行为安全概念及基于其的自动驾驶能力认可方法

陈 龙[1]，高鲁涛[1]，徐晓庆[2]，曹建永[3]，李楚照[4]

（1. 北京华为数字技术有限公司，北京 100085；2. 北京镝石数据科技有限公司，北京 100176；
3. 上海机动车检测认证技术研究中心有限公司，上海 201805；4. 清华大学，北京 100084）

摘 要：针对自动驾驶接受准则难以统一并量化的问题，通过综述现有相关法规、标准和评价方法最优实践，凝练5种安全概念及相互关联关系。着重阐释了以合理可预见且可避免为核心的行为安全概念及研究现状。结合场景数据统计和驾驶员紧急反应机制，提出了合理可预见和可避免的量化研究方法框架。结合事故和实践经验，给出了行为安全在自动驾驶评价中的使用方法，及基于其概念的覆盖事前事后的闭环认证认可流程。可为主管机构、行业第三方、研发企业围绕行为安全建立相关自动驾驶研发、测试流程提供参考。

关键词：合理可预见；驾驶员紧急反应模型；自动驾驶认证认可；行为安全

中图分类号：U471.15 **文献标志码：**A **DOI：**10.3969/j.issn.2095–1469.2024.05.04

Concept and Capability Approval Method of Behavioral Safety for Automated Driving Systems

CHEN Long[1]，GAO Lutao[1]，XU Xiaoqing[2]，CAO Jianyong[3]，LI Chuzhao[4]

（1. Beijing Huawei Digital Technologies Co.，Ltd.，Beijing 100085，China；
2. Beijing Dishi Data Technology Co.，Ltd.，Beijing 100176，China；
3. Shanghai Motor Vehicle Inspection Certification & Technology Innovation Center Co.，Ltd.，Shanghai 201805，China；
4. Tsinghua University，Beijing 100084，China）

Abstract: Aiming at acceptance criteria for automated driving, this paper reviews the best practices in relevant regulations, standards, and evaluation methods, identifying five safety concepts and their interrelationships. The paper focuses on the concept and research status of behavioral safety, centered around "reasonably foreseeable and preventable" behaviors. By combining scenario data statistics with the driver's emergency response mechanism, a quantitative research framework is proposed for reasonably foreseeable and preventable situations. Finally, combining traffic accident data and practical experience, the paper provides a method for using behavioral safety in autonomous driving evaluation, along with a closed-loop certification and approval process based on this concept. The research in this paper serves as a reference for authorities, third parties, and R&D companies to establish relevant R&D, testing, and processes centered around behavioral safety.

Keywords: reasonably foreseeable; careful and competent driver model; type approval of automated driving system; behavior safety

收稿日期：2023–04–28 **改稿日期：**2023–07–24 **网络首发日期：**2024–05–30

参考文献引用格式：

陈龙，高鲁涛，徐晓庆，等. 行为安全概念及基于其的自动驾驶能力认可方法[J]. 汽车工程学报，2024，14(5)：781–790.

CHEN Long，GAO Lutao，XU Xiaoqing，et al. Concept and Capability Approval Method of Behavioral Safety for Automated Driving Systems[J]. Chinese Journal of Automotive Engineering，2024，14(5)：781–790.(in Chinese)

在自动驾驶技术发展之初，“自动驾驶比人安全”的论述主要是站在自动驾驶解决了人类缺陷引发事故的角度上进行分析，但这论据并不全面。诚然，人类驾驶员有诸多缺陷，但其优势也是现有自动驾驶算法所无法比拟的，比如对行人、其他驾驶员面部表情、手势的理解，这是自动驾驶短期或者永远也达不到的能力，而正是这样的能力使车辆在混杂的路口穿梭自如，避免不必要的交互风险造成交通事故。因此，行业不应仅对比露出在水面的冰山（事故），更应关注那些被人合理处理掉的沉在水面下的大量冰山（安全风险或者正常行驶）。

自动驾驶并非一个传统意义上可以无限试错的APP，其安全基线就是要能替代人类处理路上千奇百怪却又常见的行驶风险[1]。自动驾驶不是可以“任性甩锅”的辅助驾驶功能，其开发商将从交通运输工具提供商变为直接交通事故责任者以及整体交通协调的责任者，既要考虑自车安全，又要兼顾整体交通的安全和效率。这种改变使原有的汽车和驾驶员管理相关部门职能交织在一起，因此，自动驾驶需要适应各方管理要求。

在美国交通部发布的自动驾驶安全愿景2.0[2]中，就提到了13种企业需要声明的安全要素，分别是：系统安全、设计运行条件（Operational Design Domain，ODD）、目标和事件检测与响应（Object and Event Detection and Response，OEDR）、接管（最小风险状态）、验证方法、人机交互（Human-Machine Interaction，HMI）、车辆网络安全、耐撞性、事故后行为、数据记录、消费者教育与培训及对国家、州和地方法规的遵守。美国政府鼓励自动驾驶公司按照这13方面阐述企业实际情况。虽然目前NHTSA网站中已有多达28家企业提交了安全报告，但报告内容相似度较高，内容比较空洞。

