

基于 GCN-LSTM 的多交叉口信号灯控制^①

徐东伟^{②*} 朱宏俊^{③**} 郭海锋^{*} 周晓刚^{***} 汤立新^{****}

(* 浙江工业大学网络空间安全研究院 杭州 310023)

(** 浙江工业大学信息工程学院 杭州 310023)

(*** 宁波宁工交通工程设计咨询有限公司 宁波 315010)

(**** 浙江沪杭甬高速公路股份有限公司 杭州 311500)

摘要 强化学习(reinforcement learning, RL)由于其解决高度动态环境中复杂决策问题的能力,成为信号灯控制中一种具有前景的解决方案。大多数基于强化学习的方法独立生成智能体的动作,它们可能导致交叉口的动作冲突、道路资源浪费。因此,本文提出了基于图卷积网络和长短期记忆(graph convolution network-long short-term memory, GCN-LSTM)的多交叉口信号灯控制方法。首先,基于二进制权重网络对多交叉口进行构图。其次,通过图卷积网络聚合周围交叉口的空间状态信息,利用长短期记忆(long short-term memory, LSTM)获得交叉口的历史状态信息。最后,通过基于竞争网络框架的 Q 值网络进行动作的选择,实现对交叉口相位控制。实验结果表明,与其他强化学习方法相比,本文方法在多交叉口的信号灯控制中能够减少交叉口的队列长度,并使道路网络中的车辆获得更少的等待时间。

关键词 智能交通系统;交通信号灯控制;多智能体强化学习;长短期记忆;图卷积网络

交通信号灯控制(traffic light control, TSC)是一个关键且具有挑战的现实性问题,它可以在城市道路资源有限的情况下最大限度地提高交通效率,并避免交叉口的冲突。找到合适的交通信号控制方法可以缓解交通拥堵,并带来显著的经济、社会和环境效益。

传统的交通信号控制方法有定时控制、驱动控制和自适应控制方法。定时信号控制是一种重复模式,无论实时交通如何变化,它都会持续其周期。驱动控制方法根据环路检测器的实时数据控制交通信号。尽管具有交通响应性,但驱动控制方法并不能完全满足具有波动的交通量,特别是在高度饱和的交通量中,优化效果较差。自适应信号是一种更有效的解决方案,因为它具有适应交通变化的能力。

与传统的控制方法不同,基于强化学习(reinforcement learning, RL)^[1-3]的方法通常采用与外部环境持续交互训练获得的模型来预测最佳交通信号切换时间或者相位,并取得了显著成果。强化学习算法可以分为基于值、基于策略和基于演员-评论员(actor-critic, AC)的方法。基于值的方法^[4]很容易在 TSC 系统中实现,例如 Q-Learning 和 SARSA(state-action-reward-state-action)等,但其收敛性在很大程度上依赖于平稳的马尔科夫决策过程(Markov decision process, MDP)。基于策略的方法由于在 TSC 系统中计算奖励需要在一轮结束之后,所以很难正确定义策略评估。基于演员-评论员^[5-6]的方法结合了两种方法的优点,减少了基于策略方法的偏差和方差。

① 国家自然科学基金(62373325, 6190334, 52072343)和浙江省自然科学基金(LY21F030016, LY20E080023, LQ16E080011)资助项目。

② 男,1985年生,博士,副教授;研究方向:智能交通,大数据挖掘;E-mail: dongweixu@zjut.edu.cn。

③ 通信作者,E-mail: 2366374226@qq.com。

(收稿日期:2023-09-14)

