Open Journal of Transportation Technologies 交通技术, 2024, 13(3), 192-203
Published Online May 2024 in Hans. https://www.hanspub.org/journal/ojtt
https://doi.org/10.12677/ojtt.2024.133023

Hans汉斯

基于多智能体的大规模路口交通信号灯协同控制研究

夏格格，吴小花，靳增锐，孙军艳

陕西科技大学机电工程学院，陕西 西安

收稿日期：2024年3月12日；录用日期：2024年5月17日；发布日期：2024年5月31日

摘　要

随着城市化建设加快，车辆数量增加，道路负载变大，交通拥堵问题日益严重，目前大多数的控制方法仍局限于孤立的路口，整体路网拥挤程度仍存在较大的优化空间。因此，本文以交通信号灯控制为研究对象，深度强化学习算法为基础，针对基于多智能体的交通信号灯协同控制问题展开研究。首先将信号灯协同控制问题抽象为马尔可夫决策过程，基于Python开发平台搭建智能体交互环境，设计大规模路口信号灯决策下的DQN算法并进行调试运行，结果表明算法训练出了有效的红绿灯控制策略，并在各个路口都能够学习到公平策略，能够提高路网整体通行效率。最后通过与传统的信号固定配时方案进行对比实验，验证算法具有良好的优化性能。通过调整算法的超参数对训练结果进行对比分析，研究不同超参数对网络训练的影响，以及超参数对项目研究的重要性。

关键词

交通信号灯协同控制，强化学习，深度神经网络，DQN算法

Research on Multi-Agent Based Cooperative Control of Traffic Lights at Large-Scale Intersections

Gege Xia, Xiaohua Wu, Zengrui Jin, Junyan Sun

College of Mechanical and Electrical Engineering, Shaanxi University of Science and Technology, Xi'an Shaanxi

Received: Mar. 12th, 2024; accepted: May 17th, 2024; published: May 31st, 2024

文章引用: 夏格格, 吴小花, 靳增锐, 孙军艳. 基于多智能体的大规模路口交通信号灯协同控制研究[J]. 交通技术, 2024, 13(3): 192-203. DOI: 10.12677/ojtt.2024.133023

Abstract

With the continuous acceleration of urbanization construction, the number of vehicles continues to increase, the road load gradually increases, and the problem of urban traffic congestion is becoming increasingly serious. Currently, most control methods are still limited to isolated intersections, and there is still significant room for optimization of the overall road network congestion level. Therefore, this paper takes the traffic light control as the research object, and based on the Deep reinforcement learning algorithm, carries out relevant research on the multi-agent based traffic light cooperative control problem. Firstly, the problem of signal light collaborative control is abstracted as a Markov decision process. Based on the Python development platform, an intelligent learning and interaction environment is built, and the DQN algorithm for large-scale intersection signal light decision-making is designed and debugged. The results show that the algorithm has trained effective red and green light control strategies, and fairness strategies can be learned at each intersection, which can improve the overall traffic efficiency of the road network. Finally, through comparative experiments with traditional signal fixed timing schemes, the algorithm was verified to have good optimization performance. Compare and analyze the training results by adjusting the hyperparameters of the algorithm, study the impact of different hyperparameters on network training, and the importance of hyperparameters for project research.

Keywords

Traffic Light Cooperative Control, Reinforcement Learning, Deep Neural Network, DQN Algorithm

Copyright © 2024 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

目前我国驾驶人员总量和机动车保有量均居世界第一，道路建设供不应求，在现有的城市交通体系下，交通拥堵等问题日益严重，道路通行能力主要受交叉路口的影响，而路口通行能力主要受到现有红绿灯控制率的限制，仍然存在较大的优化空间[1]。随着科技的进步，传统的交通信号灯控制技术已不能持续满足当下复杂多变的交通环境。

如今，机器学习广泛地应用于各个领域，强化学习在交通信号控制领域也得到了应用和支持。但大多研究都局限于孤立的路口，如何有效协调多个路口提高出行效率，已成为当前应用研究的重点和难点。因此，本文在现有信号灯控制方法的基础上研究多个路口信号灯的协同优化以有效提高大规模路口整体通行能力具有重要的研究价值。主要以深度强化学习为基础，基于 Python 开发平台搭建交通系统环境，设计了基于 DQN 下的多路口信号灯控制决策算法以优化整体路网拥挤程度。

2. 搭建交通系统环境模型

2.1. 环境配置

在算法设计开始之前，我们需要正确配置深度强化学习算法的环境变量，从而让操作系统能够正确地运行模型，使得智能体可以在环境中不断地进行交互学习完成训练。本文主要基于 Python 平台进行开发环境，利用 SUMO 工具配合 Python 完成环境变量的配置。首先我们从 Python 和 SUMO 的官方网站中

获取相应 Windows 操作系统的版本安装包，按照相关操作步骤进行安装。SUMO 选择 Version 1.17.0，Python 选择 Version 3.7.9 在电脑 Windows 10 家庭中文版上配置好相应的环境变量。

2.2. 路网模型搭建及参数设置

本文通过多个交叉路口模型研究深度强化学习算法机理，路口规模为 2*2，呈井字形，如下图 1 所示。每个路口的通行规律由一个信号灯控制，每个信号灯作为一个独立的智能体，4 个智能体交互环境变量，对所有路口信号灯进行协同控制实现多路口联动。共同决策路网整体行车延误和排队长度，将简化问题以聚焦于路口通行规律和交通流的耦合机理以及智能体间决策的协同，研究背景设置在双向六车道的道路工况下，交叉路口结构为十字路口，路口间道路长度设置为统一值。其中，路网各交通参数设置如表 1 所示。

