

文章编号: 1671-7449(2024)04-0371-07

无人机辅助智能交通系统中基于C-DQN的信息年龄最小化研究

李登治, 王宁*

(山西大学 物理电子工程学院, 山西 太原 030006)

摘要: 无人机(Unmanned Aerial Vehicle, UAV)由于高机动性和低运营成本的优势,有望在未来智能交通系统(Intelligent Transportation System, ITS)中被部署为车辆数据收集器。为了保证所采集数据的时效性,以信息年龄(Age of Information, AoI)来衡量无人机从车辆接收到的数据的新鲜度。因此,提出一种基于信息年龄最小化的无人机辅助智能交通系统方案,通过联合优化无人机轨迹和无人机与车辆的关联策略,最小化所有车辆的信息年龄加权和。将该优化问题建模为一个马尔可夫决策过程(Markov Decision Process, MDP),并采用Curiosity-Driven Deep Q-learning Network(C-DQN)算法来求解。大量仿真结果表明,该算法在探索能力和收益性能方面优于传统Deep Q-learning Network(DQN)算法。

关键词: 无人机; 智能交通系统; 信息年龄; 马尔可夫决策过程; C-DQN

中图分类号: TP393 **文献标识码:** AA **doi:** 10.3969/j.issn.1671-7449.2024048

引用格式: 李登治, 王宁. 无人机辅助智能交通系统中基于C-DQN的信息年龄最小化研究[J]. 测试技术学报, 2024, 38(4): 371-377.

LI Dengzhi, WANG Ning. Research on age of information minimization based on C-DQN in UAV assisted intelligent transportation system[J]. Journal of Test and Measurement Technology, 2024, 38(4): 371-377.

Research on Age of Information Minimization Based on C-DQN in UAV Assisted Intelligent Transportation System

LI Dengzhi, WANG Ning*

(College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

Abstract: Unmanned aerial vehicle (UAV) have the potential to be deployed as vehicle data collectors in future intelligent transportation systems (ITS) due to their high maneuverability and low operational costs. To ensure the timeliness of the collected data, the freshness of the data received by UAVs from vehicles is measured in terms of age of information (AoI). Therefore, this paper proposes a UAV-assisted ITS scheme based on minimizing the Age of Information, by jointly optimizing the UAV's trajectory and the association strategy between the UAV and vehicles, to minimize the weighted sum of AoI for all vehicles. The optimization problem is modeled as a MARKOV decision process (MDP), and the curiosity-driven deep Q-learning network (C-DQN) algorithm is employed to solve it. Extensive simulation results demonstrate that this algorithm outperforms the traditional deep Q-learning network (DQN) algorithm in terms of exploration capability and performance gain.

Key words: Unmanned aerial vehicle; intelligent transportation system; age of information; markov decision process; C-DQN

收稿日期: 2023-07-22

作者简介: 李登治(1996-), 男, 硕士生, 主要从事5G资源分配、强化学习等研究。E-mail: 13393478022@qq.com。

* 通信作者: 王宁(1989-), 女, 副教授, 主要从事量子通信与量子光学等研究。E-mail: 992angning@sxu.edu.cn。

0 引言

目前,无人机(UAV)辅助智能交通系统已经在学术界和工业界引起了极大的关注^[1-3]。然而,大部分现有的研究工作都集中在提高无人机的通信覆盖率^[4]、系统吞吐量^[5]以及减少无人机的传输延迟^[6]等方面。随着5G技术的商业化^[7],新兴应用领域如智能农业、森林火灾监测和铁路巡检等将得到广泛应用。这些应用对传输质量和延迟要求极高,并对信息的时效性非常敏感。特别是在无人驾驶和森林火灾监测领域,过时的信息可能导致错误的控制决策,甚至引发严重事故。因此,信息新鲜度对于这些应用至关重要。

为了有效地量化信息新鲜度,提出了信息年龄(AoI)^[8]的概念,其定义为自信息生成后到达接收器的时间间隔。通过引入AoI作为衡量指标,能够更加准确地评估信息的时效性,较低的AoI值表示信息更新速度更快,接收器可以更及时地获取到最新的信息。在无人机辅助智能交通系统中,这意味着可以更准确地做出决策和调整,以提高交通管理的效率和安全性。

