

微小卫星集群协同观测构形调整规划

汪志军¹, 陈 鹏², 赵艳斌³, 程月华^{4,5†}, 廖 鹤¹, 姜 斌^{4,5}

(1. 南京航空航天大学 航天学院, 江苏 南京 210016; 2. 上海航天技术研究院北京研发中心, 北京 100081;

3. 上海卫星工程研究所, 上海 200240; 4. 南京航空航天大学 自动化学院, 江苏 南京 211106;

5. 江苏省物联网与控制技术重点实验室, 江苏 南京 211106)

摘要: 以微小卫星集群实现小行星探测为背景, 研究局部信息交互的空间目标观测任务构形调整. 针对难以直接求取集群构形调整的全局最优解问题, 利用通信协调图, 将全局协调决策分解成多个局部求解问题, 并引入强化学习机制实现求解. 针对集群全局协调决策问题, 通过设计基于Max-plus算法的全局协调决策算法来实现全局协作; 针对单星局部优化问题, 设计基于神经网络的局部Q学习算法来实现单星动作调整规划. 仿真结果表明, 本文所提的协作规划算法能自主有效地将集群调整至期望构形, 实现协同观测任务.

关键词: 微小卫星集群; 构形调整; 协调图; 强化学习; Max-plus算法

引用格式: 汪志军, 陈鹏, 赵艳斌, 等. 微小卫星集群协同观测构形调整规划. 控制理论与应用, 2019, 36(12): 2047 – 2054

DOI: 10.7641/CTA.2019.90500

Configuration adjustment planning for cooperative observation of microsatellite cluster

WANG Zhi-jun¹, CHEN Peng², ZHAO Yan-bin³, CHENG Yue-hua^{4,5†},
LIAO He¹, JIANG Bin^{4,5}

(1. College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 210016, China;

2. Beijing R&D Center, Shanghai Academy of spaceflight Technology, Beijing 100081, China;

3. Satellite Engineering Institute of Shanghai, Shanghai 200240, China;

4. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 211106, China;

5. Jiangsu Key Laboratory of Internet of Things and Control Technology, Nanjing Jiangsu 211106, China)

Abstract: This paper investigates the problem of configuration adjustment of spatial target observation tasks for local information interaction during the deep space exploration mission of the microsatellite cluster. In order to solve the problem of directly solving the global optimal difficulty of cluster configuration adjustment, this paper uses the communication coordination diagram to decompose the global coordination strategy into multiple local solution problems, and the reinforcement learning mechanism is introduced to realize solution. Aiming at the global coordination decision-making problem of cluster, the cooperative interaction in the local learning process is realized by designing the global coordinated decision algorithm based on Max-plus algorithm. Aiming at the local optimization problem of individual satellite, a local Q learning algorithm based on neural network is designed to realize the adjustment action planning of each satellite. The simulation results show that the proposed planning algorithm can independently and effectively adjust the cluster to the desired configuration to realize the observation task collaboratively.

Key words: microsatellite cluster; configuration adjustment; coordination graph; reinforcement learning; max-plus algorithm

Citation: WANG Zhijun, CHEN Peng, ZHAO Yanbin, et al. Configuration adjustment planning for cooperative observation of microsatellite cluster. *Control Theory & Applications*, 2019, 36(12): 2047 – 2054

收稿日期: 2019-06-29; 录用日期: 2019-11-26.

†通信作者. E-mail: chengyuehua@nuaa.edu.cn; Tel.: +86 25-52832301.

本文责任编辑: 王大轶.

国家重点研发计划项目(2016YFB0500803), 研究生创新基地(实验室)开放基金项目(kfjj20180321), 江苏省高校优势学科建设工程项目资助.

Supported by the National Key Research and Development Plan (2016YFB0500803), the Postgraduate Innovation Base (Laboratory) Open Fund (kfjj20180321) and the Jiangsu Province University Advantageous Subject Construction Project Funding Project.

1 引言

近年来,国内外有关深空探测的研究成果显著.例如中国“嫦娥四号”首次实现在月球背面软着陆,美国“洞察号”登陆火星等.现代微小卫星的优势越来越明显,成百上千颗微小卫星构成的集群灵活性高、鲁棒性强,能完成大卫星无法完成的任务,例如小行星探测等,已成为当今国际航天领域的一个研究热点.微小卫星集群在小行星探测任务中,通过构形调整对空间目标实施观测.

微小卫星集群组成的观测系统应用于空间目标观测时,可以充分利用集群优势,完成对目标的多角度观测、监视与识别,获取更加详细的目标特征信息.为了达到协同观测的效果,集群内各颗卫星需自主规划并协调各自的策略,通过协作完成构形调整以达到任务要求的观测构形.