对于大规模 TSC 系统来说,使用单个智能体来优化一组交叉口^[7]时,由于状态空间和动作空间非常大,容易发生维度诅咒问题。因此,为每个交叉口分配一个智能体的多智能体强化学习方法^[8-9]成为研究的热点。例如,基于多智能体 A2C (multi-agent advantage actor-critic, MA2C) 的方法通过将相邻智能体的最新策略与智能体当前状态一同输入深度神经网络 (deep neural networks, DNN), 实现了有限通信, 提高了算法的稳定性与收敛性; 基于协作的深度 Q 学习 (collaborative deep Q-learning, Co-DQN) 方法采用置信上限法则进行智能体的动作探索行为, 在学习过程中, 应用基于平均场理论保证模型收敛和降低复杂度, 共享状态信息来提高训练过程的稳定性; 基于信息交换深度 Q 网络 (information exchange deep Q-network, IEDQN)^[10] 的方法使用上一时刻的观测信息代替当前时刻的观测信息, 提高了智能体之间通信的鲁棒性; 基于考虑博弈的多智能体强化学习 (multi-agent reinforcement learning based on the game, G-MARL)^[11] 方法引入博弈论中混合策略纳什均衡的概念, 实现了快速响应路网交通需求不均衡和波动的现象, G-MARL 在单位行程时间和单位车均延误方面比独立动作的多智能体强化学习 (independent action multi-agent reinforcement learning, IA-MARL) 方法分别改善 59.94% 和 81.45%; 基于双评论家的多智能体深度确定性策略梯度 (multi-agent deep deterministic policy gradient method based on double critics, MADDPG-DC)^[12] 方法, 通过在双评论家网络上的最小值操作来避免价值高估, 促进了智能体学习最优的策略。在合成的时长为 1 h 的道路交通网络的训练结果中, MADDPG-DC 方法的平均队列长度小于其他基线方法, 在模拟时间为 2 700 s 时, MADDPG-DC 方法下的平均队列长度达到峰值, 约为 0.63 辆。对于其他基线方法, MADDPG-DC 方法在 2 980 s 时达到约为 1.41 辆的峰值, MADDPG-DC 方法在 2 980 s 时的峰值在 0.92 辆以上。图网络模型是强大的捕获系统中流量相关性的工具, 已被用于聚集相邻交叉口的状态, 实现了交叉口之间信息共享。例如, 基于交通信号协作的 Colight 方法使用图注意力网络来改进智能体之间的通

信, 实现了高效的控制性能; 基于图注意网络的区域感知协作策略方法^[13] 通过动态获取和集成其他相邻交叉口的状态, 并改进状态、策略和奖励, 减少了车辆的总行驶时间; 基于归纳图强化学习 (inductive graph reinforcement learning, IG-RL)^[14] 的方法将交叉口中的对象建模为拓扑结构图中的节点, 对象之间的物理连接关系建模为拓扑结构图中的边, 采用图卷积网络 (graph convolution network, GCN) 在相同类别的节点间、同交叉口中的对象间共享参数, 实现了更好的泛化能力、更高的可迁移性。基于封建多智能体深度强化学习^[15] 方法将交通网络分为区块, 实现了在保证可扩展性的同时能够全局协调, 并使旅行延迟相对于 MA2C 方法减少 17%。

既有的交通控制方法在对路网进行构图时, 将交叉口视为一个整体作为图网络的节点, 而实际上交叉口内部各路段之间的状态也会对信号灯行为产生影响, 而以往的研究由于更加细致的构图导致更加复杂的状态输入, 所以没有考虑路段间的影响。同时, 过往时刻的路网状态也对当前的决策有一定影响。本文通过二值权重交叉网络对多交叉口场景进行建模, 并使用图卷积网络聚合周围道路的隐藏状态, 利用长短期记忆 (long short-term memory, LSTM) 网络获取当前交叉口的历史隐藏状态, 最终通过 Q 值网络实现动作的选择。

1 多交叉口信号灯控制问题的描述

本节将交通信号控制问题建模为一个马尔可夫决策过程。路网中的信号灯都由智能体控制, 每个智能体都只能观察到部分系统状态。智能体通过观察到的部分系统状态决定当前路口的相位, 以减少路网中车辆的平均等待时间。具体来说, 这一过程由 $(S, O, A, P, r, \pi, \gamma)$ 组成:

S 为环境状态集, $s \in S$ 表示路网整体所处状态, 即环境的一种内部状态。对于智能体来说, 环境状态是不完全可观测的。

O 为观测空间集, 在红绿灯控制系统中, 智能体无法获取环境状态, 只能对本身所控制的交叉口进行观测。 $o_i \in O$ 表示路网中第 i 个交叉口所观测到

的状态,对于有 K 条入口道路的交叉口,其观察状态可以表示为 $\mathbf{o}_i = [o_{i_1}, \dots, o_{i_k}]$ 。其中, o_{i_k} 表示第 i 个交叉口的第 k ($k \in 1, 2, \dots, K$) 条道路上的状态。图 1 表示一个交叉口,包含 4 条入口道路,每条入口道路包含 2 条车道。 o_{i_k} 可以用许多变量来表示,如车道占用率、平均速度、停车占用率和车道通行状态等。

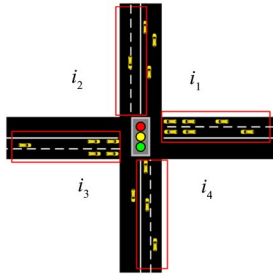


图 1 道路交叉口示意图

每个入口道路包含 B 条车道,则每个进道占用率可以表述为 B 条车道占用率的平均,车道占用率即车道长度中车辆所占的比值,对于第 i 个交叉口的第 k 条入口道路的车道占用率 L_{i_k} 表示如下。

$$L_{i_k} = \frac{1}{B} \left(\frac{\sum_{l=1}^B \sum_{v=1}^{\zeta} \text{len}(v)}{\sum_{l=1}^B \text{len}(l)} \right) \quad (1)$$

