第 62 卷　第 1 期
2026 年 1 月

南京大学学报(自然科学)
JOURNAL OF NANJING UNIVERSITY
(NATURAL SCIENCE)

Vol. 62, No. 1
Jan., 2026

DOI:10.13232/j.cnki.jnju.2026.01.006

基于Shapley值拥堵归因与多智能体深度强化学习的交通信号控制优化方法

车　倩[1,2],王　群[1],王义晶[3],刘　晓[2],王万元[2],宋沫飞[2*]

(1.江苏警官学院计算机信息与网络安全系,南京,210031;2.东南大学计算机科学与工程学院,南京,211189;
3.军事科学院战略评估咨询中心,北京,100091)

摘　要:城市化进程加速带来了日益严峻的交通拥堵问题,亟须研发高效的智能交通信号灯控制方法.基于深度强化学习(Deep Reinforcement Learning,DRL)的交通信号控制方法能依据实时交通信息来调整信号策略,但现有方法在单智能体建模中缺乏协同性,在多智能体建模中则面临复杂性与可扩展性不足的挑战.为此,提出一种基于Shapley值拥堵归因与多智能体深度强化学习的交通信号控制优化方法.首先,引入合作博弈中的Shapley值进行拥堵归因分析,将路口信号策略视为博弈参与者,路网拥堵状态作为合作结果,以量化各路口对拥堵的责任度.其次,提出归因辅助的DRL框架:(1)在多智能体同步决策中,仅联合训练Top-k个高责任度路口智能体,逼近全网络联合训练效果;(2)针对同步决策的稳定性缺陷,提出归因辅助的顺序异步决策方法,其中,决策顺序的选择依据基于Shapley值的归因分析结果.实验结果证明了基于Shapley值的拥堵归因的有效性,与基线方法相比,提出的框架在提高训练效率和整体交通效率方面具有优越性能.

关键词:交通信号控制,信用分配,Shapley值,深度强化学习,拥堵归因

中图分类号:TP399　　**文献标志码:**A

Shapley value-based congestion attribution: A practical multiagent reinforcement learning for traffic signal control

Che Qian[1,2], Wang Qun[1], Wang Yijing[3], Liu Xiao[2], Wang Wanyuan[2], Song Mofei[2*]

(1. Department of Computer Information and Network Security, Jiangsu Police Institute, Nanjing, 210031, China;
2. School of Computer Science and Engineering, Southeast University, Nanjing, 211189, China;
3. Center for Strategic Assessment and Consulting, Academy of Military Science, Beijing, 100091, China)

Abstract: The rapid urbanization process has significantly exacerbated traffic congestion in metropolitan areas, creating an urgent need for intelligent traffic management solutions. In this context, DRL (Deep Reinforcement Learning) has emerged as a prominent research focus due to its superior dynamic adaptability in complex traffic environments. However, existing approaches face critical limitations: single-agent DRL models lack coordination capabilities among intersections, while multi-agent systems often suffer from high computational complexity and poor scalability. To address these challenges, this paper proposes a novel DRL-based traffic signal control framework that integrates congestion attribution analysis with optimization strategies. First, the Shapley value from cooperative game theory is applied to analyze congestion attribution. Considering intersection signal strategies as game players and road network congestion as the cooperative result, it quantifies each

基金项目:国家自然科学基金(12201619),江苏省产业前瞻与关键核心技术重点项目(BE2021001),山东省自然科学基金(ZR2023MA031),江苏省高等学校自然科学研究面上项目(25KJB520007)

收稿日期:2025－08－25

*通信联系人,E-mail:songmf@seu.edu.cn

intersection's contribution to congestion. Secondly, it proposes a Shapley value-based attribution-assisted DRL optimization framework. During multi-agent synchronous decision-making, it jointly trains only the Top-k high-contribution intersection agents, approaching full network joint-training performance. To address synchronous decision-making's stability issue, it develops an attribution-assisted sequential decision-making approach, where decision-order selection is based on Shapley value-based attribution analysis results. Experimental results verify the effectiveness of Shapley value-based congestion attribution. Compared with baseline methods, the proposed framework improves the training efficiency and the overall traffic efficiency.

Keywords: intelligent traffic signal control, credit assignment, Shapley value, Deep Reinforcement Learning, congestion attribution

交通拥堵对社会和经济产生了重大影响[1],智能交通信号控制基于实时交通信息动态优化路口的车辆通行情况,能够有效缓解城市交通拥堵[2]. 目前的交通拥堵归因分析主要侧重于通过专家经验[3]、启发式方法[4]或人工智能方法[5]来识别路口的拥堵根源,却忽略了路口之间的协同特性,也缺乏对信号控制策略如何影响拥堵的定量分析. 此外,现有基于数据驱动的归因方法往往可解释性较差. 博弈论中的Shapley值分解方法[6]能公平估计每个参与者的贡献[7],已被证明是能够满足特征归因方法所有理想特性的可解释模型[8]. 具体地,Shapley值分解基于严格数值推导客观量化拥堵责任,避免了启发式方法中权重设置的主观偏差问题. 通过博弈论框架精确计算边际贡献,捕捉传统方法忽略的协同效应. 并且,Shapley值以数值形式明确量化信号控制策略对拥堵的责任度,具备良好的可解释性. 因此,本文引入Shapley值来衡量信号控制策略对拥堵的责任度,并采用蒙特卡罗采样方法全面考虑交通状况,以提高计算效率.

