

博弈 5.0：基于平行系统和机器博弈的社会认知平行博弈

李亚玲¹, 杨林瑶², 葛俊¹, 覃缘琪¹, 王晓^{2,3}

(1. 之江实验室智能社会治理研究中心, 浙江 杭州 311100;

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. 青岛智能产业技术研究院, 山东 青岛 256200)

摘要: 机器博弈是一种机器模仿人的方式进行思考和决策的人工智能, 是当前人工智能领域非常具有挑战性的研究方向, 其研究现状很大程度上代表了人工智能的发展水平。总结了机器博弈的基本概念、研究现状和典型应用, 探讨了机器博弈研究面临的挑战和发展趋势, 分析了与机器博弈相关的研究在社会治理上可能的应用, 进而提出基于平行系统和机器博弈的社会认知平行博弈方法, 以期为社会治理相关的研究提供新的思路。

关键词: 机器博弈; 人工智能; 平行系统; 社会治理

中图分类号: TP301

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202151

Game 5.0: social cognition parallel game based on the parallel systems and machine game

LI Yaling¹, YANG Linyao², GE Jun¹, QIN Yuanqi¹, WANG Xiao^{2,3}

1. Research Center for Intelligent Society and Governance, Zhejiang Laboratory, Hangzhou 311100, China

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

3. Qingdao Academy of Intelligent Industries, Qingdao 256200, China

Abstract: Machine game is a kind of artificial intelligence in which machines imitate the way people think and make decisions. It is currently the most challenging research direction in the field of artificial intelligence, whose research status largely represents the development level of artificial intelligence. The basic concepts, research status and typical applications of machine game were summarized, its challenges and development trends were discussed, and the possible applications of machine game on the social governance were analyzed. Then, a social cognitive parallel game method based on the parallel systems and the machine game were proposed, aiming at providing new ideas for the research of social governance.

Key words: machine game, artificial intelligence, parallel system, social governance

1 引言

机器博弈是指构建和训练计算机系统, 使之能够模仿人类的方式进行信息获取、信息分析、智能决策和自动学习, 进而成为一个博弈智能体^[1]。机器博弈是计算机展现其智能水平的窗口, 是检验人

工智能发展水平的手段之一^[2-3]。机器博弈研究的很多问题与现实生活中的商业竞争、战争等情况十分相似, 对此类问题的决策具有很好的参考价值。

博弈贯穿人类文明的发展进程, 早在封建社会早期就诞生了中国象棋等博弈游戏, 战国时期的田忌赛马更是作为一个经典的博弈案例被世人所熟

收稿日期: 2021-11-03; 修回日期: 2021-11-22

通信作者: 王晓, x.wang@ia.ac.cn

基金项目: 之江实验室中心自设科研项目 (No.115011-212004); 浙江省科学技术情报学会科研项目 (No.2021qbxh009)

Foundation Items: Center-initiated Research Project of Zhejiang Laboratory (No.115011-212004), Research Project of Zhejiang Society for Scientific and Technical Information (No.2021qbxh009)

知。随着人类社会的发展,博弈越来越广泛地存在于人类活动中。例如,日常生活中的棋牌、球类竞赛和游戏,经济领域的贸易谈判、营销活动,政党之间的政治斗争,国家之间的外交谈判和战争等都可以见到博弈的影子。棋类游戏自古以来就是人类智慧的象征,其规则清晰、胜负容易评估、受众面广泛。因此,让机器通过游戏的方式与人类博弈,是研究人类思维和实现机器思维较好的实验载体,是检验机器智能水平的有效手段。例如,图灵测试就是通过让机器进行一场模仿游戏,来判断机器是否具备所谓的智能。图灵测试在游戏中可以被描述为:当玩家和机器同时游戏时,如果玩家无法准确区分人类玩家和机器玩家,就说明机器通过了相应的图灵测试^[4]。机器博弈也是人工智能领域特别具有挑战性的研究方向之一,是机器智能、兵棋推演、智能决策系统等重要基础。

1928年,约翰·冯·诺依曼(John von Neumann)提出极大极小定理,证明了博弈论的基本原理。1944年,在他和奥斯卡·摩根斯特恩(Oskar Morgenstern)合著的《博弈论和经济行为》中,他们从经济行为出发,系统而全面地建立了博弈的公理化描述,奠定了机器博弈研究的基础和理论体系。机器博弈研究兴起于20世纪50年代,1950年,信息论创始人克劳德·艾尔伍德·香农(Claude Elwood Shannon)在论文^[5]中提出了象棋博弈的编程方案。1958年,Bernstein A等人^[6]在IBM704计算机上开发了第一个达到孩童水平的国际象棋程序。1962年,人工智能的创始人之一——亚瑟·塞缪尔(Arthur Samuel)编写了一个国际跳棋程序,并于1962年击败了美国的一个州冠军。随着电子计算机的诞生和软硬件技术的不断发展,机器的智能化水平不断提升,其在多个更加复杂的游戏中击败了人类顶尖选手。1997年,国际商业机器(IBM)公司开发的基于知识规则引擎和强大计算机硬件的“深蓝”战胜了世界国际象棋棋王加里·卡斯帕罗夫(Garry Kasparov)。2011年,IBM公司开发的基于自然语言理解和知识图谱的Watson在美国智力问答竞赛节目中战胜人类冠军。2016年,谷歌公司开发的AlphaGo^[7]战胜了围棋世界冠军李世石。这3次人机大战都引起了公众的广泛关注,推动了人工智能技术的快速发展。

近年来,随着人工智能的发展,机器学习面临的场景愈发复杂,在拟合数据分布的基础上,机器

学习开始被广泛应用于决策任务中。研究人员开始将机器学习与博弈论结合,提出基于机器学习模拟人类决策的机器博弈方法。因此,已有的机器博弈方法可以分为3类^[8]:一是基于博弈论基本理论对博弈树进行搜索,进而找到给定状态下的全局或局部最优解;二是基于机器学习,特别是强化学习方法,训练从感知到动作的博弈智能体;三是将两者结合的方法。已有的机器博弈研究主要针对围棋、象棋、军棋等棋类游戏,扑克、桥牌等牌类游戏以及《星际争霸》《王者荣耀》等动作策略类游戏等。按照所涉及游戏的信息完整性、零和性及合作性等特征,机器博弈可以分为是否完全信息、是否零和博弈、是否合作博弈。完全信息博弈是指博弈参与方了解其他参与方的策略集、收益等信息,而非完全信息博弈双方信息不完全透明,存在一定的欺诈信息。典型的完全信息博弈包括中国象棋、围棋、五子棋等,常见的非完全信息博弈有德州扑克^[9]、桥牌、斗地主等。零和博弈是指所有博弈参与方的收益和损失之和永远为零、一方收益必然导致另一方损失的博弈问题,已有的针对游戏的博弈研究大多属于零和博弈。合作博弈是指博弈参与方可以通过同盟、合作方式进行博弈对抗的博弈问题。已有的机器博弈研究大多针对非合作的单智能体博弈,但是,合作博弈在城市建设、无人机集群军事对抗、指挥控制、企业间的博弈决策等领域发挥着关键作用,亟待进一步研究。

随着机器学习、物联网、大数据、云计算等技术的发展,人工智能已成为新一轮科技革命和产业变革的重要驱动力,并对经济社会发展和人类生活产生了深远影响。将人工智能同社会治理结合,能够有效提升社会治理的现代化水平^[10]。赵宜萱等人^[11]分析了人工智能在人力资源管理中的应用逻辑、应用架构和具体实践,并提出基于人工智能与社会认知和自我决定理论的人力资源管理理论模型,为提升人力资源管理效率提供了有效方法;参考文献[12-13]从博弈的角度分析了社会组织与其他社会主体在社会治理演化下的关键作用和策略选择。社会治理的关键在于在多元化的格局中争取组织利益、维护各方关系,其核心是博弈问题。因此,将机器博弈关键技术与社会治理结合,基于人工社会^[14-15]虚拟系统模拟不同主体在社会博弈中的利益目标和策略选择,基于系统演化数据自适应搜索和学习最优的管理和控制策略,进而引导现实社会的运行,有助于

实现社会关键问题的解决和社会服务水平的提升。

本文梳理了机器博弈的发展历程,总结了机器博弈的关键技术和发展现状,并且从社会治理的角度对机器博弈在社会认知博弈方面的应用前景进行了分析,提出了基于平行系统和机器博弈技术的社会认知平行博弈方法,以期为社会治理领域的研究人员提供新的参考和借鉴。

