基于图注意力网络的异构多智能体系统动态任务分配方法

李中杨[1,2],曹筱可[2],蔡奕辰[2,3],孙贵宾[2*],刘克新[2]
(1. 北京航空航天大学沈元学院,北京 100191;2. 北京航空航天大学自动化科学与电气工程学院,北京 100191;
3. 北京航空航天大学人工智能学院,北京 100191)

摘 要: 异构多智能体系统任务分配问题是多智能体领域的核心问题之一。该问题要求将具有不同能力类型的异构智能体合理分配到需多智能体协作完成的任务中,在实际应用场景中存在的任务新增、智能体失效等动态事件,进一步增加了问题的复杂性。针对现有方法计算代价高昂、难以有效建模异构个体与任务间的复杂依赖关系,以及动态场景自适应决策能力差的问题,本文提出了一种基于图注意力网络的异构多智能体系统动态任务分配方法。该方法引入了动态图构建机制建模异构智能体与任务间的复杂交互关系,并通过节点与边的实时更新实现对动态变化场景的表征。同时本文设计了基于图注意力机制的编解码架构,通过为不同边分配独立的注意力通道解耦异构节点的特征语义,并结合指针式解码器实现了能力与需求的匹配及对变长输入的适应。针对大规模任务分配下的稀疏奖励难题,本文提出了涵盖任务规模与环境动态性双维度的多阶段课程学习策略,通过平滑优化曲面引导策略逐步收敛。仿真实验结果表明,所提方法在动态场景下保持100%的成功率,完成时间较基于学习的对比方法降低了4%~8%,较贪婪算法降低约23%,在大规模场景下仍能保持毫秒级决策速度和高质量的分配结果,验证了方法在动态适应性、规模扩展性和分配方案质量方面的综合优势。

关键词: 异构多智能体系统;动态任务分配;图注意力网络;深度强化学习;课程学习
基金项目: 国家自然科学基金(No.62503028,No.62373019);北京市自然科学基金(No.QY25228)
中图分类号: TP18;TP242 **文献标识码:** A **文章编号:** 0372-2112(2026)03-0927-11
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20251257

Dynamic Task Allocation Method for Heterogeneous Multi-Agent Systems Based on Graph Attention Networks

LI Zhongyang[1,2], CAO Xiaoke[2], CAI Yichen[2,3], SUN Guibin[2*], LIU Kexin[2]
(1. *Shenyuan Honors College*, *Beihang University*, *Beijing* 100191, *China*;
2. *School of Automation Science and Electrical Engineering*, *Beihang University*, *Beijing* 100191, *China*;
3. *School of Artificial Intelligence*, *Beihang University*, *Beijing* 100191, *China*)

Abstract: The task allocation problem in heterogeneous multi-agent systems is one of the core issues in the multi-agent domain. This problem requires the rational allocation of heterogeneous agents with distinct capabilities to tasks that demand multi-agent collaboration. Moreover, dynamic events in real-world applications, such as the arrival of new tasks and agent failures, further exacerbate the complexity of this problem. To address the limitations of existing methods—such as high computational costs, difficulties in effectively modeling the complex dependencies between heterogeneous agents and tasks, and poor adaptive decision-making capabilities in dynamic scenarios—this paper proposes a dynamic task allocation method for heterogeneous multi-agent systems based on Graph Attention Networks. This method introduces a dynamic graph construction mechanism to model the complex interaction relationships between heterogeneous agents and tasks, explicitly characterizing dynamically evolving scenarios through real-time updates of nodes and edges. Furthermore, an encoder-decoder architecture based on graph attention mechanisms is designed. By assigning independent attention channels to different interaction edges, it decouples the feature semantics of heterogeneous nodes. Combined with a pointer-based decoder, this architecture achieves precise matching between capabilities and requirements, as well as adaptation to variable-length inputs. In addition, to overcome the sparse reward challenge in large-scale task allocation, this paper proposes a multi-stage curriculum learning strategy covering both task scale and environmental dynamics dimensions, which guides the policy to converge progressively by smoothing the optimization landscape. Simulation results demonstrate that the proposed

收稿日期:2026-01-26;录用日期:2026-02-24;责任编委:温广辉
*通信作者:孙贵宾

method maintains a 100% allocation success rate across various dynamic scenarios. The task completion time is reduced by 4% to 8% compared to learning-based baselines, and by approximately 23% compared to the greedy algorithm. Even in large-scale scenarios, the method maintains millisecond-level decision-making speeds and yields high-quality allocation results, thereby verifying its comprehensive advantages in dynamic adaptability, scalability, and the quality of allocation schemes.

Keywords: heterogeneous multi-agent system; dynamic task allocation; graph attention network; deep reinforcement learning; curriculum learning

Foundation Item(s): National Natural Science Foundation of China (No.62503028, No.62373019); Beijing Natural Science Foundation (No.QY25228)

0 引言

异构多智能体系统凭借其功能互补、鲁棒性强等优势,被广泛应用于应急救援[1]、智能仓储[2]以及协同探测[3]等领域。在这些实际应用场景中,如何将复杂任务高效分配给能力各异的智能体是决定系统性能的关键[4]。与此同时,真实环境往往具有高度的动态性与不确定性(如突发的新任务需求或执行过程中的智能体故障),这要求任务分配方法具备快速响应和自适应调整的能力。

针对异构多智能体任务分配问题,传统的解决方法主要分为精确优化方法、市场拍卖方法和启发式方法[5]。精确优化方法[6]在问题的解空间中进行全局搜索,能获得高质量的任务分配方案,但其计算复杂度随问题规模呈指数级增长,难以满足大规模动态场景的实时性需求。市场拍卖方法[7]通过智能体间的竞标机制将计算负担分散至各个智能体,但由于其复杂的竞标过程以及冲突解决机制,面临收敛时间长且难以保证全局最优性的问题。启发式方法[8]通过人工设计的局部规则或迭代式随机搜索机制逼近最优解,相较于优化方法降低了计算复杂度,但解的质量通常不稳定,且在面对动态变化的异构约束条件时,其适应性与灵活性不足。

与传统方法相比,基于学习的方法通过数据驱动的方式捕捉问题内在结构,能够处理复杂的异构约束,且在训练完成后具备毫秒级的推理决策能力,契合动态场景的实时性要求。然而,将此类方法应用于动态异构任务分配仍面临两大核心挑战。一是如何有效表征异构智能体与任务之间复杂的交互关系。现有的方法多采用卷积神经网络[9]或循环神经网络[10]处理固定维度的状态输入,忽略了智能体与任务间的图结构联系,导致模型难以适应节点数量动态变化的场景,且缺乏对异构能力匹配度的建模。二是如何在大规模组合空间中实现高效的训练。多智能体任务分配的状态与动作空间巨大,且奖励信号稀疏,直接在复杂动态场景下进行训练,极易导致策略网络收敛困难甚至陷入局部极值[11]。

针对上述挑战,本文提出了一种基于图注意力网络的动态任务分配方法,主要贡献在于:

(1)提出了一种面向异构任务分配的图注意力编解码网络架构,该架构针对动态分配过程中智能体与任务间的复杂依赖关系,通过独立的注意力通道实现了能力供给与任务需求的特征解耦与匹配,从而解决异构特征融合及变长输入处理的难题。仿真结果表明,该方法在各规模动态场景下均保持100%的分配成功率,且任务完成时间相较于传统方法降低20%以上。

(2)设计了涵盖任务规模与环境动态性的多阶段课程学习训练策略,通过将训练过程划分为从小规模静态场景到大规模高动态场景的多个阶段,引导网络在早期快速捕捉基础分配逻辑,并在后期逐步适应环境的动态变化。消融实验验证了该策略的有效性。

1 相关工作

本文聚焦于利用深度强化学习解决多智能体任务分配问题。因此,本节将对基于学习的任务分配方法进行综述,并介绍图注意力网络的核心机制及相关领域应用。

1.1 基于学习的任务分配方法

深度强化学习凭借通用性强、决策速度快等优势[12],在多智能体任务分配领域展现出巨大潜力。Nazari 等人[13]利用基于注意力的网络[14]构建解序列,有效解决了变长输入问题。Wang等人[15]将问题建模为顺序决策问题,通过网络迭代添加新边来生成解决方案,有效解决了时空约束问题。然而,上述方法将分配问题简化为单智能体的序列生成过程,缺乏对多智能体间协作关系的建模。针对需要紧密协作的场景,Jose 等人[16]利用模仿学习训练多机器人的子团队动态组建与主动等待策略,但该方法在训练阶段依赖大量专家示范数据。Dai 等人[17]提出了面向异构场景的分散式强化学习框架,通过约束闪前机制缓解了训练死锁问题,但其决策核心仍依赖局部的序列生成范式,在面对任务新增或智能体失效动态事件时能力受限。综上所述,现有方法在异构协作、动态场景的适应以及复杂约束下的训练收敛性等方面仍存

在不足。针对上述问题,本文提出一种基于图注意力网络的任务分配方法,使网络能够感知场景变化,适应异构协作需求。

1.2 图注意力网络及其应用

图注意力网络[18]是处理拓扑空间数据的关键技术。与传统图卷积网络对所有邻居节点采用固定权重聚合不同,图注意力机制能够根据节点特征自适应地计算邻居节点的重要性,从而实现更具选择性的信息聚合。其核心在于捕捉输入序列或图节点间不同位置的依赖关系[19]。对于图中节点 i,其更新后的特征 $\boldsymbol{h}_i'$ 由邻居节点的加权聚合得到。具体而言,首先对各节点特征 $\boldsymbol{h}$ 通过可学习矩阵 $\boldsymbol{W}$ 进行线性变换,然后计算节点 i 与其邻居节点 $j \in N^i$ 间的注意力系数 $\alpha_{ij}=\mathrm{softmax}\left((\boldsymbol{W}\boldsymbol{h}_i)(\boldsymbol{W}\boldsymbol{h}_j)^{\mathrm{T}}/\sqrt{d}\right)$,该系数反映了邻居节点 j 对节点 i 的重要程度。节点 i 的输出特征 $\boldsymbol{h}_i'$ 通过对邻居节点特征的加权聚合,并经非线性激活函数 σ 得到:

$$\boldsymbol{h}_i'=\sigma\left(\sum\alpha_{ij}\boldsymbol{W}\boldsymbol{h}_j\right) \tag{1}$$

凭借其优越的特征提取能力,图注意力网络已被广泛应用于智能交通与推荐系统等领域。例如,Zhang等人[20]提出的门控注意力网络(Gated Attention Networks,GaAN)网络通过门控图注意力机制捕捉路网中复杂的动态时空相关性,显著提升了交通流预测精度;Wang等人[21]设计的异构图注意力网络通过处理包含“用户-物品-商家”的异构图,成功解决了推荐系统中异构数据的语义融合难题。近年来,图注意力网络在多智能体任务分配领域也取得了重要进展。Peng等人[22]提出基于图注意力的去中心化任务分配器图注意力任务分配器(Graph Attention Task AllocatoR,GATAR),通过图注意力算子实现异构无人机与地面车辆的协同目标定位,但其性能在大规模动态场景下降严重。Du等人[23]提出的异构图注意力网络通过层次化建模智能体间的异构关系,在多智能体协作任务中表现优异,但未考虑动态场景下的任务分配问题。此外,Zhang等人[24]通过图归一化技术实现了对变规模输入的泛化,但其方法尚不支持异构智能体。为了解决现有方法在异构建模和动态适应性上的不足,本文针对异构场景设计了四组独立参数的图注意力模块,实现了不同类型节点间交互逻辑的解耦。在此基础上,本文提出了动态图构建机制以支持节点的实时增删,从而适应任务新增与智能体失效等动态事件。

2 异构多智能体系统任务分配模型

2.1 问题假设

在阐述异构任务分配问题之前,首先给出本文的基本假设:(1)假设系统具备高带宽、低延迟的通信网络,智能体间能够实时共享位置与状态信息,适用于智能仓储等具备完善通信设施的场景;(2)假设智能体间的局部避障与防碰撞问题由底层路径规划算法处理,本文专注于上层任务分配策略的优化。现有研究表明[25],若在分配阶段显式建模碰撞约束,问题复杂度将大幅上升,难以兼顾求解效率与可扩展性,而在智能体稀疏或任务空间分布较广的场景下,碰撞概率较低,底层避障算法足以保证执行安全性,对分配效果影响有限。

在上述假设下,考虑将 k_m 个任务 $M=\{m_1,m_2,\cdots,m_{k_m}\}$ 分配给 k_n 个智能体 $N=\{a_1,a_2,\cdots,a_{k_n}\}$。智能体分为 v 种类型,类型集合记为 $S=\{s_1,s_2,\cdots,s_v\}$,每种类型 s_i 包含 n_{s_i} 个智能体。不同类型智能体具备不同的能力,同类型的智能体能力相同。分别用 $N^{s_1},N^{s_2},\cdots,N^{s_v}$ 表示各类型智能体所构成的集合。每个智能体 a_i 的能力由能力向量 $\boldsymbol{c}_{a_i}=[c_1^i,c_2^i,\cdots,c_b^i]$ 表示,其中 b 为能力维度,不同类型的智能体具备不同维度的能力。同类型智能体从同一出发点出发,完成所有任务后需返回出发点。在任务分配过程中,智能体处于以下三种状态之一:(1)行进状态,正在从当前位置前往目标任务;(2)等待状态,已到达任务位置,等待其他联盟成员到场以启动任务;(3)执行状态,联盟已满足任务需求,正在执行任务。

