人机对抗智能技术

黄凯奇*, 兴军亮, 张俊格, 倪晚成, 徐博

中国科学院自动化研究所智能系统与工程研究中心, 北京 100190

* 通信作者. E-mail: kqhuang@nlpr.ia.ac.cn

收稿日期: 2019-02-27; 接受日期: 2019-06-06; 网络出版日期: 2020-04-10

摘要 人机对抗作为人工智能研究的前沿方向, 已成为国内外智能领域研究的热点, 并为探寻机器智能内在生长机制和关键技术验证提供有效试验环境和途径. 本文针对巨复杂、高动态、不确定的强对抗环境对智能认知和决策带来的巨大挑战, 分析了人机对抗智能技术研究现状, 梳理了其内涵和机理, 提出了以博弈学习为核心的人机对抗智能理论研究框架; 并在此基础上论述了其关键模型: 对抗空间表示与建模、态势评估与推理、策略生成与优化、行动协同与控制; 为复杂认知与决策问题的可建模、可计算、可解释求解奠定了基础. 最后, 本文总结了当前应用现状并对未来发展方向进行了展望.

关键词 人工智能, 人机对抗, 机器学习, 智能博弈, 认知决策

1 引言

人类社会的发展史就是自身矛盾不断变化演进的历史. 比较、对比、竞争、斗争, 各种不同形式的对抗广泛存在于人类社会中的经济、政治、金融和生活等各个领域和行业之中. 对抗双方是人类不同的个体与个体、个体与群体以及群体与群体之间, 对抗贯穿于整个人类文明发展历史当中, 也在不断地促进人类智能的提升和进化. 随着社会不断发展和技术不断进步, 对抗的复杂度, 如速度、强度等都在呈指数级别增长. 人类的感知和决策能力由于受自身生理和心理方面的限制, 面对复杂、动态、对抗环境下智能认知与决策需求, 迫切需要以机器为载体的人工智能技术的辅助与支撑.

人工智能技术的发展通常可分为计算智能、感知智能和认知智能 3 个阶段^[1]. 传统的人工智能以强大的计算能力为主要特征, 其特点是快速计算和记忆存储能力大幅超越人类. 近年来, 以机器学习为代表的人工智能技术得到了飞速发展, 使得机器获取、处理和分析信息的水平得到了大幅提升. 感知智能算法在语音处理、图像识别、文本翻译等诸多问题中都已经达到甚至局部超过人类的感知能力^[2~4]. 在认知决策领域, 以 AlphaGo^[5]、冷扑大师^[6] 等为代表的智能博弈算法在边界确定、规则固定的人机对抗问题中已经战胜了人类顶级专业选手, 为人工智能技术由感知智能向认知智能跨越式发展带来了新的曙光.

引用格式: 黄凯奇, 兴军亮, 张俊格, 等. 人机对抗智能技术. 中国科学: 信息科学, 2020, 50: 540–550, doi: 10.1360/N112019-00048
Huang K Q, Xing J L, Zhang J G, et al. Intelligent technologies of human-computer gaming (in Chinese). Sci Sin Inform, 2020, 50: 540–550, doi: 10.1360/N112019-00048

图 1 人机对抗的发展历史

Figure 1 The development history of human-computer gaming

人工智能从诞生之始就以对抗人类智能为衡量准则 (著名的图灵测试)^[7]. 作为人工智能研究的前沿方向, 人机对抗技术一直是国内外人工智能研究的热点, 以人机对抗为主要形式的人工智能研究为探寻机器智能的内在生长机制和关键技术原理提供了一个极佳的试验环境和验证途径. 基于人类现有的知识和智能水平, 通过研究和实现新型的智能学习算法, 使得机器智能在可控的范围内不断接近、达到甚至超过人类的智能水平. 整个过程不仅能够让机器更加智能地为人类服务、将人类从一些繁复复杂的任务中解脱出来, 而且人类也能够借鉴机器智能的发展过程, 提升自身的智能水平, 更深刻地理解和掌握智能的内在本质和产生机理, 进而推动整个社会由信息化向智能化发展.

2 发展历史

从 1936 年人工智能之父阿兰·图灵提出著名的“图灵测试”之后, 人和机器之间进行智能对抗就成为了衡量机器智能发展水平的最重要标准^[8]. 从实现最初的简单智能到不断进化成更为复杂的智能, 对抗贯穿于人工智能中计算、感知和认知的 3 个阶段 (图 1), 从某种意义上讲, 就如人类与自然环境的对抗促进了人类的进化, 机器与人的对抗也必将让机器智能不断发展进化.

计算智能主要以科学运算、逻辑处理、统计查询等形式化规则化运算为主要任务, 表现是能存会算会查找, 20 世纪 50 年代, 计算机的发明和普及为人工智能的发展注入了最初的驱动力, 使得计算机很快在计算速度和精度上超过了人类, 标志性事件就是 1997 年 IBM 的深蓝计算机大战国际象棋世界冠军卡斯帕罗夫 (Kasparov). 深蓝计算机的国际象棋 AI 程序基于人类棋手的经验, 总结了大量规则, 借助于计算机强大的计算和搜索能力战胜了人类国际象棋冠军.

