

基于行人安全的交通信号灯智能控制算法研究

张乾隆, 胡智群, 肖海林

(湖北大学 计算机与信息工程学院, 武汉 430062)

摘要: 提出了一种基于深度确定性策略梯度 (DDPG, deep deterministic policy gradient) 的行人安全智能交通信号控制算法; 通过对交叉口数据的实时观测, 综合考虑行人安全与车辆通行效率, 智能地调控交通信号周期时长, 相位顺序以及相位持续时间, 实现交叉路口安全高效的智能控制; 同时, 采用优先经验回放提高采样效率, 加速了算法收敛; 由于行人安全与车辆通行效率存在相互矛盾, 研究中通过精确地设计强化学习的奖励函数, 折中考虑行人违规引起的与车辆的冲突量和车辆通行的速度, 引导交通信号灯学习路口行人的行为, 学习最佳的配时方案; 仿真结果表明在动态环境下, 该算法在行人与车辆冲突量, 车辆的平均速度、等待时间和队列长度均优于现有的固定配时方案和其他的智能配时方案。

关键词: 交通信号灯; 动态配时; 强化学习; 行人安全; 车辆效率; 优先经验回放

Research on Intelligent Control Algorithm of Traffic Light Based on Pedestrian Safety

ZHANG Qianlong, HU Zhiquan, XIAO Hailin

(School of Computer and Information Engineering, Hubei University, Wuhan 430062, China)

Abstract: An intelligent traffic signal control algorithm based on Deep Deterministic Policy Gradient (DDPG) with Pedestrian Safety is proposed. Through the real-time observation of intersection data, the pedestrian safety and vehicle traffic efficiency are comprehensively considered, and the cycle duration, phase sequence and phase duration of traffic signals are intelligently controlled, safe and efficient intelligent control of intersections is realized. Meanwhile, priority empirical replay is adopted to improve sampling efficiency and accelerate algorithm convergence. Due to the contradiction between pedestrian safety and vehicle traffic efficiency, the reward function of reinforcement learning is inaccurately designed, the pedestrian-vehicle conflicts caused by pedestrian violations and the speed of vehicles is considered, traffic light is guided to learn pedestrian behaviors at intersections, and the best timing scheme is learned. The simulation results show that in the dynamic environment, the algorithm in terms of the number of collisions between pedestrians and vehicles, the average speed of vehicles, waiting time and queue length are better than the existing fixed timing schemes and other intelligent timing schemes.

Keywords: traffic signal light; dynamic timing; reinforcement learning; pedestrian safety; vehicle efficiency; prioritized experience replay

0 引言

随着城市规模的不断扩大, 城市交通拥堵问题愈加严重。为了解决拥堵问题, 如何提高车辆通行效率近几年来被广泛地进行研究。然而, 除了驾驶车辆, 行人也是使用道路的主要元素。追求车辆通行效率所带来的潜在行人安全问题也不容忽视。据交通事故统计^[1] 22%的行人死亡事故发生在有标志的人行横道上, 超过 30%的行人死亡可以归因于非法穿越行为。因此, 综合考虑车辆效率和行人安全的交通信号控制机制变得至关重要。

传统的基于效率的交通信号控制通常是基于历史车流

量预设的固定配时方案, 或者基于当前车流状态调整交通信号时长, 或者基于人工智能算法智能调整信号配时方案。例如 SCOOT^[2] 和 SCATS^[3] 系统已在全世界广泛使用。但是这些信号控制方案缺乏自适应性和预见性, 甚至导致大量的人力负担。近年来, 人工智能技术在交通控制领域的应用有效地解决 SCOOT, SCATS 系统所面临的问题。文献 [4-5] 提出利用 Q 学习算法将交通信号控制问题转换为马尔可夫决策问题, 通过与交通环境的不断交互反馈, 无需复杂的数学模型, 不断积累经验学习最佳的控制策略。由于 Q 学习需要通过表来存储所学习到的 Q 值, 当网络状态复杂且高维时, 需要占用较大的内存资源, 无法适应于复

收稿日期: 2021-11-08; 修回日期: 2021-12-01。

基金项目: 国家自然科学基金(61901163)。

作者简介: 张乾隆(1996-), 男, 河南商丘人, 硕士研究生, 主要从事机器学习、智能交通方向的研究。

通讯作者: 胡智群(1989-), 女, 湖北龙穴人, 博士, 副教授, 研究生导师, 主要从事无线网络性能分析与算法设计的相关方向的研究。

肖海林(1976-), 男, 湖北麻城人, 博士, 教授, 博士生导师, 主要从事车载无线通信的相关方向的研究。

引用格式: 张乾隆, 胡智群, 肖海林. 基于行人安全的交通信号灯智能控制算法研究[J]. 计算机测量与控制, 2022, 30(4): 114-120.