多智能体强化学习(Multi-Agent Reinforcement Learning,MARL)[9]通过智能体不断与环境交互来解决信息不完备以及环境动态情况下的决策问题,可以在无需人工干预的情况下显著提升路网效率. MARL将智能交通信号灯建模为智能体,并利用路口之间的实时交互来实现整体交通流量的协同优化. 对于训练过程中的策略更新机制,现有方法多采用同步更新,即所有智能体同时更新它们的策略[10-11]. 然而,随着交通信号灯数量的增加,采用集中式训练时产生计算复杂度高和可扩展性差的问题[12]. 此外,随着系统复杂性的增加,同步更新还可能引发非平稳性问题. 异步更新则允许智能体独立更新策略,提高了灵活性和可扩展性. 然而,Wang et al[13]指出,错误的更新顺序降低了异步更新的性能,导致不稳定性和收敛速度变慢. 目前关于更新顺序的研究主要依赖于随机方法、专家经验或贪心算法[14-15],不适用于复杂且动态的交通场景.

为了解决以上局限性,本文提出一种基于Shapley值归因辅助的MARL优化框架,根据交通信号控制策略对拥堵的影响程度来更新控制策略. 对于同步更新方式,提出"部分决策优化"方法,以选择最关键的交通信号灯子集,通过保持较少的智能体更新来提高效率,缓解大规模网络中同步更新的弊端. 对于异步更新方式,提出"顺序决策优化"方法,根据智能体对拥堵的责任度来指导其更新顺序,提高训练的稳定性并促进性能的提升. 本文的主要贡献如下.

(1)采用Shapley值进行拥堵归因,为智能交通信号控制策略对交通拥堵的影响提供一种可解释且定量的评估方法.

(2)设计一种基于归因辅助的MARL优化框架,利用优先级来优化同步和异步更新策略.

(3)实验验证了Shapley值在拥堵归因方面的有效性,并证明归因辅助的MARL优化框架在提高整体交通效率方面优于现有的基线方法.

1 相关工作

1.1 交通信号控制方法 传统的交通信号控制方法,如定时控制[16],依赖于预先设定的相位顺序和定时方案,在应对实时交通流量波动时缺乏灵活性. 自适应控制方法,包括模糊逻辑控制[17]、遗传算法[18]以及模型预测控制[19],根据实时交通流量动态调整信号配时,但这些方法通常依赖于预

先设定的规则或模型,难以有效应对复杂且不可预测的交通场景.相比之下,强化学习(Reinforcement Learning,RL)不需要预先设定规则,直接通过与环境交互来学习最优策略,具有更强的适应性和鲁棒性[20].此外,由于交通信号控制涉及多个需要协同的路口,MARL可以利用路口之间的复杂交互来优化整体交通流量[21-22].

然而,现有的方法鲜有根据交通信号对拥堵的责任度来区分其优先级,也没有基于此优先级提供有针对性的优化策略.此外,虽然Shapley值已在一些MARL方法中得到应用[23-24],但主要集中在分解和优化奖励函数上,而不是指导宏观层面的多智能体强化学习框架的设计与优化.

1.2　交通拥堵归因方法　交通拥堵归因分析旨在找出导致交通拥堵的各种因素[24],本文关注的是在现有路网结构内对拥堵责任进行定量分析.传统方法依赖于定性分析并结合专家经验[9].随着交通大数据采集等技术的广泛应用,更多的研究转向数据驱动的方法,利用启发式方法[10]或人工智能方法[11,25]来分析交通拥堵的成因.Winter[6]首先基于历史检测器收集的数据提出新的指标来评估交通拥堵的严重程度,然后通过指标的时空变化图来识别拥堵首次出现的道路,利用变换的累积曲线来分析拥堵的激活时间和持续时间,最终将拥堵首次出现时所定位的道路视为拥堵产生的根本原因.Lundberg and Lee[8]采用先预测后解释的方法,选取多种变量对路网中车辆的平均行驶速度建立预测模型,并基于最好的预测模型进行影响因素重要性的分析,以此发现导致路网陷入拥堵的原因.然而,目前基于启发式或人工智能的归因方法缺乏可解释性,且未能充分考虑交通信号灯之间的协同关系.

2　预备知识

2.1　交通信号控制

定义1　道路网　路网中的路口可以表示为$I=\{I_i\}_{i=1}^{n}$,道路$R_{i,j}\in R$是连接路口I_i和I_j的边.每条道路由多个分支车道l_i组成.

定义2　车辆运行　车辆运行(l_i,l_j)定义为车辆从车道l_i行驶到车道l_j的过程.

定义3　信号相位　信号相位SP_i表示在路口I_i允许的一组交通运行情况.

定义4　车辆密度　车道的车辆密度定义为$\frac{x(l)}{x_{\max}(l)}$,其中,$x(l)$是车道l上的实际车辆数,$x_{\max}(l)$是该车道允许通行的最大车辆数.