感知智能主要以语音识别、图像理解、文本翻译等为主要任务, 表现是能听会说、能看会认. 21 世纪伊始, 以深度学习为代表的机器学习技术在这些任务中取得了巨大突破^[9,10], 标志性事件包括微软公司在 2008 年前后推出的基于深度学习的语音识别引擎大幅降低了识别错误率^[2], 2012 年的 ImageNet 竞赛中基于深度学习的图像分类模型以绝对优势取得第 1 名^[3], 以及谷歌公司推出的高精度实时语言翻译系统^[4]等.

认知智能主要以理解、推理、思考和决策等为主要任务, 表现是能理解、会思考决策. 虽然计算

机在认知智能方面目前才刚刚起步, 还未达到人类的水平, 但是认知智能的应用前景广阔, 影响深远, 因此认知智能相关的研究如火如荼^{[5,6,11,12]1)2)}. 近几年来, 以谷歌旗下的 DeepMind 公司为代表, 在认知决策智能方面取得了一些突破性进展, 其中标志性进展就是 AlphaGo 智能围棋博弈程序^[5] 在 2016 年 4 : 1 战胜人类围棋冠军选手李世石, 标志着人工智能的发展重点逐渐由感知智能向认知智能过渡. 2017 年, 加拿大阿尔伯塔大学 (University of Alberta) 和美国卡内基梅隆大学 (Carnegie Mellon University) 提出的德州扑克 AI 算法^[5,6], 在 1 对 1 无限注德州扑克中, 击败了多名顶尖的职业德州扑克选手. 2018 年 OpenAI 公司开发的 OpenAI-Five 智能程序, 以 2 : 1 击败了由 5 名人类职业选手组成的队伍, 2019 年谷歌 AI 程序 AlphaStar 以 10 : 1 战胜人类职业选手. 鉴于决策等认知智能的重要性, 国内外高科技公司和机构围绕着博弈对抗进行了大量布局, 以此作为发展通用人工智能的路径之一. 2016 年谷歌发布了基于星际争霸的游戏 AI 训练平台^[13], 同年 OpenAI 公司发布了通用人工智能训练测试平台 Universe³⁾, 2017 年 Facebook 公司开源了游戏学习平台^[14], 国内腾讯公司在 2017 年全球合作伙伴大会上发布了 AI 战略, 围绕游戏、社交、内容 3 大场景, 并将博弈对抗 AI 做为第重要板块⁴⁾.

3 内涵及机理

虽然人机对抗智能技术获得极大的关注, 并在不同应用领域展现出应用前景, 但其理论及相关技术目前还没有一个系统性的描述, 本文将从内涵、机理对其进行阐述并在此基础上相关模型及关键技术.

3.1 人机对抗内涵

人机对抗研究计算机博弈中机器战胜人类的方法, 其更一般性的含义是指在强对抗博弈环境下, 以人机对抗为主要形式研究机器智能战胜人类智能的内在机理和技术原理, 它是智能领域研究的核心方向, 通过人、机和环境之间的交互学习, 探索巨复杂、高动态和不确定的强对抗环境下机器智能快速增长的内在机理和途径, 以期最终达到或者超过人类的智能水平. 在人机对抗研究体系中, 人机对抗是促使机器智能快速增长的主要方式, 在人机对抗过程中, 人和机器之间相互借鉴学习. 根据不同的应用, 人机对抗技术又可以表现为人机协同、人机融合, 以至达到最终的人机共生. 人机对抗的能力增强表现为首先从特定领域战胜单个人, 再到战胜群体, 最后在该领域超越整个人类群体的智能水平.

总之, 人机对抗是以机器和人类对抗为途径, 以博弈学习等为核心技术来实现机器智能快速学习进化的研究方向. 作为“图灵测试”的重要手段, 人机对抗是验证机器智能的“试金石”, 为探寻机器智能的内在生长机制和关键技术的验证提供有效试验环境、评价标准, 具有重要科学研究意义和应用价值.

3.2 人机对抗机理

人机对抗机理研究对抗各要素及其相互联系, 以及相互作用的运行规律与原理. 要素是最小的组成单元, 人机对抗涉及的要素包括人 (机器的对手)、机器 (对抗的 AI) 和环境 (对抗的规则和条件等的集合体), 我们将之抽象为人、机、物 3 要素. 在强对抗环境下的人机对抗过程中, 对抗使博弈这个

1) Open AI Dota Five. <https://blog.openai.com/openai-five>.

2) AlphaStar. <https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/>.

3) OpenAI Universe. <https://openai.com/blog/universe/>.

4) 腾讯 2017 年全球合作伙伴大会. <https://tgpc.qq.com/2017/mobile/index>.

概念进一步扩展. 人或机器去感知和适应环境的过程也是人或机器与环境进行博弈对抗的过程. 由于环境本身的高度复杂性、动态变化性和随机不定性, 在与之博弈的过程中, 人或机器需要对它的静态特性进行建模, 同时还需要对它的动态特性进行学习.

