

脑注意力机制启发的群体智能协同避障方法

项羽铭¹, 陈焜¹, 赵志峰^{1,2}, 李荣鹏¹, 张宏纲¹

(1. 浙江大学信息与电子工程学院, 浙江 杭州 310027;

2. 之江实验室, 浙江 杭州 311121)

摘要: 现有群体智能系统执行探索任务时往往获取、计算和传输大量冗余信息, 造成群智系统资源利用效率低下。设计任务驱动的计算和通信资源高效融合利用机理成为亟待解决的关键科学问题。基于此, 提出了一种脑注意力机制启发的群体智能协同避障方法。受脑注意力机制启发, 群智系统引入基于深度 Q 网络的传感器智能选择模块, 实现在探索未知环境时高效选择传感器的工作状态, 利用尽量少的传感器开销获取和计算关键信息; 以最佳交互碰撞避免算法为基础, 单一智能体融合邻居智能体的少量关键信息, 驱动传感器智能选择模块, 在完成集群协同避障的同时, 大幅降低传感器获取和计算信息的冗余度。在仿真平台及 Kehepera IV 机器人实际场景中验证, 结果显示, 所提方法可以显著降低集群系统传感器的信息冗余度, 并且随着智能体数量以及信息交互量的增加, 性能增益更加显著。

关键词: 群体智能; 资源利用效率优化; 深度 Q 网络; 最佳交互碰撞避免

中图分类号: TP18

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202215

Crowd intelligent cooperative obstacle avoidance method inspired by brain attention mechanism

XIANG Yuming¹, CHEN Kun¹, ZHAO Zhifeng^{1,2}, Li Rongpeng¹, ZHANG Honggang¹

1. College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou 310027, China

2. Zhejiang Lab, Hangzhou 311121, China

Abstract: A crowd intelligent (CI) system often acquires, calculates, and transmits a large amount of redundant information during the performing of exploration tasks, which inevitably results in inefficient use of the limited resources. Therefore, it emerges a strong incentive to design a task-driven mechanism for efficient utilization of computing and communication resources. A crowd intelligent cooperative obstacle avoidance method inspired by brain attention mechanism was proposed. Inspired by brain attention mechanism, the CI system introduced an intelligent selection module based on the deep Q network, by efficiently tuning the working state of sensors exploring the unknown environment and realized the acquisition and calculation of key necessary information with as little sensor overhead as possible. Meanwhile, based on the optimal reciprocal collision avoidance algorithm, a single agent fuses a small amount of limited information from neighbor agents to drive the intelligent selection module, so as to greatly reduce the redundancy of sensor acquisition and information calculation required for the obstacle avoidance task. The effectiveness of this proposed method was verified through extensive simulation analyses and practical realization empowered with Kehepera IV robots.

收稿日期: 2021-12-15; 修回日期: 2022-01-14

通信作者: 李荣鹏, lirongpeng@zju.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61731002, No.62071425); 浙江省重点研发计划资助项目 (No.2019C01002, No.2019C03131); 浙江省自然科学基金资助项目 (No.LY20F010016); 华为合作项目

Foundation Items: The National Natural Science Foundation of China (No.61731002, No.62071425), The Key Research and Development Program of Zhejiang Province (No.2019C01002, No.2019C03131), The Natural Science Foundation of Zhejiang Province (No.LY20F010016), Huawei Cooperation Project

The results show that the proposed method can significantly reduce the redundancy of sensor information in the CI system. More importantly, as the number of agents and the amount of information interaction increase, there also emerged a clear trend in the increase of performance gains.

Key words: crowd intelligent, optimization of resource utilization efficiency, deep Q network, optimal reciprocal collision avoidance

0 引言

智能移动机器人集环境感知、动态决策与规划、行为控制与执行等多功能于一体,并且涵盖了计算机与电子工程、自动化控制、信息处理等众多领域的研究成果,在物流分拣、工业制造、抢险救灾等应用中发挥了巨大的作用,也能代替人进行一些极端环境下的作业^[1-5]。在未知环境下进行路径选择以及自主避障是移动机器人进行路径规划必不可少的基本能力之一。在早期,国内外围绕单个移动智能体导航及避障的研究已经取得了很多不错的成果^[6-8]。然而很多算法都依赖于已知条件下的环境,在实际应用中有很大的局限性。近年来,基于深度学习的算法不断被提出,在未知环境中也能发挥出良好的效果^[9-10]。但随着技术的进一步发展,单个移动机器人的局限性越来越突出,人们迫切需要一种作业效率和可靠性更高的智能系统。因此将目光逐渐从单智能体转向群体智能,在比较密集的环境中进行群体智能协同避障也成为研究的热点及难点之一。

按照控制方式,现有的关于群体智能协同避障问题的算法可大致分为两种:集中式控制和分布式控制。集中式控制将运动控制问题定义为优化问题,中央控制器获取所有智能体的当前位置、速度和终点信息,最终目标是成功引导所有智能体到达它们所期望的位置,在避免发生碰撞的同时优化整体所需要的能量或时间等指标。集中式控制虽然能得到全局较优的结果,但它的弊端也非常明显:当智能体数量比较庞大时,其计算开销会变得非常高。此外,在实际的大规模或密集环境智能体控制问题中,通常很难建立一个通信网络使得所有智能体与中央控制器之间可以进行可靠的信息传输,并且在突发情况下单个智能体不具备自治能力,稳定性较差。而分布式控制能很好地解决这些问题。分布式控制将计算、感知以及决策的工作下发到每个智能体上。与集中式控制相比,分布式控制在计算量、运行成本、鲁棒性以及扩展性等方面都有巨大的优势。在分布式控制的群智协同避障场景中,

Berg J V D 等人^[11]提出了最佳交互碰撞避免(optimal reciprocal collision avoidance, ORCA)算法,能够使每个智能体在观测周围智能体的位置和速度后,通过求解一个线性规划问题计算出下一个时间范围内的局部最优速度,从而避免发生碰撞。由于其出色的性能,ORCA 算法已经成为常用的底层避障算法之一,同时 Li H 等人^[12]对 ORCA 算法做了一定拓展,使得其可以应用于非线性速度控制的智能体(如自行车运动模型的智能体规划),大大提升了 ORCA 算法的适应性。国内也有相当多的研究聚焦在分布式的群体智能控制上,如朱杰斌等人^[13]采用分布式编队控制策略设计了无人机编队飞行时的编队控制器,刘华罡等人^[14]在人工势场理论的基础上构造了一种变权重自适应分布式多智能体编队与避障控制算法。

