强化学习在城市交通信号灯控制方法中的应用

刘义[1],何均宏[2*]

1. 深圳市公安局交通警察局,深圳 518035
2. 华为技术有限公司,深圳 518080

摘要 悉尼自适应交通控制系统(SCATS)、绿信比-周期-相位差优化技术(SCOOT)及Smooth采用自适应交通信号灯控制方法,对城市道路口的交通信号灯进行了有效控制。随着深圳城市交通流量急剧增长,深圳交警在自主研发Smooth信号控制式基础上,提出实时、分布式、自适应调控要求,联合创新了人工信号控制方案TrafficGo,探索基于深度神经网络的强化学习,通过在线学习各种流量负荷,实时推理计算信控时段、相位、相序、信号周期、绿信比、相位差,进一步优化了交通信号灯的控制模式。介绍了在交通信号灯控制中运用的强化学习模型,实地测评表明,其取得了一定改进效果。

关键词 交通信号控制;强化学习;人工智能;通行效率

深圳作为中国最早建立信号配时日常优化机制的城市,自20世纪80年代开始建立专业团队进行信号控制优化工作,通过设置相位配时方案及一系列控制策略参数保障信号控制效果。随着深圳市信号控制路口数量的快速增长、配时复杂程度不断上升,工作量和工作难度成倍增长,城市交通面临5项挑战:(1) 专业技术应用方面:交通流运行需求超出系统能力,对信号配时或相位设计提出实时更新、路口适配要求;(2) 单点信号控制方面:位置偏远的信控设施缺乏及时有效的数据支撑,方案制定及实施需要到路口信号机现场进行反复调试、观测,信号优化工作难度大,经常被投诉;(3) 协调控制方面:随着协调范围规模扩大,方案需要根据交通流量和路网结构的变化进行快速调整,但由于现有人力、工具的限制,难以及时巡检、仿真和优化,导致协调方案调整滞后,影响了路口通行效率;(4) 设施资源利用方面:仍有大量系统数据需要采集、深入分析、挖掘应用,需要提高各类设施、数据的应用水平;(5) 工作机制方面:需要基于全量数据,提升现有信号控制路口事前预评估和信号控制方案实施后的效果评价能力,建立交通信号配时方案辅助决策机制、平台、服务。

悉尼自适应交通控制系统(Sydney coordinated adaptive traffic system,SCATS)是一种代表性的方案选择式区域协调实时配时控制系统,得到了澳大利亚、中国、美国、爱尔兰、新西兰、菲律宾、墨西哥等几十个国家的几百个大中型城市的广泛应用。SCATS作为早期系统的最大特点是通过对预置方案的优化选择实现低成本信号控制,已经不适合越来越大的交通流量和响

收稿日期:2019-01-14;修回日期:2019-01-29

作者简介:刘义,工程师,研究方向为交通管理,电子信箱:liuyi@stc.gov.cn;何均宏(通信作者),高级工程师,研究方向为人工智能,电子信箱:hejunhong@huawei.com

引用格式:刘义, 何均宏. 强化学习在城市交通信号灯控制方法中的应用[J]. 科技导报, 2019, 37(6): 84-90; doi: 10.3981/j.issn.1000-7857.2019.06.011

应时间要求[1-3]。

绿信比-周期-相位差优化技术(split cycle offset optimizing technique,SCOOT)系统是代表性的方案生成式实时信号控制系统。信号控制方案设计者无需事先准备任何既定配时方案,也不需要事先确定配时参数与交通负荷之间的对应关系;实时优化算法将根据建立好的交通数学模型,利用实时检测到的交通数据进行预测分析,对信号配时作出优化调整[1-3]。

SCATS和SCOOT基本具备了一定自适应控制能力,但在实时、大规模地进行交通流量数据采集、分析、建模、协同、控制方面,面临交通数据、模型迭代、计算能力的瓶颈和挑战[1-3]。

针对上述问题,基于深度学习、强化学习的方法,联合创新交通智能体,应对上述一系列挑战。

1　深圳交通信控技术发展概述

1.1　深圳交通流采集技术发展

1)线圈型传感器。深圳于1989年引进日本ATC(area traffic control)交通信号控制系统时,在各子区的关键路口布设线圈传感器,实现交通数据的采集,支撑系统对"周期/绿信比/相位差"参数的自动选择。由于技术方案的历史局限性,传感器仅实现过车信号的采集,不具备数据的计算能力。

