

基于语言推理和认知记忆的自动驾驶决策模型

王 祥¹, 谭国真¹, 彭衍飞¹, 任 浩², 李健平¹

(1. 大连理工大学 计算机科学与技术学院, 辽宁 大连 116081; 2. 清华大学 精密仪器系, 北京 100084)

摘 要: 针对传统自动驾驶系统安全性能不足、学习效率低等问题, 提出了一种可持续学习和理解语言信息的自动驾驶安全决策模型。该模型参考人类驾驶的推理决策和经验积累过程, 以大型语言模型(LLM)作为决策智能体, 将思维链推理、两阶段注意力机制和认知记忆存储与检索整合到驾驶过程的上下文安全学习中; 同时, 采用运动学模块将 LLM 决策转化为可操作的驾驶命令, 实现安全驾驶经验的持续学习。实验结果表明, 本文决策模型相较于基于规则、强化学习和知识的方法, 在安全、效率方面有显著提升, 并具备持续学习和根据人类指令调整驾驶行为的能力, 可为类人自动驾驶提供参考。

关键词: 车辆工程; 自动驾驶; 持续学习; 大语言模型; 思维链推理; 两阶段注意力机制

中图分类号: U495; TP391 **文献标志码:** A **文章编号:** 1671-5497(2025)12-3918-10

DOI: 10.13229/j.cnki.jdxbgxb.20240606

Autonomous driving decision-making model based on language reasoning and cognitive memory

WANG Xiang¹, TAN Guo-zhen¹, PENG Yan-fei¹, REN Hao², LI Jian-ping¹

(1. School of Computer Science and Technology, Dalian University of Technology, Dalian 116081, China; 2. Department of Precision Instrument, Tsinghua University, Beijing 100084, China)

Abstract: To address the issues of insufficient safety performance and low learning efficiency in traditional autonomous driving systems, an autonomous driving safety decision-making model capable of continuous learning and understanding linguistic information was proposed. Referring to the reasoning decision-making and experience accumulation processes in human driving, this model leverages a large language model (LLM) as the decision-making agent, integrating chain-of-thought reasoning, a two-stage attention mechanism, and cognitive memory storage and retrieval into the contextual safety learning of the driving process. Meanwhile, a kinematic module is employed to convert LLM decisions into executable driving commands, enabling the continuous learning of safe driving experiences. Experimental results demonstrate that the proposed decision-making model significantly improves safety and efficiency compared to rule, reinforcement learning, and knowledge-based approaches, and possesses the capability of continuous learning and adapting driving behaviors based on human instructions, providing a reference for human-like autonomous driving.

收稿日期: 2024-05-31.

基金项目: 国家自然科学基金重点项目(U1808206).

作者简介: 王祥(1997-), 男, 博士研究生. 研究方向: 人工智能, 自动驾驶. E-mail: DrWangXiang@mail.dlut.edu.cn

Key words: vehicle engineering; autonomous driving; continuous learning; large language model; chain-of-thought reasoning; two-stage attention mechanism

0 引言

自动驾驶车辆通过集成先进的传感器技术、高精地图、深度学习算法及车辆控制系统,能够实现环境感知、路径规划、自动控制,提供安全、高效、舒适的出行体验。然而,随着技术的不断进步,自动驾驶面临的挑战也日益复杂,其中可持续学习、驾驶安全和指令响应决策能力是确保自动驾驶技术可靠性和安全性的关键。

传统的自动驾驶算法通常采用模块化设计,由单独的组件负责不同任务。模块化方法虽能提供一定的可靠性和安全性,但系统组件间的解耦设计可能导致关键信息在转换过程中丢失,还可能产生冗余计算^[1]。同时,基于规则的决策系统设计和维护难度日益增加;强化学习方法不依赖标记的驾驶数据,而是在现实世界或仿真环境中通过试错学习^[2];端到端自动驾驶通过消除多模块间的集成误差、减少冗余计算,增强视觉和感官信息的表达^[3],同时保证更高效率。但这些系统在决策过程中通常存在“黑箱”问题^[4],使分布外检测(Out-of-distribution detection, OOD)推理和解释变得尤为困难。此外,机器学习方法依赖预先定义的训练数据集,在动态变化的交通环境中适应性不足。而可持续学习技术能使自动驾驶系统从新数据中持续学习并更新模型,从而保持高效性和准确性。自动驾驶系统除了通过冗余设计和多层次安全机制保障可靠性外,还需与其他车辆、基础设施及交通管理系统进行有效协同。

大型语言模型(Large language model, LLM)展现了迈向通用人工智能的初步迹象,在OOD推理、常识理解、知识检索、通过自然语言与人类沟通等方面具有突出能力。这些能力与自动驾驶、机器人技术等领域的的需求高度契合。LLM在机器人操纵、多模态理解、终身学习等领域展现出强大潜力^[5],在上下文理解、答案生成和复杂任务处理方面也表现非凡。将LLM集成到自动驾驶中已引起广泛关注,有效增强了自动驾驶汽车的决策能力^[6]。LLM通过思维链^[7]学习人类驾驶行为和轨迹规划,这种方法使LLM能像人类驾驶员一样理解复杂场景。将LLM集成到决策模块,可

以显著提高用户的信任,并将驾驶经验推广到各种驾驶场景。本文利用LLM的短时学习能力和多源输入理解能力,探索其在持续学习技术中的潜力,助力自动驾驶系统快速、有效地适应不断变化的驾驶环境,实现安全最大化和持续学习的目标。

1 决策框架

1.1 模型架构

基于神经网络的自动驾驶系统缺乏与人类先验知识的直接兼容性,限制了其利用先验知识提高驾驶性能的潜力。为了应对这一挑战,本文提出了一种新的基于语言推理和认知记忆的决策方法(Language reasoning and cognitive memory method, LRCMM)。文本和符号具有内在的逻辑推理、知识检索和人类交流适用性^[8],是发挥LLM能力的绝佳媒介。因此,本文以文本作为统一接口,连接神经网络和经验知识。模型总体结构如图1所示,主要包括:①与智能体交互的仿真环境;②具有回忆、推理能力的决策器;③保存和读取驾驶经验的记忆组件。智能体获取环境信息,查询存储模块的经验并执行决策,收集到的信息和决策进一步用于更新认知记忆。本文采用以安全为中心的方法,利用LLM在面对预测不确定性时的优势,为低级模型预测控制(Model predict control, MPC)制定安全约束。

