

多 Agent 强化学习下的城市路网自适应 交通信号协调配时决策研究综述

夏新海

(广州航海学院 港口与航运管理学院, 广东 广州 510725)

摘要: 相对于传统的交通信号配时决策方法, 多 Agent 强化学习及其协调方法能更好地适应城市路网交通环境的变化。为探讨其在城市路网自适应交通信号配时决策中的应用, 系统地总结了多 Agent 强化学习及协调机制的研究方法, 详细地分析了国内外研究现状, 并指出现有研究中存在的问题, 在此基础上对未来研究进行了展望。研究结果表明, 既有研究主要针对规模较小的路网, 存在维数灾难问题, 强化学习与协调机制结合研究还不够深入, 相关学习参数分析不够细致, 仿真环境和情景现实性不强。未来研究可以引入马尔科夫博弈提高决策协调性, 嵌入混合交通流、公交优先等交通管理思想增强决策实用性, 引入先验知识及其他学习技术加快学习速度, 融入物联网、主动管理、大数据等先进理念和前沿技术增加决策的实时性, 与交通诱导等集成提升决策的系统性。

关键词: Agent; 强化学习; 交通信号; 交叉口; 信号配时

中图分类号: U491.51

文献标识码: A

文章编号: 2095-9931(2017)02-0017-07

Review of Study on Adaptive Traffic Signal Coordinated Timing Decision of Urban Road Network under Multi-Agent Reinforcement Learning

XIA Xin-hai

(Department of Port and Shipping Management, Guangzhou Maritime University, Guangzhou 510725, China)

Abstract: Compared with the approaches of traditional traffic signal timing decision, multi-agent reinforcement learning and its coordination method can better adapt to the variation of traffic environment of urban road network. In order to explore its application in adaptive traffic signal coordinated timing decision of urban road network, the research methods of multi-agent reinforcement learning and its coordination mechanism were systematically summarized, the research status at home and abroad were extensively analyzed, and the existing research problems were put forward. Finally, the directions of future research on this topic were discussed. The study results show that the existing research mainly aims at small scale road network and exists the problem of dimension disaster. The research of combination of reinforcement learning and coordination mechanism isn't deep enough. The relevant learning parameter analysis isn't meticulous enough, and the reality of simulation environment and scene aren't strong

收稿日期: 2017-02-08

基金项目: 广东省自然科学基金项目(2016A030310104); 广东省科技计划项目(2015B010129017)

作者简介: 夏新海(1978—), 男, 湖北团风人, 副教授, 博士, 研究方向为交通运输规划与管理。

E-mail: xiaxinhai@126.com。

enough. The future research can introduce the Markov game to improve coordination, embed traffic management idea such as mixed traffic flow and bus priority to enhance practicability, add the priori knowledge to accelerate the learning speed, combine advanced concepts and cutting-edge technology such as the Internet of Things, active management, big data to increase the real-time performance of decision, and integrate traffic guidance to promote the systematicness of decision.

Key words: Agent; reinforcement learning; traffic signal; intersection; signal timing

0 引言

随着社会经济的发展,城市化和汽车化速度的加快,交通拥挤、交通事故、交通污染、能源短缺等问题已经成为制约我国社会经济全面、协调、可持续发展的重要因素。目前,解决交通问题的方法通常有两种:一是加强基础设施建设,如提高路网的通行能力、扩建桥梁与道路等外延设施;二是提高交通控制和管理水平,采用先进、有效的智能决策方法,合理使用现有交通设施,充分发挥其能力。近年来,随着信息技术的迅速发展,后一种办法受到了人们的普遍重视。而作为城市道路管理的核心部分,城市交通信号配时决策系统是实现城市道路交通流有效运行的重要保障^[1-2]。

