控制与决策
Control and Decision

鱼群涌现机制下集群机器人运动强化的迁移控制

刘磊, 张浩翔, 陈若妍, 高岩, 王富正, 王亚刚

引用本文:
刘磊,张浩翔,陈若妍,高岩,王富正,王亚刚. 鱼群涌现机制下集群机器人运动强化的迁移控制[J]. 控制与决策, 2023, 38(3): 621−630.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2021.1554

鱼群涌现机制下集群机器人运动强化的迁移控制

刘 磊[1,2†], 张浩翔[2], 陈若妍[1], 高 岩[1], 王富正[1], 王亚刚[2]

(1. 上海理工大学 管理学院，上海 200093；2. 上海理工大学 光电学院，上海 200093)

摘 要: 采用鱼群模型驱动多智能体可以涌现出优良的运动特性,但是,由于机器人与真实鱼类相比具有较大的差异性,使得鱼群模型难以应用于真实机器人系统. 为此,提出一种结合深度学习与强化学习的迁移控制方法,首先,使用鱼群运动数据训练深度网络(deep neural network, DNN)模型,以此作为机器人成对交互的基础;然后,连接强化学习的深度确定性策略梯度方法(deep deterministic policy gradient, DDPG)来修正DNN模型的输出,设计集群最大视觉尺寸方法挑选关键邻居,从而将DNN+DDPG模型拓展到多智能体的运动控制. 集群机器人运动实验表明:所提出方法能使机器人仅利用单个邻居信息就能形成可靠、稳定的集群运动,与单纯DNN直接迁移控制相比,所提出DNN+DDPG控制框架既可以保存原有鱼群运动的灵活性,又能增强机器人系统的安全性与可控性,使得该方法在集群机器人运动控制领域具有较大的应用潜力.

关键词: 集群机器人；鱼群交互模型；迁移控制；强化学习；生物涌现；智能控制

中图分类号: TP242.6 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2021.1554

引用格式: 刘磊,张浩翔,陈若妍,等. 鱼群涌现机制下集群机器人运动强化的迁移控制[J]. 控制与决策, 2023, 38(3): 621-630.

Transfer control of swarm robotics motion reinforcement employing fish schooling emergency mechanism

LIU Lei[1,2†], *ZHANG Hao-xiang*[2], *CHEN Ruo-yan*[1], *GAO Yan*[1], *WANG Fu-zheng*[1], *WANG Ya-gang*[2]

(1. School of Management，University of Shanghai for Science and Technology，Shanghai 200093，China；2. School of Optical-electrical，University of Shanghai for Science and Technology，Shanghai 200093，China)

Abstract: The multi-agent system driven by the fish schooling model can emerge excellent characteristics of motion. However, due to the individual differences between robots and real fish, it is difficult for a fish schooling model to be directly applied to the actual robotics system. Hence, a transfer control method combined with deep learning and deep reinforcement learning is proposed. Firstly, a deep neural network (DNN) model is trained by the data of fish schooling, which is the basement for the interactive control of robots. Then, a deep reinforcement learning method, named deep deterministic policy gradient (DDPG), is connected to the output of the DNN model. Finally, based on the above DNN+DDPG model, a key neighbor selection method of the maximum group visual size is designed to expand the DNN+DDPG model to multi-agent motion control. Collective motion experiments show that the proposed method can formulate reliable and stable collective motion of the robots via individual information. Compared with the pure DNN transfer control, the proposed DNN+DDPG control frame not only preserves the flexibility of the collective motion of fish schooling, but also enhances the safety and controllability of the robotics system. Thus, there exists strong potential application of the proposed method for the swarm robotics motion control.

Keywords: swarm robotics；interaction model of fish schooling；transfer control；reinforcement learning；biological emergence；intelligent control

0 引 言

多机器人协同作业是解决复杂工程问题的有效手段,也是提升任务效率的有效途径. 其中采用生物集群自组织的方法进行集群机器人控制具有集群规模可缩放、集群运动强鲁棒等特点[1],受到了学术界的广泛关注. 研究人员期望通过精心设计机器人的社会交互模型来涌现出所需的集群行为[2]. 例如Ning等[3]利用“锐角规则”形成各向异性交互网络,

收稿日期: 2021-09-05；**录用日期:** 2021-12-30.

基金项目: 国家自然科学基金项目(72071130)；上海市自然科学基金项目(22ZR1443300).

†通讯作者. E-mail: liulei@usst.edu.cn.

实现了集群运动的涌现. 但是,上述研究涉及的交互模型普遍具有较强的主观性,虽然被控多智能体具有生物集群的灵活性,却难以有目的地优化集群行为涌现.

随着人工智能方法的不断进步,智能控制算法逐渐成为解析生物集群运动交互模型的关键[4]. 早期的工作包括利用神经网络控制E-puck机器人汇合在安全区域,重现蟑螂的汇聚过程[5]. 进化算法也被用来对鸟群运动模型进行优化,最终实现了30架无人机的自组织飞行[6]. 这些工作表明,智能控制为集群机器人的协同行为涌现开辟了新的研究方向,但是,经典的人工智能方法难以分析较复杂的生物集群运动,例如红鼻剪刀鱼单个集群可达几百条,同时具有极强的组织纪律性[7],抽取该类鱼群的社会交互模型具有相当的复杂性. 现有研究表明,鱼类社会交互主要基于视觉感官[8],一些工作已成功地将鱼群的视觉交互机制应用到集群机器人的运动控制中[9-10],取得了较好的运动控制效果.

红鼻剪刀鱼群游数据具有较好的质量,适于深度学习数据驱动建模. 另外,该类鱼群的运动轨迹可以分解为一组线段,可以较好地弥补轮式机器人的非完整性约束限制,因此,可以利用剪刀鱼群的运动数据对集群轮式机器人进行仿生控制. 为了获得更精细的红鼻剪刀鱼群社会交互模型,刘磊等[11]利用深度神经网络(DNN)技术,抽取两鱼运动数据特征,构建了单体的社会交互模型,并仿真验证了大规模的集群运动涌现. 但是,轮式机器人相较于鱼类具有一定的性能落差,具体表现为动力学约束、刚体轮廓约束等,所以采用真实鱼群数据训练的DNN模型很难直接迁移到集群机器人中,易造成单体与边界或邻居的碰撞.

模型迁移的目的是从一个或多个源领域任务中提取有用知识,并将其用于新的目标任务中[12]. 目前,结合强化学习对原有DNN模型进行迁移的研究逐渐成为热点[13],因此,本文借助强化学习中经典的深度确定性策略梯度算法(DDPG)[14]对鱼群运动数据训练的DNN交互模型进行辅助修正,以提升机器人的实际任务性能. 正如集群机器人领域的权威专家Dorigo等[15]最近指出:不应仅在集群机器人上使用生物启发控制模型,还需兼顾整个系统的任务性能. 所以借助DDPG方法修正生物数据训练的DNN控制模型,是一种生物模型工程优化的有益探索.

本文贡献在于:

1) 利用强化学习DDPG网络修正鱼群数据训练的DNN模型(生物模型),以下简称DNN+DDPG模型,从而使原有生物模型可以迁移控制真实机器人,使机器人集群能够享有生物集群控制的灵活性.

