·自动驾驶技术专题·

基于强化学习算法的自适应巡航控制系统测试场景生成方法

赵德银[1]　高艳[1]　秦熠琳[1]　张培兴[2]　张家铭[1]

（1.中国第一汽车股份有限公司研发总院，长春 130013；2.吉林大学汽车工程学院，长春 130025）

【欢迎引用】赵德银，高艳，秦熠琳，等．基于强化学习算法的自适应巡航控制系统测试场景生成方法[J]．汽车文摘，2026(2)：32-36.

【Cite this paper】ZHAO D Y, GAO Y, QIN Y L, et al. Adaptive Cruise Control System Test Scenario Generation Method Based on Reinforcement Learning Algorithm[J]. Automotive Digest (Chinese), 2026(2): 32-36.

【摘要】为了提升自适应巡航控制系统的测试效率和完善性，提出了一种基于深度确定性策略梯度算法的测试场景生成方法。首先，设计了典型工况场景，构建了包含场景初始化、状态变量与控制变量定义、智能体交互机制及奖励函数的测试场景模型。然后，基于强化学习算法设计智能体，生成具有挑战性的测试场景，设计了综合曝光率、对抗性因素和不合理碰撞处罚的奖励函数，引导智能体生成既具有较高危险性又符合真实交通规则的场景。最后，通过左/右侧切入、切出等场景的仿真试验，验证了所提方法的有效性。结果表明，该方法能够自动生成ACC的关键危险场景，提高测试覆盖度和针对性，对实现系统的安全测试具有重要价值。

关键词：自适应巡航控制系统；测试场景生成；强化学习；DDPG算法

中图分类号：U461.91　　**文献标志码：**A　　**DOI:** 10.19822/j.cnki.1671-6329.20250156

Adaptive Cruise Control System Test Scenario Generation Method Based on Reinforcement Learning Algorithm

Zhao Deyin[1], Gao Yan[1], Qin Yilin[1], Zhang Peixing[2], Zhang Jiaming[1]

(1. Global R&D Center, China FAW Corporation Limited, Changchun 130013; 2. College of Automotive Engineering, Jilin University, Changchun 130025)

【Abstract】In order to improve the testing efficiency and comprehensiveness of the Adaptive Cruise Control (ACC) system, this paper proposes a test scenario generation method based on the Deep Deterministic Policy Gradient (DDPG) algorithm. Firstly, typical driving scenarios are designed, and a test scenario model is constructed, which includes scenario initialization, the definition of state and control variables, the agent-environment interaction mechanism, and a reward function. Then, an agent is designed based on a reinforcement learning algorithm, and generate challenging test scenarios. a reward function integrating exposure rate, adversarial factors, and unreasonable collision penalties is developed, guiding the agent to generate scenarios that are not only highly risky but also consistent with real traffic rules. Finally, simulation experiments on left/right cut-in and cut-out scenarios are conducted to validate the effectiveness of the proposed method. The results demonstrate that this approach can automatically generate critical hazardous scenarios for ACC, improve test coverage and specificity, and provide significant value for ensuring system safety testing.

Key words: Adaptive cruise control system, Test scenario generation, Reinforcement learning, DDPG algorithm

0　引言

自适应巡航控制系统（Adaptive Cruise Control System，ACC）[1]是自动驾驶辅助系统中的关键功能，通过在纵向自动调节车速和横向保持车辆在车道内行驶，实现车辆智能跟驰和车道保持等L2级自动驾驶能力。在复杂多变的道路交通环境中，前方车辆的切入（他车从相邻车道并线到自车前方）和切出（前车从

本车道驶离）场景十分常见，对ACC系统反应能力提出了更高的要求。

为了保障ACC系统在此类场景下的安全可靠运行，需要利用大量且多样化的测试场景对其功能进行充分验证[2]。然而，实际道路中危险工况（如突然切入）发生概率极低[3]，依赖人工设计或随机测试难以覆盖所有长尾场景，基于场景的虚拟仿真测试体系是现阶段主要的测试手段[4-5]。

