



交通运输系统工程与信息

Journal of Transportation Systems Engineering and Information Technology

ISSN 1009-6744, CN 11-4520/U

《交通运输系统工程与信息》网络首发论文

题目：大语言模型协同强化学习的自动驾驶决策方法
作者：王祥，任浩，谭国真，李健平，王珏，王妍力
收稿日期：2025-03-20
网络首发日期：2025-05-15
引用格式：王祥，任浩，谭国真，李健平，王珏，王妍力. 大语言模型协同强化学习的自动驾驶决策方法[J/OL]. 交通运输系统工程与信息.
<https://link.cnki.net/urlid/11.4520.u.20250514.1912.006>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

大语言模型协同强化学习的自动驾驶决策方法

王祥¹, 任浩^{*2}, 谭国真¹, 李健平¹, 王珏³, 王妍力¹

(1.大连理工大学, 计算机科学与技术学院, 大连, 116024; 2. 清华大学, 精密仪器系, 北京, 100084; 3.大连理工大学, 控制科学与工程学院, 大连, 116024)

摘要: 针对当前自动驾驶系统的高层决策缺乏具体执行细节和持续学习能力的问题, 本文围绕大语言模型 (Large Language Model, LLM) 在细化自动驾驶决策环节的应用展开研究。基于 LLM 强大的推理能力和强化学习 (Reinforcement Learning, RL) 的探索能力, 提出了一种 LLM 协同 RL 细化决策的方法: 首先, 基于 RL 输出的高级动作, 利用 LLM 的推理能力预测自车的未来轨迹点, 然后将 RL 模型的输出和当前状态信息相结合, 对下一个状态做出安全、无碰撞且可解释的预测, 最后将上述驾驶决策过程向量化后存储到记忆模块作为驾驶经验, 驾驶经验定期更新以此实现持续学习。LLM 预测的轨迹点为 PID (Proportional-Integral-Derivative) 控制器提供了详细的运动路径, 为其调整车辆加速度和速度提供依据, 确保车辆沿预定路径行驶。此外, 轨迹预测还能评估并规避潜在碰撞风险, 通过分析交通状态和历史数据规划出安全路径。

闭环实验结果表明: 本文决策方法在各项评估指标上均优于其他模型, 相对于 RL、单纯基于 LLM 的决策方法和基于 LLM 的跟车模型的驾驶分数提高了 35.12, 14.33 和 12.28, 拥有记忆模块的方法比没有记忆模块的方法的驾驶分数提高了 25.56。

关键词: 智能交通; 自动驾驶; 大语言模型; 强化学习; 持续学习; 轨迹预测

中图分类号: U495

文献标志码: A

Autonomous Driving Decision-making Method Based on Cooperative Reinforcement Learning of Large Language Model

WANG Xiang¹, REN Hao^{*2}, TAN Guozhen¹, LI Jianping¹, WANG Jue³, WANG Yanli¹

(1. School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China; 2. Department of Precision Instrument, Tsinghua University, Beijing 100084, China 3. School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: Aiming at the problems that the high-level decision-making of the current autonomous driving system lacks specific execution details and continuous learning ability, this paper conducts research on the application of the Large Language Model (LLM) in refining the decision-making process of autonomous driving. Based on the powerful reasoning ability of the LLM and the exploration ability of Reinforcement Learning (RL), a method of combining the LLM and RL to refine decision-making is proposed. Firstly, based on the high-level actions output by the RL, the reasoning ability of the LLM is used to predict the future trajectory points of the host vehicle. Then, the output of the RL model is combined with the current state information to make a safe, collision-free and interpretable prediction of the next state. Finally, the above driving decision-making process is vectorized and stored in the memory module as driving experience, and the driving experience is updated regularly to achieve sustainable learning. The trajectory points predicted by the LLM provide a detailed motion path for the

收稿日期: 2025-03-20

修回日期: 2025-04-26

录用日期: 2025-05-06

基金项目: 国家自然科学基金重点项目/Key Program of the National Natural Science Foundation of China (U1808206)。

作者简介: 王祥 (1997-), 男, 辽宁盘锦人, 博士生。 *通信作者: renhao@mail.tsinghua.edu.cn

引用格式: 王祥, 任浩, 谭国真, 等. 大语言模型协同强化学习的自动驾驶决策方法[J]. 交通运输系统工程与信息, 2025, 25(3): 00-00. [WANG X, REN H, TAN G Z, et al. Autonomous Driving Decision-making Method Based on Cooperative Reinforcement Learning of Large Language Model[J]. Journal of Transportation Systems Engineering and Information Technology, 2025, 25(3): 00-00.]

Proportional-Integral-Derivative(PID) controller, providing a basis for adjusting the vehicle's acceleration and speed to ensure that the vehicle travels along the predetermined path. In addition, the trajectory prediction can also evaluate and avoid potential collision risks, and plan a safe path by analyzing the traffic state and historical data. The results of the closed-loop experiment show that the decision-making method in this paper outperforms other models in all evaluation indicators. Compared with the RL, the decision-making method based solely on the LLM, and the LLM-based car-following model, the driving scores are increased by 35.12, 14.33 and 12.28 respectively. The method with the memory module increases the driving score by 25.56 compared with the method without the memory module.

Keywords: intelligent traffic; autonomous driving; large language model; reinforcement learning; continual learning; trajectory prediction

0 引言

近年来,自动驾驶技术取得了显著进展,有望提高交通效率,减少行车事故,已经成为人工智能、汽车工业和智能交通灯交叉领域的重要研究方向。现有的自动驾驶决策系统主要依赖于规则和模型驱动的方法,这些方法在处理复杂和动态变化的驾驶环境时,往往显得不够灵活。

黄志清等^[1]提出了一种端到端自动驾驶方法,利用深度学习模型从传感器数据直接生成动作,这类方法具有实现简洁和训练高效的优点,然而,这类方法对数据质量和模型容量要求极高,难以应对所有交通场景,尤其是在突发状况和未知环境下的表现有限。Deng 等^[2]将整个自动驾驶过程分解为多个子模块,每个模块由专门的算法和模型负责,这种方法在一定程度上提高了系统的可解释性和灵活性,但存在模块间耦合复杂和综合优化难度大等问题,无法完全捕捉真实驾驶场景中的动态复杂性和不可预测行为。宋晓琳等^[3]将模仿学习(Imitation Learning, IL)、用于换道决策,再在快速部署方面表取得了显著成果。Peng 等^[4]提出 RTA-IR,让人类专家经验参与决策过程,但是 IL 的决策能力取决于训练数据的质量和多样性,主要限制源于对训练示例的过拟合,处理新情况的困难,以及继承数据中的偏差。周卫林等^[5]提出了分段学习方法,证明了 IL 难以捕捉人类驾驶行为的细微差别,导致在意外情况下做出次优甚至不安全的决策。

