



计算机科学

COMPUTER SCIENCE

即时策略博弈在线对抗规划方法综述

罗俊仁, 张万鹏, 陆丽娜, 陈璟

引用本文

罗俊仁, 张万鹏, 陆丽娜, 陈璟. 即时策略博弈在线对抗规划方法综述[J]. 计算机科学, 2022, 49(6): 287-296.

LUO Jun-ren, ZHANG Wan-peng, LU Li-na, CHEN Jing. [Survey on Online Adversarial Planning for Real-time Strategy Game](#)[J]. Computer Science, 2022, 49(6): 287-296.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[面向人机协同的物体姿态估计帧间稳定性优化方法](#)

Optimization Method for Inter-frame Stability of Object Pose Estimation for Human-Machine Collaboration
计算机科学, 2021, 48(11): 226-233. <https://doi.org/10.11896/jsjcx.201200095>

[双人博弈问题中的蒙特卡洛树搜索算法的改进](#)

Improvement of Monte Carlo Tree Search Algorithm in Two-person Game Problem
计算机科学, 2018, 45(1): 140-143. <https://doi.org/10.11896/j.issn.1002-137X.2018.01.023>

[基于预案模板的 HTN 规划知识建模方法及其应用](#)

Method of Modeling Knowledge for HTN Planning Based on Emergency Plan Template and its Application
计算机科学, 2010, 37(10): 202-206.

即时策略博弈在线对抗规划方法综述

罗俊仁 张万鹏 陆丽娜 陈璟

国防科技大学智能科学学院 长沙 410073

(luojunren17@nudt.edu.cn)

摘要 即时策略博弈在线规划是多智能体学习领域的挑战性问题,在博弈对抗过程中,面对不确定性威胁环境和非平稳性对手,智能体需要在有限时间内根据博弈对抗态势推理对方的行动,在巨大的状态空间和动作空间中快速做出己方行动规划,组织对抗规划。即时策略博弈平台是研究在线对抗规划问题的理想测试床。文中首先借助一个典型的即时策略博弈模型引出即时策略博弈对抗问题,并将其分类成3个层次和2种操作控制方法,从5个子方向梳理了即时策略博弈面临的5个挑战;其次从战术对抗、策略对抗和混合对抗3个角度对当前在线对抗规划方法进行了全面的综述分析;最后从对手及玩家建模、人机协同在线临机规划、基于学习的规划方法3个方面点明了未来需要研究的重点问题。

关键词: 即时策略博弈; 在线对抗规划; 蒙特卡洛树搜索; 层次任务网络; 对手建模; 人机协同

中图法分类号 TP181

Survey on Online Adversarial Planning for Real-time Strategy Game

LUO Jun-ren, ZHANG Wan-peng, LU Li-na and CHEN Jing

College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China

Abstract Real-time strategy game online adversarial planning is a challenging problem in the field of multi-agent learning. In the process of game confrontation, in the face of an uncertain threat environment and non-stationary opponents, the agent needs to reason about the opponent's actions within a limited time according to the game situation, make your own action plan quickly and perform adversarial planning in the huge state space and action space. The real-time strategy game platform is an ideal testbed for studying online adversarial planning problems. This paper firstly uses a typical real-time strategy game model to elicit the real-time strategy game confrontation problems, and classifies them into three levels and two operation control methods, and sorts out the five challenges faced from five sub-directions. Secondly, the current online adversarial planning methods are comprehensively reviewed and analyzed from three perspectives of tactical adversarial planning, strategic adversarial planning and mixed adversarial planning. Finally, the key issues that need to be studied in the future are pointed out from three key aspects: opponent and player modeling, human-machine collaborative online ad hoc planning, and learning-based planning.

Keywords RTS game, Online adversarial planning, Monte Carlo tree search, Hierarchical task networks, Opponent modeling, Human-machine collaboration

1 引言

机器博弈(Computer Game)是人工智能的果蝇^[1],为人工智能 AI 程序提供了丰富的测试环境。人工智能技术的发展,特别是针对多智能体学习方法的研究与机器博弈密切相关。近年来,随着经济社会的发展,即时策略(Real Time Strategy, RTS)博弈对抗类游戏,如星际争霸(StarCraft),吸引了大量学者的关注,成为人工智能研究的新热点,并取得了一系列的先进研究成果,特别是在2019年,星际争霸 AI^[2]在人机对抗比赛中战胜了人类顶级职业选手。2020年5月,美国国防部高级研究计划局(DARPA)通过其官网宣布了“打破

游戏规则”(Gamebreaker)的人工智能探索项目^[3],其中包括数款即时策略博弈平台(StarCraft II, SpringRTS, MiniRTS, DeepRTS, μ RTS),其希望找出一种通用的方法来评估和操纵博弈对抗均衡,以便于在国防部模拟战争推演和训练实战作战人员中运用人工智能算法。当前,通用人工智能技术的研究目标是构建与人类同水平的 AI 系统,使其能够充分整合人类的各種能力,如实时稳健响应、与环境自主智能交互、以自然语言交互、能进行常识推理、有创造力等。其中能否实时稳健响应就是要求智能体能够在短时间内根据自身所处环境进行自适应调整应对。

规划是指决定行动序列以完成给定目标的过程。根据

到稿日期:2021-06-21 返修日期:2021-08-13

基金项目:国家自然科学基金(61702528,61806212);湖南省研究生科研创新项目(CX20210011)

This work was supported by the National Natural Science Foundation of China(61702528,61806212) and Hunan Postgraduate Innovation Project(CX20210011).

通信作者:陈璟(chenjng001@vip.sina.com)

目标的不同,规划可以分为3类:对抗规划(adversarial planning),规划的目标是攻击和摧毁敌人,并阻止敌人实现其目标;中立规划(neutral planning),规划的目标并不影响其他规划的目标,或者影响可以被忽略;友方规划(friendly planning),规划的目标是提升环境状态或者帮助其他规划实现目标。即时策略博弈在线对抗规划(online adversarial planning)是多智能体学习领域的挑战性课题,在博弈对抗过程中,智能体面临不确定性(uncertain)^[4]威胁环境和非平稳性(non-stationary)^[5]对手,它需要在有限时间内根据博弈对抗态势推理对方的行动,在巨大的状态空间和动作空间中快速做出己方行动规划。在线对抗规划可划分为两个相互交融的子过程:对抗推理(adversarial reasoning)^[6]和应对规划(counter planning)^[7],主要表现为处于对抗环境中,智能体通过推理对手的心智状态、意图和行为,在尽量短的时间内制定己方应对计划对抗对手的动态规划过程。作为智能规划(intelligent planning)^[8]的前沿课题,聚焦对抗规划的计划行动意图识别(Plan Activity and Intention Recognition, PAIR)^[7,9-10]相关研究在人工智能领域受到了广泛关注。DARPA在2004年就开展了实时对抗情报和决策项目(Real-time Adversarial Intelligence and Decision-making, RAID)^[6,11],帮助美军筹建了CADET系统,其已经被集成在旅级地面作战规划系统里,其中重要的一环就是对抗推理和欺骗推理,重点关注如何将敌人在军事行动中的对抗性行为纳入己方规划考虑范畴。当前对抗规划的相关问题包括信念和意图识别、对手策略预测、计划识别、欺骗计划发现、欺骗计划和计划生成等。

本文首先剖析即时策略博弈问题,借助典型模型对智能博弈问题进行了分类,指出即时策略博弈面临的挑战;其次从战术对抗、策略对抗和混合对抗3个角度对当前对抗规划方法进行了综述分析;最后探讨了对抗规划的未来研究方向。

2 即时策略博弈对抗

即时策略博弈的状态空间和动作空间往往比象棋或围棋等棋类游戏高出很多数量级,如在仅具有10个被控对象,每个被控对象具有5个动作的情况下,潜在的分支因子数为 $5^{10} \approx 1000$ 万,从状态空间来看,国际象棋的状态空间约为 10^{47} 、围棋的状态空间约为 10^{171} ,而StarCraft的状态空间约为 10^{1685} ,这远远超出了很多标准算法能够处理的范围。当前即时策略博弈面临一些严峻的挑战:1)实时性,规划算法必须在很短的时间内给出每个当前状态下的最优行动以满足实时性需求,否则决策所依据的信息可能会过时;2)不确定性,即并发行动带来的不确定性,因为对方在当前时刻所采取的行动是难以预知的,所以行动后果带有不确定性,如射击行为可能因为对方移动到了其他位置而无效,在环境部分可观的情况下,战场信息本身是不确定的,如对方实体的数量、位置等并不能及时获取到;3)动作持续性,即每个动作需要一段时间才能完成,导致构建博弈树变得相当困难。即时策略博弈平台是研究在线对抗规划问题的理想测试床。即时策略博弈在线对抗规划主要研究博弈对抗环境中多智能体在线推理与规划方法。即时策略博弈对抗过程中,玩家需要控制不同类型的军事单元。一般博弈的初始状态是从少量的