根据人机物 3 要素分析方法, 3 要素相互作用分别形成一元博弈、二元博弈和三元博弈. 一元博弈是指博弈中只有一个要素变化, 而其他两个要素固定的博弈过程. 最常见一元博弈过程就是环境和人都固定的机机博弈. 二元博弈是指博弈中有两个要素变化的博弈过程, 包括人机博弈、机物博弈和人物博弈. 机物博弈可以理解机器如何去对环境的动态属性进行建模并学习环境的动态变化性. 而三元博弈是指博弈中 3 个要素都发生变化的博弈过程, 可进一步分为有限边界的三元博弈和开放环境的三元博弈. 随着博弈中变化要素数目的增多, 博弈就愈加复杂.

博弈思想也存在于机器学习过程中, 传统的监督学习从给定的训练数据集中学习出一个函数 (模型参数), 当新的数据到来时, 即根据这个函数预测结果. 其中模型的训练过程可以理解成单个模型依照给定环境 (数据集) 不断通过博弈调整模型参数的过程, 近年来基于产生式模型和判别式模型博弈的对抗生成对抗网络 (generative adversarial networks, GAN) 也成为继深度学习之后的人工智能研究热点^[15]. 谷歌 DeepMind 公司设计的星际争霸游戏 AI 程序 AlphaStar 中使用的基于种群的强化学习技术包含了 500 多个模型之间的相互博弈学习.

人机对抗智能的科学问题可概括为博弈学习的可建模、可计算与可解释. 具体来说, 可建模是指博弈过程可以基于数学模型进行抽象或者表征学习, 它贯穿于人机对抗的整个过程. 以 AlphaGo 智能程序^[5]为例, 围棋博弈被建模为完全信息静态博弈, 并使用蒙特卡洛 (Monte Carlo) 树搜索进行博弈求解, 而 Libratus 智能程序中, 德州扑克博弈被建模为不完全信息动态博弈, 并使用反事实后悔值最小化方法进行博弈求解. 可计算是指博弈行为模型可以进行计算与学习, 并能在此基础上进行大幅优化, 它为人机对抗策略生成及使用提供了保障. 以 Libratus 程序为例, 最初版本在与人对抗时需要使用匹兹堡 (Pittsburgh) 超级计算机进行计算, 而后改进版本 Modicum 利用深度有限求解方法优化其计算性能, 实现了在普通笔记本上就可以与人进行博弈^[16]. 可解释是指博弈行为的计算结果可以解释, 它为博弈过程中第三方的介入提供了足够可以理解的信息. 目前以深度学习为代表的人工智能模型可解释性较差, 已有不少研究尝试从特征可视化^[17]、损失界面分析^[18]等不同角度对深度学习模型进行分析, 而可解释性对以认知决策为重点的人机对抗技术更为关键, 其研究为进一步的人机混合智能提供了基础保障.

4 人机对抗模型及关键技术

和感知智能有所不同, 人机对抗通常关注更为复杂的时序决策等认知智能, 对其过程建模是一个高度复杂问题, 因此认知决策建模是整个人机对抗中的核心关键环节. 已有学者从不同领域对认知决策过程进行建模, 其中 Simon 等^[19]从管理学角度提出的决策环理论将其归纳为“情报收集 – 方案设计 – 评价方案 – 实施方案”, 经典指挥控制模型中的 SHORE 模型则称其为“激励 – 假设 – 选择 – 响应”^[20], 著名的 OODA 环模型将其表示为“感知 – 判断 – 决策 – 行动”^[21], 我国军事指挥专家胡晓峰^[22]则将其归纳为“态势获取 – 态势理解 – 态势预测 – 决策 – 行动”.

本文将面向强对抗环境下的人机对抗决策流程归纳为感知、推理、决策和控制. 感知是基于大量环境数据, 对态势的判读和理解, 例如目标聚类、威胁评估等. 推理是在态势感知基础上, 对态势的深度分析、认知和预测, 包括对意图的判定, 对抗结局的预测等. 决策和控制是在感知和推理之后的策略决策和方式实施等. 整个决策环路过程在局部和整体上都不断进行循环迭代, 从而实现认知决策的迭

图 2 人机对抗的过程建模和关键技术

Figure 2 Process modeling and key technology of human-computer gaming

代增强, 对抗能力的自主提升. 本文提出的人机对抗模型具有一定的普适性. 美军在深绿计划中提出 OO-DA 模型^[23] 和在第三次抵消战略中提出的 OPA (observe-prediction-act)^[24] 可以看成是本文模型的一些特例和变种. 本文将人机对抗关键技术归纳为对抗空间表示与建模、态势评估与推理、策略生成与优化、行动协同与控制 4 部分; 通过对抗态势判读理解、认知预测、策略决策和行动实施, 局部整体不断循环迭代增强, 自主提升对抗能力. 人机对抗涉及的关键技术如图 2 所示.