李伟东等^[6]通过高层次横向决策来学习安全且高效的变道策略,在攻克自动驾驶汽车导航的复杂性难题上展现出良好前景,使智能体能够通过与环境交互的试错方法来学习最佳行为,RL 可以适应不断变化的交通场景,并从过去的经验中学习。李传耀等^[7]基于深度强化学习的生态驾驶策略,通过设计融合局部车辆动态与全局信号交叉口信息的状态空间,解决了互联自动驾驶车辆在信号交叉口场景下的纵向速度控制与横向换道协同优化问题。唐斌等^[8]利用柔性演员-评论家算法与 RL 的自适应学习能力,使自动驾驶车在应对动态且不可预测的道路交通环境时具有显著优势。自动驾驶训练方法和环境通常涉及受控的、可重复的场景,虽然有利于初始学习,但可能无法充分覆盖现实世界的复杂性。因此,利用 RL 模型进行自动驾驶训练在极大程度上依赖所提供训练场景的多样性与复杂性,于此类环境中专门训练的 RL 模型,很可能在面对不可预测的道路交通动态时,暴露出准备不充分的问题。

RL 在应用过程中,通常会面临大量迭代运算以及采样效率低下等难题,而 LLM 作为一项极具前景的技术,正被广泛应用于解决诸多下游任务,诸如分类、预测以及优化等领域。LLM 基于上下文学习的方法有 3 个优势:(1) LLM 可以执行上下文学习,无需任何额外的模型训练或参数微调,节省大量人力;(2) 基于 LLM 的上下文优化可以通过简单调整提示,快速扩展到新的任务或目标;(3) LLM 可以为其优化决策提供合理的解释,帮助人类理解复杂网络系统。

贺正冰^[9]分析了 LLM 能减少用户与结果间技术障碍、助力模型适应实际需求、实现交通视频自动理解、减轻文本梳理负担和推动自动驾驶技术发展。王祥等^[10]将 LLM 应用在高速公路和十字路口驾驶场景中生成相应的决策建议,从而提升自动驾驶系统的适应性。FU 等^[11]提出

了 DriveLikeHuman，运用 LLM 来模拟人类驾驶模式，让 LLM 深度参与自动驾驶的决策进程，显著增强自动驾驶系统对于复杂驾驶环境的理解能力以及决策的精准度与科学性。然而，在决策过程中并未有效利用过往的驾驶经验记忆，每一次面临决策时，模型都需重新开启完整的思考流程，这极大地增加了决策所需的时间成本。针对跟驰预测任务的应用及可解释性方面研究不足等问题，CHEN 等^[12]提出 GenFollower 框架，将跟驰行为转化为语言建模问题，利用 LLM 的推理能力和思维链技术，结合特定的提示工程，实现对车辆跟驰行为的有效预测和可解释性分析，为自动驾驶系统中跟驰行为建模提供了新的解决方案。HUANG 等^[13]基于语言推理的端到端方法在开环环境中进行训练和评估解决了部分问题，在数据集中根据专家操作生成操作和评估，而不是在仿真环境中闭环执行，对闭环评估导致对累积误差和动作时间不一致等关键问题的考虑不足，这使开环自动驾驶决策方法难以在短时间范围内扩展和部署。

针对上述问题，本文提出一种结合 LLM 和 RL 的自动驾驶闭环安全决策模型（Safety decision-making model based on LLM and RL, SDLLMRL），目标是增强自动驾驶系统的决策能力、安全性、模型可解释性和可持续学习能力。RL 作为高级决策器，负责根据当前环境状态和历史信息制定驾驶策略，LLM 作为验证器和预测导航点的功能模块可以根据自然语言描述和环境信息进行决策验证，提高决策过程的透明度和解释性，LLM 还能辅助 RL 模型进行更加精准和安全的路径规划。

1 语言推理与强化学习决策模型

1.1 SDLLMRL 总体架构

SDLLMRL 基于当前状态和过去的经验，生成未来轨迹的预测，如图 1 所示，首先 RL 智能体通过传感器观察环境车辆，为驾驶决策提供高级动作，其次 LLM 智能体检索经验库并进行推理，检索记忆采用相似度搜索(Facebook AI Similarity Search, FAISS)算法用于高效相似度搜索和密集向量聚类，提供接下来的六个轨迹点，最后 PID 控制器根据轨迹点实时调整车辆行为。

LLM 智能体的任务是在 RL 智能体的高级动作输入的指导下预测的下一个轨迹点，当 RL 智能体给出变道信号时，准确预测变道轨迹点。在执行任何操作之前，LLM 智能体也会作为 RL 策略的安全层，LLM 与 RL 智能体一起预测高级动作。仅当 LLM 和 RL 智能体给出相同的动作时，动作才会执行。否则，系统默认保持当前车道，并避免潜在的操纵风险，安全保障模块采用安全临界性碰撞时间(Time to Collision, TTC)和统一风险指数来评估车辆动作的安全性。

1.2 改进强化学习模型

在马尔可夫决策过程上下文中，RL 智能体的目标是优化策略最大化随时间推移的预期奖励之和^[3]，LLM 通过提供额外的上下文和高层次指导，使奖励函数在提升决策精确度和加速学习过程方面的作用得到增强，折扣回报 R_t 为

$$R_t = E \left[\sum_{t=0}^{\infty} \gamma^t (r(s_t, a_t) + \text{LLM}(s_t, a_t)) \right] \quad (1)$$

式中： E 为求数学期望， γ ($0 \leq \gamma < 1$) 为折扣因子，用于控制未来奖励的重要性， $r(s_t, a_t)$ 表示在状态 s_t 下执行动作 a_t 后，智能体获得的即时奖励， $\text{LLM}(s_t, a_t)$ 表示 LLM 提供的附加项，根据状态和动作进行奖励调整，引入 LLM 后的 Q 值更新方式为

$$y_i^{(\text{LLM})} = r + \gamma [Q_{\text{RL}} + (1 - \alpha(a_t, s_t)) Q_{\text{LLM}}] \quad (2)$$

$$Q_{\text{RL}} = Q \left(s', \arg \max_{a'} Q(s', a'; \theta); \theta' \right) \quad (3)$$

$$Q_{\text{LLM}} = Q \left(s', \arg \max_{a'} Q(s', a'; \theta), \text{context}; \theta \right) \quad (4)$$

式中： Q_{RL} 为 RL 模型计算得到的 Q 值，表示在状态 s' 下，按照当前策略选择使 Q 值最大的动作 a' 时的 Q 值，使用目标网络参数计算。 $\alpha(s_t, a_t)$ 是一个动态权重函数，可以基于当前的状态自适应调整，LLM 在当前状态和动作下的不确定性估计，其取值范围在 $[0, 1]$ 之间。当不确定

