

深度强化学习驱动的应急车辆优先信号控制策略

吴鉴庭, 佟宁, 董芃

(大连交通大学 轨道智能工程学院, 辽宁 大连 116052)

摘要: 为了保障救护车、消防车等应急车辆安全快速通过路口, 提出一种深度强化学习驱动的应急车辆优先通行交通信号控制算法 EMV-DQN。通过构建应急车辆的状态表示、设计合理的奖励函数, 综合评估当前交通流状态与应急车辆的实时通行需求, 动态调整信号灯相位, 实现应急车辆优先通行。基于 SUMO 仿真平台的实验表明, 在多方应急车辆同时到达交叉路口的场景中, EMV-DQN 在平均行驶时间、等待时间等通行性能指标上显著优于传统控制方法和现有强化学习算法。相较于固定时序控制, 能降低 29% 的应急车辆延误; 相较于无应急响应机制的基线算法, 应急车辆的综合通行效率提升 4.9%。该方法在保障应急车辆优先通过的同时, 减少了对社会车辆通行的影响, 有效提升了整体交通效率。

关键词: 深度强化学习; 应急车辆; 车路协同; 交通信号控制; 智慧交通

DOI: 10.11907/rjdk.251029

扫描二维码阅读全文:

中图分类号: TP399; U491

文献标识码: A

文章编号: 1672-7800(2025)012-0008-10



Deep Reinforcement Learning-Driven Priority Signal Control Strategy for Emergency Vehicles

WU Jianting, TONG Ning, DONG Peng

(School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian 116052, China)

Abstract: In order to ensure the safe and fast passage of emergency vehicles such as ambulances and fire trucks through intersections, a deep reinforcement learning driven emergency vehicle priority traffic signal control algorithm EMV-DQN is proposed. By constructing a state representation of emergency vehicles and designing a reasonable reward function, comprehensively evaluating the current traffic flow status and real-time traffic demand of emergency vehicles, dynamically adjusting the phase of signal lights, and achieving priority traffic for emergency vehicles. Experiments based on the SUMO simulation platform show that in the scenario where emergency vehicles from multiple directions arrive at intersections simultaneously, EMV-DQN significantly outperforms traditional control methods and existing reinforcement learning algorithms in terms of traffic performance indicators such as average travel time and waiting time. Compared to fixed timing control, it can reduce emergency vehicle delays by 29%; Compared to the baseline algorithm without emergency response mechanism, the comprehensive traffic efficiency of emergency vehicles has increased by 4.9%. This method ensures the priority passage of emergency vehicles while reducing the impact on social vehicle traffic, effectively improving overall traffic efficiency.

Key Words: deep reinforcement learning; emergency vehicle; cooperative vehicle infrastructure; traffic signal control; intelligent transportation

0 引言

随着城市交通需求的快速增长与机动车保有量的持

续增加, 交叉口作为城市交通中的重要节点, 面临着日益严峻的交通拥堵和事故频发问题^[1,2]。在此背景下, 应急车辆(如救护车、消防车)在城市突发事件中的响应速度直接关系到公共安全和社会经济稳定。然而, 现有交叉口信

收稿日期: 2025-01-13

基金项目: 国家自然科学基金青年科学基金项目(61902051); 辽宁省教育厅基本科研项目(JYTMS20230011)

作者简介: 吴鉴庭(1999-), 男, 大连交通大学轨道智能工程学院硕士研究生, 研究方向为智慧交通; 佟宁(1981-), 女, 博士, 大连交通大学轨道智能工程学院副教授、硕士生导师, 研究方向为区块链技术、智慧交通; 董芃(2001-), 男, 大连交通大学轨道智能工程学院硕士研究生, 研究方向为智慧交通。本文通讯作者: 佟宁。

号系统在面对突发应急任务时调节机制不够灵活,使得应急车辆经常受到常规交通流与交通灯信号调控的双重干扰,导致救援延误甚至交通冲突^[3]。

近年来,随着车路协同系统(Cooperative Vehicle Infrastructure System, CVIS)与车联网技术的迅速发展,车辆之间及车辆与路侧设备之间的信息互联成为可能,为实现应急车辆与信号系统的协同控制提供了基础支持^[4]。王吟松等^[5]、田芳等^[6]基于实时通信提出车速引导模型与自适应控制策略,以提升应急通行效率。Neelakandan等^[7]结合物联网技术与改进的群优化算法,实现了多维度交通数据的精确分类与拥堵预测控制。肖宇舟等^[8]基于交通波理论,通过绿灯延长、红灯提前结束与相位插入策略,动态调整信号配时以消散前方排队车辆,实现了应急车辆的快速通过。

现有研究在车路协同技术的基础上,针对应急车辆的“时空路权”问题,主要从两方面入手:一是规划应急车辆的行驶路径;二是优化交叉口的信号控制。在路径优化方面,引入逆向左转或可变导向车道策略,以保障应急车辆在特定场景下快速通过^[9,10]。罗倩等^[11]结合车道级导航与信号优先控制的策略,提前清空应急车道实现低干扰通行。Zhang等^[12]、张艺还^[13]利用实时高精度数据动态调整应急车辆路径,避开拥堵区域。然而,这些方法多用于结构简单或车流量较低的区域,在多变交通环境下缺乏鲁棒性与泛化能力。在交叉口的信号控制优化方面,早期研究采用全球定位系统(Global Positioning System, GPS)、射频识别(Radio Frequency Identification, RFID)等技术实现触发式信号优先控制。Miyawaki等^[14]利用红外信标检测应急车辆的到达。杨兆升等^[15]基于RFID技术提出模糊控制策略以实现简易的车路协同。

近年来,研究者们开始构建协同式信号控制方法。通过车载与路侧设备的数据交换,使信号系统提前感知应急需求,以实时调整配时策略。Zhong等^[16]设计三阶段信号控制策略,首先按需信号配时以降低路段饱和度,其次结合混合信号抢占以确保应急车辆无阻通行,最后利用线性规划恢复周期以快速恢复路网平衡。Boutilier等^[17]结合需求预测与随机森林模型的调度策略,提升了急救车辆部署与路径规划效率。然而,上述方法大多依赖高精度、实时数据的支撑,感知设备覆盖范围、通信延迟等问题限制了应用范围。

传统交通信号控制方法多基于规则设计,虽具有一定灵活性,但在流量波动剧烈或交叉口饱和度较高的情况下,调节效果有限且难以适应复杂场景^[18]。相较而言,强化学习因具备自适应与数据驱动能力,在动态交通环境中的信号控制优化性能明显。Cao等^[19]通过构建包含九项约束条件的奖励函数与平衡采样机制,将深度强化学习应用于应急信号控制场景。Su等^[20]结合多智能体强化学习与动态路径规划方法,提出基于协作策略的优先通行机

