基于稀疏学习的连续型机械臂自适应控制器

江 达[1], 蔡志勤[1†], 刘忠振[1], 彭海军[1,2], 吴志刚[2]

(1. 大连理工大学 工程力学系，辽宁 大连 116024；2. 大连理工大学 工业装备国家重点实验室, 辽宁 大连 116024)

摘 要: 探讨了空间连续型机械臂执行在轨操作任务过程中的自适应轨迹跟踪控制器设计问题。首先，对于具有显著非线性特征的连续型机械臂动力学模型，考虑运动过程中存在的建模误差和外部干扰因素，设计了变结构动力学控制器。其次，基于深度强化学习(Deep reinforcement learning, DRL)对变结构控制器参数进行在线调整，实时优化控制器性能。同时，提出一种针对强化学习网络稀疏训练方法，训练过程中采用具有随机稀疏拓扑结构的稀疏连接层代替神经网络的全连接层，并以一定概率对连接薄弱的网络进行迭代剪枝，使DRL的策略网络由初始稀疏拓扑结构演化成无标度网络,在不降低训练精度的基础上压缩了网络规模。仿真结果表明，所提出的基于强化学习的自适应控制器能有效地进行连续型机械臂的跟踪控制，通过稀疏学习的方法，控制器在保证控制精度的同时，双隐层网络节点参数量下降了99%，大幅降低了计算成本。

关键词: 空间机械臂；连续型机械臂；动力学控制；强化学习；稀疏学习；自适应控制

中图分类号: TP273.2 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2021.1936

引用格式 : 江达，蔡志勤，刘忠振，彭海军，吴志刚. 基于稀疏学习的连续型机械臂自适应控制器[J]. 控制与决策.

Sparse-learning-based Adaptive Controller for the Space Continuum Manipulator

JIANG Da[1], *Cai Zhi-qin*[1†], *Liu Zhong-zhen*[1], *Peng Hai-jun*[1,2], *Wu Zhi-gang*[2]

(1. Engineering Mechanics, Dalian University of Technology, Liaoning Dalian 116024, China；2. State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Liaoning Dalian 116024, China)

Abstract: In this paper, an adaptive tracking controller is designed for the space continuum manipulator in the on-orbit manipulation task. Firstly, considering the modeling error and external disturbance factors, a variable structure controller is designed for the continuum manipulator's dynamic model with typical nonlinear character. Then, a deep reinforcement learning (DRL)algorithm is adapted to adjust the controller's parameters and to optimize the control performance online. In addition, a sparse training method for DRL is proposed. In the training process, the origin network's fully-connected layer is replaced with sparsely-connected layer in a random sparse topology. The weak connections are iteratively pruned in a certain probability, which evolves DRL's policy network from an initial sparse topology into a scale-free network. Therefore, the network's dimensions are significantly compressed without reducing the training accuracy. Simulation results show that the proposed DRL-based adaptive controller can effectively carry out the tracking control of the continuum manipulator. Through the sparse training method, the quantity of the two hidden layers'parameters is reduced by 99% on the premise of maintenance of the control accuracy. The calculation is effectively reduced.

Keywords: space manipulator；continuum manipulator；dynamics control；reinforcement learning；sparse training；adaptive control

0 引 言

随着空间探索研究的深入，在轨作业的任务要求和难度不断提高，机械臂已逐渐成为空间操作研究的主流，在空间主动碎片清除、在轨维修及装配等任务中具有广阔的应用前景[1-2]。

相较于传统刚性机械臂，连续型机械臂具有灵活柔软特点，可以通过主动变形在受限的空间内完成复杂的动作。研究表明连续型机械臂在医疗手

收稿日期: 2021-11-09；**录用日期:** 2022-05-17.
基金项目: 国家自然科学基金项目(91748203),国家自然科学基金优秀青年项目(11922203).

†通讯作者. E-mail: zhqcai@dlut.edu.cn.