构形调整普遍采用的方法是先规划出最优轨迹,然后对该轨迹进行跟踪控制. Mark E等人^[1]通过求解哈密顿-雅可比-贝尔曼(Hamilton-Jacobi-Bellman, HJB)方程生成一系列从初始稳定构形到最终构形的机动动作集合,然后根据任务约束从该集合中选择机动动作,并通过轨道动力学模型来优化推力切换时间以确定最终的控制策略. Tillerson M等人^[2]针对编队构形调整中的协调问题,提出了基于顶层编队规划器和底层构形保持算法的两层结构优化框架,并通过线性整数规划方法实现了轨迹规划和构形保持. Beard R W等人^[3]提出了一种编队构形机动的协调体系结构,其中最高层为任务的动态转换,中间层用于协调每个航天器的运动以获得期望构形,最底层为每个航天器的单独控制层.

现有研究大多基于最优控制理论将系统上层规划和底层单星控制分离进行设计,但随着参与任务的卫星数目的增多,迭代寻优次数也会越来越多,计算量陡增.另一方面按照上层预规划化的结果实施构形调整,无法应对任务过程中的实际运行情况及突发情况,缺乏自主性.集群中的卫星具有物理分散的特点,需要自主规划并协调各自的策略,按照任务需求在规定时间内完成构形调整以协同完成观测任务.集中式规划决策无法运用,需寻求分布式协作规划方法.

随着Agent技术的发展,目前越来越多的研究将多智能体系统(multi-agent system, MAS)应用于分布式卫星系统(distributed satellite system, DSS)中,采用了MAS描述DSS组织结构.高黎^[4]在分析DSS系统自主协作运行模式的基础上,针对分布式对地观测卫星系统建立了一种面向自主协作的MAS体系结构,并基于协商机制研究了任务优化分配问题.陈琪峰等人^[5]针对DSS构形自主保持规划问题,提出了基于多Agent协商的最优构形调整策略.张正强等人^[6]将MAS技术引入分布式卫星系统的规划与调度问题中,建立了

DSS的分层分布式多Agent结构模型,并详细设计了各层次Agent的功能,然后基于合同网协议进行多Agent协商来解决规划与协调调度问题.

目前MAS协作规划问题求解主要是从全局考虑,Agent之间基于协商机制实现协作规划,但难以直接求解MAS全局最优策略.本文针对此问题,借助图模型表示多智能体系统中复杂的依赖关系,将构形调整的全局优化问题分解为单星局部优化,为求解MAS全局最优策略提供了一条有效途径;同时,引入强化学习算法,利用单星的局部信息交互规划求解得到最优行为策略,从而实现集群的自主构型调整.

2 基于MAS协作的集群构形调整规划模型

本节先对集群构形调整规划问题进行描述,然后根据问题的特点建立以Agent自主规划为基础的MAS协作规划框架,最后分析了集群中的卫星在协作规划过程中的任务约束条件和性能指标.

2.1 问题描述

由多颗卫星组成的集群对空间目标实施观测时,必须按照观测任务需求形成特定构形,以协同完成观测.因此,需要规划各颗卫星的相对运动转移轨迹以及相应的控制,如图1所示.

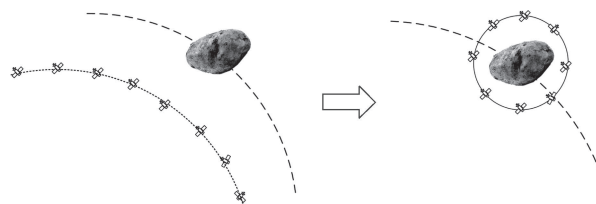


图1 集群协同观测构形调整

Fig. 1 Cluster coordinated observation and configuration adjustment

集群协同观测的构型调整是一类典型的MAS协作问题^[7-9],可描述为多阶段优化决策问题:给定集群的初始状态和观测任务要求下的期望构形状态、每颗卫星实施构形调整可用的动作集合,将规划时间划分为 N 个决策阶段,对集群中每颗卫星顺序规划每个决策阶段的调整动作,将各个决策阶段的最优调整动作组合以获得构形调整策略,实施调整策略以达到观测任务期望的构形.对于每个规划决策阶段,在满足观测任务约束的条件下,各个卫星Agent根据要实现的任务目标以及待优化的性能指标进行局部优化,并通过协作交互与其他Agent进行全局协调,自主决策规划出从初始状态调整到目标构形状态的最优动作序列.上述过程可用一个七元数组进行描述

$$M = \langle \text{SAT}, \text{Config}, S, A, \text{Rule}, \Pi, C \rangle, \quad (1)$$

其中: $\text{SAT} = \{\text{sat}_1, \text{sat}_2, \dots, \text{sat}_m\}$ 为集群中的卫星集合, m 为集群中的卫星总数; $\text{Config} = \{\text{Initial}, \text{Goal}\}$ 为集群构形状态, Initial 为集群初始构形状态,

Goal为集群协同观测的期望构形状态; S 为集群系统所有状态的集合; $A \rightarrow A_1 \times A_2 \times \cdots \times A_m$ 为构形调整问题中所有卫星的有限动作集合, A_i 为每个卫星Agent可用的调整动作集合; Rule为集群的全局协调决策策略; $\Pi: S \times A \rightarrow \Pi$ 为集群中所有卫星Agent的调整策略,

$$\Pi = \{\pi_1, \pi_2, \cdots, \pi_m\},$$

π_i 为第 i 颗卫星Agent的调整动作序列(即调整策略); C 为构形调整协同规划问题中的任务约束以及待优化的性能指标.