在智能交通系统中,无人机可以通过轨迹规划为更多的车辆用户提供服务。深度强化学习结合了深度学习和强化学习的优势,能够高效建模复杂环境并做出自主决策。在处理轨迹规划问题上,深度强化学习具有天然的优势^[9]。例如,张明等^[10]使用Q-Learning算法对无人机轨迹进行规划,以在飞行期间给用户提供最优化的总数据速率;朱晓荣等^[11]提出了基于内容感知的无人机轨迹规划和资源分配联合优化方法,来最大化提高用户平均服务速率;李延儒等^[12]使用Deep Q-Learning Network(DQN)算法来进行无人机智能航路规划,以实现自主避障。然而,尽管以上算法可以进行有效路径规划,但有时它们可能缺乏探索能力^[13]。

本文提出了一种无人机辅助智能交通系统中基于Curiosity-Driven Deep Q-Learning Network(C-DQN)算法的所有车辆AoI加权和最小化系统模型,通过联合优化无人机轨迹和无人机与车辆的关联策略,最小化所有车辆的信息年龄加权和,C-DQN算法是基于内在奖励的算法,具有比DQN更强的探索能力^[14]。

1 系统模型

1.1 系统架构

如图1模拟实际中城市区域内一个井字形的路况。考虑了UAV辅助ITS场景。其中车辆在道路上以随机速度行驶,车辆的速度遵循一个截断的高斯分布,其范围从 μ_{min} 到 μ_{max} ^[15]。在该区域中,无人机从起点到目的地巡航,旨在收集车辆的状态信息以使其尽可能新鲜。

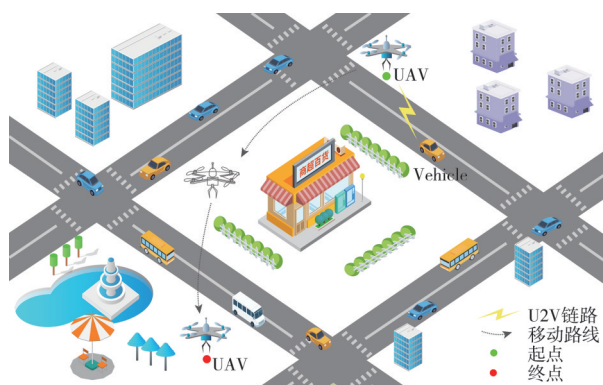


图1 系统场景

Fig. 1 System scenario

1.2 UAV-车辆通信模型

假设每个时隙相等且一个时隙的持续时间为 Δt , $T = \{0, \dots, t, \dots, T\}$ 表示时隙集。 $V = \{1, 2, \dots, \nu\}$ 表示该区域内所有车辆的集合, $q'_v = (x'_v, y'_v, 0)$ 表示车辆 v 在时隙 t 的道路坐标。请注意在考虑的时间范围内,车辆进出该区域,这会导致区域内车辆数目的变化,规定区域内每个时隙的车辆数目是固定的。 $q(t) = (x(t), y(t), H)$ 表示UAV时隙 t 的坐标,其中UAV在固定高度 H 上运动。 $q(0) = (x_{orig}, y_{orig}, H)$ 表示UAV的初始位置, $q(T) = (x_{dest}, y_{dest}, H)$ 表示UAV的目的地。因此,根据车辆和UAV位置坐标可以计算UAV和车辆 v 之间的距离为

$$d_{v,u}^t = \sqrt{(x'_v - x(t))^2 + (y'_v - y(t))^2 + H^2}, \quad \forall v \in V. \quad (1)$$