定义5　交通压力　车辆运行的压力定义为驶入车道和驶出车道的车辆密度之差:

$$w(l_i,l_j)=\frac{x(l_i)}{x_{\max}(l_i)}-\frac{x(l_j)}{x_{\max}(l_j)} \tag{1}$$

定义6　路口I_i的交通压力ρ_i　路口I_i的交通压力ρ_i定义为该路口所有交通流压力的绝对值之和,记为:

$$\rho_i=\sum_{(l_i,l_j)\in i}\left|w(l_i,l_j)\right| \tag{2}$$

2.2　多智能体建模　给定一个由多个路口$I=\{I_i\}_{i=1}^{n}$组成的路网,将交通信号灯控制问题定义为一个部分可观测马尔可夫决策过程,表示为元组$\langle S,A,P,r,N,\gamma\rangle$.其中,$S$表示环境的状态,每个智能体$i\in N:=1,2,\cdots,N$独立地从其动作空间中选择一个动作$a^i\in A^i$,依据转移函数$P(s'|s,a)$:$S\times A^N\times S\to[0,1]$,将状态$s$转换为新状态$s'$.环境提供一个全局奖励:$r(s,a):S\times A^N\to R$.最终目标是最大化长期累积奖励$\sum_{i=0}^{\infty}\gamma^i r(s^i,a^i)$,其中,$\gamma\in[0,1]$表示折扣因子.每个智能体独立采用一种信号控制策略$\{\pi^1,\cdots,\pi^n\}$.状态价值函数和状态-动作价值函数定义如下:

$$V_\pi(s)=E\left[\sum_{t=0}^{\infty}\gamma^t r_t\middle|s_0=s\right]$$

$$Q_\pi(s,a)=E\left[\sum_{t=0}^{\infty}\gamma^t r_t\middle|s_0=s,a_0=a\right]$$

优势函数表示为:

$$A_\pi(s,a)=Q_\pi(s,a)-V_\pi(s)$$

2.2.1　状态　在时间步$t\in N$,每个智能体i接收部分观测值s_t^i,包括驶入和驶出车道的车辆密度以及路口i当前交通信号相位SP_t^i.全局状态$S_t=\prod_{i=1}^{n}s_t^i$汇集所有智能体的部分观测值.

2.2.2　动作　将决策间隔设置为10 s,即智能体每隔10个时间步选择一个信号相位SP_t^i作为它

的动作 a_t^i. 所有智能体的联合动作表示为 $u_t=\{a_t^1, a_t^2, \cdots, a_t^n\}$.

2.2.3 奖励 路口压力反映驶入车道和驶出车道之间的不平衡状况,这会导致交通拥堵. 奖励函数是基于路口压力设计的,总体奖励为每个智能体的局部奖励之和:

$$R=\sum_{i=1}^{n}\rho_i \tag{3}$$

3 基于Shapley值的交通信号控制拥堵归因方法

提出一种基于Shapley值分解的交通信号控制拥堵归因方法,以量化每个交通信号灯对路网中潜在拥堵的责任度,为后续的优化策略提供依据. 为了计算Shapley值,将交通信号控制建模为合作博弈,其中,各个信号灯智能体协同工作以最大化交通效率. 随后,推导每个交通信号灯的Shapley值计算式,并采用蒙特卡罗采样方法,全面考虑各种交通流量场景,以提高计算效率.

3.1 博弈建模与Shapley值分解 将交通信号控制系统建模为一个合作博弈 $G=(N,v)$,其中,参与者集合 $N=\{1,2,\cdots,n\}$ 表示系统中的 n 个交通信号智能体. 特征函数 $v: 2^N\rightarrow\mathbb{R}$ 表示任意交通信号子集 $U\subseteq N$ 协同工作时整个路网能达到的交通效率,可以量化为整个路网的平均行程时间. 博弈的目标是最小化 $v(U)$.

为了计算Shapley值,需要衡量每个路口对整体的边际贡献,可以通过评估一个路口参与协同控制前后系统拥堵程度的变化来计算. 具体地,ϕ_i 量化了交通信号智能体 i 对整体交通拥堵的责任度:

$$\phi_i=\sum_{U\subseteq N\setminus\{i\}}\frac{|U|!(n-|U|-1)!}{n!}\left[v(U\cup\{i\})-v(U)\right] \tag{4}$$

其中,$U\subseteq N\setminus\{i\}$ 表示不包含智能体 i 的交通信号灯子集,$|U|$ 表示集合 U 中的元素个数. $v(U\cup\{i\})-v(U)$ 表示在交通效率方面的提升(即拥堵程度的降低),这是智能体 i 加入集合 U 产生的边际贡献的平均值. $\frac{|U|!(n-|U|-1)!}{n!}$ 表示集合 U 在所有交通信号可能组合中出现的概率.

3.2 基于蒙特卡罗采样的估计 为了全面分析交通系统,需要考虑在各种交通流量条件下各路口智能体的边际贡献. 然而,直接计算Shapley值需要遍历所有可能的联盟组合,会产生较高的计算成本. 因此,本节采用蒙特卡罗采样方法,考虑潜在的交通流量状态,并针对每个交通流量状态对不同的联盟状态进行采样,以计算每个路口的综合Shapley值:

$$\hat{\phi}_i=\frac{1}{M}\sum_{m=1}^{M}\frac{1}{K}\sum_{k=1}^{K}\left[v(U_{m,k}\cup\{i\})-v(U_{m,k})\right] \tag{5}$$

其中,$\hat{\phi}_i$ 是智能体 i(即路口的交通信号灯)的估计Shapley值,M 是使用蒙特卡罗方法采样的交通流量状态数量,K 是针对每个交通流量状态采样的联盟状态数量,$U_{m,k}$ 是在第 m 个交通流量状态下采样的第 k 个联盟状态(不包括智能体 i). $v(U_{m,k}\cup\{i\})-v(U_{m,k})$ 是智能体 i 加入联盟 $U_{m,k}$ 带来的交通效率提升(即拥堵程度的降低).