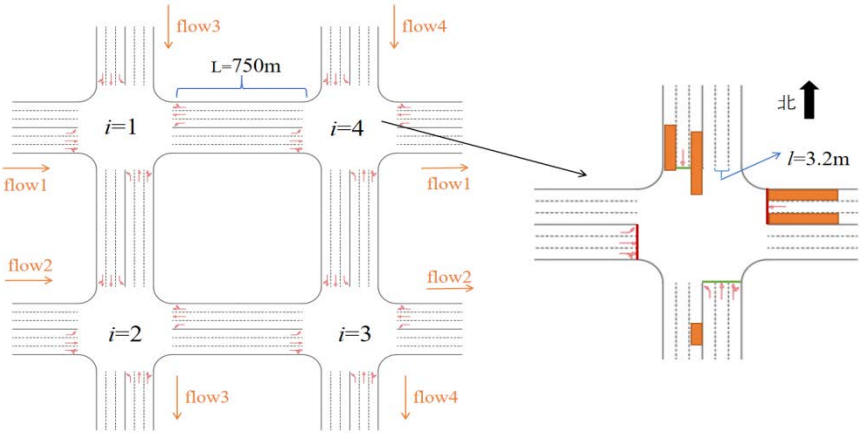

Figure 1. 2*2 scale intersection traffic system structure
图 1. 2*2 规模路口交通系统结构

Table 1. Traffic parameters for the experimental road network configuration
表 1. 实验路网配置的交通参数

参数(单位)	符号/值
交叉口个数(个)	$i = 4$
路段数(个)	$Y = 12$
每条路段的车道数(个)	$y = 3$
平均路段长度(m)	$L = 750$
车道宽度(m)	$l = 3.2$

2.3. 车辆运动模型搭建及参数设置

车辆运动模型使用由 Stefan Krauss 提出的时间上离散、空间上连续的安全车速跟驰模型。假设驾驶员有一秒钟左右的反应时间，并且使用以下 5 个参数 a：车辆最大加速度能力(m/s^2)；b：车辆最大减速度能力(m/s^2)；l：车辆长度(m)；$V_{\max}$：最大车辆行驶速度(km/h)；ξ：驾驶员处理期望的车速时的缺陷 $\xi \in (0,1)$，可以理解为驾驶熟练程度，值越小表示驾驶员越熟练[2]。该模型在计算安全车速时使用如下公式：

$$v_{safe} = v_l(t) + \frac{g(t) - v_l(t)\cdot\Gamma}{(v_l + v_f)/(2\cdot b) + \Gamma} \tag{1}$$

$v_l(t)$：t 时刻前车的速度；

v_f：t 时刻后车的速度；

$g(t)$：t 时刻前车和后车的间距；

Γ：驾驶员的反应时间(一般取 1 秒)。

在 SUMO 仿真软件中，使用如下公式来实现该模型：

$$v_{safe} = -\Gamma \cdot b + \sqrt{(\Gamma \cdot b)^2 + v_l^2 + 2b \cdot g(t)} \tag{2}$$

由于式(2)计算得到的安全车速可能会超过道路允许使用的最大车速，或超过车辆本身能够达到的最大车速，该模型实现时去上述三个值中最小的一个，并称之为期望车速，其表述为

$$v_l(t) = \min\left[v_{\max}, v + a \cdot t_l, v_{safe}\right] \tag{3}$$

其中，t_1 为仿真步长。如果驾驶员的熟练度不足无法使车辆达到预期速度，那么他们的车速应该是预期速度减去车辆加速能力、随机数以及驾驶员缺陷值的乘积，并且要保证车速大于零，因此最终的跟驰模型表述为：

$$v(t) = \max\left[0, v_{des} - \xi \cdot a \cdot rand(\,)\right] \tag{4}$$

计算下一仿真的车辆位置的公式为：

$$P_{t+1} = P(t) + v(t) \cdot t_l \tag{5}$$

车辆跟驰模型各参数设置如表 2 所示，其中车流量大小设置为 1000 辆，车流量规模大小动态可调，车辆直行的概率设为 75%，车辆转向的概率设为 25%。

Table 2. Parameter settings for car-following models
表 2. 车辆跟驰模型各参数设置

参数/单位	符号/值
车辆最大加速度能力(m/s^2)	accel = 1.0
车辆最大减速度能力(m/s^2)	decel = 4.5
车辆最大行驶速度(km/h)	maxSpeed = 25
车辆长度(m)	length = 5
车间最小间距(m)	minGap = 2.5
驾驶员处理期望的车速时的缺陷	sigma = 0.5
总车流量(辆)	n_cars_generated = 1000

2.4. 信号灯模型搭建

交叉路口的通行规律由信号灯相位和信号周期两个参数决定，相位的种类有限，是一种对车流的离散化控制。信号灯的相位选择和信号周期配时是本文所研究决策问题的核心。在本文研究的双向六车道工况下，4 个交叉口共设置 4 个信号灯，每个信号灯代表一个 agent，信号灯相位一共有两种，对应东西通行和南北通行。在这两种信号相位下，位于中间车道的车辆可以直行，位于右侧车道的车辆可以右转，位于左侧车道的车辆可以左转，更加增强了学习环境对现实场景的契合度。通过 output_til 函数定义信号灯相位对应的通行规律，同时定义当信号灯两个相位不同时自动触发黄灯的机制。绿灯相位最小持续时长设置为 15 秒。