性接近 1，会减小对总和的影响，反之则增加其影响。 Q_{LLM} 是 LLM 对未来状态和动作的额外估计值，使用在线网络参数 θ 选择动作，使用目标网络参数 θ' 计算 Q 值，使用 LLM 对未来状态和动作的估计值进行补充。context 为上下文信息包含当前驾驶场景的详细描述、历史驾驶经验。每次迭代通过优化网络权值 θ_i ，最小化损失函数 $L_i^{(LLM)}(\theta_i)$ 为

$$L_i^{(LLM)}(\theta_i) = E \left[\left(y_i^{(LLM)} - Q(s, a; \theta_i) \right)^2 \right] \quad (5)$$

引入 LLM 后的梯度下降公式同样遵循这一原则，尽管目标值发生了变化，但损失函数的形式和导数计算的基本步骤没有改变，做一次梯度下降更新模型参数 θ_i 为

$$\nabla_{\theta} L_i^{(LLM)}(\theta_i) = E \left\{ \left[y_i^{(LLM)} - Q(s, a; \theta_i) \right] \times \nabla_{\theta} Q(s, a; \theta_i) \right\} \quad (6)$$

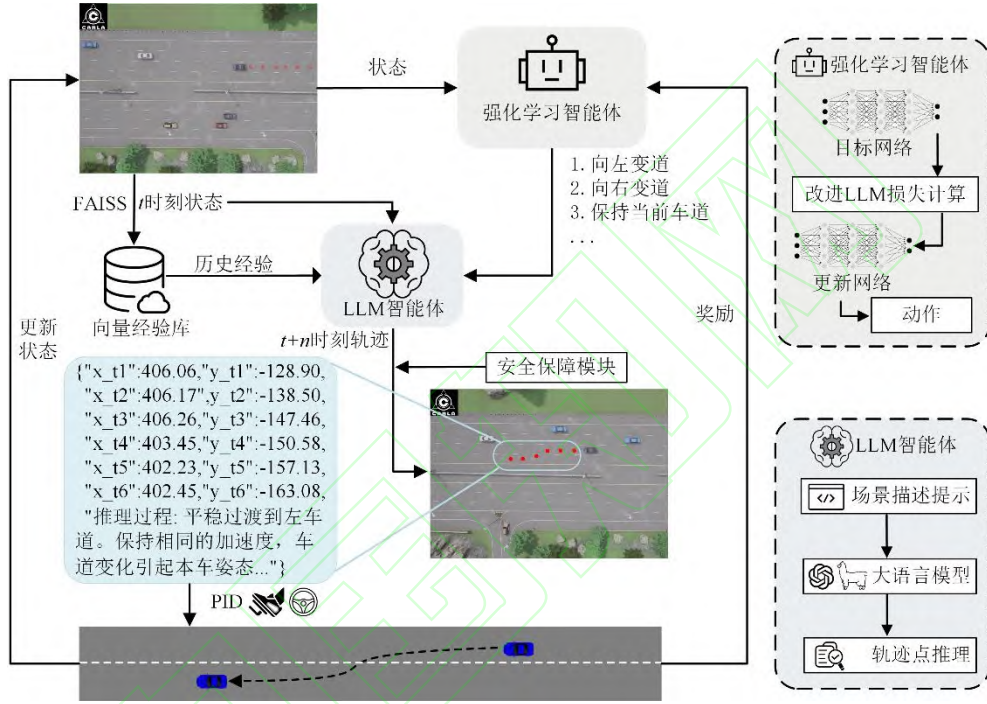


图 1 SDLLMRL 模型总体架构

Fig.1 Overall Architecture of the SDLLMRL model

1.3 奖励函数设计

奖励函数与速度、位置，与障碍物距离，平稳驾驶，是否行驶在道路中央有关。速度奖励鼓励车辆保持场景可行驶的最佳速度，如果车辆行驶太慢或太快，对车辆进行惩罚，鼓励车辆保持在指定的速度范围内，且尽可能的以最大安全速度行驶，速度奖励 R_v 为

$$R_v = \begin{cases} -\kappa_1 |v - v_{\min}| / v_{\min} & \text{if } v < v_{\min} \\ 0 & \text{if } v = v_{\min} \\ -\kappa_1 |v - v_{\max}| / v & \text{if } v > v_{\max} \\ \kappa_1 v / v_{\max} & \text{if } v_{\min} < v < v_{\max} \end{cases} \quad (7)$$

式中： v_{\min} 表示周围车辆的最小平均速度， v_{\max} 表示车道限速， κ_1 为速度奖励权重。

位置奖励值鼓励车辆靠近目标位置，计算方式为 $R_{\text{pos}} = e^{-\kappa_2 (d_{\text{pos}}^2 - \kappa_3)}$ ，其中， κ_2 和 κ_3 为位置奖励权重。距离奖励值鼓励车辆使用传感器数据与障碍物保持安全距离，如果传感器测量的距离 d 小于最小安全距离，给予惩罚，距离奖励 R_d 为

$$R_d = \begin{cases} -\kappa_4 (d_{\text{safe}} - d) / d_{\text{safe}} & \text{if } d < d_{\text{safe}} \\ 0.5\kappa_4 & \text{if } d \geq d_{\text{safe}} \end{cases} \quad (8)$$

式中: d_{safe} 为最小安全距离, κ_4 为距离奖励权重。

平稳驾驶奖励值 R_{sm} 鼓励车辆转向平稳变化, 使用最后四个转向动作 a_{list}^4 的标准差来衡量平滑度, $R_{\text{sm}} = -\kappa_5 \times \text{std}(a_{\text{list}}^4)$, std 为取标准差, κ_5 为平稳驾驶奖励权重。

碰撞惩罚 R_c 是一种关键的反馈机制, 当车辆发生碰撞时, 给予较大的负奖励, 进而调整策略, 以降低碰撞风险, 计算方式为 $R_c = -0.8\kappa_6$, κ_6 为碰撞惩罚权重。行驶车道线中央奖励值 R_{cen} 鼓励车辆利用车道线传感器检测是否行驶在车道线中央, 通过行驶在道路中央来引导车辆稳定地保持在车道中心位置行驶, 最大程度地增加与相邻车道车辆的横向安全距离, 减少因车道偏离而引发碰撞的可能性, R_{cen} 计算方式为

$$R_{\text{cen}} = \begin{cases} 0.7\kappa_7 & \text{if on_center} \\ -0.7\kappa_7 & \text{if no_center} \end{cases} \quad (9)$$

式中: κ_7 为偏离奖励权重, on_center 表示车辆行驶在车道线中央, no_center 表示车辆偏离车道线中央。综合驾驶奖励 R_{total} 的计算方式为 $R_{\text{total}} = R_v + R_{\text{pos}} + R_d + R_{\text{sm}} + R_c + R_{\text{cen}}$ 。

