

《指挥与控制学报》网络首发论文

题目: 任务级兵棋智能决策技术框架设计与关键问题分析
作者: 张驭龙, 范长俊, 冯旻赫, 张龙飞, 刘忠
收稿日期: 2021-05-07
网络首发日期: 2021-11-08
引用格式: 张驭龙, 范长俊, 冯旻赫, 张龙飞, 刘忠. 任务级兵棋智能决策技术框架设计与关键问题分析[J/OL]. 指挥与控制学报.
<https://kns.cnki.net/kcms/detail/14.1379.TP.20211108.0851.002.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

任务级兵棋智能决策技术框架设计与关键问题分析

张驭龙^{1,2} 范长俊¹ 冯旻赫¹ 张龙飞¹ 刘忠¹

摘要 在梳理总结当前兵棋智能决策方法研究现状的基础上,明确指出了开展任务级兵棋智能决策研究的八项难点,而后从博弈活动中智能体定位、信息流转与知识流转等问题出发,设计给出了任务级兵棋智能博弈框架、智能策略生成优化框架以及多智能体策略协同演进框架,进而分析了高维度下长时序的态势信息数据综合、异步策略分布式学习优化、基于语义模板的任务指令生成、多智能体训练营博弈调度等4项关键问题,并给出了基本解决方案,点明了开展任务级兵棋智能决策研究的技术要点与基本路线,为后续任务级兵棋智能体设计提供了有益参考。

关键词 兵棋; 博弈; 决策; 智能; 强化学习

引用格式 张驭龙, 范长俊, 冯旻赫, 张龙飞, 刘忠. 任务级兵棋智能决策技术框架设计与关键问题分析[J]. 指挥与控制学报

DOI

Technical Framework Design and Key Issues Analysis in Task-Level Wargame Intelligent Decision Making

ZHANG Yu-Long^{1,2} FAN Chang-Jun¹ FENG Yang-He¹ ZHANG Long-Fei¹
LIU Zhong¹

Abstract After a brief literature review of decision-making in wargame, we design three frameworks, which are the intelligent gaming framework, the intelligent strategy generating and optimizing framework, and the multi-agent intelligent strategy co-evolutionary framework of task-level wargame. The three frameworks illustrate relationship between human experts and agents, the information flow and knowledge flow in the process of intelligent gaming. Besides, four key issues of these framework are presented, which includes the high-dimension long duration situational data processing, the off-policy distributed learning, the semantic-template task instruction generating, the multi-agent training league match operating, and then basic methods of the issues are given. The framework and key issues basic methods may provide a useful reference for the future agent design in task-level wargame.

Key words wargame; gaming; decision-making; intelligent; reinforcement learning

Citation ZHANG Yu-Long, FAN Chang-Jun, FENG Yang-He, ZHANG Long-Fei, LIU Zhong. Technical framework design and key issues analysis in task-level wargame intelligent decision making[J]. Journal of Command and Control

以Deepmind团队成功研发AlphaGo击败人类
顶级围棋选手为标志,国际互联网科技巨头^[1-3]纷
纷将研究目标对准了难度更高的竞技类游戏,其

中AlphaStar^[1]首次在大中型即时战略(RTS)游戏
“星际争霸II”达到“宗师”水平,超过99.8%
人玩家,将基于强化学习的智能博弈研究带入了

收稿日期 2021-05-07

Manuscript received May 7, 2021

国家自然科学基金(71701205, 62073333)资助

Supported by National Natural Science Foundation of China (71701205, 62073333)

1. 国防科技大学系统工程学院 长沙 410073 2. 解放军 31002 部队 北京 100000

1. School of Systems Engineering, National University of Defense Technology, Changsha 410073, China 2. Unit 31002 of PLA, Beijing 100000, China