工人开始,这些工人能够收集资源和建造建筑物。AI研究者的主要研究目标是如何指挥一方兵力摧毁对手控制的所有军事单元。

2.1 典型即时策略博弈对抗模型

μ RTS是一个典型的多智能体即时策略博弈对抗研究平台^[12]。在 μ RTS中,对抗双方同时控制多个异构单元,需要在给定时间内快速作出反应,双方同时行动并且所有动作都是持续性的。 μ RTS提供了两种博弈模式,即环境部分可观和完全可观。 μ RTS内设置了各种类型的地图,规模从 2×2 到 256×256 不等,不同的地图有不同的作战配置。图1给出了在 10×10 规模地图中 μ RTS的运行截图。

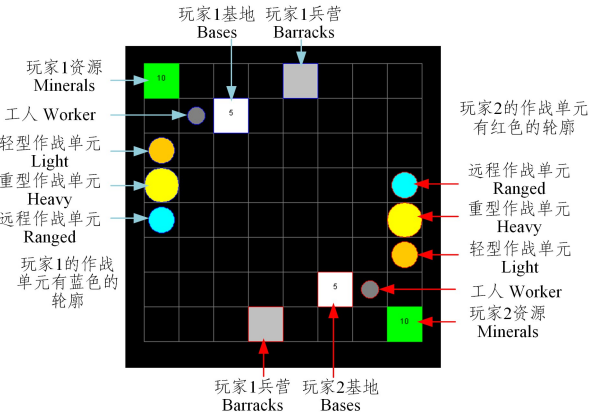


图1 μ RTS运行示意图(电子版为彩图)

Fig. 1 Operation diagram of μ RTS

红蓝对抗双方可描述为max和min,一般约定max为我方,min为敌方。每方都可以控制6种类型的军事单元,分别是工人(Worker)、轻型攻击者(Light Attacker)、重型攻击者(Heavy Attacker)、远程攻击者(Ranged Attacker)、兵营和基地。所有类型的单元都具有相同的状态属性(归属方、位置、生命值、是否可执行动作),但是具有不同的原子动作。基地主要生产工人,工人既可以收集资源,也可以建造兵营,还可以攻击敌人。轻型攻击者、重型攻击者和远程攻击者都属于作战单元,主要负责消灭敌方单元,但是它们具有不同的生命值、攻击力和攻击范围。兵营负责生产上述3种作战单元。所有的移动单元都可执行移动操作,所有单元都可执行等待操作,等待时长默认为10个决策周期。资源为双方共有,任意方都可以使用。

即时策略博弈模型可以采用一般元组来表示: $G=(P, S, A, L_u, L_p, T, W, s_{init})$ 。

(1) $P=\{\max, \min\}$ 表示博弈玩家;

(2) S 表示可能的状态。 $units(p, s)$ 表示在状态 S 属于玩家 P 的作战单元;

(3) A 表示作战单元可以执行的有限的单元动作集;

(4) $L_u(u, a, s) \rightarrow \{\text{true}, \text{false}\}$ 表示在状态 S 执行单元动作 a 是否可以返回单元 u ;

(5) $L_p(p, a, s) \rightarrow \{\text{true}, \text{false}\}$ 表示玩家 P 在状态 S 是否可以执行动作 a ;

(6) $T(s_t, \alpha_{\min}, \alpha_{\max}) \rightarrow S$ 是确定性转移函数,表示在时间 t 给定状态 $s_t \in S$ 和每个玩家的动作(α_{\min} 和 α_{\max}),返回在时间 $t+1$ 将到达的状态;

(7) $W: S \rightarrow \{\max \text{ wins}, \min \text{ wins}, \text{draw}, \text{ongoing}\}$ 决定博弈

的胜利方,结果可以是博弈正在进行、某方胜利或者平局;

(8) $s_{\min} \in S$ 表示有限状态。

2.2 即时策略博弈对抗问题分类

即时策略博弈对抗问题一般可划分成 3 个层次,对应 2 类操作控制任务。如图 2 所示,策略层主要负责收集、建造和发展等战略元素,其包含基础建设、技术树管理(构建顺序)和资源管理等与经济相关的任务、兵力组成和高层次的进攻时机;中间层为战术层,主要涉及不同单位之间的协调配合;最底层为反应层,是关于单个兵力单位的反应控制。微观(micro)操作控制是指单独控制单个智能体的能力,大致与反应层和部分战术层相对应,一个良好的微操玩家可以保证其兵力存活很长时间;宏观(macro)操作控制主要包括兵力生成和再造等,与图中策略层及部分战术层相对应,一个良好的宏观操作控制玩家通常会拥有大量的兵力单位。

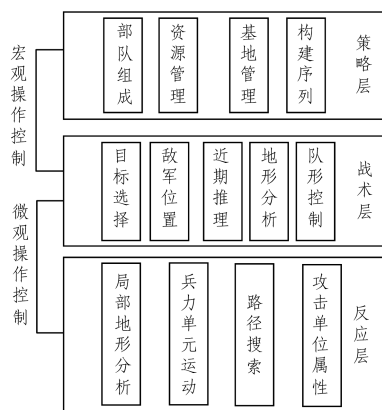


图 2 即时策略博弈问题分类

Fig. 2 Problem classification of real-time strategy game

博弈对抗过程中,高阶智能体必须小心地控制所有单位来最大化作战效能。控制多个兵力单位机动到优势位置来包围敌人,控制单个个体来攻击敌方弱势单位个体或者逃避攻击。这种对个体单位的短期战术控制通常被称为微观操作控制,即“战术规划”。此外,玩家需要依托多种兵力单位发动作战;同时还包括单位和建筑的生产来完成资源收集、单位的生产 and 升级,以保证有充足的单位来确保自己在现阶段的攻击或防守能够取得成功和在随后的阶段有足够的资源和升级优势来获取胜利;而且还必须决定整个博弈对抗过程中生产哪些单位和建筑,升级哪些科技以保证在需要的时间有合适的单位可以利用。这种中长期的高层规划,对应宏观操作控制,即“策略规划”。

2.3 即时策略博弈对抗面临的挑战

当前主流研究方法是将博弈分解为若干个具有不同抽象层次的较小的子问题,并分别处理每个子问题^[13],主要是通过将整个问题分解成可解决的部分,并专注于这些部分,通过集成每个部分来构建 AI 智能体。当前智能体即时策略博弈面临的主要挑战如下:

(1) 对抗规划:在对抗的环境下,面对无平稳策略的对手,如何实时在线生成确保获得胜利的最佳行动序列。

(2) 非完美信息:关于博弈状态的不完美信息会对智能体的行为产生负面影响。不完美源于缺乏关于环境和/或对手行为的足够信息,为了有效地对抗规划,智能体

必须将不确定性最小化。

(3) 空间与时间推理:空间推理主要分析在哪里放置单位和扩展基地会对智能体的策略结果有影响。空间推理通常是从地形分析模块中推断出来的,并用于优化布局和战术小组推理。时间推理主要分析“什么时候”和“在哪里”执行一个动作,尤其是在有持续性动作的情况下。有些动作的效果要到博弈后期才会显现。

(4) 知识迁移与利用:智能体可以从有用的数据中提取知识和经验,以辅助其智能行为决策。这些知识可用专家日志、策略指南和编码在脚本中的专家策略来表示。

(5) 策略有效学习:离线时,智能体可以有效利用博弈对抗历史经验重放数据来学习博弈对抗策略。在线时,智能体可以基于离线策略,结合在线交互数据进行自适应的在线调整学习。

