

基于自然驾驶数据构建仿真测试变道切入场景库的方法

宋华^{1,2}, 吴贤静^{1,2}, 吴琼^{1,2}, 陈波^{1,2}, 丁钊^{1,2}

(1. 安徽江淮汽车集团股份有限公司, 合肥 230601; 2. 安徽省汽车智能网联技术重点实验室, 合肥 230601)

摘要: 智能网联汽车采用虚拟仿真测试已成为行业共识, 虚拟仿真测试依赖仿真场景。基于采集的高速公路自然驾驶场景数据, 建立自动化变道切入场景提取策略, 引入感知风险系数参数, 通过单因素方差分析非连续型要素, 皮尔逊相关性分析连续型要素, 揭示不同场景因素与风险的关系, 确认关键影响因素。采用 K-Means 聚类法和肘部法建立典型场景, 保留关键参数, 在 PreScan 仿真平台完成构建。研究表明, 该方法可根据真实采集数据完成变道切入关键场景的有效提取, 并基于关键场景聚类成典型场景, 在仿真平台搭建此场景, 可以为自动驾驶系统测试提供科学支持。

关键词: 自然驾驶数据; 变道切入; 感知风险系数; 单因素方差; 皮尔逊相关性; K-Means 聚类; 肘部法

中图分类号: U495; TP389.1 文献标志码: A DOI: 10.3969/j.issn.2095-1469.2025.04.10

Building a Lane-Change Cut-in Scenario Library for Simulation Testing Based on Naturalistic Driving Data

SONG Hua^{1, 2}, WU Xianjing^{1, 2}, WU Qiong^{1, 2}, CHEN Bo^{1, 2}, DING Zhao^{1, 2}

(1. Anhui Jianghuai Automobile Group Co., Ltd., Hefei 230601, China;

2. Anhui Provincial Key Laboratory of Automotive Intelligent Connected Technology, Hefei 230601, China)

Abstract: Virtual simulation testing has become the industry-wide norm for intelligent connected vehicles. Such testing demands a rich set of simulation scenarios. Based on data collected from naturalistic highway-driving scenarios, the paper develops an automated strategy for extracting lane-change cut-in events. A perception risk coefficient is introduced, and one-way ANOVA is used to analyze discrete factors while Pearson correlation analysis is employed for continuous factors, revealing the relationships between scenario elements and risk and identifying the key influencing factors. Typical scenarios are then derived with K-Means clustering and the elbow method, key parameters are retained, and each scenario is constructed on the Prescan simulation platform. The results show that this method can effectively extract critical lane-change cut-in scenarios from real-world data, cluster them into typical scenarios, and reconstruct these scenarios on the simulation platform, providing scientific support for autonomous-driving system testing.

Keywords: natural driving data; lane-change and cut-in; perception risk coefficient; one-way ANOVA; pearson correlation; K-Means clustering; elbow method

收稿日期: 2025-01-06 改稿日期: 2025-02-14 网络首发日期: 2025-03-03

参考文献引用格式:

宋华, 吴贤静, 吴琼, 等. 基于自然驾驶数据构建仿真测试变道切入场景库的方法[J]. 汽车工程学报, 2025, 15(4): 528-538.

SONG Hua, WU Xianjing, WU Qiong, et al. Building a Lane-Change Cut-in Scenario Library for Simulation Testing Based on Naturalistic Driving Data[J]. Chinese Journal of Automotive Engineering, 2025, 15(4): 528-538. (in Chinese)



自动驾驶系统在减少事故、降低环境污染以及缓解驾驶员压力方面展现出显著效果^[1]。然而，安全性仍是高级别自动驾驶技术面临的主要挑战之一，要实现商业化量产，必须经过严格的测试流程。目前，行业内普遍采用“三支柱”测试方法，即仿真测试、封闭场地测试及开放道路测试^[2]。

变道切入作为驾驶过程中最频繁的操作之一，在中国这样一个人口众多且各地交通状况差异巨大的国家，会对驾驶安全造成较大威胁。据统计，2015年我国因变道切入引发的交通事故占比达到4.9%^[3]；而在美国，每年约有24~61万起交通事故是由不当的变道行为造成的^[4]。

传统的实车测试不仅耗时费力，还难以高效地重现复杂多变的实际路况^[5]。因此，虚拟仿真场景成为提升自动驾驶汽车技术验证效率的关键手段，也是当前各大车企和一级供应商研究的重点方向^[6]。

在仿真软件中手动搭建虚拟场景多数是以测试系统的边界为目的，对于合规性的测试可信度不足^[7]。

姚明菊等^[8]提出，目前虚拟测试场景生成方法存在相对误差大、绝对误差大和逼真度低的问题。

刘志强等^[9]利用美国高速公路数据集来开展驾驶变道风格的分析。中国由于人口密度大、交通场景复杂等特点，与美国在交通规则、驾驶习惯等方面有较大差异。在中国有交通信号灯的地方，行人或车辆通常会严格遵守信号灯的指示，在没有交通信号灯的路口，行人或车辆容易因无序通行引发事故。变道和超车行为也较频繁，尤其是在拥堵路段，会出现频繁穿插车道的现象。而在美国，一般会遵守交通信号灯，闯红灯现象较少，变道和超车行为也相对规范。所以建立符合中国交通特点的场景库更重要。

本研究收集了安徽省省内部分高速公路环境下的真实自然驾驶数据，并从中筛选出涉及变道切入的部分进行详细分析，制定关键场景筛选策略，通过量化切入场景的危险程度，计算各参数与感知危

险程度间的关联度来评估其对变道切入的影响，提取影响变道切入的关键参数。此外，利用K-Means聚类算法对关键参数下的采集数据进行聚类分析，得出变道切入典型场景。随后借助PreScan软件创建相应的模拟场景，探索如何有效地在仿真环境中再现真实的变道切入情境，以满足大规模测试的需求。

