

智能控制：从学习控制到平行控制

王飞跃^{1,2†}, 魏庆来¹

(1. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190;

2. 青岛智能产业技术研究院, 山东 青岛 266109)

摘要: 20世纪60年代, 学习控制开启了人类探究复杂系统控制的新途径, 基于人工智能技术的智能控制随之兴起. 本文以智能控制为主线, 阐述其由学习控制向平行控制发展的历程. 本文首先介绍学习控制的基本思想, 描述了智能机器的架构设计与运行机理. 随着信息科技的进步, 基于数据的计算智能方法随之出现. 对此, 本文进一步简述了基于计算智能的学习控制方法, 并以自适应动态规划方法为切入点分析非线性动态系统自学习优化问题的求解过程. 最后, 针对工程复杂性与社会复杂性互相耦合的复杂系统控制问题, 阐述了基于平行控制的学习与优化方法求解思路, 分析其在求解复杂系统优化控制问题方面的优势. 智能控制思想经历了学习控制、计算智能控制到平行控制的演化过程, 可以看出平行控制是实现复杂系统知识自动化的有效方法.

关键词: 智能控制; 平行控制; 学习控制; 人工智能; 自适应动态规划; 平行智能

引用格式: 王飞跃, 魏庆来. 智能控制: 从学习控制到平行控制. 控制理论与应用, 2018, 35(7): 939 – 948

中图分类号: TP273 **文献标识码:** A

Intelligent control: from learning control to parallel control

WANG Fei-yue^{1,2†}, WEI Qing-lai¹

(1. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;

2. Qingdao Academy of Intelligent Industries, Qingdao Shandong 266109, China)

Abstract: In the 1960s, learning control opened up new ways for humans to explore the control of complex systems, and intelligent control emerged based on artificial intelligence technologies thereafter. In this paper, intelligent control is employed as the main line, which describes the development process of intelligent control from learning control to parallel control. First, the basic idea of learning control is introduced, and then the architecture design and operating mechanism of intelligent machines are described. With the advancement of information science, data-based computational intelligence approaches are emerged. Then, the learning control methods based on computational intelligence are described. Adaptive dynamic programming, as one of the most well-known self-learning optimal control methods, is described to show how to solve the optimal control problems for nonlinear dynamic systems. Finally, considering the control problem of complex systems with both engineering and social complexities, parallel control is introduced and its advantages in solving complex optimization control problems are illustrated. Intelligent control is developed from learning control to computational intelligence control and then to parallel control. It can be seen that parallel control is an effective method to realize the knowledge automation of complex systems.

Key words: intelligent control; parallel control; learning control; artificial intelligence; adaptive dynamic programming; parallel intelligence

Citation: WANG Feiyue, WEI Qinglai. Intelligent control: from learning control to parallel control. *Control Theory & Applications*, 2018, 35(7): 939 – 948

1 引言(Introduction)

现代工业革命之后, 控制理论与方法在工业生产中发挥了越来越重要的作用. 大机器、大系统时代的到来更是极大地促进了控制理论与技术的发展创新.

工业革命成为了控制系统发生本质性变化的源动力, 促使系统控制方案不断改进与升级^[1]. 其发展过程可分为5个阶段, 分别为机械化、电气化、信息化、网络化和目前的平行智能化, 如图1所示.

收稿日期: 2018-05-02; 录用日期: 2018-07-13.

†通信作者. E-mail: feiyue.wang@ia.ac.cn; Tel.: +86 10-82544521.

本文责任编辑: 王鞠维.

“2017年智能控制研讨会”论文, 黄琳院士推荐.

国家自然科学基金项目(61722312, 61533019), 北京市科技计划项目(Z181100008918007), 英特尔智能网联汽车大学合作研究中心项目(ICRI-IACV)资助.

Supported by the National Natural Science Foundation of China (61722312, 61533019), the Beijing Municipal Science & Technology Commission (Z181100008918007) and the Intel Collaborative Research Institute for Intelligent and Automated Connected Vehicles (ICRI-IACV).

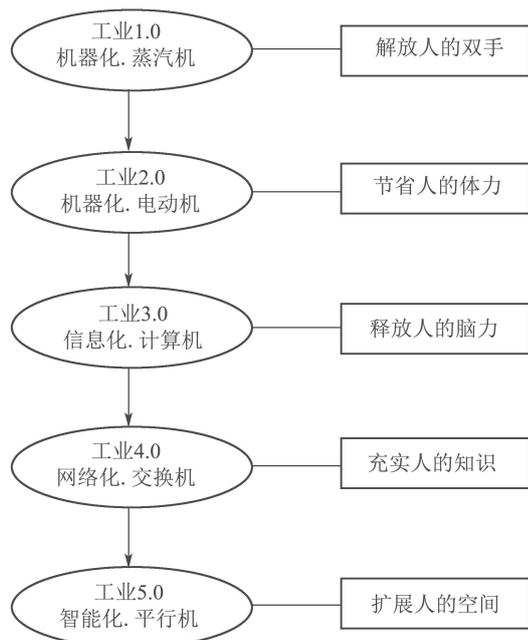


图1 工业进化过程

Fig. 1 The evolutionary process of industrialization

第一次工业革命以机械控制为主, 控制蒸汽机进行工业生产. 第二次工业革命是以电力为主导的产业革命, 提高了对控制精度、准度等方面的要求. 第三次工业革命的特点是信息化. 与第一和第二次工业革命的传统闭环控制不同, 第三次工业革命要求提供复杂、非线性、时变、时滞严重等问题的控制策略, 传统控制方法很难实现. 因此, 智能控制方法应运而生, 成为了第三次工业革命的代表性技术. 智能控制的思想起源于20世纪60年代, 傅京孙首次把启发式推理融入到了学习控制中, 开启了智能学习控制的研究大门^[2]. 随后George N. Saridis等人先后提出了智能机器的概念, 利用人工智能、操作研究和控制理论形成了智能控制的一般框架^[3-5]. 然而, 传统的智能控制侧重于工业设备工艺的“小”闭环控制, 缺少人等社会性要素, 因此其控制过程是不全面的, 难以正确处置各种情况变化及非正常状况^[6]. 第四次工业革命是以网络化为代表的工业4.0, 核心设备路由器出现, 将众多计算机互联, 进入了网络化时代, 使得信息和物理系统融合加深, 形成了信息物理系统 (cyber-physical systems, CPS). 针对CPS系统, 人工智能与知识学习成为解决复杂系统优化控制问题的热点技术方案. 在神经网络研究的基础上, 融合动态规划和强化学习的优势, 美国基金会Werbos博士等提出的自适应动态规划方法, 为求解复杂的非线性系统问题提供了一个有效的解决方案^[7]. 作为深度学习的代表, AlphaGo打败围棋名将李世石标志着人工智能技术的新发展高度^[8]. 机器有了自学习的能力, 基于快速大量的数据计算, 就能够快速对目标行为做出决策. 另外, 计算智能技术^[9]、无人控制技术^[10]、量子信息技术^[11]以及虚拟现

实^[12]技术等相继出现, 建立了众多针对CPS系统求解策略.

