

游戏智能中的 AI——从多角色博弈到平行博弈

沈宇^{1,2}, 韩金朋³, 李灵犀⁴, 王飞跃^{1,2,5}

- (1. 中国科学院大学人工智能学院, 北京 100049;
2. 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;
3. 西安交通大学软件学院, 陕西 西安 710049;
4. 印第安纳大学-普渡大学印第安纳波利斯分校电子与计算机工程系, 美国 印第安纳波利斯 IN 46202;
5. 青岛智能产业技术研究院, 山东 青岛 266109)

摘要: 总结了国内外人工智能技术在游戏领域的研究进展, 分析了游戏领域的研究进步对于现实社会的意义。针对强化学习中免模型方法存在的仿真与真实的鸿沟、基于模型的方法缺乏通用性的问题, 提出平行博弈的思想和方法, 介绍了平行博弈在解决现有单角色博弈和多角色博弈问题上的先进之处。认为虚实结合的平行博弈方法将成为迈向通用人工智能的基石。

关键词: 游戏智能; 多角色博弈; 平行博弈; 深度强化学习; 人工智能

中图分类号: TP305

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202023

AI in game intelligence—from multi-role game to parallel game

SHEN Yu^{1,2}, HAN Jinpeng³, LI Lingxi⁴, WANG Fei-Yue^{1,2,5}

1. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
2. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
3. School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China
4. Department of Electrical and Computer Engineering, Indiana University-Purdue University Indianapolis, Indianapolis IN 46202, USA
5. Qingdao Academy of Intelligent Industries, Qingdao 266109, China

Abstract: The domestic and overseas research progress of artificial intelligence technology in the field of games was summarized and the significance of the research progress in the field of games for real life was analyzed. In view of the gap between simulation and reality in model based methods and the lack of generality of the model-based approach in reinforcement learning, the idea and method of parallel game were put forward, and the advance of parallel game in solving the existing problems of single-role game and multi-role game was introduced. The parallel game method will be the cornerstone of the general artificial intelligence.

Key words: game intelligence, multi-role game, parallel game, deep reinforcement learning, artificial intelligence

1 引言

2019年7月, 由脸书和卡耐基梅隆大学合作开发的新型人工智能系统 Pluribus 扑克机器人, 在6人无限制德州扑克比赛中击败了15名顶尖选手, 其中包括多位世界冠军^[1]。这是人工智能 (artificial intelligence, AI) 首次在超过两人的复杂对局中击败人类

顶级玩家, 突破了过去 AI 仅能在国际象棋等二人游戏中战胜人类的局限, 成为机器在游戏中战胜人类的又一个里程碑性的工作, 被《科学》杂志评为2019年的十大科学突破之一。《科学》杂志的论文显示, Pluribus 与13名德州扑克高手进行了1万次不限注对局的六人桌比赛, 且 Pluribus 在与5名人类选手的对抗中获胜。同时, 在由5个 Pluribus 和1名人类选手

组成的对局中,Pluribus 分别在 5 000 手对局中先后击败了德州扑克世界冠军达伦·伊莱亚斯和克里斯·弗格森。这不是机器与人对抗的第一次胜利,早在 2016 年,国际顶尖期刊《自然》杂志就报道了由谷歌旗下的 DeepMind 公司创始人戴密斯·哈萨比斯领衔的团队开发的名为“阿尔法围棋(AlphaGo)”的人工智能系统,该系统分别击败世界围棋冠军李世石和欧洲围棋冠军樊麾^[2]。AlphaGo 专注于一对一(1v1)的围棋模式,系统可分为 3 个部分:策略部分用于判断“下一步走哪”,价值部分用于判断“当前局势如何”,组合策略部分则用于判断上一步策略是否正确。

多智能体对抗博弈策略在实际领域具有应用价值,美国军方连续几年发布无人系统路线图,将多无人系统在战场上的协同作战列为重点研究方向,并开展多项以多机器人系统或集群作战为内容的军事研究项目。俄罗斯军方也已将多无人系统应用于战场。

2 游戏智能

根据前瞻产业技术研究院的统计数据,2011—2018 年我国的游戏市场销售收入逐年增长,2018 年年底实际的销售收入达到 2 144.4 亿元。从细分类型来看,动作角色扮演类游戏收入最高。作为移动互联网产业的重要组成部分,游戏市场的规模不断发展壮大,其发展的重要推动力便是当下以智能机器人、增强现实(augmented reality, AR)、语音识别、模式识别等为代表的人工智能技术。游戏的英文名是“game”,它又有博弈的意思。

传统的游戏智能技术包括有限状态机、脚本语言、模糊逻辑、决策树和遗传算法。然而,它们通常依赖于大量的人工干预(例如人工设计的特征与动作空间),同时对模型的性能以及游戏的结果提升也有一定的限制^[3]。DeepMind 的创始人、英国《商业评论》的作者 Hassabis 等人^[4]基于世界冠军和人工智能之间的竞争——深蓝(Deep Blue),分析了用于国际象棋的人工智能。深蓝的利弊在 Hassabis 这位前国际象棋世界冠军自己的观点中得到了体现,深蓝还被拿来与更先进的人工智能 AlphaGo 相比,后者更智能、更通用,他认为人工智能的发展正在促进如今的技术进步。参考文献[5]提供了一个国际象棋引擎,它使用自我对战来发现其领域内特定的知识,而程序员只提供了有限的知识。他们还开发了另一种基于概率搜索框架的机器学习系统,该系统具有更高的效率,从而提高了游戏强度。该

