基于深度强化学习的有轨电车信号优先控制

王云鹏 [1] 郭 戈 [2, 3]

摘 要 现有的有轨电车信号优先控制系统存在诸多问题, 如无法适应实时交通变化、优化求解较为复杂等. 本文提出了一种基于深度强化学习的有轨电车信号优先控制策略. 不依赖于交叉口复杂交通建模, 采用实时交通信息作为输入, 在有轨电车整个通行过程中连续动态调整交通信号. 协同考虑有轨电车与社会车辆的通行需求, 在尽量保证有轨电车无需停车的同时, 降低社会车辆的通行延误. 采用深度 Q 网络算法进行问题求解, 并利用竞争架构、双 Q 网络和加权样本池改善学习性能. 基于 SUMO 的实验表明, 该模型能够有效地协同提高有轨电车与社会车辆的通行效率.

关键词 有轨电车, 信号优先, 马尔科夫决策过程, 深度强化学习

引用格式 王云鹏, 郭戈. 基于深度强化学习的有轨电车信号优先控制. 自动化学报, 2019, **45**(12): 2366−2377

DOI 10.16383/j.aas.c190164

Signal Priority Control for Trams Using Deep Reinforcement Learning

WANG Yun-Peng[1] GUO Ge[2, 3]

Abstract Current trams-priority signal control systems have many problems, such as low adaptability to real-time traffic changes and high complexity in optimization solutions, etc. In this paper, an active signal priority control model is proposed for the trams based on deep reinforcement learning. Considering the traffic demands from tram and general vehicles, it can reduce the traffic delay of general vehicles while minimizing the need for trams to stop at the intersection. Real-time traffic information is used to dynamically adjust the sequence of traffic signals throughout the whole passing process of the tram, without relying on the complex traffic modeling. We use deep Q-network algorithm for problem-solving, and adopt dueling network, double Q network, and prioritized experience replay to improve the learning performance. Experiments based on SUMO have demonstrated that the proposed model can excellently improve the efficiency of trams and general vehicles simultaneously.

Key words Trams, signal priority, Markov decision process, deep reinforcement learning

Citation Wang Yun-Peng, Guo Ge. Signal priority control for trams using deep reinforcement learning. *Acta Automatica Sinica*, 2019, **45**(12): 2366−2377

现代有轨电车在我国应用十分广泛[1], 其与公共汽车相比具有更大的容量和更快的运行速度. 但实际中应用效率还相对较低, 一个关键原因便是有轨电车在轨道/公路交叉道口的延误严重影响了其运行可靠性[2]. 因此, 对轨道/公路交叉道口交通信号控制的研究成为城市智能交通系统的一个重要问题. 如同其他公交车辆, 利用公交信号优先 (Transit signal priority, TSP) 来提高有轨电车系统运行效率是一种主要的研究思路. 目前, 对于公交信号优先的研究主要分为两类: 被动优先和主动优先[3].

被动优先是一种离线策略. 典型的被动优先策略由应用于普通交叉口的 MaxBand 信号配时策略[4]发展而来. 文献 [5] 在保证有轨电车固定绿波带的前提下, 提出了一种最大化社会车辆绿波带的多路口信号配时优化模型. 文献 [6] 在设计公共汽车绿波带的优化模型中协同设计了交通信号与站点驻留时间. 文献 [7] 通过建立不同站点的驻留时间概率分布模型, 设计可变的绿波带来处理公共汽车驻留时间的变化. 虽然设计良好的被动优先策略能有效提高公交系统运行效率. 但是其更适用于公共汽车系统. 因为有轨电车的通行频率相对更低, 如大连 202 路有轨电车间隔约为 10 分钟, 造成绿波带浪费. 与此同时, 基于被动优先策略的交通信号优化设计均考虑交通流为固定的情况, 优化计算依赖于公交车辆与社会车辆通行量的历史数据, 无法处理

收稿日期 2019-03-15 录用日期 2019-09-02
Manuscript received March 15, 2019; accepted September 2, 2019
国家自然科学基金 (61573077, U1808205) 资助
Supported by National Natural Science Foundation of China (61573077, U1808205)
本文责任编委 阳春华
Recommended by Associate Editor YANG Chun-Hua
1. 大连理工大学控制科学与工程学院 大连 116024 2. 东北大学流程工业综合自动化国家重点实验室 沈阳 110819 3. 东北大学秦皇岛分校控制工程学院 秦皇岛 066004
1. School of Control Science and Engineering, Dalian University of Technology, Dalian 116024 2. State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819 3. School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004

公交车辆与社会车辆行驶中的实时随机变化.

主动优先是一种在线策略. 公交车辆在接近交叉口时主动向信号控制器发出信号优先请求. 随后控制器对信号进行实时调整, 使其能够不停车通过交叉口. 目前, 主动优先策略在有轨电车系统中应用广泛, 如墨尔本、多伦多和犹他州[8]. 主动优先能够行之有效地解决有轨电车在交叉口停车与延误的问题, 但会对社会车辆通行造成额外的延误, 特别是在近饱和通行状态下[9]. 因此, 近期的研究成果中, 均通过对交通信号的优化来兼顾有轨电车与社会车辆的通行需求. 文献 [2] 提出了一种基于有轨电车时刻表对主动优先策略进行优化的方法. 文献 [10] 基于历史数据对有轨电车站点驻留时间进行预测, 并对于主动优先策略中的绿灯时长和相位偏移通过非线性规划进行设计. 据我们所知, 现有的有轨电车信号主动优先控制均采用固定信号相序. 在响应信号优先请求时, 通过优化模型对某些特定信号相位进行优化. 由于优化计算提前于有轨电车到达路口, 因此主动信号优先控制同样需要预测有轨电车与社会车辆在信号请求发起后的运行状态.

上述的被/主动信号优先控制方法均可以视为“基于模型” 方法, 需要通过模型对交通运行状态进行预测, 在交通环境较为稳定时有很好的效果. 然而当存在不可预测实时随机变化时 (驾驶员的随机性, 交通突发情况等), 模型预测会发生较大的偏差, 导致实际控制效果变差. 同时, 非线性优化模型能够直接利用的交通信息非常受限, 如无法直接利用交通流信息而只能利用从中计算得到最大排队长度与消散时间, 无法通过更微观的信息来更好地把握交通变化. 因此, 针对于存在无法预测实时随机变化的交通场景, 需要一种无需交通建模的信号优先控制方法, 能够从更大范围更高维度的交通信息中提取特征以在控制中对实时的交通变化做出反应.

作为一种 “数据驱动” 的无模型自学习算法, 强化学习 (Reinforcement learning)[11−12] 因其在实时决策方面的优势, 成为交通信号控制研究的热点之一. 由于传统强化学习自身的限制, 早期成果如文献 [13], 无法处理高维且连续的交通状态信息. 而随着深度学习的快速发展[14], 借助深度神经网络 (Deep neural network, DNN) 在高维连续数据处理方面的优势, 近期一些研究成果在更高维系统的交通信号控制方面, 取得了很好的结果[15−16]. 然而, 在公交信号优先控制研究中, 强化学习的应用极少. 文献 [17−18] 中提出了基于 Q 学习的高频公共汽车信号主动优先控制模型, 然而其决策范围仅限单个信号周期且只能应用低维离散的交通信息. 目前将高维 − 连续的实时交通信息作为输入的公交信号优先控制研究仍为空白.

本文提出了一种基于序贯决策框架的有轨电车交叉口信号优先控制模型, 将有轨电车通过交叉口整个过程中的交通信号多步控制问题建模为马尔科夫决策过程. 不设置固定的信号循环并优化某些特定相位配时, 而是在更大的时域内不断动态调整交通信号. 无需对交通环境建模并预测交通状态, 只根据实时交通信息, 通过深度强化学习迭代得到最优信号相位切换策略. 本文的主要贡献如下: 1) 针对存在无法预测实时随机变化的交通场景, 建立了一种新的 “基于数据驱动” 的信号优先控制模型. 在有轨电车接近并通过路口的过程中, 通过采集的实时交通变化信息, 对交通信号进行连续多步决策, 以实现有轨电车信号优先并协同考虑路口社会车辆的通行. 2) 提出一种改进的深度 Q 学习方法框架用于信号优先控制策略求解. 首次将高维 − 连续的实时交通信息作为控制模型的直接输入, 能够从更大范围、更高维度的交通信息中提取特征以改善控制效果.

