基于分层自主决策和 DQN 的自适应牧羊控制方法

赵 江[1]，杨 智[1]，池 沛[2†]，王英勋[2]

(1. 北京航空航天大学 自动化科学与电气工程学院，北京 100191；
2. 北京航空航天大学 无人系统研究院，北京 100191)

摘 要: 牧羊控制方法逐渐被应用于机场鸟群驱离、无人机放牧、空地协同监视和引导等大规模集群运动协调问题. 以牧羊无人机为例, 提出基于分层自主决策和深度 Q 网络 (DQN) 的自适应牧羊控制方法. 首先, 考虑离群个体活跃度衰减等因素, 建立牧羊控制问题的感知和运动模型; 然后, 针对个体滞留和离群问题, 提出基于全局质心的弧形轨迹 (GCM-Arc) 控制方法和避障策略, 提升羊群受控个体占比; 最后, 建立分层自主决策模型, 结合 GCM-Arc 控制方法与深度 Q 网络, 提出分层 GCM-Arc 控制方法, 以实现控制模式自适应切换和参数自适应调整. 数字仿真实验表明, 所提出方法在牧羊任务时间、无人机总路程、羊群平均半径、单体离群率和牧羊任务成功率方面, 明显优于经典的两种牧羊控制方法.

关键词: 牧羊控制；无人机；分层自主决策；深度 Q 网络；自适应；路径规划

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2024.0634

引用格式: 赵江, 杨智, 池沛, 等. 基于分层自主决策和 DQN 的自适应牧羊控制方法 [J]. 控制与决策.

Hierarchical autonomous decision and DQN based self-adaptive shepherding control method

ZHAO Jiang[1]，*YANG Zhi*[1]，*CHI Pei*[2†]，*WANG Ying-xun*[2]

(1. School of Automation Science and Electrical Engineering， Beihang University， Beijing 100191， China；
2. Institute of Unmanned System，Beihang University，Beijing 100191，China)

Abstract: The shepherd control method is gradually being applied to address large-scale collective motion coordination problems, such as bird dispersal at airports, drone herding, as well as air-ground coordinated surveillance and guidance. Taking UAV herding as an example, a self-adaptive shepherding control method based on a deep Q-network (DQN) and hierarchical autonomous decision-making is proposed. Firstly, considering the factors such as the decay of the activity of outlying individuals, a perception and motion model of the shepherding control problem is established. Then, a global center of mass arc (GCM-Arc) control method and an obstacle avoidance strategy are proposed to improve the percentage of controlled individuals in the flock for the individual stagnation and outlier problem. Finally, a hierarchical autonomous decision-making model is established, and a hierarchical GCM-Arc control method is proposed by combining the GCM-Arc control method and the DQN, which realizes adaptive switching of control mode and adaptive adjustment of parameters. Simulation experiments demonstrate that the proposed method outperforms classical GCM-V (V-shaped trajectory based on global center of mass) and Arc-Formation shepherding control methods significantly in terms of shepherding task completion time, total drone distance traveled, average radius of the sheep herd, individual outlier rate, and sheep herding task success rate.

Keywords: shepherding control； unmanned aerial vehicle (UAV)； hierarchical autonomous decision； deep Q-network (DQN)；self-adaptive；path planning

0 引 言

多智能体系统 (MAS) 是由多个具备一定感知和通信能力的智能体组成的集合, 个体间通过协调其知识、目标、技能和规划, 能够共同采取行动解决问题[1-2]. 围绕自然界各类生物的集群现象, 通过牧羊犬与羊群间相互作用产生的牧羊 (shepherding) 行为

收稿日期: 2024-05-27；**录用日期:** 2024-11-25.
†通讯作者. E-mail: peichi@buaa.edu.cn.

引起了关注, 即牧羊犬通过对羊群施加作用以影响其运动[3]. 多智能体协同的许多研究工作均是在引入基于行为的控制范式后开展的, 这些基于行为的范式源自仿生学, 且对于多智能体系统的设计具有指导作用[4], 如牧羊行为可应用于机场鸟群驱离[5]、水面泄漏石油清理[6] 以及空地协同监视和引导[7].

牧羊任务是指使用一只或多只牧羊犬控制一群羊从起点移动至目标区域的过程[8]. 而牧羊控制问题是指如何控制牧羊犬完成牧羊任务. 现有的解决牧羊控制问题的方法可分为两类: 基于规则的控制方法和基于机器学习的控制方法[9]. 基于规则的控制方法可由羊群系统的状态参数根据自定义的规则计算出下一时刻智能体的运动参数. 如文献 [10] 提出了一种基于规则的路线图控制方法, 采用了 MAPRM (medial-axis probabilistic roadmap) 方法, 其生成的路线图与 PRM (probabilistic roadmap) 方法[11] 相比更适用于群体在障碍物环境中的运动; 文献 [12] 将牧羊控制问题转化为旅行商问题, 无人机按照规划的顺序靠近分布在空间中的羊群个体附近并将其聚拢; 文献 [13] 提出了一种基于 GCM (global center of mass) 的启发式算法, 无人机可根据羊群的离散状态决定聚拢羊群或朝目标方向驱赶羊群; 文献 [14] 提出了一种无人机按照 V 形轨迹运动的算法, 并将其与文献 [13] 中的 GCM-Targeting 控制方法相结合得到基于全局质心的 V 形轨迹 (GCM-V)控制方法; 文献 [15] 提出了一种以弧形环绕羊群运动从而完成驱赶的弧形轨迹 (Arc-Formation) 控制方法, 通过制定的规则完成环绕羊群的运动方向的切换; 文献 [16] 证明了牧羊系统简化后的模型与三维的非完整航行器系统等价, 且当航行器系统上给定偏移点P趋向目标点时, 羊群会收敛于目标区域; 文献 [17] 构建了一种基于图的羊群模型, 通过几何划分的方式规定了无人机需要控制的子群, 并将其与粒子群优化算法相结合来求解无人机的目标位置; 文献 [18] 在文献 [17] 的基础上增加了无人机路径规划功能, 避免了无人机扰动羊群; 文献 [19] 设计了一种只关注羊群最外层个体的牧羊控制方法, 根据局部视野中距离终点最远的个体计算无人机放牧位置, 且证明了该控制方法能够使得无人机在满足给定感知范围条件下持续追踪羊群; 文献 [20] 通过设计描述无人机间、无人机与羊群间相互作用的势场函数, 从而使得无人机集群实现了对羊群的包围, 并引导羊群向终点移动; 文献 [21] 提出了一种基于无人机局部视野内羊群密度的牧羊控制方法, 能够完成多无人机对羊群的合围, 并讨论了对于一定规模的羊群合围所需的最少无人机数量; 文献 [22] 设计了一种可减少通讯负荷的多无人机牧羊控制方法, 无人机间不需要交换各自获取到的信息, 只需根据局部羊群信息以及相邻无人机位置即可进行决策; 文献 [23] 提出了一种隐式控制 (IC) 方法, 其通过对于羊群和无人机状态进行数值理论分析, 从而得到合适的无人机输入; 文献 [24] 在 IC 方法的基础上, 采用一类分布式的卡尔曼滤波器对于羊群和无人机的状态进行估计, 作为 IC 方法的输入, 提出了协调隐式控制 (CIC) 方法. 这一类基于规则的控制方法在简单的牧羊任务中通常更易于实现, 但是依赖具体数学模型和人对问题的考虑, 缺乏灵活性.

