

基于自适应动态规划的城市交通信号优化控制方法综述

赵冬斌¹ 刘德荣¹ 易建强¹

摘要 针对日益受到重视的自适应动态规划 (Adaptive dynamic programming, ADP) 的方法和应用进行了概述. 详细分析了自适应动态规划的关键问题, 如收敛性、稳定性和协调性的研究现状和趋势. 介绍了城市交通信号控制问题的特点和目前采用的控制方法, 以及自适应动态规划方法在街区路口和快速路入口匝道的交通信号优化控制的应用现状和前景.

关键词 自适应动态规划, 收敛性, 稳定性, 协调性, 交通信号控制

中图分类号 TP274

An Overview on the Adaptive Dynamic Programming Based Urban City Traffic Signal Optimal Control

ZHAO Dong-Bin¹ LIU De-Rong¹ YI Jian-Qiang¹

Abstract This paper surveys the algorithms and application of a hot spot adaptive dynamic programming (ADP). Some key research issues, including convergence, stability, and coordination of adaptive dynamic programming, are extensively analyzed. The problems of urban city traffic signal control and current control schemes are introduced, as well as the present and potential applications of adaptive dynamic programming in the optimization of traffic signal control both in surface way intersections and freeway ramp metering systems.

Key words Adaptive dynamic programming (ADP), convergence, stability, coordination, traffic signal control

自适应动态规划 (Adaptive dynamic programming, ADP) 本质上基于强化学习原理, 模拟人通过环境反馈进行学习, 近年来被认为是一种非常接近人脑智能的方法^[1]. 1977 年, Werbos^[2] 首次提出自适应动态规划的思想, 命名为 ACD (Adaptive critic design). 其主要原理基于人工神经网络能以任意精度逼近非线性函数的特性, 通过单步计算估计动态规划一段时间序列的性能指标函数, 有效地解决了动态规划计算“维数灾”的难题, 为高维复杂系统的最优控制提供了一种切实可行的理论和方法. 随后受到人们的广泛重视, 也产生了很多同义名称, 如 Approximate dynamic programming^[3]、Asymptotic dynamic programming^[4]、Neuro-dynamic programming^[5] 等. 2006 年美国科学基金会组织的“2006 NSF Workshop and Outreach Tutorials on Approximate Dynamic Programming”研讨会上, 建议将该方法统一为 Adaptive/Approximate dynamic programming. 自适应动态规划是人工神经网络、最优

控制和强化学习 (Reinforcement learning) 相融合产生的交叉领域, 也可以认为是离散领域的强化学习在连续领域的扩展, Bertsekas 等^[5] 则将其定义为现代版的强化学习. 自适应动态规划在各种复杂控制领域得到了广泛应用, 包括飞机的自动降落控制^[6]、多个电梯的调度控制^[7]、无线通讯网络的控制^[8]、智能交通系统的控制与管理^[9] 等.

另一方面, 随着近年来交通需求的快速增长, 尽管道路越建越多, 可交通却越来越堵, 交通拥堵似乎成了城市交通的典型缩影. 除去交通规划与设计不合理、公众交通意识淡薄等方面的因素外, 另一方面的主要原因在于现有的城市交通信号控制系统没有充分发挥合理的交通指挥和疏导作用. 目前国内城市所采用的城市交通信号控制系统多为集中式控制系统, 如北京采用的 SCOOT 系统, 可以实现一个区域内交通信号的协调控制, 但仍然存在单路口交通信号控制对交通流变化的适应性差、交通分区不合理和多路口交通信号协调性差等问题. 可以预见, 随着交通设施的快速发展和人们对交通需求的不断提高, 城市交通信号控制的重要性将日益突出. 由于交通信号控制系统本质上是分布式控制系统, 控制系统结构趋于采用分层递阶结构, 将每个交通路口作为一个代理 (Agent), 通过分布式多代理系统 (Multi-agent systems, MAS) 的协调优化理论和方法来实现区域交通信号的协调优化调节. 无论是单个代理的性能优化, 还是 MAS 的协调优化, 强化学习方法的应用研究逐渐增多. 这是由于通常情况下,

收稿日期 2008-12-18 收修改稿日期 2009-02-25
Received December 18, 2008; in revised form February 25, 2009
国家自然科学基金 (60874043, 60621001) 资助
Supported by National Natural Science Foundation of China (60874043, 60621001)

1. 中国科学院自动化研究所复杂系统与智能科学重点实验室, 北京 100190

1. Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

DOI: 10.3724/SP.J.1004.2009.00676

代理的优化和协调无法获得导师信号, 只能从环境中得到反馈信息, 通过不断强化改善性能. 强化学习中研究成果最多的是 Q 学习^[10] 方法, 而自适应动态规划则是一种更普遍意义的强化学习方法, 在交通信号控制与管理领域也日益受到重视.

