

基于多智能体强化学习算法的机械臂优化控制研究

杨德青

郑州商学院，信息与机电工程学院，郑州，中国，451200

摘要：本次研究针对机械臂运动控制方法无法对其进行良好控制的问题，提出基于多智能体强化学习算法的机械臂优化方法。该方法先对机械臂空间运动捕捉问题进行分析，建立机械臂动力学模型，然后以此为基础，在 Unity 平台的支持下完成机械臂仿真模型的构建。之后将多智能体与 PPO 算法相结合，对机械臂进行智能运动的控制。仿真实验结果，所提方法可较好的控制机械臂运动，具有一定应用价值。

关键词：多智能体；强化学习算法；机械臂；智能控制

1 引言

机械臂生产过程中需要借助自动化技术、电力能源、计算机通信等多种技术，其对其控制的精度越高，代表其应用性能越好。机械臂在给定时刻可运行到特定工作空间中从事搬运、生产等工作，工作效率优于人工操作，且其还可扩展到人类无法到达的危险环境中工作，故其在几年前获得广泛应用。在该种背景下，应积极引用现代技术优化其控制性能。目前出现众多相关研究，比如文献[1]基于非线性微分方程分析机械臂运动位置控制问题，首先对机械臂动力学进行分析，获取动力学目标函数，构建机械臂相关的非线性微分方程并进行方程计算，将计算结果作为参数输入神经网络中，构建机械臂运动控制模型实现研究目的；文献[2]提出了一种基于深度强化学习的机械臂控制方法，引入深度确定性策略梯度算法设计了控制器，然后结合机械臂状态计算力矩，有效降低运动速度，实现柔顺捕获。但上述相关研究存在机械臂控制算法控制精度较低的问题，基于此，本次研究设置仿真环境，建立仿真模型，并将多智能体的概念引入其中，将强化学习中的 PPO 算法与多智能体相结合，对机械臂运动决策进行了规划，并利用神经网络完成了训练，从而达到了收敛的目的，完成了机械臂运动控制决策的制定。希望通过本次研究，为后续机械臂的优化控制提供文献参考。

2 问题描述与系统建模

2.1 机械臂运动分析

本文研究对象是一种小尺寸、在轨工作的六自由度机械手，其一端连接平台底座，其构造见图 1。

为实现机械臂优化控制，对其研究环境进行如下

假设：

1) 把被测目标想象成均匀的小球，并在机器人的近距离空间内作均匀直线移动，在模拟的起始点且在一段时间内，不离开机械臂的工作空间。2) 本次研究不考虑机械臂终端捕获行为，只考虑机械臂的关节运动。3) 将机械臂底座平台抽象成零控均质刚性体，在不考虑机器人移动时所受全部外部作用力及力矩的情况下开展研究^[3]。

基于此，完成机械臂仿真所用的 D-H 参数设置如下表 1 所示。

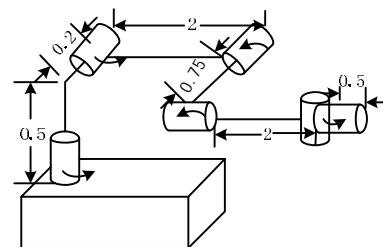


图 1. 空间机械臂

表 1. 空间机械臂 D-H 参数

机械臂连杆序号	连杆长度	扭转角	关节偏移
1	0	$\frac{p}{2}$	0
2	2	0	0
3	0	$-\frac{p}{2}$	$\frac{p}{2}$
4	0	$\frac{p}{2}$	0
5	0	$-\frac{p}{2}$	0

6	0	0	0
---	---	---	---

在不工作时，机械臂的状态是折叠的，因此将折叠状态作为机械臂的初始状态。在初始状态下，只有

关节 2 的振幅存在一个刚性约束，其它关节的振幅没有约束。这样，将关节 2 的转动角度的范围设为 $[0, p]$ ，而其他关节转动角度的范围设为 $[-p, p]$ 。机器臂的动力参数设置如表 2 所示。

表 2. 机械臂动力学参数

机械臂连杆序号	$I_x / (kg \cdot m^2)$	$I_y / (kg \cdot m^2)$	$I_z / (kg \cdot m^2)$	$T_c / (N \cdot m)$
1	0	0.36	0	0.396
2	0.13	0.54	0.54	0.128
3	0.08	0.09	0.02	0.132
4	0	0.02	0	0.169
5	0.004	0.004	0.004	0.014
6	0.005	0.005	0	0.010
平台	157.6	157.6	157.6	0