当前, 人类社会已经步入第五次工业革命, 即工业5.0的时代. 工业5.0的概念在文献[13-14]中提出, 并预言第5个技术发展阶段的特征-平行化. 文献[15]对工业5.0进行了深入研究, 探讨了平行控制在核电系统中的应用. 在工业5.0时代, 网络化应用的加深, 特别是“互联网+”的出现, 进一步加深了信息和物理系统、工业与人类社会的相互融合, 形成了工程复杂性与社会复杂性相耦合的复杂系统, 该系统可以概括为社会物理信息系统 (cyber-physical-social systems, CPSS). 一般情况下, CPSS系统不能建立解析的系统行为模型, 并且此类系统结构不明确、边界不确定. 因此以往的系统分析方法往往难以刻画系统部分之间的相互关系. 平行控制的提出为复杂CPSS系统控制问题提供了理念指导^[16]. 平行控制首先根据实际CPSS系统建立虚拟人工系统, 其次利用人工系统进行计算实验, 探究实际系统的演化及相互作用关系, 然后虚实控制系统平行执行, 获得CPSS系统优化决策. 可以看到, 工业5.0的运行模式将引领工业进入平行智能化时代.

本文以工业系统的进化历程为背景, 以智能控制的研究发展为主线, 系统性的梳理了智能控制由学习控制到平行控制的发展历程. 希望借此抛砖引玉, 为复杂系统的智能控制与优化提供新的解决思路. 论文的主要结构总结如下: 第2节描述了学习控制与智能控制的发展; 第3节阐述了智能机器的研究方法; 第4节讲述了计算智能控制的发展历程; 第5章以自适应动态规划方法为切入点, 叙述了计算智能优化控制的实现过程; 第6节描述了平行控制理念下的知识自动化方法. 最后, 在第7节对文章进行了总结与展望.

2 学习控制与智能控制(Learning control and intelligent control)

20世纪60年代, 学习控制的思想在当时成为了研究热点. 其原因是学习控制能够基于自身的学习能力来识别处理被控对象状态变化和外界环境改变, 并根据自身的特性不断学习改进, 从而适应被控对象的变化, 在控制方法中体现出强大的优势. 美国普渡大学傅京孙教授首先将启发式推理规则用于学习控制系统, 为智能控制的发展做出了重要的贡献^[17]. 在文献[17]中, 傅京孙教授列举了学习控制的基本求解方案: 针对一个最优控制系统, 如果已知并能精确描述控制过程的全部先验知识, 可使用确定性优化技巧设计优化控制器; 当只能统计描述全部或部分先验知识时, 需要使用统计或随机方法设计其优化控制器; 如先验知识未知或无法全面获取时, 经典的优化设计方法(如动态规划等)就无法实现控制器设计. 此时存在

两种求解思路: 1) 基于已知部分知识的控制器设计. 该方法忽略部分未知信息, 是不完整的次优求解方法. 2) 设计能估计未知信息的优化控制器. 当估计信息逐渐接近于真实值时, 控制器逐渐逼近最优理想控制器.

学习控制的关键是能否成功逼近系统的未知知识, 而未知部分可能是系统参数或者描述系统的不确定性或随机函数. 在评估学习过程中, 随着控制器获取了系统的更多未知信息, 控制律也随之更新. 图2描述了典型的学习控制系统结构, 其中被控对象存在未知或部分未知的系统扰动, 因此需要设计相应的最优控制器以预测未知信息. 需要指出的是, 实际获得的控制器一般为次优的控制器, 但是在学习过程中次优控制器会不断趋近于最优理想控制器. 监督部分的作用是评估控制器的性能并指导学习过程, 从而不断改善系统的控制性能.

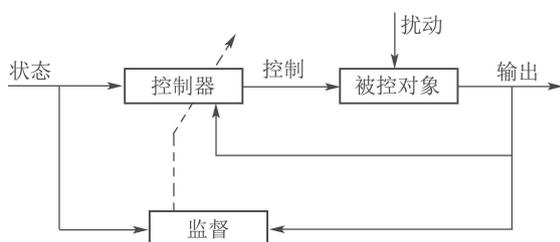


图2 学习控制结构图

Fig. 2 Structure of learning control

学习过程可分为两种: 1) 监督式学习(或离线式学习). 训练目标已知, 系统可根据目标值改进系统性能. 2) 无监督式学习(或在线式学习). 在无外界目标时, 一种方法是基于贝叶斯学习求解全部的可能解. 另一种方法则是利用性能评估指导训练过程. 后者在不同的控制场景中获取不同的控制经验知识. 相近的控制经验被用于相近的控制场景中用以提升系统的整体性能. 随着经验知识的积累, 不同场景的最优控制律便可累积得到. 文献[17]中的学习控制框架打破了经典控制中人为设计控制器的思路, 为后续智能控制的发展奠定了理论研究基础. 1971年, 在学习控制的基础上, 傅京孙教授正式提出了智能控制的概念, 认为智能控制就是人工智能与控制理论的结合, 这是在学术文献中第一次将智能控制做为一个独立的学科进行描述^[18]. 随后, Uchiyama于1978年首先提出迭代学习控制理论, 通过反复应用先前试验得到的信息来获得能够产生期望输出轨迹的控制输入, 以改善控制质量. 它仅需较少的先验知识和计算量就能够处理不确定度相当高的动态系统, 具有适应性强, 易于实现的功能. 随着研究的不断深入, 越来越多研究人员投入到迭代控制算法的研究中, 并提出了很多有效的控制算法, 如增强学习和自适应动态规划等等. 此外, 基于学习控制的智能控制方法也得到了广泛的关注, 取得了很多的成就, 如模糊算法、模拟退火算法、粒子群算

法、遗传算法、人工神经网络算法等, 这些算法在求解复杂非线性问题上展现了自身的独特魅力. 目前, 这些算法已经广泛的应用于机器人控制、工业生产等领域中.

3 智能控制与智能机器 (Intelligent control and intelligent machines)

智能控制具有智能的信息处理、反馈和控制决策能力, 使其在求解复杂系统的控制问题中具有强大优势. 实现智能控制思想的典型代表就是智能机器. 在文献[3-5]中, Saridis教授定义了智能机器的结构, 包括组织层(organization level)、协调层(coordination level)和执行层(execution level)三个部分. 智能机器结构可由图3表示. 组织层是智能机器的大脑, 其职能是根据任务目标进行分析、判断与决策. 协调层作为组织层和执行层的中间层, 其作用是协调硬件任务分配. 执行层是底层结构, 其功能是根据控制任务需求在硬件上精确执行相应的控制命令.