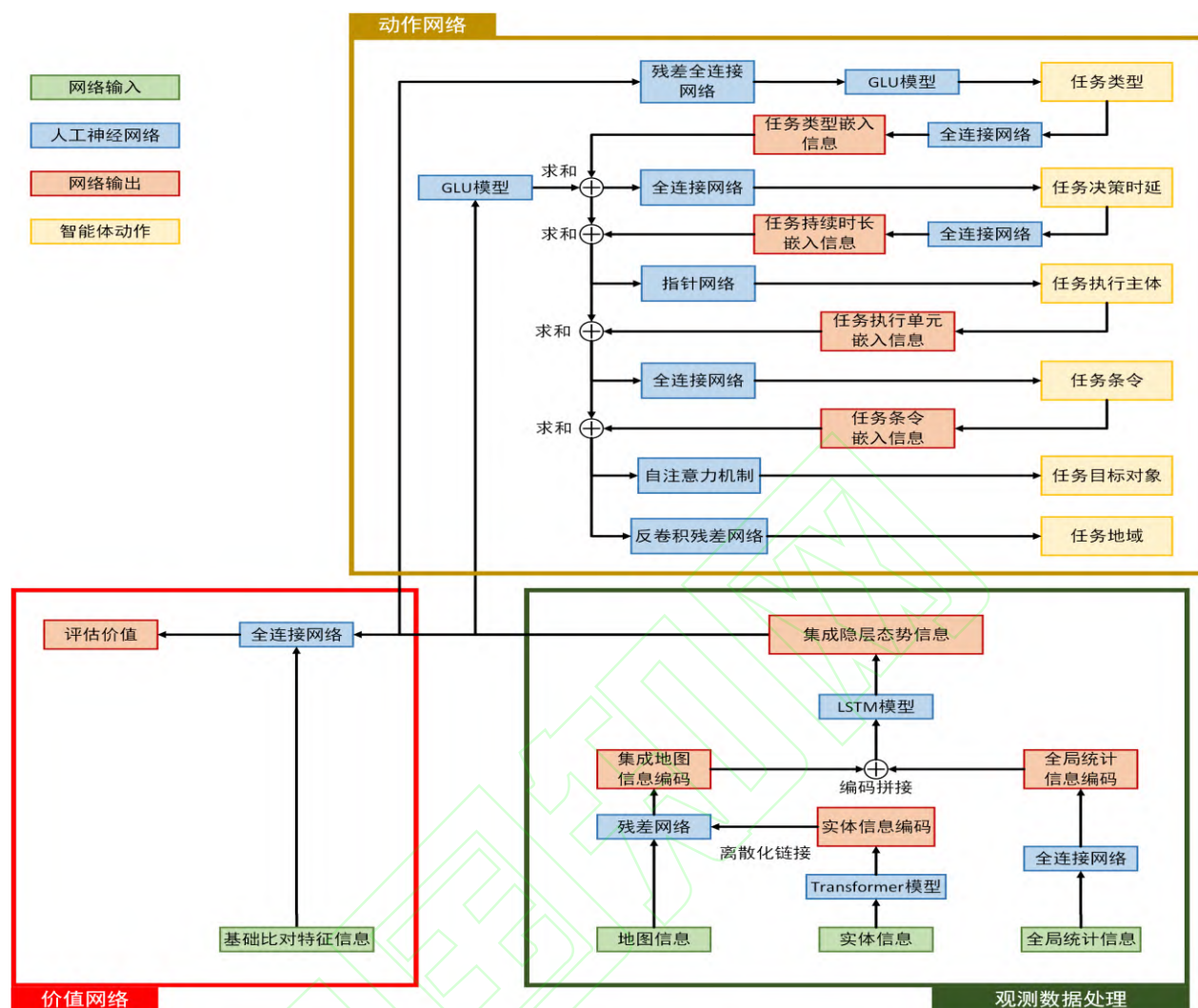


图2 任务级兵棋智能策略生成优化框架

Fig. 2 The framework of intelligent strategy generating in task-level wargame

设置等内容,且存在武器挂载等任务预规划问题,使得动作空间结构复杂性进一步提高。

四是博弈实体规模更大。与星际争霸博弈具有实体数额上限(200)不同,基于任务级兵棋的一次中等规模的战役背景下的对抗想定中,实体数目一般在数千以上,极大地增加了实体间的协同控制难度,因而对智能体算法提出了更高的要求。

五是决策持续时间更长。星际争霸一次对抗持续时间一般在30分钟至1小时以上,玩家需要做出数以千计的决策。而任务级兵棋一次推演仿真时长一般在1.5小时至3小时左右,相较于星际争霸,任务级兵棋玩家需要做出数倍决策才能最终确定胜负,因而训练数据更加稀疏。