汽车行业新技术的发展与应用往往是基于综合各供应商的最优实践开展新增模块的应用和标准的制定。驾驶辅助由于只是在人的驾驶能力出现不足时予以补充，其标准制定思路仍是以实践为主导，仅需测试机构和研发机构权衡，利用简单工况订立门槛驱逐劣币。但驾驶能力的最优实践并非某家企业的驾驶技术水平，而是由低概率故障水平的车+有概率性错误的人类组成的驾驶水平。“最优实践决定标准”这条路貌似很难走通，自动驾驶安全要求不再是个研发技术水平问题，而是从社会接受度、法规、权责等角度去综合讨论得出对企业开发管理流程和质量的要求。

一种常见的理论是自动驾驶汽车安全行驶的里程越多，它就越安全[3]。但事实上，飞机的自动辅助驾驶功能在简单巡航场景上安全行驶数十亿公里，也不能代表它能成功处理1次安全的起降过程；安全行驶里程数需要在某一特定场景复杂度下，才可以成为界定自动驾驶安全能力的指标。

另一种常见的理论是自动驾驶车辆驾驶员主动干预次数越少，它安全驾驶的能力就越强。但是这种理论与基于里程的安全理论一样，也与区域场景的复杂度（不同的位置、不同的时间等）有很大关系。选择简单场景的路段，车辆接管次数就少；反之接管次数就多。

基于交通法规的安全理论认为：只要把所有的道路交通规则数字化，然后严格保证按照交通规则[4]驾驶，就可以达到安全驾驶的目标。但事实上，在某些情况下，即使自动驾驶车辆严格按照道路交通规则行驶，交通事故也难以完全避免。

事故避免理论的支持者认为：自动驾驶车辆在任何时候任何场景下都应该避免任意类型的碰撞事故。但是事实上，这种理论过于理想化。在有些情况下，即使是经验丰富的人类司机也难以处理这种场景。

覆盖率驱动验证是通过自动驾驶对场景参数范围的避免事故覆盖程度衡量其安全性。这种方式适合在研发过程中使用，方便观察自动驾驶的能力边界。但这种方法无法确定安全基线。

预期功能安全[5]标准中将场景分为4种象限。按照标准的要求，要将已知不安全的象限控制到足够小。其给出了以危险事件的平均运行里程来界定自动驾驶安全性能力。

UNECE R157[6]所采用的评价标准是自动驾驶系统在其运行过程中不能带来可预见可避免的风险

（事故），并且满足属地国的交通法规（除需要违规才可避免发生事故的情况），即综合了交通法规、事故避免、覆盖率3种理论，通过人类驾驶员模型来定义“可避免”的风险边界。但如何定义“合理可预见”仍没有定论。

针对自动驾驶接受准则难以统一并量化的问题，本文通过综述现有相关法规、标准和评价方法最优实践，综合社会接受度、法规、权责等角度，凝炼出5种安全概念及相互关联关系。着重阐释了以合理可预见且可避免为核心的行为安全概念及研究现状。结合场景数据统计和驾驶员紧急反应机制，提出了合理可预见和可避免的量化研究方法框架。最后，结合事故和实践经验，给出了行为安全在自动驾驶评价中的使用方法，及基于其概念的覆盖事前事后的主管机构-行业第三方-研发企业间闭环的认证认可流程。

1　自动驾驶安全概念

本节从事故风险展开，探讨各安全概念内涵及定义。

1.1　安全概念定义

交通法规符合度、行为安全、ODD合理性、人机交互安全、残留风险认可等安全概念均与事故风险直接相关。这些概念在事故风险方面的关系如图1所示。

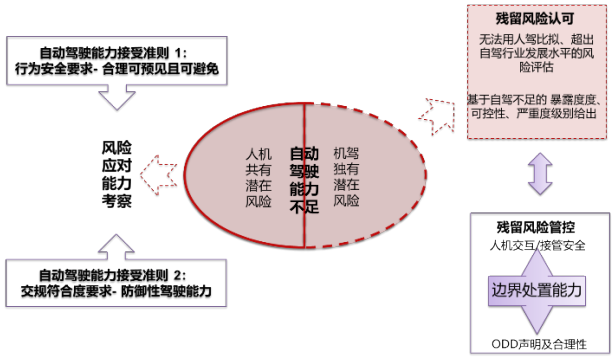

图1　各安全概念的关联关系

首先讨论图1中人类驾驶缺陷和自动驾驶（车）缺陷重合部分。人类驾驶员虽然在获得驾驶证的过程中经历过系统的道路交通安全相关法规规范培训学习，但通过与客观交通现实不同时间磨合与理解，形成了多种行为范式理解和安全风险处理逻辑，如本车道内超越同向运动行人。人类并非无问题的机器，会综合一定风险、文明要求、有意和无意失误等方面因素实现驾驶操作。这些要素耦合形成了由于人类驾驶缺陷造成的事故。自动驾驶功能（车）作为现有交通现实的少数分子，其应能适应人类驾驶员及弱势道路使用者组成的行为范式理解和安全风险处理逻辑，但应对仍有极限，这些客观极限条件仍可导致事故。对于缺陷重合部分，目前已经形成较统一的自动驾驶能力接受准则，即行为安全和交通法规符合度。