随着人工智能的发展, 近年来的研究工作开始关注基于机器学习的控制方法. 文献 [25] 将强化学习与模仿学习相结合, 依靠人类专家的演示来学习奖励函数结构, 通过优化人类专家的策略, 避免产生类似于人类专家的失误; 文献 [9] 将深度强化学习与 PRM 方法相结合, 训练出能够在障碍物环境中运行的模型. 这一类基于机器学习的控制方法相较于基于规则的控制方法更加灵活, 但是其训练用时较长, 且模仿学习等算法需要人类专家演示的数据集, 耗时耗力.

本文主要内容如下: 1) 考虑离群个体活跃度衰减等问题, 建立羊群模型; 2) 针对个体滞留和离群问题, 提出 GCM-Arc 控制方法, 在牧羊任务时间、无人机总路程、羊群平均半径、单体离群率和牧羊任务成功率方面, 明显优于经典的 GCM-V[14] 与 Arc-Formation[15] 控制方法; 3) 针对控制模式切换条件固定和参数无法实时调节问题, 将所提出 GCM-Arc 控制方法与深度 Q 网络相结合, 通过建立分层自主决策模型, 提出分层 GCM-Arc 控制方法, 与所提出 GCM-Arc 控制方法在羊群平均半径、单体离群率和牧羊任务成功率方面保持一致的情况下, 节省了牧羊任务时间, 减少了无人机总路程.

1 预备知识

1.1 感知和运动模型

羊的模型如下所示:

$$\begin{cases}\dot{\boldsymbol{p}}_i = \xi \cdot \boldsymbol{v}_i,\\ \dot{\boldsymbol{v}}_i = \boldsymbol{u}_i,\end{cases} \quad i = 1, 2, \ldots, N. \tag{1}$$

其中: $\boldsymbol{p}$、$\boldsymbol{v}$分别为羊的位置和速度; i为羊的序号; N为羊的数量; ξ为个体活跃度衰减率, 当个体离群后等于 0.99, 否则为 1. 而羊群内部相互作用的建模是基于经典的 Boid 模型[26], 即

$$\boldsymbol{u}_i = \boldsymbol{u}_i^{\text{sep}} + \boldsymbol{u}_i^{\text{align}} + \boldsymbol{u}_i^{\text{coh}} + \boldsymbol{u}_i^{\text{avoid}}, \tag{2}$$

这里$\boldsymbol{u}_i^{\text{sep}}$、$\boldsymbol{u}_i^{\text{align}}$、$\boldsymbol{u}_i^{\text{coh}}$、$\boldsymbol{u}_i^{\text{avoid}}$分别为分离、对齐、聚集、排斥控制分量. 然后, 在地面坐标系下建立如下无人机模型:

$$\begin{cases} \dot{x} = v \cdot \cos\theta, \\ \dot{y} = v \cdot \sin\theta, \\ \dot{v} = a, \\ \dot{\theta} = \omega. \end{cases} \tag{3}$$

其中: x、y分别为无人机横、纵坐标, v、θ分别为无人机速度大小和航向角, a、ω分别为无人机速度大小变化率和航向角变化率. 本文对羊、无人机的感知、运动能力做出如下假设.

1) 羊的感知区域为以羊的坐标为中心, 半径为r_{perceive}的圆形区域. 当障碍物、无人机以及边界栅栏出现在羊的感知区域内时, 羊将受到排斥作用. 为了模拟真实环境中的不确定性, 用高斯噪声对排斥作用对应控制量的方向角进行扰动.

2) 无人机能够获取当前时刻所有羊的位置信息(x_i, y_i)以及静态障碍物的位置和尺寸. 本文假设无人机飞行高度h_{UAV}大于羊的高度h_{sheep}, 低于障碍物高度h_{fence}, 且当无人机与障碍物发生碰撞后将停止运动.

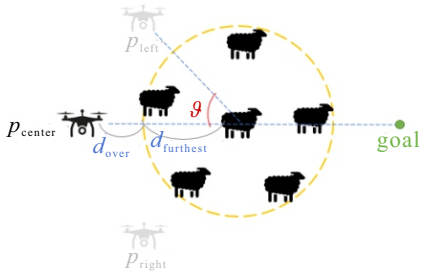
1.2　GCM-V 控制方法

GCM-Targeting 控制方法[13] 将牧羊任务划分为两个子任务: 聚拢羊群和驱赶羊群. 当羊群轮廓半径小于给定阈值时, 无人机尾随在羊群后方将羊群朝目标点方向驱赶; 当羊群轮廓半径大于给定阈值时, 无人机飞至距离羊群质心最远的羊附近, 将其朝羊群质心方向驱赶.

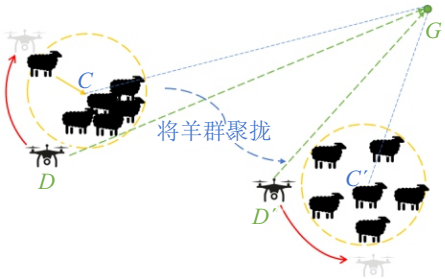
GCM-V 控制方法[14] 在 GCM-Targeting 控制方法的基础上进行了改进, 先计算出所有羊与羊群质心的距离, 取其中的最大值d_{furthest}作为羊群半径, 如图 1(a) 所示 (ϑ和d_{over}由人为给定). 当羊群半径大于给定阈值时, 由目标点经羊群质心引出一条射线, 在该射线上取一点p_{center}, 使得p_{center}与羊群质心间的距离为$d_{\text{furthest}} + d_{\text{over}}$, 然后由$\vartheta$确定$p_{\text{right}}$和$p_{\text{left}}$, 无人机从$p_{\text{right}}$和$p_{\text{left}}$中随机选择一点, 并移动至该点, 随后在 3 点之间循环移动, 本文选取$\vartheta = 60°$, 同时规定d_{over}等于 10 个单位长度大小; 当羊群半径小于给定阈值时, 无人机在羊群后距质心固定距离d_{drive}处将其朝目标点处驱赶.

1.3　Arc-Formation 控制方法

Arc-Formation 控制方法[15] 的流程如图 1(b) 所示. 以直线连接无人机D与目标点G, 参考无人机视角, 判断羊群质心C位于直线DG的左侧或右侧, 若



(a)　V 形轨迹示意图



(b)　Arc-Formation 控制方法示意图

图1　经典牧羊控制方法示意图

位于左 (右) 侧, 则无人机将按照顺 (逆) 时针环绕羊群运动.

