智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于改进深度Q网络的智能网联汽车路径规划

文家燕, 王怡博, 辛华健, 谢广明

引用本文:
文家燕, 王怡博, 辛华健, 等. 基于改进深度Q网络的智能网联汽车路径规划[J]. 智能系统学报, 2026, 21(1): 226-235.
WEN Jiayan, WANG Yibo, XIN Huajian, et al. Intelligent connected vehicle path planning based on optimized deep Q-network[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(1): 226-235.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202502010

DOI: 10.11992/tis.202502010
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20251224.0954.002

基于改进深度 Q 网络的智能网联汽车路径规划

文家燕[1,2], 王怡博[1,2], 辛华健[3], 谢广明[4]
（1. 广西科技大学 自动化学院, 广西 柳州 545616; 2. 广西科技大学 智能协同与交叉应用研究中心, 广西 柳州 545616; 3. 广西工业职业技术学院 广西 南宁 530001; 4. 北京大学 工学院, 北京 100871）

摘　要: 针对非结构环境中的智能网联汽车路径规划问题, 传统的深度 Q 网络 (deep Q-network, DQN) 算法存在规划效率低、收敛速度慢、泛化性差等问题, 本文提出了一种结合注意力机制和经验分类的 DQN 规划方法。通过结合注意力机制设计经验回放池, 通过动态权重分配解决多目标优化冲突, 提升相似环境中的经验利用率, 降低规划时间, 加快收敛; 构建非稀疏奖励约束, 结合交通环境特性优化状态空间, 以便适应多目标场景和实现多场景泛化。仿真表明, 优化后的算法平均规划速度提升了 28.6%, 行进路程较优化前缩短了 25.2%, 且在不同场景下通过载入训练数据, 首次规划成功的耗时缩短了 32.8%。

关键词: 智能网联汽车; 路径规划; 非结构化环境; 注意力机制; 经验回放; 避障; 深度 Q 网络; 深度强化学习
中图分类号: TP183; TP2　　**文献标志码:** A　　**文章编号:** 1673−4785(2026)01−0226−10

中文引用格式: 文家燕, 王怡博, 辛华健, 等. 基于改进深度 Q 网络的智能网联汽车路径规划 [J]. 智能系统学报, 2026, 21(1): 226–235.
英文引用格式: WEN Jiayan, WANG Yibo, XIN Huajian, et al. Intelligent connected vehicle path planning based on optimized deep Q-network[J]. CAAI transactions on intelligent systems, 2026, 21(1): 226–235.

Intelligent connected vehicle path planning based on optimized deep Q-network

WEN Jiayan[1,2], WANG Yibo[1,2], XIN Huajian[3], XIE Guangming[4]
(1. School of Automation, Guangxi University of Science and Technology, Liuzhou 545616, China; 2. The Research Center for Intelligent Cooperation and Cross-application, Guangxi University of Science and Technology, Liuzhou 545616, China; 3. Guangxi Vocational and Technical College of Industry, Nanning 530001, China; 4. College of Engineering, Peking University, Beijing 100871, China)

Abstract: Aiming at the path planning problem of intelligent connected vehicles in unstructured environment, the traditional deep Q-network (DQN) algorithm has problems such as low planning efficiency, slow convergence speed, poor generalization, etc. This paper proposes a DQN planning method combining attention mechanism and empirical classification. The experience playback pool is designed by combining the attention mechanism, and the multi-objective optimization conflict is solved by dynamic weight allocation, so as to improve the experience utilization rate in similar environments, reduce the planning time, and accelerate the convergence; Build non sparse reward constraints, and optimize the state space in combination with the characteristics of the traffic environment, so as to adapt to multi-objective scenarios and achieve multi scenario generalization. The simulation shows that the average planning speed of the optimized algorithm is increased by 28.6%, and the travel distance is shortened by 25.2% compared with that before optimization. In addition, the time for the first successful planning is shortened by 32.8% by loading training data in different scenarios.

Keywords: intelligent connected vehicles; path planning; unstructured environment; attention mechanism; experience replay; obstacle avoidance; deep Q-network; deep reinforcement learning

近些年来, 智能网联汽车 (intelligent connected vehicles, ICVs) 的快速发展很大程度上影响着智能交通系统的发展。这种高度自动化的车辆, 通过集成先进的传感器、控制系统及通信技术, 能够实现与其他车辆 (vehicle-to-vehicle, V2V)、基础设施 (vehicle-to-infrastructure, V2I) 以及行人等道路使用者之间的信息交换和协同操作[1]。通过获

收稿日期: 2025−02−24.　**网络出版日期:** 2025−12−24.
基金项目: 国家自然科学基金 (62541306, 619630060); 广西科技重大专项 (桂科 AA24206054).
通信作者: 辛华健. E-mail: 13659619535@163.com.

取交通信息、环境信息, 可以保障车辆于城市中的高效通行, 减少交通堵塞与交通事故的发生。而复杂的城市道路环境中快速变化的环境与交通信息, 将对智能车辆路径规划提出更高要求。针对车辆的路径规划问题, 已有不少学者展开了探索[1-5], 传统的路径规划算法主要有 A*算法[6-10]、RRT(rapidly exploring random tree) 算法[11-13]、蚁群算法[14-17]、粒子群优化算法 (particle swarm optimization, PSO)[18-20] 等。而针对交通信息的特点, 这类算法难以适用于变量较多的非结构化环境。因此, 针对多变的交通环境, 设计符合实际约束的智能网联汽车路径规划方法亟待探究。

随着机器学习算法的发展, 强化学习逐渐被用于解决智能车辆路径规划问题。但现实生活中车辆处于一个连续空间, 并且利用传感器采集的数据往往是高维信息, 难以用简单的表格形式表示。沿着这个思路, Sutton 等[21] 采用参数化函数作为 Q 值函数的逼近器, 有效缓解了传统表格法面临的状态空间爆炸问题。但需针对特定环境类型定制化设计函数架构参数, 普适性较弱。Mnih 等[22-23] 提出了结合深度神经网络与强化学习的深度 Q 网络算法 (deep Q-network, DQN), 能够直接从高维感知输入 (如屏幕图像) 中学习策略, 通过随机采样和存储过去的交互经验来提高学习的稳定性; 夏雨奇等[24] 采用经验分类的方法, 提出了基于经验分类的 DQN(sorted replay memory DQN, SRMDQN), 将探索获得的经验分类储存, 通过使用现有的有限存储空间和计算量提升神经网络训练效率; 李宗刚等[25] 提出一种角度搜索 (angle searching) 和 DQN 相结合的算法 (AS-DQN), 通过规划搜索域, 控制搜索方向, 减少栅格节点的遍历, 提高路径规划的效率。