杂的场景。文献[6-7]对 Q 学习进行改进, 将 DQN 算法应用在交通信号控制, 通过深度神经网络对 Q 值进行近似, 选取最佳交通信号灯动作。由于 DQN 算法依赖于找到最大近似 Q 值, 当处理连续动作的交通场合时, 需要对每个步骤迭代优化, 随着迭代的增加, 量化误差越来越大, 无法在连续复杂的交通网络中找到最优策略。文献[7]提出利用 DDPG 算法解决 Q 学习应用在交通场合时动作空间不连续问题, 利用策略网络和目标网络组成的双网络结构, 在与交通环境交互探索时对所选 Q 值动作不断进行梯度计算调整, 使得智能体倾向于生成车辆效率更好的动作选择, 不断引导智能体找到最优策略。这些策略方法根据路口车流量自适应的调整配时方案, 能够有效地提高路口的通行效率, 缓解交通拥堵问题。但是这些方案只注重驾驶车辆, 没有考虑行人因素。由于车辆和行人速度相差较大, 只考虑车辆拥堵现象, 会导致绿灯期间行人不能完全通过马路, 与车辆产生冲突, 安全受到威胁。据文献[8]研究发现 79% 的行人在交通灯变红之前没有通过交叉路口。

在有关行人安全过马路的研究中, 文献[9]提出一种行人手动激活控制方案, 在有过马路需求时, 行人通过手动按下激活按钮改变交通信号灯状态, 以保障安全通过路口。但是经过研究发现^[10], 这种将行人与车辆分离的控制方案, 实际上大大影响了交通信号灯控制系统的整体效率。Guo 等人^[11]和 Midenet 等人^[12]的研究发现车辆和行人信号协调控制对交叉口安全具有积极影响。

为了进一步探讨车辆与行人行为对交通信号控制的影响, 通过仿真、数学模型等手段研究协调控制被广泛研究。文献[13]考虑行人因素的前提下, 提出一种基于遗传算法的交通信号灯控制方案。通过适应度函数表示车辆与行人的等待数量, 评价候选染色体的有效性来调控信号灯。然而所考虑的仅仅是减少行人或者车辆的等待队列长度, 没有考虑车辆和行人整体协调, 在遇到两者不公平的情形下, 设定车辆优先级大于行人。文献[14]综合考虑了车辆和行人的队列长度, 提出一种基于分布式多智能体强化学习的交通信号控制方案, 实现车辆与行人等待时间最小化。但是忽略了行人在绿灯期间可能无法完全通过路口, 造成与车辆冲突的问题。文献[15]提出了一种考虑车辆总延迟和行人安全的全网络交通信号策略。通过建立行人-车辆混合流网络宏观模型智能调控交通信号灯, 过程中考虑到绿灯闪烁期间行人违规率较高, 在每个绿灯闪烁期间结束时引入一个动态全红阶段, 并根据违规行人的数量自适应调整其持续时间。但是动作空间里固定了绿灯闪烁时间而且持续时间较长, 由于行人往往忽略绿灯闪烁剩余时间强行通过路口, 会导致每一动态全红阶段持续时间都较长, 大大影响了车辆通行效率。

为了协调车辆效率和行人安全, 本文提出了一种基于行人安全的交通信号灯智能控制方案, 通过对交叉口流量的实时观测, 综合考虑行人违规和车辆效率存在的矛盾, 精准设计强化学习架构中的奖励函数引导智能体学习, 智

能地调控交通信号周期时长, 相位顺序以及相位持续时间, 实现交叉路口安全高效的智能控制, 同时, 采用基于优先经验的样本采样和方差缩放, 加速了算法收敛。研究结果表明, 基于行人安全配时方案在行人与车辆冲突量, 车辆的排队长度、等待时间和平均速度等方面均优于固定配时方案和 DQN 配时方案。

1 系统模型和算法设计

1.1 DDPG 算法

在交通信号灯控制中, 自适应地调整周期时长、相位顺序和相位持续时间可以看成典型的马尔可夫决策问题, 它很好的符合强化学习 (RL, reinforcement learning) 框架。智能体通过和交叉路口环境交互, 获取环境状态 s_t , 随后采用“试错”机制寻找最佳策略 π 。在每次探索过程中, 智能体根据所选策略执行动作 a_t , 动作执行后将积累的奖励值 r_t 和交叉口下一状态 s_{t+1} 返还给智能体, 智能体以奖励值为导向对动作评估, 不断更新策略, 如此循环以寻找最大预期值 Q 。如公式 (1):

$$Q^{\pi}(s, a) = E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots | s_t = s, a_t = a, \pi] = E\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t = s, a_t = a, \pi\right] \quad (1)$$

式中, $Q(s, a)$ 定义为智能体接收到交叉路口车辆和行人状态 s 时, 交通信号灯采取动作 a 时所获得的未来累计奖励; $\gamma \in [0, 1]$ 为折扣因子, 表明了未来交通信号灯配时方案回报相对于当前配时方案回报的重要程度。

随后通过贝尔曼公式不断更新值函数直至逼近最优值函数, 由此智能体学习最优策略 π^* , 如公式 (2):

$$Q^*(s, a) = E_t[r_t + \gamma \max_{a'} Q^*(s', a') | s, a] \quad (2)$$

DDPG 算法是一种基于策略梯度的 RL 算法, 可以通过与交通系统在线互动输出连续的动作。算法采用行动者-评判家 (AC, actor-critic) 框架, 通过相同的神经网络架构, 构成当前 Actor 网络和 Critic 网络, 目标 Actor 网络和 Critic 网络, 协作完成算法的决策与更新。