本文采用MARL方法,对所有交通信号智能体进行联合训练,直至模型收敛,从而估计每个状态下的协同相位策略. 当一个智能体退出联盟时,相应的交通信号将恢复至原始的相位控制策略.

4 归因辅助的DRL交通信号控制优化方法

获取路网中各路口信号控制策略对交通拥堵的影响大小序列后,将此序列作为辅助信息,引入后续的多智能体强化学习建模中. 首先,将该知识引入多智能体同步决策策略(Multi-Agent Traffic Control Strategy,MATCS),在人工构建的路网环境下,分别依据路口的Shapley值大小,选取前Top-k个路口进行智能体建模并联合训练. 考虑同步决策时,智能体面临不能观察其他智能体的变化而带来的非稳定性问题,将多智能体顺序决策的方法应用其中,提出基于Shapley值的顺序异步控制优化策略(Sequential Traffic Control Strategy,SeTCS),其中决策顺序的选择依据前文提出的归因分析方法.

4.1　归因辅助的部分同步决策　提出一种基于Shapley值的部分同步决策优化方法,旨在利用策略同步的优势,通过减少控制数量来克服可扩展性和非平稳性问题.

具体地,利用Shapley值选择最重要的k个交通信号智能体进行控制优化,从而降低训练开销.为所有智能体训练一个全局价值函数(Critic网络),Critic网络接收来自所有智能体的观测信息,而每个智能体独立接收局部观测信息,并通过共享的Actor网络生成动作概率.

定义$i_{1:k}\subseteq I=top-k\left(\left\{\hat{\phi}_1,\cdots,\hat{\phi}_n\right\}\right)$,其中,$\phi$表示Critic网络的所有参数.每个智能体$i\in i_{1:k}$按照共享策略行动,并生成个体轨迹$\tau_i=\left\{s_t^i,a_t^i,r_t^i\right\}_{t=1}^{T}$,其中,$r_t^i$是智能体$i$的局部奖励值.

对于Actor网络,通过最大化目标函数来更新共享策略π_θ:

$$J(\theta)=E\left[\min\left(l_t^i(\theta)\hat{A}_t^i,clip\left(l_t^i(\theta),1-\varepsilon,1+\varepsilon\right)\hat{A}_t^i\right)\right] \tag{6}$$

其中,θ表示Actor网络的所有参数,ε是人工设置的裁剪系数,$l_t^i(\theta)=\dfrac{\pi_\theta\left(a_t^i\middle|s_t^i\right)}{\pi_{\theta_{\text{old}}}\left(a_t^i\middle|s_t^i\right)}$,$\pi_{\theta_{\text{old}}}$表示与环境交互的旧策略,$\pi_\theta$表示本轮学习要优化的策略.$\hat{A}_t^i=\sum_{l=0}^{h}(\gamma\lambda)^l\delta_{t+l}^i$是基于当前价值函数$\hat{V}_\phi$的GAE估计器,时间差分定义为$\delta_t=r_t+\gamma V_\phi(s_{t+1})-V_\phi(s_t)$,其中,$\gamma$是折扣因子,$\lambda$是平滑因子.

对于Critic网络,通过最小化损失函数来更新:

$$L(\phi)=E\left[\left(V_\phi(s_t)-R_t\right)^2\right] \tag{7}$$

其中,$R_t=\sum_{t'=t}^{T}\gamma^{t'-t}r_{t'}$.

4.2　归因辅助的顺序异步决策　4.1中的同步决策采用同时更新智能体的方式,不能观察其他智能体的变化,可能导致非稳定性问题.本节提出基于Shapley值的顺序异步决策优化方法,根据Shapley值来确定智能体策略更新的顺序.

每个智能体采用独立的Actor网络,同时采用一个共享的全局Critic网络来评估整个系统.依据Shapley值决定每个智能体的策略更新顺序,通过迭代方式改进各个智能体的策略.

在任意多智能体马尔可夫决策过程中,给定一个联合策略π,对于任意状态s,都满足:

$$A_\pi^{i_{1:m}}\left(s,a^{i_{1:m}}\right)=\sum_{j=1}^{m}A_\pi^{i_j}\left(s,a^{i_{1:j-1}},a^{i_j}\right) \tag{8}$$

在顺序更新过程中,联合优势函数可以分解为每个智能体的局部优势之和[28].如果能够最大化$A_\pi^{i_j}$的每一项,就可以实现$A_\pi^{i_{1:m}}$的最大化.