2.2 MAS协作规划框架

Agent自主规划是MAS协作的基础. 将集群中的卫星抽象为Agent, 则本文所讨论的Agent是一个具有感知、信息处理、决策、通讯与执行能力的软硬件结合体. 每个Agent的内部规划结构可分为信息交互模块、规划决策模块以及用于执行规划方案的硬件驱动

模块, 如图2所示. MAS协作规划即: 在单Agent自主规划的基础上, 基于集群通信结构进行协作交互和全局协调决策, 以达到协同的目的.

信息交互模块、通信模块和硬件驱动模块负责与系统环境和其他Agent进行交互. 信息交互模块用于感知环境状态信息并对Agent自身状态信息以及与之协调的其他Agent的信息进行处理, 然后生成协调信息输入到规划决策模块; 规划决策模块根据规划算法和协调信息进行局部优化决策, 生成规划方案, 并将局部规划信息提供给通信模块用于与其他Agent协调交互; 硬件驱动模块一方面为规划模块提供Agent自身状态信息, 另一方面也接收规划方案信息, 并控制相关组件执行动作作用于环境. Agent执行调整动作后, 环境状态发生改变, 形成一个状态-动作对序列, 此时信息交互模块感知到新的环境状态以及反馈的调整动作效果, Agent重复上述过程, 不断迭代规划, 直到完成任务目标.

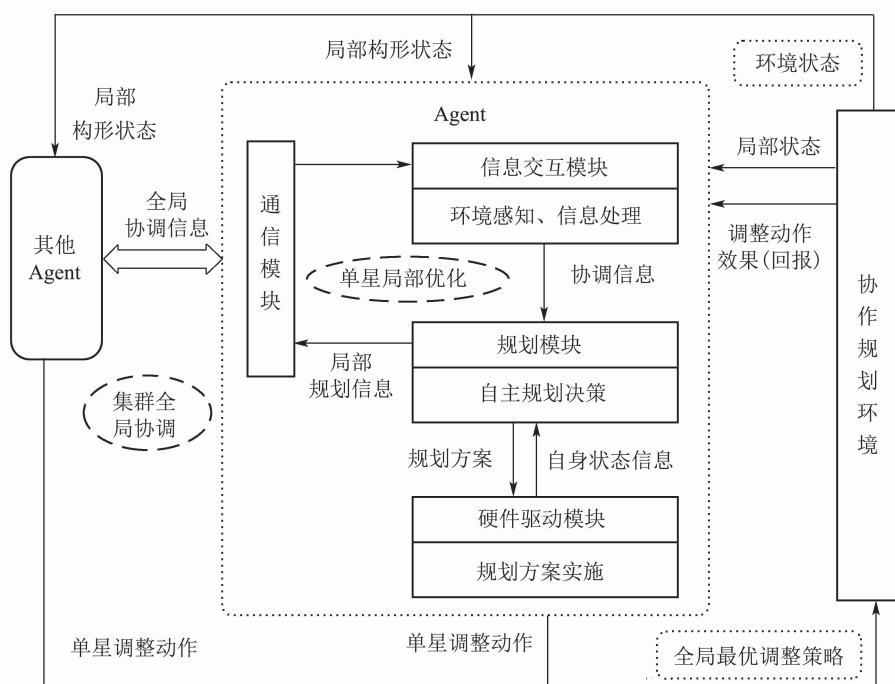


图 2 基于Agent自主规划的MAS协作规划结构

Fig. 2 MAS collaborative planning structure based on Agent self-planning

2.3 任务约束及性能指标

下面讨论集群中的卫星在协作规划过程中的任务约束以及优化的性能指标函数. 为便于界定, 首先做一些相关定义.

定义 1 本文只考虑轨道平面内的构形调整, 因此定义每颗卫星 i 的相对运动状态为

$$X_i = [x_i \ y_i \ \dot{x}_i \ \dot{y}_i]. \quad (2)$$

定义 2 根据定义1, 则集群系统状态可表示为

$$SW = \{X_1, X_2, \cdots, X_m\}, \quad (3)$$

其中 m 为集群中的卫星数量.

定义 3 为了方便对观测构形约束进行处理, 同时也为了直观地描述卫星之间的相对运动, 定义集群的构形状态由每颗卫星相对于观测目标的距离以及该星与近邻卫星之间的相对距离描述. 本文所说的近邻指通信关系.