1 自然驾驶场景数据的采集与分析

1.1 数据的采集

本研究依托江淮汽车的数据集开展，该数据集是在贴近实际驾驶环境的多样化场景中，通过1辆集成高级传感器技术的专用数据采集车辆搜集得到，确保了数据的广泛代表性和高度真实性。该采集车配备1套数据采集系统，包括6个摄像头、1个前视摄像头、6个毫米波雷达、1个GPS定位器以及1个采集系统主机，具体参数见表1。通过CAN总线记录车辆状态信息。其中，GPS提供实时的经纬度位置，并作为时间同步的基准；毫米波雷达则用于检测目标物体的信息，如尺寸、距离和速度等；前视摄像头安装在挡风玻璃中上方，而6个摄像头则布置于车辆前方及左右两侧，以提供全方位的视觉覆盖，确保获取到更全面的视频资料。

本研究基于采集系统采集到的8 241 km自然驾驶数据，主要为安徽省内部高速公路的实际行驶情况，为了实现全国范围内适用的变道切入场景库，目前正在全国不同范围、不同路况的场景下采集，旨在通过多样化的数据源来提高自动驾驶系统开发与测试的准确性和可靠性。

1.2 数据预处理

由于数据采集系统的精度以及自然场景下的干扰等原因，采集的数据存在异常值或数据丢失等情况。在进行数据分析前，需对数据进行预处理，以消除数据异常带来的影响。

本研究采用线性插值法对缺失值进行预处理，取缺失值前后数值，进行平均化作为当前值，再使用卡尔曼滤波方法进行噪声抑制、数据平滑，并保

表 1 采集系统传感器参数

| | 摄像头 | 前视摄像头 | 毫米波雷达 |
|----------|---|---------|---|
| 安装位置 | 前视: 前挡风玻璃; 前、后、左、右: 车辆后视镜下沿; 后视: 车辆后视镜顶部。 | 前挡风玻璃中间 | 正前: 前保险杠中间; 前左、前右、后左、后右: 车辆前左、前右、后左、后右; 正后: 后保险杠中间。 |
| 分辨率 | 1 600×1 200 | | |
| 探测距离/m | 160、80、10 | 0~260 | 0~260 |
| 探测角度/(°) | 52、112、192 | 100 | 近距 120/远距 18 |

留数据的有效性, 通过试验和误差分析不断优化参数, 最终设置过程噪声协方差矩阵 Q 为 2、测量噪声协方差矩阵 R 为 5。图 1 为数据预处理前后的对比。

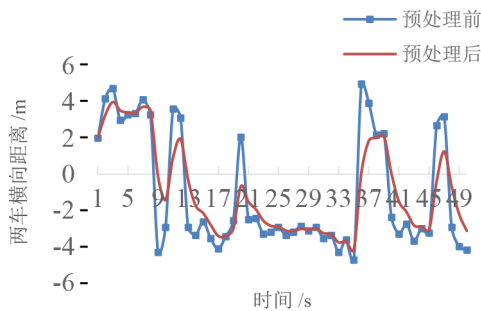


图 1 数据滤波前后对比

2 切入场景的提取

2.1 切入场景的定义

丁洁云等^[10]对切入行为的开始与结束进行了定义。切入的开始定义为目标车辆接近主车所在车道的时刻。切入的结束定义为目标车辆越过主车所在车道中心线的时刻。

本文参考此定义, 将目标车在主车相邻车道内, 且靠近主车侧前轮与两车中间车道线距离定义为 ΔS , 当前轮接触两车中间车道线时 ΔS 为 0, 从 ΔS 为 0 前推至 ΔS 值最大时刻, 定义为切入开始时刻。目标车车身中心的地面落点位于两车中间车道线上, 定义为目标车越线时刻。目标车穿过越线时刻, 车身完全处于主车车道上, 定义为切入结束时刻。图 2 为切入场景。

2.2 切入场景的提取策略

本研究基于前车前方目标物的变化筛选切入场

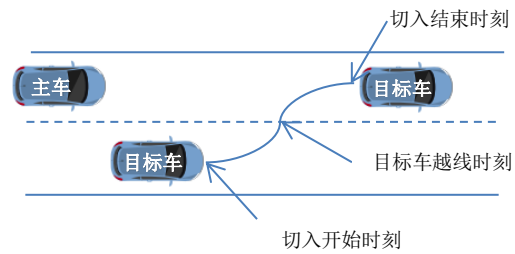


图 2 切入场景

景, 并从横向距离、纵向距离^[11]、两车车头时距 THW^[12] 等要素考虑切入场景的进一步提取策略。定义主车前方自车道内最近目标车辆为 CIPV, 当 1 个目标车辆由非 CIPV 变为 CIPV, 则说明该车辆由相邻车道侵入, 且距离主车最近, 这个过程为前车切入。根据 CIPV 进行切入场景初筛, 初筛出切入场景 1 000 例。为了保证场景的有效性, 此次初筛已经过初步处理, 删除两车纵向距离 >150 m、THW >5 s 的场景。

再根据车道曲率、主车行驶曲率、横向距离、纵向距离、车头时距 THW 对初筛的场景进行滤波, 排除因曲率过大导致的误判等逻辑, 筛选出关键的场景, 最后根据目标车横向绝对速度判断切入开始时刻, 截取切入场景。具体关键场景筛选策略为如下 4 点。