UAV-to-Vehicle(U2V)无线信道遵循自由空间路径损耗数学模型,因此在时隙 t 内信道增益可以表示为

$$h_{v,u}^t = \zeta_0 (d_{v,u}^t(q(t)))^{-2}, \quad \forall v \in V, \quad (2)$$

式中: ζ_0 是参考距离1 m处的信道增益。

无人机和车辆 v 之间在时隙 t 时的信噪比可以用 $SINR_{v,u}^t$ 表示为

$$SINR_{v,u}^t = \frac{p_u h_{v,u}^t}{\sigma^2}, \quad (3)$$

式中: p_u 为 UAV 的发射功率, σ^2 表示车辆 v 处自然噪声功率。因此, 在时隙 t 时车辆 v 与 UAV 之间可实现的数据速率

$$R_{v,u}^t = B \log_2(1 + SINR_{v,u}^t), \quad (4)$$

式中: B 为信道带宽。

1.3 AoI 模型

AoI 的概念从接收者(即 UAV)的角度描述了信息的新鲜度, 用 $D_{v,u}$ 表示在时隙 t 时 UAV 从车辆 v 上接收信息的数据量。

$$D_{v,u}^t = c_{v,u}^t R_{v,u}^t \Delta t, \quad (5)$$

式中: $c_{v,u}^t$ 表示 UAV 和车辆 v 的调度策略, 即 $c_{v,u}^t \in \{0, 1\}, \forall v \in V, t \in T$, 其中 $c_{v,u}^t = 1$ 表示在时隙 t 时 UAV 为车辆 v 提供服务, 否则 $c_{v,u}^t = 0$ 。同时, 利用时分多址技术, 即一个时隙 UAV 最多只能为该区域内一个车辆提供信息采集服务, 并且 UAV 在采集车辆信息时处于悬停状态, 即有

$$\sum_{v=1}^V c_{v,u}^t \leq 1, \forall v \in V, t \in T. \quad (6)$$

用 D_{\min} 表示成功恢复或解码接收到的数据所需的最小数据量, 当 $D_{v,u} \geq D_{\min}$ 时, 更新车辆 v 的信息年龄为 $A_v^t = 1$, 这意味着车辆 v 的当前状态信息成功恢复或解码到 UAV。如果 $D_{v,u} < D_{\min}$ 时表示数据恢复或解码失败, 此时车辆 v 的信息年龄为 $A_v^t = A_v^{t-1} + 1$, 其细节如图 2 所示。

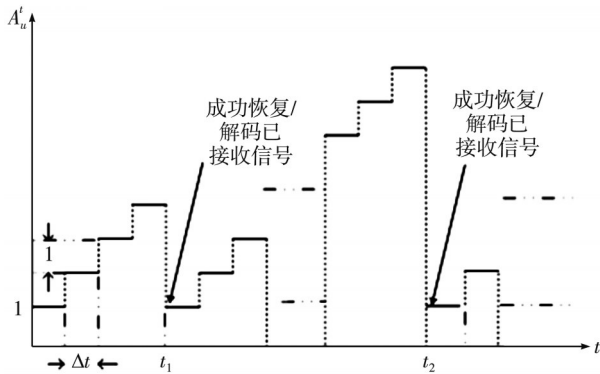


图 2 AoI 演进图

Fig. 2 AoI evolutionary chart

2 问题表述

通过联合优化 UAV-车辆调度策略 $c_{v,u}^t$ 和

UAV 的轨迹 $q(t)$, 使 UAV 整个飞行周期内的长期 AoI 加权和最小。问题表述为

$$\min: \sum_{t=1}^T \sum_{v=1}^V \varpi_v A_v^t, \quad (7)$$

$$\text{subject to } q(0) = (x_{\text{orig}}, y_{\text{orig}}, H), \quad (7a)$$

$$q(T) = (x_{\text{dest}}, y_{\text{dest}}, H), \quad (7b)$$

$$c_{v,u}^t \in \{0, 1\}, \forall v \in V, t \in T, \quad (7c)$$

$$\sum_{v=1}^V c_{v,u}^t \leq 1, \forall v \in V, t \in T, \quad (7d)$$

$$\|q(t) - q(t-1)\| \leq U_{\max} \Delta t, \quad (7e)$$

式中: ϖ_v 为车辆 v 的权重, 代表不同车辆的重要性不同; 式(7a)和式(7b)规定了 UAV 的初始和最终的位置, 式(7c)和式(7d)是调度策略, 表示一个时隙内 UAV 最多只能为一个车辆进行服务, 式(7e)是无人机的速度约束, 其中 U_{\max} 是 UAV 的最大速度。

问题(7)需要随着时间的推移依次做出一系列决策。传统的方法, 如静态优化和博弈论, 无力处理该问题, 它们试图根据当前状态搜索最优/次优解, 以最大限度地提高即时回报。基于上述观察, 将式(7)建模为马尔可夫决策过程(MDP)。将 MDP 用一个元组表示: $\langle S, A, P, r \rangle$, 每个元素具体描述如下:

S 是环境的状态空间。包括: 第 t 个时隙时的所有车辆坐标 $q_v^t = (x_v^t, y_v^t, 0)$; UAV 在时隙 t 时的坐标 $q(t) = (x(t), y(t), H)$; 在时隙 t 时所有车辆的 AoI 值; 无人机初始和最终的位置。

A 是环境的动作空间。包括: 无人机-车辆调度策略 $c_{v,u}^t$ 和 UAV 的方向(前, 后, 左, 右及其组合)和速度 U 。

P 是状态转移概率函数。包括无人机坐标根据 $q(t) = U \Delta t + q(t-1)$ 进行转移, $U \leq U_{\max}$, 其中 U 是无人机的飞行速度。而车辆在下一个时隙的坐标由车辆的当前位置、速度和加速度决定。

r 是即时奖励函数, 定义为

$$r = \begin{cases} -\sum_{v=1}^V \varpi_v A_v^t, & \text{if } q(T) \neq (x_{\text{dest}}, y_{\text{dest}}, H), \\ -\sum_{v=1}^V \varpi_v A_v^t + \Omega, & \text{otherwise,} \end{cases} \quad (8)$$

式中: Ω 是一个正数, 用于将无人机诱导到最终位置。

之后采用深度强化学习算法即 C-DQN 来求解上述 MDP 问题。

3 基于C-DQN算法的设计

传统的强化学习(RL)算法主要是通过最大化外部奖励(环境提供的奖励)来不断进行学习,从而尽可能地实现目标任务。然而,这些现有的RL方法存在一定的缺陷,它们在应对相对复杂问题时大多缺乏探索能力^[16]。为了弥补这一缺陷,使算法具有更强大的探索能力,提出使用C-DQN来解决该问题。如图3所示,该算法主要由两大部分组成:上面一部分是传统DQN模块;下面一部分是基于好奇心的内在驱动模块,它是算法具有探索能力的关键。相比于只能通过与外部环境交互而获得奖励的DQN算法,C-DQN的最大区别在于智能体可以由自身的内在好奇奖励所驱动,就像机器有了人的好奇心一样,可以自行的探索,这使得算法拥有更快的收敛速度,收敛值也会更大。

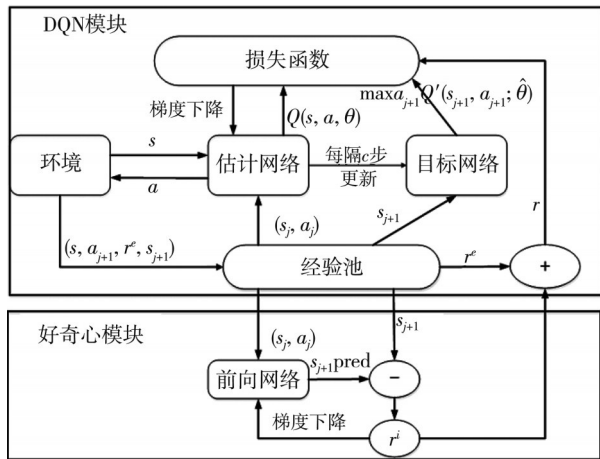


图3 C-DQN算法

Fig. 3 C-DQN algorithm

在图3显示的C-DQN算法框架中,内在奖励信号 γ^i 由好奇心网络输出,传统的DQN则输出一个基于训练过的状态-动作值函数 $Q(s, a)$ 的动作。在每次的学习步骤中,智能体在状态 s 下采取动作 a ,并通过自身内部探索奖励 r^i 和外部环境驱动奖励 r^e 使学习过程被强化。

3.1 DQN模块

在无人机与环境交互模型中,无人机在 s_t 状态下,根据 ϵ -贪心策略随机选择动作 a_t ,获得到外部奖励 r_t^e 和下一个状态 s_{t+1} 。

$$r^e = - \sum_{v=1}^V \omega_v A_{v0}^t \quad (9)$$

将 $(s_t, a_t, r_t^e, s_{t+1})$ 存储到经验池中,用于神经网络的训练。训练时,从经验池中任意抽取一小块经验值 (s_j, a_j, r_j, s_{j+1}) ,根据贝叶斯公式得到标签

$$y_j = \begin{cases} r_j, & j+1 \text{ 步结束,} \\ \gamma \max_{a_{j+1}} Q'(s_{j+1}, a_{j+1}; \hat{\theta}) + r_j, & \text{其他,} \end{cases} \quad (10)$$