本文结构如下: 第 1 节设计了基于马尔科夫决策过程的有轨电车信号优先控制模型; 第 2 节基于深度强化学习对有轨电车信号优先控制进行求解; 第 3 节为实验验证与分析; 第 4 节为结论.

1 问题描述

本文旨在寻求一种交叉口信号灯相位多步控制策略以实现有轨电车主动信号优先控制. 通过提供信号优先权, 使有轨电车在通过交叉口时中能够尽可能地避免停车; 并同时提高交叉口社会车辆的通行效率, 减少平均通行延误.

在根据控制目标建立具体模型之前, 对基本的马尔科夫决策过程以及各组成部分描述如下.

1.1 马尔科夫决策过程

一个典型的马尔科夫决策过程以四元组 $\langle S, A, R, P\rangle$ 定义如下:

1) S: 所有状态的集合. s 表示一个特定的状态 $(s \in S)$;

2) A: 所有动作的集合. a 表示一个特定的动作 $(a \in A)$;

3) R: 所有标量奖励的集合. $r_{s,a}$ 表示在状态 s 时采取动作 a 所获得的奖励;

4) P: 所有状态之间的状态转移函数.

设初始状态为 s_0, 控制器从 A 中选择动作 a_0. 执行 a_0 后状态变为 s_1, 并获得奖励 r_0, 以此类推.

从长远来看, 控制器的目标便是通过一系列的动作 $\{a_0, a_1, a_2, \cdots\}$, 获得最大化的累积奖励,

$$G = r_0 + \gamma r_1 + \gamma^2 r_2 + \cdots = \sum_{t=0}^{\infty} \gamma^t r_t \tag{1}$$

其中, $\gamma \in [0,1)$ 为折扣因子. 折扣因子的作用是对未来的不同时间步获得的奖励赋予不同的价值权重.

控制策略 π 可以看成是状态空间到动作空间上的映射. 对于有限步决策, 设 T 步, 用离散时间步 $t, t \in [0,T]$ 表示控制策略 π 下的第 t 步动作, 则在状态 s 时根据 π 选取动作 a 后所获得的累积奖励由式 (2) 中的状态 − 动作值函数 $Q^{\pi}(s,a)$ 定义

$$Q^{\pi}(s,a) = \mathrm{E}\left[r_t + \gamma r_{t+1} + \cdots | s_t = s, a_t = a, \pi\right] = \mathrm{E}\left[\sum_{k=0}^{T-t} \gamma^k r_{t+k} \,|s_t = s, a_t = a, \pi\right] \tag{2}$$

最终的目标便是寻找最优控制策略 $\pi*$. 如果在当前状态下, 能够得知后续状态的最佳 Q 值, 则最优策略便是选择能够实现最高累积奖励的控制动作. 因此, 最佳的 $Q^{\pi}(s,a)$ 是根据后续状态的最佳 Q 值计算得到, 用贝尔曼最优性方程表示为

$$Q^{\pi *}(s,a) = \mathrm{E}_{s'}\left[r_t + \gamma \max_{a'} Q^{\pi^*}(s',a') \,|s,a\right] \tag{3}$$

当系统模型已知时, 即转移函数空间 P 已知, 式 (3) 可以利用动态规划法进行求解, 但需要状态数有限, 以使得整体的计算复杂度可控. 而当环境模型未知时, 可基于蒙特卡罗法以及时间差分法通过迭代逼近来近似求解[11].

1.2 有轨电车信号优先控制模型建立

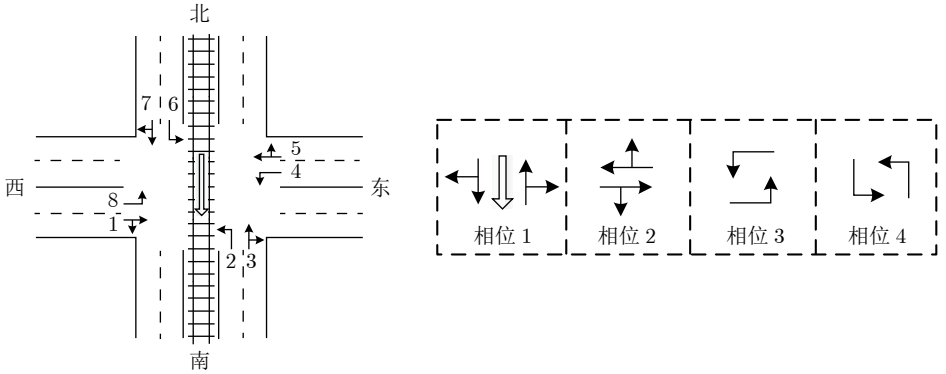
本小节将基于马尔科夫决策过程, 建立有轨电车主动信号优先控制模型.

首先对有轨电车与社会车辆混合通行的轨道/公路交叉道口进行建模. 如图 1(a) 所示, 轨道/公路交叉道口分别连接了 8 条普通车道与 1 条电车轨道. 路口中 8 条车道以 $i = 1, \cdots, 8$ 分别进行编号. 有轨电车的行驶方向为从北向南. 路口中有轨电车与不同方向社会车辆的通行由实际交通中常见的四相位交通信号灯控制, 相位信息如图 1(b) 所示.

本文定义当有轨电车行驶至距交叉口一定距离 L 时, 有轨电车信号优先控制启动, 即有轨电车向交通信号灯发送信号优先请求, 信号灯切换至信号优先控制模式. 而当有轨电车通过后, 信号优先控制模式结束. 下面对有轨电车信号优先控制问题进行拆解, 根据本文控制目标 (即尽可能同时减少有轨电车平均停车次数以及社会车辆的路口平均延误), 分别定义状态向量、动作空间和奖励函数.

1.2.1 状态向量

不同于现有的研究成果, 为了更好地把握实时交通变化, 本文将更大范围更微观的高维交通流特性作为模型输入. 基于有轨电车以及各车道社会车辆的行驶信息来定义交通环境的状态向量 s. 通过车联网, 所有车辆的速度与位置均可以实时收集得到. 每条车道均分为相同长度的分段. 假定交叉口附近社会车辆行驶信息的收集范围为 b, 均分的车道小段长度为 l, 则每个车道的划分的分段数量为 $c = b/l$. 对于每一个车道分段, 其实时状态定义为一个二元组, $\langle n, v\rangle$. 其中, n 为该车道分段中社会车辆的数量, 为非负的离散值. v 为该车道分段中社会车辆的平均行驶速度, 为连续值. 特别地, 当分段没有社会车辆存在时, v 的取值定义为 -1. b 的取值越大, 则表示对社会车辆车流信息采集的范围越



(a) 轨道/公路交叉道口
(a) Rail/road crossing

(b) 信号灯相位示意图
(b) Signal phase sequence

图 1 路口示意图

Fig. 1 Intersection diagram

大, 即在控制中同时考虑更多的社会车辆通行. 与此同时, l 的长度取值越小 (最小不能小于社会车辆的长度), 则对社会车流的信息采集得越微观, 在实际应用中便能对实时社会车流的变化掌握更充分. 同时, 为了在模型中考虑更详细的有轨电车运行信息, 有轨电车的实时行驶状态同样定义为二元组 $\langle d_{Tm}, v_{Tm}\rangle$. 其中 d_{Tm} 与 v_{Tm} 分别表示有轨电车距交叉口停车线的距离及其实时行驶速度, 两者均为连续值, 且速度值保持非负.