3. 深度强化学习算法设计

3.1. 大规模路口信号灯决策下的 DQN 算法伪代码设计

算法设计逻辑如图 2 所示，运行前文搭建的交通系统环境模型，通过订阅 E2 检测器获取初始时刻路网中车辆位置信息和行驶速度等信息，并将其作为状态变量输入至深度神经网络之中，在网络中根据路网拥堵程度最优的目标，估算状态 Q 值，输出动作的执行概率，选择动作是否继续保持当前相位还是切换到下一相位。然后在环境中执行，使得下一时刻环境的状态发生改变，之后环境就会将当前动作执行情况的好坏反馈给目标网络，然后更新 Q 表。改变后的下一时刻状态将又再次输入神经网络中，网络输出动作给环境，环境反馈执行情况，如此往复不断循环迭代是得算法学习出有效的红绿灯控制策略。

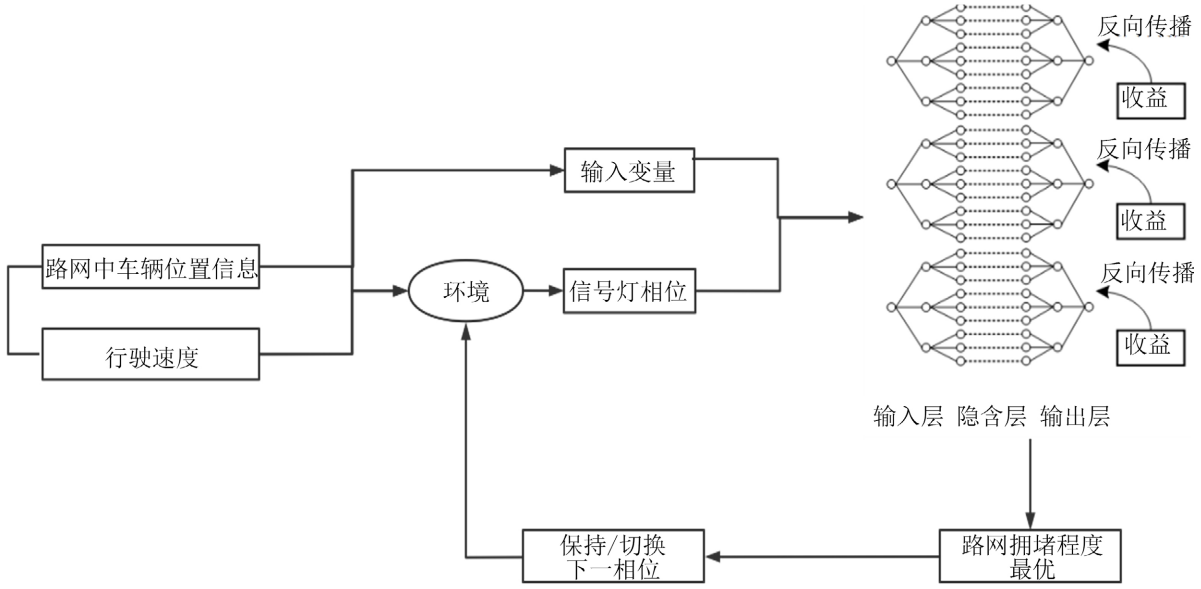

Figure 2. Logical flowchart of algorithm design
图 2. 算法设计逻辑流程图

对于交通灯协调控制来说，总回报最大的目标函数 $Q_\pi(s,a)$ 必然是让整体路网拥堵程度达到最优[3]，长期减少交叉路口停车时间或排队长度等因素。现实问题中，根据马尔可夫决策过程，某一时刻的智能体执行动作不仅会对当下的路网状况产生影响，同时也会对未来路网状况产生影响。但是当前的决策对当前的路网状况的影响一定是最大的。因此我们将路网的目标函数写成所有 reward 的累加，同时添加上衰减系数 $\lambda(\lambda<1)$ 用升幂表示路网当前决策对当前时刻的影响最大，同时对未来的影响不断衰减，因此定义 Q 值为如下累计回报：

$$\begin{aligned} Q_\pi(s,a) &= E\left\{R_t+\gamma R_{t+1}+\gamma^2 R_{t+1}+\cdots \middle| S_t=s,A_t=a,\pi\right\} \\ &= E\left\{\sum_{k=0}^{k} R_{t+k} \middle| S_t=s,A_t=a,\pi\right\} \end{aligned} \tag{6}$$

DQN 算法的前置更新公式如下：

$$Q^*(s,a)=Q(s,a)+\alpha\left[r+\gamma\max_a Q(s',a')-Q(s,a)\right] \tag{7}$$

算法的伪代码设计如表 3 所示：

Table 3. Pseudocode design of DQN algorithm for decision-making of large-scale intersection traffic lights
表 3. 大规模路口信号灯决策下的 DQN 算法伪代码设计