卫星 i 相对于观测目标的距离为

$$d_i = \sqrt{x_i^2 + y_i^2}, i = 1, 2, \dots, m. \quad (4)$$

卫星*i*与近邻卫星*j*的相对距离为

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (5)$$

其中: $j \in \Gamma(i)$, $\Gamma(i)$ 表示卫星*i*的近邻; 若*i, j*互为近邻关系, 则 $d_{ij} = d_{ji}$.

则集群构形可表示为

config =

$$\{d_1, d_2, \dots, d_m, d_{1\Gamma(1)}, d_{1\Gamma(2)} \dots, d_{m\Gamma(m)}\}. \quad (6)$$

规划时刻*t*的环境状态 s^t 由该时刻的集群构形状态描述

$$s^t = \text{config}^t, t = t_0 + k\Delta t, \quad (7)$$

其中 Δt 为一个决策周期的时间.

1) 任务约束.

① 观测构形约束.

协同观测的目标构形是由任务需求决定的集群卫星之间的空间相对几何关系, 本文的任务目标要求实现集群中各颗卫星均匀分布在目标轨道周围. 集群中每颗卫星的最终相对状态是未知的, 只由卫星相对目标的距离以及相位差约束.

根据定义3中对集群构形的描述, 本文采用卫星*i*相对于观测目标的期望距离 D_i 以及卫星*i, j*之间的期望相对距离 D_{ij} 来表达目标构形约束, 如图3所示. 则期望构形可表达为

$$\text{config}^d = \{D_1, D_2, \dots, D_m, D_{1\Gamma(1)}, D_{2\Gamma(2)}, \dots, D_{m\Gamma(m)}\}. \quad (8)$$

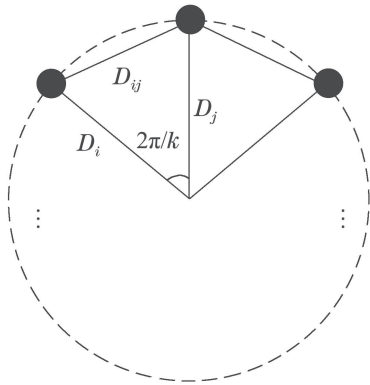


图3 观测构形约束

Fig. 3 Observed configuration constraint

构形调整的目标要求集群最终构形状态 S^{t_f} 达到期望的观测构形, 但在实际情况中不可能完全符合, 因此本文将观测构形约束表示为

$$\|S^{t_f} - \text{config}^d\| \rightarrow \epsilon, \quad (9)$$

其中 ϵ 为允许的构形偏差, 通过欧氏距离来描述.

② 星间碰撞避免约束.

集群中的多颗卫星同时调整自身的相对运动, 为了避免星间碰撞情况发生, 一般采用星间安全距离 d_{safe} 来约束. 本文只考虑平面运动, 则

$$[x_i(t) - x_j(t)]^2 + [y_i(t) - y_j(t)]^2 \geq d_{\text{safe}}^2, \quad (10)$$

其中: $t \in [t_0, t_f]$, *i, j*为集群中的两颗不同卫星.

③ 控制约束.

继电型推力模型下, 轨控发动机推力加速度为常值, 则推力器每作用一次所能获得的加速度为

$$|a_x| = |a_y| = u_{\text{max}}. \quad (11)$$

2) 性能指标.

由于每颗卫星星上微推进系统所能提供的推力有限, 为了观测任务成功实施, 以集群整个构形调整过程中所需的总速度增量为优化的性能指标

$$\min \sum_{i=1}^m (|\Delta v_x^i| + |\Delta v_y^i|), \quad (12)$$

其中: Δv_x^i 为卫星*i*在*x*轴上的速度增量, Δv_y^i 为卫星*i*在*y*轴上的速度增量.

3 基于通信协调图的全局协调决策

强化学习是一种无模型可在线学习的方法, 通过不断试验和试错获得环境反馈信息, 以此评判所采取的动作对环境的适应程度, 不断迭代形成状态到动作的映射策略, 直到逼近最优行为策略, 实现卫星的自主规划.

本文采用多智能体强化学习(multi-agent reinforcement learning, MARL)的方法求解全局最优的动作调整. 由于期望的任务目标只约束了集群的整体构形, 而每颗卫星最终的状态信息是不确定的, 且每颗卫星的局部调整动作都将引起集群构形状态动态变化. 直接求解全局最优策略比较困难.

针对该问题, 作者借鉴了Jelle R. Kok等人^[10]和Nikos Vlassis^[11]提出的基于消息传递机制的Max-plus算法, 对组合动作下的全局*Q*学习进行分解, 将全局学习任务分解到各个Agent进行局部学习. 当局部动作取得最优时, 所有卫星的组合动作调整也取得最优.