强化学习具备自主探索和对抗优化能力，在自动驾驶测试场景生成领域应用广泛。Koren等[6]将车辆状态描述为马尔可夫决策过程，采用蒙特卡洛树搜索和前馈神经网络衍生危险测试场景。Du等[7]通过增加不同轨迹的奖励提高了生成危险测试场景的多样性。Chen等[8]考虑到边缘测试场景通常分布在不同的聚类区域，通过非参数贝叶斯方法对危险测试场景聚类，结合深度确定性策略梯度强化学习模型衍生交通车切入主车车道的危险测试场景。Feng等[9]利用密集强化学习建模交通车智能体并执行对抗性策略，能够高效生成大量的危险场景。

深度确定性策略梯度（Deep Deterministic Policy Gradient，DDPG）算法[10]是一种典型的深度强化学习方法，采用Actor-Critic架构能够处理连续动作决策问题，适用于交通场景中连续加速度控制的建模。基于上述问题，本文将DDPG强化学习算法应用于ACC测试场景生成，以高级自适应巡航控制（Supervisory Adaptive Cruise Control，SACC）典型工况（切入或切出）为例，利用DDPG智能体建模交通车对抗行为，学习产生对SACC具有高挑战性的测试场景，验证本文方法的有效性。本研究不仅可用于自动生成本文特定场景类型下SACC的危险测试场景，提高测试效率和覆盖度，也适用于任意场景类型下危险测试场景生成任务。

1 SACC测试场景设计及DDPG智能体设计

1.1 SACC测试场景设计

切入和切出是ACC测试中的典型场景，本节从场景初始化、场景状态变量和控制变量定义、场景与智能体交互、奖励函数4个方面阐述设计过程。

1.1.1 场景初始化

在场景初始化阶段，需设定最大仿真步数、主车与交通车的初始位置、速度及相对距离等参数，同时明确切入或切出的触发距离与横向速度。为了增强场景泛化能力，在每一回合仿真时，对部分初始参数进行重新随机化。

1.1.2 场景状态变量和控制变量定义

场景状态变量主要包括主车与交通车的横向和纵向距离、主车与交通车的速度、智能驾驶模型（Intelligent Driver Model，IDM）输出的主车加速度、车头时距、切入或切出状态以及碰撞标志。控制变量为DDPG智能体输出的交通车加速度。上述变量在设定的初始条件下，能够依据车辆的运动规律，推动场景动态变化。

1.1.3 场景与智能体交互

在交互过程中，DDPG智能体根据当前状态输出交通车的加速度，场景模型据此更新车辆的速度、位置等相关状态。当满足切入或切出条件时，交通车执行横向位移模拟变道动作，主车则根据IDM模型调整自身加速度。每一步交互后，场景会输出新的状态变量、奖励值及终止标志，从而实现智能体与环境的闭环交互。

1.1.4 奖励函数

奖励函数是DDPG的重要组成元素，其直接影响智能体学习、进化的方向。为使生成的场景兼顾场景曝光频率和危险性，需要从曝光奖励、对抗奖励2个方面设计奖励函数。

（1）场景曝光频率是衡量智能驾驶在自然驾驶环境中遭遇场景的关键指标。设x表示某具体场景在自然驾驶环境中发生的概率，即智能驾驶在自然道路环境中行驶，在经历多次设计运行域θ所定义的场景后，其遭遇该具体场景$(x|\theta)$的可能性。曝光频率奖励衡量了测试场景的自然驾驶特性，可表示为：

$$R_1(s_t,a_t)=\begin{cases}\varepsilon, & p(x|\theta)>1\times10^{-3}\\ 0, & \text{其他}\end{cases} \tag{1}$$

式中：R_1为曝光频率奖励，s_t为状态，a_t为动作，ε为满足条件下的奖励值，$p(x|\theta)$为曝光频率。

（2）场景危险性衡量安全性评价指标下的场景危险程度，制动强度越大说明背景车的影响越大，场景越具有挑战性和危险性。因此被测系统紧急制动过程中的制动加速度可以量化智能体的对抗性奖励。对抗性奖励可表示为式（2），以加速度的导数衡量测试场景整体的危险程度，此时，对抗性奖励R_2可表示为：

$$R_2(s_t,a_t)=\begin{cases}|a_b(t)\times\dot{a}_b(t)|, & \text{紧急制动}\\ a_b(t), & \text{其他}\end{cases} \tag{2}$$

