

## 面向无人机集群协同感知的多智能体资源分配策略

王志宏<sup>1,2</sup>, 冷甦鹏<sup>1,2</sup>, 熊凯<sup>1,2</sup>

(1. 电子科技大学信息与通信工程学院, 四川 成都 611731; 2. 电子科技大学(深圳)高等研究院, 广东 深圳 518110)

**摘要:** 在智能物联网技术发展的推动下, 无人机集群已广泛用于应急、救援等场景的感知监测。无人机在任务区域自动感知发现任务目标, 邻近无人机组成协作感知与协作计算任务群组, 协同完成数据的感知、采集和处理。然而, 重复的感知资源以及多任务间的计算资源供需不平衡, 会造成额外的计算与通信开销, 增大端到端处理时延。为了应对这一挑战, 提出了一种结合仿生学和多智能体独立强化学习的多任务资源分配策略, 基于局部的任务信息进行资源协同分配决策。该方法用任务情景信息浓度表示各个任务的资源需求, 并通过情景信息在各任务群组间的扩散, 动态更新各任务异构资源需求。同时, 结合独立多智能体强化学习方法进行智能决策, 以对各任务异构资源进行智能协同分配。仿真结果表明, 所提方案不仅能够有效缩短任务执行时间, 还可显著提高计算资源利用率。

**关键词:** 无人机集群; 资源分配; 独立强化学习; 仿生学; 多智能体

**中图分类号:** TP393

**文献标志码:** A

**doi:** 10.11959/j.issn.2096-3750.2023.00326

## Multi-agent resource allocation strategy for UAV swarm-based cooperative sensing

WANG Zhihong<sup>1,2</sup>, LENG Supeng<sup>1,2</sup>, XIONG Kai<sup>1,2</sup>

1. School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

2. Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518110, China

**Abstract:** Driven by the development of intelligent internet of things (IoT) technology, unmanned aerial vehicle (UAV) swarms have been widely used for sensing and monitoring in emergency and rescue scenarios. The UAVs automatically sense and discover mission targets in the mission area, recruiting neighboring UAVs to form perception and computation task groups to collaboratively complete the perception, acquisition and processing of data. However, repetitive sensory data and imbalance in the supply and demand of computational resources between multiple tasks cause additional computational and communication overheads and increase the end-to-end processing latency. To address this challenge, a multi-task resource allocation approach combining bionics and multi-agent independent reinforcement learning was proposed, making collaborative resource allocation decisions based on local task information. The method represents the resource requirements of individual tasks as situational information concentrations and dynamically updates the heterogeneous resource requirements of each task by spreading the situational information across task groups. At the same time, it combines multi-agent independent reinforcement learning methods for intelligent decision making in order to collaboratively allocate the heterogeneous resources of each task. Simulation results show that this solution can not only effectively reduce the task execution time, but also significantly improve the computational resource utilization.

**Key words:** UAV swarm, resource allocation, independent reinforcement learning, bionics, multi-agent

收稿日期: 2022-06-22; 修回日期: 2022-12-29

通信作者: 冷甦鹏, spleng@uestc.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62171104)

**Foundation Item:** The National Natural Science Foundation of China (No.62171104)

## 0 引言

随着智能物联网技术的发展,无人机作为一种新兴技术已被广泛应用于多种应急场景,如森林火灾监测、紧急救援、应急通信等<sup>[1-4]</sup>。无人机具有灵活、小型化的特点,可携带多种小型传感设备,在危险区域执行环境数据搜集任务,并通过本地或者边缘设备进行计算处理,为决策方提供决策依据<sup>[5-6]</sup>。然而,单个无人机的感知范围和感知精度有限,无法在有效时间内对感知目标形成多角度、高精度的数据搜集,使得单个无人机搜集到的数据存在较大误差。同时,单个无人机的计算能力有限,无法在任务有效时间内快速处理大量的感知数据。因此,考虑采用多无人机协作感知与处理的方案,由多架无人机对同一目标进行协作感知与处理。此外,由于各任务区域在空间上离散分布,各任务区域都需要多无人机协同组成任务群组,在对应区域执行协作感知与处理任务。

在复杂的多任务(multi-task)环境中,多无人机协同感知与协同计算又面临新的挑战。首先,各任务之间低效的异构资源调度会阻碍任务的高效执行。任务对感知资源的需求不仅受感知目标规模和感知精度的限制,同时受计算资源的限制,过度的感知数据搜集虽然可以提高感知精度,降低感知数据搜集时间,但是会使得感知数据的处理时延增大,同时带来较高的数据传输代价;此外,通信资源需求既要考虑感知数据规模,又要考虑任务群组内计算资源的分布,即计算资源需要根据各任务的感知数据规模和计算资源需求进行动态调整。在基于拍卖和共识的多任务资源分配研究<sup>[7-8]</sup>和基于博弈的多任务资源分配研究中<sup>[9-10]</sup>,研究者仅研究了各资源需求间相互独立的多任务分配场景,但是在面对各资源相互耦合时,这些方法未考虑各资源间的联系,往往会带来各任务间无效的异构资源调度过程。其次,在复杂动态的任务环境中,任务的资源需求随时空动态变化,任务初始分配到的异构资源可能无法在未来有效支撑任务执行。因此,需要根据任务执行情况动态调整任务资源需求,在任务执行过程中动态地调整任务资源,以保证各任务有效完成。基于分布式博弈的任务分配研究<sup>[9-10]</sup>在理论上可以收敛到达成分布式共识的任务分配策略,但是只有在各任务资源需求静态时才可保证该方法的收敛性。同时,基于二次规划的任务分配方法<sup>[11]</sup>

可根据各任务动态变化的资源需求进行任务分配优化方案求解,但是该方法由于计算复杂度高,无法满足任务分配的实时性要求。最后,高动态无人机集群在各无人机上维护全局的任务信息时,会引入较高的通信交互代价,而局部的任务信息无法保证有效的异构资源协同过程。例如,在基于拍卖和共识进行多任务分配的工作中<sup>[7-8]</sup>,无法基于局部任务信息达成共识,使得任务分配算法无法收敛。