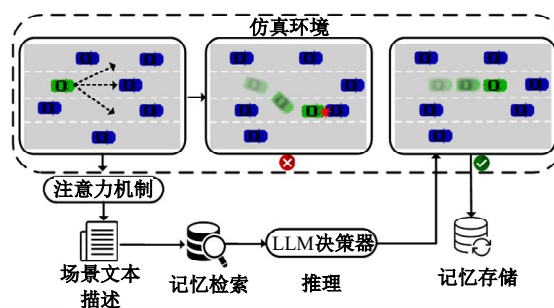


图1 模型总体架构

Fig. 1 Overall architecture of the model

1.2 两阶段注意力机制

态势感知包括对场景的深刻理解和常识性推理。在复杂场景中,模型应能过滤信息,仅关注与决策相关的信息,即智能体应重点关注与规划路

线接近或存在冲突的车辆。车辆状态采用位置、速度等特征描述,设本车状态为 S_0 , N 辆环境车的联合状态可用车辆状态特征 S'_i 表示:

$$S'_i = [x'_i, y'_i, v'_x, v'_y, L_j]^T, \quad i \in [0, N] \quad (1)$$

式中: i 为所选车辆标号; t 为时间; x 和 y 分别为横向、纵向位置信息; v'_x 和 v'_y 分别为横向、纵向速度; L_j 为车道信息, j 为所在车道线。

为适配注意力层输入,统一多车辆的多维度状态特征规格,送入多注意力头堆叠的自注意力层,最终输出唯一与本车相关的特征聚合结果。

自注意力头结构如图 2 所示,其中块 L_q, L_k 和 L_v 表示线性层。本车根据环境筛选车辆子集,发布查询 $Q_a = [q_0] \in \mathbb{R}^{1 \times d_k}$, 其中, d_k 为自注意力机制中查询和键的特征维度。定义 $K_a = [k_0, k_1, \dots, k_N] \in \mathbb{R}^{N \times d_k}$, 其中包含每辆车的特征 k_i 。 q_0 与 k_i 之间的相似性通过 $q_0 k_i^T$ 评估^[9]。通过注意力矩阵收集一组输出值 $V_a = [v_0, v_1, \dots, v_N]$, 每个头的注意力计算公式为:

$$\text{Output} = \sigma \left(\frac{Q_a K_a^T}{\sqrt{d_k}} \right) V_a \quad (2)$$

式中: σ 为 softmax 函数,用于对缩放后的结果进行归一化处理。

所有注意力头的输出与线性层结合,将结果张量添加到残差网络的编码器中。该过程排列不变,虽然排列会改变键 K_a 和 V_a 的顺序,但是会保持二者的对应关系不变。最终结果为值和键的点积,与排序无关。 K_a 和 V_a 由所有车辆连接而来,查询 Q_a 仅由本车发布。

为了确保 LLM 提示信息的准确性,将一阶车

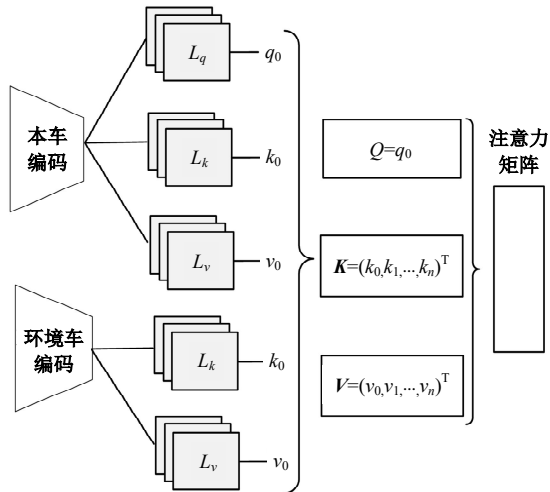


图 2 自注意力头结构

Fig. 2 Architecture of ego-attention head

辆信息作为文本提示信息输入 LLM 中。LLM 通过评估与 $V'_i = \{V'_1, V'_2, \dots, V'_N\}$ 相关的信息,预测这些车辆在 t 时刻的意图,并最终判断它们是否与本车运动产生冲突。在 $[t, t+w]$ 时段内,对周围车辆的每个元素 V'_i 预测其意图 P'_i , 如式(3)所示。 $M'_i \in [0, 1]$ 表示 LLM 判定 V'_i 是否值得关注,判定逻辑如式(4)所示。

$$P'_i = \text{LLM}(S'_i, \dots, S'^{t-w}_i, S'_N, \dots, S'^{t-w}_N, L'_{\text{line}}, E'_{\text{env}}) \quad (3)$$

$$M'_i = \text{LLM}(S'_i, P'_i, L'_j, E'_{\text{env}}) \quad (4)$$

式中: E'_{env} 为 t 时刻环境信息,包括限速、行驶规则等; L'_{line} 为道路信息; P'_i 为 LLM 在 $[t, t+w]$ 时段内预测的 V'_i 意图。

根据 LLM 识别的车辆,为 MPC 创建环境信息矩阵 O'_n , 即:

$$O'_n = \lambda_0 (S'_i, I'_i, L'_j, E'_{\text{env}}) M'_i \quad (5)$$

式中: n 为 MPC 环境信息矩阵的第 n 行; λ_0 为 MPC 计算观测矩阵的算子,确保 MPC 仅关注所选车辆。

最终输出特定感知范围内环境的文本描述,包括车辆的相对位置、相对速度、与其他智能体碰撞时间估计及其他车辆的意图预测。

1.3 决策模块

MPC 可基于车辆运动学模型表示为 $S_0^{t+1} = f(S_0^t, A^t)$, 用于指导本车行动,设本车行动为 A_0^t 。行动指导控制包括加速信号、减速信号、转向信号等,其中 $f: S \times A \rightarrow A$ 为两轮自行车运动学模型。预测模块可预测周围环境车辆 V'_i 在时间步长 t 的未来状态 S'_i 。考虑不确定性,预测的位置结果为特定时间步长上的区间,即 $[\bar{x}'_z, \tilde{x}'_z]$ 和 $[\bar{y}'_z, \tilde{y}'_z]$, 且 $x'_0 \in [\bar{x}'_0, \tilde{x}'_0]$, $y'_0 \in [\bar{y}'_0, \tilde{y}'_0]$ 。由于 MPC 具有视界后退特性,需调用预测模块获取 MPC 安全约束条件的预测结果。