当前 Actor 网络采用确定性策略 μ 输出当前动作 a_t , 表示为 $a_t = \mu(s_t | \theta^{\mu})$, 然后通过目标函数 J 评价策略 μ , 引导训练寻找最优策略, 即 $\mu = \arg \max_{\mu} J(\mu)$ 。其中 θ^{μ} 是产生确定性动作的网络参数, 用于交叉路口车辆和行人状态到交通信号灯配时方案的映射。当前 Actor 网络通过链式规则更新, 如公式 (3):

$$\nabla_{\theta^{\mu}} J \approx E[\nabla_{\theta^{\mu}} Q(s, \mu(s | \theta^{\mu}) | \theta^Q) | \theta^Q] = E[\nabla_{\theta^{\mu}} Q(s, \mu(s | \theta^{\mu}) | \theta^Q) \nabla_{\theta^{\mu}} \mu(s | \theta^{\mu})] \quad (3)$$

当前 Critic 网络用于拟合并参数化 Q 函数为 $Q(s, a | \theta^Q)$, 并采用均方误差形式完成对当前交通信号灯配时方案的实时评估, 具体形式如公式 (4):

$$L = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i | \theta^Q))^2 \quad (4)$$

式中, $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1} | \theta^{\mu}') | \theta^Q)$, μ' 和 Q' 分别对应目标 Actor 网络和目标 Critic 网络。其中, 目标 Actor 网络

用于输出估计的交通信号灯配时方案, 目标 Critic 网络用于输出对估计的交通信号灯配时方案评估。

当前 Actor 网络和 Critic 网络通过从记忆库中采样实现更新, 而目标 Actor 网络和 Critic 网络的更新采用软更新的形式, 能够使得目标网络参数变化小, 训练更易于收敛。软更新的具体形式如式 (5):

$$\theta^i \leftarrow \tau \theta^i + (1 - \tau) \theta^i \quad (5a)$$

$$\theta^q \leftarrow \tau \theta^q + (1 - \tau) \theta^q \quad (5b)$$

1.2 DDPG 算法改进

1.2.1 交叉口行人交通背景

行人在穿越路口时, 往往不会注意绿灯剩余时间, 经常会发生还没有完全通过路口, 信号灯就变成了红灯的情况; 老年人和儿童在通过路口的时候, 往往会因为速度过慢而不能在绿灯期间通过路口; 携带较多、较重物品的行人在通过路口时同样也会遇到相同的情况。这些情况下, 行人安全难以保证。行人在规定绿灯时间内, 不能按时通过交叉口, 极易和车辆产生冲突, 一方面行人安全难以保证, 另一方面造成交通阻塞。以前的交通信号灯智能控制方案往往只专注于提高车辆效率, 而忽略了行人, 这实际上有很大的安全隐患。因此, 智能交通有必要将行人不确定因素考虑进去, 尤其是行人较多的例如学校、医院等旁边的交叉路口。

通过实际行人交通背景我们可以知道, 行人经常会发生不能在规定时间内通过交叉口的情况^[16-17]。因为强化学习特殊的“试错”学习机制, 这实际上可以作为一条条宝贵的“学习经验”供智能体学习, 以保证智能体能够有效的学习行人行为, 进而保障行人安全。

1.2.2 PRE-DDPG 算法设计

DDPG 算法中当前 Actor 网络和 Critic 网络在采样更新时认为记忆库中所有经验都一样重要, 因此它等概率地从经验池中选取一定数目的样本来训练网络。而实际上, 通过行人交通背景分析知道, 在交叉口行人不能按时通过时, 智能体理应获取一个较差的“经验”, 以避免这种情况的发生。因此, 为了更有效地保证行人安全, 合理处理交叉口行人与车辆行为实际经验, 结合奖励值函数的精准设计, 提出一种基于优先经验回放的 DDPG (PER-DDPG, DDPG with prioritized experience replay) 行人安全算法, 并在记忆库满了后对方差进行 0.999 的缩放, 以提高优先级大的样本的利用率, 进而提高学习的效率, 加快算法收敛。TD 误差 $\delta(t)$ 计算如式 (6) 所示:

$$\delta(t) = R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t) \quad (6)$$

定义抽取样本的概率如式 (7) 所示:

$$p_i = |\delta(t)| + \epsilon \quad (7)$$

其中: ϵ 用于防止概率为 0。

为了避免每次抽样都需要针对 p 对所有样本排序, 采用 SumTree 的树形结构对数据进行存储。SumTree 是一种树形结构, 每片树叶存储每个样本的优先级 p , 每个树枝

节点只有两个分叉, 节点的值是两个分叉的和。数据采样时, 将优先级 p 的总和除以抽样数, 得到区间数, 然后在每个区间里随机选取一个数, 将此数从 SumTree 的根节点开始按照一定规律向下搜索, 最后将搜索得到的优先级 p 与样本数据相对应, 即可实现更为高效的经验回放。SumTree 结构如图 1 所示。