1.4 LLM 推理

虽然 RL 提供了左转、右转等高层次决策, 但决策并不包含车辆具体如何执行这些动作的详细信息。LLM 通过预测未来轨迹点, 细化这些高层次决策, 转化为具体的路径规划, 使车辆能顺利、安全地执行高层次决策动作。LLM 作为固定参数的推理引擎, 基于预训练知识和提示工程生成决策, 不进行权重更新。

提示模版设计: “你是一名自动驾驶决策专家, 当前驾驶场景为高速公路, 限速为 120km/h, 允许车辆变道。自车与环境车状态包括速度、位置、加速度及所在车道 ID 及相对方位, 强化学习输出的高级动作是向左变道。请结合历史经验检索得到的相似场景轨迹参考, 遵循最小安全距离、碰撞风险阈值及车道保持要求等安全约束, 根据强化学习的高级动作规划未来 3 秒内共 6 个轨迹点, 每 0.5 秒 1 个。输出需包含各时刻(x,y)坐标, 并阐述决策推理, 内容应涵盖强化学习动作执行逻辑、避障策略及历史经验复用依据。”

自车和环境状态由多维向量 \mathbf{S} 表示, 自车状态信息 \mathbf{S}_t 为

$$\mathbf{S}_t = [x, y, v_x, v_y, a_x, a_y, \text{Line}_j] \quad (10)$$

式中: 状态包含当前位置 x 和 y , 以及速度 v_x 和 v_y , 以及 x 方向加速度 a_x 和 y 方向加速度 a_y , 行驶的车道线信息 Line_j , j 为车道线标号, 周围车辆信息 \mathbf{S}_{vec} 为

$$\mathbf{S}_{\text{vec}} = \begin{bmatrix} [x_1, y_1, v_{x1}, v_{y1}, a_{x1}, a_{y1}, \text{Line}_j] \\ \vdots \\ [x_n, y_n, v_{xn}, v_{yn}, a_{xn}, a_{yn}, \text{Line}_j] \end{bmatrix} \quad (11)$$

LLM 的输入为当前状态的过去相似轨迹, 记为 $\tau_{\text{experience}}$, 轨迹预测问题表述为

$$P_S = \text{LLM}(\mathbf{S}_t, \mathbf{S}_{\text{vec}}, \tau_{\text{experience}}) \quad (12)$$

式中: P_S 表示时间 $t+1$ 的预测状态, $\text{LLM}(\cdot)$ 表示 LLM 推理函数。为了使 LLM 智能体的输出可转换为车辆动作的兼容格式, 将 LLM 的函数调用工具使输出格式转换为

$$P_{S-\text{LLM}} = [Px_{t+n}, Py_{t+n}, P_{\text{reason}}] \quad (13)$$

式中: $n \in \{1, 2, 3\}$, Px_{t+n} 为预测第 $t+n$ 秒 x 方向的位置, Py_{t+n} 为预测第 $t+n$ 秒 y 方向的位置, P_{reason} 为推理过程。

1.5 记忆存储与检索

SDLLMRL 集成了一个持续学习框架, 通过扩展记忆模块来提高算法性能。为了从 LLM 智能体中获得期望的输出, 使用提示工程技术和函数调用来限制系统输出。LLM 考虑给定的状态信息以及过去类似的轨迹, 并要求它通过避免任何碰撞来预测下一个状态, 以确保安全行驶和

平稳的速度转换。

LLM 智能体需要驾驶知识来提供对未来状态的合理预测, 本文设计一个由高速公路车辆轨迹组成的大型向量驾驶经验知识库^[10]。经验库 \mathcal{E} 基于亚琛工业大学汽车工程研究开发的自动驾驶数据集 HighD Datasets (HDD), HDD 收集了六个不同地点的高速公路真实车辆轨迹, 为 LLM 智能体提供经验检索。记忆模块通过检索经验库中的历史经验辅助决策, 但其静态特性限制了系统对动态交通环境的适应能力。随着驾驶场景的不断变化, 经验库中存储的旧经验会逐渐失效或降低驾驶性能, 而新经验的缺乏会导致系统在面对新场景时表现不佳。为此, 提出一种驾驶经验更新机制, 通过动态维护经验库, 确保系统始终存储高价值的驾驶经验。为了评估每条经验的效用, 综合考虑时间衰减、使用频率和场景相关性, 定义经验效用评分函数 $U(E_i)$

$$U(E_i) = \alpha_1 e^{-\lambda(t_{\text{current}} - t_i)} + \alpha_2 \frac{N_{\text{use}}(E_i)}{N_{\text{total}}(E_i)} + \alpha_3 S_{\text{Sim}}(E_i, \mathcal{E}_{\text{current}}) \quad (14)$$

式中: t_i 为经验 e_i 的存储时间戳, λ 为时间衰减系数, $N_{\text{use}}(e_i)$ 为经验被检索次数, N_{total} 为总检索次数, $S_{\text{Sim}}(E_i, \mathcal{E}_{\text{current}})$ 表示经验 E_i 与当前经验 $\mathcal{E}_{\text{current}}$ 的余弦相似度, $\alpha_1, \alpha_2, \alpha_3$ 为加权系数, 满足 $\alpha_1 + \alpha_2 + \alpha_3 = 1$ 。为了维护经验库的容量上限 ε_{max} , 采用动态阈值淘汰策略, 当经验库容量 $|\mathcal{E}_{\text{num}}| \geq \varepsilon_{\text{max}}$ 时, 计算所有经验的效用评分 $U(E_i)$ 。按 $U(E_i)$ 升序排列, 删除排名后 $k\%$ 的低效用经验。若存在 $t_{\text{current}} - t_i > t_{\text{max}}$ 的经验, 其中 t_{max} 为最大保存周期, 旧经验直接淘汰。

新增经验需满足以下条件之一: (1) 总奖励 R_{total} 大于奖励阈值 $R_{\text{threshold}}$ 的成功经验; (2) 在统一风险指数大于 0.8 的高风险场景下成功驾驶的风险规避经验; (3) 与现有场景经验的平均相似度 $S_{\text{Sim}}(E_{\text{new}}, E_i) < \delta$ 的新经验。存储时将新经验编码为 $E_{\text{new}} = \text{Encoder}(\tau_{\text{new}}, S_t, A_{RL})$ 。

经验库每次更新后需要重建倒排文件索引来平衡查询效率与存储开销, 新的索引 I' 为

$$I' = I \setminus \{\tau_i | e_i \in \text{deleted}\} \cup \{\tau_{\text{new}}\} \quad (15)$$

