中国科学: 技术科学 2023 年 第 53 卷 第 2 期: 200 ~ 209
SCIENTIA SINICA Technologica techcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
论 文 群体智能激发汇聚及应用专辑

一种基于PIO改进强化学习的航天器复杂多约束姿态机动规划新方法

华冰*, 孙胜刚, 吴云华, 陈志明

南京航空航天大学航天学院, 南京 210016
*E-mail: huabing@nuaa.edu.cn

收稿日期: 2021-08-04; 接受日期: 2021-10-14; 网络版发表日期: 2022-07-14
中央高校基本科研业务费(编号: NS2021063)资助项目

摘要 本文针对多个姿态约束条件下的航天器姿态机动规划问题进行了研究, 提出了一种基于鸽群算法的改进的策略梯度强化学习算法(PIOPGRL). 首先, 针对强制指向约束和禁止指向约束, 建立了基于角度的姿态约束模型, 根据约束模型建立了强化学习的回报函数. 然后, 使用适应度函数替代策略评价函数, 将鸽群算法与强化学习相融合. 针对策略梯度强化学习算法计算量大、收敛速度慢的问题, 使用鸽群算法求解策略梯度, 极大减少了计算量. 仿真结果表明, 相比于策略梯度强化学习算法, 基于自PIO改进强化学习的航天器姿态机动规划算法(PIOPGRL)在极大减少计算量的同时, 有更优的规划结果, 更小的机动代价, 适用于微小航天器解决多个姿态约束条件下的姿态机动规划问题.

关键词 姿态机动, 姿态约束, 路径规划, 强化学习, 鸽群算法, 航天器

1 引言

随着航天任务愈发多样, 航天器执行任务时要进行大量的姿态机动. 在姿态机动过程中, 航天器需要满足一定的约束条件. 例如, 为了获取能量, 航天器的光伏电池阵列必须始终保持面对太阳; 而对于某些精密器件, 在姿态机动过程中却需要避免太阳光直射. 除了上述姿态约束, 还需要考虑通信天线指向、光学敏感器杂光抑制、机动能力有限等约束. 因此, 在航天器执行姿态机动任务时, 需要在复杂约束条件下进行姿态机动规划, 合理的规划是任务执行的重要保障.

对于航天器姿态机动规划问题, 国内外学者展开了大量研究. 传统思路有两种. 一种是在线规划法, 如半定规划法[1]、约束检测法[2]和随机规划法[3], 但是上述方法普适性差, 计算量较大, 难以投入工程实用. 武长青等人[4]将约束机动问题归纳为非凸二次约束二次规划问题, 利用线性松弛结合评价函数进行迭代, 求出姿态优化最优解. Xu等人[5]提出了一种基于动态迭代的多目标规划(DIMP)方法, 但该方法将约束线性化处理, 放宽了对路径中间节点高精度的要求. Kjellberg等人[6]和Tanygin[7]分别使用A*算法寻找远离姿态禁区的

引用格式: 华冰, 孙胜刚, 吴云华, 等. 一种基于PIO改进强化学习的航天器复杂多约束姿态机动规划新方法. 中国科学: 技术科学, 2023, 53: 200–209
Hua B, Sun S G, Wu Y H, et al. A spacecraft attitude maneuvering path planning method based on PIO-improved reinforcement learning (in Chinese). Sci Sin Tech, 2023, 53: 200–209, doi: 10.1360/SST-2021-0346

最优路径, 但是A*算法在搜索前期效率较差, 容易使航天器做无效的姿态机动.

另一种思路是确定性解析法, 目前该思路的主流方法是势函数法. 武长青等人[8]提出了一种基于对数势函数的多约束姿态机动规划方法. 冯振欣等人[9]在势函数的基础上, 引入自适应干扰估计律, 增强了控制器的鲁棒性. 马广富等人[10]设计了新型的凸势函数, 并设计了基于反步法的控制率. 势函数方法存在固有不足: 极易陷入局部最小值, 且存在目标附近目标不可及问题(goals nonreachable with obstacles nearby)[11]. 相关研究大多简化了排斥势函数的存在条件, 参数设置不当, 易造成不必要的姿态机动[12]. 目前, 国内外对于姿态机动问题的势函数研究大多停留在虚拟空间内静态的指向约束, 约束数量也仅限于2~3个.

将姿态规划和传统路径规划进行相比, 在轨的航天器的姿态规划实际上是姿轨耦合问题, 而且姿态运动的三自由度耦合性较高, 相互影响较大. 另外, 航天器的姿态机动规划相对于路径规划来说, 机动能力较弱但精度要求更高, 这对控制机构和规划算法提出了更严格的要求.

近年来, 以机器学习为代表的人工智能技术在航天领域取得了极大的应用与进展. 强化学习是机器学习的范式之一, 不需要复杂繁琐的问题建模过程, 不需要系统完全可知, 便于解决非线性问题[13]. 多约束条件下航天器姿态机动规划问题, 属于非线性高维度的最优化问题, 适合运用强化学习求解. 考虑到星载计算机的计算能力, 其中基于策略梯度的强化学习算法计算量较小, 适合航天器使用, 但是存在策略梯度收敛速度慢的缺点.

本文选择群体智能优化算法中的鸽群算法(pigeon-inspired optimization, PIO)来计算策略梯度, 进一步减少强化学习算法的计算量, 加快收敛. 本文研究了多姿态约束条件下的航天器姿态机动规划问题. 针对现有姿态机动规划模型复杂、通用性较差和求解精度较差等问题, 提出了一种基于PIO改进强化学习的航天器姿态机动规划方法(PIOPGRL). 仿真结果表明, 在基本策略梯度强化学习算法的基础上, 引入鸽群算法计算策略梯度, 规划结果成功规避多个动态姿态约束区域, 不仅大幅度降低了计算量, 同时得到了更好的规划结果.