因此, 本文希望通过对自适应动态规划的定义、研究现状、关键问题, 以及在交通信号优化控制中的应用现状和前景的介绍, 促进自适应动态规划的方法研究及其在城市交通信号控制中的应用.

1 自适应动态规划

1.1 定义

自适应动态规划通常由三个模块组成: 评价模块 (Critic), 模型模块 (Model) 及控制模块 (Action), 如图 1 所示. 模型模块和控制模块就是传统的系统模型和控制器. 评价模块的输出为性能指标函数 $J(t)$, 训练误差定义为

$$e_c(t) = \gamma J(t) - J(t-1) + r(t) \quad (1)$$

其中, γ ($0 < \gamma < 1$) 是折扣因子, r 是回报函数 (Reinforcement signal), 用来评价控制作用的好坏. 假设评价模块的误差趋于零, 可以递推得到

$$J(t) = \sum_{k=t+1}^{\infty} \gamma^{k-t-1} r(t) \quad (2)$$

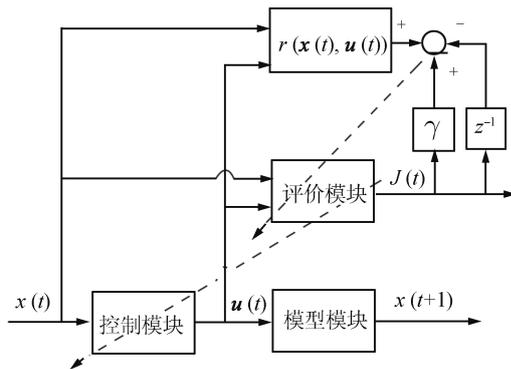


图 1 自适应动态规划的结构和训练示意图
(实线为信号流, 虚线为参数训练流)

Fig. 1 Schematic diagram of the structure and training of ADP (The solid lines represent signal flow, while the dashed lines are the paths for parameter tuning.)

式 (2) 与动态规划的性能指标函数的定义相同, 因此, 评价模块可以用来定量指导控制模块的优化. 即通过改变控制模块的参数, 使评价模块的输出达到最大值, 所得到的控制模块能给出最优或近最优的控制信号. 根据评价模块输出的不同, 可以将自适应动态规划分为以下三类^[11]: 基于性能指标函数误差的启发式动态规划 (Heuristic dynamic

programming, HDP)、基于误差导数的双重启发式动态规划 (Dual heuristic programming, DHP) 以及基于误差和误差导数结合的全局双重启发式动态规划 (Globalized dual heuristic programming, GDHP). 若控制模块的输出作为评价模块的输入, 又衍生出控制倚赖 (Action dependent, AD) 的三种方法, 如 ADHDP、ADDHP 和 ADGDHP.

1.2 收敛性

自适应动态规划应用的关键问题之一在于评价模块和控制模块训练时的收敛性, 将直接影响对控制模块的优化性能. 由式 (1) 和 (2) 及神经网络特性可知, 收敛性受到回报函数 r 和折扣因子 γ 等参数、关于系统的先验知识、训练方法和策略等的影响.