3 在线对抗规划方法

规划领域的主流研究集中在单个 Agent 的中立规划任务,并不考虑对抗性的存在。但实际中的很多领域都有对抗性的存在,对手会积极主动地阻断我方完成目标,同时试图完成自身的目标,而这些目标往往是针锋相对和冲突的。除了对手之外,实时条件和隐藏信息约束的存在也让规划过程变得更加复杂。在线对抗规划同时包含这两方面的内容,在军事仿真领域和机器博弈领域都有着广泛的应用。在棋牌类博弈中,代表性的就是 Alpha-Beta 博弈树搜索,尤其是在国际象棋比赛中,AI 已经可以与人类顶级棋手进行对抗,甚至表现得更好。AI 智能体的目标是在每次博弈循环中基于当前博弈态势生成动作,其引导智能体战胜任意对手。换句话说,智能体必须在很大的状态和动作空间中,实时地找出获胜的方案计划。对抗规划方法主要有三种形式:基于脚本的对抗规划、基于搜索的对抗规划和基于学习的对抗规划^[12]。其中基于脚本的方法^[14]也称反应式规划方法,主要通过利用固化的规则脚本匹配查询,虽然速度快,但无法应对新情况;基于搜索的方法^[15-16]也称自主式规划方法,主要利用智能规划技术完成对最优动作的探索;基于学习的方法主要通过深度学习^[17]、强化学习^[18]等方法来学习最优动作,扩展性强,但需要大量博弈对抗样本。当前面向即时策略博弈的在线对抗规划方法主要可分为三大类:战术对抗规划、策略对抗规划和混合对抗规划。

3.1 战术对抗规划

3.1.1 极大极小树搜索

极大极小树搜索算法^[15]将敌我双方的行为都以树状结构进行表示,枚举每一种可能的情况,比较每个动作可能导致对手采取的行动效果,并搜索至终止情况的残局状态进行评估,得到当前应当采取的行动。由于此种方法构建的博弈树比较大,从根节点搜索到叶子节点,时间上无法接受,因此大多数博弈树搜索方法采用深度受限搜索方法,借用估值函数估算辅助决策。在极大极小博弈树搜索过程中,理性假设各个玩家都会选择可以达到对己方有利的博弈态势,max 玩家总是希望博弈态势估值较大,min 玩家希望博弈态势变成对 max 玩家来说估值较小的态势,其本质就是对博弈树做有限深度的展开,从而寻找最优策略路径,详细过程如图 3 所示。

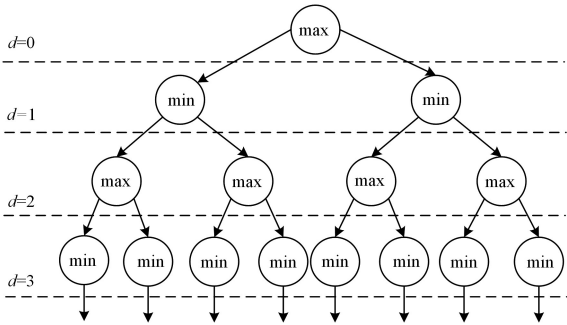


图3 极大极小树搜索

Fig. 3 Max-min tree search

极大极小树搜索的基本原理就是对博弈树做一定层数的展开,寻找最优策略路径,并走出路径中的第一步策略。Churchill等基于Alpha-Beta算法,提出了ABCD(Alpha-Beta Consider Duration)算法^[19]和UCTCD算法^[20]。其中ABCD算法解决了RTS博弈对抗过程中存在的动作同时执行以及动作持续执行两个问题。

3.1.2 蒙特卡洛树搜索

蒙特卡洛树搜索(Monte Carlo Tree Search, MCTS)本质上是一种基于最佳优先搜索和蒙特卡洛估值的前向树搜索算法。其通过随机采样的方式来获得相关数据,构建博弈树模型以用于搜索,基于统计结果得出最终决策。针对RTS中环境动态变化的问题,MCTS以实时搜索的方式产生在线决策^[21]。基本的MCTS算法首先在预先设置的计算资源内(一般指迭代次数或者运行时间)迭代地构建一棵非对称的搜索树,然后停止搜索并返回根节点上的最优动作。搜索树中的每个节点都表示状态,指向子节点的定向链接表示导致后续状态的动作。图4给出了MCTS执行过程的主要步骤。

重复 n 次

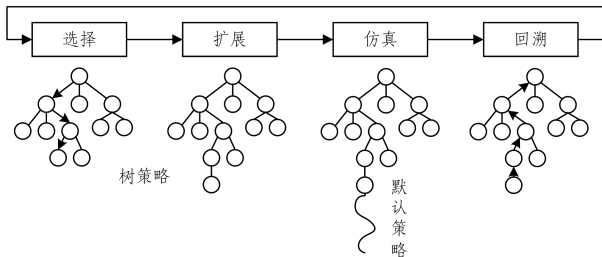


图4 蒙特卡洛树搜索的主要步骤

Fig. 4 Main steps of MCTS

对于每次算法迭代,MCTS主要执行4个步骤:1)选择(Selection),从根节点开始,递归运用子节点选择策略(不同的实现方法有不同的选择规则),直到到达最紧急的可扩展节点。如果一个节点表示非终止状态并且具有未访问(即未扩展)的子节点,则该节点是可扩展的。2)扩展(Expansion),根据当前节点可用动作集,添加一个(或多个)子节点扩展搜索树。3)仿真(Simulation),从新节点开始,运行默认的仿真策略,直到结束,产生仿真结果,最简单的仿真是随机策略。4)回溯(Backpropagation),通过反向传播更新路径上每个节点的统计信息,用以将来的策略决策。MCTS主要包括两种不同的策略:1)树策略(Tree policy),从搜索树中已经包含的节点中选择或创建叶节点,也就是节点选择和扩展策略。2)默认策略(Default policy),从一个给定的非终止状态开始

预演生成值估计,也就是仿真策略。MCTS是否成功主要取决于树策略^[22],在MCTS相关研究中,Chung等^[23]提出了一个用于战术规划的蒙特卡洛规划框架,并在基于ORTS框架^[24]的即时策略博弈中进行了测试。Churchill^[25]将MCTS应用于战斗中单个单元的底层控制。Balla等^[26]应用MCTS算法在即时策略博弈Wargus中进行战术攻击决策。类似地,Soe-mers^[27]在星际争霸中使用UCT进行战术规划。Uriarte等^[28]在星际争霸中使用MCTS进行军事演练。到目前为止,已经有大量研究工作关注于提高多智能体对抗环境中MCTS的性能。

Ontañón^[12]提出了利用NaiveMCTS来提高采样效率。其基于朴素采样假设,将从组合臂获得的奖励视为从每个底层赌博臂获得的个人奖励之和,其在 μ RTS中性能良好。当分支因子增加时,NaiveMCTS的实验结果显示出令人满意的性能,但仅在小场景中适用。Uriarte等^[28-31]使用抽象博弈状态来减少星际争霸中的分支因子,但是这种方法不能胜过脚本内置的AI。Shleyfman等^[32]指出,在计算时间接近结束时产生的分支没有足够的开发时间来产生良好的估计,这可能会误导搜索,因此他们建议在候选项生成阶段和候选项评估阶段之间分配计算资源,两阶段的采样算法在候选项生成阶段收集了关于每个值的辅助信息,该辅助信息是对每个值的回报贡献的估计,假设是线性的,接下来,使用这些信息基于熵生成组合臂,剩余的时间用于寻找最佳的组合臂。Ontañón^[33]提出了一种由信息引导的MCTS,它利用先验信息来改善MCTS的性能。其试验了两个不同的模型,校准的朴素贝叶斯(CNB)和行动型相互依赖模型(AIM),后者由于更好地表示了合法和非法行动之间的依赖关系而获得了更好的性能。但是这种方法需要使用大量专业博弈对抗数据进行预训练。Sironic等^[34]通过自适应调整MCTS参数来提高智能体在游戏中的性能。

3.1.3 组合多臂赌博机

UCT(Upper Confidence Bound for Trees)采用UCB1^[35]作为树策略,用以平衡探索和利用,是目前最著名的MCTS实现方法之一。在UCT算法中,它将每个子节点的选择看成多臂赌博机问题(multi-armed bandit problem)^[36],子节点的值是蒙特卡洛仿真得到的近似期望回报,因此这些回报是分布未知的随机变量。如果多个子节点具有相同的最大值,则通常进行随机选择^[37]。针对分支因子变得非常大时,UCT性能下降的问题,Ontañón^[38]提出的解决方案是将选择阶段建模成组合多臂赌博机(CMAB)问题,该算法比较适合即时策略博弈的组合决策空间,在CMAB中,智能体必须在每个周期中激活一个赌博臂,以期望组合臂获得最大的预期回报。

