

面向应急响应群智感知的异构群体协作任务分配方法

路文浩, 赵勇, 季雅泰, 张琪, 许凯, 朱正秋

(国防科技大学系统工程学院, 湖南 长沙 410000)

摘要: 近年来, 异构群体协作感知成为群体智能领域的重要研究方向, 主要探讨不同类型的智能体 (如人类、无人机和无人车) 如何协同工作以感知和理解环境, 在应急救援等活动中被广泛应用。现有针对异构群体协作感知的任务分配算法大多难以平衡任务分配效果和求解效率, 且未能实现异构群体的深度协作。针对应急救援活动中环境状态的部分可观条件, 提出了一种“硬协作”异构群体协作模式, 并建立了部分可观异构群体协作感知任务分配问题模型。为求解该问题, 构建了多智能体协作框架, 并在此基础上提出了环境状态部分可观条件下的异构群体协作感知任务分配算法。实验结果表明, 相较于基线算法, 所提方法在任务完成率上更具优势, 4个场景中的平均任务完成率为 $84.40\% \pm 4.74\%$, 远高于目前最优基线算法的 $65.98\% \pm 4.97\%$ 。此外, 所提方法展现出良好的鲁棒性, 即使在感知场景变化时仍能保持较高的任务完成率, 显示出在动态环境中的应用潜力。

关键词: 应急响应; 异构群体协作感知; 任务分配; 多智能体强化学习

中图分类号: TP249

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2025.00488

Task allocation method for collaborative crowdsensing of heterogeneous entities in emergency response

LU Wenhao, ZHAO Yong, JI Yatai, ZHANG Qi, XU Kai, ZHU Zhengqiu

College of Systems Engineering, National University of Defense Technology, Changsha 410000, China

Abstract: Collaborative crowdsensing of heterogeneous entities has emerged as a pivotal research area in the field of collective intelligence in recent years, primarily focusing on how diverse agents, such as humans, drones, and unmanned vehicles, can collaboratively perceive and interpret their environments, which holds significant promise for applications in the emergency rescue and urban operations. However, existing task allocation algorithms for collaborative crowdsensing often struggle to balance allocation effectiveness with solution efficiency, and they frequently fail to facilitate deep collaboration among diverse entities. To address the partially observable conditions of environmental states during emergency rescue operations, a “hard collaboration” model for heterogeneous entities was introduced and a model for task allocation problem was formulated. To address this challenge, a multi-agent collaboration framework was developed and a task allocation algorithm for heterogeneous entities with partially observable environment states was proposed. Extensive experiments across four scenarios show that the proposed method outperforms the baseline algorithms in task completion rates, achieving an average of $84.40\% \pm 4.74\%$ compared with the best baseline's $65.98\% \pm 4.97\%$. Moreover, the proposed method exhibits strong robustness, maintaining commendable task completion rates even amid changes in perception scenarios and underscoring its potential for deployment in dynamic environments.

Key words: emergency response, collaborative crowdsensing of heterogeneous entity, task allocation, multi-agent reinforcement learning

收稿日期: 2025-01-27; 修回日期: 2025-03-18

通信作者: 朱正秋, zhuzhengqiu12@nudt.edu.cn

基金项目: 国防科技大学青年自主创新基金资助项目 (No. ZK-2023-21)

Foundation Item: The Youth Independent Innovation Foundation of NUDT (No. ZK-2023-21)

0 引言

自然灾害、工业事故等突发事件^[1-2]通常在短时间内造成重大的人员伤亡与财产损失,处置不当甚至会引发二次灾难。特别是,如果救援不及时,个体生存几率会显著降低。因此,在突发事件发生后,开展迅速且有效的应急救援至关重要^[3-4]。这需要及时的获取灾害或事故现场最新、最全面的环境信息,而这又依赖于对突发事件环境的快速、有效地感知^[5]。

传统的环境感知方式依赖于固定传感器组成的静态监测网络^[6-7]。这种网络主要通过预先布置在固定位置的传感器节点进行环境监测,存在部署维护成本高、监测范围有限等缺陷。近年来,随着物联网应用的规模扩大、移动智能终端的快速普及和群智计算模式的兴起,出现了一种基于移动设备(智能手机、个人穿戴设备、车载终端等)感知能力的获取方式——移动群智感知(MCS, mobile crowdsensing)^[8-9]。移动群智感知使用户成为感知过程中的一个节点,构成分布式传感网络,以众包^[10-11]的形式完成感知数据的收集、处理、计算和上传等工作,具有时空覆盖范围广、感知成本低和部署灵活等优势,已在智慧交通、环境治理等领域有了广泛的应用^[8]。

近年来,以无人机、地面无人平台为代表的无人智能终端逐渐发展成熟,能够执行感知任务的资源逐渐丰富。为提升群智感知效能,部分学者引入无人智能终端,与传统人类参与者组成异构感知群体,协同执行各类感知任务。Qi等^[12]将无人机引入群智感知活动,提出了一种无人机辅助数据收集框架。Zhu等^[13]将行人与地面无人车辆结合,提出了一种人车协作的感知方法。

任务分配^[14-15]是群智感知的核心研究问题,其本质是实现感知任务与感知主体之间的有效匹配,分配方案的质量将直接影响感知活动的效果和感知主体的协作效能。现有的任务分配求解方法主要包括精确方法、近似方法和深度强化学习方法。

精确求解方法指的是利用运筹优化等方法实现对任务分配方案的精确求解。精确求解方法^[16]得到的解质量高,理论上在时间无限时可以求得最优任务分配方案,但面临着求解时间长、复杂度高等问题。近似方法主要是通过模拟自然过程^[17-18]或基于

问题的特性^[19]来优化求解任务分配方案,目标是在有限时间内得到一个满意的可行解,包括贪婪算法、启发式算法和元启发式算法等。这类方法通常可以保证求解速度,但经常陷入局部最优,难以保证解的质量。深度强化学习方法^[20-23]将任务分配问题建模为一个多阶段的序贯决策过程,通过智能体与环境的交互来学习最优的策略。这类方法求解精度高、速度快,是解决任务分配问题的理想方法。Wang等^[20]研究了无人机辅助移动群智感知中的数据收集问题,提出了DRL-UCS框架,考虑在能源补给有限的条件下,平衡兴趣点数据收集新鲜度和分布,规划无人机的轨迹。Zhao等^[21]提出了一个基于异构图强化学习的任务分配框架HGRL-TA,有效地处理了依赖感知多任务分配问题中的多种复杂关系,如技能匹配、任务依赖和空间相邻等。此外,Zhao等^[22]提出了一种基于扩散模型的层次化MADRL方法gMADRL-VCS,用于优化空地车辆感知任务中感知主体的导航和路径规划策略,其中,扩散模型能够通过渐进式去噪过程,更好地捕捉多智能体环境中的复杂动态和策略交互。Han等^[23]构建了地震背景下,无人机、感知车辆和人类协作搜索救援的移动群智感知场景,提出了一种结合全局和局部信息处理的多智能体强化学习算法MANF-RL-RP。这是首个将异构感知主体协作感知应用到灾难响应中的方法。然而,这项工作没有考虑城市环境中感知主体的部分可观测性。

上述成果从异构主体协作感知建模、任务分配求解等方面研究了群智感知任务分配问题,具有一定的指导意义,但应急救援场景中的异构群体协作感知任务分配还存在很大的探索空间,以下问题尚未完全解决:1)应急救援场景需要异构感知主体更高层次的协作。应急救援场景环境复杂,仅依靠人类感知存在感知效率和精度低、感知范围不足以及安全风险高等问题;2)应急救援场景存在环境部分可观条件,感知主体无法了解全部环境状态,仅凭当前局部观察难以做出最优决策;3)应急救援活动对任务分配方案的时效性、准确性和泛化性要求较高,现有求解方法难以实现求解效率与分配效果的平衡,且泛化性能一般。

为解决上述3个问题,本文研究环境状态部分可观条件下的异构群体协作感知任务分配方法。针对问题1),本文建立多智能体协作框架(MACF, multi-

agent cooperative framework), 采用人类、无人机和地面无人平台3类主体协作执行感知任务, 并提出了异构群体协作的“硬协作”模式, 即3类感知主体独立执行感知任务, 地面无人平台负责为无人机补充电量。当地面无人平台探测到附近无人机电量不足时, 立刻前去为无人机换电, 而非继续执行感知任务。现有异构群体协作感知任务分配的前沿工作^[24-25]仅利用无人机这一单一感知主体执行感知任务, 人、地面无人平台等感知主体作为无人机感知的辅助, 这在很大程度上限制了感知任务的执行速度。此外, 鲜有工作研究地面无人平台为无人机补充电量与自身执行感知任务冲突时如何处理的问题, “硬协作”模式首次为这一问题提出了解决方案。针对问题2), 本文基于分布式部分可观马尔可夫决策过程(Dec-POMDP, decentralized partially observable Markov decision process)^[24]对异构群体协作感知过程进行建模, 并在算法部分建立集中式训练、分布式执行机制, 分别在全部环境状态已知和部分可观场景中进行训练和测试。针对问题3), 本文提出基于QTRAN(Q-Transformation)的异构群体协作感知应急救援(HECS-QTRAN-ER, heterogeneous entities cooperative sensing QTRAN emergency rescue)算法, 设计了环境信息输入模块和感知主体决策模块。有效地实现了分配效果与求解效率的平衡, 该算法还展示了一定的泛化性能。