每个任务对智能体的能力与数量有特定的需求,采用能力需求向量 $\boldsymbol{q}_m=[q_1,q_2,\cdots,q_k]$ 表示任务 m 对各种能力智能体的需求数量,其中 q_k 表示为完成任务 m 所需的具备第 k 种能力的智能体的数量。任务 m 的位置为 (x_m,y_m),不失一般性,将任务分布区域归一化为 $[0,1]\times[0,1]$ 的二维空间。任务间不存在时间优先级或执行顺序依赖,可并行执行。当多个智能体被分配到同一任务时,它们组成执行该任务的联盟 L。联盟的能力向量为各成员能力向量的逐元素求和:$\boldsymbol{c}_L=\sum\boldsymbol{c}_a$。当且仅当联盟能力满足任务需求 $\boldsymbol{c}_L\geqslant\boldsymbol{q}_m$ 且联盟中所有智能体均已到达任务位置时,任务 m 能够被启动执行,其中 $\geqslant$ 表示逐元素大于等于。以应急救援场景为例,针对人员搜救任务,若需求向量为 $\boldsymbol{q}_m=[1,1,1,0,0]$,分别对应空中侦察、地面运输和医疗救护三种能力,则最少需要一架侦察无人机、一辆地面运输车和一个医疗机器人组成联盟协同完成,即完成该任务的最少异构智能体数量 $N_{\min}=3$。在任务执行期间,联盟中的所有智能体需停留在任务位置直到任务完成。

本文考虑动态场景,在任务执行过程中可能发生以下两类动态事件:(1)任务新增,在时刻 t,新任务

m_{new}加入任务集合M,空闲智能体需根据更新后的任务集合重新进行分配决策;(2)智能体失效,在时刻t,智能体a_i因故障退出系统,已分配到该智能体目标任务的其他智能体需调整分配策略。

2.2 优化目标

本文将优化目标设置为最小化所有任务的完成时间。上述任务分配问题可描述为以下形式:

$$\min_{\Phi} T=\min_{\Phi}\max_{m\in M} t_m \tag{2}$$

$$\text{s.t. } x_{ij}\in\{0,1\},\ \forall a_i\in N,\ \forall m_j\in M \tag{3}$$

$$\text{sum}(\boldsymbol{c}_{a_i})\geqslant \boldsymbol{q}_{m_j},\ \forall m_j\in M \tag{4}$$

其中:$\Phi=\{\phi_{a_1},\phi_{a_2},\cdots,\phi_{a_{k_a}}\}$为所有智能体的路径集合,$\phi_{a_i}=(p_{a_i},m_{i_1},m_{i_2},\cdots,m_{i_l},p_{a_i})$为一个智能体的任务执行路径,$p_{a_i}$为智能体$a_i$的出发点,$m_{i_k}$为其依次执行的任务;$t_m$为任务$m$的完成时间;$x_{ij}\in\{0,1\}$为决策变量,若智能体$a_i$被分配执行任务$m_j$则$x_{ij}=1$,否则$x_{ij}=0$;$L_j=\{a_i: x_{ij}=1\}$为执行任务$m_j$的联盟。不等式约束确保分配给每个任务的联盟能力满足该任务的需求。

3 基于图注意力网络的任务分配

在本文提出的异构多智能体系统动态任务分配方法中,所有异构智能体共享同一套图注意力网络参数,这种参数共享机制使得网络能够学习到通用的任务分配策略,而非针对特定智能体的局部策略。在决策过程中,每个智能体将当前场景中的任务和其他智能体的状态信息以图的形式输入网络,通过图注意力机制聚合邻域信息后,输出当前要执行的目标任务。当出现任务新增和智能体失效的动态事件时,网络通过更新图结构中的节点和边,对任务分配策略进行实时调整,从而适应环境的动态变化。任务分配算法的伪代码如算法1所示。

3.1 动态图构建

在异构多智能体系统任务分配问题中,任务与智能体之间存在复杂的依赖与交互关系。本文通过构建图$\boldsymbol{G}=(V,E)$,对任务与智能体之间的关系进行建模。其中,$V=V_t\cup V_a$为节点集合,包含任务节点集V_t和智能体节点集V_a。$E=E_{ta}\cup E_{tt}\cup E_{aa}$为边集合,包含任务-智能体边$E_{ta}$、任务-任务边$E_{tt}$和智能体-智能体边$E_{aa}$。

任务节点$v_j\in V_t$特征定义为$\boldsymbol{x}_j^t=[\boldsymbol{s}_j,\boldsymbol{r}_j,d_j,\tau_j,\Delta\boldsymbol{p}_j,f_j]$,各分量含义如下:$\boldsymbol{s}_j\in\mathbb{Z}^5$为任务当前剩余需求向量;$\boldsymbol{r}_j\in\mathbb{Z}^5$为任务原始需求向量;$d_j\in R^+$为任务执行时长;$\tau_j\in R^+$为当前决策智能体到达该任务的预计行程时间;$\Delta\boldsymbol{p}_j\in[-1,1]^2$为任务相对于当前决策智能体的位置偏移;$f_j\in\{0,1\}$为任务是否已满足分配条件的标志。任务节点特征总维度为15维。

算法1 任务分配算法

输入:任务节点特征集合x^t、智能体节点特征集合x^a、边集合E
输出:当前决策智能体的目标任务v^*

```
BEGIN
  IF v_new非空 THEN   %发生任务新增
    V_t ← V_t ∪ {v_new},更新x^t与E
  ELSE IF a_fail非空 THEN   %发生智能体失效
    mask[a_fail] ← 0
  END IF
  H^t, H^a ← 嵌入层(x^t, x^a)
  h_i^{a'} ← GAT_aa(H^a, E_aa) + GAT_ta(H^t, H^a, E_ta)
  H_t' ← GAT_at(H^a, H^t, E_at) + GAT_tt(H^t, E_tt)
  s_i ← CrossAttn(h_i^{a'}, H_t')
  P(v_j | s_i) ← Softmax((W_q s_i)^T · (W_k H_t'))
  P ← P⊙mask   %应用掩码屏蔽已完成任务
  IF 推理模式 THEN
    v* ← arg max_j P(v_j | s_i)
  ELSE
    v* ← Sample(P)
  END IF
  RETURN v*
END
```

智能体节点$v_i\in V_a$特征定义为$\boldsymbol{x}_i^a=[\boldsymbol{c}_i,\tau_i,w_i^r,w_i^w,\Delta\boldsymbol{p}_i,b_i]$,各分量含义如下:$\boldsymbol{c}_i\in\mathbb{Z}^5$为智能体能力向量;$\tau_i\in\mathbf{R}^+$为该智能体到达当前目标任务的剩余行程时间;$w_i^r\in\mathbf{R}^+$为当前任务剩余工作时间;$w_i^w\in R^+$为当前等待时间;$\Delta\boldsymbol{p}_i\in[-1,1]^2$为该智能体相对于当前决策智能体的位置偏移;$b_i\in\{0,1\}$为智能体是否已被分配任务的状态标志。智能体节点特征总维度为11维。