2)线圈型交通数据采集器。2003年Smooth交通信号控制系统研制并开始推广应用,将线圈型传感器升级为线圈型交通数据采集器。不仅可以提供过车"脉冲"信号,而且可以提供1 min、5 min等不同时间间隔的统计交通数据(流量、占有率、速度)。还可以根据信号控制机的起止指令,计算周期级的交通数据,支撑自适应控制策略的实施。

3)电警、卡口的数据复用。随着高清电子警察、卡口设备的布设,此类设备同时具备交通数据采集的功能,为数据的收集和利用提供了新的渠道。

4)视频、地磁、微波车辆检测器。2018年初,交警投入建设的"智慧交通一期"项目中,分别采用了视频、地磁、微波车辆检测器,希望通过应用实践,获得实用、可靠、先进、经济的交通数据采集手段。

1.2　深圳交通信号控制技术发展

1989年,深圳引进了日本ATC交通信号控制系统,控制路口数量约50个。实现了绿波控制、自动方案选择控制、感应控制、多时段控制、步伐保持控制等功能。

到1999年,日本ATC系统的控制半径小、维护难等制约因素逐步显现。针对高饱和度、高复杂度、高期望值的交通管理现状,深圳交警开始组织研发新一代Smooth交通信号控制系统。该系统在充分学习并借鉴国外交通信号控制系统经验的基础上,进行了大胆的技术创新,采用分布式控制模式、三层体系结构、大型数据库、多服务器协同处理、GPRS(general packet radio service)/3G/4G无线通信等技术手段,基于交通状态识别的多目标决策控制策略,形成了较完善的分布式多目标信号控制解决方案。Smooth系统自2003年推广使用以来,一直在深圳交通信号控制系统中占据主导地位,控制路口超过2200个。

自2013年起,深圳市进一步引入竞争机制,开始试用海信HICON(Hisense control)交通信号控制系统,目前控制路口约60个。2017年,深圳建设龙华有轨电车项目时,沿线27个平交路口采用了"华通有轨电车交通信号控制系统"。

1.3　Smooth系统的技术特点

2003年,深圳交警组织研发新一代Smooth交通信号控制系统,成为中国最早大规模推广使用国产交通信号控制系统的城市之一。目前,全市共有信号控制路口2346个,其中联网路口2332个,联网率99.4%,区域协调控制路口达1715个,联网协调率为73.5%。2008—2012年期间,该系统运行自适应控制的路口140余个,取得较好的控制效果。其特点如下。

1)系统功能较为完善,具备绿波控制、子区连接控制、防溢出控制、警卫线路控制、匝道控制、自适应控制、感应控制、行人按钮请求控制、公交优先控制等。

2)体系架构较为灵活,可支撑创新应用,如潮汐车道、借道左转、移位左转、可变车道、多功能灯控、排阵式控制等。

3)支持3G/4G无线通信方式,使系统的联网控制范围不受光缆资源限制,始终保持几乎100%的联网率,为区域协调控制、干线协调控制、远程参数维护提供了有力支持。

系统采用了基于交通状态识别的多目标决策控制策略。在"O-Q"平面上,将交通需求识别为闲散、自由、受控、拥挤、堵塞、队列等6种不同负荷的交通状态(图1)。针对每一种交通状态设定对应的控制目标(表1)。

4)系统采用"战略/战术/队列"检测分布的方案

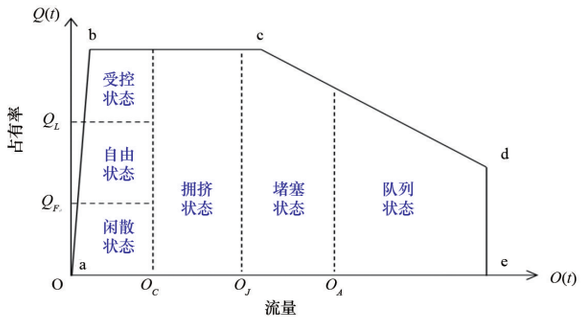

图1 “*O*-*Q*”平面示意

Fig. 1 “*O*-*Q*” plane

表1 交通状态及控制目标

Table 1 Traffic conditions and control objectives

符号	名称	控制目标
Z_1	闲散态特征量	安全通行
Z_2	自由态特征量	延误最小
Z_3	受控态特征量	延误最小
Z_4	拥挤态特征量	通行能力最大
Z_5	堵塞态特征量	通行能力最大
Z_6	队列态特征量	防止溢出