近年来,基于分布式控制的群体智能系统凭借其天然优势,受到越来越多的学者和专家的关注,并得到了许多可喜的研究成果,但依旧存在许多亟待解决的问题。例如在实际的机器人场景中,受到自身的计算资源、环境中的通信带宽资源以及信号的传输距离等限制,机器人在共同完成一些复杂的任务时,特别是当系统中的机器人数量十分庞大时,除了需要考虑网络拓扑结构的限制,还需要考虑自身能量的损耗和通信带宽的利用率等。此外,机器人在未知的环境中执行任务时,只有通过对环境具有感知功能的传感器才能进行探索、判断以及决策。目前移动机器人常用的传感器主要包括摄像机、超声波、红外线、惯性测量单元、光电编码器等。结合多传感器丰富的环境感知能力的优势,让移动机器人可以在复杂的环境中保持稳定的性能并执行特定任务,是目前机器人应用领域需要解决的重点和难点。例如,如何在群智协同工作时对传感器数据融合进行优化以及如何降低传感器工作时的冗余度,从而减少智能体受限的能量,解决计算资源的瓶颈问题,仍缺乏行之有效的解决方法。

在大脑认知科学中,注意力机制(attention mechanism)指的是人们大脑在面对大量信息时会

选择性地将注意力聚焦在重要信息而忽略其他信息,把信息处理资源分配给最需要的部分意味着其可以在整体计算能力受限时进行合理的资源分配,极大地提高人类视觉等信息处理的效率。近几年在深度学习中应用的注意力机制可以被看作人类认知行为的一种模拟,其由于突出的表现而被广泛应用于在自然语言处理、图像识别及语音识别等不同类型的机器学习任务中^[15]。受此启发,笔者利用注意力机制来优化群体智能系统中有限能量或计算资源的分配问题,减少信息贡献度较低的传感器的工作,降低能量消耗。

因此,本文提出了一种脑注意力机制启发的群体智能协同避障方法,旨在在保证机器人完成任务的同时,降低执行任务时的传感器获取信息的冗余度。目前流行的注意力机制^[16]的核心是在众多信息中通过训练加权系数,给不同的信息给予不同的权重,将更多的关注放在更关键的信息上,从而实现对输入的理解和处理。本文的脑注意力机制实质上是对注意力这一概念的模仿,通过智能体对当前状态进行分析,在获得执行任务所需要的关键信息后,将有限的信息处理资源分配给最需要的部分,从而节省整体开销。与目前流行的注意力机制相比,本文的脑注意力机制不需要采用复杂的结构,更偏向于认知科学概念中的注意力,直接通过神经网络来判断执行协同避障任务时的关键传感器信息,进而实现资源的倾斜。本文所考虑的群体智能协同避障环境如图1所示,多个机器人在运动过程中与环境进行双向的信息交互,同时在一定距离范围内可与其他机器人进行少量信息交互,并通过机器人内部的处理模块达到群智协同避障的效果。具体而言,机器人在执行协同避障任务时,将自身信息(已知环境信息、当前速度、位置等)以及与近距离内的机器人交互得到的信息作为初始信息,设计基于深度Q网络(deep Q network, DQN)的传感器智能选择模块分析初始信息,并选择当前需要打开的关键传感器。随后将关键传感器探索到的环境信息与初始信息整合,通过避碰决策模块规划出各机器人需要执行的动作,从而减少传感器的开销,并将资源集中在关键信息的处理上,实现类似脑注意力的效果。最后,笔者将该算法部署到仿真环境和由3个Khepera IV机器人构成的实际场景中,验证所提方法在实际场景中的有效性。

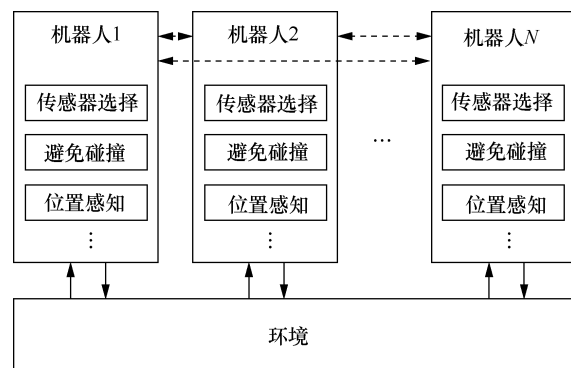


图1 群体智能协同避障环境

本文的主要贡献如下。

(1) 本文提出了一种基于DQN的群智注意力机制,赋能传感器智能选择模块,通过对周围的环境信息进行不同粒度的划分,合理规划各传感器的使用,降低获取信息的冗余度。

(2) 本文对ORCA算法的使用进行扩展,通过结合传感器智能选择模块的结果以及一定距离内邻居机器人的信息,动态规划集群不发生碰撞的速度,并将结果转换为双轮式差速驱动运动模型来实现避障。

(3) 本文使用仿真实验环境以及Khepera IV机器人实际环境进行实验,实验结果证明了机器人集群协同工作的高效性及必要性。

1 相关工作

1.1 DQN

在AI领域中,强化学习(reinforcement learning, RL)让智能体基于环境而行动,以最大化智能体在环境中获得的总奖励值(reward),其重要目标是通过在智能体与环境的交互过程中不断试错来修正其决策机制,使得智能体能适应特定的环境,做出利于自身或全局目标最优的决策^[17]。近年来,随着结合了强化学习和深度神经网络(deep neural network, DNN)的深度强化学习(deep reinforcement learning, DRL)^[18]在各个领域大放异彩,许多先进的DRL算法大大提高了智能体从环境中学习的效率,例如DQN^[19]。在智能体避障和决策领域,有很多基于DQN的算法取得了很好的成果。Xin J等人^[20]和Yue P Y等人^[21]均提出了利用DQN进行自主导航与避障的算法;罗鹏等人^[22]在DQN的基础上提出了一种智能车辆决策框架,解决了高速场景下的车辆决策问题。

在DQN中,总的状态空间为 S ,当前状态

和次态分别为 s 、 s' ，动作空间和当前动作分别记为 A 和 a 。传统的 Q -learning 算法使用 Q 表来记录并迭代更新在状态 s 下采取动作 a 时的 Q 值，以求出准确的值函数。 Q -learning 在更新时将当前状态下的 Q 值和奖励值与次态 s' 中最大的 Q 值进行加权求和，其更新计算式如下（其中 α 、 γ 为参数）：

$$Q(s, a) \leftarrow Q(s, a) + \alpha(\text{reward} + \gamma \max_{a'} Q(s', a') - Q(s, a)) \quad (1)$$

不难看出，当状态空间和动作空间维数较大或为连续空间时，值函数无法用一张表格来表示，而神经网络刚好可以从高维原始数据中学习，并通过拟合复杂的函数来逼近值函数，DQN 就是在此基础上被提出来的。