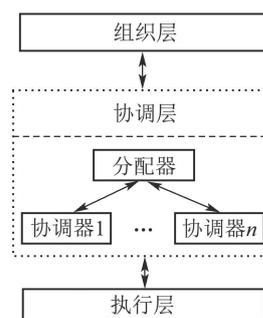


图3 智能机器结构图

Fig. 3 Structure of intelligent machines

协调层内部基于协调控制理论^[19], 由分配器(dispatcher)和协调器(coordiantor)两个部分组成. 分配器的职能是处理协调器的控制和通信. 分配器需要根据系统状态采用控制网络把给定的任务信息解析成一系列的控制决策行为, 并分配给对应的协调器. 然后协调器将处理的反馈信息和数据再经过分配器汇总上传. 分配器自身具有通信能力, 数据处理能力, 任务处理能力以及学习能力. 对于协调器而言, 每个协调器都要处理一系列的硬件设备数据, 因此协调器可以认为是特定领域内的任务专家. 在给定的任务空间及时间条件下, 结合分配器发出的任务状态和当前的任务信息, 协调器具体决定底层硬件的操作执行问题. 协调器可根据底层硬件的特性生成并下达底层任务控制命令, 并将执行结果返回上层结构. 分配器和协调器具有相同的结构, 如图4所示. 其中, 数据处理器的主要功能是提供当前系统信息和需要执行的任务信息, 可以分为任务描述、状态描述和数据描述3个部分. 任务处理器生成需要底层执行的任务命令, 主要包括任务规划、任务解析和任务构建3个步骤. 学习器

用来提高任务处理器的能力,有助于减少决策及信息处理过程的不确定性.上述结构为智能机器的基本执行结构.智能机器为现代工业,如智能机器人系统及智能生产制造,提供了一种有效的执行、决策与控制模型.

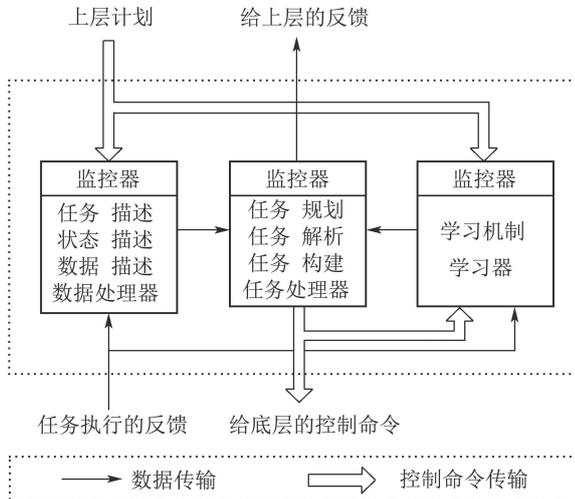


图4 分配器和协调器的结构图

Fig. 4 Structure for dispatcher and coordinators

4 计算智能控制 (Computational intelligence based control)

当前,控制理论在工业生产和航空航天等众多领域都取得了良好的应用效果,然而其应用的控制问题往往需要被控对象具有精确的数学机理模型.因而对被控对象的精确建模成为控制问题求解的关键因素.随着社会的进步与科技的发展,被控对象的复杂程度越来越大,在时间广度和空间维度上的扩展使其存在强烈的非线性、时滞性、不确定性、强耦合性以及各种随机测量问题.这些问题导致无法对被控对象进行精确的数学建模,更难以进行准确的控制求解.对此,研究人员在获取了大量现场数据基础上,开始了基于数据的计算智能控制研究.

在数据获取过程中,被控对象的描述,包括对象特征、控制策略、控制效果等,很多都采用了自然语言的方法.自然语言区别于传统的定量控制的表达形式,具有一定的模糊性,然而控制器一般无法直接对自然语言进行处理.这使得模糊控制理论在自然语言处理中获得了极大的应用.1965年,L. A. Zadeh提出了模糊控制理论^[20].该方法将现场采集到的自然语言描述数据转化为计算机能够接受和处理的算法语言.模糊控制方法简化了控制问题的复杂性,不必建立传统的数学模型,使用自然语言利用计算机进行控制,开启了计算智能控制的里程碑.之后,通过模拟达尔文自然选择的遗传学机理,J. Holland在1975年提出了遗传算法^[21].遗传算法无需已知的确定规则,基于遗传的

规律进行随机化的演化搜索获得优化控制策略.基于固体物质的退火过程特性,S. Kirkpatrick等于1983年提出了一种通用的概率推演算法^[22],即模拟退火算法.该算法基于计算数据从任一位置开始求解,渐进收敛到全局最优解.遗传算法与模拟退火算法都是计算智能控制的关键算法,在机器学习、自适应控制中有广泛的应用.然后,J. Kennedy和R. C. Eberhart等于1995年提出了粒子群算法^[23].粒子群算法是一种基于迭代计算的优化算法,粒子们追随当前的最优粒子在解空间中不断寻优.由于粒子群算法比遗传算法更为简单,因而该算法在函数优化等方面得到广泛应用.

基于现代医学对大脑的研究成果,科研人员建立了神经网络的数学模型.经过数代的发展和改进,神经网络已成为当前计算智能控制中重要的逼近工具和执行器.1943年,W. McCulloch和W. Pitts二人提出了McCulloch-Pitts(MP)模型^[24].1982年,J. J. Hopfield将计算能量引入到神经网络,提出了Hopfield网络,从而开启了神经网络在优化计算方面的新前景^[25].基于环境数据,神经网络通过对自身的权值进行调整,实现了对环境的适应性变化.随着半个多世纪的研究与发展,神经网络基于自身的非线性处理特性,在模式识别、智能控制等领域取得了非常多的研究成果,成为当前的热门研究方向之一.

计算机数据处理技术的发展以及大数据的便捷化快速运算处理使得求解复杂、非线性严重的控制与决策问题成为可能.基于多层神经网络的深度学习方法,谷歌公司Demis Hassabis、David Silver、黄士杰等人设计了一款人工智能的围棋程序,即阿尔法狗(Alpha Go)^[8],并分别在2016年与围棋九段选手李世石和2017年中国选手柯洁的比赛中获胜.AlphaGo主要由Move Picker和Position Evaluation两个不同的多层神经网络组成.Move Picker网络寻找最优的控制策略,Position Evaluation网络则通过对给定棋局位置的评估预测获胜的概率,从而帮助Move Picker网络寻找最优的策略.在阿尔法狗的基础上,DeepMind相继公布了AlphaGo Zero^[26]与Alpha Zero^[27].其中,AlphaGo Zero打破人类基于先验知识学习的方式,实现了在无任何人类知识指导条件下进行学习并超越了人类的能力.Alpha Zero在AlphaGo Zero基础之上实现了通用智能学习,掀起了人类对基于计算智能决策研究的新高潮.

5 计算智能最优控制: 自适应动态规划 (Computational intelligence based optimization: adaptive dynamic programming)

随着市场竞争的日趋激烈与能源的日益枯竭,人们对控制系统性能,如节能、降耗等的要求越来越高.在控制系统时,通常会要求工程人员根据被控对象的

有限信息做出高水平的决策, 使控制系统性能达到最优化. 计算智能控制技术的迅猛发展, 为解决复杂非线性动态系统的优化问题提供了新思路. 在众多的复杂系统优化控制方法中, 自适应动态规划(adaptive dynamic programming, ADP)采用自学习优化的方式, 成为一类解决复杂系统优化控制的强有力工具^[28], 其核心思想与强化学习类似, 都是通过系统从环境到行为映射的学习实现强化信号的函数值最大. 自适应动态规划的理论基础是动态规划原理. 动态规划(dynamic programming, DP)是求解决策过程最优化的数学方法. 20世纪50年代初美国数学家R. E. Bellman等人在研究多阶段决策过程的优化问题时, 提出了著名的最优性原理, 把多阶段过程转化为一系列单阶段问题逐个求解, 创立了解决这类过程优化问题的新方法—动态规划^[29]. 1957年出版了他的名著:《动态规划》, 这是该领域的第1本著作. 该方法与庞特里雅金(L. S. Pontryagin)最大值原理和卡尔曼(R. E. Kalman)滤波理论被称为现代控制理论中的3个里程碑.