制。Garg等^[21]利用高清摄像头捕捉车流的多维特征,通过能处理复杂视觉输入的强化学习控制方案,提升交叉口调控效率。尽管现有研究在优化应急通行方面取得了一定进展,但多数方法仍聚焦于应急车辆延误最小化,较少考虑社会车辆的通行效率、等待延时及燃油消耗等综合因素,存在优化目标单一的问题。

为此,本文提出一种深度强化学习驱动的应急车辆优先信号控制策略,主要贡献包括:①构建融合车路协同机制的信号控制系统框架,将多应急车辆场景下的优先需求嵌入深度强化学习流程,实现交通信号的智能调控;②提出基于部分可观测交通状态的深度强化学习方法,在实际路网与真实车流数据条件下实现高效决策,提升了策略稳定性与泛化性;③设计多目标优化奖励函数,动态权衡应急车辆与社会车辆的通行需求,在保障应急效率的同时兼顾了整体通行的公平性与系统性能。

1 基于车路协同的应急车辆优先通行控制系统

本文设计了一种基于车路协同的交叉口应急车辆优先通行控制系统,如图1所示。该系统由车载单元、路侧单元和智能信号控制系统三大模块组成。车载单元安装于应急车辆上,用于采集车辆动态数据并与路侧单元通信;路侧单元部署于交叉口,负责数据感知、融合及信号优化;智能信号控制系统作为核心决策模块,结合实时数据动态优化信号灯相位^[22]。

当应急车辆进入交叉口的感知范围后,车载单元实时采集应急车辆动态信息,通过车联网技术与路侧单元进行数据交互。路侧单元在接收应急车辆信息的同时,同步采集交叉口整体交通流数据,并将整合后的数据传输至智能信号控制系统进行处理。系统利用信号控制算法对当前交通状态进行综合评估,动态生成最优信号相位策略,实

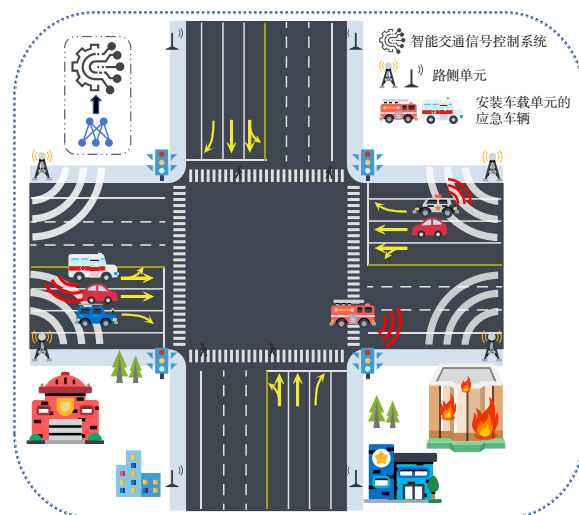


Fig. 1 The signal control system framework that supports priority passage for emergency vehicles

图1 支持应急车辆优先通行的信号控制系统框架

现应急车辆优先通行与社会车辆效率的平衡。

2 问题描述

2.1 研究问题定义

在城市交通交叉口场景中,应急车辆的优先通行面临着以下核心挑战:①不同方向上的应急车辆同时请求通行时,传统固定配时策略难以实时响应;②布设于交叉口的路侧单元的感知范围受限,无法精准感知整个交叉口及其临近道路的全局状态;③信号控制系统既要保障应急车辆快速通行,又要避免严重干扰社会车辆正常通行的秩序。

综上,现有规则模型、静态规划等方法因依赖高精度全局数据或缺乏动态优化能力,在面对突发性通行需求和动态交通流时的适应性不足。为此,本文基于有限的感知信息构建适应动态变化的交叉口交通状态表征模型,基于深度强化学习设计了一种数据驱动的信号控制策略,以实现应急车辆优先通行的同时,最大化减少对社会车辆通行效率的负面影响。

2.2 基于MDP的信号控制建模

在实际应用中,交通信号控制是一个高度动态且复杂的系统,涉及交通流、交通事故、天气变化等因素,以及多种相位信号的切换决策。传统规则控制或优化方法难以在复杂环境下实时变化,因此需要一种自适应学习最优决策的框架。马尔可夫决策过程(Markov Decision Process, MDP)作为强化学习中的核心模型之一,广泛应用于决策问题的建模,特别适用于在不确定性和动态环境中作出决策的交通信号控制场景。

一个典型的MDP由状态空间、动作空间、奖励函数、状态转移概率和折扣因子等要素组成。在应急车辆优先通行的场景下,MDP为交通信号控制问题提供了一个数学框架,通过抽象环境的动态变化与信号灯相位切换决策,使智能体能在复杂环境中学习最优策略。本文主要关注MDP的核心建模过程,并不显式建模状态转移概率和策略优化过程,而是由环境动态生成交通流状态和信号控制状态。

MDP建模过程中涉及到的变量符号如表1所示。其中,环境 E 表示交叉口的整体交通流信息,包括实时车流量、道路拥堵情况等;信号控制系统作为智能体 G ,需基于感知的交通状态选择最优的信号相位;动作 A 定义智能体的决策空间,包括所有可能的信号灯相位切换方案;交叉口的交通状态、应急车辆位置及信号灯相位等信息共同构成状态 S ^[23,24];奖励 R 作为智能体衡量当前决策的优劣的依据,以最大化奖励为优化目标,通过强化学习机制持续优化信号控制策略。

2.2.1 状态空间

在交通信号控制任务中,信号控制系统需依赖高频、实时更新的交通数据为信号决策提供依据。状态空间 S

Table 1 Notation descriptions

表1 符号说明

符号	含义
E	环境
G	智能体
A	动作
S	状态
R	奖励
Δt	动作间隔
P_c	当前信号相位
Q_i	车道 i 上的队列长度
W_i	车道 i 上车辆累计等待时间
EMV_i	车道 i 上的应急车辆状态

用于描述交叉口整体交通流的状态,包括驶入车道的交通信息和应急车辆的实时特征,这些信息直接影响信号控制系统对当前交通状态的感知能力,并决定决策质量。

通常情况下,所有驶入车道的车辆信息均可通过部署于交叉口的传感设备实时获取。在实际应用场景中,受限传感器的覆盖范围、通信延迟及检测精度,系统可能无法完整获取所有驶入车辆的信息,从而影响了信号控制的准确性和效率。为了更贴近现实应用,本文引入有限感知机制,假设智能体的感知范围为交叉口周围200 m内的车道。在此范围内,传感设备能采集驶入车道的部分车辆状态信息,超出感知范围的车辆则视为不可观测,这种机制不仅提升了算法的可部署性,还能更准确地模拟实际交通环境中的感知约束。

