计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(8):125−132 [doi: 10.15888/j.cnki.csa.008600]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于决策知识学习的多无人机航迹协同规划①

曾　熠[1], 刘丽华[2], 李　璇[2], 杜溢墨[1], 陈丽娜[1]

[1](解放军 31008 部队, 北京 100091)
[2](国防科技大学 系统工程学院, 长沙 410073)
通信作者: 刘丽华, E-mail: gogonudt@126.com

摘　要: 考虑无人机群体行为决策与状态变化的内在驱动, 从信息处理角度提出基于决策知识学习的多无人机航迹协同规划方法. 首先, 基于马尔科夫决策过程对无人机的行为状态进行知识表示, 形成关于连续动作空间的决策知识; 然后, 提出基于知识决策学习的深度确定性策略梯度算法, 实现无人机在决策知识层次上的协同规划. 实验结果表明: 在研发设计演示系统的基础上, 所提方法通过强化学习能够得到一个最优航迹规划策略, 同时使航迹综合评价和平均奖励收敛稳定, 为无人机任务执行提供了决策支持.

关键词: 多无人机; 决策知识; 知识学习; 航迹协同规划; 工业互联网; 人工智能

引用格式: 曾熠,刘丽华,李璇,杜溢墨,陈丽娜.基于决策知识学习的多无人机航迹协同规划.计算机系统应用,2022,31(8):125–132. http://www.c-s-a.org.cn/1003-3254/8600.html

Trajectory Collaborative Planning of Multi-UAV Based on Decision-making Knowledge Learning

ZENG Yi[1], LIU Li-Hua[2], LI Xuan[2], DU Yi-Mo[1], CHEN Li-Na[1]

[1](PLA 31008 Unit, Beijing 100091, China)
[2](College of Systems Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Considering the internal driving mechanism of behavior decision-making and state changes of multiple UAVs, a collaborative trajectory planning method based on decision-making knowledge learning is proposed from the perspective of information processing. Firstly, the behavior states of UAVs are represented by knowledge on the basis of the Markov decision process, and the decision-making knowledge on continuous action space is developed. Then, a deep deterministic policy gradient (DDPG) algorithm based on decision-making knowledge learning is presented to achieve the collaborative planning of UAVs on the decision-making knowledge level. The experimental results reveal that on the basis of developing a demonstration system, the method can obtain an optimal trajectory planning strategy by reinforcement learning and can simultaneously achieve the convergence and stability of the comprehensive evaluation and average reward of trajectories, which provides decision-making support for mission execution of UAVs.

Key words: multi-UAV; decision-making knowledge; knowledge learning; trajectory collaborative planning; industrial Internet; artificial intelligence

航迹协同规划[1]是实现多无人机自主行为导航与控制的关键技术, 也是对环境感知决策的具体体现形式, 其目的是规划出最优的航迹策略, 以解决目标搜索、飞行避碰、编队控制等问题. 现有关于航迹协同规划方法主要有启发式方法[2]、Voronoi 方法[3]、遗传算法[4]、粒子群算法[5]等, 由于外界复杂环境影响, 行为变化的不确定性对航迹规划提出了更高的任务需求[6,7].

从无人机群体行为决策[8,9]与状态变化的内在驱动机制看, 复杂的群体行为通过简单的局部交互知, 需

① 收稿时间: 2021-10-29; 修改时间: 2021-11-29; 采用时间: 2021-12-08; csa 在线出版时间: 2022-06-01

要遵循一定的标准知识才能保证整个系统可控性. 决策知识[10] 是实现自然语言与环境信息交互的一种接口, 它采用标准化的规则格式实现机器指令与外界信息的交互理解, 是目前智能机器领域的研究热点[11]. 文献 [12] 采用知识本体的思维构建了任务规划的概念层次, 给出了决策知识学习在无人机航迹协同规划上的逻辑推理应用. 但该方法只描述了外部环境的概念形式, 缺少对无人机动作和状态内部驱动的知识表示. 文献 [13] 运用层级式表达方式对无人机环境信息进行概念抽取, 在航迹序列点位置上部署决策点, 并赋予基于决策树的知识学习方法. 但该方法计算航迹代价较高, 容易陷入局部最优状态, 较难保证全局航迹规划最优. 文献 [14] 使用神经网络指导无人机建立了一个决策知识框架, 用于推理目标搜索中的环境知识和状态, 从而获得最优策略. 但该方法未考虑事件触发与系统内在关系, 较难保证任务背景中的知识学习能力[15,16].

综上所述, 从信息处理的角度探讨性地提出了一种基于决策知识学习的多无人机航迹协同规划方法. 该方法基于马尔可夫决策过程, 重点构建决策知识库, 形成基于事件触发-知识驱动的群体决策机制, 通过引入意义接受性学习理论增强决策知识学习的相关性, 以获取多无人机航迹规划的最优策略.