表 2 中， I_x, I_y, I_z 分别表示转动惯量在 x, y, z 轴的分量， T_c 表示机械臂的关节转矩。

2.2 机械臂动力学模型

本文以拉格朗日法为基础完成机械臂动力学模型的建立。机械臂系统的总动能为各部件动能之和，每个杆件和关节动能可由其质心线速度动能和转动角速度动能组成^[4]，则自由漂浮空间机械臂系统的总动能为

$$T = \frac{1}{2} \sum_{i=0}^{12} (m_i v_i^T v_i + w_i^T I_i w_i) \quad (1)$$

式中： m_i 表示机械臂的质量； v_i 表示受到惯性影响下机械臂运动地速度向量； I_i 表示机械臂在运动过程中的转动惯量； w_i 表示机械臂运动过程中的角速度向量。将机械臂的运动学方程式待入其中，即可得出：

$$T = \frac{1}{2} \begin{bmatrix} v_0^T & w_0^T & \Phi^T \end{bmatrix} \begin{bmatrix} ME & M_r^T & J_{Tw} \\ Mr & H_w & H_{wj} \\ J_{Tw}^T & H_w^T & H_j \end{bmatrix} \begin{bmatrix} v_0 \\ w_0 \\ \Phi \end{bmatrix} \quad (2)$$

式中， v_0 、 w_0 分别为惯性效应下平台的速度矢量与角度向量， Φ 为联合角速度， M 为机械臂质量， E 为补齐的单位矩阵， r 为由机械臂重心朝向平台重心的位置矢量， J_{Tw} 为平台切向转动惯性矩阵， H_w 为平台的惯性张量矩阵， H_{wf} 为组合体耦合的惯性张量矩阵， H_r 为基座机械臂的惯性张量矩阵，可被表达为：

$$H_r = \sum_{i=1}^6 (J_{TLi}^T I_i J_{TLi} + m_i J_{RLi}^T J_{RLi}) \quad (3)$$

式中， J_{TLi} 代表机械臂的切向转动惯量； I_i 代表机械臂的转动惯量， m_i 代表部件质量， J_{RLi} 代表机械臂径向转动惯量，则机械臂系统的拉格朗日动力学方程为：

$$\begin{bmatrix} H_w & H_{wj} \\ H_{wj}^T & H_j \end{bmatrix} \begin{bmatrix} \Phi \\ \Phi \end{bmatrix} + \begin{bmatrix} c_b \\ c_m \end{bmatrix} = \begin{bmatrix} F_b \\ t_m \end{bmatrix} \quad (4)$$

式中： c_b 代表平台本体牵连速度的非线性项； c_m 为机械臂本体牵连速度的非线性项； F_b 为机械臂受到的外力力矩； t_m 为机械臂运动关节的力矩。

基于此，完成机械臂动力学模型的构建。

3 多智能体强化学习的运动控制

3.1 多智能体与 PPO 算法结合

在机械臂仿真模型的基础上，利用多智能体增强学习方法，实现机械臂手臂的最优运动控制。首先，将几个相同并且各自独立的智能体设定为一个单位，在这些智能体之间，由于其可直接共享器奖惩信号，加速收敛速度，提高训练效率。

多智能体之间存在着各种各样的关系，本文根据机械臂选取智能体之间的合作关系进行训练，由于每个智能体之间的利益都是相同的，故可通过合作来完成奖励的获得。将多智能体的学习方式看作中心化训练，并进行无中心化执行。将 Agent 的数量设定为 n ，第 i 个 Agent 的动作为 a_i 、观测值为 o_i 、奖励值为 r_i 。每一个主体都有其自身的战略网 $p(a^i | o^i, q^i)$ 和价值网 $q(o_i | a; w^i)$ 。对于所有智能体的观察值、行为和策略的集合见公式 (5)：

$$\begin{cases} o = [o^1, o^2, \mathbf{K}, o^n] \\ a = [a^1, a^2, \mathbf{K}, a^n] \\ p = [p^1, p^2, \mathbf{K}, p^n] \end{cases} \quad (5)$$

该中心控制系统由 n 个价值网路组成, 在训练时, 中心控制系统会得到每一个智能体的动作、观测值以及奖励。在执行期间, 不会再利用中心控制器的值网, 其网路结构见图 2:

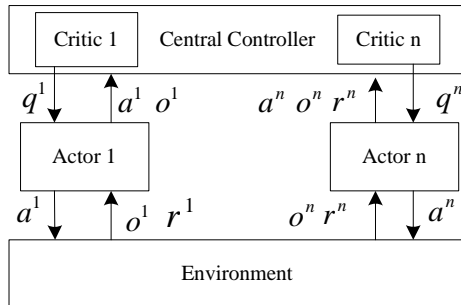


图 2. 多智能体与 PPO 结合的网络结构

3.3 训练神经网络

利用一个 4 层神经网络来对机械臂运动进行训练, 其输入为一个 10 维的向量, 输出为一个 6 维向

量, 其结构具体如图 3 所示:

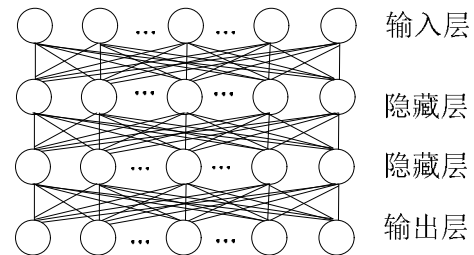


图 3. 机械臂智能体训练网络

由此, 完成训练网络的构建, 利用该训练网络对其模型进行控制, 实现机械臂的优化控制。

4 实验分析

本次实验在一台操作系统为 Windows 10 的电脑中进行, 借助 Unity 引擎平台, 构建机械臂数学孪生仿真模型, 然后对其进行测试。为了对本文控制方法的实用性进行充分地检验, 选择了文献[1]方法和文献[2]方法作为比较方法, 使用了三种不同的方法来对机械臂进行控制, 获得对比结果如图 4 所示。

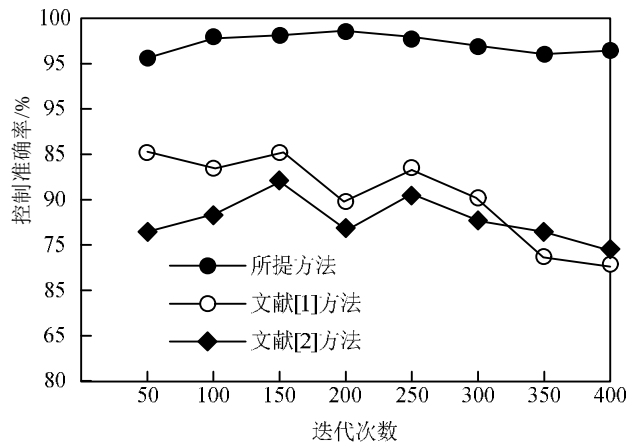


图 4. 对比结果

如图 4 所示, 采用所提方法进行控制的准确率可达到 95%以上, 且其随着迭代次数增加, 控制精度较高且较为稳定, 而对比方法的控制精度均低于 85%, 且随着系统迭代次数的增加, 控制精度大幅下跌, 故采用所提方法对机械臂进行运动决策的制定, 相比于

传统方法, 可在保持其稳定性的同时提高其控制精度, 由此证明所提方法应用价值较高。

5 结束语

机械臂操作灵活, 在工业装配、安全防爆等领域

得到了广泛应用，但机械臂系统是一个非常复杂的非线性系统，具有多输入多输出、强耦合以及非线性等特征，因此对其进行控制研究存在一些难度。本次研究针对现有方法无法对机械臂进行准确控制的问题，基于多智能体强化学习算法，开展机械臂控制优化设计。首先，通过对机械臂系统的动态建模，建立机械臂系统的模拟模型，然后利用多 Agent 增强学习算法对其进行训练，从而达到系统的最优控制。最后通过实验证明所提方法可完成机械臂运动地控制，控制准确率可达到 95% 以上且稳定性较高，其控制效果远远

优于文献[1]方法和文献[2]方法，具有一定应用价值。

参考文献

- [1] 宁海成. 基于非线性微分方程的机械臂运动位置控制方法. 机械与电子. 2022, 40(11), 46-50.
- [2] 文闻, 周元子, 周晓东, 等. 基于深度强化学习的空间机械臂柔顺捕获控制方法研究. 空间控制技术与应用. 2022, 48(01), 1-8.
- [3] 杨照, 黄少伟, 陈颖. 基于多智能体强化学习的多园区综合能源系统协同优化运行研究. 电工电能新技术. 2021, 40(08), 1-10.
- [4] 张永超, 于智伟, 丁丽林. 基于强化学习的煤矸石分拣机械臂智能控制算法研究. 工矿自动化. 2021, 47(01), 36-42.