2) 设计了改进最大视觉角度的关键邻居选择方法,通过该方法可以使DNN+DDPG模型拓展到集群机器人的集群控制. 该机制充分利用了生物模型的涌现性,为大规模集群运动的强化控制提供了新思路.

3) 将DNN与DDPG相互嵌合,通过简单交互即可控制集群机器人安全沿墙运动,为集群机器人的实际应用奠定基础.

1 集群运动智能控制方法

集群机器人属于典型的复杂系统,刘磊等[11]利用生物集群运动数据训练DNN交互模型,通过软件仿真验证了大规模集群运动,如图1虚线部分所示. 但是,如果直接将DNN模型生成的生物控制信号驱动真实机器人,则会经常性地出现群体阻塞与环境碰撞,无法形成稳定的集群运动. 其中机器人实验采用课题组具有自主知识产权的微型机器人系统[16],实验环境及单体机器人如图2所示.

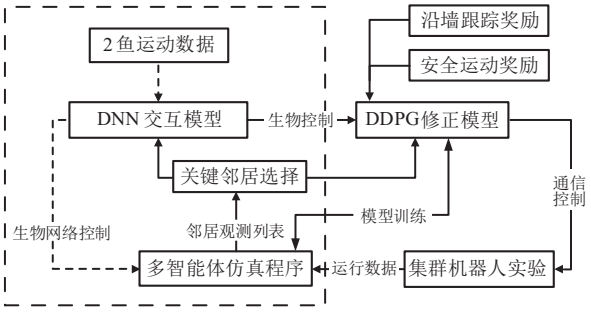


图1 集群运动系统智能控制框架

机器人能够沿轨道灵活安全运动是集群机器人实现大规模应用的基础,虽然先期工作已证明鱼群运动数据训练的DNN控制能够实现聚集运动[11],但是,对于真实机器人群体的安全控制以及边界调控仍缺乏必要的迁移优化过程. 为此,引入深度强化学习模型,对原有DNN网络输出进行修正,以提升运动安全与墙壁跟踪性能,具体控制框架如图1所示.

自主研发的微型机器人尺寸为40 mm×40 mm×60 mm,鱼群实验采用的剪刀鱼平均身长30 mm,鱼群与机器人群的实验环境均采用圆形围墙约束,机器人群围墙直径1 m,鱼群围墙直径0.5 m. 本课题组采集了30 h的2鱼运动实验数据[7]用于训练DNN交互模型,11 h的5鱼运动实验数据[10]用于评估集群机器人的运动性能,图3(a)显示了5条鱼的集群运动实验.

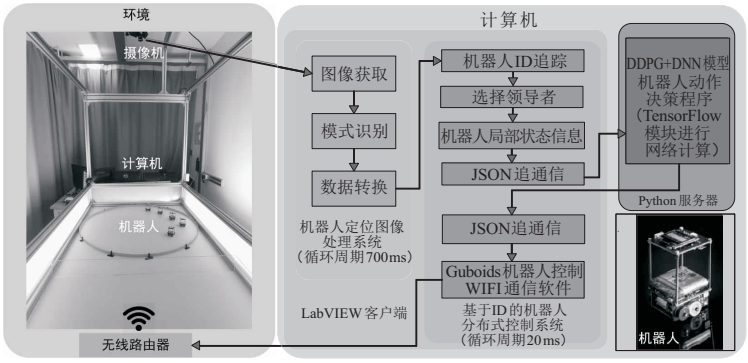


图 2　自制集群机器人实验平台

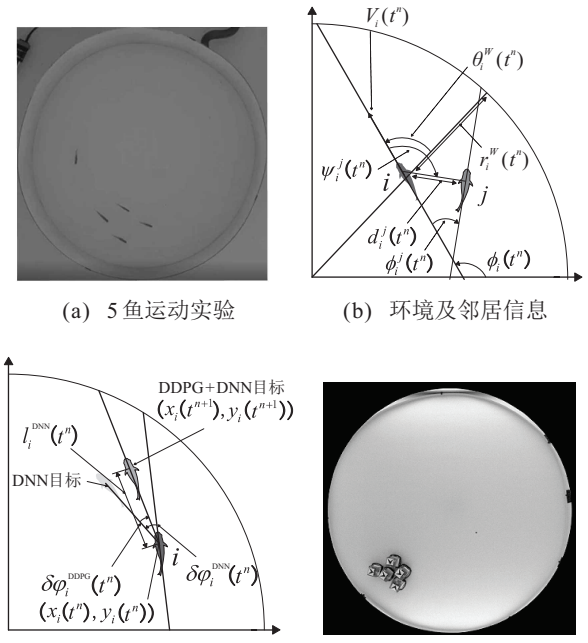


图 3　鱼类群游实验环境及其运动控制变量

1.1　智能体成对交互控制模型

实验鱼群运动可以抽象为原地转向与直线滑行两阶段过程,利用2鱼运动数据分别构建“转向决策”与“直行决策”DNN交互模型,如图4所示[11]. DNN网络的输入变量包括两部分: 1)环境交互变量: 单体 i 相对于墙的距离 r_i^W,角度 θ_i^W; 2)社会交互变量: 邻居 j 相对于单体 i 的距离 d_i^j,视角 ψ_i^j 以及航向角差 $\phi_i^j = \phi_i - \phi_j$,如图3(b)所示. “转向决策”网络输出为 $\delta\phi_i^{\mathrm{DNN}}(t^n)$,调整角度 ϕ_i 后,会将调整后的变量输入到“直行决策”网络,产生运动步长 $l_i^{\mathrm{DNN}}(t^n)$,如图4所示. 图3(c)显示在 t^n 时刻焦点鱼(深色)的朝向改变为 $\theta_i(t^n)+\delta\phi_i^{\mathrm{DNN}}(t^n)$,然后直线滑行 $l_i^{\mathrm{DNN}}(t^n)$ 距离(浅灰色),接着再次触发 t^{n+1} 时刻的DNN决策,形成仿真运动.

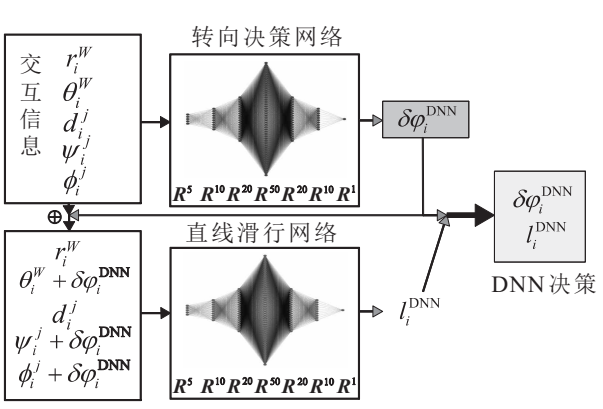


图 4　深度神经网络交互模型结构

1.2　基于视觉机制的多智能体控制

智能体的DNN交互模型仅给出了两智能体之间的相互作用,如果要推广到多智能体控制,则需要焦点单体从集群中选择合适的邻居进行成对交互,然后再融合所有交互结果以供焦点单体的运动决策[17]. 文献[10]发现: 鱼类单体仅利用一到两个邻居的信息就能形成集群运动,该结论表明了集群内部信息传播、处理的精简性. 因此,文献[11]中利用视觉感知来挑选单一关键邻居,即最大视觉角度(压力)法,实现了大规模的集群运动仿真,其中视觉角度定义为从焦点鱼观测邻居身体所占据的视觉角度. 但是,由于近处邻居的视觉角度过大,会出现遮挡远处鱼群的现象,从而导致焦点单体出现“一叶障目”的问题,与近处邻居交互会使单体脱离集群,不利于集群运动的涌现.