术[2-4]、空间操作[5-7]等复杂工作环境中具有独特的优势。然而，在具有实时、高精度需求的空间在轨操作场景中，连续型机器人具有的非线性动力学特征增加了对应控制器的设计难度。

近年来，学者们进行了大量的关于连续型机械臂跟踪控制的研究，此类非线性动力学系统的控制问题，对控制系统的鲁棒性和实时性等性能提出了较高的要求，传统的运动学控制很难满足高精度的复杂操作。Melingui[8]为机器人设计了的自适应神经网络控制器，实时跟踪所需路径。该方法采用多层感知器离线学习最优控制器并进行在线控制，但在部分情况下出现了跟踪失效、调节时间过长的现象。Yip[9]提出了一种肌腱驱动机器人的任务空间闭环控制器，通过实时估计机器人运动学参数，使其以稳定的方式与受约束的环境交互，但跟踪精度较低。Morteza[10]利用最优控制框架，通过Cosserat杆理论推导出了空间连续型机器人的动力学控制方程并进行了实验验证，该连续型机器人模型为单臂纯弹性杆结构，自由度较低，运动形式较为简单，不适合拓展到更柔软灵活的连续型机器人控制问题中。机器学习运算速度快、训练方便，对高维数据的处理能力使其在基于动力学的控制问题中具有明显的优势。为进一步提升控制效果，大量工作通过引入机器学习方法来自适应调整控制器参数，以实时优化控制器性能。Tan[11]采用归零神经动力学网络对管状连续型机器人进行建模，并据此设计轨迹跟踪控制器，该方法不需要驱动器输入和传感器输出的信息。Frazelle[12]采用Actor-Critic框架的策略搜索方法实现了对连续型机器人进行运动学控制，该方法采用离散状态和动作空间下的学习框架，限制了控制精度，导致该方法不适用于被控模型更为复杂的动力学控制问题。

机器学习方法为自适应控制方法提供了思路，但往往需要较高的计算量来保证良好的性能。这限制了其在空间在轨操作环境中有限内存设备上的应用。稀疏训练在降低网络维度方面有良好的效果，它利用了深度神经网络参数的冗余性，对稀疏拓扑与网络权值进行并行优化[13]。同时相关工作[13-15]表明,经稀疏训练的网络具有保持了与原网络相当的精度。在本文工作中，我们将稀疏训练引入到空间连续型机械臂控制器设计中，提出一种针对强化学习的训练机制，在不降低精度的前提下，通过稀疏化神经网络的网络结构来降低训练过程的计算和内存成本，以完成空间连续型机械臂的在轨操作。

综上，本文针对有较高精度要求的连续型机器人空间运送零件的操作任务，基于平面3臂线驱动连续型机械臂模型，结合强化学习原理，设计了自适应动力学跟踪控制器，通过深度确定性策略梯度的强化学习方法实时调整控制器参数，来优化控制器能力。为在不影响精度的前提下进一步降低强化学习的计算成本，提出了一种神经网络稀疏训练方法来压缩网络规模，并应用到强化学习的策略网络中，最终实现空间主动操作过程中连续型机械臂的自适应跟踪控制。

1 空间连续型机械臂动力学模型

考虑一个做平面运动的3节串联空间线驱动连续型机械臂的运送零件的操作过程，机械臂以3D塑料制成的柔性支撑骨架位变形主体，各节臂等距地被固定节盘分成三小段，驱动线由节盘孔从底部电机连接至整个机械臂来驱动柔性支撑进行变形，末端执行器固定在尾部节盘处。假设机械臂已被送至目标点附近的工作区域内，机械臂操作任务为通过主动变形来使末端执行器所携带的零件运送至目标点。本文研究的平面连续型机械臂单节臂构型如图1。其中α_i为第i节臂的弯曲角($i=1,2,3$)，β_i和(x_i,y_i)分别为臂i所在局部坐标系$O_iX_iY_i$在全局坐标系OXY下的转角和坐标。

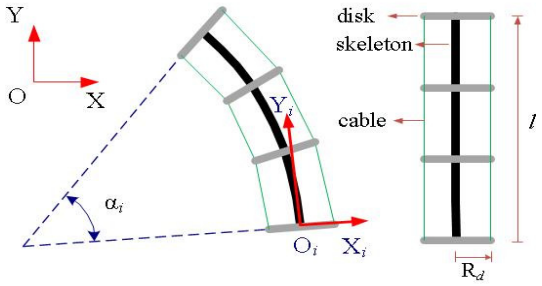


图 1 连续型机械臂单节构型

本文采用文献[16]描述的空间连续型机械臂动力学模型。设广义坐标$\boldsymbol{q}=[\alpha_1,\alpha_2,\alpha_3]^{\mathrm{T}}$，机械臂系统的质量阵为$\boldsymbol{M}$，将末端执行器计入到末端节盘处，则机械臂系统动能$T$由柔性支撑动能$T^s$和节盘动能$T^d$组成：

$$T=T^s+T^d=(1/2)\dot{\boldsymbol{q}}^{\mathrm{T}}\boldsymbol{M}\dot{\boldsymbol{q}} \tag{1}$$

连续型机械臂系统虚功由弹性力$\boldsymbol{Q}_e$和驱动力$\boldsymbol{Q}_a$对应的虚功组成：

$$\delta W=-\boldsymbol{Q}_e^{\mathrm{T}}\delta\boldsymbol{q}+\boldsymbol{Q}_a^{\mathrm{T}}\delta\boldsymbol{q} \tag{2}$$

由此可得空间连续型机械臂的动力学方程：

$$\boldsymbol{M}\ddot{\boldsymbol{q}}=-\boldsymbol{Q}_e+\boldsymbol{Q}_a+\boldsymbol{Q}_v \tag{3}$$

其中$\boldsymbol{Q}_v=-\dot{\boldsymbol{M}}\dot{\boldsymbol{q}}+\partial T/\partial \mathbf{q}$。为了更好地表示驱动力向量$\boldsymbol{\tau}=[\tau_1,\tau_2,\tau_3]^{\mathrm{T}}$，将式(3)整理为：

$$\boldsymbol{M}_0(\boldsymbol{q})\ddot{\boldsymbol{q}}+\boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})=\boldsymbol{\tau}+\boldsymbol{f}(t) \tag{4}$$

式中$\boldsymbol{M_0}(\boldsymbol{q})\in\boldsymbol{R}^{3\times3}$，$\boldsymbol{C_0}(\boldsymbol{q},\dot{\boldsymbol{q}})\in\boldsymbol{R}^{3\times1}$。$\boldsymbol{f}(t)\in\boldsymbol{R}^{3\times1}$为外部干扰$\boldsymbol{d}(t)$和建模误差$\boldsymbol{\Delta M}_0\ddot{\boldsymbol{q}}+\boldsymbol{\Delta C}_0$对应项：

$$\boldsymbol{f}(t)=\boldsymbol{d}(t)+\boldsymbol{\Delta M}_0\ddot{\boldsymbol{q}}+\boldsymbol{\Delta C}_0 \tag{5}$$

基于此，可以通过控制驱动力向量$\boldsymbol{\tau}$来改变各臂的弯曲角度，从而控制整个机械臂的运动。

2 空间连续型机器人变结构控制器

定义$\boldsymbol{q}_d$为期望轨迹向量，则机械臂系统跟踪误差为$\boldsymbol{e}=\boldsymbol{q}-\boldsymbol{q}_d$。定义切换函数为：

$$\boldsymbol{h}=\dot{\boldsymbol{e}}+\delta\boldsymbol{e}+\sigma\boldsymbol{e}^{p/q} \tag{6}$$

式中δ是正常数；p和q均为正奇数，$p>q$。由局部线性化原理可得：

$$\begin{aligned}\dot{\boldsymbol{h}}&=\ddot{\boldsymbol{e}}+\delta\dot{\boldsymbol{e}}+\sigma\boldsymbol{e}^{(p-q)/q}\,p/q;\\&=\boldsymbol{M}_0{}^{-1}(\boldsymbol{\tau}+\boldsymbol{f}-\boldsymbol{C}_0)-\ddot{\boldsymbol{q}}_d+\delta\dot{\boldsymbol{e}}+\sigma\boldsymbol{e}^{(p-q)/q}\,p/q\end{aligned} \tag{7}$$

因此本文提出的变结构跟踪控制律：

$$\boldsymbol{\tau}=\boldsymbol{M}_0[\ddot{\boldsymbol{q}}_d-\delta\dot{\boldsymbol{e}}-\sigma\boldsymbol{e}^{(p-q)/q}\,p/q\,]+\boldsymbol{C}_0-\boldsymbol{f}_c \tag{8}$$

式中$\boldsymbol{f}_c$为$\boldsymbol{f}$的估计向量。选择Lyapunov函数：

$$V=\frac{1}{2}\boldsymbol{h}^{\mathrm{T}}\boldsymbol{h} \tag{9}$$

由于$\dot{V}=\boldsymbol{h}^{\mathrm{T}}\dot{\boldsymbol{h}}$，将(9)代入(8)有

$$\begin{aligned}\dot{\boldsymbol{h}}&=\ddot{\boldsymbol{e}}+\delta\dot{\boldsymbol{e}}+\delta\boldsymbol{e}^{(p-q)/q}p/q\\&=\boldsymbol{M}_0^{-1}(\boldsymbol{\tau}+\boldsymbol{f}+\boldsymbol{C}_0)-\ddot{\boldsymbol{q}}_d+\delta\dot{\boldsymbol{e}}+\delta\boldsymbol{e}^{(p-q)/q}p/q\\&=\boldsymbol{M}_0^{-1}(\boldsymbol{f}-\boldsymbol{f}_c)\\\dot{V}&=\boldsymbol{h}^{\mathrm{T}}\dot{\boldsymbol{h}}=\boldsymbol{h}^{\mathrm{T}}\boldsymbol{M}_0{}^{-1}(\boldsymbol{f}-\boldsymbol{f}_c)\end{aligned}$$

定义$\bar{\boldsymbol{f}}_c=\boldsymbol{M}_0^{-1}\boldsymbol{f}_c$，$\bar{\boldsymbol{f}}=\boldsymbol{M}_0^{-1}\boldsymbol{f}$，假设$\bar{\boldsymbol{f}}$有界：

$$\bar{\boldsymbol{f}}\le\bar{\boldsymbol{f}}_u$$

通过选取$\bar{\boldsymbol{f}}_c=(\boldsymbol{\kappa}+\bar{\boldsymbol{f}}_u)sgn(\boldsymbol{h})(\boldsymbol{\kappa}>\boldsymbol{0})$，有：

$$\begin{aligned}\boldsymbol{h}^{\mathrm{T}}(\bar{\boldsymbol{f}}-\bar{\boldsymbol{f}}_c)&=\boldsymbol{h}^{\mathrm{T}}[\bar{\boldsymbol{f}}-(\boldsymbol{\kappa}+\bar{\boldsymbol{f}}_u)sgn(\boldsymbol{h})]\\&=\boldsymbol{h}^{\mathrm{T}}\bar{\boldsymbol{f}}-(\boldsymbol{\kappa}+\bar{\boldsymbol{f}}_u)\,|\boldsymbol{h}|\le-\boldsymbol{\kappa}\,|\boldsymbol{h}|\le0\end{aligned}$$

则$\dot{V}<0$。综上，系统稳定性得证。

3 基于稀疏训练强化学习的变结构控制器

根据变结构控制理论可知，适当调整其切换函数的参数，能够优化系统向切换面的渐近收敛速度，使系统快速、稳定地收敛到平衡状态。为了能够自适应地实时调节控制器的切换参数，本文选用δ和σ作为强化学习的调整参数，采用深度确定性策略梯度(Deep deterministic policy gradient, DDPG)的强化学习算法[17]，通过合理地设计网络结构和奖励函数，使得训练后的策略网络能根据当前的跟踪状态调整参数，从而改善控制器动态性能。考虑一个包含状态向量$\boldsymbol{s}$，动作向量$\boldsymbol{a}$，状态转移概率分布$\boldsymbol{p}(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t,\boldsymbol{a}_t)$和奖励函数$r$的标准马尔科夫过程。DDPG包含两套Actor网络和Critic网络，对应主网络参数θ^Q和θ^μ通过滑动平均法更新至目标网络的$\theta^{Q'}$和$\theta^{\mu'}$。主Critic网络$Q(\boldsymbol{s},\boldsymbol{a}|\theta^Q)$更新依据代价函数：

$$\begin{aligned}L&=\frac{1}{N_b}\sum\nolimits_{j=1}^{N_b}\left(y_{Qj}-Q(\boldsymbol{s}_j,\boldsymbol{a}_j|\theta^Q)\right)^2\\y_{Qj}&=r_j+\gamma min[Q'(\boldsymbol{s}_{j+1},\mu'(\boldsymbol{s}_{j+1}|\theta^{\mu'})|\theta^{Q'})]\end{aligned} \tag{10}$$

其中N_b为经验池$\boldsymbol{R}$中储存的批次数据量。主Actor网络$\mu(\boldsymbol{s}|\theta^\mu)$由确定性策略梯度原理更新：

$$\nabla_{\theta^\mu}J=\frac{1}{N_b}\sum_{i=1}^{N_b}\nabla_{\boldsymbol{a}}Q(\boldsymbol{s},\boldsymbol{a}|\theta^Q)|_{\boldsymbol{s}_i,\mu(\boldsymbol{s}_i)}\nabla_{\theta^\mu}\mu(\boldsymbol{s}|\theta^\mu)|_{\boldsymbol{s}_i}$$