式中: $\text{len}(\cdot)$ 为车辆或车道的长度, ζ 为车道 l 上的车辆集, v 为车辆集中的车辆。

交通控制的目标是尽量减少道路上的拥堵,即减少车辆在道路上的停顿,考虑车道被停车占用的程度,定义停车占用率 H_{i_k} 为

$$H_{i_k} = \frac{1}{B} \left(\frac{\sum_{l=1}^B \sum_{a=1}^{\eta} \text{len}(a)}{\sum_{l=1}^B \text{len}(l)} \right) \quad (2)$$

式中: η 表示速度接近于 0 的车辆集, a 表示车辆集中相应的车辆。

与其他特征相似,将平均速度泛化到 $[0, 1]$,即车道上所有车辆速度的平均值与最大限速速度的比值 MS_{i_k} , 即:

$$MS_{i_k} = \frac{1}{B} \left(\frac{\sum_{l=1}^B \sum_{v=1}^{\zeta} \min(\text{spe}(v), \text{maxspe}(l))}{\sum_{l=1}^B \sum_{v=1}^{\zeta} \text{maxspe}(l)} \right) \quad (3)$$

式中: $\text{maxspe}(l)$ 为道路上允许行驶的最大速度,

$\text{spe}(v)$ 为车辆的行驶速度。

因此,每一个道路的观测状态为 $o_{i_k} = \{L_{i_k}, H_{i_k}, MS_{i_k}, P_{i_k}\}$, 交叉口的状态则为 $\mathbf{o}_i = \{o_{i_1}, o_{i_2}, \dots, o_{i_k}\}$

A 为动作集, $a \in A$ 表示智能体能够采取的行动,在智能控制系统中,该动作为交通信号灯可以控制的相位。在实验中将每个动作相位的绿灯执行时间设置为 30 s,黄灯时间为 5 s,用于让剩余车辆离开交叉口。

P 为条件转移概率, $P(s_{t+1} | s_t, a_t)$ 描述了当前环境从状态 s_t 执行动作 a_t 后转换到下一个潜在状态 s_{t+1} 的概率。

R 为奖励集,即智能体采取行动后从环境中得到的反馈,用于评估所执行动作的好坏,通常与强化学习需要优化的目标有关。在红绿灯控制系统中,可以队列长度、等待时间等作为奖励函数。在本文中使用的是将等待时间作为奖励 r , 即

$$r = \sum_{i=1}^{\Lambda_t} w_{i,t} - \sum_{i=1}^{\Lambda_{t+1}} w_{i,t+1} \quad (4)$$

式中: Λ_t 表示到第 t 个时间步为止相应的车辆总数, $w_{i,t}$ 表示车辆 i 在时间步 t 的等待时间,式中描述的奖励为 2 个相邻时间步之间等待时间的变化。

γ 为折扣因子, $\gamma \in [0, 1]$ 表示智能体相对于当前奖励对未来奖励的折扣程度。

2 基于图卷积网络的合作学习

在多交叉口的红绿灯控制过程中,智能体的决策与周围交叉口的状态和自身历史的状态信息密切相关,因此,使用二值权重交叉网络对道路网络进行构图,利用图卷积网络取得周围交叉口的空间信息,同时使用 LSTM 网络获得自身的时间信息。过程如图 2 所示。

2.1 基于二值权重交叉网络的路网构建

在 SCATS (sydney coordinated adaptive traffic system) 系统中,只有入口道路有道路交通状态数据(流量和占用率)。因此,基于二值权重交叉网络构建城市道路的复杂网络,可以充分利用 SCATS 系统环路检测器收集的数据,获得更多有用的交叉口信息。带有交叉口的城市道路网络模型(urban road