本文的结构如下:第2节概括机器博弈的发展历程、主要成果和关键事件;第3节总结机器博弈中的典型技术及其发展现状;第4节基于平行系统和机器博弈的已有技术提出社会认知平行博弈方法;第5节对全文进行总结,并展望未来。

2 机器博弈发展概述

机器博弈从模仿人类的思考和决策方式出发,以游戏为研究载体,通过模拟人类在游戏之中的思维过程构建接近人类智能的信息处理系统,并扩展到其他相关领域,进而解决现实中难以解决的复杂问题^[16]。

2.1 机器博弈发展历程

从20世纪50年代开始,许多著名学者都开始涉足机器博弈的研究工作,并对其做出了突出贡献。1953年,艾伦·麦席森·图灵(Alan Mathison Turing)设计了一个能下国际象棋的指令,实现了第一个国际象棋的程序化博弈。1958年,IBM公司推出第一台能够与人类进行国际象棋对抗的计算机IBM704。1962年,亚瑟·塞谬编写的国际跳棋程序击败美国的一个州冠军,成为机器博弈发展历程中的一个里程碑事件。在机器博弈研究的早期阶段,机器主要通过穷举博弈双方所有可能的策略,对不同策略的结果进行评价并产生相应的策略。因此,该阶段研究的主要内容是如何建立有效、快速的评价函数和评价方法,使评价的效率更高、花费的时间和空间代价更小,以及如何在生成的博弈树上更准确高效地找到最优解^[17-18]。

20世纪80年代,随着计算机软硬件水平的不断提升,机器博弈理论日趋完善,机器博弈研究进入快速发展的阶段。特别是传统的符号学习算法的完善和反向传播算法^[19]的成功应用,使机器学习的基础理论研究引起了人们的重视,掀起了基于统计模型的机器学习热潮。1987年,美国卡内基梅隆大学设计的国际象棋程序“深思”在与世界国际象棋棋王加里·卡斯帕罗夫进行的“人机大战”中以0:2

败北,它具有每秒75万步的处理速度,其水平相当于拥有2450国际等级分的棋手。1995年,IBM公司设计了全球第一台采用并行计算系统、以RS/6000SP为基础、集成了总计256块处理器以及480颗VLSI加速芯片的国际象棋计算机“深蓝”。1996年,“深蓝”在美国费城败给了加里·卡斯帕罗夫。IBM公司将“深蓝”升级为“更深的蓝”,将计算速度提升了2倍,并丰富了芯片中的象棋知识,使它能够识别不同的棋局,并从可行方案中找出最佳的行棋方案。“更深的蓝”能够搜寻和估算出当前局面往后的12步行棋方案,最多可达20步,而人类棋手的极限是10步。1997年,“更深的蓝”凭借精准的计算战胜加里·卡斯帕罗夫,在世界范围内引起了轰动。“深蓝”成功的秘诀主要有两点:一是将通用计算机处理器和象棋加速结合,采用混合决策的方法快速进行推理;二是在软件设计上采用了超大规模知识库结合优化搜索的方法,存储了70万份国际特级大师的棋谱,利用知识库在开局和残局阶段节省处理时间,并得出更合理的行棋方案,同时采用Alpha-Beta剪枝算法和基于规则的方法对棋局进行评价,通过缩小搜索空间对棋局进行更有效的评价。

20世纪末期,研究人员在机器博弈搜索和评估方面进行了大量深入探索和实质性研究,极大极小算法、Alpha-Beta剪枝算法、并行搜索、遗传算法、人工神经网络等日趋成熟,支持向量机、Boosting算法、最大熵算法等相继被提出,机器博弈研究进入新的发展阶段。1992年,中山大学教授陈志行开发的围棋程序“手谈”获得国际电脑围棋赛冠军;东北大学徐心和团队研发的“棋天大圣”^[20]在2006年和2007年两次获得中国象棋比赛的世界冠军;北京邮电大学刘知青团队开发的“本手”能够战胜高水平的业余围棋选手;北京理工大学团队在2009年和2011年获得六子棋比赛的世界冠军。

2006年,辛顿(Hinton)提出无监督逐层贪心训练神经网络的方法,掀起了深度学习研究的热潮。并行计算、深度学习等技术的突破性进展成功解决了机器博弈的抽象认知难题,使深度学习等技术被成功应用于机器博弈及相关领域中,将机器博弈水平带上了一个新的台阶。谷歌公司的DeepMind团队开发的AlphaGo是基于深度学习和并行计算的机器博弈算法的典型代表,该系统通过对16万局人类高手棋谱中的3000万手盘面进行监督学习和强化学习,使用策略网络和估值网络实现落子决策

和局势评估。通过与蒙特卡洛树搜索算法结合, AlphaGo 极大地改善了搜索决策的质量。此外, AlphaGo 还提出了一种异步分布式并行算法, 使其可运行于 CPU/GPU 集群。

AlphaGo 的原理示意图如图 1 所示, 其对弈知识是基于离线深度学习从大量数据中学习的。AlphaGo 的离线学习过程可以分为 3 个阶段: 第一阶段以 3 万多幅人类高手棋谱为输入, 基于全局特征和深度卷积神经网络训练策略网络, 进而学习给定状态下的落子概率和走棋策略; 第二阶段是利用第 t 轮训练的策略网络与之前训练好的策略网络进行对抗, 通过自我博弈的增强学习修正第 t 轮的策略网络; 第三阶段通过大量的自我对弈训练一个价值网络, 进而实现对全局态势的评估, 以判定每步棋对全局胜负的影响, 优化走棋策略。这些技术使得 AlphaGo 初步具备了既可以考虑局部得失, 又可以考虑全局态势的能力。

2017 年 5 月, AlphaGo 的升级版 AlphaGo Master 以 3:0 战胜世界排名第一的柯洁, 取得了更加辉煌的战绩。AlphaGo Master 更多地依靠自我对弈产生的高质量数据, 从自己的搜索中学习, 同时使用了算力更强的 TPU 资源, 棋力和棋速均得到了

有效增强。2017 年 10 月, DeepMind 推出 AlphaGo Zero, 该版本将 AlphaGo 中的策略网络和价值网络合并为一个网络, 能够同时输出所有动作的选择概率和状态的价值, 同时, AlphaGo Zero 没有使用任何人工经验, 而是完全基于强化学习自学习的方式进行训练。经过 3 天的训练, AlphaGo Zero 便以 100:0 的战绩击败了 AlphaGo。AlphaGo 系列的成功充分验证了深度学习与机器博弈结合的价值, 同时证明了机器智能水平迈上了新的台阶。

除棋类游戏外, 近年来, 基于深度学习的机器博弈技术还被成功应用于扑克、电子游戏对抗等场景中。2017 年, 美国卡内基梅隆大学开发的人工智能 Libratus 与 4 名人类顶尖德州扑克选手进行了为期 20 天的对战, 最终人工智能取得胜利。2019 年 7 月, 脸书和卡内基梅隆大学合作研发的 Pluribus 扑克机器人基于在线搜索算法和自学习算法, 有效提升了处理能力。在不需要先验信息的情况下, Pluribus 以自我博弈的方式不断提升, 在 6 人无限制德州扑克比赛中击败了 15 名顶尖选手^[21], 成为机器博弈的又一个里程碑工作。2018 年, DeepMind 开发的《星际争霸》机器博弈程序 AlphaStar^[22]以 5:0 的战绩战胜了职业选手。AlphaStar 是将游戏看

