基于仿人自适应DDPG算法的机械臂运动规划研究

汪红,解伟,田莎莎*,帖军,昂寅
(中南民族大学 计算机科学学院 & 农业区块链与智能管理湖北省工程研究中心 & 湖北省制造企业智能管理工程技术研究中心,武汉 430074)

摘 要 现代大棚高架栽培技术中作物种植密度大,机械臂采摘容易受到旁边植株影响.针对这一问题,将人手臂的运动约束应用到机械臂的DDPG运动规划算法中,从而让机械臂具有人手臂的灵活性.所提出的仿人自适应DDPG算法新增自适应采样策略来对经验缓冲池中轨迹经验的优先级进行自适应动态调整,从而避免算法陷入局部最优,加快算法的训练速度.将仿人自适应DDPG算法在PyBullet仿真平台上进行实验,经过300轮训练,最终机械臂的抓取成功率高达96.5%,且机械臂在执行抓取操作的过程中各关节的运动范围皆符合人手臂运动时的关节范围,证明机械臂具有人手臂的运动特性,能更好地应用于大棚高架栽培作物的采摘.

关键词 机械臂;奖励函数;运动约束;自适应
中图分类号 O625.67;O643.3 **文献标志码** A **文章编号** 1672-4321(2023)03-0334-09
doi:10.20056/j.cnki.ZNMDZK.20230307

Research on motion planning of manipulator based on humanoid adaptive DDPG algorithm

WANG Hong, XIE Wei, TIAN Shasha*, TIE Jun, ANG Yin
(College of Computer Science & Hubei Provincial Engineering Research Center of Agricultural Blockchain and Intelligent Management & Hubei Provincial Engineering Research Center for Intelligent Management of Manufacturing Enterprises, South-Central Minzu University, Wuhan, 430074, China)

Abstract In the modern greenhouse elevated cultivation technology, the crop planting density is high, and the picking of the robotic arm is easily affected by the plants next to it. In response to this problem, the motion constraints of the human arm is applied to the DDPG motion planning algorithm of the robotic arm, so that the robotic arm has the flexibility of a human arm. The human-like adaptive DDPG algorithm proposed adds an adaptive sampling strategy to dynamically adjust the priority of the trajectory experience in the experience buffer pool, so as to avoid the algorithm from falling into local optimum and speed up the training speed of the algorithm. The humanoid adaptive DDPG algorithm was tested on the PyBullet simulation platform. After 300 rounds of training, the final grasping success rate of the robotic arm was as high as 96.5%, and the range of motion of each joint during the grasping operation of the robotic arm was consistent with the joint range of the human arm movement proves that the robotic arm has the movement characteristics of the human arm, and can be better applied to the picking of elevated cultivation crops in greenhouses.

Keywords mechanical arm; reward function; motion constraint; adaptive dynamic adjustment

随着智慧农业的普及,各地农企对智能大棚的投入越来越多,智能大棚的相关技术也日益成熟.草莓、番茄等果蔬的大棚高架栽培技术[1-2]已经遍及全国各地.果蔬的高架栽培技术,符合人体工学,可显著降低劳动强度,提高工作效率,能将生产人员从繁重的劳动中解放出来,因而日益受到关注.为

收稿日期 2022-05-25 *** 通信作者** 田莎莎,研究方向:机器学习,E-mail: shashatian77@mail.scuec.edu.cn
作者简介 汪红(1968-),女,副教授,研究方向:机器学习,E-mail: wanghong_2010@foxmail.com
基金项目 国家民委中青年英才培养计划(MZR20007);湖北省科技重大专项(2020AEA011);武汉市科技计划应用基础前沿项目(2020020601012267);中央高校基本科研业务费专项资金资助项目(CZQ21026)

了进一步解放劳动力，提高劳动效率，不少现代化大棚已经开始使用机械臂来采摘果蔬.由于大棚高架栽培的特殊性，植株高度一般在1~2 m之间，横向种植密度较大，机械臂的采摘路径容易受到两边植株的影响，不能进行大幅度的横向移动，因此一种更高效率的针对大棚高架栽培的采摘机械臂路径规划成为智慧农业的研究热点.

以往，大多数研究者结合人类手臂运动规律和机器人运动学来实现仿人机械臂.LIU等人[3]提出一种改进的快速探索随机树算法用于人形机械臂路径规划，仿真结果显示，所提出的方法可以有效地减少路径规划时间和路径长度.PAUS等人[4]提出了一种结合机器人位置和覆盖路径规划的通用方法，该方法考虑了避免碰撞和静态稳定性等约束，使用仿人机器人ARMAR-III进行实验并达到了很好的表现.YANG等人[5]开发了一个框架，使机器人能够从人类导师那里学习运动和刚度特征，所提出的框架可以有效地实现刚度泛化和运动泛化，在双臂Baxter机器人上进行了实验测试，验证了所提出方法的有效性.ROSELL等人[6]使用动作捕捉系统采集人在不同物体上的抓握动作，再映射到机械臂上，使用基于双向多目标采样的规划器规划运动，实现了机械臂的仿人抓握.GONG等人[7]通过分析逆运动学方法将HAMP映射到机器人关节角度以控制机械臂并在iiwa机器人上进行物理实验，证明了所提方法的有效性.GARCIA等人[8]用磁性跟踪器和感应手套在执行操作任务时对操作员手腕的位置和方向进行采样，通过运动学求解将捕获的数据映射到双臂机器人系统，结合RRT算法实现了更加接近人类的机械臂运动规划.