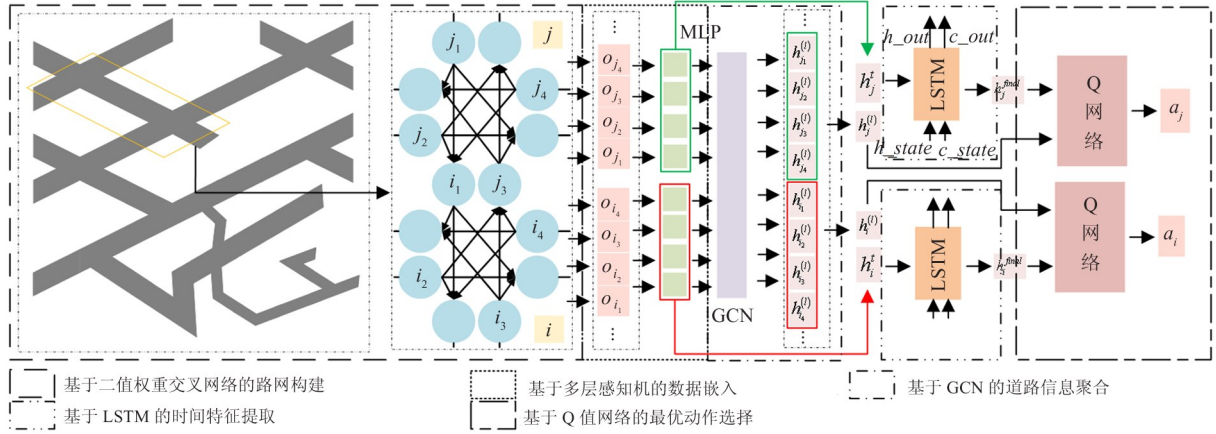


图2 算法框架图

network, URN) 可以描述为 $G = (V, E)$, V 为路网图中的节点集, E 为路网图中的边, 其中 $V = \{v_1, v_2, \dots, v_n\}$ 表示所有十字路口控制的道路, $E = \{e_{ij} \parallel i, j \in V\}$ 表示道路之间的连接关系, 如果 2 条道路之间能够通过交叉口到达, 则 2 条道路之间存在连接关系。每条入口道路上的交通状态数据表示为 o_{i_k} 。

2.2 道路信息的聚合

在收集到的每条入口道路的状态信息后, 即车道占用率、停车占用率、平均速度与通行状态, 通过多层感知机 (multi-layer perceptron, MLP) 将得到的数据嵌入到潜在空间中:

$$h_{i_k}^{(0)} = Embed(o_{i_k}) = \sigma(o_{i_k} \mathbf{W}_e + \mathbf{b}_e) \quad (5)$$

式中: o_{i_k} 是第 i 个交叉口的第 k 条入口道路的观察状态, \mathbf{W}_e 是权重矩阵, \mathbf{b}_e 为偏置向量, $Embed$ 表示对观察状态进行嵌入计算, σ 是 ReLU 函数。生成的隐藏状态 $h_{i_k}^{(0)}$ 表示第 i 个交叉口的第 k 个入口道路的当前交通状态。

利用图卷积网络有效地将每条入口道路的状态信息处理成嵌入向量, 以对入口道路空间结构产生的状态之间的相关性建模, 聚合相邻入口道路的特征, 聚合过程可以用式(6)表示。

$$h_{i_k}^{(l)} = \sum_{j \in \mathcal{N}(i_k) \cup \{i_k\}} \frac{1}{\sqrt{deg(i_k)} \sqrt{deg(j)}} \cdot (\Theta \cdot h_j^{(l-1)}) \quad (6)$$

式中: $h_j^{(l-1)}$ 是第 $l-1$ 层邻居入口道路的特征, $j \in \mathcal{N}(i_k) \cup \{i_k\}$ 表示 i_k 的邻居节点, $\sqrt{deg(i_k)}$ 、 $\sqrt{deg(j)}$ 为节点的度, Θ 为权重矩阵, $h_{i_k}^{(l)}$ 是第 l 层

聚合邻居入口道路后第 i 个交叉口的第 k 条入口道路交通状态。由此, $Concat$ 表示拼接其他交叉口的入口道路的第 i 个交叉口的状态, h_i^{GCN} 表示聚合后的节点特征。

$$h_i^{GCN} = Concat(h_{i_1}^{(l)}, h_{i_2}^{(l)}, \dots, h_{i_K}^{(l)}) \quad (7)$$

目前的交通灯控制方法多数只考虑了智能体之间的空间关系, 缺少一种合作方法捕获时间信息。与空间信息不同, 时间信息捕捉交叉口历史状态对当前策略的影响。由此, 将交叉口控制的道路信息通过编码器进行编码, 然后与获得的前一时刻的隐藏状态输入 LSTM 获得当前时刻的隐藏状态:

$$h_i = Concat(h_{i_1}^{(0)}, h_{i_2}^{(0)}, \dots, h_{i_K}^{(0)}) \quad (8)$$

$$h_i^t, c_{out}, h_{out} = LSTM(h_i, c_{state}, h_{state}) \quad (9)$$

式中: $LSTM$ 表示对状态信息进行长短期记忆网络提取时间特征, c_{state} 与 h_{state} 为 $LSTM$ 的细胞状态与隐藏状态, 即前一时刻的 c_{out} 与 h_{out} , 在初始时间步, 都将被赋予一个初始值。 h_i 为交叉口 i 的原始交通状态, h_i^t 为包含历史信息的时间步 t 的交通状态。