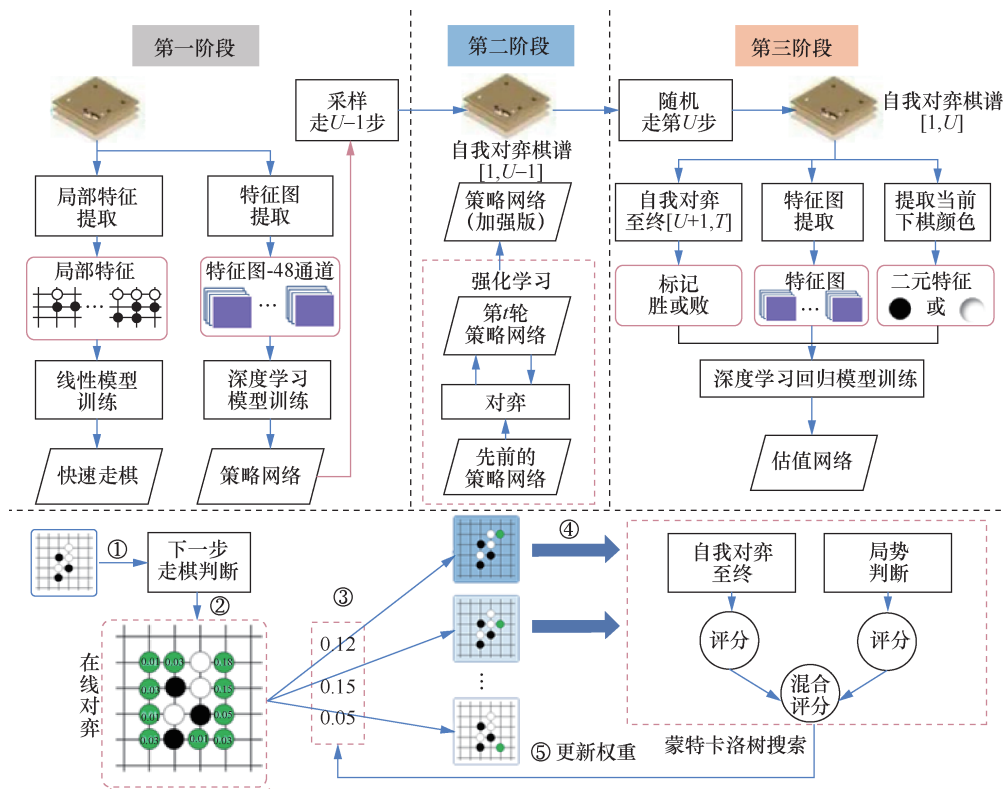


图1 AlphaGo 的原理示意图

作长序列建模学习任务的强化学习智能体, 模型从游戏接口接收游戏中的单位列表及对应单位的属性, 经过深度神经网络计算后输出在游戏中执行的指令。AlphaStar 的训练策略与 AlphaGo 类似, 首先使用人类比赛过程的回放进行有监督的模仿学习, 使模型快速学习到高水平玩家在《星际争霸》天梯中使用的基础策略和微操, 然后, 通过强化学习的自我对弈, 继续提升模型的博弈水平。中国科学院自动化研究所张俊格团队研发的 CSEbot 在 2018 年的《星际争霸》AI 比赛中获得季军。2020 年, 启元 AI “星际指挥官” 以 2:0 完胜全国冠军黄慧明及中国星际争霸超人族选手、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, 表现出了与一个冷静且智商和情商极高的人类相当的智能水平。经过多年的理论研究, 在国际象棋、围棋、扑克等人机对抗场景中, 从最初人类完胜电脑, 到如今电脑击败人类顶级高手, 机器的智能水平迅速上升。

2.2 机器博弈发展现状

寻找博弈问题的理论解, 一直是机器博弈研究的重要目标。近年来, 随着一些机器博弈系统达到或超过人类大师的水平, 部分博弈问题的理论解已经被成功解决。其中, 对于部分复杂度过高、难解的博弈问题, 只能采用渐近法估算其计算复杂性。国内外学者对部分机器博弈的计算复杂性进行了研究^[23], 证明了国际象棋属于指数时间完全 (EXPTIME-complete) 问题, 围棋属于多项式空间难 (PSPACE-hard) 问题。常见棋类游戏的状态空间复杂度和博弈树复杂度的估算值见表 1。针对井字棋、国际跳棋、国际象棋等复杂度较小的二人零和完全信息博弈问题, 搜索理论已经比较系统完善, 基于极大极小算法等搜索算法已经能够较好地解决其策略搜索问题。然而, 针对博弈树复杂度较大的搜索问题, 以极大极小算法等为基础的策略搜索算法由于其时间和空间开销巨大, 难以做到很深的搜索, 在实际应用中仍然存在很多问题无法解决。

常见的机器博弈问题可以分为是否完全信息、是否零和、是否合作等, 不同机器博弈游戏的分类如图 2 所示。按照博弈参与方对其他参与方的特征、策略和收益等信息的掌握程度, 机器博弈可以分为完全信息博弈和不完全信息博弈。国际象棋、围棋等棋类游戏及已有的动作策略类电子游戏大多属

于完全信息博弈, 所有博弈参与方都按照公开的策略集、收益和输赢目标进行博弈, 针对此类问题, 可以有效构建博弈树, 并且通过对策略的搜索和评估得到有效的博弈策略。一般而言, 完全信息博弈的难度和计算复杂度仅与博弈树的复杂度和搜索深度相关。对于斗地主、德州扑克等不完全信息博弈问题, 机器不仅需要推理对手的资源 and 策略, 需要在不同的子博弈之间寻找平衡, 还需要学会使用“诈唬”等心理技巧, 寻找让对手无法预测的策略。因此, 不完全信息博弈问题的难度更大。很长一段时间以来, 最好的算法也只能解决不完全信息博弈中的玩具问题。直到 2007 年, 阿尔伯塔大学计算机扑克研究小组的研究人员提出反事实后悔最小化算法^[24], 并于 2015 年成功解决有限注德州扑克问题。2017 年, 该研究小组开发的人工智能 DeepStack^[25]在与职业扑克玩家的对抗中击败了 11 位职业玩家, 该系统采用了集成深度神经网络的蒙特卡洛树搜索方法, 通过对当前公共状态的本地策略计算、使用任意状态的价值函数实现有限深度的前瞻以及预测动作的受限集合, 实现了在无限注德州扑克上超越人类智能。由此可见, 对于不完全信息博弈, 已经初步形成了可行的解决方案。但是, 机器仍然无法很好地理解人类情绪, 在现实生活中, 不完全信息博弈问题中的智能水平也还没有得到有效验证。

表 1 常见棋类游戏的状态空间复杂度和
博弈树复杂度的估算值

游戏名称	棋盘大小	状态空间复杂度	博弈树复杂度
井字棋	3×3	10 ³	10 ⁵
国际跳棋	8×8	10 ²¹	10 ³¹
国际象棋	8×8	10 ⁴⁶	10 ¹²³
五子棋	15×15	10 ¹⁰⁵	10 ⁷⁰
围棋	19×19	10 ¹⁷²	10 ³⁶⁰

零和博弈即参与博弈过程的各方, 一方的收益必然导致另一方的损失, 在这种情况下, 博弈参与者不存在合作共赢的情况, 一个游戏参与者的收益会演变成其他参与者的损失。双人零和博弈由于其普遍性得到了广泛的研究, 是机器博弈的重要研究方向, 人们熟悉的大多数游戏属于双人零和博弈。当前, 双人零和博弈问题的解决方法本质是通过策略搜索算法对特定局面下的博弈树进行搜索, 以获得最优行动策略。博弈树是对博弈局面状态变化的

结构化表示,针对国际象棋、围棋等复杂度高的博弈问题,往往只能构建有限深度的博弈树。已有的机器博弈算法大多采用带有启发式信息的启发式搜索算法解决复杂度较高的双人零和博弈问题,常用算法包括极大极小算法、Alpha-Beta 剪枝算法等。在日常生活中,除了零和博弈,更多的场景是非零和博弈,特别是第二次世界大战以来,零和游戏的观念正逐步被双赢的观念所取代,能够使博弈双方都得益的正和博弈成了人们的追求。

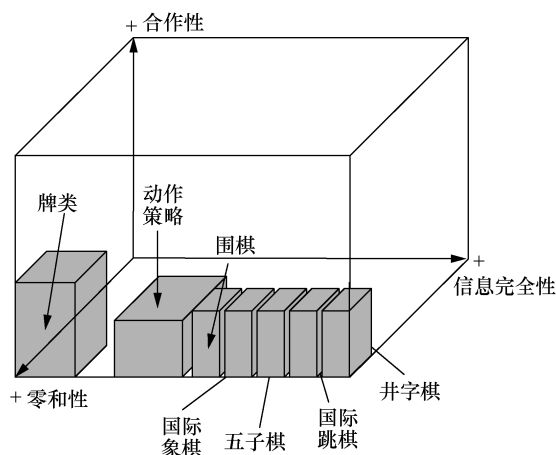


图2 不同机器博弈游戏的分类

非零和博弈问题大多是基于合作博弈的群体决策问题,合作博弈强调集体理性,在合作博弈的场景下,由于联盟的整体收益大于每个个体单独博弈的收益之和,且联盟内的参与者都能获得比不加入联盟更高的收益,因此合作博弈中参与者会互相协调,进行联盟,其主要求解方法是多智能体方法。多智能体机器博弈的核心问题是基于均衡、协同和合作等学习目标,建立一个策略交互模型,通过交互式强化学习等算法,达到令博弈玩家都满意的均衡解,如纳什均衡、斯塔克尔伯格(Stackelberg)均衡和元均衡等^[26]。按照博弈玩家的目标是否相同,非零和博弈可以分为共同利益博弈和不同利益非零和博弈。其中,共同利益博弈一般基于单智能体强化学习算法,主要有团队博弈^[27]、势博弈^[28]和去中心化的部分可观察马尔可夫决策^[29]等。不同利益非零和博弈一般为有限一般和博弈,其求解难度比零和博弈困难很多,已有的求解算法大多将经典Q学习作为中心控制器,进而运用不同均衡方法引导智能体在迭代中逐渐收敛。从单智能体博弈向多智能体博弈过渡存在维度爆炸、环境非平稳性、信度分配等问题,随着智能体数量增多,系统很难