问题 1 证明用于轨迹规划的 MPC 是解决式(6)~式(9)所示优化问题的。运动学模型约束可以保证车辆运动的真实性和可行性;道路边界约束可以确保车辆不超出道路边界,行驶在安全道路范围内;安全约束可以避免与其他车辆发生碰撞。

$$\min_{u_t, u_{t+1}, \dots, u_{t+a}} -x_{t+a} + \sum_{z=t}^{t+a-1} \|A_{z+1} - A_z\|_2 \quad (6)$$

$$S_{z+1} = (S'_z, A_z), \quad z \in [t, t+w] \quad (7)$$

$$y_{\text{left}} \leq y_z \leq y_{\text{right}}, \quad z \in [t, t+w] \quad (8)$$

$$\begin{cases} |x'_0 - \bar{x}'_0| - L \geq 0 \\ |x'_0 - \tilde{x}'_0| - L \geq 0 \end{cases}, \text{Lane}[\bar{x}'_0, \tilde{x}'_0] = \text{Lane}(y_z) \quad (9)$$

式中: y_{left} 和 y_{right} 分别为道路左、右边界; $\text{Lane}(y_z) \in j$ 为指示函数,通过车辆横向位置确定行驶车道,其中 $j=0,1,2,3$ 分别表示“左1”“左2”“右1”“右2”车道。目标函数旨在以最小控制输入最大化纵向速度。

问题2 为了降低问题复杂性,利用LLM的推理能力和常识知识,通过向LLM提供场景文本描述,要求其输出带有MPC约束的决策,以确定MPC的行驶车道。当时间步长为 t 时,LLM条件下MPC在问题1基础上解决如下优化问题:

$$\begin{cases} |x'_0 - \bar{x}'_0| - L \geq 0 \\ |x'_0 - \tilde{x}'_0| - L \geq 0 \end{cases}, \text{Lane}[\bar{x}'_0, \tilde{x}'_0] = \text{Lane}(\text{LLM}) \quad (10)$$

$$(11)$$

问题1比问题2难解,增加的复杂性源于约束条件 $\text{Lane}[\bar{x}'_0, \tilde{x}'_0] = \text{Lane}(y_z)$, 其中 $\text{Lane}(y_z)$ 是不确定的,可以从 $\{0,1,2,3\}$ 中选择。问题1属于混合整数非线性规划问题,在实际应用中通常难以求解。利用LLM的知识推理,删除问题2中的整数决策变量,改为 $\text{Lane}(\text{LLM})$, 由LLM输出决策,将复杂的轨迹规划分解为容易解决的两阶段优化问题。

问题3 LLM从若干相似场景与记忆中筛选最相似的场景和历史行为作为经验,利用在1.2节注意力分配过程中收集的信息及LLM推理结果,定义特征 $F'_i = \{S'_i, P'_i, M'_i\}$ 表征与 V'_i 相关的信息,则有:

$$C' = \text{LLM}(C'_1, \dots, C'_i, L'_j, E'_{\text{env}}) \quad (12)$$

式中: C' 为选择情况,LLM根据选择情况提供加速、减速和保持速度的动作指导 A'_0 。

同时,将场景描述和安全决策动作存储到向量数据库,作为认知记忆。

$$A'_0 = \text{LLM}(F'_1, \dots, F'_k, L'_j, E'_{\text{env}}, C') \quad (13)$$

1.4 认知记忆

人类驾驶员凭借驾驶常识和经验完成驾驶决策,如遵守当地交通规则、借鉴类似情况下的驾驶经验等^[10]。然而,这种能力不适用于传统的感知-预测-规划框架,而本文通过认知记忆的存储和读取解决该问题。对于每个驾驶场景,系统以收集到

的环境信息作为查询,在认知记忆中搜索类似的过往经验,以辅助决策。认知记忆由人类知识和过往驾驶经验构成,包含常识记忆和经验记忆。

常识记忆涵盖在道路上安全驾驶需要的基本知识,如交通规则、危险行为识别等。常识存储器基于文本构建,且支持自由配置,用户可以向存储器中写入不同类型的知识,以定制适配不同驾驶条件的决策信息。经验记忆包含一系列过往驾驶场景和决策,其中每个场景都由当时的环境信息和1.3节所述的驾驶决策组成。通过检索最相似的经验并参考当时的驾驶决策,系统能够做出更安全、高效的驾驶决策。

本文采用两阶段记忆搜索方法查询历史记忆。第1阶段搜索受到向量数据库的启发,将输入查询和记忆库中的每条记忆编码为 Embeddings, 然后通过嵌入空间中的K近邻检索(K-nearest neighbor, KNN) Top-K个相似记录。由于场景的多样性,搜索受到编码方法的限制,导致泛化能力不足。第2阶段采用基于LLM的模糊搜索方法,如图3所示。