最终, 本文模型的状态向量定义为 $\boldsymbol{s}=[d_{Tm}, v_{Tm}, n_{1,1}, v_{1,1}, \cdots, n_{i,j}, v_{i,j}, \cdots, n_{8,c}, v_{8,c}]^{\mathrm{T}}$, $i=1,\cdots,8$, $j=1,\cdots,c$. 其中, i 对应不同的社会车辆车道, j 对应不同的车道分段. 值得注意的是, 状态向量的维度由 b 与 l 的取值决定, b 越大且 l 越大, 则状态向量维度越高, 对应模型越复杂.

1.2.2 动作空间

信号灯需要根据实时交叉口通行状态选择适当的动作来不断引导社会车辆的通行, 减少其路口排队以及在有轨电车接近时给予信号优先. 每步控制的动作为交通信号灯在下一个离散步长中的相位. 本文考虑交通环境中存在实时随机变化的情况. 为了提高多步控制的灵活性以对交通变化进行及时响应, 本模型中各交通信号相位的顺序与时长并不固定, 均根据实时交通变化动态调整, 没有固定的交通信号循环. 信号控制动作的选择有两种情况出现: 保持当前信号相位与切换到其他信号相位. 最终, 动作空间定义为, $A=\{$相位 1, 相位 2, 相位 3, 相位 4$\}$. 在每一个离散时间步 t, 信号灯选择一个特定的相位, 控制对应方向的车辆在 $t\to t+1$ 中通过交叉口.

1.2.3 奖励函数

作为试错 − 改正机制的重要部分, 实时奖励的设计用于指导控制器进行有效地学习以获得最佳的动作选择策略. 作为标量值, 奖励需要设计为环境状态的函数, 以指导学习过程达到最终的控制目标. 根据本文的控制目标, 奖励包括两个部分: 1) 由有轨电车在到达路口时是否停车来进行确定; 2) 由 8 个方向车道上的社会车辆排队状态变化来进行确定. 奖励函数设计为

$$r_t = r_{\mathrm{trm},t} + r_{\mathrm{veh},t} \tag{4}$$

其中, r_{trm} 表示来自有轨电车运行状态的奖励. 假定有轨电车在通过交叉口的过程中仅会因为红灯信号才减速并停车. 有轨电车在行驶过程中没有发生停车, 则无需反馈奖励值. 而当有轨电车在行驶过程中发生停车时, 则会直接反馈一个较大的负值奖励, 即惩罚.

$$r_{\mathrm{trm},t}=\begin{cases}0, & \text{未停车}\\ -20, & \text{停车}\end{cases} \tag{5}$$

式 (4) 中 r_{veh} 表示来自交叉口社会车辆排队状态变化的奖励, 体现上一控制动作对社会车辆实时排队的影响, 包括两个部分:

$$r_{\mathrm{veh},t}=\phi_1\cdot\sum_{i=1}^{8}\kappa_i\cdot(q_{i,t-1}-q_{i,t})+ \phi_2\cdot\begin{cases}2, & q_{i,t}<5,\ \forall i=1,\cdots,8\\ 0, & \text{其他}\end{cases} \tag{6}$$

其中, ϕ_1 与 ϕ_2 分别为两部分的权重参数, $\phi_1,\phi_2\geq 0$ 并且 $\phi_1+\phi_2=1$. $q_{i,t}$ 与 $q_{i,t-1}$ 分别表示车道 i 在当前离散决策时刻 t 与前一个离散决策时刻 $t-1$ 的社会车辆排队长度 (即排队车辆数). κ_i 是离散权重参数, 用于在实时奖励计算中对不同的拥堵等级区别对待. 奖励函数中考虑当前控制动作对交叉口各车道社会车辆排队的影响时, 将会侧重有更长排队长度的车道. 当不同车道的初始排队长度与驶入车流量不相同时, 这种设计能够更均衡地考虑所有的车道, 而不会出现某一车道因为初始排队长度大或驶入车流量大而导致平均延误更大的情况. 对于车道 i, 定义如下:

$$\kappa_i=\begin{cases}0.1, & q_{i,t-1}\in[0,5]\\ 0.3, & q_{i,t-1}\in(5,10]\\ 0.7, & q_{i,t-1}\in(10,15]\\ 1, & q_{i,t-1}\in(15,20]\\ 1.5, & q_{i,t-1}\in(25,30]\end{cases} \tag{7}$$

特别地, 从式 (7) 中可以看出, κ_i 的取值由 $t-1$ 的社会车辆排队长度来进行确定, 即动作执行前而非执行后. 这样的设计能够让奖励函数对拥堵等级的变化更敏感. 式 (6) 的第 1 部分表示: 控制动作实施后, 综合加权考虑交叉口各车道社会车辆排队长度的变化, 总体减少量越大, 则奖励越大; 式 (6) 的第 2 部分表示: 如果采取所选控制动作后, 各车道上的社会车辆排队长度均较小, 将会追加一个额外的奖励. 这样有助于更好地实现减少社会车辆排队延误的控制目标.

从式 (5) 中可以看出, 根据有轨电车运行状态确定的奖励较为稀疏, 在多数离散决策步中均为零. 然而, 如式 (2) 中所示, 每个决策步选择最优动作时并不是只考虑当前决策步所能获得的奖励最大, 而是整个过程所有决策步奖励加权之和的期望最大.

因此, 即使在前期的决策中有轨电车还没有接近交叉口, 控制器的动作仍然可以视为向接近交叉口的有轨电车提供信号优先前的准备工作. 前期决策的目的同样是使得整个过程奖励加权之和的期望最大. 由于可以从过往经验中学习, 在前期动作选择时便知道在某一步后续动作中需要为有轨电车提供信号优先, 因此会提前疏解与有轨电车存在冲突车道方向的社会车辆排队. 从整个通行过程平均来看, 有轨电车接近路口时, 其信号优先权对冲突车道方向的社会车辆存在的负面影响便被抵消了. 本文的这种交通信号多步决策模型设计, 综合考虑了有轨电车通行的全过程, 因此可以有效地实现在保证有轨电车信号优先的同时减少社会车辆的平均通行延误.

在实时随机变化的交通环境中, 系统状态有很强的不确定性与不可预测性, 直接精确建模有轨电车信号优先对交叉口社会车辆行驶所造成的影响非常困难. 因此无法得到马尔科夫决策过程直接求解所需的状态转移函数 P. 同时, 为了更好地对实时交通变化进行响应, 本文通过车联网收集得到有轨电车与大范围社会车辆实时行驶信息. 通过处理高维 − 连续的微观交通信息来实现对交通实时变化更好地把握, 状态空间维度较高且包括了连续状态变量. 因此式 (3) 的求解将很难借助于传统方法. 本文采用深度强化学习方法, 利用交叉口信号控制的历史经验数据, 通过不断交互学习求解马尔科夫决策问题并得到最优控制策略.

2 有轨电车信号优先控制问题求解

本文将有轨电车信号优先控制问题定义为无模型马尔科夫决策问题. 由于需要采用高维 − 连续的交通状态信息作为输入, 状态空间的维度极大, 经典强化学习算法无法应用, 因此采用深度强化学习算法 (即深度 Q 网络) 来进行问题求解. 此外, 本文结合几种最新技术对深度 Q 网络算法进行改进, 以得到更快的收敛速度和更好的稳态控制效果.

2.1 深度 Q 网络

由于无法得到精确的交通模型, 因此式 (3) 求解中的最优 Q 值无法通过直接计算获得, 而是需要通过不断学习得到其估计值. 本文模型中, b 取 300 米, l 取 20 米, 状态向量的维度达到 242×1, 且一半的状态变量连续, 无法通过经典 Q 学习[19] 中表的方式来表示 Q 值. 因此采用深度神经网络 (其网络参数向量定义为 $\boldsymbol{\theta}$) 作为非线性函数逼近器来近似表示 Q 值, 即将深度神经网络与经典 Q 学习结合起来, 形成深度 Q 网络 (Deep Q network, DQN)[20].