Algorthm 大规模路口信号灯决策下的 DQN
1) 输入：初始化主网络 q^0，目标网络 q^1，设定相关超参数
2) 初始化经验回放 D 至容量 N 的经验池中
3) 随机权重 θ 初始化动作价值函数 Q 的权重 $\theta'=\theta$，设定相关超参数
4) 对于每个回合：
5) 初始化交通系统环境中交通流及每个智能体的状态 S_0；
6) 如果回合未结束：
7) 在 ε 的情况下选择随机动作 a_t 进行决策；
8) 否则：根据环境 S_t 和主网络选择动作 $a_t=\arg\max_a q^0(s_t,a_t,\theta)$ 进行信号灯决策继续保持当前相位还是切换下一相位； 根据得到动作在环境中执行相位；如果相位变化则先执行 3s 黄灯， 周期结束根据交通流状态 s_t，得到奖励 r_t，和新状态 s_{t+1}；
9) 将四元组 (s_t,a_t,r_t,s_{t+1}) 存入到经验池 D 中；
10) 从经验池 D 中随机采样出一小批经验；
11) 计算当前目标估计值 $y_i=\begin{cases}r_t, & \text{迭代至最后一步}\\ r_t+\gamma\arg\max_a q^0(s_t,a_t,\theta'), & \text{否则}\end{cases}$；
12) 更新主网络 q^0，以 $(y_i-Q(s_t,a_t,\theta))^2$ 为目标值做梯度下降；
13) 将新的交通系统环境状态定义为当前状态 $s_{t+1}=s_t$；
14) 在一定步数后，用主网络的权重更新目标网络 $q^1=q^0$；
15) 直到迭代结束，当前交通系统环境 s_t 是终止状态
16) 输出：训练后的主神经网络 q^0

最后，路网拥挤程度路口由路网中车辆平均等待时长、平均排队长度来综合体现。因此算法主要从状态、动作、奖励函数及神经网络的架构四个方面进行定义、设计。

3.2. 状态 State

状态表示智能体所感知到的环境信息，以及因其自身行为改变而带来的变化。它是做出决策和评估累积回报的依据，状态设计的好坏直接决定了深度强化学习算法的收敛性和稳定性。

将状态定义为：进入车道上车辆的位置 $position[l]$ 和行驶速度 $velocity[l]$ 以及信号灯状态 L。Agent 每个 step 的开始时观测卡口状态 $S_t=(P,V,L)$

$$s_{t,i}=\{position[l],velocity[l]\},i\in\{0,1,2,3\},l\in\{0,1,\cdots,n*n-1\} \tag{8}$$

其中，i 路网的交叉口，l 表示车辆驶入所在车道；$position[l]$ 表示路网中交叉口 i 的驶入车道 l 中车辆的位置矩阵，$velocity[l]$ 表示交叉口 i 的驶入车道 l 中车辆的速度矩阵。

3.3. 动作 Action

交叉口的二元动作设置为：智能体是保持在当前相位还是切换到下一相位。由于多交叉口情形应该考虑到动作的连续输出，所以此时定义动作相位切换时无需考虑控制下一相位的时长，动作均以相同的相位时长持续切换：

$$A_{t,i}=\{0,1\} \tag{9}$$

A_t：$A_t=0$，表示继续保持当前相位不变；$A_t=1$，表示切换下一相位。

3.4. 奖励函数 Reward

总目标奖励函数：对于交通信号灯协调控制问题来说，总目标应该是最大化优化路网整体拥挤程度，即奖励函数应表示为使得路网整体拥挤程度越小回报越大(因此，加入路口平均等待时间，平均排队长度指标共同构建 Reward)，如以下公式所示：

$$r_{t,i,l} = -\sum_i queue_{t+\Delta t} + a \cdot wait_{t+\Delta t} \tag{10}$$

式中：$queue_{t+\Delta t}$ 为路口 i 的驶入车道 l 拥堵队列的长度 $wait_{t+\Delta t}$ 为路口 i 的驶入车道 l 拥堵队列的延误时间 a 为权重参数，体现对拥堵队列长度和等待时间赋予的不同偏重。

3.5. 深度神经网络架构设计

利用深度神经网络处理通过环境获取的高维图像数据提取空间特征，即同时提取路网不同位置上的各项所需信息数据，形成图片类型状态的特征数据。观察空间位置信息与数据之间的联系，找出其分布特征规律。本文中首先将路网上所有车辆的位置，行驶速度等图片信息输入深度神经网络中使其转化为路网状态矩阵。即网络输入路口状态 St 后使用两层卷积层滤波器并采用 ReLU 作为激励函数提取主要空间特征，然后通过后两层的全连接层总结神经网络输入变量每个部分的特征以生成一个具有所需类数的分类器。如图 3 中输出层的设计为两个神经元分别对应输出红绿灯{0, 1}动作控制，0 为保持当前相位不变概率，1 为切换到下一相位。通过引入的目标 Q 网络不断更新得出最优动作策略执行，判断当前相位继续保持还是切换到下一相位，最后输出的值表示动作选择概率，决策出最优动作进行预测识别实现红绿灯相位控制以达到整个路网的拥堵程度最优。信号灯的相位输出动作的选择往往会选择最大的概率执行，但网络中引入了经验回放，也会根据一定概率去采样其他动作，提高了算法的可靠性。深度神经网络的结构设计如图 3 所示：