将智能体i更新后的策略记为$\bar{\pi}^i$.在更新智能体i时,联合策略可表示为:

$$\hat{\pi}^i=\bar{\pi}^1\times\cdots\times\bar{\pi}^i\times\bar{\pi}^{i+1}\times\cdots\times\bar{\pi}^n \tag{9}$$

为了避免分布偏移,对先更新的智能体采用策略优化方法,利用从联合策略π中收集的样本去近似$\hat{A}^{\pi^{i-1}}$:

$$A^{\pi,\hat{\pi}^{i-1}}(s_t,a_t)=\delta_t+\sum_{k\geqslant1}\left(\prod_{j=1}^{k}\lambda M_{t+j}^{i-1}\right)\delta_{t+k} \tag{10}$$

其中,

$$M_{t+j}^{i-1}=\min\left(1.0,\frac{\hat{\pi}^{i-1}\left(a_{t+j}\middle|s_{t+j}\right)}{\pi\left(a_{t+j}\middle|s_{t+j}\right)}\right)$$

$$\delta_t=r(s_t,a_t)+\gamma V(s_{t+1})-V(s_t)$$

在优化过程中,除π_t之外的策略都是固定的,所以更新优势函数绝对值更大的智能体对于优化的贡献更大.而Shapley值更大的智能体,对优化更新的贡献也更大,因此,可以根据Shapley值来确定智能体的选择规则.具体地,基于Shapley值排序,定义为$i_s=\left\{i_{k_1},i_{k_2},\cdots,i_{k_n}\right\}$,顺序更新过程可表示为:

$$\begin{aligned}&\hat{\pi}^{k_i}=\bar{\pi}^{k_1}\times\cdots\times\bar{\pi}^{k_i}\times\pi^{k_{i+1}}\times\cdots\times\pi^{k_n}\\&\pi=\hat{\pi}^{k_0}\to\hat{\pi}^{k_1}\to\cdots\to\hat{\pi}^{k_n}=\bar{\pi}\end{aligned} \tag{11}$$

对于i_s中的智能体i,目标是最大化以下裁剪目标:

$$L_{\hat{\pi}^{i-1}}(\theta)=E\left[\min\left(l(s,a)A^{\pi}(s,a),clip\left(l(s,a),1\mp\varepsilon^i\right)A^{\pi,\hat{\pi}^{i-1}}\right)\right] \tag{12}$$

其中,$l(s,a)=\dfrac{\bar{\pi}^i\left(a^i\middle|s\right)}{\pi^i\left(a^i\middle|s\right)}$.

5 实验结果及分析

5.1 实验设置 选取 Cityflow[26]作为实验平台，选择的路网包括一个不规则的真实路网 roadnet_1 和一个模拟路网 roadnet_2，如图 1 所示．其中，roadnet_1 是根据南京城区区域路网由人工构建而成．roadnet_2 是由 12 个交叉路口构成的规则性路网，每条道路包含左转、直行和右转三条车道．车道长度和宽度均保持一致．每个路口的信号配时策略采用相同的八相位切换模式，每隔 30 s 切换下一个相位，中途会持续 5 s 的红灯时间．

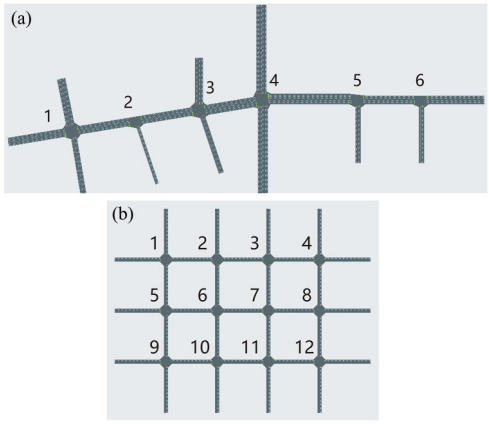

图 1 路网示意图：(a) roadnet_1；(b) roadnet_2

Fig.1 Network diagram：(a) roadnet_1，(b) roadnet_2

使用的评价指标如下．

(1)车道平均排队长度(Queue)，即车道上处于等待状态的车辆数量．该指标直接反映交叉路口进口道的排队状况，其值越高，交通流畅性越差．

(2)车辆平均行驶时间(Travel Time)，即车辆从进入路网到离开所花费的平均时间，用于整体评估路网通行效率，时间越长，整体交通性能越低．

(3)交通拥堵率(Traffic Congestion Rate，*TCR*)，用来综合量化路网中的交通拥堵程度，定义如下：

$$TCR=\frac{1}{|R|}\sum_{i=1}^{|R|}\frac{\overline{V_i}}{V_i} \tag{13}$$

其中，V_i 表示第 i 条道路上当前的平均车速，$\overline{V_i}$ 是在无外部环境限制条件下第 i 条道路上车辆的自由行驶速度．*TCR* 越高，交通拥堵越严重．通常认为，当 $TCR>3.3$ 时，表示道路出现严重拥堵．

使用的对比方法如下．

(1) Presslight[27]：一种基于最大压力理论设计的 MARL 交通信号控制方法，通过压力函数奖励机制优化信号策略，在缓解局部拥堵方面具有明显优势．

(2) IDQL[25]：一种结合 DQN 和策略学习的 MARL 方法．每个交叉路口由独立的智能体控制，智能体根据各自交叉口收集的信息独立更新 DQN 的参数，并采用同步策略更新机制．

