doi: 10.7690/bgzdh.2023.09.018

作战 Agent 的学习算法研究进展与发展趋势

王步云，刘　聚
（海军大连舰艇学院作战软件与仿真研究所，辽宁 大连 116023）

摘要：针对作战 Agent 适应性问题，梳理遗传算法、强化学习、神经网络等方法在实现作战 Agent 适应性方面的成果，总结每种方法的特点；介绍深度强化学习方法在实现作战 Agent 适应性方面的应用情况，讨论深度强化学习在该方面应用的发展趋势和研究重点。该研究可为后续相关研究提供参考。

关键词：作战 Agent；适应性；强化学习；深度学习；神经网络
中图分类号：TP183　　**文献标志码：**A

Research Progress and Development Trend on Learning Algorithm of Combat Agent

Wang Buyun, Liu Ju
(*Operation Software and Simulation Institute*, *Dalian Naval Academy*, *Dalian* 116023, *China*)

Abstract: Aiming at the problem of combat Agent adaptability, this paper reviews the achievements of genetic algorithm, reinforcement learning, neural network and other methods in achieving combat Agent adaptability, and summarizes the characteristics of each method.It also introduces the application of deep reinforcement learning in achieving combat Agent adaptability and discusses the development trend and research focus of deep reinforcement learning in this area. This study can provide a reference for the follow-up study.

Keywords: combat Agent; adaptability; reinforcement learning; deep learning; neural network

0　引言

现代战争是典型的复杂适应系统，而基于多 Agent 建模仿真（agent-based modeling and simulation，ABMS）方法一直以来是研究这类系统的有效手段和方法。作战 Agent 是各类作战实体（层次、粒度可能不同）在仿真系统中的映射，也是 ABMS 的核心要素。Holland 的复杂适应系统理论（complex adaptive system，CAS）认为适应性造就复杂性，相对应的，在采用 ABMS 研究作战问题时，在系统微观层面上，要求 Agent 必须具备动态环境的自适应能力，即 Agent 能通过与环境和其他 Agent 的交互，不断积累经验，提高自己和所属团队在环境中的生存能力。应用机器学习技术赋予战争模拟系统中的智能 Agent 适应能力，始终是基于 CAS 理论的战争复杂性研究的基础内容之一，比如海战仿真中，舰艇 Agent 的许多决策内容都涉及到学习和适应，包括航路选择、协同防空、火力分配、目标搜索等。作战 Agent 需要通过自身知识的不断积累来逐步提高自身能力。

长期以来，很多学者已经围绕作战 Agent 的学习方法开展了卓有成效的研究。笔者重点梳理遗传算法、强化学习、神经网络等方法在这方面的成果，介绍了深度学习方法在实现作战 Agent 适应性方面的应用情况，并从智能博弈平台、学习算法构建、双方共同进化等方面讨论了深度强化学习在作战 Agent 学习领域的发展趋势和研究重点，以期起到抛砖引玉的作用。

1　多 Agent 作战建模与机器学习

基于多 Agent 作战建模仿真是利用 Agent 对作战复杂系统中各个实体构建模型，通过对作战 Agent 个体及其相互之间（包括与作战环境）的行为进行刻画，描述作战复杂系统的宏观行为。相对于传统面向过程和面向对象的仿真技术，基于多 Agent 仿真方法对复杂系统的行为具有更强的建模与仿真能力、更高的抽象性和表达能力、更强的仿真动态性和灵活性，实现了微观行为和宏观行为的有机结合，是研究作战复杂系统的重要手段[1]。

基于多 Agent 仿真是传统人工智能技术在仿真领域中的一个典型应用，因此在讨论 Agent 进化、Agent 学习时，往往绕不开人工智能中机器学习的概念。杨炳儒[2]认为：学习是系统中的任何改进，这种改进使得系统在重复同样的工作或进行类似的