得到收敛结果^[30],某些特殊收敛点对应的策略也不符合实际情况。此外,在很多实际场景下,多智能体非零和博弈最优解未必存在。当前,非零和合作博弈还需要在收敛性、可解释性等方面有所突破,以实现从个体智能到群体智能决策的飞跃。

尽管人工智能在图像识别、语音识别、机器翻译等领域都取得了重大突破甚至接近或超越了人类水平,目前的人工智能仍然属于弱人工智能。现阶段的人工智能仍是根据从大量数据中学习到的统计规律和规则进行决策,不具备真正的推理能力,没有自主意识,无法像人类一样不断适应新的环境并涌现出新的功能。能够自适应地应对新的问题、具有自我意识的强人工智能还需要长时间的研究。

在大数据智能的视角下,各类人工智能算法的目标是学习训练数据集上的最优拟合解,即在训练数据上的误差损失最小。如此,机器博弈的策略必然是在训练样本中胜率最高的策略,但是,在现实问题中,对手的策略是动态变化的,人工智能算法的学习目标应当是在动态情况下也能取得较好效果的均衡解^[31]。

当前,人工智能算法特别是深度神经网络的时间复杂度较高,在步数较少的完全信息博弈中很容易通过搜索算法求得纳什均衡解。但是,现实中面临的完全信息博弈往往状态和动作空间非常大,难以直接估计当前决策对最终局势的影响。因此,针对大规模、高复杂度问题的高效机器学习算法有待进一步研究。相比于完全信息博弈,不完全信息博弈更符合现实场景。但是,在不完全信息博弈的求解中,对手采取的行动不一定可见,导致博弈树的规模可能非常大。因此,为了缩减搜索空间,需要基于状态合并、剪枝等方法对博弈树进行压缩,将不完全信息博弈问题转化为相对简单的抽象问题,进而求解该问题的纳什均衡解。总体而言,人工智能在开放域环境下的自适应推理和决策还需要进一步研究。

已有的人工智能大多为单体智能,近年来,人们开始将研究重点转向多智能体协同的群体智能。2017年,国务院印发的《新一代人工智能发展规划》中明确提出群体智能的发展方向,以通过群体智能的主动感知与发现、知识获取与生成、协同与共享、评估与演化、人机混合与增强及群体智能的协同与控制等,实现群体智能数据-知识-决策自动化的完

整流程。实现群体智能系统主要有两种途径: 一是每个智能体独立决策, 不考虑其他智能体的决策, 但容易导致算法不收敛; 二是基于多智能体博弈在竞争、合作等博弈形式下跟踪其他智能体的行为, 通过多智能体联合行为学习纳什均衡解。2016 年, DeepMind 与美国游戏公司暴雪娱乐合作, 在《星际争霸 II》游戏环境的基础上开发了人工智能学习环境, 吸引了大量研究人员投入多智能体博弈的研究中。然而, 由于存在实时对抗、群体协作、非完全信息博弈、庞大的搜索空间、多复杂任务和时间空间推理等难点^[32], 多智能体机器博弈仍然是当前人工智能领域非常具有挑战的难题。因此, 要实现多人机协同的群体智能和强人工智能仍然需要很长一段时间。

3 机器博弈的典型技术

博弈树的搜索是机器博弈的核心, 是决定博弈结果和博弈效率的关键。典型的搜索算法按照搜索方向可以分为深度优先搜索和宽度优先搜索, 按照控制策略可以分为盲目搜索和启发式搜索, 按照搜索范围可以分为穷尽搜索和剪枝搜索。一般而言, 策略搜索的深度越深, 博弈能力越强, 但是计算时间也会大幅度提升。

穷尽搜索即对当前状态下所有可能的策略进行枚举的搜索方法, 该方法的典型代表是极大极小算法。极大极小算法基于递归找到在一定递归深度内最小化对手的最大收益的策略, 可以找到对于博弈双方都是最优的博弈值。极大极小算法奠定了机器博弈的理论基础, 但是, 该算法需要遍历整个博弈树, 时间复杂度较高。

剪枝算法是机器博弈中最常用的算法, 其典型代表是 Alpha-Beta 剪枝算法。Alpha-Beta 剪枝算法是在极大极小算法基础上的改进算法, 用以减少极大极小算法搜索树的节点数。当算法评估出某策略的后续走法比之前的策略还差时, 就会停止计算该策略的后续策略。通常, 将 Alpha-Beta 剪枝算法与置换表技术结合, 以减小博弈树规模, 提高搜索效率。

启发式搜索算法在许多规划和优化问题(如线性规划、非线性规划、旅行商问题、计划调度等问题)中具有广泛的应用。常用的启发式搜索算法包括爬山法、分支定界法、模拟退火、遗传算法、蚁群算法等。但是, 以上算法针对单一决策主体, 具

有明确的目标函数和约束条件, 只能处理静态规划和优化问题。而博弈问题的决策主体有多个, 且各方的目标函数有很大的区别。因此, 博弈中常用的启发式搜索是在 Alpha-Beta 剪枝算法的基础上通过排序让搜索树尽可能接近最小树, 从而优先搜索好的策略, 该类算法主要包括置换表启发^[33]、历史启发^[34]和杀手启发^[35]等。

随机搜索是针对状态或动作空间过大情况的策略搜索方法, 其典型算法是蒙特卡洛搜索。蒙特卡洛搜索是运用蒙特卡洛模拟进行博弈策略评估的启发式搜索算法, 该算法主要由两部分组成: 一是上限信心界应用树^[36]算法, 根据众多模拟结果优化下一次的决策; 二是基于蒙特卡洛模拟评估, 通过多次随机探索博弈树的方式评估每个节点的优劣^[37]。蒙特卡洛搜索在计算机围棋等完全信息博弈、多人博弈问题上获得了成功应用。

在博弈树的搜索过程中, 需要计算每个节点的估值才能找到最优的决策。此时, 需要穷举所有的状态, 从终止状态开始逆向计算每一个中间节点的估值。然而, 对于绝大多数的游戏来说, 决策空间相当庞大, 几乎不可能穷举所有的决策, 因此, 需要设计相应的评估函数, 通过有限规模的搜索, 对当前局面进行评分。在机器博弈系统中, 局面评估越准确, 获胜的概率越高。在完全信息博弈中, 评估函数主要根据规则和经验人为设计; 在构建非完全信息博弈的评估函数时, 除了已知的规则和己方信息, 还需要猜测对手的情况。通常, 针对当前局面, 先将评估函数应用到后面几步可能的局面中, 然后将获得的评估值通过极大极小算法反向传播回当前局面节点, 从而产生当前棋局下所有可能行棋方式的评分, 能够使对当前局面的评分更加精准。

强化学习是近年来机器学习领域研究的热点, 已被广泛应用于机器博弈领域。强化学习的基本思想是基于试错的方式最大化智能体从环境中获得的累计收益, 以学习序列决策任务的最优策略。并且, 随着多层感知机、深度信念网络、卷积神经网络等模型的提出, 人工神经网络研究取得了重大的进展, 因此, 研究人员将具有感知能力的深度学习与具有决策能力的强化学习结合^[38], 形成了能够在高维空间端到端感知决策的深度强化学习方法。深度强化学习可以分为基于值函数的方法和基于策略梯度的方法, 基于值函数的深度强化学习方法需