为解决上述问题,本文提出一种基于最大集群视觉角度的改进邻居选择方法,如图5所示. γ 为邻居 j 的视觉角度,当多个邻居出现重叠时,需要先确定最大邻居群,然后找出该群的最大视角邻居. 图5显示了4个邻居 j、k、l、m. 其中: 邻居 j、k 有重叠,形成了邻居群1; l、m 有重叠,形成了邻居群2. 两群的视觉角度分别为 α 和 β. 因为 $\alpha > \beta$,所以先选择邻居群1,再选择该群中具有较大视觉角度的邻居 j 作为关键邻居. 为防止碰撞,在焦点单体 i 的运动方向上

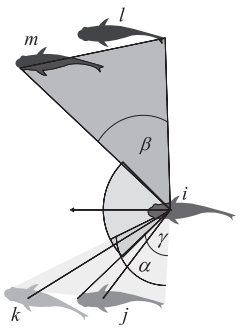


图5 关键邻居选择方法

建立一个角度范围为π/2、半径为2倍身长的扇形警戒区域,如图5所示,当邻居位于该区域时,选择距离最近的邻居作为关键邻居.该设置体现了鱼群运动成对交互的各向异性[7],即后方邻居的影响弱于前方邻居. 警戒区域的邻居距离小于1倍身长时,焦点单体保留转向运动,停止直线运动进入等待状态;当超过5 s仍然没有解除等待状态时,就重新进行决策,直到警戒区域内没有距离过小的邻居后恢复直行运动控制.

新改进的方法具有更强的集群聚合能力,与文献[11]相比,可以实现更大规模、更高速地集群聚合,如图6所示.

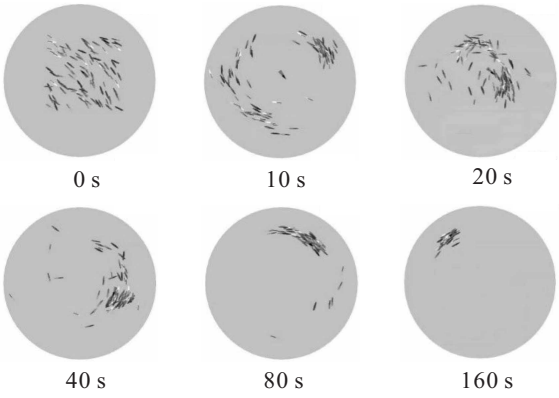


图6 基于最大集群视觉角的100智能体DNN交互控制

直接将上述集群视觉机制应用于DNN模型来迁移控制集群机器人,会因为机器人与鱼类差异而造成群体阻塞,如图3(d)所示,为此,建立基于强化学习的DDPG网络来修正生物启发DNN模型的转向决策. 仅利用DDPG修正DNN模型的转向是因为运动中采用方向调整方式比减速控制更具避障时效性[18],有利于单体适应群体复杂的运动态势. 如图3(c)所示,单体在t^n时刻的航向角改变分为两部分:一是DNN模型输出的$\delta\phi_i^{\mathrm{DNN}}(t^n)$(到达浅灰色目标);另一个是DDPG网络输出的$\delta\phi_i^{\mathrm{DDPG}}(t^n)$修正(到达灰色目标).

1.3 强化网络修正模型

深度确定性策略梯度(DDPG)方法可以应对连续动作空间的策略强化,是一类典型的强化学习控制方法. 近期研究表明,直接应用深度强化学习方法难以实现多智能体的集群运动[19],这是由于单体处于动态环境,基于“经验回放”机制的强化过程难以有效还原全部状态空间. 本课题组曾试图直接利用MADDPG[20]与QMIX[21]等多智能体深度强化学习方法控制图2所示的集群机器人平台,但是算法输出无法收敛,群体阻塞严重,表明通过设定宏观指标来强化训练微观交互模型,该类问题解算难以收敛. 为此更换思路,将强化学习串联在DNN模型后端,再利用视觉机制迁移控制集群机器人,应是解决复杂集群运动优化控制的可行路径.

根据图1框架所示,需要设计沿墙跟踪与安全运动的集群控制目标,由于这两项任务所关注的信息不同,前者主要关注环境信息,而后者着眼于社交信息,难以将两种任务奖励设计在一起. 本课题组曾使用加权平均的方法设计奖励函数,以期利用单一DDPG网络确定航向决策修正,但是,由于环境信息属于静态变量,而社交信息却属于动态变量,DDPG模型无法将上述静态、动态任务奖励进行解耦. 为此,建立两个DDPG网络,如图7所示,分别为沿墙跟踪强化网络(DDPG-WALL网络,简称DW网络)以及邻居安全交互强化网络(DDPG-Neighbor网络,简称DN网络).

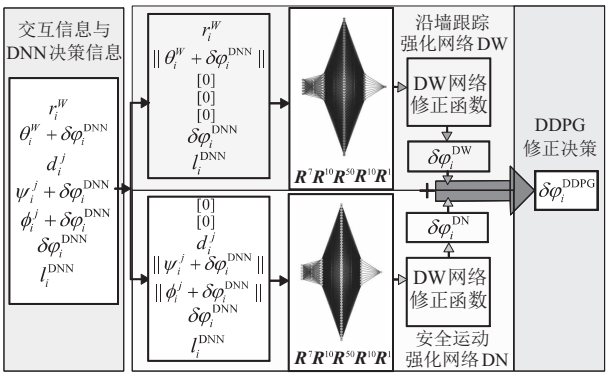


图7 DDPG网络模型

为简化训练过程,DW与DN内部各包含2对Actor、Critic网络,这些网络选用相同的全连接结构. 输入层至输出层的神经元个数依次为[7, 10, 50, 10, 1]. 两套网络的输入使用相同结构:$[r_i^W, \theta_i^W + \delta\phi_i^{\mathrm{DNN}}, d_i^j, \psi_i^j + \delta\phi_i^{\mathrm{DNN}}, \phi_i^j + \delta\phi_i^{\mathrm{DNN}}, \delta\phi_i^{\mathrm{DNN}}, l_i^{\mathrm{DNN}}]^{\mathrm{T}}$,DW与DN网络的输出分别为$\delta\phi_i^{\mathrm{DW}}$、$\delta\phi_i^{\mathrm{DN}}$,相加得到DDPG网络输出,即

$$\delta\phi_i^{\mathrm{DDPG}} = \delta\phi_i^{\mathrm{DW}} + \delta\phi_i^{\mathrm{DN}}. \tag{1}$$

该值用于对航向决策$\delta\phi_i^{\mathrm{DNN}}$进行修正,如图3(c)所示.

1.3.1 沿墙跟踪强化网络参数设计

强化学习设计的关键在于确定马尔可夫决策过程的状态空间、动作空间以及奖励函数,其中DW网络参数设计如下.