假设当前无人机按照顺时针环绕羊群运动, 以直线连接无人机D与羊群质心C, 参考无人机视角, 选取无人机左侧距直线DC最远的羊, 无人机将沿着圆弧移动至这只羊附近. 为了避免无人机一直按照顺时针 (或逆时针) 环绕羊群运动, 这里给定改变环绕方向的阈值$\varphi = 60°$, 则当满足下式时, 进行换向操作:

$$180° - |\langle \overrightarrow{CD}, \overrightarrow{CG} \rangle| \geqslant \varphi. \tag{4}$$

因此, 无人机将会环绕羊群顺时针或逆时针交替运动, 且与羊群质心的距离为$d_{\text{furthest}} + d_{\text{over}}$. 本文规定$d_{\text{over}}$的取值与前文 GCM-V 控制方法的相同.

1.4　DQN 算法

深度 Q 网络算法[27] 是一种经典的深度强化学习算法. 给定策略π, 将动作价值函数$Q_\pi(s, a)$定义为智能体在状态s下采取动作a后继续使用策略π与环境进行交互获得奖励的期望值, 其表达式为

$$Q_\pi(s, a) = E_\pi\Big[\sum_{i=0}^{\infty} \gamma^i r_{t+i} \Big| s_t = s,\ a_t = a\Big]. \tag{5}$$

其中: γ为折扣因子, $E_\pi(\cdot)$为在策略π下的数学期望, r_{t+i}为在$t+i$时刻得到的奖励. 在决策过程中, 智能体将在动作空间中选择使得在当前状态s_t下动作价值函数$Q_\pi(s, a)$取值最大的动作a_t, 即

$$a_t = \arg\max_a Q_\pi(s_t, a). \tag{6}$$

DQN 的核心步骤包括经验回放、目标网络稳定

Q值计算、策略网络更新趋近价值函数, 以及通过ε-贪婪策略平衡探索与利用, 从而实现稳定高效的深度强化学习[28].

2 分层 GCM-Arc 牧羊控制方法

本节针对个体滞留和离群问题, 提出了 GCM-Arc 控制方法和避障策略. 为了实现控制模式自适应切换和参数自适应调整, 建立了分层自主决策模型, 提出了分层 GCM-Arc 控制方法.

2.1 个体滞留和离群问题

GCM-V 控制方法以及 Arc-Formation 控制方法的相同点是当羊群半径大于给定阈值后无人机以某种形状的轨迹 (弧形或 V 形) 环绕羊群运动, 将处于羊群轮廓边界的个体向羊群中心驱赶, 同时, 根据目标的方位调整无人机运动轨迹, 使得羊群朝向目标运动, 但是上述控制方法有可能失效, 具体如下.

1) 当存在离群个体时, 羊群中其他个体将不会对该离群个体产生任何影响, 假设此时羊群半径在不断增加, 由于规则的固定性, 无人机只能沿着 V 形 (或弧形) 轨迹运动, 即使无人机对于离群个体仍然有驱赶作用, 但是无人机只在极短的时间内于运动轨迹上与该离群个体相遇, 其产生的影响可能不足以使得该离群个体返回羊群.

2) 当环境中存在障碍物时, 羊群在障碍物凹槽内的滞留可能会导致离群个体出现. 随着羊群半径不断增加, 无人机运动轨迹的跨度将增加, 使得无人机往返运动周期增加, 进而导致牧羊任务失败. 即使羊群并未在障碍物中滞留, 无人机在阻碍羊群接触障碍物的过程中, 可能会造成羊群个体局部密度不均匀, 从而导致羊群轮廓发生变化, 甚至出现羊群个体离群现象.

3) 当环境中存在障碍物时, 无人机在驱赶羊群过程中可能会滞留在障碍物的凹槽内. 当出现该情况后, 无人机仍然将根据羊群位置计算运动轨迹, 而由一组固定规则产生的轨迹很可能会与障碍物重合, 使得无人机再次与障碍物碰撞, 当无人机找不到合适路径时, 会滞留在障碍物凹槽中, 导致牧羊任务失败.

2.2 分层 GCM-Arc 控制架构

分层 GCM-Arc 控制方法架构如图 2 所示. 无人机决策模块由两个决策网络、GCM-Targeting 控制器、Arc-Formation 控制器、羊群运动路径规划器以及信息提取和整理模块构成, 其根据所有羊的位置、无人机位置、目标点位置、障碍物位置, 计算出无人机的目标位置; 而无人机运动控制律模块采用经典的比例调节器, 负责根据无人机当前位置、速度以及无人机目标位置计算无人机控制量; 无人机运动模型则根据控制量更新无人机的速度、位置信息; 羊群模型由所有羊的模型通过分布式的信息交互构成, 其根据无人机位置, 更新所有羊的速度和位置信息.

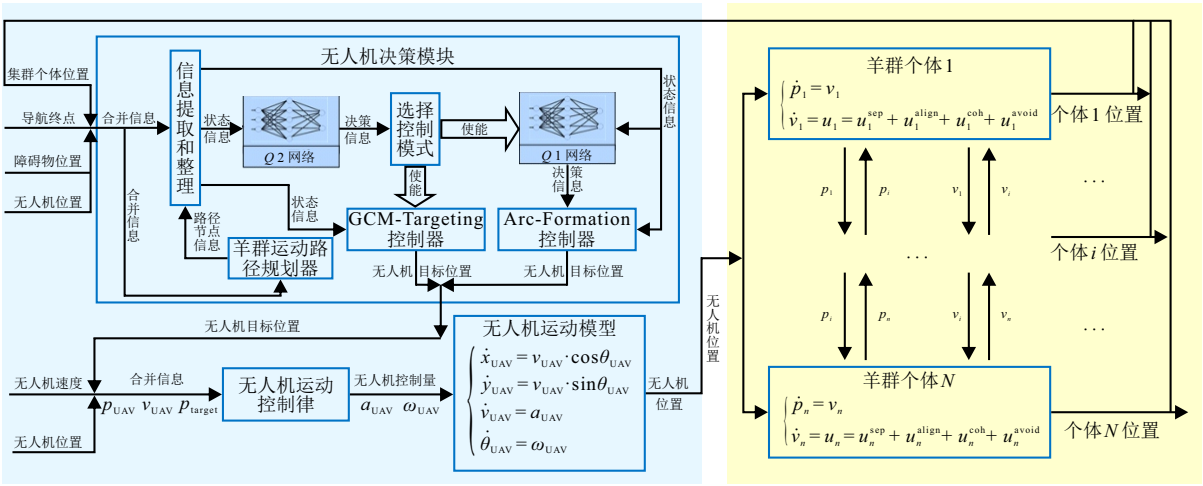


图2 分层 GCM-Arc 控制方法架构

2.3 GCM-Arc 控制方法

本节针对滞留和离群问题, 考虑根据羊群状态切换控制模式、提升羊群受控个体占比、规划羊群运动路径以及建立无人机避障策略, 设计如下方案.