要维护一个价值表或价值函数,并通过这个价值表或价值函数选取价值最大的动作进行决策,典型的算法包括深度 Q 学习^[39]、DDQN (double DQN) 算法、A3C (asynchronous advantage actor-critic) 算法等;基于策略梯度的深度强化学习方法直接学习一套动作策略,进而优化给定状态下所需要采取的动作,代表性算法包括 Reinforce 算法、Actor-Critic 算法等。近年来,研究人员将深度强化学习成功应用于机器博弈,为复杂状态下的博弈决策提供了有效的解决思路。例如,DeepMind 将深度卷积神经网络和 Q 学习结合^[39],在 Atari 视频游戏上达到了人类玩家的控制效果;Wu Y 等人^[40]提出批量 A3C 算法,批量更新多智能体模型,有效提升了数据的利用效率和模型收敛速度,进而在 VizDoom (Visual Doom AI Competition) 比赛中获得了最佳名次;Peng P 等人^[41]提出一种多智能体 Actor-Critic 模型,构建全局和个体奖励,使多个智能体单元能够协同作战,在《星际争霸》的局部对抗任务中取得了良好的效果。针对复杂任务下强化学习策略学习效率低、模型仅能适用一个单一的场景、单智能体决策能力不足、模型记忆与推理能力不足等问题,研究人员在分层深度强化学习、多任务迁移深度强化学习、多智能体深度强化学习、基于记忆与推理的深度强化学习等领域进行了探索,基于强化学习的通用人工智能取得了飞速发展。随着机器博弈研究的深入,传统算法难以满足复杂游戏任务的需求,需要更多深度强化学习算法,向更复杂的博弈任务发起挑战^[42]。

4 社会认知平行博弈

随着移动互联网时代的来临,新兴媒体的影响越来越大,网络虚拟世界已经成为人们日常活动的重要空间,社会化媒体营销和舆情传播对社会管理、经济稳定和国家安全产生了重大影响。以微博、微信、抖音等为代表的社交媒体的蓬勃发展极大地提升了民众参与社会治理的热情,同时,也对政治、经济、文化的安全和稳定提出了新的严峻挑战^[43]。以美国为首的西方军事强国,以社交媒体舆情操纵等为手段,大力推进社会认知作战^[44],以期通过非军事行动达到预期的战略目的。2010 年的“阿拉伯之春”事件,使得大规模网民的行为、组织和运动成为各国面临的严峻社会现实;2011 年美国的“占领华尔街”运动,

在社交媒体脸书和推特的推动下,演变成席卷上千城市的社会运动,造成了严重的流血冲突。在社交媒体成为主要传播渠道的情况下,信息具有碎片化、过载的显著特点,信息的传播速度呈现几何级增长,依靠传统的人工审核等手段难以进行及时、有效的分析和处理。同时,社会认知攻击存在隐蔽性强、无法准确获取对手意图、难以即时溯源等突出特点,为社会认知治理带来了极大的挑战。因此,必须研究以社交媒体为载体、以开源情报伪装欺骗为主要手段、以多智能体机器博弈为核心、以引导塑造受众认知为目标的智能社会认知博弈系统,以实时化、一体化、常态化方式,开展物理域、信息域和认知域的筹划、准备、引导和评估,从而实施有效的社会认知博弈对抗,提升我国的社会认知治理水平。

社交媒体是一个包含物理、信息和心理因素的社会物理信息系统 (cyber-physical-social systems, CPSS)^[45],是一个复杂巨系统,存在不可准确预测、难以拆分还原、无法重复实验^[46]等难题,传统的机器博弈方法难以直接应用于社会认知对抗分析和实验。然而,人们面对的社会认知博弈问题是“不可分却不得不分、不可知还需要知”的难题,针对此类问题,可行的解决方法是基于 ACP 方法的平行系统。

平行系统^[14]是中国科学院自动化研究所王飞跃研究员原创的复杂系统管理与控制理论,其理论框架如图 3 所示。平行系统的核心是 ACP 方法,即人工系统 (artificial systems)、计算实验 (computational experiments) 和平行执行 (parallel execution) 的有机统一。平行系统基于实际系统采集的小数据,借助知识表示、机器学习等数据分析方法,全面、准确地刻画个体的特征、行为和交互机制^[47],实现对复杂系统整体的建模。针对复杂系统无法预测、无法实验的问题,平行系统基于人工系统中设置的虚拟场景,通过计算实验分析对应场景下各类参数、技术方案的效果^[48],进而得到对应的最优控制方案。最终,基于人工系统与实际系统的虚实交互、闭环反馈和双向引导,实现数据驱动的描述智能、实验驱动的预测智能、虚实互动反馈的引导智能^[49],并基于人工系统和实际系统的对比、分析、预测和评估,实现人员的管理和培训、方案的实验和评估、系统的管理与控制。

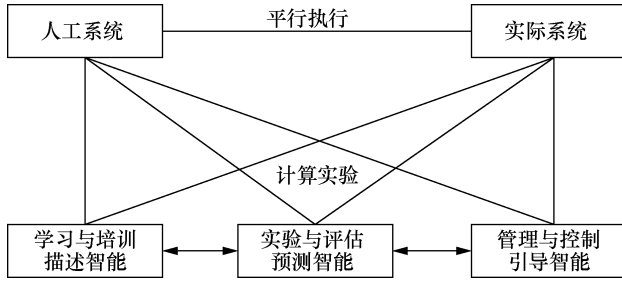


图3 基于 ACP 方法的平行系统理论框架

因此，可以构建社会认知博弈平行系统，基于爬虫等社会传感器跨媒体采集各类开源数据，开展用户画像、虚假信息识别、事件关联分析等情报分析，进而基于软件定义的人工系统建模和描述个体和各类社会组织，借助舆情信息的网络传播模型建模个体与邻居的交互规律，从而构建包含网民个体、社会媒体、攻击者等在内的社会认知博弈人工虚拟系统；进而通过构建大量场景，在人工系统中开展相应场景的可实施、可定量、可重复的计算实验，模拟各种社会认知攻击场景下的事件发展规律，预测网民群体的社会认知动态演化，进而基于自我博弈、强化学习等算法从人工场景和大量演化数据中训练社会认知管控智能体，优化其管控策

略，从而获得具体场景下的社会认知作战最佳的博弈策略；当社交媒体上发生恶性谣言传播和社会认知攻击时，即可根据相应的特征数据，从人工系统产生的社会认知博弈知识库中匹配相应的管控方案，基于平行执行方法将对应的最优方案以舆情引导扩散等方式实施到实际的社交媒体和互联网中，从而及时预测和控制谣言的传播，减少网络谣言和社会认知攻击的消极影响，进而维持社会稳定，保障社会的和谐稳定发展。

社会认知平行博弈框架如图4所示，社会认知平行博弈的第一步是构建人工系统，其主要方法是基于从社交媒体采集的小数据构建多智能体描述可计算、可重构、可编程的软件定义的对象、软件定义的流程、软件定义的关系^[50]等，进而组合成软件定义的人工系统。其中，可以借助玩家建模^[51]技术，基于有监督学习、模仿学习^[52]和强化学习等方法对网民用户和社会认知攻击者等个体进行建模，通过对真实的网民用户个体的属性和行为数据的分析，构建智能体模型，以模拟网民等各类智能体在不同场景下的行为和心里计算模型。通过获取主流社交平台的账号信息及账号之间的好友、转

