

## 基于深度强化学习的智能交通信号控制

蔡铭铭<sup>1</sup>, 何少雄<sup>2</sup>, 刘志滨<sup>3</sup>

<sup>1</sup>河南财经政法大学计算机与信息工程学院 河南省郑州市

<sup>2</sup>东北财经大学管理科学与工程学院 辽宁省大连市

<sup>3</sup>齐齐哈尔大学计算机与控制工程学院 黑龙江齐齐哈尔

**【摘要】**随着人工智能的日益发展和我国城市化战略的实施, 交通信号已经成为一个重要的因素和热点问题。交通问题严重影响城市的可持续发展, 给城市的人们带来时间和安全问题的困扰。交通与人工智能技术相结合, 利用人工智能和交通相关知识处理交通问题成为趋势。基于强化学习解决智能交通信号控制问题, 解决现有工作中的不足。一个是公平性问题, 另一个是多交叉口控制中的通信问题。通过深度强化学习的方式, 使通更加智能化, 工作效率提高。

**【关键词】**人工智能; 深度学习; 交通信号; 智能交通

### Intelligent Traffic signal control based on deep reinforcement learning

Mingming Cai<sup>1</sup>, Shaoxiong He<sup>2</sup>, Zhibin Liu<sup>3</sup>

<sup>1</sup>School of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou, China

<sup>2</sup>School of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian, Liaoning Province, China

<sup>3</sup>College of Computer and Control Engineering, Qiqihar University, Qiqihar, Heilongjiang

**【Abstract】** With the development of artificial intelligence and the implementation of urbanization strategy, traffic signal has become an important factor and hot problem. Traffic problems seriously affect the sustainable development of the city, and bring time and safety problems to the people in the city. The combination of traffic and artificial intelligence technology, using artificial intelligence and traffic-related knowledge to deal with traffic problems has become a trend. To solve the problem of intelligent traffic signal control based on reinforcement learning and solve the shortcomings in the existing work. One is the fairness problem and the other is the communication problem in multi-intersection control. Through the way of deep reinforcement learning, make Tong more intelligent and improve work efficiency

**【Keywords】** Artificial intelligence; Deep learning; Traffic signals; Intelligent transportation

#### 1 公平性和多路口控制中的通信问题

##### 1.1 公平性问题

公平性问题是控制策略为了最大化总体通行效率, 可能会忽视小车流的车辆队伍, 优先放行车流量较大的队伍, 从而造成小车流车队中的车辆长时间的等待(即“饥饿”情况), 如图 1 所示。其中, W-E 方向的几辆车可能很长时间无法通过路口, 因为 N-S 方向有更大的车流量, 优先放行该方向(设置信号为相应的相位)可以获得更大的通行

效率提升。针对这一问题, 本课题拟采用无线网络领域中的 proportional fair scheduling (PFS) 方法来解决这一问题, 因为 PFS 可在效率和公平性之间实现很好的折中, 并且可以自适应地调整调度的优先级。同时结合深度强化学习, 可以实现具有公平感知的智能交通信号控制。

##### 1.2 多路口控制中的通信问题

通信问题是指在多路口信号控制中, 智能体之间通过通信来进行信息交互, 但是已有的工作多是

直接交换智能体观测到的局部环境信息, 这种做法需要很高的通信代价, 而且直接观测到的局部环境信息往往是复杂的, 不利于接收者的信息提取和学习。本课题拟采用信息独立的数据交换模式, 为不同的智能体传输不同的信息, 因为对于某个智能体来说, 其上游路口和下游路口智能体对其观测到的局部环境信息的专注点是不一样的, 笼统的将全部观测数据传输给所有相邻路口的智能体不仅增加了通信的代价而且加大了接收方提取有用信息的难度, 也增加了学习的复杂性。

## 2 研究方法

本文采用使用一个名为 SUMO (Simulation of Urban MObility) 的仿真平台进行模拟实验。利用该模拟器, 可以方便地获得实时车辆状态, 并且可以改变交通信号来控制交通。对于公平性问题的研究, 使用以下指标来衡量习得的控制策略的效率和公平性。

- (1) 通行时间: 车辆进入路口和离开路口的时间差, 车辆在路口消耗的时间。
- (2) 延迟: 车辆穿过路口的实际时间与顺畅通

行情况下的预期时间的差值。

(3) 驾驶体验得分: 一种新的测量方法, 称为驾驶经验评分(DES)来量化驾驶员的满意度。

不同于已有方法只使用相应指标的平均值来评价模型的好坏, 还考虑的所有车辆对于相应评价指标的分布情况, 可以更加全面地衡量模型的好坏<sup>[1]</sup>。对于通信问题的研究, 通过设置不同的筛选粒度对智能体观测的原始数据进行筛选, 并对比不同粒度下习得的控制策略的性能、模型的收敛性等指标, 来确定能够保障传输必要信息的最小粒度(最小的通信代价), 最终和已有的基于完整数据传输的方法进行对比。