(1) 传统的交通信号配时决策系统及方法不能适应实时多变的较大规模路网的交通流特征

传统的交通信号配时决策系统如 SCATS、TRANSYT 和 SCOOT 需要建立复杂的交通模型,其不能充分处理过饱和和交通状态,并且传统的智能交通信号配时决策方法如专家系统、模糊控制、神经网络、进化算法等在一定程度上不能适应实时多变的较大规模路网的交通流特征,如模糊逻辑不能充分表示实时交通的不确定性,更适合孤立交叉口的交通信号配时决策。在大规模路网的交通信号配时决策下神经网络及遗传算法所需计算量大,优化参数难确定。

(2) 多 Agent 强化学习更能适应城市路网交通环境的变化

由于多 Agent 系统与城市交通控制系统在结构上、机理上具有很大的相似性,如自学习、交互式等,因此国内外很多学者将多 Agent 系统及其自学习机制应用在城市自适应交通信号配时决策中。强化学习不需要交通环境的精确模型,在随机交通环境下的自学习闭环最优交通信号配时决策方面具

有很好的潜力,因此多 Agent 强化学习非常适合于解决分布式交通信号配时决策问题,并能适应城市路网交通环境的变化。

(3) 引入协调机制能更有效地进行路网自适应交通信号配时决策

对于显著缓解城市路网的交通拥挤问题,自适应交通信号配时决策有很大的应用潜力,而目前其在中国各大城市的应用程度并不高。由于城市路网中各交叉口处的交通流是相互关联和影响的(特别是在较高饱和度交通条件下),各交叉口控制策略间存在博弈现象。因此为了更有效地进行多交叉口交通信号配时决策,有必要引入协调机制。

因此,融合多 Agent 强化学习和协调机制来进行路网自适应交通信号配时决策研究是多年来的一个热点问题。本文对多 Agent 强化学习及协调机制研究方法分类、国内外研究现状、存在的问题及未来研究方向进行了探讨,以期促进多 Agent 强化学习及其协调机制的研究和其在城市交通信号配时决策中的应用。

1 多 Agent 强化学习及协调机制研究方法分类

国内外一些学者已对多 Agent 强化学习及协调方法进行了较深入的理论研究。从强化学习的角度来看,多 Agent 强化学习研究方法的分类见表1。

表1 多 Agent 强化学习研究方法的分类

分类	问题空间	算法准则
共享多 Agent 强化学习	分布、同构、合作环境	提高学习收敛速度
基于对策的多 Agent 强化学习	同构或异构、合作或竞争环境	理性和收敛性
最佳响应多 Agent 强化学习	异构、竞争环境	收敛性和不遗憾性

其中共享多 Agent 强化学习更多地强调如何利

用分布式强化学习来提高学习速度;基于对策的多Agent强化学习算法必须满足两个性质——理性和收敛性;最佳响应多Agent强化学习方法着重研究在其他Agent采用某种策略的情况下,算法如何获得最优策略。

多Agent协调机制研究方法的分类见表2。其中显式协调即协商,基于对策论的协商研究是多Agent协商研究的主要方法。隐式协调中的社会规则(如交通管理与控制中的公交优先)等可以降低协调的费用,更可以体现多Agent系统的社会性,平衡个体利益与群体利益的矛盾。

表2 多Agent协调机制研究方法分类

一级分类	二级分类	三级分类	四级分类
显式协调	完全集中的协调	-	-
	集中与分布结合的协调	-	-
	完全分布的协调	基于对策论的协调	无通信的协调 基于对策论的协调
		基于观察的协调	-
		其他协调方法	-
隐式协调	社会规则、规范等	-	-
	过滤策略	-	-

2 国外研究现状

2.1 自适应交通信号配时决策研究快速发展

从20世纪60年代起,美国、英国、澳大利亚等国家便开始了自适应交通信号配时决策的相关研究^[3-6],其发展历程见表3。

表3 自适应交通信号配时决策的发展历程

等级	系统	时间
6	超级系统;人工智能技术 (基于经验的自学习能力)	现今
5	具有增强特征的等级4的系统 (OPAC V, RHODES)	1992
4	分布式自适应控制系统 (OPAC III, 基于动态规划法)	1983
3	集中控制,在线优化方案生成 (SCOOT)	1981
2	分布式控制,知识库选择(SCAT) 在线优化方案选择	1979
1	基本系统(TRANSYT) 定时控制及感应式控制的混合控制	1969