尽管基于深度强化学习的路径规划算法在智能网联汽车领域展现出显著优势, 但在实际非结构化场景部署中仍面临多重挑战, 在复杂动态交通环境下的应用常存在以下几个问题:

1) 智能网联汽车在复杂多变的交通环境中, 需通过多模态传感器阵列实时捕获多源异构数据流, 同时受限于毫秒级实时响应窗口, 对路径规划算法的计算效率与决策可靠性提出严苛要求。

2) 现有算法的学习速度仍有提升空间, 且因地图的差异导致需要重新进行训练, 其迁移性和普适性较差。

3) 避障极限距离会随着车辆种类与环境的不同而发生改变, 现有的研究工作大多只针对某一类型环境进行相应约束, 泛化能力不强。

综上, 为解决现有算法学习速度慢且泛化性不足的问题, 本文提出了一种结合注意力机制的改进深度 Q 网络的智能网联车路径规划方法。该方法可在非结构化场景中, 进行探索和学习, 最终根据不同的需求规划出从任意起点到目标点的有效路径, 同时所获得的数据能够应用于不同地图环境。

1　智能网联汽车模型及算法设计

1.1　车辆运动模型

在实际情况中, 车辆的运动学模型可以通过单车模型进行有效模拟[26]。单车模型能够简化车辆的动力学特性, 将车辆的运动视为一个整体, 从而在路径规划和控制策略中提供足够的精度和效率。其模型构建基于以下假设:

1) 忽略车辆垂直方向的运动, 视为一个二维平面上的运动物体, 不考虑车辆在 z 轴的位移;

2) 车辆的前轮作为转向轮并作为直接输入;

3) 车辆的转向轮有相同的转向角与转速, 从动轮具有相同转速, 在运动中, 其状态保持一致;

4) 不考虑空气动力对车辆行驶的影响。

根据上述假设, 车辆的转向轮和从动轮在单独时刻内具有相同的状态, 从而可以将车辆视为单车模型进行分析。单车模型如图 1 所示。

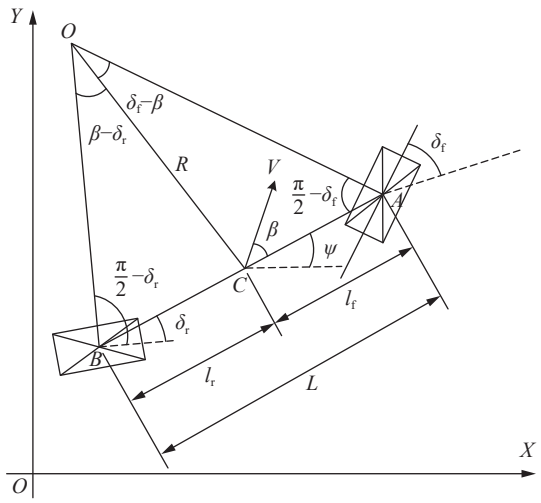


图 1　单车模型图

Fig. 1　Single vehicle model

在图 1 中 C 点为车辆重心, 坐标为 (x_{rb}, y_{rb}); l_r、l_f 为后轮/前轮到车重心点的距离; δ_f、δ_r 为前后轮的转向角; A、B 为单个轮子中心点; 点 O 为车辆的瞬时旋转中心, AO、BO 与转轮方向垂直, β 为车辆速度矢量与车辆纵轴夹角, 滑移角; ψ 为车辆航向角; 航迹角 $\gamma=\psi+\beta$; V 为车辆的线速度。

车辆模型的相关信息可表示为

$$\beta=\tan^{-1}\left(\frac{l_{\mathrm{f}}\tan(\delta_{\mathrm{r}})+l_{\mathrm{r}}\tan(\delta_{\mathrm{f}})}{l_{\mathrm{f}}+l_{\mathrm{r}}}\right)$$

$$x_{\mathrm{rb}}(t+1)=x_{\mathrm{rb}}(t)+v_{\mathrm{rb}}\cos(\psi(t)+\beta)\mathrm{d}t$$

$$y_{\mathrm{rb}}(t+1)=y_{\mathrm{rb}}(t)+v_{\mathrm{rb}}\cos(\psi(t)+\beta)\mathrm{d}t$$

$$\psi_{\mathrm{rb}}(t+1)=\psi_{\mathrm{rb}}(t)+\frac{v_{\mathrm{rb}}(t)\sin\beta}{l_{\mathrm{r}}}\mathrm{d}t$$

式中：$x_{\mathrm{rb}}(t)$、$x_{\mathrm{rb}}(t+1)$表示车辆在t和t+1 时刻 x 轴方向位置，$y_{\mathrm{rb}}(t)$、$y_{\mathrm{rb}}(t+1)$表示车辆在t和t+1 时刻的 y 轴方向位置，$v_{\mathrm{rb}}(t)$表示车辆在t时刻的线速度，$\psi(t)$表示车辆在t时刻的偏航角。

1.2　车辆信息感知

智能网联汽车环境感知示意图如图 2 所示。

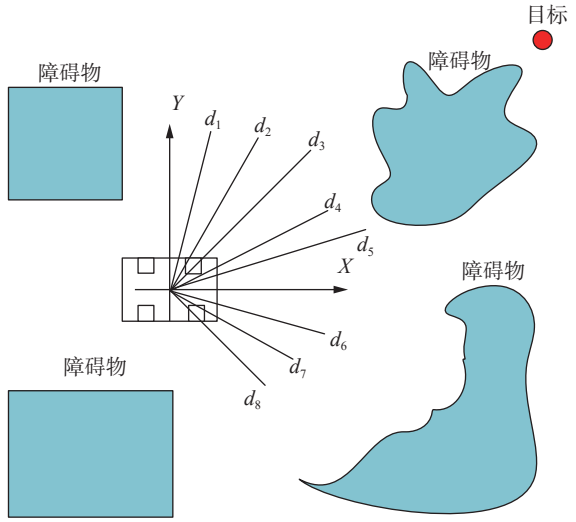


图 2　智能网联汽车环境感知示意

Fig. 2　Schematic of environmental perception of intelligent connected vehicles

图 2 中的障碍物包括规则和不规则障碍物，汽车能够获取目标与自身距离、方位等数据信息。汽车可以对周围的障碍物进行感知并对距离信息$d_{\max}$进行反馈，感知范围内时反馈相应距离，超出感知距离时反馈$d_{\max}$。在汽车与障碍物之间的距离小于安全距离d_{rb}时，表示汽车与障碍物发生碰撞，回合结束 (任务失败)；在汽车与目标距离小于安全距离时，表示达到目标位置，回合结束 (任务完成)。