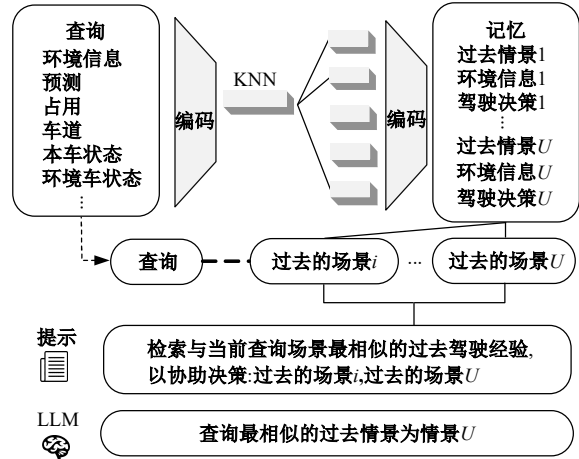


图3 两阶段记忆查询

Fig. 3 Two-stage memory query

在第1阶段,对每个过去场景的本车状态 $e_u \in \mathbb{R}^{n_e}$ 、任务目标 $g_u \in \mathbb{R}^{n_g}$ 和历史动作 $a_u \in \mathbb{R}^{n_a}$ 进行矢量化,为每个场景 u 生成矢量化键 $q_u \in \mathbb{R}^{n_e + n_g + n_a}$ 。经验记忆中 U 个过去场景共同构建张量 $K_s \in \mathbb{R}^{U \times (n_e + n_g + n_a)}$, 即:

$$K_s = \{[e_u, g_u, a_u], u = \{1, 2, \dots, U\}\} \quad (14)$$

将查询场景矢量化为:

$$Q_s = [e, g, a] \in \mathbb{R}^{(n_e + n_g + n_a)} \quad (15)$$

计算查询场景 Q_s 与过去场景 K_s 之间的相似

度分数 $\text{Score} \in \mathbb{R}^U$, 即:

$$\text{Score} = \mathbf{Q}_s \mathbf{A} \mathbf{K}_s^T \quad (16)$$

式中: $\mathbf{A} = \text{diag}(\lambda_e, \lambda_g, \lambda_a) \in \mathbb{R}^{(n_e+n_g+n_a) \times (n_e+n_g+n_a)}$ 。

最终选取相似性得分最高的 Top-K 个样本, 作为第 2 阶段搜索的候选样本。

在第 2 阶段, 利用基于 LLM 的模糊搜索, 以自然语言文本作为统一接口, 使环境信息与人类知识保持一致, 从而提高系统兼容性。以第 1 阶段选取的 Top-K 个历史驾驶场景的自然语言描述作为提示信息输入 LLM, 其任务是理解这些场景的文本描述, 并确定与查询场景最相似的过去驾驶场景。所选场景对应的决策信息也被检索, 为当前决策提供参考。通过向量搜索方法和基于 LLM 的模糊搜索方法, 可以有效地检索到最相似的过往驾驶经验, 而这些过往经验和决策可对当前决策过程提供有力支持。

1.5 思维链推理

LLM 作为推理引擎, 对车辆当前状态、环境信息、过往经验、目标和常识进行全面分析和推理, 做出下一步决策。这一复杂过程需要精细化的处理策略, 使 LLM 在输出最终决策前, 可以通过生成一系列中间输出, 将复杂问题拆解为若干更易处理的子问题, 并逐步解决。LLM 通过生成中间步骤, 可在每个阶段进行检查和调整, 从而减少直接生成最终决策可能引发的错误处理。这种分步推理方法适用于动态驾驶场景, 因为这些场景本质上具有高度复杂性和变化性。

若没有中间步骤, 则 LLM 可能生成不安全的决策。因此, LLM 作为决策者的输出被严格约束为低级 MPC 以选择目标车道和具体驾驶动作。这些低级决策包括目标车道选择, 以及车辆加速、减速、右转、左转等控制动作(见算法 1), 旨在确保车辆在复杂动态环境中做出安全且准确的动作响应。

1.6 指令指导学习安全驾驶行为

出于安全考虑, 本车的控制输入必须经过验证模块的安全验证, 验证结果将返回给 LLM, 用于重新评估行为决策。不可行的 MPC 将导致潜在的不安全规划路径点, 将这些路径点反馈给 LLM, 以重新评估场景并重新生成另一种行为, LLM 的语境安全学习如图 4 所示。由于 LLM 存在固有的幻觉缺陷, 有时决策模块生成的驾驶行为可能是不可行的, 因此不能保证安全性。在这

算法 1 思维链推理

1. while 车辆正常行驶 do
2. 第一阶段采用 h 个头的注意力机制收集环境信息
3. 环境信息 $S_i^t = [x_i^t, y_i^t, v_i^x, v_i^y, L_{ij}]^T, i \in [0, N]$
4. 将存储信息转化为文本并输入 LLM, 开始第 2 阶段注意力机制
5. 第二阶段利用 LLM 识别第 1 阶段需要注意的车辆
6. 输出 output 并解码
7. 生成场景文本信息
8. if 场景信息! = null
9. 获取场景的自然语言描述
10. 将场景信息编码为向量
11. 计算查询场景和历史场景的相似度分数 $\text{Score} = \mathbf{Q} \mathbf{A} \mathbf{K}^T$
12. 选取相似性 Top-K 个样本, 作为第 2 阶段检索的候选样本
13. LLM 理解场景文本描述, 确定与查询场景最相似的历史场景
14. 将场景、记忆和行为信息组合成提示
15. Return(提示信息)
16. else
17. 根据历史经验执行操作
18. end if
19. 对提示信息进行推理
20. 决策 = LLM(提示信息)
21. 将决策解码为具体的行动, 通过运动学模型执行行为转换
22. end while

种情况下, 设计自动驾驶系统切换到故障安全模式, 其核心行为为保持当前车道行驶, 并通过刹车或者减速维持与前车的最小安全距离。安全作为最高优先级, 车辆可以应用保守的硬制动方式降低速度。



Fig. 4 In-context learning guiding decision-making

2 仿真实验设计

2.1 仿真环境及车辆运动学模型

实验基于 HighwayEnv^[11] 搭建了高速公路和十字路口仿真环境。在高速公路场景中, 本车的任务是在确保不与附近车辆相撞的前提下实现高

速行驶。在十字路口场景中,本车的任务是安全通过繁忙的十字路口,且不发生碰撞。为了控制车辆,HighwayEnv 提供了 5 种元操作,包括保持当前车道和速度、左变道、右变道、加速和减速,车辆在每个时间步只能执行其中一个动作。实验环境配置:操作系统为 Ubuntu20.4;运行内存为 64 GB;GPU 为 RTX 4090S;Pytorch 版本为 1.13.0。