行为安全要求自动驾驶功能在无自动驾驶特有缺陷的情形下，在合理可预见的场景下达到一般驾驶经验且注意力集中的驾驶员避免事故的水平。

交通法规符合度要求自动驾驶功能在无自动驾驶特有缺陷的情形下，遵从其声明可运行区域及边界内交通法规中的明确行为规范，做到预防安全风险。

对于已知的自动驾驶（车）特有缺陷，我们希望企业能参照功能安全和预期功能安全进行安全管理和风险评估，并将无法处理的风险暴露给行业，由行业统一进行风险定级。若企业无法达到行业订立的级别要求，则需将相关缺陷导致风险通过企业声明的ODD及相关接管功能合理剥离。针对这一部分缺陷尚未统一，本文推荐安全概念为缺陷定级和ODD设定合理性。

残余风险认可要求对已知各项无法完全克服的功能不足进行分析、评估、确认，遵从其公认的暴露度（Exposure）、可控性（Controllability）、严重度（Severity）级别要求。

ODD合理性要求除去严重故障外禁止因直接碰撞风险发出接管请求，合理设置接管时间，避免人机共驾环节的存在，遵从交通法规、道路、环境等客观条件。

在接管和人类驾驶员误用滥用时，自动驾驶应处理好与人类驾驶员交互功能，合理分摊事故风险及约束其他潜在风险，即人机交互安全。

人机交互安全要求自动驾驶功能可通过接管将无法应对的潜在风险，在合理考虑人接管的缺陷特性，安全交接给人类驾驶员，自动驾驶功能也应能约束驾驶员及乘客不合理的行为，避免由此造成潜

在风险。另外，自动驾驶行为应符合人类行为范式，易于驾驶员及其他的道路使用者理解其行为模式。

通过上述5个安全概念的定义和内涵论述可知，虽然自动驾驶的安全要求复杂，但企业也能较清晰拆解评估管控自动驾驶各类风险点，监管机构也可简化认证认可难题。

为了将自动驾驶与人类驾驶水平拉齐，针对常常被忽视的“水下冰山”（人类能轻松应对的情景），我们也需要定义自动驾驶能力可接受的准则。目前行业共识是自动驾驶需要避免设计运行范围内或者自动驾驶激活状态下合理可预见且可避免的事故发生。这就与剩余风险的接受准则有明显的差异：风险量化方式是概率，能力量化方式是参数范围。

1.2 行为安全概念介绍

本节围绕行为安全，梳理与其相关的所有要求，得出其核心内涵“合理可预见及可避免”，进一步展开对其的综述及阐释。

1.2.1 行为安全概念汇总

本文将现有相关法规和标准（如文献[7]、UNECE R157、文献[8]）中关于行为安全的内容全部提取出来。通过归类整理，行为安全共分为4个环节：支撑能力要求、行为模式要求、性能要求、验证方法。这4部分的关系及细化要求如图2所示。

感知与决策在自动驾驶中是强耦合的。在行为安全概念的要求下，感知作为支撑能力，应能支持行为模式的实现，且不应成为性能达成的障碍。行为模式应重点关注与目标物的交互行为，做到逻辑架构可解释且符合其他道路使用者（Other Road User，ORU）的预期。此外，自动驾驶功能不得因短时间内可预期交互风险而失活，即自动驾驶应承担其与目标物交互的全部行为责任，不得发出接管请求，造成人机混驾的伦理风险。从行为安全角度看，对行为模式（决策逻辑）要求较模糊，在企业产品设计中，应考虑如何将“符合其他道路使用者的预期”具象化，比如参照交通法规或者机动车驾驶人安全文明操作规范，这又跟交通法规符合度相耦合，本节不做深入探讨。不同企业具体产品行为