本文使用的 DQN 利用卷积神经网络来近似值函数，并将两个结构完全相同的网络作为主网络和目标网络，分别预测当前状态和次态的 Q 值。与 Q -learning 类似，DQN 将次态中最大的 Q 值与奖励值的加权值作为主网络训练的标签，从而最小化均方误差。

DQN 中的损失函数计算式如下：

$$L(\theta_i) = E \left[\left(\text{reward} + \gamma \max_a Q(s', a'; \theta_i^-) - Q(s, a; \theta_i) \right)^2 \right] \quad (2)$$

其中， $Q(s, a; \theta_i)$ 和 $Q(s', a'; \theta_i^-)$ 分别是主网络和目标网络的预测值，为了提高算法的稳定性，主网络的参数每次迭代均会更新，而目标网络的参数每隔一定步数再进行更新。

由于通过强化学习采集到的数据前后具有较强的关联性，而顺序训练会在一定程度上影响训练结果，因此 DQN 采用了经验回放（experience replay）的方式进行训练，即把所有环境探索时采集到的数据（当前状态、次态、当前动作、奖励值）存储到一个数据库中，再利用均匀随机采样从数据库中批量抽取数据来训练网络，从而打破数据间的相关性。同时为了避免最终结果落入局部最优，网络训练时也采用了 ε -greedy 探索策略以增强对环境的探索能力，可以在已知的信息中最大化奖励的同时发掘环境中更多的信息。

1.2 ORCA 算法

ORCA 算法^[11]是一种可以保证大量机器人在杂乱的工作空间中进行局部无碰撞运动的算法，它基于速度障碍（velocity obstacles, VO）^[23]区域的

思想，将避障问题转化为非常高效的线性规划问题，其基础流程如下。

由于机器人感知的环境有限，ORCA 算法的首要目标为保证各机器人在时间窗 τ 内不发生碰撞。记每个机器人当前位置为 p ，当前速度为 v ，半径为 r ，最大速度为 $v^{\max}$ ，首选速度（即指向机器人目标点的速度）为 v^{pref} ，且对于两个机器人 A 和 B ，将 A 相对于 B 的速度记为 $v_{A-B} = v_A - v_B$ ， B 对 A 产生的速度障碍区域记为 $\text{VO}_{A/B}^\tau$ ，这是所有会导致 A 和 B 在时间窗 τ 内发生碰撞的 v_{A-B} 的集合，其正式定义如下：

$$\text{VO}_{A/B}^\tau = \{v \mid \exists t \in [0, \tau], tv \in D(p_B - p_A, r_A + r_B)\} \quad (3)$$

其中， $D(p_B - p_A, r_A + r_B)$ 表示以点 $p_B - p_A$ 为中心、半径为 $r_A + r_B$ 的圆盘区域。此时可获取其相对速度 v_{A-B} 到速度障碍区域 $\text{VO}_{A/B}^\tau$ 边界的最短向量 u ：

$$u = (\arg \min_{v \in \partial \text{VO}_{A/B}^\tau} \|v - (v_A - v_B)\|) - (v_A - v_B) \quad (4)$$

u 为 A 和 B 在时间窗 τ 内避免碰撞所需要最小速度变化量，一般情况下采用两个机器人速度均改变 $u/2$ 的方式，使得整体达到能避免碰撞的状态（当 B 为静止障碍物时， A 将负责全部的速度变化量）。记 u 的单位向量为 n ，则 A 在与 B 避障时的可行速度集合 $\text{ORCA}_{A/B}^\tau$ 为从点 $v_A + u/2$ 开始的指向 n 方向的半平面：

$$\text{ORCA}_{A/B}^\tau = \left\{ v \mid \left(v - \left(v_A + \frac{1}{2}u \right) \right) \cdot n \geq 0 \right\} \quad (5)$$

当 A 与多个机器人进行避障时， A 与周围每个机器人都生成一个可行速度的半平面集合，将这些集合与 A 的最大速度取交集，即最终的可行速度集合 ORCA_A^τ 为：

$$\text{ORCA}_A^\tau = D(0, v_A^{\max}) \bigcap_{B \neq A} \text{ORCA}_{A/B}^\tau \quad (6)$$

此时在可行速度集合中选取速度便可以保证 A 不会与周围机器人发生碰撞，同时为了让 A 能选择更好的路线，选择最接近其首选速度 v_A^{pref} 的新速度 v_A^{new} ：

$$v_A^{\text{new}} = \arg \max_{v \in \text{ORCA}_A^\tau} \|v - v_A^{\text{pref}}\| \quad (7)$$

由此可见， ORCA_A^τ 是由多个可行速度的半平面的线性约束产生的凸区域。因此可以通过线性规划，计算出避免碰撞的新速度。

1.3 Khepera IV 机器人

Khepera IV 机器人是一款由 K-Team 公司在 2015 年 1 月发布的微型移动机器人。该机器人周身配置 12 个红外传感器, 5 个超声波传感器, 高精度的三轴陀螺仪、三轴加速度计, 两个高效率、高质量的直流电机以及适配的编码器里程计 (odometry); 无线通信通过蓝牙或 Wi-Fi 完成; 正前方有 1 个彩色摄像头, 两侧有高增益的扬声器。

在本文的实验场景中, Khepera IV 机器人移动并使用传感器扫描周围环境, 因此机器人需要实时计算自身的位置信息。但由于存在两方面的限制, 精度相对较低的 Wi-Fi 和蓝牙等定位模型也不适用: 一方面由于成本等因素并没有在机器人内部集成类似于 GPS 等的定位模块, 另一方面本文使用的是室内的实验场景。因此本文将里程计作为位置估计的方法, 利用机器人移动过程中传感器的数据来估计机器人的位置。此外, 推算定位 (dead reckoning) 算法是一个借助先前的位置信息、速度信息, 推导出当前位置的算法。双轮式差速驱动运动模型如图 2 所示, 本文将里程计信息与推算定位算法结合, 完成机器人位置估计。

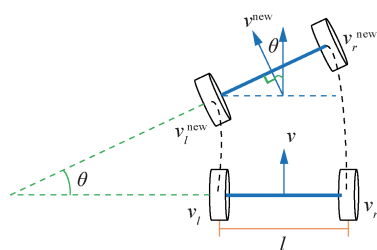


图2 双轮式差速驱动运动模型

如图 2 所示, 通过左右轮里程计编码器在一个单位时间间隔内的数值增量, 可以得到左右轮的速度大小 v_l 和 v_r , 由双轮式差速驱动运动模型可以求得机器人的速度大小 v 和角速度大小 w :

$$\begin{aligned} v &= \frac{v_r + v_l}{2} \\ w &= \frac{v_r - v_l}{l} \end{aligned} \quad (8)$$