式中: r_j 不同于传统的DQN网络,它包括与环境相关的外部奖励 r_j^e 和内部奖励 r_j^i ,其中 $r_j = r_j^e + r_j^i$, $\gamma \in [0, 1]$ 表示折扣因子, Q' 由权值为 $\hat{\theta}$ 的目标网络状态动作函数值近似等于(经过几步后更新为 $\hat{\theta} = \theta$),参数 θ 沿损失函数梯度下降的方向更新,即DQN算法通过梯度下降的策略最小化损失函数来优化DQN神经网络的参数 θ 。

损失函数

$$Loss(\theta) = E(y_j - Q(s_j, a_j; \theta))^2, \quad (11)$$

式中: $Q(s_j, a_j; \theta)$ 为参数 θ 的网络估计状态动作函数, y_j 为目标值,该策略可以表示为

$$\theta \leftarrow \theta - \alpha_q (y_j - Q(s_j, a_j; \theta)) \cdot \nabla_{\theta} Q(s_j, a_j; \theta), \quad (12)$$

式中: α_q 是DQN的学习率,需满足 $\alpha_q \in (0, 1)$ 。

估计网络设置固定的时长对目标网络进行参数更新,以便更好地消除时间相关性,保证数据的准确。

3.2 基于好奇心的内在驱动模块

本文对原始的好奇心模型进行了修改,删除它的反向模型,只留下前向模型,在原始模型中,反向模型的作用只是预测智能体从当前状态变化到下一个状态时将要采取的动作,但在本文的场景中,智能体在采取下一个动作时,对当前的状态空间已经进行了筛选,那些无效的状态提前被去除。下一状态则可以通过前向模型进行预测,最后将状态的预测值与下一状态的实际值相比较,就可以计算内在奖励的大小。

根据从经验池中抽取的 a_j 和 s_j ,先将初始状态 s_t 编码为特征向量 $\phi(s_j)$,再通过前向模型对这两个输入进行预测,得到与参数为 ϑ 相关的下一状态预测值为

$$\phi_{\text{pred}}(s_{j+1}) = f(a_j, \phi(s_j); \vartheta), \quad (13)$$

式中: $\phi_{\text{pred}}(s_{j+1})$ 为 $\phi_{s_{j+1}}$ 的预测估计,函数 f 为前向模型,此时损失函数为

$$Loss(\vartheta) = \frac{1}{2} \|s' - \phi(s, a, \vartheta)\|_2^2 \quad (14)$$

前向模型通过对式(13)使用梯度下降策略来优化更新好奇心神经网络的参数 ϑ , 即求损失函数的最小值, 该策略可以表示为

$$\vartheta \leftarrow \vartheta - \alpha_q (y - \phi_{pred}(s_{j+1})) \cdot \nabla_{\vartheta} \phi_{pred}(s_{j+1}), \quad (15)$$

式中: α_q 是好奇心驱动模块的学习率, 需满足 $\alpha_q \in (0, 1)$ 。为了延长好奇心的时长, ϑ 在经过几个步骤后就会被更新一次, 前向模型的预测误差也被用来衡量好奇心, 在研究中将其定义为好奇心驱动的内在奖励, 它与预测的下一个状态和实际状态有关, 表示为

$$r_i = \frac{\sigma}{2} \|\phi_{pred}(s_{j+1}) - \phi(s_{j+1})\|_2^2, \quad (16)$$

式中: σ 是比例因子, 需满足 $\sigma > 0$ 。 r_i 鼓励智能体的好奇心, 并推动它更有效地探索环境, 当内在奖励趋于 0 时, 智能体探索的好奇心已经消失, 该模块已完成参数的训练更新。

4 仿真结果分析

在本节中, 给出了所提算法的仿真结果。仿真在基于 Python 的模拟器上实现, 其中软件环境

为 TensorFlow1. 15. 0, Python3. 6. 5。硬件环境为基于 CPU 的服务器, 具有 8GB 3200MHz DDR4, 3. 0 GHz AMD Core R5 和 512G 内存。