1) 曲率。车道曲率和主车行驶曲率低于 0.01 时, 判断可能由车辆行驶曲率过大导致的误判, 本研究暂不考虑, 以曲率 ≥ 0.01 为筛选条件, 进行选取。

2) 横向距离。中国道路的标准宽度为 3.5 m, 为保证切入行为的完整性, 两车在切入开始时的相对横向距离 (绝对值) 保持在 1.75~5.25 m 之间。

3) 纵向距离。当两车在距离较近, 目标车的

行为让主车感到有危险时，主车会下意识地采取动作来避免可能发生的危险。但是，当两车的纵向距离超过120 m时，主车不会做出相应动作^[13]。结合毫米波雷达的点云疏密程度，最终把两车的纵向距离定义为70 m。

4) 车头时距 THW。为了让目标车的切入行为对主车的影响最大，以 $THW < 2$ s 作为切入过程的选取标准。

根据切入场景筛选条件，本研究从 8 241 km 自然驾驶场景数据中，截取 197 个关键切入场景，用于要素与场景聚类分析。关键切入场景占有切入场景的比例为 19.7%。这 197 个关键切入场景基于关键场景标准提取，代表了具有挑战性和高应用价值的驾驶环境。

通过从海量的自然驾驶数据中提取出具有代表性的关键切入场景，能有效减少对所有场景的全面测试需求，节省时间和资源，更加高效地进行系统验证和优化。

3 切入场景的要素分析

3.1 切入场景的危险感知系数

相邻车道车辆突然切入会影响主车行车安全，为量化该危险程度，参考 OLSEN 等^[14]提出的切入场景危险感知系数 RP，同时考虑碰撞时间 TTC 和车头时距 THW 的影响。

TTC 根据两车的相对距离与相对速度比值计算，当 TTC 低于一定值时，说明两车间距过近，存在碰撞风险。RP 计算式为：

$$RP = \frac{5}{T_{TTC}} + \frac{1}{T_{THW}} = \frac{5V_{rel} + V_{ego}}{D_{rel}} \quad (1)$$

式中：RP 为危险感知系数； T_{TTC} 为 TTC 值； T_{THW} 为 THW 值； D_{rel} 为两车相对距离； V_{rel} 为两车相对纵向速度； V_{ego} 为主车的纵向速度。

对提取到的切入场景进行分析，如图 3 所示，在 95% 的置信水平下，RP 值遵循正态分布 ($P=0.07 > 0.05$)，可以进行统计学分析。

本研究选取车辆类型、切入方向、转向灯使用情况、天气情况、主车纵向速度、主车纵向加速

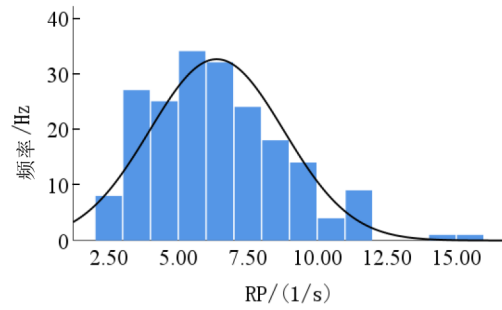


图3 PR值分布

度、相对纵向速度、两车纵向距离、两车横向距离、切入时间共 10 个要素分别进行显著性影响分析，判断是否为切入场景的显著影响因素。其中，前 4 个因素为离散型数据，后 6 个因素为连续型数据。

3.2 不同要素与 RP 的单因素方差分析

3.2.1 车辆类型

本文将采集到的切入车辆类型按大类研究。I 类为乘用车，包括乘用车、SUV、MPV、面包车等蓝牌车辆，II 类为商用车，包括大、中型载货车辆，大、中型城际车辆。

由图 4 可知，I 类车的 RP 值较 II 类车更集中，且较大。表 2 为车辆类型方差分析， F 值为组间均方与组内均方的比值，查表比较 F 值与显著性水平 $P=0.05$ 下的 F 临界值，当 F 值较大，意味着组间方差较大，组内方差较小，认为组间存在显著差异。在显著性水平为 0.05 时， $F=8.457$ ， $P=0.004 < 0.05$ ，结果表明，切入开始时刻车辆类型对 RP 值的影响是显著的。

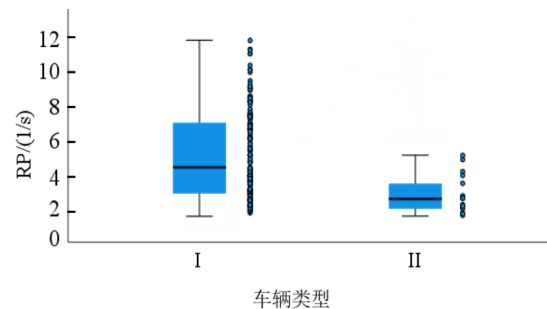


图4 不同切入车辆类型下的PR值分布

3.2.2 切入方向

在切入场景中，目标车的切入方向分左侧切入

表 2 车辆类型方差分析表

| | 平方和 | 自由度 | 均方 | F | 显著性 |
|----|-----------|-----|--------|-------|-------|
| 组间 | 48.259 | 1 | 48.259 | 8.457 | 0.004 |
| 组内 | 1 112.754 | 195 | 5.706 | | |

和右侧切入 2 种。由图 5 可知, 左切入的 RP 平均值大于右切入, 说明从左侧切入的风险大于右边。分析可知, 我国是左侧为主驾驶位, 对于驾驶员来说, 左侧视野更清晰, 切入时更大胆。

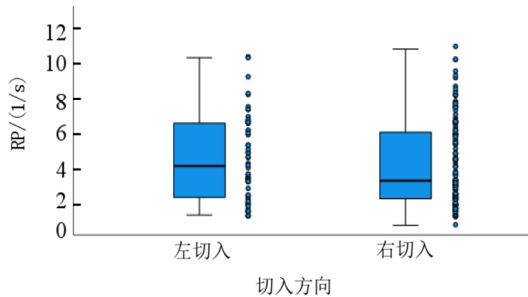


图 5 不同切入方向下的 PR 值分布

表 3 为切入方向方差分析, $F=0.495$, $P=0.483 > 0.05$ 。分析结果表明, 切入方向对危险感知系数 RP 没有显著影响。

表 3 切入方向方差分析表

| | 平方和 | 自由度 | 均方 | F | 显著性 |
|----|-----------|-----|-------|-------|-------|
| 组间 | 2.939 | 1 | 2.939 | 0.495 | 0.483 |
| 组内 | 1 158.074 | 195 | 5.939 | | |