式中：a_b为制动加速度，被测系统的制动加速度绝对值大于3 m/s^{2}时认为进入紧急制动。

（3）SACC测试场景的奖励函数设计公式为：

$$R(s_t,a_t)=w_1\cdot R_1(s_t,a_t)+w_2\cdot R_2(s_t,a_t) \quad (3)$$

式中：R为SACC测试场景的整体奖励函数；w_1，w_2为权重系数，分别设置为0.3和0.7，以确保生成场景满足场景曝光频率的同时，更具危险性，满足ACC系统的测试需求。

1.2 DDPG智能体设计

与场景交互的智能体采用DDPG强化学习算法，以下从Actor网络、Critic网络和DDPG算法的具体流程3个方面进行说明。

1.2.1 Actor网络

Actor网络根据当前环境状态输出动作分布的均值和标准差，并从该高斯分布中采样得到实际动作。网络通过2层非线性全连接结构提取状态特征，再输出动作的均值和对数标准差。为了保证分布稳定性，将对数标准差限制在[-2,2]区间，并通过指数变换得到正的标准差值。在动作采样过程中，引入tanh映射将动作范围压缩到[-1,1]，随后缩放到[-5,5]以满足交通车加速度的控制需求，Actor网络结构如图1所示。

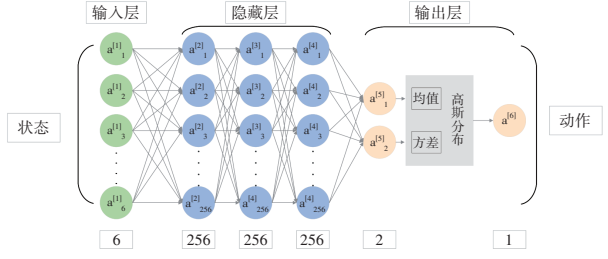


图1 Actor网络结构

1.2.2 Critic网络

Critic网络用于估计当前状态和动作对应的预期回报。在输入阶段，首先将状态和动作在最后一个维度上拼接为一个整体输入向量；随后通过2层隐藏全连接网络提取特征，并最终输出标量Q值，Critic网络结构如图2所示。

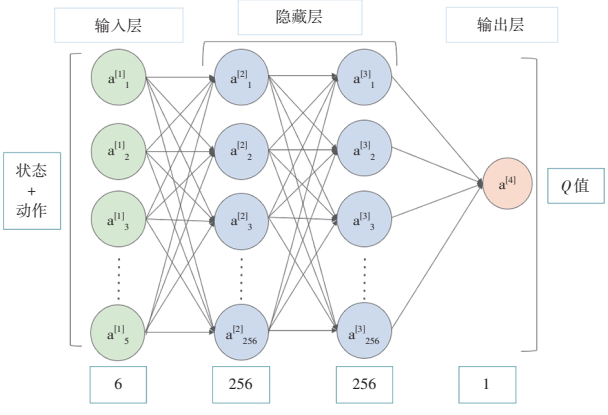


图2 Critic网络结构

1.2.3 DDPG算法的具体流程

首先对Actor和Critic网络的参数以及经验回放池进行初始化；然后Actor网络中的策略网络基于行为策略选择随机化的加速度动作，并在确定性策略中引入随机噪声，从而使动作决策从确定性过程变为随机过程，再从该随机过程中采样得到动作下达给环境执行；环境执行加速度动作后会获得相应的奖励，并进入下一状态，同时将状态转换过程信息存入经验回放池中。循环训练时，从经验回放池中随机采样一批数据，通过损失函数最小化更新Critic网络参数，通过策略梯度更新Actor网络策略。最后更新Actor和Critic的目标网络参数，如此循环直至达到收敛条件。DDPG算法伪代码如下。