通过将 proportional fair scheduling 和深度强化学习进行结合, 有效地缓解了已有工作中可能存在的“饥饿”情况, 未来还需拓展至多路口信号控制。

## 3 基于强化学习的交通信号控制方法

不同的人工智能技术来控制交通信号, 例如遗传算法、群体智能以及强化学习。其中在这些技术中, 强化学习在近年来更具趋势。

### 3.1 基于强化学习的交通信号控制框架

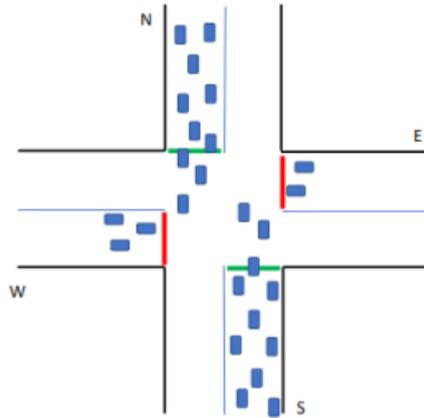


图1 “饥饿”情况图示

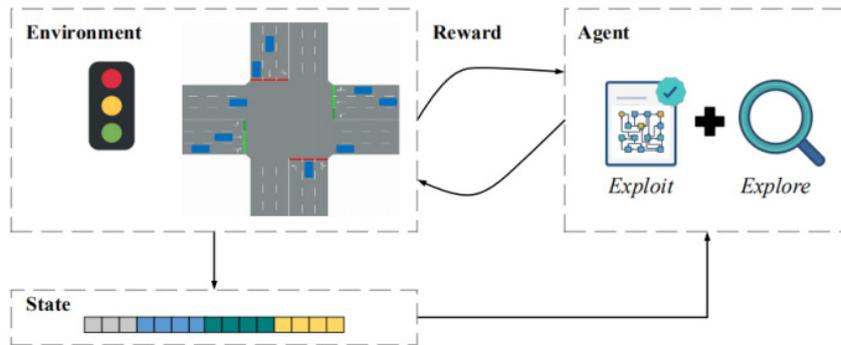


图2 基于强化学习的交通信号控制框架

使用强化学习解决单一路口的信号控制的模型框架,如图 2。其中路口的交通状况以及信号灯。在每个时间步骤  $t$  中,智能体将生成一个对于环境的描述(例如信号相位、汽车等待时间、汽车排队长度和汽车位置)。为了最大化预期奖励,智能体将会预测该状态下的最优动作。然后这个动作会被作用到信号灯上(例如,继续保持当前相位或者变化到其他相位)并返回一个奖励,这种奖励通常是根据十字路口的交通状况来确定的。在训练过程中,代理采用一种结合了对所学策略的利用和对新策略的探索的策略。

### 3.2 基本要素

使用强化学习来解决交通信号控制问题要先确定以下几个基本要素:

(1) 奖励设计: 由于强化学习是以最大化累计奖励为目标来学习的,所以奖励的选择决定了学习的方向<sup>[2]</sup>。在交通信号控制问题中,虽然最终目标是尽量减少所有车辆的通行时间,但由于几个原因,旅行时间很难直接作为 RL 的有效奖励。首先,车辆的行驶时间不仅受红绿灯的影响,还受车辆自由流动速度等其他因素的影响<sup>[3]</sup>。第二,当交通信号控制器事先不知道车辆的目的地时,优化道路上所有车辆的行驶时间变得尤为困难。在这种情况下,只有在车辆完成几个动作后完全离开交叉口后,才能测量车辆的通过时间。已有工作的奖励设计通常是基于一些可以直接在一个动作后测量的指标的加权和。例如,等待车辆的队列长度、车辆等待时间、速度、累计延迟、路口的吞吐量、车辆平均停车次数、信号变化频率以及路口的压力(Max-pressure 中定义的 pressure)等<sup>[4]</sup>。

(2) 状态表示: 以一种数值化的形式来描述路口的交通状况,描述的越全面越有利于学习到最优策略,通常使用多个要素组合来描述交通状况。

(3) 动作选择机制: 动作选择机制决定了以何种方式来控制信号灯,不同的动作机制有不同的影响。主要可以总结为以下四种方式:

①确定当前相位时长: 在这中动作选择机制下,智能体学习通过从预定义的候选时间段(比如,10秒、15秒、20秒等)中选择来设置当前相位的持续时间。

②确定基于周期的相位比: 这种方式定义的动作

作为下一个周期的相位分裂比(phase split ratio)通常,给出总周期长度,并预先定义一个包含一些相位比的候选集。

③保持或改变当前相位: 这种方式也是基于周期性的信号计划,通常一个二进制数来定义动作。例如,1表示保持当前相位,0表示变换到下一相位。

④选择下一相位: 这种方式直接从待选相位序列中选择一个相位并变化到该相位,其中相位序列不是预定的。因此,这种信号控制方式更加的灵活,智能学习在不同状态下选择最优的相位,而不假设信号会以循环的方式改变。