具体而言,状态空间由交叉路口各驶入车道的状态向量组成,单个车道的状态向量由以下信息构成:当前信号相位 P_c 表示车道 i 是否对应当前绿灯方向(1表示绿灯,0表示红灯);排队长度 Q_i 描述车道 i 上车辆的排队长度(单位:m);累计等待时间 W_i 表示车道 i 上所有车辆的等待时间总和(单位:s);应急车辆存在状态 EMV_i 指车道 i 上是否存在应急车辆(1表示存在,0表示不存在);应急车辆位置 EMV_{pos} 表示车道 i 上离路口最近的应急车辆与路口的距离(单位:m);应急车辆速度 EMV_{speed} 表示车道 i 上离路口最近的应急车辆的行驶速度(单位:m/s)^[25]。

基于上述定义,车道 i 的状态向量可表示为:

$$S_i = [P_c, Q_i, W_i, EMV_i, EMV_{pos}, EMV_{speed}] \quad (1)$$

交叉口整体状态由所有驶入车道的状态向量组合构成,记为: $S = \{S_1, S_2, \dots, S_n\}$, n 为交叉路口的驶入车道数量。

2.2.2 动作空间

在交通信号控制任务中,动作空间 A 定义智能体在每个决策时刻可执行控制操作,即不同的信号灯相位切换方案。本文参考现实中大多数道路路口的信号灯控制方式,将动作空间设置为四相位控制策略:

$$A = [NS, NSL, WE, WEL] \quad (2)$$

式中: NS 、 NSL 、 WE 、 WEL 分别对应南北直行,南北左转,东西直行和东西左转4个方向。

相较于传统控制方法采用固定顺序的相位切换规则, 本文采用动态决策机制使智能体根据当前交通状态灵活选择最优相位, 不仅满足了一般交通流优化需求, 更能在紧急情况下响应驶入的应急车辆, 提升通行优先级^[26]。在每个时间步 Δt 中, 智能体根据当前观测到的交通状态, 从动作空间 A 中动态选择一个最优相位执行信号切换操作, 使相应方向的车辆获得通行权。

2.2.3 奖励函数

奖励函数 R 用于评估智能体每次决策的效果, 是强化学习信号控制策略优化的核心。本文在交通信号控制问题中, 综合考虑了应急车辆的优先通行和社会车辆的通行效率, 设计了一种加权线性组合的奖励函数。

$$R = \alpha \cdot R_e + \beta \cdot R_s \quad (3)$$

式中: α, β 为奖励模式优先系数, 分别表示应急车辆和社会车辆的优先权重, 系数越大对应的模式优先度越高, 进而直接影响信号控制策略的倾向性; R_e, R_s 分别表示应急车辆奖励和社会车辆奖励, 用于评估不同类型车辆通行的优化程度。

应急车辆奖励 R_e 用于评估应急车辆的通行效率, 旨在减少应急车辆的停车、等待和减速, 确保其快速、安全地通过交叉口。通过实时获取应急车辆在智能体感知半径内的动态数据, 基于其行程时间、累计等待时间和行驶速度进行计算。^[19]

$$R_e = \frac{1}{n} \sum_{i=1}^n (-e_1 \cdot T_i - e_2 \cdot W_i + e_3 \bar{V}_i - e_4 \cdot L_i) \quad (4)$$

式中: e_1, e_2, e_3, e_4 为应急车辆奖励中各指标的权重参数, 用于平衡不同优化目标的重要性; n 为通过路口的应急车辆总数; T_i 为应急车辆 i 通过路口的行程时间, 即从进入智能体感知半径到通过路口所花费的时间; W_i 为应急车辆 i 通过交叉口前的累计等待时间; $\bar{V}_i$ 为应急车辆 i 通过交叉口的平均行驶速度; L_i 为应急车辆 i 低于预期车速而产生的时间损失。

社会车辆奖励 R_s 用于评估社会车辆的通行效率。在

正常交通场景中, 社会车辆的通行需求占据主导地位, 应急车辆的出现仅为小概率事件。本文为了优化整体交通效率并兼顾车道间的通行公平性, 设计了社会车辆奖励函数。

$$R_s = -s_1 \sum_{j=1}^m (C \cdot W_j - \bar{W})^2 - s_2 \cdot \bar{W} - s_3 \cdot \bar{Q}_i + s_4 \bar{V} \quad (5)$$

式中: s_1, s_2, s_3, s_4 为社会奖励中各项指标的权重参数; s_1 为各车道平均等待时间与整体平均等待时间的偏差程度权重, 作为公平性惩罚项; C 为容忍常量, 允许车道平均延误与整体平均延误存在一定范围的波动, 抑制极端车道延误情况的发生; j 为交叉口驶入车道编号; m 为交叉口驶入车道总数; $W_{avg}, Q_{avg}, V_{avg}$ 分别为当前驶入车道上所有车辆的平均等待时间、平均队列长度和平均行驶速度。

3 应急车辆优先通行控制算法

3.1 框架概述

本文在 MDP 建模的基础上, 提出一种基于深度强化学习的应急车辆优先通行交通信号控制算法 EMV-DQN, 整体框架结构和工作流程如图 2 所示。

由图 2 可见, 整个系统主要包括环境感知、状态表示、智能体决策与优化 3 个核心模块: ① 环境感知模块由部署在交叉口的路侧单元和应急车辆上的车载单元共同组成, 路侧单元负责实时采集交叉口交通流信息, 包括信号灯状态、排队长度、累计等待时间等, 车载单元通过车联网技术向系统上报应急车辆的实时位置和行驶速度, 作为状态输入; ② 状态表示模块基于有限感知机制, 将采集到的交通数据转换为高维状态向量, 并输入至智能体进行分析和决策; ③ 智能体决策优化模块采用深度 Q 网络 (Deep Q-Network, DQN) 进行信号控制决策。

