研究与开发

多头图注意力驱动特征聚合的巨型遥感星座任务规划算法

佘雨璇[1]，周康燕[2]，代奥[1]，魏子翔[1]，周家喜[1]，肖丽霞[1]
（1. 华中科技大学网络空间安全学院第六代移动通信研究中心，湖北 武汉 430074；
2. 西安电子科技大学电子工程学院，陕西 西安 710071）

摘　要：遥感卫星星座规模的持续扩大，极大地增强了其在资源普查、环境监测和灾害应急等领域的综合应用能力。然而，星座规模增长也显著增加了任务规划的复杂性。多智能体强化学习是解决大规模星座任务规划的有效途径，但卫星数量激增也带来智能体状态空间维度爆炸难题。为此，提出一种多头图注意力驱动特征聚合（multi-head graph attention driven feature aggregation，MGADFA）的巨型遥感星座任务规划算法。该方法利用多头图注意力捕捉卫星间的动态交互权重，通过特征聚合机制将复杂的多体交互转化为个体与群体的关联，在保留协作特征的同时，将网络参数规模从指数级降为线性级。仿真结果表明，所提算法在任务接受量和负载均衡性等方面均表现更优。

关键词：巨型遥感星座；多智能体强化学习；图注意力网络；特征聚合；任务规划
中图分类号：TN927.2；TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2026119

Multi-head graph-attention driven feature aggregation for large-scale satellite mission planning

She Yuxuan[1], Zhou Kangyan[2], Dai Ao[1], Wei Zixiang[1], Zhou Jiaxi[1], Xiao Lixia[1]
1. School of Cyber Science and Engineering, 6G Research Center, Huazhong University of Science and Technology, Wuhan 430074, China
2. School of Electronic Engineering, Xidian University, Xi’an 710071, China

Abstract: The continuous expansion of remote-sensing satellite constellations has significantly enhanced their comprehensive application capabilities in resource surveys, environmental monitoring, and emergency disaster response. However, the burgeoning scale of these constellations markedly increases the complexity of mission planning. While multi-agent reinforcement learning is an effective approach for large-scale mission planning, the surge in satellite numbers leads to the challenge of dimensionality explosion in the state space. To address this bottleneck, a multi-head graph attention driven feature aggregation (MGADFA) algorithm for mission planning in giant remote-sensing constellations was proposed. This method utilized multi-head graph-attention to capture dynamic interaction weights be-

收稿日期：2025−12−02；修回日期：2026−01−27
通信作者：肖丽霞，lixiaxiao@hust.edu.cn
基金项目：国家自然科学基金资助项目（No.62571202）
Foundation Item: The National Natural Science Foundation of China (No.62571202)

tween satellites and employed a feature aggregation mechanism to transform complex multi-body interactions into individual-population associations. While preserving key cooperative features, this approach reduced the joint decision space from an exponential scale to a linear dimension. Simulation results demonstrated that the proposed algorithm exhibited superior performance in terms of task throughput and load balancing.

Key words: large-scale remote sensing constellation, multi-agent reinforcement learning, graph attention network, feature aggregation, task scheduling

0 引言

随着卫星遥感技术的持续进步，以及对遥感任务要求的不断提高，构建由上百甚至上千颗卫星组成的巨型遥感星座正逐步成为全球航天发展的重要方向。这类星座在资源普查、环境监测、灾害应急以及军事侦察监视等领域展现出巨大应用潜力[1]。然而，随着星座规模的不断扩大，以及光学、合成孔径雷达（synthetic aperture radar，SAR)和电子侦察等载荷类型的日益多样化，其运行过程中所面临的任务规划复杂度急剧上升。在多用户、多任务、多约束等复杂需求条件下，实现海量卫星的高效协同调度成为一项关键任务。

在遥感卫星数量较少时，数学规划方法凭借严谨的建模能力，通过动态规划等方法[2]能够为任务规划问题提供全局最优解。然而，其计算复杂度随卫星数量和任务规模呈指数增长，难以满足大规模星座对实时性和动态适应性的要求[3]。启发式算法在计算效率上具备一定优势，能够在有限时间内获得较优解，满足一定程度的实时性要求。因此，搜索效率更高的启发式算法逐渐被广泛应用于较大规模的遥感星座。为提升求解效率，遗传算法、粒子群优化和模拟退火算法等启发式方法[4-9]被引入多星协同调度问题，在一定程度上缓解了计算压力，并通过设计合理的适应度函数来解决多目标下的优化问题。这类启发式算法在应对大量遥感卫星任务协同规划方面有一定的优势，但其依赖经验规则或策略设计，难以对解空间进行系统性搜索，易受初始条件与参数设定的影响，性能对问题实例较为敏感，导致解的稳定性和鲁棒性不足。同时，数学规划算法和启发式算法将时间窗口、存储容量等约束视为静态的约束，预设任务为固定集合。面对突发的新增任务，这种批处理机制难以实现实时响应，通常需要对全局方案进行高代价的重规划。

强化学习通过智能体与环境的持续交互不断优化决策策略，展现出在动态复杂任务场景下的优越性能，为大规模遥感星座的智能任务调度提供了全新的解决思路。强化学习通过训练好的模型进行推理，可以实现对动态任务的实时响应。文献[10-14]尝试采用分布式架构、分层决策或融合启发式规则的强化学习方法，通过设计合理的奖励函数来实现多目标优化，在一定程度上提升了动态环境下的调度效能。然而，上述单智能体的强化学习方法不具备多智能体协同决策能力，难以充分挖掘智能体间的复杂协作潜力，导致整体性能受限。

相比之下，多智能体强化学习（multi-agent reinforcement learning，MARL）能够通过多个智能体间的自主协商与协同演化，实现更高效的分布式资源调度与多目标均衡。文献[15]针对星座系统的协同优化问题，引入多智能体近端策略优化（multi-agent proximal policy optimization，MAPPO）算法，以提升多卫星系统在资源约束下的协同规划能力。当任务规划扩展至包含数百颗卫星的大规模星座时，传统MARL框架[16]的扩展性瓶颈日益凸显。传统集中式训练架构通常将所有智能体的状态向量和动作集合进行简单拼接后输入神经网络。在典型的MARL算法下，网络通常需要接收所有智能体的状态作为输入。具体

而言，若单个智能体的状态维度为d_s，隐藏层维度为H。当智能体规模为N时，单个网络的输入维度为Nd_s，其网络的参数规模达到$O(Nd_s \times H)$，整个多智能体系统参数规模为$O(N^2 d_s \times H)$。在巨型星座下，N往往达到数百甚至上千量级，这使得网络输入维度和参数规模急剧膨胀。在这种机制下，计算与存储的开销显著增加，且神经网络难以收敛，严重制约了训练效率。

本文针对多智能体强化学习算法在巨型遥感星座任务规划时面临的维度爆炸难题，提出基于多头图注意力驱动特征聚合（multi-head graph attention driven feature aggregation，MGADFA）的智能规划算法，极大降低了算法的复杂度，提升了巨型遥感星座任务规划的效率。本文的主要贡献如下。

（1）提出基于多头图注意力的星间耦合关系动态建模机制。针对巨型遥感星座中卫星间交互关系的复杂性、时变性及异构性，本文利用多头图注意力网络（multi-head graph-attention network，MGAT）对星间协作关系构建结构化模型，自适应刻画星间动态交互强度，为后续多星特征聚合提供精确关系表征基础。

（2）通过平均场（mean-field，MF）对MGAT学习到的动态权重进行动态聚合，将具有指数复杂度的N个智能体交互问题，简化为个体与加权群体行为的交互，在保证建模精度的同时，将联合网络参数规模降至线性维度，解决了大规模场景下的计算复杂度大和训练不稳定难题。