仿真中用 SUMO 软件在某个城市十字道路上随机生成车辆, 并生成车辆轨迹。无人机从起点出发, 在该区域上空巡航用于接收车辆用户的数据, 在时隙 T 时回到终点。无人机-车辆无线信道遵循自由空间路径损耗数学模型, 其飞行高度 H 为固定值 (50~150 m), 传输功率 P_u 设置为 0. 01 W, 带宽 $B=0. 9$ MHz。每个时隙的长度 $\Delta_t = 1$ s。学习参数设置为折扣因子 $\gamma = 0. 95$, 经验池大小 20 000, 一次训练所选取样本数为 256。学习率讨论如下:

图 4 是 C-DQN 算法不同学习率的收敛性能, 其中无人机高度 $H=100$ m, 车辆数目 $V=4$ 。从图中可以看出, $\alpha_q = 0. 0006$, 在图中 1 500 个回合左右就达到了收敛, 而 $\alpha_q = 0. 0003$ 在 2 500 回合左右才达到了收敛。当 $\alpha_q = 0. 001$ 时, 虽然前期探索期奖励值比 $\alpha_q = 0. 0006$ 时快, 但是后期收敛速度较慢, 2 000 回合才达到收敛, 且收敛后整体奖励值低于 $\alpha_q = 0. 0006$ 时的奖励值。因此, 当学习率为 0. 0006 时学习效果最佳, 将学习率固定为 $\alpha_q = 0. 0006$ 。

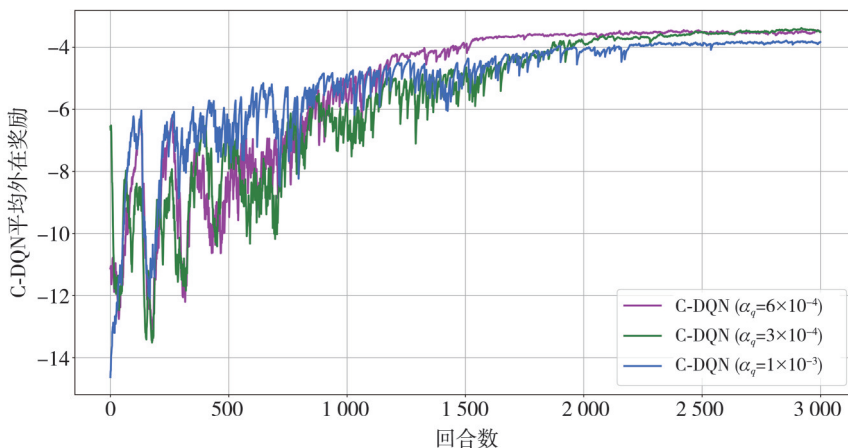


图 4 不同学习率比较

Fig. 4 Comparison of different learning rates

图 5 显示了 C-DQN 算法与 DQN 算法的收敛性能比较, 图中横轴表示回合数, 左边纵轴表示每个回合积累的平均外部奖励, 右边纵轴表示内在奖励, 其中无人机的高度 $H=100$ m, 车辆数目 $V=4$, $\alpha_q = \alpha_a = 0. 0006$ 。从图中可以看出, 智能体的好奇心一开始很高, 随着智能体对环境的不断探索, 对环境逐渐熟悉, 好奇心逐渐下降, 并最终趋于零。C-DQN 算法在 1 500 回合左右达到收敛, 而 DQN 算法在 2 500 回合左右才达到收敛,

且 DQN 算法的波动性明显高于 C-DQN 算法的波动性, 并且整体平均奖励低于 C-DQN 算法奖励值, 因此 C-DQN 算法相比传统 DQN 性能更佳。

图 6 显示了每个回合的 AoI 加权和与 UAV 飞行高度的关系, 其中车辆数目 $V=4$, $\alpha_q = \alpha_a = 0. 0006$ 。我们注意到, 当无人机的高度增加时, AoI 值增加。由于车辆到无人机的信道增益主要取决于两者之间的距离, 因此在带宽和发射功率一定的情况下, 飞行高度越高, 信道条件越弱, 传