(图2),根据路口的交通需求状况选择合理的布设方案。战略检测器分布于距停止线150~200 m处,解算各入口方向断面的流量、占有率、平均车速、队列到达率及队列驻留时间等数据,用于识别交通状态,优化战略控制参数“周期”“相位差”。战术检测器分布于导向车道入口处,约距停止线线30~50 m,采集左、直、右各流向的流量、占有率及车间时距数据,用于辨识各流向的交通强度,并实现绿信比的战术分配。队列检测器分布于距停车线300 m或更远处,解算队列断面的流量、占有率、平均车速、队列到达率及队列驻留时间等数据,响应特殊的队列冲击或交通事件。

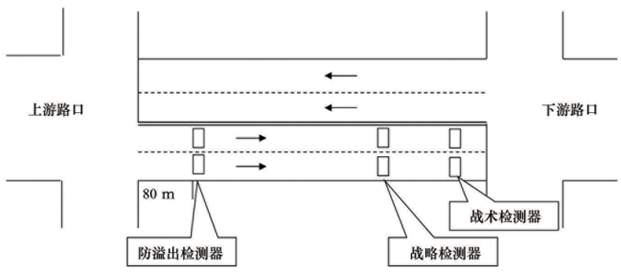

图2 “战略/战术/队列”检测分布方案

Fig. 2 “Strategic/tactical/queue” detection distribution scheme

5) 系统实现了防溢出控制功能。系统将易发生溢出现象的上下游路口组成“防溢出控制单元”,在上游路口的出口处布设“防溢出检测器”,采集占有率、速度参数,识别队列排至溢出警戒线时,启动防溢出预案,通过增加下游绿灯时长、缩短上游绿灯时长、前置相位差等策略防止“车队溢出”现象的发生。

6) 系统实现了公交及特种车辆优先功能。系统通过读取安装在特种车辆上的RFID(radio frequency identification)卡,识别公交、急救等特种车辆。对于公交车辆,采用“绿尾延长”“红灯早断”等策略给予优先;对于急救车辆,采用“安全插入通行相位”等更加迅捷的优先手段。

7) 系统还有丰富的数据分析功能,便于交通工程师从数据深度观测交叉口的运行情况(图3)。

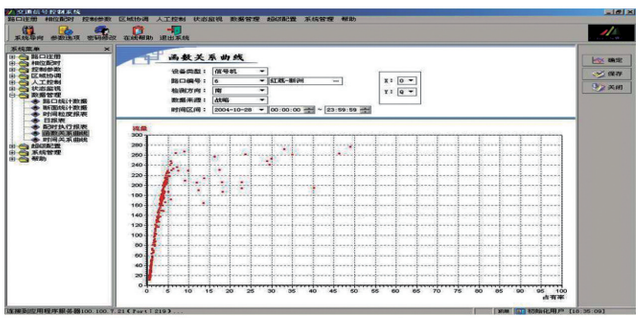

图3 系统数据分析功能

Fig. 3 System function of data analysis

1.4 深圳交通信号控制技术发展方向

近20多年来,国内外城市交通信号控制系统的研究飞速发展,在城市交通控制系统的研究中逐渐重视控制协作与控制方式、系统体系结构及系统控制的优化模型等多个方面。随着大数据、云计算、人工智能等新技术与基础理论的不断发展,城市交通信号控制系统也逐渐向集成化、智能化、多模式化的方向发展,逐步形成具有分布协同功能的集成化城市交通智能控制系统。顺应技术发展趋势,深圳交警联合多方力量成立“深圳信号控制实验室”,通过资源有机融合,推动信号控制向更高效、更智能的方向发展。

2 TrafficGo©:人工智能信号控制联合创新

TrafficGo©为Traffic Go的简称,是华为公司基于多智能体强化学习技术,结合深圳交警-华为公司联合创新需求研发的人工智能(AI)信号控制创新方案品牌名称。