2　问题模型构建

本文为复杂多约束条件下的航天器保持低可见性制定姿态机动策略, 要求航天器在满足太阳能发电的对日定向姿态要求下, 通过姿态机动使自身携带的敏感器规避姿态禁区. 航天器所面临的姿态约束分为强制指向约束和禁止指向约束.

2.1　姿态模型

航天器本体系$O\mathbf{x}_B\mathbf{y}_B\mathbf{z}_B$定义为: O为坐标系原点, 位于航天器质心. $\mathbf{x}_B$轴, $\mathbf{y}_B$轴和$\mathbf{z}_B$轴分别与航天器的三个惯性主轴重合.

质心轨道坐标系$O\mathbf{x}_O\mathbf{y}_O\mathbf{z}_O$定义为: 坐标系原点位于航天器质心, $\mathbf{x}_O$轴指向地心, $\mathbf{y}_O$轴在轨道平面内, 与$\mathbf{z}_O$轴垂直并且指向航天器飞行的方向.

本文使用姿态角描述航天器姿态, 姿态角包括滚转角、俯仰角和偏航角, 分别代表航天器绕$\mathbf{x}_B$轴, $\mathbf{y}_B$轴和$\mathbf{z}_B$轴逆时针旋转的角度(图1).

2.2　姿态约束模型

姿态约束包括强制指向约束和禁止指向约束两类.

强制指向约束要求包括航天器对日的能量获取约束和对地指向约束. 本文中航天器的太阳能帆板朝向与$-\mathbf{y}_B$轴一致, 通信天线朝向与$\mathbf{z}_B$轴一致.

能量获取约束要求在航天器本体系中, $-\mathbf{y}_B$轴和太阳位置矢量$\mathbf{R}_{\text{sun}}$的夹角小于α_1:

$$\cos\langle -\mathbf{y}_B, \mathbf{R}_{\text{sun}}\rangle = \frac{-\mathbf{y}_B\cdot\mathbf{R}_{\text{sun}}}{|\mathbf{R}_{\text{sun}}|} > \cos(\alpha_1). \tag{1}$$

对地指向约束要求在航天器本体系中, $\mathbf{z}_B$轴和地心位置矢量$\mathbf{R}_{\text{earth}}$的夹角小于$\alpha_2$:

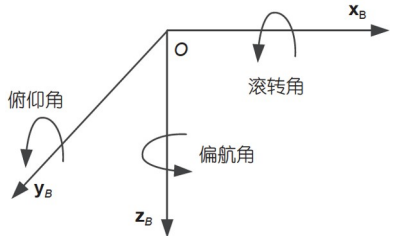

图 1　姿态角示意图

Figure 1　Schematic diagram of attitude angle.

$$\cos\langle \mathbf{z}_B, \mathbf{R}_{\text{earth}} \rangle = \frac{\mathbf{z}_B \cdot \mathbf{R}_{\text{earth}}}{|\mathbf{R}_{\text{earth}}|} > \cos(\alpha_2). \tag{2}$$

禁止指向约束针对敏感器禁止指向，某些星载敏感器在工作时要求规避强光强热和抑制杂光，因此对于航天器来说存在敏感器姿态禁区．本文假设敏感器中心轴的指向代表敏感器指向，敏感器姿态禁区是圆形区域，航天器质心与该圆形区域圆心的连线定义为敏感器禁止指向.

禁止指向约束要求：敏感器中心轴矢量与敏感器禁止指向矢量之间的夹角大于最小约束角．设第i个敏感器的中心轴指向在航天器本体坐标系下的位置矢量为$\mathbf{r}^B_{f-i}$，敏感器禁止指向j相对于航天器的位置矢量在航天器本体坐标系下表示为$\mathbf{r}^B_{m-j}$，本文要求$\mathbf{r}^B_{f-i}$和$\mathbf{r}^B_{m-j}$的夹角大于最小约束角β:

$$\cos\left\langle \mathbf{r}^B_{f-i}, \mathbf{r}^B_{m-j} \right\rangle = \frac{\mathbf{r}^B_{f-i} \cdot \mathbf{r}^B_{m-j}}{\left|\mathbf{r}^B_{f-i}\right|\left|\mathbf{r}^B_{m-j}\right|} < \cos\beta. \tag{3}$$

3 基于PIO改进的策略梯度强化学习姿态机动规划方法

强化学习可分为三类，包括基于策略、基于价值、基于策略和价值．考虑到星载计算机的计算能力，本文采用的基于有限差分的策略梯度方法(PGRL)属于基于策略的强化学习算法.

3.1 鸽群算法

鸽群算法[14]是受到鸽群在归巢中的导航过程启发而创造的，算法包括地图指南针算子和地标算子.

在地图指南针算子阶段，鸽群中每个个体通过当前种群中的最优解个体与自身的位置和速度进行更新，地图指南针算子公式如下:

$$\mathbf{V}_i(t) = \mathbf{V}_i(t-1) \cdot \mathrm{e}^{-Rt} + rand \cdot (\mathbf{X}_g - \mathbf{X}_i(t-1)), \tag{4}$$

$$\mathbf{X}_i(t) = \mathbf{X}_i(t-1) + \mathbf{V}_i(t), \tag{5}$$

式中，t是迭代次数，R是地图和指南针因子，是一个[0,1]的常数．$rand$是[0,1]的随机数，$\mathbf{V}_i(t)$和$\mathbf{X}_i(t)$分别是个体i在第t代的速度和位置，$\mathbf{X}_g$是当前种群所有个体的最佳位置.