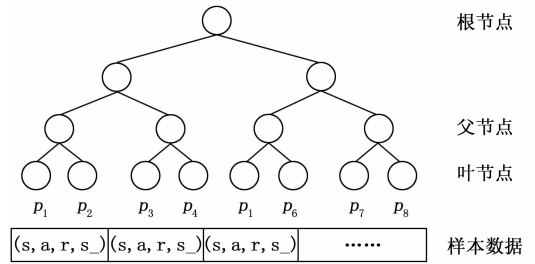


图 1 SumTree 结构图

1.3 基于行人安全智能交通灯控制模型

本小节主要阐述行人安全智能交通灯控制算法中状态、动作和奖励的设计和信号灯控制模型的建立。

1.3.1 状态

状态为在每一时间步中, 智能体接收到关于交叉路口环境的定量描述。本设计将路口车辆等待数量, 行人与车辆的冲突量作为状态输入。定义状态表达式如下所示:

$$s_t = \{w_1, \dots, w_n, v_1, \dots, v_n\} \quad (8)$$

式中, w_1, \dots, w_n 表示道路 nm 上车辆的平均等待数量; v_1, \dots, v_n 表示人行横道 n 上行人与车辆的冲突量。

1.3.2 动作

本设计将动作设置为交通信号灯周期可变、相位可变和相位持续时间可变。定义动作如式 (9) 所示:

$$a_t = \{c, p_1, \dots, p_6, d_1, \dots, d_m\} \quad (9)$$

式中, cc 代表决定下一周期持续时间的时长, 为避免周期过大或过小, 将周期限制在内 $[c_{\min} T, c_{\max} T], c_{\min}, c_{\max}, T$ 分别取 0.6, 2, 90; p_1, \dots, p_6 表示下一周期相位持续顺序; d_1, \dots, d_m 表示下一周期内各阶段的持续时长的比例因子, 通过与环境交互后由 softmax 归一化处理确定。

1.3.3 奖励

为有效保证行人安全, 我们设定只有在绿灯亮起范围内通过路口的行人是合法的, 其他各阶段行人出现在人行横道上均视为违规行为, 视为与车辆正常行驶发生冲突, 此时行人数量称为行人与车辆的冲突量。由此, 在评价行人过路口是否安全时, 我们以每一时刻行人与车辆的冲突量多少来反映, 智能体通过学习减少冲突量来保障行人安全。在评价车辆通行效率时, 评价指标为每条道路上车辆的平均速度。

综上所述, 本文充分考虑行人安全和车辆通行效率存在的矛盾及行人与车辆不同行为所带来的不同经验, 定义奖励函数如下所示:

$$Reward = \frac{1}{1 - e^{-r}} (1 - e^{-br}) \times C \quad (10)$$

式中, 当发生行人与车辆冲突时, $k = 0$, 当未发生行人与车辆冲突时, 随着车辆通过路口速度增大, k 值增大, 对速度归一化处理使 $k \in (0, 1]$; $k \in (0, 1]$; r 表示幂函数的参数, r 值越大, 在行人与车辆未发生冲突时, 奖励值大小受车辆平均速度大小的影响效果越大, 实验时合理设置 $r > 0$ 即可, 本实验选取 $r = 10$; C 表示波动因子, 用于平衡南北一东西两方向车辆通行速度。具体的 $C = \frac{\min(V_1, V_2)}{\max(V_1, V_2)}$, V_1, V_2 分别为东西和南北两方向上车辆的平均速度, 当两方向的通行速度相似时, $Reward$ 基本不受影响, 而当两方向通行速度相差过大时 CC 值很小, 从而引入的波动因子有效的平衡了东西一南北两方向车辆通行速度。

1.3.4 基于行人安全的交通信号灯控制模型

本模型综合考虑了交叉口行人安全和车辆效率, 通过共享行人与车辆状态, 赋予路口智能体更为广阔的观测视野, 同时, 采用优先经验回放提高采样效率, 加速算法收敛, 使智能体能够通过观测的全局状态及时调控自身策略, 导出全局最优 Q 值, 基于行人安全的交通信号控制结构如图 2 所示。

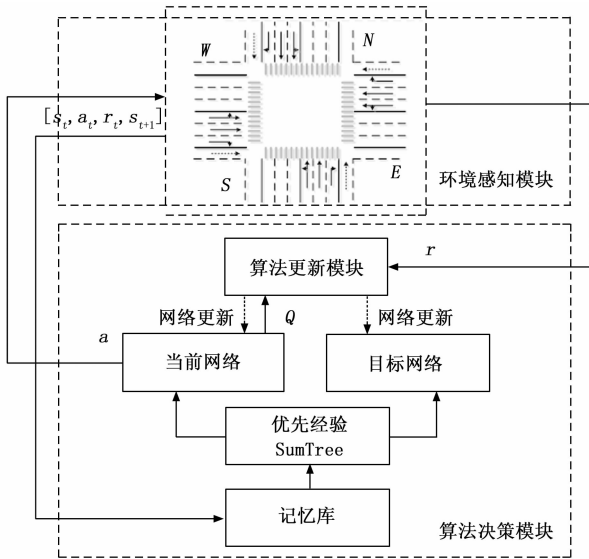


图 2 基于行人安全的交通信号控制结构示意图

每一回合开始, 读取相同的行人与车辆流量文件, 以一个周期为频率更新信号配时。当前 Actor 网络的输出确定下一周期的时长、相位顺序及相位持续时间。当一个周期结束以后通过目标 Actor 网络和目标 Critic 网络的输出值函数与当前 Critic 网络输出的值函数计算出 TD 误差, 并按概率大小存储到优先经验回放池, 当前 Critic 网络从优先经验池中按概率抽样更新, 进而指导当前 Actor 在下一时刻作出包含更少冲突量和更高车辆效率的配时方案, 以不断完善策略, 如此循环直至达到最终仿真时长。基于行人安全的交通灯智能控制模型伪代码如算法 1 所示。