在每个时间步, 智能体根据当前状态选择最优的信号相位, 通过强化学习策略不断优化其决策以提升应急车辆的通行效率。其中, 框架的核心部分是智能体利用深度 Q

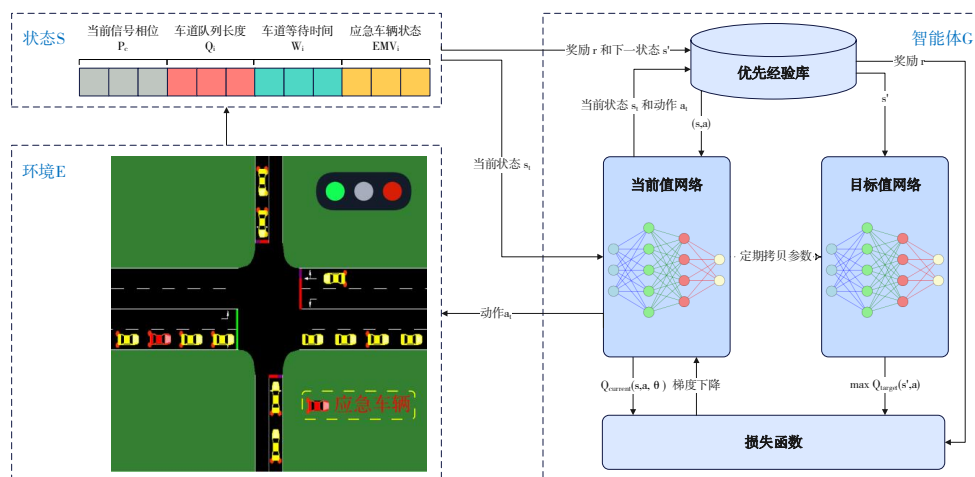


Fig. 2 EMV-DQN framework

图 2 EMV-DQN 框架

网络近似Q值函数,不断与环境进行交互,优化信号灯相位切换策略。

3.2 深度Q网络实现

Q学习是一种典型的无模型强化学习方法,智能体通过与环境持续交互,学习状态—动作对的预期回报(Q值),逐步构建最优Q值函数。然而,在高维、连续状态空间中,传统Q学习面临存储需求大、计算复杂度高的问题,难以直接应用于复杂交通场景。为此,深度Q网络结合深度神经网络和Q学习,通过神经网络对Q值函数进行近似,有效解决了上述瓶颈问题。

图2为智能体的决策流程。智能体从环境 E 中获取当前状态 S ,该状态向量包括信号灯相位 P_c 、车道队列长度 Q_i 、车辆累计等待时间 W_i 及应急车辆的实时状态 EMV_i 等多维交通特征信息。这些信息通过路侧感知设备实时采集,并作为输入传递至值网络进行处理。当前值网络基于深度神经网络近似Q值,根据 ϵ -贪心策略选择最优动作 a_t ,即信号相位的切换方案。执行动作 a_t 后信号相位发生变化,从而影响当前交叉口的交通状态,环境则反馈即时奖励 r_t ,以及进入下一时刻的状态 s_{t+1} 。

在智能体与环境不断交互的过程中,部分历史数据(s_{t+1}, a_t, r_t)被存储至优先经验回放池。经验回放机制通过随机采样打破了样本间的时间相关性,提升了模型的训练效率和泛化能力^[27]。此外,为了避免Q值剧烈波动,DQN采用目标网络计算目标Q值 $\max_{a'} Q(s_{t+1}, a'; \theta^-)$,其中目标网络的参数 θ^- 由当前值网络的参数 θ 定期复制,以提升训练稳定性。智能体通过神经网络的损失函数,最小化预测Q值与目标Q值的误差,不断更新当前值网络的Q值,以自适应地优化信号控制策略,在复杂动态的交通环境中实现应急车辆优先通行的目标。Q值的更新公式为:

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q_{\text{target}}(s', a') - Q(s, a)] \quad (6)$$

式中: α 为学习率; γ 为折扣因子; r 为即时奖励; Q_{target} 为目标网络的参数。

4 仿真实验与分析

4.1 仿真环境

本文采用Python编程语言和城市交通仿真平台SUMO (Simulation of Urban Mobility, SUMO)构建了交通信号控制系统及应急车辆交通流模型。SUMO为一款开源的微观交通模拟工具,具备时间离散、空间连续的特点,能精确模拟大规模交通网络的动态变化,现已广泛应用于智能交通研究^[28]。

为了验证EMV-DQN算法的有效性,基于交通信号控制强化学习基准框架(Reinforcement Learning Benchmarks for Traffic Signal Control, RESCO),选取基于德国因戈尔施塔特市交通网络的“InTAS”场景进行实验^[29]。该场景因

交通流量复杂、动态变化显著被广泛应用于智能信号控制研究,路网结构如图3所示。



Fig. 3 Traffic network and InTAS scenario road structure in Ingolstadt city

图3 因戈尔施塔特市交通网络与实验场景路网结构

图4为社会车辆的到达分布情况,反映了高峰期不同时间段交通流量的动态变化。为了模拟城市突发事件的随机性,假设应急车辆以相等概率从交叉口的4个驶入车道方向随机出现,并按照固定时间间隔随机插入一条交通流。

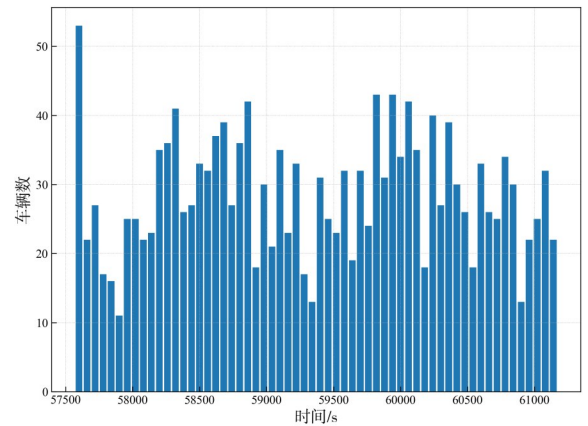


Fig. 4 Distribution of social vehicle arrivals in InTAS scenarios

图4 InTAS场景中社会车辆的到达分布情况

为了充分评估算法在应急车辆优先通行场景下的优化效果,实验设定应急车辆数量占总体社会车辆数量的比例为5%(高于实际城市高峰期中通常不超过1%的水平),以更具挑战性地检验信号控制策略在高强度应急交通需求下的适应能力,确保实验场景的代表性和通用性。

4.2 仿真参数

为了确保仿真环境能真实反映城市交通运行规律,仿真参数的设定综合实际交通流统计数据、相关研究经验及实验优化结果。

在仿真环境构建方面,单回合仿真时长设为3 600 s,对应一个完整的高峰时段交通流,以观察信号控制策略在长期运行下的稳定性和适应性。为了保证智能体充分训练并收敛、兼顾计算成本,仿真总回合数设为100。仿真步长设为10 s,并在信号相位切换时采用3 s黄灯缓冲期,确保信号控制系统在合理时间范围内完成信号决策,减少因信号切换时而引发的交通冲突风险。

在强化学习智能体的超参数方面,初始探索率设定为 1,保证训练初期具备足够的探索能力,防止智能体过早陷入局部最优解。最小探索率设定为 0.01,确保训练后期模型稳定收敛,避免策略因探索过度而产生剧烈波动。折扣因子设定为 0.99,使智能体在优化决策时更关注长期收益,在整体交通效率和应急车辆优先通行之间达到平衡。主要参数设置如表 2 所示。