为了应对上述挑战,本文提出了一种结合仿生学和多智能体(multi-agent)独立强化学习的分布式多任务异构资源协同分配方法,以根据各任务执行状态,进行异构资源的高效协同调度。该方法将各个任务的资源需求以任务情景信息浓度表示,并通过任务情景信息的局部扩散,更新各无人机候选任务的选择策略,从而结合独立强化学习方法,基于局部的任务信息进行资源分配决策。本文的主要贡献如下。

1) 基于仿生学设计了无人机集群中任务异构资源需求动态更新方法。借鉴火蚁群通过控制臭迹信息物质释放浓度调整离群觅食的工蚁数量的现象,本文利用任务情景信息浓度表示各个任务的异构资源需求,并以任务情景信息的扩散、衰减和叠加特性动态更新任务情景信息浓度,从而动态描述任务感知、计算和通信资源需求。

2) 基于任务情景信息的扩散更新,设计不同任务间的异构资源动态调整过程。各任务群组依据任务优先级和任务执行情况,动态调整任务情景信息的释放浓度,以调整任务情景信息在邻近的任务群组中的浓度,协助异构资源动态调整过程的高效进行。该方法以任务群间信息交互的形式,避免了全局任务信息维护带来的通信开销,保证了任务资源动态调整过程的高效进行。

3) 结合独立强化学习与仿生学,设计多智能体独立强化学习方法,进行多任务驱动下的异构资源协同分配。各无人机作为智能体基于感知到的任务情景信息浓度进行候选任务的选择,并基于独立强化学习方法在候选任务和所在任务间进行资源分配决策,从而实现在面对动态任务需求与局部任务信息时的高效资源协同分配过程。

## 1 系统模型

本文研究的任务驱动下的无人机集群异构资源协同场景如图1所示。 $L$ 架无人机部署在指定的

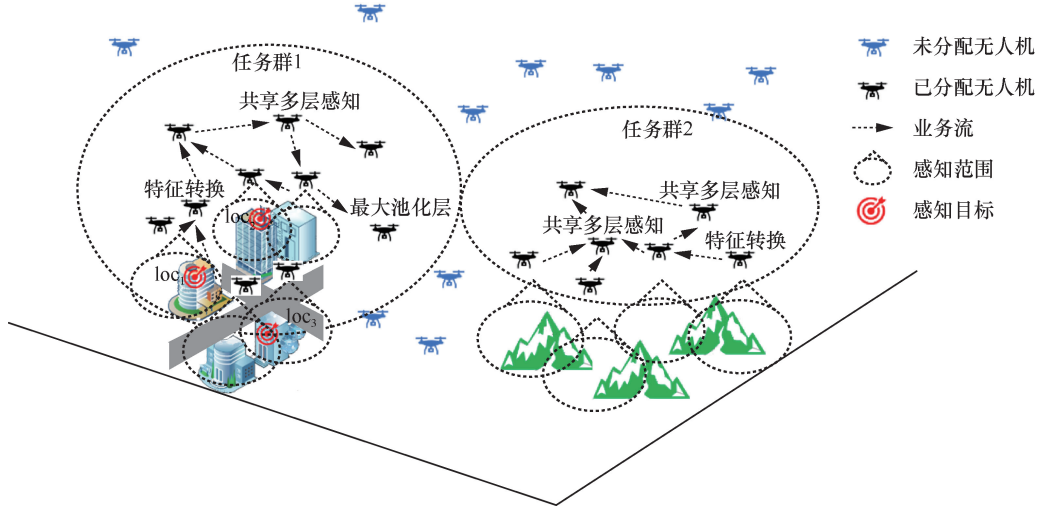


图 1 任务驱动下的无人机集群异构资源协同场景

区域进行三维点云数据搜集与处理任务 ( $\mathcal{L} = \{1, 2, \dots, L\} \in N^+$ ), 各无人机都具有感知能力、计算能力与存储能力。整个任务区域由边长为  $d$  的单元格构成,  $k_i, k_j (k_i, k_j \in K = \{1, 2, \dots, K\})$  表示任务区域的行和列的序号, 第  $k_i$  行、第  $k_j$  列的中心坐标表示为  $(x_{k_i k_j}, y_{k_i k_j}, 0)$ , 且  $p_{k_i k_j}$  表示该区域任务发生概率。各无人机在任务区域内巡航并检测潜在任务。同时, 发现任务的无人机需要根据任务异构资源需求迅速召集邻近无人机形成协同感知与计算任务群组, 并由各任务群组分工完成三维点云数据的搜集与处理。

### 1.1 任务模型

以  $\mathcal{T} = \{ta_1, ta_2, \dots, ta_H\}$  表示在任务区域中发现的  $H$  个任务构成的任务集合, 且定义  $ta_j = \langle id_j, o_j, n_j^s, n_j^f, n_j^p, t_j^r, t_j^e \rangle$  表示任务  $j$ 。其中,  $id_j$  为任务  $j$  的 ID 标识;  $o_j$  为任务  $j$  的优先级;  $n_j^s$  为任务  $j$  感知目标数量;  $n_j^p$  为无人机获得的每帧点云数据的大小;  $t_j^r$  为该任务的任务执行时间上限;  $t_j^e$  为任务  $j$  的初始任务群组召集时间;  $n_j^f$  为任务  $j$  每个感知目标需要获取的帧数, 满足  $(1 - P_{i, k_i, k_j})^{n_j^f} \leq \hat{p}$ , 其中  $P_{i, k_i, k_j}$  为无人机  $i$  正确感知区域  $k_i k_j$  的概率, 且可表示为  $P_{i, k_i, k_j} = e^{-\lambda d_{i, k_i, k_j}(t)}$  [12-15],  $\lambda$  为感知准确率的参数;  $d_{i, k_i, k_j}$  为无人机  $i$  和区域  $k_i, k_j$  之间的欧氏距离,  $\hat{p}$  为感知错误率上界。任务  $j$  的无人机集群组表示为  $U_j$ 。在任务  $j$  执行时, 将  $U_j$  中执行感知任务的无人机均匀分配给感知目标集合