车辆运动学模型考虑了位移、速度、角度等运动学状态,用于建立控制参数与车辆在状态空间中行为的映射关系。运动学模型适用于轮胎与地面摩擦近似恒定的运动分析场景。该模型主要依赖车辆运动过程中的几何关系,不涉及受力影响。Chang 等^[12]通过实车实验验证,即使车辆速度高于正常情况,动力学模型和运动学模型之间也不存在明显的路径跟随误差。考虑到动力学模型的因素复杂性和计算时间成本,本实验采用运动学模型,如图 5 所示。

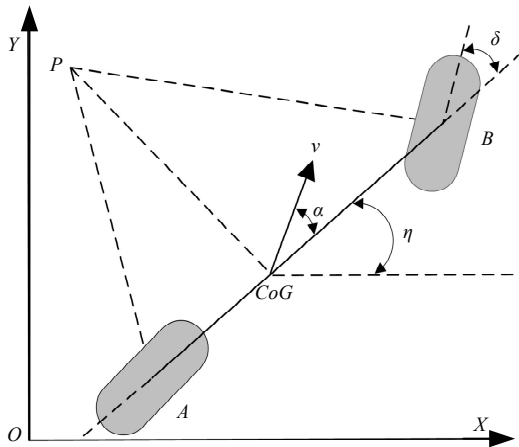


图 5 车辆运动学模型

Fig. 5 Vehicle kinematics model

如图 5 所示, XOY 为公路环境的全局坐标系。其中,车辆质心位于 CoG 点; P 为转向曲率圆心;两个后轮由 A 点(中央后轮)表示;同理 B 点定义为中央前轮; α 为车辆侧偏角; δ 为车轮转角; η 为车辆横摆角。定义前轮、后轮距车辆质心的距离分别为 L_a 和 L_b , 轴距为 $L = L_a + L_b$, 建立车辆运动学模型为:

$$\begin{cases} \dot{x} = v \cos(\eta + \alpha) \\ \dot{y} = v \sin(\eta + \alpha) \\ \dot{\eta} = \frac{v}{L} \cos \alpha \tan \delta \\ \alpha = \tan^{-1} \left(\frac{L_b}{L} \tan \delta \right) \end{cases} \quad (17)$$

2.1.3 环境车模型设置

为使环境车辆能对周围交通形势的变化做出相应反应,采用智能驾驶员模型(Intelligent driver model, IDM)对其进行控制^[13],其数学模型为:

$$\text{acc}_{\text{target}} = \text{acc}_{\text{max}} \left[1 - \left(\frac{v_y}{v_{\text{target}}} \right)^{a_c} - \left(\frac{d_{\text{target}}}{d} \right)^2 \right] \quad (18)$$

$$d_{\text{target}} = d_{\text{min}} + \left(t_{\text{min}} + \frac{\Delta v}{10} \right) v_y \quad (19)$$

式中: $\text{acc}_{\text{target}}$ 为目标加速度; acc_{max} 为最大加速度; v_y 为车辆的纵向速度; v_{target} 为车辆纵向期望速度; a_c 为加速度指数; d_{target} 为环境车的安全纵向车距阈值; d_{min} 为车辆纵向最小距离; t_{min} 为车辆最小碰撞时间; Δv 为与前车的相对速度。

2.2 Llama2 部署

调用 Llama2 模型的 API 需申请并获批后才能使用,否则 API 调用会返回错误状态码。Llama2 提供了 7B、13B 和 70B 三个版本,包括预训练版本和针对对话场景的微调版本。Llama2 基于 2T token 进行训练,将上下文长度从 2 048 扩展到 4 096,从而能够理解更长的文本,并在多个公开基准测试中的表现均优于现有开源模型。通过高质量数据微调和基于人类反馈的强化学习训练, Llama2 系列模型具有较高的可靠性和安全性,其数据流如表 1 所示。

表 1 Llama 2 数据流

Table 1 Llama 2 data flow

数据流	参数	类型	说明
输入	model	string	llama2-7b-chat-v2
	messages	list	输入的内容
	result_format	string	用户返回的内容类型
输出	request_id	string	llama2-7b-chat-v2
	output	list	调用结果信息
	usage.input_tokens	int	用户输入文本转换为 Token 后的长度
	usage.output_tokens	int	模型生成回复转换为 Token 后的长度

2.3 评价指标

实验采用碰撞次数、交通流量、行驶效率、惩罚系数、任务成本等指标衡量驾驶行为的安全性和效率。交通流量通过速度差因子 ξ 计算,其表达式为:

$$\xi = \frac{1}{N} \sum_{i=1}^N (v_i^{\text{max}} - v_i) \quad (20)$$

式中: v_i^{max} 为车辆最大速度; v_i 为车辆当前速度。

以本车到达目标区域所需的时间作为本车行驶效率指标。惩罚系数根据目标车辆与本车的距离及目标车辆的减速情况计算,具体计算公式如下:

$$P_D = w_D \sum_{i=1}^N f_D(\gamma) \quad (21)$$

$$P_d = w_d \sum_{i=1}^N f_d(d_0 - d_i, \gamma) \quad (22)$$

$$f_D(\gamma) = \begin{cases} 0, & \gamma \leq 0 \\ \gamma, & \gamma > 0 \end{cases} \quad (23)$$

$$f_d(d_0 - d_i, \gamma) = \begin{cases} 0, & \gamma \leq 0 \\ d_0 - d_i, & \gamma > 0 \end{cases} \quad (24)$$

式中: $\gamma = D_i - D_0$, 其中 D_i 为减速度; P_D 为减速惩罚; P_d 为距离惩罚, 较小的跟车距离和较高的减速度均为不安全的驾驶行为; d_i 为与前车的距离; $w_D = 100$; w_d 为调节因子; $D_0 = 1 \text{ m/s}^2$; $d_0 = 50 \text{ m}$ 。

惩罚因子针对每辆车单独计算, 最终汇总得到总惩罚分数。对上述指标加权计算, 得到成本函数 C_n 为:

$$C_n = 0.3 \times \xi + 0.1 \times t + 5 \times P_D + 0.5 \times P_d \quad (25)$$