由于即时策略博弈环境的巨大复杂度,在计算时间有限的情况下,巨大的状态空间和动作空间使得目前大多数MCTS方法都只在小规模场景中适用。这类聚焦战术层行动的对抗规划方法没有全局视野,需要更高层级的策略来引导。

3.2 策略对抗规划

对于策略对抗规划方法的研究,一些工作引入了抽象机制^[39],包括动作抽象和状态抽象,将原始动作替换为脚本动作,如Justesen等^[40]提出的Script-UCT方法大大降低了搜索空间。

3.2.1 组合贪婪搜索

Churchill等^[20]提出了基于爬山算法的组合贪婪搜索

(Portfolio Greedy Search, PGS)方法,其不是为了优化每个单元选择采取哪种行动,而是为每个单元选择最优的脚本来控制单元动作。与其他搜索算法相比,该算法在大中型单元战斗中的性能明显优于纯脚本玩家,并且显示出巨大

的控制效率。如图 5 所示,初始脚本动作默认为 NOK-AV,统一迭代阶段固定一方不变来优化另一方,采用 Playout 方法进行效果评估,直到决策时间结束,输出局部最优解。

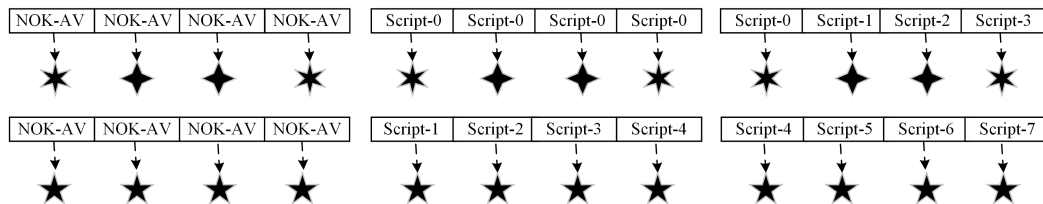


图 5 组合贪婪搜索方法

Fig. 5 Portfolio greedy search method

Moraes 等^[41]发现了 PGS 中的不一致问题,由于 PGS 在快速推理中依赖脚本化的对手模型,其可能无法返回可能的最佳动作,因此他们提出了一种嵌套贪婪搜索(Nest Greedy Search, NGS)方法,其工作方式与 PGS 相同,不同之处在于 NGS 处理对手建模的方式。NGS 贪婪地寻找对手对每个单位/动作组合的最佳反应,并计算针对它的最佳动作。

3.2.2 组合在线演化

Wang 等^[42]提出了基于演化迭代的策略选择组合在线演化(Portfolio Online Evolution, POE)方法,实验结果表明其效果优于 PGS。如图 6 所示,其主要流程包括设置初始基因组、在变异阶段对基因组进行随机变异操作、在交叉阶段对基因组进行随机两两交换、在选择阶段采用 Playout 方法筛选效果好的基因组至决策时间结束。

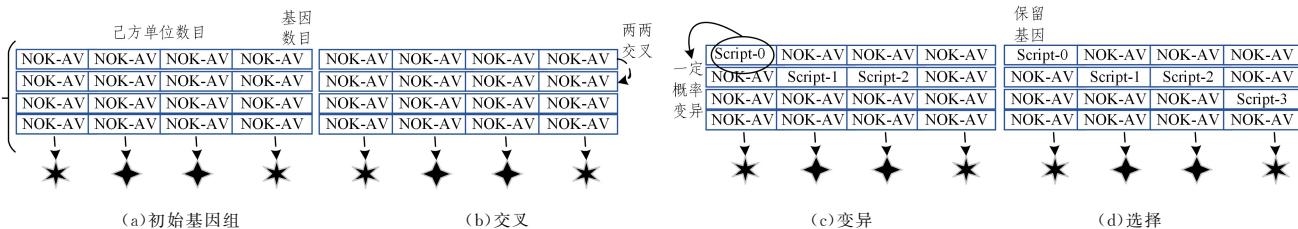


图 6 组合在线演化方法

Fig. 6 Portfolio online evolution method

3.2.3 分层策略选择

Lelis^[43]提出了分层策略选择(Stratified Strategy Selection, SSS)方法,该方法使用类型系统将玩家的单位划分为不同类型,并假定相同类型的单位必须遵循相同的策略。将一个脚本分配给一组被判断为属于同一类型的单元。一个类型系统决定如何使用属性来划分单位,如当前的生命值、攻击范围和武器伤害,策略数量越多,策略越精细,数量越少,策略越

粗糙。SSS+是一个变体,使用自适应类型系统和元推理方法来平衡搜索之间的策略粒度。Lelis^[43]在 SparCraft 的不同战斗场景中测试了这两种 SSS 方法,结果显示其相对于 PGS 和 POE 优势明显。如图 7 所示,SSS 算法采用类型系统,SSS+采用自适应类型系统,初始设定均默认为 NOK-AV,两种算法均依据类型进行种群划分,然后进行逐群体迭代交替至决策时间结束。

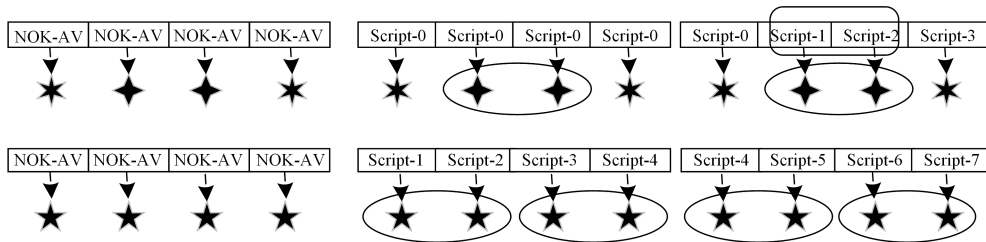


图 7 分层策略选择方法

Fig. 7 Stratified strategy selection method

3.2.4 非对称抽象采样

由于分支数非常高,基于 MCTS 的规划方法不能很好地扩展到更大的即时策略博弈场景,因此 Moraes 等^[44]提出利用行动抽象方法来提升 NaiveMCTS 的扩展性,并设计了两种方法。在第一种方法 A1N 中,其从一组脚本中进行朴素采样,而不是从一组低层级动作导出的抽象动作子集中采样。在第二种方法 A2N 中,单位动作是从两组不同的脚本中采样的。这两种方法均依赖动作抽象,其主要缺点是硬编码的

策略将智能体的行为限制在脚本动作的集合中,对对手策略的动态适应探索不够,这些都会使智能体容易被对手利用。

3.2.5 投票式策略创建

Silva 等^[45]试图通过投票式策略创建(Strategy Creation via Voting, SCV)方法解决以上两个问题,该方法包括一个新颖的策略生成机制,该机制建立在一组硬编码脚本的基础上,通过投票机制生成新的策略。此外,SCV 实施了一种策略适应技术,使用从不同策略之间的匹配数据集训练的逻辑回归

模型,在每一个固定数量的决策周期后,它会检测对手采用的策略,并相应地调整自己的策略。为了找到一个最优的抽象以显著提高智能体的策略水平,Marino 等^[46]使用进化算法(Evolutionary Algorithm, EA)来进化一组由脚本诱导的抽象,并找到最佳抽象,与 SCV 类似,EA 通过改变每个策略的参数,从一个较小的集合中产生一个较大的策略库。

3.3 混合对抗规划

策略对抗规划通常会牺牲短期的战术表现来获得更好的长期策略决策,而战术对抗规划方法则相反。近年来,国内外学者还提出了很多混合对抗规划方法,主要是通过将两大类方法集成到一种方法中来执行两种类型的规划^[47]。这种混合方式很好地模仿了人类进行策略和战术规划的方式。