1) 当羊群离散时采用 GCM-Targeting 控制方法; 为了提升羊群受控个体占比, 当羊群聚拢后采用 Arc-Formation 控制方法, 使得无人机贴合羊群轮廓运动. 值得注意的是, 这里判断羊群是否离散的依据并非传统的“羊群半径是否过大”, 而是离群个体是否存在. 其流程如下: 首先求出距羊群质心最远的羊与质

心间的距离d_{furthest}, 然后计算所有羊与质心间的平均距离d_{average}, 当$d_{\text{furthest}} \geqslant \mu \cdot d_{\text{average}}$时, 判定存在离群个体 ($\mu$由人为给定), 此时, 无人机会运动至距羊群质心最远的离群个体附近, 随后将离群个体朝羊群质心处驱赶; 反之, 则无人机沿弧形轨迹环绕羊群运动.

2) 规划一条远离障碍物的羊群路径, 在满足与障碍物保持给定距离的前提下, 为了避免无人机能耗过大, 路径长度不能过大. 值得注意的是, 在任务过程中将通过比较羊群质心与接下来所有路径节点间距, 动态选取下一个更近的路径节点.

3) 建立无人机避障策略, 在无人机采用 Arc-Formation 控制方法牧羊过程中, 与障碍物发生碰撞时立即改变环绕羊群运动的方向, 以达到避障效果; 在无人机采用 GCM-Targeting 控制方法牧羊过程中, 将与障碍物发生碰撞时会规划出一条避开障碍物前往离群个体附近的路径, 无人机沿此路径运动过程中, 当其与离群个体距离小于给定阈值时, 将继续采用 GCM-Targeting 控制方法牧羊.

2.4　分层自主决策模型

2.4.1　控制模式自适应切换和参数自适应调整

1) 所提出 GCM-Arc 控制方法实现了聚拢与驱赶控制模式的切换, 但是其切换条件是固定的, 可能会导致无人机在未考虑到的场景采用错误控制模式. 如当羊群被障碍物分割时, 应停止驱赶, 将羊群聚拢, 但是由于羊群只是被分为两部分, 未出现离群现象, 无人机只会继续驱赶羊群, 而这将导致羊群继续分裂.

2) GCM-V 和 Arc-Formation 控制方法均有其可调参数. 如在 Arc-Formation 控制方法中用于判断无人机是否符合方向切换条件的阈值φ, 还有决定无人机离羊群质心距离的参数d_{over}等, 而上述参数在无人机执行牧羊任务过程中均是固定的, 但是实际上, 对于不同的羊群离散程度以及障碍物相对于羊群的位置均有其适合的数值.

2.4.2　模型框架

本节在所提出 GCM-Arc 控制方法的基础上, 构造分层自主决策模型框架如图 3 所示.

为了解决在训练过程中网络模型难以收敛的问题, 图 3 框架仍然预先规划羊群运动路径, 同时, 其实现了分层自主决策, 在第 1 层决策中控制模式决策网络$Q2$负责选择 GCM-Targeting 或 Arc-Formation 控制方法生成无人机目标位置, 在第 2 层决策中轨迹参数决策网络$Q1$负责调整 Arc-Formation 控制方法参数.

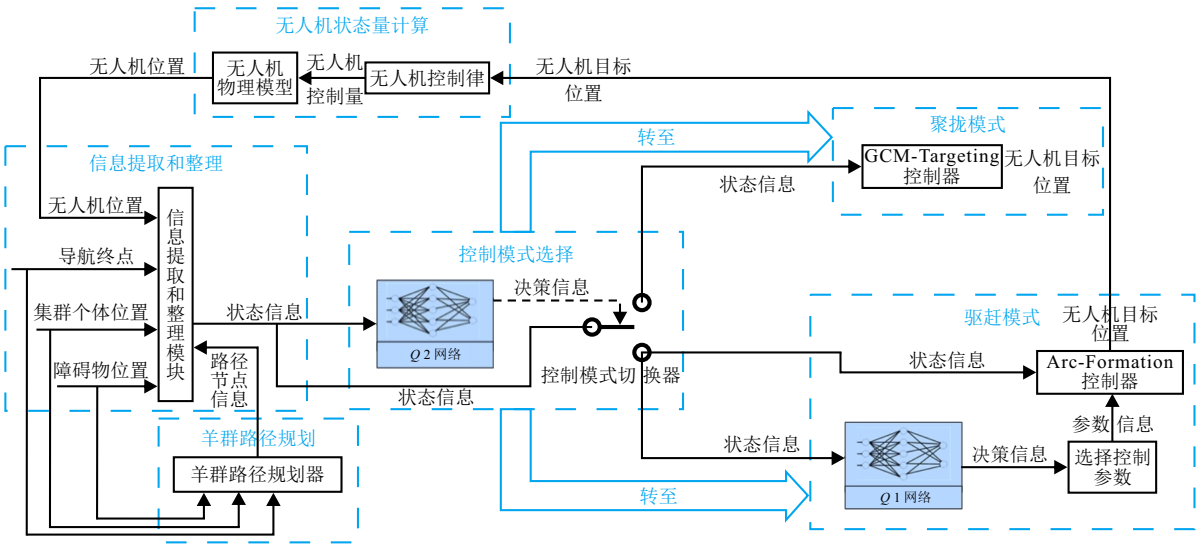


图3　分层自主决策模型框架

2.4.3　轨迹参数决策网络$Q1$的训练

本文分层自主决策模型框架中两个决策网络分别承担不同的任务, 因此, 本节采取不同的训练方法. 首先, 在 Arc-Formation 控制方法执行牧羊任务的场景下训练轨迹参数决策网络$Q1$; 然后, 将$Q1$与 GCM-Arc 控制方法相结合; 最后, 在此基础上训练控制模式决策网络$Q2$.

$Q1$网络负责调节 Arc-Formation 控制方法的参数, 而在文献 [15] 中提到 Arc-Formation 控制方法对于决定无人机何时改变环绕羊群运动方向的参数非常敏感, 这里将环绕的方向 (顺时针或逆时针) 直接作为控制器参数进行学习. 本文对于$Q1$决策网络的动作集合设置如表 1 所示.