引擎已被评估并证明与官方评级在前 2.2%的国际象棋锦标赛选手相当。Ruben Rodriguez T 等人^[6]描述了通用的电子游戏 AI(general video game AI, GVGA),它提供了一种对许多应用领域使用特定描述语言编写的游戏的 AI 算法进行基准测试的方法。他们还将 GVGA 与 Open AI Gym 环境结合,实现了针对多款游戏的深度强化学习算法。作者还对结果进行了分析,并对不同智能体(agent)在不同游戏中的表现进行了评价。参考文献[7]介绍了 3 个主要的年度《星际争霸》人工智能竞赛:学生《星际争霸》人工智能比赛(student starcraft AI tournament, SSCAIT)、游戏中的计算智能竞赛(computational intelligence in games competition, CIG)以及人工智能和交互式数字娱乐(artificial intelligence and interactive digital entertainment, AIIDE)大会。作者分析了这些竞赛的现状以及人工智能机器人在竞赛中的竞争。参考文献[8]分析了 AlphaStar,这是一个针对游戏《星际争霸 II》的专业 AI。AlphaStar 最有趣的方面包括:使用拉马克进化、竞争性共同进化和质量多样性。作者希望提供一个进化计算领域和主要 AI 系统之间的枢纽。参考文献[9]从 AI 的角度定义了即时策略(real-time strategy, RTS)游戏,并总结了 RTS 游戏 AI 中存在的开放性问题以及针对这些问题所做的工作。作者认为 RTS 游戏可以被看作在有限且更小的世界中对真实复杂动态环境的模拟。在 RTS 游戏中寻找有效的技术来解决这些问题,这将有益于其他 AI 学科和应用领域。参考文献[10]试图研究 AI 和视频游戏之间的关系。其讨论内容包括一般的电子游戏如何成为人工智能的理想测试平台,以及加入更高级的人工智能技术后游戏的表现形式。

图灵在 1950 年提出著名的“图灵测试”,用于检验机器是否具有智能^[11]。“图灵测试”在游戏中可以被描述为:当玩家与机器在同时游戏时,如果玩家通过任何手段都无法判断哪个角色是玩家,哪个角色是机器,就说明机器通过了“游戏中的图灵测试”。与棋类游戏相比,多角色游戏中多智能体 RTS 游戏的研究难度更大,主要体现在如下几点:

- 多玩家共存及多异构智能体博弈;
- 实时对抗及动作持续性;
- 非完整信息博弈和强不确定性;
- 巨大的搜索空间和多复杂任务。

参考文献[12]总结了多智能体 RTS 游戏的难点与人工智能研究挑战的对应关系,如图 1 所示。

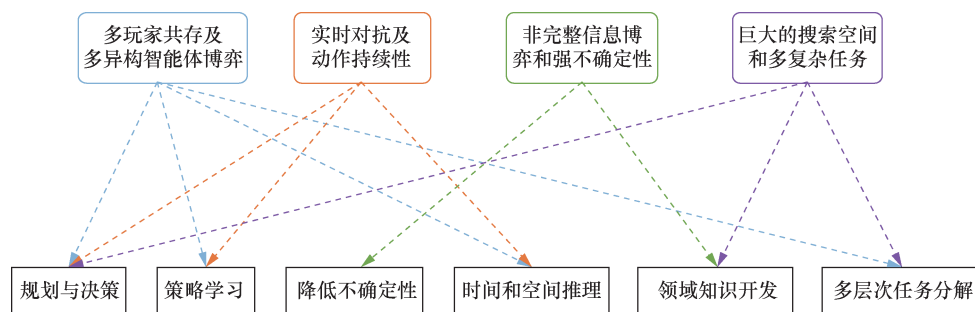


图1 多智能体 RTS 游戏的难点与人工智能研究挑战的对应关系

多智能体博弈游戏具有实时对抗、群体协作、非完全信息对称、搜索空间巨大、时间-空间推理任务复杂等特点。

2.1 单角色博弈

谷歌的 DeepMind 使用强化学习玩打砖块的游戏。模型将图像像素作为输入，在无任何先验知识的条件下指导 AI 学习，模型中的 agent 被随机初始化后与环境进行交互，得到的分数被作为奖励，在进行多次循环后得到一系列训练样本并被用于训练神经网络，最终得出决策模型^[13]。

卡耐基梅隆大学之前开发过 Libratus，这是一个玩无限德州扑克的人工智能程序，也是 Pluribus 的前身^[14]。Libratus 进行的是 1v1 的比赛，没有用到专业的牌局进行神经网络的训练，不同于 AlphaGo 用了大量的棋局做训练，这个模型用随机生成的牌局（随机产生公共牌、底池筹码、玩家拿牌概率）和尝试性的动作带来的结果（在随机生成的输入情况下模拟玩家跟牌后的结果）作为训练数据。Libratus 不再积累大量对局数据，而是平衡风险与收益，使其达到纳什均衡，从而选择策略。Libratus 有一个独特的自学习模块，可以随着时间的推移逐渐增强策略蓝本的计算水平，当部分博弈树中对手有可能在它的策略中发现潜在的漏洞时，它可以计算出更接近纳什均衡的近似解。模型中同样未加入强化学习，Libratus 使用蒙特卡洛反事实遗憾最小化（counterfactual regret minimization, CFR）的一种变体，计算策略蓝本，处理实时的子博弈求解。CFR 是一种自我博弈的算法，与强化学习类似，但是 CFR 会更多地考虑在自我对局中未被选择的那些选项，并假设选了之后会有什么样的投入产出。