如图2所示，稀疏化训练主要是基于神经网络的冗余性来简化对影响微弱的网络连接，最终获得低参数量的网络结构。

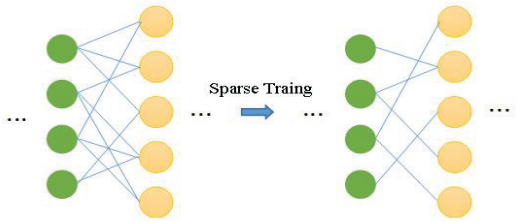


图 2 稀疏训练

在稀疏化训练过程中，首先采用基于ε的Erdòs-Rényi(ε-ER)随机图模型[18]初始化的稀疏连接层替代DDPG网络中的隐藏层。假设相邻的两个隐藏层L_a、L_b中神经元数量分别为h_a、h_b，则两层中各自任意神经元m、n之间的连接概率如式(11):

$$p_{mn}=\varepsilon\frac{h_a+h_b}{h_ah_b} \tag{11}$$

其中ε为表征连接稀疏程度的超参数。初始化过程中，稀疏网络的节点参数由原参数与二进制掩膜的乘积表示，则稀疏化训练就转化为动态寻找最优掩膜矩阵的问题。这样，在随后的训练过程中，在每经过t_d轮参数更新后，对系统影响较小网络连接（即参数绝对值极小的连接）进行剪枝操作，修剪比例为η，可通过归零掩膜矩阵对应项来实现。在这个过程中，每次优化时自适应调整剪枝参数比例，这样降低了计算量，同时使得网络的稀疏化逐渐稳

定。同时，在网络每次训练前的初始化过程中随机增加部分新的连接，以降低训练相近数据样本时删除重要连接的潜在风险，从而提高网络对数据的泛化能力。最终，通过不断迭代将强化学习的神经网络由随机网络演化成无标度网络，实现整个稀疏化流程。我们将稀疏化训练的方法应用到DDPG算法中，形成的基于稀疏训练的DDPG算法如算法1。并由此形成自适应变结构控制器，通过策略网络在线优化变结构控制器的控制参数。

算法1 稀疏训练的DDPG算法

基于ε-ER图随机初始化Actor/Critic稀疏网络
初始化目标网络和经验池$\boldsymbol{R}$
for episode=1 to $\mathrm{T_E}$ **do**:
  初始化随机过程N_a用于动作探索
  初始化环境，接收初始状态s_1
  for step=1 to $\mathrm{T_S}$ do:
    基于当前策略网络和N_a获得动作a_t
    执行动作，获得预期奖励r_t和状态s_{t+1}
    更新(s_t,a_t,r_t,s_{t+1})至经验池$\boldsymbol{R}$
    从经验池R选取随机批次数据
    通过最小化L来更新Critic网络
    根据确定性策略梯度更新Actor网络
    for each t_d steps:
      根据稀疏学习修剪Actor/Critic网络
    end for
  end for
end for

4 仿真校验

提出的基于DDPG的自适应变结构控制器控制流程如图3所示，其中Actor和Critic网络采用4层神经网络，隐藏层各自包含250个节点，隐藏层之间的稀疏连接由ε-ER图初始化。强化学习动作向量$\boldsymbol{a}$包括控制器调节参数δ和σ。状态向量$\boldsymbol{s}$信息包含连续型机械臂运动过程中各臂角度$\boldsymbol{q}$、角速度$\dot{\boldsymbol{q}}$、角度跟踪误差$\boldsymbol{e}$、角速度跟踪误差$\dot{\boldsymbol{e}}$及下一时刻预期角度$\boldsymbol{q}_d$、角速度$\dot{\boldsymbol{q}}_d$、角加速度$\dot{\boldsymbol{q}}_d$。奖励值r设置为：

$$r=-\Delta\left\|\boldsymbol{e}\right\|+g$$

式中g表示持续跟踪目标轨迹时给予的激励奖励。仿真过程中，当各臂角度跟踪期误差保持在0.02rad内超过25步时获得额外奖励。基于此，在控制初期，误差降低最快时奖励值最大；后续随着控制的推进，当控制跟踪持续趋于稳定时，奖励值最大，使得强化学习持续朝着快速而稳定的方向优化控制器参数。