2.2 多Agent强化学习在路网自适应交通信号配时决策应用中取得重要进展

国外一些学者将不同的强化学习应用到路网自适应交通信号配时决策研究中,并且取得了一系列重要成果,但对强化学习方法、强化学习要素、动作选择方法等学习参数对路网自适应交通信号配时决策的影响评价的定量精细分析还不够。

(1) 强化学习算法方面,常用到的强化学习算法主要有Q-学习(如Abdulhai等^[9]、Arel等^[14]、Balaji等^[38]、Lu等^[40])、类似Q-学习(如Oliveira等^[11])、修正Q-学习(如Richter等^[13]、Salkham等^[15])、基于模型的Q-学习(如Wiering^[10])、SARSA(λ)(如Thorpe^[39])。

(2) 强化学习要素方面,用到的状态变量通常有两类,一类为车辆数类,如通向交叉口的路径车辆数和位置、排队长度、最近时间步总排队长度的变化;另一类为时间类,如当前周期长度、当前相位持续时间、交叉口总延误或者通向交叉口每一车道的相关延误等。用到的回报函数也通常有两类,一类为车辆数类,如交叉口前等待的车辆数,最近时间步设法通过交叉口的车辆数与仍在等待的车辆数的差值、最近时间步总排队长度的变化等;另一类为时间类,如放行经过路网固定交通量所需的时间、连续决策点间发生的总延误、延误的减少值等。

(3) 动作选择方法方面,常采用 ϵ -Greedy(如Thorpe^[39]、Wiering^[10]、Lu等^[40]、Balaji等^[38]、Arel等^[14]、Oliveira等^[11])和Softmax(如Abdulhai等^[9]、Richter等^[13]、Salkham等^[15])。

对于多Agent强化学习下的自适应交通信号配时决策,国外的研究主要集中在三个方面。

(1) 应用于较小路网的自适应交通信号配时决策

现有多Agent强化学习在主干道的绿波带信号协调配时方面研究较深入,对于较大路网自适应交通信号配时决策还在不断探索之中。Mannion等^[7]提出将启发式预见性建议融入到强化学习中来进行单交叉口的交通信号配时决策,加快了学习收敛速度,减少了延误和排队长度,但对于推广到多交叉口的交通信号协调配时决策中还存在计算复杂性的

问题。Prashanth等^[8]提出了将Q-学习算法与函数估计相结合的交通信号配时决策方法,但采用了基于交通灯的状态描述方法,遭遇了由于状态选取不合理而产生的状态空间爆炸问题,此算法只能用于较小路网而不能应用到较大路网。

(2) 基于完全状态表示强化学习的自适应交通信号配时决策

Abdulhai等^[9]、Wiering^[10]、Medina等^[16]在自适应交通信号配时决策中应用强化学习时需要完全状态表示,但随着交叉口和车道的个数增加,其复杂性呈指数增长,于是就会遭受维数灾难问题,因此在中等大小的网络路网中也很难实现。

(3) 基于独立强化学习的自适应交通信号配时决策

Oliveira等^[11]将强化学习应用扩展到多个交叉口,其能处理由于交通动态性而产生的随机交通模式问题,Mannion等^[12]将并行计算应用于交叉口之间的强化学习,但两者均强调各个交叉口之间进行独立学习。虽然在Medina等^[6]、Richter等^[13]、Arel等^[14]、Salkham等^[15]、Medina等^[16]进行的自适应交通信号配时决策研究中分别考虑了相邻交叉口的状态、相对交通量、延误、拥挤水平等信息,但是这些方法没有包含任何外在协调机制。