本文的主要贡献如下:

1) 提出了一种多智能体协作框架MACF来解决环境状态部分可观条件下的异构群体协作感知任务分配问题。考虑环境状态部分可观条件, 本文给出了异构群体协作感知任务分配问题的形式化定义和表述, 并证明了其NP难问题特性。通过将该问题处理为Dec-POMDP过程, 可实现问题的求解。

2) 提出了“硬协作”这一全新的异构感知主体协作模式, 并证明了人类、无人机和无人平台3者的完全协作关系。3类感知主体都可以独立执行感知任务, 若地面无人平台探测到有无人机因为电量不足而无法移动, 那么下一时刻无人平台将会立刻为无人机进行换电, 而非执行感知任务。

3) 提出了一种新的异构群体协作感知任务分配算法HECS-QTRAN-ER。通过构建环境信息输入模块抽取环境特征, 并利用感知主体决策模块优化提升智能体自主决策水平。大量实验表明, HECS-

QTRAN-ER算法相比基线算法在任务完成率上表现最优, 且泛化性能更好。

1 场景建模与问题定义

1.1 应急救援环境建模

本文使用随机网格方法对应急救援环境进行简化建模, 将感知区域划分为一个个小的网格, 并在其中表示障碍物、任务等环境信息, 基于随机网格的应急救援环境建模如图1所示。

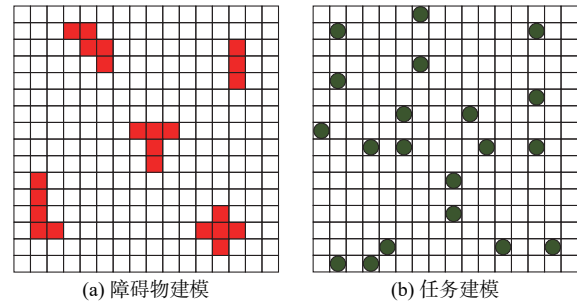


图1 基于随机网格的应急救援环境建模

将城市环境建模为一个二维区域集合 $P = \{p_1, p_2, \dots, p_m, \dots\}$, 集合的每一个元素代表了一个区域网格。 p_m 代表城市中的第 m 个区域, 每一个区域都可以成为一个障碍物或任务点, 有 $p_m = \langle po_m, pt_m \rangle$ 。 po_m 表示区域 m 的障碍物标识, 若存在障碍物, $po_m = 1$; 否则, $po_m = 0$ 。 pt_m 表示区域 m 的任务标识, 若存在任务, $pt_m = 1$; 否则, $pt_m = 0$ 。

为方便对感知过程进行刻画, 这里将整个感知过程划分为 TimeLimit 个感知周期 $T = \{t_1, t_2, \dots\}$, 感知周期之间以时刻划分, 每个感知周期的持续时间为 t_d , 感知周期为任务分配的最小时间单位。感知任务数量为 N_0 , 每个感知任务 $T = \langle Ttype, Tloc, Tdur \rangle$ 有 3 个属性, 分别表示任务类型、任务位置和任务执行时间。在异构群体协作感知过程中, 设定高空感知、地面快速感知和细节感知 3 种不同任务, 分别由无人机、地面无人平台和人单独执行, 每类感知任务只能由特定类别的感知主体完成。

1.2 感知主体建模

人类、无人机和地面无人平台具有不同的属性特点, 可在感知任务执行过程中承担不同的任务。3类感知主体的数量总和为 F 。人类 $w = \langle wLoc', wRge' \rangle$ 有 2 个属性, 分别表示当前时刻的位置与下一时刻的可移动范围; 无人机 $d = \langle dLoc', dPow', dCsp, dRge' \rangle$ 有 4 个属性, 分别表示当前时刻的位

置与电量、移动所消耗的电量 and 下一时刻的可移动范围；地面无人平台 $g = \langle gLoc', gRge', gDet' \rangle$ 有 3 个属性，分别表示当前时刻的位置、下一时刻的可移动范围和无人机的探测范围，地面无人平台只能探测到处于 $gDet'$ 范围内的无人机电量。WTRA、DTAR 和 GTRA 分别表示不同感知主体的路线集合。以 WTRA 为例，其组成元素为所有人类感知主体每一时刻的位置，可以表示为 $WTRA = \{\dots, wLoc_{k_1}^t, \dots, wLoc_{k_1}^{TimeLimit}, \dots\}$ ， k_1 表示任意一个人类感知主体。

无人平台与无人机采用一种“硬协作”的能源补充模式。“硬协作”模式是指如果无人平台 k_3 探测到无人机 k_2 因为电量不足而无法移动，那么下一时刻无人平台 k_3 将会立刻为无人机 k_2 进行换电，而非执行感知任务，如式(1)所示，其中， $gLoc_{k_3}^{t+1}$ 为无人平台 k_3 在 $t+1$ 时刻的位置， $gDet_{k_3}^t$ 为无人平台 k_3 在 $t+1$ 时刻可以探测到无人机电量的范围， $dCsp_{k_2}$ 为无人机 k_2 移动所消耗的电量。

$$\begin{aligned} gLoc_{k_3}^{t+1} &= dLoc_{k_2}^t, \\ dLoc_{k_2}^t &\in gDet_{k_3}^t \wedge dPow_{k_2}^t < dCsp_{k_2} \end{aligned} \quad (1)$$

对于无人机 k_2 的电量变化，有如下约束：如果无人平台 k_3 对无人机 k_2 进行换电，则无人机 k_2 电量变为 1。若无人平台 k_3 未对无人机 k_2 进行换电，且无人机 k_2 的电量小于下一感知阶段移动所需要的电量，则无人机 k_2 不会移动，电量保持不变。如果无人机 k_2 的电量可以支撑下一阶段的移动，则无人机 k_2 移动，电量相应减少。式(2)中， $dLoc_{k_2}^t = gLoc_{k_3}^t$ 指的是无人机 k_2 与地面无人平台 k_3 处于相同位置。

$$dPow_{k_2}^{t+1} = \begin{cases} 1, & dLoc_{k_2}^t = gLoc_{k_3}^t \\ dPow_{k_2}^t - dCsp_{k_2}, & dLoc_{k_2}^t \neq gLoc_{k_3}^t \\ & \wedge dPow_{k_2}^t \geq dCsp_{k_2} \\ dPow_{k_2}^t, & dLoc_{k_2}^t \neq gLoc_{k_3}^t \\ & \wedge dPow_{k_2}^t < dCsp_{k_2} \end{cases} \quad (2)$$

1.3 优化问题定义

本文中的异构群体协作感知任务分配问题可以建模为一个优化问题，优化目标为：在一定的时间内，最大化感知任务完成率，如式(3)所示，其中， N_t 表示 t 时刻未完成的任务数。

$$\max \frac{N_0 - N_t}{N_0} \quad (3)$$

在感知任务执行过程中，受障碍物阻挡的

影响，感知主体无法进入存在障碍物的区域，如式(4)所示；受移动速度的影响，感知主体在一个感知阶段内的移动范围是有限的，如式(5)所示。 $(w/g/d)Loc_k^t.po$ 是感知主体 k 在 t 时刻所处位置区域的障碍物标识。

$$(w/g/d)Loc_k^t.po = 0 \quad (4)$$

$$(w/g/d)Loc_k^{t+1} \in (w/g/d)Rge_k^t \quad (5)$$

在城市环境抽象的基础上，规定任务点和障碍物不能位于同一区域，如式(6)所示。

$$po_m + pt_m \leq 1 \quad (6)$$

x_{ki}^t 表示感知周期 t 内感知主体与感知任务间的匹配关系， $x_{ki}^t = 1$ 表示第 k 个感知主体执行第 i 个感知任务； $x_{ki}^t = 0$ 则表示不执行。感知主体在一个感知周期最多只能执行一个感知任务，而一个感知任务在一个感知周期内最多也只能被一个感知主体执行，如式(7)和式(8)所示，其中， F 为感知主体的数量总和。

$$\sum_{i=1}^{N_i} x_{ki}^t \leq 1, k = 1, 2, \dots \quad (7)$$

$$\sum_{k=1}^F x_{ki}^t \leq 1, i = 1, 2, \dots \quad (8)$$

同一感知主体在某一任务区域的连续停留时间等于该感知任务执行时间是感知任务完成的充分必要条件，用式(9)表示，其中， $Tloc_i$ 指的是感知任务 i 的位置。不考虑同一类别的不同感知主体在不同感知周期共同执行相同感知任务的情形。

$$\begin{aligned} N_t &= N_{t-1} - 1, \\ \exists k, \forall t \in [t, t + Tdur_t], (w/g/d)Loc_k^t &= Tloc_i \end{aligned} \quad (9)$$

应急救援场景中的异构群体协作感知任务分配问题可以表述为：对于给定的离散区域集合、感知主体集合和时间约束，确定感知主体的路线集合，使感知任务完成率最高，如式(10)所示。

$$\begin{aligned} &\text{确定 } WTRA, DTAR, GTRA, \\ \max & \frac{N_0 - N_t}{N_0}, \\ \text{s.t.} & \text{ 式(4)~(9)} \end{aligned} \quad (10)$$

2 基于 Dec-POMDP 的异构群体协作感知任务分配问题建模

本文选择 Dec-POMDP 模型建模应急救援场景中的异构群体协作感知任务分配问题主要有以下 3 点考虑。首先，在 Dec-POMDP 模型中，每一个智能体都有自己的局部观测空间，这与应急救援场景中感知主体的部分可观测性相符合。通过定义合理的

局部观测空间，可以准确地描述感知主体在应急救援场景中的感知能力。此外，Dec-POMDP考虑了多个智能体之间的相互作用，每个智能体的动作不仅影响自身的状态，还会对其他智能体的状态与环境状态产生影响，这适应异构感知主体的协作需求。最后，应急救援任务通常需要感知主体在多个时间步内进行序贯决策，而Dec-POMDP模型是一个动态过程模型，通过动作空间、状态转移概率和奖励函数等设定对智能体在不同时间步上的决策问题进行处理，考虑了智能体动作对环境状态的长期影响。

环境状态部分可观的异构群体协作感知任务分配问题可以用7元组 $\langle S, F, A, P, R, \gamma, O \rangle$ ^[25-26]表示，其中， S 为系统全局状态空间， F 为感知主体数量， A 为感知主体可行动作集合， P 为系统状态间的转移函数， R 为奖励函数， γ 为折扣因子， O 为感知主体的局部观测空间。

2.1 状态空间

状态空间如图2所示。在应急救援场景中，异构群体协作感知任务分配涉及诸多感知主体和感知任务点间的关系匹配。在任务分配时不仅要考虑感知主体间的相互作用关系，也要考虑城市障碍物、任务点和感知主体的位置关系。因此， t 时刻系统的状态可由式(11)表示。

$$s_t = \{ \text{obstDist}_t, \text{taskDist}_t, \text{agentDist}_t, \text{taskAttr}_t, \text{taskDur}_t, \text{taskAgent}_t \} \quad (11)$$

其中， obstDist_t 是障碍物的位置分布，对应图2(a)； taskDist_t 是任务点的位置分布，对应图2(b)； agentDist_t 是感知主体的位置分布，对应图2(c)；

taskAttr_t 是任务的属性，表示感知任务只能由某一类感知主体完成，对应图2(d)； taskDur_t 是任务的剩余执行时长，对应图2(e)； taskAgent_t 是任务与感知主体的匹配情况，即感知任务正在或者已经被哪一个感知主体完成，对应图2(f)。

2.2 局部观测空间

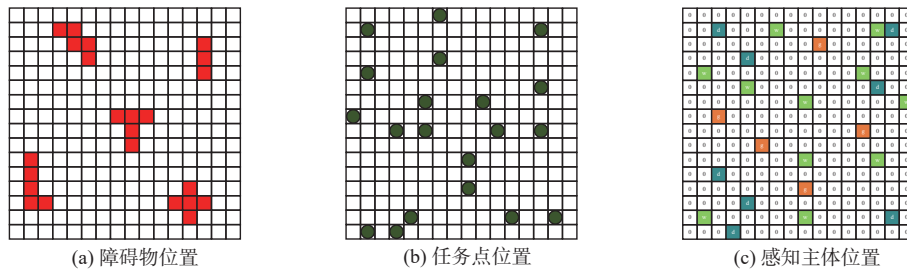
局部观测空间如图3所示。感知主体之间的通信条件是受限的，感知主体只能观察到可达范围内的有限信息。因此， t 时刻感知主体 k 的局部观测空间 o_k^t 如式(12)所示，其中， $(w/g/d)\text{Loc}_k^t$ 表示感知主体 k 当前时刻的位置，对应图3(a)； $(w/g/d)\text{Rge}_k^t$ 表示感知主体 k 下一时刻可以到达的范围，对应图3(b)； urge_k^t 表示感知主体 k 当前时刻的电量信息，包括当前时刻的剩余电量与单步移动所消耗的电量，对应图3(c)，人和无人平台没有这一属性； ID_k 表示感知主体 k 的编号，对应图3(d)。

$$o_k^t = \{ (w/g/d)\text{Loc}_k^t, (w/g/d)\text{Rge}_k^t, \text{urge}_k^t, \text{ID}_k \} \quad (12)$$

编号是感知主体的唯一标识。为了方便计算，编号采用独热编码，第 k 个感知主体编号的第 k 位为1，其余位置为0，通过编号中值为1的位置判断感知主体的类别。

2.3 动作空间

动作空间如图4所示。传统智能体路径规划问题中的动作空间一般为智能体下一时刻运动的方向，如图4(a)所示，智能体在8个方向中选择一个作为下一时刻移动的方向。但在本文中，考虑智能体的速度变化，设置智能体的动作空间为下一时刻智能



任务	T_1	T_2	T_3	T_4	T_5	...	T_N
(d) 任务属性						...	
(e) 任务时间						...	
(f) 任务执行主体	w_1	d_1	g_1	g_2	d_2	...	d_8

图2 状态空间

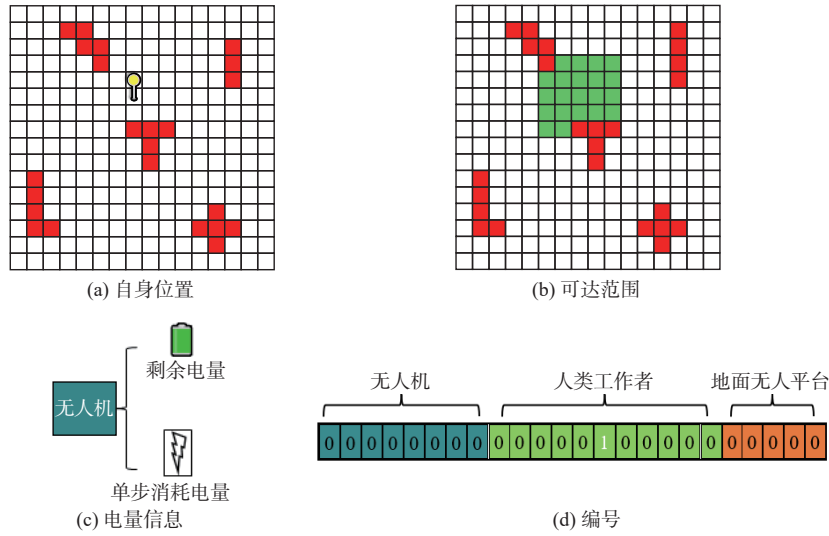


图3 局部观测空间

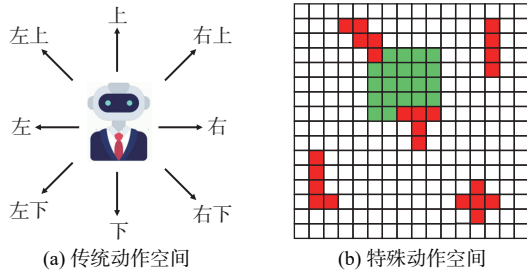


图4 动作空间

体能够到达的位置，由传统的运动方位选择转变为到达位置的选择，如图4(b)所示，智能体在绿色区域中选择一个格子作为下一时刻到达的区域。

2.4 状态转移与奖励函数

状态转移概率如式(13)所示。

$$p(s_{t+1}|s_t, \{a'_0, a'_1, \dots, a'_k, \dots\}) \quad (13)$$

将一个时刻感知任务的完成数量作为动作集合 $\{a'_0, a'_1, \dots, a'_k, \dots\}$ 的奖励 r_t ，如果感知主体所选择的动作超出了其可移动的范围，那么 r_t 设置-10以示惩罚，如式(14)所示。

$$r_t = \begin{cases} N_{t-1} - N_t, & \forall a'_k \in (w/g/d)Rge'_k \\ -10, & \text{其他} \end{cases} \quad (14)$$

在实际场景中，场景面积远比感知主体的移动范围要大，所以奖励值为正的概率将会极小，这导致了奖励稀疏问题，使模型难以训练^[23]。在实际的动作选择过程中，加入动作筛选机制可以很好地解决这一问题。在做出动作之前，感知主体先对动作进行逻辑检查，判断这一动作是否在移动范围内，如果不在这一范围内，则重新选择动作。

2.5 问题证明

引理1 异构群体协作感知任务分配问题是NP难的。

证明 设式(10)所表示的问题为问题1。对问题1进行简化，假设只有无人机可以执行感知任务，且任务执行过程不受人类和无人平台的影响。从而可以将问题1表示为问题2：

确定 **DTRA**,

$$\max \frac{N_0 - N_t}{N_0},$$

s.t. 式(6)~(8),

$$dLoc_{k_2}^t \cdot po = 0, \quad (15)$$

$$dLoc_{k_2}^{t+1} \in dRge'_{k_2},$$

$$N_t = N_{t-1} - 1,$$

$$\text{如果 } \exists k_2, \forall t \in [t, t + Tdur_t],$$

$$dLoc_{k_2}^t = Tloc_i$$

问题2是问题1的一个特例，若假设无人机集合中只有一架无人机，那么问题2可以转化为问题3：

确定 **dTRA**₀,

$$\max \frac{N_0 - N_t}{N_0},$$

s.t. $dLoc_0^t \cdot po = 0,$

$$dLoc_0^{t+1} \in dRge'_0,$$

$$po_0 + pt_0 \leq 1, \quad (16)$$

$$\sum_{i=1}^{N_0} x_{1i} \leq 1,$$

$$N_t = N_{t-1} - 1,$$

$$\text{如果 } \forall t \in [t, t + Tdur_t], dLoc_0^t = Tloc_i$$

问题3表示时间序列下的子集选择问题，是问题2的一个特例。由于子集选择问题是一类经典的NP难问题，所以问题3是一个NP难问题，进而可以证明问题1是一个NP难问题。证毕。

引理2 在问题1中，无人机、无人平台和人类工作者3者是完全协作关系。

证明 从任务执行的目标上来看，无人机、无人平台和人类工作者执行感知任务的最终目标都是完成尽可能多的任务，最终目标是一致的。且每一时刻完成的任务数量与感知主体在该时刻的位置是正相关的，即 $N_0 - N_t \propto \{(w/g/d)Loc'_k\}$ 。

同时，由于无人平台还负责为无人机换电，所以无人平台和无人机之间也是协作关系。故异构感知主体间是完全协作关系。证毕。

本文所提的HECS-QTRAN-ER算法和基线算法都是建立在引理2之上的。

3 异构群体协作感知任务分配框架与算法

为解决问题1，本文首先提出了MACF结构。在MACF中，每一个智能体负责控制一个人、无人机或地面无人平台^[27-28]。接着，设计环境信息输入模块和感知主体决策模块，提出HECS-QTRAN-ER算法。

3.1 多智能体协作框架MACF

MACF结构如图5所示。MACF主要包括环境信息输入模块、感知主体决策模块和混合网络模块。环境信息输入模块对环境状态和感知主体局部观察进行特征抽取；感知主体决策模块输出每一个智能体的 Q'_k 值；混合网络模块将 Q'_k 值作为输入，输出联合 Q'_{total} 值。MACF采用Actor-Critic架构^[29]，

所有网络均存在评价网络与目标网络两种网络。两种网络采用异步更新策略，评价网络的参数每次训练都进行更新，而目标网络每隔一定周期才会更新。

图5(a)部分为环境信息输入模块。为了提高智能体在复杂环境中的学习效果和算法收敛速度，本文构建了两个结构相同、参数不同的卷积神经网络^[30]CNN₁和CNN₂，分别用于抽取环境状态和感知主体局部观察特征。 s' 和 O' 经抽取特征得到 s'_c 和 O'_c 。卷积神经网络包括全连接层、卷积层和池化层等结构，采用的激活函数为ReLU。

图5(b)部分为感知主体决策模块。为了使智能体充分利用历史信息，本文为每一类智能体单独设置了一个决策模块。模块决策过程基于循环神经网络^[31]，包括动作价值生成、动作选择和动作筛选等过程。循环神经网络内嵌隐藏值 h_k^{t-1} ，代表了智能体 k 的历史记忆。智能体的决策过程分为3步，首先，模块输入智能体的局部观察特征 $O'_{c,k}$ 和 h_k^{t-1} ，计算得到当前时刻的隐藏值 h'_k 与所有动作的 Q_k 值；然后，依据 ϵ -greedy策略^[32]进行动作选择；最后，对选中动作进行过滤，如果其在可行动作范围内，则输出 Q'_k 值，否则调整隐藏值，重新计算并选择动作 Q_k 值。

图5(c)部分为混合网络模块。混合网络模块负责对系统全局进行价值评估，包含联合动作价值网络 Q_{tot} 、差值网络 V 和一个求和器。 Q_{tot} 和 V 以 s'_c 、 O'_c 和 h'_c 为输入，输出 Q'_{total} 和 V' 。求和器将所有智能体的 Q'_k 求和得到 Q'_{total} 。

3.2 HECS-QTRAN-ER算法

基于MACF，本文提出了HECS-QTRAN-ER

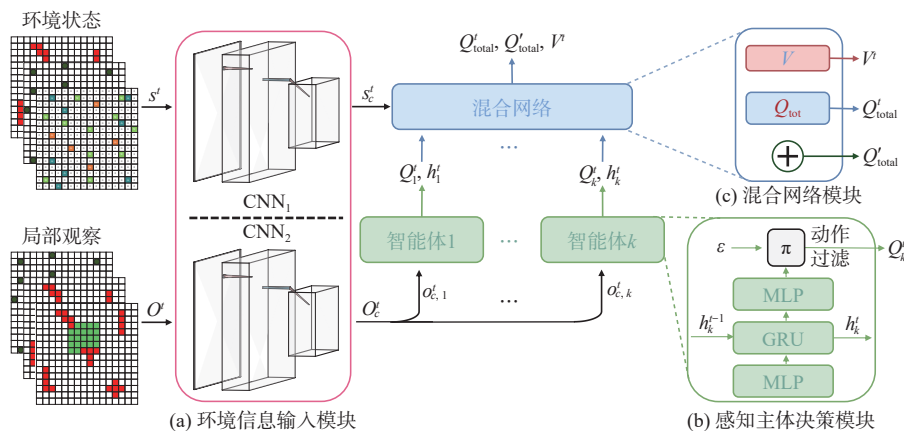


图5 MACF结构

算法, 如算法1所示。算法采用集中式训练, 分布式执行的架构^[33], 训练阶段存在智能体决策模块和混合网络模块, 按照算法1方式训练网络参数, 环境状态信息作为已知信息经过CNN₁和CNN₂输入混合网络中。执行阶段只有智能体决策模块, 智能体依据训练好的策略独立选择动作, 不存在环境状态信息的输入过程, 也不存在混合网络模块。

在训练阶段, 算法通过模拟环境与智能体的交互来学习最优策略。具体步骤为: 首先, 初始化相关神经网络和经验池参数, 并将评价网络的参数复制到目标网络。然后, 进行仿真循环, 在有限时间内, 智能体与环境交互产生经验数据, 并存储在经验池中。每个时刻, 智能体根据当前状态和历史信息选择动作, 环境根据动作更新状态并返回奖励。最后, 当训练轮次达到规定数量时, 在经验池中随机抽取数据, 基于梯度下降策略更新评价网络参数, 以最小化损失函数。每隔一段时间, 将评价网络的参数值复制到目标网络, 以稳定训练过程。

算法采用的损失函数如式(17)所示。 L_{td} 为真值估计损失, 采用DQN算法的更新方式^[34], 通过最小化 Q'_{total} 的时间差分误差来估计真实的联合动作价值, θ 是评价网络参数, θ^- 是目标网络参数。 L_{opt} 和 L_{nopt} 为函数分解损失, 用于确保函数分解满足个体-全局最大 (IGM, individual-global-max) 条件^[35], 具体的数学证明较为复杂, 详见文献[36]及其附件。 λ_{opt} 和 λ_{nopt} 为两个损失的权重系数。 Q_{tot} 为混合网络的评价网络, Q'_{tot} 为求和网络, \hat{Q}_{tot} 为混合网络的目标网络。 y^{dqn} 是通过深度Q网络 (DQN, deep-Q-network) 算法计算的目标值, 与减号前一项真值共同构成时间差分误差。具体计算方式可以参考文献[34]。 A^t 是所有智能体当前时刻的动作, \bar{A}^t 是所有智能体当前时刻除最优动作之外的非最优动作集合。

$$\begin{aligned}
 L &= L_{td} + \lambda_{opt} L_{opt} + \lambda_{nopt} L_{nopt}, \\
 L_{td} &= (Q_{tot}(s^t, A^t) - y^{dqn}(r^t, s^{t+1}; \theta^-))^2, \\
 L_{opt} &= (Q'_{tot}(s^t, \bar{A}^t) - \hat{Q}_{tot}(s^t, \bar{A}^t) + V(s^t))^2, \\
 L_{nopt} &= (\min [Q'_{tot}(s^t, A^t) - \hat{Q}_{tot}(s^t, A^t) + V(s^t), 0])^2
 \end{aligned}
 \tag{17}$$

算法1 HECS-QTRAN-ER

输入 环境信息, 仿真时长

输出 空间卷积网络 $cnnSpace_{1,2}$, 智能体评估网络 $evalAgent_l (l = 1, 2, 3)$, 混合网络 $evalTotal$, 差

值网络 V

- 1) 初始化 $cnnSpace_{1,2}$, $evalAgent_l$, $evalTotal$, V , 参数更新频率 P , 经验池 E ;
- 2) 分别将 $evalAgent_l$ 和 $evalTotal$ 的网络参数复制到各自的目标网络 $tgtAgent_l$ 和 $tgtTotal$;
- 3) 训练轮次 $TrainNum = 0$;
- 4) **while** $cnnSpace_{1,2}$, $evalAgent_l$, $evalTotal$ 和 V 没有收敛 **do**
- 5) **for** 仿真时间从开始到中止 **do**
- 6) 初始化经验池单局存储 E_{ji} (第 j 局, t 时刻, 第 k 个智能体);
- 7) 获得 s^{jt} 和当前时刻所有智能体各自的观察 $\{o_k^{jt}\}_{k=1}^F$;
- 8) $s_c^{jt} = cnnSpace_1(s^{jt})$;
- 9) $\{o_{c,k}^{jt}\}_{k=1}^F = cnnSpace_2(\{o_k^{jt}\}_{k=1}^F)$;
- 10) $\{Q_k^{jt}\}_{k=1}^F = evalAgent_l(\{o_{c,k}^{jt}\}_{k=1}^F, \{h_k^{jt}\}_{k=1}^F)$;
- 11) 基于 ϵ -greedy 策略由 $\{Q_k^{jt}\}_{k=1}^F$ 获得 $\{a_k^{jt}\}_{k=1}^F$;
- 12) 基于 $\{a_k^{jt}\}_{k=1}^F$ 更新智能体与环境信息;
- 13) 参照式(14)获得 r_{jt} ;
- 14) 参照行 7)~9) 获得 $s_c^{j,t+1}$ 和 $\{o_{c,k}^{j,t+1}\}_{k=1}^F$;
- 15) **if** 仿真时间终止 **then**
- 16) $te^{jt} = 0$;
- 17) **else**
- 18) $te^{jt} = 1$;
- 19) **end if**
- 20) 将 s_c^{jt} , $s_c^{j,t+1}$, $\{o_{c,k}^{jt}\}_{k=1}^F$, $\{o_{c,k}^{j,t+1}\}_{k=1}^F$, $\{h_k^{jt}\}_{k=1}^F$, $\{h_k^{j,t+1}\}_{k=1}^F$ 和 r^{jt} , te^{jt} 加入 E_{ji} ;
- 21) **end for**
- 22) 将 E_{ji} 加入 E_j ;
- 23) $j = j + 1$;
- 24) 在 E 中随机抽取训练数据; //以下基于训练数据
- 25) $\{Q_k^t\}_{k=1}^F = evalAgent(\{o_{c,k}^t\}_{k=1}^F, \{h_k^t\}_{k=1}^F)$;
- 26) $Q'_{total} = evalTotal(s_c^t, \{h_k^t\}_{k=1}^F, \{a_k^t\}_{k=1}^F)$;
- 27) Q_{total} 的目标网络值 $tgtQ'_{total} = tgtTotal(s_c^{t+1}, \{h_k^{t+1}\}_{k=1}^F, \{a_k^t\}_{k=1}^F)$;

- 28) 根据式(17)计算 L ;
- 29) 基于 L 采用梯度下降策略更新 $cnnSpace_{1,2}$, $evalAgent_i$, $evalTotal$ 和 V ;
- 30) $TrainNum = TrainNum + 1$;
- 31) **if** $TrainNum \% P == 0$ **then**
- 32) 分别将 $evalAgent_i$ 和 $evalTotal$ 的网络参数复制到 $tgtAgent_i$ 和 $tgtTotal$;
- 33) **end if**
- 34) **end while**

其中, 16)的 te 为仿真时间是否终止的指示符。如仿真时间终止, 则 te 值为0, 反之则为1。

4 仿真实验

本节在仿真生成的感知场景中开展实验, 实验场景见表1。

4.1 实验设置

本文共设置3组实验, 其中, 实验1对比同构和异构感知主体的任务执行效果; 实验2与基线算法进行对比, 验证本文所提算法的有效性; 实验3探索算法的鲁棒性。

由于神经网络设计本身并不是论文研究的重点, 所以, 对于所有的神经网络, 在控制实验中都采用相同的结构, 网络每层节点的具体数量取可以使其

收敛的最小值。实验参数见表2, 其中, M 指的是经验回放池的容量, 输入通道数和输出通道数分别指的是卷积神经网络的输入尺寸和输出尺寸, P 指的是参数更新频率。对于一个感知任务, 越早被执行越好, 因此, 这里设置衰减系数为一个较小的值, 使感知主体更关注于当下感知任务的执行。

实验中对比的基线算法有以下4个。

1) Greedy算法: 对文献[23]提出的Greedy-SC-RP算法稍作改进以适配本问题, 应用贪婪方法依次为3类感知主体分配感知任务。

2) MANF-RL-RP算法^[23]: 一种用于应急灾难救援的人、无人机和车辆协同感知路径规划强化学习方法, 这是目前异构群体应急救援场景中最新且感知效果最好的算法。

3) HECS-QMIX-ER算法: 将HECS-QTAN-ER算法的混合网络部分替换为QMIX算法^[35]的混合网络部分。

4) HECS-QTRAN-ER-temp算法: 移除HECS-QTRAN-ER算法中的循环神经网络。

4.2 感知主体对比实验

为了验证异构群体协作感知效果与同构群体相比是否存在显著提高, 本文分别在单一感知主体(场景11)与异构感知主体(场景2)场景中使用HECS-

表1 实验场景

场景编号	感知主体初始位置	感知主体数量	场景尺寸	任务数量(分布)	任务类型数量	任务执行时间(数量)/个	障碍数量	移动半径	电量消耗
1	相同	6/15/3	16×16	120(随机)	30/75/15	1/2(96/24)	20	8/3/5	0.3
2	随机	6/15/3	16×16	120(随机)	30/75/15	1/2(96/24)	20	[7, 9]、[2, 4]、[4, 6]	[0.2, 0.4]
3	Check-in	6/15/3	16×16	120(随机)	30/75/15	1/2(96/24)	20	[7, 9]、[2, 4]、[4, 6]	[0.2, 0.4]
4	Check-in	6/15/3	16×16	120(Check-in)	30/75/15	1/2(96/24)	20	[7, 9]、[2, 4]、[4, 6]	[0.2, 0.4]
5	随机	4/10/2、6/15/3、8/20/4	16×16	120(随机)	30/75/15	1/2(96/24)	20	[7, 9]、[2, 4]、[4, 6]	[0.2, 0.4]
6	随机	6/15/3	12×12、16×16、20×20	120(随机)	30/75/15	1/2(96/24)	20	[7, 9]、[2, 4]、[4, 6]	[0.2, 0.4]
7	随机	6/15/3	16×16	96/120/144(随机)	24/60/12、30/75/15、36/90/18	1/2(96/24)	20	[7, 9]、[2, 4]、[4, 6]	[0.2, 0.4]
8	随机	6/15/3	16×16	120(随机)	40/40/40、30/75/15	1/2(96/24)	20	[7, 9]、[2, 4]、[4, 6]	[0.2, 0.4]
9	随机	6/15/3	16×16	120(随机)	30/75/15	1/2(72/48)、1/2(96/24)、1/2/3(72/36/12)	20	[7, 9]、[2, 4]、[4, 6]	[0.2, 0.4]
10	随机	6/15/3	16×16	120(随机)	30/75/15	1/2(96/24)	20/40/60	[7, 9]、[2, 4]、[4, 6]	[0.2, 0.4]
11	随机	0/15/0	16×16	120(随机)	0/75/0	1/2(96/24)	20	[2, 4]	—

表2 实验参数

参数	值	参数	值
学习率	0.000 1	折扣因子 γ	0.7
经验回放池容量 M	5 000	探索率 ϵ	1/0/1/32
优化器	RMSprop	输入通道数	7/3
卷积核尺寸	3	输出通道数	10
参数更新频率 P	200	批大小	32

QTRAN-ER 算法进行 10 次任务分配实验，比较了感知任务完成率。感知主体效果对比见表 3。在场景设置上，场景 11 去除了除人之外的其他感知主体及其任务。实验结果显示，在相同条件下，异构群体协作比同构群体协作具有更高的任务完成率；相同类别的感知主体采用异构群体协作感知方式的任务完成率远高于同构感知方式。这说明异构群体协作感知具有更好的感知效果。

表3 感知主体效果对比

感知时间上限	单一感知主体任务完成率	异构感知主体中人类任务完成率	异构感知主体总任务完成率
6	49.64%	68.20%	64.21%
9	52.39%	753.20%	72.49%
12	55.14%	79.32%	71.86%

4.3 算法有效性实验

为了比较 HECS-QTRAN-ER 算法与 4 个基线算法的性能，本文在 (1/2/3/4) 4 个感知场景中，设定不同的随机数种子，分别开展 10 次感知时间上限为 9 的仿真实验，算法在每个场景中结果的平均值与 95% 置信区间如图 6 所示。

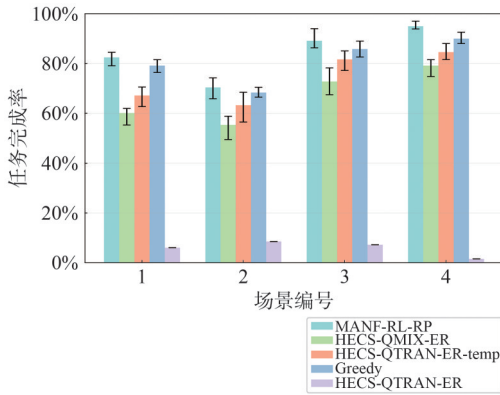


图 6 算法在每个场景中结果的平均值与 95% 置信区间

在所有场景中，HECS-QTRAN-ER 算法的任务完成率最高，尤其在第 4 个场景中达到了最高值，接近 96%。MANF-RL-RP 算法在 4 个场景中的任务

完成率均低于 HECS-QTRAN-ER 算法。Greedy 算法的任务完成率最低，在所有场景中都低于 10%。HECS-QMIX-ER 算法的任务完成率在所有场景都处于中等水平，略低于 HECS-QTRAN-ER 算法，但高于 MANF-RL-RP 算法。HECS-QTRAN-ER-temp 算法的任务完成率在所有场景中都较高，与 HECS-QTRAN-ER 算法相近，但略低。总体来看，HECS-QTRAN-ER 算法在所有场景中的表现最佳，而 Greedy 算法的表现最差。MANF-RL-RP 算法、HECS-QMIX-ER 算法和 HECS-QTRAN-ER-temp 算法的表现则介于两者之间。

从置信区间的分析来看，HECS-QTRAN-ER 算法在所有场景中的置信区间均较窄，表明其性能稳定且具有较高的可信度。相比之下，MANF-RL-RP 算法的稳定性较低，在场景 2 中，其置信区间达到了 49.47% 到 58.85%，波动较大。HECS-QMIX-ER 和 HECS-QTRAN-ER-temp 算法的置信区间宽度适中，表明它们的性能在多次实验中具有一定的波动性，但总体表现仍优于 MANF-RL-RP 算法。

场景 1 和场景 2 中除 Greedy 算法外其他算法表现不如场景 3 和场景 4 的原因为：场景 1 和场景 2 中感知主体初始时刻分别处在相同位置与随机位置，初始位置分布没有规律，算法难以取得较高的任务完成率，且方差较大；场景 3 和 4 中的感知主体分布满足 Check-in^[35]分布，更贴近真实环境，算法能够获得比较高的任务完成率，方差较小。

此外，为了验证 HECS-QTRAN-ER 算法与 MANF-RL-RP 算法相比在任务完成率上具有显著差异，本文在 4 个场景中分别对两个算法进行方差分析。结果显示，两个算法在 4 个场景中具有显著性差异假设下的 p 值分别为 5.63×10^{-3} 、 1.17×10^{-3} 、 1.48×10^{-4} 和 2.18×10^{-5} ，均小于 0.05，这说明 HECS-QTRAN-ER 算法与 MANF-RL-RP 算法在任务完成率上具有显著性差异。

本文算法在 4 个场景中相比基线算法均获得了一定的性能提升，特别是在感知主体位置随机分布的场景 2 中，仍获得了 70% 左右的任务完成率。HECS-QTRAN-ER 算法与 HECS-QTRAN-ER-temp 算法相比平均性能提升了 3.5%，这说明在局部观察问题中，感知主体结合历史信息可以做出更好的动作选择，验证了设置感知主体决策模块的有效性。HECS-QTRAN-ER 算法在 4 个场景中多次实验的平均任务

完成率为 $84.40\% \pm 4.74\%$ ，远高于 MANF-RL-RP 算法的 $65.98\% \pm 4.97\%$ ，显示出所提算法的卓越表现。

4.4 鲁棒性实验

为了更好地研究算法的性能，本节以 5/6/7/8/9/10 共 6 组感知场景为基础对所提的任务分配算法进行了鲁棒性分析。

鲁棒性实验具体过程为：首先，在单一场景中进行算法训练，然后，通过控制变量的方式在该场景的基础上生成 50 组全新的随机实验场景，并在其中做性能测试，得到算法在新实验场景中的任务完成率。感知时间上限统一设置为 9，算法鲁棒性实验结果见表 4。

总体上看，在控制变量的随机场景中，MANF-RL-RP 和 HECS-QTRAN-ER 两个算法的效果均存在不同程度的下降，但在每个实验场景中，HECS-QTRAN-ER 算法的实验结果总是高于 MANF-RL-RP 算法。这说明在环境不确定的情况下，HECS-QTRAN-ER 算法仍能保持较好的性能，显示出更强的鲁棒性。

具体来看，与其他参数相比，两种算法对任务属性和感知主体位置表现得尤为敏感，效果下降程度比较严重，而对任务执行时间和障碍物位置的改变较为稳健。

原因如下：任务属性规定了完成该任务的感知

主体类别，任务属性的改变大大增加了感知主体的决策难度；初始位置变化使感知主体需要重新分析环境信息；任务执行时间和障碍物位置的变化在本质上没有改变原有的训练场景，因此实验结果变化不大。

4.5 算法复杂度与训练过程分析

HECS-QTRAN-HECS 算法的复杂度如式(18)所示，其中， H 为循环迭代次数， T 是一个 episode 的时间步长， B 为所有神经网络的参数总和，神经网络的前向和后向传播复杂度为 $O(B)$ ， E 为经验池的容量，经验池的更新操作为线性复杂度 $O(E)$ ， G 为梯度计算的复杂度。

$$O(H \times (T \times (B + E) + G)) \quad (18)$$

基于场景 2，对 HECS-QMIX-ER 算法和 HECS-QTRAN-ER 算法的训练过程进行分析。训练效果展示如图 7 所示，展示了在模型训练期间，随着训练轮次的增加，损失值和任务完成率的变化。可以在曲线中观察到两个明显的特征。

1) 从图 7(a) 可以看出，在前 6 000 轮次中，两种算法的训练损失值已经达到了较低的水平。然而，如图 7(b) 所示，在该点后，两种算法的收敛速度趋缓。因为该阶段智能体未实现对环境空间的完全感知，基于局部观察做出的决策难以达到最优水平。

2) 从图 7(b) 可以看出，在训练初期，与 HECS-QTRAN-ER 算法相比，HECS-QMIX-ER 算法的收

表 4 算法鲁棒性实验结果(任务完成率)