算法 1 基于行人安全的交通信号灯控制模型

初始化: 随机初始化当前 Actor 神经网络 $\mu(s_t | \theta^a)$ 和当

前 Critic 神经网络 $Q(s, a | \theta^q)$, 初始化目标网络 μ' 和 Q' , 记忆库 R 和高斯噪声 N 。

算法:

- 1) for episode = 1, M do:
- 2) 在 SUMO 平台上导入行人与车辆流量文件
- 3) 随机初始化分布 N 用作动作探索
- 4) 初始化初始状态 s_0
- 5) for step = 1, 10800 do:
- 6) Actor 网络获取当前路口行人与车辆冲突量信息和路口车辆通行速度信息, 并根据当前的策略选取动作传递给交通信号灯执行, 即 $a_t = \mu(s_t | \theta^a) + N$
- 7) Actor 网络将获取的路口行人与车辆数据 (s_t, a_t, r_t, s_{t+1}) 存储在记忆库中, 以作为网络训练学习的数据集。
- 8) 计算记忆库 R 中样本抽样概率
- 9) $p_t = |\delta(t)| + \epsilon$
- 10) 将搜索得到的优先级 p 与样本数据相对应, 作为当前 Actor 网络和当前 Critic 网络的训练数据
- 11) $y_t = r_t + \gamma Q'(s_{t+1}, \mu'(s_{t+1} | \theta^q) | \theta^q)$
- 12) 记忆库存储满后对方差进行 0.999 的缩放
- 13) 根据公式(3)更新当前 Actor 网络
- 14) 根据公式(4)更新当前 Critic 网络
- 15) 更新目标网络
- 16) $\theta^a \leftarrow \tau \theta^a + (1 - \tau) \theta^a$
- $\theta^q \leftarrow \tau \theta^q + (1 - \tau) \theta^q$
- 17) end for
- 18) end for

2 仿真实验

2.1 仿真环境与参数设置

交通仿真环境采用 SUOM1.9.0 (simulation of urban mobility, 城市交通仿真) 搭建, 并利用 TraCI 接口完成与算法的在线交互。算法环境采用 python3.6.6 + TensorFlow1.1.2 搭建完成。

2.1.1 单路口属性设置

针对单路口进行仿真分析, 道路为 4 车道, 每条路长 150 m, 从里到外分别为: 左转或直行, 直行, 右转, 人行道。其中车道限速为 34 km/h, 人行道限速为 1 m/s, 人行横道限速为 1.3 m/s。在仿真中不考虑车辆右转的情况。

2.1.2 交通灯初始属性设置

交通灯属性设置按 SUMO 官方推荐四车道设置, 具体见表 1。

表 1 交通信号灯初始相位设置

交通信号灯初始相位	持续时间/s
直行、左/右转(行人绿灯)	33
直行、左/右转(行人红灯)	7
黄灯	5

2.1.3 流量设置

车辆和行人的到达服从泊松分布, 根据在不同时间段

更改车辆和行人到达概率有效模拟出“早高峰”、“晚高峰”和“超低峰”等场景，从而反映出一天内真实的流量环境。将一天的交通流量进行压缩至 3 个小时 (10 800 s) 进行仿真实验。分为 0~1 350 s, 1 350~3 150 s, 3 150~4 050 s, 4 050~5 400 s, 5 400~6 300 s, 6 300~7 650 s, 7 650~8 550 s, 8 550~10 800 s 共 8 个时间段对流量进行统计。各时间段内的到达流量如图 3 所示。车辆随机进入交叉口的入口并提前选择所在车道，驶出路口时随机选择目标车道，行人随机进入交叉口的入口，直行通过路口。

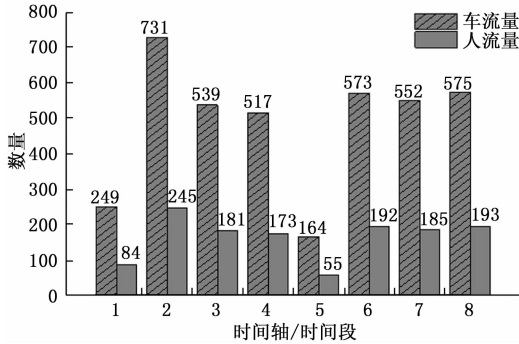


图 3 车辆与行人流量分布柱状图

2.1.4 参数选取

神经网络参数及其余参数选取参考文献[18-19]，实际使用中根据学习情况作适当调整，具体参数如表 2 和表 3 所示。

表 2 神经网络参数

Actor 网络参数	层数	维度	激活函数
输入层	1	24	NA
隐藏层	2	200,200	Relu,Relu6
输出层	1	6	Tanh
Critic 网络参数	层数	维度	激活函数
输入层	1	30	NA
隐藏层	2	200,200	Relu,Relu6
输出层	1	6	Tanh

表 3 其余参数

参数名	数值
Actor 网络学习率 LR-A	0.000 001
Critic 网络学习率 LR-C	0.000 001
折扣因子 λ	0.95
软交换系数 τ	0.99
批处理大小	32
奖励值参数 r	10
周期系数 C_{min}, C_{max}	0.6, 2
固定周期时长 T	90