在地标算子阶段，鸽群跟随熟悉地标的精英个体飞行，不熟悉地标的个体将被逐渐舍弃，鸽群的中心位置成为个体速度的参考方向．地标算子的数学表达式如下:

$$N_p(t) = \frac{N_p(t-1)}{2}, \tag{6}$$

$$\mathbf{X}_c = \frac{\sum_{i=1}^{N_p} \mathbf{X}_i(t) \cdot fitness(\mathbf{X}_i(t))}{N_p \sum_{i=1}^{N_p} fitness(\mathbf{X}_i(t))}, \tag{7}$$

$$\mathbf{X}_i(t) = \mathbf{X}_i(t-1) + rand \cdot (\mathbf{X}_c(t) - \mathbf{X}_i(t-1)),$$

式中，$N_p(t)$为第t次迭代的个体数目，$\mathbf{X}_c$是剩余鸽群的中心位置，$fitness(\mathbf{X}_i(t))$是个体i在第t次迭代时的适应度函数.

3.2 基于PIO改进的策略梯度强化学习

策略梯度强化学习的基本思想[15]是基于策略价值函数对的策略进行优化，经过策略的多次迭代逐步逼近并求出最优解．多次迭代评估策略耗时较多，而基于PIO改进的策略梯度强化学习方法使用鸽群算法评估并改进策略，大大加快了收敛速度并且能够探索到更好的策略，提高求解精度.下面阐述使用基于PIO改进的策略梯度强化学习方法进行航天器姿态机动规划的基本步骤.

为建立强化学习数据库，首先定义两个辅助坐标系$O\mathbf{x}_{\text{earth}}\mathbf{y}_{\text{earth}}\mathbf{z}_{\text{earth}}$和$O\mathbf{x}_{\text{sun}}\mathbf{y}_{\text{sun}}\mathbf{z}_{\text{sun}}$，定义分别如式(8)和(9)所示.

$$\begin{cases} \mathbf{z}_{\text{earth}} = \mathbf{z}_O, \\ \mathbf{y}_{\text{earth}} = \mathbf{z}_{\text{earth}} \otimes \mathbf{x}_B, \end{cases} \tag{8}$$

$$\begin{cases} \mathbf{y}_{\text{sun}} = -\dfrac{\mathbf{R}_{\text{sun}}}{|\mathbf{R}_{\text{sun}}|}, \\ \mathbf{z}_{\text{sun}} = \mathbf{x}_B \otimes \mathbf{y}_{\text{sun}}, \end{cases} \tag{9}$$

式中，$\mathbf{R}_{\text{sun}}$为太阳相对航天器的位置矢量，$\otimes$代表向量叉乘.

以不同的对日、对地定向重要性考虑上述两个辅助坐标系，则有:

$$\begin{cases} \mathbf{y}_c = \dfrac{\kappa_{\text{sun}}\mathbf{y}_{\text{sun}} + \kappa_{\text{earth}}\mathbf{y}_{\text{earth}}}{|\kappa_{\text{sun}}\mathbf{y}_{\text{sun}} + \kappa_{\text{earth}}\mathbf{y}_{\text{earth}}|}, \\ \mathbf{z}_c = \dfrac{\kappa_{\text{sun}}\mathbf{z}_{\text{sun}} + \kappa_{\text{earth}}\mathbf{z}_{\text{earth}}}{|\kappa_{\text{sun}}\mathbf{z}_{\text{sun}} + \kappa_{\text{earth}}\mathbf{z}_{\text{earth}}|}, \end{cases} \tag{10}$$

式中, $\kappa_{\text{earth}}, \kappa_{\text{sun}} \in [0,1]$为权重系数.

在式(8)~(10)的基础上最终建立强化学习数据库如式(12)所示:

$$\begin{cases} E1 = \arctan\dfrac{\mathbf{y}_c \cdot \mathbf{z}_O}{\mathbf{z}_c \cdot \mathbf{z}_O}, \\ E2 = \arcsin(-\mathbf{x}_B \cdot \mathbf{z}_O), \\ E3 = \arctan\dfrac{\mathbf{x}_B \cdot \mathbf{y}_O}{\mathbf{x}_B \cdot \mathbf{x}_O}, \end{cases} \tag{11}$$

$$\begin{cases} Ed1(n) = \dfrac{1}{N} \cdot n \cdot E1, \\ Ed2(n) = \dfrac{1}{N} \cdot n \cdot E2, \quad n = 1 \sim N, \\ Ed3(n) = \dfrac{1}{N} \cdot n \cdot E3. \end{cases} \tag{12}$$

因航天器有对日和对地定向需求, 本文以航天器不同权重满足对地或对日定向的姿态集组成N组数据的数据库. 本文使用姿态角描述航天器姿态, 故强化学习的策略u定为航天器的姿态角. 设当前强化学习迭代次数为k=1, 当前时刻m=1.

然后设定鸽群的种群数目为N_p, $N_p = N + k - 1$. 根据数据库, 种群中第i只鸽子的初始位置为$u_i(m)$, 则初始化种群分别为$u_1(m)$, $u_2(m)$, $\cdots$, $u_{N+k-1}(m)$.

考虑到鸽群算法的特点和强化学习的计算过程, 本文改进的关键是使用鸽群算法中的地标算子加速策略梯度的收敛, 并将强化学习概念中的策略评价函数选为鸽群算法的适应度函数. 适应度函数表达式为

$$fit[u_i(m)] = \sum_{n=1}^{m_f} \gamma(m) r[u_i(m)], \tag{13}$$

式中, m_f是终止时刻, $\gamma(m)$为强化学习概念中的折扣因子. $r[u_k(m)]$是策略的总回报函数, 与航天器对地和对日定向精度以及禁止指向姿态约束相关:

$$r[u_i(m)] = R_{d,k}(m) + R_{m-i-j,k}(m), \tag{14}$$

$$\begin{cases} R_{d,k}(m) = -\mu_{\text{earth}} \dfrac{1}{\cos^2\langle \mathbf{z}_{B,k}(m), \mathbf{z}_{O,k}(m) \rangle} \\ \qquad\qquad -\mu_{\text{sun}} \dfrac{1}{\cos^2\langle -\mathbf{y}_{B,k}(m), \mathbf{R}_{\text{sun},k}(m) \rangle}, \\ R_{m-i-j,k}(m) = -\sum\limits_{i=1}^{N_o} \sum\limits_{\text{j}=1}^{N_J} \mu_{f\text{-j-}i} \cos^2\langle \mathbf{r}^B_{f-i,k}(m), \mathbf{r}^B_{m\text{-j},k}(m) \rangle, \end{cases} \tag{15}$$

式中, 下标k表示当前迭代次数, $R_{d,k}(m)$为与对地和对日定向相关的回报函数, N_o代表航天器敏感器的个数, N_J代表姿态禁区个数. $R_{m-i-j,k}(m)$是第i个敏感器中心轴与第j个禁止指向的夹角的回报函数, μ_{f-j-i}为回报函数权重系数.