输速率越低,且同等高度下,C-DQN相比于DQN 学习效果更好。

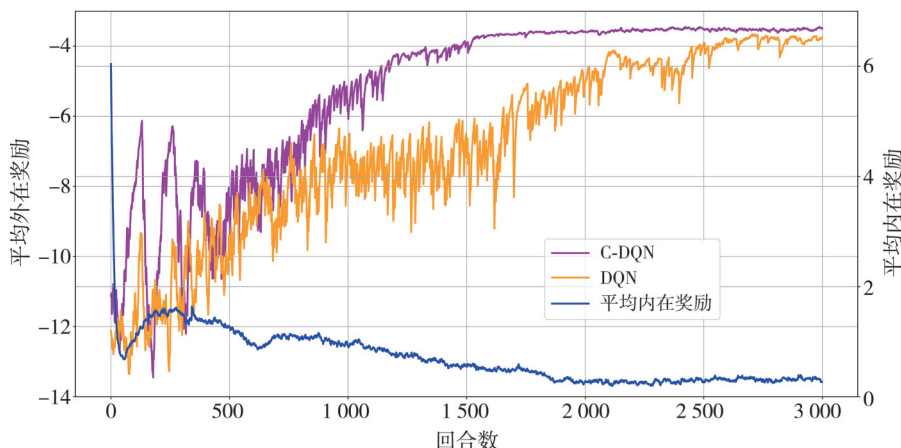


图5 C-DQN与DQN收敛性能比较图

Fig. 5 Comparison graph of convergence performance between C-DQN and DQN

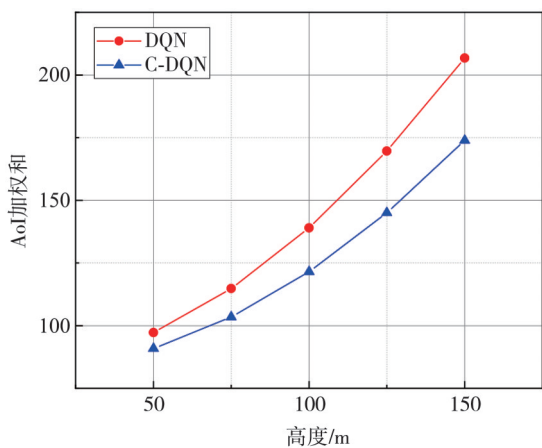


图6 不同高度下AoI加权和

Fig. 6 Sum of AoI at different heights

图7显示了每个回合的不同车辆数目下AoI加权和,其中无人机高度 $H=100\text{ m}$, $\alpha_q = \alpha_a = 0.0006$ 。

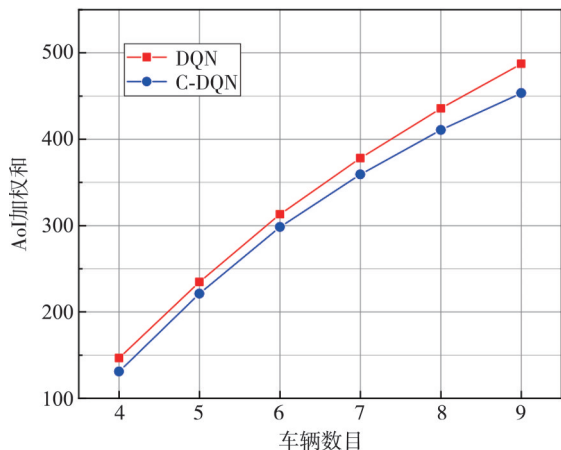


图7 不同车辆数目的AoI加权和

Fig. 7 Sum of AoI for different numbers of vehicles

研究发现,随着车辆数目的增加,AoI加权和显著增加。这是因为无人机在每个时隙最多连接一个车辆,道路上车辆越多,平均每个车辆享受的服务越少,AoI之和也随之增加。此外,从图7中可以看出,在同样车辆数目下,C-DQN算法比DQN算法的AoI加权和要小,证明了C-DQN算法比传统DQN更强的探索性。

5 结语

本文提出了基于信息年龄最小化的无人机辅助智能交通系统模型,通过联合优化无人机轨迹和无人机与车辆的关联策略,最小化所有车辆的信息年龄加权和。将该联合优化问题建模成马尔可夫决策过程,采用C-DQN来解决。仿真结果表明该算法比传统DQN算法有更强的探索能力和收益性能。

参考文献:

[1] 闫铭. 无人机智能交通系统[J]. 科技风, 2020(22): 12-13.
 YAN Ming. UAV intelligent transportation system of unmanned aerial vehicle[J]. Technology Wind, 2020 (22): 12-13. (in Chinese)

[2] 赵立峰. 基于Adaptive MRSMC多旋翼无人机自主飞行控制系统研究[D]. 南京: 南京信息工程大学, 2016.