2.1　Smooth对SCATS、SCOOT的发展

不论是KATNET(日本京山公司区域交通信号控制系统)、SCATS还是SCOOT,均开发于20世纪70—80年代,主要面对的是当时非饱和、两相位的交通需求,其控制机理不完全适应中国高饱和度、高混合度、高复杂度的交通需求特征。这些系统均采用了集中式控制模式,一旦中央计算机或通信链路发生故障,信号机即失去优化的参数、指令来源,只能降级为缺乏实时数据支持的无电缆协调、简单的感应控制或定时控制。

Smooth系统继承了KATNET系统识别交通状态的方法,汲取了SCOOT系统"临近预测"的策略,引入了SCATS系统战术微调的手段,针对中国的交通现状和发展趋势,提出了基于交通状态识别下的多目标决策控制策略。系统采用了成熟的分布式控制模式,系统的控制策略、控制功能由"中央控制管理系统-信号控制机-车辆检测器"分布实施、协同处理完成,可靠性、灵活性更高;Smooth系统采用了先进的3层体系结构,由"数据库主机群-应用服务器群-客户机"组成,满足可扩展性、开放性、安全性等要求。

动态优化控制:信号机具备连接地埋式感应线圈检测器实现本路口自适应控制的功能,无须依赖于中心系统,可以实时处理本路口检测器的交通数据,并实时调整相应的交通参数,以适应交通流时变的要求。在单点控制策略中,信号机构造了单路口动态优化控制的功能。动态优化建立在平交路口O-Q-V(流量-占有率-平均速度)交通模型的基础上,要求为各受控相位配置战略检测器。信号机采集检测器自主解算的二次交通数据,对信号控制周期和绿信比进行合理配平和适度调节,以期达到安全行驶、路口通行能力最大或拥挤度最小。动态优化控制是单路口非常有效的一种控制模式。

相位差协调控制:建立中心系统后,信号机接收中心系统的相位差协调指令,实现侧重单向绿波、干线相位差协调及区域相位差协调控制。通过有线或无线通信方式和中央控制系统建立通信链路,定时将车辆检测器解算的二次交通数据、根据动态优化控制算法决策的单点控制参数(周期、绿信比)、设备的运行状态等上报中央系统,接收中央系统根据区域协调控制算法调节的控制参数(周期-绿信比-相位差),实现大区域的信号协调控制。

远程指令控制:建立中心系统后,信号机可以接收中心系统的指令实现黄闪、灭灯、灯色保持、临时方案等中央特殊控制功能。

交通拥挤度发布:建立中心系统后,信号机接收中心系统的数据,向动态诱导路牌发布相邻路口、路段的交通状态信息。

路口参数设置及状态显示:用户可以通过菜单选择输入路口的形状、是否有渠化岛等几何参数,以路口图形为背景,监视信号灯色的运行状况,同时提示信号周期长、当前相位及当前步伐等信息。

2.2　基于交通视频深度学习的流量采集

交通视频图像分析技术,早期是通过视频虚拟线圈感知流量、违法行为,闯红灯也是通过触发线圈后摄像机取证。近2年随着视频图像分析技术的发展,人工智能(artificial intelligence,AI)在视频目标分类、轨迹跟踪方面的准确率越来越高,相较于传统的违法检测技术,视频单一技术不需要埋设线圈、免去了土木工程作业,可以复用现有的摄像头,无缝添加检测点。从而有摄像头的地方,就能感知流量和事件,最大化摄像头的传感器功能。

基于深度神经网络的发展而产生的深度学习(deep learning),是机器学习的一个分支,它主要基于多层的神经网络结构和方法快速训练模型,在收敛方面更加可靠[4-6]。