当前，应用 AI 的大型、对抗性游戏主要有以《星际争霸》为首的 RTS 游戏和以《DOTA》为首的多人在线战术竞技（multi-player online battle arena，

MOBA）游戏。《星际争霸》是一款 RTS 游戏，其单个单元较为简单，主要学习兵力的组合和进攻策略，游戏规则就是造农民，采矿，造建筑、兵和武器，升级科技，最后拆掉对方建筑；而在《DOTA》游戏中，选手通过打怪兽、击杀对手获得经验升级等级和金钱升级装备，基于选手的操作和团队配合推翻敌方战略要地，获得胜利，单个英雄操作复杂，玩家主要学习如何操控以及团队配合。《星际争霸》和《DOTA》游戏的共同特点都是由于战争迷雾导致的部分信息不可知问题。

我国在游戏智能领域的研究起步较国外较晚，但是也取得了很好的成绩。中国科学院自动化研究所的张俊格团队设计开发的 CSE bot 在 2018 年的第八届 AIIDE《星际争霸》AI 比赛中以 87.11% 的胜率取得大赛季军。在《星际争霸》游戏中，玩家需要收集自然资源建造防御和作战单元，并在最短时间内消灭对手取得胜利。自 2010 年首届《星际争霸》AI 比赛开始，参赛者提交《星际争霸》AI bot，这些人工智能体将会自主进行厮杀。开发者使用 BWAPI（Brood War API）控制《星际争霸》AI 智能体，BWAPI 提供了一个 C++ 的编程接口，可以自由编程，生成动态链接库 DLL 格式的 AI 文件。

在 2020 年 6 月 21 日的北京智源大会上，启元 AI “星际指挥官”以 2:0 完胜人类选手《星际争霸 I/II》全国冠军黄慧明，以及中国星际最强人族选手、黄金总决赛三连冠的李培楠。在比赛过程中，AI 不断使用新的打法和策略，稳定发挥，其表现与一个极其冷静且双商极高的人类类似。启元 AI 分别从工程和算法层面进行深层次优化，最终仅用 1% 的算力便达到了国外顶尖机器智能的水平。

腾讯 AI Lab 利用深度强化学习在《王者荣耀》1v1 游戏虚拟环境中构建“绝悟”AI，开发高扩展、低耦合的强化训练系统，使得 AI 能够具有进攻、诱导、防御、欺骗和技能连招释放的能力^[15]。不同

于星际 RTS 研究智能体之间的协作策略,“绝悟”AI 更关注 agent 复杂的动作控制,MOBA 一对一类型的游戏的复杂性来自巨大的动作和状态空间以及游戏机制。主要可概括为:

- agent 需要在部分可观测环境中完成决策过程;
- 小兵、炮塔等单元给 agent 目标选择带来挑战,因此需要更精细的决策和动作控制;
- 不同英雄玩法差别大,因此需要进行统一的、具有鲁棒性的建模;
- 缺乏高质量数据,监督学习方法变得不可行。

“绝悟”AI 借助控制依赖解耦、动作屏蔽、目标注意力和双裁剪最近策略 (dual clip proximal policy algorithm, PPO), 训练用于对 MOBA 动作控制进行建模的 actor-critic 神经网络, 实现对大规模多智能体竞争环境的有效探索。

如图 2 所示,“绝悟”AI 系统架构由 4 个模块组成: 游戏环境 AI 服务器实现 AI 模型与环境交互的方式; 调度模块用于样本收集、压缩和传输; 内存池存储数据, 为强化学习器提供训练样例; 强化学习器训练模型, 并以点对点的方式同步到游戏环境 AI 服务器中。“绝悟”AI 使用 384 块 GPU, 日均对战局数相当于人类 440 年的对战局数。在训练半个月以上之后, 其在与大量顶级业余玩家的 2 100 场

对战中取得了 99.81% 的胜率。

2016 年, 基于游戏的 ViZDoom AI 竞赛诞生, 作为第一人称射击类游戏 (first-person shooter game, FPS), 该比赛首先搜索输入像素级视觉信息, 直接输出 AI 控制策略的强化学习算法。比赛共分为两个挑战: Track 1 为单人闯关模式, 考核标准为在最短时间内闯最多的关, 不同于以往的死亡竞赛, 该部分需要 AI 能同时完成搜索路径、收集装备、躲避陷阱、寻找出口等诸多复杂任务, 对 AI 的任务理解和环境认知能力要求极高; Track 2 为随机对战模式, 这是 ViZDoom 的传统项目, 采用死亡竞赛模式, 要求参赛选手在同一个地图里对杀 10 min, AI 需要在保存自身实力的同时, 尽量多杀敌人。单玩家场景如图 3 所示。

