干线动态协调控制的深度Q网络方法*

郭瑝清[1],陈　锋[1,2]

(1.中国科学技术大学 信息科学技术学院,安徽 合肥 230027;
2.安徽中科龙安科技股份有限公司,安徽 合肥 230088)

摘　要: 为有效降低城市交通干线的车均延误与停车次数,将深度Q网络引入干线协调控制,给出了一种干线动态协调控制的DDDQN(Dueling Double Deep Q Network)方法。该方法结合双重深度Q网络与基于竞争架构深度Q网络,并将干线作为整体处理,通过深度神经网络挖掘干线各交叉口协调控制的相关性,基于Q学习进行交通信号控制决策。通过仿真实验,在近饱和流量和干线存在初始排队的情况下,将DDDQN方法与现有绿波方法,以及经典深度Q网络、双重深度Q网络、基于竞争架构深度Q网络的干线协调控制算法进行对比,实验结果表明基于DDDQN的干线动态协调控制算法性能优于其他四种方法。

关键词: 城市交通;干线协调控制;深度Q网络;双重深度Q网络;基于竞争架构深度Q网络

中图分类号: TP181　　**文献标识码:** A　　**DOI:** 10.19358/j.issn.2096-5133.2020.06.001

引用格式: 郭瑝清,陈锋. 干线动态协调控制的深度Q网络方法[J].信息技术与网络安全,2020,39(6):1-6.

A deep Q network method for dynamic arterial coordinated control

Guo Huangqing[1], Chen Feng[1,2]

(1.School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China;
2.Anhui LoongSec Science and Technology Ltd., Hefei 230088, China)

Abstract: In order to effectively reduce the average delay and number of stops for urban traffic trunk roads, a deep Q network is introduced to arterial coordinated control, and a DDDQN (Dueling Double Deep Q Network) method is presented in this paper. This method combines the double deep Q network and the dueling deep Q network, and views the trunk road as a whole. The deep neural network is applied to find the correlation of the coordinated control for each intersection in the trunk road, and Q learning makes these decisions for traffic signal control. Through simulation platform, in the condition of near saturation and initial queue, the method proposed in this paper is compared with the existing green wave method, the arterial coordinated methods respectively based on deep Q network, double deep Q network, and dueling deep Q network. The experimental results show that the performance of DDDQN algorithm is better than the other four methods.

Key words: urban traffic; arterial coordinated control; deep Q network; double deep Q network; dueling deep Q network

0 引言

随着城市人口的增多与经济的快速发展,我国汽车保有量不断增长,城市交通拥堵问题日益严峻。而城市交通干线是城市交通的动脉,实现干线各交叉口间交通信号的动态协调,有效地疏导干线车辆,对于缓解城市交通拥堵具有重要意义。

目前,城市主干道多交叉口的协调控制,主要采用Maxband和Multiband法以及图解法、数解法等绿波方法。LITTLE J D C等人[1]最早提出最大绿波带宽Maxband模型;GARTNER N H等人[2]在Maxband模型的基础上,提出复合绿波带宽Multiband模型;陈昕等人[3]对图解法进行了优化,基于绿波带的中心线交点,设计了一种新的双向绿波图解法;卢凯等人[4]在绿灯中心点型双向绿波协调设计数解法的基础上,建立了一种绿灯终点型的双向绿波数解法,从而减少了干线车队的延误时间;曲大义等人[5]在绿波协调中考虑了公交车辆的影响,并通过增加绿

* 基金项目:安徽省对外科技合作项目(1804b06020376)

信比与对公交车辆适当的提速,进一步提升了交叉口的通行效率。

现有的绿波方法难以准确地描述复杂的城市干线交通流状态,且采用静态的控制模式,无法有效地协调时变的干线交通流。随着人工智能的不断发展,采用深度强化学习实现城市交通信号优化控制已成为研究的热点。HA-LI P 等人[6]为提高交叉口通行能力,提出了一种基于深度强化学习算法的单交叉口信号优化控制方法;GAO J 等人[7]提出一种深度强化学习算法,从实时的交通流数据中自动提取有用特征,实现单交叉口交通流的自适应控制,并采用经验回放和目标网络技术[8],提高了算法的稳定性;LI C C 等人[9]为提高城市路网通行能力,提出了一种用于区域交叉口交通信号控制的深度强化学习算法,通过多智能体学习最佳的交通信号控制策略;VAN DER POL E[10]采用 Max-plus 算法和基于深度强化学习的多智能体方法,实现城市交通区域协调控制。

在深度强化学习领域,目前对于城市交通信号控制的研究,多以单交叉口为研究对象,而对于多交叉口的协调处理,普遍采用多智能体的协调控制。本文结合了双重深度 Q 网络(Double Deep Q Network, Double DQN)[11]与基于竞争架构深度 Q 网络(Dueling Deep Q Network, Dueling DQN)[12],设计了基于 DDDQN (Dueling Double Deep Q Network, DDDQN)的干线动态协调控制算法。通过将干线多交叉口的交通信号作为一个整体进行处理,相比于采用多智能体协调控制,减轻了智能体间通信协调的负担,且智能体通过获取多交叉口的实时状态,掌握干线全局信息,并使用 Dueling DQN 网络结构模型,能更充分地发挥网络提取干线交通流特征的能力,挖掘出多交叉口间协调控制的相关性。实验结果表明,本文方法相比于现有绿波方法、经典的深度 Q 网络(Deep Q Network, DQN)[13]、以及 Double DQN 与 Dueling DQN,能够更有效地降低城市主干道的车均延误和车辆的停车次数等重要的交通评价指标。

1 DDDQN 介绍

1.1 Q 学习算法

Q 学习算法是由 WATKINS C J C H 等人[14]提出的一种无模型强化学习算法。在 Q 学习中,智能体通过与环境交互获得奖励来进行学习,以使得自身能根据当前的状态选择最优的动作。具体过程描述如下:在与环境交互过程中,t 时刻智能体观测到的状态为 s_t,当执行了某一动作 a_t 后,环境转移到下一状态 s_{t+1},此时智能体会获得一个相应奖励 r_t。智能体根据所有记录 (s_t, a_t, r_t, s_{t+1}),更新状态动作的 Q 值,即 $Q(s_t, a_t)$。

1.2 DQN 算法

Q 学习通过表格的形式来存储 Q 值,而对于城市干线复杂的交通流,其状态空间巨大,表格形式的存储显然无法满足需求。结合深度学习方法,采用 DQN 算法,通过深度神经网络来拟合 Q 值函数。

DQN 算法引入经验回放和目标网络两大技术[8]。智能体在与环境交互中,通过经验缓冲区存储(s_t, a_t, r_t, s_{t+1})形式的样本,在学习的过程中,再从经验缓冲区中抽取样本,经过深度神经网络的训练学习,调整网络参数,以此达到拟合 Q 值函数,实现最优策略的选取。当 t 时刻输入状态 s_t 时,输出的目标 Q 值为:

$$y_t = r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1} | \theta') \tag{1}$$

式中 $\gamma \in (0, 1)$ 表示折扣因子,θ' 表示目标网络参数。

1.3 DDDQN 算法

DQN 算法的目标 Q 值是通过贪婪法得到的,虽然这样可以让 Q 值逼近可能的优化目标,但是容易导致过估计。为了解决这一问题,Double DQN 算法被提出[11],该算法通过解耦目标 Q 值动作的选择与目标 Q 值的计算,达到消除过估计问题。Double DQN 的目标 Q 值计算如式(2)所示:

$$y_t' = r_t + \gamma Q(s_{t+1}, \arg\max_{a_{t+1}} Q(s_{t+1}, a_{t+1} | \theta) | \theta') \tag{2}$$

而 WANG Z Y 等人[12]针对 DQN 的网络结构模型进行优化,提出了 Dueling DQN。该算法将深度神经网络分成两部分,一部分表示价值函数,这部分仅仅与状态 s_t 有关,与具体采用的 a_t 无关;而另外一部分称为优势函数,同时与状态 s_t 和动作 a_t 相关。所以最终的 Q 值函数可以表示为:

$$Q(s_t, a_t, \omega, \mu, \beta) = V(s_t, \omega, \mu) + A(s_t, a_t, \omega, \beta) \tag{3}$$

式中 $V(s_t, \omega, \mu)$ 表示价值函数,$A(s_t, a_t, \omega, \beta)$ 表示优势函数,ω 为公共部分的网络参数,μ 为价值函数独有部分的网络参数,β 为优势函数独有部分的网络参数。