目前仅有少数学者如Wiering^[10]、Kuyer等^[17]、Alvarez等^[18]在自适应交通信号配时决策中考虑了交叉口在学习过程中的协调机制。Kuyer等^[17]在Wiering^[10]研究的基础上使用协调图来考虑交叉口交通信号配时决策Agent间的外在协调机制,通过在相连的交叉口交通信号配时决策Agent间发送局部最优信息,利用Max-plus算法来估计最优联合控制方案。Medina^[16]也利用Max-plus算法来寻求全网的交通信号配时协调策略。但是,Max-plus算法对计算要求较高,并且其利用了基于模型的强化学习方法,相对于非模型的强化学习方法(如Q-学习)来说增加了不必要的复杂性,其仅仅适用于树形结构网络,对于一般的循环网络,它不能确保收敛到最优策略。Alvarez等^[18]、Clemptner等^[19]将每一交叉口看作非合作对策,利用马尔科夫决策过程为交叉口交通信号配时决策建模,在每一迭代过程中求

得Stackelberg平衡和Nash平衡,但其研究的是孤立交叉口。

3 国内研究现状

国内一些学者如首艳芳等^[20]、杨兆升^[21]、陆化普等^[22]、刘智勇^[23]进行了有关自适应交通信号配时决策的理论和应用研究,取得了可喜的成果。近年来北京工业大学、清华大学、北京交通大学、上海交通大学等著名高校就多Agent强化学习技术在自适应交通信号配时决策中的应用开展了卓有成效的探索性研究。

3.1 路网自适应交通信号配时决策中强化学习与协调机制结合研究还不够深入

国内学者也已经意识到自适应交通信号配时决策中协调机制研究的重要性,但研究还不够深入。现有交通系统和技术的交通信号配时协调机制通常应用在主干道的绿波带信号配时方面,而在允许交叉口间进行网络范围的协调方面应用还不够。首艳芳等^[20]通过引入群体动力学来进行交叉口群协调控制机制研究,但未结合强化学习研究。湛永荣等^[24]研究了区域信号配时模型的非平衡交通分配算法,采用遗传算法来求解,但遗传算法存在早熟收敛、全局搜索能力不强的问题,从而无法快速找到最佳配时方案。闫飞等^[25]研究了城市区域交通信号迭代学习控制策略,但未引入协调机制。张邻等^[26]建立了基于动态信号配时的非线性规划模型,考虑了各路段机动车的流量特征,但主要针对单交叉口,没有考虑相邻交叉口的关联性。Li等^[36]通过建立深度神经网络来学习强化学习的Q函数,但主要应用在单交叉口的交通信号配时决策中。

3.2 路网交通信号配时决策中博弈思想的应用还处于起步阶段

目前,博弈论在交通领域中的应用大多集中在交通诱导和交通管理方面,而在交通信号配时决策中,博弈思想的应用还处于起步阶段。较多学者在多Agent强化学习中尝试利用博弈论初步探讨了交叉口交通信号配时决策Agent间的协调机制。马寿峰等^[27]、赵晓华等^[28]、朱铭琳等^[29]利用二人非零和合作型对策来进行交叉口间的信号协调控制,但假定交叉口交通信号配时决策Agent的效用值为公共

知识，即完全信息假设；杜荣华等^[30]、李振龙等^[31]应用博弈论的 Nash 平衡理论建立了交叉口之间的交互模型并给出协调算法，但对各个交叉口间的交互约束关系没有进行研究；李建明等^[32]通过引入演化博弈理论和选择机制来对单交叉口交通信号配时决策进行优化，但其研究的是孤立交叉口，没有推广至多交叉口。马赞甫等^[33]利用斗鸡博弈来进行单交叉口的信号配时决策，但其没有引入强化学习，也没有考虑平衡策略的唯一性问题。由于博弈论中 Nash 平衡点具有不唯一性，利用博弈论来研究自适应交通信号配时决策都存在多个均衡控制策略问题，并且没有很好地解决此问题，从而也影响了协调性能。