交叉口 i 的交通状态最终可以表示为

$$h_i^{final} = Concat(h_i^t, h_i^{GCN}) \quad (10)$$

2.3 最优动作的选择

从上面的步骤中通过图卷积获得了包含周围交叉口状态的空间信息, 以及通过 LSTM 获得的包含自身历史状态的时间信息, 将会被输入 Q 值网络用于动作的选择。

不同的状态动作对的值函数是不同的, 但是在某些状态下, 值函数的大小与动作无关, 例如交叉口无车时。于是引入竞争网络来解决上述问题, 该网

络通过状态价值 V 与动作价值之和来计算 Q 值。

$$Q(s, a; \theta) = V(s; \theta) + \left(A(s, a; \theta) - \frac{1}{|A|} \left(\sum_{a'} A(s, a'; \theta) \right) \right) \quad (11)$$

式中: $A(s, a; \theta)$ 为优势函数, 表示当前状态 s 下, 执行动作 a 能够获得的有益程度, 并且当前状态 s 下的优势值之和等于 0, θ 为网络参数。 V 表示状态价值函数, 即在状态 s 下能够获得的未来累计奖励, 也是当前状态 s 下 Q 值的期望。

为了解决训练过程中参数更新时引起的策略波动, 引入目标网络。参数为 θ 的主 Q 网络在状态 s 下采取动作 a 取得的 Q 值记作 $Q(s, a; \theta)$, 相应的参数为 θ^- 的目标网络记为 $Q_{\text{tar}}(s, a; \theta^-)$, 采用均方误差更新网络

$$L = E \left[\left(r_i + \gamma \max_{a'} Q_{\text{tar}}(s', a'; \theta^-) - Q(s, a; \theta) \right)^2 \right] \quad (12)$$

式中: L 为损失函数, E 为奖励的期望。

更新网络的过程中目标网络使用了贪婪算法, 导致了过度估计的问题。于是通过解耦目标 Q 值的选择和计算这两步, 来消除这一问题。在这里目标 Q 值的获取不再是直接在目标网络里面找各个动作中的最大 Q 值, 而是先在主 Q 网络中找出最大

Q 值对应的动作, 然后利用这个选择出来的动作在目标网络里计算目标 Q 值。最终, 网络的更新可以表示为

$$L = E \left[\left(r_i + \gamma Q_{\text{tar}}(s', \arg \max_{a'} Q(s', a', \theta), \theta^-) - Q(s, a; \theta) \right)^2 \right] \quad (13)$$

3 实验结果

3.1 对比方法

本节在由德国宇航中心开发的微观、连续道路交通仿真平台 (simulation of urban mobility, SUMO) 上进行实验。实验场景为包含 16 个交叉口的 4×4 网格场景, 包含 7 个交叉口的 Cologne 场景与包含 21 个交叉口的 Ingolstadt 场景, 如图 3 所示。其中, 4×4 网格场景为自定义的简单场景, 路网中的出行需求为随机生成, 信号灯之间的联系相对简单, 能够大致反映算法的有效性。Cologne 场景描述了科隆市内一整天的交通情况, 原始需求数据源自 TAPAS。截取其中的一个区域作为实验场景, 该场景结构相对简单, 数据为真实的出行数据。Ingolstadt 场景为