其中, l 表示机器人左右轮间的距离。由此即可计算出机器人的中心位置相较于上一时刻的位置增量 Δx 和 Δy :

$$\begin{aligned} \Delta x &= v \Delta t \cdot \cos(\theta_0 + w \Delta t) \\ \Delta y &= v \Delta t \cdot \sin(\theta_0 + w \Delta t) \end{aligned} \quad (9)$$

其中, θ_0 表示上一时刻机器人的方向。至此, 便可求出机器人当前的位置信息。

2 系统算法

2.1 算法框图

在分布式控制的密集障碍物群智协同避障实验中, 每个机器人在受限于通信距离和传感器感知范围的情况下, 需要根据获取到的状态信息做出收益尽可能大的决策。因此, 笔者建立如下系统模型: 在一定区域范围内放置若干个机器人, 将此设置为机器人的初始状态; 以整个区域的中心点为对称中心, 机器人以此对称中心设置与初始状态相对应的目标状态; 并假设该区域中存在静止的实体障碍物。实验目标为: 机器人在通信距离和传感器感知范围受限且对环境信息未知的初始状态下, 向目标状态转移, 期间不与其他机器人和静止的实体障碍物发生碰撞, 与此同时, 尽可能减少传感器的开销。

基于此, 笔者设计了脑注意力机制启发的群体智能协同避障方法框图, 如图 3 所示。该方法主要可以分为两个部分: 传感器智能选择模块和避碰决策模块。传感器智能选择模块以深度强化学习中的 DQN 为基础, 旨在根据机器人在前一时刻的状态以及与其他机器人交互的信息合理地使用传感器, 并做出决策, 从而减少机器人在探索未知环境的过程中使用过多的传感器所带来的额外资源开销; 避碰决策模块将机器人自身获取的环境信息和与其他机器人交互得到的信息融合作为 ORCA 算法的输入, 计算不发生碰撞的速度输出, 进一步通过双轮式差速驱动运动模型将此输出结果转换为机器人左右轮的速度并执行。传感器智能选择模块和避碰决策模块相互配合实现了在不与其他机器人和静止障碍物发生碰撞的前提下, 降低传感器的使用数量, 减少部分冗余的传感器资源的开销。

2.2 传感器智能选择模块

虽然传感器是机器人获取环境信息的重要手段, 但是无法保证每个传感器获取的环境信息均对当前任务有帮助。受到注意力机制的启发, 希望机器人在工作时也将有限的能量与计算资源分配到关键任务上, 因此笔者设计了传感器智能选择模块来避免传感器上的能量资源浪费, 进而在能量受限的情况下保障关键任务的执行。

需要注意的是, 笔者关注的是认知科学中的脑注意力机制概念, 并非深度学习算法中的注意力结

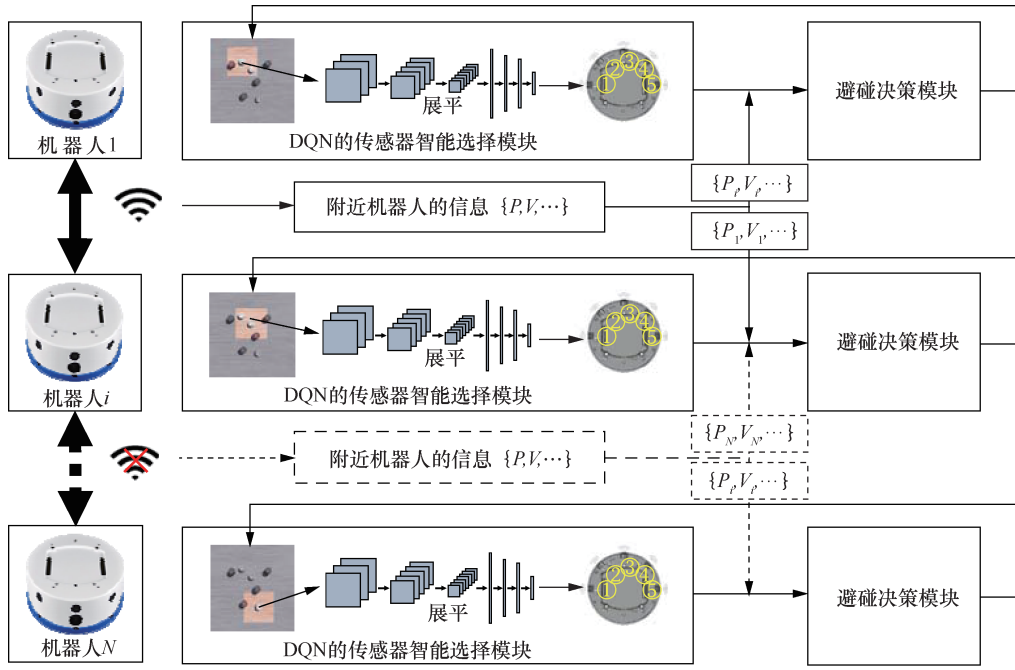


图3 脑注意力机制启发的群智协同避障方法框图

注：图中的实线代表机器人间距离较近，可以相互通信并传递信息；虚线代表机器人间距离较远，无法建立有效通信； P 、 V 分别代表机器人的位置信息和速度信息

构，智能体通过神经网络对当前状态进行分析，来判断执行协同避障任务时的关键传感器信息，并做出资源分配决策，从而节省整体开销。在该模块中，该注意力机制的输入为当前状态，而神经网络是该注意力机制进行分析的主体部分，并输出需要采取的动作，同时将针对各状态和动作计算得到的奖励值作为衡量该注意力机制所做出的决策质量的指标。

其中，笔者设计的网络与奖励函数均仅针对单个机器人，在训练完成后，各机器人将独立采用相同的网络进行传感器的控制，进而通过避障算法做出决策，并未进行整体的共享与协调。在未来的工作中，笔者将尝试针对整个机器人集群进行传感器智能选择网络的设计与实现^[24-25]。

(1) 问题定义

在本问题中，当前状态 s 为当前时刻机器人周围的局部地图，可将该地图等分为 $n \times m$ 个子区域，每个子区域 i 有一个区域类型 c_i ，且 $c_i \in \{U, E, O\}$ ，其中 U 、 E 、 O 分别表示未探索、空旷、障碍物 3 种类型，此时 s 可表示成 $s = \{c_i\}^{n \times m}$ ，其为 $n \times m$ 的地图信息矩阵。当前动作 a 为机器人设置的各传感器开关状态， k 个传感器中每个传感器 i 均有一个状态 d_i ，且 $d_i \in \{Y, N\}$ ，其中 Y 、 N 分别表示开和关状态，此时 a 即可表示成 $a = \{d_i\}^{1 \times k}$ ，其为 $1 \times k$ 的动

