

自动驾驶的生命防线：如何科学评估无人车的风险

来源：黄坚源 发布时间：2025-11-13 17:21:14

当你坐在一辆自动驾驶汽车里，看着方向盘自动转动、油门自动调节时，你是否想过：这台车究竟有多安全？是什么在保护我们的生命？答案可能出乎意料——不是司机的反应速度，也不是人类的驾驶经验，而是一套看不见但无处不在的“风险评估系统”。这套系统正在成为自动驾驶技术能否走进千家万户的关键钥匙。

近期，来自同济大学汽车学院的研究团队在学术期刊《自主智能系统》上发表了一篇综合性文献综述，系统性地梳理了自动驾驶风险评估的最新研究进展。这项研究不仅为学术界提供了重要参考，更为普通人理解自动驾驶安全机制打开了一扇窗。

看不见的威胁：自动驾驶从哪里来的风险？

你可能想象不到，一辆自动驾驶汽车面临的风险来自三个完全不同的地方——感知、预测和决策。这是自动驾驶系统的三大薄弱环节，也是所有安全事故的潜在源头。

首先是感知的风险。想象一下，一辆自动驾驶汽车在下雨天行驶。此时，摄像头、激光雷达、毫米波雷达这些“眼睛”可能会迷茫——雨水遮挡了视线，传感器捕捉到的信息变得模糊不清。有时候，汽车甚至会把一个部分被遮挡的行人识别错误，或者误判障碍物的距离。根据同济大学的研究，感知不确定性是“安全链条中最薄弱的一环”，因为一旦感知出错，这个错误会像病毒一样传播到后续的预测和决策系统，最终导致严重后果。



其次是预测的风险。自动驾驶汽车需要预测周围车辆、行人的下一步行动。但人类行为往往充满了随意性和互动性。一个行人可能突然冲过马路，一辆车可能在没有信号的情况下突然变道。自动驾驶系统需要理解这些“未知意图”，而这几乎是不可能完成的任务。加上那些罕见的、长尾的场景——比如特殊恶劣天气下的驾驶——现有的预测模型往往无法覆盖。

最后是决策的风险。在一个复杂的路况中，自动驾驶系统需要在瞬间做出决策：是加速，还是减速？是左转，还是右转？这些决策往往存在内在的冲突——提高安全等级可能会降低行驶效率，追求舒适可能会增加风险。而且，系统对决策后果的评估总是不完美的。

四种武器：科学家如何对风险进行“量化”？

如果风险是一种看不见的敌人，那么风险评估方法就是我们对付它的武器。同济大学的研究将这些武器分为四大类，每一种都代表了人类技术进步的不同阶段。

规则型方法：最古老但最可靠的防线

最早期的风险评估就像一份严格的规则清单。系统根据预先设定的规则进行判断：如果车与前车的距离小于一定值，就立即减速；如果某个传感器出现故障，就立即启动备用方案。这种方法的优点是逻辑清晰、易于验证，就像遵守交通规则一样直接有效。

最著名的规则型系统是“RSS理论”(Responsibility-Sensitive Safety)。它就像一份法律文件，明确定义了在各种场景下安全的驾驶距离和反应时间。如果你违反了这些规则，系统就知道你处于危险中。但问题是，现实世界远比规则复杂。并非所有风险都能用几条规则来定义。

概率统计方法：用数据说话

当规则不够用时，科学家们转向了数据。概率统计方法通过分析大量真实驾驶数据，建立风险的概率分布模型。比如，系统会计算在特定条件下发生碰撞的概率有多高。

这种方法的妙处在于它承认不确定性的存在。系统不会说“一定安全”或“一定危险”，而是说“有80%的概率是安全的”。这更接近现实世界的真相。研究者使用蒙特卡洛模拟、贝叶斯推理等高级数学工具，能够从历史数据中提取出风险的真实特征。

等价模型方法：借用物理规律的智慧

有一类聪明的方法把自动驾驶的风险转化为物理问题。比如著名的“人工势场法”，它把行驶中的车辆想象成在一个力场中运动。目标地点是一个“吸引力”，障碍物是“排斥力”。风险就像一个“能量场”——接近障碍物时能量升高，说明风险增加。

这类方法的好处是直观易懂，而且计算效率高。系统可以实时评估车周围的风险分布，就像看一张热力图——红色区域风险高，绿色区域风险低。

深度学习方法：让机器自己学会评估风险

最新、最强大的方法是使用深度神经网络。系统不需要人类明确告诉它什么是风险，而是通过学习海量的驾驶数据，自己去发现风险的规律。这就像一个学徒通过看一千次驾驶视频，逐渐学会了识别危险的能力。

深度学习方法特别擅长处理高维、非线性的问题。比如，通过分析驾驶行为的时间序列，系统可以识别驾驶员即将犯错的信号。通过强化学习，系统甚至可以自己制定最优的驾驶策略——既安全又高效。但这类方法也有弱点：它们需要大量标注数据，而且往往缺乏可解释性。