4.1 对抗空间表示与建模

构建有效的知识表示模型, 准确刻画对抗空间的决策要素构成、属性特征及要素之间的交互关系, 是实现人机对抗的基础. 巨复杂、高动态、强对抗环境具有决策要素海量高维、要素影响高度耦合、决策关键信息不完全等特性, 使得对抗空间的定量表示极富挑战, 包括: (1) 超高维度决策要素及其高度耦合的关联影响, 使得对抗空间的特征表示、提取、学习和预测十分困难; (2) 对抗双方行为与场景特征高度依赖, 对抗空间表示需要具备跨场景、可迁移的能力; (3) 对手信息缺失甚至带有高度欺骗性, 如何表征缺失及不可靠的对手信息, 依据不完全信息推理整体态势, 也是极富挑战性的问题.

有效的知识表示模型如知识图谱是研究对抗空间表示与建模的主流研究方法^[25]. 同时, 基于模式识别的特征表示方法如多通道图像的对抗空间张量表示方法也被广泛采用^[5,11]. Racaniere 等^[26] 使用序列预测对环境进行建模, 并通过前向推演来改进策略, 提高了算法的性能和鲁棒性. Huang 等^[27] 使用自举法来衡量环境模型的可信度, 并增加对模型预测可信度低的状态的探索次数, 从而有效减少了模型的潜在误差. 这方面可开展的研究包括: (1) 对抗空间要素实体 - 关系表示, 研究对抗空间中各类不同实体、实体属性及其交互关联关系, 构建决策要素的表示模型; (2) 对抗空间特征张量表示学习, 分析实体属性关系耦合和结构拓扑对个体及群体对抗能力的影响, 构建可解释的对抗空间的高维张量描述; (3) 层级化聚合演算规则建模, 融合经验表示与数值计算, 定义多因素、层级化的对抗态势

和对抗能力聚合演算规则,形成环境-我方-对手多元耦合的可计算表征体系;(4)基于异构信息网络的抽象通用空间表示,基于能力演算规则,研究对抗要素的抽象态势表示,减轻对抗场景依赖带来的噪声和数据稀疏影响。

4.2 对抗态势评估与推理

对抗态势是指对抗各方通过实力对比、调配和行动等形成的状态和趋势^[28],态势的评估与推理为后续对抗策略生成与优化提供了依据。面对人机对抗空间各种要素非线性的交叉影响,态势评估与推理面临着以下挑战:(1)训练态势认知和预测模型的高水平对抗数据往往非常有限;(2)对手信息往往是凌乱的、不完全的,仅依据对手部分信息进行整体局势评估准确性较差;(3)复杂对抗环境导致可用于态势评估的信息众多,难以有效进行融合以形成多角度层次化态势。

针对挑战,目前国内外已开展相关研究,谷歌 DeepMind 公司的研究人员^[5,11]基于自我博弈技术在围棋对抗过程中产生数据以进行估值网络的训练。Hu 等^[29]借助于深度神经网络学习状态与动作之间投影,进而实现对手信息估计。Lowe 等^[30]将其他智能体的策略估计嵌入到整体的强化学习过程中,可以实现协作与对抗环境下其他智能体策略的有效估计。Shen 等^[31,32]提出了深度态势匹配算法,可以有效识别态势之间的匹配度。此类可开展的研究包括:(1)高质量对抗数据的生成,通过自主博弈或者生成式对抗网络^[15]等方法生成用于人机对抗建模与分析的高质量人机对抗数据;(2)小样本学习,研究在数据较少情况下通过迁移或自适应等方法实现对抗态势的直接评估;(3)对方信息估计,研究结合历史信息与当前对抗环境实现从底层到高层的对手动作估计、意图识别与策略估计;(4)态势层次化认知,研究结合多源异构信息进行态势的多角度层次化评估与推理。

4.3 对抗策略生成与优化

对抗策略主要涉及多智能体协同的任务规划,解决群体与单体的行动规划问题,其技术挑战在于:不完全信息使得对手位置、行为、企图不能完全知道,对手行为概率模型未知导致策略选择保守,需要进行不完全信息下博弈策略选择;宏观决策收益反馈滞后,使得宏观决策的效应需要经过较长时间才能体现,导致决策行为与效益之间难以形成有效映射;行动能力与环境深度耦合,忽略了局部环境因素可能导致策略分析严重偏差,过度详细分析又导致对抗空间难以约减。

针对策略生成和优化方面的挑战,利用策略游戏如星际争霸作为平台开展研究是国际上较为公认的方式^[13,14]。相关方法包括使用学习的方法来选择特定的宏观运营策略^[33,34]和采矿、建造等特定微观操作^[35],利用构建顺序^[36]、自动策略生成^[37]等开展策略优化方法。可开展的研究包括:(1)宏观策略生成,针对面向使命任务的全局博弈对抗问题,构建分层任务分解与任务协同机制,实现复杂群体博弈对抗问题向低维空间约减;(2)微观策略生成,针对面向战斗任务的局部博弈对抗问题,构建微型群体局部策略自适应机制,实现微型群体的强博弈对抗能力与环境迁移能力;(3)策略优化方法,针对策略能力需要自主提升的问题,构建博弈策略的评价机制和学习型策略演进机制,实现博弈策略的自主进化与能力提升。

