基于 GPR 和深度强化学习的分层人机协作控制

金哲豪 [1]　刘安东 [1]　俞　立 [1]

摘　要　提出了一种基于高斯过程回归与深度强化学习的分层人机协作控制方法, 并以人机协作控制球杆系统为例检验该方法的高效性. 主要贡献是: 1) 在模型未知的情况下, 采用深度强化学习算法设计了一种有效的非线性次优控制策略, 并将其作为顶层期望控制策略以引导分层人机协作控制过程, 解决了传统控制方法无法直接应用于模型未知人机协作场景的问题; 2) 针对分层人机协作过程中人未知和随机控制策略带来的不利影响, 采用高斯过程回归拟合人体控制策略以建立机器人对人控制行为的认知模型, 在减弱该不利影响的同时提升机器人在协作过程中的主动性, 从而进一步提升协作效率; 3) 利用所得认知模型和期望控制策略设计机器人末端速度的控制律, 并通过实验对比验证了所提方法的有效性.

关键词　深度强化学习, 高斯过程回归, 人体控制策略感知, 分层人机协作

引用格式　金哲豪, 刘安东, 俞立. 基于 GPR 和深度强化学习的分层人机协作控制. 自动化学报, 2022, **48**(9): 2352−2360

DOI　10.16383/j.aas.c190451

Hierarchical Human-robot Cooperative Control Based on GPR and Deep Reinforcement Learning

JIN Zhe-Hao[1]　LIU An-Dong[1]　YU Li[1]

Abstract　In this paper, a hierarchical human-robot collaboration control problem is investigated by Gaussian process regression and deep reinforcement learning approaches, and a ball and beam system controlled jointly by human and robot is used to verify the proposed method. The main contributions are as follows: 1) To deal with the problem that the classical control method can not be directly used in the human-robot collaboration scenario without a known model, a deep reinforcement learning algorithm is adopted to design an effective nonlinear suboptimal policy without the system model, and this suboptimal policy is considered as the expected control policy to guide the Human-robot collaboration process; 2) To weaken the negative influences caused by the unknown and random human-control strategies, the Gaussian process regression method is used to fit the human-control strategies and build the cognitive model of robot for human control behaviors, which can improve the efficiency of collaboration by enhancing the initiative of the robot through the Human-robot collaboration process; 3) A controller for the end-effector velocity is designed based on the cognitive model and the expected control policy, and the effectiveness of the proposed method is verified by experimental comparison.

Key words　Deep reinforcement learning, Gaussian process regression, human-control strategy perception, hierarchical human-robot collaboration

Citation　Jin Zhe-Hao, Liu An-Dong, Yu Li. Hierarchical human-robot cooperative control based on GPR and deep reinforcement learning. *Acta Automatica Sinica*, 2022, **48**(9): 2352−2360

近年来, 随着机器人技术的高速发展, 机器人在工业生产中替代了大量的人力资源. 然而, 对于一些复杂的任务, 机器人往往无法和人类一样灵活的操作与控制. 人机协作 (Human-robot collaboration, HRC) 研究如何利用人的灵活性与机器人的高效性, 使机器人与人协同高效、精准地完成复杂任务, 因此受到了国内外学者的广泛关注[1].

人机协作按机器人在协作过程中的角色可分为人主−机器人从、机器人主−人从、人机平等 3 类. 第 1 类人机协作中机器人接收人发出的命令并执行, 主要完成一些负重类的任务. 如文献 [2] 中人与机器人共同搬运一个物体, 其中人决定了运动轨迹, 而机器人作为跟随者负责轨迹跟随并承担重物. 在这一类人机协作任务中的一大难点是如何将人的想法正确的传递给机器人. 文献 [3−4] 研究了在人与机器人共同操作一个对象时, 如何消除传递给机器人旋转与平移命令之间歧义的方法. 第 2 类人机协作的研究相对较少, 文献 [5] 将人建模为一个被动的旋转关节模型, 并且用实验证明了在机器人主导的情况下如何使用该模型将物体维持水平. 以上两类人机协作方法虽然能一定程度上结合人与机器人自身的优点, 但过于注重单方面的性能, 如人类的灵活性或机器人的高效性, 从而导致协作的整体效率不高.

人机平等形式的人机协作考虑人与机器人以平等的关系完成复杂任务, 这要求协作双方对对方的操作规律有一定的了解. 由于人的智能性, 对于人而言这种能力可以很方便地获得, 但机器人无法自然获取这种能力, 因此如何为机器人建立有关人的运动规律模型是非常重要的. 其中较为常用的方法假设是人的运动规律满足最小抖动模型[6], 并根据该模型预测人的运动轨迹. 文献 [7] 在人与机器人协作抬一根长杆的场景中, 使用加权最小二乘实时估计最小抖动模型中的参数, 并利用变种阻抗控制器使机器人跟踪最小抖动模型的预测值, 从而达到使机器人主动跟随人运动的效果. 文献 [8] 利用扩展卡尔曼滤波估计最小抖动模型中的参数, 并在一维的点到点运动中证明该方法的有效性. 文献 [7−8] 均证明了在人机协作中使用以上基于最小抖动模型的方法能在一定程度提升人的舒适度. 然而, 基于最小抖动模型生成人的运动轨迹需要事先了解人运动轨迹起止时间与起止位置, 这在一些任务中过于苛刻. 文献 [9] 表明最小抖动模型在一些特别的协作任务中会失效, 如一些协作任务中人的轨迹存在大量的干扰与抖动, 或者人在协作过程中多次决定改变其运动轨迹. 文献 [10−11] 假设人在运动过程中其加速度变化较小, 利用卡尔曼滤波器预测人下一时刻的位置, 并根据预测精度加权融合机器人主被动控制器, 从而提高机器人协作时的主动性以及协作的鲁棒性. 该方法在人机协作抬桌子的场景中得到了验证. 文献 [12] 使用基于与文献 [10−11] 相同的运动模型的扩展卡尔曼滤波预测人下一时刻的位置, 但是其使用基于强化学习的方法设计机器人的速度控制律, 并且利用扩展卡尔

收稿日期 2019-06-11　录用日期 2019-12-02

Manuscript received June 11, 2019; accepted December 2, 2019

NSFC-浙江两化融合联合基金 (U1709213) 和国家自然科学基金 (61973275) 资助

Supported by NFSC-Zhejiang Joint Foundation for the Integration of Industrialization and Informatization (U1709213) and National Natural Science Foundation of China (61973275)

本文责任编委 程龙

Recommended by Associate Editor CHENG Long

1. 浙江工业大学信息工程学院 杭州 310023

1. College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023

曼滤波的预测值减小强化学习算法搜索的动作空间范围, 提升了机器人的协调能力, 同时加强了机器人在协作任务中的主动性. 也有一些工作[13−14] 将人的控制量作扰动处理.

以上方法均属于较为经典的人运动轨迹建模方法, 有较强的可解释性. 然而一些复杂的人机协作任务中, 人的运动轨迹往往很不规律, 如人手在 3 维空间中到达某些不同目标位置时形成的轨迹[15]、人在完成装配任务时的运动轨迹[16] 等. 此时用概率分布去建模这些轨迹显然更加合适, 因此一些基于学习和统计的轨迹建模方法往往更加有效. 文献 [15]利用高斯混合模型 (Gaussian mixture model, GMM) 与高斯混合回归 (Gaussian mixture regression, GMR) 建立人手到达不同目标位置所形成的轨迹概率分布模型, 该模型被用来提升人机协作过程中的安全性以及机器人的自主性. 文献 [16] 通过人拖动机器人完成装配任务的方式将人的运动轨迹转化为机器人末端的轨迹, 并利用 GMM/GMR 建立机器人末端的轨迹概率分布模型以达到示教学习的目的. 文献 [17] 利用高斯过程回归 (Gaussian process regression, GPR) 拟合包含人在内的球杆系统的前向传播模型, 并利用基于模型的 RL 算法设计次优控制律, 极大地提升了对数据的利用率. 文献 [18] 使用卷积神经网络学习人在完成零件装配任务时的动作与意图. 文献 [19] 使用触觉数据作为输入, 利用基于隐马尔科夫模型的高层控制器估计人的意图并生成相应的机器人参考轨迹, 并在机器人与人握手的场景中验证了该方法的有效性. 另外, 部分可观马尔科夫模型[20] 以及贝叶斯神经网络[21] 也被用来预测人下一时刻的行为.