算法1 DDPG算法

1：初始化Actor网络和Critic网络的参数：θ^Q和θ^μ
2：初始化目标网络中的参数$\theta^{Q'}=\theta^Q$，$\theta^{\mu'}=\theta^\mu$初始化经验回放池V，设最大训练次数为M，每回合最大时间步数为T
3：以下过程循环M步
4：初始化乌伦贝克-奥恩斯坦（Uhlenbeck-Ornstein, UO）随机过程N_t
5：获取初始场景信息
6：以下过程循环T步
7：根据策略和探索噪声选择动作
8：执行动作，得到奖励r_t和新的行驶状态s_{t+1}
9：将状态转换过程信息(s_t,a_t,r_t,s_{t+1})存入经验回放池V
10：从V中随机采样N个过程数据
11：通过最小化损失函数更新策略网络参数
12：计算Q网络的策略梯度更新Actor网络策略
13：更新目标网络参数
14：结束循环
15：结束循环

2 基于DDPG的ACC测试场景生成试验

通过DDPG智能体与场景的交互，可生成主车和交通车的运动轨迹。场景仿真的时间步长为0.1 s，总时长为30 s。将车辆纵向坐标、横向坐标和时间作为坐标轴，以三维图形的形式可视化主车和交通车的轨迹变化。

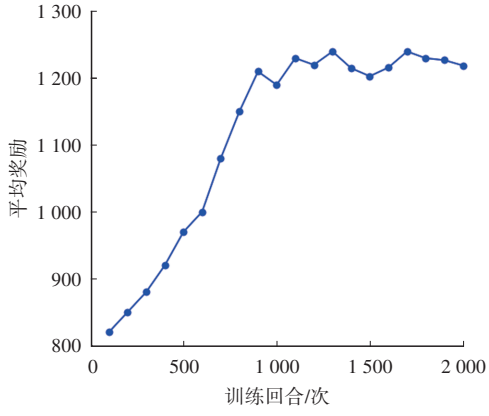
2.1 切入场景

2.1.1 左侧切入场景

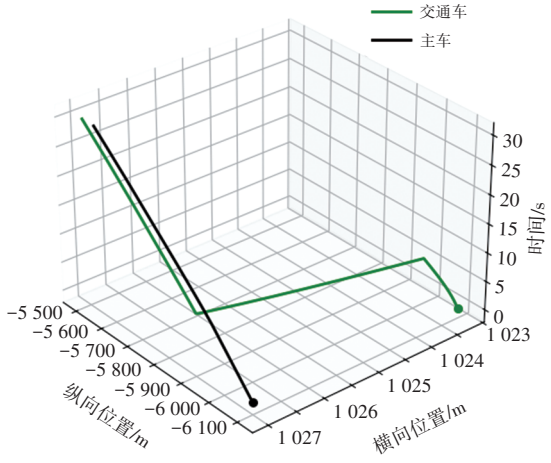
左侧切入场景中，主车的初始坐标为(1 027, -6 100)，初始车速为16.03 m/s；交通车的初始纵向偏移为-9.19 m，初始横向偏移为-3.76 m，初始车速为9.74 m/s。