1.3　DQN 算法设计

自主智能体的学习范式可建模为马尔可夫决策过程 (Markov decision processes, MDP)。本文采用五元组对马尔可夫决策过程进行表示：

$$<S,A,P,R,\gamma>$$

其中：S表示有限状态集，A表示有限动作集，P表示状态转移概率矩阵，R表示奖励函数，$\gamma\in[0,1]$表示折扣因子。在 MDP 中，任意时刻智能体的状态为$s_t(s_t\in S)$，根据所选策略Ψ决定的动作为$a_t(a_t\in A)$，智能体的状态s_t在动作执行之后根据状态转移概率矩阵$\boldsymbol{P}$转移到S，从而得到回报$r_t(r_t\in R)$。在路径规划过程中，车辆不断进行 MDP 直到到达目标位置。根据此决策过程可以对 DQN 算法结构进行搭建，其网络结构如图 3 所示。

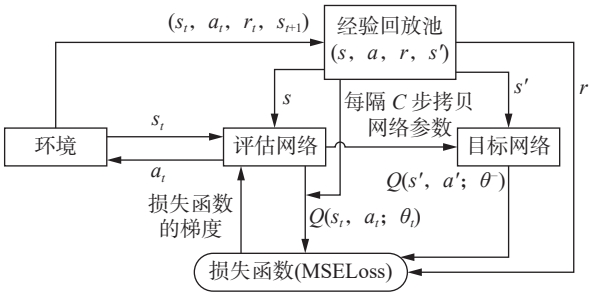


图 3　DQN 算法结构

Fig. 3　DQN algorithm structure

DQN 算法通过定义评估网络$Q(s_t,a_t;\theta_t)$和目标网络$Q(s',a';\theta^-)$来进行智能体的决策、固定参数及更新评估网络与状态，这种方式使 DQN 算法更加稳定，且便于优化。

DQN 算法通过训练网络来近似最优 Q 值函数$Q^*(s,a)$，即对每个状态s和动作a，预测在当前状态下执行该动作后能够获得的期望积累奖励：

$$Q^*(s,a)=\max_{\pi}\mathrm{E}\left[\sum_{t=0}^{\infty}\gamma^t r_{t+1}|s_t=s,a_t=a,\pi\right]$$

若下一个状态s_{t+1}的所有可能动作a_{t+1}均已知，则当前策略的优化目标为使目标函数为其选择最优的动作下取得的值：

$$Q^*(s,a)=\mathrm{E}[r_t+\gamma\max_{a'}Q^*(s_{t+1},a_{t+1})|s_t,a_t]$$

此后使用 Bellman 方程，对评估网络进行迭代更新：

$$Q_i(s_t,a_t)=\left[r_t+\gamma\max_{a'}Q^*(s_{t+1},a_{t+1})|s_t,a_t\right]$$

最终迭代收敛于最优动作价值函数。Q 网络的训练通过评估网络中定义的损失函数来实现：

$$L(\theta)=\mathop{\mathrm{E}}_{s,a,r,s'\sim D}[(y_i-Q(s_i,a_i;\theta))^2]$$

式中：$\mathop{\mathrm{E}}_{s,a,r,s'\sim D}$为期望值，为经验回放缓冲区 D 中采样的状态、动作、奖励和下一状态的期望；$y_i=r_i+\gamma\max_{a'}Q(s'_t,a';\theta^-)$，为第 i 次迭代目标。最终得到损失函数的梯度下降公式：

$$\nabla_\theta L(\theta)=\mathop{\mathrm{E}}_{s,a,r,s'\sim D}[2(y-Q(s_t,a_t;\theta))\nabla_\theta Q(s_t,a_t;\theta)]=$$

$$\mathop{\mathrm{E}}_{s,a,r,s'\sim D}2[r_t+\gamma\max_{a'}Q(s'_t,a';\theta^-)-$$

$$Q(s_t,a_t;\theta)]\times\nabla_\theta(Q(s_t,a_t;\theta))$$

更新参数θ的公式为

$$\theta\leftarrow\theta-\alpha\nabla_\theta L(\theta)$$

其中α是学习率，控制每次参数更新的步长。

DQN 算法将智能体与环境交互获得的信息存储在经验池中

$$< s_t, a_t, r_t, s_{t+1} >$$

式中: s_t表示智能体在t时刻的状态, a_t表示智能体在t时刻状态下将采用的动作, r_t表示智能体在t时刻状态下采用动作a_t后获得的回报, s_{t+1}表示智能体在$t+1$时刻的状态。此后需要判断智能体在t时刻之后回合是否结束。

智能体主要学习过程为从经验池 D 中随机抽取n个样本进行学习, 主网络对经验值 Q 进行预测, 在目标网络中进行计算, 获得损失, 并对主网络和目标网络的权重进行更新。经验池的存在能够提高算法的稳定性, 并且不限制学习方式, 提高了样本的复用率。

2 基于改进 DQN 的路径规划算法

针对智能车辆在非结构化动态交通场景中存在的决策响应时效性不足、策略收敛稳定性差及场景迁移鲁棒性不足等核心问题, 构建了融合多模态交通特征的状态表征空间, 并基于深度强化学习框架设计混合奖励机制, 提出优化后的 DQN 算法。

2.1 状态空间和动作空间的优化

智能网联汽车从环境信息中获得车辆与目标地点的相对位置$[d_x, d_y]$、车辆的速度v_{rb}、偏航角ψ。针对仿真环境中车辆模型与算法模型的泛化性与简化性, 将汽车的移动速度v_{rb}设定为固定值。针对路径规划的问题, 信息的输入一般为车辆当前的位置信息或车辆的网格信息, 在不同的地图中可以通过相应坐标来确定位置。在优化的算法之中, 前 2 个维度表示车辆与重点之间的位置关系用来描述车辆当前位置与目标地点的位置关系, 从而规划车辆的前进方向; 后 6 个维度表示车辆行进方向所代表的值与奖励损失值, 对环境进行描述的同时, 让车辆进行障碍躲避。获得的状态空间表示为

$$V_i = M(p'_{\Lambda}(a_i)), i \in [1,6]$$

$$O = [d_x, d_y, V_1, V_2, V_3, V_4, V_5, V_6]$$

式中: d_x、d_y表示车辆与目标地点之间 x、y 轴的相对距离; $M(\cdot)$表示环境的状态函数, 用于输出当前位置的状态值; a_i表示车辆动作空间的动作值; p'_{Λ}表示采取相应动作之后车辆的位置。