动态规划的原理可以叙述如下. 对于一个给定离散时间非线性系统

$$x_{k+1} = f(x_k, u_k), \quad (1)$$

其中: $x_k \in \mathbb{R}^n$ 表示系统状态, $u_k \in \mathbb{R}^m$ 表示控制变量. 设性能指标函数为

$$J(x_k) = \sum_{i=k}^{\infty} U(x_i, u_i), \quad (2)$$

其中: $U(x_i, u_i)$ 称为效用函数. 最优控制的目标是使性能指标函数(2)最小化. 令 k 时刻的代价函数表示为: $U(x_k, u_k) + J^*(x_{k+1})$. 根据Bellman最优性原理, k 时刻最优性能指标函数可以表示为

$$J^*(x_k) = \min_{u_k} \{U(x_k, u_k) + J^*(x_{k+1})\}. \quad (3)$$

因此, k 时刻的最优控制 u_k^* 可以表示为

$$u_k^* = \arg \min_{u_k} \{U(x_k, u_k) + J^*(x_{k+1})\}. \quad (4)$$

然而, 动态规划方法却很难直接应用于实际系统求解系统的最优控制. 从式(3)中可以看出, 如果本文要求解 k 时刻的最优性能指标函数 $J^*(x_k)$, 那么首先要获得 $k+1$ 时刻的最优性能指标函数 $J^*(x_{k+1})$. 然而 $J^*(x_{k+1})$ 却是未知的, 因此动态规划需要时间后向进行求解最优控制问题. 这种方式在求解步骤以及求解空间增加时, 其计算量将指数增长. 这就是动态规划方法中著名的“维数灾”问题. 为了解决维数灾问题, 1977年Werbos博士首先提出了自适应动态规划方法求解贝尔曼方程^[7]. 该方法利用梯度下降法调整多层神经网络的权值参数近似最优控制策略与性能指标函数. 在文献[7]中, Werbos博士提出了启发式动态规划(heuristic dynamic programming, HDP)方法, 求解连续状态空间非线性系统最优控制问题. 随后Werbos

博士又提出了执行依赖启发式动态规划(action dependent heuristic dynamic programming, ADHDP)方法, 双启发式动态规划(dual heuristic dynamic programming, DHP)方法和执行依赖双重启发式动态规划(action dependent dual heuristic programming, ADDHP)方法. 全局双重启发式动态规划(globalized heuristic dynamic programming, GHDP)方法和执行依赖全局双重启发式规划(action dependent globalized dual heuristic programming, ADGDHP)方法由Prokhorov和Wunsch在1997年提出^[30]. 以上这6种方法是自适应动态规划方法的基础实现结构, 自适应动态规划的基本结构可由图5表示.

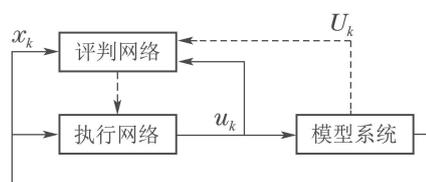


图 5 自适应动态规划结构图

Fig. 5 Structure of adaptive dynamic programming

自适应动态规划方法整个结构主要由3部分组成: 动态系统、执行(action)模块和评判(critic)模块. 每个部分均可由神经网络代替, 其中动态系统可以通过神经网络进行建模, 执行模块用来近似最优控制律, 评判模块用来近似最优性能指标函数. 后两者的组合相当于一个智能体, 控制/执行模块作用于动态系统(或者被控对象)后, 通过环境(或者被控对象)在不同阶段产生的奖励/惩罚(reward/penalty)来影响评判函数. 评判模块的参数更新是基于Bellman最优原理进行的, 实现性能指标函数达到最优. 上述过程构成了一种自学习计算智能优化控制方法.

下面采用数学语言来描述自适应动态规划的求解过程:

- i) 初始化.
 - a) 初始化性能指标函数;
 - b) 给定初始状态;
 - c) 给定计算精度.
- ii) 对 $k = 0, 1, 2, \dots$, 更新控制律如下:

$$u(x_k) = \arg \min_{u_k} \{U(x_k, u_k) + J(x_{k+1})\}. \quad (5)$$

- iii) 更新性能指标函数如下:

$$J(x_k) = \min_{u_k} \{U(x_k, u_k) + J(x_{k+1})\} = U(x_k, u(x_k)) + J(f(x_k, u(x_k))). \quad (6)$$

- iv) 如果 $\|J(x_k) - J(x_{k+1})\| < \varepsilon$, 则停止.

从上述算法可以看出, 自适应动态规划的迭代算法由两部分组成, 分别对应了执行网络的更新和评价

网络的更新. 自适应动态规划算法并不是逆序的根据每个阶段实际的代价精确计算的, 而是正序的从任意一个初始值开始不断的根据Bellman方程进行修正的, 同时整个过程是自学习的而不需要人为设计, 这是自适应动态规划的主要优点之一. HDP是自适应动态规划方法中最基本也是使用最普遍的一种^[31-36], 其结构如图6所示.

在图6的HDP方法中, 如果3个神经网络, 包括模型网络、执行网络和评判网络, 各自均采用3层BP网络, 那么既可以将HDP看成一个9层深度神经网络, 如图7所示. 从这个角度可以看出, HDP其实可以理解成深度神经网络的结构. 对于其他自适应动态规划结构, 例如双启发式动态规划, 全局双重启发式动态规划和执行依赖双重启发式动态规划等, 都可以采用类似于

图7的结构求解, 神经网络训练的目标可以统一为使如下误差:

$$e_k = U_k(x_k, u_k) + J_{k+1}(x_{k+1}) - J_k(x_k) \quad (7)$$

达到极小值.