1 系统设计

1.1 任务描述

多无人机协同航迹规划问题是将每台无人机同时从不同的起点到相同目的地或侦察点, 生成可行的飞行轨迹, 这些轨迹由一组协同全局最小代价的优化准则和约束条件定义, 包括最小化无人机被摧毁的风险, 以及无人机内外部环境限制和威胁动态. 如图 1 所示, 多无人机航迹协同规划任务描述中, 任务空间中有 3 台无人机和 6 个威胁区域以及部分地形障碍, 需要通过系统状态不断调整优化动作, 对每台无人机形成一个动作序列, 每个动作又形式化表示为协同决策和任务优化问题.

1.2 系统框架

针对无人机航迹协同规划的连续动作空间特征, 将知识决策框架分为数据支持层、模型生成层和策略控制层. 如图 2 所示, 它是整个系统的基本框架.

(1) 数据支持层: 主要将空间数据库的信息和无人机传感器获取的环境信息、威胁信息、历史经验等进行知识的实例化表示, 对情景任务进行有效分析, 形成具有图结构的决策知识库, 同时赋予了相关事件属性、动作模板和关系条件, 其功能包括事件触发、行为动作和状态转移等.

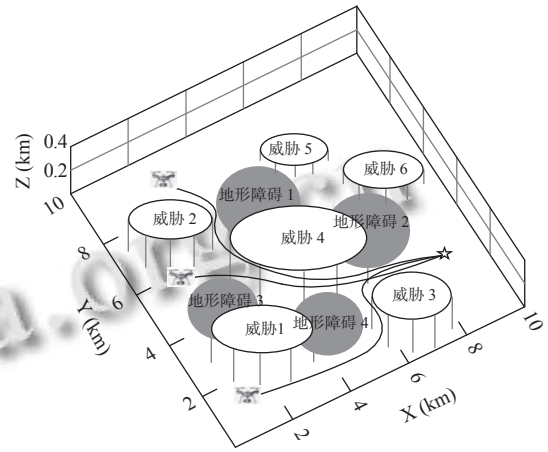


图 1 多无人机航迹协同规划任务描述

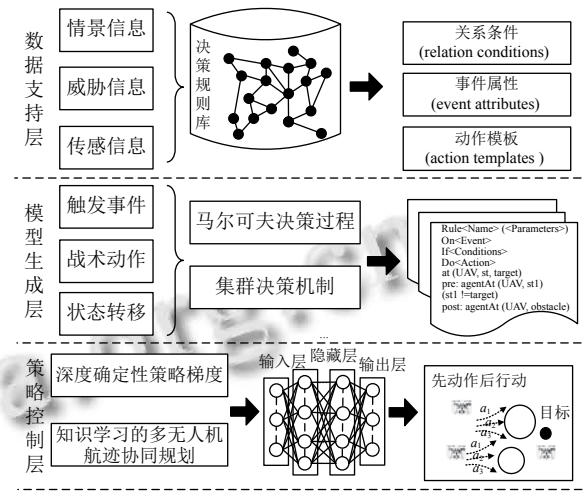


图 2 基于决策知识学习的多无人机航迹协同规划系统框架

(2) 模型生成层: 主要利用马尔科夫决策过程对无人机群体的状态和动作进行建模, 得出最优状态-动作值, 产生最优策略; 通过群体决策机制对无人机当前状态、情景和动作进行分析, 形成决策知识与无人机系统的信息交互, 为航迹规划的策略控制提供数据支撑.

(3) 策略控制层: 主要采用深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 对当前无人机群体的动作和状态进行训练, 通过引入意义接受性学习理论, 提出基于知识决策学习的深度确定性策略梯度算法, 不断调整选择最优策略, 将新的群体协同决

策经验知识映射存储于至知识库, 以提高航迹规划的准确性.

2 决策过程

2.1 基于马尔科夫的决策过程

马尔科夫决策[17] 过程对序贯决策问题进行了数学定义, 为多无人机决策和任务优化提供了一种端到端的学习框架. 根据马尔科夫决策过程, 将无人机航迹规划表示为一个五元组模型(S,A,P,R,γ). 其中, S为无人机在当前航迹序列下可以到达的所有状态的集合, A 表示无人机可以在环境中选择的所有动作的集合, P 和 R 表示从状态s到状态s'执行动作a的概率和奖励 ($a\in A$且$s,s'\in S$), $\gamma\in[0,1]$为决定当前或未来奖励重要性的折扣因素.

在每个时间步长 t, 无人机的状态为s_t, a_t为无人机在该状态下执行的动作, 无人机从环境交互中获得奖励r_t, 并在下一时间步到达状态s_{t+1}. 同时, 无人机在每个时间步长选择的动作由策略集π决定, 一个包含在策略π的元素$\pi(a|s)$表示无人机在某个状态s所采取行动a的概率. 在所有策略中, 有一个最优策略π^*, 当无人机遵循该策略时, 可以获得最大奖励R_t, 从步长 t 的开始时间到结束时间, 累积奖励表示为:

$$R_t=\sum_{1}^{N}\sum_{t'=t}^{T}\gamma^{t'-t}r(s_{t'},a_{t'}) \tag{1}$$

状态-动作值函数$Q^{\pi}(s,a)$计算为$Q^{\pi}(s,a)=E[R_t|s_t=s,a_t=a,\pi]$表示无人机在当前状态下执行动作的过程. 当执行最佳策略π^*时, 则$Q^*(s,a)=\max\limits_{\pi}E[R_t|s_t=s,a_t=a,\pi]$为最优状态-动作值函数满足贝尔曼最优性方程[17].

$$Q^*(s,a)=E[r_{t+1}+\gamma\max_{a_{t+1}}Q^{\pi}(s_{t+1},a_{t+1})|s_t=s,a_t=a] \tag{2}$$

状态-动作值函数将通过不断迭代等式最终收敛$Q^*(s,a)$产生最优策略$\pi^*=\arg\max\limits_{a\in A}Q^*(s,a)$.

2.2 决策知识库

根据任务区域内所有实体在空间上的布局, 将区域内的每个实体进行知识表示, 初始化为一个图形化的决策知识库 SD_Net. 存储在 SD_Net 中的每个知识结构对应无人机的不同危险程度和航迹序列位置, 其作用是将当前无人机采样的环境信息和系统状态与知识库的知识进行信息交互, 供无人机系统学习训练. SD_Net 的结构为, 其中为概念层次的层次结构分为系统状态 (由马尔可夫决策过程生成的状态网络层次)、触发事件 (由历史不确定事件所构成的网络层次) 和环境知识 (由历史态势环境的背景知识构成的网络层次); E 为链接各概念层次的关系; I 为具体的实例, 存储了所有不同任务背景下的案例知识; At 为马尔可夫决策过程形成的行为动作和奖励. SD_Net 模型如图 3 所示, 以 Protégé平台[11] 进行构建, 封装为基于 SPARQL 语言的 OWL 模型[12], 存有 500 余个概念层次实体和 6 000 余个案例知识, 用 SWRL 调试分析无人机群体决策的情景分析, 形成决策知识.

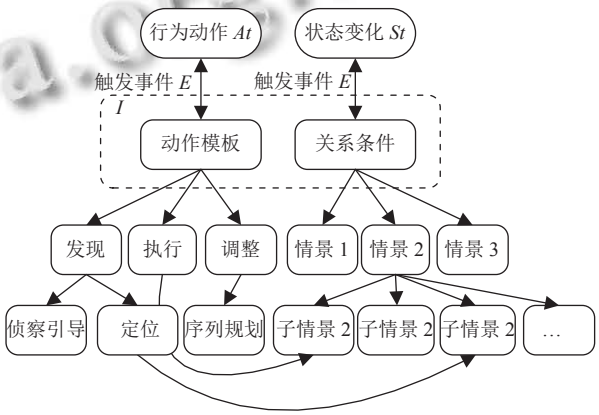


图 3 无人机决策知识库模型

知识库 SD_Net 作为无人机航迹规划的初始规决策知识, 是当前系统知识认知结构中已有的概念层次, 但在动态任务的决策过程中由于外部威胁区域的不确定性, 需要利用当前状态和动作不断进行调整优化, 学习到最优策略达到航迹规划效果.

2.3 群体决策机制

无人机群体作为一个行为可控的复杂系统, 在马尔科夫模型的状态和动作基础上, 关键在于任务平台能够针对特定的触发事件, 以决策知识为驱动, 需要自主进行行为决策与状态变化, 因此提出基于事件触发-知识驱动决策机制. 其中, 事件触发为无人机外部触发条件, 通过事件检测器对底层数据进行事件提取, 与所建立的知识库进行匹配, 构建事件与动作行为的映射关系. 作为背景知识的一部分, 事件触发与任务区域的不确定环境进行交互, 在一定程度上提高了决策过程的可解释性[18], 同时节约了存储和计算资源; 知识驱动是内部驱动机制, 根据协同航迹规划需遵循的规则, 群体行为通过局部交互知识产生, 逐渐扩散到全局知识, 以发现、执行和调整 3 种动作规则, 支撑自主行为决

策和状态变化的概念化、形式化的知识表示, 从而明确任务空间相关要素与系统状态转移的关联关系. 如图 4 所示, 多台无人机进入指定任务区域执行任务, 在情景 T 的触发下产生状态, 根据 SD_Net 知识库中的概念层次, 分别使用案例规则知识, 形成不同状态下的行为选择, 从而规划出一个连续动作空间.

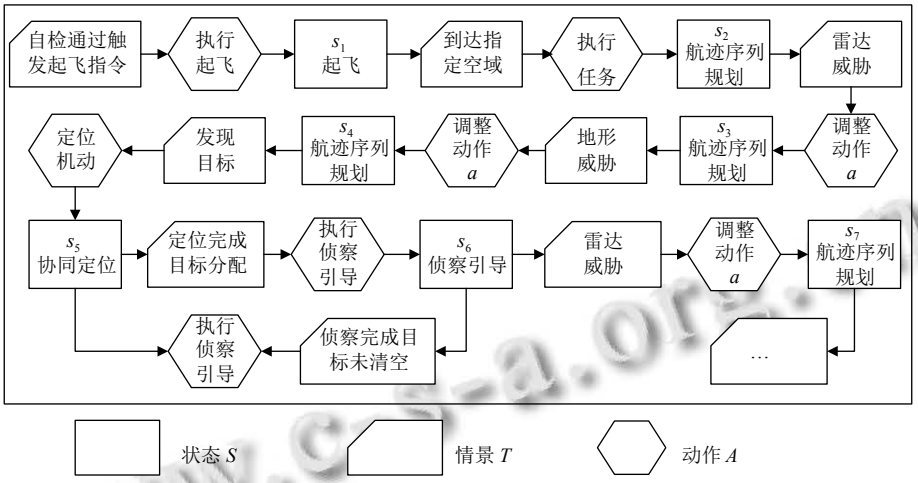


图 4 多无人机决策知识情景分析

3 航迹规划方法

3.1 深度学习框架

在群体决策的知识分析后, 多台无人机航迹协同规划形成了一个连续动作空间, 需要寻找优化速度较快且适用于大规模数据的函数逼近器[17], 使群体决策达到最优策略. 将深度 DDPG 算法[19] 应用于连续动作空间, 主要由态势环境、经验池、行动者网络和评论家网络 4 个部分组成. 其中, 态势环境和经验池的功能分别用于生成和存储知识, 在不断与环境交互的过程中, 无人机系统获取环境知识并将其存储于经验池中以备学习, 并实时映射于知识库 SD_Net; 行动者网络用于确定无人机系统选择行动的概率, 当无人机与环境交互时, 它会根据行动者网络选择行动; 评论家网络根据环境状态评估无人机系统选择的操作, 行动者网络基于评估修改选择行动的概率. 为使行动者网络更具稳定性, 在行动者网络的输出中加入了探测噪声N_t, 即样本大小, 使每个动作为$a_t=\mu(s_t|\theta^{\mu})+N_t$, 其中$\theta^{\mu}$表示在行为者网络中显示评论家网络的参数, 行为者网络通过策略梯度近似计算评论家网络参数:

$$\nabla_{\theta^{\mu}}J\approx\frac{1}{N}\sum_{i}\nabla_{a}Q(s,a|\theta^{Q})|_{s=s_i,a=\mu(s_i)}\nabla_{\theta^{\mu}}\mu(s|\theta^{\mu})|_{s_i} \tag{3}$$

其中, θ^{Q}表示行动者网络中评估参数, 通过最小化损失函数更新评估参数:

$$L(\theta^{Q})=\frac{1}{N}\sum_{i}(y_i-Q(s_i,a_i|\theta^{Q}))^2 \tag{4}$$

$$y_i=r(s_i,a_i)+\gamma Q'(s_{i+1},\mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'}) \tag{5}$$

其中, $\theta^{Q'}$为评论家网络中的目标网络, N 为学习经验数量. 评论家网络的参数更新方法采用软更新的方法来提高训练的稳定性$\theta'\leftarrow\tau\theta+(1-\tau)\theta',\tau\in[0,1]$用于确定更新程度.

3.2 知识相关性学习

为提高训练效率, 将意义接受性学习理论 (meaningful receptive learning, MRL)[20] 引入协同任务规划的学习训练中, 其原理是将无人机所需学习的态势环境内容与 SD_Net 中的背景知识和概念层次进行关联.

首先, 设计一个知识相关性函数$f_r(s)$评估当前状态与知识库 SD_Net 中知识的相关性.

$$f_r(s)=\rho_1\left(-\frac{D_{\text{current}}}{D_{\text{maximum}}}\right)+\rho_2\left(-\frac{\varphi_{uT}+\alpha_{uT}}{8}\right)+\rho_3 d_{\text{ave}} \tag{6}$$

其中, D_{maximum}为无人机与环境中目标之间的理论最大距离, ρ_1、ρ_2和ρ_3为常数. 首先, 在每个时间步长, 计算群体决策知识的相关性, 将其存储于知识库 SD_Net 中, 筛选出与当前学习状态最相关的知识用于任务规划. 然后, 按照 DDPG 算法的时序差分差错 (temporal difference error, TD-error)[21], 来评估评论家网络的目标网与行动者的行动网之间的差异, 选择与当前状态最相似的知识. 最后, 采用先学习后选择动作方式输出最优策略. 如图 5 所示, 以二维离散场景为例, 无人机在当前时间

步长有左转一定角度 a_1、前进 a_2、右转一定角度 a_3 等 3 个离散动作可供选择. 假设在当前状态 s 下, 神经网络将输出动作 a_1, 这将导致任务失败. 在每个学习时间步骤, DDPG 算法会根据当前状态选择一个动作 a_2 并执行, 在调整学习和动作选择的顺序后, 本文所提算法根据当前状态 s 选择最适合学习的经验, 在学习过程结束后, 神经网络的参数会发生一定程度的变化, 由于学习到的知识是与状态 s 相似的经验, 参数更新后的神经网络输出 a_3, 使无人机能够安全避开威胁区域.

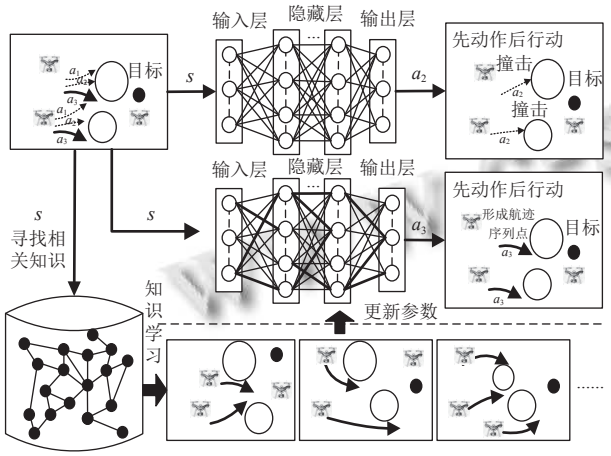


图 5 知识相关性学习

3.3 算法实现

在每个时间步长 t 中, 计算每个状态s_t的经验相关性函数$f_r(s_t)$, 将其存储于知识库 SD_Net 中, 提出基于决策知识学习的深度确定性策略梯度算法 (knowledge learning decision-PPDG, KLD-PPDG). 其过程是知识库 SD_Net 中的知识结构从$(s_t,a_t,r_t,s_{t+1},E,I,)$变化为$(s_t,a_t,r_t,s_{t+1},E,I,f_r(s_t))$, 在每个时间步长 t 中, 根据知识库中每个知识选择当前状态N_t个知识$(s_i,a_i,r_i,s_{i+1},E,I,f_r(s_i))_{i=1,2,\cdots,N_{td}}$进行排序; 然后, 根据每个采样经验的当前状态$f_r(s_t)$和$f_r(s_i)$,形成一个最小值阈$\Delta f_r=|f_r(s_t)-f_r(s_i)|$, 根据 TD-error 的$\delta_i$更新这些知识选择概率; 最后, 这些知识用于更新网络的参数. 具体算法如算法 1 所示.

算法 1. KLD-PPDG

1.初始化知识库容量 D, 无人机数量 N, 样本大小 N_t, 重播周期 K, 训练集 M, 以及范例 α 和 β;
2.随机初始化评论家网络$Q(s,a|\theta^Q)$和行动者网络$\mu(s|\theta^\mu)$, 它们的权重分别为θ^Q和θ^μ;
3.初始化目标网络Q'和μ', 权重分别为$\theta^{Q'}\leftarrow\theta^Q$和$\theta^{\mu'}\leftarrow\theta^\mu$;
4.初始化经验池 R;
5.**for** 集合=1, M **do**
6. 初始化一个随机进程 N 进行动作探索;
7. 接收初始观察状态 s_1 并设置 p_1 =1;
8. **for** t=1, T **do**
9. 计算经验相关值$f_r(s_t)$;
10. **if** t=0 mod K **then**
11. 根据样本概率 $i\sim P(i)=p_i^\alpha/\sum_j p_j^\alpha$, 取样$N_t$个经验$(s_i,a_i,r_i,s_{i+1},E,I,f_r(s_i))_{i=1,2,\cdots,N_{td}}$;
12. 根据每个采样经验的当前状态$f_r(s_t)$和$f_r(s_i)$, 形成一个最小值阈$\Delta f_r=|f_r(s_t)-f_r(s_i)|$;
13. 计算重要性采样权重$\omega_i=(D\cdot P(i))^{-\beta}/\max_j\omega_j$, 根据式 (4) 设置
$$y_i=r(s_i,a_i)+\gamma Q'(s_{i+1},\mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$$
14. 计算 TD-error 值$\delta_i=y_i-(s_i,a_i|\theta^Q)$;
15. 更新知识优先值$p_i\leftarrow|\delta_i|$, 通过式 (1) 计算奖励值 R, 通过式 (2) 最小化损失函数更新评价网络的参数, 通过策略梯度近似评估行动者网络参数式 (3);
16. 更新目标网络
$$\theta^{Q'}\leftarrow\tau\theta^Q+(1-\tau)\theta^{Q'}$$
$$\theta^{\mu'}\leftarrow\tau\theta^\mu+(1-\tau)\theta^{\mu'}$$
17. **endif**
18. 基于当前策略和探测噪声选择动作$a_t=\mu(s_t|\theta^\mu)|_{s_i}+N_t$; 产生最优策略$\pi^*=\arg\max_{a\in A}Q^*(s,a)$;
18. 执行动作a_t并观察奖励值r_t和下一个状态s_t;
19. 存储$(s_t,a_t,r_t,s_{t+1},f_r(s_t))$至 SD_Net 中;
20. **endfor**
21.**endfor**

KLD-PPDG 算法利用 MRL 理论计算连续知识相关性选择适合不同时间的学习知识, 还调整算法中学习和动作选择的顺序, 增强历史相关经验知识对当前状态下规划决策的影响, 提高算法的收敛速度.

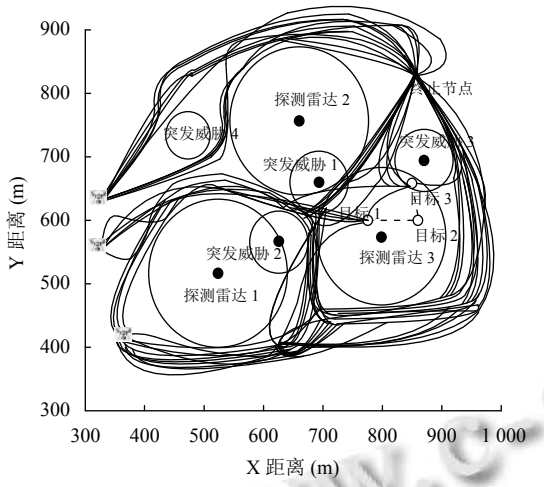
4 实验分析

结合军民融合研究项目, 本文主要通过 Netlogo 平台验证所提出方法的有效性.

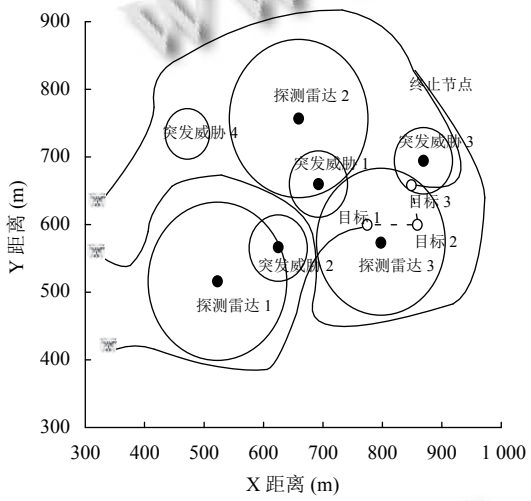
4.1 航迹规划分析

航迹规划分析的实验如图 6 所示, 采用 3 台无人机在某作战区域执行侦察任务, 在 Netlogo 中实时导入知识库 SD_Net 的概念层次, 针对该区域的探测雷达和突发威胁, 描述为马尔可夫决策过程的行为状态. 由图 6(a) 可知, 生成多条航迹序列点作为历史航迹经验知识. 由图 6(b) 可知, 运行 PPDG 算法后, 形成的航迹规划效果. 由图 6(c) 可知, 运行 KLD-PPDG 算法后, 3 台无人机从各初始位置触发, 将无人机所需学习的态势环境内容与 SD_Net 中的背景知识和概念层次进行关联, 然后以先学习后动作方式, 执行每个航迹序列点上的状

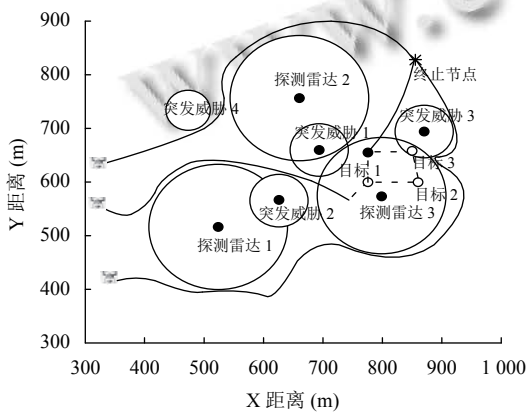
态转移和 TD-error 计算, 重复以上过程, 使知识得到充分利用, 最终形成一个最优的轨迹规划策略.