智能网联汽车作为智能体, 通常更倾向于在接近目标地点时选择斜向动作, 而非直线运动。然而, 在实际交通环境中, 车辆需要遵循相应的交通规则, 且变道行为的选择会影响相应路段的交通安全性, 斜向动作并不适合实际情况, 因此, 本文对斜向动作进行限制, 使车辆通过增加转弯次数不断调整位姿以接近目标地点。动作的选择使用ε–贪心策略, 用公式表示为

$$a = \begin{cases} \arg\max\limits_{a} Q(s,a,\theta), & c \geqslant \varepsilon \\ \text{随机}, & c < \varepsilon \end{cases}$$

式中: $Q(\cdot)$ 表示 Q 值网络, θ表示网络参数, s、a表示网络的输入状态与动作, c为 (0,1) 的随机数。ε常随着训练的进行逐渐减少, 本文采用指数衰减, 计算公式为

$$\varepsilon(t) = f + (g - f)\mathrm{e}^{-\lambda t}$$

其中: λ为衰减速率, 控制ε减少的快慢; f为ε的非 0 最小值, 保证模型具有脱离局部最优的随机性; g为ε的最大值, 常为 1; t为训练步数。

2.2 奖励函数的优化设计

奖励函数的设计直接影响智能体在环境中的行为选择, 合理的奖励设计能够有效提升学习效率, 并做出更优的决策。任务最终能否成功与奖励函数的设计密切相关。当强化学习任务难度较低时, 基于任务完成情况来设计奖励函数的思路虽能完成设定的学习任务, 但缺乏泛化能力。在面对复杂任务时, 这种奖励设计方法通常导致学习速度缓慢, 且策略空间探索易陷入局部最优陷阱的情形。

为了提升学习效率与决策的优越性, 本文中提出了一种非稀疏的奖励设定, 总奖励由方向奖励r_{vehicle}、避障奖励$r_{\mathrm{avoidance}}$、目标奖励r_{target}、约束奖励$r_{\mathrm{constraint}}$共同组成。

方向奖励r_{vehicle}的设计公式表示为

$$r_{\mathrm{vehicle}} = \begin{cases} \lambda_1\left(\dfrac{\pi}{2} - |\theta_n|\right), & 0 \leqslant |\theta_n| \leqslant \dfrac{\pi}{2} \\ \lambda_2\left(\dfrac{\pi}{2} - |\theta_n|\right), & \dfrac{\pi}{2} < |\theta_n| \leqslant \pi \end{cases}$$

式中: λ_1表示车辆行进方向和上一动作夹角在$\left[0, \dfrac{\pi}{2}\right]$内时的系数, λ_2表示车辆行进方向和上一动作夹角在$\left(\dfrac{\pi}{2}, \pi\right]$内时的系数, θ_n表示车辆当前行进方向与上一动作方向的夹角。该奖励表示当小车前进方向与上一动作的方向不冲突, 即夹角小于 90°时, 获得最大奖励; 当小车前进方向与上一动作的方向产生冲突时, 夹角大于 90°, 获得负值奖励。常规情况下$0 < \lambda_1 < \lambda_2 < 1$, 这样的设计避免了车辆在行进过程中不能稳定避障和突然掉头产生交通隐患的情况。

避障奖励$r_{\mathrm{avoidance}}$计算公式为

$$r_{\text{avoidance}}=\begin{cases}c_{\text{collision}}, & R_{\text{vehicle}}\leqslant\chi_{\text{obstacle}}\\ c_{\text{convention}}, & R_{\text{vehicle}}>\chi_{\text{obstacle}}\end{cases}$$

式中：$c_{\text{collision}}$表示车辆碰撞到障碍物获得的奖励，R_{vehicle}表示车辆的碰撞半径，χ_{obstacle}表示车辆与最近障碍物之间的距离。当车辆碰撞半径小于最近障碍物之间的距离时，表示车辆发生碰撞，其他情况下获得的奖励值为$c_{\text{convention}}$。

目标奖励r_{target}计算公式为

$$r_{\text{target}}=\begin{cases}c_{\text{target}}, & \chi_{\text{target}}\leqslant R_{\text{target}}-R_{\text{vehicle}}\\ 0, & \chi_{\text{target}}>R_{\text{target}}-R_{\text{vehicle}}\end{cases}$$

式中：c_{target}表示车辆到达目标地点时获得的常数奖励，R_{target}表示目标地点半径。车辆与目标的距离χ_{target}小于车辆碰撞半径的时候表示车辆到达目标地点，获得奖励c_{target}，其他情况下获得的奖励值为 0。

约束奖励$r_{\text{constraint}}$计算公式为

$$r_{\text{constraint}}=\begin{cases}\beta_1(\lambda_1\left(\dfrac{\pi}{2}-|\theta_n|\right)+c_1), & 0\leqslant|\theta_n|<\dfrac{\pi}{2}\\ \beta_2(\lambda_1\left(\dfrac{\pi}{2}-|\theta_n|\right)+c_2), & |\theta_n|=\dfrac{\pi}{2}\\ \beta_3(\lambda_2\left(\dfrac{\pi}{2}-|\theta_n|\right)+c_3), & \dfrac{\pi}{2}<|\theta_n|\leqslant\pi\end{cases}$$

式中：β_1、c_1分别为车辆正向无转向情况下的相应系数与常数奖励，β_2、c_2分别为车辆正向转向情况下的相应系数与常数奖励，β_3、c_3分别为车辆反转行进方向情况下的相应系数与常数奖励。车辆更改转向方向时会影响自身速度，在相同距离下，更少的转弯次数使得车辆平均车速更稳定，能够更快的到达目的地。常规情况下$0<c_3<c_2<c_1<1$，$0<\beta_1<\beta_2<\beta_3<10$，这样的奖励设计能够减少车辆转向次数，使得车辆在避障情况下，获得更优越的结果。

最终的奖励函数计算公式为

$$r=r_{\text{vehicle}}+r_{\text{avoidance}}+r_{\text{target}}+r_{\text{constraint}}$$

相较于稀疏奖励回报方式，本文的奖励函数设计更加合理，并切合实际情况。其中，r_{vehicle}的设计能够减少车辆折返情况出现的次数，使得车辆更快地找到目标；$r_{\text{constraint}}$的设计减少车辆行为方式改变的情况，降低转向影响车辆速度的程度，缩减训练时长并提高训练成果的概率。