数据的诞生：如何获得评估风险所需的信息？

风险评估再先进也需要数据支撑。如果没有正确的数据输入，就像盖房子没有地基。那么，这些宝贵的数据是从哪里来的呢？

第一个来源是真实世界的驾驶数据。研究者从Waymo、百度、特斯拉等公司收集的真实行驶记录中挑选出“关键时刻”——那些几乎导致事故的场景。通过物理模型和机器学习，科学家们能从成千上万的无趣数据中筛选出那些最有价值、最能代表风险的片段。

第二个来源是模拟环境。现实中很难收集到大量的危险数据（因为真正的事故太少了），所以研究者使用高保真的驾驶模拟器——比如著名的CARLA平台——来生成那些罕见但关键的高风险场景。通过因果贝叶斯网络等算法，系统可以从真实事故数据中学习规律，然后在虚拟环境中合成出类似的危险场景。

第三个来源是人类反馈。研究者通过眼动仪、脑电图等生物传感器，记录真实驾驶员在危险时刻的生理反应。系统学习人类是如何感知和应对风险的，然后把这些知识融入到自动驾驶的决策过程中。这种方法虽然成本高，但最接近真实的安全需求。

系统框架的完整性：如何把风险评估整合到整个系统中？

得出风险评估结果只是第一步。关键的第二步是把这些结果真正整合到自动驾驶汽车的决策系统中。同济大学的研究强调，风险评估不能是一个孤立的环节，而要贯穿感知、规划和控制的整个流程。

感知阶段，系统评估传感器数据的可靠性，如果检测到可疑目标或异常数据，就提高警觉。规划阶段，系统利用风险评估的结果来选择最安全的路线和驾驶策略。控制阶段，系统根据实时风险反馈来调整加速度、转向等控制指令。

特别重要的是“功能安全”（SOTIF）的概念。这不是指硬件故障的安全防护，而是指系统的核心功能本身是否足够安全。比如，即使传感器工作正常，检测算法也可能出错。SOTIF框架要求系统不仅要预防故障，还要预防“设计缺陷”——那些在正常工作情况下仍然可能导致事故的功能不足。

三大困境：当前自动驾驶风险评估面临的瓶颈

尽管学术界付出了巨大努力，风险评估仍然面临三个根本性的困境，这些困境可能是阻碍自动驾驶大规模商用的关键因素。

困境一：缺乏统一的“风险语言”

目前，学术界还没有达成共识关于如何定义和衡量风险。有人用“碰撞时间”（TTC）来表示风险，有人用“碰撞能量”来计算风险，还有人引入驾驶员的主观感受。这就像各国说着不同的语言——一个说“安全”，另一个说“有点危险”，但谁都不知道彼此的标准是什么。

这个问题看似技术性，实际上影响深远。因为没有统一的标准，不同的自动驾驶系统采用不同的风险评估方法，导致它们在同样的场景中做出完全不同的决策。最终用户和监管部门都无法判断哪辆车更安全。

困境二：忽视了系统自身的“实力”

这是一个容易被忽视但极其重要的问题。想象两辆都在同一场景中的自动驾驶汽车。如果一辆车的感知能力强、算法更智能、反应更快，它实际面临的风险就远低于另一辆性能差的车。但目前大多数风险评估方法只看场景本身的危险程度，而忽视了系统的应对能力。

这就像说“一个十级台阶对所有人都等地危险”，但实际上这对运动员和对老年人来说风险是完全不同的。未来的风险评估必须能够根据自动驾驶系统的实际能力动态调整风险等级。

困境三：数据标注的噩梦

最后一个困难是实践性的：获取大规模、高质量、准确标注的风险数据极其困难。通过传感器筛选数据需要复杂的算法；通过模拟生成数据虽然数量充足但真实性不足；通过人工标注数据虽然准确但成本天文数字。

这就像在沙漠里寻找金矿——即使知道大致方向，实际开采仍然困难重重。没有充足的高质量数据，再先进的算法也如同纸上谈兵。

未来的光亮：如何突破这些困境？

面对这些困难，研究者提出了一些希望的方向。第一，需要建立一个跨学科、国际化的标准委员会，就像制定交通规则一样，为自动驾驶的风险评估制定统一的国际标准。这样每个系统都能用同一套语言表达风险。

第二，风险评估必须融入系统能力评估。每次评估风险时，系统都要同时评估自己的感知、预测和决策能力。当系统能力下降时，风险等级自动上调；当系统经过训练能力提升时，风险等级适度下调。这样的评估才是真正“适配”的。

第三，开发更高效的数据生成和标注方法。研究者建议结合模拟、真实数据和众包的力量——在虚拟环境中大规模生成数据，用真实驾驶数据验证模型，再用众包让普通驾驶员标

注危险场景。这样可以在成本和数据量之间取得平衡。

结语：安全的自动驾驶不是梦

当一辆自动驾驶汽车在城市街道上平稳行驶时，在每一个路口、每一次加速、每一次转向，都有一套复杂的风险评估系统在后台运行。这套系统正在变得越来越聪慧、越来越可靠。

同济大学的这项综合研究表明，自动驾驶的安全性不是凭空而降的，而是通过科学、系统的风险评估框架一步步建立的。从感知到预测再到决策，每一个环节都有对应的评估方法。从规则型到深度学习，每一代技术都在进步。