清华大学 TSAIL 团队与腾讯 AI Lab 合作^[16], 获得 2018 年比赛的初赛和决赛冠军。在 Track 1 中, TSAIL 团队提出环境信息引导的分层强化学习技术, 在对环境信息进行有效感知的基础上, 融合环境反馈和强化学习的奖励信号, 引导分层强化学习训练; 在 Track 2 中, TSAIL 团队针对目前强化学习中普遍存在的动作空间大和奖励信号稀疏问题, 更改适配轻量级物体的检测框架 YOLO-V3, 并与强化学习算法有效融合, 极大地提高了强化学习的训练效率。

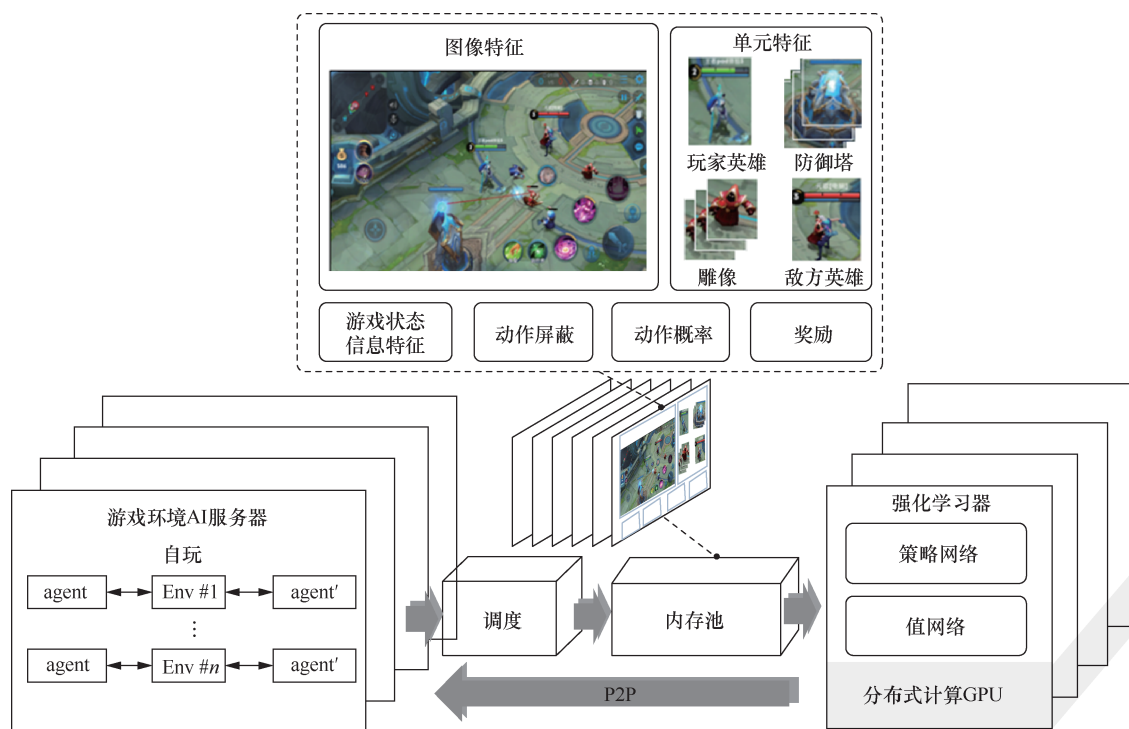
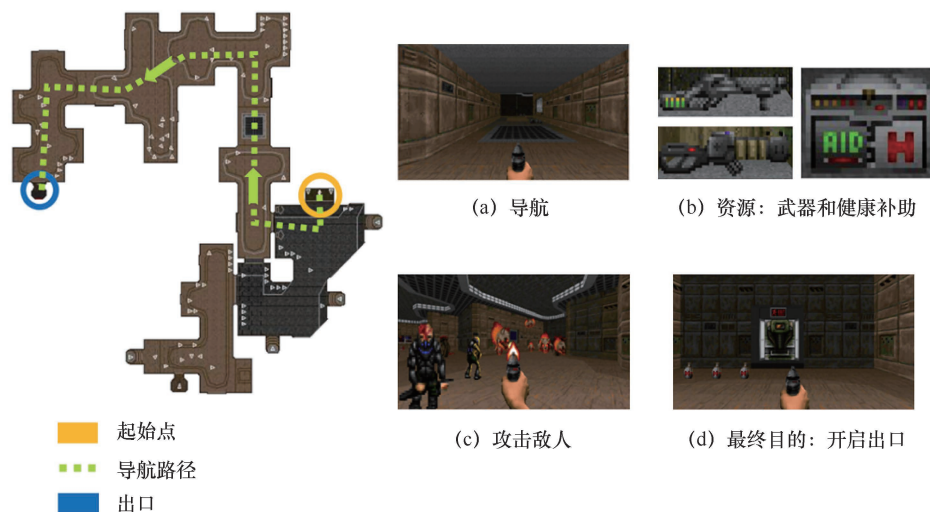


图2 “绝悟”AI 系统架构^[15]

图3 单玩家场景^[16]