然而, 上述方法几乎都是对人在一段时间内的运动轨迹进行建模, 很少有文献直接对人的控制策略建模. 与人运动轨迹建模不同, 针对人体控制策略建模主要为了预测人在遇到某个状态时可能执行的动作, 从而为机器人对人的控制行为建立更加直观的认知模型. 本文提出了一种基于 GPR 与深度强化学习 (Deep reinforcement learning, DRL) 的两层人机协作控制方法, 不仅设计了一种次优的非线性控制律, 还对人体控制策略建模, 从而降低了人为不确定因素的不利影响, 增强了协作系统的稳定性, 并解决了传统主从式人机协作中效率较低的问题. 本文以人机协作控制球杆系统为例验证该方法的可行性. 首先, 针对顶层期望控制律的设计问题, 利用深度确定性策略梯度算法 (Deep deterministic policy gradients, DDPG)[22] 得到了一种次优的非线性控制器. 其次, 本文使用 GPR 建立球杆系统的人体控制策略模型, 解决了协作过程中由人为不确定因素所导致的系统不稳定问题. 然后, 根据期望控制律和人体控制策略模型设计机器人的控制律以提升人机协作的效率. 最后, 通过实验验证了该方法的可行性与有效性.

1　问题描述

本文以球杆系统为例设计分层人机协作控制方法, 考虑如图 1 所示的人机协作球杆系统.

图 1 中, 人与机械臂各执长杆一端以控制长杆倾角, 使小球快速, 平稳地到达并停留在目标位置 (虚线小球位置). 在人机协作环境下, 由于长杆的倾角变化幅度较大, 使得在平衡点附近线性化模型后设计相应控制器的方法效果不佳. 因此, 如何针对该球杆系统设计一种有效的非线性控制器是本文的一大难点. 然而, 常规的非线性控制方法对模型精度依赖较高, 而一些复杂协作任务往往很难精确建模, 甚至无法建模. 因此, 本文基于 DRL 算法设计球杆系统的控制器. DRL 算法不依赖环境模型, 其通过不断

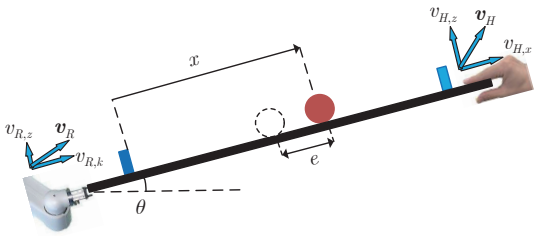


图 1　人机协作控制球杆系统示意图

Fig.1　Schematic diagram of the human-robot collaboration task

与环境交互, 以寻找一种使累积奖励最大化的控制策略. 由于 DRL 利用神经网络设计控制器, 并通过迭代的方式更新参数, 易陷入局部最优. 因此, 基于 DRL 的非线性控制器是一种次优控制器.

使用 DRL 设计控制器需要先将球杆系统建立成马尔科夫决策模型. 马尔科夫决策模型由 5 元组 $(S,\ A,\ P,\ r,\ \gamma)$ 表示. 其中 S 表示状态空间, 是对环境状况的一种数学描述; A 表示动作空间, 是智能体影响环境的手段; P 表示状态转移概率, 表示在当前状态受到某个动作后下一个状态的概率分布, 也可以理解为环境模型; r 表示奖励函数, 是环境对当前状态施加某个动作后的一个奖惩反馈; γ 表示折扣因子, 是调节智能体关注长远利益程度的参数.

控制器的设计问题可以转化为解马尔科夫决策模型问题, 即设计一个最优策略 $\pi^*:\boldsymbol{s}\mapsto a$ 使智能体获得的累积奖励最大化. 对于任意的 $\boldsymbol{s}\in S$, $\pi^*(\boldsymbol{s})$ 满足:

$$\pi^*(\boldsymbol{s})=\arg\max_{\pi}\mathrm{E}_{\pi}\left[\sum_{t=0}^{\infty}\gamma^t r\left(\boldsymbol{s}_t,\ \pi(\boldsymbol{s}_t)\right)|\boldsymbol{s}_0=\boldsymbol{s}\right] \tag{1}$$

式中, π^* 可以通过强化学习算法设计. 由于球杆系统状态空间连续的特性, 使得处理离散状态空间马尔科夫决策模型的传统强化学习算法无法为其设计最优策略. 因此, 对于这类状态空间连续的马尔科夫决策模型常常使用基于估计的强化学习算法 (如 DRL). 为了取得更好的控制效果, 本文考虑连续的动作空间, 这使处理离散动作空间的基于值函数的 DRL 方法[23−24] 失效. 本文使用的 DDPG 算法利用 Actor-Critic 结构, 能在连续的动作空间中寻找一种次优控制策略.

另外, 在主从式协作中, 从方往往不做决策, 只承担跟随或执行主导方发出的命令的任务. 因此, 该模式的协作效率往往较低, 即系统进入稳态所需的控制时间较长. 本文考虑人机平等的协作方式, 即人与机器人均为完成任务作出控制决策, 而人的高随机性行为将为机器人控制器设计带来极大的不确定性. 因此, 如何为机器人建立人体控制策略预测模型, 增强机器人在协作过程中的主动性, 从而提高协作效率与协作鲁棒性是本文的第 2 个难点. 考虑到人体控制策略的随机性 (即使同一个人面对相同状态, 其采取的控制行为也可能不同, 本文假设该行为服从高斯分布), 本文利用 GPR 拟合人体控制策略. 与传统回归算法不同的是, 对于一个特定的输入, GPR 模型的输出并不是一个固定的值, 而是一个高斯分布, 即 $\hat{\pi}_H(\boldsymbol{s})\sim\mathrm{N}(a,\ \delta)$. 并且, GPR是一种非参数估计方法, 因此不会有过拟合的风险.

由于协作过程中只有机械臂的行为是可控的, 本文的

目标是为机械臂设计合适的末端速度控制律以使小球在人机协同控制下快速, 平稳地到达并停留在指定位置. 本文以基于 DRL 的次优非线性控制策略为期望控制策略, 以拟合的人体控制策略预测模型作为机器人对人控制行为的认知模型, 设计机器人的控制律, 从而使人机协作的整体控制效果趋向于期望控制策略的控制效果.

2 人机协作控制

本节将设计基于 GPR 与 DRL 的分层人机协作控制方法, 具体分为顶层与底层的设计. 其结构如图 2 所示:

顶层利用 DDPG 算法为非线性球杆系统设计一种次优的高效控制律, 并作为人机协作过程中的期望控制策略. 底层主要分为两部分: 1) 基于 GPR拟合人体控制策略, 为机械臂建立人控制行为的认知模型; 2) 根据期望控制策略以及认知模型设计机械臂的末端速度控制律, 从而使人机协作下的控制行为趋向于期望控制策略的控制行为.

2.1 顶层设计

本节主要介绍如何利用 DDPG 设计球杆系统的期望控制策略. 在此之前, 必须先将球杆系统建立成马尔科夫决策模型, 主要包括状态空间、动作空间和奖励函数的设计.