作向量。将根据动作打开传感器后更新的地图信息矩阵作为次态 s' 。状态空间 S 和动作空间 A 的具体定义如下：

$$\begin{aligned} S &= \{c_1 \times \cdots \times c_{n \times m} \mid c_i \in \{U, E, O\}, i \in \{1, \cdots, n \times m\}\} \\ A &= \{d_1 \times \cdots \times d_k \mid d_i \in \{Y, N\}, i \in \{1, \cdots, k\}\} \end{aligned} \quad (10)$$

本问题的最终目标是让机器人能够根据周围环境的探测情况自行选择需要使用的传感器，即对当前状态 s 进行分析，并采取动作 a 来最大限度地找出周围潜藏的障碍物或可以前进的空旷区域，同时尽量减小传感器的能量开销。

(2) 奖励函数设计

奖励函数是强化学习中非常重要的一部分。在本问题中，需要优先保证机器人不与障碍物发生碰撞，并在此基础上减小传感器的能量开销，即需要选择的传感器给机器人提供的信息足以使其完成避障。最终的奖励函数如下：

$$\text{reward} = \varphi(s) \cdot I(s, s') - E(a) \quad (11)$$

其中， $I(s, s')$ 为环境信息探索量，该探索量可以比较当前状态与次态中探索环境的数量，直接体现了使用传感器带来的信息量增加情况； $E(a)$ 为传感器的能量消耗代价，这是为了保证每个传感器的能量消耗带来足够的收益； $\varphi(s)$ 为环境信息比例因子，该值与

当前状态 s 中的未知环境信息总量呈正相关, 这意味着在当前状态已知的环境信息较少时, 探索环境的优先级高于节省传感器的能量, 可以用于鼓励机器人打开更多的传感器来获取信息, 避免碰撞。

(3) 神经网络结构与训练

传感器智能选择模块神经网络架构如图4所示(图4中 W 、 H 、 C 分别表示各卷积层输出的宽、高、通道数, N 表示各全连接层输出的尺寸), 由3层卷积层(conv)和3层全连接层(FC)构成, 其中使用 ReLU 激活函数; 以 60×60 的周围环境地图为输入, 每个像素有 U 、 E 、 O 3 种取值, 分别代表未探索、空旷、障碍物; 由于实验中最多使用机器人的5个传感器, 动作类型共 $2^5=32$ 种, 因此可将 1×32 的动作作为输出, 将其中的最大值作为最终动作类型; 并采用经典的 DQN 架构, 训练时设置主网络及目标网络, 采用 ε -greedy 策略来更新经验池。此时的状态空间 S 为:

$$S = \{c_1 \times \dots \times c_{60 \times 60} \mid c_i \in \{U, E, O\}, i \in \{1, \dots, 60 \times 60\}\} \quad (12)$$

传感器智能选择网络训练算法如算法1所示。

算法1 传感器智能选择网络训练算法

输入: 折扣因子 γ , 学习率 α , 环境因子参数 β , 传感器能量参数 λ , 批大小 m , 迭代次数 M , $\hat{Q}$ 网络更新频率 C ;

输出: θ ;

1: 机器初始化回放缓存区 B , 容量为 N ;

2: 随机初始化 Q 网络的参数 θ ;

3: 随机初始化 $\hat{Q}$ 网络的参数 $\hat{\theta} = \theta$;

4: for episode=1 to M do

5: 随机生成 60×60 地图作为状态 s ;

6: 根据概率 ε 选择随机动作 a ;

7: 否则 $a = \arg \max_a Q(s, a)$;

8: 智能体打开动作 a 对应的传感器感知周围环境, 并得到状态 s' ;

9: 计算状态 s 和 s' 中的未探索区域大小 $\text{Unk}(s)$ 和 $\text{Unk}(s')$, 以及动作 a 对应的传感器数量 $\text{Num}(a)$;

10: 设置

$$\varphi(s) = \beta \cdot \text{Unk}(s),$$

$$I(s, s') = \text{Unk}(s) - \text{Unk}(s'),$$

$$E(a) = \lambda \cdot \text{Num}(a)$$

11: 得到环境反馈的奖励

$$\text{reward} = \varphi(s) \cdot I(s, s') - E(a);$$

12: 将 $(s, a, \text{reward}, s')$ 保存到 B ;

13: 从 B 中随机采样 m 个四元组

$$(s_j, a_j, \text{reward}_j, s'_j), j = 1, 2, \dots, m;$$

14: 设置 $y_j = \text{reward}_j + \gamma \max_{a'} \hat{Q}(s'_j, a')$;

15: 对损失函数 $(y_j - Q(s_j, a_j))^2$ 使用 Adam 算法更新 Q 网络参数 θ ;

16: 每隔 C 步更新 $\hat{Q}$ 网络的参数 $\hat{\theta} = \theta$;

17: end for

2.3 避碰决策模块

在传统的 ORCA 算法中, 在计算机器人之间的相互避碰速度区间时, 需要明确且固定的机器人数量才能计算出相对准确的避碰速度。但在本文的实验场景中, 受到通信条件的限制, 无法保证机器人能及时收到所有障碍物的信息。因此, 笔者考虑将时间因素

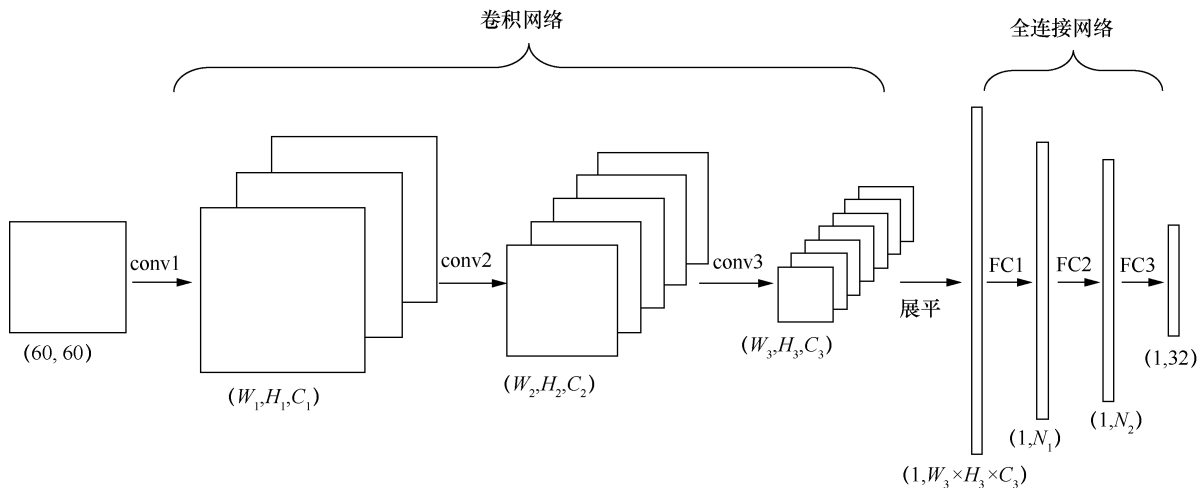


图4 传感器智能选择模块神经网络架构

作为 ORCA 算法的另一个参数,不同时间段每个机器人对应的附近机器人的数量是不同的;同时,笔者也将第 2.2 节中传感器智能选择模块的结果加入 ORCA 算法中,使得避碰的效果更加准确。