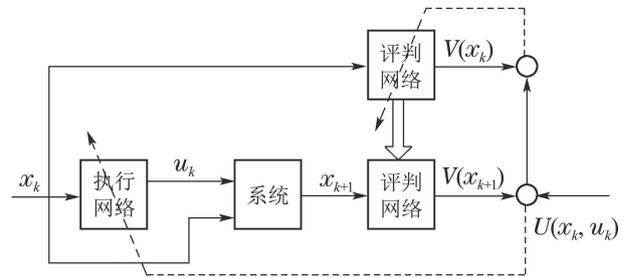


图6 启发式动态规划的结构图

Fig. 6 Structure of HDP

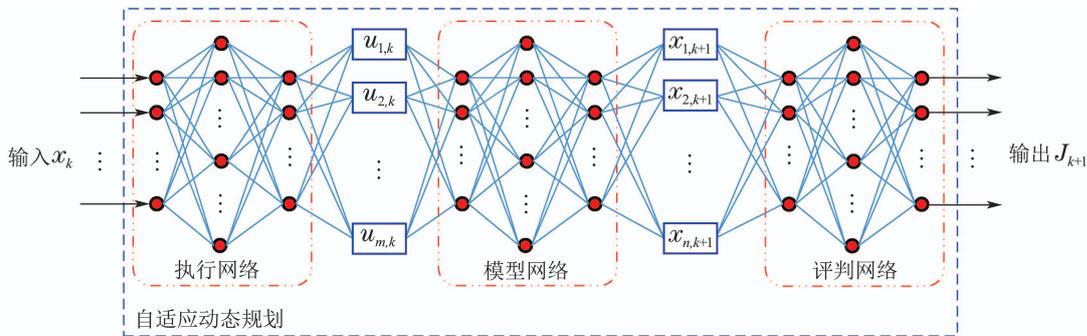


图7 自适应动态规划的深度神经网络架构

Fig. 7 Deep neural network structure of ADP

近年来, 自适应动态规划凭借自身的强大优势, 已经取得了长足的进展. 2005年, Abu-Khalaf和Lewis提出了带有饱和执行器的策略迭代方法^[37]. 在文献[38]中, 自适应动态规划解决了多控制器非零和对策问题. 在模型无法获得的条件下, 提出了基于数据的Q学习自适应动态规划方法, 直接根据系统输入输出数据获得系统最优控制策略^[39-40]. 在文献[41-42]中, 自适应动态规划解决了多智能体最优一致性控制问题. 在文献[43]中提出了基于局部数据的迭代自适应动态规划方法, 在每次迭代过程中只需要根据局部数据便可以更新迭代策略并证明了最优性. 可以看到, 自适应动态规划方法为求解复杂系统优化控制提供了有效的解决方案.

6 平行控制: 从自适应动态规划到知识自动化 (Parallel control: from adaptive dynamic programming to knowledge automation)

随着人类社会向工业5.0的时代的逐步迈进, 工业系统呈现出明显的社会化, 同时涉及工程复杂性和社会复杂性两个方面, 呈现动态性、开放性和交互性等特征, 变得越来越难以控制. 当前对自适应动态规划

方法等计算智能优化控制方法的研究虽然取得了长足的进步, 但是主要的研究领域仍集中在CPS系统的优化控制. 对于CPSS系统, 现有的计算智能优化控制方法, 如自适应动态规划方法等, 求解方式就显得相形见绌. 首先, 在本质上, 一个复杂系统的整体行为不可能通过对其部分行为的独立分析而完全确定^[44]. 因此, 复杂系统需要进行整体的分析和模拟. 而对于复杂系统, 例如大型生产中过程控制系统、交通系统、制造执行系统以及企业资源规划等众多CPSS系统, 随着系统的集成程度的不断提高以及人的参与, 其机理模型几乎不可能精确的建立. 在这种情况下, 对于如图6中的自适应动态规划, 如果在系统模型无法建立的条件, 那么系统的“Model Network”即会失效, 这直接导致了图7中神经网络结构的失效致使自适应动态规划无法运行. 因此, 系统的优化控制就难以获得. 其次, 对于基于数据的自适应动态规划方法来说, 一般需要系统数据的完备性, 使得控制律在全局有效. 然而, 对于复杂系统而言, 相对于任何有限资源, 一个复杂系统的整体行为不可能预先在大范围内完全确定^[45]. 因此, 几乎无法获得完整的复杂系统的实际数据. 而另一方面, 现有的自适应动态规划方法非常依

赖数据的完整性. 当已知数据量不够大或在数据不完备时, 自适应动态规划训练出的最优控制策略不能体现整体系统特征, 更无法获得复杂系统的最优控制方法. 上述矛盾给自适应动态规划方法带来了巨大的挑战, 必须要提出新的理念与优化控制方法^[46]. 对此, 平行动态规划的思想孕育而生. 平行动态规划、自适应动态规划与动态规划方法的关系可以由图8进行描述.

平行动态规划是基于数据的复杂系统优化控制方案^[47]. 首先, 通过对实际世界的观察, 收集状态-执行-奖/惩信号, 建立人工系统, 如图9所示. 由于实际系统的复杂性以及人类行为等高未知性因素, 不能将实际数据直接用来进行复杂系统控制器的设计, 因而人工系统的建立是必要的. 另外, 人工系统并不是传统意义上对实际系统的仿真或重构. 对一个实际复杂系统, 可以构造多个人工系统, 并与实际系统进行互动. 举例来说, 根据需要, 一个实际系统可同时或分时地与理想人工系统、试验人工系统、应急人工系统、优化人工系统、评价人工系统、培训人工系统等平行交互. 这些平行的人工系统可以统称为人工系统. 此外, 从某种意义上来说, 经典的仿真系统或半实物仿真系统也可以看作是特殊的平行系统, 它们能够有效地模拟真实系统的状态, 这与平行系统很相似; 然而, 他们却无法控制真实的系统, 无法实现真实系统与人工系统的双向平行控制, 因此, 仿真系统只是平行系统的一个特例.

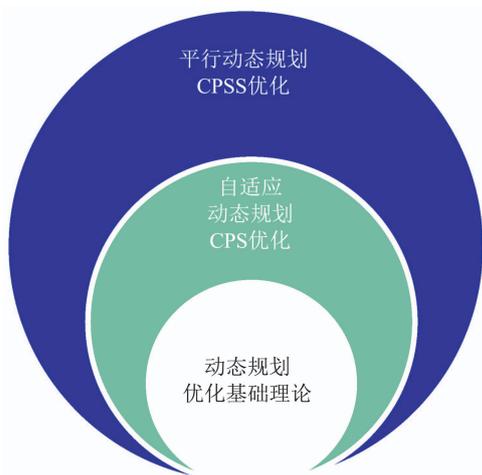


图 8 平行动态规划关系图

Fig. 8 Diagram of parallel dynamic programming

人工系统建立之后, 可以通过人工系统生成大量数据, 包括实际系统中没有或无法获得的数据. 然后在人工系统中进行计算实验, 优化求解. 然而, 人工系统上的优化控制策略并不一定适应于实际系统. 因此, 将人工系统的优化控制作为评价实际系统优化效果的一条准则并用来指导实际系统进行优化. 而后, 对实际系统进行控制与优化, 将实际系统的自适应动态

规划与人工系统自适应动态规划方法平行执行, 两个系统不断交互升级与进步, 保持控制策略的有效性与最优性. 整个过程可由图9进行表示.

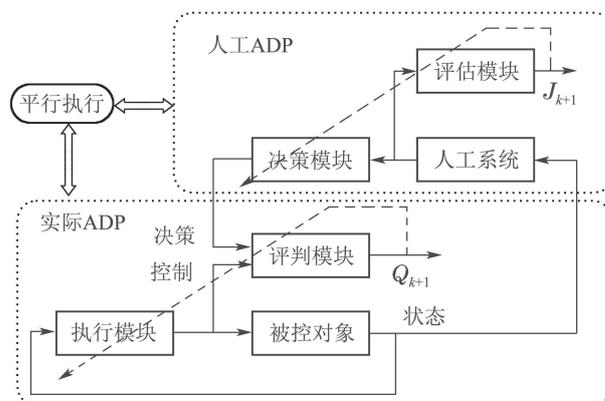


图 9 平行动态规划运行结构图

Fig. 9 Operation structure for parallel dynamic programming

此外, 平行动态规划不局限于单层人工自适应动态规划结构, 可以采用多层自适应动态规划结构进行实现. 在多层平行动态规划结构中, 每个人工系统可以根据预测分析获得数据, 并结合实际数据进行计算实验, 获得每个系统的优化控制策略. 具体来说, 对于图10, 假设存在 n 个平行的人工系统, 本文在每个人工系统中均采用自适应动态规划方法获得系统优化控制策略, 即可获得 n 个优化控制策略. 这样很好的避免了在实际系统上的控制风险, 体现了平行动态规划的优势. 在平行执行过程中, 基于动态规划的最优性原理, 评判网络将根据计算实验中的优化结果, 择优做出实际系统的决策, 并将系统运行结果反馈给人工系统进行下一次的迭代. 可以看到, 通过实际与虚拟系统的平行执行, 系统的性能可以不断被优化, 并最终获得复杂系统的优化控制策略.