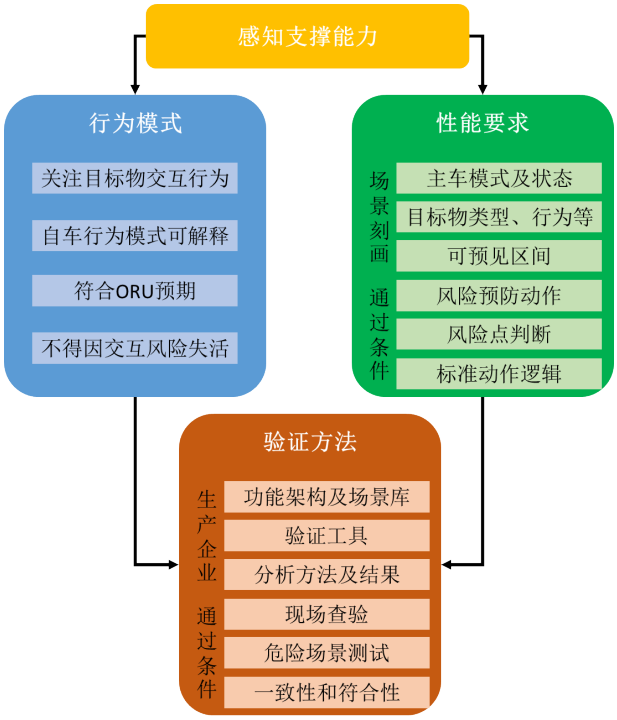

图2 行为安全概念在不同环节的落实要求及方法

逻辑会有所差异，这也导致难以形成类似于驾驶辅助系统可通过标准化关键参数实现性能要求和验证方法的一致性。

由于上述问题，目前对于行为安全，考虑的性能要求是避免合理可预见且可避免的事故，可拆分为两方面：一是从自车行为模式及内部状态和目标物行为模式耦合的角度，构建综合预判原则、行为全覆盖、可预见参数范围的交互场景；二是对于危险难以处理的场景，需要建立人类驾驶员行为模型作为通过参考条件。人类驾驶员行为模型的构建方法需要标准化，比如风险预防动作模式、风险出现点定义方法、风险应对标准动作方式。

验证方法分为生产（研发）企业内部验证和主管机构认证两类。企业为证明对行为安全概念的遵从需要做到以下4点：

（1）给出功能架构图，并提供材料证明其满足其他道路使用者的行为预期；

（2）给出场景覆盖的逻辑方法，并展示出基于此建设的场景库；

（3）需介绍其验证工具，并给出工具置信度证明材料；

（4）需给出分析方法，并证明结果满足通过条件。

主管机构认证方法主要是通过现场审查、危险场景测试等重点评估企业提供材料的一致性和对行为安全要求的符合性。

1.2.2　合理可预见及可避免阐释

（1）合理可预见及可避免的内涵

从行为安全概念定义上，可以将其拆解成3个层面：适用条件、测试范围、通过要求。

适用条件：在无自动驾驶特有缺陷的情形下。如，由于轮胎制动力不足（已达到标准要求），仍无法避免碰撞，那么，这种情形对于人类驾驶员和自动驾驶均是车辆结构所限的行业认可的缺陷，并非自动驾驶特有的缺陷。例如，感知算法的漏检属于自动驾驶的自有缺陷，不在行为安全考虑范围内。

测试范围：在合理可预见的场景下。自动驾驶是以代码程序呈现，仅能实现对一定参数范围内的场景应对。同样，人类驾驶员也是基于已有的行为范式去分析实时遇到的场景，先进行交互目标物行为分类，结合其潜在意图（行为范式预判），进行目标轨迹预测。行业及企业需要枚举交互行为范式，达到行为范式的高覆盖度，并定义每种行为范式的标准参数化表达方式，通过数据量化参数的大概率取值空间。

通过要求：达到一般驾驶经验且注意力集中的驾驶员避免事故的水平。这项要求自动驾驶系统能参照人类驾驶员水平处理风险场景。为了实现这一要求的量化，我们需要将人类处理风险场景的动作进行标准化。在标准化过程中，需要考虑人类驾驶的如下3个特点。

1）防御性驾驶能力：当代社会防御性驾驶能力已经是一般驾驶员的普遍要求。人类驾驶员可对周围环境进行提前10 s的意图预判，比如道路前方公交车站有静止公交车开启左转灯，那么对于该情景的预判是，公交车将要起步向左变道，社会车辆人类驾驶员的正确动作是在公交车起步前，及时制动让其变道，或者自车向左变道。

2）能灵活遵守交通法规：在交通法规中有层级关系，为避免事故发生或为特殊车辆让行的情况下，人类驾驶员可以突破交通法规其他条款要求。

3）可采取多种规避风险的手段：在UNECE R157中，驾驶员对于切入场景的应对措施为制动跟随，但真实人类驾驶员规避手段还有鸣笛、灯语、变道等，这些手段均需要进行建模评价。

（2）合理可预见的研究现状

合理可预见这是对“水面下”所有情形的语义级和参数级的量化。这就引出了基于场景的自动驾驶能力评价，参考标准是ISO 34502[9]。

该标准针对语义级量化主要做了两方面贡献。其将功能场景分为3类：交通干扰相关场景、感知相关场景、车辆控制相关场景。交通干扰相关场景和感知相关场景探讨的重点都是目标物交互的事故，而车辆控制关注于单车事故。由于单车事故产生及预防机理过于复杂，本文不做探讨。

针对交通干扰场景，标准通过标准化障碍物类别、位置、行为、动作、道路特征结构，进行组合枚举，筛选出有交互风险的场景类别。感知盲区场景对于自动驾驶和人类驾驶员并无差异，均是由于遮挡导致目标物被较晚识别。这类场景可以和交通干扰场景合并处理。

通过对道路实际情形的采集，统计得到逻辑场景的参数分布的统计值，通过“小概率事件”概念找到合理可预见的参数范围。在ISO 34502的参考文献14中，采用了广泛意义的切入场景为例介绍了两参数联合分布区间统计方法，给出3倍标准差和5倍标准差作为参考合理可预见的量化[10]。

企业应用该标准时，需要进一步细化功能场景，避免功能场景宽泛，导致参数范围过大，造成自动驾驶能力的过约束。从实际企业应用实践来看，路上的切入行为可以进一步细分，如慢速切入、快速切入、慢速切入后制动、切入中止等行为。通过对每一种危险交互行为进行参数边界刻画，可以降低对自动驾驶能力的过约束程度，并降低测试验证的成本。