3.2.3 转向灯使用情况

目标车切入开始时刻, 开启转向灯对相邻车道后方车辆起到警示作用。图 6 中, 未开启转向灯的均值大于开启转向灯的均值。在提取出的切入场景中, 转向灯的使用比例为 25%。远远低于戴思阳^[15]提出的高速公路上转向灯开启率 77.8%。同样也低于 OLSEN 等^[14]研究的美国转向灯使用率 53.3%。分析可知, 高速公路上人流稀少, 车辆速度普遍较高, 为了保持安全距离, 两车之间也会刻意拉开距离。此时主车驾驶员通过外后视镜观察到相邻车道上没有车辆或者距离较远时, 会在不开启转向灯的情况下直接切换到相邻车道。

表 4 为转向灯方差分析, $F=4.608$, $P=0.03 < 0.05$ 。结果表明, 切入开始时刻, 转向灯的状态对危险感知系数 RP 有显著性影响。

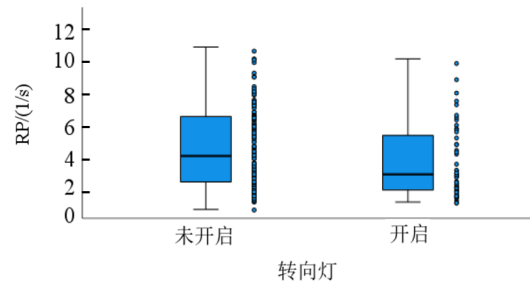


图 6 转向灯状态与 PR 值分布

表 4 转向灯方差分析表

| | 平方和 | 自由度 | 均方 | F | 显著性 |
|----|-----------|-----|--------|-------|-------|
| 组间 | 26.805 | 1 | 26.805 | 4.608 | 0.033 |
| 组内 | 1 134.208 | 195 | 5.816 | | |

3.2.4 天气情况

天气情况会对驾驶员视线、路面摩擦力等造成影响, 从而影响行驶安全性。图 7 中晴天的均值大于雨天的均值, 晴天时, 驾驶员的视线较好, 切入的行为比雨天更容易操作。

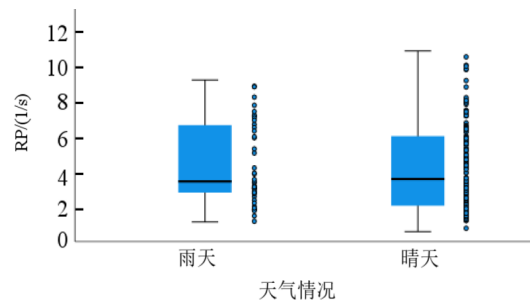


图 7 不同天气情况下的 PR 值分布

表 5 为天气情况方差分析, $F=0.343$, $P=0.559 > 0.05$ 。结果表明, 切入开始时刻, 环境因素中的天气情况对危险感知系数 RP 无显著性影响。

表 5 天气情况方差分析表

| | 平方和 | 自由度 | 均方 | F | 显著性 |
|----|-----------|-----|-------|-------|-------|
| 组间 | 2.040 | 1 | 2.040 | 0.343 | 0.559 |
| 组内 | 1 158.973 | 195 | 5.943 | | |

3.3 基于皮尔逊相关性检验的连续型场景要素分析

皮尔逊相关系数 (Pearson Correlation Coefficient), 简记为 r , 是一种用于度量 2 个定量变量 X 和 Y 之间线性相关程度的统计指标。其值域

介于-1~1之间。计算式为：

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} \sqrt{\sum(Y_i - \bar{Y})^2}} \quad (2)$$

式中： X_i 和 Y_i 分别为变量 X 和 Y 的第 i 个观测值； \bar{X} 和 \bar{Y} 分别为变量 X 和 Y 的样本均值。

该式反映了2个变量的协方差与它们标准差的乘积之比。

3.3.1 主车纵向速度

由图8可知，目标车在变道切入的过程中，2个参数存在正相关关系。经Pearson相关性检验， $r=0.232$ ，进一步验证了两者关系。显著性系数 $P=0.001<0.05$ ，说明两者之间有显著的相关性。

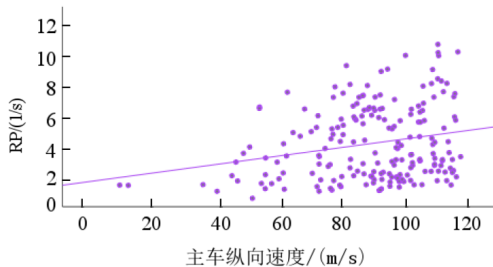


图8 主车纵向速度 & PR 散点

3.3.2 主车纵向加速度

由图9可知，主车纵向加速度范围主要处于 $-0.5\sim 0.5\text{ m/s}^2$ 之间，且主车纵向加速度的变化没有引起RP的明显变化。表明在目标车切入过程中，有些驾驶员会轻踩制动，而有些驾驶员反而会轻踩油门。分析原因为有些主车驾驶员观察到目标车有切入的可能时，不太想让出车道，加速动作会让切入的车辆感到危险，从而放弃切入行为。而有些主车驾驶员为了确保两车之间的安全，会主动制动。这些动作也可能与驾驶熟练度有关。

