Journal of System Simulation

Volume 36 | Issue 1 Article 12

1-20-2024

Combat Effectiveness Evaluation Method of Homogeneous Cluster Equipment System Based on RLoMAG+EAS

Guohui Zhang
Department of Information and Communication, Academy of Army Armored Force, Beijing 100072, China, zgh8002@126.com

Ang Gao
Joint Operations College, National Defence University, Beijing 100091, China, 15689783388@163.com

Ya'nan Zhang
Department of Information and Communication, Academy of Army Armored Force, Beijing 100072, China

Combat Effectiveness Evaluation Method of Homogeneous Cluster Equipment System Based on RLoMAG+EAS

Abstract

Abstract: The equipment system is the reflection of the combat system from the perspective of equipment. The research on the combat effectiveness evaluation of the equipment system is of great practical significance for the optimization, construction, and development of the equipment system. Cluster equipment combat system confrontation is characterized by large-scale, highly dynamic and strong confrontation, and it is difficult to directly evaluate combat effectiveness with traditional methods. Aiming at the single task homogeneous cluster equipment system (such as UAV reconnaissance swarm and ground unmanned platform fire assault cluster), this paper regards the confrontation process of equipment system as the Markov game process of multi-agent system from the perspective of multi-agent game theory. The combat effectiveness evaluation method of equipment system based on reinforcement learning of multi-agent game (RLoMAG) is proposed. Firstly, the principle of evaluation method is analyzed and the model of equipment system confrontation is built. Secondly, the framework of the combat effectiveness evaluation method of the equipment system is given, including conducting the agent modeling, game algorithm design and combat effectiveness index design of the equipment system, carrying out exploratory system confrontation simulation, solving the game optimal strategy of the equipment system, and analyzing the combat effectiveness index of the equipment system under the optimal strategy. Finally, based on the base defense combat scenario, an application example of the combat effectiveness evaluation method of the UAV swarm equipment system is given to verify the effectiveness of the method.

Keywords

equipment system, combat effectiveness evaluation, reinforcement learning of multi-agent game, optimum strategy

Recommended Citation

Zhang Guohui, Gao Ang, Zhang Ya'nan. Combat Effectiveness Evaluation Method of Homogeneous Cluster Equipment System Based on RLoMAG+EAS[J]. Journal of System Simulation, 2024, 36(1): 160-169.

第 36 卷第 1 期 系统仿真学报© Vol. 36 No. 1
2024 年 1 月 Journal of System Simulation Jan. 2024

基于RLoMAG+EAS的同构集群装备体系作战效能评估方法

张国辉[1]，高昂[2*]，张雅楠[1]
(1. 陆军装甲兵学院 信息通信系，北京 100072；2. 国防大学 联合作战学院，北京 100091)

摘要：装备体系是作战体系在装备视角的反映，研究装备体系作战效能评估问题，对装备体系优化、建设发展具有重要现实意义。集群装备作战体系对抗，具有大规模、高动态、强对抗特点，传统方法难以对其作战效能直接进行评估，*针对单一任务同构集群装备体系（如无人机侦察蜂群、地面无人平台火力突击集群等），从多智能体博弈理论的视角，将装备体系对抗过程看作为多智能体系统马尔可夫博弈过程，提出一种基于多智能体博弈强化学习（reinforcement learning of multi-agent game，RLoMAG）的装备体系作战效能评估方法。分析了评估方法原理，建立了装备体系对抗模型。给出了装备体系作战效能评估方法框架，包括智能体建模、博弈算法设计、装备体系作战效能指标设计，开展探索性体系对抗仿真，求解装备体系博弈最优策略，分析最优策略下的装备体系作战效能指标等步骤。以基地防御作战场景为背景，给出了无人机蜂群装备体系作战效能评估方法应用示例，验证了方法的有效性。*

关键词：装备体系；作战效能评估；多智能体博弈强化学习；最优策略
中图分类号：TN94; E919; TP391.9　文献标志码：A　文章编号：1004-731X(2024)01-0160-10
DOI: 10.16182/j.issn1004731x.joss.22-1034
引用格式: 张国辉, 高昂, 张雅楠. 基于RLoMAG+EAS的同构集群装备体系作战效能评估方法[J]. 系统仿真学报, 2024, 36(1): 160-169.
Reference format: Zhang Guohui, Gao Ang, Zhang Ya'nan. Combat Effectiveness Evaluation Method of Homogeneous Cluster Equipment System Based on RLoMAG+EAS[J]. Journal of System Simulation, 2024, 36(1): 160-169.

Combat Effectiveness Evaluation Method of Homogeneous Cluster Equipment System Based on RLoMAG+EAS

Zhang Guohui[1], *Gao Ang*[2*], *Zhang Ya'nan*[1]
(1. Department of Information and Communication, Academy of Army Armored Force, Beijing 100072, China;
2. Joint Operations College, National Defence University, Beijing 100091, China)

Abstract: The equipment system is the reflection of the combat system from the perspective of equipment. The research on the combat effectiveness evaluation of the equipment system is of great practical significance for the optimization, construction, and development of the equipment system. Cluster equipment combat system confrontation is characterized by large-scale, highly dynamic and strong confrontation, and it is difficult to directly evaluate combat effectiveness with traditional methods. *Aiming at the single task homogeneous cluster equipment system (such as UAV reconnaissance swarm and ground unmanned platform fire assault cluster), this paper regards the confrontation process of equipment system as the Markov game process of multi-agent system from the perspective of multi-agent game theory. The combat effectiveness evaluation method of equipment system based on reinforcement learning of multi-agent game (RLoMAG) is proposed. Firstly, the principle of evaluation method is*

收稿日期：2022-09-01　修回日期：2023-05-14
第一作者：张国辉(1980-)，男，副教授，博士，研究方向为智能指挥决策。E-mail：zgh8002@126.com
通讯作者：高昂(1988-)，男，讲师，博士，研究方向为作战仿真。E-mail：15689783388@163.com

analyzed and the model of equipment system confrontation is built. Secondly, the framework of the combat effectiveness evaluation method of the equipment system is given, including conducting the agent modeling, game algorithm design and combat effectiveness index design of the equipment system, carrying out exploratory system confrontation simulation, solving the game optimal strategy of the equipment system, and analyzing the combat effectiveness index of the equipment system under the optimal strategy. Finally, based on the base defense combat scenario, an application example of the combat effectiveness evaluation method of the UAV swarm equipment system is given to verify the effectiveness of the method.