1) 状态空间: DW网络的控制目的是使智能体运动靠近墙壁并保持稳定距离,相当于根据环境构建任务轨道. 为修正生物数据训练的DNN网络输出,采用环境信息与DNN决策信息编码作为输入,其中环境信息为焦点单体i与墙壁的距离r_i^W和决策后角度$\theta_i^W+\delta\phi_i^{\mathrm{DNN}}$. 决策信息包括DNN网络输出的决策转角$\delta\phi_i^{\mathrm{DNN}}$和直线步长$l_i^{\mathrm{DNN}}$,由于DW与DN网络具有相似的结构,便于强化学习网络的调试、复用,为此,定义DW网络的具体输入结构为$S_{\mathrm{DW}}[r_i^W,\theta_i^W+\delta\phi_i^{\mathrm{DNN}},0,0,0,\delta\phi_i^{\mathrm{DNN}},l_i^{\mathrm{DNN}}]^{\mathrm{T}}$,其中用0来补充DN网络使用的社交信息空位.

2) 动作空间: DW网络模型的输出用来修正机器人的决策转向角度,所以将动作输出定义为$a_{\mathrm{DW}}=\delta\phi_i^{\mathrm{DW}}$. 为防止网络输出过大破坏DNN网络的涌现性,将DW网络的动作空间约束在$a_{\mathrm{DW}}\in[-\pi/6,\pi/6]$之内.

3) 奖励函数: 为使焦点单体与墙壁保持稳定距离,降低奖励函数的复杂度,经过多次参数对比实验,该模型奖励函数设定如下:

$$r_{\mathrm{DW}}=\begin{cases}-5|r_i^W-r_e|,\ r_i^W<r_e;\\ -|r_i^W-r_e|,\ r_i^W>r_e.\end{cases}\tag{2}$$

其中r_e为智能体i与墙壁间的期望距离,设定为50 mm. 当智能体与墙壁的距离小于r_e时,因存在碰撞风险,所以设计较大的损失增益5;当智能体与墙壁间距大于r_e时,设计损失增益为1.

1.3.2 邻居安全交互网络参数设计

1) 状态空间: DN网络的控制目的是使焦点单体既要靠近邻居(形成集群聚合),又要保持安全距离. 因此,DN网络输入需要包含关键邻居的距离d_i^j、视角$\psi_i^j+\delta\phi_i^{\mathrm{DNN}}$以及航向角差$\phi_i^j+\delta\phi_i^{\mathrm{DNN}}$,其中关键邻居利用最大视觉角策略来选择,详情见1.2节. 最后,把DNN网络输出的转角$\delta\phi_i^{\mathrm{DNN}}$和步长$l_i^{\mathrm{DNN}}$也输入到DN网络. 具体的结构为$[0,0,d_i^j,\psi_i^j+\delta\phi_i^{\mathrm{DNN}},\phi_i^j+\delta\phi_i^{\mathrm{DNN}},\delta\phi_i^{\mathrm{DNN}},l_i^{\mathrm{DNN}}]^{\mathrm{T}}$,其中环境信息输入位置用0占位.

2) 动作空间: 该模型与墙壁强化网络的动作空间相同,动作为$a_{\mathrm{DN}}=\delta\phi_i^{\mathrm{DN}}$,动作空间被同样限制在$a_{\mathrm{DN}}\in[-\pi/6,\pi/6]$区间范围.

3) 奖励函数: 为了使焦点单体与邻居保持稳定的距离,降低奖励函数的复杂度,经过多次参数对比实验,设定该模型的奖励函数如下:

$$r_{\mathrm{DN}}=\begin{cases}-5|d_i^j-d_e|,\ d_i^j<d_e;\\ -|d_i^j-d_e|,\ d_i^j>d_e.\end{cases}\tag{3}$$

其中d_e为焦点单体i与重点邻居j的期望距离. 根据2鱼实验数据分析,单体与邻居距离的高密度分布可近似为短径35 mm、长径65 mm的椭圆形,如图8所示,即焦点单体前后方邻居的高密度位置间距为65 mm,两侧的高密度位置间距为35 mm. 因此,根据邻居视角ψ_i^j的不同,建立如下焦点单体(椭圆中心)到关键邻居(椭圆边界)的期望距离公式:

$$d_{\mathrm{oval}}=\sqrt{(1+\tan^2\psi)\frac{a^2}{1+(a^2/b^2)\tan^2\psi}},\tag{4}$$

其中a、b分别为椭圆的长、短半径. 由于集群机器人采用的是异步决策,为保证运行安全,保守设计$d_e=2d_{\mathrm{oval}}$. 采用椭圆安全区域的优点在于,轮式机器人具有非完整性约束,其在前后方向需要较大的安全空间以防止碰撞与追尾,而在侧向区域的安全距离可以适当缩短. 与鱼类运动一致,当焦点单体与关键邻居的距离小于d_e时,存在碰撞风险,所以需设计较大的损失增益5;当距离大于d_e时,设计幅度较小的损失.

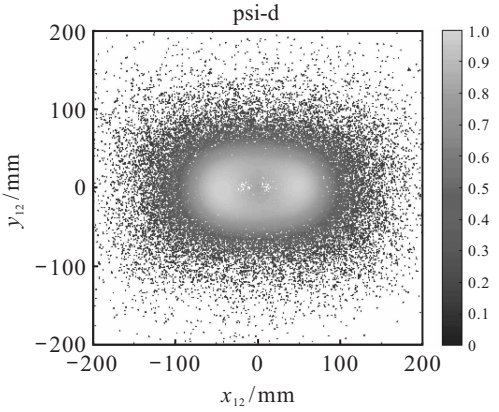


图 8　2鱼实验中焦点单体的邻居位置分布

1.3.3 DDPG网络的辅助控制方案

多次实验发现,DDPG网络难以应对强烈的临界非线性输出. 文献[7]的转向模型表明,鱼类的转向决策与相对墙壁角度θ_i^W有关,但是,在动态环境中,DW网络受DNN模型干扰,无法稳定收敛出正确的转向方向. 因此,根据文献[7]的结论,对DW输出附加如下的环境转向功能:

$$\delta\phi_i^{\mathrm{DW}}=-\mathrm{Sign}(\theta_i^W)\times|\delta\phi_i^{\mathrm{DW}}|.\tag{5}$$

即单体顺时针运动时,需逆时针转向靠近墙壁;反之,则需顺时针转向靠近墙壁. 因为图像识别具有时滞

特性,会造成机器人转角过冲,同时前方邻居遮挡直行路径也会触发新的网络决策,造成上周期直行过程被迫中断,导致运行距离偏短,这些都会客观地增加机器人的运动曲率,使得单纯DNN控制的机器人群普遍远离边界,所以设计功能辅助输出(5),从而保证DW网络可以较好地贴近边界. 而对于DN网络,由于真实鱼群交互存在各向异性[11],表现为焦点单体对前方的关注大于后方,为模拟该交互机制,需要对焦点单体后方的邻居进行屏蔽,然后才能进行关键邻居选择与DN网络输入. 另外,由于环境曲率对DN网络的修正也有影响,当焦点单体i在圆形环境中逆时针运动时($\theta_i^W > 0$),为躲避前方邻居碰撞,DN网络应给予单体i远离墙壁的转向修正,即加深逆时针转动幅度(正值$\theta_i^{\rm DW}$)以促进内圈超越,所以邻居安全交互辅助设计如下:

$$\delta\phi_i^{\rm DN} = -{\rm Sign}(\theta_i^W) \times |\delta\phi_i^{\rm DN}|. \tag{6}$$

值得指出的是:如果$\delta\phi_i^{\rm DN}$过大,则会使得焦点单体i与关键邻居j脱离;如果$\delta\phi_i^{\rm DN}$过小,则会被DNN模型的不当聚合能力吸引,造成集群碰撞阻塞. DN网络的功能是使机器人之间长时间保持安全距离,当邻居缓慢时,也可以实现弯道“超车”,并最终与新的邻居进行交互.