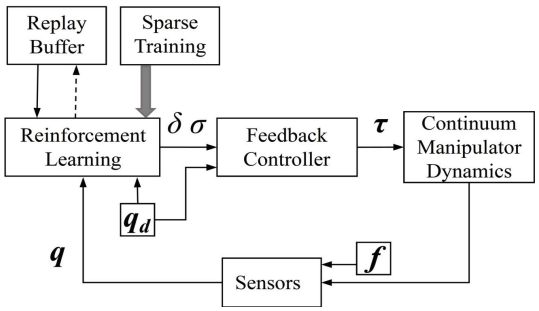


图 3 基于稀疏训练DDPG的自适应变结构控制器流程

表1 各节臂物理参数

符号	参数	值
m_0	质量	1kg
l	长度	0.2m
R_d	节盘半径	0.02m
I_z	惯性矩	$7.85\times10^{-9}\mathrm{kg\cdot m^2}$
E	弹性模量	$6.2\times10^{9}\mathrm{Pa}$

连续型机械臂的物理参数如表1。实际空间在轨操作环境中，存在着电磁干扰、观测噪声等外部干扰；同时，连续型机械臂的动力学模型也不可避免地存在建模误差，会导致动力学模型与实际操作时存在计算误差。这些噪声均会对在轨操作造成影响，降低控制器的控制精度。因此网络测试过程中，针对式(5)的建模误差和外部干扰采用文献[19]的有界噪声，并加入了高斯噪声$N(0,0.03)$，以更贴近在轨操作中的不确定性干扰。训练过程中，在策略网络输出处引入随机噪声来进一步增加智能体探索环境的能力。同时在控制过程中执行滚动优化方案来降低强化学习自适应调参问题中奖励迟滞现象[20]的影响，其中预测域值设置为5。训练持续600代，每代包括200个时间步，采样步长为0.1s，每代起始时随机初始化期望轨迹。前10代采用随机策略，仿真结果作为经验池初始数据。每经过600步进行稀疏化操作，初始修剪比例η为5%，为获得稳定稀疏训练过程，修剪比例随训练整体进程匀速收敛至零。训练过程采用Adam优化器，学习率为0.001。训练过程中网络隐藏层稀疏程度（定义为稀疏连接占总连接的比例）如图4示。可以看出经过一段时间迭代稀疏训练优化后，Actor和Critic网络的稀疏程度在不断加深，最终Actor和Critic主网络隐藏层间的网络连接分别被稀疏了99.2%和98.9%，至400代后稀疏程度趋于稳定，显示出网络参数得到了大幅压缩。