一个神经网络就是一个带有一系列参数的机器学习模型,模型的输入是一个M维的向量$\boldsymbol{x}$,$\boldsymbol{x}$通过一系列的隐藏层和激活函数,最后形成一个K维的输出向量$\boldsymbol{y}$(图4)。一个神经网络包含全链接层,每一层会在输入$\boldsymbol{x}$和权重$\boldsymbol{w}$之间计算一个线性的映射,并且添加一个偏置项b,通过一个非线性的激活函数和结果映射[4-6]。例如,一个向量$\boldsymbol{x}$通过一个权重为$W_0\in\theta$的单层隐藏层,偏置项为$b_0\in\theta$,非线性误差为h_0的网络,结果为

$$\vec{x}=h_0\left(W_0\vec{x}+b_0\right) \tag{3}$$

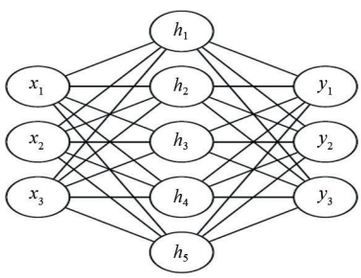

图4　一个简单的神经网络模型

Fig. 4　A simple neural network model

基于视频进行深度学习的流量采集的关键在于目标分类,将交通的各类参与元素区分开来,交通目标主要是3类:行人、机动车和非机动车。目前,白天和有路灯照明的分类采集准确率约为98%。

2.3 信号周期方案优化和马尔科夫决策过程

SCATS的信号周期的选择通常是以控制子区为单位进行,即在一个控制子区之内,所有交叉口共用同一个信号周期长度。控制子区信号周期长度的选择,主要将依据子区实时交通量指数的大小而定。

设计者可以将子区的交通量指数排列成一系列级别,每一个级别对应于一个确定的信号周期长度,使任何一种交通量状况都有一个信号周期长度与之对应。

通常需要为整个控制区域规定15个不同大小的信号周期长度,考虑到一个控制子区所包含的范围有限,交通状况的差异性不会很大,因此,为各控制子区预先规定一个较小的参数选择范围即可。一般情况下,每个控制子区可以从15个信号周期长度取值中挑选出5个作为其信号周期长度的选择范围。

优化算法也可以选择饱和度或类饱和度作为确定控制子区信号周期长度的主要依据:(1) 当区域内饱和度很高时,控制子区应选取最大的信号周期长度;(2) 当区域内饱和度适中时,控制子区可选取适中的信号周期长度;(3) 当区域内饱和度很低时,控制子区则可选取最小的信号周期长度。

SCTAS的理论假设先进,但系统假设受限于20世纪70年代末的计算系统能力,是一种马尔科夫的低成本初级实现。Smooth系统继承了SCATS、KATNET系统识别交通状态的方法,汲取了SCOOT系统"临近预测"的策略,提出了基于交通状态识别下的多目标决策控制策略。本研究组考虑进一步发挥马尔科夫决策过程(Markov decision process,MDP)能力,进行周期计算优化。因为交通信号周期优化是针对某个特定环境规定的,目标是让智能体达到某种期望。基于MDP定义了交通智能体一系列环境状态、所采取的一系列动作(actions),为在某些状态而采取某些行动所分配的回报函数(reward function),描述环境状态变化的转换函数(transition function)[7-9]。在交通智能体转换函数仅依赖于当前状态s和动作a时,就满足了马尔科夫特性[10]。在动作a仅依赖于当前状态时,概率值从s转换为s',即

$$P\left(s_{t+1}\middle|s_t,a_t,r_t,s_{t-1},a_{t-1},\cdots,r_1,s_0,a_0\right)=P\left(s_{t+1}\middle|s_t,a_t\right) \quad (1)$$

通常,一个MDP是一个四元组<S,A,R,T>,其中S是可能的状态空间;A是可能的动作空间;$R^a_{ss'}$代表一个回报函数,确定在状态s时采取行动a,并在状态s'结束行动的回报r;$T^a_{ss'}$代表一个转换函数,确定在状态s时采取行动a,并在状态s'结束行动的概率值。

SCATS预置15种模型,而本研究组AI信号控制创新方案TrafficGo交通智能体基于模型(model-based)的强化学习,可参考这些模型进行学习;SCOOT不预置模型,而TrafficGo交通智能体也可采用无模型(model-free)强化学习方法,其智能体同时跳跃T和R,通过大规模、分布式的连续计算估计T和R,寻找一个最优化的策略。