不失一般性,设编号为 A 的机器人向目标状态转移的过程中,由于受到通信距离以及传感器感知范围的限制, A 获取的其他机器人的信息 $R(s_i)$ 以及由自身传感器探知到的环境信息 $O(s_i)$ 会随着时间 t 不断变化,其中 $O(s_i)$ 为探知到的周围圆柱形静止障碍物的集合。

对于障碍物 $C \in O(s_i)$, 结合式 (3)、式 (4), 可以得到在 τ 时间内避免碰撞所需要的 A 与 C 的最小速度变化量,同时由于 C 为静止障碍物,因此 A 将负责全部的速度变化量并生成对应的可行速度集合 $ORCA_{A|C}^\tau$, 进而对 $O(s_i)$ 中每个障碍物生成的可行速度集合与 A 的最大速度限制取交集即可获得针对传感器探知环境信息的可行速度集合 $ORCA_{A|O(s_i)}^\tau$:

$$ORCA_{A|O(s_i)}^\tau = D(0, v_A^{\max}) \cap_{C \in O(s_i)} ORCA_{A|C}^\tau \quad (13)$$

类似地,可以求得 $ORCA_{A|R(s_i)}^\tau$, 因此机器人 A 在 t 时刻的最终可行速度集合 $ORCA_A^\tau(t)$ 如下:

$$ORCA_A^\tau(t) = ORCA_{A|R(s_i)}^\tau \cap ORCA_{A|O(s_i)}^\tau \quad (14)$$

通过式 (14)、式 (7) 共同得到机器人 A 在时间窗 τ 内不发生碰撞的新速度 v_A^{new} 以及角度变化量 $\Delta\theta$:

$$\begin{aligned} v_A^{\text{new}} &= (v_x^{\text{new}}, v_y^{\text{new}}) \\ \Delta\theta &= \arctan\left(\frac{v_y^{\text{new}}}{v_x^{\text{new}}}\right) - \theta_0 \end{aligned} \quad (15)$$

由于本文实验使用的 Khepera IV 机器人为双轮式差速驱动运动学模型,因此需要将上述 ORCA 算法计算的新速度 v_A^{new} 转换为机器人左右轮的速度 $v_w^{\text{new}} = (v_l^{\text{new}}, v_r^{\text{new}})$ 。将式 (8) 以及式 (15) 结合,可计算 v_w^{new} 如下:

$$\begin{aligned} v_r^{\text{new}} &= \|v_A^{\text{new}}\| + \frac{\Delta\theta \cdot l}{2\tau} \\ v_l^{\text{new}} &= \|v_A^{\text{new}}\| - \frac{\Delta\theta \cdot l}{2\tau} \end{aligned} \quad (16)$$

式 (16) 同时考虑运动学约束和传感器使用约束,实现平面内导航的多个双轮式差速驱动机器人模型与 ORCA 算法的结合。避碰决策算法如算法 2 所示。

算法 2 避碰决策算法

输入: 附近所有机器人的信息 R , 机器人的环境信息 O ;

输出: 机器人的速度 v_w^{new} ;

1: 机器人根据信息 R 和 O 将附近机器人与障碍物信息合并,生成附近全部信息 $N = R \cup O$;

2: 机器人根据 N 计算与附近所有机器人和障碍物的相对距离;

3: 机器人根据当前与目标位置的方向计算首选速度 v^{pref} ;

4: 机器人根据 N 分别计算对应可行速度集合 $ORCA^\tau$;

5: 机器人使用线性规划从 $ORCA^\tau$ 中选择最接近首选速度 v^{pref} 的 v^{new} ;

6: 机器人将 v^{new} 通过双轮式差速驱动运动模型转换为 v_w^{new} 。

综上,本文使用整体算法如算法 3 所示。

算法 3 脑注意力机制启发的群智协同避障算法

1: 初始化 ORCA 算法避障的相关参数,设置每轮执行速度变化的时间窗 τ ;

2: 初始化机器人各传感器、通信等功能模块,机器人的环境信息 $O(s_i)$ 和附近其他机器人的信息 $R(s_i)$;

3: repeat

4: 机器人将当前获取的环境信息 $O(s_i)$ 作为传感器智能选择网络的输入,得到需要执行的传感器选择动作 a ;

5: 机器人根据 a 使用对应的传感器功能来更新当前状态的环境信息 $O(s'_i)$;

6: 机器人将当前的环境信息 $O(s'_i)$ 与在通信范围内其他机器人的信息 $R(s_i)$ 融合作为算法 2 的输入;

7: 机器人通过算法 2 计算机器人不发生碰撞的速度 v_w^{new} ;

8: 机器人的电机模块执行 v_w^{new} 并等待时间窗 τ ;

9: until 机器人从初始状态转移到目标状态。

3 实验结果与分析

本节使用仿真软件和实际环境分别验证算法 3 的有效性,同时在实验过程中根据实际情况对传感器智能选择网络的训练过程进行了一定调整。在仿真环境中,笔者使用 Webots 仿真平台验证不同机器人数量以及静止障碍物对模型的影响;在实际环

境中,笔者考察算法在3个Khepera IV机器人中的性能。

在介绍上述性能之前,首先简述算法中使用到的DQN算法的训练过程。

3.1 传感器智能选择网络训练

本节笔者结合实际情况对传感器智能选择网络的训练过程进行了调整。在实际网络训练中笔者注意到,若以 1×1 为最小单位对输入的 60×60 的地图进行区域类型划分,输入一共有 $3^{60 \times 60}$ 种可能性(每个划分区域为空旷、障碍物、未探索3种类型之一),由于状态空间过于庞大,模型的收敛变得非常困难。考虑到在实际情况中的区域类型在局部范围内有一定的聚集性,笔者在训练时采取了更大尺度的划分,将周围环境按 10×20 的粒度划分,即把 60×60 的地图划分为18个区域,每个区域内均设置为相同的类型,以此为网络训练的输入数据。训练完成后,部分传感器智能选择网络的输出结果如图5所示,可见机器人会对周围环境和传感器的能量消耗进行权衡,做出合理的决策。图5中,已探测区域用黑色表示,未探测区域用灰色表示,传感器感知到的区域用白色表示。