本文结合 Double DQN 与 Dueling DQN 算法,对 DQN 存在的过估计问题和深度神经网络结构同时优化,并根据干线交通流特性,设计了基于 DDDQN 的干线动态协调控制算法。基于 DDDQN 算法设计的目标 Q 值计算如式(4)所示:

$$y_t^{\text{DDDQN}} = r_t + \gamma Q(s_{t+1}, \arg\max_{a_{t+1}} Q(s_{t+1}, a_{t+1}, \omega, \mu, \beta), \omega', \mu', \beta') \quad (4)$$

式中,ω、μ、β 为当前网络的网络参数,ω'、μ'、β' 为目标网络的网络参数。DDDQN 算法深度神经网络定义的损失函数如式(5)所示:

$$\text{Loss} = (y_t^{\text{DDDQN}} - Q(s_t, a_t, \omega, \mu, \beta))^2 \quad (5)$$

网络训练时,还对优势函数部分做了中心化处理,得到如下公式:

$$Q(s_t, a_t, \omega, \mu, \beta) = V(s_t, \omega, \mu) + (A(s_t, a_t, \omega, \beta) - 1/|\mathscr{A}| \sum_{a' \in \mathscr{A}} A(s_t, a', \omega, \beta)) \quad (6)$$

式中 $\mathscr{A}$ 表示优势函数输出的动作空间价值。

2 基于 DDDQN 的干线动态协调控制设计

本节将针对城市交通干线的动态协调控制,定义 DDDQN 算法的状态空间、动作空间、奖励函数,以及对深度神经网络结构进行设计。

2.1 状态空间

如图 1 所示,以经典的三个交叉口组成的城市交通干线为例。将干线上各交叉口所有入口道车辆的位置信息、速度分布情况以及各交叉口当前相位作为深度神经网络的输入。考虑第一个交叉口入口道 a,距离停车线长度为 L 的路段,将每个车道等分为长度为 c 的多个小路段,每段记为 Cell。假设车辆占用了长度为 c' 的 Cell 空间,则该 Cell 的值为 c'/c,所有的 Cell 值构成了入口道 a 的位置矩阵,而在速度矩阵中,取每个 Cell 内所有车辆速度的平均值。干线的位置矩阵和速度矩阵由干线上所有入口道的位置矩阵和速度矩阵组成。

入口道 a 的车道划分如图 2 所示,对应的位置矩阵、速度矩阵分别如图 3、图 4 所示。

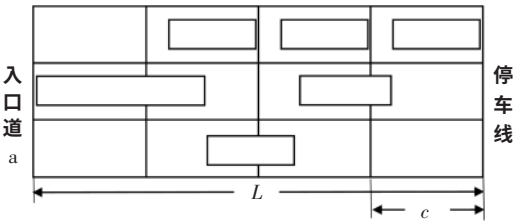

图 2 入口道 a 车道划分示意图

0	5/6	5/6	5/6
1	1/2	2/3	1/6
0	1/2	1/3	0

图 3 入口道 a 位置矩阵图

0	0	0	0
7	7	5	5
0	8	8	0

图 4 入口道 a 速度矩阵图

同时将干线上各个交叉口当前相位形成的相位矩阵作为深度神经网络的输入。假设干线上 3 个交叉口的相位均为典型的 4 相位,其相位空间如图5所示,定义相位空间 $p=\{\text{EW}, \text{EWL}, \text{SN}, \text{SNL}\}$,分别表示东西直行、东西左转、南北直行、南北左转 4 个相位。且假设当前干线上 3 个交叉口的交通信号状态均为东西直行,则相位矩阵表示为[EW, EW, EW],并在输入深度神经网络前,转换为该相位在相位空