场景编号	训练效果		改变任务执行时间		改变任务属性		改变障碍位置		改变感知主体位置	
	MANF-RL-RP	HECS-QTRAN-ER	MANF-RL-RP	HECS-QTRAN-ER	MANF-RL-RP	HECS-QTRAN-ER	MANF-RL-RP	HECS-QTRAN-ER	MANF-RL-RP	HECS-QTRAN-ER
5-1	51.3%	72.2%	32.5%	44.7%	18.4%	28.4%	56.7%	67.7%	18.9%	30.8%
5-2	63.2%	81.0%	41.5%	56.1%	25.9%	33.2%	66.3%	75.1%	25.6%	35.9%
5-3	66.3%	84.4%	52.6%	64.3%	29.1%	36.8%	70.4%	78.9%	21.6%	35.8%
6-1	63.1%	86.3%	46.2%	63.5%	28.4%	36.5%	62.4%	77.3%	43.2%	43.0%
6-2	62.5%	85.0%	40.9%	57.0%	25.1%	33.6%	61.9%	77.8%	22.6%	32.2%
6-3	61.7%	81.3%	43.2%	55.3%	19.0%	31.0%	60.4%	76.3%	21.0%	29.3%
7-1	60.9%	84.0%	46.6%	61.0%	24.3%	35.6%	68.8%	76.6%	25.6%	36.0%
7-2	52.5%	77.0%	47.3%	56.0%	22.5%	31.5%	61.5%	71.9%	29.8%	32.7%
7-3	63.2%	81.3%	31.5%	45.6%	16.3%	29.1%	67.3%	71.7%	30.0%	31.9%
8-1	49.8%	62.3%	32.6%	43.6%	15.4%	21.6%	50.1%	57.8%	24.3%	30.3%
8-2	69.0%	79.0%	41.6%	51.0%	13.6%	28.0%	72.1%	75.3%	30.5%	30.4%
9-1	39.6%	64.3%	30.9%	46.2%	20.9%	26.8%	50.5%	59.9%	21.5%	30.3%
9-2	62.8%	82.8%	43.2%	55.6%	24.2%	30.9%	65.4%	76.7%	25.6%	31.2%
9-3	53.6%	71.0%	28.1%	36.2%	21.6%	29.9%	61.5%	68.7%	29.4%	33.8%
10-1	69.4%	83.2%	47.3%	58.9%	33.5%	35.8%	72.4%	78.7%	25.7%	34.2%
10-2	70.1%	78.4%	48.0%	55.5%	25.9%	31.3%	63.5%	71.9%	24.1%	32.9%
10-3	52.4%	73.2%	46.3%	53.1%	21.3%	29.0%	54.3%	64.6%	20.9%	31.8%

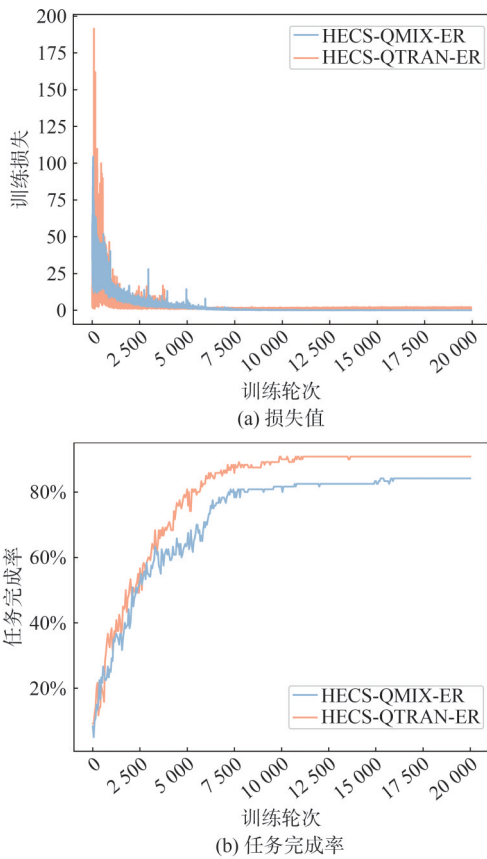


图7 训练效果展示

敛速度更快；但在训练后期，HECS-QTRAN-ER 算法收敛的任务完成率更高。由于 HECS-QMIX-ER 算法的混合网络结构较简单，所以在训练初期收敛迅速。HECS-QTRAN-ER 算法的混合网络结构较复杂，可以更好地捕捉感知主体间的交流与协作，在训练后期达到了更高的任务完成率。

5 讨论

本文研究了面向应急救援场景的异构群体协作感知任务分配方法。具体而言，本文提出了 MACF，采用人、无人机和地面无人平台协作执行感知任务，其中，3 类感知主体都可以独立执行感知任务，地面无人平台负责为无人机补充电量。针对地面无人平台电量补充与执行感知任务冲突的情况，提出了异构群体的“硬协作”模式。当发生冲突时，地面无人平台立刻前往为无人机换电，而非继续执行感知任务。然后，针对应急救援场景存在环境部分可观测的条件与多步序贯决策的需求，本文使用 Dec-POMDP 模型对感知过程进行数学建模，并证明了两个关键引理。最后，本文提出了基

于 MACF 框架的异构群体协作感知任务分配算法——HECS-QTRAN-ER 算法，并通过与基线算法对比，证明了该算法的性能。

但本文研究还存在以下不足。首先，场景建模与现实条件还有一定差距。现实场景是三维连续空间，二维网格并不能很全面地表示现实场景的信息。然后，在感知主体移动的建模表示上，仅确定了感知主体移动的目标位置，缺乏移动过程中的连续路径规划。此外，本文未考虑任务执行过程中感知主体的安全性问题。无人机和无人地面平台如何进行安全充电以及无人机如何进行安全避障在本文中均被简化。此外，作为一种离线任务分配方法，人类参与者的隐私保护也是一个重要问题，本文尚未对该问题进行考虑。

针对以上不足，未来工作可从以下几个方面展开。首先，研究更细致的场景建模与感知主体路径规划，考虑三维立体空间中的异构群体协作感知问题并为其移动规划具体的路径。具体而言，利用实景三维建模^[37-38]方法获取真实世界某一区域的深度点云数据，并通过语义标注、点云分割等方法实现对障碍物和感知任务的建模表示。将感知主体的动作空间建模为连续动作空间，并把启发式路径规划算法的原理嵌入深度强化学习方法中的奖励函数，实现对感知主体的连续路径规划。其次，探索提高算法泛化性能的方法，使其在一个场景中训练后可以快速迁移到其他多个场景（甚至更复杂的场景）中。将 DeepSeek-R1^[39]等推理大模型与深度强化学习方法结合，采用大小模型协同的方法，通过知识蒸馏，将大模型的推理能力迁移到深度强化学习模型中，从而提高方法的泛化能力。此外，对感知过程中感知主体的安全性问题进行研究。考虑利用人工势场法和栅格法等方法实现感知主体避障，同时设计感知主体意外退出备选机制。如果某一感知主体因安全问题无法继续执行感知任务，考虑就近安排同类感知主体接力完成剩余感知任务。最后，考虑使用多种方法保护参与者的隐私信息。对于提前获取的参与者能力信息，可以通过隐去姓名，身份证号等关键因素实现信息保护；对于任务执行过程中的位置信息，可以采取区块链、数据加密等方法^[40-41]实现信息保护。

6 结束语

本文研究了环境状态部分可观条件下的异构群

体协作感知任务分配问题, 设计了“硬协作”异构群体协作模式, 证明了3类感知主体是完全协作的。构建了多智能体协作框架MACF, 并基于此框架提出了HECS-QTRAN-ER任务分配算法, 实现了分配效果与求解效率的平衡。仿真实验结果表明, 本文所提方法在任务完成率上相较于基线算法更具优势, 4个实验场景中的平均任务完成率为 $84.40\% \pm 4.74\%$, 远高于目前最好的基线算法的 $65.98\% \pm 4.97\%$, 且当感知场景变化时, 仍能保持较好的任务完成率, 是一个有效且鲁棒的任务分配算法。

参考文献:

- [1] KUGLITSCH M M, PELIVAN I, CEOLA S, et al. Facilitating adoption of AI in natural disaster management through collaboration[J]. *Nature Communications*, 2022, 13: 1579.
- [2] SOLTANZADEH A, YARANDI M S, JAZARI M D, et al. Incidence investigation of accidents in chemical industries: a comprehensive study based on factor analysis[J]. *Process Safety Progress*, 2022, 41(3): 531-537.
- [3] FENG Y, CUI S Z. A review of emergency response in disasters: present and future perspectives[J]. *Natural Hazards*, 2021, 105(1): 1109-1138.
- [4] QU C Y, SORBELLI F B, SINGH R, et al. Environmentally-aware and energy-efficient multi-drone coordination and networking for disaster response[J]. *IEEE Transactions on Network and Service Management*, 2023, 20(2): 1093-1109.
- [5] SHI K Z, PENG X P, LU H, et al. Application of social sensors in natural disasters emergency management: a review[J]. *IEEE Transactions on Computational Social Systems*, 2023, 10(6): 3143-3158.
- [6] WANG Z N, CAO Y, JIANG K, et al. When crowdsensing meets smart cities: a comprehensive survey and new perspectives[J]. *IEEE Communications Surveys & Tutorials*, 2025, 27(2): 1101-1151.
- [7] BELLINI P, NESI P, PANTALEO G. IoT-enabled smart cities: a review of concepts, frameworks and key technologies[J]. *Applied Sciences*, 2022, 12(3): 1607.
- [8] SUHAG D, JHA V. A comprehensive survey on mobile crowdsensing systems[J]. *Journal of Systems Architecture*, 2023, 142: 102952.
- [9] KIM J W, EDEMACU K, JANG B. Privacy-preserving mechanisms for location privacy in mobile crowdsensing: a survey[J]. *Journal of Network and Computer Applications*, 2022, 200: 103315.
- [10] HETTIACHCHI D, KOSTAKOS V, GONCALVES J. A survey on task assignment in crowdsourcing[J]. *ACM Computing Surveys*, 2022, 55(3): 1-35.
- [11] ANG K L M, SENG J K P, NGHARAMIKE E. Towards crowdsourcing Internet of things (crowd-IoT): architectures, security and applications[J]. *Future Internet*, 2022, 14(2): 49.
- [12] QI H Q, LI Y, XU Y Q, et al. UAV-assisted mobile crowdsensing systems empowered by wireless power transfer and adaptive compression techniques[J]. *IEEE Wireless Communications Letters*, 2024, 13(9): 2487-2491.
- [13] ZHU Z Q, ZHAO Y, CHEN B, et al. A crowd-aided vehicular hybrid sensing framework for intelligent transportation systems[J]. *IEEE Transactions on Intelligent Vehicles*, 2023, 8(2): 1484-1497.
- [14] WANG C Q, MEI D, WANG Y, et al. Task allocation for multi-AUV system: a review[J]. *Ocean Engineering*, 2022, 266: 112911.
- [15] POUDEL S, MOH S. Task assignment algorithms for unmanned aerial vehicle networks: a comprehensive survey[J]. *Vehicular Communications*, 2022, 35: 100469.
- [16] WANG E, YANG Y J, WU J, et al. An efficient prediction-based user recruitment for mobile crowdsensing[J]. *IEEE Transactions on Mobile Computing*, 2018, 17(1): 16-28.
- [17] HAN D, JIANG H, WANG L F, et al. Collaborative task allocation and optimization solution for unmanned aerial vehicles in search and rescue[J]. *Drones*, 2024, 8(4): 138.
- [18] SHEN X N, XU D, SONG L Y, et al. Heterogeneous multi-project multi-task allocation in mobile crowdsensing using an ensemble fireworks algorithm[J]. *Applied Soft Computing*, 2023, 145: 110571.
- [19] WANG L, YANG D Q, YU Z W, et al. Task scheduling in three-dimensional spatial crowdsourcing: a social welfare perspective[J]. *IEEE Transactions on Mobile Computing*, 2023, 22(9): 5555-5567.
- [20] WANG H, LIU C H, YANG H M, et al. Ensuring threshold AoI for UAV-assisted mobile crowdsensing by multi-agent deep reinforcement learning with transformer[J]. *IEEE/ACM Transactions on Networking*, 2024, 32(1): 566-581.
- [21] ZHAO Y N, LIU C H, YI T J, et al. Energy-efficient ground-air-space vehicular crowdsensing by hierarchical multi-agent deep reinforcement learning with diffusion models[J]. *IEEE Journal on Selected Areas in Communications*, 2024, 42(12): 3566-3580.
- [22] ZHAO Y, ZHU Z Q, GAO C, et al. Heterogeneous graph reinforcement learning for dependency-aware multi-task allocation in spatial crowdsourcing[EB]. arXiv preprint, 2024, arXiv: 2410.15449.
- [23] HAN L, TU C Y, YU Z W, et al. Collaborative route planning of UAVs, workers, and cars for crowdsensing in disaster response[J]. *IEEE/ACM Transactions on Networking*, 2024, 32(4): 3606-3621.
- [24] YE Y X, WANG H, LIU C H, et al. QoI-aware mobile crowdsensing for metaverse by multi-agent deep reinforcement learning[J]. *IEEE Journal on Selected Areas in Communications*, 2024, 42(3): 783-798.
- [25] XU C H, SONG W. Decentralized task assignment for mobile crowdsensing with multi-agent deep reinforcement learning[J]. *IEEE Internet of Things Journal*, 2023, 10(18): 16564-16578.
- [26] PETI M, PETRIC F, BOGDAN S. Decentralized coordination of multi-agent systems based on POMDPs and consensus for active perception[J]. *IEEE Access*, 2023, 11: 52480-52491.
- [27] 夏家伟, 朱旭芳, 张建强, 等. 基于多智能体强化学习的无人艇协同围捕方法[J]. *控制与决策*, 2023, 38(5): 1438-1447.
- [28] XIA J W, ZHU X F, ZHANG J Q, et al. Research on cooperative

hunting method of unmanned surface vehicle based on multi-agent reinforcement learning[J]. Control and Decision, 2023, 38(5): 1438-1447.

- [28] 王万良, 陈浩立, 李国庆, 等. 基于深度强化学习的多配送中心车辆路径规划[J]. 控制与决策, 2022, 37(8): 2101-2109.
WANG W L, CHEN H L, LI G Q, et al. Deep reinforcement learning for multi-depot vehicle routing problem[J]. Control and Decision, 2022, 37(8): 2101-2109.
- [29] DUTTA D, UPRETI S R. A survey and comparative evaluation of actor-critic methods in process control[J]. The Canadian Journal of Chemical Engineering, 2022, 100(9): 2028-2056.
- [30] YUAN F N, ZHANG Z X, FANG Z J. An effective CNN and Transformer complementary network for medical image segmentation[J]. Pattern Recognition, 2023, 136: 109228.
- [31] SHIRI F M, PERUMAL T, MUSTAPHA N, et al. A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU[EB]. arXiv preprint, 2023, arXiv: 2305.17473.
- [32] DANN C, MANSOUR Y, MOHRI M, et al. Guarantees for epsilon-greedy reinforcement learning with function approximation[EB]. arXiv preprint, 2022, arXiv: 2206.09421.
- [33] AMATO C. An introduction to centralized training for decentralized execution in cooperative multi-agent reinforcement learning[EB]. arXiv preprint, 2024, arXiv: 2409.03052.
- [34] HUANG L Q, YE M, XUE X S, et al. Intelligent routing method based on dueling DQN reinforcement learning and network traffic state prediction in SDN[J]. Wireless Networks, 2024, 30(5): 4507-4525.
- [35] RASHID T, SAMVELYAN M, DE WITT C S, et al. Monotonic value function factorisation for deep multi-agent reinforcement learning[J]. Journal of Machine Learning Research, 2020, 21(1): 7234-7284.
- [36] SON K, KIM D, KANG W J, et al. QTRAN: learning to factorize with transformation for cooperative multi-agent reinforcement learning[EB]. arXiv preprint, 2019, arXiv: 1905.05408.
- [37] HAN L, YU Z W, YU Z Y, et al. Online organizing large-scale heterogeneous tasks and multi-skilled participants in mobile crowd-sensing[J]. IEEE Transactions on Mobile Computing, 2023, 22(5): 2892-2909.
- [38] LIU W Q, ZANG Y, XIONG Z Y, et al. 3D building model generation from MLS point cloud and 3D mesh using multi-source data fusion[J]. International Journal of Applied Earth Observation and Geoinformation, 2023, 116: 103171.
- [39] GUO D Y, YANG D J, ZHANG H, et al. DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning[EB]. arXiv preprint, 2025, arXiv: 2501.12948.
- [40] SELVARAJAN S, SRIVASTAVA G, KHADIDOS A O, et al. An artificial intelligence lightweight blockchain security model for security and privacy in IIoT systems[J]. Journal of Cloud Computing, 2023, 12(1): 38.
- [41] DANKAN GOWDA V, KAWALE S R, PRASAD K, et al. Technologies for comprehensive information security in the IoT[C]//

Proceedings of the 2023 International Conference for Advancement in Technology (ICONAT). Piscataway: IEEE Press, 2023: 1-5.

[作者简介]



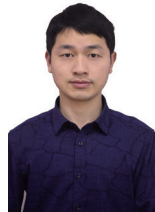
路文浩(2002–), 男, 国防科技大学系统工程学院博士生, 主要研究方向为移动群智感知、深度强化学习、具身智能。



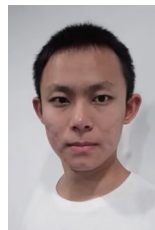
赵勇(1997–), 男, 国防科技大学系统工程学院博士生, 主要研究方向为群智感知、人机交互、具身智能。



季雅泰(1998–), 男, 国防科技大学系统工程学院博士生, 主要研究方向为具身智能、气体源搜索、无人机目标搜索。



张琪(1988–), 男, 博士, 国防科技大学系统工程学院助理研究员, 主要研究方向为仿真智能兵力行为建模、行为树。



许凯(1990–), 男, 博士, 国防科技大学系统工程学院讲师, 主要研究方向为复杂系统建模与仿真、人工智能和认知建模。



朱正秋(1994–), 男, 博士, 国防科技大学系统工程学院讲师, 主要研究方向为群智计算、具身智能、大模型智能体。