Study on driving style clustering based on k-means and gaussian mixture model [J]. China Safety Science Journal, 2019, 29(12): 40.
- [11] 张志鸿.基于OBD数据分析的驾驶行为研究[D].西安:长安大学,2017.
ZHANG Zhihong. Research on driving behavior based on OBD data analysis [D]. Xi'an: Chang'an University, 2017.
- [12] 吴紫恒,吴仲城,张俊,等.基于模糊C均值和神经网络的驾驶行为评价研究[J].计算机系统应用,2018,27(3):263.
WU Ziheng, WU Zhongcheng, ZHANG Jun, *et al.* Research on driving behavior evaluation based on fuzzy C-means and neural network [J]. Application of Computer Systems, 2018, 27(3): 263.
- [13] 李经纬,赵治国,沈沛鸿,等.驾驶风格K-means聚类与识别方法研究[J].汽车技术,2018(12):8.
LI Jingwei, ZHAO Zhiguo, SHEN Peihong, *et al.* Study on K-

- means clustering and recognition of driving style [J]. *Automobile Technology*, 2018(12): 8.
- [14] 王海星, 王翔宇, 王招贤, 等. 基于数据挖掘的危险货物运输风险驾驶行为聚类分析 [J]. *交通运输系统工程与信息*, 2020, 20(1): 183.
- WANG Haixing, WANG Xiangyu, WANG Zhaoxian, *et al.* Clustering analysis of dangerous goods transportation risk driving behavior based on data mining [J]. *Journal of Transportation Systems Engineering and Information*, 2020, 20(1): 183.
- [15] 黄曦. 城市下穿隧道交通安全分析 [D]. 成都: 西南交通大学, 2013.
- HUANG Xi. *Traffic Safety Analysis of Urban Underpass Tunnel* [D]. Chengdu: Southwest Jiaotong University, 2013.
- [16] 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]. 西安: 长安大学, 2015.
- XU Shuquan. *Research on driving behavior safety and energy saving evaluation method based on vehicle operation monitoring system* [D]. Xi'an: Chang'an University, 2015.
- [17] 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: 北京交通大学, 2016.
- XIA Jie. *Evaluation method of driving behavior safety and energy saving based on safety production management data of road transportation enterprises* [D]. Beijing: Beijing Jiaotong University, 2016.
- [18] 彭江琴. 基于GID的车联网保险UBI费率与驾驶行为评分研究 [D]. 南京: 南京邮电大学, 2016.
- PENG Jiangqin. *Research on UBI rate and driving behavior score of Internet of vehicles insurance based on GID* [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2016.
- [19] 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.
- REN Huijun, XU Tao, LI Xiang. *Analysis of bus driving safety based on on-board GPS trajectory data* [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(6): 739.
- [20] 朱爽. 车联网环境下基于UBI的车险费率厘定模式与方法研究 [D]. 北京: 北京交通大学, 2015.
- ZHU Shuang. *Research on the model and method of car insurance rate determination based on ubi in the environment of internet of vehicles* [D]. Beijing: Beijing Jiaotong University, 2015.
- [21] 李福林. 疲劳驾驶的危害及预防 [J]. *产业与科技论坛*, 2013, 12(6): 127.
- LI Fulin. *Harm and prevention of fatigue driving* [J]. *Forum of Industry and Science & Technology*, 2013, 12(6): 127.
- [22] BESTEL Robert, DAUS Andreas W., Thielemann Christiane. A novel automated spike sorting algorithm with adaptable feature extraction [J]. *Journal of Neuroscience Methods*, 2012, 211(1): 168.
- [23] HINNEBURG Alexander, AGGARWAL Charu C., KEIM Daniel A. What is the nearest neighbor in high dimensional spaces? [C]//*Proceedings of the 26th International Conference on Very Large Data Bases*. San Francisco: Morgan Kaufmann Publishers Inc. 2000: 506-515.
- [24] 苏飞. 高维数据的聚类分析研究及应用 [D]. 郑州: 华北水利水电大学, 2017.
- SU Fei. *Research and application of cluster analysis of high-dimensional data* [D]. Zhengzhou: North China University of Water Resources and Electric Power, 2017.