(3) IPPO[7]：一种基于近端策略优化(PPO)算法的分布式 MARL 方法．各智能体依据各自交叉路口的信息独立更新模型网络参数，采用同步策略进行更新．

(4) HAPPO[14]：一种基于 PPO 算法的 MARL 方法，引入了随机顺序的异步策略更新机制，可在部分观测环境下实现高效协作．

5.2 基于 Shapley 值的拥堵归因结果 为了验证 Shapley 值在分析交通信号灯对拥堵责任中的有效性，首先扰动交通信号灯的相位持续时间，以模拟次优信号控制策略引起的拥堵．随后，使用 MARL 训练交通信号灯并计算其 Shapley 值，通过将引起拥塞的信号灯的 Shapley 值与其他信号灯的 Shapley 值进行比较，验证 Shapley 值能否有效地识别导致拥塞的交通信号灯．

信号灯的固定配时策略是基于工作日早晨 7:00－8:00 的真实信号控制策略生成的．在路网 roadnet_1 中，将路口 4 东西方向的相位时长减少 30 s，使 *TCR* 从 2.13 上升到 4.61，达到严重拥堵的状态．路网 roadnet_2 中，将路口 6 的相位切换频率由 30 s 调整为 60 s，使 *TCR* 从 2.58 上升到 6.22．图 2 展示了这两个路网计算得到的 Shapley 值．如图 2a 所示，在路网 roadnet_1 中，路口 4 的平均 Shapley 值最高，且明显高于其他路口，表明其对网络拥堵负主要责任．此外，路口 4 相邻路口的 Shapley 值比距其较远的路口(如路口 1)的 Shapley 值更高，这是因为相邻路口之间的交通情况相互影响，因而相邻路口的信号相位策略也需要作出调整．在图 2b 的路网 roadnet_2 中也可以观察到类似的情况，其中，路口 6 的 Shapley 值最大．

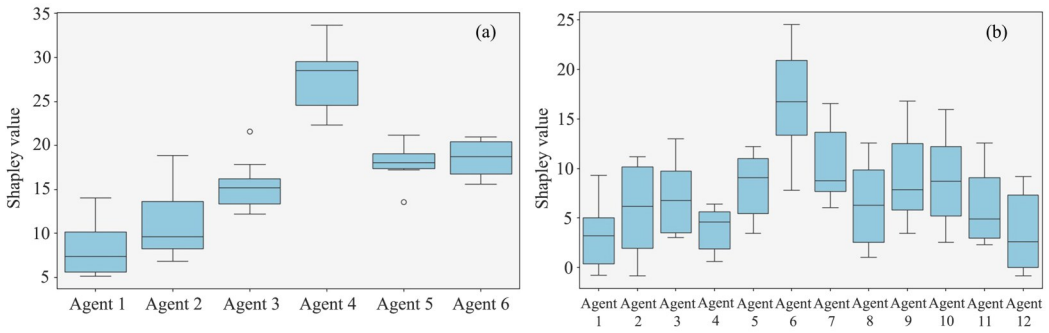

图2　智能交通信号灯Shapley值:(a) roadnet_1; (b) roadnet_2

Fig.2　Shapley values of traffic signal lights: (a) roadnet_1, (b) roadnet_2

5.3　归因辅助的同步决策结果　对基于Shapley值选择交通信号灯以进行同步策略更新的有效性进行评估.

首先,计算在典型交通流量场景下所有交通信号灯的Shapley值,并按照Shapley值的降序选择前k个关键路口信号灯作为智能体进行联合训练. 图3展示了MARL在两个具有不同智能体数量的路网中的性能表现. 由图3b可知,选择对拥堵影响更大的路口信号作为关键智能体,当k=6时,模型的性能接近对所有智能体进行联合训练(k=12)时的性能,差异仅为2.3%,同时,还节省了约28.49%的训练时间.

由于路网roadnet_1中的智能体数量有限,图4a展示了roadnet_2的训练收敛时间,由图可知,模型收敛所需的训练时间随着智能体数量的增加而增加. 保持路网结构不变,将roadnet_2扩

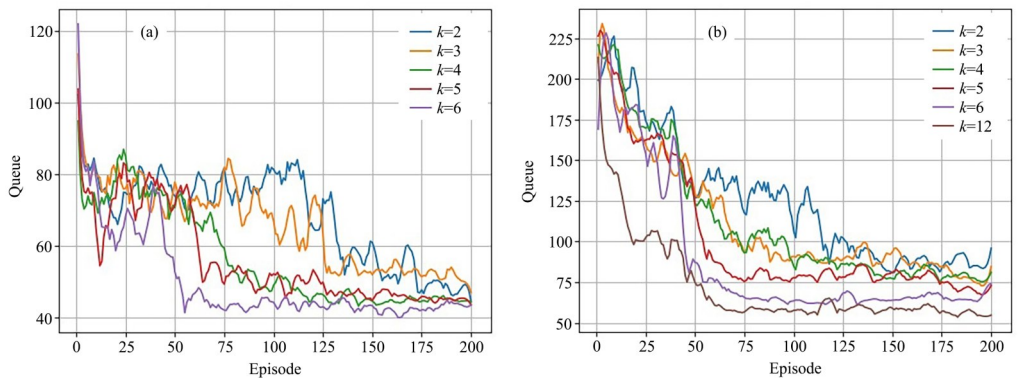

图3　不同智能体数量下同步决策行驶时间:(a) roadnet_1; (b) roadnet_2

Fig.3　Synchronized decision-making travel time under different numbers of agents: (a) roadnet_1, (b) roadnet_2