（3）构建集成MGADFA的多智能体深度确定性策略梯度（multi-agent deep deterministic policy gradient，MADDPG）算法协同训练架构。采用集中式训练分布式执行（centralized training with decentralized execution，CTDE）范式，将特征聚合生成的低维群体表征嵌入MADDPG算法的网络中，实现了巨型异构星座场景下稳定且高效的协同策略学习。仿真实验表明，本文所提算法在收敛速度、任务接受量和资源负载均衡度等多目标指标上均表现最优。

1　系统建模

遥感卫星任务规划是一个复杂的多约束优化问题，旨在最大化星座系统的整体观测效益。这需要在目标区域位置、任务优先级、卫星资源类型和时间窗口等多重约束下[17-18]合理分配卫星资源，制订最优的任务–卫星匹配方案。任务规划的核心在于评估卫星的资源状态和任务的可执行性。在任务规划过程中，资源状态需要考虑卫星的实时情况，包括剩余运行时间是否低于预设阈值、可用存储容量能否满足任务需求等。与此同时，还需要从任务对载荷类型的需求是否与卫星实际搭载载荷相匹配、任务时间窗口与卫星可见时间窗口之间的匹配度，以及目标切换所带来的时间开销等多个方面对任务可执行性进行评估。在上述多重约束的共同作用下，系统需要动态决策任务是否执行，并选择最合适的卫星进行指派，从而在提升任务响应效率的同时，实现全局观测收益的有效平衡。

1.1　遥感任务及卫星定义

遥感任务集合定义为$\mathrm{Task}=\{\mathrm{task}_1, \mathrm{task}_2, \cdots, \mathrm{task}_M\}$，其中$M$表示用户提交的遥感任务总量。每个任务$\mathrm{task}_m\ (1 \leqslant m \leqslant M)$包含六元组属性特征：$\mathrm{task}_m = <\mathrm{Geo}_m, \mathrm{Load}_m, \mathrm{Win}_m, \mathrm{Dur}_m, \mathrm{Mem}_m, \mathrm{Prio}_m>$。具体而言，$\mathrm{Geo}_m$表征目标地理位置坐标，由纬度$\mathrm{Lat}_m$和经度$\mathrm{Lon}_m$构成二元组；$\mathrm{Load}_m$指明任务所需的专用载荷类型；$\mathrm{Win}_m$以时间区间$[\mathrm{WinS}_m, \mathrm{WinE}_m]$形式定义任务可执行时段约束，其中$\mathrm{WinS}_m$为最早启动时刻，$\mathrm{WinE}_m$为最迟终止时刻，任务执行过程必须完全包含在该时间窗口内且持续时长为Dur_m；Mem_m

量化任务执行时占用的存储空间资源；Prio_m则表征任务在调度过程中的优先等级参数。

卫星集合定义为$\mathrm{Sat}=\{\mathrm{sat}_1,\mathrm{sat}_2,\cdots,\mathrm{sat}_N\}$，每颗卫星代表一个智能体且其属性可由一个多元组表示$\mathrm{sat}_i=<\mathrm{Stor}_{\max,i},\mathrm{Time}_{\max,i},\mathrm{LOAD}_i,\mathrm{WINS}_i>,1\leqslant i\leqslant N$。其中，$\mathrm{Stor}_{\max,i}$表示当前卫星最大的存储容量；$\mathrm{Time}_{\max,i}$表示当前卫星最大可运行时间；$\mathrm{LOAD}_i$表示当前卫星所携带的载荷；$[\mathrm{WinS}_i,\mathrm{WinE}_i]$记录卫星已被分配的不可变观测时段集合，作为硬性约束条件，后续任务时间窗口必须满足对应的互斥要求。

1.2 决策变量

决策变量集合定义为$X=\{x_{im}|\forall i\in\mathrm{Satellite},m\in\mathrm{Task}\}$，其中决策变量$x_{im}\in\{0,1\}$为二元变量，表征任务$m$与卫星$i$之间的映射关系。当$x_{im}=1$时，表征卫星$i$与遥感任务$m$之间建立有效的指派关系，即任务$m$被调度至卫星$i$执行；反之，则表明两者未形成可行映射关系。

1.3 约束条件

决策变量的取值需要严格满足卫星资源约束与任务时空冲突规避条件，约束条件包括：

（1）当前卫星剩余容量需要大于等于当前任务所需容量：

$$\mathrm{remain\_stor}_i\geqslant\mathrm{Mem}_m \tag{1}$$

（2）当前卫星剩余执行时间需要大于等于当前任务所需执行时间：

$$\mathrm{remain\_time}_i\geqslant\mathrm{Dur}_m \tag{2}$$

（3）当前卫星载荷需要与任务所需载荷匹配：

$$\mathrm{LOAD}_i=\mathrm{Load}_m \tag{3}$$

（4）当前任务时间窗口$[s,e]$与当前卫星已分配窗口交集为空：

$$\forall[s,e]\in\mathrm{WINS}_i,[\mathrm{WinS}_m,\mathrm{WinE}_m]\cap[s,e]=\varnothing \tag{4}$$

（5）由于卫星在执行任务之后，有一定的转换时间，因此，每两个时间窗口之间需要一定的间隔：

$$\forall[s,e]\in\mathrm{WINS}_i,[\mathrm{WinS}_m-\delta,\mathrm{WinE}_m+\delta]\cap[s,e]=\varnothing \tag{5}$$

其中，δ为一个固定的常量值。

（6）每个任务只能被执行一次，即对于任务m满足：

$$\sum_{i=1}^{N}x_{im}\leqslant 1,m\in M \tag{6}$$

2 MGADFA算法

2.1 算法结构

MGADFA算法是一种基于多头图注意力驱动特征聚合的MARL框架，用于解决巨型遥感星座任务规划中，由卫星数量激增导致的高维空间问题。MGADFA框架由MGAT和特征聚合（feature aggregation，FA）两个核心模块构成。

MGAT模块作为核心的关系表征层，旨在解决巨型星座中星间交互的复杂性、时变性及异构性。相较于传统单头机制，多头图注意力结构能够并行地从多个特征子空间提取依赖信息，动态计算交互权重，从而提供精确、多模态的协作关系表征，显著提升局部策略对环境的感知能力。每颗卫星将自身的观测状态输入MGAT模块，动态计算其与邻居卫星之间的交互权重，从而更准确地刻画卫星之间影响的强度与方向，提升局部策略对环境的感知能力。其中，根据关注维度的不同，注意力头可以分为4组：第1组注意力头重点关注卫星的可用资源，包括卫星剩余可用时间和存储容量，该组注意力头会向资源充足的卫星分配更多的权重，提高系统的负载均衡；第2组注意力头关注卫星剩余存储与任务所需存储之间的匹配关系，若卫星剩余存储与任务所需存储之间的差值越大，该卫星分配到的权重就越大；第3组注意力头关注卫星剩余可用时间与任务所需执行时间之间的匹配关系，若卫星剩余可用时间与任务所需执行时间之间的差值越大，该卫星分配到的权重就越大；第4组注意力头对多种状

态维度进行综合权衡，不只专注于单一策略，而是学习一种混合权重，综合考虑以上多种因素。

FA模块利用MGAT学习到的动态权重构建特征聚合模型，克服了传统平均场假设邻居贡献均等的局限性。在MGAT的基础上，FA利用注意力权重对其他智能体的状态与动作进行加权汇聚，将原本高维的多智能体交互信息压缩为固定维度的平均场表征。该过程有效过滤了与当前决策关联度较低的冗余交互信息，使网络在训练过程中只需要关注个体和群体层面的关系。这不仅缓解了大规模场景下状态空间膨胀导致的“维度爆炸”，而且在保证对星间动态协作关系建模精度的同时，将网络参数规模降至线性维度。