2.2 对比方案分析

为了测试模型的有效性，将本文方法与以下方法进行对比。各方案配备相同的路口属性和行人与车辆流量，如

2.1.1 和 2.1.3 所示。

1) 固定配时方案。交通灯的相位及相位持续时间设置如 2.1.2 节，方案中利用 SUMO 的 TraCI 接口直接获取每一时刻行人与车辆冲突量，车辆的平均速度等数据来反映路口车辆通行效率和行人安全度。

2) DQN 算法配时方案^[20]。状态设定为路口车辆位置-速度矩阵和行人位置-速度矩阵；初始动作设定为每个子阶段 10 s，学习后智能体根据路口流量动态地改变相位持续时间 (可以为 0 s)；奖励值函数设定同行人安全配时方案保持一致。方案中相关参数设置见表 4 所示。

表 4 DQN 配时方案相关参数

参数名	数值
网络学习率 α	0.000 01
贪婪率 ϵ	0.95
记忆库大小 R	80 000
状态维度	56 * 20
动作维度	6
批处理大小	32
初始子阶段时长	10

2.3 实验

为了验证本研究的有效性和可用性，需要进行仿真实验回答以下两个问题：

问题 1：本研究是否能够提高行人安全性？即在相同的实验环境下与对比方案相比，平均冲突量是否有明显降低？

问题 2：本研究是否兼顾了行人安全和车辆效率？即在相同的实验环境下与对比方案相比，路口的总体性能如何？

强化学习中学习率的选取至关重要，实验设计验证前先进行学习率的确定，然后进行对比实验。

2.3.1 PRE-DDPG 算法学习率选取

图 4 描述了单路口下基于行人安全配时方案在不同学习率下的收敛结果。从图 4 中可以看出学习速率过大或者过小都可能带来性能的损失。当 LR-A=LR-C=0.000 001 时，算法表现出较好的性能优势。因此在本小节后续的仿真中，算法学习率均选择 LR-A=LR-C=0.000 001。

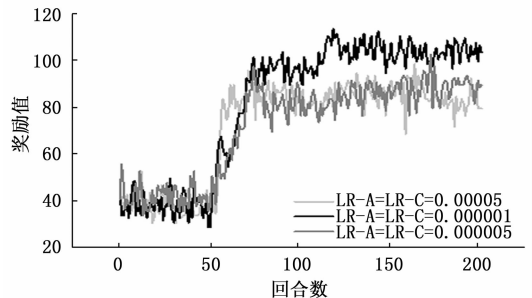


图 4 不同学习率的行人安全配时方案训练过程中回合奖励值性能对比

2.3.2 实验设计及结果分析

为了回答以上 2 个问题，本小节进行了如下实验设计：

1) 针对问题 1, 进行对比实验 1:

为了验证交叉口行人安全是否得到提高, 即各配时方案中行人与车辆的平均冲突量情况, 在同一实验环境下将所提行人安全配时方案与其它对比方案作对比实验。

(1) 各配时方案行人与车辆平均冲突量对比

图 5 即为各配时方案中行人与车辆的平均冲突量对比。从图 5 可以看到, 在相对于其他配时方案, 基于学习的配时方案能够根据人流量、车流量的情况实时地调整最佳配时方案, 有效地减少行人与车辆的相互冲突。相对于 DQN 配时方案, 我们所提出的行人安全配时方案平均冲突量为 0.66, 减少了 16.40% 的行人与车辆的事故率, 有效地改善了行人过马路的安全性。

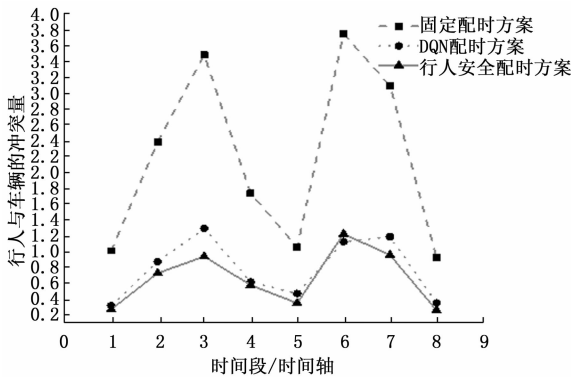


图 5 行人与车辆的平均冲突量对比

2) 针对问题 2, 进行对比实验 2:

为了验证所提行人安全配时方案的总体性能, 在同一实验环境下对各方案作摘除行人因素的对比实验。

(1) 考虑行人安全下, 各配时方案车辆平均速度对比。

图 6 为各配时方案中车辆在平均速度的对比情况。从图中可以发现, 本文所提出的方案在平均速度性能上优于对比方案, 特别是在高峰车流量下 (例如, 时间段 3)。通过对各时段的数据进行平均处理, 我们可以得到行人安全配时方案中每条道路上车辆平均速度为 21.40 km/h。相对于 DQN 配时方案均速度提高了 40.28%, 相对于固定配时方案平均速度提高了 29.62%, 车辆效率有效得到改善。同时可以看到, DQN 配时方案相对于固定配时方案车辆平均速度减少了 5.40%, 平均冲突量减少了 64.04%, 可以明显看出 DQN 配时方案为保证行人安全通过交叉口牺牲了一定的车辆通行效率。