在检索经验前需要对场景中的自车与环境车辆状态进行匹配, 从驾驶经验知识库中筛选出与当前车辆及环境状态高度相似的历史状态。当前状态向量 $\mathbf{S}_{\text{current}} = [x, y, v_x, v_y, a_x, a_y, \mathbf{S}_{\text{vec}}]$, 为精确衡量当前状态与历史状态的相似程度, 采用欧氏距离度量两个状态向量空间距离。 $\mathbf{S}_{\text{current}}$ 和 $\mathbf{S}_{\text{history}}^{(i)}$ 之间的欧氏距离 d_E 为

$$d_E(\mathbf{S}_{\text{current}}, \mathbf{S}_{\text{history}}^{(i)}) = \sqrt{\sum_{j=1}^m (\mathbf{S}_{\text{current}}^{(j)} - \mathbf{S}_{\text{history}}^{(i,j)})^2} \quad (16)$$

式中: m 为状态向量的维度, $\mathbf{S}_{\text{current}}^{(j)}$ 和 $\mathbf{S}_{\text{history}}^{(i,j)}$ 分别为当前状态向量的第 j 个元素和第 i 个经验状态向量的第 j 个元素。欧氏距离的值越小, 表明两个状态在空间上越接近, 即相似度越高。

知识库中存储的历史轨迹转换为向量后, 矢量化给定的时间步长和状态。使用 l_2 距离度量, 检索与给定 \mathbf{S}_t 密切匹配的最相关的驾驶经验。轨迹被格式化为 3 秒内的 6 个路径点, 计算驾驶场景中轨迹的 l_2 误差, 将 l_2 均值最小的轨迹作为 LLM 的输入。

2 仿真验证与结果分析

2.1 仿真配置

(1) 仿真环境

本文实验环境为 Ubuntu20.04, GPU 为 NVIDIA GeForce GTX 4090 24G×2, 仿真平台使用 Carla 0.9.10。利用机器人操作系统(Robot Operating System, ROS)发布自车的状态信息, 并将 LLM 嵌入到 ROS 中。创建一个 ROS 节点负责从 Carla 仿真环境中接收车辆状态信息, 通过 RL 和 LLM 智能体生成安全的车辆动作和预测导航点。在 ROS 节点中调用各个版本 LLM 的 API, 将车辆的当前状态和环境信息输入到 LLM 中, 生成动作指令和预测的导航点后将发布到 ROS 话题中, Carla 中的车辆将订阅该话题并执行动作。

(2) 车辆运动学约束

车辆横向控制采用 PID 控制器计算横向加速度, 控制自车到达期望的位置。车辆纵向控制

采用智能驾驶员模型考虑前车与自车的距离和速度，纵向控制车辆加速度，通过观察距离并在距离接近阈值时降低速度来防止碰撞。自车的速度应满足：（1）使车辆尽快到达目标点；（2）确保在未来的道路上避免碰撞；（3）考虑自车的运动学约束。

2.2 实验设置

训练场景为 Carla Town 04 的高速公路部分，高速公路部分呈环形结构，车辆可在其中不断循环行驶，模拟类似真实世界中的环形高速公路场景，双向八车道，车道宽度 3.5~4 m。回合定义为一次完整的驾驶任务，发生以下情况之一视为回合终止，（1）从起点成功行驶到终点；（2）发生碰撞；（3）超时未完成驾驶任务。

高速公路测试场景将现实世界的交通动态纳入仿真环境，采用 SUMO 模拟真实交通场景，包括道路布局、车流量和交通规则等。路面宽度 26 m，车道宽度 3.25 m，双向八车道，最大时速 120 km/h，测试路线长度为 2 km，在距离自车 300 m 内的随机车道生成 30 辆环境车，速度在 100-110 km/h 之间。环境车不会主动发生碰撞，并且环境车辆具有激进、保守和正常的驾驶行为来模拟动态驾驶任务，进一步验证智能体的泛化能力。

十字路口测试场景定义直行、转弯和通过红绿灯等驾驶任务，实验按照任务设置在 Carla Town 05 场景中对算法进行测试。测试任务采用从起始点按照指定路线行驶至终止点的方式，测试场景中环境车辆一共为 45 辆，速度为 25-35 km/h。行车性能指标利用等 Shao 等^[14]提出的路线完成度和驾驶分数验证各方法的有效性。

对比方法为：（1）IL 方法；（2）RL 方法；（3）仅采用大模型决策的方法（Large Language Model Decision-making, LLMT），不使用 RL 智能体；（4）结合 RL 和在线专家训练控制层的方法（Reinforcement learning combined with online experts, RL_COE）；（5）GenFollower（GF）方法；（6）DriveLikeHuman（DLH）方法。

2.3 训练过程分析

SDLLMRL 的目标是在行驶时与其他车辆不发生碰撞，不发生不必要的变道，最大化完成任务线路，并尽可能多的得到更高的驾驶分数。LLM 用于解析交通规则、驾驶指南和环境描述，实现了动态应对复杂交通状况，结合 RL 算法在仿环境中寻找最优驾驶策略。在训练的早期阶段，策略网络通常呈现出较高的随机性，这致使智能体在选择行动时极有可能采取不安全的动作。在这种情况下，LLM 能够为策略网络提供一个大致的更新方向，从而有效应对 RL 初期中的高随机性和抵消学习等问题。随着训练的逐步深入，策略网络的损失逐渐增加，RL 中的损失项开始起到主要作用。智能体在不断探索的过程中，逐步优化和完善策略，从而提升其整体性能。通过引入上述辅助奖励项和改进的 Q 值更新公式，RL 能够在学习过程中利用 LLM 的预测结果，加快奖励函数的收敛速度。

实验共记录 4000 个回合，每 5 个回合计算一个平均值奖励值，回合平均奖励曲线如图 2 所示。SDLLMRL 的奖励值相对较高，尤其在多个回合能够达到大约 35 的奖励水平，说明该方法在场景下能够获得较为丰厚的回报。RL-COE 的奖励值较稳定，在 30 左右波动。DRL 的奖励值相对较低，在 10 到 20 之间波动，说明在相同的环境下，该方法获得的奖励相对较少。SDLLMRL 在早期就能够快速上升到一个较高的奖励水平，并且在后续的回合中也能保持相对稳定，能够迅速找到有效的策略以获得高奖励，表明 SDLLMRL 算法能够利用 LLM 对未来状态和奖励的估计，从而提升估计的准确性和收敛速度。RL-COE 的收敛速度相对较慢，需要经过 1100 回合才逐渐稳定在大约 30 的奖励水平，说明该方法在探索和优化策略的过程中花费了更多的时间。RL 呈现出较为缓慢的收敛趋势，奖励值在较低水平波动了较长时间，说明该方法的收敛速度较慢，需要更多的回合数来不断调整和改进策略。