4 多 Agent 强化学习下的城市路网自适应交通信号协调配时决策研究存在的问题

目前，自适应交通信号协调配时决策中应用多 Agent 强化学习存在如下问题。

4.1 交叉口交通信号配时决策 Agent 间缺乏协调机制，均衡策略不唯一

很多交通拥挤的大城市均具有网格状的拓扑结构，而基于现有交通系统和技术的交通信号配时协调机制通常应用在主干道的绿波带信号配时方面，并局限于保持主要通道间的协调，而允许交叉口间进行网络范围的协调还不够。先前大部分的研究主要考虑独立强化学习，并且存在多个均衡策略问题。虽然少数学者考虑了两级协调，但协调机制不够深入。一些学者虽然引入了博弈论中的 Nash 平衡理论进行协调，但存在多个均衡策略，从而难以达到很好的协调。

4.2 维数灾难问题

虽然目前存在一些基于协调的多 Agent 强化学习方法，但都遭遇维数灾难问题，这是因为状态空间随着 Agent 的数目增加呈指数增长。除了维数灾难问题，这些方法需要每个 Agent 观察整个系统的状态，这在运输网络情况下是不可行的。

4.3 强化学习的参数选择对自适应交通信号配时决策影响的定量评价分析不够深入

虽然强化学习在自适应交通信号配时决策研究

中取得了很大进展，但主要针对强化学习算法本身的要素设计，而缺乏对一些学习参数对自适应交通信号配时决策的性能影响的精细分析，如学习方法、交通状态表示、动作选择方法、交通信号相位方案、回报函数的定义、到达交叉口流量的变化等的定量评价分析。

4.4 缺乏通用的设计框架

在目前自适应交通信号配时决策研究中，各种多 Agent 强化学习因满足各自不同的应用需求而独立发展，但对不同类型的多 Agent 强化学习之间的逻辑关系研究不够深入，没有形成一个统一的框架。

4.5 仿真环境和情景过于简单，不具有现实性

一些研究仅考虑简单仿真环境及假设的交通流，其并不一定能模拟现实的交通网络，缺乏对现实大规模的城市交通仿真网络的验证。

5 城市路网交通信号协调配时决策研究展望

为了提高城市交叉口交通信号配时决策效率，减少车辆延误时间，提高路网通行效率，提出如下研究展望。

5.1 深入分析评价学习参数的影响，提高决策的动态性

对于学习参数对自适应交通信号配时决策方法性能影响进行深入的定量评价分析，如学习方法比较（Q-学习 vs. SARSA vs. TD (λ)）、交通状态表示比较（Queue Length vs. Queues and Arrivals vs. Delay）、动作选择方法比较（ ϵ -Greedy vs. Softmax vs. ϵ -Softmax）、交通信号相位方案比较（可变 vs. 固定）、回报函数的定义比较（Delay vs. Cumulative Delay vs. Balancing Queues）、到达交叉口流量变化比较（Uniform vs. Variable Arrival Rates）等。

5.2 改进研究方法，加强决策的可靠性

（1）嵌入其他交通思想及要素

考虑混合交通流、司机驾驶行为、公交优先等的影响，将行人及非机动车识别功能、公交优先等管理思想、司机驾驶行为影响融入到交通流占有率预测及强化学习算法中^[41-43]。

(2) 提高协调性能

鉴于马尔科夫对策对建立多 Agent 系统协调模型具有较大的优势, 可在此基础上引入博弈协调机制, 加强了交叉口 Agent 间的交互, 考虑路网的整体效益, 避免了陷入局部最优, 并引入深度强化学习求解博弈问题, 提高协调性能^[44]。

(3) 加快强化学习速度和过程

用预先的人工交通管理经验, 结合其他学习技术(如深度强化学习、连接树算法^[45]等)来加快强化学习速度和过程。在许多需要感知高维度原始输入数据和决策控制的任务中, 深度强化学习方法已经取得了实质性的突破, 其深度卷积网络具有自动学习高维输入数据抽象表达的功能, 可以有效解决复杂任务中领域知识表示和获取的难题。

5.3 融入先进理念, 增加决策的现实性

为了实现动态信息的实时采集, 可将低碳交通、车路协同、主动管理、物联网、云计算、大数据等先进理念融入到未来的研究中, 从而对人、车、路、环境等交通要素进行更大范围的整合, 对交通数据进行更加广泛的采集与深入挖掘。