六是多方策略博弈更多元。与双人对战的棋类游戏相比,星际争霸可进行多方对抗,而任务级兵棋是进一步支持敌我友民多方力量对抗合作的仿真,在此基础上的策略博弈更为复杂。

七是循环策略博弈更多样。与星际争霸相

同,任务级兵棋的策略集中不存在特定最优策略,其中不同策略在训练过程中相互克制,最终形成一个闭环的策略压制关系。因此,二者在训练过程中,无法简单地使用自博弈方法。同时,由于任务级兵棋中想定场景更为多样、实体种类更为多样、优化问题更为多样,造成任务级兵棋的循环策略博弈的闭环分析难度更大。

八是人类专家数据更稀缺。星际争霸属于国际范围内流行的竞技博弈游戏,服务端有数十万盘专业玩家数据可供智能体进行模仿学习。而任务级兵棋专业性强、普及度低,不具备使用大量人类专家数据对智能体进行训练的条件,因而更强调算法对现有专业知识抽取与运用。

从以上8个难点可以看出,面向棋类与竞技游戏类的智能博弈方法并不能完全满足任务级兵棋智能决策的需要,应开展针对性研究,提出面向任务级兵棋的智能决策框架。

2 任务级兵棋智能决策框架设计

开展任务级兵棋智能决策研究,应找到智能体在任务级兵棋博弈中的定位,找准智能体与兵棋仿真平台、人类专家间的关系,找出明确信息流转路径与知识流转路径。基于以上考虑,本节给出了任务级兵棋智能博弈框架、智能策略生成优化框架以及多智能体策略协同演进框架。

2.1 任务级兵棋智能博弈框架

任务级兵棋智能博弈框架如图 1 所示,主要以“墨子”联合作战推演系统为例说明智能体完整博弈决策过程,主要包含以下 3 个步骤:

第 1 步,态势信息分析。指将兵棋仿真引擎推送的整体态势数据按三类信息整编处理。

①地图信息:即主屏幕中的态势地图,主要描述了当前态势下敌我双方各实体的空间位置;

②实体信息:主要包括当前实体的航速航向等运动信息、经纬高等位置信息、作战任务条令与电磁管控条令信息、武器挂载与油耗余量信息等,主要描述当前各实体的状态特征;

③全局统计信息:主要包括作战时间进程、敌我双方各类实体损耗数、各类弹药消耗数、评分值等,主要描述了当前战场全局的整体进度与敌我双方的力量对比。

以上三类信息进行综合处理后,直接作为智能体对当前状态的观测值,以支撑智能体决策。

第 2 步,任务指令生成。智能体依据当前状态的观测值,依据自身策略生成所属实体需要执行的任务要素,主要包括:①任务类型、②任务地域、③任务执行主体、④任务目标对象、⑤任务条令、⑥任务决策时延,共计 6 类信息要素。最终转化为任务指令并发送至兵棋仿真引擎。

第 3 步,环境实时推演。兵棋仿真引擎接收智能体的任务指令后,进行实时推演,形成新的态势供智能体分析,并进入下一轮决策循环。

基于以上 3 个步骤进行循环迭代,实现智能体在任务级兵棋中的决策,从而产生大量仿真数据。在此数据基础上,智能体不断更新自身神经网络参数,实现策略的不断优化。

2.2 任务级兵棋智能策略生成优化框架

任务级兵棋智能策略生成优化框架如图 2 所示,其中绿色代表网络输入,蓝色代表网络模型,红色代表网络输出,黄色代表执行动作。该框架主体参照 A3C 算法框架设计,由观测信息处

理、价值网络与行动网络三个模块组成。

①观测信息处理模块,主要作用是压缩当前时刻地图、实体以及全局统计三类信息,并最终得出综合多帧态势的时序化状态信息。实现方法是,针对地图类、实体类、全局统计类信息,分别使用残差网络(ResNet)、Transformer 模型、全连接网络(FC Net)进行编码,经离散连接(Scatter Connection)与拼接操作后形成单帧态势信息,再由 LSTM 处理转化为时段内集成隐层态势编码(Integrated Hidden Layer Situational Encoder)供值网络评估与行动网络决策使用。