1) 状态空间: 球杆系统的控制目的是使小球快速, 稳定地到达指定位置, 因此位置误差信号 e 被用来构建状态. 另外, 据经验可知, 人在控制球杆的时候还会关注小球的速度 $\dot{x}$ 以及长杆的倾角 θ. 同时, 为了不使小球离开长杆, 小球的位置 x 也被用来构建状态. 因此, 马尔科夫决策模型状态被定义为 $\boldsymbol{s} = [e \quad x \ \dot{x} \quad \theta]^{\mathrm{T}}$.

2) 动作空间: 本文以长杆的旋转角速度作为控制量, 因此, 动作被定义为 $a = \dot{\theta}$.

3) 奖励函数: 为了使小球快速, 稳定地到达指定位置, 本文设计的损失函数为 $c = [e \quad \dot{x} \quad \dot{\theta}]W_c\,[e \quad \dot{x} \quad \dot{\theta}]^{\mathrm{T}}$, 其中 $\boldsymbol{W}_c$ 为权重矩阵, 令奖励函数 $r = -c$. 另外, 小球离开长杆被认为是控制失败, 因此, 一但检测到小球离开长杆, 环境将给予一个幅值较大的损失函数并重新开始实验.

DDPG 算法可以用来为状态以及动作空间连续的马尔科夫决策模型寻找次优策略, 主要包含 Actor、Actor 目标网络、Critic、Critic 目标网络 4 个神经网络. 记这 4 个神经网络的参数分别为 $\boldsymbol{\theta}^{\mu}$、$\boldsymbol{\theta}^{\mu'}$、$\boldsymbol{\theta}^{Q}$、$\boldsymbol{\theta}^{Q'}$. Critic 神经网络用来估计动作值函数 $Q(\boldsymbol{s}, a)$, 即对于马尔科夫决策模型在状态 $\boldsymbol{s}$ 执行动作 a 的价值, 并利用 Bellman 方程来构建其损失函数:

$$\begin{cases} L(\boldsymbol{\theta}^{Q}) = \mathrm{E}_{\boldsymbol{s}\sim\beta}\left[\left(Q\left(\boldsymbol{s}_t,\, a_t|\boldsymbol{\theta}^{Q}\right) - y_t\right)^2\right] \\ y_t = r\left(\boldsymbol{s}_t,\, a_t\right) + \gamma Q'\left(\boldsymbol{s}_{t+1},\, \mu'(\boldsymbol{s}_{t+1}|\boldsymbol{\theta}^{\mu'})|\boldsymbol{\theta}^{Q'}\right) \end{cases} \tag{2}$$

式中, β 是一种随机策略, 用来探索未知环境. Actor 神经网络以 $\boldsymbol{s}$ 作为输入, 以 a 作为输出, 负责学习控制策略, 其参数更新规则较为复杂. 根据文献 [25] 给出的确定性策略梯度理论, Actor 网络在策略 μ 下, 目标函数对 $\boldsymbol{\theta}^{\mu}$ 的梯度为:

$$\begin{aligned} \nabla_{\boldsymbol{\theta}^{\mu}} J\left(\mu\right) = \mathrm{E}_{\boldsymbol{s}\sim\beta}\Big[& \nabla_{\boldsymbol{\theta}^{\mu}} \mu(\boldsymbol{s}|\boldsymbol{\theta}^{\mu})|_{\boldsymbol{s}=\boldsymbol{s}_t} \\ & \nabla_a Q^{\mu}\left(\boldsymbol{s},\, a|\boldsymbol{\theta}^{Q}\right)|_{\boldsymbol{s}=\boldsymbol{s}_t,\, a=\mu(\boldsymbol{s}_t|\boldsymbol{\theta}^{\mu})}\Big] \end{aligned} \tag{3}$$

设立目标网络是为了促进神经网络收敛, 目标网络与原网络之间采用软更新原则:

$$\begin{cases} \boldsymbol{\theta}^{\mu'} = \tau\boldsymbol{\theta}^{\mu} + (1-\tau)\,\boldsymbol{\theta}^{\mu'} \\ \boldsymbol{\theta}^{Q'} = \tau\boldsymbol{\theta}^{Q} + (1-\tau)\,\boldsymbol{\theta}^{Q'} \end{cases} \tag{4}$$

另外, 受到深度 Q 网络 (Deep Q network, DQN) 的启发, DDPG 还设立的回放缓冲区 M 储存过去的数据, 并从中随机抽样训练 Actor 与 Critic 神经网络. 使用 DDPG 设计球杆系统期望控制策略的算法如下所示:

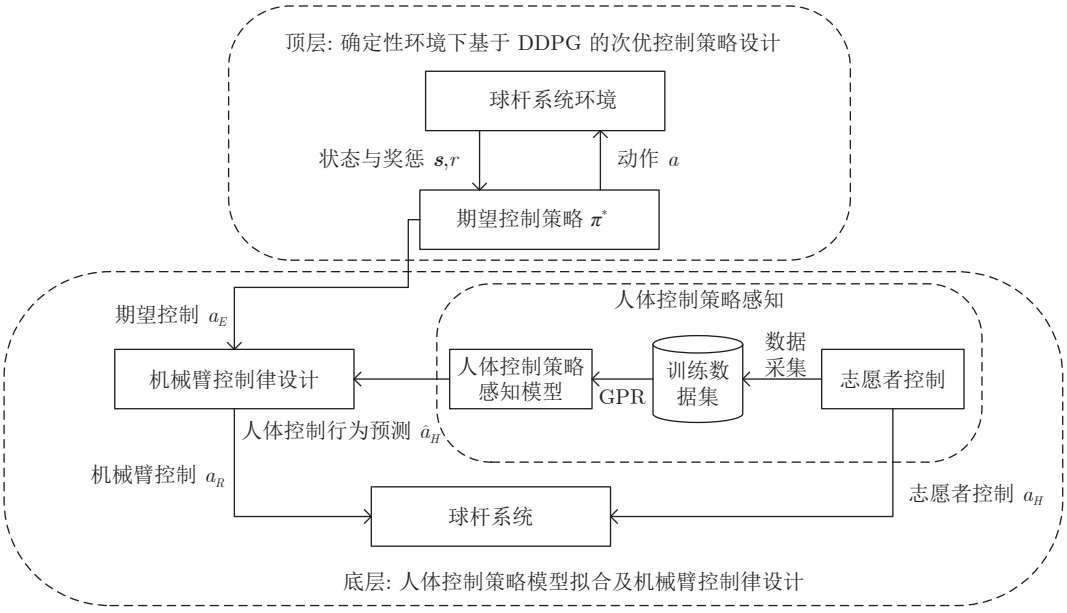


图 2 分层人机协作球杆结构示意图

Fig. 2 Schematic diagram of hierarchical human-robot collaboration

算法 1. 基于 DDPG 的球杆系统期望控制策略设计.

1) 随机初始化 Actor 和 Critic 网络参数 $\boldsymbol{\theta}^{\mu}$和$\boldsymbol{\theta}^{Q}$;
2) 将 Actor 和 Critic 网络参数复制到目标网络;
3) 初始化回放缓冲区 M;
4) for $episode = 1, \cdots, n$ do;
5) 初始化一个随机噪声生成器 ℵ;
6) 观测初始球杆系统初始状态 $\boldsymbol{s}_1$;
7) for $t = 1, \cdots, T$ do;
8) 选择并执行动作 $a_t = \mu(\boldsymbol{s}_t|\boldsymbol{\theta}^{\mu}) + \epsilon_t$;
9) 观测奖惩反馈 r_t 与下一时刻状态 $\boldsymbol{s}_{t+1}$;
10) 将数据 $(\boldsymbol{s}_t, a_t, r_t, \boldsymbol{s}_{t+1})$ 存入 M;
11) 从 M 中随机抽取 N 对 $(\boldsymbol{s}_i, a_i, r_i, \boldsymbol{s}_{i+1})$, 并根据式 (2) 和式 (3) 训练 Actor 和 Critic 网络;
12) 根据式 (4) 更新目标网络参数;
13) end for;
14) end for.