成本评估能够全面反映自动驾驶汽车系统的性能, 成本越低表示性能越好。

2.4 实验条件设置

2.4.1 认知记忆初始化

为了便于智能体构建记忆模块, 需人工初始化内存。先选择若干典型场景, 将场景描述相应的决策过程和推理过程写入认知记忆; 再选择部分场景和错误决策, 将相应的反馈写入记忆模块。在每次迭代过程中, 智能体可从记忆模块检索相关内存记忆做出决策。

2.4.2 注意力机制网络设置

第一阶段注意力机制参数设置如表 2 所示, 包含 1 个编码器[64, 64]、2 个注意力头、1 个解码器[64, 64], 输入大小为 $[\cdot, 6]$, 其中 $\cdot$ 表示输入长度可变, 每个输入元素有 6 个特征。该网络支持可变输入和排列不变量。

2.4.3 对比实验设置

本文选择基于规则的方法(Rule-based meth-

表 2 注意力机制参数设置

Table 2 Attention mechanism parameter setting

结构	取值
输入	$[\cdot, 6]$
层数	编码器[64, 64]; 2 个注意力头; $d_k=32$; 解码器[64, 64]
参数量	3.4×10^4

od, RBM)^[14]、无模型强化学习方法(Model-free reinforcement learning method, MFRLM)^[15]以及知识驱动的自动驾驶决策方法(Knowledge-driven method, KDM)^[8]与 LRCMM 进行比较。

实验设计两种场景, 即高速公路 4 车道场景和没有红绿灯的十字路口场景, 每种场景分别设置 100 回合, 统计 4 种方法的决策效果。为了模拟不同车流量, 实验设定高速公路场景和十字路口场景的每个仿真回合中环境车辆数量分别为 20 辆和 15 辆。为模拟环境车辆的随机行为, 添加车辆时采用不同驾驶风格的车辆类型, 包括激进、保守及平衡 3 种驾驶员风格, 并通过随机配置环境中的 IDM 模型参数生成差异化驾驶行为。

实验过程中对碰撞案例进行统计, 若车辆不能在指定 40 s 内到达目标区域或行驶过程中发生碰撞视为任务失败。当车辆在高速公路场景行驶时, 目标区域被定义为该车道的尽头。当车辆在十字路口场景行驶时, 目标区域被设置为目标车道的终点。

3 仿真实验结果及分析

3.1 决策结果分析

决策的定量结果如表 3 所示, 所有指标均越低越好, 最优结果用加粗字体标注。决策过程如图 6 和图 7 所示。RBM 首先采用 A* 算法搜索时空轨迹, 然后采用优化方法平滑轨迹。在密集的交通流中, RBM 发出控制信号需要花费更多的时间, 碰撞率提升, 如图 6(c) 所示。在十字路口场景中, RBM 更容易停车等待其他车辆, 导致通过效率较低, 如图 7(c) 所示。

表 3 决策评价指标

Table 3 Decision evaluation index

场景	方法	C	ξ	t	P_D	P_d	C_n
高速公路环境	LRCMM	0	12.8	26.4	1.35	1.31	13.89
	RBM	4	26.8	19.7	3.25	3.13	27.83
	MFRLM	8	35.2	15.7	3.84	3.45	46.95
	KDM	3	14.6	29.5	2.47	2.31	20.84
十字路口	LRCMM	0	32.8	47.2	1.84	0.91	24.22
	RBM	2	75.1	32.1	4.21	2.64	48.11
	MFRLM	9	68.4	30.5	5.28	1.06	50.50
	KDM	7	42.5	58.5	3.98	3.61	40.31

KDM 采用单注意力机制收集环境信息, 每帧均调用 LLM 给出驾驶动作, 然后由对应的行为智能体完成驾驶动作。在 25 次实验中, KDM 在 10~

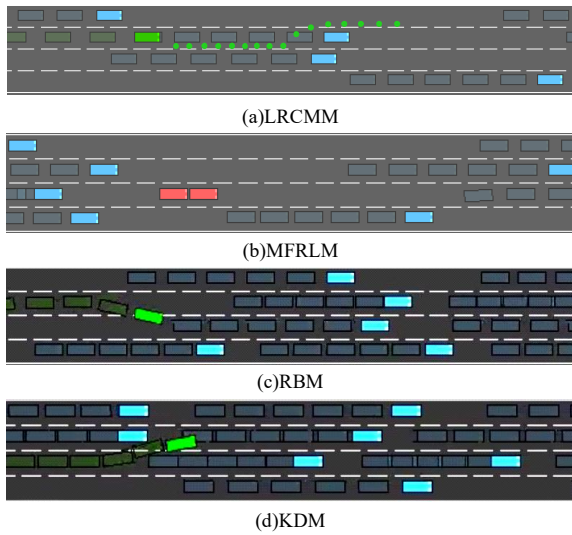


图 6 高速环境决策过程

Fig. 6 Highway decision process

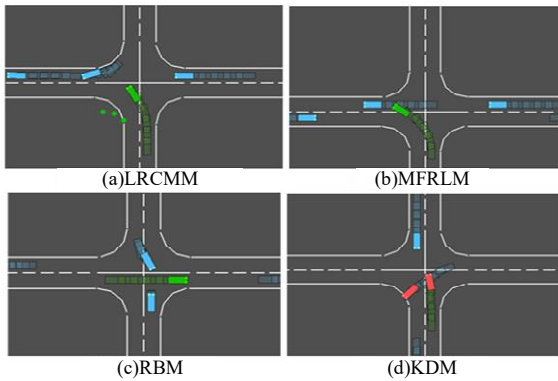


图 7 十字路口决策过程

Fig. 7 Intersection decision process

40步区间内发生了2次碰撞。由于KDM使用简单的启发式规则进行安全检查,未考虑车辆运动约束,因此无法保证行驶安全,如图7(d)所示。

MFRLM由于缺乏安全保障机制,因此碰撞次数最多。LRCMM虽然完成时间略长于MFRLM,但是在安全性、行驶效率、总体成本等驾驶性能上优于其他3种方法,总体成本最低,为13.89。在高速环境下,LRCMM的平均速度为33.8 m/s,而基线方法的仅为25.0 m/s。这是因为LRCMM在LLM条件下的目标函数中最大化了纵向速度,且可以智能地切换车道保持和变道行为以优化整体性能。虽然LRCMM可能导致运行时间轻微增加,但在提高交通流量效率和减少安全处罚方面效果显著。