（3）可避免的研究现状

可避免的模型在UNECE R157中共提到3种：基于场景及控制关键参数定义的模型、基于紧急反应特性的驾驶员模型、基于模糊控制的驾驶员模型。

基于场景及控制关键参数定义的模型可参考UNECE R157的01版5.2.4和5.2.5条款。该条款约

束了自动驾驶系统对静止目标物（静止车辆、其他道路使用者、阻断的车道）、前车制动、邻车切入、无遮挡的前方行人横穿4类场景的能力要求。针对静止目标物，应能在接触目标前，达到刹停状态，避免与之碰撞；针对前车制动，应避免与全力制动的前车碰撞；针对邻车切入，应避免某一TTC和相对速度关系假设的条件下的碰撞；针对无遮挡的前方行人横穿，应避免某一行人横穿速度和预设碰撞点条件下的碰撞。

至于没有在5.2.4和5.2.5条款描述的场景通过条件，则参考基于紧急反应特性的驾驶员模型和基于模糊控制的驾驶员模型进行设定。

ALKS的交互场景被分为可避免和不可避免场景。两者的阈值是基于仿真的注意力集中的一般驾驶员的反应设定的。由于UNECE R157的00版法规，设定的运行速度为60 km/h以下，此条件下，人类驾驶员的避撞能力主要通过制动行为表征。故在其附录中，提出了可应用于邻车切入、前车切出、前车制动3个逻辑场景（非5.2.4和5.2.5条款的用例级场景）的标准制动模型。

驾驶员模型被分为3段：感知、决策、执行。

这个模型需要标定的参数有风险感知出现点、感知时间、决策时间、最大制动减速度、制动效能提升时间。总体方法清晰明了，但仍有很多难以确定和标准化的点。

关于风险感知出现点，ALKS给出了3种不同场景的判定方式，但方式并不统一，也不符合一般人驾驶认知。以邻车切入为例，出现点为邻车横向偏移量超过自然驾驶居中行驶的横向摆动中位数。那么如果邻车已经提前开启了转向灯，是否出现点应为邻车开始有横向位置变化的时间点。前车切出的出现点仍是摆动量算，但这个场景的真正风险在于前前车慢速行驶，因此，这个时间点应与前前车的出现有关，而与前车的切出动作无直接关系。

目前，ALKS针对不同场景的风险感知时间和决策时间均是一致的，这也不符合一般人驾驶认知，应与风险预判有关。还是以打转向灯为例，相同切入场景参数设定，一种邻车未打灯，一种提前打灯，从邻车切入动作开始到驾驶员踩下制动踏板，那么驾驶员在两种情形下的反应时间很明显会不同。如果切入车辆为大型车辆，那么驾驶员可能在未发生明显动作时，已经松开加速踏板，反应时间更短。

考虑到实际评价应用，该模型只给出了驾驶员踩踏板的逻辑。从踏板解析到施加控制，每家企业的策略并不一致，这样模型并不能直接求解对应时刻车辆的运行状态，无法唯一的量化需要避免的场景参数。同样也存在制动实际效能的解析问题，这与路面附着系数、轮胎温度、制动建压时间有关，并不只取决于制动踏板的开度。

在基于模糊控制的驾驶员模型中，假设驾驶员能预料碰撞风险并施加相应的制动。模型考虑3种反应：横向安全检测、纵向安全检测、执行动作。这个模型相比注意力集中的一般驾驶员的反应模型，更贴近真实驾驶员思考逻辑。但具体的量化过程过于复杂，所用数学式达10余个，难以获得通俗理解，数学式中的数据更多并无合理量化依据。

本文将借鉴注意力集中的一般驾驶员的反应模型和ISO 34502的交通干扰场景参数刻画，阐述“合理可预见”和“可避免”的获取及建模过程。

2 合理可预见及可避免的获取及建模过程

2.1 合理可预见场景获取过程

（1）行为安全场景枚举：从自车行为模式及内部状态和目标物行为模式耦合的角度，构建综合预判原则、行为全覆盖的交互场景。

（2）功能场景筛选：依据系统运行设计条件，筛选对自动驾驶系统行驶风险最高或最常见的场景作为功能场景。应避免过于保守地制定功能场景。

（3）逻辑场景定义：需明确变量之间的因果关系，确定能描述场景的独立变量并对此进行清晰描述，并在参数范围定义时，应保全参数所有可能空间，避免出现单边分布的情况。

（4）数据获取：为了获取更真实的行驶里程数据，综合使用有较高精度的车端自然驾驶数据、无人机航拍数据、路测摄像头数据。这些采集数据应能满足场景定义的参数在范围、精度方面的需求，尽可能全面覆盖运行自动驾驶功能的国家各地域的广泛驾驶特征。

（5）场景提取：在对功能场景进行参数化后，选择合理的参数对逻辑场景进行简洁有效的定义，并清晰描述逻辑场景起点、终点及其运动过程，避免出现参数间存在相关、加速减速过程描述不清楚等情况。