2.3　经验学习的优化

经典的 DQN 算法在处理车辆的路径规划问题时，能够将智能体在环境中的状态、动作、奖励等信息进行存储，并随机抽取提高学习效率。但在经验池中存储的数据存在不均匀的问题。在训练初期，能够到达与不能到达目标地点的经验数据数量有较大差别，同时对经验的利用率较低，容易覆盖学习过程中的优异经验，导致学习速度更加缓慢。针对此问题提出将注意力机制与经验分类结合的 DQN(attention mechanism integrated experience replay deep Q-network, AMERDQN) 算法，该算法将车辆于环境中探索获得的数据分类，同时结合注意力机制，对能够在约束条件下完成任务的经验数据进行存储，针对具有不同优越样本的数据进行加权处理并进行特征提取，更新相应的优越级。

侧重注意力的经验分类 DQN 算法经验池结构如图 4 所示。

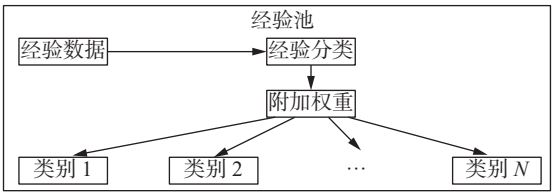


图 4　经验池结构

Fig. 4　Structure of experience pool

针对总奖励值和约束奖励值，结合相应策略特征进行综合评分，计算公式为

$$\mathrm{S}(r,r_{\text{constraint}},\mathrm{Tag})=\omega\cdot r+(1-\omega)r_{\text{constraint}}$$

式中：ω为动态权重，根据约束的违反次数进行动态调整；Tag 为相应的策略特征，做为训练经验存储的区分标签。

在本文中，经验主要含有以下几类：1) 靠近目标且避障半径没有障碍物；2) 靠近目标且避障半径存在障碍物；3) 远离目标且避障半径没有障碍物；4) 远离目标且避障半径存在障碍物；5) 完成任务且步长$\delta_{\text{step}}\in[0,c_{\text{mid}}]$；6) 完成任务且步长$\delta_{\text{step}}\in[c_{\text{mid}},c_{\text{max}}]$。

通过结合注意力机制对训练经验进行相应的评估，评估误差计算公式为

$$V_\delta=\left|r+\gamma\max_{a'}Q_{\text{tar}}(s',a')-Q(s,a)\right|$$

式中：r为执行动作a后获得的即时奖励；γ为折扣因子$(0\leqslant\gamma\leqslant1)$，用于衡量未来奖励；$Q_{\text{tar}}(s',a')$为目标网络对下一状态$s'$的最大 Q 值估计；$Q(s,a)$为主网络对当前状态−动作的 Q 值估计。

通过对状态−动作价值密度进行评估，来判断是否存在相似经验的，增加低密度经验的探索，保证对未充分探索区域的主动学习，评估原理计算公式为

$$V_d=\frac{1}{N}\sum_{i=1}^{N}f((s_i,a_i)\in\mathcal{N}(s,a))$$

式中：$\mathcal{N}(s,a)$为状态−动作空间中以(s,a)为中心的局部邻域；f为指示函数，当(s_i,a_i)属于邻域时取

1, 否则取 0; N为经验池中的样本总数。通过对密度的检测判断是否需要增加采样权重, 避免模型过度关注高密度区域。

针对高价值经验的优先级采样概率计算公式为

$$P(i)=\frac{(S_i+\epsilon)^{\alpha}}{\sum_j (S_j+\epsilon)^{\alpha}}$$

式中: α为温度系数, 用来控制概率分布的尖锐程度, 其数值越大, 高评分经验的优先级越突出; ϵ为一个极小的正数, 防止出现某个经验的评分非常低甚至为 0 时, 其优先级概率不为 0; S_i、S_j为第i、j条经验的综合评分。

相较于经典 DQN 算法, 侧重注意力的经验分类 DQN 算法通过加权分类的方式, 重构经验回放过程。对获取的经验数据进行分类的同时, 进行评估加权, 在训练神经网络时根据不同权重, 抽取相应的样本数据混合后进行训练, 从而提升算法的学习效率。

算法 1　侧重注意力机制的经验分类 DQN 算法伪代码

1) 初始化经验池 D、经验池样本总数为 N

2) 随机初始化评估网络 $Q(s_t,a_t;\theta_t)$、目标网络 $Q(s',a';\theta^-)$

3) 设置神经网络训练时抽取的样本数目 n

4) 设置初始行走的步数 δ_{step}

5) **for** episode $\in [1,M]$ **do**

6) 读取当前状态 s_t

7) **for** $t=1,2,\cdots,T$ **do**

8) **if** $\delta_{\text{step}}>0$ **then**

9) $a=$ random step

10) $\delta_{\text{step}}=\delta_{\text{step}}-1$

11) **else**

12) 以 ε 的概率随机选择动作

13) **else**

14) $a_t=\arg\max_a(Q(s_t,a;\theta))$

15) **end of**

16) 执行动作 a_t 并将 $<s_t,a_t,r_t,s_{t+1}>$ 经过分类、加权后分别存储在经验池 D 中相应部分

17) 每走 δ_{learn} 步随机从经验池 D 中选取 n 个样本, 打乱后训练神经网络

18) 根据损失函数 $(y_i-Q(s_i,a_i;\theta))^2$ 使用梯度下降法更新相应的评估网络 $Q(s_t',a';\theta^-)$

19) 每学习 n_{mid} 步就将评估网络中的参数 θ 赋予目标网络 $Q(s,a;\theta')$

20) **end for**

21) **end for**

3　仿真实验与分析

为验证算法在非结构化环境中的路径规划合理性, 本文基于 Python 架构搭建仿真环境。在该仿真环境中, 智能体能在闭环控制架构下实时生成符合物理约束的运动轨迹, 并通过 Bezier 曲线进行路径平滑处理。本研究主要对优化后的 DQN 算法、DQN 算法、深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 算法和 SRM-DQN 算法在规划结果、仿真数据参数等方面进行比较, 从而验证优化后的算法在车辆路径规划问题中的可行性与优势。

3.1　实验环境搭建与参数设计

本文智能网联汽车的路径规划的仿真实验地图如图 5 所示。

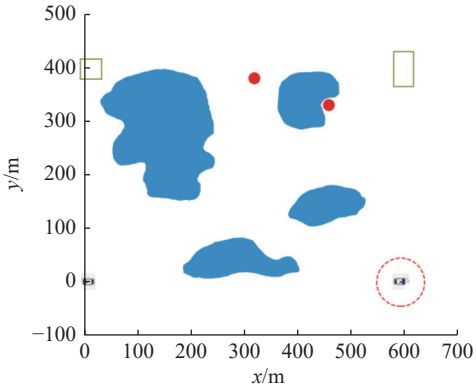


图 5　仿真实验地图

Fig. 5　Simulation experiment map

环境中存在多个不规则形状的蓝色静态障碍物以及在限定范围内随机行动的红色动态障碍物, 据此构成车辆的非结构化仿真环境; 红圈为车辆的避障敏感范围, 存在极限碰撞半径; 绿色方框为目标地点识别范围, 到达绿色方框内视为任务完成。