2.4 绿信比优化和扁平的Q学习

SCOOT信号控制方案,是一种方案生成式实时优化算法设计。设计者一般无需事先准备任何既定的配时方案,也不需要事先确定配时参数与交通负荷之间的对应关系。SCOOT制式通过方案生成式实时优化算法,将根据建立好的交通数学模型,利用实时检测到的交通数据进行预测分析,对信号周期、绿信比、相位差等配时参数作出优化调整,最后生成较为精确的区域信号协调配时方案。

SCOOT各交叉口绿信比优化调整的基本原则可以设置为:确保交叉口的总饱和度最小,即关键车流的总延误时间最少。与此同时,车辆受阻排队长度、道路拥挤程度及相位最短绿灯时间限制等,也都应作为进行绿信比优化时必须予以考虑的因素。各交叉口各信号相位绿信比的优化具有相对独立性,即无需考虑交叉口之间的关联性,与其他两项配时参数(公共信号周期时长和相位差)的优化调整相比,绿信比的优化调整是最为频繁的。

SCOOT在相邻的两个信号周期内,考虑到实时交通量经常出现随机波动的现象,为了使绿信比的调整不致于受到这种随机波动的干扰,维持一种平稳的变化,适宜根据实时交通负荷程度,采用上述"连续微量调整法"对绿信比进行实时优化调整。

Smooth吸收发扬了SCOOT、SCATS的绿信比算法,在数据采集和处理的基础上,将交通需求识别为闲散、自由、受控、拥挤、堵塞、队列6种不同负荷的交通状态,在流量-占有率平面上,对应不同的目标区域,进一步将状态区域平面映射为周期-状态平面,为不同的状态区域建立周期优化函数及其边界约束。根据前周期与交通需求的适应性,函数曲线可沿其值域边界滑动或

上下平移,从而优化生成预执行的周期参数。

本研究组采用Q学习(Q-learning)来进一步提升绿信比计算效率,这是一种无模型的强化学习算法。它不需要根据环境的转换和回报函数来建立自己的模型,而是直接估计在状态s下采取一个行动a的值[10-13]。在传统的Q学习中,智能体对状态s和a对采用一个查找表,并使用如下目标函数直接对Q值估计进行更新:

$$Q_{t+1}(s,a)=Q_t(s,a)+a\left[r_t+\gamma\left[\max_{a'}Q_t(s_{t+1},a';\theta_t)\right]-Q_t(s,a)\right] \quad (2)$$

对当前状态s和a对的估计和实际值之间存在差异,智能体利用当前的回馈信号和最大化下一个状态的Q值来作为对真实值的替代,如此循环迭代,计算出具有最优交通质量目标函数值的绿信比[13-16]。

3　实验结果测评

3.1　实验试点情况概述

本次试验的试点为深圳市坂田片区,其中包括五和张衡、五和稼先、五和隆平、五和贝尔、冲之张衡、冲之稼先、冲之隆平、冲之贝尔路口,共8个路口。

3.2　实验评价方式

1) 实验评价指标主要是第三方厂商(世纪高通)提供的浮动车数据,主要根据对浮动车的GPS(global position system)定位坐标进行跟踪,通过"车辆进入某路段的GPS时刻-车辆离开某路段的GPS时刻"获得单个车辆的行车时间;通过"车辆进入某路段的GPS位置"和"车辆离开某路段的GPS位置"之间的距离获得单个车辆的通行距离;通过"单车通行距离/单车通行时间"就可以获得单车通行速度。车辆的平均速度则由"所有通行车辆的单车通行速度之和/采样数"得到。车辆的平均通行时间由单位时间内"某路段的路长/采样数"得到。车均行驶延误由单位时间内"车辆平均通行时间-凌晨时段的车辆平均通行时间"得到。