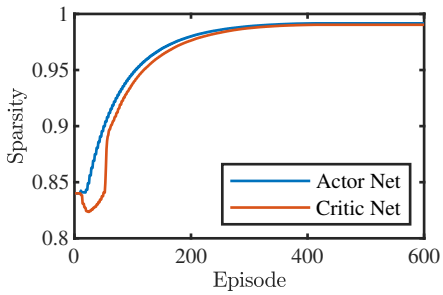


图 4 强化学习主网络平均稀疏度

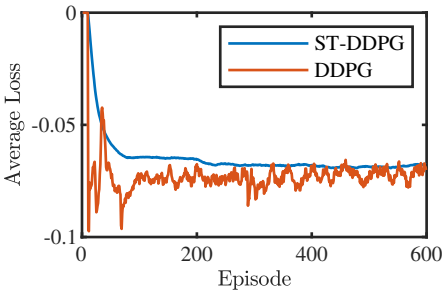


图 5 训练过程中平均Loss值

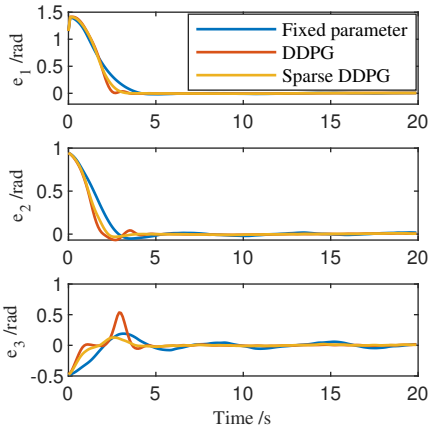


图 6 训练过程中平均Loss值

由于在控制过程中，机械臂可能很快跟踪上目标轨迹额外获得激励奖励值，平均奖励值不能直观地反映训练的效果，因此本文对比了经稀疏化与未经稀疏化的DDPG强化学习训练过程中每代策略网络Actor网络的平均Loss值，如图5所示（前10代为随机策略，未进行训练）。通过对比可以看出，两种方法训练过程中的平均Loss值逐渐降低，基于稀疏训练的DDPG算法(Sparse training-deep deterministic policy gradient, ST-DDPG)和未经稀疏训练的DDPG算法分别在250、400代后逐渐平稳，最终稳定在相近的平均loss值下，显示出本文提出的基于稀疏DDPG自适应控制器表现出了与未经稀疏化DDPG自适应控制器相近的训练效果。两种算法训练过程中的平均loss值存在一定的差异，这是由于自适应参数调整问题本身存在多解性，奖励迟滞的情况导致在部分情况下策略网络不可避免地陷入局部最优解，影响网络性能；同时所提出的稀疏化训练DDPG算法的参数更新和网络稀疏并行处理，而不是对训练好的策略网络进行再稀疏处理，因此两种方法的策略网络存在差异；此外，每代随机初始化的不同轨迹也对loss值有一定的影响。为检验网络参数规模大幅压缩后，网络性能是否得到保持，本文对比了所提出的控制器与基于DDPG的自适应变结构控制器、固定参数$(\delta,\sigma)=(2,2)$变结构控制器[21]的跟踪效果，如图6所示。假设一个连续型机械臂执行在轨操作的运送零件任务中，机械臂行进路径为如式(12)的随机轨迹。仿真结果表明，在存在外部扰动和建模误差的情况下，三种控制器都顺利地能完成连续型机械臂轨迹跟踪任务。但相比于固定参数变结构控制器所存在部分稳态误差的控制效果，两类基于强化学习自适应变结构控制器均于5s左右达到了稳定的跟踪状态，表现出更短的调节时间和更低的超调量。同时也可以看出该控制器对外部扰动和建模误差同样具有一定的抑制能力，体现了控制系统的鲁棒性。

$$\begin{cases} q_{d1}=(\pi/9)\sin[(\pi/9)t]+\pi/4 \\ q_{d2}=(\pi/9)\sin[(\pi/9)t]+\pi/5 \\ q_{d3}=(\pi/9)\sin[(\pi/9)t]+\pi/4 \end{cases} \tag{12}$$

5 结论

本文提出了动态稀疏训练的方法，应用在深度强化学习算法上，用来处理空间在轨操作任务中连续型机械臂的动力学控制问题。对于具有显著非线性特征的连续型机械臂动力学模型，考虑运动过程中存在的建模误差和外部干扰因素，采用DDPG算法实时优化变结构控制器控制参数，并应用一种稀疏训练方法，由ε-ER图初始化的稀疏连接层代替强化学习网络的全连接层，对强化学习主网络进行迭代剪枝,在不降低训练精度的基础上显著压缩了网络规模。仿真结果表明，所提出的基于强化学习的自适应控制器能有效地进行连续型机械臂的跟踪控制，相较于固定参数的变结构控制器具有更低的调节时间和超调量，通过稀疏训练的方法，控制器与未经稀疏化的自适应控制器具有相同精度水平的同时，网络节点参数下降了99%，大幅降低了计算成本。未来的工作将集中在规划控制一体化的组合控制方法，以及地面实验验证。

参考文献(References)

[1] 刘宏, 刘冬雨, 蒋再男. 空间机械臂技术综述及展望[J]. 航空学报, 2021, 42(1): 524164-524164.
(Liu H, Liu D, Jiang Z. Space manipulator technology:

Review and prospect [J]. Acta Aeronauticaet Astronautica Sinica, 2021, 42(1): 524164-524164.)

[2] Conrad B L, and Zinn M R. Interleaved continuum-rigid ma-nipulation: An approach to increase the capability of minimally invasive surgical systems [J]. IEEE/ASME Transactions on Mechatronics, 2017, 22(1):29-40.

[3] Burgner-Kahrs J, Rucker D C, and Choset H. Continuum ro-bots for medical applications: A survey [J]. IEEE Trans. Robot. 2015, 31(6): 1261-1280.

[4] Nicolae D, Leonard C, Cristian C, et al. Theoretical and exper-imental study methods for a robotic system with deformable elements used in minimally invasive surgery [J]. Mechanism and Machine Theory. 2022, 167:104459.

[5] Wang M F, Dong X, Ba W M, et al. Design, modelling and validation of a novel extra slender continuum robot for in-situ inspection and repair in aeroengine [J]. Robotics and Comput-er-Integrated Manufacturing.2020, 67:102054.