回报函数 r 和折扣因子 γ 是需要人为设置的变量, 其对收敛性的影响是至关重要的, 但对收敛性的作用还不是十分清楚, 有采用神经网络进行逼近的方法^[12]. 通常定义回报函数 r 为二值离散变量, 即 0 和 1. 一些比较研究表明, 如取 r 为系统状态与理想状态之间偏差的绝对值或平方项, 可以加快算法的收敛性. 折扣因子 γ 则决定了未来的回报函数在性能指标中的作用. 由于通常情况下评价模块和控制模块的训练是逐步收敛的, 因此在训练初期, 评价模块逼近性能指标的误差很大, 应尽量减少折扣因子的取值, 以减少所带来的误差干扰; 而在训练后期, 可以逐渐增加折扣因子的取值, 为此可以尝试引入模拟退火等方法, 对折扣因子的取值自适应调整. 由于 DHP 和 GDHP 的评价模块的输出采用了性能指标函数对状态变量求导的一组向量, 因而在学习效率等方面优于 HDP (其离散形式为 Q 学习) 和动态规划等方法^[13].

控制系统的先验知识可以用来指导评价模块和控制模块的设计, 有利于提高收敛速度. 可以采用模糊系统的方法利用模糊规则提供一些专家知识^[14], 提高整个系统的收敛性; 也可以利用训练过程的历史数据, 借鉴时间差分 TD(λ)^[15] 方法, 构造出自适应动态规划资格迹的方法 ADHDP(λ)^[16], 该方法在标准测试平台 Pendubot 的倒立平衡控制问题上的收敛性和收敛速度明显优于 ADHDP.

人工神经网络的训练通常采用梯度下降误差反传的方法, 因此通常 (ADHDP 除外) 需要系统模型, 对收敛速度和误差有一定影响; 另一方面, 误差反传要求控制模块属于前向网络结构, 在一定程度上限制了自适应动态规划在不同结构控制器上的应用. 神经网络、模糊控制等的训练也可以采用日益成熟的演化算法, 如遗传算法、粒子群算法、蚁群算法等. 与梯度下降方法相比, 演化算法采用启发式搜

索的方式, 因此具有无需系统模型、不局限于控制器结构、训练效率高、收敛性好等特点^[17].

1.3 稳定性

自适应动态规划的稳定性是另外一个关键问题, 是指通过学习得到的控制模块与控制对象所构成的系统是稳定的. 自适应动态规划的优势在于对所控制系统无先验知识的前提下, 通过训练可以得到一个有效的控制器, 显示出这种方法强大的学习能力. 但是, 通常采用前向神经网络构造的控制器, 很难证明控制系统的稳定性, 因而限制了该方法或控制模块在实际系统中的应用. 而 Lendaris 等^[14] 则提出了一种完全不基于前向神经网络的方法, 这为研究人员提供了更广阔的思路: 控制模块的设计采用模糊系统或其他控制方法, 引入相关方法的稳定性证明成果, 如采用输入变量和控制变量线性表示模糊规则后件的 T-S 模糊方法、径向基神经网络方法^[18]、递归神经网络^[19] 等.

1.4 协调性

分布式控制系统的协调优化既是自适应动态规划领域的研究难题, 又是典型的应用问题. 通常将每个控制器看成是一个代理 (Agent), 利用 MAS 的求解方法如博弈论 (Game theory, GT) 来实现协调优化. 在传统博弈论中, 每个 Agent 都是一个理性的参赛者, 通过选择合适的策略使回报达到一个平衡状态, 称为纳什均衡 (Nash equilibrium, NE)^[20]. 然而, 静态纳什均衡很难反映实际系统的动态特性, 其对参赛者的理性要求也过于严格. 演化博弈论 (Evolutionary game theory, EGT) 则将上面的条件放松, 不要求每个参赛者了解其他参赛者的决策, 他会有多种决策, 但不知道哪一种是最优的, 通过不断地迭代学习来优化自己的行为并获得最多的回报. 演化稳定策略 (Evolutionary stable strategies, ESS) 是演化博弈论的动态均衡点, 已证明获得演化稳定策略的条件要比纳什均衡更为严格, 为演化博弈论的收敛性提供了足够的理论支持^[21]. 由于演化博弈论描述了动态系统的演变过程, 因此, 很适合于实现 MAS 的动态协调优化^[22]. 近来的理论证明, 两个 Agent 的连续时间的交互 Q 学习的动力学特性, 与演化博弈论的动力学特性是一致的^[23], 为自适应动态规划在分布式控制系统协调优化中的应用奠定了有利的基础.