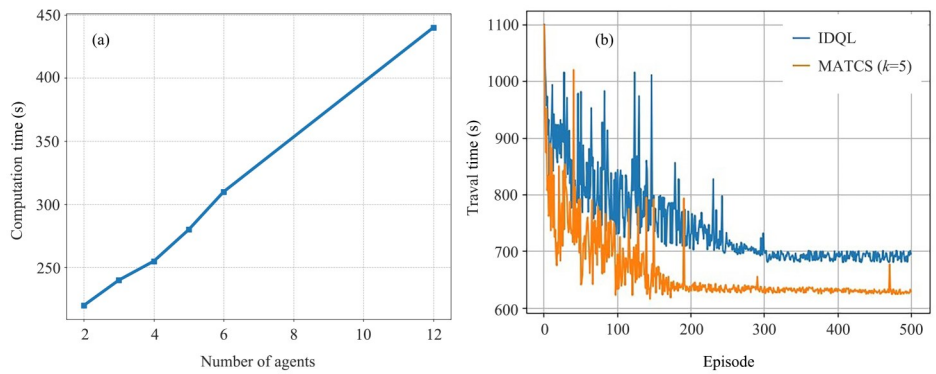

图4　5×5路网环境下的车辆平均行驶时间对比:(a) roadnet_2;(b) roadnet_3

Fig.4　Average travel time in 5×5 road network: (a) roadnet_2, (b) roadnet_3

展为更大的5×5路口的路网roadnet_3场景,以测试模型在大规模场景下的性能.图4b展示了本文的方法(记为"MATCS,k=5")与IDQL方法的性能.

由图可见,对于大规模路网,仅选择一部分关键路口信号作为智能体进行训练比使用IDQL方法对所有智能体进行联合训练的结果更好,这是因为大量智能体的同步策略更新可能导致训练难度增加,并可能收敛到局部最优解.

5.4 归因辅助的顺序决策结果 分析采用Shapley值在异步策略更新中指导智能体更新顺序的有效性.首先,比较本文的方法与采用不同更新顺序的异步更新方法HAPPO的性能.具体地,根据智能体的Shapley值,按降序(SeTCS-descending)和升序(SeTCS-ascending)排列智能体,并将它们与随机更新的HAPPO算法(HAPPO-random)进行比较.

图5a展示了三种算法的性能.由图可见,随机排序方法在通行时间上波动较大,说明其稳定性较差.基于Shapley值排序的SeTCS-ascending和SeTCS-descending都表现更好,说明依据责任度对策略更新进行优先级排序有助于提高训练稳定性和算法性能.

图5b展示了所有算法的性能.本文提出的两种决策优化方法(部分同步决策和顺序异步决策)的性能都显著优于其他基线方法,其中Shapley值辅助的异步更新方法表现最佳.此外,与基线方法相比,本文提出的方法收敛速度更快,稳定性更强,验证了拥堵归因辅助的多智能体强化学习框架的有效性.

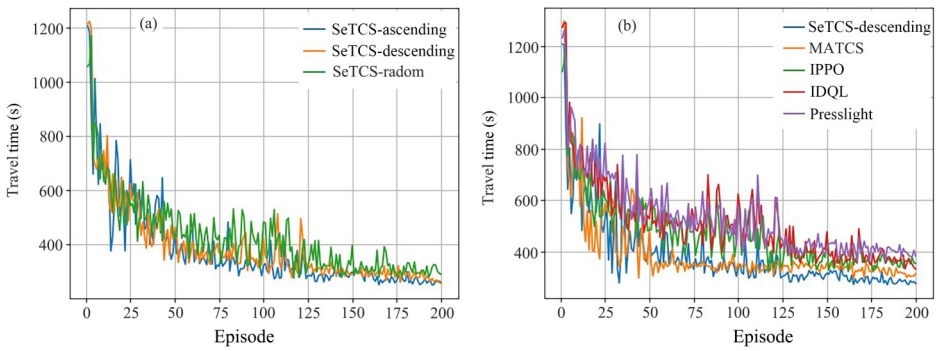

图5 顺序异步决策下平均行驶时间:(a)不同排序方式的训练过程;(b)不同算法的训练过程

Fig.5 Average travel time under sequential asynchronous decision-making: (a) training process with different sorting methods, (b) training process with different algorithms

6 结论

本文提出一种基于Shapley值拥堵归因的多智能体强化学习交通信号灯控制优化框架,利用Shapley值以可解释的方式量化每个智能交通信号灯对路网拥堵的责任度.基于Shapley值的拥堵归因为MARL中的同步和异步策略更新均起到指导作用,提高了训练效率和稳定性.对于同步更新方式,引入了部分决策优化方法,选择关键的Top-k个交通信号灯进行部分决策优化;对于异步更新方式,提出了顺序决策优化方法,根据拥堵责任度对智能体的更新进行优先级排序.实验结果验证了Shapley值在拥堵归因方面的有效性,并证明了本文的优化框架在提高整体交通通行效率方面的性能.

参考文献

[1] Sarwatt D S, Lin Y J, Ding J G, et al. Metaverse for intelligent transportation systems (ITS): A comprehensive review of technologies, applications, implications, challenges and future directions. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(7): 6290−6308.

[2] Villarreal M, Poudel B, Pan J, et al. Mixed traffic control and coordination from pixels// 2024 IEEE International Conference on Robotics and Automation. Yokohama, Japan: IEEE, 2024: 4488−4494.