2.2　底层设计

本节介绍如何利用 GPR 拟合人体控制策略以及如何根据期望控制策略和人体控制策略模型设计机械臂的控制律.

2.2.1　人体控制策略感知

本节使用 GPR 拟合人体控制策略, 训练集记为 $X = [\boldsymbol{s}_i \quad v_{H,i}]_{i=1,\cdots,N}$, 其中 N 为数据集的大小, $\boldsymbol{s}_i$ 表示球杆系统的状态, 作为 GPR 的特征, $\boldsymbol{v}_{H,i} = [v_{H,x,i} \quad v_{H,z,i}]$ 表示长杆人控制端的速度, 作为 GPR 标签, 如图 1 所示.

高斯过程的先验均值函数理论上可以随意选择, 本文选取高斯过程的先验均值函数为 0, 即 $m(x) = 0$. 真正对 GPR 的预测效果起较大影响的是高斯过程的先验协方差函数. GPR 利用核函数来构建先验协方差函数, 考虑到平滑性, 本文使用高斯核函数:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^{\mathrm{T}} W^{-1}(\boldsymbol{x} - \boldsymbol{x}')\right] \tag{5}$$

因此, GPR 的超参为 σ_f 以及核协方差矩阵 $\boldsymbol{W}$. 这些超参可以在训练集上通过最小化边缘似然来优化.

对一个测试样本 $\boldsymbol{x} = \boldsymbol{s}^*$ 以及训练集 X, 记 "测试–测试协方差"、"测试–训练协方差" 以及 "训练–训练协方差" 分别为 $K(\boldsymbol{x}, \boldsymbol{x})_{1\times1}$、$K(\boldsymbol{x}, X)_{1\times N}$ 和 $K(X, X)_{N\times N}$, 其中 $K(\boldsymbol{x}, X)_{1j} = k(\boldsymbol{s}^*, \boldsymbol{s}_j)$. GPR 的输出为一个高斯分布, 其均值和协方差表示为:

$$\boldsymbol{v}_H^* = K(\boldsymbol{x}, X)\left(K(X, X) + \sigma^2 I\right)^{-1} V \tag{6}$$

$$\Sigma^* = K(\boldsymbol{x}, \boldsymbol{x}) + \sigma^2 I - K(\boldsymbol{x}, X)\left(K(X, X) + \sigma^2 I\right)^{-1} K^{\mathrm{T}}(\boldsymbol{x}, X) \tag{7}$$

式中, $V = [v_{H,1}, \cdots, v_{H,N}]^{\mathrm{T}}$, σ^2 表示数据集的测量方差, 同样可以作为超参被优化.

2.2.2　机械臂控制

本节在期望控制策略与人体控制策略预测模型的基础上, 设计机械臂末端速度的控制律.

机械臂的控制目标是使长杆在机器人末端速度 $\boldsymbol{v}_R$ 与人控制端速度 $\boldsymbol{v}_H$ 的作用下, 其旋转角速度趋向于期望值 $\dot{\theta}$, 其中 $\dot{\theta}$ 可由顶层 Actor 网络前向传播得到, $\boldsymbol{v}_H$ 的估计值 $\hat{\boldsymbol{v}}_H$ 可由人体控制策略预测模型预测得到, 本文使用高斯分布的均值作为估计值. 如图 1 所示, $\boldsymbol{v}_H$ 和 $\boldsymbol{v}_R$ 可以分别分解成 $v_{H,x}$、$v_{H,z}$、$v_{R,x}$、$v_{R,z}$, 其中 $v_{H,x}$ 和 $v_{R,x}$ 不会影响长杆的旋转速度, 考虑到协作过程中人的舒适性, 令 $v_{R,x} = \hat{v}_{H,x}$, $v_{H,z}$, $v_{R,z}$ 与 $\dot{\theta}$ 之间满足 $\dot{\theta} = (v_{H,z} - v_{R,z})/L$, 因此, $v_{R,z} = \hat{v}_{H,z} - \dot{\theta}L$, 其中 L 为长杆长度.

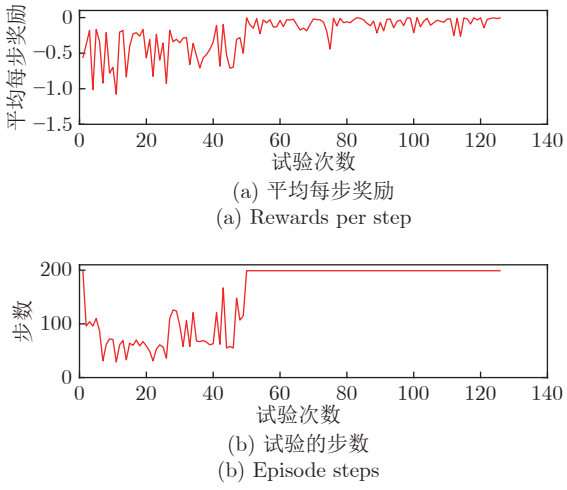
3　仿真与实验

本节通过仿真与实验验证了所设计的人机协作控制方法的有效性, 共分为 3 个部分: 1) 介绍 DDPG 中各神经网络的架构及超参数的设计, 并在仿真环境中训练各神经网络以得到顶层期望控制策略. 同时, 通过与基于值函数的 DRL 算法对比, 证明了在实际控制任务中使用基于策略的 DRL 算法 (如本文使用的 DDPG 算法) 来设计顶层期望控制策略的必要性. 2) 通过相机采集人控制球杆系统的实验数据以构建训练集, 介绍并分析了利用 GPR 拟合人体控制策略预测模型的结果. 基于得到的期望控制策略与人体控制策略预测模型. 3) 在实际场景中通过人机协作控制球杆系统与人单独控制球杆系统的控制效果作对比, 证明了所提控制方法确实能提升效率与控制精度.

3.1　基于 DDPG 的期望控制策略设计

本节分析 DDPG 学习期望控制策略的过程与结果. 首先介绍 DDPG 中神经网络的架构与超参设置. DDPG 共包含 4 个神经网络, 由于球杆系统的复杂程度相对较低, 本文将 Actor 与 Actor 目标网络设置成 3 层全连接网络, 隐藏层单元个数为 30; 将 Critic 与 Critic 目标网络设置为 4 层全连接网络, 隐藏层单元个数分别为 30 和 40. Actor 与 Critic 网络的学习率均为 0.001. 回放缓冲区大小为 10 000 对 $[\boldsymbol{s}_k \quad a_k \quad r_k \quad \boldsymbol{s}_{k+1}]$, 每次训练采样 64 对数据. 目标网络软更新参数为 $\tau = 0.01$. 损失函数中的权重矩阵 $\boldsymbol{W}_c$ 取对角阵, 对角元分别为 {5, 0.1, 0.001}. 神经网络优化器使用 Adam 优化器.

仿真环境中忽略球杆系统摩擦力, 具体模型参考文献 [26], 控制周期设置为 0.033 s (与第 3.2 节中通过相机采样志愿者控制数据的采样周期保持一致), 每次试验最长为 200 步. DDPG 的训练过程如图 3 所示.