3.3.1 规划知识引导

一些基于领域知识的策略规划方法采用由专家提供的各种编码知识。分层任务网络(Hierarchical Task Networks, HTN)^[48-49]是即时策略博弈中最常用的领域可配置规划器。HTN 是一种分层结构,由一个目标节点和多个子节点组成,这些子节点代表原始任务或非原始任务,智能体可以立即执行基本任务,而非基本任务必须使用方法分解成基本任务。一个方法必须满足其前提条件才能分解一个非基本任务。当所有树叶都是原始任务时,HTN 就完全分解了,其中领域定义决定了任务、方法和排序约束。

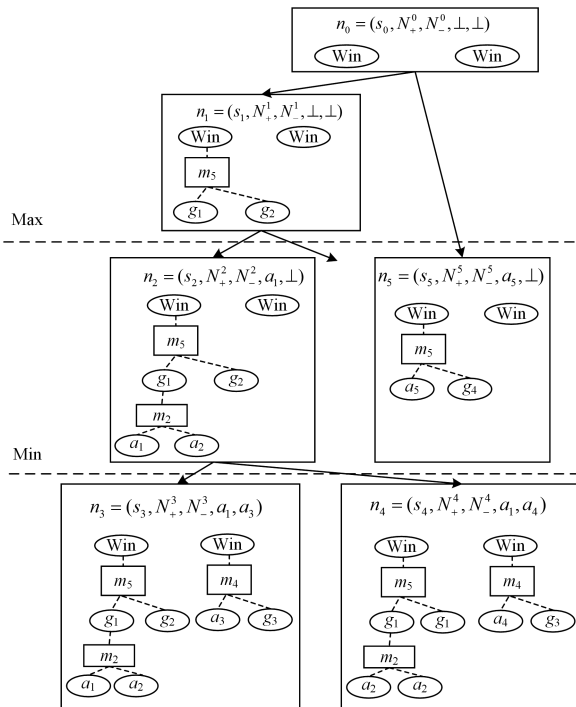


图 8 对抗分层任务网络

Fig. 8 Adversarial hierarchical task network algorithm

Ontañón 等^[49]首次提出了结合 HTN 规划和博弈树搜索的对抗性分层任务网络(Adversarial HTN, AHTN)规划算法,该方法为博弈对抗方法构造了一个 HTN,采用极大极小搜索方法,在每次迭代中,相应玩家的 HTN 被扩展,并且原始动作被执行,AHTN 适用于持续性和同时性动作,且性能优于 μ RTS 中的内置智能体。如图 8 所示,AHTN 算法采用 Alpha-Beta 剪枝方法为每个分支加入一个评估,并利用递归函数对未来行动进行预估。

由于 AHTN 没有考虑任务失败的概率,忽略了任务之间的关系,因此 Sun 等^[50]在 AHTN-R 中解决了这些问题,提出了一种任务修复机制,可以检测失败的任务并进行重规划,同时从预定义的列表选择一个替代任务。Sun 等^[51]在 AHTNCO 中采用对手策略来模拟对手行为生成博弈树的 min 节点。

Neufeld 等^[52]提出了一种基于 HTN 与 MCTS 的混合规划方法。其利用 HTN 进行策略规划,利用 MCTS 进行战术规划。每个 HTN 基元任务都用一个评估函数来表示,该函数被传递给 NaiveMCTS。每个评估函数都旨在引导 NaiveMCTS 生成特定任务,其中函数的权重是手动调整的。实验结果表明,相较于 AHTN 和 NaiveMCTS,这种混合规划方法的性能有了很大的提高。

3.3.2 行动语法识别

组合分类语法(Combinatory Categorical Grammars, CCG)是一种高效的、语言表达的语法形式主义,在自然语言处理中用于生成和识别自然语言句子。CCG 把特定语言的结构(词汇)从语法中分离出来。由于其具有识别性和生成性,因此非常适合识别和生成计划。Geib 等^[53-54]利用 CCG 作为计划识别方法。计划识别是从一系列观察到的行动中确定智能体目标的过程。Kantharaju 等^[55]提出了 LexLearn 算法,通过给一系列抽象动作分配类型标签的方式,可以学习关于动作的词典。为了识别长时间段计划,Geib 等使用一种新的学习算法 LexGreedy 来泛化 CCG,该算法通过贪婪的方法缩减了抽象的数量。计划识别过程中,由于 CCG 采用了广度优先搜索(BFS)方法,对于更大的场景,词汇贪婪学习词典将仍然包含每个动作类型的大量类别,这将影响 BFS 的识别性能,因此 Kantharaju 等^[56]提出使用 MCTS 方法进行大规模计划识别。针对即时策略博弈,Kantharaju 等^[57]提出了对抗组合分类语法(Adversarial CCG)规划算法。

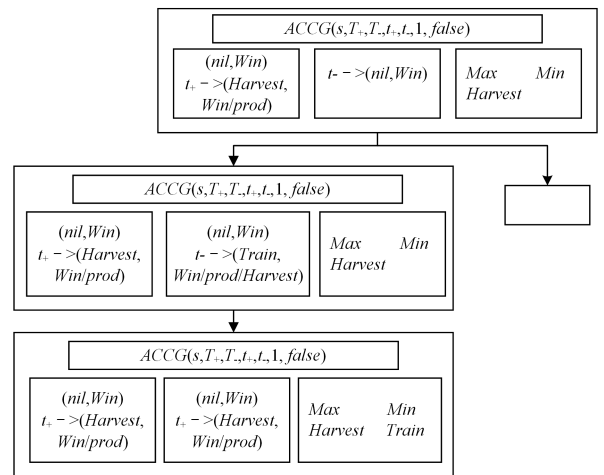


图 9 对抗组合分类语法

Fig. 9 Adversarial combinatory categorical grammars algorithm

如图 9 所示,在规划领域中使用 CCG 从一组即时策略博弈对抗回放日志中识别计划,根据识别的对手计划,基于极大极小算法采用类似 AHTN 的方法生成己方计划^[58]。

3.3.3 木偶采样搜索

Barriga 等^[59]提出了木偶搜索(puppet search)方法,这是一种新型对抗搜索框架,其依赖于在预定义的非确定性脚本

中插入许多公开的选择点,该脚本可以是硬编码的、基于机器学习的、基于搜索的或基于规则的智能体。它通过将选择点暴露给前瞻搜索过程的脚本来减少搜索空间,且搜索仅在暴露的选择点集合上执行,这显著降低了分支因子。为了提高“木偶搜索”的战术性能,Barriga 等提出策略战术 (Strategy Tactics,ST)算法^[59],试图将其与 NaiveMCTS 结合起来,利用卷积神经网络模型预测木偶搜索的输出,并生成策略,然后利用 NaiveMCTS 为备选策略生成战术规划。Moraes 等提出的非对称抽象方法 A3N^[60],可以看作一种混合规划方法,而 NaiveMCTS 被修改为在两个抽象层次中搜索。其中战术规划采用非受限搜索,策略规划采用受限搜索。Yang 等^[61]提出首选 (first choice) 和混合脚本 (mixed scripts) 的引导朴素采样 (Guided Naive Sampling, GNS) 方法。

3.3.4 学习引导搜索

Yang 等^[61-63]提出一种决策树集成的 MCTS 方法。首先利用历史数据学习基于决策树的策略,然后上层采用 Bagging 和随机森林等集成学习的方法引导下层搜索。训练后的模型可通过朴素采样预测每个单位的合法行为概率,其工作方式与 AlphaGo 非常相似,其采用的朴素采样可很好地应对即时策略博弈的多分支因子。Tavares 等^[64]提出了基于抽象的强化学习方法。Lu 等^[65]提出了基于 MAXQ 引导的分层蒙特卡洛在线规划方法,通过在 MCTS 算法中引入 MAXQ 分层任务结构,减少搜索空间并指导搜索过程,从而显著降低计算开销,而计算开销的降低使 MCTS 能够在有限的时间内执行更深入的搜索,以找到更优的规划策略。

面对指数级行动空间,此前 Lelis^[66]提出使用一个整体通用框架 (GEX),将即时策略在线对抗规划方法分为均匀行动抽象、常规行动采样和不对称行动抽象。综合来看,可将各类在线对抗规划方法作如下区分,如表 1 所列。