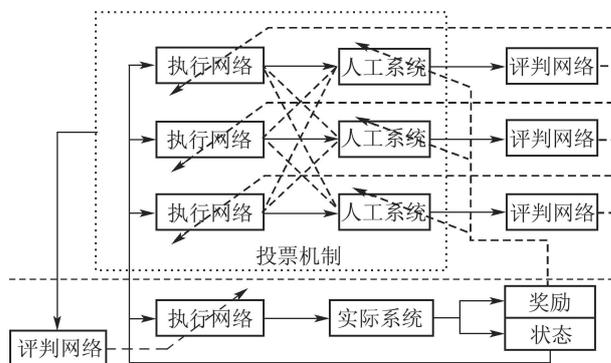


图 10 平行动态规划的运行示意图

Fig. 10 Operation schedule for parallel dynamic programming

从上面的分析可以看出, 平行动态规划方法针对复杂系统优化控制问题的求解步骤可以归结为图11所示, 即首先建立人工系统, 然后在人工系统上进行

计算实验,并根据计算实验的结果指导实际系统运行,最后采用平行执行的方式使得人工系统与实际系统不断交替进化.这与复杂系统平行控制的ACP方法^[16,48]的思想完全吻合.

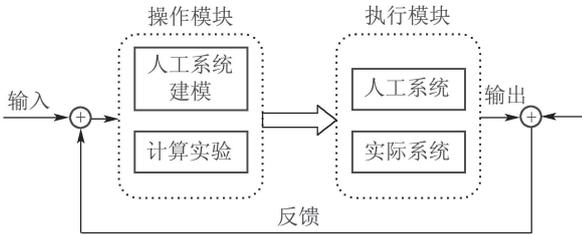


图 11 平行控制结构框架

Fig. 11 Operation schedule of parallel control

ACP方法是复杂系统平行控制方法的核心思想^[16,44-45]. ACP方法由三步曲组成:第1步,利用人工系统对复杂系统进行建模.一定意义上,可以把人工社会看成是科学游戏,就是用类似计算机游戏的技术来建模.第2步,利用计算实验对复杂系统进行分析 and 评估.一旦有了针对性的人工系统,就可以把人的行为、社会的行为放到计算机里面,把计算机变成一个实验室,进行计算实验,通过实验来分析复杂系统的行为,评估其可能的后果.第3步,将实际与人工系统并举,通过实际与人工系统之间的虚实互动,以平行执行的方式对复杂系统的运行进行有效地控制和管理.

上述的平行控制方法给出了求解复杂系统优化控制问题的基本求解思路.平行控制的主要目的,是通过实际系统与人工系统的相互连接,对二者之间的行为进行对比和分析,完成对各自未来的状况的“借鉴”和“预估”,相应地调节各自的管理与控制方式,达到实施有效解决方案以及学习和培训的目的.主要实现功能如图12所示,具体描述如下:

1) 实验和评估:在这一过程中,实际系统以规则为基础进行运行,而人工系统将数据转换为知识,可以理解为知识驱动的系统.人工系统主要用来进行计算实验,分析了解各种不同的人工系统上的实验行为和反应,并对不同的解决方案的效果进行评估,作为选择和支持管理与控制决策的依据.

2) 学习与培训:在这一过程中,人工系统主要是被用来作为一个学习和培训管理及控制复杂系统的中心.通过将实际与人工系统的适当连接组合,可以使管理和控制实际复杂系统的有关人员迅速地掌握系统的各种状况以及对应的行动.在条件允许的情况下,应以与实际相当的管理与控制方式来运行人工系统,以期获得更佳的真实效果.同时,人工系统的管理与控制方案也可以作为实际复杂系统的备用方案,增加其运行的可靠性和应变能力.

3) 管理与控制:在这一过程中,人工系统试图尽可能地模拟实际复杂系统,对其行为进行预估,从而为寻找对实际系统有效的解决方案或对当前方案进行改进提供依据.进一步,通过观察实际系统与人工系统评估的状态之间的不同,产生误差反馈信号,对人工系统的评估方式或参数进行修正,减少差别,并开始分析新一轮的优化和评估.

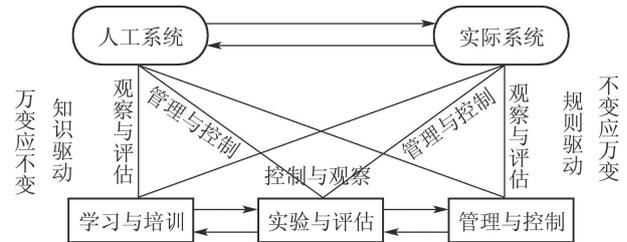


图 12 平行控制应用于复杂系统优化问题基本框架

Fig. 12 Parallel control for complex systems

由图12可以看出,平行控制过程是知识自动化过程的具体体现,是复杂系统知识自动化与智能优化的有效方法.综上,平行控制为求解复杂系统的优化控制问题提供了有效的解决方案.

7 结论与展望(Conclusion and future work)

控制理论经历了从经典控制到现代控制,再到智能控制,最后到平行控制的发展过程.经典控制最主要的贡献就是将反馈引入到控制中.经典控制的系统可由微分或差分等解析方程来描述,由于实际系统可以精确获得并且动态特性简单,故直接针对被控对象的机理设计控制器达到反馈控制的效果^[49].

随着控制系统复杂度的增加,实际系统解析模型难以获得,于是出现了自适应控制及内模控制方法.此时,系统控制已经从直接针对实际系统变成采用模型重构或预测的方式获得有效的控制策略.可以看到,现今许多智能控制方法与学习方法,如自适应动态规划方法^[50]、神经网络控制方法与模糊控制方法等,均可由自适应控制方法进行描述.

当系统变成复杂系统,尤其在包含社会复杂性与工程复杂性时,几乎无法建立可以逼近实际系统的模型,因此只能利用独立的人工系统,使实际与人工系统相互趋近,将人工系统、计算实验、平行执行成为独立组成部分获得复杂系统的优化策略,平行控制思想得到充分利用,其控制方法如图13所示.

现今阶段是人工智能高速发展的时期,同时社会的发展进步导致人们对控制系统的要求越来越高.随着社会与技术的发展,大数据时代已经来临,传统控制方法很难处理当前的海量数据,因此社会变革迫切需要一种数据驱动型的新型控制理念,这就是知识自动化.平行控制作为学习控制、智能机器、人工智能理念的总结升华,为求解复杂系统控制问题的提供新

的求解方案. 现今, 平行控制在交通系统^[51-52]、电网系统^[53]、过程控制系统^[54-55]等复杂CPSS系统中均取得了良好的控制优化效果.