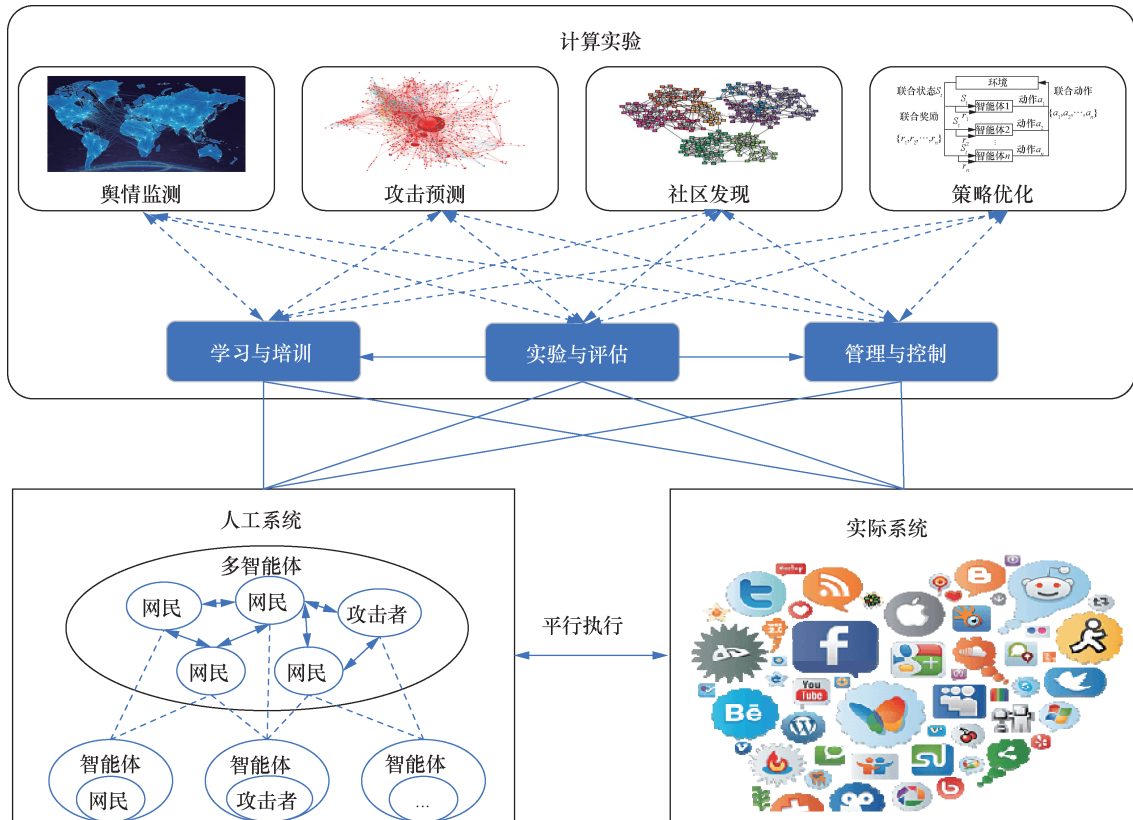


图4 社会认知平行博弈框架

发、点赞等历史数据,可以在人工系统中建立智能体之间的连接关系,模拟出相应的社会网络结构。舆情的传播模型是模拟个体交互、揭示社会认知博弈动态演变的关键。舆情传播问题已经得到了广泛的关注和研究,形成了众多的理论方法和数学模型。常用的舆情传播模型包括戴雷-肯德尔(Daley-Kendall, DK)模型^[53]、易感-传染-易感(susceptible infected susceptible, SIS)模型^[54]、易感-传染-移除(susceptible infected removed, SIR)模型^[55]、易感-暴露-患病-康复(susceptible exposed infectious recovered, SEIR)模型^[56]等,通过模拟病毒在人群中接触感染的过程,即可推演出信息在社交媒体中的传播过程,进而演化产生不同场景、不同参数下的网民认知变化,并产生相应的人工大数据。

人工系统实现了从物理状态到描述状态的转变,当实际社交媒体与人工认知博弈系统匹配执行之后,人工系统实时以社会传感数据为驱动,通过对数据的分析与评估,识别实际系统中的主体和潜在或已经发生的事件,评估实际系统的变化对人工系统生成的管控方案的影响,从而为下一时间步状态转变的决策活动提供依据。

社会认知平行博弈的第二步是基于机器博弈技术进行计算实验,通过演化博弈方法对人工系统产生的动态社会认知博弈过程进行计算分析,进而基于强化学习等机器博弈方法针对博弈过程中社交媒体环境和人员行为反馈的多种可能策略进行计算推演,从而自动从候选策略方案中生成最优的社会认知管控策略序列。可行的管控策略包括社会威慑、目标免疫和正面消息^[57]等,其中,社会威慑是指在社交媒体中通过法律约束、经济制裁等手段限制舆情的传播;目标免疫是通过控制意见领袖的舆论传播行为,进而控制整个网络的舆情传播;正面消息是通过官方媒体抑制虚假舆情消息的传播。社会认知博弈是一个典型的不完全信息合作博弈问题,社会认知攻击者的策略难以全部可知,涉及大量主体,如社交媒体、自媒体、政府部门等,且主体之间可以结成联盟形成组合策略。同时,社会认知博弈涉及海量信息的处理、开放动态策略的选择,因此,社会认知博弈是一个十分困难的问题。对此,可以基于多智能体机器博弈方法,从人工系统大量的动态演化数据中学习、优化模型参数,使多智能体在同一环境中自主交互、协同决策,从环境中感知学习端到端的最优策略。

计算实验实现了从描述状态到预测状态的转变,该过程以事件为驱动,根据事件预测系统状态的演化,生成相应的系统管控方案,推演预估方案的效果,进而确定最佳的方案。

社会认知平行博弈的第三步是平行执行,即将对应的最优社会认知博弈策略在真实的社交媒体上实施,从而引导实际系统向人工系统演化,达到更优的系统状态,实现有效的社会认知管控。

平行执行包含两个子阶段:第一阶段是从预测状态到引导状态的演变,该阶段以方案为驱动,在获得计算实验得到的最优管控方案之后,实施方案的执行分解和指令开发,包括方案计划生成、计划推演、指令开发和指令分发等,从而通过具体的指令指导物理实体开展行动,驱动实际系统向期望的方向演化;第二阶段是从引导状态到物理状态的演变,该阶段以指令驱动为主,通过监控物理实体执行指令的过程(包括计划进程监控、行动监控和主体状态监控并根据实时监控数据修正指令^[58])保证系统管控方案得到充分执行及系统管控目标顺利实现。

平行系统的上述过程不是一蹴而就的,而是迭代的、持续进行的,需要实际系统与人工系统持续虚实交互,人工系统需要根据实际系统的状态演变调整系统参数和相应的控制方案,实际系统则根据人工系统产生的动态方案不断调整,两者动态调整、持续进化^[59]。平行系统的运行方式如图5所示,其可以总结为多环嵌套,基于传感器采集的数据,人工系统实时评估向实际系统下发的计划指令是否需要改变,若计划指令不变,则继续执行,构成从描述状态到物理状态的闭环;若评估后计划指令不变但计划实现方式需要调整,则调整计划指令,继续执行,构成从描述状态到引导状态的闭环;若评估后确认计划需要改变,则改变计划,重新预测规划,形成从描述状态到预测状态的闭环。

基于平行系统和机器博弈的社会认知平行博弈系统技术架构如图6所示。

对象包括社交媒体中的网民个体及其组织、社会热点事件及其地点以及网民参与社会舆情传播所使用的设备等。

感知层负责采集社会认知博弈建模与分析所需的数据,数据的来源包括手机、台式机、笔记本电脑、智能手环、GPS、摄像头等。

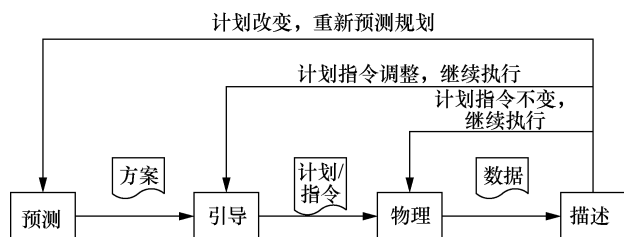


图5 平行系统的运行方式

存储层涉及对采集的海量、多模态数据的处理、融合与分析，主要流程为数据清洗、数据脱敏、特征工程、数据标记、数据存储和数据可视化等，其中关键技术包括大数据分析技术、云计算存储技术和多源异构传感器数据融合技术等。

算法层的核心是支撑平行系统和实现社会认知博弈实验的算法，主要包含三部分算法。其中，人工系统建模所需的算法包括自然语言处理、知识工程、模仿学习、情感分析、数学模型等，主要负责从真实社交媒体采集的社会传感^[60]数据中学习建模社会认知博弈中的各类智能体及其交互模式；计算实验基于强化学习、极大极小算法、Alpha-Beta剪枝算法、启发式搜索、蒙特卡洛搜索等机器博弈算法在给定的场景下学习产生最优的控制策略；平行执行模块基于自适应动态规划等算法实现对实际系统的控制和实时、在线优化，并基于增强现

实、统计分析等算法对策略运行效果进行分析及可视化。

应用层基于算法层提供的智能分析和计算，通过系统提供的接口和界面实现谣言识别、事件溯源、攻击检测、策略生成、自动发帖、舆情引导等功能，进而实现自适应优化的智能社会认知博弈，保障国家的安全和稳定。

5 结束语

机器博弈是以人机对弈方法使机器模拟人类决策、学习人类智能，进而使机器智能达到或超越人类智能的技术，是能够体现人工智能研究水平的极具挑战性的研究领域。近年来，人工智能在国际象棋、围棋等多个人机对抗游戏中战胜了人类，为更智能、优化地解决更复杂、普遍的博弈问题提供了一种可行的解决方案。本文首先总结了机器博弈的发展历程和关键技术，进而从现有的技术角度分析了机器智能的发展水平。

为了有效应对社交媒体、自媒体等带来的碎片化、海量舆情传播问题及日益普遍、严峻的社会认知战等新型威胁挑战，本文以平行系统方法和机器博弈技术为基础，提出一种社会认知平行博弈方法，通过构建社会认知平行博弈系统，基