DQN 通过样本池和目标 Q 网络来提高训练效率并减少波动. 样本池用于保存交互中产生的样本, 使得控制器可以从过往行动经验中进行学习. 定义估计状态 − 动作值函数的神经网络为主 Q 网络 (参数为 $\boldsymbol{\theta}$). 在训练主 Q 网络时, 随机从样本池中取出一定数量 (m 组) 的样本进行梯度下降计算, 以消除样本数据间的关联性, 提高学习效率. 设置目标 Q 网络, 每一次迭代学习的目标值由目标 Q 网络 (参数为 $\boldsymbol{\theta}^-$) 计算得到, 以此降低训练中因为策略未收敛而产生的波动. 在第 i 步, 迭代学习的目标值表示为

$$y_i = r + \gamma \max_{a'} Q\left(s', a'; \boldsymbol{\theta}^-\right) \tag{8}$$

状态动作 − 值函数迭代逼近形式为

$$\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}_{i-1} + \alpha\left[y_i - Q\left(s, a; \boldsymbol{\theta}_i\right)\right] \nabla_{\theta_i} Q\left(s, a; \boldsymbol{\theta}_i\right) \tag{9}$$

其中, $\alpha \in [0,1)$ 为学习因子. y_i 为第 i 步迭代的目标值. 主 Q 网络中的参数每隔 n 步便更新至目标 Q 网络 ($\boldsymbol{\theta} \to \boldsymbol{\theta}^-$).

最终得到状态 − 动作值函数的最优估计参数 $\boldsymbol{\theta}^*$ 以及交通信号最优控制策略为

$$a_t^* = \pi^*(s_t) = \arg\max_{a\in A} Q\left(s_t, a; \boldsymbol{\theta}^*\right) \tag{10}$$

其中, s_t 为 t 时刻采集得到的真实交通状态. a_t^* 为 t 时刻计算得出的最优信号相位.

2.2 基于竞争架构与加权样本池的改进深度双 Q 网络

由于本文有轨电车信号优先控制问题的复杂性, 深度 Q 网络在训练中仍然存在较大的波动且需要较长的训练时间. 为了得到更好的控制性能, 本文应用竞争网络架构、双 Q 网络以及加权样本池对深度 Q 网络算法进行改进, 以进一步提高其训练速度并改善其稳态控制效果.

通过竞争网络架构 (Dueling network)[21], 可以将状态 − 动作值函数以更详细的方式进行定义, 以更好地区分不同动作对未来预期累积奖励的潜在价值. 新的动作 − 值函数定义为

$$Q^\pi(s, a; \boldsymbol{\theta}) = V^\pi(s; \boldsymbol{\theta}) + A^\pi(s, a; \boldsymbol{\theta}) \tag{11}$$

其中, $V^\pi(s;\boldsymbol{\theta})$ 为仅基于状态的值函数, 用来衡量处于某个特定状态是否有利于获得未来预期累积奖励. $A^\pi(s,a;\boldsymbol{\theta})$ 为基于状态 − 动作对的优势函数, 用来表示某一特定动作对某个特定状态的重要性. 根据 Q 学习中对于预期奖励的定义可知

$$\mathrm{E}\left[Q(s,a)\right] = V(s) \tag{12}$$

这表明对于特定状态 s, 如果某动作 a 对应的

$Q(s,a)$值高于状态单独对应的$V(s)$值, 则说明这个动作有较高的价值. 因此在所有动作中, 如果一个动作的优势函数为正数, 则这个动作对于获得更多的累积奖励有促进作用, 超过了在这个状态下所有可能动作的平均表现. 相反, 如果动作的优势函数为负数, 则表示行动的预期累积奖励低于平均水平. 同时, 从式 (12) 中可知, $A(s,a;\boldsymbol{\theta})$的期望值必须为零. 因此在实际应用中, 为了实现网络的输出稳定并得到更好的稳态性能, 定义

$$Q(s,a;\boldsymbol{\theta})=V(s;\boldsymbol{\theta})+A(s,a;\boldsymbol{\theta})-\frac{\sum_{a'}A\left(s,a';\boldsymbol{\theta}\right)}{|A|} \tag{13}$$

DQN 中通过设置目标 Q 网络来解决策略未收敛所带来的训练波动. 将式 (8) 展开后可得

$$y_i=r+\gamma\max_{a'}Q\left(s',\max_{a'}Q\left(s',a';\boldsymbol{\theta}^-\right);\boldsymbol{\theta}^-\right) \tag{14}$$

从式 (14) 中可知, DQN 算法中采用目标 Q 网络计算每一步迭代的目标值时, 最优动作选择与目标值计算仍然使用了同样的目标 Q 网络, 这将由于相关性导致对状态 − 动作值函数产生过高估计. 因此, 本文将目标值计算的两部分进行分离, 采用双 Q 网络 (Double Q network)[22] 结构, 利用主 Q 网络完成最优动作的选择以避免过估计

$$y_i=r+\gamma\max_{a'}Q^{\pi}(s',\max_{a'}Q^{\pi}(s',a';\boldsymbol{\theta});\boldsymbol{\theta}^-) \tag{15}$$

在 DQN 的训练中, 每次随机从样本池中取出一定数量的样本用于梯度下降计算, 对样本中不同动作产生的效果差异学习效率较低. 因此, 引入加权样本池 (Prioritized experience replay)[23] 对不同样本设置不同的采样权值 (即被选择的概率). 样本中某动作表现得越差, 即相对于估计值有更大的偏差, 则其被选择的概率就越大, 以此来提高学习效率. 本文采用分阶的方法来计算样本的采样权值. 样本i的估计误差为

$$\delta_i=\left|Q(s,a;\boldsymbol{\theta})_i-Q(s,a;\boldsymbol{\theta}^-)_i\right| \tag{16}$$

样本的采样权值设为样本估计误差的倒数, 则可以得到样本i的被采样概率为

$$P_i=\frac{\dfrac{1}{\delta_i^{\tau}}}{\sum\limits_{k\in M}\dfrac{1}{\delta_k^{\tau}}} \tag{17}$$

其中, τ 表示采样权值的作用大小. 特别地, 当$\tau=0$时, 加权采样就变成普通随机采样.

此外, 为了在模型迭代学习的初期, 能够尽可能产生更多的有效样本, 模型中采用了可变ε贪婪算法在不同训练阶段平衡动作选择的探索和利用. 本文设置ε在前 200 次动作选择中取值为 1. 随着训练次数的逐步增加, ε值将按照特定速率 ($\Delta\varepsilon$/次) 逐渐减小. 当ε值到达 0.001 时, ε值将不再减小并在后续的训练过程中保持不变.

最终, 得到改进后的有轨电车信号优先多步控制策略求解算法: 基于竞争架构与加权样本池的深度双 Q 网络 (Dueling double deep Q network with prioritized experience replay, 3DQN), 详见算法 1. 同时, 最终的深度神经网络结构见图 2.

算法 1. 基于改进深度 Q 网络有轨电车信号优先多步决策求解算法

输入. 奖励折扣因子γ, 学习因子α, 目标 Q 网络更新周期n, 探索概率ε.

初始化容量为N的样本池D;

初始化容量为m的采样暂存区C;

初始化状态 − 行动模型网络 Q 和对应参数$\boldsymbol{\theta}$;

初始化目标 Q 网络$\hat{Q}$和对应参数$\boldsymbol{\theta}^-$;

Begin

1: **For** episode =1, M **do**

2:　初始化交叉口环境, 得到初始交通环境状态s_1

3:　**For** t=1, T **do**

4:　　以ε的概率随机选择一个信号控制动作a_t, 或根据模型选择当前最优信号控制动作$a_t=\max_a Q(s_t,a;\boldsymbol{\theta})$

5:　　执行信号控制动作a_t, 得到新一轮的交通环境状态s_{t+1}和奖励r_{t+1}

6:　　如果样本池满, 则从其中删除最早的样本记录

7:　　将四元组$\{s_t,a_t,r_{t+1},s_{t+1}\}$作为一个样本储存到$D$中

8:　　更新样本的采样权值与被采样概率

9:　　从D中采样m组样本放入C中

10:　更新交通环境系统状态$s\leftarrow s'$

11:　根据 C中数据计算目标值:

　　If s_{j+1}为最终状态

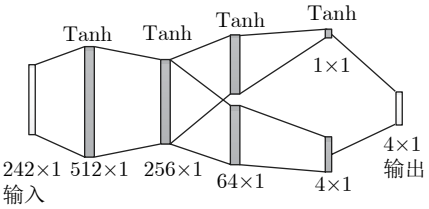


图 2　深度神经网络结构图

Fig. 2　The structure of DNN

$y_j = r_{j+1}$

else

$y_j = r_{j+1} + \gamma \max_{a'} Q(s', \max_{a'} Q(s', a'; \boldsymbol{\theta}); \boldsymbol{\theta}^-)$

12: 根据目标函数: $J(\boldsymbol{\theta}) = E_\pi\left[\left(y_j - Q(s_j, a_j; \boldsymbol{\theta})\right)^2\right]$进行梯度下降法求解

13: 每隔 n 时间步, 更新目标 Q 网络参数 $\boldsymbol{\theta} \to \boldsymbol{\theta}^-$

14: 更新 ε

15: **End for**

16: **End fo**

End.