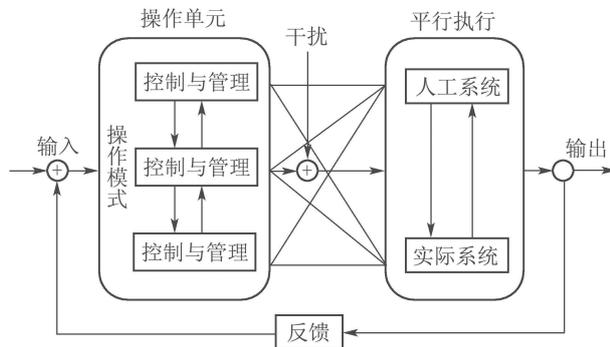


图 13 平行控制过程

Fig. 13 Parallel control diagram

本文以智能控制与学习优化的视角, 给出了平行控制在智能学习与优化控制方面应用发展的初步技术架构. 平行控制思想是从学习控制到人工智能控制的发展中形成一种学习控制方法论. 平行控制将实际系统与人工系统相结合, 用人工系统的计算实验完善实际系统优化控制策略, 帮助实现对复杂系统的有效控制. 现在, 平行控制理论的研究仍处于初始阶段, 如何将平行控制理论应用到实际生活中, 使得工业、交通、电网等领域内的控制更加智能化, 将是未来平行控制研究的重要方向. 随着平行控制理论不断丰富和完善, 平行控制思想将成为未来社会指导求解复杂系统问题的方法论, 平行控制将在复杂系统智能优化与控制方面将迎来更大的发展空间.

参考文献(References):

- 王飞跃. X5.0: 平行时代的平行智能体系 [J]. 中国计算机学会通讯, 2015, 11(5): 10 - 14.
- FU K S. Learning control systems and intelligent control systems: an intersection of artificial intelligence and automatic control [J]. *IEEE Transactions on Automatic Control*, 1971, 16(1): 70 - 72.
- SARIDIS G N. *Self-Organizing Control of Stochastic Systems* [M]. New York: Marcel Dekker, Inc, 1977.
- SARIDIS G N. Intelligent robotic control [J]. *IEEE Transactions on Automatic Control*, 1983, 28(5): 547 - 557.
- SARIDIS G N. Foundations of the theory of intelligent controls [C] // *Proceedings of IEEE Workshop on Intelligent Control*. Miami: IEEE, 1985: 23 - 28.
- WANG Feiyue. Parallel system methods for management and control of complex systems [J]. *Control and Decision*, 2004, 19(5): 485 - 489.
(王飞跃. 平行系统方法与复杂系统的管理和控制 [J]. 控制与决策, 2004, 19(5): 485 - 489.)
- WERBOS P J. Advanced forecasting methods for global crisis warning and models of intelligence [J]. *General Systems Yearbook*, 1977, 22(6): 25 - 38.
- CHEN J X. The evolution of computing: AlphaGo [J]. *Computing in Science & Engineering*, 2016, 18(4): 4 - 7.
- WEI Q, LIU D, LIU Y, et al. Optimal constrained self-learning battery sequential management in microgrids via adaptive dynamic programming [J]. *IEEE/CAA Journal of Automatica Sinica*, 2017, 4(2): 168 - 176.
- ZHU Huayong, NIU Yifeng, SHEN Lincheng, et al. State of the art and trends of autonomous control of UAV systems [J]. *Journal of National University of Defense Technology*, 2010, 32(3): 115 - 120.
(朱华勇, 牛轶峰, 沈林成, 等. 无人机系统自主控制技术的研究现状与发展趋势 [J]. 国防科技大学学报, 2010, 32(3): 115 - 120.)
- LIAO S K, LIN J, REN J G, et al. Space-to-ground quantum key distribution using a small-sized payload on tiangong-2 space lab [J]. *Chinese Physics Letters*, 34(9): 18 - 23.
- YIN Lei, HAN Jing, WANG Ye, et al. Study on virtual hand control for virtual reality [J]. *Journal of System Simulation*, 2009, 21(2): 448 - 456.
(殷磊, 韩静, 王烨, 等. 虚拟现实环境下虚拟手控制技术的研究 [J]. 系统仿真学报, 2009, 21(2): 448 - 456.)
- 王飞跃. 工业5.0到智业1.0: 平行的本源、现状、展望 [R]. 北京: 航天智慧产业专题研讨会, 2014.
- 王飞跃. 复杂性研究与智能产业: 平行企业与工业5.0 [R]. 上海: 控制工程师峰会, 2014.
- DENG Jianling, WANG Feiyue, CHEN Yaobin, et al. From industries 4.0 to energy 5.0: concept and framework of intelligent energy systems [J]. *Acta Automatica Sinica*, 2015, 41(12): 2003 - 2016.
(邓建玲, 王飞跃, 陈耀斌, 等. 从工业4.0到能源5.0: 智能能源系统的概念、内涵及体系框架 [J]. 自动化学报, 2015, 41(12): 2003 - 2016.)
- WANG Feiyue, LIU Derong, XIONG Gang, et al. Parallel control theory of complex systems and applications [J]. *Complex Systems and Complexity Science*, 2012, 9(3): 1 - 12.
(王飞跃, 刘德荣, 熊刚, 等. 复杂系统的平行控制理论与应用 [J]. 复杂系统与复杂性科学, 2012, 9(3): 1 - 12.)
- FU K S. Learning control systems - review and outlook [J]. *IEEE transactions on Automatic Control*, 1970, 15(2): 210 - 221.
- GILSTAD D W, FU K S. Two-dimensional adaptive model of a human controller using pattern recognition techniques [J]. *IEEE Transactions on Systems Man & Cybernetics*, 1971, 1(3): 261 - 266.
- WANG F Y, SARIDIS G N. A coordination theory for intelligent machines [J]. *Automatica*, 1990, 23(8): 235 - 240.
- ZADEH L A. Fuzzy sets [J]. *Information and Control*, 1965, 8: 338 - 365.
- HOLLAND J H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Application to Biology, Control, and Artificial Intelligence* [M]. Ann Arbor, MI: University of Michigan Press, 1975: 439 - 444.
- KIRKPATRICK S, GELATT C D, VECCHI M P. Optimization by simulated annealing [J]. *Science*, 1983, 220(4598): 671 - 680.
- EBERHART R, KENNEDY J. A new optimizer using particle swarm theory [C] // *Proceedings of the Sixth International Symposium on IEEE Micro Machine and Human Science*. Nagoya: IEEE, 1995: 39 - 43.
- MCCULLOCH W S, PITTS W. A logical calculus of the ideas immanent in nervous activity [J]. *The Bulletin of Mathematical Biophysics*, 1943, 5(4): 115 - 133.
- HOPFIELD J J. Neural networks and physical systems with emergent collective computational abilities [J]. *Proceedings of the National Academy of Sciences*, 1982, 79(8): 2554 - 2558.
- SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of go without human knowledge [J]. *Nature*, 2017, 550(7676): 354 - 359.
- SILVER D, HUBERT T, SCHRITTWIESER J, et al. Mastering chess and shogi by self-play with a general reinforcement learning algorithm [J]. *Artificial Intelligence*, 2017, arXiv:1712.01815.