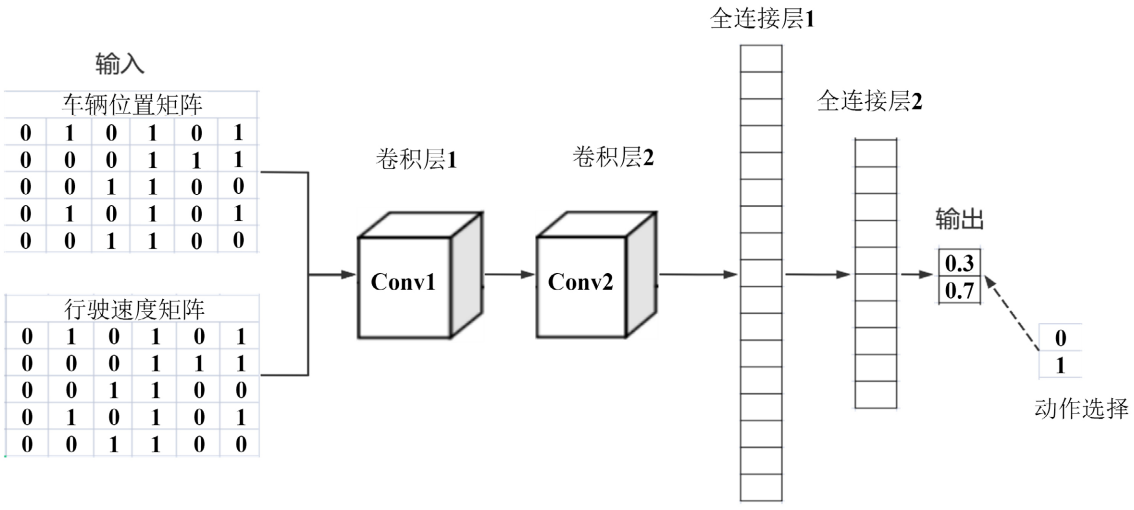

Figure 3. Structural design of deep neural networks
图 3. 深度神经网络的结构设计

4. 算法验证与结果分析

4.1. 评价指标设计

由于本文中的奖励函数由车辆平均排队长度和车辆平均延误时间两个变量共同构建的，所以下文将

使用车辆平均排队长度 $queue_{t+\Delta t}$ 和车辆平均延误时间 $wait_{t+\Delta t}$ 两个指标来对路网通行能力及路网整体优化情况进行评价。

1) 车辆平均排队长度 $queue_{t+\Delta t}$

指通行的车辆在交叉路口因为交通灯信号变化而致使排队车辆无法继续行驶的数量的平均值。

2) 车辆平均延误时间 $wait_{t+\Delta t}$

是指车辆在行驶途中，驾驶员受到其他意外车辆的影响、或是交通控制设旋等无法掌握的因素，而造成运行时间的增加。

4.2. 算法训练结果分析

在深度强化学习算法中，除了状态、动作、奖励等的定义会对直接影响结果外，很多参数都会间接地影响到算法运行的结果和速度，因此需要不断地调整参数以寻找最优结果。在本例中 DQN 算法中各参数的设置如表 4 所示。

Table 4. Parameter settings for the DQN algorithm
表 4. DQN 算法的参数设置

参数	值
隐藏层的节点数量 num_layers	4
抽样批次 batch_size	100
学习率 learning_rate	0.001
总的尝试次数 training_epochs	800
最小存储数量 memory_size_min	800
最大存储数量 memory_size_max	20000
输出的动作数量 num_actions	4
折扣因子 gamma	0.75
迭代轮数 total_episodes	200

1) 整体路网车辆平均排队长度

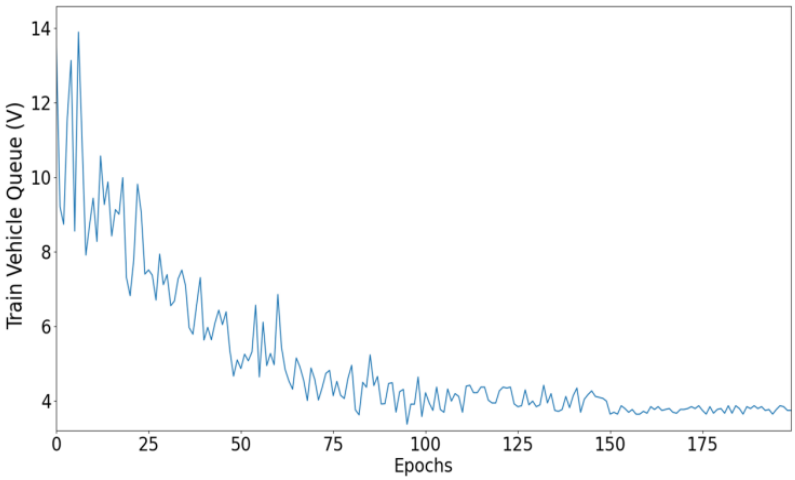

Figure 4. Line chart of the average queuing length of vehicles in the overall road network
图 4. 整体路网车辆平均排队长度折线图