图6 基于平行系统和机器博弈的社会认知平行博弈系统技术架构

于计算实验学习和搜索针对社会认知攻击的最优博弈策略,实现对真实社交媒体的自适应优化管控。该方法为解决高度复杂、难以实验的社会治理问题提供了一种新的手段,可以预见,以机器博弈为核心的人工智能方法与以平行系统为基础的复杂系统管理与控制方法的结合将能够有效应对大数据时代的社会治理挑战,提升我国的社会治理水平。

参考文献:

- [1] 徐心和, 邓志立, 王驷, 等. 机器博弈研究面临的挑战[J]. 智能系统学报, 2008, 3(4): 288-293.
XU X H, DENG Z L, WANG J, et al. Challenging issues facing computer game research[J]. CAAI Transactions on Intelligent Systems, 2008, 3(4): 288-293.
- [2] 张小川, 唐艳, 梁宁宇. 采用时间差分算法的九路围棋机器博弈系统[J]. 智能系统学报, 2012, 7(3): 278-282.
ZHANG X C, TANG Y, LIANG N N. A 9 × 9 Go computer game system using temporal difference[J]. CAAI Transactions on Intelligent Systems, 2012, 7(3): 278-282.
- [3] 张小川, 王宛宛, 彭丽蓉. 一种军棋机器博弈的多棋子协同博弈方法[J]. 智能系统学报, 2020, 15(2): 399-404.
ZHANG X C, WANG W W, PENG L R. A multi-chess collaborative game method for military chess game machine[J]. CAAI Transactions on Intelligent Systems, 2020, 15(2): 399-404.
- [4] 沈宇, 韩金朋, 李灵犀, 等. 游戏智能中的 AI: 从多角色博弈到平行博弈[J]. 智能科学与技术学报, 2020, 2(3): 205-213.
SHEN Y, HAN J P, LI L X, et al. AI in game intelligence—from multi-role game to parallel game[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(3): 205-213.
- [5] SHANNON C E. Programming a computer for playing chess[J]. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1950, 41(314): 256-275.
- [6] BERNSTEIN A, ARBUCKLE T, DE V ROBERTS M, et al. A chess playing program for the IBM704[C]//Proceedings of the Western Joint Computer Conference: Contrasts in Computers. [S.l.:s.n.], 1958: 157-159.
- [7] 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.
- [8] 唐平中, 朱军, 俞扬, 等. 动态不确定条件下的人工智能[J]. 中国科学基金, 2018, 32(3): 266-270.
TANG P Z, ZHU J, YU Y, et al. AI under dynamics and uncertainties[J]. Bulletin of National Natural Science Foundation of China, 2018, 32(3): 266-270.
- [9] 李翔, 姜晓红, 陈英芝, 等. 基于手牌预测的多人无限注德州扑克博弈方法[J]. 计算机学报, 2018, 41(1): 47-64.
LI X, JIANG X H, CHEN Y Z, et al. Game in multiplayer no-limit texas Hold'em based on hands prediction[J]. Chinese Journal of Computers, 2018, 41(1): 47-64.
- [10] 文习明. 人工智能时代的社会治理: 机遇、挑战与总体框架设计[J]. 岭南学刊, 2019(3): 83-88.
- WEN X M. Social governance in the age of artificial intelligence: opportunities, challenges and overall framework design[J]. Lingnan Journal, 2019(3): 83-88.
- [11] 赵宜萱, 赵曙明, 栾佳锐. 基于人工智能的人力资源管理: 理论模型与研究展望[J]. 南京社会科学, 2020(2): 36-43.
ZHAO Y X, ZHAO S M, LUAN J R. Human resource management based on artificial intelligence: a conceptional model and future research[J]. Nanjing Journal of Social Sciences, 2020(2): 36-43.
- [12] 朱燕丹, 靖鸣. 传播与权力的博弈: 新媒体视域下社会治理的问题与对策[J]. 江西师范大学学报(哲学社会科学版), 2016, 49(4): 50-56.
ZHU Y D, JING M. The game of communication and power: the research about the problems and the countermeasures of social governance from the perspective of new-media[J]. Journal of Jiangxi Normal University (Philosophy and Social Sciences Edition), 2016, 49(4): 50-56.
- [13] 易轩宇. 社会协同治理中社会组织的博弈评价与优化对策[J]. 甘肃社会科学, 2014(6): 190-194.
YI X Y. Game evaluation and optimal countermeasures of social organizations in social cooperative governance[J]. Gansu Social Sciences, 2014(6): 190-194.
- [14] 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 societies, computational experiments, and parallel systems: a discussion on computational theory of complex social-economic systems[J]. Complex Systems and Complexity Science, 2004, 1(4): 25-35.
- [15] 杨林瑶, 韩双双, 王晓, 等. 网络系统实验平台: 发展现状及展望[J]. 自动化学报, 2019, 45(9): 1637-1654.
YANG L Y, HAN S S, WANG X, et al. Computational experiment platforms for networks: the state of the art and prospect[J]. Acta Automatica Sinica, 2019, 45(9): 1637-1654.
- [16] 王亚杰, 邱虹坤, 吴燕燕, 等. 计算机博弈的研究与发展[J]. 智能系统学报, 2016, 11(6): 788-798.
WANG Y J, QIU H K, WU Y Y, et al. Research and development of computer games[J]. CAAI Transactions on Intelligent Systems, 2016, 11(6): 788-798.
- [17] KORF R E. Depth-first iterative-deepening: an optimal admissible tree search[J]. Artificial Intelligence, 1985, 27(1): 97-109.
- [18] ROIZEN I, PEARL J. A minimax algorithm better than Alpha-Beta? Yes and No[J]. Artificial Intelligence, 1983, 21(1/2): 199-220.
- [19] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.
- [20] 徐心和, 王驷. 中国象棋计算机博弈关键技术分析[J]. 小型微型计算机系统, 2006, 27(6): 961-969.
XU X H, WANG J. Key technologies analysis of Chinese chess computer game[J]. Journal of Chinese Computer Systems, 2006, 27(6): 961-969.
- [21] BROWN N, SANDHOLM T. Superhuman AI for multiplayer poker[J]. Science, 2019, 365(6456): 885-890.
- [22] ARULKUMARAN K, CULLY A, TOGELIUS J. AlphaStar: an evolutionary computation perspective[C]//Proceedings of the Genetic and Evolutionary Computation Conference Companion. [S.l.:s.n.], 2019: 314-315.
- [23] ETESSAMI K, LOCHBIHLER A. The computational complexity of