任务-智能体边E_{ta},建模任务需求与智能体能力之间的匹配关系。当智能体a_i能够为任务t_j的剩余需求提供有效贡献时边$e_{ji}^{ta}\in E_{ta}$存在:

$$e_{ji}^{ta}=\begin{cases}1, & \text{if } \sum_{k=1}^{K}\min(c_i^{(k)},s_j^{(k)})>0\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

任务-任务边E_{tt},建模任务之间的空间邻近关系,当任务空间距离小于阈值δ时边$e_{jk}^{tt}\in E_{tt}$存在:

$$e_{jk}^{tt}=\begin{cases}1, & \text{if } \|\Delta\boldsymbol{p}_j-\Delta\boldsymbol{p}_k\|<\delta\\ 0, & \text{otherwise}\end{cases} \tag{6}$$

智能体-智能体边E_{aa},建模异构智能体之间的协作关系。当两个智能体能力互补时边$e_{il}^{aa}\in E_{aa}$存在:

$$e_{il}^{aa}=\begin{cases}1, & \text{if } \boldsymbol{c}_i\cdot\boldsymbol{c}_l=0\\ 0, & \text{otherwise}\end{cases} \tag{7}$$

动态图构建机制赋予了网络对场景变化的实时感知与自适应能力。图结构的更新频率与系统决策频率保持一致,即在每个决策时间步,算法根据

当前最新任务状态和智能体状态重新计算式(5)~式(7),从而实时刷新边集合。当任务新增时,根据异构特征匹配规则建立新任务节点与相关智能体之间的连接边,实现图结构的动态扩展;当智能体失效时,利用有效性掩码对失效节点进行标记,将其关联边的聚合权重置0,从而实现对失效个体的逻辑隔离。

3.2　图注意力网络

本文设计的图注意力网络如图1所示,主要由两个组件组成:信息编码器和任务选择概率解码器。信息编码器通过堆叠的图注意力模块对任务节点和智能体节点进行特征编码,得到任务分配场景的信息。任务选择概率解码器根据当前决策智能体的状态,输出覆盖所有可选任务的动作概率分布。

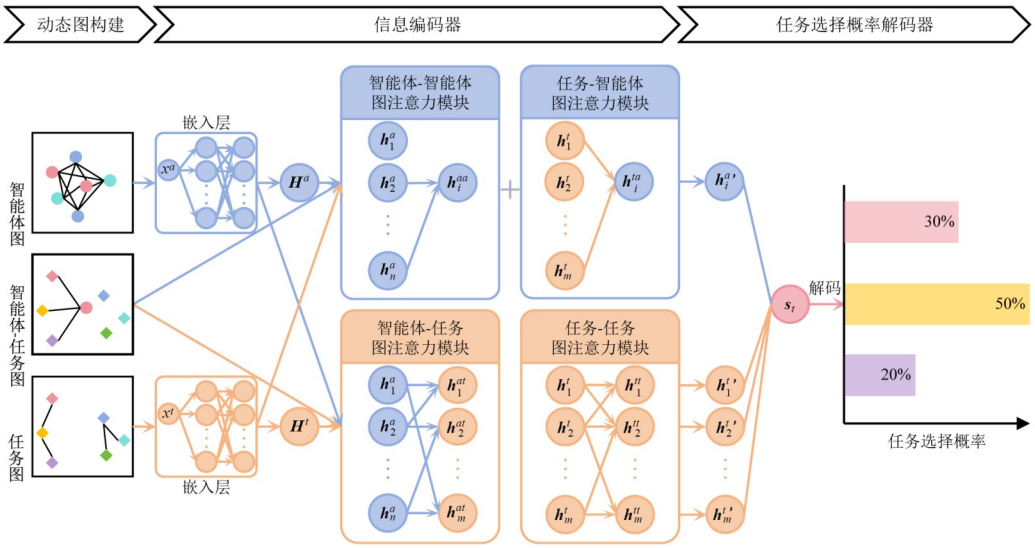

图1　网络结构示意图

Figure 1　Schematic diagram of network architecture

3.2.1　信息编码器

任务节点和智能体节点具有不同的特征维度和语义含义,为了实现节点间的注意力交互,编码器通过两个独立的嵌入层将不同维度的任务和智能体节点的原始特征映射到同一维度的嵌入空间。两个嵌入层保持相同的网络结构但不共享参数,从而在保持节点语义特性的同时实现特征空间的统一。在获取嵌入后的智能体特征 $\boldsymbol{H}^a=\{\boldsymbol{h}_1^a,\boldsymbol{h}_2^a,\cdots,\boldsymbol{h}_n^a\}$ 和任务特征 $\boldsymbol{H}^t=\{\boldsymbol{h}_1^t,\boldsymbol{h}_2^t,\cdots,\boldsymbol{h}_m^t\}$ 后,编码器通过图注意力模块进行节点间的信息融合。为了建模不同节点间的依赖逻辑,本文针对图中的任务和智能体的不同交互方向设计了四组独立参数的图注意力计算模块。具体而言,智能体-智能体模块负责学习团队成员间的隐式协作模式,任务-任务模块用于提取任务分布的空间聚类特征,而智能体-任务及任务-智能体模块则分别从供给和需求两个方向量化能力与任务的契合度。具体计算时,注意力机制仅对存在有效边连接的节点进行特征交互,这种基于稀疏邻接关系的局部交互机制避免了全局注意力计算带来的冗余噪声与高计算复杂度。最终将智能体-智能体和智能体-任务对应的模块输出聚合为智能体信息 $\boldsymbol{h}_i^{a\prime}$,将任务-任务和智能体-任务对应的模块输出聚合为任务信息 $\boldsymbol{H}_t'=\{\boldsymbol{h}_1^{t\prime},\boldsymbol{h}_2^{t\prime},\cdots,\boldsymbol{h}_m^{t\prime}\}$。在参数设置层面,本文设置编码器嵌入维度 $d_{\text{model}}=128$,包含 $L=3$ 层堆叠的图注意力模块。

3.2.2　任务选择概率解码器

为了使智能体在决策时能感知全局任务状态,解码器引入交叉注意力机制解码当前智能体的状态。在这一过程中,$\boldsymbol{h}_i^{a\prime}$ 通过线性变换获得查询矩阵 $\boldsymbol{Q}^a$,$\boldsymbol{H}_t'$ 通过线性变换获得键与值矩阵 $\boldsymbol{K}^t$、$\boldsymbol{V}^t$,通过注意力计算获得智能体状态 $\boldsymbol{s}_i$:

$$\boldsymbol{s}_i=\text{softmax}\left(\frac{\boldsymbol{Q}^a(\boldsymbol{K}^t)^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V}^t \tag{8}$$

随后,解码器计算智能体与每个候选任务的匹配分数 $u_j=(\boldsymbol{W}_q\boldsymbol{s}_i)^{\mathrm{T}}\cdot(\boldsymbol{W}_k\boldsymbol{h}_j')$。最终,通过 softmax 函数对匹配分数进行归一化,生成覆盖所有可选任务的动作概率分布:

$$p(v_j|\boldsymbol{s}_i)=\frac{\exp(u_j)}{\sum_{k=1}^{m}\exp(u_k)} \tag{9}$$

训练阶段采用随机采样机制,使智能体能够遍历动作空间,避免策略过早收敛于局部极值;而推理阶段则执行确定性的贪婪搜索,选择概率最高的任务以

保证解的稳定性。解码器的输出并非固定维度的分类标签，而是直接指向输入序列中动态元素的概率指针，使网络能够在不改变网络结构或重新训练的情况下，适应由任务新增导致的动作空间维度的实时变化。

3.3 多阶段课程学习训练策略

本文采用 REINFORCE（REward Increment = Non-negative Factor × Offset Reinforcement × Characteristic Eligibility）[26]策略梯度算法训练策略网络。奖励函数定义为所有任务的完成时间 $R=-T$。但在训练初期，智能体极易因无法在最大时限 $T_{\max}$ 内完成所有任务而获得无效的奖励信号。针对这一稀疏奖励问题[27]，本文设计了渐进式奖励机制，引入任务完成率 $\rho=k_{\text{done}}/k_m$ 与参数 λ，当智能体未在时限内完成所有任务时，将奖励重塑为 $R=-T_{\max}-\lambda(1-\rho)$，从而确保网络在此时也能获得指导优化的梯度信号。为了缓解策略梯度的高方差问题并提升训练稳定性，本文基于多最优策略优化（Policy Optimization with Multiple Optima，POMO）[28]的思想，对同一场景进行 B 次采样，将平均奖励作为基线 $b_{\text{shared}}=\sum R_i/B$，最终损失函数的梯度定义为

$$\nabla_\theta J(\theta)\approx \frac{1}{B}\sum_{i=1}^{B}\sum_{t=0}^{T}\nabla_\theta \ln \pi_\theta(a\,|\,s_i^t)(R_i-b_{\text{shared}}) \quad (10)$$

上述机制缓解了梯度估计的方差与稀疏性问题，但在面对大规模动态异构场景中指数级增长的状态-动作空间时，仍难以在合理时间内实现训练收敛。针对这一挑战，本文设计了涵盖任务规模与环境动态性的多阶段课程学习机制，将训练过程划分为多个难度递增的课程阶段。在规模维度上，设定了从小规模（10~15 任务）逐步过渡到大规模（35~50 任务）的四个数量级；在动态性维度上，构建了从静态环境到高度动态（含 30% 新任务及 15% 智能体失效）的四个扰动梯度。阶段切换遵循以下量化标准：采用滑动窗口机制评估策略性能，窗口大小设为 $W=512$ 个回合，以减少短期性能波动对阶段切换判断的影响。当策略在评估窗口内的平均任务完成率超过升级阈值 $\theta_{\text{up}}=90\%$ 时，晋升至下一难度等级。若完成率跌破 $\theta_{\text{down}}=20\%$，则回退至上一阶段以重新巩固基础策略。同时设置降级冷却期 $T_{\text{cool}}=512$ 个回合，防止策略在阶段边界处频繁震荡。

在训练稳定性方面，为确保策略在不同难度阶段间的平滑过渡，本文采用统一的超参数配置策略。学习率初始值设为 $\eta_0=5\times10^{-5}$，并采用阶梯式衰减策略，每隔 10^4 步将学习率乘以 0.98。为了抑制大规模场景下可能出现的梯度爆炸问题，本文将梯度裁剪阈值设为 $\gamma_{\text{clip}}=1.0$。这种统一超参数+自适应课程的训练模式有效平滑了高维参数空间中的非凸优化曲面，使策略无需针对不同难度阶段手动调参，引导策略在早期快速捕捉基础异构匹配逻辑，进而逐步习得应对复杂动态变化的泛化能力，提升算法的训练效率与解的质量。

4 仿真结果与分析

为评估所提出算法的有效性、规模扩展性、动态适应性和分配方案质量，本章在不同设置下进行仿真。仿真在配备 AMD Ryzen 99950X，NVIDIA GeForce RTX 5080，32.0 GB RAM 的计算机上进行。

4.1 方法有效性验证

为验证所提方法在异构多智能体动态任务分配中的有效性，本文进行了如图 2 所示的仿真验证。

仿真中部署了四种具有不同能力的异构智能体，分别以红色、蓝色、橙色和紫色圆形节点表示，各类智能体初始分布于场景四角。任务节点呈现为方形区域，其内部由多个彩色槽位组成，每个槽位的颜色对应所需智能体类型，槽位数量表示该类型智能体的需求数量。任务状态通过边框颜色区分，灰色边框表示待分配状态，绿色边框表示已完成状态。

图 2 第一行展示了静态仿真下的初始任务分配流程。帧（1）中，12 个智能体分布于场景四角，6 个任务随机分布于中央区域。帧（2）~（4）展示了分配进程，图注意力网络通过计算智能体能力向量与任务需求向量间的异构匹配权重，驱动各类智能体向最优目标任务移动。当具有互补能力的多个智能体到达同一任务位置时，组成执行联盟并启动任务执行。在帧（5）中，所有初始任务均已完成。第二行展示了任务新增事件的响应过程。在初始任务执行过程中，帧（6）处场景动态新增了 3 个新任务。本文提出的动态图构建机制实时扩展图结构，将新任务节点及其关联边纳入决策空间。在帧（7）~（9）中，网络重新计算匹配权重，空闲智能体被分配至新增任务。在帧（10）中，包括新增任务在内的所有任务均已完成。第三行展示了智能体失效事件的响应过程。在帧（11）中再次新增任务，任务分配进程进行至帧（12）时，一个紫色智能体突然失效。网络通过更新图节点的有效性掩码，使失效节点不再参与权重计算，并触发剩余智能体与任务间匹配权重的重计算。在帧（13）中，同类型的空闲智能体被重新分配以接管失效智能体的任务。帧（14）~（15）展示了动态调整后的任务执行与最终完成状态。该仿真验证了所提方法对场景动态变化的实时感知与自适应调整能力。