②值网络模块,主要作用是评估当前状态价值,以供行动网络学习优化。其结构相对简单,主要依据集成隐层态势编码与基础比对特征值,经全连接网络计算得出状态价值。需要指出,此处基础比对特征值,是指当前状态下一个或多个统计值向量与历史数据均值向量之差,如当前智能体拥有的主力装备数目与历史赛事中均值的差值。

③行动网络模块,主要作用是生成任务指令。该模块采用了多头输出结构,不同任务要素参考了不同信息,并使用了不同神经网络。其中,确定任务的类型主要使用了具有残差连接的全连接网络与门控线性单元(GLU);确定任务的决策时延与条令主要使用了结构简单的全连接网络;确定任务的执行主体使用了可解决组合优化问题的指针网络(Pointer Network);确定任务的目标使用了适于找寻关注点的自注意力机制(Self-attention);确定任务的地域则使用了适于图像处理的反卷积残差网络(Deconv ResNet)。

以上框架,只适用于训练单一策略的智能体,无法应对循环策略的学习。因而,还需要进一步引入相对完善的多智能体训练机制,使其不仅能快速收敛于一个相对完善的策略,更能“记住”每个特定策略的应对方法,从而破解智能体循环策略训练难题。

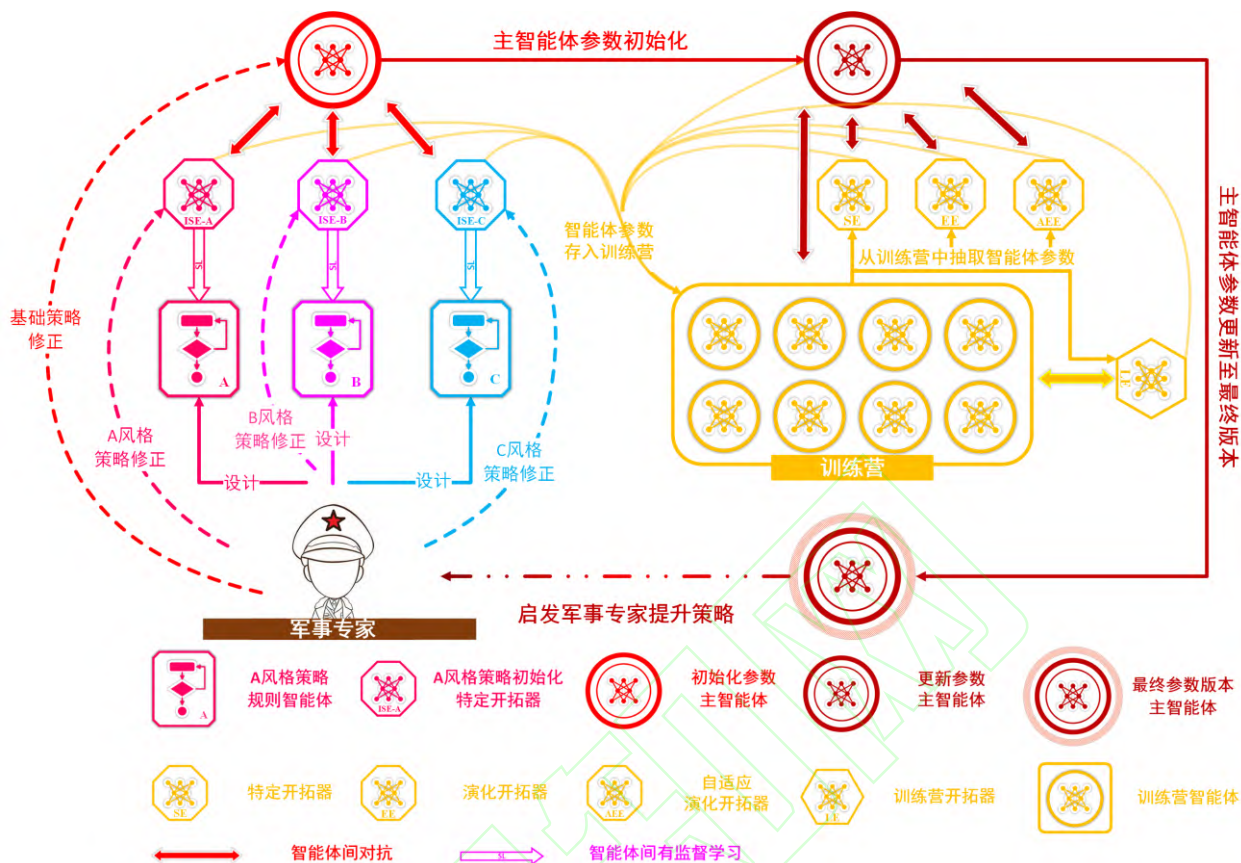


图3 任务级兵棋多智能体策略协同演进框架

Fig. 3 The framework of multi-agent strategy evolution in task-level wargame

2.3 任务级兵棋多智能体策略协同演进框架