[3] 刘晓锋, 彭仲仁, 张立业, 等. 面向交通信息采集的无人飞机路径规划[J]. 交通运输系统工程与信息, 2012, 12(1): 91-97.
 LIU Xiaofeng, PENG Zhongren, ZHANG Liye,

- et al. Unmanned aerial vehicle route planning for traffic information collection [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2012, 12(1): 91-97. (in Chinese)
- [4] LIU X, LI Z, ZHAO N, et al. Transceiver design and multihop D2D for UAV IoT coverage in disasters [J]. *IEEE Internet of Things Journal*, 2019, 6(2): 1803-1815.
- [5] ZENG Y, ZHANG R, LIM T J. Throughput maximization for UAV-enabled mobile relaying systems [J]. *IEEE Transactions on Communications*, 2016, 64(12): 4983-4996.
- [6] LI J, HAN Y. A traffic service scheme for delay minimization in multi-layer UAV networks [J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(6): 5500-5504.
- [7] LIU Y, XIONG K, NI Q, et al. UAV-assisted wireless powered cooperative mobile edge computing: Joint offloading, CPU control, and trajectory optimization [J]. *IEEE Internet of Things Journal*, 2020, 7(4): 2777-2790.
- [8] KAUL S, YATES R, GRUTESER M. Real-time status: How often should one update? [C]//2012 Proceedings IEEE INFOCOM, 2012: 2731-2735.
- [9] CHOI H C, DENG C, HWANG I. Hybrid machine learning and estimation-based flight trajectory prediction in terminal airspace [J]. *IEEE Access*, 2021, 9: 151186-151197.
- [10] ZHANG M, DONG C, FENG S, et al. Adaptive 3D routing protocol for flying ad hoc networks based on prediction-driven Q-learning [J]. *China Communications*, 2022, 19(5): 302-317.
- [11] 朱晓荣, 张文锦, 赵凌宇, 等. 基于内容感知的无人机轨迹规划和资源分配联合优化方法 [J]. *电子与信息学报*, 2023, 45(5): 1644-1650.
- ZHU Xiaorong, ZHANG Wenjin, ZHAO Lingyu, et al. UAV trajectory planning and resource joint optimization method based on content-aware [J]. *Journal of Electronics & Information Technology*, 2023, 45(5): 1644-1650. (in Chinese)
- [12] 李延儒, 左铁东, 王婧. 基于 DQN 深度强化学习的无人机智能航路规划方法研究 [J]. *电子技术与软件工程*, 2022(18): 5-8.
- LI Yanru, ZUO Tiedong, WANG Jing. Research on UAV intelligent route planning method based on DQN depth reinforcement learning [J]. *Electronic Technology & Software Engineering*, 2022(18): 5-8. (in Chinese)
- [13] FU F, JIAO Q, YU F R, et al. Securing UAV-to-vehicle communications: a curiosity-driven deep Q-learning network (C-DQN) approach [C]//2021 IEEE International Conference on Communications Workshops (ICC Workshops), 2021: 1-6.
- [14] LEE S, LEE H. MUSK-DQN: multi-UBS selective-K deep Q-network for maximizing energy-efficiency [C]//2023 IEEE 20th Consumer Communications & Networking Conference (CCNC), 2023: 438-442.
- [15] KADOTA I, SINHA A, MODIANO E. Optimizing age of information in wireless networks with throughput constraints [C]//IEEE INFOCOM 2018-IEEE Conference on Computer Communications. 2018: 1844-1852.
- [16] 沈冰, 许士勋, 卢小凤, 等. 无人机中基于 C-DQN 的资源分配和轨迹优化研究 [J]. *测试技术学报*, 2022, 36(2): 141-146.
- SHEN Bing, XU Shixun, LU Xiaofeng, et al. Research on resource allocation and trajectory optimization based on C-DQN in UAV [J]. *Journal of Test & Measurement Technology*, 2022, 36(2): 141-146. (in Chinese)