3 实验与验证

3.1 仿真训练与参数设置

基于微观交通仿真软件 SUMO[24] 搭建交叉口仿真实验环境. 交叉口面积为 3 000 m×3 000 m. 假定有轨电车距离路口 1 500 m 处时, 发起信号优先请求, 此时定义为多步决策过程起始时刻. 当有轨电车通过交叉口后, 多步决策过程结束. 一次完整通行定义为一次训练. 每 20 次训练定义为一个训练集. 本文中, 深度强化学习模型共经过 450 个训练集, 模型参数如表 1 所示.

本文考虑随机变化的交通环境, 在每次训练开始随机设置系统初始状态. 初始信号相位以及相位剩余时间均为随机选择; 各车道社会车辆初始排队长度为离散整数值, 在 0～30 辆中随机选择; 有轨电车初始速度为连续值, 在 50～70 km/h中随机选择. 在有轨电车接近路口过程中, 其速度存在随机波动, 范围为 50～70 km/h, 且有 20 % 的概率出现突然加减速情况; 不同车道中社会车辆平均到达率均不同, 且存在随机波动, 分别对应为: 300±20 辆/h, 310±20 辆/h, 300±30 辆/h, 310±30 辆/h, 290±20 辆/h, 310±20 辆/h, 290± 30 辆/h, 300±30 辆/h. 社会车辆行驶速度同样存在波动, 范围为 50～60 km/h, 且有 30 % 的概率出现突然加减速情况.

表 1 模型参数

Table 1 Model parameters

参数	取值
N	20 000
m	32
$\Delta\varepsilon$	−0.001
γ	0.99
α	0.001

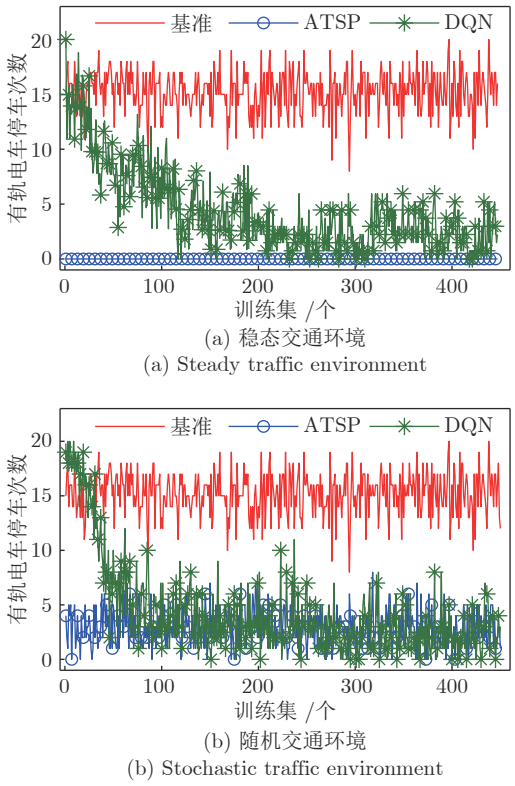
同时, 为了进行对比分析, 我们定义稳态交通环境, 即各车辆速度与到达率保持不变.

3.2 实验结果与讨论

实验中通过 3 种性能指标来评估实验结果: 有轨电车平均停车次数、平均累积奖励以及社会车辆平均等待时间. 为了评估深度 Q 网络与改进的深度 Q 网络模型性能, 在实验中分别与以下两种策略进行对比: 第 1 种是固定时长策略: 各相位持续时间固定为 40 s, 实验中作为基准策略. 第 2 种是传统的有轨电车信号主动优先控制策略 (Active transit signal priority, ATSP)[25]. 其中需要基于有轨电车运动模型来预测有轨电车交叉口到达时间. 与此同时, 对于本文应用的两种深度强化学习模型, 同样从收敛速度与稳态性能两方面进行对比研究.

3.2.1 深度 Q 网络模型实验

首先, 评估深度 Q 网络、固定相位时长策略以及文献 [25] 中 ATSP 策略三者之间分别在稳态交通环境和随机可变交通环境中的性能差异. 图 3 和图 4 分别对应表示稳态交通环境与随机变化交通环境下, 三种控制策略下的有轨电车平均停车次数与平均累积奖励随训练时长 (即训练集数增加) 的变化