表 1 在线对抗规划方法

Table 1 Online adversarial planning method

类别	分类	方法
战术对抗规划	极大极小树搜索	Max-min ^[15] , ABCD ^[19] , UCTCD ^[20]
	蒙特卡洛树搜索	MCTS ^[22] , NaiveMCTS ^[12,29] , HMCTS ^[33]
	组合多臂赌博机	CMAB ^[37]
策略对抗规划	组合贪婪搜索	PGS ^[20] , Script-UCT ^[40] , NGS ^[41]
	组合在线演化	POE ^[42]
	分层策略选择	SSS ^[43] , SSS+ ^[43]
	非对称抽象采样	A1N, A2N ^[44]
混合对抗规划	投票式策略创建	SCV ^[45]
	规划知识引导	AHTN ^[48-49] , AHTN-R ^[50] , AHTNCO ^[51] , HTN-MCTS ^[52]
	行动语法识别	CCG ^[53] , Adversarial CCG ^[54]
木偶采样搜索	木偶采样搜索	Puppet Search ^[59] , A3N ^[60] , GNS ^[61]
	学习引导搜索	Ensemble Tree ^[61-63] , MAXQ ^[65] , GEX ^[66]

4 未来研究方向

当前,即时策略博弈在线对抗规划方法研究已然取得了一些进展^[67]。围绕对抗推理和应对规划两个子过程,未来我们可以利用对手及玩家建模方法,分析对手的行动特点,推理对手可能采取的行动;借助人机协同方法,利用人在回路的方式进行在线临机规划;充分发挥数据驱动的学习范式的优势辅助在线对抗规划生成。

4.1 对手及玩家建模

对手建模 (Opponent Modeling, OM) 是智能体理解其他

智能体行为的主要方式,是一种典型的认知行为推理方法^[68]。对手建模的目标是学习对手模型,这个模型可以利用博弈对抗数据集进行离线学习^[69],或者在博弈对抗过程中进行在线学习^[70]。通过对手建模,我们可以利用对手模型训练出更高阶的智能体。常用的对手建模方法大致可以分为计划行动意图识别方法^[9-10]、行为分类与类型推理方法^[71-72]、策略重构方法^[73]、交互建模^[74]与推理方法^[75]等。其中在线对手建模^[76]和有限理性对手建模^[77]仍是大难题。玩家建模 (Player Modeling, PM) 主要是学习玩家多种不同的可能模型,如博弈风格^[78-79]、博弈行为^[79-80]、性格特征^[81]、博弈技能表现^[82]等。

4.2 人机协同在线临机规划

人机协同规划是人工智能实用化的必然要求,同时人工智能为人机协同提供了强大支撑。传统的规划按照不同的时机可以划分为“两个阶段、三种规划”,即博弈对抗前的预先规划、临战规划和博弈对抗时的临机规划。其中临机规划直面瞬息万变的博弈对抗态势,相关成果将直接影响战局发展,凭借简单的方法难以有效支撑高层决策,因而成为当前规划中亟需解决的瓶颈问题。采用人机协作的方式可以提升智能体的规划能力,打通人与智能体之间的知识循环。最新的一些研究探索了人与单智能体协同^[83-84]、人与多智能体组队^[85]博弈对抗问题。此外,一些学者研究如何利用自然语言指令引导博弈对抗^[86],这项研究或许能催生出更懂人类口头或书面指令的博弈智能体。

4.3 基于学习的规划方法

规划和学习是人工智能的两个著名范式和子领域。规划注重设计行动序列以实现指定的目标或一组目标,而学习则侧重于交互提升。虽然这两种范式在博弈对抗领域都取得了巨大成功,但它们都存在一些缺点。规划方法需要环境模型以及规划任务的准确描述来规划,这在更复杂的环境中可能不容易获得。对于学习方法,当处理更复杂的环境时,如果没有提供任何指导,智能体可能会花费很多时间探索次优行为。然而,这两种范式可以相互补充^[87],从某种意义上说,每个范式都可以用来弥补另一个范式的缺点。

尽管目前搜索方法的研究成果众多,但即时策略博弈战术层对抗规划的实时性要求极高,在战斗单位数量众多时,抽象机制的存在也仅能保证搜索有限的动作空间,导致搜索所得解的最优性仍不足。可以借鉴 Muzero 的算法思路,将学习到的模型融入在线搜索过程,由学习模型引导搜索过程^[61,63]。最新的一些研究采用分层、分布式并行技术和机器学习的方法提升 MCTS^[89],采用深度学习评估博弈对抗态势^[90],学习博弈态势表示^[91];采用监督学习、战术搜索与强化学习方法^[92]学习行动策略;采用多样性对手建模与强化学习的方法学习全局行动策略^[93]。

结束语 本文对即时策略博弈在线对抗规划方法进行了全面的综述分析。首先,利用一类典型的即时策略对抗平台构建了博弈对抗基础模型,从策略层、战术层和反应层 3 个层次,宏观操作控制和微观操作控制 2 大类方法的角度对即时策略博弈对抗相关问题进行了分类,从对抗规划、非完美信息、空间与时间推理、知识迁移与利用和策略有效学习 5 个方面分析了智能体即时策略博弈面临的挑战;其次重点围绕在线对抗规划方法,分析了 3 类战术对抗规划方法(极大极小树

搜索、蒙特卡洛树搜索、组合多臂赌博机), 5 类策略对抗规划方法(组合贪婪搜索、组合在线演化、分层策略选择、非对称抽象采样、投票式策略创建), 4 类混合对抗规划方法(规划知识引导、行动语法识别、木偶采样搜索、学习引导搜索); 最后, 围绕在线对抗规划的未来研究方向, 简要分析了对手及玩家建模、人机协同在线临机规划和基于学习的规划方法 3 个方面的研究现状及重点。