4.4 对抗行动协同与控制

策略的执行需要多个智能体的行动协同,各智能体在自身信息获取与初步认知的基础上,利用资源贡献、信息连通、要素融合、虚拟协作、智能辅助等功能,将多个单元虚拟协同,形成整合的群体行动协同与控制。多智能体协同的难点包括:多智能体的学习目标,个体回报和团队回报的关系,学习过程中各智能体之间的作用和影响,联合状态和联合动作的获取,扩大的状态空间和动作空间导致的维

数灾难等问题。

目前相关研究工作主要集中在多智能体协同与学习方面^[38]。以星际争霸游戏为例, 代表性工作包括学习多个智能体协同进行局部战斗时的战术配合和运营建造时的动作配合^[17, 35, 39]。其中多智能体的协同控制方式包括集中式^[40]、分布式^[41]和混合式^[35, 39], 而学习的方法则主要基于不同的深度强化学习模型变种^[42]。针对上述难点, 行动协同与控制可开展的研究包括: (1) 从协同过程可分为序列策略表征、协同机制优化、异构多智能体协同以及多元协同的融合; (2) 从协同任务类型可分为同任务合作智能协同、异任务资源协调等; (3) 从理论上突破去中心化、通信中断的默契型协作方法, 任务涵盖序列化任务、多层次任务、多领域任务等, 实现场景类型全覆盖, 协同方式多元化, 为训练提供高质量协同策略。

5 应用及挑战

人机对抗智能技术从其诞生起, 就因在不断挑战人类过程中验证其能力而受到广泛的关注, 从深蓝到 IBM Watson 直至 AlphaGo 和 AlphaStar, 不同环境下的人机对抗技术的突破不断成为这一领域的代表性工作, 尤其以 2016 年 AlphaGo 击败人类顶级棋手为标志, 拉开了新一轮人机大战的序幕。人机对抗智能技术的应用领域涉及棋牌类游戏、即时策略游戏、兵棋推演等, 在多个领域内机器智能已经达到并超过了该领域的人类顶级选手, 不断刷新博弈对抗记录, 显示出了新一轮人工智能技术在认知决策方面的鲜明特点。

5.1 棋牌类策略游戏对抗

棋牌类策略游戏一直以来都被用作测试计算机智能发展水平的参考标准。这些游戏由于简单的规则和丰富的玩法而深受全球范围内的广大爱好者的喜爱。同样由于其规则的确定性, 加之游戏环境可控性和不完全信息等特性, 吸引了人工智能领域的持续研究。传统棋牌类策略游戏 AI 主要基于人工编写的规则和一些启发式搜索算法, 近期基于机器学习的博弈对抗算法在棋牌类策略游戏对抗应用中取得了突破性进展, 其中一个标志性事件就是谷歌的 DeepMind 公司 2016 年相继提出的 AlphaGo 系列技术^[5, 11]在围棋人机对抗中不仅取得了压倒性的优势, 而且完全脱离人类知识从零开始学习^[11], 学到很多人类围棋选手无法理解的定式。2017 年来自加拿大阿尔伯塔大学等机构的研究人员提出了面向德州扑克的 AI (DeepStack)^[12], 该算法利用深度神经网络学习快速估计模型, 模拟人类选手的“直觉”估计器, 能够快速向前搜索特定步数的博弈策略。在 1 对 1 无限注德州扑克中, DeepStack 成为了第 1 个击败职业选手的德州扑克 AI 程序。同年, 来自于美国卡内基梅隆大学的研究者提出了一种名为 Libratus 的德州扑克 AI 算法^[6], 在 1 对 1 无限注德州扑克中, 同样击败了多名顶尖的职业德州扑克选手。上述两种算法采用的策略都是通过约简采样和学习搜索去逼近一个近似的纳什均衡, 保证 AI 程序尽量少犯错, 这种策略对于 1 对 1 无限注德州扑克博弈问题可以得到一个较好的解决方案, 但是对于多对多无限注德州扑克博弈问题, 现有的 AI 算法在理论上被证明无法适用, 因此需要研发新型博弈算法进行技术突破。

5.2 即时策略游戏对抗

即时策略游戏是另一种常用于评估机器智能的平台, 和棋牌类游戏相比, 其整个对抗过程即时进行。通常情况下, 该类游戏包含资源采集、基地建造、发展科技等若干因素, 游戏玩家需要平衡不同因素并控制单个或者多个被控制单元以完成对抗, 经典的实时策略游戏包括星际争霸以及 Dota2