本次实验构建 650 m×450 m 的矩形区域, 本文仿真环境参数设计如表 1 所示。

表 1　仿真物理环境参数

Table 1　Simulation physical environment parameters

参数	数值
车辆线速度 v_{vehicle}/(m/s)	5
车辆碰撞半径 R_{vehicle}/m	2
车辆碰撞极限半径 R_{vmid}/m	1
车辆前轮与中心距离 f_{q}/m	1.3
车辆后轮与中心距离 f_{h}/m	1.3
车辆感知最远距离 $d_{\max}$/m	30
目的地范围半径 R_{target}/m	10
环境刷新间隔/ms	100

AMERDQN 算法参数设计如表 2 所示。

表 2　算法参数

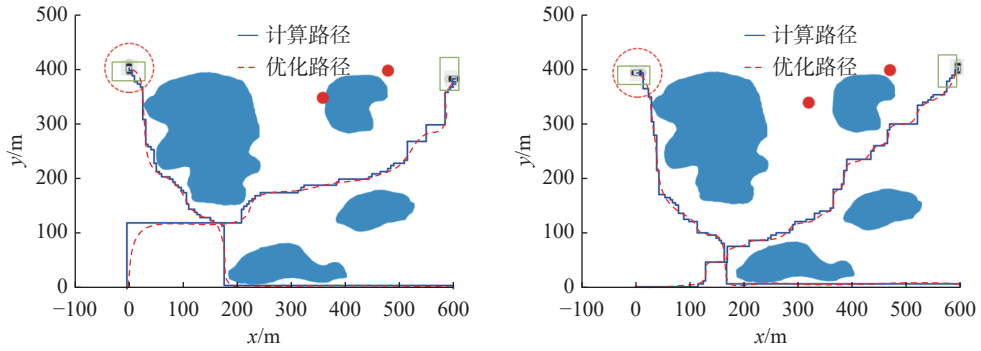
Table 2　Algorithm parameters

参数	数值
方向奖励参数λ_1	0.1
方向奖励参数λ_2	0.7
避障奖励$r_{\text{avoidance}}$	−100
目标奖励r_{target}	200
强化学习折扣系数γ	0.95
神经网络学习率l	0.000 3
初始随机步数δ_{step}	20
学习间隔n_{learn}	5
目标网络赋值间隔n_{w}	10
经验池样本总数N	500 000
训练抽取样本数目	64
最大训练回合数目M	2 000
每回合最大运行步数$T_{\max}$	500

其中，方向奖励参数用于衡量智能体在接近目的地时所获得的奖励，从而防止智能体在避障时出现原地转圈的现象；碰撞奖励和目标奖励的参数设置确保了系统的稳定收敛；通过合理设置折扣系数和贪婪因子，保证智能体在学习过程中能够实现探索−利用平衡；学习率、学习间隔、赋值间隔、抽取样本的数量以及经验池大小等参数共同保障 Q 值迭代稳定性；初始随机步数的参数设定构建初始状态−动作分布的完备性覆盖；为了防止训练陷入死循环，最大训练回合数和最大运行步数的设定确保在必要时能够及时终止当前回合的训练。

3.2　仿真实验结果与分析

根据上述车辆的物理环境参数和 AMERDQN 算法参数在地图中进行仿真，结果如图 6 所示。



(a) DQN 算法　　(b) AMERDQN 算法

图 6　DQN 算法与 AMERDQN 算法路线对比

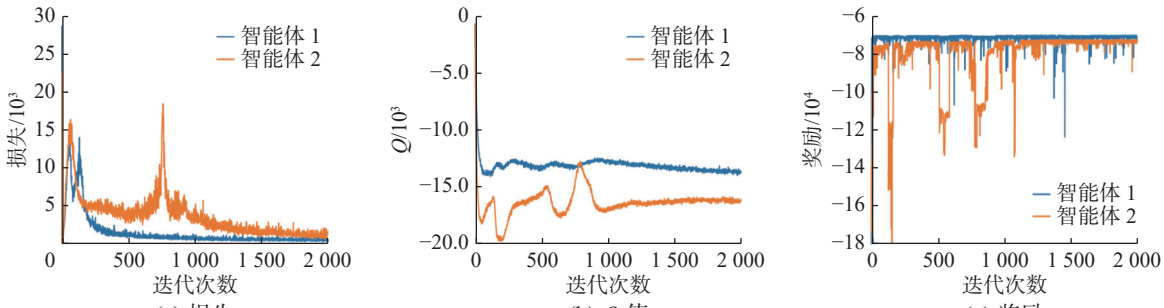
Fig. 6　Comparison between DQN algorithm and AMERDQN algorithm

对比不同算法在相同约束环境中的结果，本研究注意力机制与深度 Q 网络模型相结合的方法所规划的路径表现更加平滑，具有更少的转弯次数，符合约束要求。与传统 DQN 算法相比，本文提出的 AMERDQN 算法在相同训练参数下能在更短时间内规划出优质路径。各算法规划出的路径结果相关参数如表 3 所示，其中的数据均为多次训练的均值。DQN 算法和 AMERDQN 算法在环境中的训练结果如图 7、图 8 所示。

表 3　算法规划结果参数对比

Table 3　Comparison of parameters of algorithm planning results

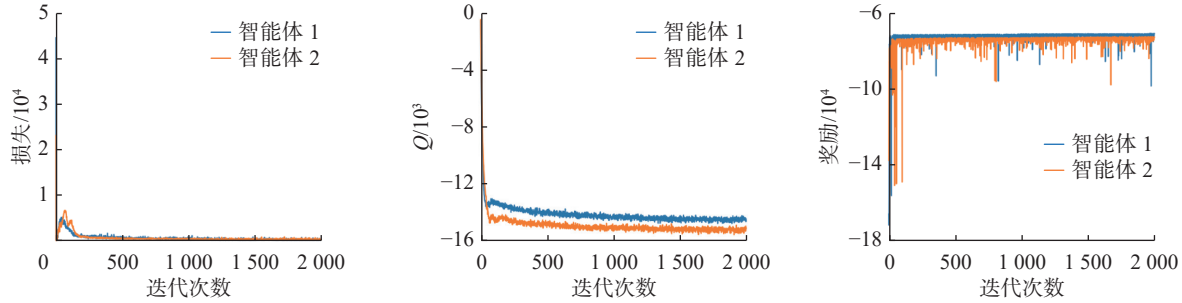
算法	路径长度/m	平均步长	转弯次数
DQN	932.23	357	67
DDPG	902.20	341	62
SRMDQN	895.73	288	56
AMERDQN	890.37	255	51