考虑到算法对大规模集群的可扩展性, 本文以通信的两颗卫星*i, j*为一个分解单位^[12], 基于 $G = \langle V, E \rangle$ 中的边 e_{ij} 对全局值函数*Q*进行分解, 每条边对应一个局部*Q*函数, 则全局*Q*函数可分解为所有局部*Q*函数之和

$$Q(s, a) = \sum_{i=1}^K Q_i(s_{ij}, a_i, a_j), i, j \in E, \quad (13)$$

其中: $s_{ij} \subseteq (S_i \cap S_j)$ 为局部状态, S_i 和 S_j 分别为 Agent i, j 的有限状态集, a_i 和 a_j 分别为 Agent i, j 的调整动作。

由式(13)可以看出, 局部 Q 函数 $Q_i(s_{ij}, a_i, a_j)$ 只取决于局部状态和边 e_{ij} 所连接的具有通信关系的两个 Agent 的调整动作组合。因此, 可以按照 Agent 之间的通信关系两两一组进行分解, 这样基于集群组合状态和动作的全局 Q 学习问题就变成了多个局部 Q 学习任务的组合, 而每个局部 Q 学习过程仅涉及具有通信关系的两个 Agent。因此, 针对大规模集群, 该分解模型也能灵活适用, 且单个局部 Q 学习的状态-动作空间不会受集群中 Agent 数目的影响。

4 基于通信协调图分解的局部优化

针对单星局部优化问题, 本节基于 Max-plus 的全局协调决策算法集成于局部 Q 学习算法中来实现集群全局协调和单星局部优化的迭代, 以此求解集群构形调整的规划问题。

4.1 效用函数及目标函数

1) 回报项。

在单星局部优化 Q 学习中, 每个 Agent 的学习过程仅依赖与之有通信关系的 Agent, 其回报矩阵表现出稀疏性。由于所有 Agent 组合动作下的全局回报 R 为所有 Agent 回报的总和, 根据通信依赖关系, 基于协调图将全局回报函数分解为局部回报函数的线性组合, 集群全局回报构成如图4所示。图4中 $r_i = r(a_i)$, $r_{ij} = r(a_i, a_j)$, $(i, j) \in E$ 且 $r_{ij} = r_{ji}$ 。

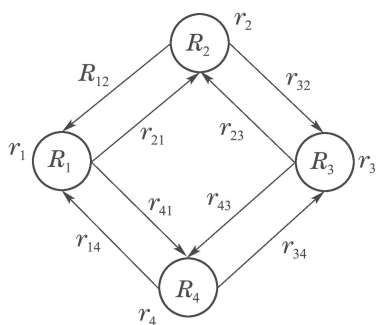


图4 集群全局回报构成

Fig. 4 Composition of cluster global returns

Agent i 的局部回报 R_i 定义为与近邻 Agent 交互产生的局部回报之和

$$R_i(a) = r(a_i) + \sum_{j \in \Gamma(i)} r(a_i, a_j), \quad (14)$$

其中: $r(a_i)$ 表示 Agent i 的调整动作所获得的回报, $r(a_i, a_j)$ 表示具有通信关系的 Agent i, j 的组合调整动作 (a_i, a_j) 给集群构形变化带来的回报。

具有通信关系的 Agent i, j 实施各自的调整动作后获得的局部回报为

$$R_{ij} = \frac{r(a_i)}{2} + \frac{r(a_j)}{2} + r(a_i, a_j) + r(a_j, a_i). \quad (15)$$

在构形调整问题中, 相对距离 d_i 的变化只与调整动作 a_i 相关, 而相对距离 d_{ij} 的变化则受组合调整动作 (a_i, a_j) 的影响。因此可利用 Agent 当前局部构形与目标构形之间的构形差 Δd 来设计回报函数

$$r(a_i) = \|d_i - D_i\|, \text{ 若 } d_{ij} \geq d_{\text{safe}}, \quad (16)$$

$$r(a_i, a_j) = r(a_j, a_i) = \begin{cases} \frac{1}{2} \|d_{ij} - D_{ij}\|, & \text{若 } d_{ij} \geq d_{\text{safe}}, \\ 1000, & \text{其他.} \end{cases} \quad (17)$$

注 1 式(17)中将碰撞避免约束综合在回报函数中进行考虑, 本文以最小化为优化方向。1000 表示惩罚项, 含义如下: 若星间距离在安全距离内, 则该动作下的惩罚项极大, 经过学习, Agent 在选择动作的过程中则可避免碰撞发生。同时, 这样处理也能避免引入非凸约束。

2) 构形调整控制量。

构形调整控制量用于评价执行某个调整动作所需的代价。根据 Agent 所执行的调整动作的推力加速度 U 可以确定所需的控制量

$$U_\epsilon(s_k, a_k) = \int_t^{t+\Delta t} \frac{1}{2} U^T U dt. \quad (18)$$

3) 效用函数。

综合式(1)–(2), 那么第 k 个决策阶段的效用函数 $U(s_k, a_k)$ 可定义为

$$U(s_k, a_k) = \beta R_{ij}(s_k, a_k) + (1 - \beta) U_\epsilon(s_k, a_k), \quad (19)$$