(a) 导入知识库 SD_Net 形成历史航迹经验知识



(b) 运行 PPDG 算法结果



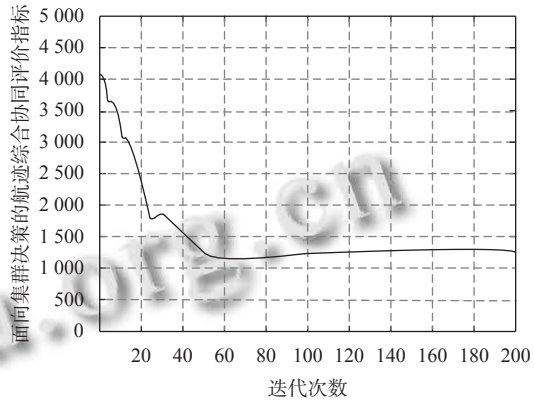
(c) 运行 KLD-PPDG 算法结果

图 6　航迹规划结果

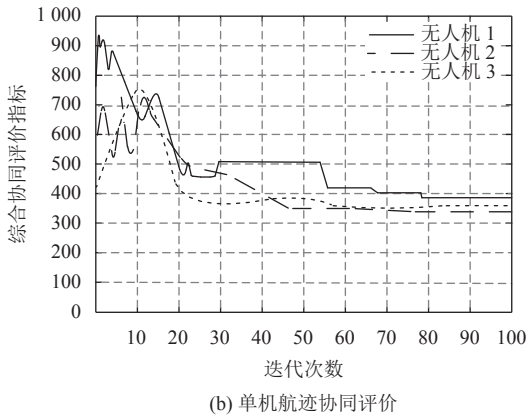
4.2　性能分析

(1) 航迹综合协同评价[22]

航迹综合协同评价指标是测量多无人机协同规划的重要评价指标, 对于任务区域和威胁区域的不断, 使无人机航迹序列点在 KDL-PPDG 学习过程中不断更新迭代, 引入航迹综合协同评价指标[4] 说明本文所提 KDL-PPDG 方法在航迹协同规划中群体决策的性能. 如图 7 所示, 3 台无人机任务航迹协同综合评价变化曲线, 在实时复杂的探测雷达和突发威胁环境态势下, 其航迹综合协同评价指标 (图 7(a)) 和单个无人机的协同评价指标 (图 7(b)), 在迭代至 50 次时其航迹代价值差距逐步变小且趋于收敛稳定, 这说明对于真实环境信息的感知, 每台无人机在经过多次知识学习后, 目标航迹序列点上的状态和动作选择趋于最优. 主要是由于本文方法在初始阶段构建了一个决策知识库 SD_Net, 体现了决策知识对航迹规划的优势, 使用 TD-error 对知识相关性进行评估, 以更新目标网络策略的方式, 不断更新知识库中的知识, 得到最优航迹规划的策略.



(a) 多机航迹协同评价



(b) 单机航迹协同评价

图 7　任务航迹代价变化曲线

(2) 平均奖励比较[19]

为进一步说明 KLD-PPDG 算法在航迹规划的有效性, 与现有遗传算法 (GA)、粒子群算法 (PSO)、PPDG 算法进行性能比较. 性能比较平台利用 Matlab 对数据进行编程, 形成各方法的导入压缩包, 从深度强化学习的奖励值这个指标衡量不同方法下的航迹规划效果. 深度强化学习的奖励值描述了在无人机群体决策过程中对威胁区域的避障效果, 表示为多台无人机在每个计算迭代次数内遵循最优策略所获得的平均奖励. 由图 8 可知, 本文所提 KLD-PPDG 算法在 500 以内的迭代次数时, 其平均奖励值时会出现微小的振幅, 这有利于算法跳出局部最优解区域, 在第 500 次迭代后平均奖励值迅速提高, 并于 3 500 次迭代后逐步收敛稳定, 奖励值固定在 16 附近, 这种情况主要受益于 PPDG 中行动者网络与评论家网络的相互作用, 使目标网络逐步靠近最优策略, 同时引入 MRL 知识相关性计算, 使无人机遇到威胁区域后采用先学习后动作的方式成功规划出新的航迹知识, 这种规划调整方式使无人机在当前状态下基于知识库历史经验做出更好的决策, 加强当前状态与知识库的联系, 提高算法的收敛速度. 而 PSO 算法在虽然在奖励最优值方面靠近 KLD-PPDG 算法, 摆脱了局部最优困扰, 但随着迭代次数的增加, 其值不稳定; GA 算法则由于采用启发式的方式进行航迹序列点计算, 计算空间较大, 导致平均奖励值的振幅较大且在短时间内无法稳定.

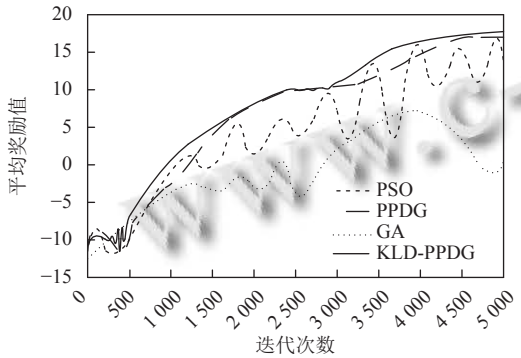


图 8 平均奖励比较

5 结论与展望

本文在分析马尔科夫决策过程的行为状态变化的基础上, 提出了基于决策知识学习的深度确定性策略梯度算法, 与其他基于深度学习的多无人机航迹协同任务规划研究不同的是, 本文将决策知识库作为深度学习经验池的知识储备, 态势环境和经验池的功能分别用于生成和存储知识, 并将意义接受性学习理论引入协同任务规划的学习训练中, 以增强决策知识的相关性学习能力. 但无人机群体航迹协同规划是一个复杂的大规模优化问题, 当无人机数量较大时会出现连续空间不稳定现象, 下一步将充分考虑空间和时序的约束, 进一步优化领域情景知识, 从多维空间数据展开研究.