4.2 规模扩展性验证

为验证本文方法的规模扩展性，本文在不同规模场景下将所提方法和贪婪启发式方法[29]、拍卖方

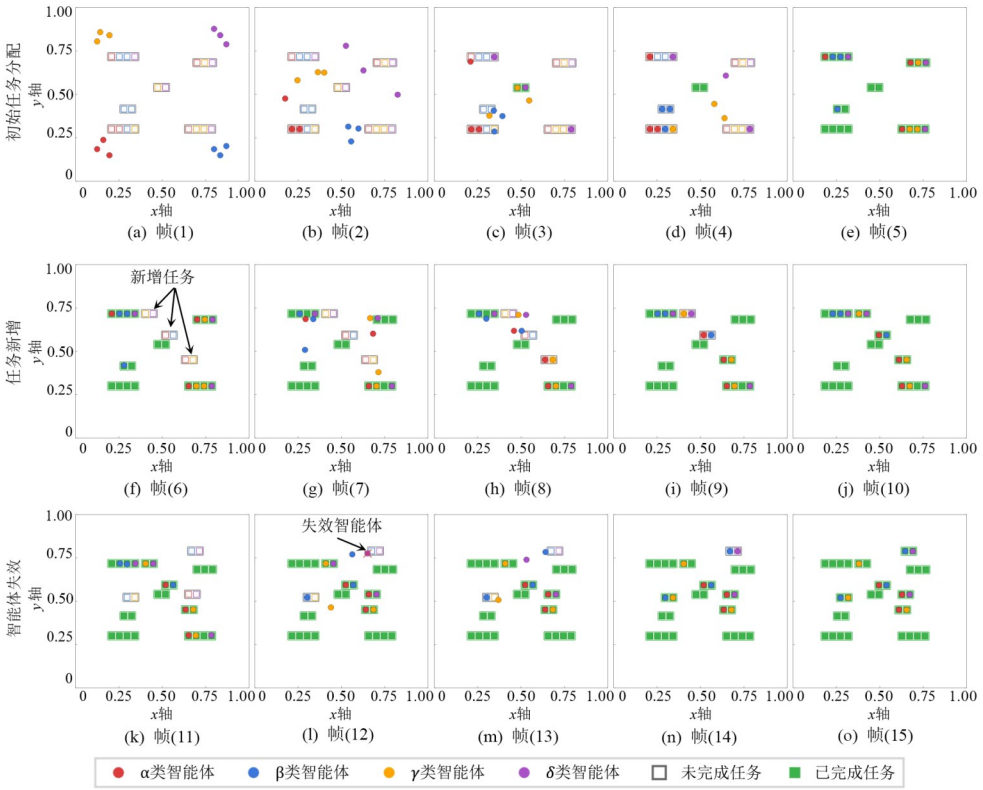

图2　异构多智能体系统动态任务分配仿真结果

Figure 2　Simulation results of dynamic task allocation for heterogeneous multi-agent systems

法[30]和采用全局注意力架构的学习方法[基于强化学习的动态联盟形成与路由的多机器人任务分配(Dynamic Coalition formation and routing for MultiRobot Task Allocation via reinforcement learning, DCMRTA)方法][11]执行单次决策的耗时情况进行统计与横向对比。场景中,智能体数量N从20递增至100,任务数量与之保持3∶1的比例,共有5种类型的异构智能体,每种类型智能体的数量均为$N/5$。在每个规模配置下,智能体与任务的位置及能力需求均随机生成,并重复执行100次独立仿真以获取稳定的统计结果。

图3为各方法决策耗时随智能体数量变化的箱线图对比,箱体表示四分位距范围,折线连接各规模下的均值。由图3可知,贪婪方法和拍卖方法在20智能体的小规模场景下具有较低的决策耗时,分别约2.2 ms和2.5 ms,这得益于其简单的计算逻辑。然而,随着规模增长,两者耗时呈现超线性增长趋势,当智能体数量由20扩展至100时,贪婪方法耗时增长至12.8 ms,拍卖方法增长至16.3 ms,这反映了传统方法较高的计算复杂度特性。相比之下,本文方法与DCMRTA方法的耗时增长明显平缓,呈现亚线性增长趋势。DCMRTA方法计算复杂度为$O\left((N+M)^2\right)$,虽然得益于神经网络的并行计算特性使其在大规模场景下仍具备扩展能力,但全局注意力架构带来的冗余计算使其决策耗时始终高于本文方法。本文引入基于能力贡献度、空间邻近度的稀疏图构建机制,确保有效边数E与总节点数$(N+M)$维持线性关系,图注意力的计算复杂度降至$O(E)$,从而在所有规模下均保持更低的决策耗时。当规模达到100个智能体时,本文方法比DCMRTA方法快31%,比贪婪方法快25%,比拍卖方法快41%,体现其在效率与扩展性上的综合优势。

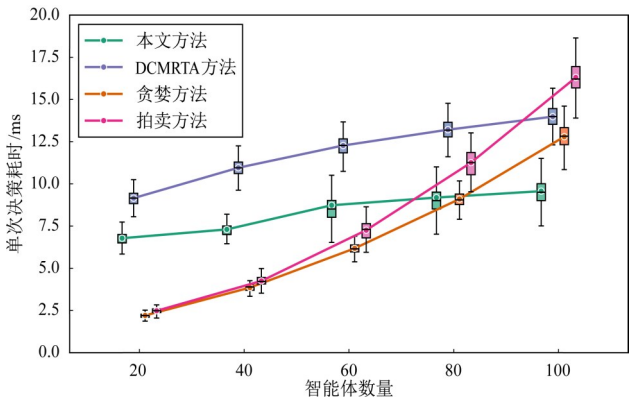

图3　单次决策耗时对比图

Figure 3　Comparison of single decision-making time

4.3 动态适应性验证

为评估所提方法的动态适应能力，本文在不同的动态场景中将所提方法和其他方法进行对比。对比结果如表1所示。

仿真实验在包含5种异构类型共20个智能体与30个初始任务的基础场景上，设计了任务新增、智能体失效和混合动态三类动态场景，并划分了轻度、中度、高度三种难度梯度，对应于占初始任务量的10%、20%、30%的任务新增和占智能体总数5%、10%、15%的智能体失效，混合场景同时包含对应级别的任务新增和智能体失效事件。动态事件的触发机制设定如下：依据难度梯度预先确定新增任务与失效智能体的数量，采用基于任务完成进度的触发策略，将事件均匀映射至任务完成进度轴，确保各方法在相同完成进度节点触发相同事件。新增任务的位置与需求随机生成，失效智能体从各类型中均匀选取。评价体系涵盖三个指标：成功率表示成功完成所有任务的场景占总场景的比例；完成时间表示在场景中完成所有任务的时间；时间增量表示动态场景下的完成时间相比无动态事件的静态场景所增加的时间。

由结果可知，本文方法在所有动态配置下均展现出最优的综合性能，不仅保持100%的成功率，且任务完成时间及其增量最小。本文方法的平均完成时间较贪婪方法降低约23%，较拍卖方法降低约50%，较DCMRTA方法降低约7%。本文方法利用动态图构建机制实时映射场景变化，并通过图注意力网络重新计算智能体与新任务之间的能力匹配权重，从而能够在全局视角下快速生成适应当前场景变化的策略，实现对全局高质量解的逼近。在混合高动态场景下，本文方法的计算开销依然保持稳定，这是因为图注意力的计算复杂度为$O(E)$，动态图构建机制确保边数E与节点数$(N+M)$呈线性关系，加之神经网络的并行计算特性，动态事件引起的节点规模变化对决策耗时影响有限，不产生显著波动。

4.4 大规模场景泛化能力验证

为验证所提方法在大规模场景下的泛化能力，本文在不同规模场景下将所提方法和其他方法进行对比，动态难度梯度设置为占初始任务量的20%的任务新增和占智能体总数10%的智能体失效的中度混合动态场景。结果如表2所示。