2 城市交通信号的优化控制

近年来, 城市交通需求的快速增长, 产生了一系列问题, 如交通拥堵、交通延误、环境污染和交通事故等. 甚至出现“道路越建越多, 车辆越来越多”的奇怪现象. 关键问题之一在于现有的城市交

通信号控制系统没有充分发挥合理的交通指挥和疏导作用. 据统计, 在美国 Arizona 州 Phoenix 市的某条道路上实施先进交通信号控制系统后, 碰撞事故减少 6.7%, 车辆行程时间减少 11.4%, 延误降低 24.9%, 停车数减少 27%, 能源消耗显著减少, 突出交通信号优化控制的重要性.

城市交通信号控制方法可以分为: 定周期、感应和自适应控制方法. 定周期控制方法采用预先设定的周期和绿信比 (绿灯时间在信号控制周期中所占的比例) 的配时方案; 感应控制方法是在交叉口进口道上设置车辆检测器, 信号配时可根据检测到的车流信息而随时改变; 自适应控制方法则把交通系统看作一个不确定系统, 通过测量状态量, 如车流量、停车次数、延误时间和排队长度等, 进行反馈实现信号配时的动态优化调整^[24]. 目前, 在全球范围内广泛应用的交通信号控制系统包括澳大利亚的 SCATS 系统、英国的 SCOOT 系统等. SCATS 系统属于方案选择式控制系统, 每个交叉口配时方案根据子系统的整体需要进行选择, 上海运行着该系统. 而 SCOOT 系统属于方案生成式实时自适应控制系统, 采用小步长渐进寻优的方法, 连续实时地调整绿信比、周期和时差三个参数, 北京引进了该系统. 该系统的不足之处是采用集中式控制结构, 难以实现较大区域的控制. 上世纪 80 年代后期开发的交通信号控制系统, 如意大利的 SPOT/UTOPIA 系统、美国的 RHODES 系统、美国的 OPAC 系统、法国的 PROLYN 系统等, 多采用分布式分层递阶的控制结构, 以方便构成区域控制系统^[25].

当代城市交通信号控制系统既包括受交叉口信号灯调节的街区交通系统, 也包括市内快速路、并通过出入口匝道与街区路网耦合在一起构成的城市交通路网的复杂的非线性大系统. 十字路口与快速路入口匝道的交通信号控制是最基本的控制单元, 如图 2 (见下页) 所示. 下面介绍一些比较重要的交通信号优化控制的智能方法. 模糊控制是研究较早而成果较多的一种智能控制方法, Pappis 等^[26] 于 1977 年提出采用模糊控制实现单路口两相位的信号控制, Trabia 等^[27] 提出了两层模糊控制的方法, Lee 等^[28] 将这种方法扩展, 设计出更复杂的调整相序和绿信比的路口间协调模糊控制方法. 为实现模糊控制的最优化或次优化调节, Bingham^[29] 将单路口两相位的模糊控制器以模糊神经网络形式表示, 通过构造另外一个评价神经网络进行优化调节, 该思路采用的就是自适应动态规划的思想. 近年来, 采用这种思路来实现交通信号的优化控制的研究日益增多. 这是由于控制器的优化训练无法获得导师信号, 只能从环境中得到反馈信息, 通过不断学习改善性能. 在这些方法中, 离散状态和动作的 Q 学习应

用得最多^[30-31]. 但也存在系统维数的问题, 尤其对于多相位、交通流大范围变化的优化控制问题, Q 值矩阵存储空间要求异常庞大, 也产生“维数灾”, 使得该方法在实际系统的应用受到限制. 为此, 近来提出了针对连续域问题采用函数逼近的自适应动态规划的求解方法, 在单路口交通信号的优化控制和快速路入口匝道的信号控制方面得到初步应用^[16, 32-33].