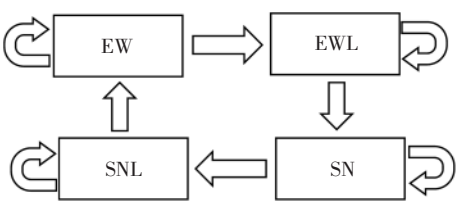

图 5 相位空间

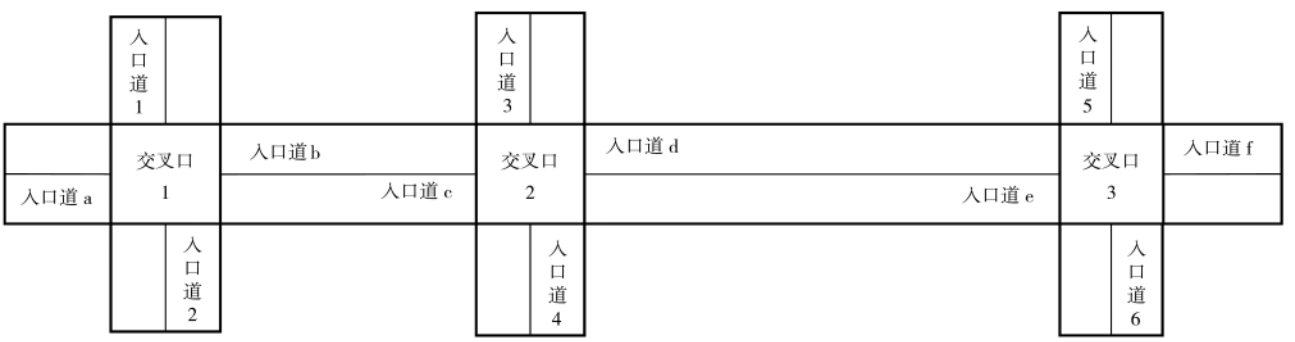

图 1 城市交通干线示意图

间的位置序号,EW 在相位空间位于第一个,即相位矩阵为[1,1,1]。

2.2 动作空间

智能体根据干线当前的交通状态,选择相应动作对各个交叉口交通信号进行控制,以达到动态协调干线交通流的目的。本文将干线作为整体进行处理。每个交叉口的动作空间如图 6 所示,定义 $a_i=\{0,1\}$。当 $a_i=0$ 时,第 i 交叉口保持当前相位 1 s;当 $a_i=1$ 时,切换至下一个相位,相位切换按照相位空间顺序循环切换。为更适用现实情况,达到更好的控制效果,对待执行动作做如下限制:当前相位持续保持时间不小于交叉口的最小绿灯时间,才允许切换至下一相位,否则继续保持当前相位;若当前相位持续保持时间大于交叉口的最大绿灯时间,则强制切换至下一相位。为进一步确保各个交叉口的交通安全性,在相位切换前,通过设置黄灯进行过渡,即当 $a_i=1$ 时,先执行一个黄灯状态,黄灯保持时间为3 s。

图 6 动作空间

每个交叉口用一位二进制对其动作编码表示,所以对于 n 个交叉口的交通干线,需要 n 位二进制,动作空间的动作总数为 2^n。

2.3 奖励函数

智能体根据获得的奖励来调整所选取的最佳动作,使得长期累积奖励最大。本文定义的奖励为:

$$r_t=\begin{cases}-1, & \max(D_t)>D\\ 1/d_t, & d_t>0\\ 1, & d_t=0\end{cases} \tag{7}$$

式中,D_t 表示 t 时刻各个支路延误的集合,在所有支路中,当存在某一支路的延误超过给定的阈值 D 时,给予智能体惩罚;否则考虑主干道是否有延误产生,奖励为主干道延误 d_t 的倒数,即延误越低,奖励越大;当主干道不存在延误,即 $d_t=0$ 时,给予智能体最大奖励。

2.4 深度神经网络结构

本文设计的深度神经网络结构如图 7 所示。将干线的车辆位置矩阵和速度矩阵以及各交叉口当前相位组成的相位矩阵作为深度神经网络的输入。位置矩阵和速度矩阵经两层卷积层提取相应的特征,并将特征展开为一维,与相位矩阵经过一个全连接层后,两者一起输入全连接层,并将最后一层全连接层分成价值 V 和优势 A 两部分,并根据式(6)融合两部分输出,最终输出 Q 值。