1.4 强化网络修正模型

1.4.1 DNN训练及实验平台

DNN交互模型训练使用2鱼运动轨迹,提取焦点单体i的局部环境信息$[r_i^W,\ \theta_i^W]^{\rm T}$、社交信息$[d_i^j,\ \psi_i^j,\ \phi_i^j]^{\rm T}$以及对应的决策$\delta\phi_i$、$l_i$. 使用均方误差作为“转向决策”网络和“直行决策”网络的损失函数,训练主机采用NVIDIA Geforce 2080Ti显卡,软件使用Python3.6配合TensorFlow-GPU-1.08,在相同训练平台上编写LabView仿真软件来强化训练DDPG网络.

采用平台顶部的工业相机对机器人的运动实验进行图像采集,使用LabVIEW的模式识别程序进行图像处理,获取各单体的位置与朝向角,再利用卡尔曼滤波器进行滤波,从而获取每台机器人的全局信息(坐标位置(x, y)与朝向角度ϕ_i);然后,将全局信息转换为焦点单体的局部观测信息,例如对于机器人i,其测量的环境信息为$[r_i^W,\ \theta_i^W]^{\rm T}$,邻居$j$的社交信息为$[d_i^j,\ \psi_i^j,\ \phi_i^j]^{\rm T}$;最后,利用关键邻居选择算法锁定关键邻居,将关键邻居的测量信息通过自制应用层通信协议打包成JSON格式,利用TCP协议传输到Python程序进行决策输出,最终输出转向角度$\delta\phi_i^{\rm DNN}$或$\delta\phi_i^{\rm DNN} + \delta\phi_i^{\rm DDPG}$、直线距离$l_i^{\rm DNN}$控制信号. 将控制信号转换成JSON数据包后,利用TCP协议回传LabView仿真平台,仿真平台通过挂接WIFI路由器,直接驱动机器人差动运行,如图2所示. 单体机器人i的运动分为两个阶段: 1) 差动旋转$\delta\phi_i^{\rm DNN}$(单纯DNN迁移控制)或$\delta\phi_i^{\rm DNN} + \delta\phi_i^{\rm DDPG}$(DNN+DDPG模型控制)的角度; 2) 直线运动$l_i^{\rm DNN}$长度. 由于采用顶部图像伺服的形式控制机器人,机器人的转向与直线运动存在大量噪声,受篇幅所限,机器人平台详细信息可参考文献[10].

1.4.2 DDPG网络参数及训练

DDPG网络是用来寻找连续系统次优策略的算法,包括4个神经网络: Actor与Actor target网络,Critic与Critic target网络,可将这4个神经网络的参数记作θ^A、$\theta^{A'}$、θ^Q、$\theta^{Q'}$. 其中Critic网络用于估计状态S下采取动作a的价值$Q(s, a)$,一般采用Bellman方程[22]构建损失函数$\mathcal{L}(\theta^Q)$,有

$$B_E = r(s_t, a_t) + \gamma Q'(s_{t+1}, A'(s_{t+1}|\theta^{A'})|\theta^{Q'}), \tag{7}$$

$$\mathcal{L}(\theta^Q) = \boldsymbol{E}_{S\sim\beta}[(Q(s_t, a_t|\theta^Q) - B_E)^2]. \tag{8}$$

其中: B_E为Bellman方程,β为正态分布随机策略.

Actor网络接收状态s映射输出动作a,其网络参数θ^A利用策略梯度进行更新,在策略A下,其性能指标$J(A)$对网络参数θ^A的策略梯度为

$$\nabla_{\theta^A}(J(A)) = \boldsymbol{E}_{S\sim\beta}[\nabla_{\theta^A}A(s|\theta^A)|_{s=s_t}\nabla_A Q(s, a|\theta^Q)|_{s=s_t, a=A(s_t|\theta^A)}]. \tag{9}$$

设立target网络是为了使模型训练数据独立同分布,能够保证算法稳定收敛,target网络与原始网络之间使用如下软更新公式:

$$\theta^{A'} = \tau\theta^A + (1-\tau)\theta^{A'}, \tag{10}$$

$$\theta^{Q'} = \tau\theta^Q + (1-\tau)\theta^{Q'}, \tag{11}$$

其中τ为软更新率. DDPG训练方法具体可参见文献[14]. DW网络与DN网络训练使用早停法以防止过拟合,训练参数如表1所示.

表1 强化网络训练参数

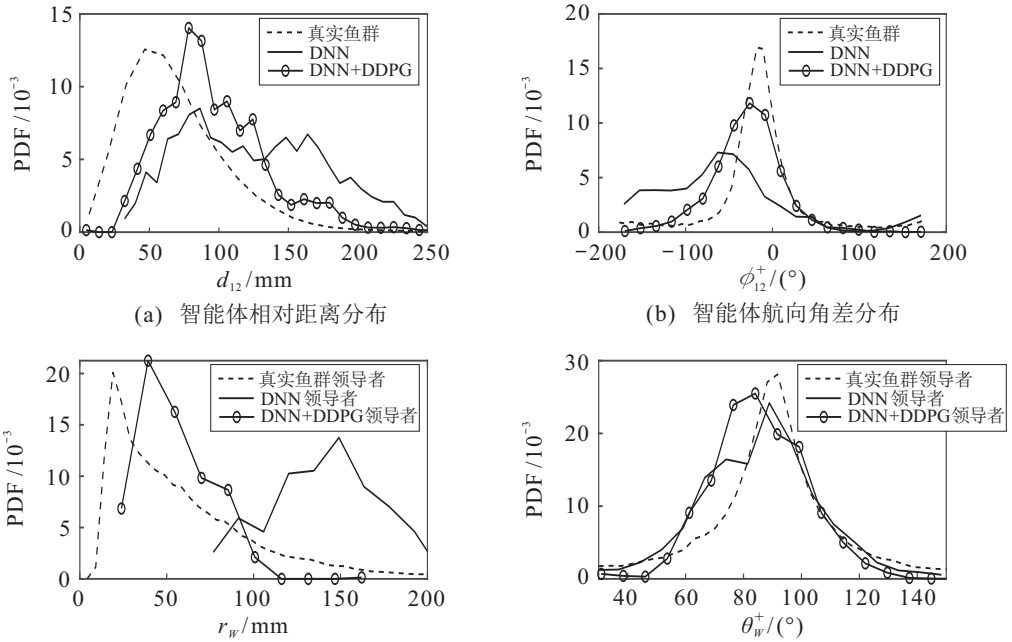
参数	数值	参数	数值
θ^A学习率	e-4	批采样个数	32
θ^Q学习率	e-6	初始探索率β	0.3
贪婪率γ	0.99	单轮迭代数	256
软更新率τ	5e-3	迭代轮数	16
记忆池容量	128		

2 机器人实验分析

2.1 2机器人运动修正效果对比

放置2台机器人,分别使用DNN生物模型迁移控制和DNN+DDPG模型控制45 min,对环境与社交

数据进行记录,概率分布如图9所示. 由于机器人与真实鱼类不同,通过将环境比例放大的方式直接使用DNN模型仿真的数据与真实鱼群运动的数据分布有显著不同. 鉴于鱼群实验半径是机器人平台半径的一半,所以将机器人运动的长度结果均乘以0.5的缩放系数以方便数据比较.