表1　$Q1$决策网络动作集合

序号	0	1
动作	顺时针旋转	逆时针旋转

$Q1$网络的状态观测集合如表 2 所示. 由于提供一些在观测量的基础上由人为计算得到的信息会使得模型在训练过程中更易收敛, 除无人机坐标、羊的坐标、当前子目标节点坐标以及障碍物信息, 本文在状态观测集合中添加了羊群质心速度矢量$\boldsymbol{v}_{\rm GCM}$以及羊群的理想速度矢量$\boldsymbol{v}_{\rm optimal}$. 其中: $\boldsymbol{v}_{\rm GCM}$为相邻两个时间步间羊群质心坐标矢量差, 而$\boldsymbol{v}_{\rm optimal}$为当前羊群路径目标节点与羊群质心坐标矢量差.

表2 $Q1$决策网络状态观测集合

项目	无人机坐标	羊群个体坐标	障碍物采样坐标	当前子目标节点
数量	1	20	264	1
项目	羊群半径	羊群质心坐标	$\boldsymbol{v}_{\rm GCM}$	$\boldsymbol{v}_{\rm optimal}$
数量	1	1	1	1

$Q1$网络训练过程中的奖惩机制可分为 3 个部分: 目标点奖励、羊群状态惩罚以及无人机移动惩罚. 其中: 羊群状态惩罚可分为离散状态惩罚和速度矢量状态惩罚, 如下所示:

$$r_t = \lambda_{\rm goal}\cdot r_t^{\rm goal} + \lambda_{\rm disp}\cdot H(R_t^{\rm flock}, R_{\rm given}) + \lambda_{\rm vec}\cdot|\langle \vec{v}_{\rm GCM}, \vec{v}_{\rm optimal}\rangle| + \lambda_{\rm route}\cdot d_{\rm route}. \tag{7}$$

这里: $\lambda_{\rm goal}$与羊群是否到达目标点有关, 若到达则$\lambda_{\rm goal}=1$, 否则$\lambda_{\rm goal}=0$; $\lambda_{\rm disp}$为离散状态惩罚系数; $R_t^{\rm flock}$为羊群在t时刻的轮廓半径; $R_{\rm given}$为期望的羊群半径; $\lambda_{\rm vec}$为速度矢量状态惩罚系数; $\langle \vec{v}_{\rm GCM}, \vec{v}_{\rm optimal}\rangle$为$\vec{v}_{\rm GCM}$与$\vec{v}_{\rm optimal}$的夹角; $\lambda_{\rm route}$为无人机移动惩罚系数; $d_{\rm route}$为无人机在单个时间步内运动的路程; $H(\cdot)$为软阈值算子, 下式给出其表达式:

$$H(u,v) = \mathrm{sign}(u)\cdot\max(0, |u|-v). \tag{8}$$

离散状态惩罚是指当集群半径过大或出现离群个体时给出的惩罚. 速度矢量状态惩罚是指当羊群质心速度矢量与理想速度矢量角度差过大时会给出惩罚.

2.4.4 控制模式决策网络$Q2$的训练

$Q2$网络负责决策无人机在当前环境状态下应选择聚拢控制模式还是驱赶控制模式. 而聚拢控制模式对应 GCM-Targeting 控制方法, 驱赶控制模式对应 Arc-Formation 控制方法. 可得到$Q2$网络的动作集合如表 3 所示.

表3 $Q2$决策网络动作集合

序号	0	1
动作	GCM-Targeting 控制方法	Arc-Formation 控制方法

$Q2$网络的状态观测集合与$Q1$相同, 如表 2 所示. $Q2$网络的奖惩机制在$Q1$网络的基础上增加了强制学习奖惩$r_t^{\rm impose}$. 强制学习惩罚是指在无人机执行牧羊任务过程中, 当羊群处于异常状态时 (羊群轮廓半径异常、羊群异常停滞等), 若$Q2$网络还不能选择正确控制模式, 则会对其进行惩罚.

$Q2$网络与$Q1$网络在训练过程中采用同一组超参数, 其数值如表 3 所示.

2.5 算法流程和伪代码

本节根据本文分层自主决策模型框架, 设计分层 GCM-Arc 控制方法, 其流程如下.

step 1: 在第 1 个时间步中根据环境信息规划羊群运动路径.

step 2: 由无人机与羊群、环境交互, 获取羊群、障碍物信息.

step 3: 将羊群、障碍物信息和羊群路径节点信息合并得到状态信息, 输入决策网络$Q2$.

step 4: 决策网络$Q2$根据状态信息选择控制模式.

step 5: 若 step 4 中决策网络$Q2$选择聚拢控制模式, 则采用 GCM-Targeting 控制方法根据状态信息生成下一时刻的无人机目标位置; 若选择了驱赶控制模式, 则将状态信息输入决策网络$Q1$, 由网络$Q1$选择 Arc-Formation 控制方法的参数, 然后由 Arc-Formation 控制方法生成下一时刻无人机的目标位置.

step 6: 无人机根据下一时刻的目标位置计算控制量, 由控制量计算状态量, 同时获取羊群、障碍物的信息并重复以上流程.

综上, 分层 GCM-Arc 控制方法的伪代码如算法 1 所示.

算法1 分层GCM-Arc控制方法.

1. 初始化羊群和无人机模型参数, 初始化时间步$t=1$;
2. 初始化无人机坐标、羊群个体数量N和位置坐标;
3. 加载轨迹参数决策网络$Q1$权重参数;
4. 加载控制模式决策网络$Q2$权重参数;
5. while 集群未到达目标区域
6. if 还未规划羊群运动路径 then
7. 规划羊群路径, 生成路径节点;
8. end
9. if 无人机到达$P_{\rm target}^t$ then
10. if 无人机处于避障模式 then
11. 选取避障路径中下一个节点作为$P_{\rm target}^{t+1}$;
12. else
13. 使用$Q2$决策网络选择控制模式;
14. if 选择聚拢控制模式 then
15. 采用GCM-Targeting控制方法计算$P_{\rm target}^{t+1}$;

16. else
17. 使用$Q1$决策网络选择控制方法参数;
18. 采用Arc-Formation计算出P_{target}^{t+1};
19. end
20. end
21. else
22. $P_{\text{target}}^{t+1} = P_{\text{target}}^{t}$;
23. end
24. 更新时间步$t = t + 1$;
25. end

3 仿真验证

本节验证 GCM-Arc 以及分层 GCM-Arc 控制方法的有效性, 并将其与 GCM-V[14]、Arc-Formation[15] 控制方法进行比较. 这里以障碍物环境中的牧羊任务为例, 设计仿真实验. 其中: 仿真环境中放置 6 个障碍物, 其中心点初始坐标分别为 (175, 200)、(400, 200)、(625, 200)、(175, 400)、(400, 400)、(625, 400).

本文以 PRM 算法为例, 在环境中均匀随机生成 100 个采样点, 然后删除与障碍物重合的采样点, 结合 A*算法[29], 以羊群质心为起点, 以目标点为终点构造羊群运动路径; 以 RRT 算法[30] 为例, 以无人机为起点, 以离群个体为终点, 构造无人机避障路径; 以 DQN 算法为例, 分别训练轨迹参数决策网络$Q1$、控制模式决策网络$Q2$, 并将训练过程数据绘制成如图 4 所示的曲线.

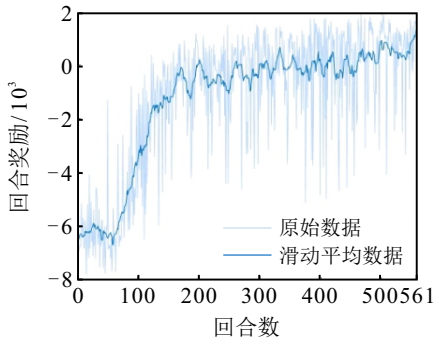
由图 4 可得出以下结论: 由图 4(a) 和图 4(c) 可知, 其回合奖励曲线均有明显上升趋势, 这表明网络参数沿正确的方向进行了更新; 由图 4(b) 和图 4(d) 可知, ε参数最后稳定在 0.05, 这表明在后面的训练过程中仅有 0.05 的几率会产生随机动作, 模型将会减少探索次数并逐步稳定.

本文在相同仿真环境中, 针对 16 种不同的羊群个体数量 $(20 \sim 35)$, 分别进行 10 组仿真实验, 共 160 组. 图 5 为其中 1 组具有代表性的无人机轨迹.

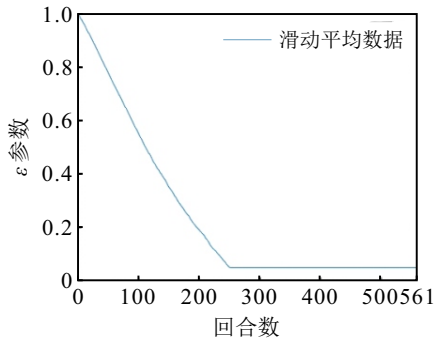
通过分析图 5 无人机轨迹可知: GCM-V[14] 以及 Arc-Formation[15] 控制方法均会出现个体滞留和离群问题; 所提出 GCM-Arc 控制方法对应无人机轨迹环绕羊群运动的弧长较大, 这种充分包围羊群的方式会使得羊群半径较小; 所提出分层 GCM-Arc 控制方法会根据羊群的状态决定无人机环绕羊群运动的弧长, 因此, 无人机轨迹中每段圆弧的弧度均是自适应的, 即使羊群平均半径较大, 但是无人机最终完成了牧羊任务, 且总路程最短. 这里将 4 种牧羊控制方法对应的指标数据平均值记录在表 4 中.

通过分析表 4 中的数据, 将所提出分层 GCM-Arc、GCM-Arc 控制方法与 GCM-V[14]、Arc-Formation[15] 控制方法进行对比, 可得出以下结论.

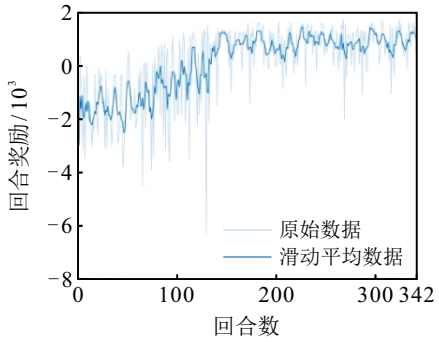
1) 所提出分层 GCM-Arc 控制方法对应平均总时间步数小于 GCM-V 与 Arc-Formation 控制方法的 45 %, 所提出 GCM-Arc 控制方法对应平均总时间步数小于其他两种现有方法的 50 %.



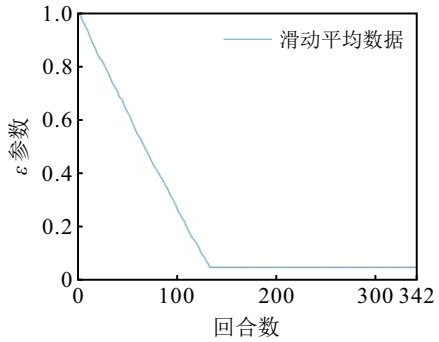
(a) $Q1$ 网络奖励曲线



(b) $Q1$ 网络ε参数变化曲线

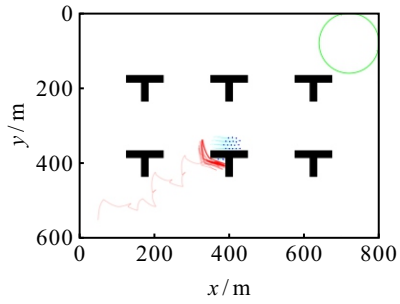


(c) $Q2$ 网络奖励曲线

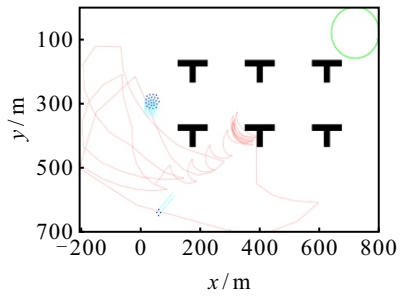


(d) $Q2$ 网络ε参数变化曲线

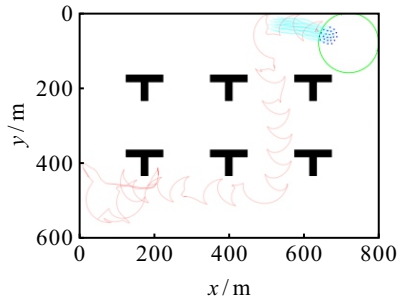
图4　决策网络训练评估曲线



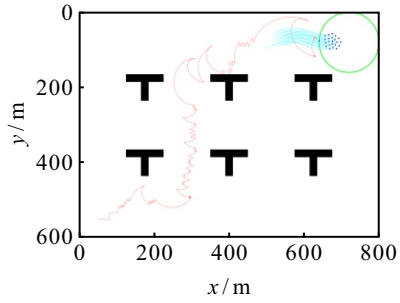
(a) GCM-V 控制方法[14]



(b) Are-Formation 控制方法[15]



(c) GCM-Arc 控制方法



(d) 分层GCM-Arc 控制方法

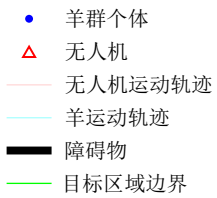


图5 牧羊任务仿真流程

表4 牧羊控制方法仿真实验数据平均值 (160 组)

	GCM-V[14]	Arc-Formation[15]	GCM-Arc	分层GCM-Arc
总时间步数	33889.581	30386.156	14272.550	12841.200
无人机总路程/m	7650.190	7617.555	4846.013	4221.207
羊群平均半径/m	366.890	210.572	39.335	49.976
单体离群率	0.794	0.598	0.194	0.166
成功率	0.081	0.413	0.900	0.825

2) 所提出分层 GCM-Arc 控制方法对应无人机平均总路程最小, 与其他两种现有方法相比, 平均总路程缩减大于 50 %.