LRCMM以遵守交通法规为前提,倾向于采取更安全、合理的驾驶行为。在两种场景中,LRCMM在所有指标上都表现良好,降低了事故率

并提高了整体性能。LRCMM首先分析了当前车道的情况,评估周围车辆的速度和位置;然后结合过往驾驶经验和当前驾驶目标,逐步生成一系列中间推理步骤,每个步骤都会经过评估和验证,确保其合理性和准确性,最终生成具体的驾驶操作。如图6(a)和图7(a)所示,LRCMM每个回合平均包含270个LLM决策过程,MPC约束中基于区间的预测包含了周围车辆的未来位置,可行的MPC和不可行时的安全模式保证了系统的安全性。

3.2 可持续学习分析

在两种场景中,分别将记忆项数量设置为10~90进行实验,结果如图8所示。在图8(a)中,随着记忆项数量从10项增加到90项,平均成功步数总体呈增加趋势。初始平均成功步数略低于18,最终上升到略高于27。如图8(b)所示,成功率随着记忆项的增加而增加,从15.6%上升到99.4%。在十字路口场景中,平均成功步数和成功率随着记忆项的增加而增加,表明LRCMM可以提高不同场景下的持续学习能力。当记忆项数

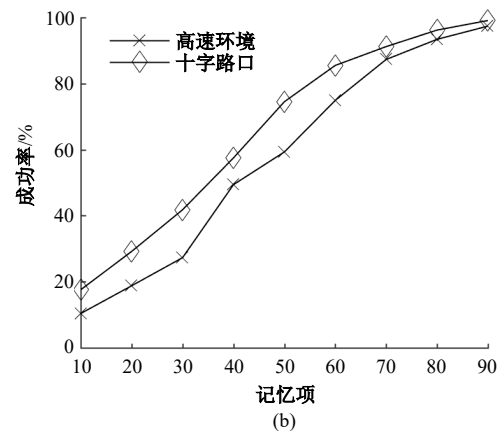
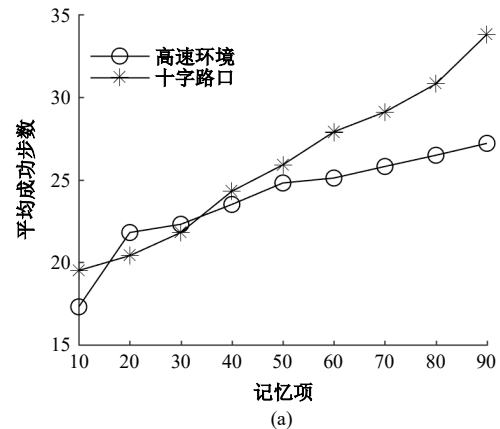


图 8 持续学习结果

Fig. 8 Continuous learning result

量从0增加到10时,平均成功步数增加约18步,说明未添加初始经验的LLM不能直接执行任务,同时证明认知记忆模块对于框架的重要意义。

3.3 消融实验分析

在无信号十字路口和高速公路两种典型驾驶场景下进行消融实验,结果如表4所示。其中,CM为记忆模块,TAM为两阶段注意力机制。

表4 消融实验

Table 4 Ablation experiment

场景	TAM	CM	C	ζ	t	P_D	P_d	C_n
高速环境	×	×	6	69.5	34.3	4.35	3.94	48.00
	✓	×	4	64.7	36.5	3.84	3.12	43.82
	×	✓	3	38.6	28.4	2.37	1.26	26.90
	✓	✓	0	14.7	27.1	1.44	1.29	14.97
十字路口	×	×	2	33.7	48.9	4.21	3.95	38.03
	✓	×	3	32.5	47.8	4.18	3.84	37.35
	×	✓	1	32.1	47.3	3.89	3.52	35.57
	✓	✓	0	31.7	46.2	3.74	3.43	34.55

当不使用TAM和CM时,模型几乎所有指标都有所下降。这一现象源于该方法无法有效感知周围车辆,导致决策效率低。此外,当使用TAM而不使用CM时,模型运行时间指标显著提高,但任务总体成本大幅降低,这表明LRCMM对过去场景的理解和决策在当前决策中发挥着关键作用。

3.4 指令指导驾驶行为分析

当智能体配备思维链时,其性能增强,这凸显了思维链在自动驾驶中的潜力。通过思维链的提示,LLM能够提供清晰地理解上下文,尤其在高速公路场景中,LLM展现出更加可靠的性能。这表明,思维链提示作为引导信号,可以确保算法推理与人类驾驶的实际考虑一致。随着时间的推移,智能体会在类似的驾驶场景中推广这些指导决策过程。LRCMM在上述约束条件下使用思维链提示,能够确保车辆在复杂的环境中行驶,既不危及乘客,也不违反法规。

在高速公路场景下进行指令指导驾驶行为变化实验,结果如表5所示。其中, $\bar{a}$ 为平均绝对值加速度; $\bar{s}$ 为平均绝对值转向角; $\bar{v}$ 为最大速度; $\bar{g}$ 为与前车的最小距离间隔; $\bar{l}$ 为平均换道次数;DMA代表更激进的驾驶行为,DMC代表更保守的驾驶行为,NEC代表没有额外指令。

LLM在执行指令前后,驾驶行为差异明显。当车辆执行DMA指令时,加速度达到 3.5 m/s^2 ,

表5 不同指令下的驾驶行为

Table 5 Driving behavior with different commands

指导命令	$\bar{a}/(\text{m}\cdot\text{s}^{-2})$	$\bar{s}/\text{rad}$	$\bar{v}/(\text{m}\cdot\text{s}^{-1})$	$\bar{g}/\text{m}$	$\bar{l}$
DMA	3.50	0.03	34.7	8.6	7
DMC	0.21	0.01	21.2	41.8	1
NEC	1.87	0.02	28.3	25.6	2