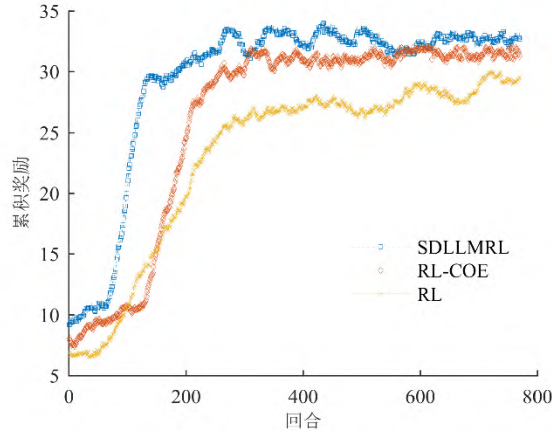


图 2 4000 个回合中每 5 个回合的平均奖励情况

Fig.2 Average reward per 5-episode interval over 4000 episodes

2.4 决策过程分析

为了评估 LLM 在驾驶安全和决策方面的性能，将 LLM 作为轨迹规划器，LLM 集成到 RL 的安全层运行，记录 100 次高速公路测试场景的碰撞次数、平均速度、路线完成度和驾驶分数。在高速公路驾驶场景中测试了 SDLLMRL 的决策性能并对比 6 种方法，结果如表 1 所示。

LLMT 作为轨迹规划器发生 2 次碰撞，平均速度为 104.7 km/h。相比之下，RL 导致平均 3.84 次碰撞，平均速度为 100.8 km/h。当 LLMT 智能体作为安全层时，平均发生碰撞次数 2 次，平均速度 104.7 km/h。将 LLM 作为轨迹规划器，由于缺乏信息理解和环境探索，在初始状态下生成的轨迹点不准确。在图 3 中展示了 SDLLMRL 在变道动作下的推理过程，RL 提供高级决策，LLM 提供一个轨迹路径点，并根据它的理解解释场景和行动，LLM 的推理可以用来解释 RL 在自动驾驶中的黑盒子性质。GF 方法在碰撞次数、平均速度、推理时间及路线完成度等指标上均表现突出，表明该方法通过 LLM 优化轨迹预测，有效平衡了决策安全性与效率，综合 SDLLMRL 取得的结果，验证了 LLM 用于驾驶决策的有效性。DLH 方法因过度依赖 LLM 生成复杂自然语言指令，导致平均推理时间显著高于其他模型，且路线完成度与驾驶分数均为最低水平。尽管 DLH 在理论上增强了系统可解释性，但其决策流程冗长且缺乏动态环境适应机制，在高速行驶场景中频繁出现因反应延迟引发的碰撞风险，整体表现未能达到实际应用需求。LLMT 的平均生成时间为 5.54 秒，而 SDLLMRL 的动作生成时间仅为 2.21 秒。此响应时间取决于检索数据库的大小和所需输出轨迹点的数量。另一方面，SDLLMRL 只提供一个作为变道动作的输出，而轨迹点以 (x, y) 的形式总共输出 6 个数据。因此，SDLLMRL 的响应速度比生成 LLMT 快两倍以上。与 LLMT 的响应时间相比，RL 提供的操作平均快 2 秒。RL 模型虽能提供高层次决策，但欠缺具体执行细节，而 LLM 通过预测未来轨迹点，可将高层次决策转化为详细的路径规划，保障车辆顺利、安全地执行动作。

表 1 不同方法的决策过程指标对比

Table 1 Comparison of decision-making process indicators for different methods

指标	RL	IL	RL_COE	GF	DLH	LLMT	SDLLMRL
碰撞次数	15	13	2	3	6	3	1
平均速度 (km/h)	101.4	108.6	113.4	115.4	105.5	104.7	117.2
平均决策时间 (s)	0.05	0.04	11.7	2.98	14.9	5.54	2.21
路线完成度 (%)	64.2	66.9	70.1	70.1	41.7	72.3	79.5
驾驶分数	24.12	36.87	47.26	47.26	28.6	45.21	59.54

RL 发起变道，LLM 提供变道的轨迹路径点和变道动作的推理。蓝色车为本辆，其他车辆为环境车辆。RL 输出的高级决策为换到左边车道。自车当前的 y 位置是 97.09，所以为了安全

地变道并保持在车道中心，在接下来的 3 个轨迹点中减小 y 以确保即使前面的车辆变道或调整速度，自车仍然有足够的空间进行安全机动和速度转换。

通过评估、反对或支持 RL 智能体提出的行动推理过程来评估 LLM 改善决策的能力。SDLLMRL 的推理过程为：“由于 RL 命令是换到左侧车道，并且考虑到高速公路的边界，将 y 位置调整为 95.5 是一个合理的选择。它确保安全行驶，避免与其他车辆碰撞，并实现平稳的速度转换。考虑到命令是换到左侧车道，给定前视距为 48.06 m。因此，预测自车将移动到左侧车道并减小 y 位置，保持与前车的安全距离下，同时保持在高速公路边界内。当前车道前方车辆的行驶速度比自车快，因此没有立即发生碰撞的危险。自车 x 位置会平稳增加，以保持安全行驶和确保平稳的速度转换。我已经减少了未来 6 个点的预测 y 位置，以便安全地移动到左边车道，同时保持行驶在高速公路安全边界内。自车目前的速度和加速度表明换到左侧车道后它将继续向同一方向运动。” RL 模型提供了高层次的策略方向，而 LLM 将这些策略方向转化为具体的操作路径，PID 控制器则执行这些操作路径以实现车辆的实际运动。

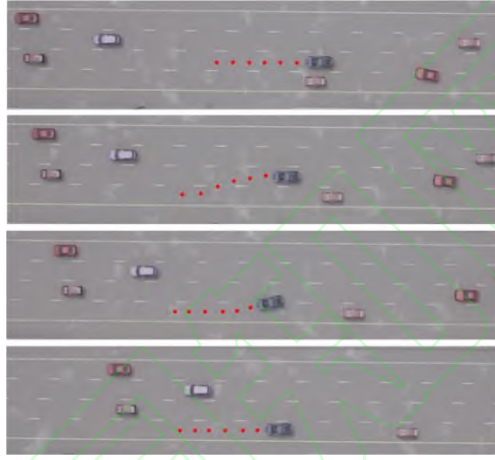


图 3 高速公路变道过程

Fig.3 Highway lane-changing process

由于经验库中没有十字路口相关决策的经验，基于高速公路驾驶经验以及由人工和 LLM 生成的 20 条十字路口驾驶经验，对十字路口场景决策性能展开测试。图 4 展示了 SDLLMRL 模型在复杂交通场景中的决策流程，其中涵盖直行通过十字路口与右转操作两个驾驶场景，包括第 21 秒-36 秒的决策过程。