$S_j = \{s_{1,j}, s_{2,j}, \dots, s_{n_j^s,j}\}$  进行点云数据搜集, 且各无人机在完成一个目标位置的数据搜集之后开始下一个目标位置的数据搜集, 各任务位置的对应的数据点集可表示为  $D_j = \{D_{1,j}, D_{2,j}, \dots, D_{n_j^s,j}\}$ , 其中  $D_{i,j}$  表示任务  $j$  的第  $i$  个感知目标对应的点云数据点集合。且从任务开始执行, 到完成感知目标  $s_{i,j}$  的点云数据搜集所需时间为

$$t_{i,j}^s = \begin{cases} \left\lceil \frac{i}{n_j^u} \right\rceil \times (n_j^f T_s + T_f), & n_j^s \geq n_j^u \\ \left\lceil \left( n_j^f / \left\lceil \frac{i}{n_j^s} \right\rceil \right) T_s \right\rceil, & \text{其他} \end{cases} \quad (1)$$

其中,  $T_s$  表示无人机搜集一帧点云数据所需时间,  $T_f$  为无人机从一个感知目标  $s_{i,j}$  飞往下一个感知目标  $s_{i,j+1}$  所需飞行时间,  $n_j^u$  为任务  $j$  中感知无人机的数量。

### 1.2 计算模型

本节基于 PointNet<sup>[16-17]</sup> 点云处理模型进行分布式点云处理复杂度分析。为了实现高效的点云处理过程, 点云数据集可由多无人机以分布式并行的方式进行计算处理。且 PointNet 模型可分为两个特征转换 (FT, feature transform) 模块和两个共享多层感知机 (Shared MLP, shared multilayer perceptron) 模块。同时, 特征转换模块和共享多层感知机都为可并行计算模块, 可以由不同的无人机携带相同的模块分布式且并行地对同一点云数据进行处理。

特征转换模块和共享多层感知机模块都是基于一维卷积进行的, 一维卷积处理单个点的计算时间复杂度可表示为  $O(c_{i-1} s_i c_i m_i)$ 。其中,  $c_i$  为第  $i$  层

一维卷积的输出通道数量,  $s_i$  为第  $i$  层的一维卷积核数量,  $m_i$  为每个通道输出的特征维度。特征转换模块由  $\alpha^c$  个一维卷积层和  $\alpha^l$  个线性层组成, 且线性层处理单个点的时间复杂度为  $O(d_i d_{i+1})$ , 其中  $d_i$  为第  $i$  个线性层的输入特征维度。则三维点云中的单个点执行一次特征变换所需的 CPU 周期数为

$$C^F = ku \left( \sum_{i=0}^{\alpha^c} c_{i-1} s_i c_i m_i + \sum_{i=0}^{\alpha^l} d_i d_{i+1} \right) \quad (2)$$

其中,  $u$  为执行一次运算所需 CPU 计算周期数,  $k$  为常数。则无人机  $i$  进行一次特征变换所需计算时间可表示为  $\tau_i^F = \frac{C^F}{f_i}$ 。

共享多层感知机由  $\alpha^s$  层一维卷积构成, 其对单个点进行处理所需 CPU 周期数可表示为

$$C^{SM} = ku \left( \sum_{i=0}^{\alpha^s} c_{i-1} s_i c_i m_i \right) \quad (3)$$

则无人机  $i$  基于共享多层感知机进行点云处理所需时间为  $\tau_i^{SM} = \frac{C^{SM}}{f_i}$ 。基于 PointNet 的点云分类模型

中包含两个特征转换处理模块和两个共享多层感知机模块, 即  $\mathcal{M} = \{FT_1, SM_1, FT_2, SM_2\}$ 。本文用  $\Gamma = \{\gamma_i | i=1, 2, 3, 4\}$  简化表示对应的 4 个计算模块,

从而点  $i$  所需的处理时间为  $\tau_i = \sum_{j=1}^4 \tau_{\delta_i^{\gamma_j}}^{\gamma_j}$ , 其中

$\delta_i^{\gamma_j} \in \mathcal{L}$  表示对点  $i$  执行处理模块  $\gamma_j$  的无人机。

### 1.3 传输模型

无人机之间的信道可看作视距无线传输, 则其路径损耗可以用自由空间损耗 (FSPL, free space propagation loss) 模型表示<sup>[18]</sup>, 从而无人机  $i$  到无人机  $j$  在时隙  $t$  的无线信道增益可表示为  $g_{i,j}(t) = \lambda_0 d_{i,j}^{-\alpha}(t)$ , 其中  $d_{i,j}(t)$  为在时隙  $t$  时无人机  $i$  和无人机  $j$  之间的距离,  $\lambda_0$  为参考距离为单位距离时视距无线信道的增益, 且  $\alpha$  为路径损耗系数<sup>[19]</sup>。各无人机采用正交频分多路接入作为传输方式进行数据传输, 则无人机  $i$  到无人机  $j$  之间的最大传输速率可表示为

$$c_{i,j}(t) = \begin{cases} b_i \log \left( 1 + \frac{p_i(t) g_{i,j}(t)}{\sigma_{i,j}} \right), & d_{i,j}(t) \leq R \\ 0, & d_{i,j}(t) > R \end{cases} \quad (4)$$

其中,  $b_i$  为无人机  $i$  获得的传输带宽,  $\sigma_{i,j}$  为无人机  $i$  和无人机  $j$  之间的系统噪声,  $R$  为传输半径,  $p_i$  为传输功率。