由图9的曲线趋势可知，主车纵向加速度与RP呈正相关，与Pearson相关性检验结果($r=0.08$)一致。 $P=0.244>0.05$ ，说明两者之间没有显著的相关性。

3.3.3 相对纵向速度

图10中，相对纵向速度小于0的值有18%。由图10可知，数据集中在 $0\sim 10\text{ m/s}$ 之间，表明在切入的过程中，目标车多数以低于主车的速度来切入车道。分析可知，目标车驾驶员在以一个较快速度

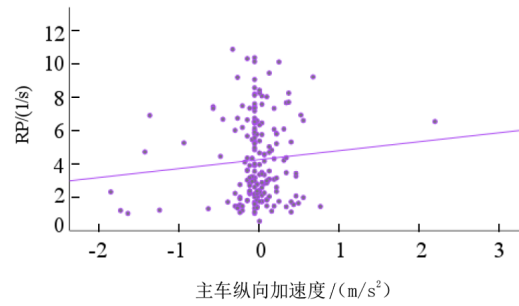


图9 主车纵向加速度 & PR 散点

超过主车，如果这时想切换车道，首先会通过后视镜观察目标车道的车辆，如果主车没有明显的加速动作，目标车通常会采取稍微降速的动作来切入目标车道。

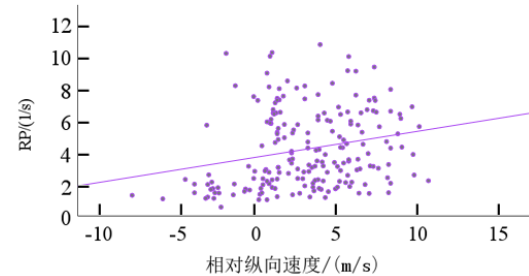


图10 相对纵向速度 & PR 散点

经Pearson相关性检验， $r=0.23$ ，说明两者呈正相关。 $P=0.001<0.05$ ，说明两者有显著的相关性。

3.3.4 两车纵向距离

图11中，可以看出两车纵向距离与RP之间存在负相关，与Pearson相关性检验结果($r=-0.853$)一致。两车纵向距离聚焦在 $15\sim 60\text{ m}$ 之间，说明在高速公路上，车辆行驶速度较快，交通流量也相对较大。为了确保行车安全，驾驶员通常会与前车保持安全距离。此外，由于高速公路上的车道线划分清晰，驾驶员可以更容易地判断前方车辆的位置和动态变化，从而做出更及时的反应。

显著性系数 $P=0.000<0.05$ ，说明纵向距离与RP的相关性是显著的。

3.3.5 两车横向距离

由图12可知，以0点为界，从相同方向切入的值会聚集在一起。不管从哪个方向切入，对RP值没有影响是不明显的，与显著性系数 $P(0.364>0.05)$ 反映的结果是相同的。

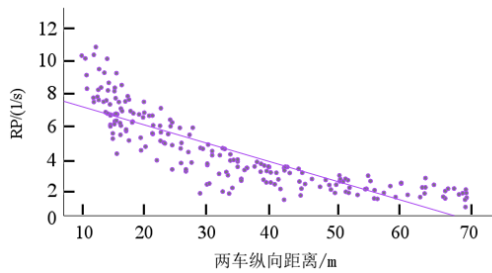


图 11 两车纵向距离 & PR 散点

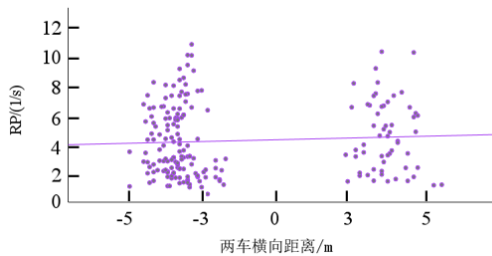


图 12 两车横向距离 & PR 散点

经 Pearson 相关性检验，横向距离与 RP 值的相关性检验结果 $r=0.065$ ，说明两者呈正相关性。

3.3.6 切入时间

从自然驾驶场景中共提取变道切入片段场景 197 个。其中，变道切入时间最小的是 2.16 s，最大的是 15.36 s。

经 Pearson 相关性检验， $r=0.035$ ，说明两者之间呈正相关。 $P=0.626>0.05$ ，说明切入时间与 RP 值的相关性是不显著的。

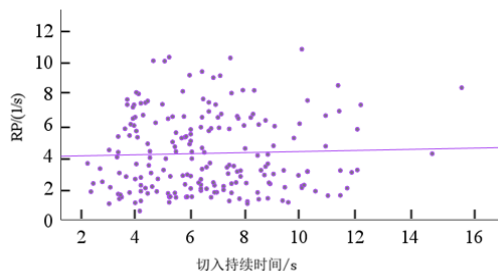


图 13 切入时间与 PR 散点

4 切入场景聚类分析

根据切入场景要素分析结果，车辆类型、转向灯、主车纵向速度、相对纵向距离、相对纵向速度 5 个因素对 RP 值有显著影响，为切入场景的主要影响因素。

聚类算法是一种数据分析技术，通过度量数据点之间的相似性，将数据自动划分为多个不同的组，同一个组内的数据点具有较高的相似性，而不同组之间具有显著的差异。可以客观地对变道切入场景进行分类，减少了人为因素的干扰。

本研究采用 K-Means 聚类算法进行聚类，K-Means 算法能根据预定义的聚类个数，分析样本之间的关联性并进行自动聚类^[16]。对提取出的关键切入场景基于 5 个场景要素进行聚类分析，得到典型变道切入场景。

4.1 数据标准化

在对场景进行聚类处理之前，需对所有数据进行标准化，以消除不同参数的量纲和数量级对结果的干扰。本研究采用数据处理中最常用的 Z-score 标准化方法，对有显著影响的参数进行标准化处理，部分数据处理见表 6。