3) 所提出方法对应羊群平均半径小于其他两种现有方法对应数值的 25 %.

4) 所提出分层 GCM-Arc 控制方法对应平均单体离群率最小, 小于其他两种现有方法的 33 %; 所提出 GCM-Arc 控制方法对应平均单体离群率小于其他两种现有方法的 50 %.

5) 所提出分层 GCM-Arc 与 GCM-Arc 控制方法对应平均成功率大于其他两种现有方法的 2 倍.

为了探究所提出分层 GCM-Arc 和 GCM-Arc 控制方法在不同羊群个体数量下相较于 GCM-V 与 Arc-Formation 控制方法的优势, 这里将在不同羊群个体数量下记录的各评价指标数据的平均值以及成功率绘制成如图 6 所示的一组曲线图. 通过分析可知: 在本文仿真环境中, 当羊群个体数量在20 ～ 35 典型范围内变化时, 所提出方法在牧羊任务时间、无人机总路程、羊群平均半径、单体离群率和牧羊任务成功率方面, 明显优于经典的 GCM-V 与 Arc-Formation 牧羊控制方法.

值得注意的是, 相较于文献 [14] 与文献 [15], 本文考虑了障碍物环境中的牧羊问题; 文献 [20] 中无人机对羊群进行驱赶的前提条件是完成合围, 而本文并未涉及合围, 只关注如何通过设计单架无人机的控制量从而使得羊群受控; 文献 [9] 中使用强化学习模型直接生成无人机速度, 而本文考虑到收敛性等因素, 采用强化学习模型对控制模式和控制参数进行选择, 从而实现牧羊控制.

4 结 论

本文研究了基于分层自主决策和深度 Q 网络的牧羊控制方法. 首先, 考虑离群个体活跃度衰减等因素, 建立了牧羊控制问题的感知和运动模型; 然后, 针对个体滞留和离群问题, 提出了 GCM-Arc 控制方法; 最后, 在 GCM-Arc 控制方法的基础上, 建立了分层自主决策模型, 通过构造轨迹参数决策网络$Q1$和控制模式决策网络$Q2$, 提出了分层 GCM-Arc 控制方法. 在障碍物环境下的 160 组仿真实验结果表明: 所提出方法相较于经典的 GCM-V、Arc-Formation 控制方法, 在牧羊任务成功率方面的提升大于 100 %, 在牧羊任务用时以及单体离群率方面低于原来的 50 %, 在无人机总路程方面的缩减大于 25 %, 在羊群平均半径方面的缩减大于 75 %, 且本文基于已提出的 GCM-Arc 控制方法进行的分层自主决策的改

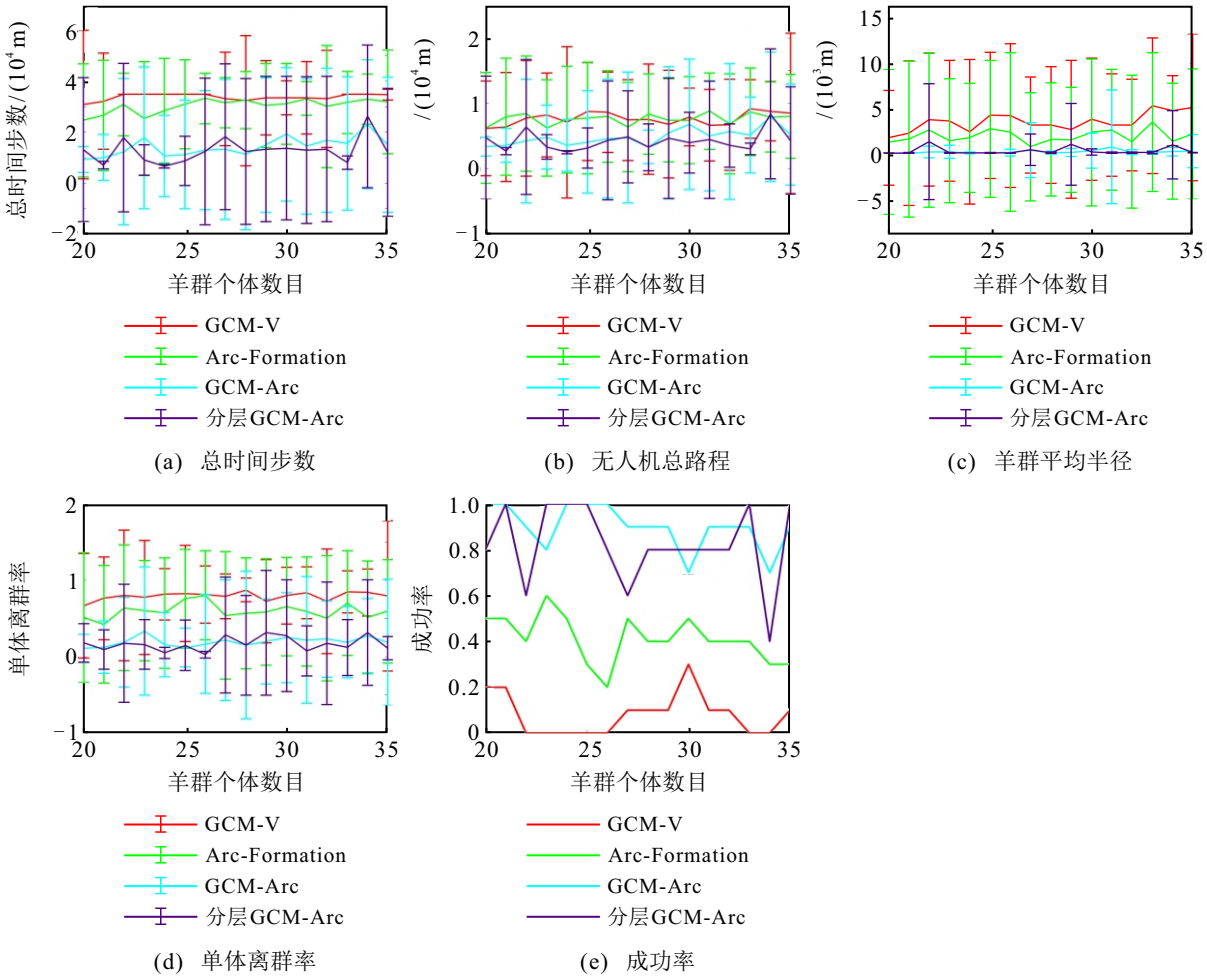


图6　牧羊控制方法性能指标对比

进, 使其在节省牧羊用时以及缩短无人机路程方面得到了进一步提升. 后续将结合分层自主决策和多智能体强化学习方法, 研究大规模羊群场景下协同多无人机分群自适应牧羊控制和导引避障问题.

参考文献 (References)

[1] Grant T J. A review of multi-agent systems techniques, with application to Columbus user support organisation[J]. Future Generation Computer Systems, 1992, 7(4): 413-437.

[2] 汤泽, 王佳枫, 王艳, 等. 非连续混合自时延多智能体系统的饱和分布式控制[J]. 控制与决策, 2023, 38(3): 670-680.
(Tang Z, Wang J F, Wang Y, et al. Consensus of discontinuous multi-agent systems with hybrid self-delays via saturated distributed control[J]. Control and Decision, 2023, 38(3): 670-680.)

[3] Lien J-M, Bayazit O B, Sowell R T, et al. Shepherding behaviors[C]. IEEE International Conference on Robotics and Automation. New Orleans, 2004: 4159-4164.

[4] Arai T, Pagello E, Parker L E. Guest editorial advances in multirobot systems[J]. IEEE Transactions on Robotics and Automation, 2002, 18(5): 655-661.

[5] Paranjape A A, Chung S J, Kim K, et al. Robotic herding of a flock of birds using an unmanned aerial vehicle[J]. IEEE Transactions on Robotics, 2018, 34(4): 901-915.

[6] Özdemir A, Gauci M, Gross R. Shepherding with robots that do not compute[C]. Artificial Life Conference Proceedings. Lyon, 2017: 332-339.

[7] Chaimowicz L, Kumar V. Aerial shepherds: Coordination among UAVs and swarms of robots[M]. Distributed Autonomous Robotic Systems 6, 2008.

[8] 姚军. 基于改进聚集策略的仿生牧羊任务研究[D]. 重庆: 重庆大学, 2021: 1-6.
(Yao J. Research on bio-inspired shepherding task based on improved aggregation strategies[D]. Chongqing: Chongqing University, 2021: 1-6.)

[9] Zhi J X, Lien J-M. Learning to herd agents amongst obstacles: Training robust shepherding behaviors using deep reinforcement learning[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 4163-4168.

[10] Bayazit O B, Lien J-M, Amato N M. Roadmap-based flocking for complex environments[C]. Proceedings of the 10th Pacific Conference on Computer Graphics and Applications. Beijing, 2003: 104-113.

[11] Kavraki L E, Svestka P, Latombe J C, et al. Probabilistic roadmaps for path planning in high-dimensional configuration spaces[J]. IEEE Transactions on Robotics and Automation, 1996, 12(4): 566-580.

[12] Imahayashi W, Tsunoda Y, Ogura M. Route design in

sheepdog system-traveling salesman problem formulation and evolutionary computation solution[J]. Advanced Robotics, 2024, 38(9/10): 632-646.

[13] Strömbom D, Mann R P, Wilson A M, et al. Solving the shepherding problem: Heuristics for herding autonomous, interacting agents[J]. Journal of the Royal Society, Interface, 2014, 11(100): 20140719.

[14] Fujioka K, Hayashi S. Effective shepherding behaviours using multi-agent systems[C]. IEEE Region 10 Conference. Singapore, 2016: 3179-3182.

[15] Bennett B, Trafankowski M. A comparative investigation of herding algorithms[C]. Proc. Symp. on Understanding and Modelling Collective Phenomena. Birmingham, 2012: 33-38.

[16] Pierson A, Schwager M. Controlling noncooperative herds with robotic herders[J]. IEEE Transactions on Robotics, 2018, 34(2): 517-525.

[17] Mohamed R E, Elsayed S, Hunjet R, et al. A graph-based approach for shepherding swarms with limited sensing range[C]. IEEE Congress on Evolutionary Computation. Kraków, 2021: 2315-2322.

[18] Mohamed R E, Hunjet R, Elsayed S, et al. Connectivity-aware particle swarm optimisation for swarm shepherding[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, 7(3): 661-683.

[19] Zhang S, Lei X K, Duan M Y, et al. A distributed outmost push approach for multirobot herding[J]. IEEE Transactions on Robotics, 2024, 40: 1706-1723.

[20] Nguyen D D K, Paul G, Alempijevic A. Decentralized multi-phase formation control for cattle herding[C]. IEEE International Conference on Robotics and Automation. Yokohama, 2024: 17948-17953.

[21] Zhang S, Pan J. Collecting a flock with multiple sub-groups by using multi-robot system[J]. IEEE Robotics and Automation Letters, 2022, 7(3): 6974-6981.

[22] Li A Y, Ogura M, Wakamiya N. Communication-free shepherding navigation with multiple steering agents[J]. Frontiers in Control Engineering, 2023, 4: 989232.

[23] Sebastián E, Montijano E, Sagüés C. Adaptive multirobot implicit control of heterogeneous herds[J]. IEEE Transactions on Robotics, 2022, 38(6): 3622-3635.

[24] Sebastián E, Montijano E, Sagüés C. On the distributed multi-robot herding[C]. Workshop on Distributed Graph Algorithms for Robotics at ICRA. London, 2023: 1-5.

[25] Nguyen H T, Garratt M, Bui L T, et al. Apprenticeship learning for continuous state spaces and actions in a swarm-guidance shepherding task[C]. IEEE Symposium Series on Computational Intelligence. Xiamen, 2019: 102-109.

[26] Reynolds C W. Flocks, herds and schools: A distributed behavioral model[C]. Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques. New York, 1987: 25-34.

[27] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518: 529-533.

[28] 陈梓豪, 胡春鹤. 基于 DQN 出价策略的多无人机目标分配拍卖算法[J]. 聊城大学学报: 自然科学版, 2024, 37(4): 23-32.
(Chen Z H, Hu C H. Multi drones target allocation auction algorithm based on DQN bidding strategy[J]. Journal of Liaocheng University: Natural Science Edition, 2024, 37(4): 23-32.)

[29] Hart P E, Nilsson N J, Raphael B. A formal basis for the heuristic determination of minimum cost paths[J]. IEEE Transactions on Systems Science and Cybernetics, 1968, 4(2): 100-107.

[30] Rodriguez, Tang X Y, Lien J-M, et al. An obstacle-based rapidly-exploring random tree[C]. Proceedings ot the IEEE International Conference on Robotics and Automation. Orlando, 2006: 895-900.

作者简介

赵江 (1986–) 男, 副教授, 博士, 主要研究方向为无人集群智能决策与控制、智能感知与自主飞行控制, E-mail: jzhao@buaa.edu.cn;

杨智 (2001–) 男, 硕士生, 主要研究方向为无人集群智能决策与控制, E-mail: zhiyang@buaa.edu.cn;

池沛 (1980–) 男, 教授, 博士, 主要研究方向为智能自主控制体系架构与原理方法、无人系统智能自主控制技术与应用, E-mail: peichi@buaa.edu.cn;

王英勋 (1964–) 男, 研究员, 主要研究方向为创新布局无人机总体设计、无人集群智能自主控制, E-mail: wangyx@buaa.edu.cn.