收稿日期：2023-05-05；修回日期：2023-06-05
作者简介：王步云（1984—），男，湖北人，博士。

工作时，能完成得更好。机器学习就是要使计算机能够模拟人的学习行为，通过学习获取新知识和新技能，不断改善性能，实现自我进化，不断适应环境。

在传统机器学习理论中，虽然监督学习方法如归纳学习、类比学习、基于解释的学习等，在很多问题上取得了很好的效果，但很难完成作战 Agent 学习，因为监督学习需大量依靠历史样本数据，学习结果也缺乏实时性和灵活性，难以满足作战 Agent 所面临的复杂的、动态的、不确定性的环境要求。对于未知环境，作战 Agent 需要从交互中学习，而从交互中获得的样本并不能保证结果的正确性和代表性；因此，以监督学习为代表传统机器学习方法很难适用作战 Agent 的学习。

随着人工智能的不断发展和进步，机器学习先后引入了进化计算、神经网络方法以及基于试错方法、动态规划和瞬时误差方法形成了强化学习理论。这些不同于传统机器学习理论的方法，能够实现动态环境中的 Agent 学习，常用来探讨作战 Agent 学习问题。

2　常用作战 Agent 学习算法

2.1　遗传算法

遗传算法是模拟生物界“优胜劣汰”而产生的一种进化算法，已被广泛应用于各个领域。由于其原理与 Agent 演化原理类似，因此很自然地成为研究 Agent 学习的主要方法之一。采用遗传算法实现 Agent 进化的模式如图 1 所示。

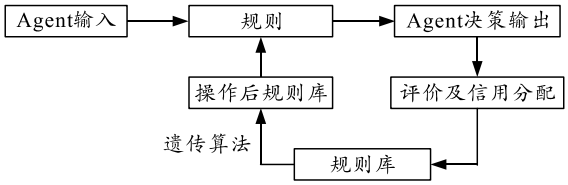

图 1　采用遗传算法实现 Agent 学习原理

文献[3]利用特定飞行动作的正反例样本建立飞行规则，条件是飞行动作的各类参数，结论是采取的飞行动作同时赋予每条规则一定适应度，在此基础上利用遗传算法对规则进行选择、交叉、变异等操作，实现了飞行规则自动获取和学习；文献[4]则利用智能 Agent 对无人机协同侦察实体进行模拟，采用遗传算法对侦查规则进行优化，为无人机的任务规划和优化提供支持；文献[5]设计了影响 Agent 决策的个性权重向量，并利用权重向量建立 Agent 的各类作战规则；在此基础上通过遗传算法对权重向量进行优化，实现了 Agent 的演化和学习，该方法在舰艇编队对海作战仿真中得到了成功应用；文献[6]建立了作战指挥 Agent 个性与行为规则的模型，使用遗传算法对个性参数进行了优化，进而实现对作战指挥 Agent 模型的优化，定量分析了指挥员指挥风格对作战过程和结果的影响。

应用遗传算法实现 Agent 学习的机制主要有 2 种：1) 采用进化计算建立 Agent 从“感知”到“动作”的映射规则，直接驱动执行机构产生相应的动作行为。其中最著名的一例是 John Holland 建立的 LCS 模型；2) 首先在 Agent 内部建立一个行为模型，然后采用进化计算驱动该模型，进而产生适应于环境的动作行为，例如采用有限状态机（finite state machine，FSM）建立 Agent 的行为模型，使用进化规划方法驱动行为模型的进化。

采用遗传算法进行 Agent 模型优化的难点在于：

1) 需要根据每种作战样式和作战任务的特点以及研究目的和需求，事前建立符合要求的 Agent 规则。遗传算法优化的是规则，研究对象不同，建立 Agent 规则不同；研究目的不同，也需要重新构建 Agent 规则。Agent 进化与 Agent 规则的紧耦合关系大大限制了算法的通用性。

2) 评价及信用分配方法对优化结果影响较大。对于 Agent 决策结果，如何对参与决策的每条规则进行合理地评价和信用分配一直是 Agent 学习的重难点问题；此外由于遗传算法本身的全局优化收敛性的理论分析尚未完全解决，在采用该方法实现 Agent 学习时还会存在学习效率较低、容易过早收敛和陷入局部最优解等不足。

2.2　强化学习法

强化学习是一种无师在线学习技术，它的基本特征是通过感知环境状态和从环境中获得不确定奖赏值来学习最优行为策略。与传统的监督学习不同，在强化学习中，没有“教师”直接给出最优解，Agent 得到的是环境对它行为的反馈，它根据这个反馈来调整自己的策略。强化学习的这种基本特点决定了它非常适用于最优策略不能预知的问题和未知的不确定的环境，因此该方法已经成为研究作战 Agent 适应性的主要方法之一。

另外，Agent 的行为不但会引起立即反馈，而且会改变环境，继而影响后继的反馈。试探交互和延迟反馈是强化学习 2 个最基本的特点。这种延迟

反馈、纠正错误的学习方式，对于战争系统中的智能 Agent 的学习过程是经常出现的。Agent 学习原理如图 2 所示。

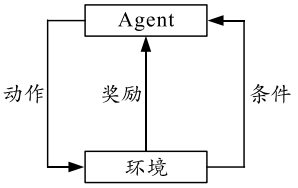

图 2 采用强化学习方法实现 Agent 学习原理

文献[7]采用强化学习的理论方法，对战争系统中指挥控制 Agent 的适应性机制进行研究。模型采用产生式规则实现指挥控制 Agent 的决策过程，但在决策规则中增加了选择权值向量用于反映 Agent 对各个决策结果的偏好程度，在此基础上根据环境反馈信息利用强化学习算法对选择权值进行修正和调整，以实现指挥 Agent 的适应性。结果表明，强化学习算法能够实现指挥控制 Agent 适应性，适合处理战争系统中不确定性和动态性环境的学习问题。

文献[8]分析了陆战 Agent 基于效果学习的本质特征，结合陆战 Agent 通信和指挥控制的特点，提出了基于知识共享的陆战 Agent PS(profit-sharing)强化学习机理模型。与一般的强化学习模型相比，该模型既解决了感知混淆和学习一致性问题，提高了学习效率，还可实现不同形式的知识共享，增强陆战 Agent 系统的整体学习和完成作战任务的能力。

文献[9]为了实现无人机航路自主规划，提出了基于多 Agent 强化学习理论的飞行路径规划算法。该算法利用分层强化学习方法，采用 2 个功能不同的 Agent，分别对应局部和全局路径规划，并对 Agent 的状态和动作空间进行划分和抽象，减少状态数量，解决强化学习的维数灾难问题，取得了较好效果。

文献[10]将多分辨率与多 Agent 理论相结合，战术级任务建立高分辨率 Agent，利用 OA 环进行协作，战略级任务建立分辨率聚合 Agent，并采用基于强化学习的方法进行决策。在具体过程中，考虑 Agent 数量、毁伤数量以及伤亡增加率等因素，采用 Q-learning 算法，构建了初始 Q 函数和瞬时奖赏函数，用以决策 Agent 是防守或是进攻。

强化学习虽然提供了一个通用框架和一组方法[11]，但对状态和动作的描述使得它在应用到复杂的现实世界中时非常困难，需要手动对状态特征进行建模，特别是对作战这类复杂系统，对状态特征建模往往非常困难；同时由于作战 Agent 的战场环境涉及变量多、复杂；因此，需要大量的参数进行描述，这样又会引起状态空间到动作空间的组合爆炸，给搜索带来繁重的任务。此外作战 Agent 都属顺序型任务，并且多个 Agent 之间往往还存在协同关系，在这种情况下，如何设计作战 Agent 的分配奖惩一直是强化学习应用的难点问题。

2.3 神经网络法

神经网络是在理解和抽样人脑结构和外界刺激响应机制后，以网络拓扑知识为理论基础，模拟人脑的神经系统对复杂信息处理机制的一种数学模型。如图 3 所示。

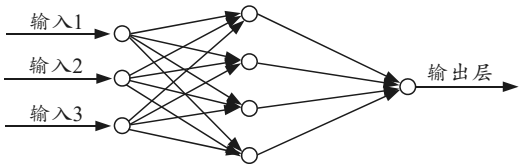

图 3 采用神经网络方法实现 Agent 学习原理

在应用神经网络学习时，首先根据样本信息，通过比对输出信息与样本信息的差异，不断修正神经元之间的权重，使之能够根据输入数据计算出期望值；再将实际得到的数据代入到训练好的神经网络中进行计算，实现一系列的功能。随着神经网络的不断发展，很多改进神经网络模型被陆续提出，如 BP 神经网络、多层神经网络、遗传神经网络等，在模式识别、图像处理、工程优化、金融预测与管理、通信等领域得到了广泛应用。

神经网络具有良好的学习能力，拓扑结构灵活、可实现高度并行计算、模拟人脑神经系统，具有非线性处理能力，这使得神经网络在作战仿真领域有着很好的应用前景。而在作战 Agent 学习方面，文献[12]以多机编队协同空战作为应用背景，综合运用神经网络、影响图等方法，在能力、信念和优先学习的基础上构建 Agent 模型，其学习方法是利用其他 Agent 的历史行为作为训练集，利用神经网络、决策知识和专家知识来修改影响图中各节点的连接关系，进而实现 Agent 的进化和学习。

文献[13]在基于效用的行为选择模型基础上对多 Agent 系统中个性建模问题进行研究，结合心理学中个性的五因素模型建立 Agent 个性神经网络，通过不同参数反映个性对效用变化的影响方式，具有更强的个性表征能力，设计梯度下降的学习算法训练 Agent 相应的个性神经网络。

文献[14]以 2V1 反隐身超视距空战路径规划为例，提出了基于神经网络和人工势场的协同博弈路径规划方法。该方法首先采用人工势场方法构建 Agent 决策模型，实现路径规划，在此基础上使用 BP 神经网络自适应调整人工势场函数，实现 Agent 路径学习。在 BP 神经网络训练过程中，为保证训练质量和数量，采用遗传算法和滚动时域优化来生成仿真样本。

神经网络可以记忆已知的信息，还具有较强的概括能力和联想记忆能力，较少依赖先验知识。由于神经网络的推理知识表示体现在网络连接权值上，表达难以理解，因此训练完成后模型被认为是黑盒，展现出的规律、特征还需要人为去总结；另外，模型的训练效率、质量通常还依赖样本质量和计算机硬件能力。

总之，无论是遗传算法、强化学习和神经网络都有各自特点和优势，很多学者试图融合多种方法解决问题，如文献[15−16]的遗传算法与神经网络、神经网络与强化学习的结合研究等。随着计算机性能的飞速提升，特别是 Alpha Go 在围棋领域取得突破，融合了神经网络、强化学习的深度学习方法得到了广泛应用，目前已成为作战 Agent 学习最为关注的一种实现方法。

3　深度学习方法

深度学习是在强化学习基础上，使用神经网络代替值函数模型或策略模型，然后通过强化学习算法训练算法模型，指导 Agent 决策。深度学习本质上是一系列反复训练调优的神经元网络，其训练过程就是整合理解的过程[17]。原理如图 4 所示。

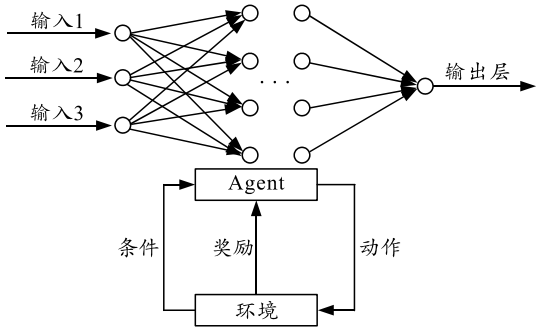

图 4　采用深度学习方法实现 Agent 学习原理

相对于传统的学习算法，深度强化学习方法本质上是“大数据+高性能计算+强化神经网络”，不仅具备自学习能力，而且具有一定的通用性，通用性意味着对解决其他问题极具参考价值，这使得以深度学习为代表的各类智能算法在目标识别[18]、态势理解[19]、辅助决策[20]等军事领域有了成功应用。

在应用深度强化模型实现作战 Agent 学习方面，近期涌现了大批成果。文献[21]针对指控系统复杂性、非线性，态势感知多源性、异构性问题，采用深度强化学习技术设计具有决策能力的作战智能体，通过多模态深度神经网络感知并处理多源异构态势，利用基于价值−策略的自学习方法不断优化智能体决策效果；在此基础上采用不同场景对作战智能体进行训练，利用作战智能体的自学习能力，在人类经验基础上、人为引导的前提下探索新战法。

文献[22−24]分别采用强化学习研究了战斗机、UCAV 等的指挥决策问题，主要思路是首先对决策空间进行离散化，并利用不同的神经网络模型对输入进行处理，最后输出决策动作。结果表明，强化学习算法不仅能够控制飞机导航决策学习，而且能为优化飞机突防、战斗机作战等决策提供手段。

文献[25−26]则分别博弈平台的角度研究了深度强化学习的应用，将人工智能技术集成到计算机兵棋推演和作战模拟仿真系统中形成智能博弈对抗系统，提出一种适应于军事博弈对抗系统智能应用研究的通用设计框架，适用于 AI 智能体实现跨系统、跨平台的泛化应用，为指挥训练提供技术支撑。

4　发展趋势

随着计算机算力不断提升以及深度学习的不断发展，应用深度学习技术解决作战 Agent 学习问题必将逐渐走向深入，为开展军事复杂系统进化仿真，进而为战法创新、装备需求论证等提供全新的技术支撑，但为了达成上述目的，还需要重点解决如下问题。

4.1　智能博弈平台

为便于各种智能技术在作战仿真领域的落地并走向实用，需要一个统一的平台为作战 Agent 的学习训练提供环境。该平台除了具备传统军事仿真平台的想定编辑、模型开发、红蓝对抗、仿真运行、态势显示、评估分析等功能外，还应具备：

1) 学习 Agent 的接入。仿真平台能够为学习 Agent 提供作战态势数据，并且识别 Agent 的动作输出，驱动仿真平台内对应的仿真实体运行。

2) 内嵌智能学习算法库。平台应提供丰富算法库供仿真人员选择或组合使用，以便平台使用人员将主要精力用于作战问题的研究；支持按需增加、

修改、管理各种算法。

3) 支持并行分布式仿真，学习算法的基础是海量大数据，仿真平台必须具备高效仿真能力，支持并行分布式仿真运行，以满足学习所需的样本数据。

4) 实现 Agent 学习。智能博弈平台的本质特点是智能学习，平台能够利用仿真产生样本数据，结合智能学习算法，完成学习 Agent 的训练，进而实现 Agent 决策的优化。

4.2 学习算法构建

深度学习只是提供了一个机器学习的通用框架，并不是一个拿来即用的算法，需要根据实际的应用场景，对模型和算法进行设计。作战是一个非完全信息条件下对抗博弈过程，特别是现代信息化条件，对手隐真示假，战争迷雾重重。对于各类学习模型，一方面是对学习算法本身的研究，包括合理地确定态势输入及动作输出空间需要反复思考和研究；另一方面是如何根据作战样式、作战规模等选择合理的学习模型也是仿真过程中需要反复试验和比较的。

此外，学习效果评估是检验学习算法有效性的基础。具体可以从 2 方面开展评估：1) 从仿真结果看，类似围棋的对弈结局，对比学习前后的仿真结果，检验学习算法的有效性；2) 从仿真过程看，观察仿真每步决策行为，通过建立合理的指标体系，划分不同智能等级，定量评估仿真模型的进化水平。

4.3 共同进化

现代战争是复杂适应系统，作战双方中的各“主体”均具有适应性，能够根据作战态势的变化、对方编队系统的变化调整自己的行为方式。正是这样的原因，使得战争具有典型的共同演化特征。在基于传统建模仿真方法对作战过程仿真研究中，实体的决策模型往往采用了固定规则，很少考虑到实体行为的演化；后来随着多 Agent 仿真方法的提出，人们开始日益重视 Agent 适应性对作战仿真过程及其结果的影响，也逐步加强了对其研究，取得了丰硕的成果；但是为了便于比较，现有的研究往往只实现某一方 Agent 的适应性，作为对抗中的另一方仍然采用了固定规则的决策模式。在仿真中如何实现红蓝双方的同时学习，这种方式是否可以收敛，对战仿真结果又有何影响都是可进一步深入研究的问题。

5 结束语

自适应性是 Agent 的特性之一，也是采用 ABMS 方法开展军事系统仿真关注的基础问题。随着科学技术的不断发展，以深度强化学习为代表的新一代人工智能技术在实现 Agent 适应性方面具有一定优势，下一步将围绕智能博弈平台构建、智能学习算法开展探索和研究，为实现作战 Agent 自适应提供支撑。

参考文献：

[1] 胡晓峰, 司光亚. 战争复杂系统建模与仿真[M]. 北京: 国防大学出版社, 2006: 3.

[2] 杨炳儒. 知识工程与知识发现[M]. 北京: 冶金工业出版社, 2000: 48.

[3] 胡飞, 徐浩军, 曹登高. 遗传算法在产生式规则获取中的应用[J]. 电光与控制, 2006, 13(3): 87−96.

[4] 杜健健. 多无人机协同侦察任务规划器的设计与实现[D]. 南京: 南京航空航天大学, 2019.

[5] 王步云, 张国. 一种适用于人工生命作战仿真的混合 Agent 结构[J]. 系统仿真学报, 2010, 22(11): 2515−2518.

[6] 杜伟. 基于 Agent 个性的指挥风格研究[J]. 火力与指挥控制, 2018, 43(8): 37−41.

[7] 李志强, 胡晓峰, 张斌, 等. 基于强化学习的指挥控制 Agent 适应性仿真研究[J]. 系统仿真学报, 2005, 17(11): 2801−2804.

[8] 韩月敏, 林燕, 刘非平, 等. 陆战 Agent 学习机理模型研究[J]. 指挥控制与仿真, 2010, 32(1): 13−17.

[9] 李东华, 江驹, 姜长生. 多智能体强化学习飞行路径规划算法[J]. 电光与控制, 2009, 16(10): 10−14.

[10] 闫雪飞, 李新明, 刘东, 等. 基于多分辨率的 multi-Agent 武器装备体系作战仿真研究[J]. 系统仿真学报, 2017, 29(1): 136−143.

[11] 高阳, 陈世福, 陆鑫, 等. 强化学习研究综述[J]. 自动化学报, 2004, 30(1): 86−100.

[12] 钟麟, 陈丽娟, 佟明安, 等. 基于影响图的多智能体学习算法[J]. 系统工程学报, 2008, 23(3): 377−380.

[13] 肖正, 张世永. 基于神经网络的 Agent 个性化行为选择[J]. 计算机工程, 2009, 24(12): 199−201.

[14] 张菁, 何友, 彭应宁, 等. 基于神经网络和人工势场的协同博弈路径规划[J]. 航空学报, 2019, 40(3): 322−493.

[15] 熊辉, 赵英凯, 丁瑶君. 基于神经网络的遗传算法优化及其应用[J]. 南京化工大学学报, 2000, 22(7): 21−24.

[16] 唐亮贵, 刘波, 唐灿, 等. 基于神经网络的 Agent 增强学习模型[J]. 计算机科学, 2007, 34(11): 156−159.

（下转第 96 页）