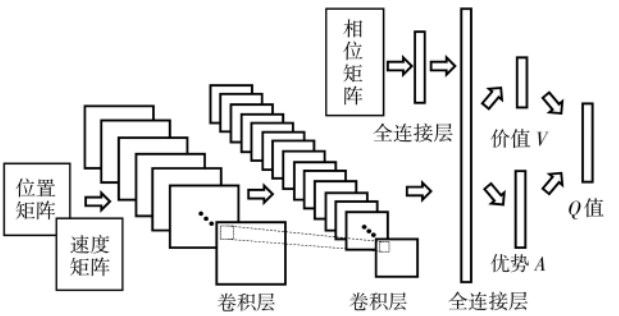

图 7 深度神经网络结构

训练时,深度神经网络采用 Adam 优化器和 ε 贪心策略进行网络输出动作的选取。ε 值随着训练次数的增加而增加,并最终以 0.999 的概率选择深度神经网络输出最大价值的动作。

3 仿真实验及分析

3.1 仿真环境与参数设置

本文的仿真实验平台选择中国科学技术大学微观交通仿真平台 2.1(USTC Microscopic Traffic Simulator 2.1,USTC-MTS 2.1)。算法模型通过 Python 实现,并采用 PyTorch 搭建深度神经网络结构。

参数设置:折扣因子 $\gamma=0.9$,经验池大小 $M=2\,000$,神经网络学习率为 0.001,卷积核大小为 5×5,批处理大小为 32,训练循环轮数为 400,车道划分取 $c=6$ m。干线上各交叉口间距如表 1 所示。

表 1 干线各交叉口间距

路段	交叉口 1-交叉口 2	交叉口 2-交叉口 3
间距/m	396	680

实际生活中,交通拥堵普遍发生在近饱和交通流量状态下,并且在进行干线协调控制时,主干道上各个交叉口往往已存在车辆排队。为使实验更贴近现实情况,本文研究在近饱和流量,且主干道形成初始排队情况下的干线协调控制。主干道各个入口道的平均初始排队长度如表 2 所示。

表 2 主干道入口道平均初始排队长度

入口道编号	a	b	c	d	e	f
排队长度/m	112	96	128	130	104	125

3.2 结果分析

在近饱和流量和主干道存在初始排队的情况下,将本文方法与现有绿波方法(用 GW 表示)、经典 DQN 算法、Double DQN 和 Dueling DQN 进行比较。

指标参数选择干线车均延误和停车次数,实验结果如图 8、图 9 所示。

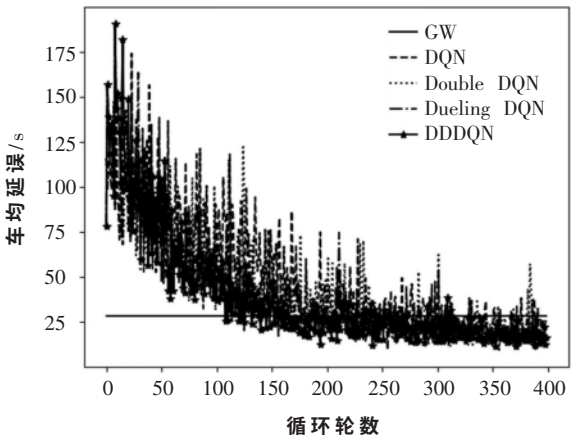

图 8 车均延误

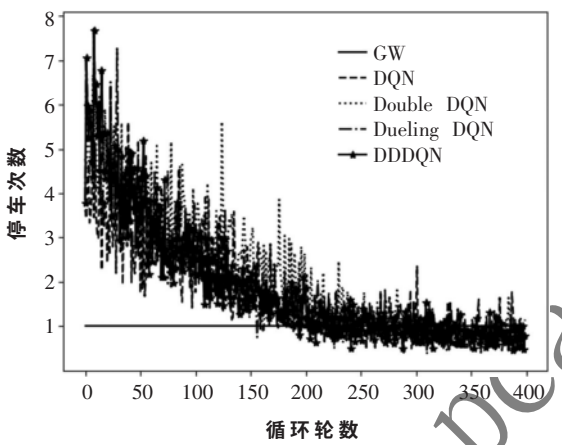

图 9 停车次数

根据实验结果,各类算法在主干道平均每轮的车均延误与停车次数如表 3 所示。

表 3 算法性能比较

	车均延误/s	停车次数
GW	27.95	1.01
DQN	22.6	0.89
Double DQN	21.38	0.88
Dueling DQN	19.61	0.86
DDDQN	18.47	0.81