其中: β 为权重, $s_k = s_{ij}^k$, $a_k = (a_i^k, a_j^k)$ 。

4) 目标函数。

采用局部 Q 学习进行单星局部优化的目标函数为规划时间段内多个决策阶段的折扣效用函数之和

$$J(s_k) = \sum_{j=0}^{N+1-k} \gamma^j U(s_{k+j}, a_{k+j}), \quad (20)$$

其中 $\gamma \in [0, 1]$ 为折扣因子。

4.2 局部 Q 学习规划结构

1) 局部 Q 函数定义与更新。

强化学习的任务是通过迭代优化所学习的目标函数, 获得最优目标函数值所对应的最优动作。根据 Bellman 最优性原理, 第 k 个决策阶段的最优目标函数值可以通过递推求解如下离散 HJB 方程获得

$$J(s_k)^* = \min_{a_k \in \Theta} (U(s_k, a_k) + \gamma J^*(s_{k+1})), \quad (21)$$

其中: $U(s_k, a_k)$ 为效用函数, $\Theta = A_i \times A_j$ 为 Agent i, j 的组合动作空间。

以上HJB方程可以采用动态规划的方法逐级逆向递推进行求解,但是动态规划方法只适用于系统完全了解并可控制的环境,且计算量十分巨大.为此本文基于Q学习方法对该HJB方程进行迭代近似求解,以获得最优调整策略.

定义局部Q函数为

$$Q(s_k, a_k) = U(s_k, a_k) + \gamma J^*(s_{k+1}). \quad (22)$$

进一步可得

$$Q(s_k, a_k) = U(s_k, a_k) + \gamma \min_{a_{k+1}} Q(s_{k+1}, a_{k+1}), \quad (23)$$

其中: $s_k = s_{ij}^k$ 为局部状态, $a_k = (a_i^k, a_j^k)$.

Q学习中,全局Q函数的更新可以表示为

$$Q(s, a) \leftarrow Q(s, a) + \alpha(U(s, a) + \gamma \min_{a' \in A} Q(s', a') - Q(s, a)). \quad (24)$$

全局效用函数可表示为局部效用函数之和:

$$U(s, a) = \sum_{i=1}^m U_i(s, a) = \sum_{i=1}^m U_{ij}(s, a). \quad (25)$$

通过Agent之间的协作交互达成集群全局协调后, $\min_{a \in A} Q(s', a)$ 可近似为

$$\min_{a \in A} U_{a \in A}(s', a) = Q_{a_i^* \in A_i}(s', a_1^*, a_2^*, \dots, a_K^*) \approx \sum_{i=1}^m Q_i(s'_{ij}, a_i^*, a_j^*). \quad (26)$$

将式(25)–(26)代入式(24)拆解后可得局部Q函

数的更新如下:

$$Q_i(s_{ij}, a_i, a_j) \leftarrow Q_i(s_{ij}, a_i, a_j) + \alpha(U_i(s_{ij}, a_i, a_j) + \gamma Q_i(s'_{ij}, a_i^*, a_j^*) - Q_i(s_{ij}, a_i, a_j)), \quad (27)$$

其中: s'_{ij} 为Agent i, j 分别执行调整动作 a_i, a_j 后的局部构形状态; a_i^*, a_j^* 为基于Max-plus的全局协调决策算法所得的最优调整动作.

2) 基于神经网络的局部Q学习.

Q学习方法一般基于Q-table查表法实现,但是随着状态和动作的维数增加,搜索空间维数呈指数级增长.针对该问题,利用神经网络可表达任意函数近似结构的特点来逼近局部Q函数,通过权值训练获得不同状态–动作下的Q值映射.

基于神经网络的Q学习方法也被称为自适应动态规划或近似动态规划(approximate dynamic programming, ADP)^[13],每个局部优化Q函数的规划结构如图5所示.整个规划包含系统模型、全局协调决策算法、评价网络3个部分,其中系统模型用于确定并输出系统的转移状态,即执行调整动作后各颗卫星的相对运动状态变化;基于Max-plus的全局协调决策算法用于协调各颗卫星的行为,输出给定状态 s_k 下的最优调整动作估计 a_k^* ; 评价网络采用BP神经网络结构,输出局部Q函数的估计值用于对调整动作进行评价.