针对目前游戏AI中出现的挑战，腾讯AI Lab提出了相应的解决方案。

- 游戏的状态空间过大。许多战略性游戏的状态空间远大于围棋空间，腾讯提出一套整合了算法、模型、计算机体系结构的方案，即腾讯大脑，其核心是利用深度神经网络建模超大规模的状态空间。

- 复杂的多玩家游戏需要多智能体协调操作。腾讯提出将强化学习的价值网络与描述宏观战略的行为树进行互相操作与融合，设计出完善的多智能体AI。

- 绝大部分游戏AI的模拟器过于理想化。为打通虚拟世界与现实世界，需要结合反向强化学习即动态探索机制，对游戏AI中的参数进行贝叶斯升级，从而保证在现实世界中用最小的成本即可完成部署。

2.2 多角色博弈

近几年AI研究者在不同的复杂游戏博弈中都取得了重要突破，比如围棋、二人扑克、《星际争霸II》和《Dota2》等。然而，为什么Pluribus被如此关注，还被视为AI领域的重大突破？因为这些游戏绝大多数被限制在二人玩家的零和博弈与完备信息框架之内。尽管基于强化学习的AI策略在《Dota2》Five和《Quake III》等多玩家游戏博弈问题上已取得了相当的进展，但对于AI来说，6个玩家的德州扑克问题依然是非常有挑战性的问题^[17]。主要原因如下。

- 必须处理不完备信息，玩家需要在不知道对手策略和资源的情况下进行决策，需要在不同的子

博弈之间寻找平衡。

- 很难达到纳什均衡，纳什均衡的复杂度随着玩家数量的增加而呈指数性增加，在算力上几乎不可能实现。

- 需要使用“诈唬(bluffer)”等心理技巧，真实比赛中需要有效推理隐含信息、寻找让对手无法预测的策略，懂得“诈唬”技巧是成功的关键之一。

Pluribus不用任何先验信息、从头开始学习，其以自玩的模式训练AI。其首次在6个玩家、无限制的德州扑克复杂问题中击败国际顶尖玩家，获得胜利。对多玩家场景下的博弈问题的研究贡献，是Pluribus成为2019年“十大科学突破”的关键。

相较于其他AI博弈系统，Pluribus具有“诈唬”手段。专业玩家承认在抵制机器的行动中遇到麻烦，Pluribus是个怪兽级的欺骗者，而且大多数情况下机器的欺骗手段是更加高效的，这也是机器难被击败的原因。尽管很多情况下明知道机器在使用欺骗手段，玩家在与之对战时还是感受到了巨大压力，Pluribus被训练得让人类难以预测。

Pluribus是Libratus的增强版本，Pluribus在Libratus的基础上增加了一个新的在线搜索算法，可以通过搜索前面几个游戏的步骤，来评估自己的下一步战术。同时，Pluribus还拥有比Libratus更快的自玩算法。在线搜索算法和自玩算法的更新与结合，使得Pluribus能用比Libratus更少的处理能力和内存进行训练。此外，Pluribus也利用动作抽象和信息抽象来推断游戏中未来几轮的下注情况以及批量计算相似的牌，它同样也使用了CFR算法进行循环推理，不断自我博弈并实现自我改进。

3 平行博弈

从广义上讲, 强化学习是序贯决策问题, 通常被纳入马尔可夫决策过程的框架下解决。根据马尔可夫决策过程的状态转移概率 P 是否已知, 强化学习可以被分为基于模型的动态规划和基于无模型的强化学习。前者源自最优控制领域, 常用模型预测控制(model predictive control, MPC)、迭代学习控制(iterative learning control, ILC)等最优控制方法解决; 后者更多发展自机器学习领域, 属于数据驱动方法, 算法通过采样估计智能体的状态、动作值函数或奖励函数, 从而优化动作策略。

免模型的数据驱动强化学习方法需要大量的采样数据。而现实生活中的机械控制却难以获取如此多的数据, 只能使用仿真环境的模拟器进行训练, 因此, 模拟器与现实世界间的鸿沟成了制约算法性能的重要因素, 直接限制了算法的泛化性能。此外, 由于没有模型, 算法的可解释性也就不强, 算法调试变得很困难, 鲁棒性不高。

基于模型的动态规划方法对模型进行优化, 可以极大地提升数据利用率, 对于环境变化的鲁棒性更高。但是该算法严重依赖模型的建立, 建模的误差会直接影响算法的收敛, 因此缺乏通用性。基于模型的方法与无模型的方法的区别可被看作基于知识的方法与基于统计的方法的区别。

计算机不能有效处理感知信息, 根本原因在于不能对真实场景的基本特征进行可靠提取, 缺乏对真实场景基本特征的一般表达以及对不同模式下

信息特征有效整合的理论。本文提出平行博弈的理论框架, 希望能对现有的强化学习方法中存在的问题提供解决思路。

平行理论是王飞跃教授在钱学森、于景元、戴汝为提出的综合集成科学思想和综合研讨体系技术的基础之上, 把信息、心理、仿真、决策融为一体, 以可计算、可操作、可实现的方式, 为研究复杂性和控制与管理复杂系统提供一种思路及方法^[18]。早在2010年的“亚洲游戏(Gameon Asia)”会上, 王飞跃教授在题为“面向科学的游戏: 基于ACP的方法(toward scientific games: an ACP-based approach)”的报告中就提出将平行方法应用于计算机游戏和博弈问题中的思想, 即平行博弈^[19], 平行博弈系统架构如图4所示。平行博弈数字孪生由真实空间和认知空间的描述系统、预测系统和引导系统构成^[20], 系统架构如图5所示。