参考文献

1 Reiter A. Optimal path and trajectory planning for serial robots. Wiesbaden: Springer-Verlag, 2020.
2 Završnik A. Drones and unmanned aerial systems. Cham: Springer, 2016.
3 Hu JW, Wang M, Zhao CH, *et al*. Formation control and collision avoidance for multi-UAV systems based on Voronoi partition. SCIENCE CHINA Technological Sciences, 2020, 63(1): 65–72. [doi: 10.1007/s11431-018-9449-9]
4 Zhou ZW, Luo DL, Shao J, *et al*. Immune genetic algorithm based multi-UAV cooperative target search with event-triggered mechanism. Physical Communication, 2020, 41: 101103. [doi: 10.1016/j.phycom.2020.101103]
5 Krishnan PS, Manimala K. Implementation of optimized dynamic trajectory modification algorithm to avoid obstacles for secure navigation of UAV. Applied Soft Computing, 2020, 90: 106168. [doi: 10.1016/j.asoc.2020.106168]
6 Zhen ZY, Chen Y, Wen LD, *et al*. An intelligent cooperative mission planning scheme of UAV swarm in uncertain dynamic environment. Aerospace Science and Technology, 2020, 100: 105826. [doi: 10.1016/j.ast.2020.105826]
7 Fan YC, Zhang QC, Tang YL, *et al*. Blitz-SLAM: A semantic SLAM in dynamic environments. Pattern Recognition, 2021, 121: 108225.
8 Shima T, Rasmussen S. UAV Cooperative Decision and Control. Pennsylvania: Society for Industrial and Applied Mathematics, 2008.
9 Meng YF, Xu JC, He JH, *et al*. A cluster UAV inspired honeycomb defense system to confront military IoT: A dynamic game approach. Soft Computing, 2021: 1–11. [doi: 10.1007/s00500-021-05881-4]
10 Domingue J, Fensel D, Hendler JA. Handbook of Semantic Web Technologies. Heidelberg: Springer, 2011. 43–75.
11 Ramirez-Amaro K, Yang YZ, Cheng G. A survey on semantic-based methods for the understanding of human

movements. Robotics and Autonomous Systems, 2019, 119: 31–50. [doi: 10.1016/j.robot.2019.05.013]

12 Hu X, Liu J. Ontology construction and evaluation of UAV FCMS software requirement elicitation considering geographic environment factors. IEEE Access, 2020, 8: 106165–106182. [doi: 10.1109/ACCESS.2020.2998843]

13 Emel 'yanov S, Makarov D, Panov AI, *et al*. Multilayer cognitive architecture for UAV control. Cognitive Systems Research, 2016, 39: 58–72. [doi: 10.1016/j.cogsys.2015.12.008]

14 杨清清, 高盈盈, 郭珏, 等. 基于深度强化学习的海战场目标搜寻路径规划. 系统工程与电子技术, 2021. http://kns.cnki.net/kcms/detail/11.2422.TN.20211117.0923.002.html. (2021-11-17)[2021-12-01].

15 郝忠孝. 时空数据库查询与推理. 北京: 科学出版社, 2010: 89–96.

16 Hu ZJ, Gao XG, Wan KF, *et al*. Relevant experience learning: A deep reinforcement learning method for UAV autonomous motion planning in complex unknown environments. Chinese Journal of Aeronautics, 2021, 34(12): 187–204. [doi: 10.1016/j.cja.2020.12.027]

17 Ning Q, Tao GP, Chen BC, *et al*. Multi-UAVs trajectory and mission cooperative planning based on the Markov model. Physical Communication, 2019, 35: 100717. [doi: 10.1016/j.phycom.2019.100717]

18 Wang C, Wu LZ, Yan C, *et al*. Coactive design of explainable agent-based task planning and deep reinforcement learning for human-UAVs teamwork. Chinese Journal of Aeronautics, 2020, 33(11): 2930–2945. [doi: 10.1016/j.cja.2020.05.001]

19 Guo T, Jiang N, Li BY, *et al*. UAV navigation in high dynamic environments: A deep reinforcement learning approach. Chinese Journal of Aeronautics, 2021, 34(2): 479–489. [doi: 10.1016/j.cja.2020.05.011]

20 Ausubel DP. Educational Psychology: A Cognitive View. New York: Holt, Rinehart and Winston, 1968.

21 白辰甲, 刘鹏, 赵巍, 等. 基于 TD-error 自适应校正的深度 Q 学习主动采样方法. 计算机研究与发展, 2019, 56(2): 262–280. [doi: 10.7544/issn1000-1239.2019.20170812]

22 沈林成, 牛轶峰, 朱华勇. 多无人机自主协同控制理论与方法. 第 2 版, 北京: 国防工业出版社. 2018: 187–189.

(校对责编: 孙君艳)