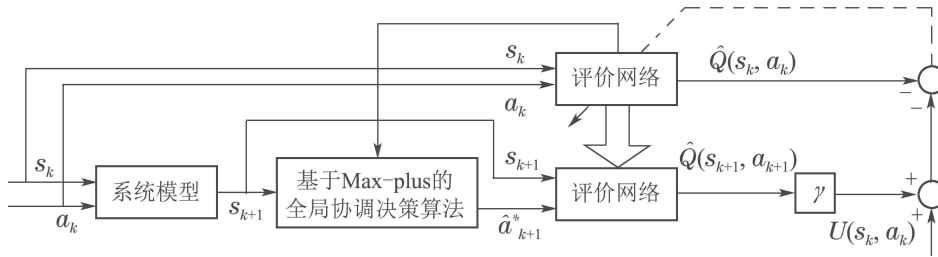


图5 基于神经网络的局部Q学习规划结构

Fig. 5 Local Q learning planning structure based on neural network

5 集群构形调整仿真分析

集群中各颗卫星的初始状态参数如表1所示,其中4颗卫星经过构形调整,均匀分布在目标参考轨道水平面投影半径200 m的圆上.各颗卫星需在0.1个轨道周期内实现从初始状态到期望观测构形状态的协作调整,参考轨道周期 $T = 5983.86$ s.

在多星同时学习的协作规划过程中,每颗星以 $\Delta t = 10$ s 为决策周期进行构形调整动作规划.协作调整后的集群构形状态如表2所示.本文以卫星的相对距离构形表达与期望的观测构形之间的欧氏距离来描述构形偏差,可以看出,调整后的构形偏

差能满足目标协同观测的任务要求.

表1 微小卫星集群的初始状态参数

Table 1 Initial state parameters of a microsatellite cluster

微小卫星	x_0/m	y_0/m	$\dot{x}_0/(\text{m} \cdot \text{s}^{-1})$	$\dot{y}_0/(\text{m} \cdot \text{s}^{-1})$
1	110	260	0	-0.02
2	250	-150	-0.1	0
3	-130	-250	0	0.02
4	-260	120	0.1	0

表 2 调整后的集群构形状态
Table 2 Adjusted cluster configuration status

参数	d_1/m	d_2/m	d_3/m	d_4/m	d_{12}/m	d_{23}/m	d_{34}/m	d_{41}/m	ϵ/m
期望值	200	200	200	200	$200\sqrt{2}$	$200\sqrt{2}$	$200\sqrt{2}$	$200\sqrt{2}$	5
实际值	199.46	198.92	199.69	198.54	283.17	280.13	281.00	281.39	4.077

协作规划下, 集群中各颗卫星的构形调整过程及相应的调整轨迹分别如图6所示. 可以看出, 本文的构形调整协作规划策略可以在规定时间内将集群从初始构形状态调整至期望的构形以实现协同观测, 且调整过程中不会发生碰撞等冲突.

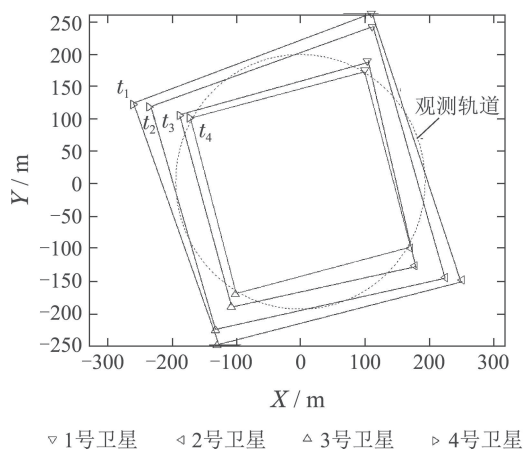


图 6 集群构形调整过程

Fig. 6 Cluster configuration adjustment process

图7为整个构形调整过程中集群构形偏差的变化情况. 可以看出, 集群的整体构形偏差一直在往全局优化的方向变化, 整个变化过程较为平滑, 说明集群中各颗卫星进行了有效的协作交互, 达到了全局协调的效果.

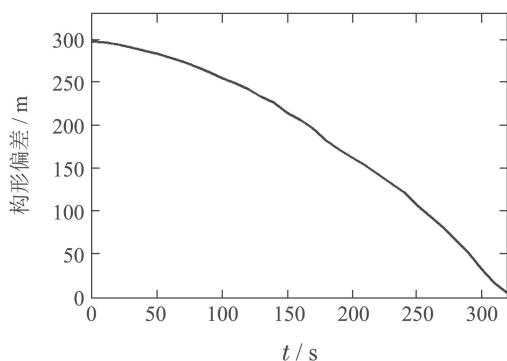


图 7 集群调整过程中的构形偏差

Fig. 7 Configuration deviation during cluster adjustment

以上图表所示的仿真结果表明, 本文所提的基于Max-plus的全局协调算法以及基于通信协调图分解的单星局部优化Q学习算法是有效的. 在满足各种任务约束的条件下, 各颗卫星同时规划决策, 可

以实现单星自主局部规划的同时与其他卫星Agent协作交互, 通过有效地全局协调达到协作的目的. 在协作所得的最优构形调整策略作用下, 集群中的多颗卫星同时进行调整, 可以将集群从初始状态调整至协同观测期望的构形.