Table 2 Main parameters
表 2 主要参数

仿真参数	设定值	算法超参数	设定值
回合数	100	初始探索率 ϵ	1
仿真时长/s	3 600	最小探索率 ϵ	0.01
总训练步数	36 000	目标网络更新步长	500
交叉口通信半径/m	200	折扣因子 γ	0.99
车流量/(辆·h ⁻¹)	2 015	经验池大小/M	10 000
应急车辆出现率/%	5	最小采样	32

4.3 奖励函数权重

由于奖励函数的权重系数直接影响信号控制策略的优化效果,决定了强化学习模型如何平衡应急车辆优先通行与社会车辆正常的通行需求。为了确保权重设置的合理性,基于应急车辆优先通行需求与交通流优化原理进行理论分析,参考相关研究中的奖励函数权重设定经验,最终确定的奖励函数权重参数如表 3 所示。

Table 3 Reward function weight parameters
表 3 奖励函数权重参数

参数符号	参数说明	数值
α	应急车辆奖励优先系数	0.7
β	社会车辆奖励优先系数	0.3
e_1	应急车辆行驶时间权重	0.1
e_2	应急车辆等待时间权重	0.5
e_3	应急车辆行驶速度权重	0.3
e_4	应急车辆低速行驶惩罚	0.1
s_1	车道等待时间公平性惩罚	0.1
s_2	社会车辆等待时间权重	0.4
s_3	车道队列长度权重	0.3
s_4	社会车辆行驶速度权重	0.2

在初步设定阶段,为保证信号控制策略优先为应急车辆提供通行权,给予应急车辆奖励模式优先系数 α 一个较大的权重范围[0.6, 0.9],社会车辆奖励模式优先系数 β 的权重范围为[0.1, 0.4],应急车辆奖励权重 e_1 、 e_2 、 e_3 、 e_4 及社会车辆奖励权重 s_1 、 s_2 、 s_3 、 s_4 的初始搜索范围设定为[0.1, 0.5]。

在参数优化阶段,采用改进型网格搜索算法在 SUMO 仿真平台上进行多目标优化实验,通过调整不同的权重组合以评估应急车辆及社会车辆的通行效率,筛选出在不同交通流情况下均能稳定收敛的最优参数组合。

4.4 训练结果分析

为了验证 EMV-DQN 算法的有效性,将其与传统交通信号控制方法及其他强化学习方法进行分析。固定时序控制是最常见的信号控制方案,按照预设周期和固定相位顺序进行信号切换,不随交通流动态调整。贪婪控制根据车道队列长度,优先为排队最长的车道分配通行权^[30]。最大压力控制计算驶入车道与驶出车道车辆数的差值(即交通压力),优先选择压力最大的作为下一绿灯相位^[31]。

同时,选取近年来在交通信号控制与强化学习领域表现优异的典型算法作为比较对象。其中,独立深度 Q 网络(Independent deep Q-networks, IDQN)引入了双网络结构与经验回放机制,增强了对复杂状态空间的泛化能力^[27]。独立近端策略优化(Independent proximal policy optimization, IPPO)作为多智能体场景中表现突出的策略优化算法,兼具稳定性与高效性,其输出层根据实际情况进行适当调整^[32,33]。DTLight 是一种基于自注意力机制(Self-Attention)的深度强化学习方法,利用知识蒸馏从教师模型中学习得到一个轻量级 Transformer 模型,降低了实现过程中的计算成本^[34]。

图 5 为 EMV-DQN 算法训练过程中奖励值、整体交通流等性能指标的变化情况。由此可见,随着训练次数增加奖励值逐步提升,平均等待时间逐步下降,表明算法在不断优化信号控制策略。训练后期,奖励值和关键性能指标趋于平稳,表明算法已收敛至较优策略。

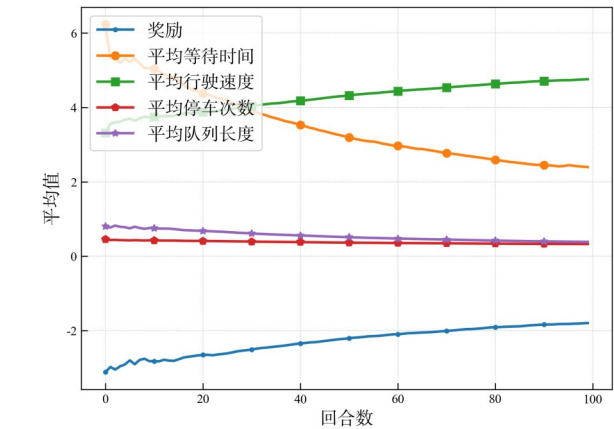


Fig. 5 Training curve of EMV-DQN algorithm

图 5 EMV-DQN 算法训练曲线

图 6 为不同控制算法下,应急车辆累计延误的变化趋势,横轴表示训练回合数,纵轴表示应急车辆的延误时间。由此可见,传统控制方法(固定时序、贪婪控制、最大压力控制)的延误时间基本稳定,表明这些方法在应急车辆优先通行的任务中适应性有限,无法动态调整信号相位来优化通行效率。EMV-DQN 在训练初期的表现与 IDQN 相近,但在第 60 回合后性能逐渐优于传统方法与基线强化学习方法,表明该算法通过智能体与环境的持续交互学习,可动态识别并优先响应应急车辆的通行需求。

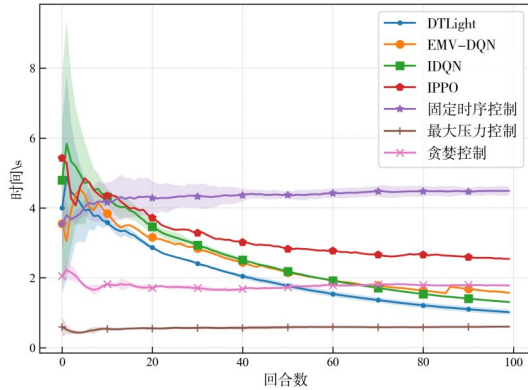


Fig. 6 Variation of cumulative emergency vehicle delay under different control methods with training episodes

图6 应急车辆累计延误随训练回合变化趋势

4.5 评估结果

4.5.1 整体车辆通行效果评估

为了评估 EMV-DQN 算法在动态交通环境中的通行优化能力,基于相同路网场景选取长时间真实交通数据,各控制方法进行 10 次独立仿真实验。评价指标包括平均队列长度、平均等待时间、平均停车次数和平均行驶速度,如表 4 所示(黑体加粗部分表示最优值,带下划线部分表示次优值)。