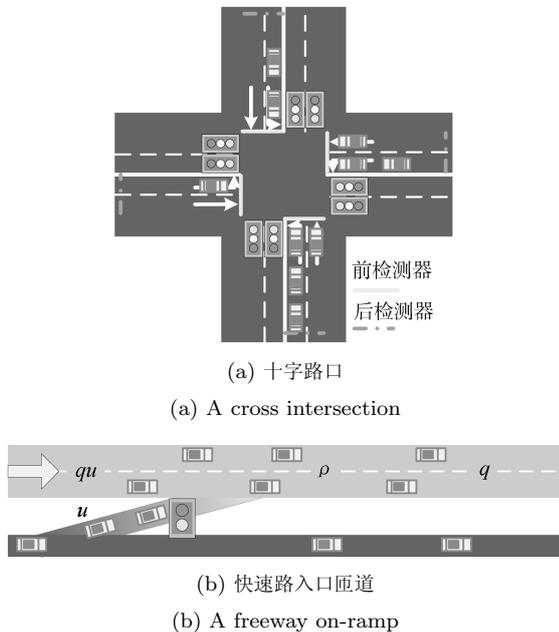


图2 十字路口和快速路入口匝道交通信号控制示意图
Fig.2 Schematic diagrams of traffic signal control for cross intersections and expressway on-ramps

对于交通干线上多路口的协调控制问题, 如图3(a)所示, 目前的主要方法是采用“绿波带”^[24]概念: 即车辆以一定速度通过干线上的交通路口时所遇到的都是绿灯, 其本质上属于集中式控制系统, 难以适应交通流的变化, 而且通常以牺牲非主要路口和非主要干线的性能指标为代价. 对于快速路多个入口匝道协调保持快速路主路的交通流速的控制问题, 传统的方法是在单入口匝道信号控制器ALINEA的基础上进行构造, 本质上属于线性控制器, 适用于交通流变化不大的情况^[34]. 对于由快速路入口匝道和出口街区路口所构成的一段快速路的通道集成控制问题, 如图4所示, 交通信号控制器之间的协调优化研究还很少.

先进的城市交通信号控制与管理系统是采用分层递阶的分布式控制方案. 为实现控制器之间的协调优化, 通常引入MAS和强化学习的理论和方法, 将每个控制器作为一个Agent, 通过多Agent的协调实现区域交通的优化控制^[35]. 最简单的方式是Agent相互之间独立、没有通讯, Agent自身采用自适应动态规划(如 Q 学习)不断进行学习, 提高控

制性能指标, 间接实现区域内交通流的优化调节^[31]. 另一种方式是Agent通过相互之间的通讯获得上游或下游的交通状态, 并用来构造自身的性能指标函数, 从而在自身性能指标函数学习优化的过程中实现相互之间的协调. 在干线多路口和快速路多入口匝道的协调控制问题中, 各个控制器在性能指标函数中的比重会在一定程度上影响单个优化控制器的收敛性和整体优化性能^[32, 36].

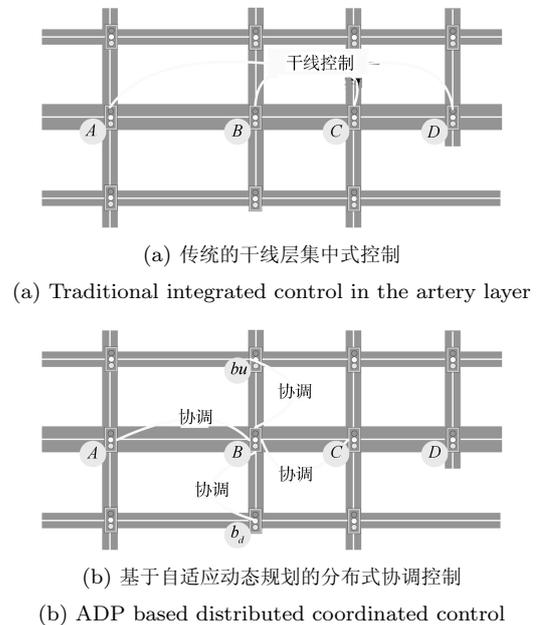


图3 街区多路口交通信号的协调控制
Fig.3 Coordinated traffic signal control for multiple street intersections

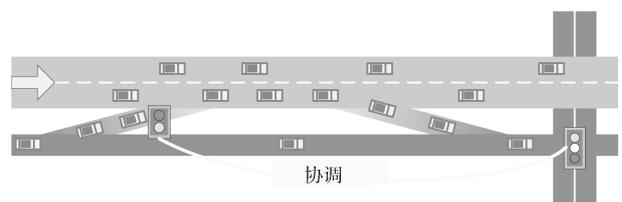


图4 包括入口匝道和出口街区路口交通信号的一段快速路的交通信号的协调控制
Fig.4 Coordinated traffic signal control for a freeway section comprising an on-ramp and an off-ramp intersection