6 结论

本文对集群协同观测的构形调整规划问题展开研究, 引入MAS强化学习实现协作规划, 通过集群全局协调决策和单星局部优化的迭代对问题进行求解. 通过构建通信协调图, 基于Max-plus进行集群全局协调决策; 通过协调图分解, 采用基于神经网络的局部Q学习进行单星局部优化. 仿真结果表明, 基于Max-plus的全局协调决策算法可以有效地对每颗卫星的调整动作进行协调, 在此基础上采用局部Q学习方法实现的单星局部优化也可以获得良好的学习效果. 通过两级优化迭代能达到集群协同观测构形调整协作规划的目的, 所求得的对每颗卫星的近似最优构形调整策略可以使集群在规定时间内调整至期望的观测构形.

参考文献:

- [1] CAMPBELL M E. Planning algorithm for large satellite clusters. *AIAA Guidance, Navigation, and Control Conference and Exhibit*. Monterey, California: AIAA, 2002: 4958 – 4966.
- [2] TILLERSON M, INALHAN G, HOW J P. Co-ordination and control of distributed spacecraft systems using convex optimization techniques. *International Journal of Robust & Nonlinear Control*, 2002, 12(2/3): 207 – 242.
- [3] BEARD R W, LAWTON J, HADAEGH F Y. A coordination architecture for spacecraft formation control. *IEEE Transactions on Control Systems Technology*, 2002, 9(6): 777 – 790.
- [4] GAO Li. *Research on earth observation task cooperation for distributed satellite system*. Changsha: National University of Defense Technology, 2007.
(高黎. 对地观测分布式卫星系统任务协作问题研究. 长沙: 国防科学技术大学, 2007.)
- [5] CHEN Qifeng, WU Wenzhao, DAI Jinhai. Autonomous formation keeping planning for distributed satellite system based on multi agent negotiation. *Journal of Astronautics*, 2008, 29(2): 517 – 521.
(陈琪峰, 吴文昭, 戴金海. 基于多Agent协商的分布式卫星自主构形保持规划研究. 宇航学报, 2008, 29(2): 517 – 521.)
- [6] ZHANG Zhengqiang, TAN Yuejin, WANG Junmin. Research on MAS-based mission planning for distributed satellite system. *Journal of System Simulation*, 2007, 19(12): 2868 – 2871.
(张正强, 谭跃进, 王军民. 基于MAS的分布式卫星系统任务规划研究. 系统仿真学报, 2007, 19(12): 2868 – 2871.)

- [7] YANG Yupu, LI Xiaomeng, XU Xiaoming. A survey of technology of multiagent cooperation. *Information and Control*, 2001, 30(4): 337 – 342.
(杨煜普, 李晓萌, 许晓鸣. 多智能体协作技术综述. 信息与控制, 2001, 30(4): 337 – 342.)
- [8] WANG Chong, JING Ning, LI Jun, et al. An algorithm of cooperative multiple satellites mission planning based on multi-agent reinforcement learning. *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. 国防科技大学学报, 2011, 33(1): 53 – 58.)
- [9] LUO Delin, ZHANG Haiyang, XIE Rongzeng, et al. Unmanned aerial vehicles swarm conflict based on multi-agent system. *Control Theory & Applications*, 2015, 32(11): 1498 – 1504.
(罗德林, 张海洋, 谢荣增, 等. 基于多agent系统的大规模无人机集群对抗. 控制理论与应用, 2015, 32(11): 1498 – 1504.)
- [10] KOK J R, VLASSIS N. Collaborative multiagent reinforcement learning by payoff propagation. *Journal of Machine Learning Research*, 2006, 7(1): 1789 – 1828.
- [11] VLASSIS N. A concise introduction to multiagent systems and distributed artificial intelligence. *Synthesis Lectures on Artificial Intelligence & Machine Learning*, 2007, 1(1): 71.
- [12] GUESTIN C, LAGOUDAKIS M G, PARR R. Coordinated reinforcement learning. *The 19th International Conference on Machine Learning*. San Francisco: Morgan Kaufmann Publishers Inc, 2002.
- [13] LIU D R. Approximate dynamic programming for self-learning control. *Acta Automatica Sinica*, 2005, 31(1): 13 – 18.

作者简介:

汪志军 硕士研究生, 目前研究方向为航天器任务规划, E-mail: jesse341@163.com;

陈鹏 工程师, 研究方向为航天器总体设计, E-mail: 250084163@qq.com;

赵艳斌 研究员, 研究方向为航天器总体设计, E-mail: zhaoyanbin_509@163.com;

程月华 副研究员, 研究方向为航天器任务规划、故障诊断与容错控制, E-mail: chengyuehua@nuaa.edu.cn;

廖鹤 副研究员, 研究方向为卫星动力学与控制, E-mail: liaohe_crane@nuaa.edu.cn;

姜斌 教授, 研究方向为故障诊断与容错控制, E-mail: binjiang@nuaa.edu.cn.