从表 3 可见,采用 DQN 算法及其相关改进算法 Double DQN、Dueling DQN 以及本文的 DDDQN 算法,相比于现有绿波方法,在城市干线的车均延误与车辆停车次数等重要交通评价指标方面均有较大的改善。说明基于 DQN 设计的算法能根据实时的交通流状态,对干线各个交叉口交通信号实现动态协调控制,相比于现有绿波方法,能够更加有效地降低主干道的车均延误与停车次数,进一步改善城市交通拥堵现象。其中使用基于 DDDQN 的干线动态协调控制算法,其干线的车均延误与停车次数均为最小,算法性能最优。且从图 8 的车均延误曲线与图 9 的停车次数曲线可以看出,相比于采用经典DQN 算法以及单独使用 Double DQN 与 Dueling DQN,采用 DDDQN 算法的车均延误曲线与停车次数曲线波动程度最小,最为稳定,且收敛速度较快,对城市干线的动态协调效果最优。

4 结论

本文将深度强化学习方法引入到城市交通干线的动态协调控制中,结合 Double DQN 与 Dueling DQN,给出了一种 DDDQN 的干线动态协调算法。通过解耦目标 Q 值动作的选取与目标 Q 值的计算,消除了 DQN 的过估计问题,同时对深度神经网络结构进行优化,将输出分为状态价值与动作优势两部分,并做了中心化处理,使得智能体能更好地进行干线交通信号决策控制。并且本文将干线作为一个整体处理,通过将干线整体的交通状态输入深度神经网络,能够更充分发挥网络挖掘干线各交叉口协调控制的相关性。实验结果表明,DDDQN 算法较现有绿波方法、经典的 DQN、Double DQN 与 Dueling DQN,有效地降低了城市干线的车均延误与停车次数。后续工作考虑将 DDDQN 算法的应用扩展至城市路网,实现区域协调优化控制。

参考文献

[1] LITTLE J D C, KELSON M D, GARTNER N H. Maxband: a program for setting signal on arteries and triangular network[J]. Journal of the Transportation Research Board, 1981, 795: 40-46.

[2] GARTNER N H, ASSMANN S F, LASAGA F, et al. Multiband-a variable bandwidth arterial progression scheme[J]. Journal of the Transportation Research Board, 1990, 1287: 212-222.

[3] 陈昕,张驰.基于绿波带中心线交点的双向绿波控制图解法[J].辽宁工业大学学报(自然科学版),2017,37(2):137-140.

[4] 卢凯,徐广辉,林观荣,等.绿灯终点型双向绿波协调控制数解算法[J].中国公路学报,2019,32(11):202-211.

[5] 曲大义,周警春,杨晶茹,等.绿波协调下公交车辆在交叉口的延误影响分析[J].山东科技大学学报

(自然科学版), 2019, 38(6): 98-104.

[6] HA-LI P, KE D. An intersection signal control method based on deep reinforcement learning[C]. 2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA). Changsha, 2017: 344-348.

[7] GAO J, SHEN Y, LIU J, et al. Adaptive traffic signal control: deep reinforcement learning algorithm with experience replay and target network[J]. arXiv preprint arXiv: 1705.02755, 2017.

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

[9] LI C C, YAN F, ZHOU Y D, et al. A regional traffic signal control strategy with deep reinforcement learning[C]. 2018 37th Chinese Control Conference (CCC). Wuhan, 2018: 7690-7695.

[10] VAN DER POL E. Deep reinforcement learning for coordination in traffic light control[D]. Amsterdam: University of Amsterdam, 2016.

[11] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q learning[C]. Association for the Advance of Artificial Intelligence, 2016: 2094-2100.

[12] WANG Z Y, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[J]. arXiv preprint arXiv: 1511.06581, 2015.

[13] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[C]. Proceedings of Workshops at the 26th Neural Information Processing Systems, 2013. Lake Tahoe, USA, 2013: 201-220.

[14] WATKINS C J C H, DAYAN P. Q-learning[J]. Machine Learning, 1992, 8(3-4): 279-292.

(收稿日期: 2020-04-20)

作者简介:

郭瑾清(1994-),**通信作者**,**男**,**硕士研究生**,**主要研究方向**:**深度强化学习**、**智能交通**。E-mail: ghq@mail.ustc.edu.cn。

陈锋(1966-),**男**,**博士**,**副教授**,**主要研究方向**:**智能交通**、**人工智能**。

版权声明