任务级兵棋多智能体策略协同演进框架如图3所示,完整地展现了任务级兵棋智能决策中知识流动过程,主要分为以下5个阶段:

①专家知识规则化。人类专家依据特定场景,结合自身经验知识,设计多个风格不同且水平较高的规则智能体,将专家知识转化为可复用的程序,用于后续智能学习。

②规则知识参数化。针对不同风格的规则智能体设置初始特定开拓者(Initializing Specific Exploiter, ISE)智能体进行有监督学习,将规则智能体中的知识提炼到特定网络之中,固化为网络参数。与此同时,专家还针对相关策略进行标注,依托RGPS^[3]方法引导不同决策风格的特定开拓者快速收敛。

③网络参数初始化。用本次博弈训练的主智能体与不同决策网络的特定开拓者智能体展开对抗,使其训练伊始即可应对多种不同风格策略,并通过博弈不断提升稳定,形成性能相对较高的初始化参数。

④初始参数最优化。在初始化参数版本主智能体的基础上进行自博弈,基于训练营机制(Training League)^[1,14]生成针对特定策略的特

定开拓者(SE)^[3]、博弈水平相对较高策略的演化开拓者(EE)^[3]、博弈水平相对较低策略的自适应演化开拓者(AEE)^[3]对主智能体进行训练,并使用训练营开拓者(LE)^[1]不断查找训练营中的策略漏洞。从而保证训练营策略集不断丰富,促进主智能体策略的持续演进,实现博弈知识的自我发现,达到并超越人类专家的水平,并生成最终参数版本的主智能体。

⑤参数知识抽象化。使用最终参数版本的主智能体再次与人类专家进行博弈,基于智能算法优异表现展开定性分析,辅助人类专家修正自身策略盲点,启发人类专家进一步改进完善自身策略,实现将智能体知识转化为人类的抽象思维。