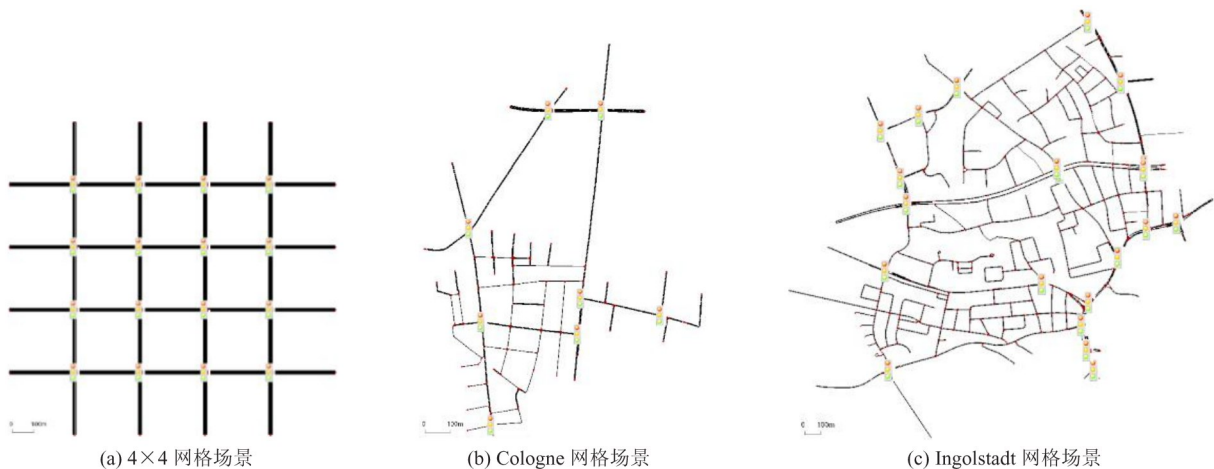


图 3 实验场景

SUMO 仿真软件自带的真实因戈尔施塔特交通场景, 截取其中一个较大区域, 作为更复杂的实验交通场景验证本方法的有效性。

本文方法与多交叉口交通信号控制中几种典型

的方法进行了比较。

MP (MaxPressure): 压力是上游排队车辆与下游排队车辆的数目差值。MP 通过计算每个相位下的压力, 比较它们大小, 最后激活压力最大的相位。

IPPO:不考虑邻居信息的单个深层 RL 方法。一个智能体控制一个交叉口,并且独立更新自己的网络。

MADDPG:通过将每个智能体的观测状态送入 critic 网络进行集中训练,并由 actor 单独执行每个智能体的动作。

FMA2C:一种分层的多智能体强化学习方法,被分为 manager 网络和 worker 网络,由 manager 网络来指导 worker 网络做出决策。

MPlight:利用“压力”概念设计来设计强化学习智能体,以实现区域级的信号协调。并且通过对奖励设计进行设计,实现独立智能体可以内隐协调,从而降低维度。

3.2 实验性能

本节在 3 个实验场景展示了与传统控制方法 MaxPressure 及强化学习控制方法 MPlight、FMA2C 等的结果对比,并从平均等待时间,平均队列长度等评估指标分析了实验结果。

表 1 总结了在 4×4 网格场景下的实验结果,MaxPressure 是根据交叉口的通行压力控制交通灯,IPPO 是将每个交叉口视为单独的智能体进行控制。FMA2C 是通过共享策略实现交叉口的通信。MADDPG 通过集中训练,分布执行实现交通灯之间合作。表 2 总结了 Ingolstadt 场景下的结果对比。从结果中可以观察到,所有分布式的控制结果都表现良好,基于 MADDPG 的控制结果表现最差,这是由于在集中训练的过程中,价值网络的输入是将每个智能体的局部观察全部输入价值网络来进行训练,这导致训练过程中网络的输入维度过高,信息繁杂,增加了训练的复杂度。因此,分布式比集中式控制效果更好。MPlight 通过共享参数能够从其他十字路口的体验中受益,但并不能获得其他交叉口的行为对自身的影响。同时所有的交叉口的体验由于在环境中所处的位置不同,所需的体验也相同,例如, 4×4 网格中边缘的交叉口的体验不一定能从中心交叉口的体验中学习到有效经验。所以,通过图卷积

表 1 4×4 网格场景下的结果对比

模型	平均队列长度/veh	平均等待时间/s	平均旅行时间/s	平均延迟时间/(s · veh ⁻¹)
本文方法	0.26	9.30	140.41	30.03
IPPO	0.50	18.93	153.96	44.01
MPlight	0.55	20.21	157.77	46.91
MaxPressure	0.56	21.64	157.65	52.55
FMA2C	1.74	71.42	209.28	99.77
MADDPG	2.24	88.41	228.26	117.92

表 2 Ingolstadt 场景下的结果对比

模型	平均队列长度/veh	平均等待时间/s	平均旅行时间/s	平均延迟时间/(s · veh ⁻¹)
本文方法	0.53	12.47	185.87	48.42
MPlight	1.47	34.56	215.71	78.15
MaxPressure	1.36	35.44	213.27	83.55
FMA2C	1.78	44.24	226.57	90.29
IPPO	1.88	46.44	233.70	93.11
MADDPG	2.25	51.33	253.32	128.79

网络获得道路级状态之间的影响,能够细致描述交叉口的交通状态,同时考虑历史状态信息,为交叉口提供更多的有效信息来学习经验。

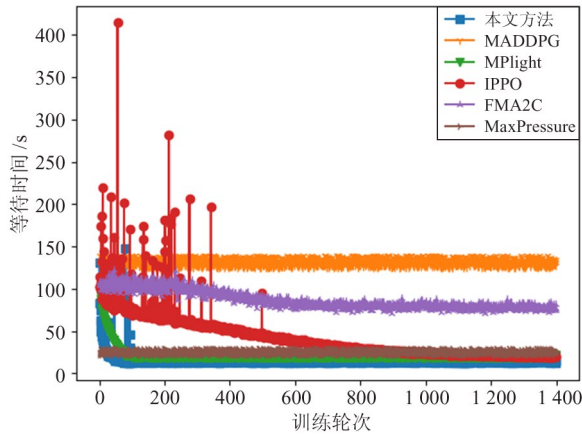
表 3 总结了在 Cologne 场景下的实验结果。与最优的强化学习方法 IPPO 相比,本文方法在平均