参 考 文 献

- [1] OMIDSHAFIEI S, TUYLS K, CZARNECKI W M, et al. Navigating the Landscape of Multiplayer Games[J]. *Nature Communications*, 2020, 11(1): 5603.
- [2] VINYAS O, BABUSCHKIN I, CZARNECKI W M, et al. Grandmaster level in StarCraft II using Multi-agent Reinforcement Learning[J]. *Nature*, 2019, 575(7782): 350-354.
- [3] DARPA. Gamebreaker AI Effort Gets Under Way, 2020[EB/OL]. (2020-05-13) [2021-05-30]. <https://www.darpa.mil/news-events/2020-05-13>.
- [4] KEITH A J. Operational Decision Making Under Uncertainty: Inferential, Sequential, and Adversarial Approaches [D]. Air Force Institute of Technology, 2020.
- [5] HERNANDEZ-LEAL P, ZHAN Y, TAYLOR M E, et al. An Exploration Strategy for Non-stationary Opponents[J]. *Autonomous Agents and Multi-Agent Systems*, 2017, 31(5): 971-1002.
- [6] ALEXANDER K. Adversarial Reasoning: Computational Approaches to Reading the Opponent's Mind [M]. Chapman & Hall/CRC Computer and Information Science Series, 2006.
- [7] GU W X, LIU Y. Adversarial Planning and Response [M]. The Science Publishing Company, 2016.
- [8] GHALLAB M, NAU D, TRAVERSO P. Automated Planning and Acting [M]. Cambridge University Press, 2016.
- [9] SUKTHANKAR G, GOLDMAN R P, GEIB C, et al. Plan, Activity, and Intent Recognition: Theory and Practice [M]. *Ai Magazine*, 2014.
- [10] MIRSKY R, KEREN S, GEIB C. Introduction to Symbolic Plan and Goal Recognition [M]. Morgan & Claypool publishers, 2021.
- [11] OWNBY M, KOTT A. Predicting Enemy's Actions Improves Commander Decision-Making [C]// CCRTS, 2016.
- [12] ONTAÑÓN S. The Combinatorial Multi-armed Bandit Problem and Its Application to Real-time Strategy Games [C]// Proceedings of the 9th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 2013: 58-64.
- [13] ONTAÑÓN S, SYNNAEVE G, URIARTE A, et al. A Survey of Real-time Strategy Game AI Research and Competition in Starcraft [J]. *IEEE Transactions on Computational Intelligence and AI in Games*, 2013, 5(4): 293-311.
- [14] BARRIGA N A, STANESCU M, BURO M. Game Tree Search Based on Non-Deterministic Action Scripts in Real-Time Strategy Games [J]. *IEEE Transactions on Computational Intelligence and AI in Games*, 2017, 10(1): 69-77.
- [15] ONTAÑÓN S. Experiments with Game Tree Search in Real-Time Strategy Games [EB/OL]. (2012-08-01) [2021-05-06]. <http://arxiv.org/pdf/1208.1940.pdf>.
- [16] STANESCU M, BARRIGA N A, BURO M. Hierarchical Adversarial Search Applied to Real-Time Strategy [C]// National Conference on Artificial Intelligence, 2014: 66-72.
- [17] ANDERSEN P A, GOODWIN M, GRANMO O C. Deep RTS: A Game Environment for Deep Reinforcement Learning in Real-time Strategy Games [C]// IEEE Conference on Computational Intelligence and Games (CIG), 2018: 1-8.
- [18] SETHY H, PATEL A, PADMANABHAN V. Real Time Strategy Games: A Reinforcement Learning Approach [J]. *Procedia Computer Science*, 2015, 54: 257-264.
- [19] CHURCHILL D, SAFFIDINE A, BURO M. Fast heuristic search for RTS game combat scenarios [C]// Proceedings of the Eighth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 2012: 8-14.
- [20] CHURCHILL D, BURO M. Portfolio greedy search and simulation for large-scale combat in Starcraft [C]// Proceedings of the IEEE Conference on Computational Intelligence in Games, 2013: 1-8.
- [21] ISHIHARA M, HARADA T, MIYAZAKI T, et al. Applying and improving monte-carlo tree search in a fighting game AI [C]// Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology, 2016: 1-6.
- [22] BROWNE C B, WHITEHOUSE D, POWLEY E, 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.
- [23] CHUNG M, BURO M, SCHAEFFER J. Monte-carlo Planning in RTS Games [C]// IEEE 2005 Symposium on Computational Intelligence and Games, 2005: 117-124.
- [24] BURO M. ORTS: A Hack-free RTS Game Environment [C]// International Conference on Computers and Games. Berlin: Springer, 2002: 280-291.
- [25] CHURCHILL D. Heuristic Search Techniques for Real-Time Strategy Games [D]. University of Alberta, 2016.
- [26] BALLA R K, FERN A. UCT for Tactical Assault Planning in Real-time Strategy Games [C]// International Joint Conference on Artificial Intelligence, 2009: 40-45.
- [27] SOEMERS D. Tactical Planning Using MCTS in the Game of StarCraft [D]. Department of Knowledge Engineering, Maastricht University, 2014.
- [28] URIARTE A, ONTAÑÓN S. Game-tree Search over High-level Game States in RTS Games [C]// Proceedings of the 10th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE, 2014: 73-79.
- [29] URIARTE A, ONTAÑÓN S. High-level Representations for Game-Tree Search in RTS Games [C]// Proceedings of the 10th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE, RTS Workshop, 2014.
- [30] URIARTE A, ONTAÑÓN S. Improving Monte Carlo Tree Search Policies in StarCraft via Probabilistic Models Learned from Replay Data [C]// Proceedings of the Tenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 2016: 100-106.
- [31] URIARTE A. Adversarial Search and Spatial Reasoning in Real Time Strategy Games [D]. University of Alberta, 2017.
- [32] SHLEYFMAN A, KOMENDA A, DOMSHLAK C. On Combinatorial Actions and CMABs with Linear Side Information [J]. *Frontiers in Artificial Intelligence & Applications*, 2014, 263: 825-830.
- [33] ONTAÑÓN S. Informed Monte Carlo Tree Search for Real-Time Strategy games [C]// IEEE Conference on Computational Intelligence and Games, 2016: 1-8.

- [34] SIRONIC F, LIU J, PEREZ-LIEBANA D, et al. Self-Adaptive MCTS for General Video Game Playing Self-Adaptive MCTS for General Video Game Playing[C]//International Conference on the Applications of Evolutionary Computation. Cham: Springer, 2018: 358-375.
- [35] AUER P, CESA-BIANCHI N, FISCHER P. Finite-Time Analysis of the Multiarmed Bandit Problem[J]. *Machine Learning*, 2002, 47(2/3): 235-256.
- [36] KOCSIS L, SZEPESVÁRI C. Bandit Based Monte-Carlo Planning[C]//European Conference on Machine Learning. Berlin: Springer, 2006: 282-293.
- [37] KOCSIS L, SZEPESVÁRI C, WILLEMSON J. Improved Monte-Carlo Search[R]. Univ. Tartu, Estonia, Tech. Rep, 2006, 1.
- [38] ONTAÑÓN S. Combinatorial Multi-armed Bandits for Real-Time Strategy Games[J]. *Journal of Artificial Intelligence Research*, 2017, 58: 665-702.
- [39] BARRIGA N A. Search, Abstractions and Learning in Real-Time Strategy Games[D]. Drexel University, 2017.
- [40] JUSTESEN N, TILLMAN B, TOGELIUS J, et al. Script and Cluster-based UCT for StarCraft[C]//Computational Intelligence and Games. IEEE, 2014: 1-8.
- [41] MORAES R O, LELIS L H S. Nested-Greedy Search for Adversarial Real-Time Games[C]//Proceedings of the 14th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. AIIDE, 2018.
- [42] WANG C, CHEN P, LI Y, et al. Portfolio Online Evolution in Starcraft [C]//Proceedings of the Twelfth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. 2016: 114-120.
- [43] LELIS L H S. Stratified Strategy Selection for Unit Control in Real-Time Strategy Games [C]//Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017: 3735-3741.
- [44] MORAES R O, MARINO J R H, LELIS L H S, et al. Action Abstractions for Combinatorial Multi-Armed Bandit Tree Search [C]//AAAI Publications. Fourteenth Artificial Intelligence and Interactive Digital Entertainment Conference, 2018.
- [45] SILVA C R, MORAES R O, LELIS L H S, et al. Strategy Generation for Multiunit Real-Time Games via Voting[C]//IEEE Transactions on Games. 2018: 426-435.
- [46] MARINO J R H, MORAES R O, TOLEDO C, et al. Evolving Action Abstractions for Real-Time Planning in Extensive-Form Games[C]//Proceedings of the Conference on Artificial Intelligence (AAAI). 2019.
- [47] BARRIGA N A, STANESCU M, BURO M. Combining Strategic Learning and Tactical Search in Real-Time Strategy Games [EB/OL]. (2017-09-11) [2021-05-06]. <http://arxiv.org/pdf/1709.03480.pdf>.
- [48] GOPALAKRISHNAN S. Learning Hierarchical Task Networks Using Semantic Word Embeddings [D]. Lehigh University, 2017.
- [49] ONTAÑÓN S, BURO M. Adversarial Hierarchical-Task Network Planning for Complex Real-Time Games[C]//Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015: 1652-1658.
- [50] SUN L, JIAO P, KAI X, et al. Modified Adversarial Hierarchical Task Network Planning in Real-Time Strategy Games[J]. *Applied Sciences*, 2017, 7(9): 872.
- [51] SUN L, ZHU A, LI B, et al. HTN Guided Adversarial Planning for RTS Games[C]//IEEE ICMA. IEEE, 2020.
- [52] NEUFELD X, MOSTAGHIM S, PEREZ-LIEBANA D. A Hybrid Planning and Execution Approach Through HTN and MCTS[J]. *The International Conference on Automated Planning and Scheduling Workshop on Integrating Planning, Acting, and Execution*, 2019, 1: 37-49.
- [53] GEIB C W, GOLDMAN R P. Recognizing Plans with Loops Represented in a Lexicalized Grammar[C]//Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI'11). 2011: 958-963.
- [54] GEIB C, KANTHARAJU P. Learning Combinatory Categorical Grammars for Plan Recognition[C]//Proc. of the AAAI Conference on Artificial Intelligence. 2018.
- [55] KANTHARAJU P, ONTAÑÓN S, GEIB C. Extracting CCGs for Plan Recognition in RTS Games[C]//Proceedings of the Workshop on Knowledge Extraction in Games. 2019.
- [56] KANTHARAJU P, ONTAÑÓN S, GEIB C. Scaling up CCG-Based Plan Recognition via Monte-Carlo Tree Search [C]//Proc. of the IEEE Conference on Games (CoG). 2019.
- [57] KANTHARAJU P, ONTAÑÓN S, GEIB C. μ CCG, a CCG-based Game-Playing Agent for μ RTS[C]//Proc. of IEEE Conference on Computational Intelligence in Games (CIG). 2018.
- [58] KANTHARAJU P. Learning Decomposition Models for Hierarchical Planning and Plan Recognition [D]. Drexel University, 2020.
- [59] BARRIGA N A, STANESCU M, BURO M. Puppet Search: Enhancing Scripted Behavior by Look-Ahead Search with Applications to Real-Time Strategy Games [C]//Proceedings of the 11th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. AIIDE, 2015.
- [60] MORAES R O, LELIS L. Asymmetric Action Abstractions for Multi-Unit Control in Adversarial Real-Time Games[C]//The Thirty-Second AAAI Conference on Artificial Intelligence AAAI-18. 2018: 876-883.
- [61] YANG Z, ONTAÑÓN S. Guiding Monte Carlo Tree Search by Scripts in Real-Time Strategy Games [C]//Proceedings of the Fifteenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-19). 2019.
- [62] YANG Z, ONTAÑÓN S. Extracting Policies from Replays to Improve MCTS in Real Time Strategy Games[C]//Proceedings of the AAAI-2019 Knowledge Extraction from Games Workshop. 2019.
- [63] YANG Z, ONTAÑÓN S. An Empirical Survey on Methods for Integrating Scripts into Adversarial Search for RTS Games[J]. *IEEE Transactions on Games*, 2021(99): 1-10.
- [64] TAVARES A R, CHAIMOWICZ L. Tabular Reinforcement Learning in Real-Time Strategy Games via Options[C]//2018 IEEE Conference on Computational Intelligence and Games (CIG). 2018.
- [65] LU L, ZHANG W, GU X, et al. HMCTS-OP: Hierarchical MCTS Based Online Planning in the Asymmetric Adversarial Environment[J]. *Symmetry*, 2020, 12(5): 719.
- [66] LELIS L H S. Planning Algorithms for Zero-Sum Games with Exponential Action Spaces: A Unifying Perspective[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20). 2020.
- [67] OUESSAI A, SALEM M, MORA A M. Online Adversarial Planning in μ RTS: A Survey[C]//2019 International Confe-