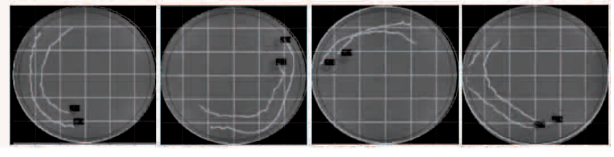
(a) 智能体相对距离分布　(b) 智能体航向角差分布

(c) 智能体相对墙壁距离分布　(d) 智能体相对墙壁角度分布

图 9　2机器人实验的环境与社交数据分布对比

图9(a)显示:两条鱼之间的距离分布具有2倍身长(剪刀鱼平均身长30 mm)的峰值,并且多数情况下两条鱼的距离较小,呈现出吸引联动的特性;而直接采用DNN模型迁移控制的2机器人相互距离分布峰值不明显,呈现出较离散的距离分布,跟随效果不紧密,这是由于机器人运动延迟误差造成的集群控制发散;经过DDPG网络修正以后,明显出现了聚集峰值,加强了机器人间的跟随特性. 图9(b)展示了两智能体运动的对齐特性,$\phi_{12}^{+} = \phi_{12} \times \mathrm{Sign}(\theta_1^W)$为智能体1测量智能体2的航向角差,真实鱼群曲线表明鱼类之间对齐行为明显(在0处具有峰值);而单纯DNN模型迁移控制的数据分布显示两机器人之间的航向角差值较大,即两机器人由于运动误差导致控制朝向基本不一致;DNN+DDPG曲线则表明通过DDPG强化可以提升机器人的对齐特性,有利于机器人之间的运动协同. 图9(c)中:真实鱼群曲线显示,2鱼运动中的领先者($|\psi_{12}| > \pi/2$)大概率地靠近墙壁,但曲线具有较长的拖尾特性,体现了鱼群领导者运动的随机性,容易将鱼群引入无序的运动状态;而DNN曲线说明单纯使用DNN模型迁移控制的领先机器人远离墙壁,可见鱼群运动的随机性以及机器人与真实鱼类之间的差异性导致机器人无法较好地跟踪墙壁轨道;经过DDPG网络修正,DNN+DDPG曲线显示了领头机器人在墙壁附近出现了分布峰值,并且分布的拖尾特性也得到明显改善,机器人运动呈现出较高的纪律性,有利于集群机器人的实际应用. 图9(d)显示:领先智能体相对围墙的角度绝对值基本保持在90°(平行于围墙运动);而单纯DNN模型的数据分布具有不规则波动,并具有更宽的分布范围;DNN+DDPG混合模型则较好地修正了领导者的对墙角度.

2机器人DNN+DDPG模型控制的轨迹如图10所示,在逆时针运动的过程中,出现了领导者(领先机器人)-跟随者(滞后机器人)交替变化的现象. 该现象与2鱼实验类似[7],表明DDPG强化控制并没有破坏原有生物模型(DNN模型)的交互灵活性,保留了鱼群运动频繁更换领导者的特性;同时也表明智能控制具有较强的鲁棒性,当进行多智能体仿真时,某些领导者由于故障被跟随者超越后,跟随者就会切换新的领导者进行交互. 图10同时还显示DDPG网络修正可以使机器人有序靠近墙壁,相比于单纯生物DNN模型迁移控制具有更强的运动纪律性.



(a) 0 s　(b) 1 min 12 s　(c) 2 min 46 s　(d) 4 min 0 s

图 10　DNN+DDPG控制的2机器人实验

2.2 5机器人运动修正效果对比

利用1.2节提出的关键邻居选择法,为焦点单体i挑选关键邻居进行交互,以拓展实现集群机器人的运动控制. 由于本课题组采集了5条鱼运动的实验数据[10],这里采用5机器人运动实验进行控制效果分析. 首先定义4个宏观运动指标来描述集群运动特征. 考虑集群重心B,其位置坐标定义为$p_B(t)=(x_B(t),y_B(t))$. 其中:$x_B(t)=\dfrac{1}{N}\sum\limits_{i=1}^{N}x_i(t)$,$y_B(t)=\dfrac{1}{N}\sum\limits_{i=1}^{N}y_i(t)$,$N$为集群中个体数目. B的速度为$\boldsymbol{v}_B(t)=(v_B^x(t),v_B^y(t))$,是由$B$的位置向后差分得到,集群重心航向角为$\phi_B(t)=\arctan(v_B^x(t),v_B^y(t))$,则集群运动的4个宏观指标可定义如下:

1) 集群重心B的位置相对于墙壁的距离$r_W\in[0,R_W]$,有

$$r_W(t)=R_W-\sqrt{x_B^2(t)+y_B^2(t)}, \tag{12}$$

其中R_W为墙壁半径.

2) 集群重心B的位置相对于墙壁角度$\theta_W^+\in[0,\pi]$,有

$$\theta_W^+(t)=|\phi_B(t)-\arctan(y_B(t),x_B(t))|. \tag{13}$$

3) 集群极性$P(t)\in[0,1]$,有

$$P(t)=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{e}_i(t), \tag{14}$$

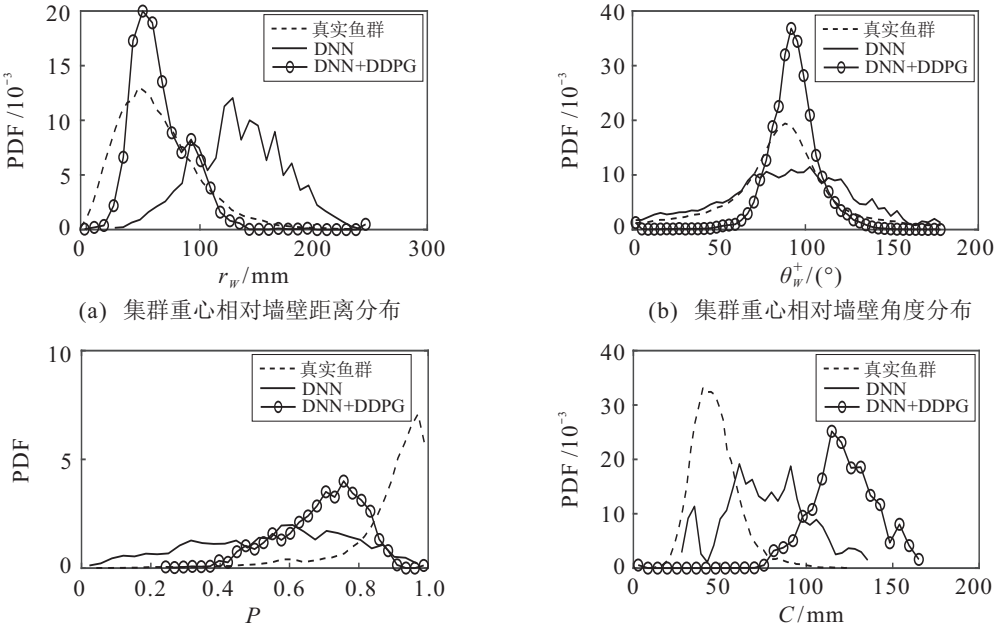
其中$\boldsymbol{e}_i=(\cos\phi_i,\sin\phi_i)$表示单体$i$航向的单位向量. 当$P$值接近1时,表示几乎所有单体的航向角相同;当$P$值接近0时,所有单体的航向角发散.