为高效利用MGADFA算法得到的特征聚合交互模型，本文构建了集成MGADFA算法的MADDPG训练方法。该方法采用集中式训练分布式执行范式，核心创新在于输入表征的重构：行动者（Actor）网络与评价者（Critic）网络不再依赖随星座规模膨胀的联合状态，而是将MGADFA算法生成的低维加权群体表征嵌入输入层。在训练阶段，Actor网络基于该低维表征输出确定性动作，Critic网络则据此评估动作价值并回传策略梯度，指导Actor更新参数，从而在巨型星座场景下有效规避了“维度爆炸”问题。同时，随着训练的不断推进，注意力权重会在策略更新和价值函数回传的共同作用下逐步发生更新，使得特征聚合结果越来越符合当前决策目标。在训练初期，聚合特征主要反映的是对多卫星交互关系的粗略刻画。随着模型不断学习，这些权重会逐渐强化对当前决策有用的信息，弱化与当前决策关联度较低的信息，从而形成更加准确的群体表征。在执行阶段，各卫星仅需要基于本地观测与群体加权特征进行决策。最终，结合资源优化筛选机制，系统在满足复杂约束的同时，向全局最优协同策略收敛。

与传统的MADDPG方法相比，MGADFA算法通过降低网络参数规模，大大提高了大规模场景中的训练效率。在传统MADDPG框架下，网络参数规模随着智能体的数量N迅速增长，其复杂度为$O(N^2 d_s \times H)$，其中，d_s表示智能体的状态维度，H表示隐藏层的大小，这使得训练难以收敛。相比之下，MGADFA算法通过引入特征聚合模块，直接将网络参数规模显著降低为线性规模。各智能体在决策过程中不再直接依赖于所有其他智能体的原始状态与动作信息，而是利用MGAT计算出的动态交互权重，再通过FA将其他智能体的状态与动作映射为一个固定维度的平均场特征向量。在该算法下，原先的高维网络参数被有效压缩，单个网络的输入维度降为$2d_s$，整个系统网络规模降为$O(2Nd_s \times H)$。因此，该算法在星座规模变大时，能够保持相对稳定的计算与存储开销，提升了算法在巨型遥感星座场景下的可扩展性，使得模型在大规模星座场景中仍能够保持稳定训练和高效推理。MGADFA模型框架如图1所示。

2.2　多星任务决策过程建模

多星任务决策过程可建模为马尔可夫博弈过程[19]。初始时刻，系统处于全局初始状态S_0。每颗卫星根据S_0做出对应的动作，系统执行联合动作A_0，随后状态转移至S_1，并获得全局单步奖励R_0。卫星们不断执行动作A，直到达到终止状态。状态空间、动作空间及奖励函数的定义如下。

（1）状态空间：全局状态空间$\boldsymbol{S}=(s_1, s_2, \cdots, s_N)$由每颗卫星的状态组合而成。对于单颗卫星来说，其状态的构成为：

$$\mathrm{s}_i=\left(\mathrm{sat}_{\mathrm{RS}}^i, \mathrm{sat}_{\mathrm{RT}}^i, \mathrm{task}_{\mathrm{ES}}^m, \mathrm{task}_{\mathrm{ET}}^m, \mathrm{task}_{\mathrm{P}}^m\right) \quad (7)$$

其中，$\mathrm{sat}_{\mathrm{RS}}^i$为当前卫星的剩余存储容量，表示可用于新任务的数据存储空间；$\mathrm{sat}_{\mathrm{RT}}^i$为当前卫星可用剩余时间，即卫星在本调度周期内尚可执行任务的剩余工作时间；$\mathrm{task}_{\mathrm{ES}}^m$为任务$m$所需的存储大小，即分配后将占用的存储空间；$\mathrm{task}_{\mathrm{ET}}^m$为任

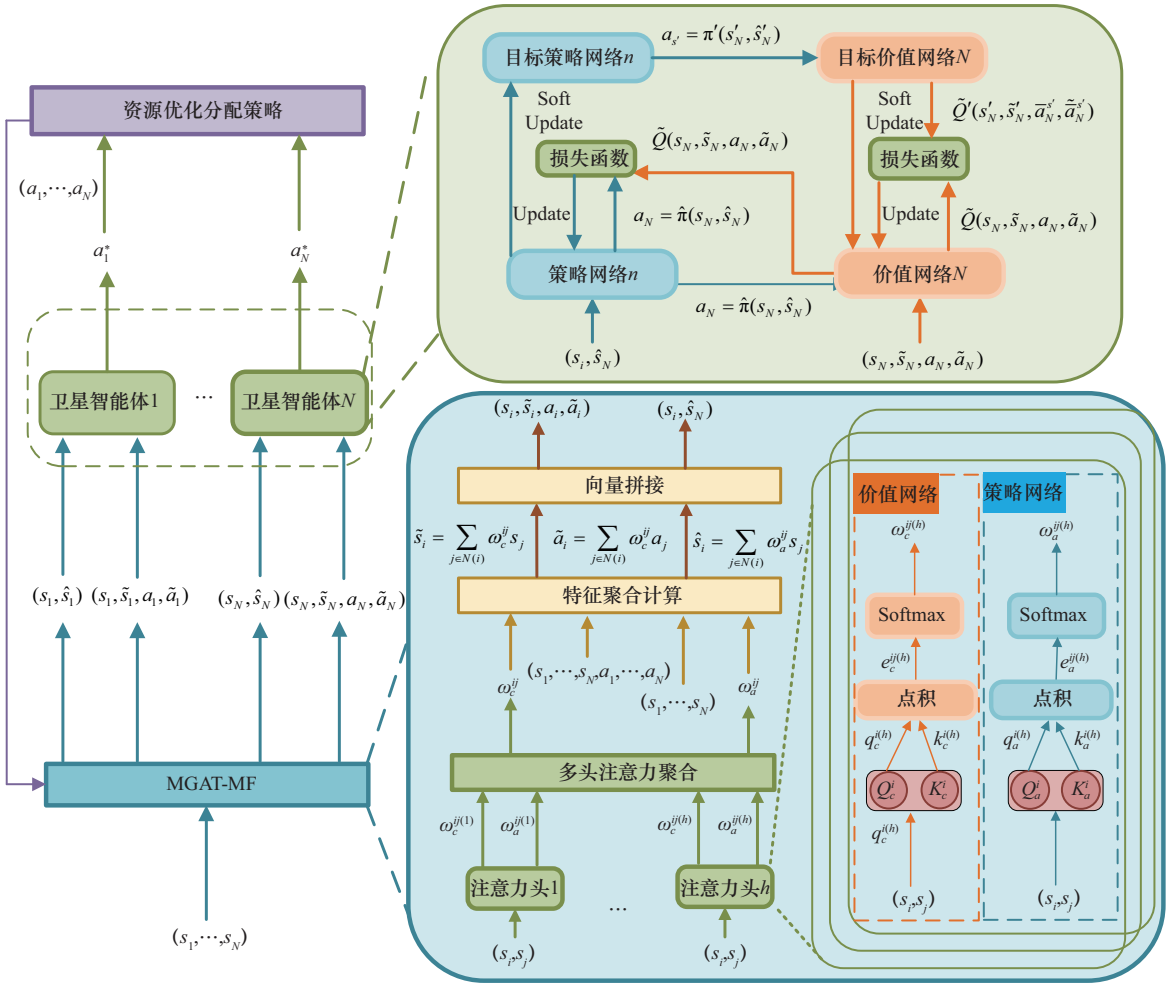

图1　MGADFA模型框架

务m所需的执行时间，即完成该任务所需的连续观测时长；$\text{task}_{\text{P}}^{m}$为任务$m$的优先级指标，数值越高，表示任务的重要性或紧急性越大，应在调度中优先考虑。