基于 Dijkstra 算法, 本文将无人机  $i$  到无人机  $j$  的  $k$  跳路径表示为  $A = \{i, \dots, h, \dots, j\}$ , 用  $B = \{0, 1, 2, \dots, |A|-1\}$  表示该路径中的每一个节点。假设点云数据中每个点的特征中包含  $b$  比特的信息, 且无人机  $i$  执行的处理模块为  $\gamma_{h_i}$ , 其中,  $h_i = \{1, 2, 3, 4\}$ , 表示无人机  $i$  执行的处理模块种类, 则单个点传输时间可表示为

$$t_{i,j}^{\gamma_{h_i}} = \begin{cases} \sum_{k=0}^{|A|-2} \frac{b f^{\gamma_{h_i}}}{c_{k,k+1}}, & B \neq \{0\} \\ 0, & B = \{0\} \end{cases} \quad (5)$$

其中,  $f^{\gamma_{h_i}}$  表示模块  $\gamma_{h_i}$  输出的点云的特征维度, 则

点  $i$  的传输时延可表示为  $t_i^c = \sum_{k=2}^4 t_{\delta_i^{\gamma_{h_k}}, \delta_i^{\gamma_{h_{k-1}}}}^{\gamma_{h_k}}$ 。

### 1.4 优化模型

由于无人机集群本身的高动态性和任务的突发性, 保障各任务在有效时间内完成变得较为困难。因此, 本文以最小化任务平均执行代价为目标, 在任务执行过程中动态调整任务异构资源, 设计了以下优化模型。

$$\begin{aligned} \min_{\mathcal{X}(t), \mathcal{B}(t)} & \sum_{j \in \{1, \dots, H\}} \frac{o_j T_{\text{ta}_j}}{t_j^r} \\ \text{s.t.} \quad \text{C1:} & \sum_{j=0}^H x_{i,j} \leq 1, i \in \mathcal{L} \\ \text{C2:} & T_{\text{ta}_i} \leq t_i^r, i \in \{1, \dots, H\} \\ \text{C3:} & \prod_i^m \left( 1 - P_{i,k_i,k_j} \right)^{n_j^f} \leq \hat{p}, i \in \mathcal{L}, k_i, k_j \in \mathcal{K} \quad (6) \\ \text{C4:} & x_{ij}(t) \in \{0, 1\}, i \in \mathcal{L}, j \in \{1, \dots, H\} \\ \text{C5:} & \sum_{i=0}^H b_i = B \\ \text{C6:} & b_i \geq 0, i \in \mathcal{L} \end{aligned}$$

其中,  $\mathcal{X}(t) = (X_1(t), \dots, X_L(t))$ , 且  $X_i(t) = (x_{i,1}, x_{i,2}, \dots, x_{i,H}(t))$ , 本文用  $x_{i,j} = 1$  表示无人机  $i$  在时隙  $t$  时参与任务  $j$ 。 $\mathcal{B} = (b_1(t), b_2(t), \dots, b_H(t))$ ,  $b_i(t)$  表示任务  $i$  在时隙  $t$  时的通信带宽。 $T_{\text{ta}_j} = \max_{i \in \{1, \dots, n_j^s\}} \left[ t_{i,j}^s + \max_{k \in D_{i,j}} (t_k^c + \tau_k + t_k^w) \right]$  表示任务  $j$  的执行时间,  $t_k^w = \sum_{i=1}^4 (\varphi_k^{\gamma_i} + \beta_k^{\gamma_i})$ ,



$\varphi_k^{\gamma_i}$  为点云处理过程中的计算处理排队时延,  $\beta_k^{\gamma_i}$  为通信队列的排队等待时延。

## 2 基于仿生的多智能体异构资源协同分配

为了实现面向多任务的分布式异构资源协同分配, 本文基于仿生学设计多智能体独立强化学习算法。其中, 本文以任务情景信息进行任务资源需求表征, 同时任务情景信息在任务群中扩散、衰减和叠加, 更新任务情景信息浓度。从而, 各智能体基于任务情景信息浓度形成候选任务选择策略, 进行候选任务的选择。各智能体基于独立强化学习算法, 在候选任务和当前所在任务之间进行局部的异构资源分配, 最大化任务执行效率。

### 2.1 任务情景信息

为了提高任务群组招募效率, 本文引入任务需求信息进行任务需求信息的扩散与更新。不同于自然界中蚂蚁留下的化学信息, 任务需求信息是由智能体产生的虚拟信息, 可由多种属性信息进行表征, 包括任务群组 ID、任务位置和任务群组所需感知无人机的数量等信息<sup>[20-24]</sup>。为此, 本文引入任务情景信息作为任务资源需求的表征。

任务情景信息在任务执行过程中产生, 依据任务执行情况动态更新异构资源需求, 任务  $j$  的情景信息可表示为  $Q_j = (\text{id}_j, \mu_j, n_j^1, \bar{r}_j^s, o_j)$ ,  $n_j^1$  为任务  $j$  在时隙  $k$  时剩余待处理点云数据规模,  $\mu_j = \frac{\bar{t}_j}{t_j^r - k\Delta t}$  为任务当前执行效率,  $\bar{t}_j = \frac{\beta_j(k)}{\sigma_j(k)/\Delta t}$  为当前任务预估剩余时间, 且  $\beta_j(k)$  表示任务  $j$  当前时隙剩余的总点云规模,  $\sigma_j(k)$  为任务群组在时隙  $k\Delta t$  内处理的点云规模,  $\bar{r}_j^s$  为任务  $j$  感知资源需求。

### 2.2 基于任务情景信息浓度的候选任务选择

本文结合任务情景信息的浓度, 进行候选任务的选择, 从而各智能体可在候选任务和当前所在任务间进行异构资源的调整。无人机网络作为任务需求信息的传播介质将任务情景信息存储在其信息缓存中, 且基于局部的通信过程对各任务群组的情景信息进行局部扩散、浓度的线性叠加和衰减, 以动态更新各任务群组的异构资源需求。无人机  $i$  根据信息缓存中各任务信息需求浓度, 形成的任务选择策略为

$$P_{ij}(t) = \frac{\eta_{ij}(t)}{\sum_{j \in Q_i(t)} \eta_{ij}(t)} \quad (7)$$

其中,  $P_{ij}(t)$  为无人机  $i$  选择任务  $j$  的概率,  $Q_i(t)$  为无人机  $i$  在时间  $t$  可感知到的任务集合,  $\eta_{ij}(t)$  为任务  $j$  在无人机  $i$  上的任务情景信息浓度。同时, 基于任务选择策略随机进行候选任务的选择, 可使得任务情景信息浓度较低的任务被选中, 避免无人机同时选中同一任务情景信息浓度最高的任务。