(2) 摘除行人因素前后, 各配时方案车辆平均时延对比。

为了更好地比较各方案中行人因素对车辆效率的影响, 分别对各方案作摘除行人因素的对比实验。图 7 为各方案中每条路上车辆平均时间延迟的对比情况。时间延迟为车辆实际行驶时间与最大行驶速度行驶时间的差值。从图 7 可以看到各配时方案中行人因素对于车辆效率的影响:

(1) 将行人安全配时方案和 DQN 配时方案分别做摘除行人因素前后车辆平均时延对比, 得到所提行人安全配时

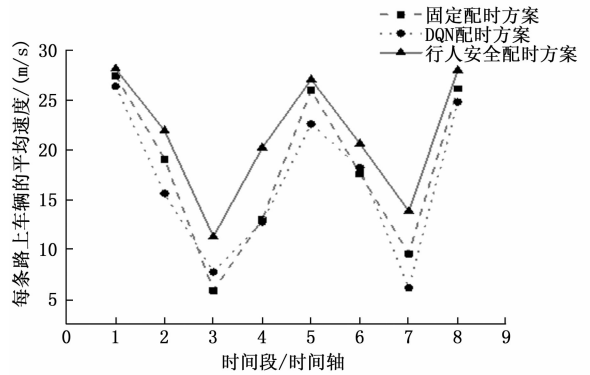


图 6 车辆平均速度性能对比

方案由不考虑行人时的车辆平均时延 2.59 s 增长到 3.32 s, DQN 配时方案由 5.46 s 增长到 7.05 s。可以看出, 交叉口考虑行人安全会对车辆效率产生一定的负面影响。

(2) 尽管考虑行人安全会对车辆效率产生一定的负面影响, 但所提基于行人安全配时方案在兼顾行人安全和车辆效率方面, 优势仍然比较明显。对比车辆平均时延可以看到, 行人安全配时方案相对于 DQN 配时方案 (不考虑行人) 有 39.16% 的提高, 相对于 DQN 配时方案有 52.90% 的提高, 相对于固定配时方案有 47.54% 的提高。

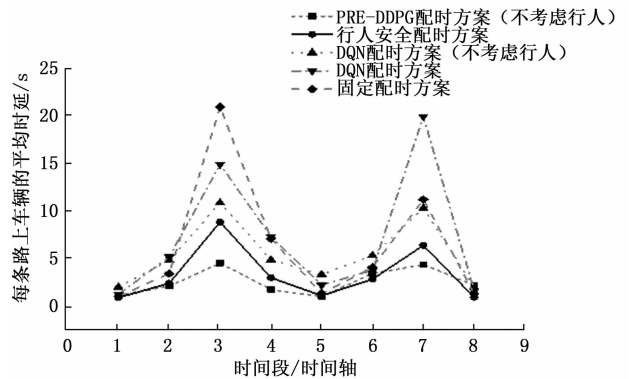


图 7 车辆平均时间延迟对比

3 结束语

提出了基于行人安全的交通信号灯智能控制方案, 折中考虑行人安全与车辆通行效率, 实现系统效用的最大化。相对于传统配时方案, 所提方案考虑行人因素并且自适应性强, 能够在连续强大的交通环境中提取合适的特征, 选取合适的动作。研究结果显示, 该算法模型具有较好的收敛性, 在兼顾行人安全和车辆效率上都明显优于固定配时方案和 DQN 配时方案。本文仅考虑了单路口时行人和车辆问题, 没有做多路口的情况, 下一步的研究主要集中在多路口协调下的行人和车辆问题。

参考文献:

[1] Road Traffic Accident Database 2009—2011; Traffic Police Department [R]. SPF, Singapore, 2013.

[2] HUNT P B, ROBERTSON D I, BRETHERTON R D, et al. The SCOOT on-line traffic signal optimisation technique [J]. *Traffic Engineering & Control*, 1982, 23 (4): 190 - 192.

[3] LUK J Y K. Two traffic-responsive area traffic control methods: SCAT and SCOOT [J]. *Traffic Engineering & Control*, 1984, 25 (1): 14 - 22.

[4] ABDULHAI B, PRINGLE R, KARAKOULAS G J. Reinforcement learning for true adaptive traffic signal control [J]. *Journal of Transportation Engineering*, 2003, 129 (3): 278 - 285.

[5] 戈 军, 周蓬英. 基于 SARSA (λ) 的实时交通信号控制模型 [J]. *计算机工程与应用*, 2015, 51 (24): 244 - 248.

[6] GENDERS W, RAZAVI S. Using a deep reinforcement learning agent for traffic signal control [J]. *arXiv Preprint. arXiv: 1611. 01142*, 2016, 1: 1 - 9.

[7] 颜文胜, 吕红兵. 基于 Deep Q Networks 的交通指示灯控制方法 [J]. *计算机测量与控制*, 2021, 29 (6): 93 - 97.

[8] PANG H, GAO W. Deep deterministic policy gradient for traffic signal control of single intersection [C] //2019 Chinese Control And Decision Conference (CCDC), IEEE, 2019: 5861 - 5866.

[9] ZHUANG X, WU C, MA S. Cross or wait? Pedestrian decision making during clearance phase at signalized intersections [J]. *Accident Analysis & Prevention*, 2018, 111: 115 - 124.