（2）动作空间：同样的，全局的动作空间$\boldsymbol{A}=(a_1,a_2,\cdots,a_N)$由每颗卫星的动作联合而成。对于第$i$颗卫星，在接收到任务$m$时有两个选择：接受任务或者拒绝任务，表示为：

$$a_i^m=\begin{cases}1,\text{接受任务}m\\0,\text{拒绝任务}m\end{cases}\tag{8}$$

该动作为离散变量，卫星i根据其局部状态s_i和策略网络Actor输出的概率分布，对a_i^m进行采样或贪心选择。全局层面上，动作向量$\boldsymbol{A}$决定了各卫星对当前任务的分配情况，进而影响系统的资源利用效率与任务完成质量。

（3）奖励函数：奖励函数作为多智能体强化学习的策略引导核心，其构建需要对齐多星任务规划系统的多目标优化需求。本问设计多维奖励机制，综合考虑任务优先级、资源均衡度及任务分配状态三重因素。针对卫星i的奖励函数，采用条件分段式设计，具体为：

$$R_i^m=\begin{cases}f_{\text{base}}+\alpha_1 r_{\text{task}}+\alpha_2\overline{T}_i+\alpha_3\overline{S}_i,\text{若}a_i^m=1\text{且}\text{flag}_m=\text{True}\\-\delta,\text{若}a_i^m=1\text{且}\text{flag}_m=\text{False}\\0,\text{若}a_i^m=0\end{cases}\tag{9}$$

其中，flag_m表示任务m是否尚未被分配。由于每个任务只能被执行一次，只有在$\text{flag}_m=\text{True}$时，卫星$i$接受任务$m$（即$a_i^m=1$）才能获得正向奖励，否则会受到一定惩罚。当$a_i^m=0$时，表示卫星$i$拒绝任务$m$，既不获得奖励，也不遭受惩罚。当且仅当$a_i^m=1$且$\text{flag}_m=\text{True}$时，卫星才会获得奖励。其中，$f_{\text{base}}$为基础奖励，$r_{\text{task}}$与任务的优先级正相关，$\overline{T}_i$表示当前卫星剩余运行时间占最大运行时间的比率，$\overline{S}_i$表示剩余存储容量占最大存储容量的比率，兼顾了任务重要性与卫星资源利用效率，有助于卫星在调度过程中权衡收益与资源消耗。α_1、α_2、α_3分别表示任务优先级、剩余时间负载均衡以及剩余存储容量负载均衡所占的权重。

2.3　算法流程

针对巨型遥感星座任务规划中的高维决策难题，本文提出的MGADFA算法利用多头图注意力网络动态捕获星间协作权重，并通过特征聚合将指数级规模的网络参数降为线性规模，最终在MADDPG框架下实现高效的集中式训练与分布式协同决策。为了在复杂的星间关系中实现精准建模，本文引入图注意力网络（graph attention network，GAT）[20]，通过在图结构中引入注意力机制，使模型能够自适应地刻画邻居节点之间的差异性，并根据其重要程度对群体信息进行加权聚合，从而实现对节点间关系的有效建模。其与传统图卷积网络[21]（graph convolutional network，GCN）在图结构中采用固定权重共享策略不同，GAT采用数据驱动方式学习节点之间的动态交互权重，从而增强对复杂拓扑结构和异构关系的适应能力。MGAT模块为每颗卫星的Actor网络和Critic网络都分别维护一个独立的查询-键值对网络架构：Actor网络包含查询网络Q_{a}^i和键网络K_{a}^i，Critic网络则对应Q_{c}^i和K_{c}^i，这种设计实现了策略生成与价值评估的解耦学习。在执行过程中，卫星i将当前观测状态s_i分别输入对应的网络模块：Q_{a}^i、K_{a}^i生成策略相关的查询向量$\boldsymbol{q}_{\text{a}}^i$与键向量$\boldsymbol{k}_{\text{a}}^i$，而$Q_{\text{c}}^i$、$K_{\text{c}}^i$则产生价值评估导向的$\boldsymbol{q}_{\text{c}}^i$和$\boldsymbol{k}_{\text{c}}^i$。通过计算卫星$i$的查询向量与除$i$外其他卫星节点$j$的键向量的余弦相似度，再对多个注意力头求平均，系统可动态生成策略网络交互权重$\boldsymbol{\omega}_{\text{a}}^{ij}$和值函数网络交互权重$\boldsymbol{\omega}_{\text{c}}^{ij}$。在训练过程中，$Q_{\text{a}}^i$、$Q_{\text{c}}^i$、$K_{\text{a}}^i$和$K_{\text{c}}^i$的网络参数随着Actor网络和Critic网络一起更新。这种端到端的参数更新机制使得交互权重$\{\boldsymbol{\omega}_{\text{a}}^{ij},\boldsymbol{\omega}_{\text{c}}^{ij}\}$能够根据环境反馈自适应调整，从而使模型逐步学习到更加合理的交互模式。MGAT获取交互权重$\{\boldsymbol{\omega}_{\text{a}}^{ij},\boldsymbol{\omega}_{\text{c}}^{ij}\}$的过程可以划分为以下3个阶段。

（1）特征空间投影。将原始节点特征通过卫星i的查询网络Q^i和卫星j的键网络K^j映射至注意力度量空间，构建查询向量$\boldsymbol{q}^{i(h)}$与键向量$\boldsymbol{k}^{j(h)}$：

$$\begin{cases}\boldsymbol{q}^{i(h)}=\boldsymbol{W}_{Q^i}^{(h)}s_i\\ \boldsymbol{k}^{j(h)}=\boldsymbol{W}_{K^j}^{(h)}s_j\end{cases},h=1,\cdots,H \quad (10)$$

其中，H表示注意力头的总数量，每个头以独立方式处理特征表示，从而学习不同的关系模式，s_i和s_j分别表示卫星i和卫星j的当前状态，$\boldsymbol{W}_{Q^i}^{(h)}$和$\boldsymbol{W}_{K^j}^{(h)}$分别表示查询网络$Q^i$和键网络$K^i$的权重，当处理Actor网络的输入时，$\boldsymbol{q}^{i(h)}=\boldsymbol{q}_{\text{a}}^{i(h)}$、$\boldsymbol{k}^{j(h)}=\boldsymbol{k}_{\text{a}}^{j(h)}$、$Q^i=Q_{\text{a}}^i$、$K^j=K_{\text{a}}^j$、$\boldsymbol{W}_{Q^i}^{(h)}=\boldsymbol{W}_{Q_{\text{a}}^i}^{(h)}$、$\boldsymbol{W}_{K^j}^{(h)}=\boldsymbol{W}_{K_{\text{a}}^j}^{(h)}$；当处理Critic网络的输入时，$\boldsymbol{q}^{i(h)}=\boldsymbol{q}_{\text{c}}^{i(h)}$、$\boldsymbol{k}^{j(h)}=\boldsymbol{k}_{\text{c}}^{j(h)}$、$Q^i=Q_{\text{c}}^i$、$K^j=K_{\text{c}}^j$、$\boldsymbol{W}_{Q^i}^{(h)}=\boldsymbol{W}_{Q_{\text{c}}^i}^{(h)}$、$\boldsymbol{W}_{K^j}^{(h)}=\boldsymbol{W}_{K_{\text{c}}^j}^{(h)}$。

（2）关联强度量化。采用点积操作度量不同节点之间的关联强度，即：

$$e^{ij(h)}=\left(\boldsymbol{q}^{i(h)}\right)^{\text{T}}\cdot\boldsymbol{k}^{j(h)} \quad (11)$$

这里采用的点积用于衡量两个向量在投影空间中的方向一致程度，从而反映在当前注意力头下节点之间的交互强弱。当处理Actor网络的输

入时，$\boldsymbol{q}^{i(h)}=\boldsymbol{q}_{\mathrm{a}}^{i(h)}$、$\boldsymbol{k}^{j(h)}=\boldsymbol{k}_{\mathrm{a}}^{j(h)}$、$e^{ij(h)}=e_{\mathrm{a}}^{ij(h)}$；当处理Critic网络的输入时，$\boldsymbol{q}^{i(h)}=\boldsymbol{q}_{\mathrm{c}}^{i(h)}$、$\boldsymbol{k}^{j(h)}=\boldsymbol{k}_{\mathrm{c}}^{j(h)}$、$e^{ij(h)}=e_{\mathrm{c}}^{ij(h)}$。

（3）注意力权重归一化与融合。通过softmax函数对原始关联强度进行概率化处理，生成归一化注意力权重：

$$\boldsymbol{\alpha}^{ij(h)}=\frac{\exp\left(e^{ij(h)}\right)}{\sum_{o\in N(i)}\exp\left(e^{io(h)}\right)} \tag{12}$$

上述注意力权重具有概率分布特性，满足$\sum_{j\in N(i)}\alpha^{ij(h)}=1$。同时，通过引入指数运算，模型能够进一步放大不同关联关系之间的权重差异，从而更加突出对关键交互关系的刻画。其中，$N(i)$为除卫星i外的卫星集合。当处理Actor网络的输入时，$\boldsymbol{\alpha}^{ij(h)}=\boldsymbol{\omega}_{a}^{ij(h)}$、$e^{ij(h)}=e_{\mathrm{a}}^{ij(h)}$；当处理Critic网络的输入时，$\boldsymbol{\alpha}^{ij(h)}=\boldsymbol{\omega}_{c}^{ij(h)}$、$e^{ij(h)}=e_{\mathrm{c}}^{ij(h)}$。

（4）最终，模型通过对所有注意力头的权重进行平均，生成一个稳定的交互表示，并且可以体现不同头部关注的多样化关系特征：

$$\boldsymbol{\alpha}^{ij}=\frac{1}{H}\sum_{h=1}^{H}\boldsymbol{\alpha}^{ij(h)} \tag{13}$$

其中，当处理Actor网络的输入时，$\boldsymbol{\alpha}^{ij}=\boldsymbol{\omega}_{a}^{ij}$；当处理Critic网络的输入时，$\boldsymbol{\alpha}^{ij}=\boldsymbol{\omega}_{\mathrm{c}}^{ij}$。

通过以上多头图注意力网络，模型能够有效建模大规模多智能体系统中复杂且异质的交互关系，既提升了表示能力，又避免了传统平均场方法在建模上可能出现的精度损失。

为了将关系权重转化为神经网络的输入特征，FA通过将其他卫星智能体的联合行为简化为一个加权平均量，从而将高维、复杂的交互问题转化为更易处理的线性维度问题。每颗卫星将其他所有卫星的行为视为一个平均场，而非单独建模。例如，在更新Q值时，仅考虑自身动作和状态与平均场的交互，而非所有卫星的联合动作：

$$Q_i(\boldsymbol{S},\boldsymbol{A})\sim\bar{Q}_i(s_i,\bar{s}_i,a_i,\bar{a}_i) \tag{14}$$

得到卫星智能体之间的交互权重后，特征聚合方法对系统的全局状态$\boldsymbol{S}=(s_1,s_2,\cdots,s_N)$以及所有卫星的联合动作$\boldsymbol{A}=(a_1,a_2,\cdots,a_N)$进行汇聚与近似，从而将原本高维网络参数，映射到仅与平均场特征相关的低维表示。这样的处理不仅兼顾了个体差异，也保留了系统整体的统计特性，使得价值网络和策略网络在网络架构设计上只需要接收固定维度的组合作为输入。网络参数规模从原本与卫星数目N指数相关，缩减为线性维度，极大降低了计算量，进而将MARL中的全局动作和全局状态近似为单颗卫星的状态/动作与其平均场的联合，大大降低了策略网络和值网络的输入维度。具体而言，在MGADFA框架下，MADDPG算法中的值网络可以近似为：

$$Q_i(\boldsymbol{S},\boldsymbol{A})\sim\tilde{Q}_i(s_i,\tilde{s}_i,a_i,\tilde{a}_i),\tilde{s}_i=\sum_{j\in N(i)}\boldsymbol{\omega}_{\mathrm{c}}^{ij}s_j,\tilde{a}_i=\sum_{j\in N(i)}\omega_{\mathrm{c}}^{ij}a_j \tag{15}$$

同时，策略网络可以近似为：

$$\pi_i(\boldsymbol{S})\sim\hat{\pi}_i(s_i,\hat{s}_i),\hat{s}_i=\sum_{j\in N(i)}\boldsymbol{\omega}_{\mathrm{a}}^{ij}s_j \tag{16}$$

在完成了这种输入维度的“轻量化”改造后，算法便能够按照“集中式训练、分布式执行”的范式展开协同学习。MADDPG在集中式训练的过程中，利用全局的状态信息$S=(s_1,s_2,\cdots,s_N)$与联合动作$A=(a_1,a_2,\cdots,a_N)$训练Critic网络，评估整体协作效果。在分布式执行过程中，每颗卫星仅根据自身局部观测s_i生成确定性动作$a_i=\pi_i(s_i)$，不需要依赖其他卫星实时状态。为了保证任务的唯一执行主体，在多星联合动作输出后引入资源优化策略，对候选卫星进行筛选。具体而言，多头注意力网络设计了基于多维度评分的选择机制，综合考量卫星的剩余资源、任务执行成本及任务收益等因素，从中选取综合得分最高的卫星，作为该任务的

唯一执行者。训练过程中，Critic网络通过梯度下降方法最小化贝尔曼误差进行更新，其损失函数为：

$$L=\sum_i\left(y-Q_i\left(s_i,\tilde{s}_i,a_i,\tilde{a}_i\right)\right)^2 \tag{17}$$

其中，$y=R+Q'(s_i',\tilde{s}_i',\dot{a}_i^{s'},\tilde{\dot{a}}_i^{s'})$，$\dot{a}_i^{s'}=\pi_i'\left(s_i',\hat{s}_i'\right)$。Actor网络根据确定性策略梯度，利用梯度上升的方法提升Critic评估的Q值，其损失函数为：

$$J=\sum_i Q_i\left(s_i,\tilde{s}_i,a_i{}^s,\tilde{a}_i{}^s\right) \tag{18}$$

其中，$a_i{}^s=\pi_i(s_i,\hat{s}_i)$。$Q_i'$和$\pi_i'$分别为$Q_i$和$\pi_i$的目标网络。

本文构建了适用于巨型低轨遥感星座任务协同规划的多智能体训练框架，MGADFA训练流程如算法1所示。首先，构建轨道星座模拟环境，生成任务集合，并计算每项任务的可执行时间窗口及其对应的可观测卫星集合；随后，对策略网络、值网络及其图注意力模块（包括查询网络与键网络）进行初始化，并同步初始化对应的目标网络。

在每轮训练开始前，重置任务状态与环境参数，并根据预设的衰减策略调整探索噪声ε。接着，各卫星依据特征聚合机制与$\varepsilon-$greedy策略生成动作并执行联合决策，同时将交互产生的状态转移信息存储至经验回放缓冲区。

算法1：MGADFA训练流程

输入：

卫星参数：轨道平面数N_planes，每平面卫星数N_sats，倾角inclination，轨道高度h，卫星智能体数量N，卫星环境env，任务集task_set

学习参数：学习率α，折扣因子γ，软更新率τ，探索噪声ε，批次大小batch_size，更新间隔update_interval，当前更新周期episode，最大更新周期MAX_EPISODES, 注意力权重$\boldsymbol{\omega}_{\mathrm{a}}^{ij}$，经验回放缓冲区$B$，策略网络$\pi$，价值网络$Q$，查询网络$Q_{\mathrm{a}}$，键网络$K_{\mathrm{a}}$以及它们的目标网络$\pi'$、$Q'$、$Q'_{\mathrm{a}}$、$K'_{\mathrm{a}}$

输出：训练后的多智能体模型

1：创建卫星环境env，生成任务集task_set并计算时间窗口

2：初始化策略网络π和价值网络Q，初始化查询网络Q_{a}和键网络K_{a}

3：创建π、Q、Q_{a}、K_{a}的目标网络π'、Q'、Q'_{a}、K'_{a}

4：初始化经验回放缓冲区B

5：**for** episode=1 to MAX_EPISODES **do**

6：　重置env和任务状态，衰减探索噪声ε

7：　**for** $i=1$ to N **do**

8：　　计算Actor网络对应注意力权重：$\boldsymbol{\omega}_{\mathrm{a}}^{ij}=\left(Q_{\mathrm{a}}(s_i)\right)^{\mathrm{T}}K_{\mathrm{a}}(s_j)$，$j$为除$i$外的其他节点

9：　　计算动作$a_i{}^*=\pi(s_i,\hat{s}_i),\tilde{s}_i=\sum_{j\in N}\boldsymbol{\omega}_{\mathrm{a}}^{ij}s_j$

10：　**end for**

11：　　资源优化策略处理联合动作$\boldsymbol{A}^*=(a_1{}^*,a_2{}^*,\cdots,a_N{}^*)$得到$\boldsymbol{A}=(a_1,a_2,\cdots,a_N)$

12：　**for** 每个任务 **in** task_set **do**

13：　　执行联合动作$\boldsymbol{A}=(a_1,a_2,\cdots,a_N)$，得到下一状态$S'$以及奖励$R$

14：　　存储(S,A,R,S')到缓冲区B中

15：　**end for**

16：　**if** episode≥最小采样回合且episode%采样间隔==0

17：　　从缓冲区B中采样

18：　　通过算法2更新网络参数

19：　end if

20：end for

21：return训练后的模型π、Q_a、K_a

当累计的经验满足一定规模后，从经验池中按批次采样样本，定期执行网络更新操作。MGADFA网络参数更新流程如算法2所示，主要采用梯度下降和梯度上升的方法，对策略网络、

评估网络以及注意力模块中的参数进行联合优化，以提升整体任务协同性能。

算法2：MGADFA网络参数更新流程

输入：卫星智能体数量N，Actor网络π，Critic网络Q，查询网络Q_a、Q_c，键网络K_a、K_c，目标网络π'、Q'、Q'_a、Q'_c,K'_a、K'_c，经验(s,a,r,s')以及软更新率τ

输出：更新后的网络π、Q、Q_a、Q_c、K_a、K_c

1：计算$y=r+Q'(s',\tilde{s}',\dot{a}_{s'},\tilde{\dot{a}}_{s'})$

其中$\dot{a}_{s'}=\pi'(s',\hat{s}')$，$\hat{s}'$为$s'$通过$Q'_a$和$K'_a$计算得到的带权平均场，$\tilde{s}'$为$s'$通过$Q'_c$和$K'_c$计算得到的带权平均场，$\tilde{\dot{a}}_{s'}$为$\dot{a}_{s'}$通过$Q'_c$和$K'_c$计算得到的带权平均场

2：计算$L=(y-Q(s,\tilde{s},a,\tilde{a}))^2$

其中，$\tilde{s}$为s通过Q_c和K_c计算得到的带权平均场，$\tilde{a}$为a通过Q_c和K_c计算得到的带权平均场

3：计算$J=Q(s,\tilde{s},a_s,\tilde{a}_s)$

其中，$a_S=\pi(s,\hat{s})$，$\hat{s}$为s通过Q_a和K_a计算得到的带权平均场，$\tilde{a}_s$为a_S通过Q_c和K_c计算得到的带权平均场

4：通过梯度下降方法最小化L，更新Q,Q_c,K_c参数

5：通过梯度上升方法最大化J，更新π,Q_a,K_a参数

6：通过软更新（参数为τ），将网络π、Q、Q_a、Q_c、K_a、K_c参数复制到π'、Q'、Q'_a、Q'_c,K'_a、K'_c中

7：返回网络π、Q、Q_a、Q_c、K_a、K_c

3 实验结果与分析

为了全面评估MGADFA算法在巨型遥感星座任务规划中的整体表现，本文从注意力头数量、算法收敛性、任务接受数量以及资源负载均衡4个关键方面展开对比分析。本文首先通过分析不同注意力头数量对模型性能的影响，并结合理论分析，确定注意力头数量；然后将MGADFA与典型启发式算法和强化学习方法进行对比，评估其在复杂任务规划问题中的效果。在巨型星座的协同调度场景下，如果仅以任务收益最大化为目标，往往会导致部分卫星被频繁选用、资源消耗过快，从而形成资源分布不均，影响星座的长期运行能力。因此，本文进一步引入资源负载均衡度作为重要评价指标，通过刻画不同卫星的时间与存储资源的分布，分析各算法在提升任务接受数量的同时，是否能够实现合理的任务分配，避免资源使用过度集中。

3.1 实验设置

仿真实验采用规模为20×30（轨道数×每轨道卫星数）的遥感星座。遥感任务的地理位置在指定区域内随机生成，其中纬度范围为-38°～38°，经度范围为−180°～180°。任务数量为1 500，每项任务的优先级在1～10随机设定，观测持续时间为0～300 s，所占用的存储容量为10～30 GB。卫星的最大存储容量设置在50～85 GB范围内，其可用于任务执行的时间上限为600～1 200 s。在该仿真设置下，选取总奖励值、任务优先级完成情况以及负载均衡度作为算法性能的主要评估指标，综合衡量系统的资源利用效率、任务完成质量与协同调度能力。本文在星座规模600颗遥感卫星的场景下进行训练。其中，与训练相关的强化学习参数见表1，仿真实验的硬件与软件平台配置见表2。

表1 与训练相关的强化学习参数

参数	数值	参数	数值
学习率lr	0.000 1	轮次episode	1 000
折扣因子gama	0.95	采样大小batch_size	512
ε初始值epsilon_start	1.0	软更新率soft_rate	0.01
ε最小值epsilon_end	0.01	采样间隔interval	10

表2　仿真实验的硬件与软件平台配置

组件	配置详情
CPU	Intel Core i7-12700KF
GPU	NVIDIA GeForce RTX 4070
内存	32GB DDR4 3200MHz
开发框架	PyTorch 1.12.1 + CUDA 11.6
编程语言	Python 3.10

3.2　注意力头数量对比

多头图注意力网络中的注意力头数量是影响模型表达能力与计算复杂度的重要超参数。理论上，增加注意力头数量有助于模型在多个特征子空间中并行建模星间协作关系，从而提升关系表示的丰富性。然而，注意力头数量的增加也会带来网络参数规模和计算开销的同步增长，尤其在巨型遥感星座场景下，过多的注意力头可能导致梯度估计噪声增大，从而影响训练效率和稳定性。

当注意力头数量为1或2时，模型在较长的训练阶段内难以学习到有效策略，奖励值长期处于较低水平，整体收敛过程明显滞后。这表明在注意力头数量受限的情况下，不同关注维度只能被压缩并混合到有限的表示空间中，模型难以同时刻画卫星资源状态与任务匹配关系等多种关键协作因素，从而制约了特征表达能力的发挥。当注意力头数量增加至4个时，模型开始能够在多个并行的表示子空间中对不同关注模式进行区分建模，训练过程中奖励的提升速度显著加快，并能在较少轮次内完成收敛，这说明多头机制在该阶段已有效增强了信息聚合和特征表达能力。当注意力头数量进一步增加至8个时，算法表现出最快的收敛速度，并能够在较短训练轮次内达到稳定且较高的奖励水平。这表明在该配置下，模型具备较为充足的表示能力，可以并行刻画多种星间交互关系，同时仍保持良好的训练稳定性。进一步将注意力头数量增加至16个时，模型的最终奖励水平与8个头配置基本接近，但由于模型参数规模和计算复杂度的增加，参数更新更新效率变低，收敛速度略有下降。综合考虑收敛速度、最终性能以及计算效率等因素，本文选择8个注意力头作为MGADFA的默认配置。

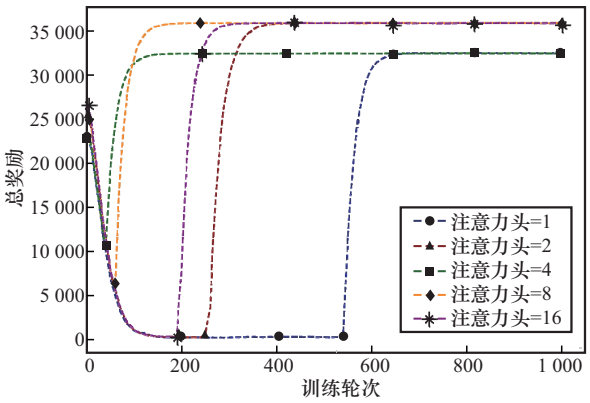

图2　不同注意力头配置下的训练总奖励对比

3.3　算法收敛性

本文将所提出的MGADFA方法与GA、SA、DQN以及MAPPO算法在相同数量的任务场景下进行了对比分析。为保证实验对比的公平性，各对比算法的参数设置均遵循统一的调参原则。对于启发式算法GA和SA，其参数包括种群规模、交叉与变异概率、初始温度及降温系数等，分别参考文献[4]和文献[8]中的相关配置，并在合理范围内进行了有限调优，以确保算法在当前实验场景下能够稳定运行并发挥其代表性性能。其中，GA种群规模为200，交叉概率为0.85，变异概率为0.15。SA初始温度为100°，降温系数为0.95。对于基于强化学习的算法DQN和MAPPO，其网络结构与训练参数分别参考文献[22]和文献[23]，并结合本实验环境对学习率、折扣因子和探索策略等少量关键参数进行了独立调节，以保证算法的收敛性。其中，DQN学习率为0.001，折扣因子为0.99，探索策略为ε–greedy策略。MAPPO学习率为0.000 3，折扣因子为0.99，探索策略为ε–greedy策略。上述算法迭代周期均为1 000。各算法均在相同的任务规模和计算资源条

件下进行训练与评估。

不同算法的训练总奖励对比如图3所示。由图3可知，传统启发式算法SA与GA在整个训练过程中奖励值基本保持稳定，受限于固定规则，其整体性能提升空间较为有限。基于强化学习的方法DQN和MAPPO随训练轮次的增加呈现逐步学习和性能提升的趋势，但其收敛速度和稳定性存在一定差异。其中，DQN收敛过程相对平缓，而MAPPO虽在中前期能够快速提升奖励水平，但训练后期仍存在一定幅度的波动。相比之下，MGADFA在初期探索阶段后表现出显著的性能提升能力，其累计奖励在较少训练轮次内迅速上升并达到稳定收敛状态，最终奖励水平明显高于其余对比算法，这表明该方法在巨型遥感星座场景下，具备较强的优化能力和良好的收敛速度。

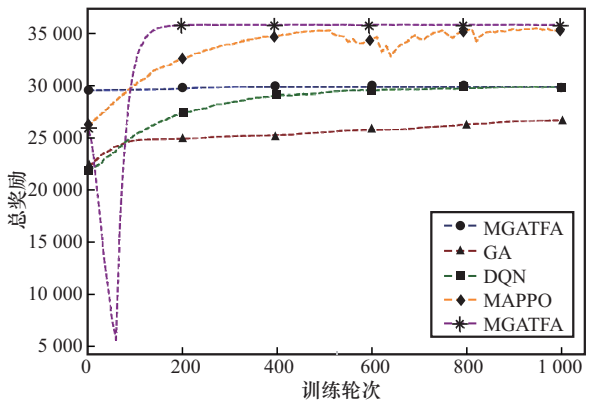

图3　不同算法的训练总奖励对比

3.4　负载均衡度对比

本文引入变异系数（coefficient of variation，CV）[24]作为负载均衡度的指标。变异系数是标准差与均值的比值，用于衡量数据相对离散程度，数值越小，表示越稳定、越均衡。

不同算法的负载均衡对比如图4所示。图4对比展示了SA、GA、DQN、MAPPO以及本文提出的MGADFA这5种算法在存储负载均衡、时间负载均衡以及任务接受数量3个维度下的性能表现。由图4可知，在存储与时间负载均衡方面，MGADFA整体优于传统启发式算法SA与GA，并与强化学习方法DQN和MAPPO的表现基本相当。其中，MGADFA在存储负载均衡度上相较SA和GA分别降低约9.2%和5.7%；在时间负载均衡度上则分别降低约6.6%和4.1%，表明其在多维资源分配上能够实现更为合理的均衡控制。在任务接受数量方面，MGADFA展现出更加明显的优势，其累计接受任务数量达到861个，高于SA、GA、DQN和MAPPO，分别提升了约5.1%、5.7%、5.0%和1.1%。综合来看，MGADFA在不牺牲负载均衡性的前提下有效提升了任务调度效率，体现出更优的全局规划能力和更强的策略泛化能力，为复杂动态任务场景下的高效调度提供了有力支撑。这些结果表明，MGADFA在提升任务接受能力的同时，能够有效改善资源分配的均衡性，从而展现出更优的全局规划能力和更强的泛化能力，为复杂任务场景下的高效调度提供了有力支撑。

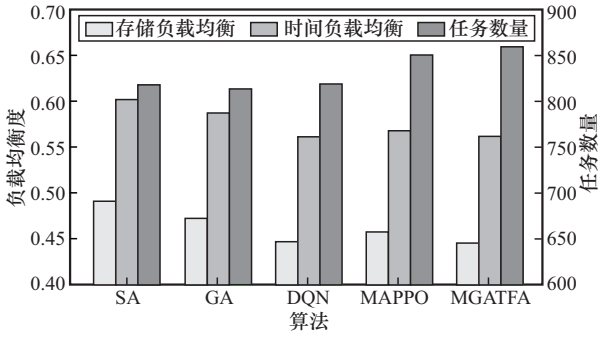

图4　不同算法的负载均衡对比

3.5　仿真平台验证

本文在自研的巨型遥感星座仿真平台上对所提出的任务规划算法进行了可视化演示与验证。首先，依据设定的星座构型参数，构建出由多轨道面、多卫星组成的大规模遥感卫星星座，真实模拟了高密度轨道部署下的空间拓扑结构；在此基础上，利用前述训练得到的MGADFA模型对任务集进行调度规划；系统自动根据卫星资源、任务约束及可见性窗口等因素，完成任务的分

配；任务分配完成后，遥感任务规划结果见表5，涵盖任务ID、目标位置、时间窗口、执行时间以及执行卫星相关信息。其中，目标位置以经纬度形式给出观测点的精确坐标；时间窗口用于展示任务规划阶段设定的起止时间范围；执行时间则记录实际完成观测的具体时刻。

表5　遥感任务规划结果

任务ID	目标位置	时间窗口	执行时间	执行卫星
senseA-TASK-1001	83.78°,−30.23°	14:51～14:58	14:55	419-B
senseA-TASK-1002	46.90°,13.85°	14:49～14:56	14:54	158-C
senseA-TASK-1003	−113.38°,34.56°	14:50～14:57	14:55	488-D
senseA-TASK-1004	−77.66°,12.34°	14:52～14:59	14:56	348-A
senseA-TASK-1005	106.66°,23.45°	14:51～14:57	14:55	419-B

4　结束语

本文提出的MGADFA算法，通过多头图注意力驱动的特征聚合机制，有效化解了巨型遥感星座在智能规划任务中面临的“维度爆炸”难题。仿真实验与平台验证结果一致表明，该方法在保留星间交互关系的同时，显著提升了联合决策的收敛效率，并在任务吞吐量与资源负载均衡性上展现出卓越的性能。这为未来超大规模星座的自主协同与智能化运行提供了一种兼顾效率与精度的可行方案。

参考文献：

[1] 胡笑旋, 王执龙, 夏维, 等. 遥感卫星任务规划技术: 现状与展望[J]. 指挥与控制学报, 2023, 9(5): 495-507.
Hu X X, Wang Z L, Xia W, et al. Mission planning technology for remote sensing satellite: status and prospect[J]. Journal of Command and Control, 2023, 9(5): 495-507.

[2] Peng G S, Song G P, Xing L N, et al. An exact algorithm for agile earth observation satellite scheduling with time-dependent profits[J]. Computers & Operations Research, 2020, 120: 104946.

[3] Zeng G M, Zhan Y F, Xie H R, et al. Resource allocation for networked telemetry system of mega LEO satellite constellations[J]. IEEE Transactions on Communications, 2022, 70(12): 8215-8228.

[4] Wei L N, Chen M, Xing L N, et al. Knowledge-transfer based genetic programming algorithm for multi-objective dynamic agile earth observation satellite scheduling problem[J]. Swarm and Evolutionary Computation, 2024, 85: 101460.

[5] 张泽华, 张加友, 张嘉凯, 等. 基于遗传禁忌算法的多星协同任务规划方法[J]. 无线电工程, 2022, 52(7): 1127-1135.
Zhang Z H, Zhang J Y, Zhang J K, et al. Multi-satellite cooperative mission planning method based on genetic tabu algorithm [J]. Radio Engineering, 2022, 52(7): 1127-1135.

[6] Chen Y, Zhang D Y, Zhou M Q, et al. Multi-satellite observation scheduling algorithm based on hybrid genetic particle swarm optimization[M]//Advances in Information Technology and Industry Applications. Berlin, HeidelbergSpringer2012: 441-448.

[7] Gu Y, Han C, Chen Y H, et al. Large region targets observation scheduling by multiple satellites using resampling particle swarm optimization[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(2): 1800-1815.

[8] Han C, Gu Y, Wu G H, et al. Simulated annealing-based heuristic for multiple agile satellites scheduling under cloud coverage uncertainty[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(5): 2863-2874.

[9] Khojah H A, Mosa M A. Multi-objective optimization for multi-satellite scheduling task: Multi-objective optimization for multi-satellite scheduling task[J]. Journal of Soft Computing Exploration, 2022, 3(1): 19-30.

[10] Liu Y C, Chen Q F, Li C Y, et al. Mission planning for Earth observation satellite with competitive learning strategy[J]. Aerospace Science and Technology, 2021, 118: 107047.

[11] Chen X Y, Tian T, Dai G M, et al. Deep reinforcement learning-based resource allocation method for multi-satellite scheduling [J]. Computers & Operations Research, 2025, 181: 107088.

[12] Ou J W, Xing L N, Yao F, et al. Deep reinforcement learning method for satellite range scheduling problem[J]. Swarm and Evolutionary Computation, 2023, 77: 101233.

[13] 李英玉, 史好迎, 赵通. 面向即时响应的卫星在轨分布式协商智能任务规划[J]. 空间科学学报, 2024, 44(1): 159-168.
Li Y Y, Shi H Y, Zhao T. On-orbit distributed negotiation intelligent mission planning for instant response[J]. Chinese Journal of Space Science, 2024, 44(1): 159-168.

[14] Herrmann A, Stephenson M A, Schaub H. Single-agent reinforcement learning for scalable earth-observing satellite constellation operations[J]. Journal of Spacecraft and Rockets, 2024, 61(1): 114-132.

[15] Zhang G H, Li X H, Hu G X, et al. MARL-based multi-satellite intelligent task planning method[J]. IEEE Access, 2023, 11: 135517-135528.

[16] Lowe R, Wu Y, Tamar A, et al. Multi-agent actor-critic for

mixed cooperative-competitive environments[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6382-6393.

[17] Liu S, Chen Y W, Xing L N, et al. Time-dependent autonomous task planning of agile imaging satellites[J]. Journal of Intelligent & Fuzzy Systems, 2016, 31(3): 1365-1375.

[18] Liu Z, Gao L, Chai Y. Research on task scheduling mechanism of distributed satellite system[J]. Journal of Computational Information Systems, 2014, 10(24): 10703-10713.

[19] Nowé A, Vrancx P, De Hauwere Y M. Game theory and multi-agent reinforcement learning[M]//Wiering M, van Otterlo M. Reinforcement Learning: State-of-the-Art. Berlin, Heidelberg: Springer, 2012: : 441-470.

[20] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[C]//Proceedings of the International Conference on Learning Representations (ICLR). 2018.

[21] Zhang S, Tong H H, Xu J J, et al. Graph convolutional networks: a comprehensive review[J]. Computational Social Networks, 2019, 6: 11.

[22] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518 (7540): 529-533.

[23] Yu C, Velu A, Vinitsky E, et al. The surprising effectiveness of PPO in cooperative multi-agent games[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: ACM, 2022: 24611-24624.

[24] Abdi H. Coefficient of variation[J]. Encyclopedia of Research Design, 2010, 1(5): 169-171.

[作者简介]

佘雨璇（2002- ），男，华中科技大学网络空间安全学院第六代移动通信研究中心硕士生，主要研究方向为卫星互联网。

周康燕（1981- ），男，西安电子科技大学电子工程学院在读博士，主要研究方向为卫星互联网。

代奥（2001- ），男，华中科技大学网络空间安全学院第六代移动通信研究中心硕士生，主要研究方向为卫星互联网。

魏子翔（2002- ），男，华中科技大学网络空间安全学院第六代移动通信研究中心硕士，主要研究方向为卫星互联网。

周家喜（1980- ），男，华中科技大学网络空间安全学院第六代移动通信研究中心教授，主要研究方向为卫星网络智能任务规划、智能资源调度与星载智能体安全等。

肖丽霞（1987- ），女，华中科技大学网络空间安全学院第六代移动通信研究中心研究员，主要研究方向为卫星网络波形设计与智能资源规划。