等. 得益于国际 AI 赛事以及企业与学术界的友好合作, 即时策略游戏 AI 取得了长足的进展, 包括 DeepMind、Facebook、OpenAI、三星、中国科学院自动化研究所、斯坦福大学 (Stanford University) 等越来越多的知名机构与高校参与到即时策略游戏 AI 的研发中. 在 2018 年星际争霸 AI 国际挑战赛中, 三星、Facebook 与中国科学院自动化研究所荣获前 3 名, AI 算法已经初步具备了战胜业余高手的水平. 相关的 AI 算法包括基于多智能体强化学习实现被控单元间微操的控制, 基于深度神经网络模型进行宏观战斗决策的制定, 以及基于状态机模型实现对抗过程的演化等. 相比于星际争霸, Dota2 的每个玩家仅需要控制一个单位, 并协助其他 4 位玩家进行对抗. 2018 年 OpenAI 开发的 OpenAI-Five 智能程序基于深度强化学习完成对抗策略生成, 击败了由 5 名人类高手 (前职业选手) 组成的队伍 (比分为 2 : 1). 2019 年, DeepMind 团队继续在即时策略 AI 中不断取得突破, 设计的星际争霸 AI 程序 AlphaStar 以 10 : 1 击败了星际争霸游戏人类专业选手, 最新提出的基于种群的强化学习算法在《雷神之锤 3: 战场》游戏中的夺旗任务中团战胜率超过人类玩家^[43]. 尽管深度强化学习技术在多个游戏中具有不俗的表现, 如何提高模型的稳定性以及性能以击败顶级人类玩家、如何实现模型的可解释以支撑学习的可信任建模仍需要 AI 算法的不断突破.

5.3 军事模拟及推演对抗

军事模拟及推演由于其战略意义, 长期以来受到国家层面的关注, 其对抗演练具有重要实用价值. 2016 年美国辛辛那提大学 (University of Cincinnati) 构建了空战智能博弈人机对抗系统 (ALPHA), 击败了美国空军顶级飞行员李上校, 拉开了无人装备对抗有人装备的序幕. 空战 ALPHA 程序主要采用了遗传模糊树的自主进化决策逻辑, 空战决策速度比人类快了 250 倍. 兵棋推演是另一种利用兵棋进行模拟战争活动的平台, 军事指挥员使用代表环境和军事力量的棋盘和棋子, 依据军事规则和概率论原理, 模拟战争对抗, 对作战方案进行过程推演和评估优化. 2017 年在首届全国兵棋推演大赛上, 中国科学院自动化研究所研制的 CASIA 先知 1.0 系统, 击败了军队四强和地方四强 8 名冠军级选手, 获得了 7 : 1 的胜利, 在对抗中采用了面向作战筹划的使命任务分析、地形分析、兵力对比、对手行为估计、作战部署算法以及面向作战管理的态势认知与预测算法、作战决策和作战计划自动生成算法, 实现了博弈对抗与迁移能力的同步提升. 随着博弈对抗规模的扩大, 对抗空间呈现指数级增长, 多兵种协同与环境耦合的问题凸显, 战争系统具有强非线性和高动态等复杂特性, 解析计算和随机逼近最佳策略都存在巨大挑战; 人机对抗需要发展对手行为意识建模和协同演化博弈策略, 以不断提升对抗能力.

以上我们只是简单列举了 3 类人机对抗的相关应用, 事实上, 博弈贯穿于人类社会的方方面面^[8, 44], 博弈和对抗在政治、经济、文化、军事等多方面都发挥着重要的作用, 人机对抗技术也将在信息获取、传输、分析、理解、推理、决策等环节发挥作用, 推动感知和认知智能的发展.

6 总结与展望

本文着眼于智能化时代的智能需求, 以利用人机对抗形成逼近人类智能为牵引, 针对人机对抗智能的关键技术进行探讨, 阐述了人机对抗智能技术研究框架下的空间表示与建模、态势评估与推理、策略生成与优化、行动协同与控制相关概念与模型, 以及各模块的重点研究内容与难点, 同时就相关对抗应用领域现状、典型算法、存在的主要问题进行了介绍. 从人机对抗技术的发展可以看到, 无论是人机对抗的基础理论, 还是主要环节的关键技术, 依然存在诸多关键难题, 通用的对抗空间表示与建模、精准的态势评估与推理、完善的策略生成与优化, 以及对手心理建模等方法与技术是迫切需要攻克的难题.

总之, 人工智能尤其是机器学习、类脑计算等领域的发展给人机对抗智能及自主进化带来了契机, 迫切需要深化人机对抗问题本质的理解与解析, 科学合理地建立机器智能与人类智能的高效协作机制, 取得人机对抗理论、技术与应用的重大突破, 并在人类社会的经济、政治、金融及生活等多领域推进智能化进程。