表1 动态场景对比实验结果

Table 1 Comparison of experimental results in dynamic scenarios

难度	方法	任务新增			智能体失效			混合场景		
		成功率/%	完成时间/ms	时间增量/ms	成功率/%	完成时间/ms	时间增量/ms	成功率/%	完成时间/ms	时间增量/ms
轻度	贪婪方法	100	49.59	+3.90	100	48.03	+2.66	100	52.76	+7.07
	拍卖方法	100	76.89	+6.48	79	73.40	+3.16	67	79.82	+9.31
	DCMRTA方法	100	41.21	+3.56	100	41.21	+3.41	100	44.77	+7.08
	本文方法	**100**	**38.56**	**+3.07**	**100**	**38.05**	**+2.34**	**100**	**41.97**	**+6.98**
中度	贪婪方法	100	53.99	+8.30	100	50.89	+5.20	100	61.32	+15.63
	拍卖方法	100	83.50	+13.10	67	75.21	+5.71	59	91.88	+21.84
	DCMRTA方法	100	45.12	+7.25	97	42.89	+5.14	100	51.63	+14.24
	本文方法	**100**	**42.25**	**+6.94**	**100**	**40.20**	**+4.86**	**100**	**48.78**	**+13.33**
高度	贪婪方法	100	58.44	+12.75	100	52.45	+6.76	100	67.68	+21.99
	拍卖方法	100	89.98	+19.58	45	78.81	+8.48	41	103.19	+33.94
	DCMRTA方法	100	48.94	+11.39	96	44.27	+6.31	96	56.47	+18.64
	本文方法	**100**	**45.07**	**+9.61**	**100**	**41.40**	**+6.26**	**100**	**54.06**	**+18.46**

注：加粗字体为本文方法测试结果。

由对比结果可知，在大规模场景中，本文方法在保持100%成功率的同时，完成时间和时间增量低于其他算法。这是因为图注意力网络的权重参数独立于节点规模，学习的是通用的异构交互逻辑而非特定场景，具备大规模场景的泛化能力。与此同时，图注意力机制在信息传递中发挥滤波作用，通过抑制低相关节点的噪声干扰有效收敛搜索空间，确保方法在解空间膨胀的大规模场景下仍能获取高质量解。

4.5 多阶段课程学习消融实验

为验证多阶段课程学习训练策略的有效性，本文开展了对比消融实验。仿真实验设置两组训练策略：第一组为本文方法，遵循课程学习框架，场景配置从小规模静态向大规模高度动态演进；第二组直接在大规模高度动态场景中进行训练。图4展示了两组训练策略在10 000个训练批次下的成功率曲线。

由结果可知，采用无课程学习的策略由于面临巨

表2　大规模场景对比实验结果

Table 2　Comparison of experimental results in large-scale scenarios

规模	方法	成功率/%	完成时间/ms	时间增量/ms
100个智能体 100个任务	贪婪方法	100	52.26	+22.22
	拍卖方法	10	94.50	+34.40
	DCMRTA方法	100	43.47	+17.11
	本文方法	**100**	**41.68**	**+15.58**
100个智能体 300个任务	贪婪方法	100	149.72	+72.97
	拍卖方法	0	—	—
	DCMRTA方法	100	113.94	+55.10
	本文方法	**100**	**108.60**	**+51.86**
200个智能体 200个任务	贪婪方法	100	53.21	+23.06
	拍卖方法	0	—	—
	DCMRTA方法	100	45.47	+18.48
	本文方法	**100**	**42.02**	**+16.71**
200个智能体 600个任务	贪婪方法	100	136.07	+60.58
	拍卖方法	0	—	—
	DCMRTA方法	100	106.32	+47.29
	本文方法	**100**	**102.92**	**+44.51**

注:加粗字体为本文方法测试结果。

大的状态-动作空间与极度稀疏的奖励信号,难以获取有效的正向反馈,导致成功率长期停滞在55%左右,陷入局部最优。相比之下,本文提出的多阶段课程学习训练策略呈现出阶梯式上升趋势。尽管在课程阶段切换节点场景突变会导致成功率下降,但得益于前期在简单场景中习得的基础分配逻辑,网络能够迅速适应新环境并实现性能提升。这说明本文所提出的多阶段课程学习训练策略通过平滑优化曲面,解决了在大规模动态任务分配场景中的收敛难题。

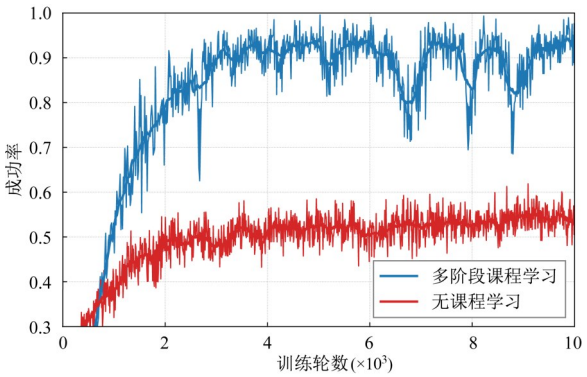

图4　两种训练策略成功率对比图

Figure 4　Comparison of success rates between two training strategies

5　结束语

本文针对异构多智能体系统动态任务分配问题,提出了一种基于图注意力的任务分配方法。通过引入动态图构建机制,实现了对任务新增与智能体失效动态事件的实时感知;利用图注意力编解码架构,通过独立的注意力通道解耦异构节点的特征语义,实现了能力供给与任务需求的匹配及对变长输入的自适应决策;配合涵盖双维度的多阶段课程学习策略,平滑了大规模任务分配下的优化曲面,解决了稀疏奖励下难以收敛的挑战。仿真结果显示,本文所提方法在各类动态场景下保持100%的分配成功率,在含30%新任务及15%智能体失效的高度动态场景下,任务完成时间较传统方法降低20%以上。在200个智能体、600个任务的大规模场景下,方法仍能保持毫秒级决策速度与100%成功率。

未来的研究工作可以围绕以下方向展开:(1)引入新的网络模块学习时间依赖性,以处理具有时间窗口或先决条件约束的任务;(2)在通信受限的场景下,本文方法可能面临信息同步延迟的挑战,可将智能体的通信与观察限定于局部范围,设计阶段性的奖励与训练机制使网络适应分布式局部观察情况下的任务分配;(3)在任务分配层面显式考虑智能体间的碰撞约束,实现任务分配与无碰撞路径规划的端到端求解。通过上述方法的持续研究,有望进一步提升图注意力网络的任务分配效率与实际部署能力。

参考文献

[1] 饶凌风, 耿娜, 张勇, 等. 不确定环境下无人机任务分配的种群交互式粒子群算法[J]. 电子学报, 2025, 53(8): 2678-2690.

Rao Lingfeng, Geng Na, Zhang Yong, et al. Population in-

teractive particle swarm optimization algorithm for UAV task allocation in uncertain environments[J]. Acta Electronica Sinica, 2025, 53(8): 2678-2690. (in Chinese)

[2] Shi Qinru, Liu Meiqin, Zhang Senlin, et al. Reinforcement learning for multi-agent path finding in large-scale warehouses via distributed policy evolution[J]. IEEE Robotics and Automation Letters, 2025, 10(8): 7843-7850.