由图 4 可知，车辆排队长度出在 0 到 25 回合的时候，各路口散点图数据离散，折线图的波动变化大，路网的整体排队指标忽高忽低，在 25 到 100 回合的时候，散点图分布效果已逐渐变好，折线图虽然仍在波动，但出现了明显的下降趋势，车辆平均排队长度有效地降低。等到 100 回合以后，我们会明显发现大量的数据点开始集中分布，折线图波动较小，特别是在 150 回合的时候数据点分布和折线几乎趋于水平，此时我们可以得出算法已逐渐趋于收敛。各路口数据分别都稳定在 3 辆车的排队长度，算法学习到公平策略。整体路网车辆平均排队长度稳定在 4 辆车左右，整体路网的通行能力得到了提高。通过车辆排队长度指标的运行结果可以表明，该算法训练出了有效红绿灯控制策略。

由上图分析可知，强化学习任意的动作都会得到不同的反馈结果，在前期时智能体随机选择动作执行，效果变化明显，随着不断训练智能体在不断地试错过程中积累经验，去选择更好的动作训练网络，逐渐训练出最佳执行策略。

2) 整体路网车辆平均延误时间

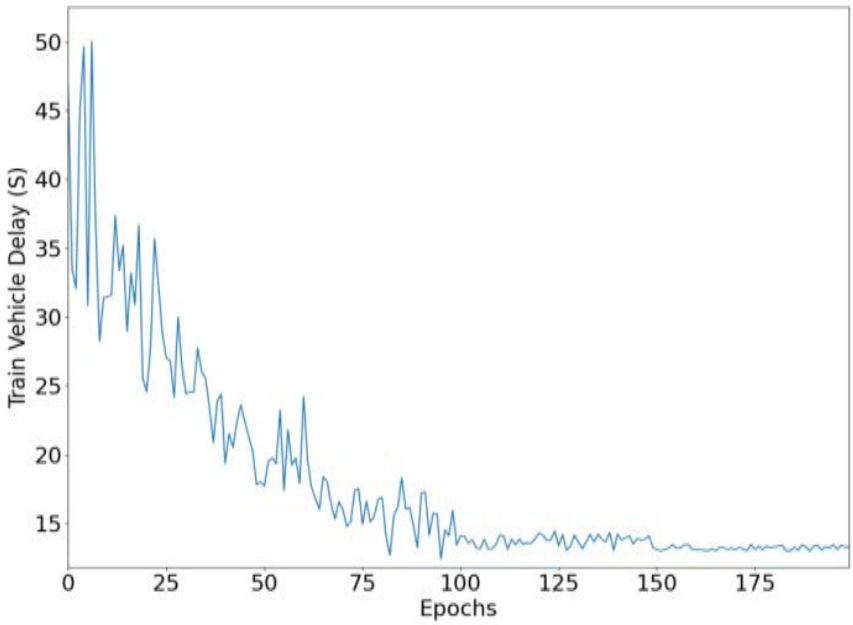

Figure 5. Line chart of average delay time for vehicles in the overall road network
图 5. 整体路网车辆平均延误时间折线图

由图 5 可知，随着不断训练，四个交叉口的车辆平均延误时间都在有效减少下降，且最后各路口数据分别都稳定在 15 秒的延误时间，算法学习到公平策略。由折线图可以看出整体路网的车辆延误时间快速减小。最后通过观察该指标的全部运行结果，表明该算法可以训练出有效红绿灯控制策略，各路口信号灯进行协同控制，实现多路口联动，使整体路网的拥挤程度得到了优化。

利用 DQN 算法实现多路口联动，总目标奖励函数训练曲线如图 6 所示。车辆停留越小，路口拥挤程度越小，回报越大。由上图可以看出算法每个回合的总 Reward 明显变好，不断增大后期逐渐趋于稳定，可以得出本文设计的算法训练出了有效的红绿灯控制策略。

综上所述，曲线的之所以呈现出这样的状态因为强化学习前期的动作选择是随机的，强化学习算法随机选择动作指导智能体运行，随着训练轮数地不断增加，前期积累的试错经验被网络学习，使得这个随机动作的触发概率不断减少，网络自发选出更好的动作去指导智能体运行。所以随机动作越来越少，说明网络有效动作或者优化动作引起的效果越来越好。

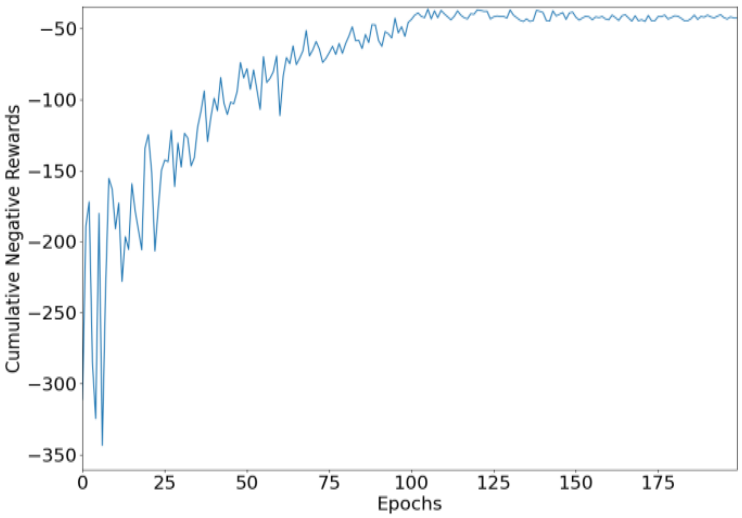

Figure 6. Training curve of DQN algorithm
图 6. DQN 算法训练曲线图

4.3. 不同配时方案的对比结果分析

目前大多数的城市交通系统的信号灯控制方案中仍采用的是固定配时，因此我们接下来通过传统的固定配时方案与本文设计的深度强化学习智能信号控制方案进行对比评价，来比较算法的优势。