当交通车相对主车的纵向距离为24.08 m时，以

1.38 m/s的横向速度向右变道，从主车前方切入，奖励曲线及轨迹示意如图3所示。



(a)左侧切入场景奖励曲线

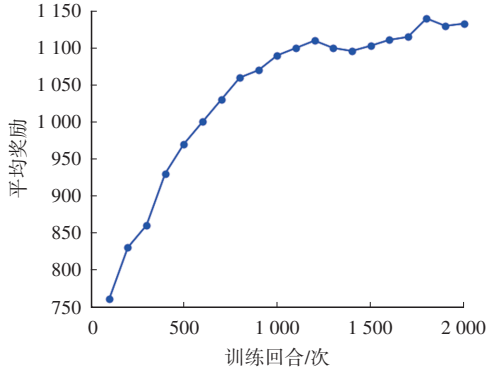


(b)左侧切入场景轨迹示意

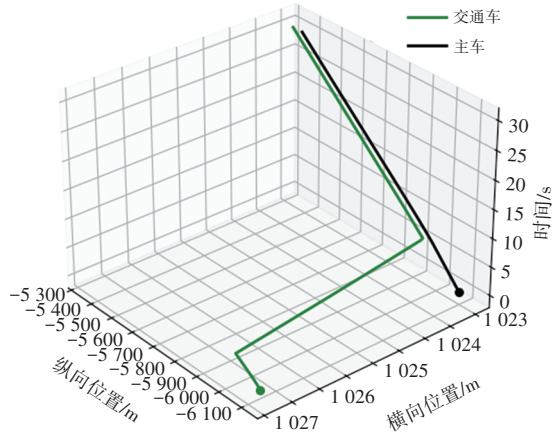
图3　左侧切入场景奖励曲线及轨迹示意

2.1.2　右侧切入场景

右侧切入场景中，主车的初始坐标为(1 023, −6 100)，初始车速为20.25 m/s；交通车的初始纵向偏移为−16.19 m，初始横向偏移为3.84 m，初始车速为25.20 m/s。当交通车相对主车的纵向距离为21.89 m时，以1.90 m/s的横向速度向右变道，从主车前方切入，奖励曲线及轨迹示意如图4所示。



(a)右侧切入场景奖励曲线



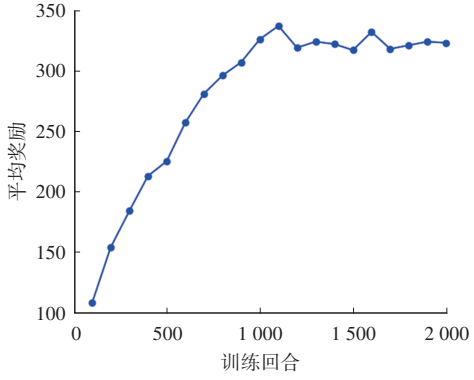
(b)右侧切入场景轨迹示意

图4　右侧切入场景奖励曲线及轨迹示意

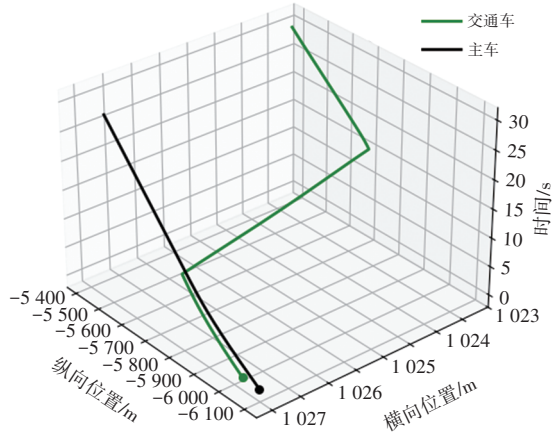
2.2　切出场景

2.2.1　左侧切出场景

左侧切出场景中，主车的初始坐标为(1 027, −6 100)，初始车速为30.31 m/s；交通车的初始纵向偏移为64.66 m，初始横向偏移为0 m，初始车速为28.48 m/s。当交通车相对主车的纵向距离为27.13 m时，以0.80 m/s的横向速度向左变道至横向偏移−3.74 m，从主车前方切出，从主车前方切出，奖励曲线及轨迹示意如图5所示。