4) 集群大小$C(t)$,有

$$C(t)=\frac{1}{N}\sum_{i=1}^{N}\|p_i(t)-p_B(t)\|, \tag{15}$$

其中$\|p_i(t)-p_B(t)\|$为单体i与集群重心B的距离. $C(t)$的值越小,集群越紧密.

图11(a)为集群重心B与墙壁距离r_W,真实鱼群重心靠近边界(如虚线所示);而单纯使用DNN模型迁移控制机器人群的重心却远离墙壁轨道(如实线所示);通过DDPG修正可以使机器人重心靠近墙壁,并且具有更狭窄的分布范围(带标记○的实线),实现了墙壁的轨道跟踪. 图11(b)为集群重心B相对于墙壁的正向角度θ_W^+,可见,几乎所有的控制策略和真实鱼群的$\theta_W^+\approx 90°$,即集群重心表现为与墙壁平行运动,但是,鱼群运动显示有较大的随机性,纯DNN迁移控制的沿墙运动随机性更加明显,DNN+DDPG模型控制则产生了最陡峭的角度分布,体现了集群运动的确定性. 图11(c)为极性P分布,相较于纯DNN迁移控制机器人,DDPG修正控制的极性更高,在$P\approx 0.8$处产生了峰值;真实鱼群具有更高的运动极性,这是因为鱼类身体细长,有利于在紧凑空间内实现超高极性. 如图11(d)所示,鱼群半径$C\approx 50$ mm,表明鱼群适于在拥挤的状态下运动,但是,如果直接将该交互策略迁移驱动机器人,则过小的间隙会导致频繁碰撞,从而触发1.2节所述的警戒停止机制,不利于机器人的工程应用. 图11(d)表明,单纯DNN迁移控制出现了交通阻塞,群体停滞导致集群半径在$C\approx 35$ mm(尺寸缩放系数0.5)附近产生了一个小的峰值;

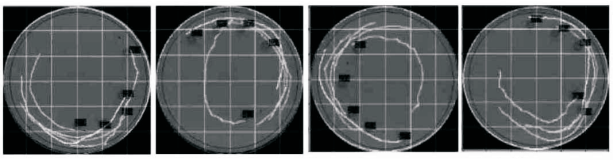


(a) 集群重心相对墙壁距离分布 (b) 集群重心相对墙壁角度分布

(c) 智能体集群极性分布 (d) 智能体集群大小分布

图11 5鱼、机器人群实验的宏观指标分布比较

而DDPG修正后,聚合集群被有目的地延展为长线队列,见图12,间接地拉大了集群半径,从而降低了集群运动的阻塞风险,但是,较高的极性峰值($P \approx 0.8$)说明DDPG的修正并未破坏集群运动的方向一致性.



(a) 0 s　(b) 1 min 30 s　(c) 2 min 38 s　(d) 3 min 52 s

图 12　DNN+DDPG模型控制的5机器人集群运动

图12展示了5机器人DNN+DDPG模型控制的运行轨迹: 当运动中的机器人因机械问题离队时,如图12(b)所示;其余单体通过关键邻居不断交替,保持了剩余4个单体的有序集群运动,如图12(c)所示;接着,落后的单体加速前进,并通过内侧超车方式,重新融入集群,如图12(d)所示. 综上,DNN+DDPG交互模型控制能充分利用生物模型的涌现机制,快速自组织出灵活的集群运动,同时DDPG修正可以优化集群机器人的安全裕度与跟踪性能,增强了集群控制的鲁棒性,使生物集群模型更适于工程应用,符合集群机器人研究的特点与发展趋势.

2.3　5机器人集群运动模型对比

为明晰DNN+DDPG模型的性能,下面将所提出的模型与相关文献中具有代表性的集群交互方法进行1 h的5机器人集群运动实验对比,包括: 1)经典Vicsek模型[17]; 2)焦点单体使用集群运动领域权威专家Guy所在团队提出的成对解耦交互模型[7]来进行最近邻居信息交互; 3)使用与2)相同的交互模型,焦点单体随机选择单个随机邻居进行交互; 4)直接迁移未修正的DNN模型[11]用于真实机器人集群运动,具体结构详见1.1节. 统计集群运动宏观指标的均值与方差,包括:集群重心与墙壁间的距离r_W、集群重心运动方向与墙壁间的角度绝对值θ_W^+、集群极性P以及最近的两个机器人间的距离$d_{\min}$,如表2所示.

表 2　不同交互模型5机器人集群运动宏观指标对比分析

模型	r_W^*/mm	θ_W^+/(°)	P	$d_{\min}^*$/mm
DNN+DDPG	58.7±24.3	92.7±15.8	0.69±0.12	76.4±16.7
DNN	134.1±38.9	89.2±37.6	0.55±0.22	51.5±15.9
最近邻	81.4±36.1	89.6±22.9	0.44±0.21	65.4±25.4
随机邻	89.6±54.0	88.7±44.8	0.63±0.23	55.9±21.6
Vicsek	43.5±22.9	93.0±46.0	0.31±0.16	35.0±5.5
真实鱼	62.7±35.8	86.2±30.3	0.88±0.13	37.5±14.5

*除真实鱼群数据外,机器人实验的长度尺寸缩放系数为0.5.

值得注意的是,当两个机器人紧贴时,采用缩放系数0.5后的距离约为25 mm.

表2数据表明:传统Vicsek模型虽然具有不错的仿真性能,但受制于机器人的性能、约束等实际限制,基本不能生成有序集群运动,集群极性最低,机器人在墙附近阻塞严重;采用Guy团队提出的解耦交互模型[7]进行最近邻居交互,集群极性较低,基本不能涌现出集群运动;选择随机邻居交互,集群涌动性能有所改观,集群极性提升至0.63,但是邻居间最小距离偏小,容易引发阻塞,另外沿墙运动性能一般,不利于集群机器人的实际应用;DNN模型尽管具有鱼群的涌现行为,但由于机器人与真实鱼类之间具有明显的结构与运动差异,导致集群运动效果较差,体现为沿墙跟踪能力低下,并且较低的$d_{\min}$数值体现了集群阻塞的出现. 本文所提出的DNN+DDPG交互模型控制的5机器人集群运动结果显示:集群重心与墙壁间距离r_W较为稳定,与墙壁间角度θ_W^+的标准差最小,基本保持与墙壁平行;集群极性P最高,一致性最好;各机器间的最小距离适中,符合强化奖励设计,能够有效防止集群机器人间的物理碰撞.