3.2 仿真环境设置

在本节中,笔者将本文提出的算法分别部署到含3个、4个和8个仿真机器人的场景中,从而考察机器人数量对性能的影响。具体而言,将每个机器人放置在区域靠近边缘的位置,每个机器人的目标点是以区域中心为对称中心的相对位置,同时在

整个区域中放置若干个能影响机器人正常行动轨迹的静止障碍物。所有机器人在整个行动过程不发生碰撞的前提下,进一步完成对传感器开销的限制。Webots仿真平台中机器人集群以及障碍物的初始状态如图6所示。如图6(a)所示,在3个机器人的实验场景中,初始状态时分别将机器人放置在左上方、右上方和左下方,4个静止的障碍物放置在整体区域的中间位置,图7(a)为该场景下3个机器人集群的运动轨迹。类似地,图6(b)、图6(c)和图7(b)、图7(c)分别为4个和8个机器人集群的初始状态和整个行动的运动轨迹。从运动轨迹中可以看出,笔者提出的算法在不同场景下均能成功地使机器人集群完成协同避障任务。图7中机器人从初始位置到目标位置的轨迹颜色由浅到深。

为了直观地反映机器人集群的传感器能量消耗,笔者将机器人从初始状态向目标状态转移的过程中打开传感器的总次数记为RSO(robot sensors overhead),并用集群中所有机器人的平均RSO来表示整体的传感器开销。

3.3 仿真实验结果

在仿真实验中,笔者使用3种传感器选择方法进行对比实验,分别是:持续采样法(实验1)、混合高斯模型选择方法^[26](实验2)以及DQN架构的脑注意力启发的传感器智能选择(DQN智能选择)方法(实验3)。其中,持续采样法以固定的时间间隔持续使用传感器采样来建模,其目的是以最稳妥的DQN智能选择方式完成基本任务,并以此



图5 部分传感器智能选择网络的输出结果



图6 Webots 仿真平台中机器人集群以及障碍物的初始状态

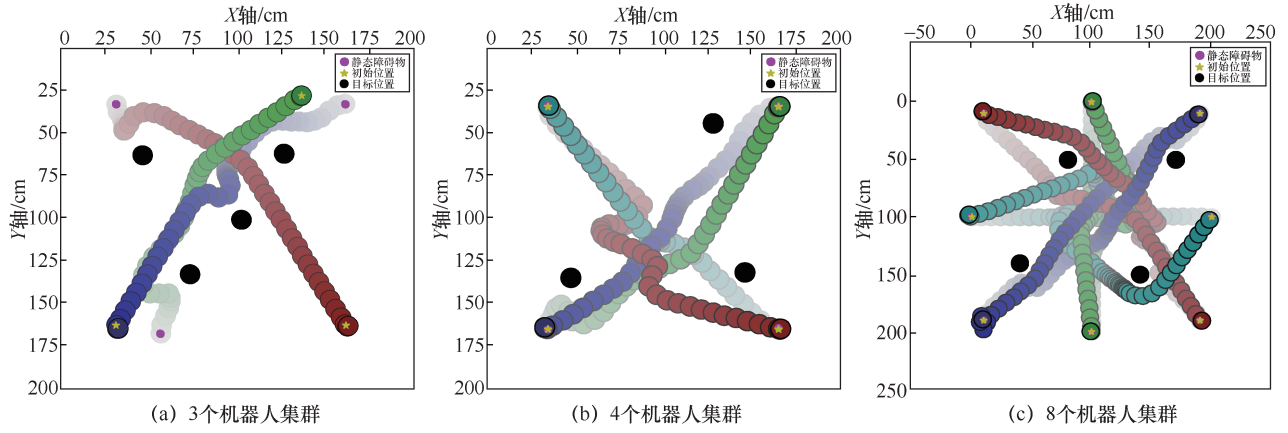


图 7 机器人集群运动轨迹

为基线分析其他方法的性能；混合高斯模型选择方法是一种多传感器采样环境信息和建模的方法，通过预测获得信息量最大的位置采样和建模，完成对环境不同信息的采集。与此同时，笔者分别在实验 2、3 的基础上，设置机器人集群可以在一定距离范围内传递各自检测到的障碍物的信息以及历史传感器获取的信息，作为实验 4、5。不同仿真场景下的机器人集群平均 RSO 及标准差如图 8 所示。机器人数量不同时各实验中的总 RSO 以及相对于基线 RSO 下降的百分比见表 1。接下来以图 8、表 1 中的 3 个机器人实验为例，详细分析实验结果。

从实验 1、2、3 可以看出，在不发生碰撞的前提下，应用本文设计的 DQN 智能选择方法约束传感器资源的使用时，3 个机器人集群的总 RSO 与不进行约束的持续采样法相比降低了 37.1%。这是因为笔者在设计 DQN 智能选择模块时，对历史传感器数据建模，并将其作为后续约束传感器使用的依据，使得对已采样的环境减少传感器的使用，从而降低 RSO。此外，与混合高斯模型选择方法相比，使用 DQN 智能选择方法的集群 RSO 降低了 17.0%。实验过程表明，机器人为了达到在运动过程中不发生碰撞这一前提，需要调整混合高斯模型中可获得最大信息量的阈值，这意味着机器人在运动过程中

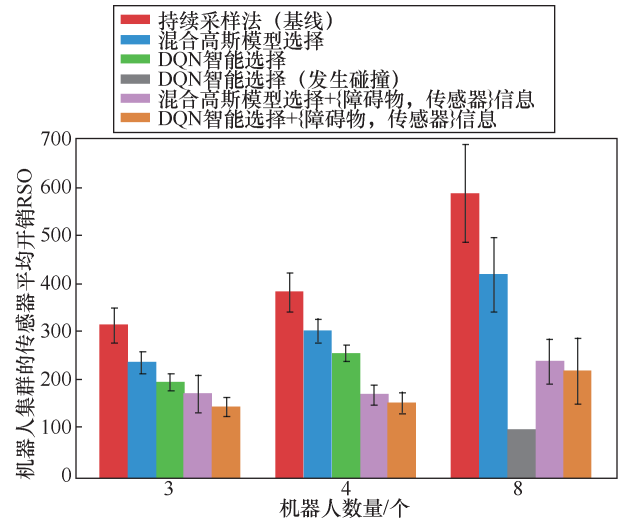


图 8 不同仿真场景下的机器人集群平均 RSO 及标准差

无法“贪婪”地做到每次采样的环境信息量都是最大的，导致采样覆盖的环境概率密度相较于 DQN 智能选择模块更高，即对已知的环境信息继续使用传感器采样，从而使传感器资源的开销增大。同时，笔者在实验过程中发现，机器人之间传递信息量的丰富程度也将影响集群任务的 RSO。因此，笔者设计实验 2、4 以及实验 3、5 对比使用相同方法而传递不同信息的结果。从表 1 中的实验数据可以看出，以一定的通信资源为