(a)左侧切出场景奖励曲线

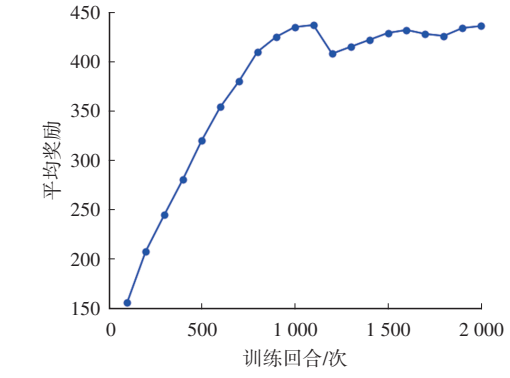


(b)左侧切出场景轨迹示意

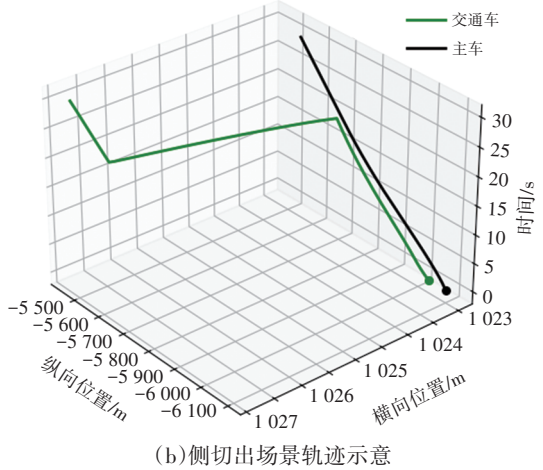
图5　左侧切出场景奖励曲线及轨迹示意

2.2.2　右侧切出场景

右侧切出场景中，主车的初始坐标为(1 023, −6 100)，初始车速为10.37 m/s；交通车的初始纵向偏移为58.50 m，初始横向偏移为0 m，初始车速为28.43 m/s。当交通车相对主车的纵向距离为28.75 m时，以0.56 m/s的横向速度向左变道至横向偏移3.83 m，从主车前方切出，从主车前方切出，奖励曲线及轨迹示意如图6所示。



(a)侧切出场景奖励曲线



(b)侧切出场景轨迹示意

图6　右侧切出场景奖励曲线及轨迹示意

3　结束语

本文针对ACC自巡航控制系统测试中的典型场景，提出了一种基于DDPG算法的智能测试场景生成方法。通过构建包含车辆动力学和决策逻辑的仿真场景模型，利用DDPG智能体自主学习背景车辆的加速度控制策略，实现自动生成对ACC具有挑战性的切入、切出场景。试验结果表明，该方法可有效触发被测ACC系统的紧急制动反应，其生成的测试场景在危险性和合理性方面达到较好平衡。同时，该方法具有良好的通用性，可进一步推广应用于其他驾驶辅助系统的场景测试。未来的研究将考虑引入更多智能交通参与者和复杂环境因素，以生成更加多样化的极端工况场景，并结合实际道路测试数据优化智能体策略，不断提升自动驾驶系统测试场景生成方法的完善性和实用价值。

参　考　文　献

[1] RAN W, LI C, SHA E H M. SACC: Secure-Cooperative Adaptive Cruise Control for Unmanned Vehicles[C]// Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data. Singapore: Springer Nature Singapore, 2024: 151-162.

[2] 朱冰, 张培兴, 赵健, 等. 基于场景的自动驾驶汽车虚拟测试研究进展[J]. 中国公路学报, 2019, 32(6): 1-19.

[3] CAO X, CHEN H, GELBAL S Y, et al. Vehicle-in-Virtual-Environment (VVE) Method for Autonomous Driving System Development, Evaluation and Demonstration[J]. Sensors, 2023, 23(11): 5088.

[4] 朱冰, 孙宇航, 赵健, 等. 自动驾驶汽车测试场景基元自动提取方法[J]. 汽车工程, 2022, 44(11): 1647-1655.

[5] 朱冰, 范天昕, 张培兴, 等. 智能网联汽车标准化建设进程综述[J]. 汽车技术, 2023(7): 1-16.

[6] KOREN M, ALSAIF S, LEE R, et al. Adaptive Stress Testing for Autonomous Vehicles[C]// 2018 IEEE Intelligent Vehicles Symposium (IV), Suzhou: IEEE, 2018: 1-7.

[7] DU P, DRIGGS-CAMPBELL K. Finding Diverse Failure Scenarios in Autonomous Systems Using Adaptive Stress Testing[J]. SAE International Journal of Connected and Automated Vehicles, 2019(2): 241-251.

[8] CHEN B, CHEN X, WU Q, et al. Adversarial Evaluation of Autonomous Vehicles in Lane-Change Scenarios[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(8): 10333-10342.

[9] FENG S, SUN H, YAN X, et al. Dense Reinforcement Learning for Safety Validation of Autonomous Vehicles[J]. Nature, 2023, 615(7953): 620-627.

[10] DANKWA S, ZHENG W. Twin-Delayed DDPG: A Deep Reinforcement Learning Technique to Model a Continuous Movement of an Intelligent Robot Agent[C]// Proceedings of the 3rd International Conference on Vision, Image and Signal Processing. New York: Association for Computing Machinery, 2019: 1-5.

（责任编辑　**梵玲**）