（6）联合参数特征统计：首先构建一维参数分布，分析参数的整体分布情况，确认场景中行为模式的统一性。进而进行二维联合分布构建，分析不同参数的耦合关系与相关性。为保全参数所有可能空间，剔除掉没有相关性的二维联合分布约束，并通过统计学方法对空间进行描述，后续依据需求进行筛选。

（7）参数空间确定：考虑自动驾驶系统运行设计条件与道路交通安全法规，依据小概率事件原则，保证场景参数覆盖$\mu\pm N\sigma$的范围，同时结合专家评审意见，最终划定参数范围。

2.2　紧急驾驶员反应模型建模过程

（1）将人类驾驶员的标准操作分为2种动作和2个阶段。2种动作为转向动作和松加速踏板-迁移-制动；2个阶段为防御性阶段和应激性阶段。

（2）根据道路交通法和安全文明驾驶规范及驾驶经验，梳理潜在风险点，有以下6种情况。

a）有减速义务的路段：隧道入口、急弯路、施工路段、坡道顶端等影响安全视距、减速车道、匝道、下坡路段。

b）有注意义务的路段：注意落石、注意横风、易滑标志、注意野生动物、路面高凸等标志。

c）有静态盲区的情况，如超越相邻车道静止车辆时，前方存在盲区，可能出现行人横穿。

d）旁车道车辆存在制动动作。

e）目标车为大型车辆。

f）目标车在动作前提前正确开启了对应的警示信号。

（3）基于本节（1）的区分和（2）潜在风险点的分类，可以针对差异化的场景，选取适合的动作组合模式，如下列2种：

a）对于无潜在风险的场景，动作模式简化为真实风险点-紧急动作模型；

b）对于有潜在风险的场景，动作模式简化为潜在风险点-滑行-真实风险点-紧急动作模型（短时间反应）。

（4）根据场景中真实风险点的诱因，确定真实风险点的取值方法有下列3种。

a）前车切出场景的风险点：以自车驾驶员能看清前车遮挡障碍物为标志。

b）小型汽车切入场景的风险点：切入车辆偏离车道中心线超过居中行驶情形下的50%分位累计概率的值为标志。

c）大型汽车切入场景的风险点：切入车辆开始发生横向偏移为标志。

（5）定义紧急反应各阶段的车辆状态的模型形式，如图3所示。

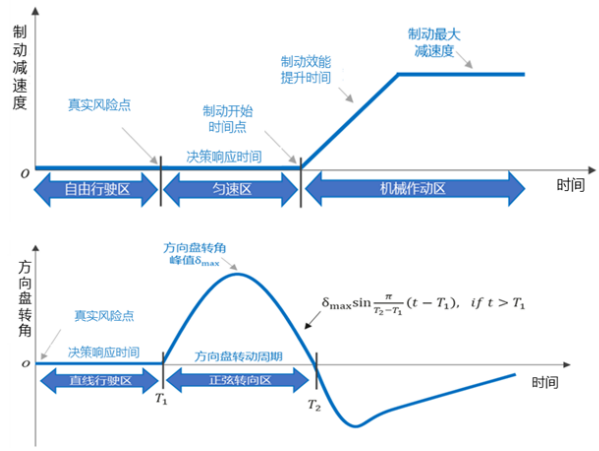

图3　紧急驾驶员反应模型及量化需求

a）将制动模型分为3个区：自由行驶区、匀速区、机械作动区。在自由行驶区，驾驶员按照驾驶需求进行驾驶，这个初始状态仍需进一步探究；在风险出现时间点后，进入匀速行驶区，驾驶员进行风险评估、决策反应、脚部迁移、开始踩踏板；在机械作动区，车辆产生明显的制动效果，通过制动效能快速提升后，达到最大。

b）将转向模型分为2个区：直线行驶区和正弦转向区（参考文献［11］，进行行为简化）。在直线行驶区，驾驶员按照驾驶需求进行驾驶，这个初始状态仍需进一步探究；在风险出现时间点后，仍处于直线行驶状态，驾驶员进行风险评估、决策反

应、手部开始握紧并转动方向盘；在正弦转向区，方向盘以正弦方式进行转动。

（6）梳理模型标定参数需求。

a）在制动模型中需要量化的参数包括：匀速时段持续时间、制动效能提升时间、制动效能提升时间、最大制动减速度。

b）在转向模型中需要量化的参数包括：直线时段持续时间、方向盘转角峰值和方向盘转动周期。

（7）基于驾驶模拟器，设计合理的试验，获取人类驾驶员在各模式下的参数，求取均值。

3 基于行为安全的自动驾驶能力认可方法

在实际应用“合理可预见且可避免”原则时，并不能单纯叠加“合理可预见”和“可避免”两种筛选条件。单纯基于自然驾驶数据进行泛化，场景数据中会产生不满足动力学约束和运动学约束的点，如横向加速度超过了 9.8 m/s^{2}或者前车出现逆行。由于场景中发生其他事故造成中断，该原则也并不适用，比如前车剐蹭前前车，前车切出情景未完成转变成了事故场景；同理，切入车辆追尾自车，也并非初始切入场景的测试目的。在后续实际应用时，应在“合理可预见”和“可避免”的基础上增加其他专家经验约束条件，以形成公认的行为安全接受准则。本文推荐的细化使用原则如 3.1 节所述。