任务群组  $j$  中的无人机可根据任务执行情况产生动态情景信息  $Q_j^d$ , 并基于无人机网络向  $k$  跳内的邻居进行扩散, 以更新任务情景信息在  $k$  跳邻接的无人机中的浓度, 从而更新其候选任务的选择策略。受蚁群觅食过程的启发, 本文将同一任务群组的任务情景信息在同一节点的浓度累积看作线性叠加过程, 以加强该任务群组对邻近节点的影响。同时, 为了保证任务群组召集的收敛性, 任务情景信息浓度在传播过程中会随着时间的任务执行情况进行衰减, 因此在无人机  $i$  收到任务  $j$  的任务情景信息后, 其任务情景信息浓度更新过程可表示为

$$\eta_{ij}(k+1) = \begin{cases} \eta_{ij}(k) + \mu_j o_j, & j \in Q_i(k+1) \\ \eta_{ij}(k)a, & \text{其他} \end{cases} \quad (8)$$

其中,  $a \in (0,1)$  表示任务需求信息浓度衰减因子, 以保证失效的情景任务信息随着时间推移而消失。

### 2.3 基于独立强化学习的局部异构资源分配

为了在候选任务和无人机当前所在任务间进行局部的资源调整, 各智能体基于独立强化学习算法<sup>[25-28]</sup>进行任务分配决策。从而, 各智能体仅根据和自身有关的局部环境状态和局部奖励值, 进行独立的学习过程, 执行局部的资源调整决策。在进行局部的决策时, 各智能体的动作空间有限, 仅包含候选任务和当前所在任务的选择, 因此本文采用深度  $Q$  网络 (DQN, deep  $Q$ -network) 作为每个独立智能体算法进行资源协同决策<sup>[29]</sup>, 具体描述如下。

1) 状态: 无人机  $i$  的状态空间可表示为  $S_i = \{(\rho_i(k), \varphi_i(k), s_i^u, f_i^u) | k=0,1,\dots,j,\dots\}$ , 其中  $\rho_i(k) = (n_{\sigma^i}^1, \mu_{\sigma^i}(k), \bar{r}_{\sigma^i}^s(k))$  为无人机  $i$  在时隙  $k$  所在任务群组  $\sigma^i$  的任务状态;  $\varphi_i(k) = (n_{\xi^i}^1, \mu_{\xi^i}(k), \bar{r}_{\xi^i}^s(k))$  表示无人机  $i$  在时隙  $k$  基于式(7)所选候选任务  $\xi^i$  的任务状态;  $s_i^u \in \{0,1\}$  表示无人机  $i$  的感知能力, 若  $s_i^u = 1$  则表示无人机  $i$  可

执行感知任务； $f_i^n$  表示无人机  $i$  的 CPU 频率。

2) 动作：无人机  $i$  的动作空间可表示为  $A_i = \{x_{ij} | x_{ij} \in \{0,1\}\}$ ，若  $x_{ij} = 1$  则表示无人机  $i$  选择加入候选任务  $j$ ，反之则无人机  $i$  保持空闲状态或继续在原有任务群组执行任务。

3) 奖励：为了最大化无人机  $i$  可感知到的任务执行效率，同时保证智能体学习过程的平稳性，奖励函数可定义为

$$r_i = \sum_{j \in \{\sigma^i, \xi^i\}} \omega_j r_{ij} \quad (9)$$

其中

$$r_{ij} = \begin{cases} -\frac{\sum_{h=k}^{k+\lceil T_f/\Delta t \rceil+1} \mu_j(h)}{\left\lfloor \frac{T_f}{\Delta t} + 1 \right\rfloor}, & x_{ij} = 1 \\ -\mu_j(k), & \text{其他} \end{cases} \quad (10)$$

$$\omega_i(k) = \sum_{j \in \{\sigma^i, \xi^i\}} \frac{o_j}{\sum_{z \in \{\sigma^i, \xi^i\}} o_z} \quad (11)$$

即无人机的奖励值来自当前可感知到的各任务的执行代价的加权平均值，使得无人机通过智能决策来最小化可感知到的候选任务和当前所在任务的执行效率。

4) 参数自主学习：DQN 深度强化学习方法是一种基于不断试错并基于经验进行学习的过程，并以深度神经网络  $Q_i(s_i, a_i | \theta_i)$  代替  $Q$  表记录智能体  $i$  的学习经验，通过不断更新参数  $\theta_i$  达到最优任务执行策略<sup>[30]</sup>。为了保证学习过程的稳定性，引入目标网络  $\hat{Q}(s_i, a_i | \hat{\theta}_i)$ ，从而  $Q_i(s_i | \theta_i)$  的更新函数表示为

$$Q_i(s_i, a_i | \theta_i) = r_i + \gamma \max_{a'} \hat{Q}(s_{i+1}, a' | \hat{\theta}_i) \quad (12)$$

其中， $\gamma$  为折扣因子， $\hat{Q}$  的参数  $\hat{\theta}_i$  在  $Q_i$  的参数更新  $k$  次后置为  $\hat{\theta}_i = \theta_i$ 。