3　结　论

机器人集群运动实验表明:所设计的最大视觉角度关键邻居选择方法搭配DNN+DDPG模型控制,可以形成灵活、鲁棒的集群机器人运动,与其他运动控制方法[7,11,17]对比,能够有效利用鱼群仿生涌现机制迁移控制集群机器人,从而将仿生控制的灵活性与鲁棒性带入到真实集群机器人控制中,再加入强化学习的优化能力,增加了集群机器人工程应用的目的性与安全性. 相比于传统深度网络迁移研究[13,23-24],利用强化网络训练来提升深度网络的迁移质量,为跨平台迁移控制领域研究提供了新的思路.

本文所提出的方法仅需要单体对周边环境进行观测,保持少量邻居社会交互,就能实现有序的集群运动,计算负载低、实时性好、集群规模可伸缩、运行灵活鲁棒,所以相应的控制方法有望在建筑物集群维护、大规模集群物流、农牧业集群作业、无人机群非攻击性辅助等领域具有潜在的应用前景.

参考文献(References)

[1] 王伟嘉, 郑雅婷, 林国政, 等. 集群机器人研究综述[J]. 机器人, 2020, 42(2): 232-256.
(Wang W J, Zheng Y T, Lin G Z, et al. Swarm robotics: A review[J]. Robot, 2020, 42(2): 232-256.)

[2] Vicsek T, Zafeiris A. Collective motion[J]. Physics Reports, 2012, 517(3/4): 71-140.

[3] Ning B D, Han Q L, Zuo Z Y, et al. Collective behaviors of mobile robots beyond the nearest neighbor rules with switching topology[J]. IEEE Transactions on Cybernetics, 2018, 48(5): 1577-1590.

[4] Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour[J]. Nature, 2019, 568(7753): 477-486.

[5] Francesca G, Brambilla M, Trianni V, et al. Analysing an evolved robotic behaviour using a biological model of collegial decision making[C]. SAB2012. Berlin: Springer, 2012: 381-390.

[6] Vásárhelyi G, Virágh C, Somorjai G, et al. Optimized flocking of autonomous drones in confined environments[J]. Science Robotics, 2018, 3(20): eaat3536.

[7] Calovi D S, Litchinko A, Lecheval V, et al. Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors[J]. PLoS Computational Biology, 2018, 14(1): e1005933.

[8] Rosenthal S B, Twomey C R, Hartnett A T, et al. Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion[J]. Proceedings of the National Academy of Sciences of the United States of America, 2015, 112(15): 4690-4695.

[9] Berlinger F, Gauci M, Nagpal R. Implicit coordination for 3D underwater collective behaviors in a fish-inspired robot swarm[J]. Science Robotics, 2021, 6(50): eabd8668.

[10] Lei L, Escobedo R, Sire C, et al. Computational and robotic modeling reveal parsimonious combinations of interactions between individuals in schooling fish[J]. PLoS Computational Biology, 2020, 16(3): e1007194.

[11] 刘磊, 孙卓文, 陈令仪, 等. 基于深度学习的仿生集群运动智能控制[J]. 控制与决策, 2021, 36(9): 2195-2202.
(Liu L, Sun Z W, Chen L Y, et al. Intelligent control of bionic collective motion based on deep learning[J]. Control and Decision, 2021, 36(9): 2195-2202.)

[12] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[13] 周涵婷, 程龙生, 乔佩蕊, 等. 基于CiteSpace的故障预测知识结构与热点迁徙研究[J]. 控制与决策, 2022, 37(4): 815-828.
(Zhou H T, Cheng L S, Qiao P R, et al. Knowledge structure and hotspots migration of prognostics based on CiteSpace[J]. Control and Decision, 2022, 37(4): 815-828.)

[14] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning[J]. Computer Sciences, 2015, 8(6): A187.

[15] Dorigo M, Theraulaz G, Trianni V. Swarm robotics: Past, present, and future[J]. Proceedings of the IEEE, 2021, 109(7): 1152-1165.

[16] 刘磊, 陶杰, 尹钟. 微型机器人以及群机器人系统[P]. 中国: CN201710441229.2. 2017-06-13.
(Liu L, Tao J, Yin Z. Micro robot and swarm robot systems[P]. China: CN201710441229.2. 2017-06-13.)

[17] Vicsek T, Czirók A, Ben-Jacob E, et al. Novel type of phase transition in a system of self-driven particles[J]. Physical Review Letters, 1995, 75(6): 1226-1229.

[18] Soudbakhsh D, Eskandarian A. Steering control collision avoidance system and verification through subject study[J]. IET Intelligent Transport Systems, 2015, 9(10): 907-915.

[19] 梁星星, 冯旸赫, 马扬, 等. 多Agent深度强化学习综述[J]. 自动化学报, 2020, 46(12): 2537-2557.
(Liang X X, Feng Y H, Ma Y, et al. Deep multi-agent reinforcement learning: A survey[J]. Acta Automatica Sinica, 2020, 46(12): 2537-2557.)

[20] Lowe R, Wu Y, Tamar A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments[J/OL]. 2017, arXiv: 1706.02275.

[21] Rashid T, Samvelyan M, Witt C S, et al. QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning[J/OL]. 2018, arXiv: 1803.11485.

[22] Sutton R S, Barto A G. Reinforcement learning: An introduction[J]. IEEE Transactions on Neural Networks, 1998, 9(5): 1054.

[23] Sunehag P, Lever G, Gruslys A, et al. Value-decomposition networks for cooperative multi-agent learning based on team reward[C]. AAMAS 2018. Stockholm, 2018: 2085-2087.

[24] 陈佳鲜, 毛文涛, 刘京, 等. 基于深度时序特征迁移的轴承剩余寿命预测方法[J]. 控制与决策, 2021, 36(7): 1699-1706.
(Chen J X, Mao W T, Liu J, et al. Remaining useful life prediction of bearing based on deep temporal feature transfer[J]. Control and Decision, 2021, 36(7): 1699-1706.)

作者简介

刘磊(1982−), 男, 副教授, 博士, 从事集群机器人控制、生物集群系统运动等研究, E-mail: liulei@usst.edu.cn;

张浩翔(1996−), 男, 硕士生, 从事集群机器人控制、生物集群系统运动的研究, E-mail: zhxei@vip.qq.com;

陈若妍(2001−), 女, 本科生, 从事自组织的研究, E-mail: 1020218308@qq.com;

高岩(1962−), 男, 教授, 博士生导师, 从事非光滑优化、生存控制等研究, E-mail: gaoyan@usst.edu.cn;

王富正(1968−), 男, 教授, 博士, 从事新能源汽车控制、模型预测控制等研究, E-mail: fcw@ntu.edu.tw;

王亚刚(1967−), 男, 教授, 博士, 从事过程控制、系统辨识等研究, E-mail: ygwang@usst.edu.cn.

(责任编辑：李君玲)