参考文献

- 1 Tan T N. Artificial Intelligence - Building an Intelligent Future With AI Technologies. Beijing: China Science and Technology Press, 2019 [谭铁牛. 人工智能 - 用 AI 技术打造智能化未来. 北京: 中国科学技术出版社, 2019]
- 2 Li D, Dong Y. Deep Learning: Methods and Applications. Boston: Now Publishers Inc., 2014
- 3 Huang K Q, Ren W Q, Tan T N. A review on image object classification and detection. Chin J Comput, 2014, 37: 1-18 [黄凯奇, 任伟强, 谭铁牛. 图像物体分类与检测算法综述. 计算机学报, 2014, 37: 1-18]
- 4 Li D, Yang L. Deep Learning in Natural Language Processing. Switzerland: Springer Press, 2018
- 5 Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search. Nature, 2016, 529: 484-489
- 6 Brown N, Sandholm T. Superhuman AI for heads-up no-limit poker: libratus beats top professionals. Science, 2018, 359: 418-424
- 7 Turing A M. Computing machinery and intelligence. In: Parsing the Turing Test. Dordrecht: Springer. 2009
- 8 Zhang W Y. Gaming and Scocity. Beijing: Peking University Press, 2013 [张维迎. 博弈与社会. 北京: 北京大学出版社, 2013]
- 9 Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. Science, 2006, 313: 504-507
- 10 Huang K Q, Chen X T, Kang Y F, et al. Intelligent visual surveillance: a review. Chin J Comput, 2015, 38: 1093-1118 [黄凯奇, 陈晓棠, 康运锋, 等. 智能视频监控技术综述. 计算机学报, 2015, 38: 1093-1118]
- 11 Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge. Nature, 2017, 550: 354-359
- 12 Moravčík M, Schmid M, Burch N, et al. DeepStack: expert-level artificial intelligence in heads-up no-limit poker. Science, 2017, 356: 508-513
- 13 Vinyals O, Ewalds T, Bartunov S, et al. StarCraft II: a new challenge for reinforcement learning. 2017. ArXiv: 1708.04782
- 14 Tian Y, Gong Q, Shang W, et al. ELF: an extensive, lightweight and flexible research platform for real-time strategy games. 2017. ArXiv: 1707.01067
- 15 Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Proceedings of Neural Information Processing Systems, 2014. 2672-2680
- 16 Brown N, Sandholm T, Amos B. Depth-limited solving for imperfect-information games. In: Proceedings of Neural Information Processing Systems, 2018
- 17 Zeiler M D, Fergus R. Visualizing and understanding convolutional networks. In: Proceedings of European Conference on Computer Vision, 2014
- 18 Zhou P, Feng J. Empirical risk landscape analysis for understanding deep neural networks. In: Proceedings of International Conference on Learning Representations, 2018
- 19 Simon H A. Administrative Behavior - A Study of Decision-Making processes in Administrative Organizations. New York: The Macmillan Company, 1947 [Simon H A, 著. 杨砾, 译. 管理行为 - 管理组织决策过程的研究. 北京: 北京经济学院出版社, 1988]
- 20 Zeng P, Wu L D, Wei Y M. Analysis, description and design for a tactical plan recognition model. Comput Digi Eng, 2006, 34: 1-4 [曾鹏, 吴玲达, 魏迎梅. 战术计划识别模型的分析、描述与设计. 计算机与数字工程, 2006, 34: 1-4]
- 21 Shi C, Li Y T, Zhang J W, et al. A survey of heterogeneous information network analysis. IEEE Trans Knowl Data Eng, 2017, 29: 17-37
- 22 Zhu F, Hu X F, Wu L, et al. From situation cognition stepped into situation intelligent cognition. J Sys Simul, 2018, 30: 761-771 [朱丰, 胡晓峰, 吴琳, 等. 从态势认知走向态势智能认知. 系统仿真学报, 2018, 30: 761-771]