5.4 引入集成, 提升决策的系统性

在系统层面上, 与交通诱导、交通组织等集成, 实现对城市交通系统的最佳管理和控制^[10]。考虑交通参与者(交通管理者、出行者)间的博弈对交通拥堵的影响, 站在交通管理者的角度, 对交通流进行控制和诱导, 减少出行者选择行为的策略集; 针对出行者, 规范其交通行为, 促进出行者朝“合作”的方向发展。

5.5 应用前沿技术, 加大决策的实时性

充分利用车联网、车路协同技术, 未来交叉口可以通过安装在车辆上的智能设备与交叉口的服务器实现无障碍通行。当车辆接近交叉口时, 车辆的速度便由交叉口服务器控制。交叉口服务器通过采集动态车辆信息, 利用精确的计算模型, 可以使所有车辆在不停车的状态下, 安全地通过交叉口。

参考文献

- [1] 李瑞敏. 城市交通信号控制系统发展分析[J]. 综合运输, 2015(7): 104-108.
- [2] HAN K, SUN Y Q, LIU H C, et al. A Bi-Level Model of Dynamic Traffic Signal Control with Continuum Approximation[J]. Transportation Research Part C: Emerging Technologies, 2015, 55: 409-431.
- [3] BALAJI P G, SRINIVASAN D, THAM C K. Coordination in Distributed Multi-Agent System Using Type-2 Fuzzy Decision Systems[C]// IEEE International Conference on Fuzzy Systems. Hong Kong: IEEE, 2008: 2291-2298.
- [4] SRINIVASAN D, CHOY M C, CHEU R L. Neural Networks for Real-Time Traffic Signal Control[J]. IEEE Transactions on Intelligent Transportation Systems, 2006, 7(3): 261-272.
- [5] BAZZAN A L C. Opportunities for Multiagent Systems and Multiagent Reinforcement Learning in Traffic Control[J]. Autonomous Agents and Multi-Agent Systems, 2009, 18(3): 342-375.
- [6] MEDINA J C, HAJBABAIE A, BENEKOHAL R F. Arterial Traffic Control Using Reinforcement Learning Agents and Information from Adjacent Intersections in the State and Reward Structure[C]// 13th International IEEE Conference on Intelligent Transportation Systems. Funchal: IEEE, 2010: 525-530.
- [7] MANNION P, DUGGAN J, HOWLEY E. Learning Traffic Signal Control with Advice[C]// Proceedings of Adaptive and Learning Agents Workshop. Istanbul: Adaptive & Learning Agents Workshop, 2015: 1-8.
- [8] PRASHANTH L A, BHATNAGAR S. Reinforcement Learning With Function Approximation for Traffic Signal Control[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(2): 412-421.
- [9] ABDULHAI B, RINGLER P, KARAKOULAS G J. Reinforcement Learning for True Adaptive Traffic Signal Control[J]. Journal of Transportation Engineering, 2003, 129(3): 278-285.
- [10] WIERING M. Multi-Agent Reinforcement Learning for Traffic Light Control[C]// Proceedings of the 17th International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc, 2000: 1151-1158.
- [11] OLIVEIRA D D, BAZZAN A L C, SILVA B C D, et al. Reinforcement Learning-Based Control of Traffic Lights in Non-Stationary Environments: A Case Study in a Microscopic Simulator[C]// Proceedings of the 4th European Workshop on Multi-Agent Systems. Lisbon: EUMAS' 06, 2006: 1-12.
- [12] MANNION P, DUGGAN J, HOWLEY E. Parallel Reinforcement Learning for Traffic Signal Control[J]. Procedia Computer Science, 2015, 52: 956-961.
- [13] RICHTER S, ABERDEEN D, JIN Y. Natural Actor-Critic for Road Traffic Optimisation[J]. Advances in Neural Infor-