(a) 稳态交通环境

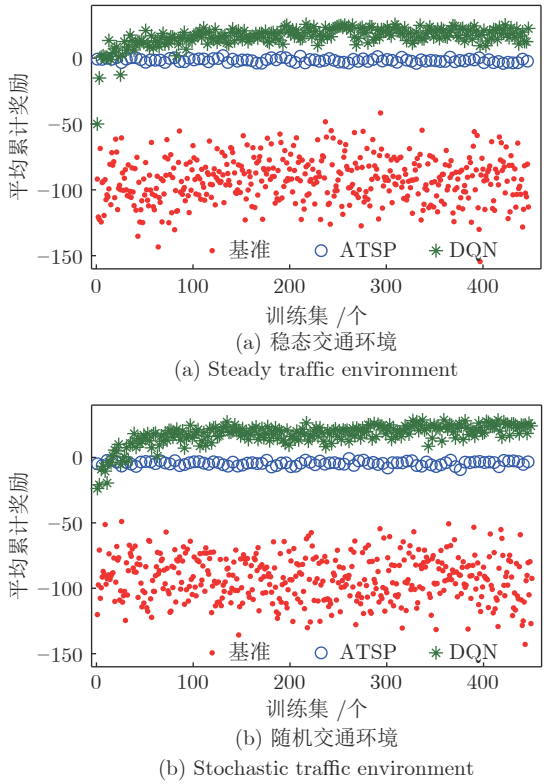
(a) Steady traffic environment

(b) 随机交通环境

(b) Stochastic traffic environment

图 3 有轨电车平均停车次数对比

Fig. 3 Comparison of tram mean stops



(a) 稳态交通环境

(a) Steady traffic environment

(b) 随机交通环境

(b) Stochastic traffic environment

图 4　平均累积奖励对比

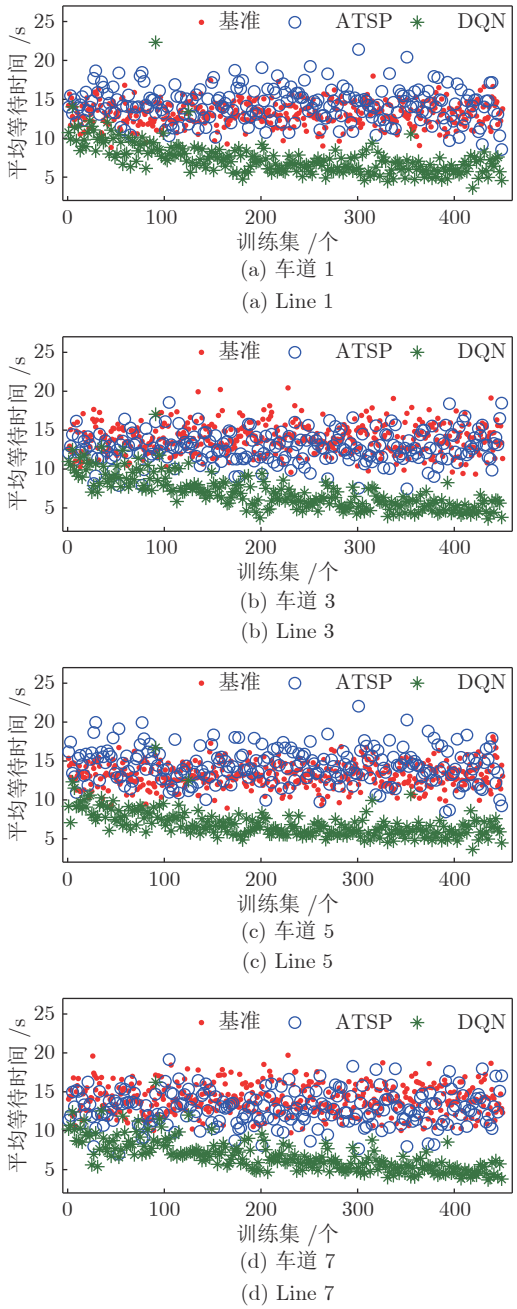
Fig. 4　Comparison of cumulative reward

曲线. 从图 3(a) 中可以看出, 当交通环境较为稳定时, 因为有轨电车的交叉口到达时间能够准确地预测, 因此 ATSP 策略有最好的效果, 能够完全使得有轨电车避免在交叉口停车. 相比之下, 深度 Q 网络仅能平均减少约 84 % 的有轨电车交叉口停车. 然而, ATSP 策略仅考虑有轨电车的信号优先, 无法降低路口社会车辆的通行延误. 由于本文中奖励的计算协同考虑了有轨电车与社会车辆的通行, 因此 ATSP 的累积奖励要低于深度 Q 网络, 如图 4(a) 所示.

对于随机变化交通环境, 从图 3(b) 中可以得知, 由于到达时间预测存在误差, ATSP 策略无法保证有轨电车完全避免停车, 仅能平均减少约 81 % 的有轨电车交叉口停车. 与之相比, 深度 Q 网络能够平均减少约 82 % 的有轨电车交叉口停车, 但是存在不稳定的问题, 即使在训练后期模型处于收敛状态, 有轨电车平均停车次数仍然有较大的波动.

深度 Q 网络与 ATSP 策略均有效地减少了有轨电车在通过交叉口时的停车次数. 然而, 从图 4 中可以得知, 深度 Q 网络通过迭代学习能够将平均奖励值最大化, 这意味着深度 Q 网络的控制策略最符合本文控制目标的要求, 即在尽可能保证有轨电车不停车通过路口的前提下, 提高路口社会车辆的通行效率和减少排队长度. 固定时长策略和 ATSP 策略由于没有从控制结果中进行反馈学习的能力, 因此无法通过模型的迭代次数来获得控制性能上的优化与提升, 有轨电车平均停车次数均保持稳定. 而深度 Q 网络则可以通过训练时长的累积, 逐步改善控制效果并达到稳定状态.

图 5 为随机变化交通环境下三种策略对交叉口



(a) 车道 1

(a) Line 1

(b) 车道 3

(b) Line 3

(c) 车道 5

(c) Line 5

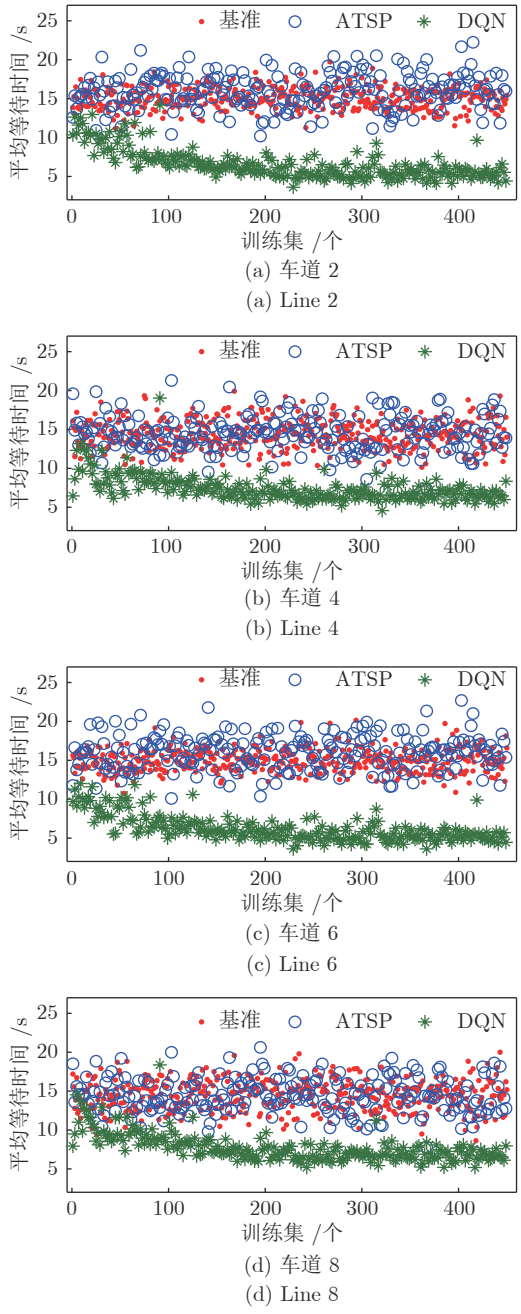
(d) 车道 7

(d) Line 7

图 5　各直行/右转车道平均停车等待时间对比

Fig. 5　Comparison of waiting time in direct/right turn lanes

直行/右转车道社会车辆的影响对比. 图 6 为随机变化交通环境下三种策略对交叉口左转车道社会车辆的影响对比. 从图中可以明显地看到 ATSP 策略对交叉口社会车辆的负面影响. 由于车道 3 和车道 7 中社会车辆的行驶与有轨电车的行驶并不发生冲突, 即由同一个信号相位控制, 因此有轨电车信号优先权对车道 3 和车道 7 中社会车辆的影响并不是负面的. 在图 5(b) 和图 5(d) 中可以明确看出, ATSP 策略 (圆圈) 下社会车辆的平均延误时间略低于基准 (实点). 而 1, 5, 2, 6 车道方向的平均延误时间是高于基准的, 分别见图 5 与图 6 中的 (a) 与 (c). 这充分说明了信号优先权只对冲突车道方向有负面影响. 因为在 ATSP 中, 为了给予有轨电车信号优先权 (信号相位 1), 会改变信号循环中相邻的两个相位长度 (相位 2 和 4). 而相位 3 对应的车道 4 和车道 8 的平均延误基本不变, 见图 6(b) 和图 6(d). 同时, 深度 Q 网络能够有效降低交叉口社会车辆的等待时间, 在各社会车辆行驶车道, 深度 Q 网络相较于基准策略均有着明显的下降, 减少量均超过 50 %.



(a) 车道 2
(a) Line 2

(b) 车道 4
(b) Line 4

(c) 车道 6
(c) Line 6

(d) 车道 8
(d) Line 8

图 6 各左转车道平均停车等待时间对比

Fig. 6 Comparison of waiting time in left turn lanes

从上述实验结果可以看出, 在存在实时随机变化的交通环境中, 虽然深度 Q 网络与 ATSP 策略在减少有轨电车交叉口平均停车次数方面的效果相近, 但是深度 Q 网络能够同时减少交叉口社会车辆平均等待时间, 这充分说明了本文基于深度 Q 网络的有轨电车信号优先控制模型的有效性. 然而, 普通的深度 Q 网络在实际应用中仍然存在一些问题, 如模型在经过训练收敛后有轨电车的平均停车次数依然存在波动较大的情况.