队列长度上减少了 27%,在平均等待时间上减少了 24%,缩短了 2% 的平均旅行时间,减少了 8% 的平均延迟。与最先进传统控制方法相比,缩短了 8% 的平均旅行时间和 23% 的平均延迟。

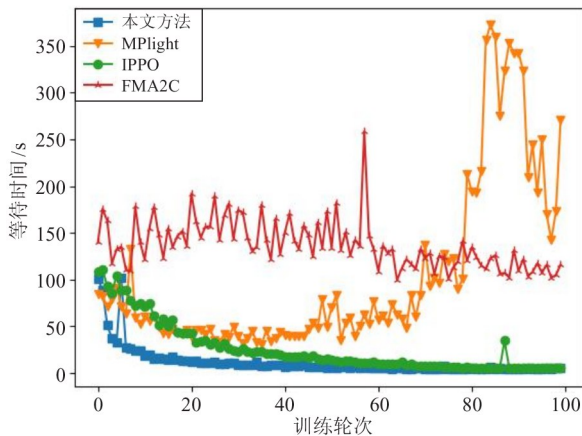
图 4(a) 显示了在 4×4 网格场景下前 100 轮的

表 3 Cologne 场景下的结果对比

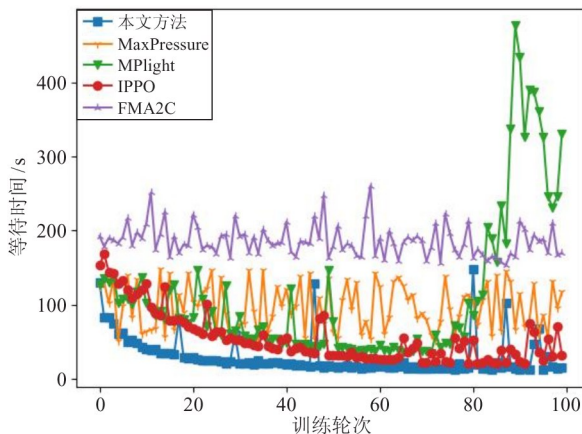
模型	平均队列长度/veh	平均等待时间/s	平均旅行时间/s	平均延迟时间/(s · veh ⁻¹)
本文方法	0.26	3.92	83.76	19.75
IPPO	0.36	5.18	85.70	21.62
MaxPressure	0.62	8.50	92.02	28.18
FMA2C	1.02	14.25	97.98	33.87
MPlight	2.32	30.33	123.92	60.41
MADDPG	3.42	36.52	136.97	69.91



(a) 4 × 4 网格场景下平均等待时间



(b) Cologne 网格场景下平均等待时间



(c) Ingolstadt 网格场景下平均等待时间

图 4 平均旅行时间结果

训练的实验结果。结果表明, MADDPG 与 FMA2C 在短时间的训练过程中,难以收敛到一个最优的策略。传统的 MaxPressure 虽然一直有较稳定的表现,但是难以通过学习过程获得更优的控制效果。图 4(b)是在 Cologne 场景下的实验结果, MPlight 与 FMA2C 在训练的后半部分难以继续收敛到最优结果,本文方法在 40 轮左右时就能具有稳定的控制效果。图 4(c)是 Ingolstadt 场景下的实验结果,所有方法在该场景下都会出现不同程度的波动,但是本文方法波动幅度更小。FMA2C 在后半部分的训练过程还会出现平均等待时间增加的情况,但是本文方法依旧可以快速收敛到最优的控制结果。

表 4 是在场景 Cologne 下,不同方法运行 100 轮所需的运行时间,可以看出,与强化学习的控制方法相比,本文方法所需运行时间最短。虽然 MaxPressure 运行时间最短,但是获得的控制效果不如本文方法。

表 4 模型的运行时间

模型	本文方法	FMA2C	MPlight	MaxPressure
仿真时间/s	1 671.77	3 694.97	2 498.20	693.02

4 结论

本文提出了一种基于 GCN-LSTM 的多交叉口交通信号控制方法。通过二值权重网络对路网进行构图,实现了路网中路段级连接关系的获取,提高了交叉口状态信息的细致程度;通过图卷积网络,实现了对路段状态信息的聚合,有效提取了路段空间状态特征;通过 LSTM 网络有效提取了路段历史状态信息;通过竞争网络框架实现交通信号灯的相位控制。在 3 个场景下进行实验对比,结果表明本文方法在各个交通场景下,能够有效减少车辆的等待时