[6] Yang J Z, Peng H J, Zhang J, et al. Dynamic modeling and beating phenomenon analysis of space robots with continuum manipulators [J]. Chinese Journal of Aeronautics. 2021, Pre-print: https://doi.org/10.1016/j.cja.2021.08.007html.

[7] Ouyang X, Meng D S, Wang X Q, et al. Hybrid rigid-continuum dual-arm space robots: Modeling, coupling analysis, and coordinated motion planning [J]. Aerospace Science and Technology. 2021 116:106861.

[8] Melingui A, Lakhal O, Daachi B, et al. Adaptive neural net-work control of a compact bionic handling arm [J]. IEEE/ASME Transactions on Mechatronics. 2015, 20(6): 1-14.

[9] Yip M C and Camarillo D B. Model-less feedback control of continuum manipulators in constrained environments [J]. IEEE Transactions on Robotics. 2014, 30(4): 880-889.

[10] Morteza G, Ali K K. Modeling and experimental analysis of a multi-rod parallel continuum robot using the Cosserat theory [J]. Robotics and Autonomous Systems. 2020, 134:103650.

[11] Tan N , Yu P , Zhang X , et al. Model-free motion control of continuum robots based on a zeroing neurodynamic approach [J]. Neural Networks. 2021, 133:21-31.

[12] Frazelle C, Rogers J, Karamouzas I, Walker I. Optimizing a Continuum Manipulator's Search Policy Through Model-Free Reinforcement Learning [C]. 2020. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, 2020,5564-5571.

[13] Decebal C M, Elena M , Peter S, et al. Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science [J]. Nature Communications, 2018, 9(1):2383.

[14] Zhuang L, Li J, Shen Z, et al. Learning efficient convolutional networks through network slimming [C]. 2017 IEEE Interna-tional Conference on Computer Vision (ICCV). Venice, 2017, 2755-2763.

[15] Hensham M, Xin W. Parameter efficient training of deep con-volutional neural networks by dynamic sparse reparameteriza-tion [C]. International Conference on Machine Learning, Long Beach, 2019, 97:4646-4655.

[16] Rone W S, and Ben-Tzvi P. Continuum robot dynamics utiliz-ing the principle of virtual power[J]. IEEE Transactions on robotics. 2014, 30(1): 275-287.

[17] Fujimoto S, Van H, Meger D. Amsterdam Machine Learning Lab. Addressing function approximation error in Actor-Critic methods [C]. Proceedings of Machine Learning Research, New York, 2018, 80: 1587-1596.

[18] Barabási A L, Albert R. Emergence of scaling in random net-works[J]. Science. 1999, 286:509-512.

[19] 艾海平, 陈力. 空间机器人捕获航天器操作的避撞柔顺复合自抗扰控制[J].控制与决策, 2021, 36(2):355-362.
(Ai H, Chen L. Collision avoidance and compliant composite active disturbance rejection control of space robot capture spacecraft[J]. Control and Decision, 2021, 36(2):355-362.)

[20] Jiang D, Cai Z, Peng H, et al. Coordinated control based on reinforcement learning for dual-arm continuum manipulators in space capture missions[J]. Journal of Aerospace Engineering. 2021, 34(6): 04021087.

[21] 雷荣华, 陈力. 空间机器人执行器部分失效故障的终端滑模容错控制[J].中国机械工程, 2019, 30(08):69-75.
(Lei R , Chen L.Terminal Sliding Mode Fault - tolerant Con-trolfor Space Robots under Partial Loss of Actuator Effective-ness[J].China Mechanical Engineering, 2019, 30(8):69-75.)

作者简介

江达 (1992−), 男, 博士研究生, 从事智能控制、机器学习等研究, E-mail: ziangdar@sina.com;

蔡志勤 (1961−), 女, 教授, 博士生导师, 从事机器人动力学与控制、最优控制等研究, E-mail: zhqcai@dlut.edu.cn;

刘忠振 (1990−), 男, 博士研究生, 从事机器人动力学控制的研究, E-mail: liuzhongzhen@mail.dlut.edu.cn;

彭海军 (1982−), 男, 教授, 博士生导师, 从事最优控制理论与计算算法、机器人动力学与控制等研究, E-mail: hjpeng@dlut.edu.cn;

吴志刚 (1971−), 男, 教授, 博士生导师, 从事飞行器动力学与控制、空间智能结构控制等研究, E-mail: wuzhg@dlut.edu.cn.