Table 4 Overall vehicle traffic performance under different control methods

表4 不同控制方法下的总体车辆通行性能

控制方法	平均队列长度/m	平均等待时间/s	平均停车次数	平均行驶速度/($\text{m}\cdot\text{s}^{-1}$)
固定时序控制	0.88	4.38	0.46	2.94
贪婪控制	0.36	3.03	0.32	4.89
最大压力控制	0.28	1.09	0.26	5.38
IPPO	0.40	4.64	0.40	4.35
IDQN	<u>0.21</u>	1.02	0.27	5.47
DTLight	0.20	<u>0.94</u>	<u>0.26</u>	<u>5.53</u>
EMV-DQN	0.22	0.92	0.25	5.56

由表 4 可知,不同控制方法在长期高流量场景下表现存在显著差异,固定时序控制和贪婪控制难以适应复杂交通流,导致平均队列长度和等待时间较高,整体通行效率偏低。EMV-DQN 在所有关键指标中有 3 项最优,相较于固定时序控制的平均行驶速度提升 89.2%,平均队列长度减少 75.2%,平均等待时间减少 79.1%,平均停车次数减少 46.3%,表明其有效提升了整体通行能力并减少延误。相较于传统方法中表现最优的最大压力控制,EMV-DQN 仍在全部指标上取得领先。

在强化学习方法中,IPPO 表现最差,甚至不及贪婪控制,表明其策略优化能力不足,未能适应动态交通流。IDQN 在整体通行效率上优于传统方法,能减少车辆的等待时间和停车次数。DTLight 在整体通行效率上表现优于 IDQN,仅在平均队列长度上领先 EMV-DQN。实验表明,

EMV-DQN 在复杂交通环境中能有效优化信号调度策略,提升整体交通流的通行效率。

4.5.2 应急车辆通行效果评估

为了评估不同控制方法在应急车辆优先通行保障方面的表现,从平均行程时间、损失时间、等待时间和停车次数等 4 个关键指标进行比较,实验结果如表 5、图 7 所示。表 5 为不同方法的实验具体数据。图 7 中的 4 个子图分别对应各项指标,柱形高度表示平均值,误差条反映标准差。

Table 5 Comparison of emergency vehicle traffic performance under different control methods

表5 不同控制方法下的应急车辆通行性能比较

控制方法	平均行程时间/s	平均损失时间/s	平均等待时间/s	平均停车次数
固定时序控制	71.12±0.9	49.07±0.9	24.59±0.6	1.99±0.0
贪婪控制	70.22±1.3	47.21±1.3	22.96±0.9	1.96±0.1
最大压力控制	54.11±0.7	31.58±0.6	12.02±0.4	1.35±0.0
IPPO	193.1±9.6	165.8±9.1	109.4±6.4	6.1±0.3
IDQN	55.63±0.8	32.89±0.7	13.16±0.5	1.37±0.0
DTLight	<u>53.18±1.5</u>	<u>30.49±1.4</u>	<u>11.38±0.8</u>	<u>1.24±0.1</u>
EMV-DQN	51.20±0.9	28.66±0.8	10.27±0.5	1.15±0.0

由表 5、图 7 可知,固定时序控制、贪婪控制和随机控制在所有指标上的表现均较差,尤其是在平均损失时间和等待时间方面,其高数值表明这些方法难以满足应急车辆的优先通行需求。最大压力控制相较于其他传统方法表现最佳,其较强的实时调节能力在一定程度上减少了应急车辆的延误,但仍受限于固定的信号切换逻辑。

在强化学习方法中,IPPO 在所有指标上均明显落后,甚至不及部分传统方法,说明其在应急车辆信号优化任务中的适应性较差。IDQN 在整体通行效率上有所优化,减少了应急车辆的等待时间和停车次数,但由于信号相位调整的灵活性不足,仍未能充分发挥应急优先控制的优势。DTLight 作为使用轻量级 Transformer 的强化学习方法之一,在应急车辆通行方面优于 IDQN,尤其是在平均损失时间、等待时间上更具优势,但 4 项指标的性能仍不如 EMV-DQN。从其标准差范围可知,DTLight 在突发应急场景的优化效果波动较大,难以输出稳定的交通信号控制策略。EMV-DQN 在所有指标上均表现最佳,有效减少了应急车辆的平均行程时间、损失时间、等待时间、停车次数,为应急车辆提供了更高效的优先通行保障。

4.5.3 综合性能分析

为了量化不同信号控制方法的综合性能,引入综合性能指标(Overall Performance Index, OPI)统一衡量各方法在多个性能指标上的表现。由于不同指标的数值范围存在较大差异,为了避免某些指标对综合性能计算的主导影响,先对数据进行对数变换以压缩其动态范围,再对部分

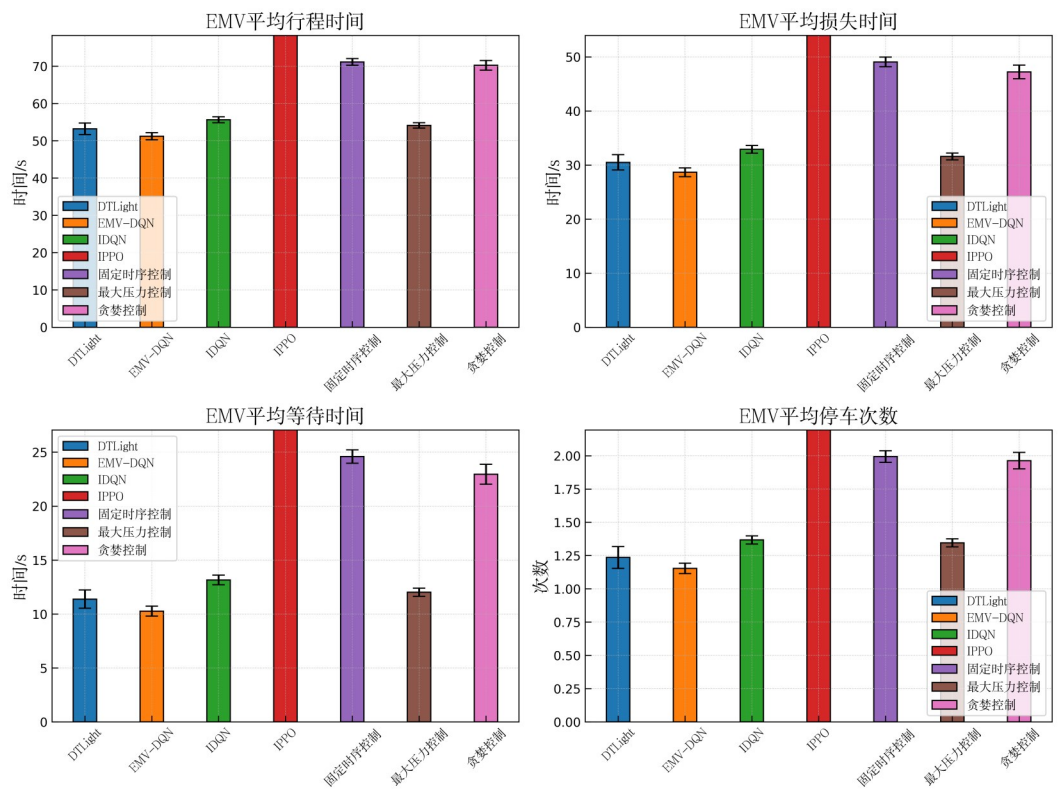


Fig. 7 Comparison of emergency vehicle traffic performance under different control methods