方案一采用传统的具有固定配时的信号方案，每个路口均采用标准四相位信号配时，信号时序方案如表 5 所示。方案二：，利用本文设计的大规模路口信号灯决策下的 DQN 算法形成深度强化学习智能信号控制方案。

Table 5. Fixed timing plan for static traffic signals
表 5. 静态交通信号固定配时方案

相位编号	相位名称	信号配时
1	东西直行	25 (绿灯) + 3 (黄灯)
2	东西左转	20 (绿灯) + 3 (黄灯)
3	南北直行	25 (绿灯) + 3 (黄灯)
4	南北左转	20 (绿灯) + 3 (黄灯)

将上述两种方案得到的各个方向的指标数据分别进行对比，整理汇总后的结果如表 6 所示：

Table 6. Comparative analysis of evaluation indicators for two schemes
表 6. 两种方案考察指标对比分析表

车道转向	平均延误(s)			平均排队(m)		
	传统固定	强化学习	提升	传统固定	强化学习	提升
东进口	19.0	14.2	25.31%	26.4	20.8	21.11%
西进口	23.5	13.5	42.78%	9.4	23.6	19.64%
南进口	21.5	15.6	27.48%	42.6	35.6	16.58%
北进口	27.7	16.0	42.31%	25.1	18.5	26.43%
总延误/总队长	22.9	14.8	35.37%	30.9	19.0	38.51%

通过分析该表可知，深度强化学习智能信号控制方案与静态交通信号固定配时方案相比，本文提出的方法在考察指标数据集上都显著优于传统的配时方法。车辆平均总延误、总队长分别提高了 35.37%，38.51%，其他方向数据都有不同程度的优化效果，我们可以得出深度强化学习算法使得整体路网的拥挤程度得到了明显的改善。这一现象证明了与传统控制方案相比，基于多智能体的大规模路口交通信号灯协同控制研究的强化学习算法可以展现出良好性能，在城市交通系统控制优化上具有一定的优越性。

4.4. 不同超参数下的算法训练结果分析

超参数是指在训练前或者训练中可以进行人为调整的参数，如学习率、折扣因子等。不同的超参数在训练中具有不同的作用，它们都会对模型的训练效果有着不同程度影响。因此我们接下来通过调整 DQN 算法的重要超参数对训练结果进行对比分析，研究超参数设置对网络训练的影响。

我们使用本文设计的 DQN 算法，载入训练模型、动作选择及搭建好的深度神经网络之后进行相关超参数设置，不变的超参数设置如前文表 5 所示。将迭代轮数分别设置为 100 轮、200 轮，研究迭代轮数设置对网络训练的影响，Reward 累积回报对比分析图如下所示：

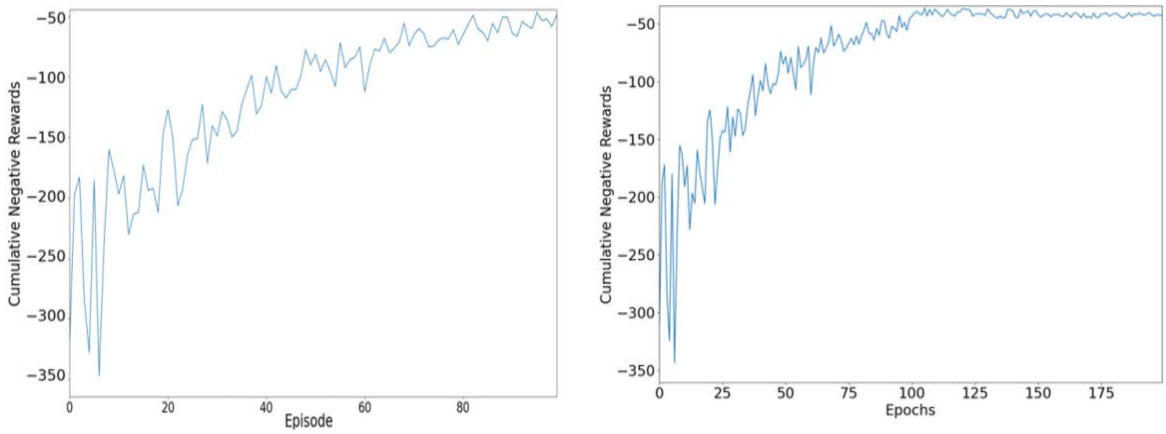

Figure 7. Comparative analysis chart of cumulative Reward under different iteration rounds
图 7. 不同迭代轮数下 Reward 累积回报对比分析图

由图 7 可知，迭代轮数分别为 100 和 200 时，Reward 累积收益总体呈上升趋势。在 0-100 轮中曲线变化趋势近乎相同，整体路网拥挤程度在该过程中得到相同程度优化。之后，迭代轮数设置为 100 的网络会停止训练 reward 不再产生值，可以明显看出结果还没有收敛。而迭代轮数设置为 200 时的网络会继续训练至 200 轮结束，在第 100 轮后 Reward 值仍在继续提高，到第 150 时网络训练开始收敛，到 175 之后并无明显变化开始收敛。