随着深度强化学习的不断发展，其优异的决策性能为解决机械臂路径规划问题提供新的思路.2015年，Deepmind提出了Deep Deterministic Policy Gradient（DDPG）算法[9].在用DDPG算法解决连续控制问题方面，研究者们也进行了大量的研究[10-13].WEN等人[14]提出了一种新的避障算法，基于现有的深度确定性策略梯度（DDPG）学习框架.使用DDPG的机械臂避障通过自学习实现，解决了高维状态输入和多个返回值带来的收敛问题.KUO等人[15]提出了一种基于机器学习和模糊逻辑的运动控制器，使用了DDPG算法允许仿人机器人自学习并自主规划其手臂的运动和关节角度.人形机器人在将模糊逻辑与DDPG算法相结合的实验中表现出令人满意的学习结果.LI[16]提出了多重经验池重播双延迟DDPG来平衡集成能源系统中的随机功率扰动.ZHANG[17]使用异步方式设计了一种经验缓冲池数据采集方案，提出了异步间歇式DDPG算法来解决复杂环境下的连续控制问题.

为了解决大棚高密度植株情况下机械臂采摘的问题，本文提出了一种基于仿人自适应DDPG算法的机械臂运动规划方法.

本文的贡献如下：

（1）从强化学习的角度，对机械臂进行路径规划，让机械臂在大棚高密度植株的情况下采摘果实时具有了人手臂的灵活性；

（2）采用了惯性传感器（IMU）获取人手臂运动的数据，并对数据进行分析，得到了人手臂运动时的规律，进一步根据规律提取出仿人手臂运动约束，最后将约束转变为奖励函数，利用强化学习算法进行训练；

（3）本文所提及的DDPG算法，在前人的基础上，加入了自适应的优先经验回放，防止算法训练后期陷入局部最优解，使算法更快速的收敛，提高了算法的训练效率.

1　人手臂运动规律分析

1.1　人手臂运动数据检测

本文使用惯性传感器来检测人手臂的运动数据，对其运动规律进行建模.惯性传感器主要由加速度计和陀螺仪构成，可以测量物体运动的加速度，角度等数据.惯性传感器可以检测每个时刻三轴方向的加速度，将加速度进行一次积分可以得到速度，进而将速度进行积分，就可以得到位移.结合位置的初始值就能递推得到每个关节在每个时刻的位置.惯性传感器测量物体位置的原理如图1所示.

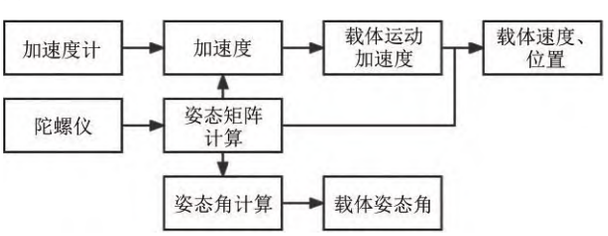


图1　基于惯性传感器的位置计算原理

Fig. 1　Position calculation principle based on inertial sensor

1.2　人手臂运动约束分析

为了更好地为DDPG算法设计奖励函数，我们将动作空间划分为三个部分，以人体手臂自然下垂

时肩关节和肘关节所在平面为分界线,分为上层空间、中层空间和下层空间,如图2所示.本文对每个空间中的人手臂运动范围进行了理论研究,最终得到了每个空间中的肩夹角和肘夹角的角度范围,为后续奖励函数的设计提供理论依据.分别在实验者的肩关节,肘关节和腕关节绑定惯性传感器模块,采集手臂运动过程中的位置信息.

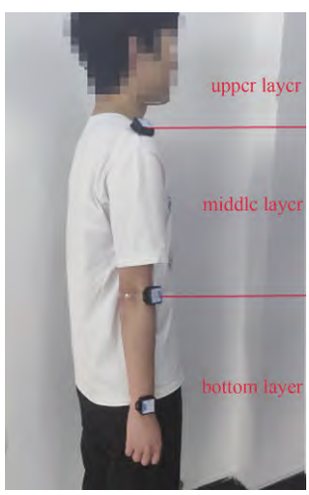


图2 空间划分

Fig. 2 Space division

对关节的坐标信息进行处理可以得到关节的角度信息.如图3所示,在空间坐标系中为各个关节进行建模,A、B、C分别表示肩关节,肘关节和腕关节.肩夹角γ为大臂与上层分界线的夹角,逆时针为正.肘夹角β为向量$\overrightarrow{AB}$与向量$\overrightarrow{BC}$的夹角.各关节位置在XOY平面内的投影分别是A'、B'、C',目标点的投影为G',肩关节和目标点投影的向量为$\overrightarrow{A'G'}$,肩关节和腕关节投影的向量为$\overrightarrow{A'C'}$,向量$\overrightarrow{A'G'}$和$\overrightarrow{A'C'}$的夹角定义为腕关节的转动角α.α、β、γ的大小分别由公式(1)、(2)和(3)给出.其中l_1为大臂长度,l_2为小臂长度.