3.2.2 基于竞争架构的双深度 Q 网络模型实验

本小节中, 我们评估改进的深度 Q 网络模型 (3DQN) 与普通深度 Q 网络模型 (DQN) 以及 ATSP 策略在随机变化交通环境下的性能差异. 图 7 为两种深度强化学习模型下有轨电车平均停车次数对比图. 从图中可以看出, 改进的深度 Q 网络模型比普通的深度 Q 网络模型有更好的收敛速度, 能够更快地到达稳定状态, 并且在收敛过程中模型波动较小. 同时, 在训练后期, 改进的深度 Q 网络模型进入稳定状态后, 有轨电车的平均停车次数均值比普通深度 Q 网络减少了约 1 % 左右, 有更好的稳态效果. 且进入稳态后没有较大的波动, 深度 Q 网络模型下平均停车次数的波动相较于普通深度 Q 网络减少近 30 %. 稳定后的改进深度 Q 网络比 ATSP 策略同样有更好的性能, 有轨电车的平均停车次数均值减少了约 2 % 左右, 且波动更小.

与此同时, 从图 8 可以看出, 改进的深度 Q 网络模型的累积奖励要比普通深度 Q 网络略大, 并远远大于 ATSP 策略. 因此, 改进的深度 Q 网络模型能够更好地实现本文的控制目标, 即兼顾有轨电车与社会车辆的通行需求.

三种策略对于社会车辆的影响可以从图 9 和图 10 中看出. 两种深度强化学习模型相对于 ATSP 策略均可以降低交叉口社会车辆的等待时间. 然而, 改进的深度 Q 网络比普通的深度 Q 网络有

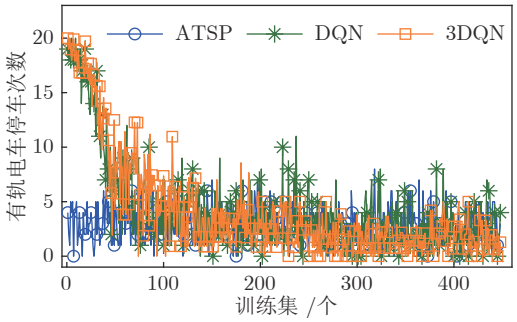


图 7　两种深度强化学习模型下有轨电车平均停车次数对比

Fig. 7　Comparison of tram mean stops under two deep reinforcement learning models

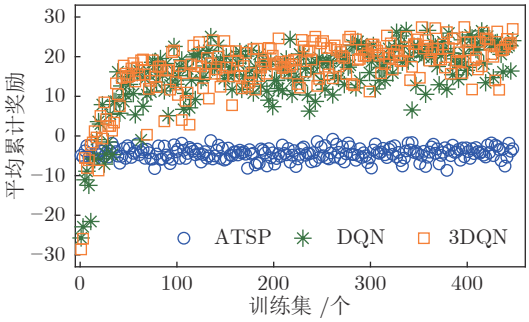


图 8　两种深度强化学习模型下累积奖励对比

Fig. 8　Comparison of cumulative reward under two deep reinforcement learning models

更好的稳态效果, 且在减少社会车辆路口延误方面远优于 ATSP 策略. 在各社会车辆行驶车道, 普通深度 Q 网络下的社会车辆平均等待时间相较于 ATSP 策略降低约为 53 %, 而改进的深度 Q 网络相较于 ATSP 策略降低约为 66 %.

综上所述, 改进的深度 Q 网络, 即基于竞争架构与加权样本池的深度双 Q 网络, 要比普通深度 Q 网络在处理随机交通变化时有更好的性能.

4　结论

本文提出了一种新的有轨电车信号主动优先控制模型来解决轨道/公路平行交叉道口交通信号多步控制问题. 通过车联网收集的有轨电车与社会车辆路口行驶信息作为模型的输入. 基于深度强化学习, 经过模型训练得到有轨电车整个通行过程中最优的交通信号连续控制策略. 通过与固定时长策略以及文献 [25] 中有轨电车信号主动优先策略进行比较, 本文模型在有轨电车与社会车辆协同控制方面取得了较好的结果. 在随机变化的交通环境下,

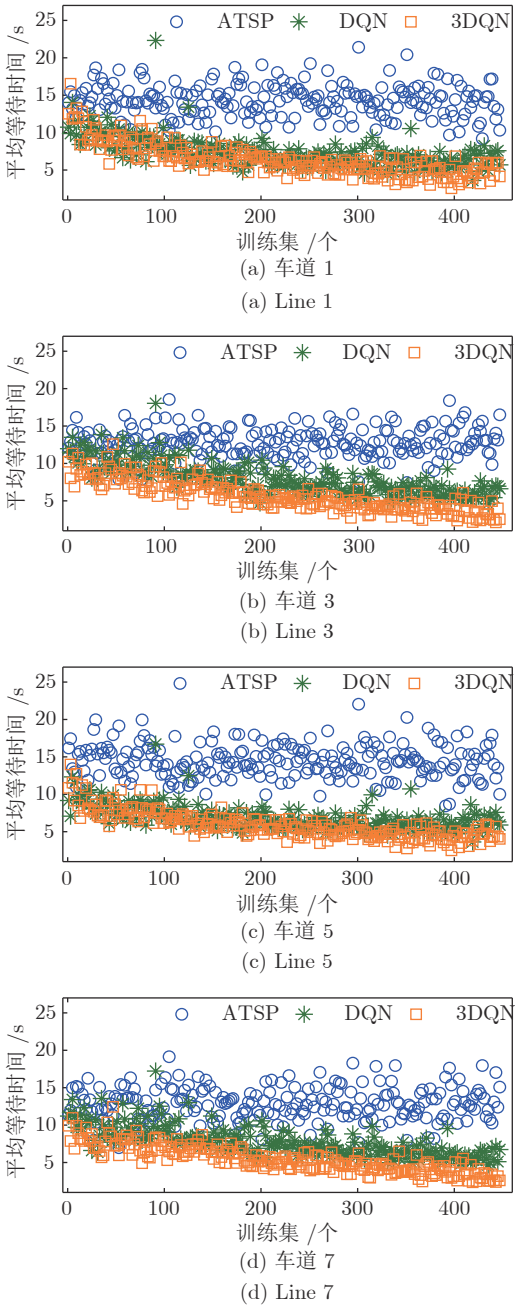
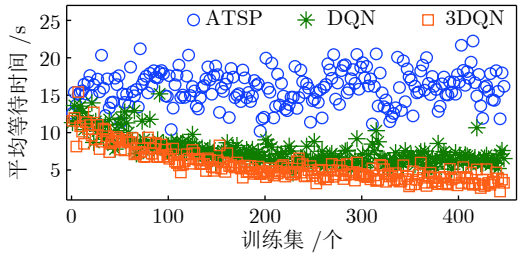


图 9　两种深度强化学习模型下各直行/右转车道平均停车等待时间对比

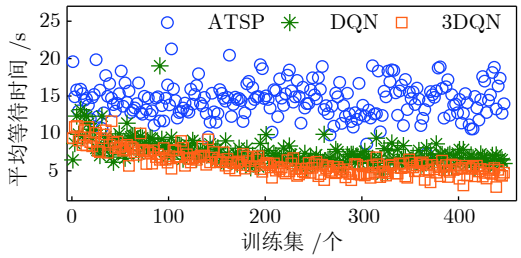
Fig. 9　Comparison of waiting time in direct/right turn lanes under two deep reinforcement learning models

能够保证 82 % 的有轨电车不停车通过交叉口并减少各车道社会车辆平均等待时间 50 % 以上. 未来的工作将专注于随机变化的连续多交叉口交通场景之中. 同时, 结合 “基于模型” 与 “数据驱动” 两种方法各自的优势对整个行程有轨电车信号优先控制问题进行建模与求解, 将是后续研究的重点.