- mation Processing Systems, 2007, 19: 1169–1176.
- [14] AREL I, LIU C, URBANIK T, et al. Reinforcement Learning-Based Multi-Agent System for Network Traffic Signal Control[J]. IET Intelligent Transport Systems, 2010, 4(2): 128–135.
- [15] SALKHAM A, CUNNINGHAM R, GARG A, et al. A Collaborative Reinforcement Learning Approach to Urban Traffic Control Optimization[C]// 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Sydney: IEEE, 2008: 560–566.
- [16] MEDINA J C, BENEKOHAL R F. Q-learning and Approximate Dynamic Programming for Traffic Control: A Case Study for an Oversaturated Network[C]// Transportation Research Board 91st Annual Meeting. Washington D. C.: Transportation Research Board, 2012.
- [17] KUYER L, WHITESON S, BAKKER B, et al. Multiagent Reinforcement Learning for Urban Traffic Control Using Coordination Graphs[C]// Proceedings of Joint European Conference on Machine Learning & Knowledge Discovery in Databases. Antwerp: ECML/PKDD, 2008: 1–16.
- [18] ALVAREZ I, POZNYAK A, MALO A. Urban Traffic Control Problem a Game Theory Approach[C]// Proceedings of the 47th IEEE Conference on Decision and Control. Cancun: IEEE, 2008: 2168–2172.
- [19] CLEMPNER J B, POZNYAK A S. Modeling the Multi-Traffic Signal-Control Synchronization: A Markov Chains Game Theory Approach[J]. Engineering Applications of Artificial Intelligence, 2015, 43: 147–156.
- [20] 首艳芳, 徐建闽. 基于群体动力学的协调控制子区划分[J]. 华南理工大学学报(自然科学版), 2013, 41(4): 77–82.
- [21] 杨兆升. 智能运输系统概论[M]. 北京: 人民交通出版社, 2009.
- [22] 陆化普, 李瑞敏, 朱茵. 智能运输系统概论[M]. 北京: 中国铁道出版社, 2004.
- [23] 刘智勇. 智能交通控制理论及其应用[M]. 北京: 科学出版社, 2003.
- [24] 湛永荣, 黄崇超. 区域信号配时模型的非平衡交通分配算法[J]. 数学杂志, 2014, 34(4): 747–751.
- [25] 闫飞, 田福礼, 史忠科. 城市区域交通信号迭代学习控制策略[J]. 控制与决策, 2015, 30(8): 1411–1416.
- [26] 张邻, 吴伟明, 黄选伟. 基于动态信号配时的非线性规划模型[J]. 公路交通科技, 2014, 31(8): 131–137.
- [27] 马寿峰, 李英, 刘豹. 一种基于agent协调的两路口交通控制方法[J]. 系统工程学报, 2003, 18(3): 272–278.
- [28] 赵晓华, 李振龙, 于泉, 等. 基于Q学习算法的两交叉口信号灯博弈协调控制[J]. 计算机仿真, 2007, 19(18): 4253–4256.
- [29] 朱铭琳, 陈阳舟. 基于博弈的交叉口交通信号配时系统设计[J]. 计算机仿真, 2013, 30(7): 151–155.
- [30] 杜荣华, 吴泉源. 城市区域交通协调控制中的多智能体博弈研究[J]. 计算机工程与科学, 2007, 29(4): 120–123.
- [31] 李振龙, 陈德望. 交通信号区域协调优化的多智能体博弈模型[J]. 公路交通科技, 2004, 21(1): 85–88.
- [32] 李建明, 余春艳. 演化博弈下单点信号灯配时优化研究[J]. 交通运输系统工程与信息, 2012, 12(4): 72–78.
- [33] 马赞甫, 刘妍珺. 基于斗鸡博弈的红绿灯时间设置[J]. 公路交通科技, 2011, 28(11): 126–130.
- [34] LI T, ZHAO D B, YI J Q. Adaptive Dynamic Programming for Multi-Intersections Traffic Signal Intelligent Control [C]// Proceedings of 11th International IEEE Conference on Intelligent Transportation Systems. Beijing: IEEE, 2008: 286–291.
- [35] SHAMSHIRBAND S. A Distributed Approach for Coordination Between Traffic Lights Based on Game Theory[J]. International Arab Journal of Information Technology, 2012, 2(2): 148–153.
- [36] LI L, LV Y S, WANG F Y. Traffic Signal Timing via Deep Reinforcement Learning[J]. IEEE/CAA Journal of Automatica Sinica, 2016, 3(3): 247–254.
- [37] Eduardo Camponogara, Werner Kraus Jr. Distributed Learning Agents in Urban Traffic Control[J]. Lecture Notes in Computer Science, 2003, 26: 324–335.
- [38] BALAJI P G, GERMAN X, SRINIVASAN D. Urban Traffic Signal Control Using Reinforcement Learning Agents[J]. IET Intelligent Transport Systems, 2010, 4(3): 177–188.
- [39] THORPE T. Vehicle Traffic Light Control Using SARSA [R]. Fort Collins: Colorado State University, 1997.
- [40] LU S F, LIU X M, DAI S Q. Q-Learning for Adaptive Traffic Signal Control Based on Delay Minimization Strategy[C]// Proceedings of IEEE International Conference on Networking, Sensing and Control. Sanya: IEEE, 2008: 687–691.
- [41] KHAMIS M A, GOMAA W. Adaptive Multi-Objective Reinforcement Learning with Hybrid Exploration for Traffic Signal Control Based on Cooperative Multi-Agent Framework[J]. Engineering Applications of Artificial Intelligence, 2014, 29(3): 134–151.
- [42] CHANLOHA P, CHINRUNGRUENG J, USAHA W, et al. Traffic Signal Control with Cell Transmission Model Using Reinforcement Learning for Total Delay Minimisation[J]. International Journal of Computers Communications & Control, 2015, 10(5): 627–642.