虽然当前还存在许多挑战，但这些困难正在被逐个克服。也许在不久的将来，每一辆上路的自动驾驶汽车都会配备一个强大的“风险卫士”，确保你的每一段旅程都是安全的。到那时，我们就可以真正放心地坐在自动驾驶汽车里，享受它带来的自由和便利。

引用信息：本文基于《自主智能系统》（Autonomous Intelligent Systems）期刊2025年第5卷第24期发表的综合综述《自动驾驶中的风险评估：风险来源、方法论与系统架构的综合调查》（Risk assessment in autonomous driving: a comprehensive survey of risk sources, methodologies, and system architectures），作者为陆冬元、杜浩阳、吴正飞、杨硕，隶属于同济大学汽车学院。

常见四六级英文词汇表（100个）

序号	单词	词性	中文含义
1	autonomous	adj.	自主的，自动的
2	perception	n.	感知，认知
3			

assessment

n.

评估, 评价

4

uncertainty

n.

不确定性

5

quantify

v.

量化, 定量表示

6

risk

n.

风险, 危险

7

collision

n.

碰撞, 碰撞事故

8

trajectory

n.

轨迹, 运动轨道

9

obstacle

n.

障碍物，障碍

10

scenario

n.

场景，情景

11

framework

n.

框架，结构

12

integrate

v.

整合，融合

13

algorithm

n.

算法，计算方法

14

inference

n.

推断，推理

15

prediction

n.

预测，预言

16

decision

n.

决策，决定

17

constraint

n.

约束，限制

18

optimization

n.

优化，最优化

19

adaptive

adj.

自适应的，适应性的

20

robust

adj.

稳健的，鲁棒的

21

complexity

n.

复杂性，复杂度

22

efficiency

n.

效率, 效能

23

reliability

n.

可靠性, 可信度

24

threshold

n.

阈值, 临界值

25

parameter

n.

参数, 参量

26

probability

n.

概率, 可能性

27

distribution

n.

分布, 分配

28

variance

n.

方差, 变异

29

statistical

adj.

统计的，统计学的

30

empirical

adj.

实验的，经验的

31

validation

n.

验证，证实

32

verification

n.

验证，核实

33

performance

n.

性能，表现

34

accuracy

n.

准确度，精确性

35

error

n.

错误，误差

36

deviation

n.

偏差，离差

37

model

n. /v.

模型；建模

38

simulation

n.

仿真，模拟

39

synthetic

adj.

综合的，人造的

40

mechanism

n.

机制，机械装置

41

propagation

n.

传播，扩散

42

amplification

n.

放大，增强

43

detection

n.

检测，探测

44

recognition

n.

识别，认可

45

classification

n.

分类，分类法

46

evaluation

n.

评估，评价

47

assessment

n.

评估，估计

48

metric

n.

指标，度量

49

indicator

n.

指示器，标志

50

benchmark

n.

基准，标准

51

calibration

n.

校准，刻度

52

noise

n.

噪音，噪声

53

signal

n.

信号，标志

54

feedback

n.

反馈

55

interpolation

n.

插值，内插

56

extrapolation

n.

外推，推断

57

regression

n.

回归，衰退

58

correlation

n.

相关性，关联

59

convergence

n.

收敛，汇聚

60

divergence

n.

发散，偏离

61

ensemble

n.

集成，整体

62

random

adj.

随机的，任意的

63

stochastic

adj.

随机的，概率的

64

deterministic

adj.

确定性的，决定的

65

heuristic

n. /adj.

启发式的

66

feasible

adj.

可行的，切实可行的

67

optimal

adj.

最优的，最佳的

68

hierarchical

adj.

分层的，等级的

69

sequential

adj.

顺序的，连续的

70

temporal

adj.

时间的，暂时的

71

spatial

adj.

空间的，空间上的

72

correlation

n.

相关性，相关系数

73

momentum

n.

动量，势头

74

velocity

n.

速度，速率

75

acceleration

n.

加速度，加快

76

dynamics

n.

动力学，动态

77

kinetic

adj.

动能的，运动的

78

potential

adj. /n.

潜在的；势能

79

energy

n.

能量，精力

80

force

n.

力，强制力

81

field

n.

场, 领域

82

gradient

n.

梯度, 斜度

83

topology

n.

拓扑学, 拓扑

84

geometry

n.

几何学, 几何形状

85

curvature

n.

曲率, 弯曲度

86

constraint

n.

约束条件, 限制

87

penalty

n.

惩罚，罚款

88

reward

n.

奖励，回报

89

loss

n.

损失，丢失

90

gradient

n.

梯度

91

convergence

n.

收敛

92

divergence

n.

发散

93

stability

n.

稳定性

94

sensitivity

n.

灵敏度，敏感性

95

specificity

n.

特异性，具体性

96

generalization

n.

泛化，一般化

97

overfitting

n.

过拟合

98

underfitting

n.

欠拟合

99

regularization

n.

正则化，规范化

100

normalization

n.

归一化，标准化

HTML版本： [自动驾驶的生命防线：如何科学评估无人车的风险](#)