[3] Chawla S, Zheng Y, Hu J F. Inferring the root cause in road traffic anomalies//2012 IEEE 12th International Conference on Data Mining. Brussels, Belgium: IEEE, 2012: 141－150.

[4] Lee W H, Tseng S S, Shieh J L, et al. Discovering traffic bottlenecks in an urban network by spatiotemporal data mining on location - based services. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(4): 1047－1056.

[5] Chen Y, Li C L, Yue W W, et al. Root cause identification for road network congestion using the gradient boosting decision trees//GLOBECOM 2020－2020 IEEE Global Communications Conference. Taipei, China: IEEE, 2020: 1－6.

[6] Winter E. The shapley valueAumann R, Hart S. Handbook of game theory with economic applications. Amsterdam, Holland: Elsevier, 2002(3): 2025－2054.

[7] MISHRA S K. Shapley value regression and the resolution of multicollinearity. Journal of Economics Bibliography, 2016, 3(3): 498－515.

[8] Lundberg S M, Lee S I. A unified approach to interpreting model predictions//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc., 2017: 4768－4777.

[9] Liu Y L, Luo G Y, Yuan Q, et al. Gplight: Grouped multi - agent reinforcement learning for large - scale traffic signal control//Proceedings of the 32nd International Joint Conference on Artificial Intelligence. Macao, China: IJCAI, 2023: 199－207.

[10] Witt C S D, Gupta T, Makoviichuk D, et al. Is independent learning all you need in the starcraft multi - agent challenge? https://arxiv. org/abs/2011.09533, 2020－11－18.

[11] Yu C, Velu A, Vinitsky E, et al. The surprising effectiveness of PPO in cooperative multi - agent games//Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2022: 24611－24624.

[12] Chu T S, Wang J, Codecâ L, et al. Multi-agent deep reinforcement learning for large - scale traffic signal control. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(3): 1086－1095.

[13] Wang X H, Tian Z, Wan Z Y, et al. Order matters: Agent-by-agent policy optimization//Proceedings of the 11th International Conference on Learning Representations. Online: ICLR, 2023: 1－35.

[14] Kuba J G, Chen R Q, Muning W, et al. Trust region policy optimisation in multi - agent reinforcement learning//ICLR 2022 10th International Conference on Learning Representations. Online: ICLR, 2022: 1046.

[15] Papageorgiou M. Overview of road traffic control strategies. IFAC Proceedings Volumes, 2004, 37(19): 29－40.

[16] Kulkarni G H, Waingankar P G. Fuzzy logic based traffic light controller//2007 International Conference on Industrial and Information Systems. Peradeniya, Sri Lanka: IEEE, 2007: 107－110.

[17] Shaikh P W, El-Abd M, Khanafer M, et al. A review on swarm intelligence and evolutionary algorithms for solving the traffic signal control problem. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(1): 48－63.

[18] Kamenev A, Wang L R, Bohan O B, et al. PredictionNet: Real - time joint probabilistic traffic prediction for planning, control and simulation//2022 International Conference on Robotics and Auto - mation. Philadelphia, PA, USA: IEEE, 2022: 8936－8942.

[19] Wei H, Zheng G J, Gayah V, et al. Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation. ACM SIGKDD Explorations Newsletter, 2021, 22(2): 12－18.

[20] Zhang Z, Yang J C, Zha H Y. Integrating Independent and centralized multi - agent reinforce - ment learning for traffic signal network optimization// Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems. Auckland, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems, 2020: 2083－2085.

[21] Zhang Y T, Zheng G H, Liu Z Y, et al. Marlens: Understanding multi-agent reinforcement learning for traffic signal control via visual analytics. IEEE Transactions on Visualization and Computer Graphics, 2024, 31(7): 4018－4033.

[22] Rizzo S G, Vantini G, Chawla S. Reinforcement learning with explainability for traffic signal control// 2019 IEEE Intelligent Transportation Systems Conference. Auckland, New Zealand: IEEE, 2019: 3567－3572.

[23] Yue W W, Li C L, Chen Y, et al. What is the root cause of congestion in urban traffic networks: Road infrastructure or signal control? IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7): 8662－8679.

[24] Wang M D, Yuan Y, Yan H, et al. Discovering causes of traffic congestion via deep transfer clustering. ACM Transactions on Intelligent Systems and Technology, 2023, 14(5): 1－24.

[25] Likmeta A, Metelli A M, Tirinzoni A, et al. Combining reinforcement learning with rule-based controllers for transparent and general decision-making in autonomous driving. Robotics and Autonomous Systems, 2020, 131: 103568.

[26] Zhang H C, Feng S Y, Liu C, et al. CityFlow: A multi-agent reinforcement learning environment for large scale city traffic scenario//The World Wide Web Conference. San Francisco, CA, USA: Association for Computing Machinery, 2019: 3620－3624.

[27] Wei H, Chen C C, Zheng G J, et al. Presslight: Learning max pressure control to coordinate traffic signals in arterial network//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage: Association for Computing Machinery, 2019: 1290－1298.

[28] 高涵, 罗娟, 蔡乾娅, 等. 一种基于异步决策的智能交通信号协调方法. 计算机研究与发展, 2023, 60(12): 2797－2805.

(责任编辑 杨可盛)