(a) 损失　　(b) Q 值　　(c) 奖励

图 7　DQN 算法训练结果

Fig. 7　DQN algorithm training results



(a) 损失　(b) Q 值　(c) 奖励

图 8　AMERDQN 算法训练结果

Fig. 8　AMERDQN algorithm training results

DQN 算法与 AMERDQN 算法在训练过程存在一定的随机性, 图 7 和图 8 所示的结果是在同一环境进行多次训练后的平均数据。实验数据对比分析表明, AMERDQN 算法在智能车辆训练中展现出显著的探索效率优势, 在高效探索环境的同时, 获得更高的平均成功率与平均奖励。

DQN 算法与 AMERDQN 算法在同仿真环境下进行训练, 成功率、回合内路程如图 9、10 所示。

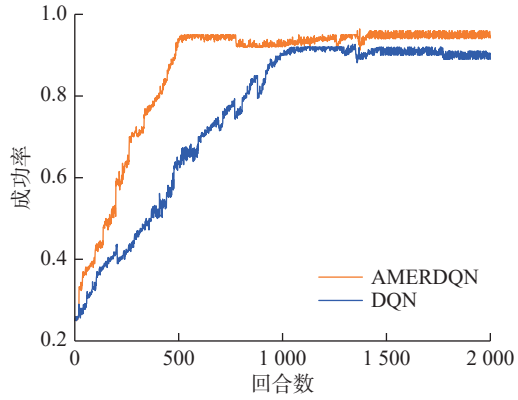


图 9　DQN 与 AMERDQN 同环境中算法成功率

Fig. 9　Algorithm success rate of DQN and AMERDQN in the same environment

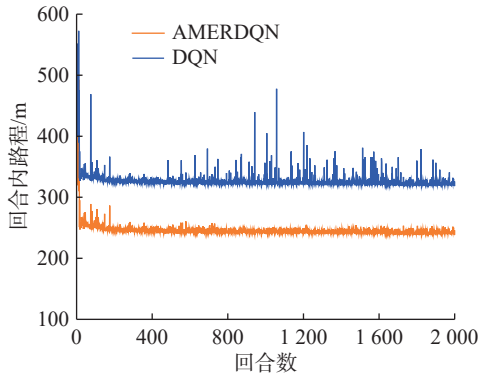


图 10　DQN 与 AMERDQN 同环境中回合内路程

Fig. 10　Path length in round between DQN and AMERDQN in the same environment

由图 9 和图 10 可知, AMERDQN 算法较于 DQN 算法, 平均成功率与收敛时间的表现更加优秀, 且 AMERDQN 算法整体的路程较 DQN 算法更短, 同时训练成果更优。

4　结束语

本文提出非结构化场景下基于改进深度 Q 网络的智能网联汽车路径规划算法。通过改进状态空间和奖励函数增强算法的场景适应能力。提出结合注意力机制的经验回放方式, 对训练过程中获得的经验进行加权分类, 有效提升了算法的学习效率, 使得路径规划速度更快, 决策路径更优质。此外, 通过载入不同环境的训练数据, 有效降低了在当前环境下达到 90% 成功率所需的回合数, 体现了一定的泛化能力。

在未来的研究工作中, 可以探索以下几个研究方向:

1) 协同感知融合。通过多车协同感知信息与多源异构传感器数据融合, 全面感知复杂动态环境, 提升决策效率与准确性, 实现多样地图环境下的优异表现。

2) 终身学习。基于跨场景经验迁移的终身学习机制, 避免算力浪费, 支持车辆在不同地图中持续优化路径规划, 实现高效能自主驾驶。

参考文献:

[1]　杨龙海 , 车婷婷 , 熊月程 , 等. 考虑智能网联车队要素的交通震荡特性研究[J]. 北京交通大学学报, 2024, 48(4): 104−114.
YANG Longhai, CHE Tingting, XIONG Yuecheng, et al. Research on the characteristics of traffic oscillations considering the elements of connected and automated vehicle platoon[J]. Journal of Beijing Jiaotong University, 2024, 48(4): 104−114.

[2]　ZHANG E, MASOUD N. V2XSim: A V2X simulator for connected and automated vehicle environment simulation[C]//2020 IEEE 23rd International Conference on Intelligent Transportation Systems. Rhodes: IEEE, 2020:

1−6.

[3] 马庆禄, 李美强, 黄光浩, 等. 智能网联汽车超车路径规划方法[J]. 控制理论与应用, 2024, 41(10): 1882−1898.
MA Qinglu, LI Meiqiang, HUANG Ghuanghao, et al. Overtaking path planning method for intelligent connected vehicle[J]. Control theory and technology, 2024, 41(10): 1882−1898.

[4] 虞立斌, 张亿, 黄磊, 等. 双向 A*路径规划算法的邻域改进方法研究[J]. 小型微型计算机系统, 2025, 46(6): 1312−1318.
YU Libin, ZHANG Yi, HUANG Lei, et al. Research on neighborhood improvement based on two-way A* path planning algorithm[J]. Journal of Chinese computer systems, 2025, 46(6): 1312−1318.

[5] 梅艺林, 崔立堃, 胡雪岩. 基于人工势场法的无人车路径规划与避障研究[J]. 兵器装备工程学报, 2024, 45(9): 300−306.
MEI Yilin, CUI Likun, HU Xueyan. Research on path planning and obstacle avoidance of unmanned vehicle based on artificial potential field method[J]. Journal of ordnance equipment engineering, 2024, 45(9): 300−306.

[6] 谢春丽, 陶天艺. 基于混合 A*算法的移动机器人路径规划研究[J]. 南京信息工程大学学报, 2025, 17(3): 340−351.
XIE Chunli, TAO Tianyi. Research on path planning of mobile robots based on hybrid A* algorithm[J]. Journal of Nanjing University of Information Science and Technology, 2025, 17(3): 340−351.

[7] 于逸然, 赖惠成, 高古学, 等. 基于遗传算法和 A*算法的多农机协同作业优化方法[J]. 系统仿真学报, 2025, 37(9): 2397−2408.
YU Yiran, LAI Huicheng, GAO Guxue, et al. Optimization method for multi agricultural machinery collaborative operation based on genetic algorithm and A~(*) algorithm[J]. Journal of system simulation, 2025, 37(9): 2397−2408.