Keywords: equipment system; combat effectiveness evaluation; reinforcement learning of multi-agent game; optimum strategy

0 引言

装备体系作战效能概念的实质含义是指装备体系实现特定任务目标的有效程度，即在给定威胁、条件、环境和作战方案下装备体系完成作战任务效果的度量[1-3]。装备体系是由多个装备相互协作、执行共同任务的统一系统，有着强大的群体协作能力，可以协同完成复杂的任务和目标[4-5]。体系对抗仿真是装备体系作战效能评估的重要手段，包括建立指标体系、选择评估方法、评估数据采集、评估模型解算、形成评估结论等步骤。目前，装备体系作战效能评估方法主要包括数学解析、复杂网络[6]、作战环[7]、探索性分析[8]、建模仿真[9]等方法。针对大规模、高动态、强对抗作战场景下的装备体系作战效能评估问题，数学解析、复杂网络类方法，难以反映体系的动态演化特性；作战环类方法，在体系规模较大时，OODA建模分析难度大，多数研究也都集中在体系的静态描述上；探索性分析，应用的难点在于给出具体的方法形式；建模仿真类方法，难点在于虚拟实体行为建模的复杂性。针对以上问题，本文介绍了一种基于多智能体博弈强化学习的装备体系作战效能仿真评估方法，包括评估方法原理、框架、应用示例，适用于典型场景、关键作战行动等背景下的由单一功能无人平台组成的任务集群的装备体系作战效能评估问题研究，例如无人机打击蜂群、地面无人平台火力突击集群的装备体系作战效能评估等。

1 方法原理

装备体系，是由功能相联、性能互补的装备系统，依据作战使命任务组成的更高层次的系统[3]，例如防空装备体系、空中突击装备体系。多智能体博弈强化学习(reinforcement learning of multi-agent game, RLoMAG)方法，将博弈论引入强化学习，优点是在很好地描述多智能体间关系的同时，可以解释收敛点对应策略的合理性，以及可以用纳什均衡解来替代最优解以求解有效策略[10]。RLoMAG方法，将装备体系中的各类装备平台，看作多智能体系统(multi agent systems, MAS)中的智能体(agent)，体系对抗过程看作为马尔可夫博弈(Markov game, MG)过程模型，以装备体系能够发挥的最大作战效能为学习对象，通过设计智能体的状态空间、收益函数、行为空间以及智能博弈算法，在体系探索最优策略的过程中，形成智能体间的复杂交互，模拟出体系的适应性、不确定性、非线性、涌现性特征。将装备体系的编组、规模、功能等作为变量，结合探索性分析仿真(exploratory simulation analysis, EAS)方法，通过探索博弈模型的最优策略，演化出大量的作战过程，分析装备体系在达到最优策略情况下的作战效能指标(例如参战双方战损比)，完成装备体系作战效能评估工作。

1.1 基于博弈的体系对抗过程

装备体系内部主要是合作与竞争关系，体系

DOI: 10.16182/j.issn1004731x.joss.22-1034

第 36 卷第 1 期 系统仿真学报 Vol. 36 No. 1
2024 年 1 月 Journal of System Simulation Jan. 2024

间为对抗关系。因此，装备体系对抗仿真过程可以建模为离散时间非合作马尔可夫博弈模型七元组：

$$(\boldsymbol{S}, \boldsymbol{SoS}_1, \boldsymbol{SoS}_2, \boldsymbol{R}_{SoS_1}, \boldsymbol{R}_{SoS_2}, p, \gamma)^{[11]},$$

其中，$\boldsymbol{S}$为当前博弈的状态空间；$\boldsymbol{SoS}_1$为体系1，$\boldsymbol{SoS}_1=\{\boldsymbol{A}_i\}_{i=1}^{N}$；$\boldsymbol{SoS}_2$为体系2，$\boldsymbol{SoS}_2=\{\boldsymbol{A}_j\}_{j=1}^{M}$；$\boldsymbol{R}_{SoS_1}$为体系1的奖励函数矩阵，$\boldsymbol{R}_{SoS_1}=\{r_i\}_{i=1}^{N}$；$\boldsymbol{R}_{SoS_2}$为体系2的奖励函数矩阵，$\boldsymbol{R}_{SoS_2}=\{r_j\}_{j=1}^{M}$。

$\boldsymbol{A}_i$为体系1中，智能体i的动作空间，其中，$\boldsymbol{A}_{SoS_1}=\boldsymbol{A}_{i=1}\times\boldsymbol{A}_{i=2}\times\cdots\times\boldsymbol{A}_{i=N}$表示$N$个智能体联合动作空间，$i\in[1,N]$；$\boldsymbol{A}_j$为体系2中，智能体$j$的动作空间，其中，$\boldsymbol{A}_{SoS_2}=\boldsymbol{A}_{j=1}\times\boldsymbol{A}_{j=2}\times\cdots\times\boldsymbol{A}_{j=M}$表示$M$个智能体联合动作空间，$j\in[1,M]$；

$$r_i:\boldsymbol{S}\times\boldsymbol{A}_{i=1}\times\cdots\times\boldsymbol{A}_{i=N}\times\boldsymbol{A}_{j=1}\times\cdots\times\boldsymbol{A}_{j=M}\to\boldsymbol{R}_{SoS_1},$$

$$r_j:\boldsymbol{S}\times\boldsymbol{A}_{i=1}\times\cdots\times\boldsymbol{A}_{i=N}\times\boldsymbol{A}_{j=1}\times\cdots\times\boldsymbol{A}_{j=M}\to\boldsymbol{R}_{SoS_2},$$

分别为智能体i、j的奖励函数，定义了智能体i、j的瞬时奖励；在体系对抗仿真中，瞬时奖励通常映射为每仿真步长内，装备平台战斗得分与战损失分的函数。$p:\boldsymbol{S}\times\boldsymbol{A}_{i=1}\times\cdots\times\boldsymbol{A}_{i=N}\times\boldsymbol{A}_{j=1}\times\cdots\times\boldsymbol{A}_{j=M}\to\Delta(\boldsymbol{S})$，为状态转移函数，定义了给定任意联合动作$\boldsymbol{a}\in\boldsymbol{A}_{SoS_1}\times\boldsymbol{A}_{SoS_2}$，从任意状态$\boldsymbol{s}\in\boldsymbol{S}$到任意状态$\boldsymbol{s'}\in\boldsymbol{S}$的状态转移概率，表征了状态随时间的随机演化，$\Delta(\boldsymbol{S})$为状态空间$\boldsymbol{S}$上的概率分布的集合；$\gamma\in[0,1)$为常数，用来表示随时间变化的奖励折现因子。

1.2 最优策略下的体系评估

(1) 状态值函数与博弈收益函数的映射

在时间步长t时刻，所有智能体同时采取动作，智能体i因采取动作$a_{i,t}$而获得的瞬时奖励为$r_{i,t}$。智能体i根据策略π_i选择当前状态下的动作，即$\pi_i:\boldsymbol{S}\to\Delta(\boldsymbol{A}_i)$，$a_{i,t}\sim\pi_i(\cdot|s_t)$，$\Delta(\boldsymbol{A}_i)$为智能体$i$动作空间概率分布的集合。

$\pi=(\pi_{SoS_1},\pi_{SoS_2})$表示全部智能体的联合策略，智能体$i$的状态值函数$v_{i,\pi}(s)$定义为$i$在策略$\pi$下的累积折扣期望奖励，即

$$v_{i,\pi}(s)=\sum_{t=0}^{\infty}\gamma^t E_{\pi,p}[r_{i,t}|s_0=s,\pi] \tag{1}$$

根据式(1)，贝尔曼方程，将智能体i的动作值函数

$$Q_{i,\pi}:\boldsymbol{S}\times\boldsymbol{A}_{i=1}\times\cdots\times\boldsymbol{A}_{i=N}\times\boldsymbol{A}_{j=1}\times\cdots\times\boldsymbol{A}_{j=M}\to\boldsymbol{R}_{SoS_1}$$

写为

$$Q_{i,\pi}(s,a)=r_i(s,a)+\gamma E_{s'\sim p}[v_\pi^i(s')] \tag{2}$$

同时，$v_{i,\pi}(s)$可以写为

$$v_{i,\pi}(s)=E_{a\sim\pi(s)}[Q_{i,\pi}(s,a)] \tag{3}$$

定义体系1的博弈收益函数为

$$U_{\pi_{SoS_1},\pi_{SoS_2}}^{SoS_1}=\sum_{i=1}^{N}v_{i,\pi}(s) \tag{4}$$

智能体j同理。

(2) 最优策略下的作战效能指标

体系博弈过程中，如果存在联合策略$\pi^*=(\pi_{SoS_1}^*,\pi_{SoS_2}^*)$，使得任意$\boldsymbol{s}\in\boldsymbol{S}$，$\pi_{SoS_1}$满足$U_{\pi_{SoS_1}^*,\pi_{SoS_2}^*}^{SoS_1}(s)\geqslant U_{\pi_{SoS_1},\pi_{SoS_2}^*}^{SoS_1}$，则$\pi^*$为最优策略。

这里假设装备体系1为所要研究的对象，装备体系2为其对手体系。在装备体系对抗仿真中，固定装备体系1面临的对手威胁(包括$\pi_{SoS_2}^*$)、作战环境、作战行动方案，基于探索性仿真分析思想，分别在不同装备体系设置条件下开展仿真实验，并计算装备体系1最优策略条件下的作战效能指标。这是因为$\pi_{SoS_2}^*$给定后，装备体系1在最优策略$\pi_{SoS_1}^*$下取得的收益值最大，可以认为装备体系1所能发挥的作战效能最大，完成任务的效果最好，作战效能指标值也最优。例如，将装备体系1与装备体系2对抗的战损比作为衡量装备体系1作战效能的指标，那么，装备体系1在最优策略下的战损比，是装备体系1在所能发挥的最大作战效能下达到的最低战损比。装备体系1与装备体系2的对抗博弈如表1所示，灰色表格为最优策略点。

2 RLoMAG+EAS模型

探索性分析是美国兰德公司针对不确定条件下复杂高层问题研究提出的分析方法，其基本思

表 1 体系1与体系2的对抗博弈

Table 1 SoS_1 versus SoS_2

SoS_2	SoS_1			
	$\pi_{SoS_2}^{(1)}$	…	$\pi_{SoS_2}^{*}$	…
$\pi_{SoS_1}^{(1)}$				
…				
$\pi_{SoS_1}^{*}$			$(U^{SoS_1}_{\pi^{*}_{SoS_1},\pi^{*}_{SoS_2}}, U^{SoS_2}_{\pi^{*}_{SoS_1},\pi^{*}_{SoS_2}})$	
…				

想是在多分辨率模型基础上，通过对各种不确定要素所产生的结果进行整体研究，在较高层次模型上进行相关探索。装备体系作战效能评估采用探索性仿真分析与智能博弈算法相结合的思路，一方面使用RLoMAG类博弈算法，求解体系对抗过程博弈模型的最优策略；另一方面利用探索性仿真分析思路，在不同装备作战体系设定条件的体系对抗过程博弈模型最优策略下，分析作战效能指标变化情况。探索性分析方法包括问题分析、探索因素分析、仿真建模、仿真实验、结果分析、得出结论6步。本文以体系对抗过程博弈模型为基础，EAS为模型基本框架，通过对装备作战体系涉及的想定空间中不确定因素分析，研究不同因素条件下的装备作战体系最大效能。具体EAS+RLoMAG模型框架如图1所示。

step 1：问题分析。明确探索性分析的研究目标，例如研究典型作战场景下，作战体系1的作战效能；固定与问题分析相关的基础信息，例如，固定威胁方(装备作战体系2)、作战环境、作战方案，表1中装备作战体系2的策略π_{SoS_2}等。

step 2：探索因素分析。找出可能对问题结果有较大影响的不确定性因素，并分析各个不确定性因素可能的取值范围。例如，将作战体系1的作战编组、规模、战技术性能等作为探索因素。

step 3：仿真建模。智能体建模，包括状态空间、动作空间、收益函数设计，以及RLoMAG博弈类算法设计，将各种不确定性因素与研究目标联系起来。

step 4：仿真实验。设计装备体系作战效能指标，根据体系对抗过程博弈模型进行探索性计算，探索各种不确定性因素导致的系统结果。这里的探索性计算是求解装备体系1博弈最优策略。

step 5：结果分析。分析不确定性因素与结果的关系，即分析装备体系1最优策略下的装备体系作战效能指标，便可以对装备体系1能够发挥的最大作战效能进行评估。判断装备体系1博弈最优策略下的作战效能指标值是否可以接受。如果是，输出装备体系作战效能指标；如果否，调整装备体系1的设置条件，如装备编组、规模、功能等，返回step 1。

step 6：得出结论。根据分析结果，提出系统优化的建议或给出适应问题不同条件的措施。

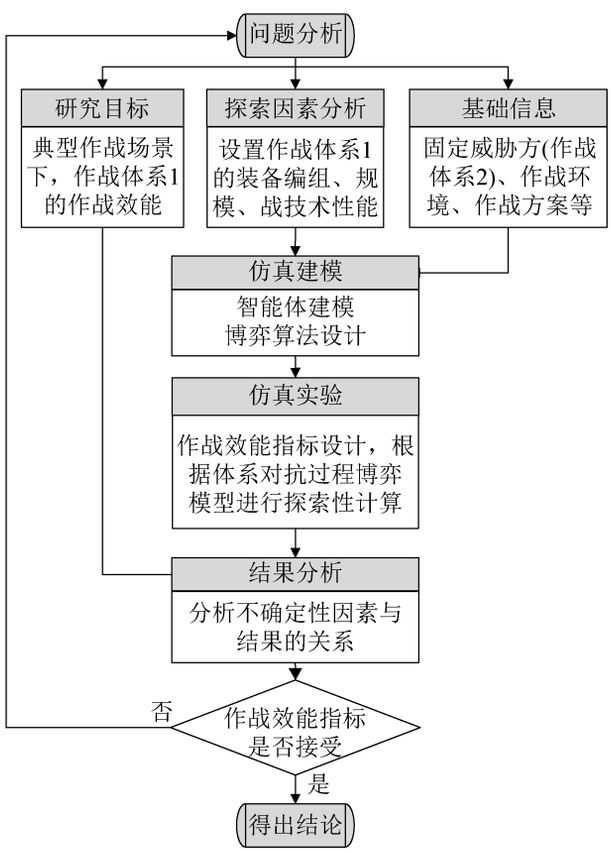

图 1 EAS+RLoMAG模型框架

Fig. 1 Framework of EAS+RLoMAG model

3 应用示例

本节为EAS+RLoMAG模型的应用示例，以基地防御作战场景为例，应用所提方法，按照问

题分析、探索因素分析、仿真建模、仿真实验、结果分析与得出结论的步骤，分析无人机蜂群装备体系作战效能。

3.1 问题分析

基地防御具体作战场景为，蓝方发射280架程序预置UAVS空袭红方地空导弹阵地，红方发射UAVS拦截，分析红方UAVS的体系作战效能，规划红方需要发射的最少UAVS数量。红蓝双方拥有完全相同的飞行约束，UAV相撞或被击毁则退出环境，蓝方目标为进入导弹阵地，红方目标为阻止蓝方入侵。在战损比，作战时长约束条件下，蓝方剩余UAV数量小于或等于1架，无法形成战斗力，红方获胜。UAVS对地空导弹阵地打击方式为饱和式攻击，且UAVS的飞行高度较低，不在地空导弹阵地的拦截范围之内。因此，可忽略蓝方地空导弹阵地的威胁范围对UAVS体系对抗过程的影响。UAV属性如表2所示。蓝方采用程序预先设定的UAVS，其体系产生的策略$\pi^{*}_{SoS_2}$作为baseline与红方对抗。

表2 UAV属性值设置
Table 2 Settings of UAV attribute values

属性	值
生命值	10
速度	2
探测范围	6
探测角度/(°)	150
攻击范围	1.5
攻击力值	2
作战范围	1 000×1 000

3.2 探索因素分析

假设红方无人机蜂群装备型号、作战行动方案是确定的，因此，影响装备体系作战效能发挥的装备战技术性能、装备编组等体系设置条件是相对固定的，这里只将装备数量规模作为作战体系1作战效能的探索因素。

3.3 仿真建模

仿真建模包括智能体建模、RLoMAG博弈类算法设计两部分，目的是将各种不确定性因素与研究目标联系起来。

3.3.1 智能体建模

智能体建模包括状态空间、动作空间、收益函数设计3部分。

(1) 状态空间设计

状态空间分别从空间、时间维度设计，如图2所示。假设旋翼无人机为1×1的矩形，毁伤半径为l，探测半径为m，以1×1为单位，将UAVS作战区域$n\times n$栅格化，如图2所示。进而探测范围近似表示为$2m\times 2m$的黄色区域，毁伤范围为$2l\times 2l$的红色区域，m，n，l为正整数。

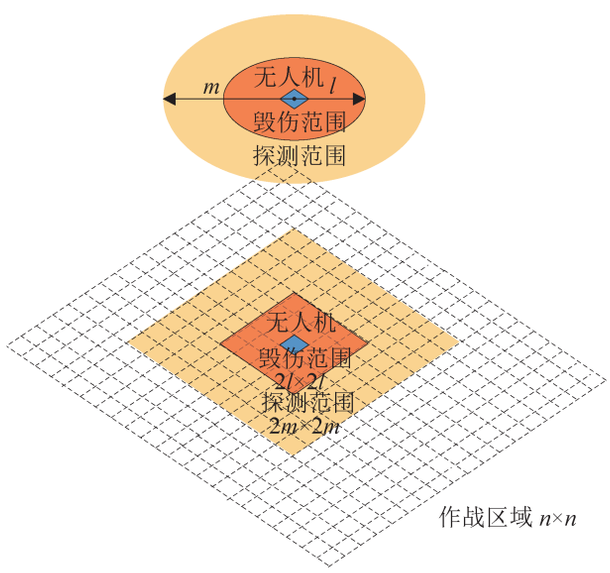

图2 UAV作战区域、毁伤范围、探测范围栅格化示意图
Fig. 2 Schematic diagram of UAV combat area, damage range, detection range grid