任务级兵棋多智能体策略协同演进框架揭示了智能决策研究的根本目标,即打通人类专家与智能算法间的知识循环,最终提升人类专家对相关问题的认知理解能力。

3 任务级兵棋智能决策关键问题与基本方法

在基本掌握任务级兵棋智能决策框架的基础上,可进一步总结梳理出以下四点关键技术问题与基本解决方案。

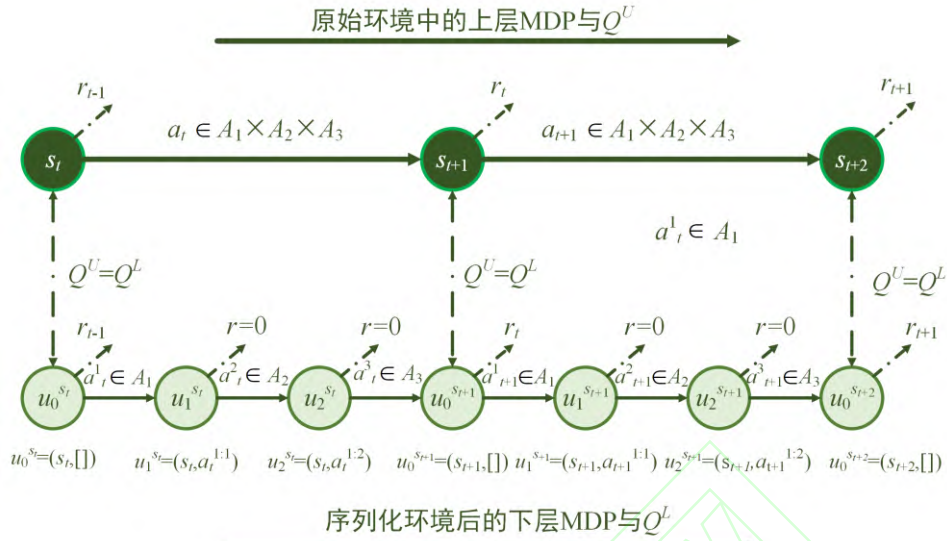


图4 Sequential DQN 中的马尔可夫决策过程结构示意图

Fig. 4 The demonstration of MDP in Sequential DQN

3.1 高维度长时段态势信息数据综合问题

高维度指任务级兵棋态势实体多、状态空间维度大的特点，要求智能体须具备信息降维分析能力；长时段指任务级兵棋决策需要时间跨度大，要求智能体须具备长时段的态势信息分析能力。参考文献[1, 3, 8]中相关成果，可以得出基于自编码器架构的解决方案，如图2中观测信息处理模块所示。针对高维信息压缩问题，首先利用Transformer模型自编码生成各实体信息编码，利用离散连接(Scatter Connection)将实体编码集成到地图数据后，经残差网络得出集成地图编码(Integrated Map Encoder)，再将其与全连接网络得出的全局统计信息编码(Statistics Encoder)进行拼接，形成压缩后的低维单帧态势信息；针对长时段态势信息压缩问题，直接使用面向非定长时序问题的LSTM循环处理单帧态势信息，得出能够综合反映一段时间内态势信息的集成隐层态势编码(Integrated Hidden Layer Situational Encoder)，完成高维度长时段态势信息数据综合处理。

3.2 异步策略分布式学习优化问题

要高效地开展任务级兵棋智能体训练，必须使用分布式强化学习架构，其核心是在于解决好基于回放寄存器(Replay Memory)的异步策略(Off-Policy)更新问题与定长采样下LSTM的初始化问题^[9-12]。其中，异步策略的更新问题可以直接使用V-trace方法解决，保证目标策略与源策略重要性比值始终大于或等于均值，促进智能体收敛。而定长采样下LSTM的初始化问题也可以直接参考使用文献[11]中的循环状态存储

(Stored State)与参数预热(Burn-in)方法，确保在可接受的Q值偏差(Q-value discrepancy)范围内实现对分布式定长采样数据的学习。

3.3 基于语义模板的任务指令生成问题

实现任务级兵棋智能决策，必须实现基于特定语义模板的任务指令控制，即明确：①任务类型、②任务地域、③任务执行主体、④任务目标对象、⑤任务条令、⑥任务决策时延，共计6类任务信息要素。基本方法可参考文献[13]中的Sequential DQN方法为基础进行分层处理。

如图4所示，Sequential DQN将决策过程分为上下两层，上层MDP将整条任务指令视作一个动作，反馈为环境反馈 r_t ，下层MDP将任务指令中每个要素决策视作一个子动作，之间反馈设为0。令上层整条任务指令的Q值与下层最后一项值相等：

$$Q^U(s_t, a_t) = Q^L(u_{N-1}^{s_t}, a_t^N) \quad (1)$$

通过自举法不断修正下层子动作的Q值，从而实现特定态势下对任务条令各信息要素值的评估，指导各任务信息要素的选择，结合图2中行动网络模块，即可生成结构化的任务指令。

3.4 多智能体训练营博弈调度问题

任务级兵棋中存在大量循环策略，易造成单智能体策略在多个循环策略之间不断变化无法提升。参考文献[14, 1, 3]有关成果，可使用多智能体训练营(training league)的方法。如图3所示，多智能体训练营的核心方法是维护一个基于胜率的支付矩阵，所有智能体对阵主智能体达到一定胜率后即将其策略参数存入训练营中。完

成网络初始参数最优化阶段后, 训练营参照胜率支付矩阵, 一方面按要求选拔特定策略开拓者、演化开拓者(胜率 50 至 70%以内)、自适应演化开拓者(胜率 20 至 50%以内)与主智能体对战, 另一方面选拔训练营开拓者与训练营全部智能体进行对战, 不断更新主智能体与开拓者策略, 逐步构建丰富完备的训练营策略集, 并最终实现主智能体策略的持续演进。