(a) 平均每步奖励
(a) Rewards per step

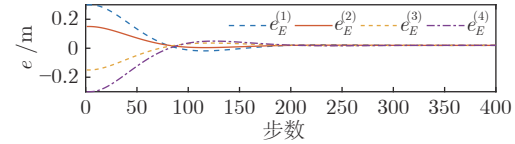
(b) 试验的步数
(b) Episode steps

图 3　DDPG 训练过程曲线图

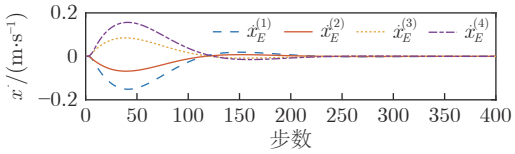
Fig.3　Training process curves of DDPG

由于环境在每一步给智能体的奖励均为负值, 而球杆系统需要长久的运行, 因此每一次试验累积的奖励值所代表的意义不鲜明. 故本文统计了每次试验在每一步的平均奖励值随训练时间的变化情况. 另外, 本文还统计了每次试验运行的时间 (步数) 以监测球杆系统在训练过程中的稳定性变化情况. 由图 3(a) 可见, 平均每一步所累积的奖励值随着训练时间的增长逐渐增加, 这说明以本文设置的奖励函数为评价标准, 控制器的表现越来越好. 最终, 平均每一步所获得的奖励值收敛于一个接近 0 的负值, 这是由奖励信号的设计方式所导致的. 图 3(b) 说明了随训练时间的增加, 球杆系统从开始的控制失败 (步数较少, 因为小球离开长杆) 逐渐变得更加稳定 (后期每次球杆系统控制时长都达到了最大值). 由图 3 可以猜测, DDPG似乎习得了一个合适的控制器.

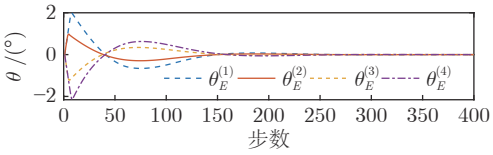
为了检验习得的期望控制策略的有效性, 在仿真环境中用该控制策略控制球杆系统 (随机选择了 4 个初始状态), 结果如图 4 所示. 其中 $e_E^{(i)}$、$\dot{x}_E^{(i)}$、$\theta_E^{(i)}$ 分别表示在期望控制策略的控制下第 i 次仿真小球位置误差, 小球速度以及长干角度的变化轨迹. 可以发现, 从任意的初始状态出发, 基于 DDPG 的期望控制策略都能高效, 稳定的完成控制任务. 另外, 该期望控制策略并没有将小球准确无误的停在目标位置, 而是存在着 2 cm 左右的误差, 这可能是 DRL 算法在学习过程中没有完美的把握 “利用与探索之间的平衡”导致的. 当然, 这也是 DRL 中公认的一大难点. 但是, 总体来说, 该期望控制策略作为一种基于神经网络的非线性控制器, 在本文设计的奖励指标上具有次优性.



(a) 误差轨迹
(a) Error trajectories



(b) 速度轨迹
(b) Velocity trajectories



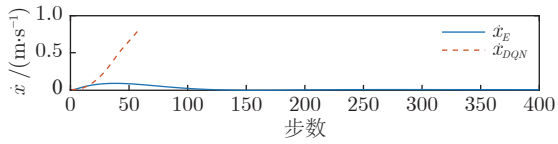
(c) 角度轨迹
(c) Angle trajectories

图 4 DDPG 控制效果图

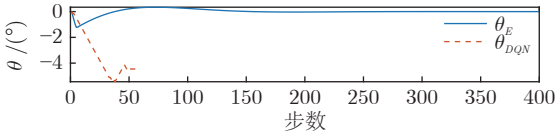
Fig.4 The control result of DDPG

另外, 本文在仿真中对比了基于 DDPG 的控制策略与基于 DQN 的控制策略的控制效果, 结果如图 5 所示. DQN 算法是一种经典的基于值函数的 DRL 算法, 其控制量是离散的. 本次仿真中 DQN的控制量属于 {5((°)/s), 0((°)/s), −5((°)/s)}. 可以发现, 由于控制量是离散且其

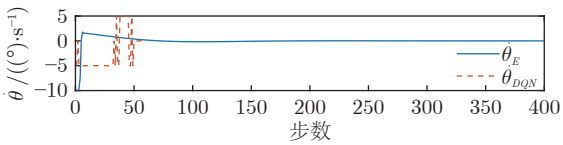
(a) 误差轨迹
(a) Error trajectories



(b) 速度轨迹
(b) Velocity trajectories



(c) 角度轨迹
(c) Angle trajectories



(d) 角速度轨迹
(d) Angular velocity trajectories

图 5 DDPG 与 DQN 的控制效果对比图

Fig.5 The comparison of control effects between DDPG and DQN

个数是有限的, 如 DQN 这种基于值函数的 DRL 方法往往很难解决实际的控制问题. 因此, 使用基于策略的 DRL 方法设计期望控制策略是必要的.

3.2 基于 GPR 的人体控制策略感知

本节分析利用 GPR 学习人体控制策略预测模型的结果. 本文通过相机检测人机协作球杆系统的状态, 具体检测环境如图 6 所示. 相机通过检测长杆两端的特征点 (分别记人端和机器人端的特征点为 $p1$ 与 $p2$) 与小球的位置 (记为 $p3$), 以确定球杆系统的实时状态.

据经验可知, 人控制球杆系统时主要根据状态 $\boldsymbol{s}=[e \quad x \quad \dot{x} \quad \theta]^{\mathrm{T}}$ 来决定旋转长杆的速度 $\boldsymbol{v}_H$. 为了获取训练数据, 本文邀请了 10 位志愿者控制球杆系统, 并利用相机记录了他们在控制过程中的控制策略数据 $(\boldsymbol{s}, \boldsymbol{v}_H)$. 由于相机检测的是位置级信息, 通过差分算法得到速度级信息时不可避免的引入高频噪声, 因此本文使用低通滤波器对数据进行滤波, 效果如图 7 所示 (本文只给出 $p1$ 点检测信息, 另外 2 点的滤波效果相似). 其中 $p1_{x,O}$、$p1_{y,O}$、$p1_{x,F}$ 和 $p1_{y,F}$ 分别表示 $p1$ 在滤波前后的横纵像素坐标, $v1_{x,O}$、$v1_{y,O}$、$v1_{x,F}$ 分别表示 $p1$ 在滤波前后的横纵像素速度. 虽然经过滤波后的数据在位置级信息中有轻微的相位落后, 但是速度级数据中的高频噪声被大幅抑制了. 因此, 利用滤波后 3 点的位置数据可以较好得到数据集 $(\boldsymbol{s}, \boldsymbol{v}_H)$. 图 8 可视化了一部分基于滤波后 3 点构建的志愿者控制球杆系统的状态轨迹.

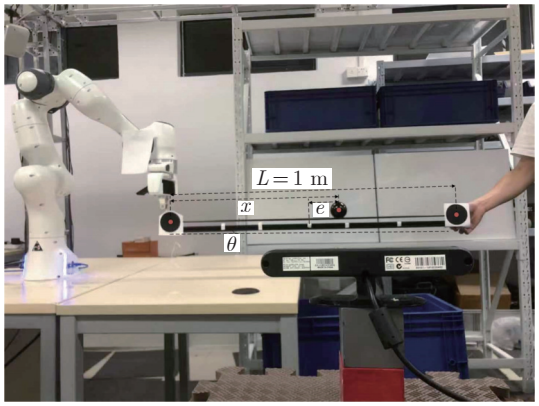
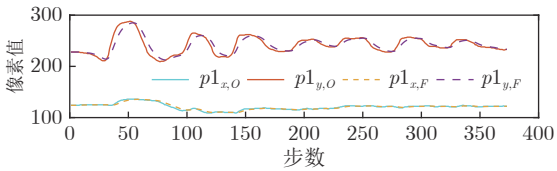


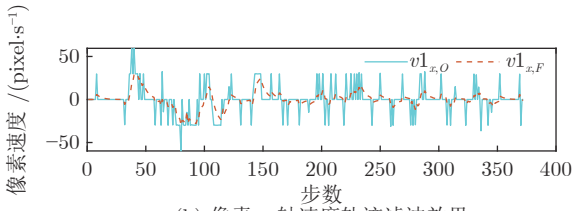
图 6　人机协作实验环境图

Fig.6　The environment of the human-robot collaboration task



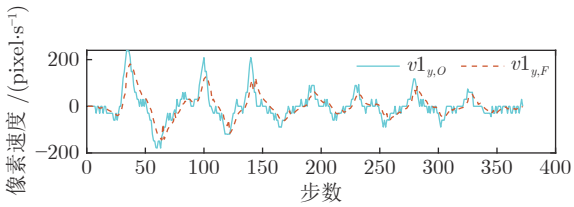
(a) 像素轨迹滤波效果

(a) Filtering effects of the pixel trajectories



(b) 像素 x 轴速度轨迹滤波效果

(b) Filtering effects of the pixel velocity trajectory in x-axis



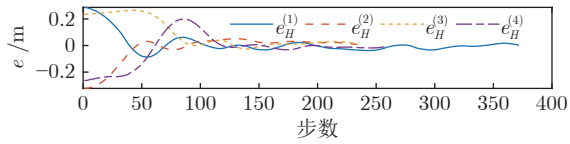
(c) 像素 y 轴速度轨迹滤波效果

(c) Filtering effects of the pixel velocity trajectory in y-axis

图 7　志愿者控制过程中产生数据的滤波结果图

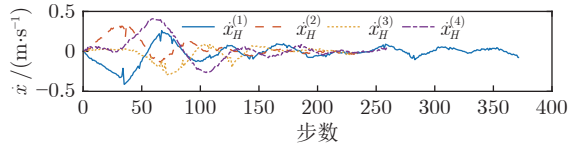
Fig.7　Filtering results of the data generated by volunteers′ control process

图 8 中的下标 H 表示这些数据是由人的控制策略产生的. 可以发现, 志愿者在控制球杆系统时并不会使小球最终严格地停在目标位置处, 而是在目标位置附近徘徊. 并且, 人在控制球杆系统时往往伴随较大的超调与一定程度的振荡. 本文认为这种现象是很自然的, 人的控制策略较为灵活与智能, 这也是人相较于机器人最大的优点. 然而, 人往往很难像数字控制器一样做到高精度, 高效率的控制. 另外, 进一步发现人的速度分量 $v_{H,\,x}$ 相对于分量 $v_{H,\,z}$ 幅值较小, 无明显规律, 更像是志愿者自己引入的随



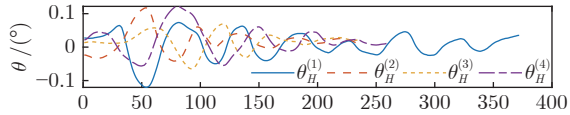
(a) 误差轨迹

(a) Error trajectories



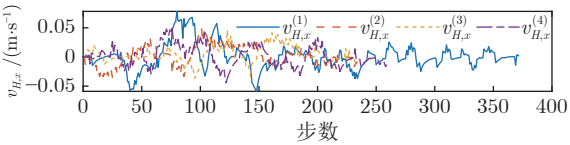
(b) 速度轨迹

(b) Velocity trajectories



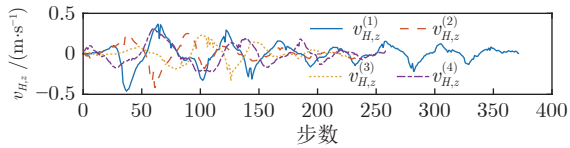
(c) 角度轨迹

(c) Angle trajectories



(d) 人 x 轴速度

(d) Human velocity in x-axis



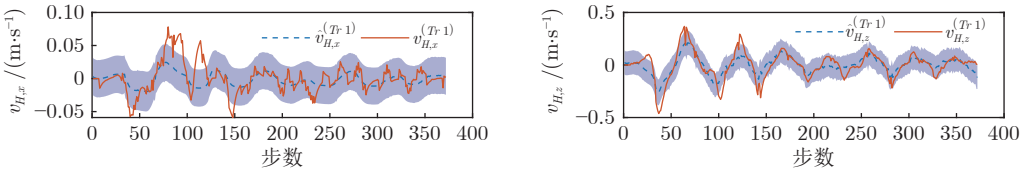
(e) 人 z 轴速度

(e) Human velocity in z-axis

图 8　志愿者控制过程中产生的部分轨迹图

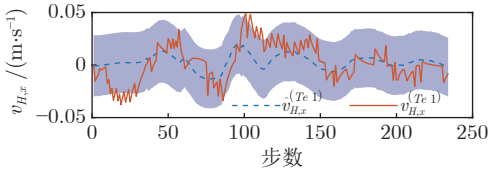
Fig.8　Some trajectories generated by the volunteers′ control process

机噪声. 本文利用 GPR 在训练数据上拟合人体控制策略预测模型, 即 $\hat{\pi}_H : \boldsymbol{s} \to N(\hat{\boldsymbol{v}}_H,\, \delta)$. 结果图 9 所示. 图 9 中阴影部分表示预测置信度为 68.2% 的区域 (GPR 的输出是 1 个高斯分布), 第 1 行的 2 幅子图分别表示在训练集中 1 条轨迹上的 $v_{H,\,x}$ 与 $v_{H,\,z}$ 的预测情况 (上标 Tr 表示). 第 2 ~ 4 行表示测试集中各速度分量的预测情况 (上标 Te 表示). 由图 9 可以看出, 无论是在训练集还是测试集中, $v_{H,\,x}$ 的预测均较差, 说明 GPR 方法较难从训练数据中寻得一种普遍规律, 这也证实了 $v_{H,\,x}$ 可能是志愿者自身引入的一种随机噪声的猜测. 而对于速度分量 $v_{H,\,z}$, 预测模型较为准确地预测了变化趋势. 由于人控制策略的高随机性与灵活性, 精确的预测其具体的幅值是不现实的. 本文得到的人体控制策略预测模型的预测值无论是在训练集还是测试集中, 其预测幅值误差均较小, 故该预测模型可使机器人在一定程度上了解人的控制规律.

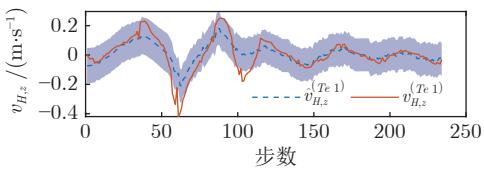


(a) 人 x 轴速度训练集轨迹 1 预测结果
(a) Prediction result for human x-axis velocity trajectory 1 in training set

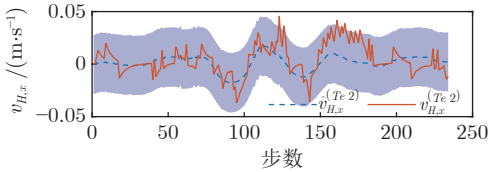
(b) 人 z 轴速度训练集轨迹 1 预测结果
(b) Prediction result for human z-axis velocity trajectory 1 in training set



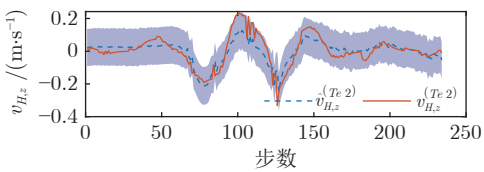
(c) 人 x 轴速度测试集轨迹 1 预测结果
(c) Prediction result for human x-axis velocity trajectory 1 in the test set



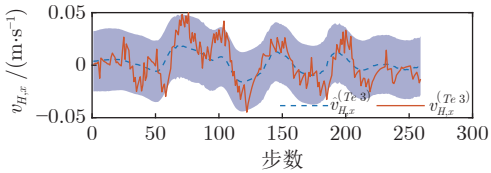
(d) 人 z 轴速度测试集轨迹 1 预测结果
(d) Prediction result for human z-axis velocity trajectory 1 in the test set



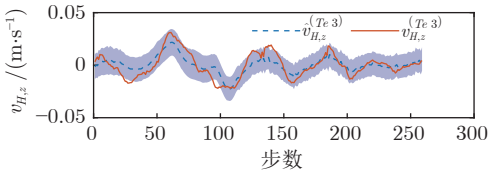
(e) 人 x 轴速度测试集轨迹 2 预测结果
(e) Prediction result for human x-axis velocity trajectory 2 in the test set



(f) 人 z 轴速度测试集轨迹 2 预测结果
(f) Prediction result for human z-axis velocity trajectory 2 in the test set



(g) 人 x 轴速度测试集轨迹 3 预测结果
(g) Prediction result for human x-axis velocity trajectory 3 in the test set



(h) 人 z 轴速度测试集轨迹 3 预测结果
(h) Prediction result for human z-axis velocity trajectory 3 in the test set

图 9 人体控制策略预测模型拟合结果图

Fig. 9 The fitting results of human-control policy prediction model

3.3 人机协作控制实验

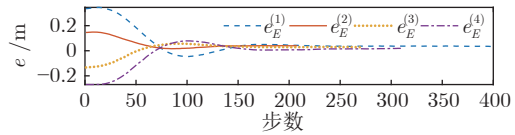
本节在图 6 所示的平台上对基于 GPR 与 DRL的分层人机协作控制方法进行实验验证.

首先验证顶层期望控制策略. 由于顶层期望控制策略是只针对非线性球杆系统设计的, 未考虑人引入的随机因素. 因此, 在该部分实验中保持 $p1$ 点固定不动, 以期望控制策略控制机器人, 即 $v_{R,x}=0$, $v_{R,z}=-\dot{\theta}L$. 其中 $\dot{\theta}$ 由期望控制策略即 Actor 网络输出得到. 实验结果如图 10 所示.

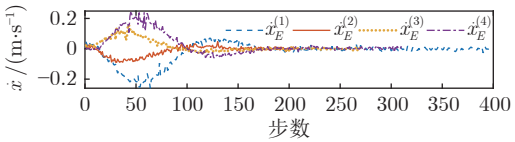
由图 10 可以发现, 无论小球从何初始位置出发, 该期望控制策略均能高效的完成控制任务 (每一步时间为 0.033 s, 故期望控制策略约在 6 s 内完成控制任务). 另外, 可以发现该实验结果与图 4 中的仿真结果非常相似, 更进一步的验证了该期望控制策略的有效性.

然而在实际人机协作任务中, 人也参与到球杆系统的控制过程中. 若仍以期望控制策略直接控制机器人, 协作任务很可能在人与机器人协同的总控制量下失败 (如人的过激控制量加上机器人的期望控制量, 使长杆的旋转速度过快, 从而使小球滚落长杆). 故本文考虑使机器人与人的总控制量趋向于期望控制策略的控制量, 即按第 2.2.2 节所述设计机器人末端速度控制律. 为了进一步突出该方法的有效性, 本文将人机协同控制的控制效果与期望控制策略和人单独控制球杆系统的控制效果作对比. 其实验效果如图 11 所示.

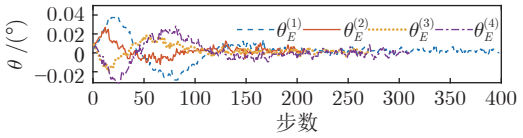
考虑到传统的主从式人机协作多为人主−机器人从模式, 即在协作任务中控制策略完全由人产生, 机器人多承担负重任务. 因此本文考虑固定机器人端 (即 $p2$ 点), 由人单独控制球杆系统来代表人主−机器人从的协作模式. 单独由人产生控制球杆系统的策略往往会带来较大幅度的振荡, 延长了整体控制时间, 降低了控制效率. 本实验的控制效率由使系统进入稳态区域的控制时间 t_s 体现, 稳态区域为稳定值正负 3 cm 所在的范围 (图 11 中的阴影部分). 如图 11 所示, 人单独控制策略下的球杆系统在 $t_{s,H}=9.57$ s 时进入稳态区域. 与顶层期望控制策略相比, 其效率明显更低, 并且最终较难精确地使小球停在目标位置处. 从振荡的角度看, 由于人在控制起始阶段往往采取过激的控制



(a) 误差轨迹
(a) Error trajectories



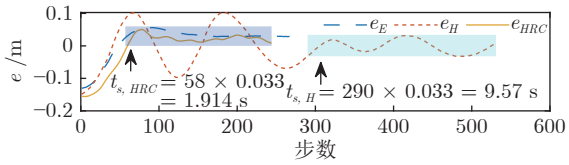
(b) 速度轨迹
(b) Velocity trajectories



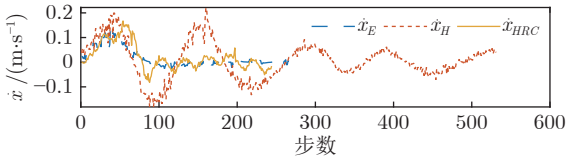
(c) 角度轨迹
(c) Angle trajectories

图 10　期望控制策略的实验验证

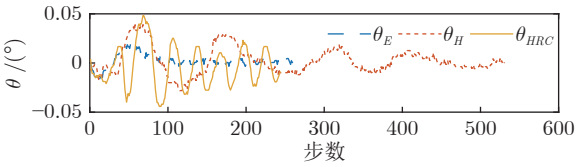
Fig. 10　The experimental validation of the expected control policy



(a) 误差轨迹
(a) Error trajectories



(b) 速度轨迹
(b) Velocity trajectories



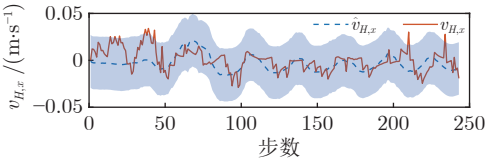
(c) 角度轨迹
(c) Angle trajectories

图 11　人机协作控制效果的实验验证

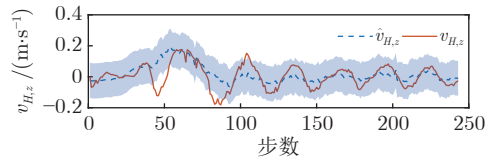
Fig. 11　The experimental validation of the HRC control

量以达到快速降低误差的目的, 其并没有考虑长远的系统变化. 而基于 DDPG 的顶层期望控制策略的目标如式 (1) 所示, 是使长远的累计奖励最大化, 其考虑到了系统的长远变化, 并在快速性与稳定性之间做出权衡, 使系统不会有过大超调. 另外, 如第 3.2 节所述, 人的控制精度相对于数字控制器较低是很自然的. 因此, 用人机协作来提高协作任务的控制效率与精度是有必要的. 可以发现, 虽然人机协作的控制效果与期望控制策略的控制效果并不是理想情况下的完全一致, 但是两者的小球位置误差与速度轨迹相差不大. 单独由人作控制决策相比, 人机协作明显提升了控制效率 ($t_{s,\,HRC} = 1.914$ s), 验证了本文方法的高效性.

进一步对比人机协作与期望控制策略之间的控制曲线可以发现, 人机协作的控制曲线存在一定的抖动, 这在长杆的倾角变化轨迹中尤为明显. 显然, 这是人体控制策略预测模型的预测误差造成的. 如图 12 所示, 可以发现虽然预测模型能较为准确地预测 $v_{H,\,z}$ 的变化趋势, 但是对于其幅值的预测存在一定的误差, 使得机器人并未完全补偿人的控制量, 从而使人机协作的总控制量中仍然包含残留着的人的控制量, 因此造成了长杆倾角抖动. 然而, 长杆倾角的抖动对球杆系统的控制目的 (使小球停在目标位置处) 并未造成较大的影响.



(a) 人 x 轴速度预测结果
(a) Prediction result for huamnin x-axis



(b) 人 z 轴速度预测结果
(b) Prediction result for huamn velocity in z-axis

图 12　人体控制策略预测模型预测结果

Fig. 12　The prediction result of the human-control policy prediction model

4　结束语

本文针对主从式人机协作效率较低的问题设计了一种基于 GPR 和 DRL 的分层人机协作控制方法. 顶层使用 DRL 算法在模型未知的情况下设计了一种有效的次优非线性控制策略, 并将其作为期望控制策略以引导人机协作控制过程. 底层使用 GPR 方法拟合人体控制策略预测模型, 为机器人建立人体行为认知模型, 从而提升机器人在协作过程中过的主动性, 提高协作效率同时降低人未知随机行为带来的不利影响. 进而, 基于期望控制策略与认知模型设计机器人的末端速度控制律. 最后由实验对比发现, 本文所提的人机协作控制方法较人主–机器人从协作控制具有更高的协作效率, 体现了本文方法的高效性.

本文用 GPR 拟合人体控制策略之后只使用了输出的均值来构建机械臂的控制律, 未利用协方差信息. 如何利用协方差信息来构建构更加具有鲁棒性的机械臂控制律是未来的一个研究要点. 另外, 如何提升在人体控制策略预测模型的预测精度也将是未来的工作之一.

References

1　Amirshirzad N, Kumru A, Oztop E. Human adaptation to human-robot shared control. *IEEE Transactions on Human-Machine Systems*, 2019, **49**(2): 126–136

2　Wojtara Y, Murayama H, Howard M, Shimoda S, Sakai S, Fujimoto H, et al. Human-robot collaboration in precise posi-

tioning of a three-dimensional object. *Automatica*, 2009, **45**(2): 333−342

3 Dumora J, Geffard F, Bidard C, Brouillet T, Fraisse P. Experimental study on haptic communication of a human in a shared human-robot collaborative task. In: Proceedings of the 2012 IEEE/ RSJ International Conference on Intelligent Robots and Systems. Vilamoura, Portugal: IEEE, 2012. 5137−5144

4 Karayiannidis Y, Smith C, Kragic D. Mapping human intentions to robot motions via physical interaction through a jointly-held object. In: Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication. Edinburgh, UK: IEEE, 2014. 391−397

5 Karayiannidis Y, Smith C, Vina F E, Kragic D. Online kinematics estimation for active human-robot manipulation of jointly held objects. In: Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, Japan: IEEE, 2013. 4872−4878

6 Burdet E, Milner T E. Quantization of human motions and learning of accurate movements. *Biological Cybernetics*, 1998, **78**(4): 307−318

7 Maeda Y, Hara T, Arai T. Human-robot cooperative manipulation with motion estimation. In: Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Maui, USA: IEEE, 2001. 2240−2245

8 Corteville B, Aertbelien E, Bruyninckx H, Schutter J D, Brussel H V. Human-inspired robot assistant for fast point-to-point movements. In: Proceedings of the 2007 IEEE International Conference on Robotics and Automation. Roma, Italy: IEEE, 2007. 3639−3644

9 Miossec S, Kheddar A. Human motion in cooperative tasks: Moving object case study. In: Proceedings of the 2009 IEEE International Conference on Robotics and Biomimetics. Bangkok, Thailand: IEEE, 2009. 1509−1514

10 Sheng W H, Thobbi A, Gu Y. An integrated framework for human-robot collaborative manipulation. *IEEE Transactions on Cybernetics*, 2015, **45**(10): 2030−2041

11 Thobbi A, Gu Y, Sheng W H. Using human motion estimation for human-robot cooperative manipulation. In: Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA: IEEE, 2011. 2873−2878

12 Deng Z, Mi J P, Han D, Huang R, Xiong X F, Zhang J W. Hierarchical robot learning for physical collaboration between humans and robots. In: Proceedings of the 2017 IEEE International Conference on Robotics and Biomimetics. Macau, China: IEEE, 2017. 750−755

13 Agravante D J, Cherubini A, Bussy A, Kheddar A. Humanhumanoid joint haptic table carrying task with height stabilization using vision. In: Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, Japan: IEEE, 2013. 4609−4614

14 Agravante D J, Cherubini A, Bussy A, Gergondet P, Kheddar A. Collaborative human-humanoid carrying using vision and haptic sensing. In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation. Hong Kong, China: IEEE, 2014. 607−612

15 Mainprice J, Berenson D. Human-robot collaborative manipulation planning using early prediction of human motion. In: Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, Japan: IEEE, 2013. 299−306

16 Maria K, Muhammad A H, Danijela R D, Axel G. Robot learning of industrial assembly task via human demonstrations. *Autonomous Robots*, 2019, **43**(1): 239−257

17 Ghadirzadeh A, Butepage J, Maki A, Kragic D, Bjorkman M. A sensorimotor reinforcement learning framework for physical human-robot interaction. In: Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems. Daejeon, South Korea: IEEE, 2016. 2682−2688

18 Wang P, Liu H Y, Wang L H, Gao R X. Deep learning-based human motion recognition for predictive context-aware human-robot collaboration. *CIRP Annals−Manufacturing Technology*, 2018, **67**(1): 17−20

19 Wang Z, Peer A, Buss M. An HMM approach to realistic haptic human-robot interaction. In: Proceedings of the World Haptics 3rd Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. Teleoperator System. Salt Lake City, USA: 2016. 374−379

20 Mainprice J, Berenson D. Learning human-robot collaboration with POMDP. In: Proceedings of the 2013 International Conference on Control, Automation and Systems. Gyeongju, South Korea: IEEE, 2013. 1238−1243

21 Hawkins K P, Vo N, Bansal S, Bobick A F. Probabilistic human action prediction and wait-sensitive planning for responsive human-robot collaboration. In: Proceedings of the 2013 13th IEEE-RAS International Conference on Humanoid Robots. Atlanta, USA: 2013. 499−506

22 Lillicrap T P, Hunt J J, Pritzel A, Heess N, Erez T, Silver D, et al. Continuous control with deep reinforcement learning. In: Proceedings of the 2016 International Conference on Learning Representations. San Juan, Puerto Rico: IEEE, 2016. 1−14

23 Mnih V, Kavukcuoglu K, Silver D, Rusu A A, Veness J, Bellemare M G, et al. Human-level control through deep reinforcement learning. *Nature*, 2015, **518**(7540): 529−533

24 Hado V H, Guez A, Silver D. Deep reinforcement learning with double Q-learning. In: Proceedings of the 2016 AAAI Conference on Artificial Intelligence. Arizona, USA: 2016. 2094−2100

25 Silver D, Lever G, Hess N, Degris T, Wierstra D, Riedmiller M. Deterministic policy gradient algorithms. In: Proceedings of the 2014 International Conference on Machine Learning. Beijing, China: 2014. 605−619

26 Espersson M. Vision Algorithms for Ball on Beam and Plate[Master thesis], Lund University, Sweden, 2010

金哲豪 浙江工业大学信息工程学院硕士研究生. 主要研究方向为人机协作. E-mail: jzh839881963@163.com
(**JIN Zhe-Hao** Master student at the College of Information Engineering, Zhejiang University of Technology. His main research interest is human-robot collaboration.)

刘安东 浙江工业大学信息工程学院讲师. 主要研究方向为模型预测控制和网络化控制系统.
E-mail: lad@zjut.edu.cn
(**LIU An-Dong** Lecturer at the College of Information Engineering, Zhejiang University of Technology. His research interest covers model predictive control and networked control system.)

俞 立 浙江工业大学信息工程学院教授. 主要研究方向为无线传感网络, 网络化控制系统和运动控制. 本文通信作者. E-mail: lyu@zjut.edu.cn
(**YU Li** Professor at the College of Information Engineering, Zhejiang University of Technology. His research interest covers wireless sensor networks, networked control systems and motion control. Corresponding author of this paper.)