UAVS联合状态空间定义为$S=\{s_i\}_{i=1}^{N}$，其中，$s_i=(\boldsymbol{W}^0_{m\times m},\boldsymbol{U}^0_{m\times m},\boldsymbol{W}^1_{m\times m},\boldsymbol{U}^1_{m\times m},\boldsymbol{V}^0_{(n/N)\times(n/N)},\boldsymbol{V}^1_{(n/N)\times(n/N)},a^{\mathrm{ID}}_{\mathrm{step}-1},r^{\mathrm{ID}}_{\mathrm{step}-1},pos^{\mathrm{ID}}_{\mathrm{step}-1})$

空间维度包括$\boldsymbol{W}^0_{m\times m}$，$\boldsymbol{U}^0_{m\times m}$，$\boldsymbol{W}^1_{m\times m}$，$\boldsymbol{U}^1_{m\times m}$，$\boldsymbol{V}^0_{(n/N)\times(n/N)}$，$\boldsymbol{V}^1_{(n/N)\times(n/N)}$6个特征矩阵，所表示的含义如下：

① $\boldsymbol{W}^0_{m\times m}$表示$m\times m$探测范围内己方UAV位置矩阵，栅格与矩阵中的元素对应，如果栅格$m_0\times m_1$内存在己方UAV，则$w^0_{m_0\times m_1}=1$，否则

$w^0_{m_0\times m_1}=0$，m_0，m_1 为正整数，且 $1\leqslant m_0,m_1\leqslant m$。$\boldsymbol{W}^0_{m\times m}$ 的设计考虑如下：每架UAV能够探测到己方UAV的位置，是UAVS根据敌情适时进行队形变换、进攻、防御，并随时保持在相对于敌方最佳的位置进行协同战斗的基础。

② $\boldsymbol{U}^0_{m\times m}$ 表示 $m\times m$ 范围内己方UAV毁伤状态矩阵，栅格 $m_0\times m_1$ 记录UAV毁伤状态 $u^0_{m_0\times m_1}$，$0<u^0_{m_0\times m_1}<1$，如果 $m_0\times m_1$ 内没有UAV，则 $u^0_{m_0\times m_1}=0$。$\boldsymbol{U}^0_{m\times m}$ 的设计考虑如下：UAV的完好状态影响UAV的协同作战能力，进而影响装备体系作战效能。

③ $\boldsymbol{W}^1_{m\times m}$ 表示 $m\times m$ 范围内敌方UAV位置矩阵，与 $\boldsymbol{W}^0_{m\times m}$ 类似。

④ $\boldsymbol{U}^1_{m\times m}$ 表示 $m\times m$ 范围内敌方UAV毁伤状态矩阵，与 $\boldsymbol{U}^0_{m\times m}$ 类似。

⑤ $\boldsymbol{V}^0_{(n/N)\times(n/N)}$ 表示己方无人机在作战区域内的数量分布矩阵。将作战区域范围 $n\times n$ 划分为 N^2 个 $(n/N)\times(n/N)$ 的小矩形区域，统计区域内己方UAV的数量占己方UAV总数量的比例。例如，在 $n_0\times n_1$ 栅格，

$$v^0_{n_0\times n_1}=\frac{n_0\times n_1\text{区域内己方UAV的数量}}{\text{己方UAV总数量}} \tag{5}$$

$0\leqslant v^0_{n_0\times n_1}\leqslant 1$。UAV的设计考虑如下：UAVS只有在一定区域内达到一定数量规模，才能协同产生涌现效果，提升装备体系作战效能。

⑥ $\boldsymbol{V}^1_{(n/N)\times(n/N)}$ 表示敌方无人机在作战区域内的分布矩阵，与 $\boldsymbol{V}^0_{(n/N)\times(n/N)}$ 类似。

UAVS的最优控制在一定程度上与之前的动作序列有关系，因此，时间维度包括上一时间步长无人机的动作 $a^{\mathrm{ID}}_{\mathrm{step}-1}$、奖励 $r^{\mathrm{ID}}_{\mathrm{step}-1}$、归一化的位置 $pos^{\mathrm{ID}}_{\mathrm{step}-1}$。

(2) 动作空间设计

在不影响验证方法有效性的条件下，将问题限定为2D平面的四旋翼类无人机。旋翼无人机的机动可描述为六自由度刚体运动，建立惯性坐标系 $OXYZ$ 和机体坐标系 $oxyz$ 描述无人机的位置与转动，如图3所示。

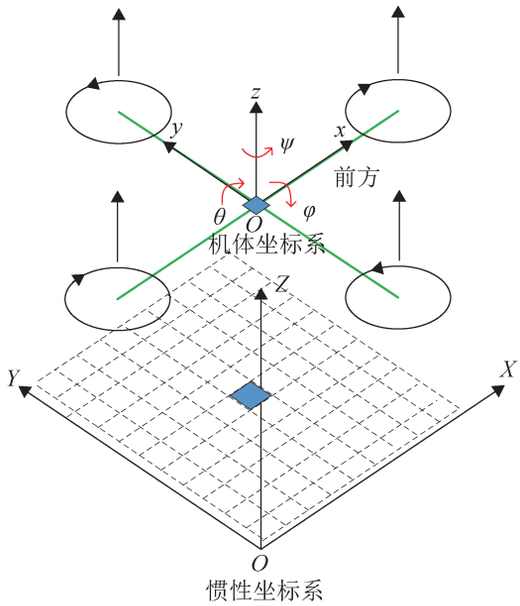

图 3 UAV坐标系建立

Fig. 3 UAV coordinate system

UAV可执行的动作包括沿 Oz 轴转动，重心沿 OXY 两个轴做线运动。因此，定义UAVS联合动作空间 $A=\{A_i\}^N_{i=1}$，$A_i=\{a_i|a_i\in A_{i,\mathrm{move}}\cup A_{i,\mathrm{attack}}\cup A_{i,\mathrm{turn}}\}$。根据UAV在单位仿真时间步长内可移动的栅格数，对移动动作空间 $A_{i,\mathrm{move}}$ 编码，根据UAV毁伤半径，对攻击动作空间 $A_{i,\mathrm{attack}}$ 编码，根据UAV可沿 oz 轴转动特性，对转向动作空间 $A_{i,\mathrm{turn}}$ 编码。

UAVS动作空间的设计考虑如下：UAVS的协同行为受生物群体智能的启发，具有规则简单、结构分布、功能涌现等特征。UAVS内部交互作用，个体行为取决于群体行为模式，而群体行为模式又根据个体行为的调整而动态变化。因此，基于平均场理论，对UAVS的动作进行整体处理，以每个UAV探测范围内所有己方UAV动作的平均来代替单个UAV的动作。

对UAV i 的动作空间 A_i 用one-hot方式编码，即 $A_i=[a^1_i,a^2_i,\cdots,a^{M_i}_i,\cdots,a^{M-1}_i,a^M_i]$ 表示 i 有 M 个动作，如果选择第 M_i 个动作，那么 $a^{M_i}_i=1$，其余元素均为0。平均动作 $\bar{a}_i$ 根据UAV i 探测范围内其他UAV的动作编码求和取平均来进行计算。每个UAV i 探测范围内的UAV j 的动作 $a_j=\bar{a}_i+\delta a_{i,j}$，$\bar{a}_i=$

$\frac{1}{N_i}\sum_j a_j$，其中，$\delta a_{i,j}$，N_i分别为UAVi探测范围内其他UAV的动作编码和平均动作编码的差，其他UAV的数量。

(3) 收益函数设计

假设UAV的毁伤状态值为u，$0 \leqslant u \leqslant 1$，$u$值越大，毁伤越严重。例如，当$u=0$时，表示UAV处于完好状态，当$u=1$时，表示UAV被完全摧毁。仿真时间步长$t$时刻，UAV$i$受到一次攻击，毁伤状态值增加$u_0$，$0<u_0 \leqslant 1$，奖励$r_{i,t}=-u_0$。这里假设$u_0$=0.2；摧毁一架UAV的奖励为$u_1$，为使UAV学会攻击行为，设置正向奖励$u_1$=3；UAV被摧毁的奖励为$u_2$，为使UAV学会躲避攻击，设置负向奖励$u_2$=−7；UAV间发生碰撞，奖励为$u_3$，为使UAV学会避让，设置负向奖励$u_3$=−6；UAV间协同攻击目标，奖励为$u_4$，为使UAV间学会协同，设置正向奖励$u_4$=10，如表3所示。

表 3 UAV奖励设置

Table 3 Settings of UAV reward values

奖励设置	值
毁伤	−0.2
摧毁	3
被摧毁	−7
碰撞	−6
协同攻击	10

对协同攻击奖励塑形的数学描述如下：任意$i \in \mathrm{UAVS_{red}}$，$j \in \mathrm{UAVS_{red}}$，$k \in \mathrm{UAVS_{blue}}$，如果在单位时间步长$t$内，$i$和$j$同时向$k$发起攻击，那么$i$和$j$均获得+10奖励值。以MAgent为实验平台，协同攻击的奖励塑形设置如下所示。

i = AgentSymbol(UAVS$_{\mathrm{red}}$, index= ‘any’)

j = AgentSymbol(UAVS$_{\mathrm{red}}$, index= ‘any’)

k = AgentSymbol(UAVS$_{\mathrm{blue}}$, index= ‘any’)

e_1 = Event (i, ‘attack’, k)

e_2 = Event (j, ‘attack’, k)

config.add_reward_rule(e_1&e_2, receiver=[a, b], value=[10, 10])

3.3.2 博弈算法设计

RLoMAG类[10,12-13]典型算法有Nash Q-learning、平均场强化学习(mean field reinforcement learning, MFRL)等，考虑到MFRL在求解多智能体博弈模型时最优策略的收敛性保证[14-15]，本节在装备体系作战效能评估方法设计中采用了RLoMAG类算法中的MFRL算法。UAV只有在各自的探测范围或火力范围内才能相互发生作用，产生协同。设置UAV的值函数只包含探测范围内UAV之间相互作用的形式，如式(6)所示：

$$Q_i(s,a)=\frac{1}{N_i}\sum_{j\in N(i)}Q_i(s,a_i,a_j)\approx Q_i(s,a_i,\bar{a}_i) \tag{6}$$

动作值函数$Q_i(s,a)$可通过式(7)迭代更新

$$Q_{i,t+1}(s,a_i,\bar{a}_i)=(1-\alpha)Q_{i,t}(s,a_i,\bar{a}_i)+\alpha[r_i+\gamma v_{i,t}(s')] \tag{7}$$

式中：α为学习率。

状态值函数定义为

$$v_{i,t}(s')=\sum_{a_i}\pi_{i,t}(a_i|s',\bar{a}_i)E_{\bar{a}_i(\boldsymbol{a}_{-i}\sim\boldsymbol{\pi}_{-i,t})}[Q_{i,t}(s',a_i,\bar{a}_i)] \tag{8}$$

式中：$\pi_{i,t}$为t时刻智能体i的动作概率分布，依赖于联合状态S及平均动作$\bar{a}_i$。MARL问题转化为求解UAVi最佳策略$\pi_{i,t}$，$\pi_{i,t}$与探测范围内UAVj的平均动作$\bar{a}_i$有关，$\bar{a}_i$计算如式(9)所示：

$$\bar{a}_i=\frac{1}{N_i}\sum_j a_j, a_j\sim\pi_{j,t}(\cdot|s,\bar{a}_{j-}) \tag{9}$$

式中：$\bar{a}_{j-}$为上一仿真时刻平均动作，UAVj的动作a_j由$\pi_{j,t}$决定，依赖于上一仿真时刻平均动作$\bar{a}_{j-}$影响。

通过式(9)可以计算出UAVj平均动作$\bar{a}_i$，$\pi_{i,t}$通过玻尔兹曼分布得到新的策略：

$$\pi_{i,t}(a_i|s,\bar{a}_i)=\frac{\exp(-\beta Q_{i,t}(s,a_i,\bar{a}_i))}{\sum_{a_i\in A_i}\exp(-\beta Q_{i,t}(s,a_{i'},\bar{a}_i))} \tag{10}$$

通过式(7)~(10)迭代更新，寻找最优策略，获得最大累积回报。

Nash Q-learning算法的核心为

$$\text{Nash } Q(s,a)=E_{s'\sim p}[r(s,a)+\gamma v^{\text{Nash}}(s')] \tag{11}$$

式中：$Q=[Q_1,Q_2,\cdots,Q_N]$，$r=[r_1,r_2,\cdots,r_N]$。文献[9]证明了$\bar{a}_i$能够收敛到唯一平衡点，并推得策略π_i收敛到纳什均衡策略。为了与Nash Q算法(11)

对应，MF-Q算法如式(12)。

$$\text{MF } Q(s,a)=E_{s'\sim p}[r(s,a)+\gamma v^{\text{MF}}(s')] \tag{12}$$

其中，$v^{\text{MF}}(s)\triangleq\left[v_1(s),v_2(s),\cdots,v_N(s)\right]$，

$v_i(s)=\sum_{a_i}\pi_i(a_i|s,\bar{a}_i)E_{\bar{a}_i(a_{-i}\sim\pi_{-i})}[Q_i(s,a_i,\bar{a}_i)]$，$i\in[1,N]$。

3.4 仿真实验

仿真实验实施步骤如下：

step 1：设定对手威胁，作战环境，作战方案；

step 2：基于MFRL求解体系对抗博弈模型。统计UAVS每迭代100轮的标准偏差$\text{stdv}=\sqrt{\frac{1}{N-1}\sum_{\text{episode}}^{\text{episode}+N-1}[R_i-\bar{R}_{(\text{episode},\,\text{episode}+N-1)}]^2}$，$i\in(\text{episode},\text{episode}+N-1)$，其中，$\bar{R}_{(\text{episode},\,\text{episode}+N-1)}=\frac{1}{N}\sum_{\text{episode}}^{\text{episode}+N-1}R_{\text{episode}}$为平均奖励，$N=100$，$\text{episode}=100k+1$，$k=0,1,2,\cdots$；

step 3：设置收敛条件$\text{stdv}<300$，如果达到收敛条件，转入step 4，如果未达到收敛条件，转入step 2；

step 4：将战损比(η=红方战损/蓝方战损)作为衡量装备体系作战效能的指标，如果满足约束条件$\eta<1$，转入step 5，否则，以10架UAV为1个数量级，调整UAVS体系设置条件，转入step 1；

step 5：输出UAVS体系设置条件，战损比。

3.5 结果分析与得出结论

以10架UAV为1个数量级，每调整体系设置条件，均开展一次仿真实验，当红方收敛至纳什均衡(Nash equilibrium, NE)策略时，表示此次仿真实验结束。在纳什均衡条件下，蓝方UAV剩余数量随红方UAV规划数量变化曲线如图4所示，可以看出，红方UAVS从170架增加至190架时，蓝方UAV剩余数量由30多架降至0架，说明红方在增加20架UAV后，装备体系作战效能有一次骤增。

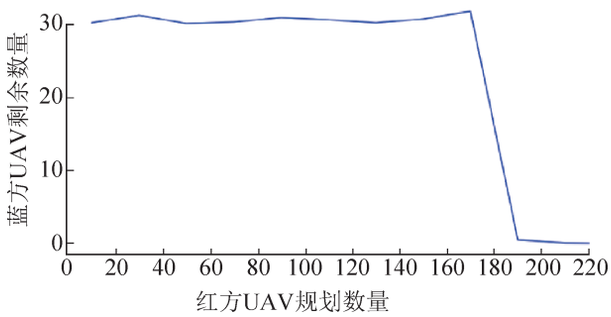

图4 NE条件下蓝方UAV剩余数量随红方UAV规划数量变化曲线

Fig. 4 Remaining number of blue UAVs varies with the planned number of red UAVs under Nash equilibrium

红方UAVS 190架时双方最终剩余数量变化曲线如图5所示，平均战损比$\bar{\eta}_{900\sim1000}\approx0.003\,9<1$。

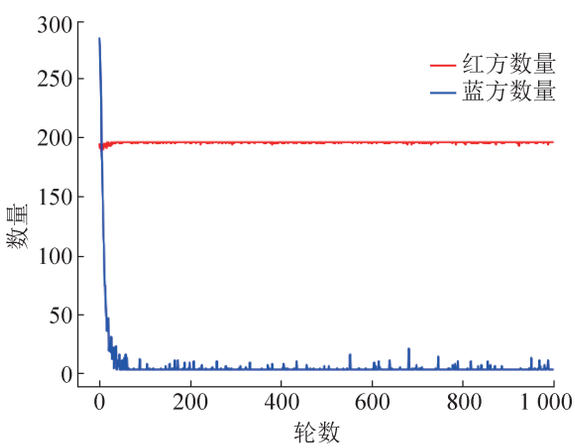

图5 红方UAVS 190架时双方最终剩余数量变化曲线

Fig. 5 Final remaining number when number of red UAVs is 190

以上数据分析表明，当红方UAVS数量为190架时，发挥出的最大装备体系作战效能，可有效拦截蓝方280架UAVS。当UAV数量为170架时，红、蓝方UAV数量变化曲线如图6所示，经计算，当$900<\text{episode}<1\,000$时，蓝方UAV数量$\text{num}\approx31.77>>1$，说明红方UAVS从190架减少至170架，装备体系作战效能发生了“塌方式”下降。

以上分析可知，在固定想定条件下，蓝方发射280架程序预置的UAV空袭蓝方导弹阵地，红方至少需要规划190架UAV，才能在最优策略下完成拦截所有红方UAV的使命任务。

第 36 卷第 1 期 系统仿真学报 Vol. 36 No. 1
2024 年 1 月 Journal of System Simulation Jan. 2024

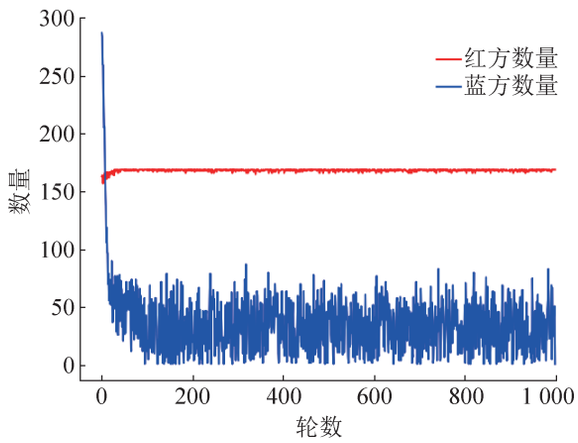

图 6 红方 UAVS 170 架时的双方最终剩余数量变化曲线
Fig. 6 Final remaining number when number of red UAVs is 170

4 结论

本文针对大规模、高动态、强对抗作战场景下的装备体系作战效能评估问题，将体系对抗过程看作为马尔可夫博弈过程，以装备体系能够发挥的最大作战效能为学习对象，通过设计智能体的状态空间、收益函数、行为空间以及智能博弈算法，在体系探索最优策略的过程中，形成智能体间的复杂交互，模拟出体系的适应性、不确定性、非线性、涌现性特征。将装备体系的编组、规模、功能等作为变量，结合探索性分析仿真方法，通过探索博弈模型的最优策略，演化出大量的作战过程，分析装备体系在达到最优策略情况下的作战效能指标，完成装备体系作战效能评估工作。最后，给出 UAVS 装备体系作战效能评估方法应用示例，验证了方法的有效性。博弈与学习结合的意义在于优势互补，博弈论提供了易于处理的解概念来描述多智能体系统的学习结果，强化学习算法提供了可收敛的学习算法，可以在序列决策过程中达到稳定和理性的均衡。随着多智能体博弈理论的不断深入研究，必将对装备体系作战效能评估领域产生更深远的影响。

参考文献：

[1] 张子伟, 郭齐胜, 董志明, 等. 体系作战效能评估与优化方法综述[J]. 系统仿真学报, 2022, 34(2): 303-313.
Zhang Ziwei, Guo Qisheng, Dong Zhiming, et al. Review of System of Systems Combat Effectiveness Evaluation and Optimization Methods[J]. Journal of System Simulation, 2022, 34(2): 303-313.

[2] 袁宏皓, 袁成. 体系效能评估技术发展综述[J]. 飞航导弹, 2019(5): 63-67.

[3] 杨克巍, 杨志伟, 谭跃进, 等. 面向体系贡献率的装备体系评估方法研究综述[J]. 系统工程与电子技术, 2019, 41(2): 311-321.
Yang Kewei, Yang Zhiwei, Tan Yuejin, et al. Review of the Evaluation Methods of Equipment System of Systems Facing the Contribution Rate[J]. Systems Engineering and Electronics, 2019, 41(2): 311-321.

[4] 梁晓龙, 胡利平, 张佳强, 等. 航空集群自主空战研究进展[J]. 科技导报, 2020, 38(15): 74-88.
Liang Xiaolong, Hu Liping, Zhang Jiaqiang, et al. Research Status and Prospect of Aircraft Swarms Autonomou Air Combat[J]. Science & Technology Review, 2020, 38(15): 74-88.

[5] 贾永楠, 田似营, 李擎. 无人机集群研究进展综述[J]. 航空学报, 2020, 41(增1): 1-11.
Jia Yongnan, Tian Siying, Li Qing. Recent Development of Unmanned Aerial Vehicle Swarms[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(S1): 1-11.

[6] 刘德胜. 基于复杂网络分析方法的作战体系评估研究综述[J]. 军事运筹与系统工程, 2020, 34(3): 66-73.
Liu Desheng. A Survey of Combat System Assessment Based on Complex Network Analysis[J]. Military Operations Research and Systems Engineering, 2020, 34 (3): 66-73.

[7] 杨圩生, 王钰, 杨洋, 等. 基于作战环的不同节点攻击策略下的作战网络效能评估[J]. 系统工程与电子技术, 2021, 43(11): 3220-3228.
Yang Weisheng, Wang Yu, Yang Yang, et al. Combat Network Effectiveness Evaluation under Different Node Attack Strategies Based on Operation Loop[J]. Systems Engineering and Electronics, 2021, 43(11): 3220-3228.

[8] 柯加山, 江敬灼, 许仁杰, 等. 联合作战体系对抗效能评估探索性分析框架[J]. 军事运筹与系统工程, 2005(4): 58-61.
Ke Jiashan, Jiang Jingzhuo, Xu Renjie, et al. The Study on the Exploratory Evaluation Framework of the Combat Effectiveness of Joint Warfare System of Systems[J]. Military Operations Research and Systems Engineering, 2005(4): 58-61.

[9] 雷永林, 朱智, 甘斌, 等. 基于仿真的复杂武器系统作战效能评估框架研究[J]. 系统仿真学报, 2020, 32(9): 1654-1663.

Lei Yonglin, Zhu Zhi, Gan Bin, et al. Combat Effectiveness Simulation Evaluation Framework of Complex Weapon System[J]. Journal of System Simulation, 2020, 32(9): 1654-1663.

[10] 罗俊仁, 张万鹏, 苏炯铭, 等.多智能体博弈学习研究进展[J/OL]. 系统工程与电子技术. (2022-06-27) [2022-08-24]. http://kns. cnki. net/kcms/detail/11.2422. TN. 20220625.1341.018.html.
Luo Junren, Zhang Wanpeng, Su Jiongming, et al. Research Progress of Multi-agent Game Theoretic Learning[J/OL]. Systems Engineering and Electronics. (2022-06-27) [2022-08-24]. http://kns. cnki. net/kcms/detail/11.2422.TN.20220625.1341.018.html.

[11] Zhang Kaiqing, Yang Zhuoran, Tamer Başar. Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms[M]//Vamvoudakis K G, Wan Yan, Lewis F L, et al. Handbook of Reinforcement Learning and Control. Cham: Springer International Publishing, 2021: 321-384.

[12] 袁唯淋, 罗俊仁, 陆丽娜, 等. 智能博弈对抗方法: 博弈论与强化学习综合视角对比分析[J]. 计算机科学, 2022, 49(8): 191-204.
Yuan Weilin, Luo Junren, Lu Lina, et al. Methods in Adversarial Intelligent Game: A Holistic Comparative Analysis from Perspective of Game Theory and Reinforcement Learning[J]. Computer Science, 2022, 49 (8): 191-204.

[13] 王军, 曹雷, 陈希亮, 等. 纯策略纳什均衡的博弈强化学习[J]. 计算机工程与应用, 2022, 58(15): 78-86.
Wang Jun, Cao Lei, Chen Xiliang, et al. Game Reinforcement Learning of Pure Strategy Nash Equilibrium[J]. Computer Engineering and Applications, 2022, 58(15): 78-86.

[14] Yang Yaodong, Luo Rui, Li Minne, et al. Mean Field Multi-agent Reinforcement Learning[C]//Proceedings of the 35th International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2018: 5571-5580.

[15] Jain M, Korzhyk D, Ondřej Vaněk, et al. A Double Oracle Algorithm for Zero-sum Security Games on Graphs[C]//The 10th International Conference on Autonomous Agents and Multiagent Systems. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2011: 327-334.

DOI: 10.16182/j.issn1004731x.joss.22-1034