图 7 不同控制方法下的应急车辆通行性能比较

值越大越优的指标取倒数进行变换。整体 OPI 和应急 OPI 的计算公式为：

$$OPI_{all} = \log((X_1 + 1)(X_2 + 1)(X_3 + 1)) + \log(\frac{1}{X_4 + 1}) \quad (7)$$

$$OPI_{emv} = \log((Y_1 + 1)(Y_2 + 1)(Y_3 + 1)(Y_4 + 1)) \quad (8)$$

式中： X_1 、 X_2 、 X_3 、 X_4 和 Y_1 、 Y_2 、 Y_3 、 Y_4 分别为整体车辆和应急车辆的各项通行指标，较低的 OPI 值表示该方法在多个通行性能指标上的表现更好，即综合通行效率更高。

表 6 为不同控制方法的综合性能得分，并以 EMV-DQN 作为基准计算其相对性能差异。由此可知，固定时序控制因信号配时固定，难以适应动态交通需求，在整体 OPI 及应急 OPI 方面均表现不佳，综合得分相较于 EMV-DQN 下降 31.9%。贪婪控制虽能对交叉口中排队最长的车道给予优先通行，在局部优化方面效果更好，但由于缺乏全局策略优化，其综合性能相较于 EMV-DQN 低 23.8%。最大压力控制在传统控制方法中表现最佳，通过计算当前驶入车道的流量压力自适应调整信号灯，综合性能相较于 EMV-DQN 低 4.25%，说明该方法在应急场景下具有一定适用性，但仍无法满足应急车辆的高优先级通行需求。

在强化学习方法中，IPPO 因其策略更新涉及更复杂的计算，在训练过程中存在较大波动，未能稳定收敛，在所有方法中表现最差，综合性能相较于 EMV-DQN 下降 67%，说明其在应急信号控制任务中的优化能力有限。IDQN 的表现略优于传统方法，但相较于 EMV-DQN 仍有 5.2% 的性能差距，表明其未能充分利用有限的观测信息进行信号调

Table 6 Comparison of comprehensive performance of different control methods

表 6 不同控制方法的综合性能比较				
控制方法	整体 OPI	应急 OPI	综合得分	性能差异/%
固定时序控制	2.985	12.529	15.514	-31.9
贪婪控制	2.165	12.403	14.568	-23.8
最大压力控制	1.386	10.914	12.300	-4.5
IPPO	2.610	17.045	19.655	-67
IDQN	1.301	11.073	12.374	-5.2
DTLight	1.241	10.765	12.005	-2.0
EMV-DQN	1.234	10.533	11.766	0

度优化。DTLight 是从教师模型中蒸馏出的轻量级模型，计算复杂度低，在相同的训练回合中收敛更快，综合性能优于 IDQN，表明其在普通交通流优化上具有一定优势，但在多应急车辆的优先通行任务上性能仍然不够，相较于 EMV-DQN 的综合性能低 2%。

综上，EMV-DQN 在综合得分上表现最佳(11.766)，远超其他传统控制方法和强化学习基线方法。实验表明，EMV-DQN 在交叉口信号控制优化中，在有效保障应急车辆快速通行的同时，优化了整体交通流，在动态交通环境下具有更强的适应性和部署价值。

4.6 讨论

分析仿真实验中的交通场景可知,交叉路口各条驶入车道的平均长度为 50~100 m,最长车道为 143.49 m,均小于智能体的可感知范围(200 m)。在实际决策过程中,智能体虽能获取局部交通信息,但受到空间和时间限制,这种局限性对强化学习算法(IDQN、IPPO)的表现产生了较大影响,尤其是当交通流密集、动态变化剧烈时,模型的决策质量可能下降。最大压力控制等基于规则的方法,因简单的决策逻辑和较低的数据依赖性,在感知范围受限的环境下稳定性更高。尽管,EMV-DQN 受到感知范围的约束,但通过高效利用局部信息,在优化应急车辆优先通行和整体交通流效率方面优势突出,表明深度强化学习在实际智能交通系统中的应用潜力巨大,能有效提升交通信号控制的智能化水平。

5 结语

本文针对多应急车辆优先通行问题,提出一种基于深度强化学习的信号控制方法 EMV-DQN。通过对局部交通状态的高效感知与信号灯控制策略的动态优化,在确保应急车辆快速、安全通行的同时,兼顾了社会车辆的通行效率。

实验表明,EMV-DQN 在优化应急车辆通行效率方面表现优越,能显著减少应急车辆的平均行驶时间、损失时间、等待时间和停车次数,有效提升了应急通行保障能力。在无应急车辆的情况下,EMV-DQN 仍能有效优化整体交通流,表明其不同交通情境下具备较强的适应性和稳定性,展现了深度强化学习在智能交通系统中的多目标优化能力和广阔的应用前景。然而,实验也发现了当前强化学习方法的局限性:如何在感知范围受限和数据采集不足条件下,充分挖掘算法潜力,提升决策适应性。

未来,将重点探索更高效的感知与决策机制,例如引入多模态信息融合技术和多智能体协同学习框架,结合更多来源的数据及协同临近路口资源,进一步增强算法在局部信息下的决策性能;针对不同城市中各种交通模式下的复杂环境,优化强化学习模型的迁移能力,为智能交通系统的实际部署与优化提供有力支持。

参考文献:

- [1] National Bureau of Statistics. China Statistical Yearbook 2024[EB/OL]. <https://www.stats.gov.cn/sj/ndsj/2024/indexch.htm>. 国家统计局. 中国统计年鉴 2024[EB/OL]. <https://www.stats.gov.cn/sj/ndsj/2024/indexch.htm>.
- [2] JING P, HUANG H, CHEN L. An adaptive traffic signal control in a connected vehicle environment: a systematic review[J]. Information, 2017, 8(3):101.
- [3] ZHAO L N, JIA X W, DAI S, et al. Characteristics of urban road traffic safety in China[J]. Urban Transport of China, 2018(3): 9-14, 20.
- [4] ZHANG L L, WANG L, LIU J D, et al. Priority control of urban emergency vehicles: overview and prospect[J]. Science Technology and Engineering, 2021, 21(34): 14484-14491.
- [5] WANG Y S, YANG X G, HUANG L Y, et al. Transit signal priority for emergency vehicle based on cooperative vehicle infrastructure system[C]// 7th Chinese Intelligent Transport Annual Meeting Symposium, 2012: 51-58.
- [6] TIAN F, WANG Q. Green light checkless system for special vehicle[J]. Journal of Heilongjiang Institute of Technology, 2009, 23(2): 44-47.
- [7] NEELAKANDAN S, BERLIN A M, TRIPATHI S, et al. IoT-based traffic prediction and traffic signal control system for smart city[J]. Soft Computing, 2021, 25(18): 1-8.
- [8] XIAO Y Z, ZHAO X, FENG L. Priority control of networked emergency vehicles based on traffic wave theory[J]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 2023(6): 1021-1029.
- [9] LONG K J, ZHANG Z G, LIU Y, et al. Cooperative optimization model for emergency vehicles using reverse left-turn lanes and traffic signals[J]. Journal of Transportation Systems Engineering and Information Technology, 2023(5): 194-201, 226.
- [10] ZHAO X, LI C, CHEN S S. Research on traffic strategy of emergency vehicle based on variable approach lane[J]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 2017, 41(1): 6-11, 16.
- [11] LUO Q, ZHAO X, CHEN X, et al. Research on signal priority control for emergency vehicle based on lane-level navigation[J]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 2023(4): 594-599.
- [12] ZHANG X, YU X, WU X. Exponential rank differential evolution algorithm for disaster emergency vehicle path planning[J]. IEEE Access, 2021, 9: 10880-10892.
- [13] ZHANG Y H. Research on route planning algorithm for emergency rescue vehicles under the environment of connected vehicle[D]. Beijing: Beijing Jiaotong University, 2019.
- [14] MIYAWAKI M, YAMASHIRO Z, YOSHIDA T. Fast emergency preemption systems (fast) [C]// International Conference on Intelligent Transportation System, 1999: 993-997.
- [15] YANG Z S, SUN X M, SUN P C. Signal priority control strategy and im-

- plementation for emergency vehicle at single intersection under traffic accidents[J]. Journal of Jilin University (Engineering and Technology Edition), 2011, 41(3): 640–644.
- 杨兆升, 孙晓梅, 孙鹏程. 非常态事件下应急车辆单点信号优先控制策略及实现方法[J]. 吉林大学学报(工学版), 2011, 41(3): 640–644.
- [16] ZHONG L, CHEN Y. A novel real-time traffic signal control strategy for emergency vehicles[J]. IEEE Access, 2022, 10: 19481–19492.
- [17] BOUTILIER J J, CHAN Y C T. Ambulance emergency response optimization in developing countries[J]. Operations Research, 2020, 68(5): 1315.
- [18] XUE N. Application scope of rule-based and model-based traffic signal control methods[J]. Traffic & Transportation, 2019, 35(4): 50–53.
- 薛凝. 基于规则和基于模型的交通信号控制方法适用范围[J]. 交通与运输, 2019, 35(4): 50–53.
- [19] CAO M, LI V O K, SHUAI Q. A gain with no pain: exploring intelligent traffic signal control for emergency vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 17899–17909.
- [20] SU H, ZHONG Y D, DEY B, et al. EMVLight: a decentralized reinforcement learning framework for efficient passage of emergency vehicles [C]// AAAI Conference on Artificial Intelligence, 2022: 4593–4601.
- [21] GARG D, CHLI M, VOGIATZIS G. A deep reinforcement learning agent for traffic intersection control optimization [C]// IEEE Intelligent Transportation Systems Conference, 2019: 4222–4229.
- [22] LONG W M, CHU D F, SHI H, et al. Algorithm research on traffic priority for emergency vehicles based on cooperative vehicle infrastructure system[J]. China Safety Science Journal, 2015(7): 141–146.
- 龙文民, 褚端峰, 施辉, 等. 车路协同环境下紧急车辆优先通行方法研究[J]. 中国安全科学学报, 2015(7): 141–146.
- [23] WANG Y P, GUO G. Signal priority control for trams using deep reinforcement learning[J]. Acta Automatica Sinica, 2019, 45(12): 2366–2377.
- 王云鹏, 郭戈. 基于深度强化学习的有轨电车信号优先控制[J]. 自动化学报, 2019, 45(12): 2366–2377.
- [24] BENEDETTI G, FANTI M P, MANGINI A M, et al. Application of deep reinforcement learning for traffic control of road intersection with emergency vehicles [C]// IEEE International Conference on Systems, Man, and Cybernetics, 2021: 182–187.
- [25] WEI H, ZHENG G, YAO H, et al. Intellilight: a reinforcement learning approach for intelligent traffic light control [C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 2496–2505.
- [26] LIU X, LI A, CHENG W. Multi-emergency vehicle signal priority control based on deep reinforcement learning[J]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 2021(6): 1056–1061.
- 刘翔, 李艾, 成卫. 基于深度强化学习的多应急车辆信号优先控制[J]. 武汉理工大学学报(交通科学与工程版), 2021(6): 1056–1061.
- [27] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529–533.
- [28] LOPEZ P A, BEHRISCH M, BIEKER-WALZ L, et al. Microscopic traffic simulation using sumo [C]// 21st International Conference on Intelligent Transportation Systems, 2018: 2575–2582.
- [29] AULT J, SHARON G. Reinforcement learning benchmarks for traffic signal control [EB/OL]. <https://openreview.net/forum?id=LqRSh6V0vR>.
- [30] MA J, WU F. Feudal multi-agent deep reinforcement learning for traffic signal control [C]// Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems, 2020: 816–824.
- [31] CHEN C, WEI H, XU N, et al. Toward a thousand lights: decentralized deep reinforcement learning for large-scale traffic signal control [C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 3414–3421.
- [32] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms [DB/OL]. <https://arxiv.org/abs/1707.06347>.
- [33] AULT J, HANNA P J, SHARON G. Learning an interpretable traffic signal control policy [DB/OL]. <https://arxiv.org/abs/1912.11023>.
- [34] HUANG X, WU D, BOULET B. Traffic signal control using lightweight transformers: an offline-to-online RL approach [DB/OL]. <https://arxiv.org/abs/2312.07795>.

(责任编辑: 刘嘉文)