[10] SULMICKI M. Don' t push the red button: a case against manual pedestrian detection in urban areas [J]. *Transportation research procedia*, 2016, 14: 4314 - 4323.

[11] GUO F, WANG X, ABDELATY M A. Modeling signalized intersection safety with corridor-level spatial correlations [J]. *Accident Analysis & Prevention*, 2010, 42 (1): 84 - 92.

[12] MIDENET S, SAUNIER N, BOILLOT F. Exposure to lateral collision in signalized intersections with protected left turn under different traffic control strategies [J]. *Accident Analysis & Prevention*, 2011, 43 (6): 1968 - 1978.

[13] TURKY A M, AHMAD M S, YUSOFF M Z M. The use of genetic algorithm for traffic light and pedestrian crossing control [J]. *International Journal of Computer Science and Network Security*, 2009, 9 (2): 88 - 96.

[14] LIU Y, LIU L, CHEN W P. Intelligent traffic light control using distributed multi-agent Q learning [C] //2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2017: 1 - 8.

[15] ZHANG Y, ZHANG Y, SU R. Pedestrian-safety-aware traffic light control strategy for urban traffic congestion alleviation [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 22 (1): 178 - 193.

[16] EVSTATIEV B, BALBUZANOV T, BELOEV I, et al. Intelligent system for improved safety of pedestrian traffic lights [J]. *Transport Problems*, 2019, 14: 35 - 43.

[17] NIVEDITHA S P, MALLESHA K M. Analysis of pedestrian crossing behavior at uncontrolled intersections [M]. *Recent Trends in Civil Engineering*, Springer, Singapore, 2021.

[18] 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.

[19] ZHANG Z, YANG J, ZHA H. Integrating independent and centralized multi-agent reinforcement learning for traffic signal network optimization [J]. *arXiv preprint. arXiv: 1909. 10651*, 2019: 2083 - 2085.

[20] 蔡含宇. 基于道路实时容量的交通灯智能控制算法研究 [D]. 西安: 西安理工大学, 2020.

[5] 卫荣恭, 金世俊. 滑动及侧滑影响下的移动机器人轨迹跟踪控制 [J]. *传感器与微系统*, 2019, 38 (3): 123 - 125, 136.

[6] 田 华, 何 翼. 基于熵质心聚类算法的大数据挖掘 [J]. *计算机应用研究*, 2020, 37 (12): 72 - 75.

[7] 陈 洲, 陆 南. 基于 K - Means 算法的 Web 日志用户聚类研究 [J]. *计算机与数字工程*, 2020, 48 (3): 140 - 144.

[8] 王立玲, 董力元, 马 东, 等. 滑动与打滑条件下的轮式移动机器人自抗扰跟踪控制 [J]. *控制理论与应用*, 2020, 37 (2): 206 - 213.

[9] 冯地平, 张 娜. 病案统计管理信息系统的数据质量控制分析 [J]. *养生保健指南*, 2019 (8): 229.

[10] 李慧真. 基于大数据处理技术的界面交互设计研究 [J]. *现代电子技术*, 2019, 42 (1): 46 - 49, 53.

[11] 黄正鹏, 王 力, 张明富. 基于云计算的海量大数据智能清洗系统设计 [J]. *现代电子技术*, 2020, 43 (3): 116 - 120.

[12] 农丽丽, 钟 琴, 卢志翔. 基于模糊 K-Means 聚类的光纤大数据分类平台设计 [J]. *激光杂志*, 2020, 41 (8): 145 - 150.

[13] 李双琴, 谢 锐, 曹文琛, 等. 基于多维分层采样的时间维度型大数据流整合系统设计 [J]. *现代电子技术*, 2020, 43 (5): 141 - 144, 148.

[14] 丁 灿. 健康教育在传染病预防控制中的作用 [J]. *饮食保健*, 2019, 6 (31): 269.

[15] 卢献华, 王洪俊. 基于大数据计算框架的分布式新闻聚类系统设计 [J]. *计算机科学*, 2019, 46 (2): 220 - 223.

[16] 于 涛, 孙汉旭, 赵 伟, 等. 一种球形滚动机器人的路径跟踪控制器设计 [J]. *计算机测量与控制*, 2019, 27 (3): 97 - 102.

[17] 顾万里, 胡云峰, 宫 洵, 等. 考虑参数不确定性的移动机器人轨迹跟踪控制 [J]. *控制与决策*, 2019, 34 (1): 81 - 88.

[18] 陈彦杰, 占巍巍, 张振国, 等. 作业型飞行机器人抓取后重心偏移的轨迹跟踪控制 [J]. *控制理论与应用*, 2020, 37 (10): 109 - 119.

[19] 张华胜, 宋旋璇, 潘礼正. 基于滑模变结构控制的路径跟踪研究 [J]. *计算机测量与控制*, 2019, 27 (11): 106 - 109, 114.

[20] 许洋洋, 王 莹, 薛东彬. 采用改进模糊神经网络 PID 控制的移动机器人运动误差研究 [J]. *中国工程机械学报*, 2019 (6): 510 - 514.

[21] 黄玉钊. 基于概率神经网络图像识别的移动机器人控制研究 [J]. *小型微型计算机系统*, 2019, 40 (4): 908 - 912.