3.1 行为安全使用原则

3.1.1 非典型情形下的行为安全准则

本文通过对高速道路上的典型事故类型和违法行为进行分析，得到非典型情形下的，即行为安全不足的可接受条件，如下列 7 点：

（1）正在发生事故的物体发生碰撞可接受；

（2）目标车失控造成的碰撞风险可接受；

（3）碰撞正在掉落及滚动过程中的遗落物体可接受；

（4）后方车辆追尾碰撞可接受；

（5）碰撞且非拥堵工况下行人及目标车横穿可接受；

（6）对方存在严重违法：前方车辆溜车、倒车、逆行、连续变道造成碰撞可接受；

（7）碰撞夜间前方未按规定开启灯光的静止车辆可接受。

相对而言，需要制定行为安全不足的不可接受准则，如下列 6 点：

（1）不应因避险发生另一起事故；

（2）不应因未给他车留出安全处置时间，导致非接触式事故；

（3）不应在自车处于非起步状态下因目标物由可见区域进入车辆自身盲区的碰撞事故；

（4）不应因匝道下的大车转弯半径问题导致自车碰撞大车事故；

（5）已发生事故或者故障的静止车辆及周围行人不可碰撞；

（6）拥堵工况下遮挡行人横穿按照合理可预见且可避免原则界定。

3.1.2 合理可预见原则使用方式

单一稳定行为（不存在相对于自车的横向位移和纵向加速度的情况，如车辆静止、匀速轧线行驶）不存在不合理的空间，根据道路限制和目标姿态求解所有可能参数组合。

稳定行为+危险行为（存在相对于自车的横向位移和纵向加速度的情况，如切入、制动）前后顺序拼接仅考虑危险行为的合理可预见空间。

单一危险行为的合理可预见空间：

（1）以各二维空间均值线为中心向两边扩充 N 倍标准差；

（2）对于切入切出类行为选择 N=3，对于制动类行为选择 N=5。

危险行为的组合空间：

（1）假设 2 种危险行为为独立事件；

（2）分别确定 2 个事件参数在各自行为二维空间下的最小的 N 值，根据 N 值确定单一行为在此 N 倍外的双边概率 P，求两个行为发生概率的乘积，利用该值反向查询 N 值，若 N 小于 3，则为合理可预见取值。

3.1.3　可避免原则的使用方式

评估场景是否对人类驾驶员有难以控制的风险，如果没有，则仅考虑合理可预见准则，相反则需进一步考虑可避免原则。

风险出现点：对于大车切入场景，选取刚开始有横向偏移动作的起点，对于小车切入场景，如果小车正确开启转向灯，则选取刚开始有横向偏移动作的起点，如果未开启灯光，则选取偏离中心线 0.375 m。

决策响应时间：如有潜在风险，则为较短时间；无潜在风险，则为较长时间。时间差异由试验结果确定。

模型介入时间点：与真实风险出现点相同。

模型介入前自车参考状态设定：对于切入切出类场景和横穿场景与自动驾驶自车的速度和相对位置相同；对于制动类工况，按照自车速度与目标车相同，距离依据法定跟车时距计算。

3.2　行为安全使用方法

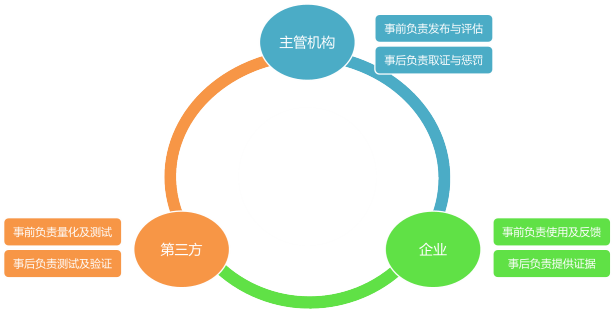

图 4　行为安全使用方法

3.2.1　事前使用方法

主管机构公布行为安全相关的各项原则、危险行为的参数及参数的均值和方差，确定好小概率事件的标准差要求及可避免模型的使用方法，并对企业提供的行为安全遵从证明材料进行评估。根据企业的使用反馈，增补修订相关原则及行为类别。

第三方根据主管机构收集到的危险行为参数及可避免模型量化需求，进行相关试验研究，将得到的结果反馈给主管机构进行裁定。并依据主管机构要求，对企业进行行为安全遵从性测试。

企业根据主管机构公布的信息，结合自身产品定义，建设行为安全遵从性验证的场景库，并进行全量仿真和边界点的场地测试验证，将结果提交给主管机构评估并进行本地封存。企业依据实际道路测试数据，给主管机构反馈危险行为参数及可避免模型量化需求。

3.2.2　事后使用方法

当发生行为安全相关事故时，主管机构现场进行简易评估，如果符合行为安全要求，则进行常规事故责任判定；如果不能判定是否符合行为安全要求，则需启动调查取证程序，进行专项立案评估。根据第三方的调查结果，依据相关法条，进行企业判罚，或者增加危险行为列表。