(下转第30页)

- ment of Origin-Destination Matrices Using Mobile Phone Call Data[J]. Transportation Research Part C: Emerging Technologies, 2014, 40(1): 63-74.
- [10] ROKIB S A, KARIM M A, QIU T Z, et al. Origin-Destination Trip Estimation from Anonymous Cell Phone and Four-square Data[C]// Transportation Research Board 94th Annual Meeting. Washington D.C.: Transportation Research Board, 2015: 1-18.
- [11] CHEN C, BIAN L, MA J. From Traces to Trajectories: How Well Can We Guess Activity Locations from Mobile Phone Traces?[J]. Transportation Research Part C: Emerging Technologies, 2014, 46(46): 326-337.
- [12] PHITHAKKITNUKON S, HORANONT T, LORENZO G D, et al. Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data[C]// Proceedings of the First International Conference on Human Behavior Understanding. Istanbul: HBU 2010, 2010:14-25.
- [13] 李纪华, 刘建玲, 肖瑞. 基于信令数据的交通出行分布异常检测[J]. 移动通信, 2015, 39(21): 17-20.
- [14] 杜翠凤, 余艺. 大数据背景下基于手机信令的道路识别研究[J]. 移动通信, 2015(13): 21-24.
- [15] 张天然. 基于手机信令数据的上海市域职住空间分析[J]. 城市交通, 2016, 14(1): 15-23.
- [16] 龚航, 孙黎, 王璞. 基于手机数据的城市交通大区 OD 分布估计——以旧金山市为例[J]. 城市交通, 2016, 14(1): 37-42.

(上接第23页)

- [43] ABDOOS M, MOZAYANI N, BAZZAN A L C. Hierarchical Control of Traffic Signals Using Q-Learning with Tile Coding[J]. *Applied Intelligence*, 2014, 40(2): 201-213.
- [44] 刘全, 翟建伟, 章宗长, 等. 深度强化学习综述[J]. *计算机学报*, 2017, 40(1): 1-26.
- [45] ZHU F, AZIZ H M A, QIAN X, et al. A Junction-Tree Based Learning Algorithm to Optimize Network Wide Traffic Control: A Coordinated Multi-Agent Framework[J]. *Transportation Research Part C: Emerging Technologies*, 2015, 58: 487-501.