- rence on Theoretical and Applicative Aspects of Computer Science (ICTAACS). 2019.
- [68] ALBRECHT S V, STONE P. Autonomous Agents Modelling Other Agents: A Comprehensive Survey and Open Problems [J]. *Artificial Intelligence*, 2018, 258: 66-95.
- [69] WEBER B G, MATEAS M. A Data Mining Approach to Strategy Prediction[C]// *Proceedings of the IEEE Symposium on Computational Intelligence and Games*. Milano, Italy, 2009: 140-147.
- [70] YAMAMOTO K, MIZUNO S, CHU C Y, et al. Deduction of Fighting-Game Countermeasures Using the k-Nearest Neighbor Algorithm and a Game Simulator[C]// *Proceedings of the IEEE Conference on Computational Intelligence and Games*. Dortmund, Germany, 2014: 437-441.
- [71] SCHADD F, BAKKES S, SPRONCK P. Opponent Modeling in Real-Time Strategy Games[C]// *GAMEON*. 2007: 61-70.
- [72] OMER E, VOLKAN S. Learning Strategies for Opponent Modeling in Poker[C]// *Workshop of the Twenty-Seventh AAAI Conference on Artificial Intelligence*. 2013.
- [73] ZHANG Y, RADULESCU R, MANNION P. Opponent Modeling using Policy Reconstruction for Multi-Objective Normal Form Games[C]// *ALA*. 2020.
- [74] DOSHI P, ZENG Y, CHEN Q. Graphical models for interactive POMDPs: representations and solutions [J]. *Autonomous Agents and Multi-Agent Systems*, 2009, 18(3): 376.
- [75] RUSCH T, STEIXNER-KUMAR S, DOSHI P, et al. Theory of Mind and Decision Science: Towards a Typology of Tasks and Computational Models[J]. *Neuropsychologia*, 2020, 146 (211): 107488.
- [76] BARD N, JOHANSON M, BURCH N, et al. Online Implicit Agent Modelling [C] // *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems*. AAMAS, 2013.
- [77] DAVID M, JAKUB C, VILIAM L, et al. Complexity and Algorithms for Exploiting Quantal Opponents in Large Two-Player Games[C]// *AAAI*. 2021.
- [78] VALLS-VARGAS J, ONTAÑÓN S, ZHU J. Exploring Player Trace Segmentation for Dynamic Play Style Prediction[C] // *Proceedings of the 11th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*. Santa Cruz, CA, USA, 2015: 93-99.
- [79] DIEGO P L, CRISTINA G R, DOMINIK J. Generating Diverse and Competitive Play-Styles for Strategy Games [EB/OL]. (2021-04-17) [2021-05-06]. <http://arxiv.org/pdf/2104.08641.pdf>.
- [80] MAHLMANN T, DRACHEN A, TOGELIUS J, et al. Predicting Player Behavior in Tomb Raider: Underworld[C] // *Proceedings of the IEEE Conference on Computational Intelligence and Games*. Dublin, Ireland, 2010: 178-185.
- [81] CHEN Z, EL-NASR M S, CANOSSA A, et al. Modeling Individual Differences through Frequent Pattern Mining on Role-Playing Game Actions [C] // *Proceedings of the 11th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*. AAAI Technical Report WS-15-23, Santa Cruz, CA, USA, 2015: 2-7.
- [82] YANNAKAKIS G N, MARAGOUAKIS M, HALLAM J. Preference Learning for Cognitive Modeling: A Case Study on Entertainment Preferences[J]. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.*, 2009, 39: 1165-1175.
- [83] LOCKHART E, BURCH N. Human-Agent Cooperation in Bridge Bidding [EB/OL]. (2020-11-28) [2021-05-06]. <http://arxiv.org/pdf/2011.14124.pdf>.
- [84] IZUMIGAWA C, LUCERO C, NANS L, et al. Building Human-Autonomy Teaming Aids for Real-Time Strategy Games[C] // *International Conference on Human-Computer Interaction*. Cham: Springer, 2020.
- [85] NAVIDI N, CHABOT F, KURANDWAD S, et al. Human and Multi-agent Collaboration in a Human-MARL Teaming Framework [EB/OL]. (2020-06-01) [2021-06-06]. <http://arxiv.org/pdf/2006.07301.pdf>.
- [86] HU H, YARATS D, GONG Q, et al. Hierarchical Decision Making by Generating and Following Natural Language Instructions[C]// *NeurIPS*. 2019.
- [87] PARTALAS I, VRAKAS D, VLAHAVAS I. Reinforcement Learning and Automated Planning: A Survey[C]// *Artificial Intelligence for Advanced Problem Solving Techniques*. IGI Global: Hershey, PA, USA, 2008: 148-165.
- [88] SCHRITTWIESER J, ANTONOGLIOU I, HUBERT T, et al. Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model[J]. *Nature*, 2020, 588(7839): 604-609.
- [89] ŚWIECHOWSKI M, GODLEWSKI K, SAWICKI B, et al. Monte Carlo Tree Search: A Review on Recent Modifications and Applications [EB/OL]. (2021-03-08) [2021-06-06]. <http://arxiv.org/pdf/2103.04931.pdf>.
- [90] STANESCU M, BARRIGA N A, HESS A, et al. Evaluating Real-time Strategy Game States using Convolutional Neural Networks[C]// *IEEE Conference on Computational Intelligence and Games (CIG)*. 2016.
- [91] HUANG S, ONTAÑÓN S. Comparing Observation and Action Representations for Deep Reinforcement Learning in μ RTS [J]. *arXiv:1910.12134*, 2019.
- [92] BARRIGA N A, STANESCU M, BESOAIN F, et al. Improving RTS Game AI by Supervised Policy Learning, Tactical Search, and Deep Reinforcement Learning[J]. *IEEE Computational Intelligence Magazine*, 2019, 14(3): 11.
- [93] HUANG S, ONTAÑÓN S, BAMFORD C, et al. Gym- μ RTS Toward Affordable Full Game Real-time Strategy Games Research with Deep Reinforcement Learning [EB/OL]. (2021-05-21) [2021-06-26]. <http://arxiv.org/pdf/2105.13807.pdf>.



LUO Jun-ren, born in 1989, Ph.D candidate. His main research interests include agent modelling, adversarial team game, and multi-agent reinforcement learning.



CHEN Jing, born in 1972, Ph.D, professor. His main research interests include artificial intelligence, intelligent decision-making, mission planning.