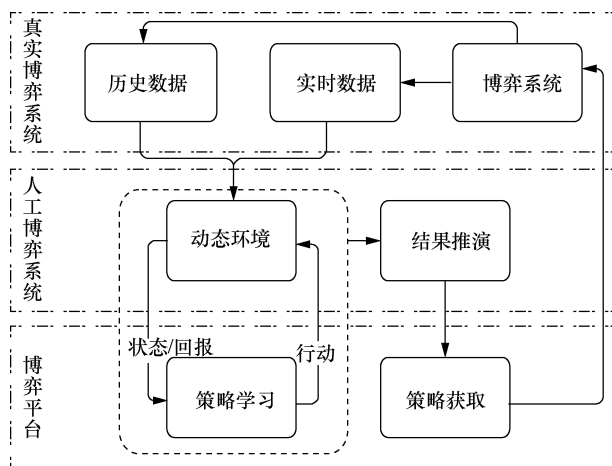


图4 平行博弈系统架构

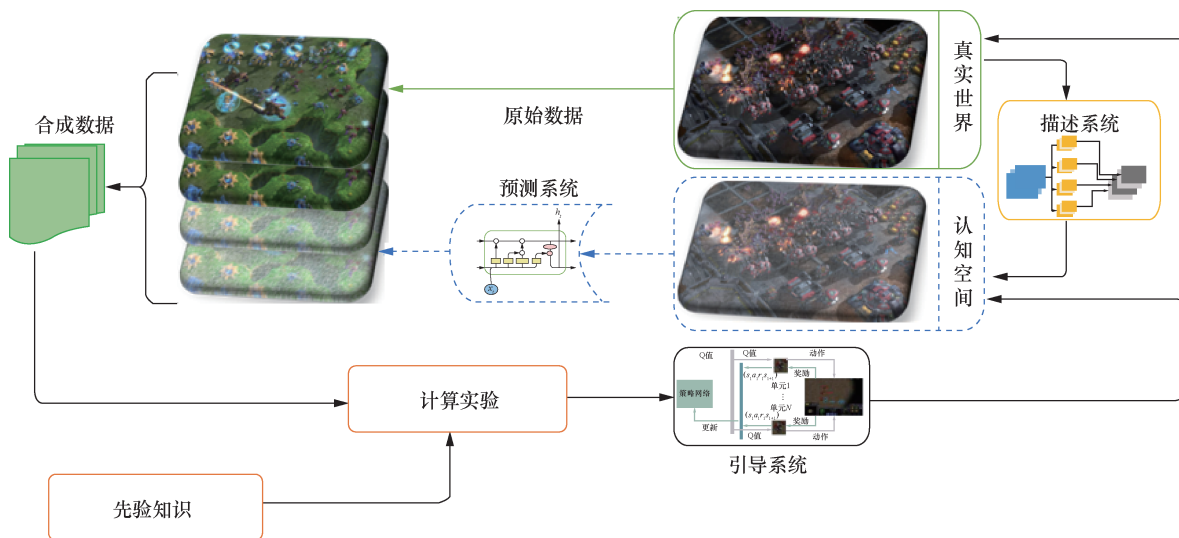


图5 平行博弈数字孪生系统架构

平行博弈将真实世界中的问题抽象到认知空间，在认知空间中建立模型，进行计算实验，用得到的结果指导真实世界中的策略执行，真实世界的执行结果反过来更新认知空间的模型，系统以这种虚实互动、平行执行的方式不断迭代，直至收敛。

3.1 描述系统

描述系统采用两种方法将真实世界的对象抽象到认知空间中。一是基于统计的无模型方法。在智能游戏领域，系统的输入为非结构化、不可以显式表达的数据，例如棋盘的牌面局势、FPS的当前场景都以图片形式进行表达。此时需要使用卷积神经网络（convolutional neural network, CNN）或者图神经网络（graph neural network, GNN）对采样得到的大量数据进行特征提取，完成对真实世界的抽象。二是基于模型的参数化方法。针对自动驾驶汽车、智能机械臂控制等问题，往往难以完成大量数据的采样，但此类问题往往容易找出对应的动力学模型。无模型的方法弱点在于对大量数据的依赖，而有模型的方法的弱点则是比较固定，不够灵活。因此针对一些可以建模也可以采集样本数据的问题，可以将这两种方法结合，将统计特征作为对模型的补充，从而提升模型在新的任务上的适应性和鲁棒性。描述系统可以被归纳为建立知识系统的过程，描述系统是不断更新的、可控的、动态的，而仿真系统则是静态的、不可控的。