所以迭代轮数的设置会影响网络训练模型的收敛性，但并不影响模型的即时奖励收益。如果迭代轮数设置太少，模型还未收敛就停止训练，会影响智能体学习的效果。而过多的迭代次数只会增加硬件工作量，并不会带来模型性能的改善。因此，我们在网络训练中一定要调整设定合适的迭代轮数，既不会影响模型的收敛，也可以减少硬件的处理工作量。

在 DQN 算法中，载入训练模型、动作选择及搭建好的深度神经网络之后进行相关超参数设置，不变的超参数设置如前文表 5 所示。改变的是学习率 learning_rate 和折扣因子 gamma，研究学习率和折扣因子的设置对网络训练的影响。这里我们设置绿色曲线超参数折扣因子 gamma = 0.5，学习率 learning_rate = 0.0001，红色曲线超参数折扣因子 gamma = 0.75，学习率 learning_rate = 0.001，进行 200 次的模拟训练后，给出如下指标训练结果对比图：

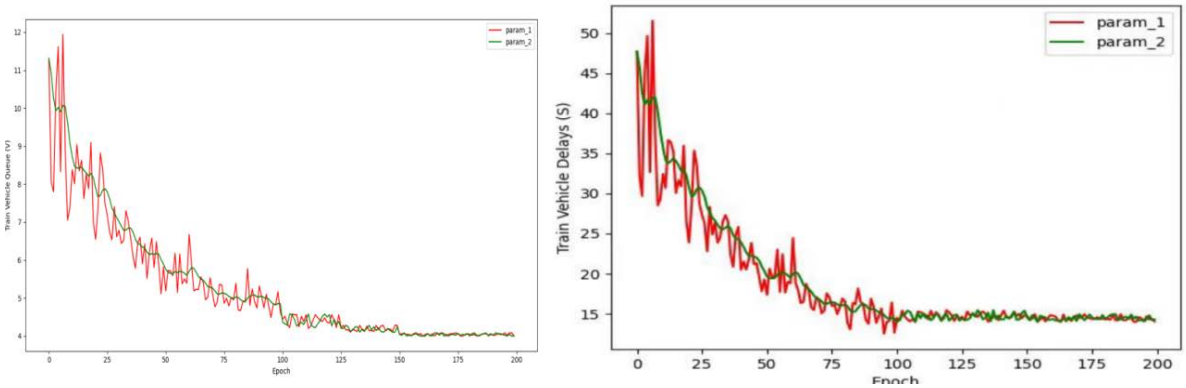

Figure 8. Comparative analysis chart of average queuing length and average delay time of vehicles under different learning rates and discount factors
图 8. 不同学习率和折扣因子下车辆平均排队长度、车辆平均延误时间对比分析图

由图 8 可知，在不同的学习率和不同的折扣因子下整体路网的车辆平均排队长度、车辆平均延误时间都呈下降趋势，说明路网整体拥挤程度都得到了明显的下降。且在迭代的最后二者均趋于稳定，说明算法得到了稳定地收敛。不同的是，红色曲线波动变化剧烈，而绿色曲线几乎一直稳定下降直至收敛，说明在绿色曲线的参数设置下，智能体在训练中选择随机动作的次数越来越少，优化的动作结果越来越好。从曲线收敛速度来看绿色曲线的两个指标均比红色曲线收敛速度快，收敛时间早。

结果说明选择一组合适超参数既能加速模型的训练，又能得到一个较优的精度。在一个训练好的网络训练模型中我们可以通过调整不同的超参数就可以使网络得到优秀的表现性能而无须重新设计网络结构，这将在项目的实际应用中带来一定的工程效益。

5. 总结

本文将深度强化学习算法应用于多智能体的大规模路口交通信号灯协同控制研究，多个智能体交互环境变量实现多路口联动以优化整体路网。首先，基于 Python 开发平台搭建强化学习的环境，然后设计了基于 DQN 下的多路口信号灯控制决策算法，以路口平均排队长度和延误时间作为评价指标验证算法的有效性，最终结果显示，在本文所提出的学习框架下，各交叉口以及整体路网数据的车辆平均延误时间、排队长度都在有效下降，且数据最终都趋于稳定，表明算法实现了多个智能体之间的协同控制，学习到了公平策略，训练出了有效的红绿灯控制策略。最后通过与固定配时方案对比实验显示，强化学习智能信号控制方案中车辆平均总延误、总队长分别提高了 35.37%，38.51%，整体路网拥挤程度明显改善，表明该算法在交通控制中能够展现出良好的优化性能。通过调整算法的超参数对比训练曲线，结果显示一组合适的超参数设置既能加速模型的训练，又能得到一个较优的精度。表明合适的超参数设置将会为项目的实际应用带来一定的工程效益。

参考文献

[1] 汪天祥. 基于多智能体深度强化学习的大规模路口信号灯协同控制研究[D]: [硕士学位论文]. 合肥: 合肥工业大学, 2021.

[2] 王文璇, 阎莹, 吴兵. 智能网联信息下车辆跟驰模型构建及行为影响分析[J]. 同济大学学报(自然科学版), 2022, 50(12): 1734-1742.

[3] Niu, L. and Pan, M. (2022) Research on Coordinated Control Method of Urban Traffic Based on Neural Network. *International Journal of Innovative Computing and Applications*, **13**, 18-26. https://doi.org/10.1504/ijica.2022.121385