- evolutionarily stable strategies[J]. *International Journal of Game Theory*, 2008, 37(1): 93-113.
- [24] BROWNE C B, POWLEY E, WHITEHOUSE D, et al. A survey of Monte Carlo tree search methods[J]. *IEEE Transactions on Computational Intelligence and AI in Games*, 2012, 4(1): 1-43.
- [25] MORAVČÍK M, SCHMID M, BURCH N, et al. DeepStack: expert-level artificial intelligence in heads-up no-limit poker[J]. *Science*, 2017, 356(6337): 508-513.
- [26] 王军, 曹雷, 陈希亮, 等. 多智能体博弈强化学习研究综述[J]. *计算机工程与应用*, 2021, 57(21): 1-13.
- WANG J, CAO L, CHEN X L, et al. Overview on reinforcement learning of multi-agent game[J]. *Computer Engineering and Applications*, 2021, 57(21): 1-13.
- [27] WANG X F. Reinforcement learning to play an optimal Nash equilibrium in team Markov games[J]. *Advances in neural information processing systems*, 2002, 15: 1603-1610.
- [28] LESLIE D S, COLLINS E J. Generalised weakened fictitious play[J]. *Games and Economic Behavior*, 2006, 56(2): 285-298.
- [29] RYU H, SHIN H, PARK J. Multi-agent actor-critic with hierarchical graph attention network[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(5): 7236-7243.
- [30] 曾隽芳, 牟佳, 刘禹. 多智能体群智博弈策略轻量化问题[J]. *指挥与控制学报*, 2020, 6(4): 381-387.
- ZENG J F, MOU J, LIU Y. Lightweight issues of swarm intelligence based multi-agent game strategy[J]. *Journal of Command and Control*, 2020, 6(4): 381-387.
- [31] 蒋胤傑, 况琨, 吴飞. 大数据智能: 从数据拟合最优解到博弈对抗均衡解[J]. *智能系统学报*, 2020, 15(1): 175-182.
- JIANG Y J, KUANG K, WU F. Big data intelligence: from the optimal solution of data fitting to the equilibrium solution of game theory[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(1): 175-182.
- [32] 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 multiagent game research[J]. *Unmanned Systems Technology*, 2019, 2(1): 5-16.
- [33] 焦尚彬, 刘丁. 博弈树置换表启发式算法研究[J]. *计算机工程与应用*, 2010, 46(6): 42-45.
- JIAO S B, LIU D. Research on translation table heuristic algorithm[J]. *Computer Engineering and Applications*, 2010, 46(6): 42-45.
- [34] SCHAEFFER J. The history heuristic and Alpha-Beta search enhancements in practice[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(11): 1203-1212.
- [35] SAKUTA M, HASHIMOTO T, NAGASHIMA J, et al. Application of the killer-tree heuristic and the lambda-search method to lines of action[J]. *Information Sciences*, 2003, 154(3/4): 141-155.
- [36] CHASLOT G M J B, WINANDS M H M, HERIK H J V D, et al. Progressive strategies for Monte Carlo tree search[J]. *New Mathematics and Natural Computation*, 2008, 4(3): 343-357.
- [37] 季辉, 丁泽军. 双人博弈问题中的蒙特卡洛树搜索算法的改进[J]. *计算机科学*, 2018, 45(1): 140-143.
- JI H, DING Z J. Improvement of Monte Carlo tree search algorithm in two-person game problem[J]. *Computer Science*, 2018, 45(1): 140-143.
- [38] SCHMIDHUBER J. Deep learning in neural networks: an overview[J]. *Neural Networks*, 2015, 61: 85-117.
- [39] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529-533.
- [40] WU Y, TIAN Y. Training agent for first-person shooter game with actor-critic curriculum learning[C]//*Proceedings of the International Conference on Learning Representations (ICLR)*. New York: ACM Press, 2017.
- [41] PENG P, WEN Y, YANG Y D, et al. Multiagent bidirectionally-coordinated nets: emergence of human-level coordination in learning to play StarCraft combat games[J]. *arXiv preprint*, 2017, arXiv: 1703.10069.
- [42] 唐振韬, 邵坤, 赵冬斌, 等. 深度强化学习进展: 从 AlphaGo 到 AlphaGo Zero[J]. *控制理论与应用*, 2017, 34(12): 1529-1546.
- TANG Z T, SHAO K, ZHAO D B, et al. Recent progress of deep reinforcement learning: from AlphaGo to AlphaGo Zero[J]. *Control Theory & Applications*, 2017, 34(12): 1529-1546.
- [43] 王晓, 韩双双, 杨林瑶, 等. 基于 ACP 的动态网民群体运动组织建模与计算实验研究[J]. *自动化学报*, 2020, 46(4): 653-669.
- WANG X, HAN S S, YANG L Y, et al. The research on ACP-based modeling and computational experiment for cyber movement organizations[J]. *Acta Automatica Sinica*, 2020, 46(4): 653-669.
- [44] 李强, 阳东升, 孙江生, 等. “社会认知战”: 时代背景、概念机理及引领性技术[J]. *指挥与控制学报*, 2021, 7(2): 97-106.
- LI Q, YANG D S, SUN J S, et al. Social cognition operations: backgrounds, concepts, mechanisms and leading technologies[J]. *Journal of Command and Control*, 2021, 7(2): 97-106.
- [45] 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.
- DENG J L, WANG F Y, CHEN Y B, 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.
- [46] 杨林瑶, 陈思远, 王晓, 等. 数字孪生与平行系统: 发展现状、对比及展望[J]. *自动化学报*, 2019, 45(11): 2001-2031.
- YANG L Y, CHEN S Y, WANG X, et al. Digital twins and parallel systems: state of the art, comparisons and prospect[J]. *Acta Automatica Sinica*, 2019, 45(11): 2001-2031.
- [47] 王飞跃, 王艳芬, 陈慧竹, 等. 联邦生态: 从联邦数据到联邦智能[J]. *智能科学与技术学报*, 2020, 2(4): 305-311.
- WANG F Y, WANG Y F, CHEN Y Z, et al. Federated ecology: from federated data to federated intelligence[J]. *Chinese Journal of Intelligent Science and Technology*, 2020, 2(4): 305-311.
- [48] 陈龙, 王晓, 杨健健, 等. 平行矿山: 从数字孪生到矿山智能[J]. *自动化学报*, 2021, 47(7): 1633-1645.
- CHEN L, WANG X, YANG J J, et al. Parallel mining operating systems: from digital twins to mining intelligence[J]. *Acta Automatica Sinica*, 2021, 47(7): 1633-1645.
- [49] 王春法, 王飞跃, 鲁越, 等. 平行博物馆: 新时代博物馆运营的智能管理与控制[J]. *智能科学与技术学报*, 2021, 3(2): 125-136.
- WANG C F, WANG F Y, LU Y, et al. Parallel museums: intelligent management and control of museum operations in the new era[J]. *Chinese Journal of Intelligent Science and Technology*, 2021, 3(2):

125-136.

- [50] 沈宇, 王晓, 韩双双, 等. 代理技术 Agent 在智能车辆与驾驶中的应用现状[J]. 指挥与控制学报, 2019, 5(2): 87-98.
SHEN Y, WANG X, HAN S S, et al. Agent-based technology in intelligent vehicles and driving: state-of-the-art and prospect[J]. Journal of Command and Control, 2019, 5(2): 87-98.
- [51] DUARTE F F, LAU N, PEREIRA A, et al. A survey of planning and learning in games[J]. Applied Sciences, 2020, 10(13): 4529.
- [52] LASKEY M, STASZAK S, HSIEH W Y S, et al. SHIV: reducing supervisor burden in DAGger using support vectors for efficient learning from demonstrations in high dimensional state spaces[C]//Proceedings of 2016 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2016: 462-469.
- [53] DALEY D J, KENDALL D G. Epidemics and rumours[J]. Nature, 1964, 204(4963): 1118.
- [54] 马知恩, 周义仓, 王稳地, 等. 传染病动力学的数学建模与研究[M]. 北京: 科学出版社, 2004.
MA Z E, ZHOU Y C, WANG W D, et al. Mathematical modeling and research on the dynamics of infectious diseases[M]. Beijing: Science Press, 2004.
- [55] ANDERSON R M, JACKSON H C, MAY R M, et al. Population dynamics of fox rabies in Europe[J]. Nature, 1981, 289(5800): 765-771.
- [56] WITBOOI P J. Stability of an SEIR epidemic model with independent stochastic perturbations[J]. Physica A: Statistical Mechanics and Its Applications, 2013, 392(20): 4928-4936.
- [57] 郭东伟, 乌云娜, 邹蕴, 等. 基于非理性博弈的舆情传播仿真建模研究[J]. 自动化学报, 2014, 40(8): 1721-1732.
GUO D W, WU Y N, ZOU Y, et al. Simulation and modeling of non-rational game based public opinion spread[J]. Acta Automatica Sinica, 2014, 40(8): 1721-1732.
- [58] 阳东升, 张维明. PREA 环及其平行智能[J]. 指挥与控制学报, 2019, 5(4): 274-281.
YANG D S, ZHANG W M. PREA loop and its parallel intelligence[J]. Journal of Command and Control, 2019, 5(4): 274-281.
- [59] HAN J P, WANG F Y, LYU Y S, et al. Efficient deployment of patrols to catch arsonists[C]//Proceedings of 2018 Chinese Automation Congress (CAC). Piscataway: IEEE Press, 2018: 792-797.
- [60] 陈虹宇, 艾红, 王晓, 等. 社会交通中的社会信号分析与感知[J]. 自动化学报, 2021, 47(6): 1256-1272.
CHEN H Y, AI H, WANG X, et al. Analysis and perception of social signals in social transportation[J]. Acta Automatica Sinica, 2021, 47(6): 1256-1272.

[作者简介]



李亚玲 (1991-), 女, 博士, 之江实验室智能社会治理研究中心助理研究员, 主要研究方向为人工智能技术的应用。



杨林瑶 (1995-), 男, 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博士生, 主要研究方向为图表示学习、知识融合。



葛俊 (1989-), 男, 之江实验室智能社会治理研究中心工程师, 主要研究方向为多智能体协同控制、强化学习等。



覃缘琪 (1992-), 女, 之江实验室智能社会治理研究中心工程师, 主要研究方向为智能交通控制、智能社会治理等。



王晓 (1988-), 女, 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副研究员, 主要研究方向为社会交通、动态网群组织、人工智能和社会网络分析。