表 6 标准化后的部分数据展示

| Z-score 主车纵向速度/(m/s) | Z-score 相对纵向距离/m | Z-score 相对纵向速度/(m/s) | Z-score 车辆类型 | Z-score 转向灯 |
|----------------------|------------------|----------------------|--------------|-------------|
| 0.43 | 0.33 | 0.87 | 1.00 | 0.00 |
| 0.63 | 0.12 | 0.49 | 1.00 | 1.00 |
| 0.48 | 1.00 | 0.65 | 0.00 | 1.00 |
| 0.58 | 0.09 | 0.63 | 1.00 | 1.00 |
| 0.70 | 0.10 | 0.70 | 1.00 | 0.00 |
| 0.66 | 0.20 | 0.71 | 1.00 | 1.00 |
| 0.65 | 0.07 | 0.47 | 1.00 | 1.00 |
| 0.62 | 0.24 | 0.78 | 1.00 | 0.00 |
| 0.80 | 0.49 | 0.58 | 1.00 | 0.00 |
| 0.82 | 0.34 | 0.65 | 1.00 | 0.00 |
| 0.48 | 0.15 | 0.63 | 0.00 | 1.00 |
| 0.43 | 0.49 | 0.74 | 0.00 | 0.00 |
| 0.45 | 0.69 | 0.66 | 0.00 | 1.00 |

4.2 确定聚类个数

曹骞等^[17]用 Silhouette 指数来确定聚类的个数，指数取值在 $-1\sim 1$ 之间，取值越接近 1，表明各类别聚类的品质越好；取值越接近 -1 ，表明各类别聚类的品质越低。许健等^[18]通过肘部法确定最

佳的聚类个数，样本聚类的紧密程度通过总误差平方和（SSE）来衡量，其值越小表示聚类样本的关系越紧密。具体步骤为：

1) 将数据集划分为 K 个簇， K 从1开始，逐步增大。

2) 计算每个 K 值下，簇内误差平方和，即所有样本点到其所属簇中心的距离平方和SSE。

$$SSE = \sum(Y_i - \bar{Y})^2。 \quad (3)$$

式中： Y_i 为观测值（真实值）； \bar{Y} 为预测值（模型预测值）。

3) 绘制 K 值与SSE的关系曲线。

4) 根据曲线拐点，选取最优 K 值，即聚类数。

针对5个显著影响因素的采集数据，计算出不同聚类个数 K 值下的SSE值，如图14所示。聚类个数从5开始，数值下降趋于稳定，过度增加 K 值会增加模型的复杂度，导致过拟合，因此，确定聚类个数为5。

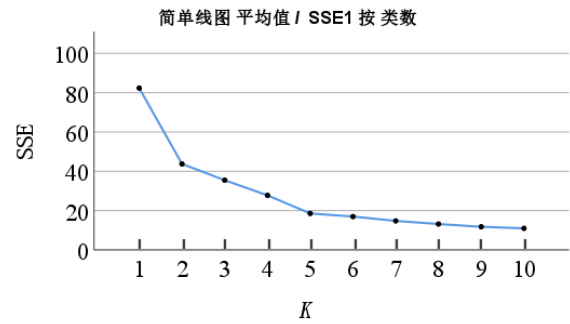


图14 肘部法确定聚类个数

4.3 K-Means 聚类结果

对提取出的所有切入场景进行聚类，聚类个数设置为5，得到5类典型切入场景，具体场景数据见表7。

4.4 聚类结果分析

通过K-Means聚类得到5类典型切入场景，由这5类典型场景可知。

表7 聚类后的典型场景

| 典型场景 | 占比/% | 主车纵向车速/(m/s) | 相对纵向距离/m | 两车相对速度/(m/s) | 车辆类型 | 转向灯 |
|------|------|--------------|----------|--------------|------|-----|
| 1 | 13 | 16.65 | 29.27 | 3.75 | 0 | 0 |
| 2 | 12 | 24.55 | 64.30 | 1.97 | 1 | 0 |
| 3 | 15 | 25.10 | 49.38 | 1.00 | 0 | 1 |
| 4 | 40 | 25.12 | 16.10 | 2.99 | 0 | 0 |
| 5 | 20 | 28.02 | 35.00 | 2.66 | 0 | 0 |

1) 乘用车发生切入行为的次数明显多于商用车，这与实际生活中的现象一致，可以推断，商用车由于特殊的运载属性，为了保证行车过程中的安全性，车速普遍保持在60~80 km/h之间，并在单一车道内保持行驶。

2) 不开启转向灯，直接切入到目标车道的行为较普遍，这与驾驶员的主动意识相关。当驾驶员透过后视镜观察到与相邻车道的后方车辆纵向距离较远时，会更随心地转动方向盘，直接切入相邻车道。

3) 切入行为发生时，主车纵向车速集中在16.65~28.02 m/s之间，推测此时主车车速相对较慢，更容易被相邻车道的其他车辆切入。

4) 相对纵向距离集中在16.1~64.3 m之间，推测这是为了保证切入过程的安全性，有些驾驶员会在距离较小时切入，这与实际生活中遇到的情况相似。

5 基于PreScan构建切入场景库

根据表7中K-Means聚类后5类典型场景，在PreScan软件中进行搭建，步骤如下。

1) 构建道路，并设置道路属性，如车道数、车道宽度、有无路牙等。

2) 放置主车与目标车，目标车辆位于主车的相邻车道后方。在主车车辆上布置TIS传感器，用于探测周围环境信息。按照聚类场景参数，设置车辆行驶策略。传感器模型参数见表8。