(4) 学习算法: 强化学习发展至今已经提出了很多不同的算法,根据估计潜在奖励和选择动作的不同可以分为以下两种:

①基于价值的方法(Value-based methods): 基于价值的方法近似于一个状态值函数或一个状态动作值函数,策略隐式地从学习到的价值函数中获得。基于值的方法(Q-learning、DQN等)直接模拟状态或状态动作值。状态和奖励可以直接输入到模型中,而无需进行额外的处理。然而,这些方法通常与  $\epsilon$ -greedy 行动选择方法结合在一起,因此,当修饰最终衰减到一个小数目时,这将导致一个几乎确定的策略。

②基于策略的方法(Policy-based methods): 基于直接更新策略的方法,以最大限度地实现预定目标)。基于策略的方法试图学习一个状态中不同动作的概率分布,基于策略的方法的优点是它不要操作是离散的。它还可以学习随机政策,并继续探索潜在的更有价值的行动。它包括基于价值的思想来学习操作概率分布的策略,控制代理行为的参与者(基于策略),以及衡量操作执行情况的批评者(基于价值)。Aslani, Mousavi, Prashanth 和 Bhatnagar 在他们的工作中使用了演员-批评家,利用价值函数逼近和策略优化的优势,在交通信号控制问题上表现出出色的性能。

(5) 协同策略: 协同策略是针对多路口场景的信号控制。在城市环境中,信号灯之间往往距离很近,离开路口的车辆可能会影响其下游路口的车辆到达模式,因此在优化一个路口的信号控制的同时,还要优化其相邻路口的信号控制<sup>[5]</sup>。已有工作实现多路口协同控制的方式可以分为以下几种:

①全局控制：一个简单的方案就是使用一个智能体来控制所有路口的信号灯。它将所有路口的交通状况作为状态输入，输出所有信号灯的联合动作。这种方法虽然可以实现全局最优，但是会导致“维度灾难”，因为状态和动作空间随着问题规模的增加（路口变多）呈指数增长。

②联合建模：这类方法对于每一个路口使用一个智能体来控制信号灯，与此同时根据各个路口之间的协同关系建立一张协同图来进行协同学习。与全局控制不同的是，它将全局 Q-function 分解为子问题的线性组合

$$\hat{Q}(o_1, \dots, o_N, \mathbf{a}) = \sum_{i,j} Q_{i,j}(o_i, o_j, a_i, a_j)$$

其中  $i, j$  表示相邻的智能体。

③独立控制：这类方法同样是对每个路口分配一个智能体，但是每个智能体之间没有显示的协作关系。每个智能体感知环境中的一部分，并以此来学习适应它。这类方法根据是否进行通信可以细分为两种。一种是智能体之间没有通信，这种方式不使用显示的通信来解决冲突问题，每个智能体只能观测到自己路口的局部信息。另一种是智能体之间通过显示的通信来传输数据及信息。

#### 4 结束语

虽然目前已经有不少关于交通信号控制的研究，但是基于强化学习的智能交通控制仍然有很多问题需要深入研究，例如公平性问题和多路口协同控制中的通信问题。一个好的控制策略应该在提高通行效率的同时能够保证每辆车所需的通行时间大致相同，也就是说，车辆通行时间的方差应该越小越好。另外在多路口的信号控制中，使用通信来交换数据可以实现协同控制。通过人工智能得结合，

基于深度强化学习的智能交通信号控制，让交通信号的控制更加灵活高效。

#### 参考文献

- [1] 范文礼. 基于复杂网络理论的连锁故障建模与预防研究[D].西南交通大学,2014.
- [2] 杨迪. 基于强化学习的虚拟机自适应迁移策略研究[D].重庆邮电大学,2020..
- [3] 张森. 城市应急物流救援物资运输优化方法研究[D].沈阳大学,2015.
- [4] 黄浩,胡智群,王鲁哈,路兆铭,温向明.基于 Sumtree DDP G 的智能交通信号控制算法[J].北京邮电大学学报,2021, 44(01):97-103.
- [5] 王娜,杨俊,李艳萍.基于自组织竞争神经网络的智能交通信号灯控制[J].电脑开发与应用,2008(10):63-65.
- [6] 李浩杰,裘国永.基于自动化渗透测试的分析[J].电子设计工程,2015,23(22):25-28.D

收稿日期：2022 年 8 月 18 日

出刊日期：2022 年 9 月 6 日

引用本文：蔡铭铭, 何少雄, 刘志滨, 基于深度强化学习的智能交通信号控制[J]. 国际计算机科学进展, 2022, 2(2): 86-89.

DOI: 10.12208/j. aics.20220014

检索信息：RCCSE 权威核心学术期刊数据库、中国知网 (CNKI Scholar)、万方数据 (WANFANG DATA)、Google Scholar 等数据库收录期刊

版权声明：©2022 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。<http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS