

基于博弈对抗的空战智能决策关键技术*

陈希亮¹ 李清伟² 孙 彧³

(1 陆军工程大学指挥控制工程学院 南京 210007)

(2 中国电子科技集团公司第二十八研究所 南京 210007)

(3 解放军31102部队 南京 210000)

摘 要: 针对空战智能决策模型生成问题,首先,分析了空战智能决策构建的意义,并对基于博弈论的方法、基于规划理论的方法以及专家系统与机器学习方法等空战决策模型生成技术进行了对比分析;然后,阐述了博弈对抗规则建模、机器学习优化、智能算法模型验证评价、大样本并行仿真推演和柔性平台集成接口设计等关键技术;最后,对智能技术空战应用的发展趋势进行了总结。

关键词: 博弈对抗;深度强化学习;智能空战

中图分类号: TP399 **文献标志码:** A **文章编号:** 1674-909X(2021)02-0001-06

Key Technologies for Air Combat Intelligent Decision Based on Game Confrontation

CHEN Xiliang¹ LI Qingwei² SUN Yu³

(1 College of Command Information System, Army Engineering University, Nanjing 210007, China)

(2 The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210007, China)

(3 Unit 31102 of PLA, Nanjing 210000, China)

Abstract: Aimed at the generation problem of air combat intelligent decision model, firstly, the significance of air combat intelligent decision construction is analyzed, and the air combat decision model generation technologies are compared, such as the methods based on game theory, the methods based on planning theory, and the expert system and machine learning methods. Then, the key technologies are described, such as the game confrontation rule modeling, the machine learning optimization, the intelligent algorithm model validation and evaluation, the large sample parallel simulation deduction, and the flexible platform integration interface design. Finally, the development trend of intelligent technology in air combat is summarized.

Key words: game confrontation; deep reinforcement learning; intelligent air combat

0 引 言

随着现代战争节奏不断加快、复杂性不断上升,军事装备和信息系统的智能滞后问题愈发凸显。为了满足未来战争快速决策、自动决策和自主决策需

求,迫切需要智能技术提升人脑辅助决策能力,以适应高速、复杂和多变的战场环境^[1]。因此,人工智能被引入作战辅助决策,而具有代表性的是美军“深绿”计划。该计划核心是基于实时态势的动态仿真,帮助人类指挥员理解各种量变在不同时空上相互交

* 基金项目:国家自然科学基金(61806221)、装备发展部“十三五”预研课题(31505550302)和国防科技重点实验室基金(6142101180304)资助项目。

收稿日期:2021-01-22

引用格式:陈希亮,李清伟,孙彧.基于博弈对抗的空战智能决策关键技术[J].指挥信息系统与技术,2021,12(2):1-6.

CHEN Xiliang, LI Qingwei, SUN Yu. Key technologies for air combat intelligent decision based on game confrontation[J]. Command Information System and Technology, 2021, 12(2): 1-6.

又产生作用后可能带来的质变,其目标是协助指挥员做出正确决策^[2-3]。此外,2013年,美国国防高级研究计划局(DARPA)启动了先进机器学习项目,旨在提高机器学习的适用性,使其他领域的专家也能使用复杂的机器学习方法解决问题。2016年,辛辛那提大学开发了智能飞行员 AlphaAI项目,采用遗传模糊搜索树方法,在空战模拟器中与人类飞行员对战并获得全胜;美军启动指挥官虚拟参谋项目,旨在将智能技术和信息系统整合,实现决策支持工具自动化,减轻指挥员的认知负担。2019年3月,美国空军公开了智能僚机项目,旨在帮助人类飞行员在任务期间通过洞悉空中战斗和复杂飞机控制,对其周围发生的一切事情做出反应,并迅速推荐行动方案。2019年5月21日,DARPA启动了空战演进项目,主要涉及:1)为局部(个人和团队战术)行为建立战斗自主性;2)建立和校准对空战局部行为的信任;3)将战斗自主性和信任扩展到全局(异构多机)行为;4)建立全面的空战试验基础设施。2020年8月,阶段性对抗试验 AlphaDogFlight中,以端到端深度强化学习为核心的智能技术在试验中全面超越人类飞行员,预计在2024年将智能算法移植到飞机上以替代人类飞行员。2019年8月9日,美空军研究实验室和DZYNE公司在杜格威试验场进行了机器飞行员(ROBOPilot)系统试验,采用摄像头采集飞机仪表盘数据,利用机械传动设备控制操纵杆、踏板和开关等,通过非侵入式手段进行无人化改装。此外,忠诚僚机、Skyborg、拒止环境中协同作战、系统集成技术与试验、小精灵、进攻性蜂群使能战术、“敏捷秃鹫”吊舱项目和“雷霆”隐身无人作战飞机等智能化空战项目均使用人工智能方法构建飞行器智能决策模型,通过自主/半自主决策来提高作战实体的快速决策能力。2020年12月,XQ-58A实现与F-22、F-35A和F-35B等有人战机协同半自主编队飞行;同月,美国空军的ARTUM无人U-2项目表明了无人系统的操作权首次超过人类飞行员,展现出自主化空中作战的趋势。

智能博弈对抗技术是智能化指挥控制、训练模拟和自主集群无人化作战等军事领域智能化建设的核心技术基础^[1]。智能空战决策模型的构建具有重要意义,体现在以下3个方面:1)训练空中编队指挥员或飞行员:智能无人空战模型可学习空战作战人员的战术经验,并可通过模拟训练中的敌方来帮助强化训练编队指挥员或飞行员;2)辅助地面领航员及空战编队的作战人员:智能无人空战的成果以软件模块化的形式嵌入战机中,进行辅助引导或空战战术辅助决策,减轻地面领航员及空战编队作战人

员的压力;3)操控无人僚机参与编队空战以降低飞行员风险:智能无人空战模型可自主操控编队僚机参与危险区域空战,降低飞行员的风险及空战成本。

目前,以深度学习和强化学习为代表的人工智能方法仍依赖于数据和计算资源,然而在面临高复杂性现实问题时,目前的数据和计算资源难以满足需求。华为英国研究中心研究成果称“Deepmind算法耗尽全球GPU算力都实现不了”^[4]。此外,空战决策面临的问题与围棋、星际争霸和Dota2等问题不同,无法简单套用深度学习和强化学习等算法,需针对具体军事问题,综合运用传统运筹、知识建模、智能搜索和机器学习等方法寻找解决方案。

1 空战智能决策模型构建技术研究现状

空战博弈对抗决策模型研究可追溯到20世纪60年代,本文对相关文献总结归纳后将空战博弈对抗决策模型方法分为基于博弈论的方法、基于优化理论的方法和专家系统和机器学习方法3个大类。空战博弈对抗决策模型生成方法如表1所示。

1.1 基于博弈论的方法

由于空战双方在对抗中产生了固有的博弈特性,因此用博弈论概念描述和求解空战博弈问题成为主要方向。其中,有代表性的为追捕模型、差分博弈模型和影响图模型^[5-6]。

1)追捕模型:指空战双方中的一方攻击、另一方逃逸的对抗模型,本质是将双方博弈策略极值化的问题,目前求解方法包括定性和定量2种方法。为了提升仿真的真实性,该模型生成了较多扩展版本,如基于追捕逃逸角色互换的双目标模型和基于离散矩阵的追捕模型等。由于追捕模型的设计理念较简单,因此目前应用较少。

2)差分博弈:是一种经典理论,本质是使用差分方程表示2个或多个决策者的控制动作,并将其应用到运动系统中以实现最佳策略。一般使用最优控制论方法求解差分博弈,在空战博弈等机动决策中,差分博弈的表现形式通常为求解交战双方的攻防策略的极值。尽管差分博弈理论成果较丰富,但在仿真中的表现仍然欠佳,原因如下:(1)差分博弈的实例化设置较困难,难以成为正常的行为理论;(2)计算复杂度较高,制约了差分博弈在高维环境中的应用;(3)结构性原因导致差分博弈模型的稳定性较差。因此,差分博弈理论虽已发展多年,但一直没有成熟的应用。

3)影响图模型:是一种直观的空战建模方法,融合了飞行员的习惯、偏好和感知等先验知识,并利用贝叶斯网络等方法进行行为推理。其中,典型应

表 1 空战博弈对抗决策模型生成方法

类别	典型方法	特性	优点	缺点	适用场景
基于博弈论的方法	追捕模型	使用极值化求解双方的博弈策略	模型简单和易于计算	理念朴素和仿真度低	单一追捕逃逸场景
	差分博弈模型	采用差分方程表示决策者动作以实现最佳策略	理论成果丰富和涉及面较广	难以实例化、计算复杂度高和仿真稳定性差	多种复杂对抗场景
	影响图模型	融合了飞行员的先验知识并采用贝叶斯网络进行推理	加入人类经验知识、态势评估直观高效和仿真度较高	复杂度较高、难于解析求解和不适用复杂环境	涉及飞行员与环境实时交互的仿真场景
基于优化理论的方法	近似动态规划(ADP)	使用函数近似方式表现回报值	仿真精度高和适用场景广	不支持多机对抗模拟仿真	实时自主的一对一空战场景
	粒子群优化(PSO)	每个粒子代表一个问题的一个可能的解,由单个粒子的相互作用来求解	搜索速度快、效率高和算法简单	难以适用离散优化问题和易陷入局部最优	离线仿真场景
	蚁群优化(ACO)	采用启发式概率搜索方式逼近最优解	鲁棒性与泛化性强以及支持并行实现	收敛速度慢、易陷入局部最优解和探索时间长	离线仿真场景
专家系统和机器学习方法	专家系统	将不同空战情况的经验提炼成规则库,并通过规则生成飞行控制指令	模型简单实用和规则库可靠性强	规则库复杂、策略固定以及针对不同机型需要调参编码	多种复杂对抗场景
	神经网络(NN)	使用 NN 存储空战规则,并在大样本中不间断学习	博弈对抗策略鲁棒性强和学习效率高	因自身网络的封闭性难以与外部经验知识融合,且精确度较低	封闭式模拟对抗场景
	强化学习(RL)	无需标记学习样本,通过与外部环境交互选择最佳动作值,并更新自身策略	可实现在线实时决策和多目标自主学习	传统 RL 只能处理小规模简单问题,单智能体 RL 无法形成良好合作策略	多种复杂对抗场景

用有基于单层和多层影响图的策略推理仿真^[7]。总的来说,影响图模型具有人类经验知识灵活辅助系统开发、显式高效评估态势威胁等空战因素以及对飞行员实际仿真度较高等优点,但同时也存在难以解析求解和无法适用于复杂环境等缺点。

1.2 基于优化理论的方法

基于优化理论的决策包括遗传算法和动态规划等优化理论,其本质是通过数学方法将空战对抗问题转化为最优化问题,并通过求解优化模型获得最优空战策略。其中,典型方法有用于实时自主的一对一空战的 ADP^[8]算法。与经典动态规划不同,ADP 方法主要使用函数近似方式表现回报值,在多个空战模拟平台上均有应用并具备相应性能。另外,PSO^[9]和 ACO^[10]等方法也广泛应用于各类模拟空战平台。由于基于优化理论方法的可扩展性较差、仿真实时性不理想且无法实现在线决策,因此该方法只能用于离线空战策略生成。

1.3 专家系统和机器学习方法

专家系统和机器学习方法主要有以下 3 个子类:

1) 专家系统方法:该方法的研究起步较早且技术思想较成熟,其核心思想是将不同空战经验提炼成规则库,并通过规则生成飞行控制指令,其中有代表性的是美军适应性机动逻辑程序(AML)^[11]。该方法的不足表现如下:(1) 规则库建立过程过于复杂;(2) 规则库中的策略较简单且固定;(3) 针对不同机型的调参和编码难度大。上述问题极大限制了

专家系统方法的应用范围和仿真真实性。

2) NN 方法:指利用神经网络存储空战规则的方法。目前在空战对抗领域应用较广泛的模型为多层反向回报网络(MBPN)^[12]。由于神经网络具备从大量空战对抗样本中不断学习的能力,因此生成的博弈对抗策略鲁棒性较高。然而神经网络因其自身的封闭性较难与人类经验知识融合,故仿真精确度饱受质疑。

3) RL 方法:与专家系统方法和 NN 方法相比,RL 方法无需标记学习样本,智能体通过与外部环境交互来选择最佳动作值,进而更新自身策略。该方法实现了在线实时决策和多目标自主学习,成为解决无模型序贯决策问题的有效方法,同时契合了空战博弈对抗任务。传统 RL 方法利用 Q 值表格进行建模和求解,但仅能处理小规模简单问题。随着深度强化学习及其具有代表性的深度 Q 网络(DQN)和深度确定性策略梯度(DDPG)算法的诞生和应用,学者们尝试将单智能体深度强化学习算法用于空战模拟环境以生成优秀的对抗策略。例如,文献[13]将 DQN 用于一对一近距空战决策的生成,但因其只适用于离散动作状态空间,故难以满足飞行模拟的实际需求;文献[14]使用 DDPG 模拟交战双方的连续机动策略模型,但因其只支持低维空间的学习过程,故仿真精度和逼真度均较低。DARPA 推出的空战演进项目(ACE)主要通过多种 RL 方法建模实现在大规模场景中的人机协同“狗斗”战术,确保人工智能与人类驾驶员在高速、高重力加速度

情况下获取更大的战斗优势。

综上,传统方法由于在学习过程中缺乏智能体间的交互,在多机博弈对抗场景下无法形成有效的合作竞争策略。因此,将深度强化学习方法应用于多智能体空战博弈对抗场景以强化优质联合策略的生成成为未来探索方向。

2 空战博弈对抗决策及其关键技术

空战博弈对抗的筹划阶段,指挥员和参谋人员依据上级作战意图和作战任务制定多套作战方案,经作战推演分析和评估比较,确定最终作战方案。作战方案的核心是各作战实体的作战行动序列,而从研究角度看则是作战实体(模拟指挥员/飞行员进行决策的智能体)的序贯决策问题。利用人工智能技术模拟指挥员思维过程,生成并优化行动序列,直接将作战方案的制定和评估整合为一个整体,借助指挥实体智能决策模型进行动作选择,实现行动序列的生成与优化。

空战博弈对抗的战斗实施阶段,指挥员/飞行员根据战场环境与作战资源的变化进行临机决策,动态调整战斗飞机。临机决策本质是指指挥实体的序贯决策问题,借助指挥实体智能决策模型进行行动选择,在临机情况下根据战场环境的变化动态调整作战行动序列,辅助指挥员快速应对不完美战场空间的高动态、不确定和复杂的变化。

空战博弈对抗过程中,智能体在与战场环境的持续交互中不断学习经验,不断更新深度神经网络来指导其连续的行为选择。训练得到的深度神经网络可理解为指挥员经验知识的非线性表达,是指指挥决策思维过程的表征^[15-16]。然而,在此过程中也面临了以下难题:战场态势复杂多变,决策过程面临高维状态-动作空间,在巨量的搜索空间进行有效探索是一个巨大挑战^[17];作战任务的目标难以直接转化为智能体可理解的形式,无法提供即时准确的反馈,导致作战行动面临反馈稀疏、延迟和不准确的问题。

综上,在基于博弈对抗的空战决策模型生成中需解决以下问题:1) 构建规则驱动的博弈对抗行为模型,实现规则驱动的行动链生成;2) 构建机器学习优化的策略模型生成机制,实现学习型策略的持续优化;3) 构建博弈对抗策略评价体系和模型,对智能决策模型的作战能力进行全过程、多视角和可追溯的评价体系;4) 构建大样本并行仿真推演技术和柔性集成平台接口,确保机机对抗产生数据的速度和仿真环境与智能算法的柔性集成。

2.1 博弈对抗规则建模

以深度强化学习为代表的人工智能技术在棋类

游戏、即时战略游戏和 Alpha Dogfight 视距内空战对抗中的胜利,展现了通用人工智能技术在解决高复杂性问题的能力。目前,能够解决的问题的主要特点是获取数据容易、决策规则完备、对抗规则清晰及评估调优目标明确等,而空战决策问题存在获取数据难度大、决策规则不完备、专业知识复杂、评估调优难度大及信息不完全且不完美等特点,因此进行对抗规则建模成为解决这些难题的基础。博弈对抗规则建模需解决以下问题:1) 针对作战条令和作战规则等非结构化数据进行建模;2) 对高级指挥员大脑中的经验知识进行建模;3) 对格式化的历史对抗数据中隐含的对抗经验知识进行建模。

针对上述问题,国内外学者展开了相关研究,在非结构化数据定性到定量转化方面,基于云推理模型的定性知识规则生成方法,依据云推理模型对定性规则知识进行表达,对客观世界知识的模糊性与随机性进行形式化表示,将不确定的和模糊的经验知识转化为计算机能够识别的规则^[18],从而实现定性知识到定量知识的转化。此外,结合模糊数学理论对博弈对抗中关键特征进行模糊化表示,将文本转化为模糊推理规则^[19];在指挥员经验知识利用方面,既可通过研讨交流及决策过程关键节点提取来辅助决策活动分析,又可通过系统与专家在线交互引导决策模型生成;在格式化历史对抗数据挖掘方面,遗传模糊系统可在不完全且不完美信息对抗条件下,利用专家知识初始化对抗模型,采用历史对抗数据对系统进行优化,实现策略规则建模。结合典型空战对抗作战样式,利用构建的决策规则驱动智能实体对抗,实现规则驱动的博弈链动态生成。

2.2 博弈对抗策略机器学习优化

以深度学习和强化学习为代表的机器学习方法在问题求解中存在数据利用率低、环境反馈稀疏、状态空间探索与利用难等问题,导致解决方案难以收敛或训练时间过长,而在机器学习框架中引入军事领域的条令条例、战法战例和专家经验等知识,以及有经验的作战人员的历史博弈对抗决策轨迹、在线指导学习过程等方法,可大幅提升强化学习算法的探索能力。但上述过程也面临以下挑战:1) 基于强化学习框架构建非完美信息博弈对抗行为模型;2) 对基于神经网络的博弈对抗策略进行建模;3) 基于数据和知识驱动实现博弈对抗策略的持续性高效优化。在此基础上,分析不同博弈策略效果,挖掘隐藏的规律,寻求对博弈策略的优化,从而帮助指挥人员对作战方案进行优化设计,对战法战术进行提炼。

以上分析可知,对人类知识的充分利用符合了人类解决复杂决策问题的一般思维方式。由于指挥

决策样本数据量小,直接采用监督学习方法训练的决策模型泛化能力较弱,因此利用小样本数据并结合先验知识,构建不完美信息条件下的高动态博弈对抗模型是解决这些难题的基础,具体内容包括:1)通过在智能体策略学习过程中引入军事领域中的条令条例、战法战例等规则知识,提高策略学习过程中探索和利用的效率;2)利用对手的条令条例、战法战例等构建对手规则,提升对抗学习的针对性;3)结合多模态深度神经网络的空间态势感知技术,对空战对抗过程中的雷达、红外和飞控等数据采用不同预处理方法,并共享决策神经网络,尽可能覆盖决策因素,实现对战场态势的多维感知建模;4)为了避免过度依赖领域知识,利用逆向强化学习、数据增强和元学习等小样本数据学习方法,获取高水平对抗数据背后的决策考量,加速学习算法收敛;5)在强化学习过程中引入专家在线评价,将蕴含人类专家对任务目标理解的评价结果应用于学习中,对回报函数进行重构,从而提升反馈的实时性和准确性。

2.3 博弈对抗人工智能算法模型验证评价

经典空战能力评估主要通过分析影响和制约空战能力的主要因素,针对战斗机生存、目标探测和火力打击等能力构建分层指标体系,通过分析计算实现空战能力评估。在对生成的算法策略进行评估方面,仅基于空战能力进行评估无法反映算法模型的优劣。人工智能算法模型验证评价面临以下挑战:1)如何构建针对算法模型的评价指标体系,从而全方位反映人工智能算法模型的特点;2)如何利用验证评价结果引导算法模型改进和优化;3)在现有神经网络模型可解释性弱的情况下,如何构建对人工智能模型的可信度和可用性的评价。

针对上述挑战,首要解决的是评价指标体系构建问题。在构建作战实体博弈对抗评价指标体系基础上,依托智能推演引擎,以评估规则为尺度,以评估方法为手段,以试验评估数据为支撑,开展面向智能生成策略的多视角全过程效能分析。针对不同的作战阶段、作战时节和作战实体,在不同评估指标下构建作战实体行为指标,采用模糊数学方法将定性评估转换为定量评估,从而得出针对各项指标的专项统计分析,并重点从任务完成情况、歼敌效果、兵力损失、资源损耗和综合效益等视角加以评估,最终得出对博弈对抗条件下智能算法模型策略的综合评估。根据专项统计分析和评估结果给出智能算法模型存在的不足和改进方向,实现博弈对抗推演智能化水平的持续提升。在可信度和可用性方面,通过模型历史对抗情况和实时推演结果,以战术推荐置信度评价、行动反馈推演和实时态势优势分析等方

式,将智能算法模型的表现情况以人类指挥员能够理解的方式呈现,从而提升人类指挥员对于其可信度和可用性的认知。

2.4 大样本并行仿真推演技术和柔性集成平台接口设计

虽然机器学习对历史数据集的需求量不大,但它对交互产生的数据的需求量较大。简单的雅达利游戏需要超过1 800万帧的交互数据才能初步解决问题,而在更复杂的空战决策应用上需要的交互数据量则更大。当智能算法模型与对抗平台交互的速度足够快,则可在可接受的时间内产生满足模型训练需要的交互数据量。基于博弈对抗的智能空战决策模型生成,就是要在高逼真度空战仿真平台上,通过高速主机自主博弈对抗,大幅减少“人在环”的比率,通过不间断的机器自我博弈快速产生海量对抗样本,充分模拟博弈对抗不确定性带来的多种可能交战过程和结局。该过程需解决以下问题:1)提升仿真推演速度;2)通过平台智能化设计或改造提升对抗性试验效率。

针对仿真推演速度的难题,主要从以下2个方面进行:1)通过对仿真平台引擎进行设计,实现无渲染情况下百倍、千倍甚至万倍的加速,以满足在可接受时间内产生足够的对抗推演样本的需求;2)在平台引擎受限及加速比无法提升的情况下,通过分布式并行训练与优化机制实现大规模并行仿真训练框架,从而加速对抗数据生成速度。分布式并行训练需解决以下2方面的问题:1)设计硬件并行架构对仿真环境进行封装,实现与对抗仿真环境的分布式交互和采样;2)在多个对抗仿真环境异步采样情况下,通过批处理、经验缓存和重要性采样等机制整合采样数据并训练决策模型。此外,机器学习算法需要大量的试验、参数和超参数,以及不同算法的对比分析,这些均需仿真平台与算法具有松耦合的柔性架构,以及能够在一定范围内满足和适应变化的算法模型和参数。因此,在设计对抗仿真平台时需通过数据规范描述、可配置算法参数和动态映射等技术,在平台归平台、算法归算法的基础上,实现二者间的柔性集成。为了实现超实时对抗试验,解决传统博弈对抗行为建模方法面临的隐性知识表示困难和数据缺乏等难题,可基于博弈对抗的作战方案和战法试验等分析数据来提升分析结果的说服力和学习结果的置信度。

3 结束语

目前,人工智能技术在空战中的应用已初露端倪,一旦人机协作投入战斗,人类飞行员的反应速度

和操作精度远跟不上智能算法,更无法监督或干预人工智能的决策。2018年8月的国际人工智能联合会议上,国际前沿的160余家人工智能技术研发公司和组织在斯德哥尔摩签署“致命性自主武器宣言”以阻止智能技术在自主武器系统中的应用,其中包括谷歌Deepmind创始人Mustafa Suleyman。然而,美国空军创建的ARTU μ 智能驾驶员不仅在技术上来自谷歌Deepmind,更直接借助谷歌Kubernetes云容器实现算法的部署和管理。

随着科学技术的不断发展,人工智能技术的军事应用已是大势所趋。与人类飞行员相比,人工智能飞行员不受人心理和生理极限的约束,机动性更强、决策速度更快且操作精准度更高。算法+无人战斗机将统治空中战场。目前,智能空战仍处于实验室仿真阶段,然而智能空战、无人空战和自主空战的奇点已然临近。在基于博弈对抗的空战智能决策模型生成方面,核心是智能算法的集智攻关与力求突破,基础是在高逼真度仿真平台建设和算法模型从仿真到作战平台的迁移,核心和基础同步发展才能率先突破奇点,占据军事智能的制高点。

参考文献(References):

- [1] 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.
- [2] 胡晓峰, 郭圣明, 贺筱媛. 指挥信息系统的智能化挑战——“深绿”计划及AlphaGo带来的启示与思考[J]. 指挥信息系统与技术, 2016, 7(3): 1-7.
- [3] 曹雷, 陈希亮, 汤伟. 智能化陆军建设[J]. 国防科技, 2019, 40(4): 14-19.
- [4] YANG Y D, TUTUNOV R, SAKULWONGTANA P, et al. α -rank: practically scaling α -rank through stochastic optimisation[C]//Proceedings of the 19th International Conference on Autonomous Agents and Multi-agent Systems. Richland: International Foundation for Autonomous Agents and Multi-agent Systems, 2020: 1575-1583.
- [5] WISHART D. Differential games: a mathematical theory with applications to warfare and pursuit, control and optimization[J]. Physics Bulletin, 1966, 17(2): 60.
- [6] POROPUDAS J, VIRTANEN K. Game-theoretic validation and analysis of air combat simulation models[J]. IEEE Transactions on Systems, Man, and Cybernetics: Part A: Systems and Humans, 2010, 40(5): 1057-1070.
- [7] VIRTANEN K, RAIVIO T, HÄMÄLÄINEN R P. Decision theoretical approach to pilot simulation[J]. Journal of Aircraft, 1999, 36(4): 632-641.
- [8] MCGREW J S, HOW J P, WILLIAMS B, et al. Air-combat strategy using approximate dynamic programming[J]. Journal of Guidance, Control, and Dynamics, 2010, 33(5): 1641-1654.
- [9] DUAN H B, LI P, YU Y X. A predator-prey particle swarm optimization approach to multiple UCAV air combat modeled by dynamic game theory[J]. IEEE/CAA Journal of Automatica Sinica, 2015, 2(1): 11-18.
- [10] LI Y, KOU Y X, LIZ W, et al. A modified pareto ant colony optimization approach to solve biobjective weapon-target assignment problem[EB/OL]. (2017-05-16) [2020-12-15]. <https://www.hindawi.com/journals/ijge/2017/1746124/>.
- [11] BURGIN G H, FOGEL L J, PHELPS J P. An adaptive maneuvering logic computer program for the simulation of one-on-one air-to-air combat: volume 1: general description[EB/OL]. [2021-01-11]. <https://core.ac.uk/display/42887441>.
- [12] 陈华, 张可, 曹建蜀. 基于PSO-BP算法的目标威胁评估[J]. 计算机应用研究, 2012, 29(3): 900-901.
- [13] LIU P, MA Y F. A deep reinforcement learning based intelligent decision method for UCAV air combat[C]//Proceedings of the 17th Asia Simulation Conference on Modeling, Design and Simulation of Systems. Singapore: Springer, 2017: 274-286.
- [14] YANG Q, ZHU Y, ZHANG J, et al. UAV air combat autonomous maneuver decision based on DDPG algorithm[C]//2019 IEEE 15th International Conference on Control and Automation (ICCA). [S. l.]: IEEE, 2019: 37-42.
- [15] 马文, 李辉, 王壮, 等. 基于深度随机博弈的近距离空战机动决策[J]. 系统工程与电子技术, 2021, 43(2): 443-451.
- [16] 张强, 杨任农, 俞利新, 等. 基于Q-network强化学习的超视距空战机动决策[J]. 空军工程大学学报(自然科学版), 2018, 19(6): 8-14.
- [17] 左家亮, 杨任农, 张滢, 等. 基于启发式强化学习的空战机动智能决策[J]. 航空学报, 2017, 38(10): 212-225.
- [18] 李晨溪, 曹雷, 陈希亮, 等. 基于云推理模型的深度强化学习探索策略研究[J]. 电子与信息学报, 2018, 40(1): 244-248.
- [19] 张可, 郝文宁, 余晓晗, 等. 基于遗传模糊系统的兵棋推演关键点推理方法[J]. 系统工程与电子技术, 2020, 42(10): 2303-2311.

作者简介:

陈希亮, 男(1985—), 副教授, 研究方向为指挥信息系统工程及深度强化学习。

李清伟, 男(1987—), 高级工程师, 研究方向为指挥控制工程及智能空战博弈。

孙 或, 男(1993—), 助理工程师, 研究方向为多智能体深度强化学习。

(本文编辑: 马 岚)