[8] 杨国, 吴晓, 肖如奇, 等. 改进 A*算法的安全高效室内全局路径规划[J]. 电子测量与仪器学报, 2024, 38(7): 131−142.
YANG Guo, WU Xiao, XIAO Ruqi, et al. Improved A* algorithm for secure and efficient indoor global path planning[J]. Journal of electronic measurement and instrumentation, 2024, 38(7): 131−142.

[9] LIU Chenguang, MAO Qingzhou, CHU Xiumin, et al. An improved A-star algorithm considering water current, traffic separation and berthing for vessel path planning[J]. Applied sciences, 2019, 9(6): 1057.

[10] 赵晓, 王铮, 黄程侃, 等. 基于改进 A*算法的移动机器人路径规划[J]. 机器人, 2018, 40(6): 903−910
ZHAO Xiao, WANG Zheng, HUANG Chengkan, et al. Mobile robot path planning based on an improved A* algorithm[J]. Robot, 2018, 40(6): 903−910.

[11] WANG Zhongshan, LI Peiqing, WANG Zhiwei, et al. APG-RRT: sampling-based path planning method for small autonomous vehicle in closed scenarios[J]. IEEE access, 2024, 12: 25731−25739.

[12] SHI Yangyang, LI Qiongqiong, BU Shengqiang, et al. Research on intelligent vehicle path planning based on rapidly-exploring random tree[J]. Mathematical problems in engineering, 2020, 2020(1): 5910503.

[13] 郭利进, 李强. 基于改进 RRT*算法的移动机器人路径规划[J]. 智能系统学报, 2024, 19(5): 1209−1217.
GUO Lijin, LI Qiang. Path planning of mobile robots based on improved RRT* algorithm[J]. CAAI transactions on intelligent systems, 2024, 19(5): 1209−1217.

[14] 陈旭飞, 胡耀炜, 丛培龙, 等. 面向路径规划的双向交互多步蚁群算法研究[J]. 计算机工程与应用, 2025, 61(3): 166−176.
CHEN Xufei, HU Yaowei, CONG Peilong, et al. Research on bidirectional interactive multi step ant colony algorithm for path planning[J]. Computer engineering and applications, 2025, 61(3): 166−176.

[15] 郑琰, 席宽, 巴文婷, 等. 基于蚁群-动态窗口法的无人驾驶汽车动态路径规划[J]. 南京信息工程大学学报, 2025(2): 256−264.
ZHENG Yan, XI Kuan, BA Wenting, et al. Dynamic path planning for autonomous vehicles based on ant colony dynamic window method[J]. Journal of Nanjing University of Information Science and Technology, 2025(2): 256−264.

[16] 蒲兴成, 冼文杰, 聂壮. 基于改进蚁群优化算法的 AUV 三维路径规划[J]. 智能系统学报, 2024, 19(3): 627−634.
PU Xingcheng, XIAN Wenjie, NIE Zhuang. Three-dimensional path planning of AUV based on improved ant colony optimization algorithm[J]. CAAI transactions on intelligent systems, 2024, 19(3): 627−634.

[17] 张志文, 刘伯威, 张继园, 等. 麻雀搜索算法-粒子群算法与快速扩展随机树算法协同优化的智能车辆路径规划[J]. 中国机械工程, 2024, 35(6): 993−999,1009.
ZHANG Zhiwen, LIU Baiwei, ZHANG Jiyuan, et al. Cooperative optimization of intelligent vehicle path planning based on PSO-SSA and RRT[J]. China mechanical engineering, 2024, 35(6): 993−999,1009.

[18] 谢金燕, 刘丽星, 杨欣, 等. 改进粒子群优化算法的果园割草机作业路径规划[J]. 中国农业大学学报, 2023, 28(11): 182−191.
XIE Jinyan, LIU Lixing, YANG Xin, et al. Orchard lawn

mower operation path planning based on improved particle swarm optimization algorithm[J]. Journal of China Agricultural University, 2023, 28(11): 182−191.

[19] 王飞, 杨清平. 基于改进粒子群算法的城市物流无人机路径规划[J]. 科学技术与工程, 2023, 23(30): 13187−13194.
WANG Fei, YANG Qingping. Route planning of urban logistics UAV based on improved particle swarm optimization algorithm[J]. Science technology and engineering, 2023, 23(30): 13187−13194.

[20] WATKINS C J C H, DAYAN P. Q-learning[J]. Machine learning, 1992, 8: 279−292.

[21] SUTTON R S, BARTO A G. Reinforcement Learning: An Introduction[M]. 2nd ed. Cambridge: MIT Press, 2018.

[22] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[EB/OL]. (2013−12−19)[2025−02−24]. https://arxiv.org/abs/1312.5602.

[23] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529−533.

[24] 夏雨奇, 黄炎焱, 陈恰. 基于深度 Q 网络的无人车侦察路径规划[J]. 系统工程与电子技术, 2024, 46(9): 3070−3081.
XIA Yuqi, HUANG Yanyan, CHEN Qia. Path planning for unmanned vehicle reconnaissance based on deep Q-network[J]. Systems engineering and electronics, 2024, 46(9): 3070−3081.

[25] 李宗刚, 韩森, 陈引娟, 等. 基于角度搜索和深度 Q 网络的移动机器人路径规划算法[J]. 兵工学报, 2025, 46(2): 30−44.
LI Zonggang, HAN Sen, CHEN Yinjuan, et al. Mobile robots path planning algorithm based on angle searching and deep Q-network[J]. Acta armamentarii, 2025, 46(2): 30−44.

[26] SNIDER J M. Automatic steering methods for autonomous automobile path tracking[EB/OL]. (2009−12−30)[2025−02−24]. https://api.semanticscholar.org.

作者简介:

文家燕, 教授, 博士生导师, 中国自动化学会青年工作委员会委员。主要研究方向为多智能体系统协同控制、智能网联汽车队列控制。现主持国家自然科学基金及省部级基金项目 8 项, 获专利授权 10 项, 发表学术论文 35 篇。E-mail: wenjiayan2012@126.com。

辛华健, 副教授, 中国仿真学会机器人专委会委员, 主持完成了广西职业教学改革重点项目 1 项, 广西教育科学规划课题重点项目 1 项, 广西中青年教师科研项目 2 项。发表学术论文 20 余篇, 主编教材 2 部。E-mail: 13659619535@163.com。

谢广明, 教授, 博士生导师, 主要研究方向为智能仿生机器人、复杂系统与多机器人控制和水下特种机器人技术, 作为核心负责人主持多项国家自然科学基金重点项目、面上项目等国家级科研课题, 获发明专利授权 10 余项, 获国家自然科学奖二等奖、教育部自然科学奖一等奖、吴文俊人工智能科学技术创新奖二等奖, 发表学术论文 200 余篇。E-mail: xiegming@pku.edu.cn。