2) 试验的统计区域为坂田片区8个路口及周边相邻的路段(图5)。通过统计试点路口及其周边路口的车辆通行效率,用以验证TrafficGo对周边道路的优化辐射作用。

3.3　信号配时实验结果

通过实验发现,试点后的平均车速有明显上升。经过统计所有路段,相比调控前,调控后的总体平均车速上升约4.2%。

图5　本次试点的统计区域

Fig. 5　Statistical area of this pilot

通过实验发现,试点后的平均通行时间有明显下降。经过统计所有路段,相比调控前,调控后的总体平均通行时间下降约5.1%。

通过实验发现,试点后的坂田片区整体车均行驶延误有明显下降。经过统计片区内所有路段,相比调控前,调控后的样本车均行驶延误最高下降约17.7%。

4　结论

深圳交警在原有传统信号控制基础上,结合近年来视频采集、运算、识别算法能力不断提升的技术背景,从视频结构化和深度学习的维度重新审视信号控制领域,并运用TrafficGo人工智能信号优化调度算法模型,具体落地深圳多个路口进行应用。从初步实验结果来看,具有进一步推广试用和深度评估的价值。未来,随着强化学习及人工智能技术的不断发展,交通信号灯控制算法也将不断得以改进,城市交通会因为新技术的引入而变得更加畅通。

参考文献(References)

[1] 陆化普. 大数据及其在城市智能交通系统中的应用综述[J]. 交通运输系统工程与信息, 2015(10): 45-51.

Lu Huapu. Big data and its applications in urban intelligent transportation system[J]. Journal of Transportation Systems Engineering and Information Technology, 2015(10): 45-51.

[2] 杨文臣, 张轮, Zhu Feng. 多智能体强化学习在城市交通网络信号控制方法中的应用综述[J]. 计算机应用研究, 2018, 35(6): 101-114.
Yang Wenchen, Zhang Lun, Zhu Feng. Multi-agent reinforcement learning based traffic signal control for integrated urban network: Survey of state of art[J]. Application Research of Computers, 2018, 35(6): 101-114.

[3] Li L, Lv Y S, Wang F Y. Traffic signal timing via deep reinforcement learning[J]. Acta Automatica Sinica, 2016, 3(3): 247-254.

[4] Hamilton A, Waterson B, Cherrett T, et al. The evolution of urban traffic control: Changing policy and technology[J]. Transportation Planning & Technology, 2013, 36(1): 24-43.

[5] Zhang J, Wang F Y, Wang K, et al. Data-driven intelligent transportation systems: A survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(4): 1624-1639.

[6] Wu X, Liu H X. Using high-resolution event-based data for traffic modeling and control: An overview[J]. Transportation Research Part C, 2014, 42(2): 28-43.

[7] Yau K L A, Qadir J, Khoo H L, et al. A Survey on reinforcement learning models and algorithms for traffic signal control [J]. ACM Computing Surveys, 2017, 50(3): 1-38.

[8] Azimirad E, Pariz N, Sistani M B N. A novel fuzzy model and control of single intersection at urban traffic network[J]. IEEE Systems Journal, 2010, 4(1): 107-111.

[9] 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.

[10] Sutton R S, Barto A G. Reinforcement learning: An introduction[J]. IEEE Transactions on Neural Networks, 1998, 9(5): 1054.

[11] Watkins C J C H, Dayan P. Q-learning[J]. Machine Learning, 1992, 8(3/4): 279-292.

[12] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[13] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518 (7540): 529-533.

[14] Genders W, Razavi S. Using a deep reinforcement learning agent for traffic signal control[J]. arXiv preprint, 2016, arXiv: 1611.01142.

[15] Tran D, Toulis P, Airoldi E M. Stochastic gradient descent methods for estimation with large data sets[J]. arXiv preprint, 2015, arXiv: 1509.06459.

[16] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning[J]. arXiv preprint, 2016, arXiv:1509.02971.

A survey of the application of reinforcement learning in urban traffic signal control methods

LIU Yi[1], HE Junhong[2*]

1. Shenzhen Traffic Police, Shenzhen 518035, China
2. Huawei Technologies Co., Ltd., Shenzhen 518080, China

Abstract The adaptive traffic signal control method is adopted to effectively control the traffic lights at the urban road junctions, with the rapid growth of the traffic flow in Shenzhen. Shenzhen traffic police asked for a real-time, distributed and adaptive control on the basis of the self-developed smooth signal control. Joint innovation has developed the reinforcement learning based on the deep neural network. Through online learning of various traffic loads, and the real-time reasoning, the information control period, phase, phase sequence, signal cycle, split and phase difference are calculated. This paper reviews the reinforcement learning model used in the traffic signal control, and makes an evaluation on the spot.

Keywords traffic signal control; reinforcement learning; artificial intelligence; pass efficiency

（责任编辑　王志敏）