Agent之间的协调也可以引入MAS的协调优化理论和方法, 如将两个或多个Agent之间的协调过程看成是一个演化博弈过程, 通过不断地迭代学习达到博弈的动态平衡, 以实现区域的协调优化^[9]. 以街区多路口交通信号的协调优化控制为例, 如图3(b)所示. 在底层单路口交通信号优化控制的基础上, 每个路口Agent都和相邻路口Agent进行协调优化, 两两之间建立演化博弈关系, 通过演化过程寻找各自的演化稳定策略, 之后, 再根据路口或路段

的重要性对演化稳定策略进行加权求和, 得到该路口交通信号的执行策略. 从上述分析可知, 每个路口 Agent 都分别与相邻路口 Agent 进行协调优化, 会增加一些计算量, 但由于并不采用集中控制方式, 便于构成大的区域协调优化系统, 理论上可以扩展到无限大的区域范围. 因此, 从发展趋势看, 演化博弈论或自适应动态规划交互学习的方法, 一定会在城市交通信号的协调优化控制中得到更多的应用.

3 讨论与展望

自适应动态规划自提出以来, 受到了科研和工程领域人员的关注和重视, 在方法构造、特性分析、理论证明和工程应用等方面都取得了不少的成绩, 但也存在着收敛性不高、稳定性无法保证、相互之间的协调有待研究等问题. 为提高自适应动态规划的收敛性, 引入资格迹的概念是一种可行方法, 其在时间差分 and Q 学习方法中的应用已经证明了这一点. 另外, 借鉴演化算法实现对控制模块和评价模块的训练, 也具有一定的优越性, 这是由于训练过程不需要对系统建模, 减少了模型误差的影响, 有利于提高收敛性. 如前所述, 采用系统的先验知识、仔细设计回报函数和折扣因子等都是有益的尝试. 而对收敛性的进一步的理论证明工作具有一定的难度, 但具有重要意义. 自适应动态规划的稳定性证明则需要结合传统的控制系统稳定分析方法进行突破. 在此基础上, 进一步研究采用自适应动态规划交互学习方法、实现多 Agent 之间的协调优化, 为分布式系统的协调控制方法的应用具有积极的现实意义.

在应用上, 交通信号的优化控制通常采用强化学习的方式, 这是由于控制器的优化训练无法获得导师信号, 只能从环境中得到反馈信息. 一些交通信号控制系统如 RHODES、OPAC、PRODYN 系统等, 采用动态规划的方法, 但受到该方法维数灾的影响, 只能采取滑动窗实现有限时间范围的优化. 而自适应动态规划方法通过函数逼近巧妙地解决了维数灾的难题, 近年来在交通信号优化控制中的研究日益增多, 预期也将在新的交通信号控制系统中得到广泛应用. 另一方面, 城市交通信号控制系统属于典型的分布式控制系统, 将每个控制器看作一个 Agent, 采用多 Agent 间的协调实现城市交通信号控制系统的协调优化, 是交通工程领域和控制领域认可的思路, 前面提到的自适应动态规划交互学习/演化博弈的方法则适合于解决这类问题. 然而, 目前的主要研究还集中在针对采用宏观模型的交通控制问题上, 尚缺少针对实际交通流以及考虑微观交通流特性, 如车辆的跟驰特性等的研究和应用. 虽然自适应动态规划是一种可以不基于系统模型的优化控制方法, 但在直接利用实际数据进行训练

过程中, 如何确保和提高该方法的收敛性和稳定性, 尚需要理论和实际工作的检验, 在一些标准交通仿真软件平台如 Paramics 和 TSIS (Traffic Software Integrated System) 等上的测试也是一种有效手段.

4 结束语

自适应动态规划解决了动态规划维数灾的难题, 其理论和方法的研究日益受到普遍关注. 本文介绍了自适应动态规划的定义、特性以及有关收敛性、稳定性、协调性的关键问题和研究现状. 而城市交通信号控制系统是典型的分布式控制系统, 基于 MAS 的协调优化控制是有效的解决方案. 自适应动态规划方法本质上适合于城市交通信号的优化控制, 对自适应动态规划交互学习和演化博弈论的深入研究, 势必会促进分布式控制系统 MAS 协调优化的理论和方法的发展, 为交通信号的协调优化控制奠定扎实的应用基础, 开拓出广阔的应用前景.