[3] Li Liuchun, Yang Bisheng, Chen Chi, et al. Intelligent multi-robot exploration in non-exposed spaces: Methods and challenges[J]. Artificial Intelligence Review, 2025, 58(12): 394.

[4] Athira K A, Divya Udayan J, Subramaniam U. A systematic literature review on multi-robot task allocation[J]. ACM Computing Surveys, 2025, 57(3): 68.

[5] Khamis A, Hussein A, ELMOGY A. Multi-robot task allocation: A review of the state-of-the-art[M]//Koubâa A, Martínez-De Dios J R. Cooperative robots and sensor networks 2015. Heidelberg: Springer, 2015: 31-51.

[6] Suslova E, Fazli P. Multi-robot task allocation with time window and ordering constraints[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE, 2020: 6909-6916.

[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] Zhang Yudong, Wang Shuihua, Ji Genlin. A comprehensive survey on particle swarm optimization algorithm and its applications[J]. Mathematical Problems in Engineering, 2015, 2015(1): 931256.

[9] Bezerra L C D, Dos Santos A M G, Park S. Learning policies for dynamic coalition formation in multi-robot task allocation[J]. IEEE Robotics and Automation Letters, 2025, 10(9): 9216-9223.

[10] Kargar E, Kyrki V. MACRPO: Multi-agent cooperative recurrent policy optimization[J]. Frontiers in Robotics and AI, 2024, 11: 1394209.

[11] Dai Weiheng, Bidwai A, Sartoretti G. Dynamic coalition formation and routing for multirobot task allocation via reinforcement learning[C]//2024 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE, 2024: 16567-16573.

[12] Kool W, Van Hoof H, Welling M. Attention, learn to solve routing problems![PP/OL]. V3.arVix (2019-02-07)[2026-01-26]. https://arxiv.org/abs/1803.08475.

[13] Nazari M, Oroojlooy A, Takáč M, et al. Reinforcement learning for solving the vehicle routing problem[PP/OL]. V2. arVix (2018-05-21) [2026-01-26]. https://arxiv. org/abs/1802.04240.

[14] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[PP/OL]. V7. arVix (2023-08-02) [2026-01-26]. https://arxiv.org/abs/1706.03762.

[15] Wang Zheyuan, Gombolay M. Learning scheduling policies for multi-robot coordination with graph attention networks[J]. IEEE Robotics and Automation Letters, 2020, 5(3): 4509-4516.

[16] Jose W J, Zhang Hao. Learning for dynamic subteaming and voluntary waiting in heterogeneous multi-robot collaborative scheduling[C]//2024 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE, 2024: 4569-4576.

[17] Dai Weiheng, Rai U, Chiun J, et al. Heterogeneous multi-robot task allocation and scheduling via reinforcement learning[J]. IEEE Robotics and Automation Letters, 2025, 10(3): 2654-2661.

[18] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[PP/OL]. V3. arVix (2018-02-04) [2026-01-26]. https://arxiv.org/abs/1710.10903.

[19] 袁丁, 李源, 孟羽倩, 等. 基于时空注意力Transformer的自动驾驶运动规划方法[J]. 电子学报, 2025, 53(7): 2418-2427.
Yuan Ding, Li Yuan, Meng Yuqian, et al. A motion planning method for autonomous driving based on spatiotemporal attention transformer[J]. Acta Electronica Sinica, 2025, 53(7): 2418-2427. (in Chinese)

[20] Zhang Jiani, Shi Xingjian, Xie Junyuan, et al. GaAN: Gated attention networks for learning on large and spatiotemporal graphs[PP/OL]. V1. arVix (2018-03-20) [2026-01-26]. https://arxiv.org/abs/1803.07294.

[21] Wang Xiao, Ji Houye, Shi Chuan, et al. Heterogeneous graph attention network[C]//Proceedings of the World Wide Web Conference. New York: ACM, 2019: 2022-2032.

[22] Peng Juntong, Viswanath H, Bera A. Graph-based decentralized task allocation for multi-robot target localization[J]. IEEE Robotics and Automation Letters, 2024, 9(11): 10676-10683.

[23] Du Wei, Ding Shifei, Zhang Chenglong, et al. Multiagent reinforcement learning with heterogeneous graph attention network[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(10): 6851-6860.

[24] Zhang Zhenqiang, Jiang Xiangyuan, Yang Zhenfa, et al. Scalable multi-robot task allocation using graph deep reinforcement learning with graph normalization[J]. Electron-

ics, 2024, 13(8): 1561.

[25] Lu Zehui, Zhou Tianyu, Mou Shaoshuai. Real-time multi-robot mission planning in cluttered environment[J]. Robotics, 2024, 13(3): 40.

[26] Williams R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning[J]. Machine Learning, 1992, 8(3/4): 229-256.

[27] 赵世钰. 强化学习的数学原理[M]. 北京: 清华大学出版社, 2024.
Zhao Shiyu. Mathematical foundations of reinforcement learning[M]. Beijing: Tsinghua University Press, 2024. (in Chinese)

[28] Kwon Y D, Choo J, Kim B, et al. POMO: Policy optimization with multiple optima for reinforcement learning[PP/OL]. V3.arVix (2021-07-13)[2026-01-26]. https://arxiv.org/abs/2010.16011.

[29] Shin H S, Li Teng, Lee H I, et al. Sample greedy based task allocation for multiple robot systems[J]. Swarm Intelligence, 2022, 16(3): 233-260.

[30] Lagoudakis M G, Berhault M, Koenig S, et al. Simple auctions with performance guarantees for multi-robot task allocation[C]//2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE, 2004: 698-705.

作者简介

李中杨　男,2003年11月出生于浙江省金华市。现为北京航空航天大学沈元学院博士研究生。主要研究方向为多智能体强化学习与机器人智能控制。
E-mail: lizhongyangx@buaa.edu.cn

孙贵宾　男,1995年1月出生于内蒙古自治区乌兰察布市。现为北京航空航天大学自动化科学与电气工程学院副教授。主要研究方向为机器人集群协同控制与决策优化。
E-mail: sunguibinx@buaa.edu.cn

曹筱可　女,2001年6月出生于陕西省延安市。现为北京航空航天大学自动化科学与电气工程学院硕士研究生。主要研究方向为集群协同任务规划。
E-mail: caoxk@buaa.edu.cn

刘克新　男,1988年12月出生于山东省聊城市。现为北京航空航天大学自动化科学与电气工程学院教授。主要研究方向为集群协同制导与控制。
E-mail: kxliu@buaa.edu.cn

蔡奕辰　男,2000年4月出生于北京市。现为北京航空航天大学人工智能学院博士研究生。主要研究方向为多智能体系统与集群协同控制。
E-mail: caiyichen@buaa.edu.cn