同时，在各智能体完成局部的资源调整决策后，任务  $j$  根据无人机分配结果和各任务执行效率，进行带宽资源的调整，任务  $j$  可获得的带宽为

$$b_j = \frac{B\mu_j}{\sum_{i=0}^H \mu_i}。$$

## 2.4 多智能体异构资源协同分配算法

为了实现在任务资源需求动态变化时的异构资源协同分配，各智能体通过多轮的迭代资源分配过程，根据异构资源的供需关系，进行智能异构资

源协同分配。在每轮异构资源协同分配中，各智能体基于任务执行状态进行任务情景信息的扩散更新，更新各任务情景信息浓度，从而更新候选任务的选择策略，并进行候选任务的选择。各智能体基于独立强化学习算法，进行候选任务和当前所在任务间的异构资源分配调整，从而实现多任务间的智能资源协同分配。

综上，本文设计了一种基于仿生学的多智能体独立强化学习 (BIO-MIRL, bionics based multi-agent independent reinforcement learning) 算法进行异构资源的协同分配。该算法根据任务执行状态，生成任务情景信息，动态更新任务异构资源需求，并通过多智能体深度强化学习算法结合任务情景信息浓度进行智能异构资源决策，敏捷高效地进行任务群组资源调整，保证各个任务在有效时间内快速完成。BIO-MIRL 算法见算法 1。

### 算法 1 BIO-MIRL 算法

```

for each episode do:
  for each task  $j \in \{1, \dots, H\}$ 
    for each agent  $i \in U_j$  do:
      产生情景信息  $Q_j = (id_j, \mu_j, n_j^1, \bar{r}_j^s, o_j)$ ;
      if  $\mu_j > 1$ 
        向非当前群组 agent 扩散  $Q_j$ ;
      end if
      基于式(7)选择任务  $\sigma^i$ ;
      agent 基于 DQN 进行资源协同决策，获得  $x_{i\sigma^i}$ ;
      更新任务群组  $U_{\sigma^i}$  和  $U_j$ ;
      根据(9)式更新  $Q$  网络;
    end for
  end for
  更新各任务带宽  $b_j = \frac{B\mu_j}{\sum_{i=0}^H \mu_i}$ ;
end for

```

## 3 仿真结果与分析

### 3.1 仿真环境

本节通过 PyTorch 仿真平台设计仿真实验，评估了 BIO-MIRL 算法在多任务环境下进行分布式异构资源协同的表现。实验参数见表 1。

表 1 实验参数

实验参数	取值
无人机移动范围	$x_i, y_i \in [0, 3000 \text{ m}], i \in \mathcal{L}$
任务区域单位长度	$d = 1000 \text{ m}$
无人机高度范围	$z_i \in [100 \text{ m}, 500 \text{ m}], i \in \mathcal{L}$
总带宽	$B = 10 \text{ MHz}$
无人机 CPU 频率	$f_i \in [1 \text{ GHz}, 2 \text{ GHz}], i \in \mathcal{L}$
点云规模	$N_j \in \{3 \times 10^4, 5 \times 10^4, 8 \times 10^4\}$
任务优先级	$o_j \in \{1, 2, 3, 4, 5\}$
任务最大执行时间	$t_j^r \in [300 \text{ s}, 1200 \text{ s}]$
折扣因子 $\gamma$	$\gamma = 0.9$
扩散跳数 $k$	3
时隙 $\Delta t$	2 s

### 3.2 实验结果分析

为了验证所提 BIO-MIRL 算法的有效性, 本文通过仿真实验对比了所提方案与两个基准方案。1) 随机资源调度策略: 在任务执行过程中, 各无人机随机选择可感知到的任务加入。2) 平均资源调度策略: 每个任务均匀分配通信带宽, 且任务内各无人机均匀分配计算带宽。

BIO-MIRL 算法收敛性所示如图 2, 测试了在算法迭代过程中多个任务的总时间代价  $\text{cost} = \sum_{j \in \{1, \dots, H\}} \frac{o_j T_{\text{ta}_j(t)}}{t_j^r}$ 。随着迭代次数的增加, 任务

总的代价在经过一段时间的缓慢降低后, 趋于稳定。这表明多个智能体在任务情景信息的协调下, 达成了共识性的策略, 且该分布式算法能较快地达到一致性收敛。同时, 由于各智能体的神经网络结构简单, 且各智能体共享经验缓存, 保证了各智能体的策略具有较高的收敛效率。

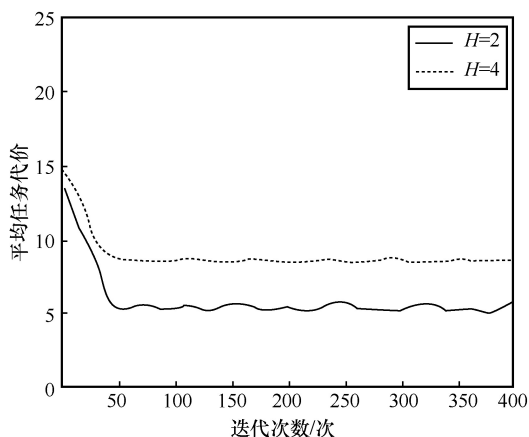


图 2 BIO-MIRL 算法收敛性

不同算法的平均任务执行代价对比如图 3 所示。从与随机资源调度策略以及平均资源调度策略的比较可以看出, BIO-MIRL 算法充分考虑了各任务执行过程中的计算负载和通信负载变化, 动态地进行异构资源调整, 因此较好地降低了总的任务执行代价, 有效地提高了多任务下的异构资源协同效率。所提算法通过敏捷的资源调整, 有效缓解了由多跳中继通信带来的通信时延和计算时延。

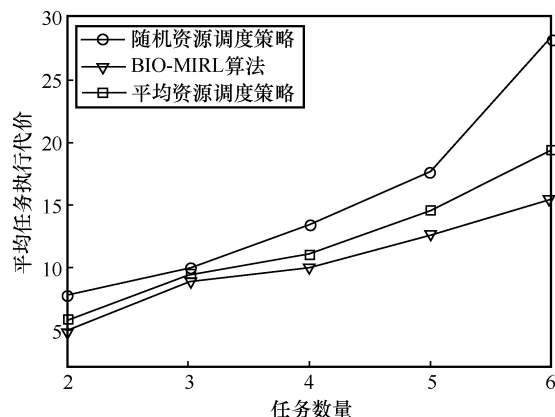


图 3 不同算法的平均任务执行代价对比

不同算法的平均计算资源利用率如图 4 所示, 即在任务执行过程中各节点的计算资源平均占用时间比例。BIO-MIRL 算法的平均计算资源利用率远远超过了随机资源调度策略和平均资源调度策略, 所提算法即使在任务数量较低时仍保持较高的计算资源利用率, 在任务动态产生的环境中体现了其独特的优越性。而平均资源调度策略的方式, 使得具有较多计算资源无人机的计算负载低, 计算资源利用率降低。因此, 动态按需的通信资源分配保证了具有不同处理能力的无人机可分配到对应的通信带宽, 保证计算资源的有效利用。

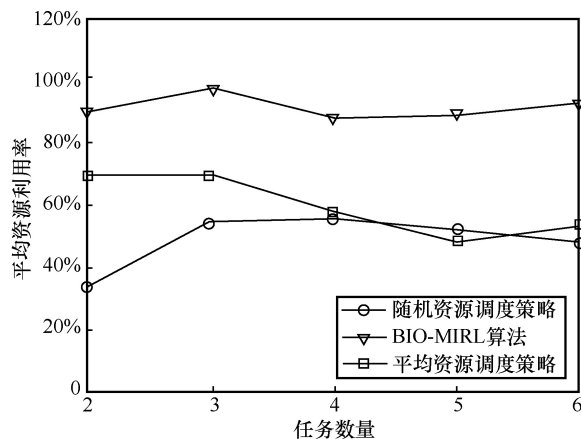


图 4 不同算法的平均计算资源利用率

不同优先级下任务执行时间和计算资源利用率对比如图5所示,对比了 BIO-MIRL 算法中优先级对任务执行时间的影响,其中任务1和任务2除优先级外其余参数相同,且任务2的优先级  $o_2=1$ 。随着任务1的优先级增大,任务1的执行时间逐渐减小,而任务2的执行时间逐渐增大。即 BIO-MIRL 算法会根据任务优先级高低进行异构资源协同,使得任务完成时间随着优先级的增大而降低。同时,随着任务1优先级的增大,无人机平均计算资源利用率降低。这是由于在 BIO-MIRL 算法中,任务1优先级的增大使得任务群组1获得更多的无人机进行计算任务。而在任务1即将完成时,任务群组1的无人机陆续退出并加入任务群组2,这会引入大量的任务切换时间,使得计算资源难以充分利用。

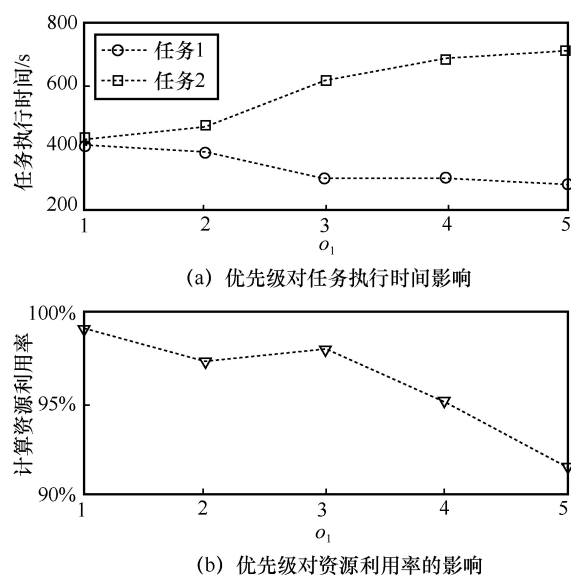


图5 不同优先级下任务执行时间和计算资源利用率对比

## 4 结束语

本文研究了基于无人机集群的点云协同感知与协同处理场景,并针对多任务驱动下异构资源协同策略展开了研究。特别地,针对低处理时延和高感知精度的业务需求,同时兼顾各任务不同的优先级,本文提出结合仿生学的多智能体独立强化学习算法,进行异构无人机网络异构资源的协同分配,快速形成协同感知与计算群组,并根据任务执行状态快速地进行任务资源调整。为了提高资源协同效率,本文引入任务情景信息,并设计任务情景信息的产生、扩散、叠加和衰减过程,动态调整各任务异构资源需求。为了最小化平均任务执行代价,本