借助地标算子更新当前迭代次数的鸽群中心位置, 更新种群位置, 淘汰远离鸽群中心位置的鸽子后, 进入下一次迭代:

$$x_c^{(k)} = \frac{\sum\limits_{i=1}^{N_p} x_i^{(k)} fit\left[x_i^{(k)}\right]}{N_p \sum\limits_{i=1}^{N_p} fit\left[x_i^{(k)}\right]}, \tag{16}$$

$$x_i^{(k+1)} = x_i^{(k)} + rand \cdot \left(x_c^{(k)} - x_i^{(k)}\right), \tag{17}$$

$$N_p^{k+1} = fix\left(\frac{N_p^k}{2}\right), \tag{18}$$

式中, $fix(\cdot)$为取整函数.

当N_p = 1时, 鸽群算法停止, 计算此时刻的策略梯度:

$$G_k(m) = -sign\left[x_c^{(k)} - G_{k-1}(m)\right]. \tag{19}$$

更新强化学习的策略, 即得到航天器下一时刻的姿态角:

$$u_i(m+1) = u_i(m) + G_k(m). \tag{20}$$

重复式(13)~(20), 直到所有时刻的姿态角计算完毕, 即完成了一次完整的强化学习迭代过程.

图2是基于鸽群算法改进的策略梯度强化学习方法进行航天器姿态机动规划的流程图. 首先确定航天器所处的轨道和初始姿态, 设置多个敏感器在航天器上的位置, 同时计算多个姿态禁区在本体系中的坐标. 根据时间和轨道信息, 可以得到初始的太阳、地心相对于航天器的相对位置矢量, 根据对日对地定向需求, 就可以计算强化学习所需的数据库. 然后, 进行基于PIO改进的策略梯度强化学习过程, 最终得到多姿态约束条件下的姿态机动轨迹.

4　仿真实验与分析

本文中针对太阳同步轨道上的航天器进行仿真分析. 设置了4个动态的敏感器禁止指向, 太阳位置矢量

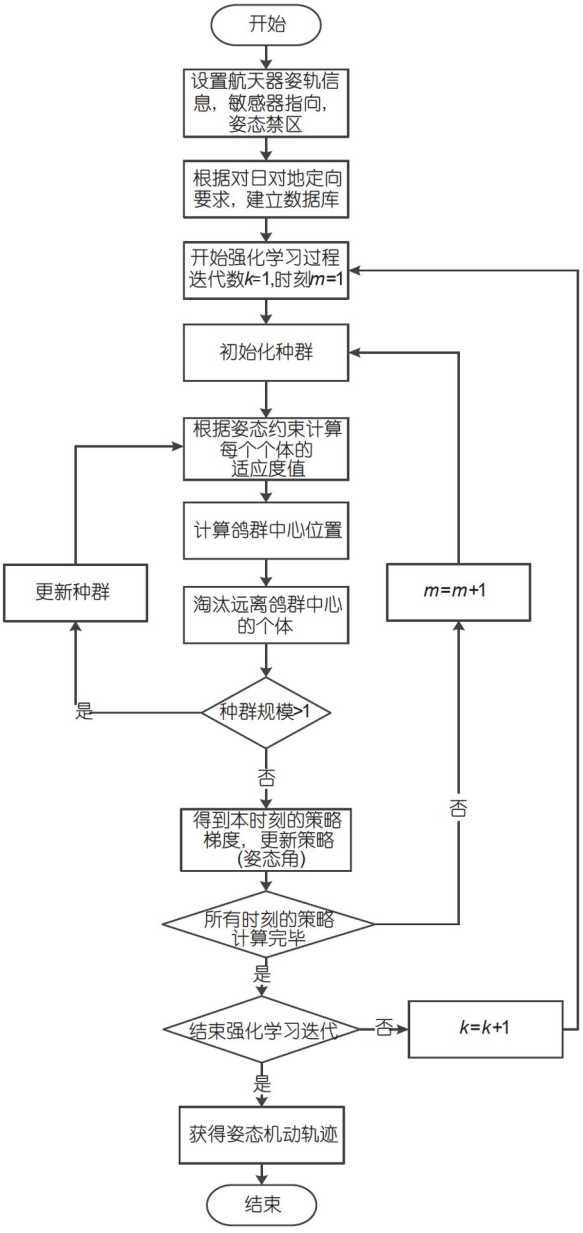

图 2 基于PIOPGRL的姿态机动规划流程图

Figure 2 Flow chart of the attitude maneuver planning based on PIOPGRL.

和敏感器禁止指向矢量均为航天器本体系下的单位矢量. 仿真时间为600 s, 强化学习迭代80次, 数据库包含数据数目N取20, 鸽群算法地标算子迭代20次. 折扣因子$\gamma(m)$取0.1, 对日定向权重μ_{sun}取3, 对地定向权重μ_{earth}取0.3, 回报函数权重系数$\mu_{f\text{-}j\text{-}l}$取20. 仿真初始条件如表1所示.

航天器初始姿态为$\mathbf{x}_B$轴指向航天器速度方向, $\mathbf{z}_B$轴指向地心, $\mathbf{y}_B$轴由右手规则确定. 航天器最大允许角速度的绝对值为0.8°/s, 最大允许角加速度绝对值为0.02°/s. $-\mathbf{y}_B$轴对日定向精度要求在30°以内, $\mathbf{z}_B$轴对地定向精度要求在10°以内, 敏感器指向与禁止指向的夹角要求大于3°.

表 1 仿真初始条件

Table 1 Initial conditions for simulation

条件参数	参数值
仿真起始时刻	9 Aug 2022 08:24:10.000
半长轴 (km)	6978.14
偏心率	0
轨道倾角 (°)	97.509
升交点赤径 (°)	227.538
近地点幅角 (°)	0
真近点角 (°)	37.957
太阳位置矢量初始值	(0.299, −0.917, −0.262)
禁止指向1初始值	(0.167, −0.904, 0.392)
禁止指向2初始值	(−0.310, −0.857, 0.409)
禁止指向3初始值	(0.133, −0.891, 0.432)
禁止指向4初始值	(0.818, −0.297, 0.491)

航天器携带8个敏感器, 所有敏感器指向在航天器本体坐标系中的单位矢量分别为

$$\begin{cases}
\mathbf{r}_{f\text{-}1}^{B}=\left[0\quad \sin(-50\,\text{deg})\quad \cos(-50\,\text{deg})\right]^{\mathrm{T}},\\
\mathbf{r}_{f\text{-}2}^{B}=\left[0\quad \sin(-25\,\text{deg})\quad \cos(-25\,\text{deg})\right]^{\mathrm{T}},\\
\mathbf{r}_{f\text{-}3}^{B}=\left[0\quad \sin(25\,\text{deg})\quad \cos(25\,\text{deg})\right]^{\mathrm{T}},\\
\mathbf{r}_{f\text{-}4}^{B}=\left[0\quad \sin(50\,\text{deg})\quad \cos(50\,\text{deg})\right]^{\mathrm{T}},\\
\mathbf{r}_{f\text{-}5}^{B}=\mathbf{C}_z(-37\,\text{deg})\mathbf{C}_y(-37\,\text{deg})[1\quad 0\quad 0]^{\mathrm{T}},\\
\mathbf{r}_{f\text{-}6}^{B}=\mathbf{C}_z(37\,\text{deg})\mathbf{C}_y(-37\,\text{deg})[1\quad 0\quad 0]^{\mathrm{T}},\\
\mathbf{r}_{f\text{-}7}^{B}=\mathbf{C}_z(-149\,\text{deg})\mathbf{C}_y(-138\,\text{deg})[1\quad 0\quad 0]^{\mathrm{T}},\\
\mathbf{r}_{f\text{-}8}^{B}=\mathbf{C}_z(149\,\text{deg})\mathbf{C}_y(-138\,\text{deg})[1\quad 0\quad 0]^{\mathrm{T}},
\end{cases}\tag{21}$$

式中, $\mathbf{C}_z(\theta)$和$\mathbf{C}_y(\theta)$分别代表绕$\mathbf{z}_B$轴和$\mathbf{y}_B$轴旋转θ角度, 逆时针旋转为正方向.

本文选择的仿真起始时刻选在2022年8月, 该月份$-\mathbf{y}_B$轴和太阳位置矢量$\mathbf{R}_{\text{sun}}$的夹角较大, 在30°左右. 选择此时刻作为仿真起始时刻, 对比策略梯度强化学习(PGRL), 更能体现本文提出的基于鸽群算法改进的策略梯度强化学习算法的有效性.

图3是使用策略梯度强化学习算法得到的姿态机动规划结果. 图4是8个敏感器的指向与4个禁止指向之间的夹角. 图5是在航天器在图2的姿态轨迹下, $-\mathbf{y}_B$轴与太阳矢量的夹角以及$\mathbf{z}_B$轴与地心矢量的夹角. 图6是航天器在图2的姿态轨迹下的姿态角速度以及角加速度.

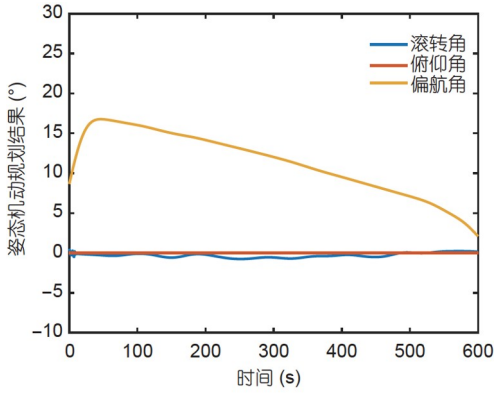

图 3　策略梯度强化学习姿态规划结果(PGRL)

Figure 3　The attitude planning result of policy gradient reinforcement learning (PGRL).

图7是使用策略梯度强化学习算法得到的姿态机动规划结果, 图8分别是8个敏感器的指向与4个禁止指向之间的夹角. 图9是在航天器在图7的姿态轨迹下, $-\mathbf{y}_B$轴与太阳矢量的夹角以及$\mathbf{z}_B$轴与地心矢量的夹角. 图10是航天器在图7的姿态轨迹下的姿态角速度以及角加速度.

表2所示为PGRL和PIOPGRL结果对比, 可以看到, 两种算法得到的规划结果中, 敏感器与禁止指向最小夹角都在5°以上, 满足了禁止指向姿态约束. 策略梯度强化学习算法的对日定向和对地定向精度较高, 极大地满足了太阳能发电和对地通信需求; PIOPGRL的这两项指标也是满足要求的.

值得注意的是, 航天器需要进行姿态机动来规避姿态禁区的时间较短. 在规避过程中的短时间内, 对

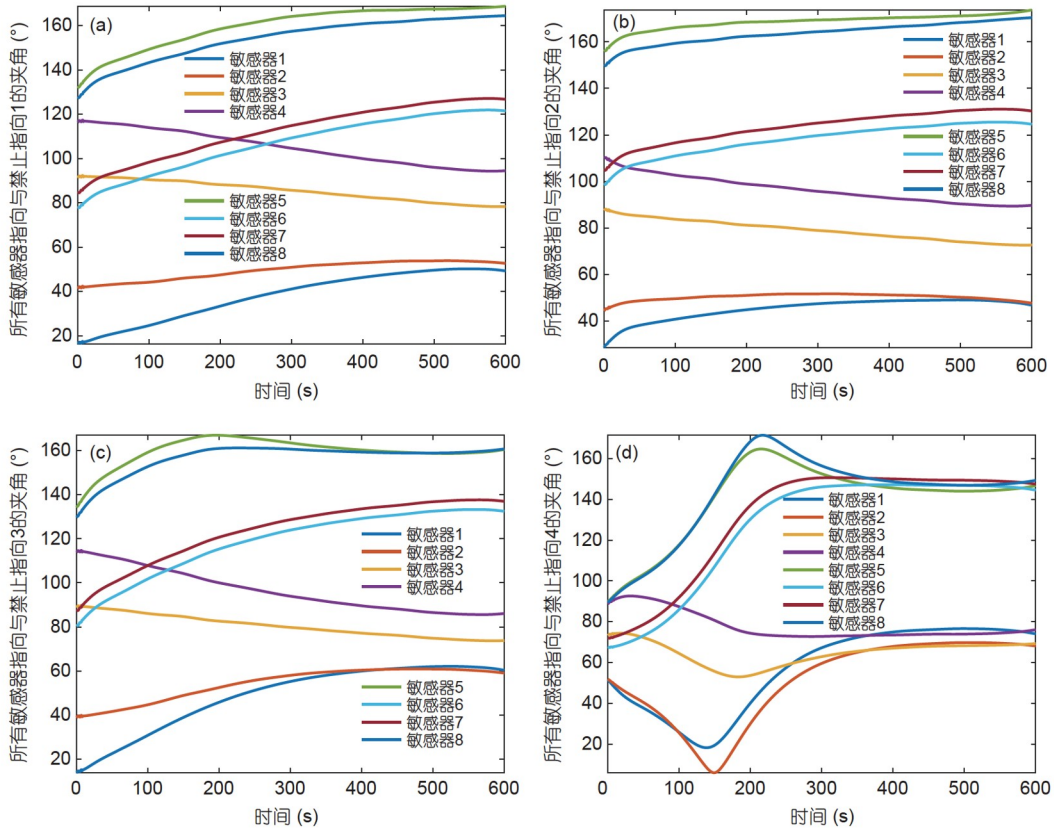

图 4　敏感器指向与禁止指向1 (a), 2 (b), 3 (c)和4 (d)的夹角(PGRL)

Figure 4　The angles between the sensor's pointing and the prohibited pointing 1 (a), 2 (b), 3 (c) and 4 (d) (PGRL).

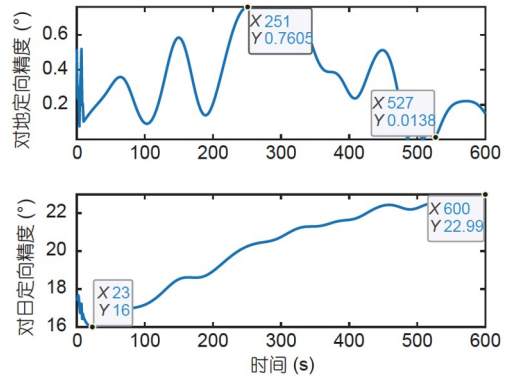

图 5 航天器对地对日精度(PGRL)

Figure 5 Orientation accuracy of spacecraft to the earth and to the sun (PGRL).

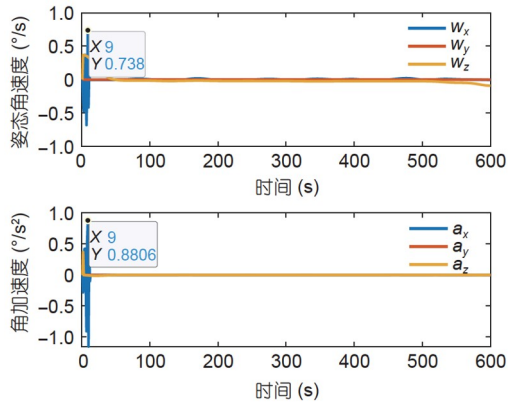

图 6 角速度和角加速度(PGRL)

Figure 6 Angular velocity and angular acceleration (PGRL).

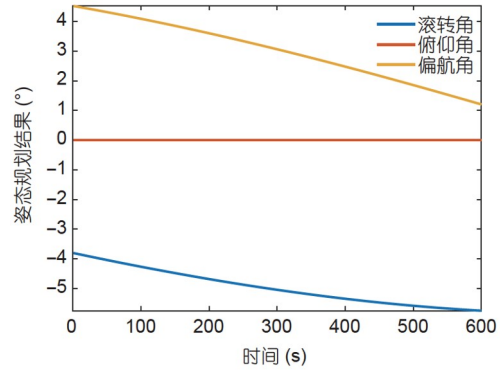

图 7 PIO改进强化学习姿态规划结果(PIOPGRL)

Figure 7 The attitude planning result of PIO improved reinforcement learning.

日定向和对地定向精度不是最重要的约束条件，最需要关注的约束是敏感器与禁止指向最小夹角，以及航天器为规避姿态禁区付出的机动代价. PGRL的姿态机动规划结果显示，偏航角的变化幅值达到了14.7°，而PIOPGRL的姿态角变化幅值仅仅为3.4°. 由于PGRL的规划结果进行了大幅度的姿态机动，其最大角速度绝对值为0.738°/s，已经接近0.8°/s的最大允许角速度; 其最大角加速度0.881°/s^{2}已经大大超过了最大允许角加速度.