- 23 Wang D X, Cui P, Zhu W W. Structural deep network embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016. 1225–1234
- 24 Martinage R. Toward a Third Offset Strategy-Exploiting U.S. Long-Term Advantages to Restore U.S. Global Power Projection Capability. Jiangyin: Knowfar Institute for Strategic and Defence Studies, 2014 [Martinage R, 著. 郑伟, 译. 迈向新抵消战略: 利用美国的长期优势恢复美国全球力量投送能力. 江阴: 知远战略与防务研究所, 2014]
- 25 Wang Q, Mao Z D, Wang B, et al. Knowledge graph embedding: a survey of approaches and applications. IEEE Trans Knowl Data Eng, 2017, 29: 2724–2743
- 26 Racaniere S, Weber T, Reichert D P, et al. Imagination-augmented agents for deep reinforcement learning. In: Proceedings of Neural Information Processing Systems, 2017
- 27 Huang W Z, Zhang J G, Huang K Q. Bootstrap estimated uncertainty of the environment model for model-based reinforcement learning. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI-19), 2019. 33: 3870–3877
- 28 Hu X F, He X Y, Tao J Y. AlphaGo's breakthrough and challenges of wargaming. Sci Technol Rev 2017, 35: 49–60 [胡晓峰, 贺筱媛, 陶九阳. AlphaGo 的突破与兵棋推演的挑战. 科学导报, 2017, 35: 49–60]
- 29 Hu Y, Li J T, Li X, et al. Knowledge-guided agent-tactic-aware Learning for StarCraft Micromanagement. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018. 1471–1477
- 30 Lowe R, Wu Y, Tamar A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments. In: Proceedings of Advances in Neural Information Processing Systems, 2017. 6379–6390
- 31 Shen S Q, Yin Q Y, Zhang J G, et al. Application and research of assistant command and control decision making method based on deep situation verification. In: Proceedings of China Conference Command Control, 2018 [申生奇, 尹奇跃, 张俊格, 等. 基于深度态势匹配的辅助指挥控制决策方法应用研究. 见: 中国指挥控制大会, 2018]
- 32 Yin Q Y, Zhan D, Zhang J G, et al. Combat unit prediction under the fog of war. In: Proceedings of China Conference Command Control, 2017 [尹奇跃, 詹东, 张俊格, 等. 不完全信息下对抗单元预测与估计模型. 见: 中国指挥控制大会, 2017]
- 33 Han X, Yan H P, Zhang J G, et al. ACM: learning dynamic multi-agent cooperation via attentional communication model. In: Proceedings of International Conference on Artificial Neural Network, 2018. 219–229
- 34 Wu H K, Zhang J G, Huang K Q. MSC: a dataset for macro-management in StarCraft. 2017. ArXiv: 1710.03131
- 35 Peng P, Xing J, Cao L, et al. Learning deep decentralized policy network by collective rewards for real-time combat game. In: Proceedings of International Joint Conference on Artificial Intelligence, 2019. 1305–1311
- 36 Churchill D, Buro M. Build order optimization in StarCraft. In: Proceedings of AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 2011
- 37 García-Sánchez P, Tondoy A, Mora A M, et al. Towards automatic starcraft strategy generation using genetic programming. In: Proceedings of IEEE International Conference on Computer Games, 2015
- 38 Busoniu L, Babuska R, de Schutter B. A comprehensive survey of multiagent reinforcement learning. IEEE Trans Syst Man Cybern C, 2008, 38: 156–172
- 39 Peng P, Yuan Q, Wen Y, et al. Multiagent bidirectionally-coordinated nets for learning to play StarCraft combat games. 2017. ArXiv: 1703.10069
- 40 Sukhbaatar S, Szlam A, Fergus R. Learning multiagent communication with backpropagation. In: Proceedings of Advances in Neural Information Processing Systems, 2016. 2244–2252
- 41 Zhang K Q, Yang Z R, Liu H, et al. Fully decentralized multi-agent reinforcement learning with networked agents. In: Proceedings of International Conference on Machine Learning, 2018
- 42 Li Y X. Deep reinforcement learning: an overview. 2017. ArXiv: 1701.07274
- 43 Jaderberg M, Czarnecki W M, Dunning I, et al. Human-level performance in 3D multiplayer games with population-based reinforcement learning. Science, 2019, 364: 859–865
- 44 Qian R Y. Command Decisions and Strategies. Beijing: Kunlun Press, 1999 [钱汝益. 指挥决策与谋略. 北京: 昆仑出版社, 1999]

Intelligent technologies of human-computer gaming

Kaiqi HUANG*, Junliang XING, Junge ZHANG, Wancheng NI & Bo XU

Center for Research on Intelligent System and Engineering, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

* Corresponding author. E-mail: kqhuang@nlpr.ia.ac.cn

Abstract At the frontier of artificial intelligence research, human-computer gaming (HCG) technology has become a research hotspot. It provides an effective experimental environment and approach to exploring the intrinsic growth mechanism and verifying key technologies of machine intelligence. Considering the mounting challenges in intelligent decision-making posed by the complex, high-dynamic, and inconclusive environment coupled with strong confrontation, this paper analyzes the research status and dissects the key elements and intrinsic gaming mechanisms of HCG. This work also proposes a game learning-based theoretical research framework for HCG. Based on these analyses, we discussed HCG's key models: gaming representation and modeling, situation assessment and reasoning, strategy generation, and optimization, as well as action coordination and control. The proposed research framework has laid a foundation for the modeling, computing, and interpreting solutions of complex cognition and decision problems. Finally, this paper summarizes the current application status and looks to future directions of development.

Keywords artificial intelligence, human-computer gaming, machine learning, autonomous intelligent gaming, command and decision making

Kaiqi HUANG was born in 1977. In 2004, he received his Ph.D. degree in signal and information processing from Southeast University, Nanjing. He is currently a professor and a Ph.D. supervisor at the Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision, computer gaming, machine learning, and artificial intelligence.

Junliang XING was born in 1984. In 2012, he received a Ph.D. degree in computer science from Tsinghua University, Beijing. Currently, he is a professor at the Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision, computer gaming, machine learning, and artificial intelligence.

Junge ZHANG received a B.S. degree from China University of Geosciences in 2008. In addition, in 2013, he received a Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences. In July 2013, he joined the Center for Research on Intelligent Perception and Computing, and he now serves as a professor by special appointment. His research interests include computer vision, pattern recognition, deep learning, and general artificial intelligence.

Wancheng NI was born in 1978. In 2008, she received a Ph.D. degree in control science and engineering from Tsinghua University, Beijing. In 2010, she then completed her postdoctoral research work at the Institute of Automation, Chinese Academy of Sciences, Beijing, where she is currently an associate professor and a master's supervisor. Her research interests include data mining, knowledge discovery, knowledge graph technology, computer gaming, and artificial intelligence.