4 结论

本文在充分梳理当下兵棋智能决策研究现状的基础上, 分析了开展任务级兵棋智能博弈研究八项难点, 进而提出了任务级兵棋智能博弈框架、智能策略生成优化框架以及多智能体策略协同演进框架, 指出了当下应重点关注的高维度长时段态势信息数据综合、异步策略分布式学习优化、基于语义模板的任务指令生成、多智能体训练营博弈调度等 4 项关键技术问题, 并给出了基本解决方案, 点明了开展任务级兵棋智能决策研究的技术要点与基本路线, 为后续任务级兵棋智能体设计提供了有益参考。

References

- VINYALS 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.
- BERNER C, BROCKMAN G, CHAN B, et al. Dota 2 with large scale deep reinforcement learning[EB/OL]. (2019-12-13) [2021-08-04]. <https://arxiv.org/pdf/1912.06680.pdf>.
- HAN L, XIONG J, SUN P, et al. TStarBot-X: an open-sourced and comprehensive study for efficient league training in StarCraft II full game[EB/OL]. (2021-04-30) [2021-08-04]. <https://arxiv.org/abs/2011.13729.pdf>.
- 陈希亮, 张永亮. 基于深度强化学习的陆军分队战术决策问题研究[J]. 军事运筹与系统工程, 2017, 31(3): 20-27.
CHEN X L, ZHANG Y L. Study on tactical decision-making problem of army detachment based on Deep Reinforcement Learning[J]. Military Operations Research and Systems Engineering, 2017, 31(3): 20-27.
- 曹雷. 基于深度强化学习的智能博弈对抗关键技术[J]. 指挥信息系统与技术, 2019, 10(5): 1-7.
Cao Lei. Key technologies of intelligent game confrontation based on deep reinforcement learning[J]. Command information system and technology, 2019, 10(5): 1-7.
- 崔文华, 李东, 唐宇波, 等. 基于深度强化学习的兵棋推演决策方法框架[J]. 国防科技, 2020, 41(2): 113-121.
CUI W H, LI D, DANG Y B, et al. Deep Reinforcement Learning based decision-making method framework of military decision-making[J]. National Defense Science and Technology, 2020, 41(2): 113-121.
- MNIH V, BADIA A P, MIRZA M, et al. Asynchronous methods for Deep Reinforcement Learning[C]// International Conference on Machine Learning. PMLR, 2016: 1928-1937.
- HAUSKNECHT M, STONE P. Deep recurrent Q-Learning for partially observable MDPs[C]// 2015 AAAI Fall Symposium Series, 2015.
- ESPEHOLT L, SOYER H, MUNOS R, et al. IMPALA: scalable distributed Deep-RL with importance weighted actor-learner architectures[C]// International Conference on Machine Learning. PMLR, 2018: 1407-1416.
- ESPEHOLT L, MARINIER R, STANCZYK P, et al. SEED RL: scalable and efficient Deep-RL with accelerated central inference[EB/OL]. (2019-10-15) [2021-08-04]. <https://arxiv.org/abs/1910.06591>.
- KAPUROWSKI S, OSTROVSKI G, QUAN J, et al. Recurrent experience replay in distributed Reinforcement Learning[C]// International conference on learning representations, 2018.
- HORGAN D, QUAN J, BUDDEN D, et al. Distributed prioritized experience replay[C]// International Conference on Learning Representations, 2018.
- METZ L, IBARZ J, JAITLEY N, et al. Discrete sequential prediction of continuous actions for deep RL[C]// International Conference on Learning Representations, 2018.
- JADERBERG M, CZARNECKI W M, DUNNING I, et al. Human-level performance in 3D multiplayer games with population-based reinforcement learning[J]. Science, 2019, 364(6443): 859-865.
- HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Conference on Neural Information Processing Systems, 2017.
- RUSU A A, COLMENAREJO S G, GULCEHRE C, et al. Policy distillation[C]// International Conference on Learning Representations, 2016.