表 2 PGRL和PIOPGRL结果对比

Table 2 Comparison of results of PGRL and PIOPGRL

算法	PGRL	PIOPGRL
敏感器与禁止指向最小夹角 (°)	5.962	5.411
姿态变化幅值 (°)	14.7	3.4
对日定向精度 (°)	15~22	23~28
对地定向精度 (°)	0.001~0.761	3.8~5.7
最大角速度绝对值 (°/s)	0.738	0.006
最大角加速度绝对值 (°/s^2)	0.881	0.004
计算时间 (s)	30	16

相比之下，PIOPGRL姿态规划结果的最大角速度绝对值为0.006°/s, 最大角加速度绝对值为0.004°/s^2, 均远远小于最大允许值. PIOPGRL在保证规避姿态禁区的同时, 付出的机动代价较小, 意味着航天器可以避免无效的姿态调整. 并且角速度和角加速度绝对值较小, 可见整个姿态机动过程比较平滑，适合微小航天器使用.

本仿真算例的运算环境为Intel i5-8300H CPU 2.30GHz, 8G RAM. 同等条件下, PGRL的运算时间在30 s左右，而PIOPGRL的运算时间仅16 s左右，运算时间几乎减少了一半. 本文设计的PIOPGRL算法，是航天器用于自主调用的算法. 由于规划必须先于执行，所以一般是提前一个轨道周期完成规划任务，如果有紧急规划任务，只需要提前10 min左右规划即可. 所以, 本文所设计的算法完全满足实时性要求.

5 总结

本文针对多个姿态约束条件下的航天器姿态机动规划问题进行了研究，提出了一种基于鸽群算法的改进的策略梯度强化学习算法(PIOPGRL). 保证了航天

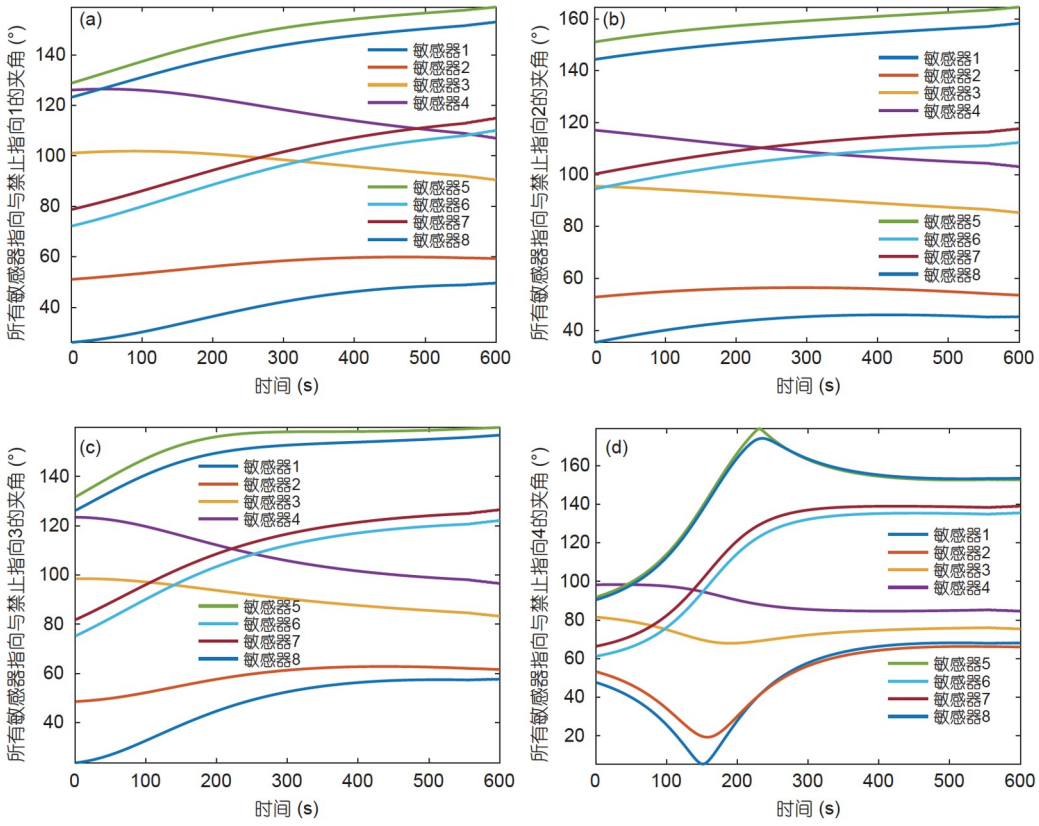

图 8　敏感器指向与禁止指向1 (a), 2 (b), 3 (c)和4 (d)的夹角(PIOPGRL)

Figure 8　The angles between the sensor's pointing and the prohibited pointing 1 (a), 2 (b), 3 (c)和4 (d) (PIOPGRL).

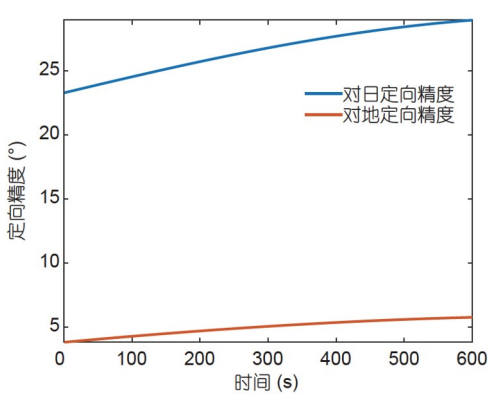

图 9　航天器对地对日精度(PIOPGRL)

Figure 9　Orientation accuracy of spacecraft to the earth and to the sun (PIOPGRL).