企业提供该事故的全量数据，并调取该事故相关的事前测试数据。

第三方根据企业提供该事故数据进行复现，若该事故中目标物的危险行为并未包含在公布的危险行为目录里，则企业无责，根据这一新行为开展事前研究，对危险行为进行增加。若该事故中目标物的危险行为包含在目录中，且参数落在可预见可避免范围内，深入研究企业的事前测试数据和相关第三方测试，查证不实之处。

4　总结

（1）针对自动驾驶接受准则难以统一并量化的问题，本文通过综述现有研究，凝练行为安全、交通法规符合度、ODD 合理性、人机交互和残留风险认可 5 种安全概念的定义及相互关联关系。

（2）本文参考合理可预见且可避免的研究现状，通过引入潜在风险点概念表征防御性驾驶能力，提出了更符合人类驾驶行为特性的量化研究方法框架。

（3）本文结合事故和实践经验，给出了非典型情形下的行为安全准则及典型情形下合理可预见原则和可避免原则的使用方式。

（4）最后给出了主管机构、行业第三方、研发企业间围绕行为安全需要开展的工作及各方应承担的义务，以实现事前事后的闭环管理。

参考文献（References）

[1] 陈龙.正确的自动驾驶“安全观”应是什么样子?[EB/OL].[2023-04-28].https://baijiahao.baidu.com/s?id=1721930222712213876&wfr=spider&for=pc.
CHEN Long.What Should Be The Correct Safety Concept for Automated Driving?[EB/OL].[2023-04-28].https://baijiahao.baidu.com/s?id=1721930222712213876&wfr=spider&for=pc.(in Chinese)

[2] NHTSA. Automated Driving Systems 2.0: A Vision for Safety[Z/OL].[2023-04-28].https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf.

[3] 中国汽车技术研究中心有限公司,同济大学,百度,等.自动驾驶汽车交通安全白皮书[Z/OL].[2023-04-28].https://www.xdyanbao.com/doc/zrcuxz9nti?bd_vid=10308723772956330160.
CATARC, Tongji University, Baidu, et al.White Paper on Traffic Safety of Automated Vehicle[Z/OL].[2023-04-28].https://www.xdyanbao.com/doc/zrcuxz9nti?bd_vid=10308723772956330160.(in Chinese)

[4] 中华人民共和国中央人民政府.中华人民共和国道路交通安全法[Z/OL].[2023-04-28].https://www.gov.cn/banshi/2005-08/23/content_25575.htm.
The State Council of the People's Republic of China. Road Traffic Safety Law of the People's Republic of China[Z/OL].[2023-04-28].https://www.gov.cn/banshi/2005-08/23/content_25575.htm.(in Chinese)

[5] Road Vehicles—Safety of the Intended Functionality: ISO 21448:2022[S/OL].[2023-04-28].https://www.iso.org/standard/77490.html.

[6] Uniform Provisions Concerning the Approval of Vehicles with Regard to Automated Lane Keeping Systems: UNECE R157[S/OL].[2023-04-28].https://op.europa.eu/en/publication-detail/-/publication/36fd3041-807a-11eb-9ac9-01aa75ed71a1.

[7] Guidelines and Recommendations Concerning Safety Requirements for Automated Driving Systems: UNECE FRAV-29-05[S/OL].[2023-04-28].https://unece.org/sites/default/files/2022-01/GRVA-12-23e.pdf.

[8] 中华人民共和国工业和信息化部,中华人民共和国公安部.关于开展智能网联汽车准入和上路通行试点工作的通知[EB/OL].[2023-04-28].https://www.miit.gov.cn/cms_files/filemanager/1226211233/attach/202210/ea67441fe7ff408e9d7ec9ff85c41d69.pdf.
Ministry of Industry and Information Technology of the People's Republic of China, Ministry of Public Security of the People's Republic of China.Notice on Carrying out Pilot Work on the Access and Road Access of Intelligent and Connected Vehicles[Z/OL].[2023-04-28].https://www.miit.gov.cn/cms_files/filemanager/1226211233/attach/202210/ea67441fe7ff408e9d7ec9ff85c41d69.pdf.(in Chinese)

[9] Road Vehicles—Automated Driving Systems—Scenario Based Safety Evaluation Framework: ISO 34502:2022[S/OL].[2023-04-28].https://www.iso.org/standard/78951.html.

[10] NAKAMURA H, MUSLIM H, KATO R, et al.Defining Reasonably Foreseeable Parameter Ranges Using Real-World Traffic Data for Scenario-Based Safety Assessment of Automated Vehicles[J].IEEE Access, 2022, 10: 37743-37760.

[11] 吴斌,朱西产,沈剑平,等.自然驾驶工况的驾驶员紧急转向变道行为[J].同济大学学报(自然科学版),2017(4):554-561.
WU Bin, ZHU Xichan, SHEN Jianping, et al.Analysis of Driver Emergency Steering Lane Changing Behavior Based on Naturalistic Driving Data[J].Journal of Tongji University (Natural Science), 2017(4): 554-561.(in Chinese)

作者简介

陈龙（1989-），男，河北廊坊人，博士，主要研究方向为自动驾驶安全概念及评估方法。
Tel：13521247551
E-mail：chenlong228@huawei.com