间,达到更优的控制效果。

参考文献

- [1] 陆丽萍, 程昱, 褚端峰, 等. 基于竞争循环双 Q 网络的自适应交通信号控制[J]. 中国公路学报, 2022, 35(8):267-277.
- [2] HAYDARI A, YILMAZ Y. Deep reinforcement learning for intelligent transportation systems: a survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(1):11-32.
- [3] LIANG X, DU X, WANG G, et al. A deep reinforcement learning network for traffic light cycle control[J]. IEEE Transactions on Vehicular Technology, 2019, 68(2):1243-1253.
- [4] CHU T, WANG J, CODECÀ L, et al. Multi-agent deep reinforcement learning for large-scale traffic signal control[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(3):1086-1095.
- [5] YANG S, YANG B, WONG H S, et al. Cooperative traffic signal control using multi-step return and off-policy asynchronous advantage actor-critic graph algorithm[J]. Knowledge-Based Systems, 2019, 183:1-19.
- [6] ASLANI M, MESCARI M S, WIERING M. Adaptive traffic signal control with actor-critic methods in a real-world traffic network with different traffic disruption events[J]. Transportation Research Part C: Emerging Technologies, 2017, 85:732-752.
- [7] PRASHANTH L A, BHATNAGAR S. Reinforcement learning with function approximation for traffic signal control[J]. IEEE Transactions on Intelligent Transportation Systems, 2010, 12(2):412-421.
- [8] WEI H, CHEN C, ZHENG G, et al. Presslight: learning max pressure control to coordinate traffic signals in arterial network[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 2019:1290-1298.
- [9] CHEN C, WEI H, XU N, et al. Toward a thousand lights: decentralized deep reinforcement learning for large-scale traffic signal control[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA: AAAI Press, 2020, 34(4):3414-3421.
- [10] XIE D, WANG Z, CHEN C, et al. IEDQN: information exchange DQN with a centralized coordinator for traffic signal control[C]//2020 International Joint Conference on Neural Networks. Glasgow, UK: IEEE, 2020:1-8.
- [11] 曲昭伟, 潘昭天, 陈永恒, 等. 考虑博弈的多智能体强化学习分布式信号控制[J]. 交通运输系统工程与信息, 2020, 20(2):76-82.
- [12] 丁世飞, 杜威, 郭丽丽, 等. 基于双评论家的多智能体深度确定性策略梯度方法[J]. 计算机研究与发展, 2023, 60(10):2394-2404.
- [13] WANG M, WU L, LI J, et al. Traffic signal control with reinforcement learning based on region-aware cooperative strategy[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7):6774-6785.
- [14] DEVAILLY F X, LAROCQUE D, CHARLIN L. IG-RL: inductive graph reinforcement learning for massive-scale traffic signal control[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7):7496-7507.
- [15] MA J, WU F. Feudal multi-agent deep reinforcement learning for traffic signal control[C]//Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems. Auckland, New Zealand: AAMAS, 2020:816-824.

Multi intersection signal light control based on GCN-LSTM

XU Dongwei^{*}, ZHU Hongjun^{**}, GUO Haifeng^{*}, ZHOU Xiaogang^{***}, TANG Lixin^{****}

(^{*} Institute of Cyberspace Security, Zhejiang University of Technology, Hangzhou 310023)

(^{**} College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

(^{***} Ningbo Ningong Transportation Engineering Design Consulting Co., Ltd., Ningbo 315010)

(^{****} Zhejiang Expressway Company Limited, Hangzhou 311500)

Abstract

Reinforcement learning (RL) has become a promising solution in signal control because of its ability to solve complex decision-making problems in a highly dynamic environment. Most methods based on reinforcement learning generate agent actions independently, which may lead to action conflicts at intersections and waste of road resources. Therefore, a multi intersection signal control method based on graph convolution network-long short-term memory (GCN-LSTM) is proposed. Firstly, multi intersections are mapped based on binary weight network. Secondly, long short-term memory (LSTM) obtains the historical state information of intersections by aggregating the spatial state information of surrounding intersections through graph convolution network. Finally, the Q value network based on the competitive network framework is used to select actions to control the intersection phase. The experimental results show that compared with some reinforcement learning methods, the queue length at intersections and the waiting time of vehicles in the road network can be reduced in the signal light control at multiple intersections.

Key words: intelligent transportation system, traffic light control, multi-agent reinforcement learning, long short-term memory, graph convolution network