References

- 1 Werbos P J. Using ADP to understand and replicate brain intelligence: the next level design. In: Proceedings of IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning. Honolulu, USA: IEEE, 2007. 209–216
- 2 Werbos P J. Advanced forecasting methods for global crisis warning and models of intelligence. *General Systems Yearbook*, 1977, **22**: 25–38
- 3 Liu De-Rong. Approximate dynamic programming for self-learning control. *Acta Automatica Sinica*, 2005, **31**(1): 13–18
- 4 Saeks R E, Cox C J, Mathia K, Maren A J. Asymptotic dynamic programming: preliminary concepts and results. In: Proceedings of the IEEE International Conference on Neural Networks. Houston, USA: IEEE, 1997. 2273–2278
- 5 Bertsekas D P, Tsitsiklis J N. *Neuro-Dynamic Programming*. Belmont: Athena Scientific, 1996
- 6 Balakrishnan S N, Biega V. Adaptive-critic-based neural networks for aircraft optimal control. *Journal of Guidance, Control, and Dynamics*, 1996, **19**(4): 893–898
- 7 Crites R H, Barto A G. Elevator group control using multiple reinforcement learning agents. *Machine Learning*, 1998, **33**(2): 235–262
- 8 Liu D R, Zhang Y, Zhang H G. A self-learning call admission control scheme for CDMA cellular networks. *IEEE Transactions on Neural Networks*, 2005, **16**(5): 1219–1228
- 9 Bazzan A L. A distributed approach for coordination of traffic signal agents. *Autonomous Agents and Multi-Agent Systems*, 2005, **10**(1): 131–164
- 10 Watkins C J C H, Dayan P. Q-Learning. *Machine Learning*, 1992, **8**(3-4): 279–292
- 11 Murray J J, Cox C J, Lendaris G G, Saeks R. Adaptive dynamic programming. *IEEE Transactions on Systems, Man, Cybernetics, Part C: Applications and Review*, 2002, **32**(2): 140–152
- 12 He P, Jagannathan S. Reinforcement learning-based output feedback control of nonlinear systems with input constraints. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2005, **35**(1): 150–154