表 8 传感器模型参数

| 类型 | 安装位置 | | | 方向 | | | 横向 FOV/(°) |
|----------|------|-----|------|-------|-------|-------|------------|
| | X/m | Y/m | Z/m | P/(°) | Y/(°) | R/(°) | |
| 长距 Radar | 3.65 | 0 | 0.40 | 0 | 0 | 0 | 60.0 |
| 短距 Radar | 3.75 | 0 | 0.50 | 0 | 0 | 0 | 120.0 |
| 摄像头 | 1.82 | 0 | 1.62 | 0 | 0 | 0 | 100.3 |

3) 运行软件, 目标车从主车相邻车道切入到主车车道, 完成整个切入过程。

本研究基于真实自然驾驶数据, 采用 K-Means 聚类算法 (聚类数 $k=5$), 以车速、加速度等作为特征向量, 提取典型驾驶场景。所有用于聚类和场景构建的参数均来源于真实数据, 确保了场景的代表性和真实性。在 PreScan 仿真环境中, 基于聚类结果搭建了对应的仿真场景。仿真场景的逼真度主要取决于仿真软件的功能和场景建模的精度。

仿真场景的逼真度反映了仿真环境在视觉、物理和行为层面与实际驾驶场景的一致性。然而, 与实际驾驶场景相比, 仿真场景在视觉渲染、物理模型和交通参与者行为等方面仍存在一定差异。例如, 仿真中的光照和纹理可能不够真实, 车辆动力学模型可能过于理想化, 交通参与者的行为可能缺乏随机性和复杂性。

为提升仿真场景的逼真度, 本研究从以下 2 个方面调整参数。

1) 仿真软件参数: 通过提高渲染质量 (如分辨率、光照效果) 和优化物理引擎参数 (如摩擦因数、重力加速度), 增强仿真的视觉和物理逼真度。

2) 场景建模参数: 使用高精度地图数据构建道路网络, 并增加动态物体 (如车辆、行人) 的多样性和随机性, 以提高场景的真实感。

6 结论

本文通过定义的切入场景提取策略, 完成对采集的自然驾驶场景数据的切入场景提取, 将场景参数分为单元素要素和连续型要素, 并引入感知危险系数参数, 进行显著性分析, 识别出对切入行为有显著影响的 5 类因素。随后, 运用 K-Means 聚类方法对 5 类显著影响因素进行聚类分析, 再通过肘部法确定场景的具体分类数。基于 PreScan 软件, 针

对确定分类数下的具体典型场景数据, 搭建了虚拟仿真场景。主要结论为如下 4 点。

1) 影响切入场景的关键参数包括: 切入车辆类型、转向灯状态、主车纵向速度、相对纵向速度以及两车纵向距离。

2) 通过 K-Means 聚类方法, 将所有场景进行分类, 采用肘部法确定最优聚类数为 5。

3) 基于 PreScan 软件搭建虚拟场景, 并依据聚类结果赋值, 从而在仿真环境中高效、精确地实现变道切入行为。

4) 本文方法避免了工程师主观设定测试场景参数, 确保了仿真场景的参数具有数据支持, 从而提高了测试的一致性和可靠性。

切入场景库的构建方法能更客观、科学地反映变道典型的切入场景, 为自动驾驶系统的测试和评估提供更加稳健的支持。

本研究主要聚焦于切入场景库的构建方法, 尚未引入自动驾驶控制算法, 也未进行完整的在环测试。因此, 存在以下 2 点局限性:

1) 无法验证场景库对自动驾驶系统性能的实际影响;

2) 缺乏对场景泛化能力的评估。

为弥补这些不足, 后续研究将着重从以下 3 个方向改进。

1) 引入算法: 结合自动驾驶控制算法, 验证通过聚类获得的典型场景对系统决策和控制的影响, 并评估场景库在不同控制算法下的适用性。

2) 开展在环测试: 通过硬件在环 (HIL) 或软件在环 (SIL) 测试平台, 结合真实驾驶数据与仿真数据, 验证场景库的实用性与可靠性。

3) 研究场景泛化能力: 扩充场景库, 增加包

括极端天气、复杂交通流等多样化的场景，评估其不同驾驶环境下的泛化能力。

这些改进将为后续研究提供重要参考，并进一步提升场景库的实用价值。

参考文献 (References)