（1）直行通过带红绿灯的十字路口

当自车并直行通过十字路口绿灯状态时，RL 智能体依据当前状态给出保持当前车道的高级动作指令。LLM 通过检索记忆模块中与当下场景相似度最高的历史经验，对未来 6 个轨迹点进行预测。一旦检测到横向白色车辆接近，LLM 会对轨迹点进行调整，以增大横向安全距离，确保自车能够以安全速度通过路口。PID 控制器依据 LLM 生成的轨迹点，动态调节车辆的加速度与转向角度，保障车辆沿预定路径行驶，与此同时，通过平稳驾驶奖励项提升乘坐舒适性。

（2）右转通过带红绿灯的十字路口

自车进入右转车道并需完成右转动作时，需要应对车道线变化、横向加速度约束以及与直行驶车辆的交互问题。RL 输出右转的高级指令，LLM 结合当前车道线位置、横向速度等信息，预测右转轨迹。通过减小 y 坐标值，确保车辆能够平滑切入右转车道，同时满足最小横向安全距离要求。利用统一风险指数对右转风险进行评估，若超过阈值，系统将触发制动或减速策略，直至风险降低。LLM 的推理包括：安全右转需减小 y 坐标以保持在车道中心，为决策提供了清晰依据，有效解决了 RL 的黑盒问题。记忆模块通过动态更新经验库，使系统在直行、右转等复杂操作中持续优化策略。在完成一次通过十字路口的操作后，系统会依据驾驶结果对记忆库中经验的效用予以评估，系统将这些经验存储至记忆库中，并采用动态阈值淘汰策略，确保记

忆库中存储的均为高价值驾驶经验。

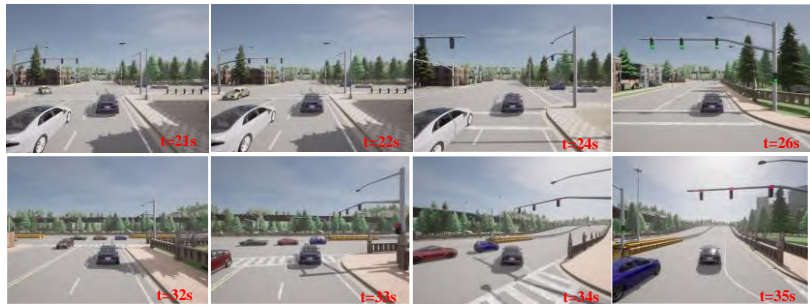


图 4 通过十字路口与右转过程

Fig.4 Through the intersection and the right turn process.

结合 HDD 提供的在高速公路上不同时间的车辆行为。利用嵌入在 LLM 中的上下文知识和有意义的先验知识、交通对象和人类驾驶行为相关的知识。SDLLMRL 结合了 RL 的全局策略优化能力和 LLM 路径规划和解释能力，轨迹预测基于 RL 的高层次决策，还考虑当前的环境状态和历史轨迹数据，从根本上保留了自然语言的可解释性，并增加了不同场景决策推理的潜力。

2.5 记忆模块消融分析

SDLLMRL 利用 LLM 的经验改进了错误的决策过程。然后将修改后的决策结果添加到驾驶员智能体的记忆模块中，作为驾驶经验，为随后的决策过程提供少量经验实例。

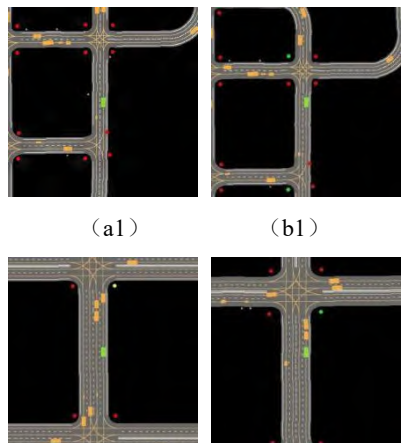
表 2 显示了有记忆 (With Memory Agent, WMA) 和无记忆 (Without Memory Agent, WOMA) 的智能体的在十字路口测试场景的驾驶性能，共测试 50 次，WMA 路线完成度和驾驶分数达到 89.21% 和 60.12，驾驶分数比 WOMA 高 37.8，路线完成度比 WOMA 高 21.06。然而，WMA 需要更长的时间做出决策，因为 WMA 需要检索更多相似的记忆项。

表 2 记忆模块消融实验指标对比

Table 2 Comparison of indicators in the ablation experiment of the memory module

指标	WMA	WOMA
路线完成度 (%)	89.21%	68.15%
平均决策时间 (s)	2.15	1.14
驾驶分数	60.12	34.53

为了更好地说明记忆模块对智能体的影响，实验选择两个十字路口场景进一步分析。如图 5 所示，在鸟瞰图视角下两个智能体在接近路口和在跟车风险升高的情况下的决策。如图 5(a1) 所示，WOMA 认为它必须停止，它在距离路口相当远的地方停了很长一段时间，最终导致驾驶任务因超时而失败。如图 5(b1) 所示，在记忆的引导下，WMA 准确的评估出在有红绿灯变红的情路下可以贴近停车线停车等待，从而顺利到达目的地。



(a2) (b2)

图 5 智能体通过十字路口与跟车 WM 和 WOM 决策分析

Fig.4 Decisions WM and WM of the agent when passing through intersections and car-following situations

如图 5 (a2) 所示, WOMA 倾向于远距离跟车, 虽然缩小了与前车的潜在碰撞风险, 但是导致多次出现无故停车的状态。相反, 如图 5 (b2) 所示, WMA 识别出前方红绿灯变绿后在执行加速动作之前选择与前车保持较小的安全距离。因此, 决策方法拥有记忆模块有益于决策过程, 使智能体的决策过程更明智, 记忆模块可成为解决可持续学习的主要方法之一。

2.6 不同 LLM 性能比较

SDLLMRL 支持集成不同类型 LLM 开展闭环驾驶仿真任务。为精准评估模型性能, 我们在 Carla Town 05 仿真环境中选取一条 2km 的典型线路作为测试路径, 对 Llama-2-7b、Llama-3.1-8b、ChatGLM-6b、Llama-3.1-70b 及 DeepSeek-R1-Distill-Qwen-7b (DRDT-7b) 驱动的驾驶智能体进行性能对比实验, 每个模型均测试 50 次, 以场景信息的自然语言文本描述作为输入。实验结果如表 3 所示, Llama-3.1-70b 驱动的驾驶员智能体在路线完成度与驾驶分数上表现突出, 分别达到 89.21%和 60.12 分。与 Llama-2-7b、Llama-3.1-8b、ChatGLM-6b、DRDT-7b 相比, 其路线完成度分别提升 29%、12.81%、26.71%和 25%; 驾驶分数也分别提高 6.86、5.71、8.87、4.45。然而, Llama-3.1-70b 的平均决策时间较长, 这是由于其执行推理时需处理大量 Token, 导致计算延迟增加; 而经过 DeepSeek-R1 模型蒸馏优化的 DRDT-7b, 凭借蒸馏技术对推理效率的提升, 成为平均决策时间最短的模型, 展现出轻量化设计在实时性上的优势。