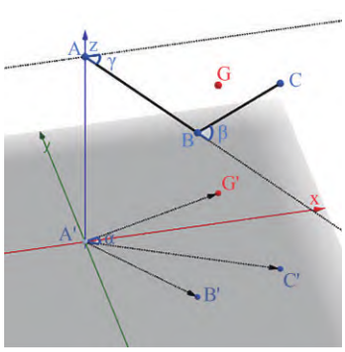


图3 各关节的空间位置关系

Fig. 3 The spatial position of each joint

$$\alpha = \arcsin \frac{\overrightarrow{A'C'} \cdot \overrightarrow{A'G'}}{\left|\overrightarrow{A'C'}\right| \cdot \left|\overrightarrow{A'G'}\right|}, \tag{1}$$

$$\beta = \arcsin \frac{\overrightarrow{AB} \cdot \overrightarrow{BC}}{l_1 l_2}, \tag{2}$$

$$\gamma = \arcsin \frac{\overrightarrow{AB} \cdot \vec{Z}}{l_1}. \tag{3}$$

如图4所示,在上层空间中肩关节的最大角度为手臂最大外展角度170°,最小角度为肘关节达到最大前屈140°,腕关节达到上层空间时的角度γ',如公式(4)所示,其中β'为肘夹角最大前屈度数.

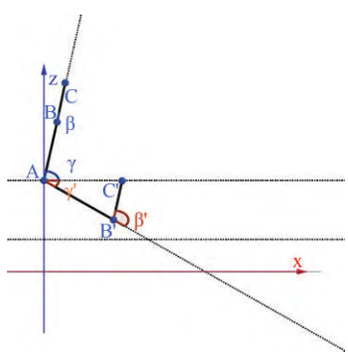


图4 上层空间肩夹角范围

Fig. 4 The range of shoulder angles in upper space

$$\gamma' = arccos \frac{\left(l_1^2 + l_2^2 - 2l_1 l_2 \cos\beta'\right)^2 + l_1^2 - l_2^2}{2l_1\left(l_1^2 + l_2^2 - 2l_1 l_2 \cos\beta'\right)}. \tag{4}$$

如图5所示,在上层空间中肘夹角的最大值为最大前屈度数140°,最小值为0.

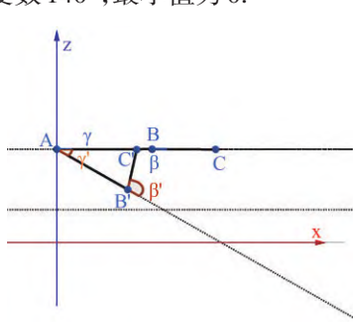


图5 上层空间肘夹角范围

Fig. 5 The range of elbow angles in upper space

如图6所示,在中层空间中肩夹角的最大值为大臂自然下垂且小臂与大臂垂直时,为90°,最小值为0.

如图7所示,在中层空间中肘夹角的最大值为最大前屈度数140°,最小值为手臂前伸时,为0.

如图8所示,在下层空间中肩夹角最大值为手臂自然下垂时的角度90°,最小值为小臂与下层空间分界线垂直时的夹角γ',此时γ'的角度由公式(5)给出.

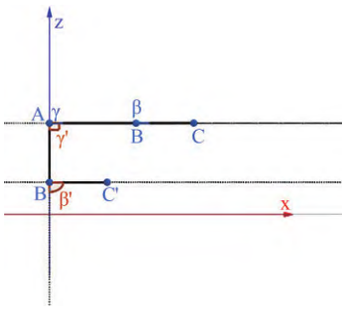


图6　中层空间肩夹角范围

Fig. 6　The range of shoulder angles in middle space

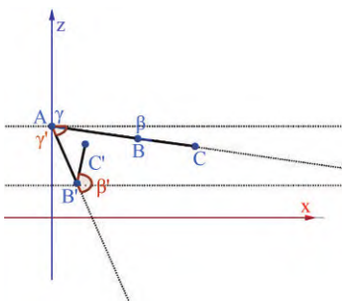


图7　中层空间肘夹角范围

Fig. 7　The range of elbow angles in middle space

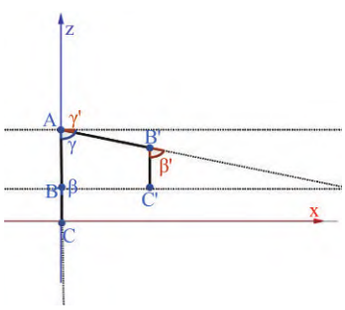


图8　下层空间肩夹角范围

Fig. 8　The range of shoulder angles in lower space

$$\gamma' = \arcsin\frac{l_1 - l_2}{l_1}. \tag{5}$$

如图9所示，在下层空间中肘夹角的最大值为腕关节达到手臂自然下垂状态时肘关节的位置时

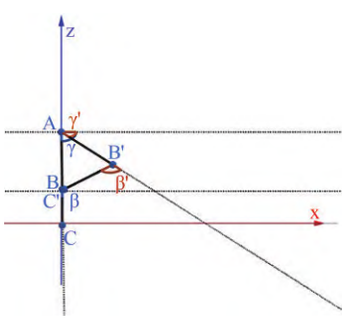


图9　下层空间肘夹角范围

Fig. 9　The range of elbow angles in lower space

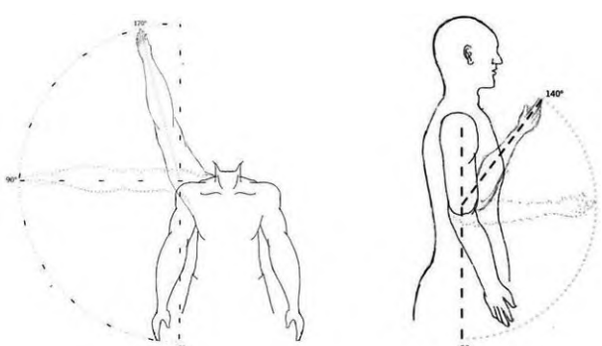
的角度β'，具体值见公式(6).

$$\beta' = \pi - \arccos\frac{l_2}{2l_1}. \tag{6}$$

另外，为了加快抓取效率，专门设计了腕关节转动角α，其定义如公式(1)所示.腕关节转动角α越小，机械臂末端执行器越接近物体，可以保证机械臂更加稳定地抓取到物体，提高了算法的训练效率.

2　仿人约束机械臂建模

人手臂的灵活得益于手臂上的三个关节，即肩关节，肘关节和腕关节，这些关节之间的配合能使人们很轻易的抓取身边的物体.在实际应用场景中，机械臂都是搭载在一个可以移动的小车上，通过移动就可以达到任何想要的采摘范围，因此本文只考虑右臂在第一卦限的情况.由于人臂每个关节都有一定的局限性，并不能达到所有的运动角度，经过大量实验测量，得出了右臂的运动范围，如图10所示，(a)显示了手臂外展时肩关节的活动范围是[0, 170°]，(b)显示了肘关节的活动范围是[0, 140°].右臂具体的运动范围在上一节中已经给出.



(a) 肩夹角范围　　(b) 肘夹角范围

图10　人手臂的肩夹角和肘夹角运动范围

Fig. 10　The angle between the shoulder and the elbow of the human arm

本文实验采用的是openMANIPULATOR-X[18]机械臂，该机械臂有5个自由度.将机械臂和人手臂的各关节进行对应，如图11所示，机械臂的joint 2对应人手臂的肩关节，joint 3对应人手臂的肘关节，joint 4对应人手臂的腕关节.

3　基于优先经验回放的自适应DDPG算法

3.1　DDPG算法

DDPG算法中，存在两个神经网络，分别是Actor网络和Critic网络.Critic网络对Actor网络的动作进行评估，Actor网络根据Critic网络的评分进行

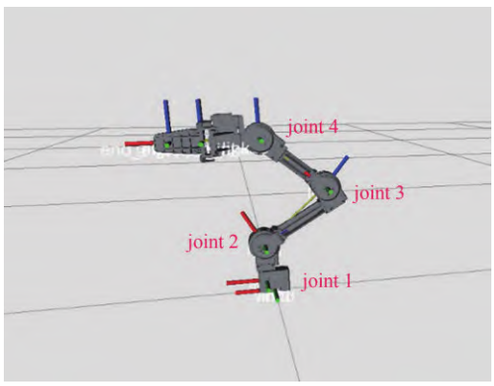


图11 人手臂和openMANIPULATOR-X之间的对应关系

Fig. 11 The correspondence between the human arm and open MANIPULATOR-X

参数更新,而Critic网络则根据环境反馈的奖励进行参数更新.

由于算法中包含两种神经网络,因此需要设计两个学习率,Actor网络的学习率为α_{Actor},Critic网络的学习率为α_{Critic}.在探索策略上,本文使用了ε-greedy策略,如公式(7)所示,以1-ε的概率选择随机动作,以ε的概率根据策略网络π选择动作:

$$a=\begin{cases} a_t^*, & 1-\epsilon \\ \text{random action}, & \epsilon \end{cases}. \tag{7}$$

DDPG算法借鉴了DQN[19]中的训练技巧,继续沿用了DQN中的target网络和经验回放机制,加入了随机噪声N,从而让训练更加的稳定.在Critic网络中,其均方差损失函数为:

$$L=\frac{1}{m}\sum_{j=1}^{m}\left[Q_w\left(s_j,a_j\right)-\left(r+\gamma Q_{\bar{w}}\left(s_{j+1},a_{j+1}\right)\right)\right]^2, \tag{8}$$

其中$a_{j+1}=\mu_{\bar{\theta}}\left(s_{j+1}\right)$,$\bar{w}$是Q-target网络的参数,$\bar{\theta}$是策略网络所对应target网络的参数.通过神经网络的梯度反向传播来更新Critic当前的网络参数w.

在Actor网络中,损失函数为:

$$J(\theta)=-\frac{1}{m}\sum_{i=1}^{m}Q_w\left(s_i,a_i\right), \tag{9}$$

其中$a_i=\mu_\theta\left(s_i\right)$,通过梯度反向传播更新策略网络当前的参数$\theta$.

而目标网络参数$\bar{w}$和$\bar{\theta}$则由每隔一段时间通过当前网络参数w和θ进行软更新得到,更新公式如下:

$$\bar{w}\leftarrow\sigma w+(1-\sigma)\bar{w}, \tag{10}$$

$$\bar{\theta}\leftarrow\sigma\theta+(1-\sigma)\bar{\theta}. \tag{11}$$

3.2 奖励函数设计

总结上文对于人手臂运动的约束分析,得到表1.

表1 肩夹角和肘夹角的范围

Tab. 1 The range of elbow angles and shoulder angles

空间位置	肩夹角(°)	肘夹角(°)
上层空间	$\left[-\arccos\frac{\left(l_1^2+l_2^2-2l_1l_2\cos\beta'\right)^2+l_1^2-l_2^2}{2l_1\left(l_1^2+l_2^2-2l_1l_2\cos\beta'\right)},170\right]$	[0,140]
中层空间	[0,90]	[0,140]
下层空间	$\left[-90,-\arcsin\frac{l_1-l_2}{l_1}\right]$	$\left[0,\pi-\arccos\frac{l_2}{2l_1}\right]$

基于表1,针对大棚高密度植株情况下机械臂采摘的问题,本文对机械臂的肩夹角和肘夹角进行了仿人约束,并将这些约束融入DDPG算法的奖励函数.

3.2.1 肩夹角和肘夹角的范围约束

如表1所示,在上层空间,中层空间,下层空间中,其肩夹角和肘夹角的角度范围也不同,当肩夹角和肘夹角满足所在空间的角度范围时,奖励函数$r_2=0$.当肩夹角和肘夹角不满足所在空间的角度范围时,此时奖励函数$r_2=-\left(k_1\tan\Delta\gamma+k_2\tan\Delta\beta\right)$.其中,$k_1$和$k_2$分别表示肩夹角和肘夹角对奖励函数的权重占比,由于肩夹角和肘夹角对奖励函数的影响同样重要,这里取$k_1=k_2=\frac{1}{2}$,$\Delta\gamma$和$\Delta\beta$分别表示当前肩夹角和肘夹角与人手臂抓取时的范围偏差.

3.2.2 肩夹角和肘夹角的大小约束

本文除对关节角度进行分析外,还分析了人手臂在抓取物体时,肩夹角和肘夹角的关系.如图12所示,经过大量的抓取实验,研究发现当目标物体处于上层空间和中层空间时,按照人类抓取物体的习惯,此时的肩夹角往往大于肘夹角.而当目标物体位于下层空间时,肩夹角和肘夹角的关系则没有特殊的规律.

当目标点在上层空间和中层空间时,人手臂在进行抓取物体的过程中,肩夹角要小于肘夹角.为了满足这一人手臂运动特性,本文设计了相应的奖励函数,当肩夹角$\beta\geqslant\gamma$时,$r_3=0$,当$\beta<\gamma$时,$r_3=$

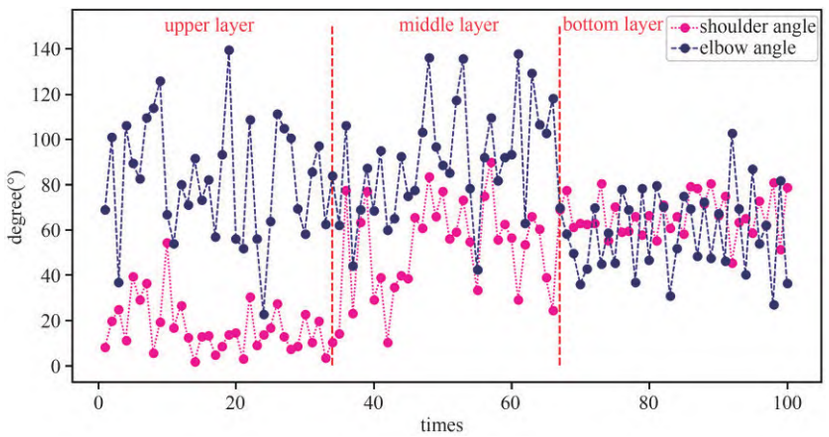


图 12　肩夹角和肘夹角的大小关系

Fig. 12　The magnitude of the shoulder angle and the elbow angle

$-e^{\arctan(\gamma-\beta)}$.

综上所述，当目标点在上层空间和中层空间时，奖励函数为 $r=\tau r_1+(1-\tau)(r_2+r_3)$，当目标点位于下层空间时，奖励函数为 $r=\tau r_1+(1-\tau)r_2$.

3.2.3　腕关节转动角

在机械臂作业过程中，当机械臂末端达到目标点，收紧末端执行器即可抓取目标，为了简化作业过程，本文只考虑机械臂末端是否到达目标点. 当机械臂末端与目标点之间的距离（即腕关节 C 点与目标点 G 之间的距离）小于 1 cm 时，则视为抓取成功. CG 之间的距离记为 dis，$\text{dis}=\sqrt{(X_G-X_C)^2+(Y_G-Y_C)^2+(Z_G-Z_C)^2}$. 我们进行了大量的右臂抓取实验，针对传感器采集的数据进行分析，实验表明，腕关节转动角大小 α 和 dis 之间存在正相关关系，如图 13 所示.

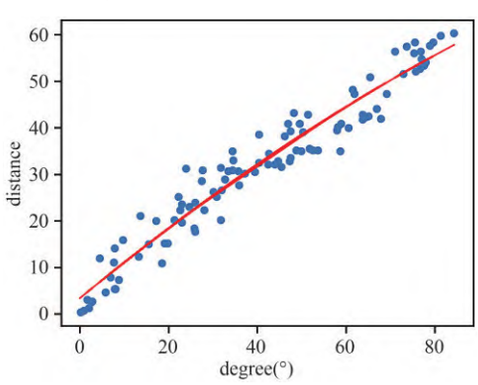


图 13　腕关节转动角和距离的关系

Fig. 13　The relationship between wrist rotation Angle and distance

随着腕关节转动角度的增大，腕关节与目标点之间的距离就越远，在设计奖励函数时，应当考虑腕关节转动角的影响. 通过腕关节转动角来一定程度上引导腕关节朝着目标点靠近，从而更加快速的接近目标点. 当 $\text{dis}<1\ cm$ 时，即视为成功抓取到物体，奖励函数 $r_1=0$，当 $\text{dis}\geqslant 1\ cm$ 时，需要同时考虑 dis 和腕关节转动角 α 的影响，此时奖励函数 $r_1=-(e^{\text{dis}}+\tan\alpha)$.

3.3　自适应经验采样策略

DDPG 算法采用经验缓冲池保存历史轨迹数据，用深度神经网络对经验缓冲池数据采样之后进行训练来得到 Actor 和 Critic 网络的参数. 而在现存的 DDPG 算法中，很多都是采用在经验缓冲池中随机抽样轨迹信息或者采样 TD-error 值较大的轨迹信息用于训练. 随机采样的方式会忽略掉重要的经验信息，而以 TD-error 为衡量标准，又会造成有些经验可能被采样很多次，而其他经验甚至从未被采样到就被移出经验池. 为了防止上述情况的发生，进一步提高 DDPG 算法的训练速度，本文提出了自适应经验采样策略. 该策略可以避免 TD-error 大的轨迹信息被多次采用造成的局部收敛问题.

在对经验缓冲池中的轨迹信息进行采样前，先根据式（12）计算出经验缓冲池中每条轨迹经验的初始优先级：

$$p_i=|\delta_i|+\epsilon\ , \tag{12}$$

之后按照式（13）在固定步数之后对经验池中每条轨迹的经验优先级进行更新：

$$P(i)=\frac{p_i+c\sqrt{\dfrac{\ln p_i}{2N+1}}}{\sum_i^M\left(p_i+c\sqrt{\dfrac{\ln p_i}{2N+1}}\right)}\ , \tag{13}$$

其中 N 表示采样到这条经验的次数，M 表示经验缓冲池中经验的条数，参数 c 是一个大于 0 的数，增加了采样的随机性，不至于每次采样都抽取到相同的经验，c 越大采样到其他经验的概率就越大. 随着采样次数 N 的增加，其采样优先级 $P(i)$ 会自适应减小，

从而保证了在训练后期同一条数据被重复采样的概率大大降低,从而避免算法的局部收敛,加快训练后期的收敛速度.为验证本文所提出的自适应经验采样策略的有效性,实验对比了传统PER算法[20]的经验采样策略和本文策略的算法收敛性能,具体效果如图14所示.

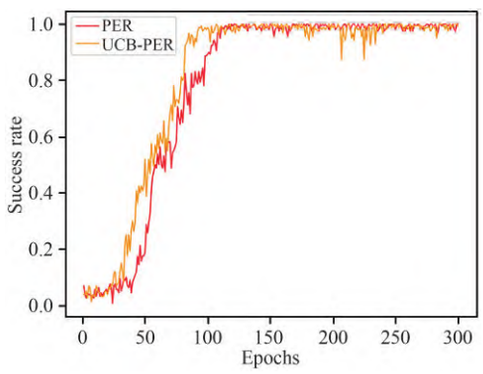


图14 两种经验采样策略的算法收敛性对比

Fig. 14 Convergence comparison of two experience extraction strategies

3.4 仿人自适应DDPG算法

本文所提出的仿人自适应DDPG算法在原基本算法的基础上,进行了如下3个方面的改进:

(1)针对大棚高密度植株情况下机械臂采摘的问题,对人手臂运动约束进行分析,将仿人约束融入DDPG算法的奖励函数中,从而解决高密度植株情况下的机械臂运动规划问题,同时提高算法的训练速度;

(2)提出自适应经验采样策略,避免DDPG算法训练后期的局部收敛问题,提高训练速度;

(3)同时,本文采用事后经验回放策略,选取目标子集G后,每次随机选择k个在这个轨迹之后的状态作为新目标.将原目标g的轨迹$(s_t\|g, a_t, r_t, s_{t+1}\|g)$进行转换,让其变成新目标$g'$的轨迹:$(s_t\|g', a_t, r_t, s_{t+1}\|g')$;

本文所提出的仿人自适应DDPG算法的伪代码如算法1所示.

当目标点在上层空间和中层空间时,奖励函数为$r = \tau r_1 + (1-\tau)(r_2 + r_3)$,当目标点位于下层空间时,奖励函数为$r = \tau r_1 + (1-\tau) r_2$.其中$r_1$是引导机械臂抓取目标点的奖励函数,$r_2$和$r_3$的作用是保证机械臂在运动时能具有人手臂运动的特征.τ为是奖励函数的权重,经过多组对比实验,$\tau = 0.5$时,机械臂的训练速度最快,如图15所示.

算法1

输入:机械臂的初始环境状态S_0和目标点的位置g.

输出:机械臂t时刻所执行的动作a_t.

1:随机初始化DDPG算法中Actor-Critic网络的参数$\theta, \bar{\theta}, w, \bar{w}$.

2:初始化经验池R.

3:for episode=1,M do:

4: 初始化PyBullet环境中的机械臂角度.

5: 合法的范围内随机生成目标点g.

6: 获取当前环境状态S_t.

7: for t=0,T-1 do:

8: 根据公式(7)选取动作a_t.

9: 执行动作a_t.

10: 与环境交互获得新的状态s_{t+1}.

11: if 目标点g在中上层空间中 do:

12: 环境反馈的奖励$r_t = \begin{cases} 0 & \text{,满足约束} \\ \tau r_1 + (1-\tau)(r_2 + r_3) & \text{,不满足约束} \end{cases}$

13: if 目标点g在下层空间中 do:

14: 环境反馈的奖励$r_t = \begin{cases} 0 & \text{,满足约束} \\ \tau r_1 + (1-\tau) r_2 & \text{,不满足约束} \end{cases}$

15: end for

16: for t=0,T-1 do:

17: 根据当前的目标点g,将当前状态s_t,动作a_t,奖励r_t和下一状态s_{t+1}存入经验池R,i.e.$(s_t\|g, a_t, r_t, s_{t+1}\|g)$.

18: 随机选取k个末端执行器的状态作为新的目标点g',将新的transition:$(s_t\|g', a_t, r_t, s_{t+1}\|g')$存入经验池$R$.

19: end for

20: 根据公式(13)采样n个transition为样本作为输入,放入Actor-Critic网络中进行训练,并根据公式(10),公式(11)更新参数。

21:end for

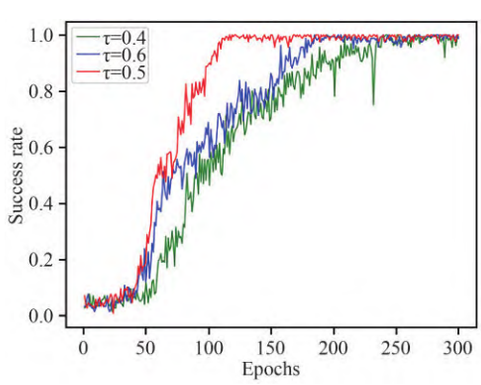


图15 τ取不同值时机械臂抓取的成功率

Fig. 15 The success rate of grasping by the mechanical arm when τ is different

4 实验验证与分析

本文采用PyBullet仿真平台对openMANIPULATOR-X机械臂进行建模,并仿照OpenAI Gym中的fetch机械臂的reach任务编写了相

应的强化学习接口，为算法训练提供环境交互基础．在reach任务中，每次初始化机械臂时，会自动随机生成一个目标点，当机械臂的末端与目标点的距离小于1 cm时，即视为完成任务．通过调用OpenAI的Gym接口对算法（详见算法1）进行训练．具体的初始网络参数见表2.

表2　部分参数的初始值

Tab. 2　Initial values of some parameters

参数	备注	初始值
epochs	训练轮次	300
cycles	每轮中最大步数	50
α_{Actor}	Actor网络学习率	0.001
α_{Critic}	Critic网络学习率	0.001
buffer_size	经验池容量	1×10^6
σ	权重软更新参数	0.05
batch_size	每次训练的样本数	256
γ	学习率	0.98

采用本文所采用的仿人自适应DDPG算法，通过300轮的训练，机械臂完成任务的平均成功率为77.65%，第100轮之后的平均成功率为98.85%，由此可见算法在前100轮的训练中基本达到了收敛，训练效率较高，证明算法能够完成任务．具体的训练结果如图14所示．

因为本文所提出的仿人自适应DDPG算法时针对密集植株情况下的机械臂路径规划问题，需要关注机械臂在完成任务过程中是否具有人手臂的运动特性．因此，本文设计了机械臂仿人性能的验证实验．在上层空间、中层空间和下层空间中分别选取20个随机目标点，对每层的20个目标点进行抓取测试．得到机械臂抓取过程中，各层空间中的各关节角度的大小范围，如表3所示．这一结果和表1中的人类手臂关节角度运动范围完全符合，从而证明了本文所提出的算法具有一定的仿人性．

表3　机械臂抓取过程中各层关节角度范围

Tab. 3　The range of joint angles of each layer during the grasping process of the robotic arm

目标点位置	肩关节角度范围	肘关节角度范围
上层空间	[2.46°,57.23°]	[21.53°,138.55°]
中层空间	[9.98°,89.77°]	[41.79°,137.97°]
下层空间	[-89.51°,-45.31°]	[24.37°,103.85°]

为了进一步证明仿人性，对随机100个中上层目标点进行抓取实验，记录每一次实验中的肩夹角和肘夹角大小，并对其进行分析．实验结果如图17所示，结果表明：在上层空间和中层空间中，机械臂的肩关节角度均小于肘关节角度，符合3.3节中肩夹角和肘夹角大小关系的约束，进一步证明了机械臂具有一定的仿人性．

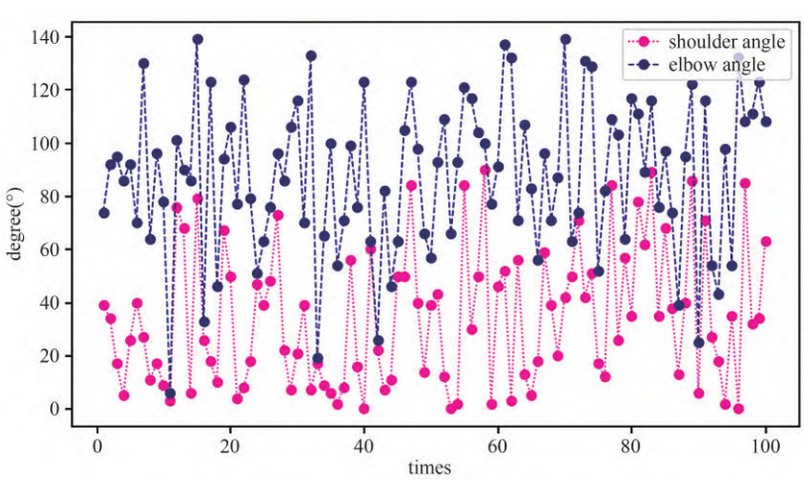


图17　中上层空间中肩夹角和肘夹角大小关系的验证

Fig. 17　The Verification of the relationship between shoulder angle and elbow angle in upper middle space

5　结论

本文以现代化大棚中的采摘场景为切入点，分析了大棚高架栽培技术下的果蔬采摘特点，针对机械臂的采摘路径容易收到两边植株的影响，不能进行大幅度的横向移动的情况，结合人手臂纵向抓取高效的运动特性提出了一种仿人自适应DDPG算法．本文通过对人右臂的运动特性分析，提出了三种手臂运动约束，并针对这三种约束为仿人自适应DDPG算法设计了不同的奖励函数引导机械臂训练，从而达到仿人手臂运动的特性．该算法采用自适应抽样策略来自适应调整经验缓冲池中轨迹经验的优先级从而提高算法的训练速度．经过实验证明，本文算法能够达到95.6%的抓取成功率，且抓取

过程中的各关节角度也满足人手臂运动时的关节角度范围,具有一定的仿人性.后续工作可以通过迁移学习等方法将该算法移植到真实的机械臂上,实现真实的仿人采摘机械臂.

参 考 文 献

[1] 桑婷,赵云霞,杨冬艳,等.栽培密度对不同品种草莓设施高架栽培生长与产量的影响[J].农业工程技术,2022,42(4):69-71.

[2] 何芬,侯永,尹义蕾,等.不同栽培模式下温室草莓根区加温环境测试与分析[J]北方园艺,2022,(1):59-64.

[3] LIU Y, ZUO G. Improved RRT path planning algorithm for humanoid robotic arm[C]//IEEE. 2020 Chinese Control And Decision Conference (CCDC). Hefei: IEEE, 2020: 397-402.

[4] PAUS F, KAISER P, VAHRENKAMP N, et al. A combined approach for robot placement and coverage path planning for mobile manipulation[C]//IEEE. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver: IEEE, 2017: 6285-6292.

[5] YANG C, ZENG C, FANG C, et al. A DMPs-based framework for robot learning and generalization of humanlike variable impedance skills[J]. IEEE/ASME Transactions on Mechatronics, 2018, 23(3): 1193-1203.

[6] ROSELL J, SUAREZ R, GARCIA N, et al. Planning grasping motions for humanoid robots[J]. International Journal of Humanoid Robotics, 2019, 16(6): 1950041.

[7] GONG S Q, ZHAO J, ZHANG Z Q, et al. Task motion planning for anthropomorphic arms based on human arm movement primitives[J]. Industrial Robot-an International Journal, 2020, 47(5): 669-681.

[8] GARCIA N, ROSELLl J, SUAREZ R. Motion planning by demonstration with human-likeness evaluation for dual-arm robots[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017, 49(11): 2298-2307.

[9] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. arXiv Preprint arXiv:1509.02971, 2015.

[10] HOU Y, LIU L, WEI Q, et al. A novel DDPG method with prioritized experience replay[C]//IEEE. 2017 IEEE international conference on systems, man, and cybernetics (SMC). Banff :IEEE, 2017: 316-321.

[11] QIU C, HU Y, CHEN Y, et al. Deep deterministic policy gradient (DDPG)-based energy harvesting wireless communications[J]. IEEE Internet of Things Journal, 2019, 6(5): 8577-8588.

[12] XU Y H, YANG C C, HUA M, et al. Deep deterministic policy gradient (DDPG)-based resource allocation scheme for NOMA vehicular communications[J]. IEEE Access, 2020, 8: 18797-18807.

[13] MATHERON G, PERRIN N, SIGAUD O. The problem with DDPG: Understanding failures in deterministic environments with sparse rewards[J]. arXiv Preprint arXiv:1911.11679, 2019.

[14] WEN S, CHEN J, WANG S, et al. Path planning of humanoid arm based on deep deterministic policy gradient[C]//IEEE. 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO). Kuala Lumpur: IEEE, 2018: 1755-1760.

[15] KUO P H, HU J, LIN S T, et al. Fuzzy deep deterministic dolicy gradient-based motion controller for humanoid robot[J]. International Journal of Fuzzy Systems, 2022: 1-17.

[16] LI J, YU T, ZHANG X, et al. Efficient experience replay based deep deterministic policy gradient for AGC dispatch in integrated energy system[J]. Applied Energy, 2021, 285: 116386.

[17] ZHANG Z, CHEN J, CHEN Z, et al. Asynchronous episodic deep deterministic policy gradient: Toward continuous control in computationally complex environments[J]. IEEE Transactions on Cybernetics, 2019, 51(2): 604-613.

[18] ROBOTIS. Open MANIPULATOR-X, 2020. [EB/OL]. [2020-07-02] https://emanual. robotis. com/docs/en/platform/openmanipulator_x/overview/.

[19] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[20] ANDRYCHOWICZ M, WOLSKI F, RAY A, et al. Hindsight experience replay[C]//MIT Press. Proceedings of the 31st International Conference on Neural Information Processing Systems. Los Angeles: MIT Press, 2017: 5055-5065.

(责编&校对 雷建云)