

面向即时响应的卫星在轨分布式 协商智能任务规划*

李英玉^{1,2} 史好迎^{1,2} 赵通³

1(中国科学院国家空间科学中心 北京 100190)

2(中国科学院大学 北京 100049)

3(北京大学计算机学院 北京 100871)

摘要 低轨遥感星座任务规划是一个复杂的多目标优化问题, 目前基于深度强化学习的卫星任务规划研究存在试验数据星座规模小、优化目标单一、任务重复安排或模型适应性差等问题. 针对上述问题, 提出 CON_DQN(Contract network and Deep Q Network) 算法, 采用主从星在轨分布式协商机制, 从星基于规划决策, 主星基于深度强化学习算法决策, 从任务优先级、资源代价和负载均衡等方面进行多目标优化, 实现面向即时响应的卫星在轨分布式协商智能任务规划. 针对用户需求高频动态到达重点观测区域的场景, 进行百星级星座不同规模任务集的仿真实验, 结果表明本文所提算法的响应速度较快且能达到较高的任务收益.

关键词 在轨任务规划, 即时响应, 分布式协商, 深度强化学习, 多目标优化

中图分类号 V4

On-orbit Distributed Negotiation Intelligent Mission Planning for Instant Response

LI Yingyu^{1,2} SHI Haoying^{1,2} ZHAO Tong³

1(National Space Science Center, Chinese Academy of Sciences, Beijing 100190)

2(University of Chinese Academy of Sciences, Beijing 100049)

3(School of Computer Science, Peking University, Beijing 100871)

Abstract The mission planning of LEO remote sensing constellation is a complex multi-objective optimization problem. At present, there are some problems in satellite mission planning research based on deep reinforcement learning, such as small scale of test data constellation, single optimization objective, repeated task arrangement and poor model adaptability. To solve the above problems, the CON_DQN (Contract network and Deep Q Network, CON_DQN) algorithm is proposed in this paper, which adopts the master-slave on-orbit distributed negotiation mechanism, the slave satellite makes decisions based on

* 中国科学院重点部署项目资助 (ZDRW-KT-2016-02)

2022-12-26 收到原稿, 2023-02-23 收到修定稿

the planning, and the master satellite makes multi-objective optimization decisions from the aspects of priority, resource cost and load balancing based on the deep reinforcement learning algorithm, and processes on-orbit distributed negotiation intelligent mission planning for instant response. Aiming at the scene where the user demand reaches the key observation area dynamically at high frequency, the simulation experiment of different scale task sets of 100-star constellation is carried out. The results show that the proposed algorithm has a fast response speed and can achieve higher task benefits.

Key words On orbit mission planning, Instant response, Distributed negotiation, Deep reinforcement learning, Multiple objective optimization

0 引言

随着卫星规模的扩大,低轨遥感星座^[1,2]在时域和空域上能够提供更强的全天候无缝覆盖能力,但也增加了卫星管控的复杂度;同时,在用户及用户需求方面,大规模星座将为全球大众用户、应急减灾专业领域用户及军方用户等多种类多层次的用户提供服务^[3],其用户需求呈现动态、并发、即时等特性.传统地面集中式卫星管控的方式在资源状态收集方面存在很高的通信代价且有更新延迟的现象^[4],并且从任务到载荷指令的最优化求解也面临规模爆炸难题,该方式不适应星座规模 and 用户需求的日益增长.近年来,基于遗传算法、模拟退火、禁忌搜索等启发式算法求解多星多任务规划的研究较多,但Wang^[5]分析该方式面临“建模难”的问题,并且Huang^[6]发现该方式在突发应急任务的即时处理问题上存在不足.无论是传统地面集中式卫星管控的方式还是启发式算法均不适合处理大量动态到达且需要即时响应的用户需求.

随着天上计算、星间通信能力增强,将卫星作为智能体,在轨动态接收用户需求,进行自主决策的在轨即时规划算法成为发展趋势.强化学习是用于处理序贯决策问题较好的方法^[7],目前,基于强化学习的卫星任务规划主要分为单星多任务和多星多任务两类.在基于强化学习的单星多任务研究中,Ma等^[8]针对密集观测场景下的敏捷成像卫星任务规划问题综合考虑多个约束条件,采用行为者评论家算法训练模型,具有较高的求解效率.Zhou等^[9]在卫星网络资源调度机制研究中,通过强化学习智能调度模块进行全局决策规划,提高资源利用率.Peng等^[10]针对密集观测场景,将注意力机制与神经网络融合,并在训练阶段采用监督学习,降低平均响应时间.在基于强化

学习的多星多任务研究中,Wang^[11]在2~4颗卫星规模下以优先级为优化目标,研究基于MADDPG分布式调度策略即时决策,但对于大规模星座训练时间较长,且存在任务重复安排的问题.Li^[12]在5~15颗卫星规模下以优先级为优化目标,通过将自注意力决策网络引入MADDPG结构缩短训练时间,并引入重复观测惩罚因子减少了任务重复安排,但仍存在资源浪费的情况.Liu等^[13]在9颗卫星规模下以优先级为优化目标,针对多星任务分配问题引入基于利益的竞争策略,通过星间传输获取其他卫星的观测数据来确定最终规划,解决了任务重复安排的问题,但模型只针对特定目标点规模,不适用于不同规模的任务集.

低轨遥感星座任务规划是一个复杂的多目标优化问题,以上基于深度强化学习的多星任务规划研究存在星座规模小、优化目标单一、任务重复安排或模型适应性差的问题.针对以上问题,本文以低轨遥感星座为研究对象,通过分析在轨任务规划问题,从任务优先级、资源代价和负载均衡等方面考虑多目标优化,采用主从星分布式协商机制^[13,14],通过星间协商避免任务重复安排,基于深度强化学习构建任务集规模无关的模型,设计在轨任务规划算法,并建立百星级星座仿真环境进行仿真实验和分析.

1 问题分析

低轨遥感星座在轨任务规划问题可以简要描述为:遥感星座中的每颗卫星具备一定的计算能力,卫星之间存在通信链路,用户产生的观测需求可通过地面站或持有的终端直接上传至门户卫星,主星接收到用户需求后根据任务特点及各卫星的时间窗、存储容量以及执行次数等约束条件进行决策,生成任务规划方案,将任务分发给执行卫星.高频动态产生的用户

需求冲突率很高;一些应急任务,例如洪涝灾害、地震救援以及森林火灾等,即时性要求很高;卫星在轨任务规划需要考虑多方面因素,是一个多目标优化问题.

1.1 用户需求冲突

低轨遥感星座直接面向广大用户服务,用户需求可以直接上传,呈现动态、并发、即时的特性,用户需求高频产生出现了极高的冲突率,这种冲突体现在两个方面:由于星座规模的扩大,某个用户需求可能存在多个卫星可以满足,这时就产生了选择哪颗卫星来完成的问题;另外,不同用户需求会产生“互斥”,即多个用户需求对卫星资源的使用无法兼容,不能同时满足.对第一个问题,本文采用主从星分布式协商来解决此问题;对第二个问题,本文采用优先级来表示不同用户需求的重要程度,作为选择的主要依据,同时兼顾资源代价和负载均衡.

1.2 即时响应

随着低轨遥感星座规模的扩大,对在轨任务规划即时性要求更高.一些特殊应急任务,在短时间内可能需要多次快速响应.这些高频动态到达的用户需求之间存在一些潜在的规律,基于规则的决策算法并不能捕获所有特征.深度强化学习算法序贯决策能力强、求解速度快,且能够在信息不完备的情况下考虑到未来信息,并进行智能决策,较适合处理面向即时响应的智能决策场景.

1.3 多目标优化

大型星座卫星数量和用户数量大大增加,任务规划需要从用户角度和系统角度等多个方面整体考虑.

用户角度主要对应不同种类的用户需求,如军用、商用及民用,可以给用户以不同优先级,计算不同的收益作为决策依据;系统角度通过资源代价和负载指标来衡量.以优先级为主要目标,同时兼顾资源代价和负载均衡.优先级表示了任务的重要程度,优先级越高表示任务越重要,用优先级来计算任务收益.执行每个任务的资源代价是不一样的,主要包括任务所需要的执行时长和存储容量.卫星的寿命与卫星在一定周期内执行任务的次数有关,通过单星任务执行次数和剩余存储来衡量单颗卫星的负载度^[15].

2 模型构建

2.1 模型框架

通过上述问题分析,采用主从星分布式协商机制设计遥感星座任务规划模型,如图1所示.假设观测目标分布在重点观测区域内,需要依据重点观测区域将卫星分为多个集群,每个集群都有一个主星和零到多颗从星,其中每个集群的主星固定,从星随时空变化.由地面站或终端将观测任务发送至上空卫星,再经互联的卫星网络传至集群主星,并由主星将任务发送至集群内各从星,从星基于规则自主决策并将结果反馈给主星,主星做智能评估选出执行从星或放弃任务.

在轨任务规划是一个复杂的计算问题,单颗卫星的计算能力有限,为了降低主星的负载,将计算分布给每个从星,由各自决策是否参与任务,克服了主星的计算瓶颈.对于一个任务可能存在多颗卫星均能完

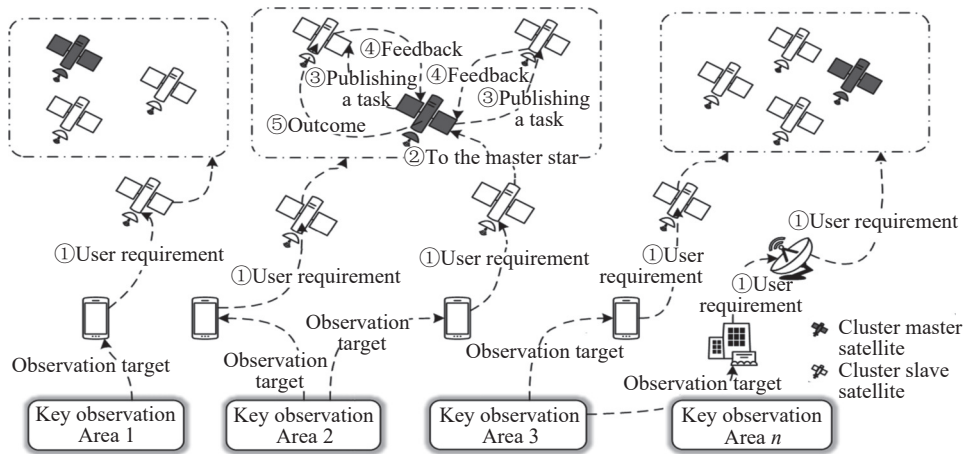


图1 遥感星座任务规划模型

Fig. 1 Remote sensing constellation mission planning model

成的情况, 为避免资源的浪费, 需要卫星间进行协商, 通过主星评标决定哪颗卫星执行任务。

通过主从星之间的交互协商, 能够解决用户需求冲突中任务重复安排的问题。另外, 无论是主星根据从星心跳获取从星的资源状态还是地面站统一进行任务规划, 都存在卫星资源状态信息不是最新的情况, 一旦已规划的卫星状态发生变化, 则需要重新规划或者放弃任务, 而卫星自行决策可以保证资源状态最新。

低轨遥感卫星在轨任务规划受多个因素的影响, 为了聚焦问题, 进行以下场景假设。

假设 1 观测目标均匀分布在几个重点观测区域内。

假设 2 卫星能够进行星间通信。

假设 3 卫星执行任务过程中不会被中断。

假设 4 卫星对用户观测任务的决策结果不会影响后续任务的生成。

假设 5 同一时间段内, 卫星只能执行一项任务。

2.2 符号定义

为了数学化描述卫星在轨任务规划, 根据问题描述, 给出了如下的定义和公式符号:

定义 1 $T=\{t_1, t_2, \dots, t_i, \dots, t_n\}$ 表示随时间动态到达的任务, 通过 $t_i=\{t_{i1}, t_{i2}\}$ 表示编号为 i 的任务, $t_{i1}=\{p_{\text{task}}, c_{\text{task}}, d_{\text{task}}, o_{\text{task}}\}$ 表示任务 i 属性集合 1, $t_{i2}=\{i_{\text{task}}, m_{\text{task}}, l_{\text{task}}\}$ 表示任务 i 属性集合 2, p_{task} 表示任务优先级, c_{task} 表示任务所需的存储, d_{task} 表示任务工作时长, o_{task} 表示任务开始执行时刻, i_{task} 表示任务编号, m_{task} 表示任务所属集群主星编号, l_{task} 表示任务的观测经纬度。

定义 2 $W_y=\{w_1, w_2, \dots, w_i, \dots, w_x\}$ 表示集群 y 当前时刻所有卫星的集合, x 表示集群 y 规模, 通过 $w_i=\{c_i, g_i\}$ 表示卫星 i 的属性信息, 其中 c_i 表示卫星 i 剩余的存储容量, g_i 表示卫星 i 在规划周期内的执行次数。

定义 3 $F_i=\{F_{i1}, F_{i2}, \dots, F_{ij}, \dots, F_{iL}\}$ 表示卫星 i 的空闲时间窗, $F_{ij}=\{o_j, e_j\}$ 表示卫星 i 第 j 个空闲时间窗, 其中 o_j 表示空闲时间窗的开始时间, e_j 表示空闲时间窗的结束时间。

2.3 从星决策策略

从星在收到主星的任务信息后, 基于一定的规则

进行自主决策。本文考虑的约束条件主要包括时间窗约束、存储容量约束和开关机次数约束。

约束 1 卫星空闲时间窗 F_i 包含任务信息 t_i 中的任务执行时间, 需要满足 $e_j - o_j \geq d_{\text{task}}$ 和 $o_j \leq o_{\text{task}}$ 。

约束 2 观测任务拍摄图像后需要占用一定的存储空间 c_{task} , 每颗卫星都有存储容量限制, 需要满足 $c_i \geq c_{\text{task}}$ 。

约束 3 一定周期内的开关机次数对卫星寿命是有影响的^[15], 每颗卫星的寿命都有限, 需要满足约束 $g_i \leq g_{\text{max}}$ 。

2.4 主星基于深度强化学习多目标决策

主星收到从星是否满足完成观测需求条件的反馈后, 基于深度强化学习算法决策是否接受该任务以及选取哪颗从星执行任务。

强化学习中智能体根据环境状态 s 采取动作 a , 得到下一个环境状态 s' 和奖励 r , 并通过奖励 r 对决策模型进行改进, 系统最终目的是在整个任务周期内获取更高的长期累计奖励, 即任务总收益。由于本文动作空间为离散状态, 因此 DQN 较适合作为决策模型。DQN 是较为经典的深度强化学习算法, 其利用卷积神经网络作为函数逼近器来逼近 Q 函数 $Q(s, a)$, 在训练过程使用梯度下降方式更新神经网络权重参数来最小化损失函数, 其损失函数如下:

$$\text{loss}(\sigma) = E\{[r(s, a, s') + \gamma \max_{a'} Q(s', a'; \sigma') - Q(s, a; \sigma)]^2\}. \quad (1)$$

式中, $r(s, a, s')$ 为智能体通过采取动作 a 其状态从 s 转变 s' 所获得的奖励函数, 另外, DQN 通过设计经验回放策略避免智能体经验之间的相关性并降低了神经网络的过拟合现象^[16]。

本文将集群主星作为智能体, 设计状态空间 $s=[t_{i1}, w_j, w_k, \dots]$ 作为输入, 包含任务信息 t_{i1} 和满足约束的卫星信息 $\{w_j, w_k, \dots\}$; 动作空间 $a=\text{action}$ 作为输出, action 的值为执行卫星编号或 -1, 其中 -1 表示拒绝任务; 奖励函数 $r=f(s, a)$ 考虑多目标优化, 以优先级为主要目标, 同时兼顾资源代价和负载均衡。

奖励函数中优先级为

$$f_1(s, a) = \frac{p_{\text{task}}(t_i) x_i}{p_{\text{task_max}}}. \quad (2)$$

式中, $p_{\text{task_max}}$ 为任务的最大优先级; x_i 为决策变量,

若 $x_i=1$, 则表示卫星接受该任务; 若 $x_i=0$, 则表示卫星拒绝该任务.

奖励函数中资源代价为

$$f_2(s, a) = \alpha_1 \frac{d_{\text{task}}(t_i)}{d_{\text{task_max}}} + \alpha_2 \frac{c_{\text{task}}(t_i)}{c_{\text{task_max}}}. \quad (3)$$

式中, $d_{\text{task_max}}$ 为任务的最大工作时长, $c_{\text{task_max}}$ 为任务的最大存储容量; α_1 为任务执行时长的权重, α_2 为任务所需存储容量的权重, 每个任务的资源代价主要由任务的执行时长和任务执行所需的存储来决定. 各个资源的权重决定了不同资源在子目标资源代价中所占的比重, 根据重要程度设置 α_1 和 α_2 .

奖励函数中负载均衡度为

$$f_3(s, a) = \frac{[g_{\max}(w_i) - g(w_i)]}{g_{\max}(w_i)} \left[\frac{c(w_i)}{c_{\max}(w_i)} \right]. \quad (4)$$

式中, $g_{\max}(w_i)$ 为任务规划周期内卫星 w_i 最多的执行次数, $g(w_i)$ 为卫星 w_i 从开始到现在所执行任务的次数, $c(w_i)$ 为卫星 w_i 剩余的存储容量, $c_{\max}(w_i)$ 为卫星 w_i 的最大存储容量. 卫星的寿命与卫星在一定周期内执行任务的次数有关. 通过单星任务执行次数和剩余存储来衡量单颗卫星的负载度.

奖励函数如下:

$$f(s, a) = \lambda_1 f_1(s, a) - \lambda_2 f_2(s, a) + \lambda_3 f_3(s, a). \quad (5)$$

式中, λ_1 为子目标任务收益的权重, λ_2 为子目标资源代价的权重, λ_3 为子目标负载均衡的权重. 各个子目标的权重决定了不同子目标在多目标优化中所占的比重, 根据实际情况设置 λ_1 , λ_2 和 λ_3 .

根据上述分析与设计, 给出 DQN 决策流程如图 2 所示, 具体步骤如下.

步骤 1 首先初始化评价 Q 网络和目标 Q 网络.

步骤 2 在一次迭代周期中, 先初始化状态 s .

步骤 3 选择动作 a , 更新卫星资源状态, 获取下一任务, 并计算下一状态 s' ; 获取奖励 r , 将 (s, a, r, s') 存储进记忆容器中, 令 $s = s'$.

步骤 4 若经验数大于学习阈值则计算损失函数 $loss$ 并更新评价 Q 网络参数, 学习次数加 1, 否则不更新网络参数, 进入步骤 3.

步骤 5 当更新频率是学习频率的整数倍时更新目标 Q 网络参数, 否则进入步骤 3.

步骤 6 若任务结束则迭代次数加 1, 否则进入步骤 3.

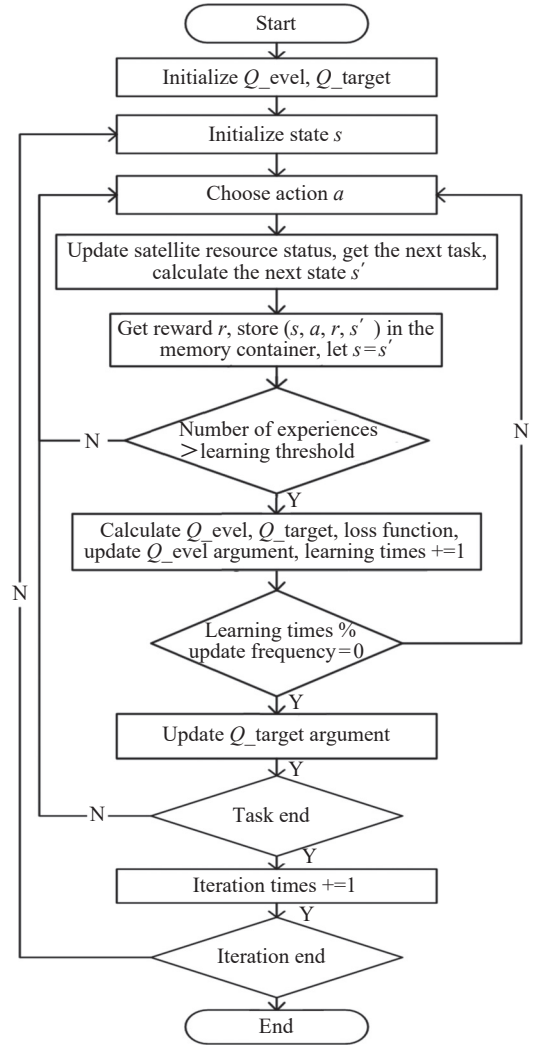


图 2 DQN 决策流程

Fig. 2 DQN decision flow chart

步骤 7 若迭代结束则进入步骤 8, 否则进入步骤 2.

步骤 8 模型训练结束.

2.5 算法设计

算法设计主要包括从星决策和主星决策两大部分, 其中从星决策伪代码如算法 1, 输入为任务信息, 输出为卫星资源信息. 主星决策伪代码如算法 2, 输入为环境状态信息, 输出为主星动作.

算法完整决策流程步骤如下.

步骤 1 主星接收到任务后, 向集群中的所有从星发布任务信息.

步骤 2 集群中的从星接收到任务信息后, 根据自身能力基于规则自主决策. 若满足 2.3 节所列的约束条件则参与, 并将卫星属性信息传递给主星; 若不

算法1: 从星决策伪代码**输入:** 任务信息 t_i **输出:** 卫星资源信息

1. if 卫星满足任务位置可见性 then:
2. if $e_j - o_j \geq d_{\text{task}}$
and $o_j \leq o_{\text{task}}$
and $c_i \geq c_{\text{task}}$
and $g_i \leq g_{\text{max}}$ then:
从星考虑时间窗约束、存储容量约束和执行次数约束
3. return $[c_i, g_i]$
4. return $[0, 0]$ # 从星不满足约束条件, 不参与投标

算法2: 主星决策伪代码**输入:** 环境状态信息 $State = [t_{i1}, w_j, w_k, \dots]$ **输出:** 动作 $Action$

1. 主星依据环境状态信息选取动作 $action$:
2. if 随机数 $< \text{EPSILON}$ then:
3. 策略选取评价网络值最大的动作
4. else:
5. 随机选取 $(0, n)$ 之间的数
6. 根据动作获取下一状态、奖励、是否结束等信息:
7. if $action = -1$ then:
8. $reward = 0$
9. 放弃任务
10. else:
11. $reward = f(action)$ # $f(action)$ 具体见公式(5)
12. 更新卫星资源状态信息, 包括存储容量、执行次数以及时间窗信息
13. 存储状态转换信息

满足则不参与。

步骤3 主星收到集群中从星的反馈后, 基于深度强化学习算法智能决策, 首先观测环境状态, 之后选择动作, 动作分为接受和拒绝两种, 最后获得奖励。若接受任务, 则获取相应奖励, 进入步骤4; 若拒绝任务, 则奖励为0, 规划结束。

步骤4 执行从星接收到主星发出的指定任务后, 修改自身资源状态, 将任务加入任务列表, 规划结束。

3 实验仿真与分析

3.1 仿真环境与数据

仿真实验采用 10×10 (轨道数 $\times$ 卫星数) 规模的

walker 星座, 计算集群卫星的星历数据, 采用卫星无需变轨直接执行任务的模式。实验选取3个重点观测区域, 分别为广东、山东和 Nebraska。任务请求频率为每秒1个, 服从泊松分布。设置任务优先级为1~10级, 所需持续工作时间为60~120 s, 任务占用存储容量为5~50 GB。单颗卫星最大存储容量为400 GB, 单颗卫星在仿真周期内最大执行次数为40次。主星决策网络模型根据实验场景设定如下: 输入层维度为20, 两个隐藏层维度都为50, 输出层维度为7。根据文献[15]的研究设计及各个资源的影响程度, 本文设计 $\alpha_1=0.7$, $\alpha_2=0.3$ 。根据文献[15]的研究设计及各个子目标的重要程度, 以任务优先级为主要目标, 同时兼顾资源代价和负载均衡, 设计 $\lambda_1=0.8$, $\lambda_2=0.1$, $\lambda_3=0.1$ 。

以卫星编号13, 14, 51, 52为例, 动态到达的任务示例如图3所示, 从图3可以看出, 随着时间变化任务不断产生并出现高冲突的现象, 即一颗卫星的多个任务资源冲突。

3.2 仿真实验及结果分析

基于上述仿真环境的设置, 采用任务总收益、任务优先级、资源代价以及负载均衡度作为算法性能的评估指标。

3.2.1 算法的收敛性和即时性

在星座规模为100, 任务规模为5000的情况下进行实验。图4给出了在进行30000轮训练 CON_DQN 算法的任务总收益变化曲线, 可以看出随着迭代次数的增加, 任务总收益逐渐趋于平稳, 达到了收敛的效果。

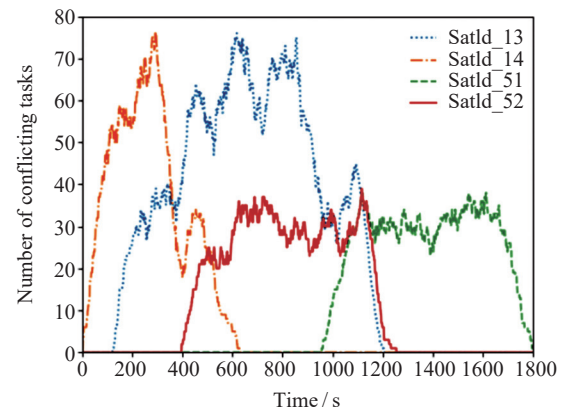


图3 动态到达的任务示例

Fig. 3 Example diagram of dynamically arriving tasks

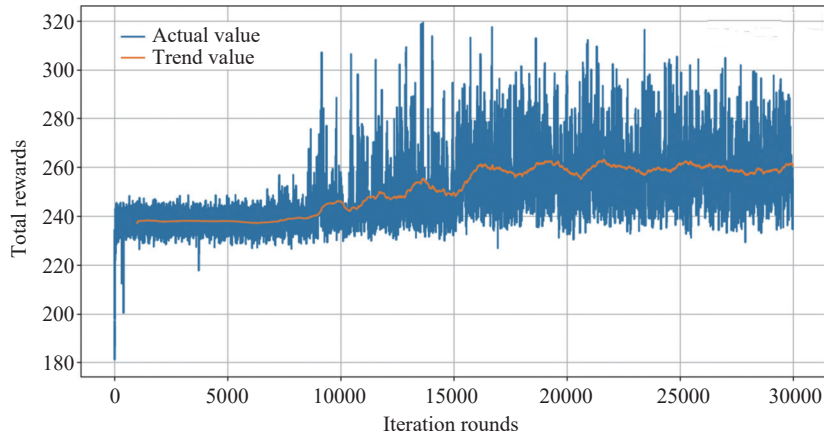


图4 CON_DQN 算法任务总收益

Fig. 4 Total task rewards of CON_DQN algorithm

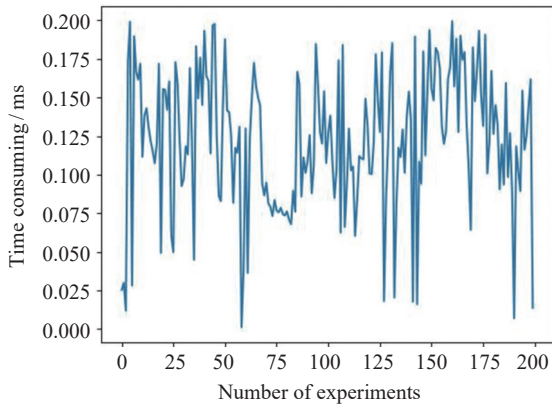


图5 单任务决策耗时

Fig. 5 Single task decision time graph

图5给出的是 CON_DQN 算法单任务决策耗时, 表示在 200 次实验中, 平均每个任务的决策耗时, 整体低于 0.200 ms, 平均耗时 0.125 ms, 说明其决策速度较快, 能够较快地处理百星级低轨遥感星座下大量动态到达的且需要即时响应的用户需求。

3.2.2 算法对比

选取的对比算法主要是基于不同指标的算法, 基于不同指标的算法主要有 FCFS 算法^[11]、INDICATOR 算法、LOADMIN 算法三种。FCFS 算法遵循先来先服务, 每次选取可执行卫星中序号最小的卫星执行任务; INDICATOR 算法每次选取可执行卫星中各项指标总收益最大的卫星执行任务; LOADMIN 算法每次选取可执行卫星中负载最小的卫星执行任务。

图6给出了在任务规模为 5000 情况下 CON_DQN

算法与其他算法各项指标的对比, 图6(a)为总收益对比, 可以看出在任务总收益方面, CON_DQN 算法是明显优于其他算法的, 与 FCFS、INDICATOR 和 LOADMIN 算法相比平均分别提高了 11.5%, 8.9% 和 9.6%。图6(b)为任务平均优先级对比, 可以看出在平均任务优先级方面, CON_DQN 算法明显高于其他算法。图6(c)为资源消耗对比, 可以看出在资源消耗方面, CON_DQN 算法整体是低于 INDICATOR 算法和 LOADMIN 算法, 高于 FCFS 算法。图6(d)为负载均衡对比, 可以看出在负载均衡方面, CON_DQN 算法与 INDICATOR 算法和 LOADMIN 算法相近, 高于 FCFS 算法。由以上分析可知, CON_DQN 算法能达到较高的任务总收益且资源消耗较少, 在负载均衡方面与其他算法相当。

图7给出的是各算法对各优先级任务的完成情况。从图7可以看出, 当优先级大于等于 4 的时候, CON_DQN 算法的任务完成数明显高于其他算法, 而在优先级等于 1 或者 2 的时候, CON_DQN 算法又明显低于其他算法, 说明 CON_DQN 算法更倾向于去完成高优先级的任务。

3.2.3 不同规模任务集下算法的效果

为进一步证明所提算法的有效性, 比较不同任务规模下各算法的完成情况 (见图8)。图8(a)为任务总收益, 随着任务规模的增大, 各算法的总收益都是增大的, CON_DQN 算法在不同任务规模条件下均高于其他算法, 且随着任务规模的增大, 差距更明显。图8(b)为算法单任务平均收益比较, CON_DQN 算

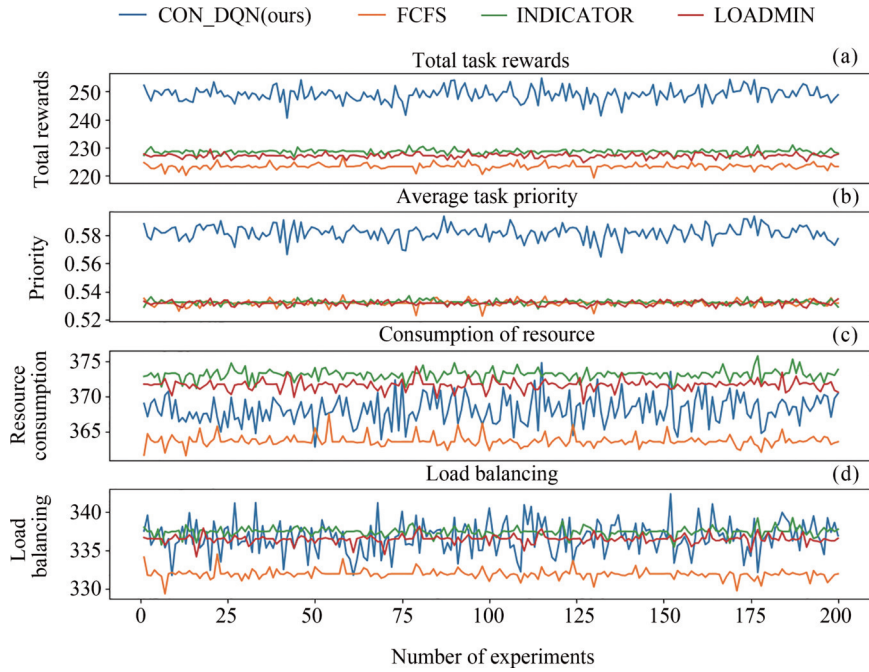


图6 CON_DQN 与其他算法各项指标对比

Fig. 6 Comparison of indexes between CON_DQN and other algorithms

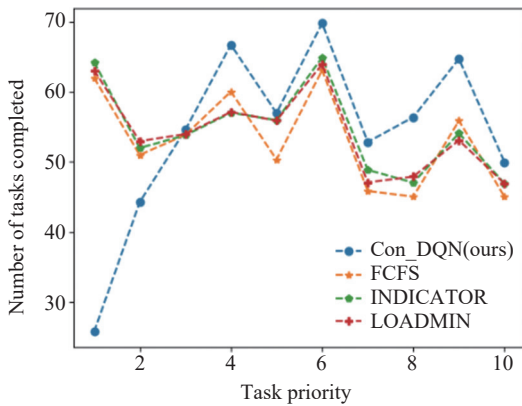


图7 各算法针对各优先级任务的完成情况

Fig. 7 Each algorithm aims at the completion of each priority task

法明显优于其他算法. 平均收益整体趋势缓慢降低, 这是因为一些卫星的负载较重导致负载均衡指标相对较低引起的. 图 8(c) 为算法单任务平均资源消耗比较, CON_DQN 算法明显低于其他算法, 平均资源消耗整体呈波动趋势, 这与任务的资源任务和分布情况相关. 图 8(d) 为算法平均负载比较, CON_DQN 算法高于其他算法, 平均负载整体趋势逐渐降低, 这是由于随着任务规模的增大, 一些卫星需要承担更多的任务, 导致负载均衡度的降低.

3.2.4 任务规划结果

图 9 给出的是 CON_DQN 算法任务规划结果, 可以看出每颗卫星在不同时间所执行的任务, 其中部分线段上方所代表的是任务编号. 由卫星任务规划结果可以看出, 受时空约束的限制, 卫星任务集中在部分卫星上. 所有的任务都没被重复安排, 说明该算法既能保证优先级又能保证任务不被重复安排.

4 结论

在百星级星座规模下, 针对高频动态到达的即时响应用户需求场景, 设计了一种 CON_DQN 算法, 采用主从星在轨分布式协商机制, 考虑任务优先级、资源代价以及负载均衡等多目标优化决策, 实现了面向即时响应的卫星在轨分布式协商智能任务规划. 实验结果表明该算法平均能够达到 0.125 ms 的响应速度, 与先来先服务、综合指标和最小负载三种决策策略相比总收益提高了 8.9%~11.5%, 该算法既能在保证任务优先级情况下兼顾资源代价和负载均衡, 又能保证任务不被重复安排, 且能灵活适应不同规模的任务集.

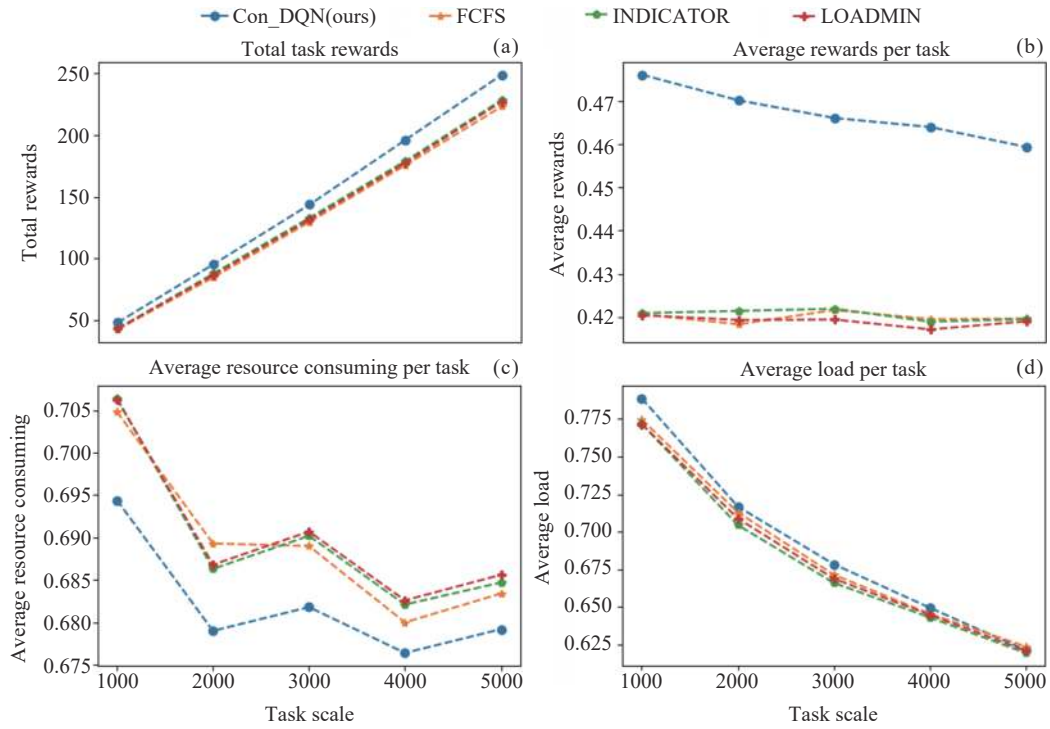


图 8 不同任务规模下各算法完成情况

Fig. 8 Completion of different algorithms in different task scales

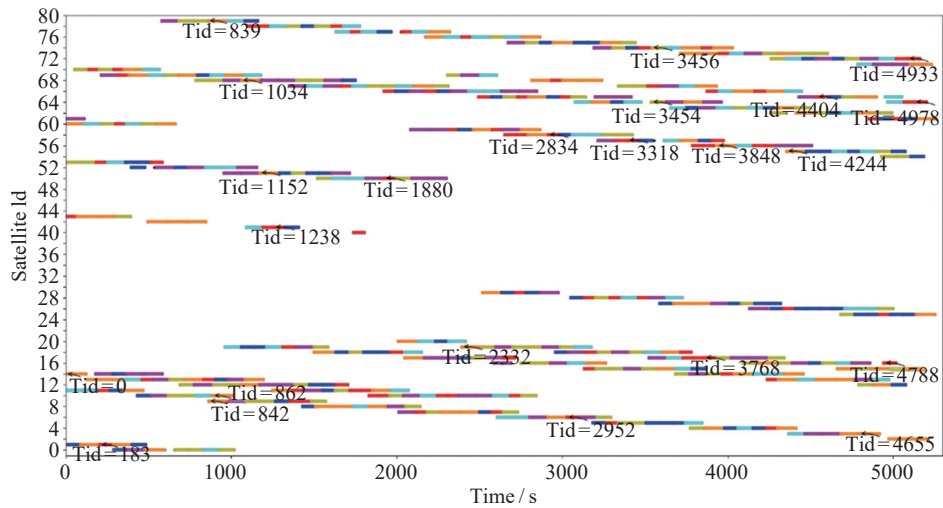


图 9 CON_DQN 算法任务规划结果

Fig. 9 CON_DQN algorithm task planning result

参考文献

- [1] WANG Jun, XIA Wei, HU Xiaoxuan, *et al.* Autonomous cooperative mission planning for remote sensing constellation based on multi-Agent[J]. *Journal of Command and Control*, 2021, 7(3): 287-294 (王俊, 夏维, 胡笑旋, 等. 基于多 Agent 的遥感星座自主协同任务规划 [J]. *指挥与控制学报*, 2021, 7(3): 287-294)
- [2] XIANG M S, DENG Q C, DUAN L S, *et al.* Dynamic monitoring and analysis of the earthquake worst-hit area based on remote sensing[J]. *Alexandria Engineering Journal*, 2022, 61(11): 8691-8702
- [3] ZHANG Luona, MA Zhongcheng, RAO Jianbing, *et al.* Development trend and market prospect of low-orbit satellite Internet[J]. *International Space*, 2020(11): 28-31 (章罗娜, 马忠成, 饶建兵, 等. 低轨卫星互联网发展趋势及市场展望)

- [J]. 国际太空, 2020(11): 28-31)
- [4] SINHA P K, DUTTA A. Multi-satellite task allocation algorithm for earth observation[C]//2016 IEEE Region 10 Conference. Singapore: IEEE, 2016: 403-408
- [5] WANG Chong, JING Ning, LI Jun, *et al.* An algorithm of cooperative multiple satellites mission planning based on multi-agent reinforcement learning[J]. *Journal of National University of Defense Technology*, 2011, **33**(1): 53-58 (王冲, 景宁, 李军, 等. 一种基于多 Agent 强化学习的多星协同任务规划算法 [J]. *国防科技大学学报*, 2011, **33**(1): 53-58)
- [6] HUANG H, SUN C Y, HU J X, *et al.* Optimization design of response satellite deployment for regional target emergency observation[C]//Proceedings of 2020 International Conference on Guidance on Advances in Guidance, Navigation and Control. Tianjin: Springer, 2022: 579-591
- [7] SUTTON R S, BARTO A G. Reinforcement Learning: An Introduction[M]. Cambridge: MIT Press, 1998
- [8] MA Yifan, ZHAO Fanyu, WANG Xin, *et al.* Agile imaging satellite task planning method for intensive observation[J]. *Journal of Zhejiang University (Engineering Science)*, 2021, **55**(6): 1215-1224 (马一凡, 赵凡宇, 王鑫, 等. 密集观测场景下的敏捷成像卫星任务规划方法 [J]. *浙江大学学报 (工学版)*, 2021, **55**(6): 1215-1224)
- [9] ZHOU Biying, WANG Aiping, FEI Changjiang, *et al.* A satellite network resource scheduling mechanism based on reinforcement learning[J]. *Computer Engineering and Science*, 2019, **41**(12): 2134-2142 (周碧莹, 王爱平, 费长江, 等. 基于强化学习的卫星网络资源调度机制 [J]. *计算机工程与科学*, 2019, **41**(12): 2134-2142)
- [10] PENG Shuang, WU Jiangjiang, CHEN Hao, *et al.* Satellite onboard observation task planning based on attention neural network[J]. *Computer Science*, 2022, **49**(7): 242-247 (彭双, 伍江江, 陈浩, 等. 基于注意力神经网络的地对观测卫星星上自主任务规划方法 [J]. *计算机科学*, 2022, **49**(7): 242-247)
- [11] WANG Haijiao. Massive Scheduling Method Under Online Situation for Satellites Based on Reinforcement Learning[D]. Beijing: University of Chinese Academy of Sciences (National Space Science Center, Chinese Academy of Sciences), 2018 (王海蛟. 基于强化学习的卫星规模化在线调度方法研究 [D]. 北京: 中国科学院大学 (中国科学院国家空间科学中心), 2018)
- [12] LI Dalin. Research on Observation Scheduling Model and Method of Astronomy Satellite[D]. Harbin: Harbin Institute of Technology, 2021 (李大林. 天文观测卫星任务规划模型与方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2021)
- [13] 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
- [14] WU Baixuan. Research on Multi-star and Multitask Planning Based on Contract Network Protocol[D]. Harbin: Harbin Engineering University, 2020 (吴白轩. 基于合同网协议的多星多任务规划方法研究 [D]. 哈尔滨: 哈尔滨工程大学, 2020)
- [15] JIANG Wei, PANG Xiuli. Collaborative Mission Planning Method for Networking Imaging Satellite[M]. Harbin: Harbin Institute of Technology Press, 2016 (姜维, 庞秀丽. 组网成像卫星协同任务规划方法 [M]. 哈尔滨: 哈尔滨工业大学出版社, 2016)
- [16] MNH V, KAVUKCUOGLU K, SILVER D, *et al.* Playing Atari with deep reinforcement learning[OL]. arXiv preprint arXiv: 1312.5602, 2013

作者简介



李英玉 女, 1976 年 11 月出生于河北省迁安市. 现为中国科学院国家空间科学中心副研究员, 硕士生导师, 主要研究方向为空间信息组织、复杂系统仿真等.
E-mail: liyingyu@nssc.ac.cn



史好迎 (通信作者) 女, 1995 年 10 月出生于河南省洛阳市. 现为中国科学院国家空间科学中心硕士, 主要研究方向为卫星任务智能规划、强化学习在信息系统中的应用等.
E-mail: shihaoying1015@outlook.com