转向输入为 0.03 rad ,获得了更高行驶速度和更小的前车跟车距离,如图9(a)所示。当车辆执行DMC指令时,LLM表现出保守的驾驶风格,降低了加速度,减少了激进的转向输入,最高速度为 21.2 m/s ,并增大了与前车的距离,反映出更安全的驾驶方式,如图9(b)所示。在未给予LLM明确指令的情况下,其驾驶行为处于平衡状态,介于保守驾驶指令和激进驾驶指令之间,如图9(c)所示。实验结果表明,LLM具有基于实时用户反馈动态调整驾驶行为的能力,能够将指令转化为以用户为中心的可操作驾驶策略。这种方法可以提高自动驾驶汽车的用户满意度和安全性。基于实时反馈不断调整驾驶行为(从保守到激进)的能力,证明了持续学习驾驶行为的潜力。随着时间的推移,在驾驶员的持续输入作用下,LLM会微调车辆的驾驶参数,使其与驾驶员的个人偏好紧密结合,从而提供量身定制的安全驾驶体验。

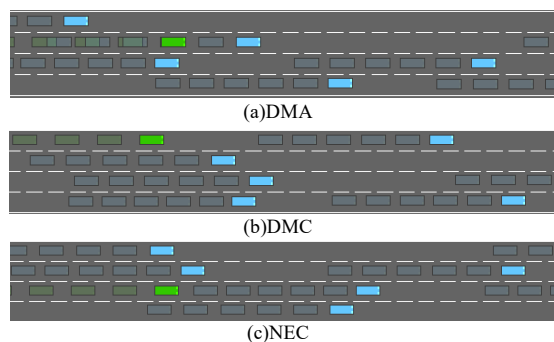


图9 驾驶行为

Fig. 9 Driving behavior

4 结束语

本文提出了一种基于LLM的可持续学习驾驶决策模型LRCMM。该模型集成了两阶段注意力机制、思维链推理引擎、认知记忆和指令引导学习模块。大量实验证明了LRCMM的优越性。与基于规则、强化学习和知识驱动的方法相比,LRCMM在通行效率和安全性方面表现出最佳的驾驶性能,并且在不同驾驶任务间具有较强的适应性。LRCMM可为开发基于LLM的安全、高效、通用自动驾驶系统奠定技术基础。

参考文献:

- [1] 马依宁,姜为,吴靖宇,等. 基于不同风格行驶模型的自动驾驶仿真测试自演绎场景研究[J]. 中国公路报, 2023, 36(2): 216-228.
Ma Yi-ning, Jiang Wei, Wu Jing-yu, et al. Self-evolution scenarios for simulation tests of autonomous vehicles based on different models of driving styles[J]. China Journal of Highway and Transport, 2023, 36(2): 216-228.
- [2] 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]. 吉林大学学报:工学版, 2025, 55(5): 1798-1805.
Li Wei-dong, Ma Cao-yuan, Shi Hao, et al. An automatic driving decision control algorithm based on hierarchical reinforcement learning[J]. Journal of Jilin University (Engineering and Technology Edition), 2025, 55(5): 1798-1805.
- [3] 朱波,张纪伟,谈东奎,等. 基于多源传感器与导航地图的端到端自动驾驶方法[J]. 汽车安全与节能学报, 2022, 13(4): 738-749.
Zhu Bo, Zhang Ji-wei, Tan Dong-kui, et al. End-to-end autonomous driving method based on multi-source sensor and navigation map[J]. Journal of Automotive Safety and Energy, 2022, 13(4): 738-749.
- [4] Zhang Q X, Zhao Y H, Wang Y J, et al. Towards cross-task universal perturbation against black-box object detectors in autonomous driving[J]. Computer Networks, 2020, 180: No. 107388.
- [5] Wang S Y, Zhu Y X, Li Z H, et al. ChatGPT as your vehicle co-pilot: An initial attempt[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(12): 4706-4721.
- [6] Cui Y D, Huang S C, Zhong J M, et al. DriveLLM: charting the path toward full autonomous driving with large language models[J]. IEEE Transactions on Intelligent Vehicles, 2023, 9(1): 1450-1464.
- [7] Kojima T, Gu S S, Reid M, et al. Large language models are zero-shot reasoners[J]. Advances in Neural Information Processing Systems, 2022, 35: 22199-22213.
- [8] 王祥,谭国真. 基于知识与大语言模型的高速环境自动驾驶决策研究[J]. 系统仿真学报, 2025(5): 1246-1255.
Wang Xiang, Tan Guo-zhen. Research on decision-making of autonomous driving in highway environment based on knowledge and large language model[J]. Journal of System Simulation, 2025(5): 1246-1255.
- [9] Peng Y F, Tan G Z, Si H W, et al. DRL-GAT-SA: Deep reinforcement learning for autonomous driving planning based on graph attention networks and simplex architecture[J]. Journal of Systems Architecture, 2022, 126: No. 102505.
- [10] 胡宏宇,张慧珩,姚荣涵,等. L3级自动驾驶接管过程驾驶员情景意识研究[J]. 吉林大学学报:工学版, 2024, 54(2): 410-418.
Hu Hong-yu, Zhang Hui-jun, Yao Rong-han, et al. Driver's situational awareness in takeover process of L3 automated vehicles[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(2): 410-418.
- [11] Nie X T, Liang Y P, Ohkura K. Autonomous highway driving using reinforcement learning with safety check system based on time-to-collision[J]. Artificial Life and Robotics, 2023, 28(1): 158-165.
- [12] Chang M K, Lee S H, Chung C C. Comparative evaluation of dynamic and kinematic vehicle models [C]//Conference on Decision and Control, Los Angeles, CA, USA, 2015: 648-653.
- [13] Treiber M, Hennecke A, Helbing D. Congested traffic states in empirical observations and microscopic simulations[J]. Physical Review E, 2000, 62(2): 1805.
- [14] Xin L, Kong Y T, Li S E, et al. Enable faster and smoother spatio-temporal trajectory planning for autonomous vehicles in constrained dynamic environment[J]. Journal of Automobile Engineering, 2021, 235(4): 1101-1112.
- [15] Li G F, Li S L, Li S, et al. Deep reinforcement learning enabled decision-making for autonomous driving at intersections[J]. Automotive Innovation, 2020, 3: 374-385.