- [1] YURTSEVER E, LAMBERT J, CARBALLO A, et al. A Survey of Autonomous Driving: Common Practices and Emerging Technologies[J]. IEEE Access, 2020, 8: 58443–58469.
- [2] 赵祥模国家重点研发计划(2021YFB2501200)团队. 自动驾驶测试与评价技术研究进展[J]. 交通运输工程学报, 2023, 23(6): 10–77.
ZHAO Xiangmo's Team Supported by the National Key Research and Development Program of China 2021YFB2501200). Research Progress in Testing and Evaluation Technologies for Autonomous Driving [J]. Journal of Traffic and Transportation Engineering, 2023, 23(6): 10–77. (in Chinese)
- [3] 中华人民共和国公安部交通管理局. 2015 中华人民共和国道路交通事故统计年报[R]. 北京: 中华人民共和国公安部交通管理局, 2016.
Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China. 2015 Road Traffic Accident Statistical Yearbook of the People's Republic of China [R]. Beijing: Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China, 2016. (in Chinese)
- [4] HOU Y, EDARA P, SUN C. Situation Assessment and Decision Making for Lane Change Assistance Using Ensemble Learning Methods [J]. Expert Systems with Applications, 2015, 42(8): 3875–3882.
- [5] 李彦睿. 基于实车试验数据的自动驾驶测试场景构建技术[D]. 重庆: 重庆大学, 2021.
LI Yanrui. Autonomous Driving Test Scenario Construction Technology Based on Real Vehicle Test Data [D]. Chongqing: Chongqing University, 2021. (in Chinese)
- [6] 邓伟文, 李江坤, 任秉韬, 等. 面向自动驾驶的仿真场景自动生成方法综述[J]. 中国公路学报, 2022, 35(1): 316–333.
DENG Weiwen, LI Jiangkun, REN Bingtao, et al. A Survey on Automatic Simulation Scenario Generation Methods for Autonomous Driving [J]. China Journal of Highway and Transport, 2022, 35(1): 316–333. (in Chinese)
- [7] 徐家正, 陈星, 王景升, 等. 自动驾驶合规性虚拟测试场景构建方法研究进展[J]. 江汉大学学报(自然科学版), 2024, 52(2): 87–96.
XU Jiazheng, CHEN Xing, WANG Jingsheng, et al. Research Progress on Automatic Driving Compliance-Related Virtual Test Scene Construction Methods [J]. Journal of Jiangnan University (Natural Science Edition), 2024, 52(2): 87–96. (in Chinese)
- [8] 姚明菊, 李志远, 羊雪玲, 等. 基于蒙特卡洛模拟的虚拟测试场景生成方法[J]. 计算机仿真, 2024, 41(2): 473–476, 481.
YAO Mingju, LI Zhiyuan, YANG Xuelling, et al. A Method for Generating Virtual Test Scenario Based on Monte Carlo Simulation [J]. Computer Simulation, 2024, 41(2): 473–476, 481. (in Chinese)
- [9] 刘志强, 马进. 基于自然驾驶数据的驾驶换道风格分析与辨识[J]. 电子设计工程, 2024, 32(9): 20–26.
LIU Zhiqiang, MA Jin. Analysis and Identification of Driving Lane Changing Style Based on Natural Driving Data [J]. Electronic Design Engineering, 2024, 32(9): 20–26. (in Chinese)
- [10] 丁洁云, 党睿娜, 王建强, 等. 驾驶人换道决策分析及意图识别算法设计[J]. 清华大学学报(自然科学版), 2015, 55(7): 769–774.
DING Jieyun, DANG Ruina, WANG Jianqiang, et al. Driver Lane Change Decision Analysis and Intention Recognition Algorithm [J]. Journal of Tsinghua University (Science and Technology), 2015, 55(7): 769–774. (in Chinese)
- [11] 郭柏苍, 雒国风, 金立生, 等. 面向自动驾驶虚拟测试的变道切入场景库构建方法[J]. 吉林大学学报(工学版), 2023, 53(11): 3130–3140.
GUO Baicang, LUO Guofeng, JIN Lisheng, et al. Construction Method of Cut-in Scenario Library for Automatic Driving Virtual Tests [J]. Journal of Jilin University (Engineering and Technology Edition), 2023, 53(11): 3130–3140. (in Chinese)
- [12] 宋越, 曾杰, 胡雄, 等. 变道切入场景下 ADAS 系统测试与评价研究[J]. 汽车工程学报, 2024, 14(3): 363–375.
SONG Yue, ZENG Jie, HU Xiong, et al. Research on Testing and Evaluation of ADAS Systems in Lane Change Cut-in Scenarios [J]. Chinese Journal of Automotive Engineering, 2024, 14(3): 363–375. (in Chinese)
- [13] 王雪松, 杨敏明. 基于自然驾驶数据的变道切入行为分析[J]. 同济大学学报(自然科学版), 2018, 46(8): 1057–

1063.
WANG Xuesong, YANG Minming. Cut-in Behavior Analyses Based on Naturalistic Driving Data [J]. Journal of Tongji University (Natural Science), 2018, 46(8): 1057-1063. (in Chinese)
- [14] OLSEN E C B, LEE S E, WIERWILLE W W, et al. Analysis of Distribution, Frequency, and Duration of Naturalistic Lane Changes [J]. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2002, 46(22): 1789-1793.
- [15] 戴思阳. 驾驶员车道变换时转向灯开启特性研究 [J]. 陕西交通职业技术学院学报, 2015(1): 12-15.
DAI Siyang. On the Characteristics of Driver's Turn Signal When Changing Lanes [J]. Journal of Shaanxi College of Communication & Technology, 2015(1): 12-15. (in Chinese)
- [16] 张益敏, 吴小兰, 张奕河, 等. 基于聚类算法的汽车行驶工况构建方法研究 [J]. 计算机应用文摘, 2024, 40(8): 82-84.
ZHANG Yimin, WU Xiaolan, ZHANG Yihe, et al. Research on the Construction Method of Automobile Driving Conditions Based on Clustering Algorithm [J]. Computer Application Abstracts, 2024, 40(8): 82-84. (in Chinese)
- [17] 曹骞, 侯乐福, 贾咸坤, 等. 基于混合聚类法的典型行驶工况构建 [J]. 重型汽车, 2024(3): 29-30.
CAO Qian, HOU Lefu, JIA Xiankun, et al. Construction of Typical Driving Conditions Based on Hybrid Clustering Method [J]. Heavy Truck, 2024(3): 29-30. (in Chinese)
- [18] 许健, 王琪, 唐海荣, 等. 基于改进 K-means 算法的指标阈值告警方法研究 [J]. 现代计算机, 2022, 28(20): 31-36.
XU Jian, WANG Qi, TANG Hairong, et al. Research on Indicator Threshold Warning Method Based on Improved K-means Algorithm [J]. Modern Computer, 2022, 28(20): 31-36. (in Chinese)

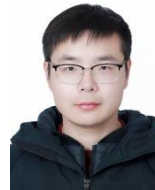
作者简介



宋华 (1976-), 男, 福建建瓯人, 学士, 工程师, 主要研究方向为智能驾驶系统开发及测试。

E-mail: songhua@jac.com.cn

通信作者



吴贤静 (1990-), 男, 安徽合肥人, 学士, 工程师, 主要研究方向为自动驾驶仿真测试。

E-mail: wxj1598@jac.com.cn