表 3 不同 LLM 驾驶性能对比

Table 3 Comparison of driving performance among different LLM					
指标	Llama-2-7b	Llama-3.1-8b	Glm-6b	DRDT-7b	Llama-3.1-70b
路线完成度	60.12%	76.4%	62.5%	64.21%	89.21%
平均决策时间	1.45	1.87	1.56	0.95	2.15
驾驶分数	53.26	54.41	51.25	55.67	60.12

3 结论

本研究提出了一种名为 SDLLMRL 的自动驾驶安全决策方法。该方法运用预训练的 RL 模型作为高级规划模块, 通过将 RL 模型的输出与当前状态信息相结合, 对交通状况及历史数据进行深入分析, 从而精准预测前车的运动趋势, 最终为自车构建出安全可靠的驾驶轨迹。经联合仿真闭环验证, 得出结论如下:

(1) 多场景安全高效决策: SDLLMRL 融合 LLM 轨迹预测与 RL 策略优化。在高速场景, 路线完成度 79.5%、驾驶分数 59.54, 较传统 RL 分别提升 15.3%、35.12; 十字路口场景, 通过动态调整轨迹点和风险评估, 统一风险指数控制在阈值下, 展现复杂交通流中的安全决策能力。

(2) 记忆模块增强持续学习: 基于 HDD 的动态经验库, 用效用评分筛选更新经验。实验显示, 有记忆模块的模型在十字路口场景路线完成度达 89.21%, 比无记忆模型提升 21.06%, 驾驶分数提高 25.56, 表明经验库更新机制提升了新场景适应能力。

(3) LLM 性能与实时性平衡: 在 Llama-3.1-70b 等 5 种 LLM 对比中, Llama-3.1-70b 路线完成度 89.21%、驾驶分数 60.12 表现最佳, 但决策时间 2.15 秒; 蒸馏优化的 DRDT-7b 模型, 驾驶分数 55.67, 决策时间缩至 0.95 秒, 满足实时需求。DeepSeek 模型经蒸馏优化后决策时间极短, 适用于复杂路况与高动态场景, 仿真发展中可降低算力成本、支撑高并发任务。本文方法通过模块化设计和语言推理弥补了黑箱决策和场景适应的不足, 未来我们将探索与端到端感知模型的深度融合, 构建更完善的自动驾驶系统。

参考文献:

[1] 黄志清, 曲志伟, 张吉, 等. 基于深度强化学习的端到端无人驾驶决策[J]. 电子学报, 2020, 48(09): 1711-1719. [HUANG Z Q, QU Z W, ZHANG J, et al. End-to-end autonomous driving

- decision based on deep reinforcement learning[J]. *Acta Electronica Sinica*, 2020,48(09):1711-1719.]
- [2] DENG Y, ZHENG X, ZHANG M, et al. Scenario-based test reduction and prioritization for multi-module autonomous driving systems[C]//*Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 2022: 82-93.
- [3] 宋晓琳,盛鑫,曹昊天,等.基于模仿学习和强化学习的智能车辆换道行为决策[J].*汽车工程*,2021,43(01):59-67.[SONG X L, SHENG X, CAO H T, et al. Lane-change behavior decision-making of intelligent vehicle based on imitation learning and reinforcement learning[J]. *Automotive Engineering*, 2021,43(01):59-67.]
- [4] PENG Y, TAN G, SI H. RTA-IR: A runtime assurance framework for behavior planning based on imitation learning and responsibility-sensitive safety model[J]. *Expert Systems with Applications*, 2023, 232: 120824.
- [5] 周卫林,王玉龙,裴锋,等.基于分段学习模型的自动驾驶行为决策算法研究[J].*中国公路学报*,2022,35(06):324-338.[ZHOU W L, WANG Y L, PEI F, et al. Decision algorithm for autonomous driving behavior based on piecewise learning model[J]. *China Journal of Highway and Transport*, 2022,35(06):324-338.]
- [6] 李伟东,马草原,史浩,等.基于分层强化学习的自动驾驶决策控制算法[J/OL].*吉林大学学报(工学版)*. [LI W D, MA C Y, SHI H, et al. An automatic driving decision control algorithm based on hierarchical reinforcement learning[J/OL]. *Journal of Jilin University (Engineering and Technology Edition)*.]
- [7] 李传耀,张帆,王涛,等.基于深度强化学习的道路交叉口生态驾驶策略研究[J].*交通运输系统工程与信息*,2024,24(01):81-92.[LI C Y, ZHANG F, WANG T, et al. Signalized intersection eco-driving strategy based on deep reinforcement learning[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2024,24(01):81-92.]
- [8] 唐斌,刘光耀,江浩斌,等.基于柔性演员-评论家算法的决策规划协同研究[J].*交通运输系统工程与信息*,2024,24(02):105-113+187. [TANG B, LIU G Y, JIANG H B, et al. Collaborative study of decision-making and trajectory planning for autonomous driving based on soft actor-critic algorithm[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2024,24(02):105-113+187.]
- [9] 贺正冰.大语言模型在道路交通领域应用：创新与挑战[J].*交通运输工程与信息学报*,2025,23(01):85-92.[HE Z B. Large language models in road transportation:innovations and challenges[J]. *Journal of Transportation Engineering and Information*, 2025,23(01):85-92.]
- [10] 王祥,谭国真,彭衍飞,等.基于语言推理与认知记忆的自动驾驶决策模型[J/OL].*吉林大学学报(工学版)*,1-10.[WANG X,TAN G Z, PENG Y F, et al. Autonomous driving decision-making model based on language reasoning and cognitive memory.[J/OL]. *Journal of Jilin University (Engineering and Technology Edition)*,1-10.]
- [11] FU D, LI X, WEN L, et al. Drive like a human: rethinking autonomous driving with large language models[C]//*2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*. IEEE, 2024: 910-919.
- [12] CHEN X, PENG M, TIU P H, et al. GenFollower: enhancing car-following prediction with large language models[J]. *IEEE Transactions on Intelligent Vehicles*, 2024.
- [13] HUANG Y, SANSOM J, MA Z, et al. DriVLM: Enhancing llm-based autonomous driving agents with embodied and social experiences[C]//*2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024: 3153-3160.
- [14] SHAO H, WANG L, CHEN R, et al. Safety-enhanced autonomous driving using interpretable

sensor fusion transformer[C]//Conference on Robot Learning. PMLR, 2023: 726-737.