3.2 预测系统

预测系统的主要功能是基于描述系统建立的模型，在认知空间使用计算试验方法仿真预测可能的未来状态。传统强化学习方法采用一种试错的方法，根据与环境交互的奖励决定策略。但是在自动驾驶以及一些涉及道德伦理的问题上，试错的方法显然是不被接受的，因此，就需要基于描述系统建立的模型做推演预测，为可行的策略集的制定提供依据。预测系统是一个将小知识点生成大数据的过程，是一个推理生成过程。对于棋牌等完备信息空间来说，预测系统是一个推理演示的过程；而对于星际争霸等由于战争迷雾导致的信息不可知问题来说，预测系统则是一个预测过程，例如绝地求生游戏中可以使用光流方法对场景中动态物体的运行进行预测。常用方法有长短期记忆（long short-term memory, LSTM）网络、变分自编码器（variational auto-encoder, VAE）以及生成式对抗网络（generative adversarial network, GAN）。预测系

统对由小知识点生成的大数据进行处理，完成状态到动作的映射，由大数据生成小智能。

3.3 引导系统

引导系统是从虚拟到现实交互和部署的重要步骤。引导系统是一个认知推理的过程，智能体探索平行系统间差异最大的动作空间，达到降低人工系统建模误差和智能体执行误差的目的。引导系统最主要的特征是反事实推理，由预测到的状态映射而来的动作策略需要经过反事实推理来剔除其中不可行的策略，将优化过的策略迁移到真实世界中指导真实系统的运行。

3.4 平行博弈总结

Pearl J 等人在参考文献[21]中把因果论分为了3个层级，称之为“因果关系之梯”：第一层级研究“关联”，第二层级研究“干预”，第三层级研究“反事实推理”。Pearl J 等人特别指出，当前的AI和机器学习只处于第一层级，只是被动地接受观测结果，考虑的是“如果我看到……会怎样”这类问题。处于第二层级的“干预”则与主动实施某个行动有关，考虑的是“如果我做了……将会怎样”“如何做”这类更高级的问题。第三个层级的“反事实推理”在现实世界里并不存在，它是想象的产物。

反事实推理处于因果关系之梯的最高层，其典型问题是“假如我做了……会怎样？为什么？”，这类问题属于反思性问题。反事实推理对于人类来说是家常便饭，但是对于机器来说就没那么容易了。历史学家尤瓦尔·赫拉利认为人类发展出描绘虚构之物的能力正是人类进化过程中的认知革命，反事实推理是人类独有的能力，也是真正的智能，借助它，人类才可以超越现实。

平行博弈的描述系统、预测系统和引导系统分别对应于Pearl J 等人提出的因果论的关联、干预和反事实推理3个部分。描述系统将现实世界的问题在虚拟世界中进行精确的描述并使之可控；预测系统则基于过去情景以及先验知识、专家经验等对未来的进程进行推演并产生相应的动作序列；引导系统需要判断预测系统给出的结果是否是可行的。整个过程在认知空间中实现，使机器具有了认知、反事实推理的能力。

4 结束语

现实社会中存在许多大型、复杂的动态系统，如交通系统、经济预测、智慧城市管理和军事指挥

控制等,对这些系统的建模和仿真存在着很大困难。复杂系统的精确解析数学模型难以被构建,仿真模拟又存在无“真”可仿的尴尬情况,假想推演极度依赖管理者的个人经验,难以进行科学化的描述。而平行系统利用描述系统能够精确仿“真”真实世界的复杂问题。平行系统由实际系统和人工系统构成,通过二者虚实互动的相互作用,实现对虚实系统的管理与控制、决策方案的试验与评估以及执行者的学习与训练。

对 AI 游戏的研究意义远不止游戏本身,对于游戏的研究可以打通虚拟世界与现实世界的藩篱,虚拟世界中的计算实验可以完成现实世界中无法完成的测试^[22]。AI 游戏可以被作为解决现实问题的试验场,通过构建与真实系统并行的人工系统,得到平行系统。通过分析数据与行为的依赖关系,将知识视为在与真实系统并行的人工系统中重建这种相互依赖的方法,并与现有机器学习方法进行整合。真实系统与人工系统的区别主要在于,对于真实系统而言,状态转移函数对应着真实的运动规律,动作价值函数来源于环境的实际反馈,是无法改变的。人工系统中的状态转移函数和动作价值函数是可控的。区分人工和真实的关键在于其状态转移函数和动作价值函数是否可控。真实系统和人工系统的集成被称为平行系统,平行系统为智能体提供了外部的空间来存储和组织计算的中间结果,使其能够灵活地存储知识和进行推理,并扩充智能体的学习能力。

该领域的研究成果可以在短期内给游戏、电竞产业带来直接推动。虚拟世界中的策略性博弈可以被应用到教育、医疗和军事等领域^[23],最终推进通用人工智能问题的探索和发展。从蒸汽机到电动机,有了“老工业技术(industrial technology, IT)”;从无线电到电话线,有了“旧信息技术(information technology, IT)”;如今,从深度学习到平行学习,以 AlphaGo 为代表的游戏智能给了人们“新智能技术(intelligent technology, IT)”。人类面临“老、旧、新”一体的“IT 革命”,人工智能向平行智能的发展将成为人类社会从工业社会向智业社会迈进的重要推动力^[24]。

参考文献:

- [1] BROWN N, SANDHOLM T. Superhuman AI for multiplayer poker[J]. Science, 2019, 365(6456): 885-890.
- [2] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484-489.
- [3] 卢卡斯·西蒙, 沈甜雨, 王晓, 等. 基于统计前向规划算法的游戏通用人工智能[J]. 智能科学与技术学报, 2019, 1(3): 219-227. LUCAS S, SHEN T Y, WANG X, et al. General game AI with statistical forward planning algorithms[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(3): 219-227.
- [4] HASSABIS, DEMIS. Artificial intelligence: chess match of the century[J]. Nature, 2017, 544(7651): 413-414.
- [5] LAI M. Giraffe: using deep reinforcement learning to play chess[J]. arXiv preprint, 2015, arXiv: 1509.01549.
- [6] RUBEN RODRIGUEZ T, BONTRAGER P, TOGELIUS J, et al. Deep reinforcement learning for general video game AI[C]// 2018 IEEE Conference on Computational Intelligence and Games. Piscataway: IEEE Press, 2018.
- [7] CERTICKY M, CHURCHILL D. The current state of starcraft AI competitions and bots[C]// The 13th Artificial Intelligence and Interactive Digital Entertainment Conference. Beijing: Tsinghua University Press, 2017.
- [8] ARULKUMARAN K, CULLY A, TOGELIUS J. AlphaStar: an evolutionary computation perspective[C]// The Genetic and Evolutionary Computation Conference Companion. [S.l.:s.n.], 2019.
- [9] ONTANÓN S, SYNNAEVE G, URIARTE A, et al. RTS AI problems and techniques[M]. [S.l.:s.n.], 2015.
- [10] TOGELIUS J. AI researchers, video games are your friends[C]// International Joint Conference on Computational Intelligence. Berlin: Springer, 2015.
- [11] TURING A M. Computing machinery and intelligence[J]. Mind, 1950, 59(236): 433-460.
- [12] 张宏达, 李德才, 何玉庆. 人工智能与“星际争霸”: 多智能体博弈研究新进展[J]. 无人系统技术, 2019, 2(1): 5-16. ZHANG H D, LI D C, HE Y Q. Artificial intelligence and starcraft: new progress in multi-agent game research[J]. Unmanned Systems Technology, 2019, 2(1): 5-16.
- [13] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [14] BROWN N, SANDHOLM T. Superhuman AI for heads-up no-limit poker: libratus beats top professionals[J]. Science, 2017: 1733.
- [15] YE D H, LIU Z, SUN M F, et al. Mastering complex control in MOBA games with deep reinforcement learning[J]. arXiv preprint, 2019, arXiv: 1912.09729.
- [16] SONG S, WENG J, SU H, et al. Playing FPS games with environment-aware hierarchical reinforcement learning[C]// The 28th International Joint Conference on Artificial Intelligence. [S.l.:s.n.], 2019: 3475-3482.
- [17] 王飞跃. 人工智能在多角色游戏中获胜[J]. 中国科学基金, 2020, 34(2): 205-206. WANG F Y. AI wins in multi-role games[J]. Bulletin of National Natural Science Foundation of China, 2020, 34(2): 205-206.
- [18] 王飞跃. 人工社会、计算实验、平行系统——关于复杂社会经济系统计算研究的讨论[J]. 复杂系统与复杂性科学, 2004, 1(4): 25-35. WANG F Y. Artificial society, computational experiment, parallel system: discussion on computational research of complex social and economic system[J]. Complex Systems and Complexity Science, 2004, 1(4): 25-35.
- [19] WANG F Y. Toward scientific games: an ACP-based approach[R]. 2010.

- [20] LI L, LIN Y L, ZHENG N N, et al. Parallel learning: a perspective and a framework[J]. IEEE/CAA Journal of Automatica Sinica, 2017, 4(3): 389-395.
- [21] PEARL J, MACKENZIE D. The book of why: the new science of cause and effect[J]. Science, 2018, 361(6405): 855.2-855.
- [22] LI L, WANG X, WANG K F, et al. Parallel testing of vehicle intelligence via virtual-real interaction[J]. Science, 2019(28).
- [23] 李宪港, 李强. 典型智能博弈系统技术分析及指控系统智能化发展展望[J]. 智能科学与技术学报, 2020, 2(1): 36-42.
- LI X G, LI Q. Technical analysis of typical intelligent game system and development prospect of intelligent command and control system[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(1): 36-42.
- [24] 叶佩军, 王飞跃. 人工智能——原理与技术[M]. 北京: 清华大学出版社, 2020.
- YE P J, WANG F Y. Artificial intelligence, principle and technology[M]. Beijing: Tsinghua University Press, 2020.

[作者简介]



沈宇（1992—），男，中国科学院大学人工智能学院博士生，主要研究方向为智能机器人平行主动感知、平行强化学习与平行博弈方法。



韩金朋（1994—），男，西安交通大学软件学院博士生，主要研究方向为安防巡逻。



李灵犀（1977—），男，博士，印第安纳大学—普渡大学印第安纳波利斯分校电子与计算机工程系副教授，主要研究方向为复杂系统的建模、分析、控制与优化，智能交通系统，智能汽车，车辆主动安全系统与人因学。



王飞跃（1961—），男，博士，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主任，中国科学院大学中国经济与社会安全研究中心主任，青岛智能产业技术研究院院长，主要研究方向为平行系统的方法与应用、社会计算、平行智能以及知识自动化。