- 18 MIKOLOV T, KARAFIÁT M, BURGET L, et al. Recurrent Neural Network based language model[C]// Eleventh Annual Conference of the International Speech Communication Association, 2010.
 - 19 VINYALS O, FORTUNATO M, JAITLY N. Pointer networks[C]// Conference on Neural Information Processing Systems, 2015.
 - 20 OH J, GUO Y, SINGH S, et al. Self-imitation learning[C]// International Conference on Machine Learning. PMLR, 2018: 3878-3887.
 - 21 LEVINE S, KOLTUN V. Guided policy search[C]// International Conference on Machine Learning. PMLR, 2013: 1-9.
 - 22 DAUPHIN Y N, FAN A, AULI M, et al. Language modeling with gated convolutional networks[C]// International Conference on Machine Learning. PMLR, 2017: 933-941.
 - 23 SUN P, SUN X, Han L, et al. Tstarbots: defeating the cheating level builtin AI in StarCraft II in the full game[EB/OL]. (2018-12-27) [2021-08-04]. <https://arxiv.org/abs/1809.07193.pdf>.
 - 24 VINYALS O, EWALDS T, BARTUNOV S, et al. Starcraft II: a new challenge for reinforcement learning[EB/OL]. (2017-08-16) [2021-08-04]. <https://arxiv.org/abs/1708.04782.pdf>.
 - 25 JADERBERG M, MNH V, CZARNECKI W M, et al. Reinforcement Learning with unsupervised auxiliary tasks[C]// International Conference on Learning Representations, 2017.
 - 26 ZAMBALDI V, RAPOSO D, SANTORO A, et al. Deep reinforcement learning with relational inductive biases[C]// International Conference on Learning Representations, 2018.
 - 27 胡晓峰, 荣明. 关于联合作战规划系统的几个问题[J]. 指挥与控制学报, 2017, 3(4): 273-280.
HU X F, RONG M. Problems about joint operation planning system[J]. Journal of Command and Control, 2018, 3(4): 273-280.
 - 28 谢苏明, 毛万峰, 李杏. 关于作战筹划与作战任务规划[J]. 指挥与控制学报, 2018, 3(4):281-285.
XIE S M, MAO W F, LI X. About operational planning and operational mission planning[J]. Journal of Command and Control, 2018, 3(4): 281-285.
 - 29 邹烨翰, 冯旻赫, 程光权, 黄金才, 刘忠. 面向军事条令条例的本体构建技术[J]. 指挥与控制学报, 2019, 5(1):47-54.
ZOU Y H, FENG Y H, CHENG G Q, HUANG J C, LIU Z. Ontology developing technology for military regulations [J]. Journal of Command and Control, 2019, 5(1):47-54.
 - 30 夏阳升, 石建迈, 陈超, 黄金才, 程光权. 车机协同多区域覆盖侦察路径规划方法[J]. 指挥与控制学报, 2020, 6(4):372-380.
XIA Y S, SHI J M, CHEN C, HUANG J C, CHENG G Q. Path planning method for multi-area reconnaissance by cooperated ground vehicle and drone[J]. Journal of Command and Control, 2020, 6(4):372-380.
 - 31 王飞跃, 刘玉超, 秦继荣, 戴浩. C2M 和 5G:新时代的智能指挥与控制[J]. 指挥与控制学报, 2019, 5(2):79-81.
WANG F Y, LIU Y C, QIN J R, DAI H. C2M and 5G: Intelligent Command and Control in the Connected and Smart Age[J]. Journal of Command and Control, 2019, 5(2):79-81.
 - 32 阳东升, 朱承, 肖卫东, 张维明, 李强, 修保新, 王飞跃. 宏观尺度 C2 过程机理:多域多 PREA 环及其冲突协调模型[J]. 指挥与控制学报, 2021, 7(1):11-27.
YANG D S, ZHU C, XIAO W D, ZHANG W M, LI Q, XIU B X, WANG F Y. Process Mechanism for Macro Command and Control:Coordinate Conflicts and Resolution Model for Multi-PREA Loop in Multi-Domains[J]. Journal of Command and Control, 2021, 7(1):11-27.
- 张驭龙 (1988—), 男, 博士研究生, 工程师, 主要研究方向为信息系统、强化学习、智能博弈。
- 范长俊 (1990—), 男, 博士, 讲师, 主要研究方向为复杂系统、图深度学习及数据挖掘。
- 冯旻赫 (1985—), 男, 博士, 副教授, 主要研究方向为因果发现与推理、主动学习及强化学习。本文通信作者。E-mail: fengyanghe@nudt.edu.cn。
- 张龙飞 (1988—), 男, 博士研究生, 主要研究方向为机器学习与强化学习。
- 刘忠 (1968—), 男, 博士, 教授, 主要研究方向为通用人工智能、多智能体系统及强化学习。本文通信作者。E-mail: Liuzhong@nudt.edu.cn