- 13 Si J, Barto A G, Powell W B, Wunsch D. *Handbook of Learning and Approximate Dynamic Programming*. New York: Wiley-IEEE Press, 2004
- 14 Lendaris G G, Shannon T T, Schultz L J, Hutsell S, Rogers A. Dual heuristic programming for fuzzy control. In: Proceedings of the Joint 9th IFSA World Congress and the 20th NAFIPS International Conference. Vancouver, Canada: IEEE, 2001. 551–556
- 15 Sutton R S, Barto A G. *Reinforcement Learning: An Introduction*. Cambridge: The MIT Press, 1998
- 16 Li T, Zhao D B, Yi J Q. Heuristic dynamic programming strategy with eligibility traces. In: Proceedings of the American Control Conference. Seattle, USA: IEEE, 2008. 4535–4540
- 17 Zhao D B, Yi J Q, Liu D R. Particle swarm optimized adaptive dynamic programming. In: Proceedings of IEEE Symposium on Approximate Dynamic Program and Reinforcement Learning. Honolulu, USA: IEEE, 2007. 32–37
- 18 Park J W, Harley R G, Venayagamoorthy G K. Adaptive-critic-based optimal neurocontrol for synchronous generators in a power system using MLP/RBF neural networks. *IEEE Transactions on Industrial Applications*, 2003, **39**(5): 1529–1540
- 19 Alanis A Y, Sanchez E N, Loukianov A G. Discrete-time recurrent neural induction motor control using Kalman learning. In: Proceedings of IEEE International Joint Conference on Neural Networks. Vancouver, Canada: IEEE, 2006. 1993–2000
- 20 Osborne M J. *An Introduction to Game Theory*. Oxford: Oxford University Press, 2004
- 21 Maynard S J. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press, 1994
- 22 Hofbauer J, Sigmund K. *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press, 1998
- 23 Tuyls K, Parsons S. What evolutionary game theory tells us about multi-agent learning. *Artificial Intelligence*, 2007, **171**(7): 406–416
- 24 Liu Zhi-Yong. *Intelligent Transportation Control Theory and Application*. Beijing: Science Press, 2003
(刘智勇. 智能交通控制理论及其应用. 北京: 科学出版社, 2003)
- 25 Lu Hua-Pu, Li Rui-Min, Zhu Yin. *Introduction to Intelligent Transportation System*. Beijing: China Railway Publishing House, 2004
(陆化普, 李瑞敏, 朱茵. 智能交通系统概论. 北京: 中国铁道出版社, 2004)
- 26 Pappis C P, Mamdani E H. A fuzzy logic controller for a traffic junction. *IEEE Transactions on Systems, Man, and Cybernetics*, 1977, **7**(10): 707–717
- 27 Trabia M B, Kaseko M S, Ande M. A two-stage fuzzy logic controller for traffic signals. *Transportation Research Part C: Emerging Technologies*, 1999, **7**(6): 353–367
- 28 Lee J H, Lee K H. Distributed and cooperative fuzzy controllers for traffic intersection group. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Review*, 1999, **29**(2): 263–271
- 29 Bingham E. Reinforcement learning in neurofuzzy traffic signal control. *European Journal of Operational Research*, 2001, **131**(2): 232–241
- 30 Yang Yu-Pu, Ou Hai-Tao. Self-organized control of traffic signals based on reinforcement learning and genetic algorithm. *Acta Automatica Sinica*, 2002, **28**(4): 564–568
(杨煜普, 欧海涛. 基于再励学习与遗传算法的交通信号自组织控制. 自动化学报, 2002, **28**(4): 564–568)
- 31 Wiering M, Vreeken J, van Veenen J, Koopman A. Simulation and optimization of traffic in a city. In: Proceedings of IEEE Intelligent Vehicles Symposium. Parma, Italy: IEEE, 2004. 453–458
- 32 Bai X R, Zhao D B, Yi J Q, Xu J. Coordinated control of multiple ramp metering based on DHP(λ) controller. In: Proceedings of the 11th IEEE International Conference on Intelligent Transportation Systems. Beijing, China: IEEE, 2008. 351–356
- 33 Cai C. An approximate dynamic programming strategy for responsive traffic signal control. In: Proceedings of IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning. Honolulu, USA: IEEE, 2007. 303–310
- 34 Papageorgiou M, Kotsialos A. Freeway ramp metering: an overview. *IEEE Transactions on Intelligent Transportation Systems*, 2002, **3**(4): 271–281
- 35 Wang F Y. Agent-based control for networked traffic management. *IEEE Intelligent Systems*, 2005, **20**(5): 92–96
- 36 Li T, Zhao D B, Yi J Q. Adaptive dynamic programming for multi-intersections traffic signal intelligent control. In: Proceedings of the 11th IEEE International Conference on Intelligent Transportation Systems. Beijing, China: IEEE, 2008. 286–291



赵冬斌 中国科学院自动化研究所副研究员. 主要研究方向为计算智能, 智能交通和机器人. 本文通信作者.

E-mail: dongbin.zhao@ia.ac.cn

(**ZHAO Dong-Bin** Associate professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers computational intelligence, intelligent transportation, and robotics. Corresponding author of this paper.)



刘德荣 中国科学院自动化研究所研究员. 主要研究方向为智能控制理论及应用, 人工神经网络, 模糊系统, 计算神经科学, 电力系统运行与控制 and 汽车发动机控制. E-mail: derong.liu@ia.ac.cn

(**LIU De-Rong** Professor at the Institute of Automation, Chinese Academy of Sciences. His research interest

covers intelligent control, neural networks, fuzzy systems, computational neuroscience, power systems, and automotive engine control.)



易建强 中国科学院自动化研究所研究员. 主要研究方向为智能控制理论及应用. E-mail: jianqiang.yi@ia.ac.cn

(**YI Jian-Qiang** Professor at the Institute of Automation, Chinese Academy of Sciences. His main research interest is intelligent control.)