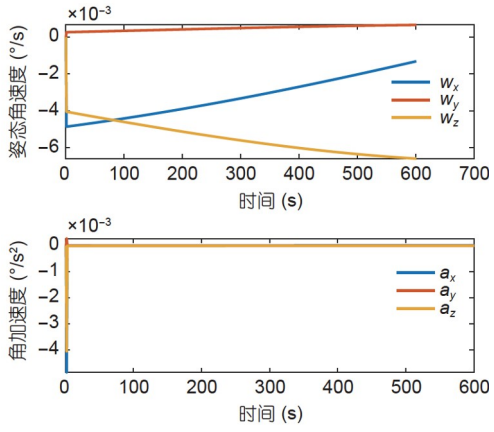

图 10　角速度和角加速度(PIOPGRL)

Figure 10　Angular velocity and angular acceleration (PIOPGRL).

器上多个敏感器规避成功多个姿态禁区，同时满足特定方向对日和对地的需求，大大减少了策略梯度强化学习的计算时间. 整个姿态机动过程满足最大角速度和最大角加速度约束要求，姿态角变化幅度小，调整过程平滑，非常适合星上计算能力和姿态机动能力有限的微小航天器使用.

参考文献

1 Hablani H B. Attitude commands avoiding bright objects and maintaining communication with ground station. J Guid Control Dynam, 1999, 22: 759–767

2 Singh G, Macala G, Wong E, et al. A constraint monitor algorithm for the Cassini spacecraft. Report. AIAA-1997-3526, American Institute of Aeronautics, 1997

3 Kim Y, Mesbahi M. Quadratically constrained attitude control via semidefinite programming. IEEE Trans Automat Contr, 2004, 49: 731–735

4 Wu C Q, Xu R, Zhu S Y, et al. Spacecraft attitude maneuver path iterative planning method under nonconvex quadratic constraints (in Chinese). J Astronaut, 2016, 37: 671–678 [武长青, 徐瑞, 朱圣英, 等. 非凸二次约束下航天器姿态机动路径迭代规划方法. 宇航学报, 2016, 37: 671–678]

5 Xu R, Wang H, Zhu S, et al. Multiobjective planning for spacecraft reorientation under complex pointing constraints. Aerospace Sci Tech, 2020, 104: 106002

6 Kjellberg H C, Lightsey E G. Discretized quaternion constrained attitude pathfinding. J Guid Control Dynam, 2016, 39: 713–718

7 Tanygin S. Fast autonomous three-axis constrained attitude pathfinding and visualization for boresight alignment. J Guid Control Dynam, 2017, 40: 358–370

8 Wu C Q, Xu R, Zhu S Y. Deep space explorer attitude planning and control method based on logarithmic potential function (in Chinese). J Deep Space Explor, 2015, 2: 365–370 [武长青, 徐瑞, 朱圣英. 基于对数势函数的深空探测器姿态规划与控制方法. 深空探测学报, 2015, 2: 365–370]

9 Feng Z X, Guo J G, Zhou J. Path maneuver planning for a microsatellite with multiple constraints (in Chinese). J Astronaut, 2019, 40: 1205–1211 [冯振欣, 郭建国, 周军. 微小卫星多约束姿态机动规划方法. 宇航学报, 2019, 40: 1205–1211]

10 Ma G F, Liu M M, Wang L Y, et al. Spacecraft backstepping attitude control considering multiple forbidden pointing regions (in Chinese). J Astronaut, 2020, 41: 1042–1048 [马广富, 柳明旻, 王靓玥, 等. 考虑多禁止指向区域的航天器反步姿态机动控制. 宇航学报, 2020, 41: 1042–1048]

11 Hu Q, Chi B, Akella M R. Anti-unwinding attitude control of spacecraft with forbidden pointing constraints. J Guid Control Dynam, 2019, 42: 822–835

12 Chen T D, Huang Y Y, Zhang Y L. Non-trap dynamic path planning based on collision risk (in Chinese). Syst Eng Electron, 2019, 41: 2496–2506 [陈天德, 黄炎焱, 张永亮. 基于碰撞危险度的无陷阱动态航路规划. 系统工程与电子技术, 2019, 41: 2496–2506]

13 Huang X X, Li S, Yang B, et al. Review of spacecraft guidance and control based on artificial intelligence (in Chinese). Acta Aeronaut Astronaut Sin, 2021, 42: 106–121 [黄旭星, 李爽, 杨彬, 等. 人工智能在航天器制导与控制中的应用综述. 航空学报, 2021, 42: 106–121]

14 Duan H, Qiao P. Pigeon-inspired optimization: A new swarm intelligence optimizer for air robot path planning. Int J Intelligent Computing Cybernetics, 2014, 7: 24–37

15 Liu J W, Gao F, Luo X L. Survey of deep reinforcement learning based on value function and policy gradient (in Chinese). Chin J Comput, 2019, 42: 1406–1438 [刘建伟, 高峰, 罗雄麟. 基于值函数和策略梯度的深度强化学习综述. 计算机学报, 2019, 42: 1406–1438]

A spacecraft attitude maneuvering path planning method based on PIO-improved reinforcement learning

HUA Bing, SUN ShengGang, WU YunHua & CHEN ZhiMing

College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Aiming at the problem of spacecraft attitude maneuver planning under multiple mandatory pointing constraints and prohibited pointing constraints, based on pigeon-inspired optimization (PIO), we proposed an improved policy gradient reinforcement learning (RL) algorithm (PIOPGRL). First, we establish an angle-based attitude constraint model, and then, we establish the reward function of RL based on the model. Then, the fitness function is used to replace the policy evaluation function, so PIOPGRL is integrated with RL. The PIOPGRL algorithm uses the PIO algorithm to solve the policy gradient, significantly reduces the amount of calculation and accelerating the convergence speed. The simulation results show that the spacecraft attitude maneuvering path planning method based on PIO-improved RL (PIOPGRL) has better planning results and lower cost of maneuver than the classical PGRL algorithm, which can solve the problem of spacecraft attitude maneuver planning under multiple pointing constraints perfectly.

attitude maneuver, attitude constraint, path planning, reinforcement learning, PIO algorithm, spacecraft

doi: 10.1360/SST-2021-0346