- [28] BERTSEKAS D P, TSITSIKLIS J N. *Neuro-dynamic Programming* [M]. Belmont, MA: Athena Scientific, 1996.
- [29] BELLMAN R E. *Dynamic Programming* [M]. Princeton, NJ: Princeton University Press, 1957.
- [30] PROKHOROV D V, WUNSCH D C. Adaptive critic designs [J]. *IEEE Transactions on Neural Networks*, 1997, 8(5): 997 – 1007.
- [31] WEI Q, LIU D, LIN H. Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems [J]. *IEEE Transactions on Cybernetics*, 2016, 46(3): 840 – 853.
- [32] WEI Q, SONG R, YAN P. Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 27(2): 444 – 458.
- [33] WEI Q, LEWIS F L, SHI G, et al. Error-tolerant iterative adaptive dynamic programming for optimal renewable home energy scheduling and battery management [J]. *IEEE Transactions on Industrial Electronics*, 2017, 64(12): 9527 – 9537.
- [34] ZHONG X, NI Z, HE H. GR-GDHP: a new architecture for globalized dual heuristic dynamic programming [J]. *IEEE Transactions on Cybernetics*, 2017, 47(10): 3318 – 3330.
- [35] LEWIS F L, VRABIE D, VAMVOUDAKIS K G. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers [J]. *IEEE Control Systems*, 2012, 32(6): 76 – 105.
- [36] ZHANG H, JIANG H, LUO C, et al. Discrete-time nonzero-sum games for multiplayer using policy-iteration-based adaptive dynamic programming algorithms [J]. *IEEE Transactions on Cybernetics*, 2017, 47(10): 3331 – 3340.
- [37] ABU-KHALAF M, LEWIS F L. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach [J]. *Automatica*, 2005, 41(5): 779 – 791.
- [38] SONG R, LEWIS F L, WEI Q. Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(3): 704 – 713.
- [39] WEI Q, LIU D. A novel policy iteration based deterministic Q -learning for discrete-time nonlinear systems [J]. *Science China Information Sciences*, 2015, 58(12): 1 – 15.
- [40] WEI Q, LEWIS F L, SUN Q, et al. Discrete-time deterministic Q -learning: a novel convergence analysis [J]. *IEEE Transactions on Cybernetics*, 2017, 47(5): 1224 – 1237.
- [41] ZHANG H, LIANG H, WANG Z, et al. Optimal output regulation for heterogeneous multiagent systems via adaptive dynamic programming [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(1): 18 – 29.
- [42] ZHANG J, ZHANG H, FENG T. Distributed optimal consensus control for nonlinear multiagent system with unknown dynamic [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, PP(99): 1 – 10.
- [43] WEI Q, LEWIS F L, LIU D, et al. Discrete-time local value Iteration adaptive dynamic programming: convergence analysis [J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, 48(6): 875 – 891.
- [44] WANG Feiyue. Parallel control: a method for data-driven and computational control [J]. *Acta Automatica Sinica*, 2013, 39(4): 293 – 302.
(王飞跃. 平行控制: 数据驱动的计算控制方法 [J]. 自动化学报, 2013, 39(4): 293 – 302.)
- [45] LIU Xin, WANG Xiao, ZHANG Weishan, et al. Parallel data: from big data to data intelligence [J]. *Pattern Recognition and Artificial Intelligence*, 2017, 30(8): 673 – 681.
(刘昕, 王晓, 张卫山, 等. 平行数据: 从大数据到数据智能 [J]. 模式识别与人工智能, 2017, 30(8): 673 – 681.)
- [46] WANG F Y. The emergence of intelligent enterprises: from CPS to CPSS [J]. *IEEE Intelligent Systems*, 2010, 25(4): 85 – 88.
- [47] WANG F Y, ZHANG J, WEI Q L, et al. PDP: parallel dynamic programming [J]. *IEEE/CAA Journal of Automatica Sinica*, 2017, 4(1): 1 – 5.
- [48] LI Li, LIN Yilun, CAO Dongpu, et al. Parallel learning — a new framework for machine learning [J]. *Acta Automatica Sinica*, 2017, 43(1): 1 – 8.
(李力, 林懿伦, 曹东璞, 等. 平行学习——机器学习的一个新型理论框架 [J]. 自动化学报, 2017, 43(1): 1 – 8.)
- [49] ZHENG Dazhong. *Linear System Theory* [M]. 2nd Edition. Beijing: Tsinghua University Press, 2002.
(郑大钟. 线性系统理论 [M]. 第2版. 北京: 清华大学出版社, 2002.)
- [50] WEI Q, LIU D, LIN Q, et al. Discrete-time optimal control via local policy iteration adaptive dynamic programming [J]. *IEEE Transactions on Cybernetics*, 2017, 47(10): 3367 – 3379.
- [51] MO Hong, HAO Xuexin. Linguistic dynamic analysis of traffic light timing design within the time-varying universe [J]. *Acta Automatica Sinica*, 2017, 43(12): 2202 – 2212.
(莫红, 郝学新. 时变论域下红绿灯配时的语言动力学分析 [J]. 自动化学报, 2017, 43(12): 2202 – 2212.)
- [52] FANG Bing, ZHANG Cuixia. Urban transportation decision support platform based on parallel simulation [J]. *Command Information System and Technology*, 2017, 8(3): 16 – 21.
(方冰, 张翠侠. 基于平行仿真的城市交通运输决策支持平台 [J]. 指挥信息系统与技术, 2017, 8(3): 16 – 21.)
- [53] LIU Jinchang, YANG Desheng, SUN Fei, et al. Parallel grid system framework research [J]. *Electric Power Information and Communication Technology*, 2016, 14(8): 7 – 13.
(刘金长, 杨德胜, 孙飞, 等. 平行电网体系框架研究 [J]. 电力信息与通信技术, 2016, 14(8): 7 – 13.)
- [54] FAN Keyu, ZHU Lin. The application of parallel control theory in NeFeB hydrogen decrepitation system [J]. *Chinese Rare Earths*, 2016, 37(6): 65 – 70.
(樊可钰, 朱林. 平行控制理论在钕铁硼氢粉破碎控制系统中的应用 [J]. 稀土, 2016, 37(6): 65 – 70.)
- [55] XIONG Gang, WANG Feiyue, ZOU Yumin, et al. Parallel evaluation method to improve long period ethylene production management [J]. *Control Engineering of China*, 2010, 17(3): 401 – 406.
(熊刚, 王飞跃, 邹余敏, 等. 提升乙烯长周期生产管理的平行评估方法 [J]. 控制工程, 2010, 17(3): 401 – 406.)

作者简介:

王飞跃 (1961–), 男, 研究员, 博士生导师, 主要研究方向为智能机器、社会计算、平行控制、知识自动化、平行智能等, E-mail: feiyue.wang@ia.ac.cn;

魏庆来 (1979–), 男, 研究员, 博士生导师, 主要研究方向为智能系统、平行控制、知识自动化、平行智能等, E-mail: qinglai.wei@ia.ac.cn.