文提出了 BIO-MIRL 算法动态协同分配异构资源。通过实验仿真分析,验证了所提算法在多任务执行效率和资源利用率方面优于随机资源调度策略和平均资源调度策略。

## 参考文献:

- [1] ERDELJ M, NATALIZIO E, CHOWDHURY K R, et al. Help from the sky: leveraging UAVs for disaster management[J]. IEEE Pervasive Computing, 2017, 16(1): 24-32.
- [2] GUPTA L, JAIN R, VASZKUN G. Survey of important issues in UAV communication networks[J]. IEEE Communications Surveys & Tutorials, 2016, 18(2): 1123-1152.
- [3] MA H J, LIU Y L, REN Y H, et al. Improved CNN classification method for groups of buildings damaged by earthquake, based on high resolution remote sensing images[J]. Remote Sensing, 2020, 12(2): 260.
- [4] LI T Y, LENG S P, WANG Z H, et al. Intelligent resource allocation schemes for UAV-swarm-based cooperative sensing[J]. IEEE Internet of Things Journal, 2022, 9(21): 21570-21582.
- [5] MEI H B, YANG K, LIU Q, et al. Joint trajectory-resource optimization in UAV-enabled edge-cloud system with virtualized mobile clone[J]. IEEE Internet of Things Journal, 2020, 7(7): 5906-5921.
- [6] SEID A M, BOATENG G O, ANOKYE S, et al. Collaborative computation offloading and resource allocation in multi-UAV-assisted IoT networks: a deep reinforcement learning approach[J]. IEEE Internet of Things Journal, 2021, 8(15): 12203-12218.
- [7] CHOI H L, BRUNET L, HOW J P. Consensus-based decentralized auctions for robust task allocation[J]. IEEE Transactions on Robotics, 2009, 25(4): 912-926.
- [8] FU X W, FENG P, GAO X G. Swarm UAVs task and resource dynamic assignment algorithm based on task sequence mechanism[J]. IEEE Access, 2019(7): 41090-41100.
- [9] BAKOLAS E, LEE Y. Decentralized game theoretic control for dynamic task allocation problems for multi-agent systems[C]//Proceedings of 2021 American Control Conference (ACC). Piscataway: IEEE Press, 2021: 3228-3233.
- [10] ZHANG C Y, LI Q Y, ZHU Y Y, et al. Dynamics of task allocation based on game theory in multi-agent systems[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2019, 66(6): 1068-1072.
- [11] BAKSHI S, FENG T H, YAN Z Y, et al. A regularized quadratic programming approach to real-time scheduling of autonomous mobile robots in a prioritized task space[C]//Proceedings of 2019 American Control Conference (ACC). Piscataway: IEEE Press, 2019: 1361-1366.
- [12] ZHANG S H, ZHANG H L, DI B Y, et al. Cellular cooperative unmanned aerial vehicle networks with sense-and-send protocol[J]. IEEE Internet of Things Journal, 2019, 6(2): 1754-1767.
- [13] HOSSAIN A, CHAKRABARTI S, BISWAS P K. Impact of sensing model on wireless sensor network coverage[J]. IET Wireless Sensor Systems, 2012, 2(3): 272.
- [14] CHAKRABORTY A, ROUT R R, CHAKRABARTI A, et al. On network lifetime expectancy with realistic sensing and traffic generation model in wireless sensor networks[J]. IEEE Sensors Journal, 2013, 13(7): 2771-2779.

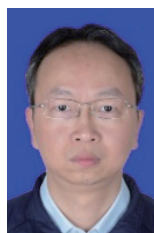


- [15] SHAKHOV V V, KOO I. Experiment design for parameter estimation in probabilistic sensing models[J]. IEEE Sensors Journal, 2017, 17(24): 8431-8437.
- [16] CHARLES R Q, HAOS, MOK C, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 77-85.
- [17] QI C R, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM Press, 2017: 5105-5114.
- [18] ZHAO H T, WANG H J, WU W Y, et al. Deployment algorithms for UAV airborne networks toward on-demand coverage[J]. IEEE Journal on Selected Areas in Communications, 2018, 36(9): 2015-2031.
- [19] WANG B W, SUN Y J, DO-DUY T, et al. Adaptive D-hop connected dominating set in highly dynamic flying ad-hoc networks[J]. IEEE Transactions on Network Science and Engineering, 2021, 8(3): 2651-2664.
- [20] VAW DYKE PARUNAK H, BRUECKNER S A, SAUTER J. Digital pheromones for coordination of unmanned vehicles[C]//Proceedings of the 1st International Conference on Environments for Multi-Agent Systems. New York: ACM Press, 2004: 246-263.
- [21] XU X, LI R P, ZHAO Z F, et al. Stigmergic independent reinforcement learning for multiagent collaboration[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(9): 4285-4299.
- [22] HOLLAND O E. Multiagent systems: lessons from social insects and collective robotics[C]//Adaptation, Coevolution, and Learning in Multiagent Systems: Papers from the 1996 AAAI Spring Symposium. [S.l.:s.n.], 1996: 57-62.
- [23] DORIGO M, BONABEAU E, THERAULAZ G. Ant algorithms and stigmergy[J]. Future Generation Computer Systems, 2000, 16(8): 851-871.
- [24] DI CARO G. AntNet: distributed stigmergetic control for communications networks[J]. Journal of Artificial Intelligence Research, 1998, 9: 317-365.
- [25] ZHANG C, JIN S, XUE W, et al. Independent reinforcement learning for weakly cooperative multiagent traffic control problem[J]. IEEE Transactions on Vehicular Technology, 2021, 70(8): 7426-7436.
- [26] LANCTOT M, ZAMBALDI V, GRUSLYS A, et al. A unified game-theoretic approach to multiagent reinforcement learning[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS). New York: ACM Press, 2017: 4193-4206.
- [27] MATIGNON L, LAUREWT G J, LE FORT-PIAT N. Independent reinforcement learners in cooperative Markov games: a survey regarding coordination problems[J]. The Knowledge Engineering Review, 2012, 27(1):1-31.
- [28] ARULKUMARAN K, DEISEWICZ M P, BRUNDAGI M, et al. Deep reinforcement learning: a brief survey[J]. IEEE Signal Processing Magazine, 2017, 34(6): 26-38.
- [29] ZHANG T K, LEI J Y, LIU Y W, et al. Trajectory optimization for UAV emergency communication with limited user equipment energy: a safe-DQN approach[J]. IEEE Transactions on Green Communications and Networking, 2021, 5(3): 1236-1247.
- [30] FOERSTER J, NARDELLI N, FARQUHAR G, et al. Stabilising experience replay for deep multi-agent reinforcement learning[C]//Proceedings of the 34th International Conference on Machine Learning. New York: ACM Press, 2017: 1146-1155.

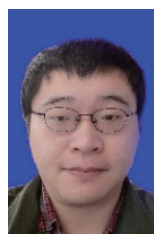
#### [作者简介]



王志宏（1997—），男，电子科技大学博士生，主要研究方向为无人机集群资源分配、语义通信和机器学习。



冷甦鹏（1973—），男，电子科技大学教授、博士生导师，主要研究方向为物联网、车联网、新一代宽带无线网络、无线自组织网络、智能交通信息系统的资源管理、介质访问控制、路由、组网与互联、智能算法理论及技术应用等。



熊凯（1991—），男，电子科技大学在站博士后，主要研究方向为车联网资源分配、移动边缘计算和机器学习。