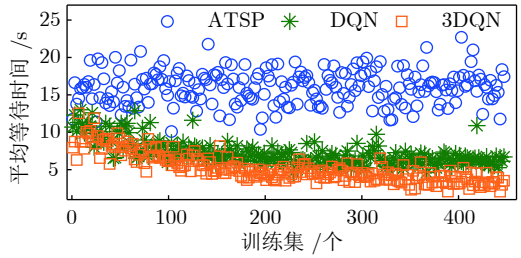
(a) 车道 2

(a) Line 2



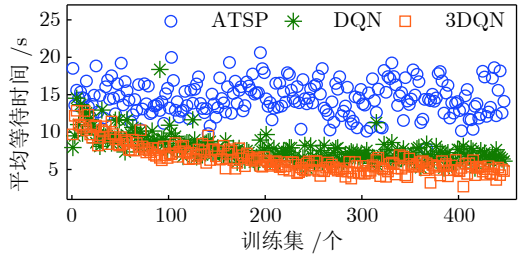
(b) 车道 4

(b) Line 4



(c) 车道 6

(c) Line 6



(d) 车道 8

(d) Line 8

图 10 两种深度强化学习模型下各左转车道平均停车等待时间对比

Fig. 10 Comparison of waiting time in left turn lanes under two deep reinforcement learning models

References

1 Ministry of tranport of China. Statistical bulletin on transportation industry development in 2018. [Online], available: http://xxgk.mot.gov.cn/jigou/zhghs/201904/t20190412_3186720.html, September 5, 2019

2 Shi J G, Sun Y S, Schonfeld P, Qi J. Joint optimization of tram timetables and signal timing adjustments at intersections. *Transportation Research Part C: Emerging Technologies*, 2017, **83**(6): 104–119

3 Ji Y X, Tang Y, Du Y C, Zhang X. Coordinated optimization of tram trajectories with arterial signal timing resynchronization. *Transportation Research Part C: Emerging Technologies*, 2019, **99**(4): 53–66

4 Little J D C, Kelson M D, Gartner N M. Maxband: a program for setting signals on arteries and triangular networks. In: Proceedings of the 60th Annual Meeting of the Transportation Research Board. Washington, USA: Transportation Research Board, 1981. 40–46

5 Jeong Y J, Kim Y C. Tram passive signal priority strategy based on the maxband model. *KSCE Journal of Civil Engineering*, 2014, **18**(5): 1518–1527

6 Ma W, Zou L, An K, Gartner N H, Wang M. A partition-enabled multi-mode band approach to arterial traffic signal optimization. *IEEE Transactions on Intelligent Transportation Systems*, 2019, **20**(1): 313–322

7 Kim H, Cheng Y, Chang G. Variable signal progression bands for transit vehicles under dwell time uncertainty and traffic queues. *IEEE Transactions on Intelligent Transportation Systems*, 2019, **20**(1): 109–122

8 Ji Y X, Tang Y, Wang W, Du Y C. Tram-oriented traffic signal timing resynchronization. *Journal of Advanced Transportation*, 2018, **2018**(1): 1–13

9 Jacobson J, Sheffi Y. Analytical model of traffic delays under bus signal preemption: theory and application. *Transportation Research Part B: Methodological*, 1981, **15**(2): 127–138

10 Yang M, Ding J, Wang W, Ma Y Y. A coordinated signal priority strategy for modern trams on arterial streets by predicting the tram dwell time. *KSCE Journal of Civil Engineering*, 2018, **22**(2): 823–836

11 Gao Yang, Chen Shi-Fu, Lu Xin. Reseacrh on reinforcement learning technology: a review. *Acta Automatica Sinica*, 2004, **30**(1): 1–15
(高阳, 陈世福, 陆鑫. 强化学习研究综述. 自动化学报, 2004, **30**(1): 1–15)

12 Bertsekas D P. Feature-based aggregation and deep reinforcement learning: a survey and some new implementations. *IEEE/CAA Journal of Automatica Sinica*, 2019, **6**(1): 1–31

13 Samah E T, Abdulhai B, Abdelgawad H. Design of reinforcement learning parameters for seamless application of adaptive traffic signal control. *Journal of Intelligent Transportation Systems*, 2014, **18**(3): 227–245

14 Duan Yan-Jie, Lv Yi-Sheng, Zhang Jie, Zhao Xue-Liang, Wang Fei-Yue. Deep learning for control: the state of the art and prospects. *Acta Automatica Sinica*, 2016, **42**(5): 643–654
(段艳杰, 吕宜生, 张杰, 赵学亮, 王飞跃. 深度学习在控制领域的研究现状与展望. 自动化学报, 2016, **42**(5): 643–654)

15 Li L, Lv Y, Wang F-Y. Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica*, 2016, **3**(3): 247−254

16 Liang X, Du X, Wang G, Han Z. A deep reinforcement learning network for traffic light cycle control. *IEEE Transactions on Vehicular Technology*, 2019, **68**(2): 1243−1253

17 Ling K, Shalaby A. Automated transit headway control via adaptive signal priority. *Journal of Advanced Transportation*, 2004, **38**(4): 45−67

18 Shu Bo, Li Da-Ming, Zhao Xin-Liang. Transit signal priority strategy based on reinforcement learning algorithm. *Journal of Northeastern University (Natural Science)*, 2012, **33**(10): 1513−1516
(舒波, 李大铭, 赵新良. 基于强化学习算法的公交信号优先策略. 东北大学学报 (自然科学版), 2012, **33**(10): 1513−1516)

19 Liang Xing-Xing, Feng Yang-He, Ma Yang, Cheng Guang-Quan, Huang Jin-Cai, Wang Qi, et al. Deep multi-agent reinforcement learning: a survey. *Acta Automatica Sinica*, 2019. DOI: 10.16383/j.aas.c180372
(梁星星, 冯旸赫, 马扬, 程光权, 黄金才, 王琦等. 多 agent 深度强化学习综述. 自动化学报, 2019. DOI: 10.16383/j.aas.c180372)

20 Zhao Ying-Nan, Liu Peng, Zhao Wei, Tang Xiang-Long. Twice sampling method in deep Q-network. *Acta Automatica Sinica*, 2019, **45**(10): 1870−1882
(赵英男, 刘鹏, 赵巍, 唐降龙. 深度 q 学习的二次主动采样方法. 自动化学报, 2019, **45**(10): 1870−1882)

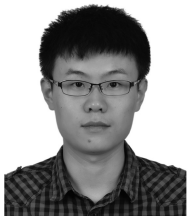
21 Wang Z Y, Schaul T, Hessel M, Hasselt H, Lanctot M, Freitas N. Dueling network architectures for deep reinforcement learning. In: Proceedings of the 33rd International Conference on Machine Learning. New York, USA: PMLR, 2016. 1995−2003

22 Hasselt H V, Guez A, Silver D. Deep reinforcement learning with double Q-learning. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence, Phoenix, USA: MIT, 2015. 2094−2100

23 Schaul T, Quan J, Antonoglou I, Silver D. Prioritized experience replay. In: Proceedings of the 2016 International Conference on Learning Representations 2016, San Juan, Puerto Rico: arXiv, 2016. 1−21

24 Lopez P A, Behrisch M, Walz L B, Erdmann J, Flotterod Y, Hilbrich R, et al. Microscopic traffic simulation using sumo. In: Proceedings of the 21st IEEE International Conference on Intelligent Transportation Systems. Hawaii, USA: IEEE, 2018. 2575−2582

25 Islam M T, Tiwana J, Bhowmick A, Qiu T Z. Design of LRT signal priority to improve arterial traffic mobility. *Journal of Transportation Engineering*, 2016, **142**(9): 04016034



王云鹏 大连理工大学控制理论与控制工程专业博士研究生. 主要研究方向为智能车路协同系统. E-mail: yunpengwang0306@163.com
(**WANG Yun-Peng** Ph.D. candidate in control theory and control engineering from Dalian University of Technology. His main research interest is intelligent vehicle infrastructure cooperative systems.)

郭 戈 东北大学教授. 1998 年获得东北大学控制理论与控制工程专业博士学位. 主要研究方向为智能交通系统, 运动目标检测跟踪网络. 本文通信作者. E-mail: geguo@yeah.net
(**GUO Ge** Ph.D., professor at Northeastern University. He received his Ph.D. degree from Northeastern University in 1998. His research interest covers intelligent transportation system, and moving target detection and tracking with network. Corresponding author of this paper.)