表 1 机器人数量不同时各实验中的总 RSO 以及相对于基线 RSO 下降的百分比

方法	3 个机器人	4 个机器人	8 个机器人
持续采样法 (基线)	315.0	383.75	586.86
混合高斯模型选择	238.67 / 24.2%	303.75 / 20.8%	419.75 / 28.5%
DQN 智能选择	198.0 / 37.1%	257.25 / 33.0%	—
混合高斯模型选择 + {障碍物, 传感器}信息 (实验 4)	174.0 / 44.8%	172.0 / 55.2%	241.0 / 58.9%
DQN 智能选择 + {障碍物, 传感器}信息 (实验 5)	147.0 / 53.3%	155.5 / 59.5%	220.86 / 62.4%

代价传递更丰富的信息,可以让机器人集群的 RSO 大幅降低。因此,即使传感器的感知范围有限,但在有限距离内传递的额外信息使得机器人在不使用传感器感知的情况下对自身周围的环境有了更多的认识,这对于降低机器人整体 RSO 具有很大的作用。

与 3 个机器人集群的实验结果类似,4 个、8 个机器人协同避障(表格纵向对比)的结果同样表明,在同等信息量的条件下约束传感器的使用,与使用持续采样法和混合高斯模型选择方法相比,使用本文提出的 DQN 智能选择方法效果更明显。此外,随着实验机器人数量的增加(表格横向对比),机器人之间基本不合作的方法(如实验 2、3)对 RSO 的影响并不大,仅仅依靠算法自身的优劣来约束传感器;而对于机器人之间有合作的方法(如实验 4、5),机器人数量的增加使得降低集群整体 RSO 的力度更大,这是由多机器人系统的分布式特性决定的。当并行执行环境采样任务的机器人数量很多时,即使在机器人和障碍物密集区域也不易增加传感器资源的开销,这也同时说明了多机协同在本实验场景中起到了很大的作用。

特别地,从表 1 中 8 个机器人集群的实验可以看出,即使多次重复实验 3,实验过程中机器人依旧不可避免地会发生大量的碰撞,这是由于传感器在机器人的行动过程中存在一定的误差以及“视野盲区”,会出现对障碍物“漏检”的情况。而在更加密集的环境中,这种“漏检”带来的问题是致命的,会直接导致机器人之间或者机器人与障碍物发生碰撞,使得任务失败。而如果机器人之间可以传递信息,多个机器人之间的信息互补使得个别机器人即使发生“漏检”的情况,也可以从其他机器人得到信息,这让整个多机器人系统发生碰撞的概率大大降低,因此可以提高实验的成功率。

3.4 实际环境设置及结果

本节将上述模型算法部署到 3 个 Khepera IV 机器人集群中,以验证本文提出的方法在实际场景中的有效性。与仿真环境设置类似,将每个机器人放置在区域的边缘,目标同样为以区域中心为对称中心的相对位置。机器人间可以通过 Wi-Fi 进行一定距离内的间接通信。受限于场景环境,在信息交互后,两个机器人之间的距离若低于一定阈值,则在各模块计算时将彼此的信息纳入考虑;否则,在计

算时直接忽略彼此交互到的信息。

由于受到机器人计算资源、能耗等因素的限制,本文使用“集中式训练,分布式执行”的方式部署模型。具体来说,即应用上述设计的模型在仿真平台中使用集中式的方式训练机器人,而后将经过良好训练的神经网络分布式地加载到每个真实的机器人上。

Khepera IV 机器人在没有静止障碍物下的协同避障运动过程及运动轨迹如图 9 所示。如图 9(a)所示,初始状态中机器人分别在左上方、左下方和右下方,图 9(b)至图 9(d)依次为机器人集群在不同时刻的运动状态,图 9(e)为该过程的运动轨迹。类似地,3 个 Khepera IV 机器人在有一个静止障碍物下的协同避障运动过程及运动轨迹如图 10 所示。从结果中可见,本文提出的方法在实际环境中也具有良好的性能。

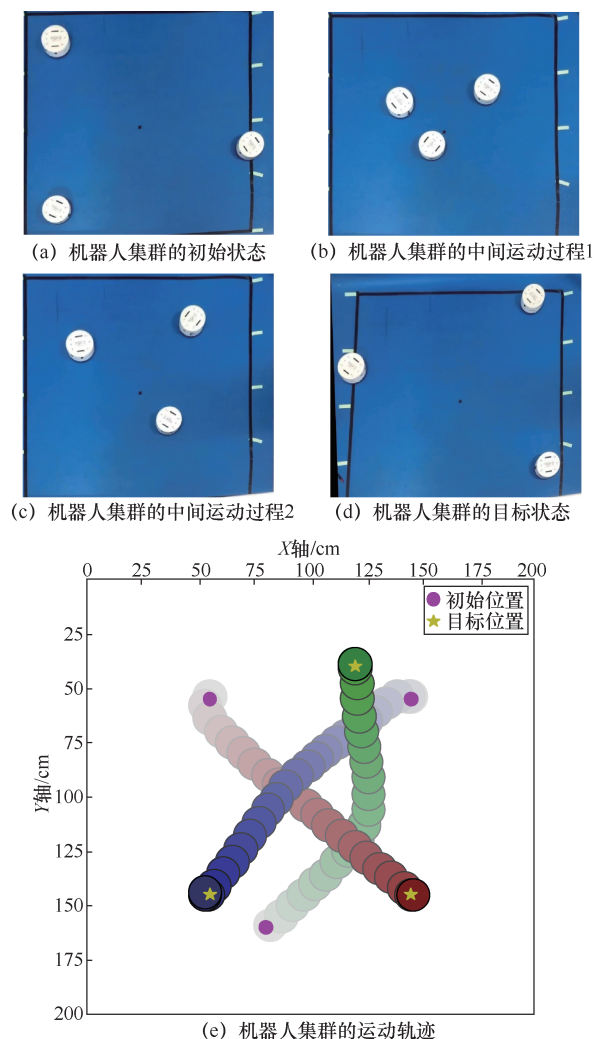


图 9 3 个 Khepera IV 机器人在没有静止障碍物下的协同避障运动过程及运动轨迹

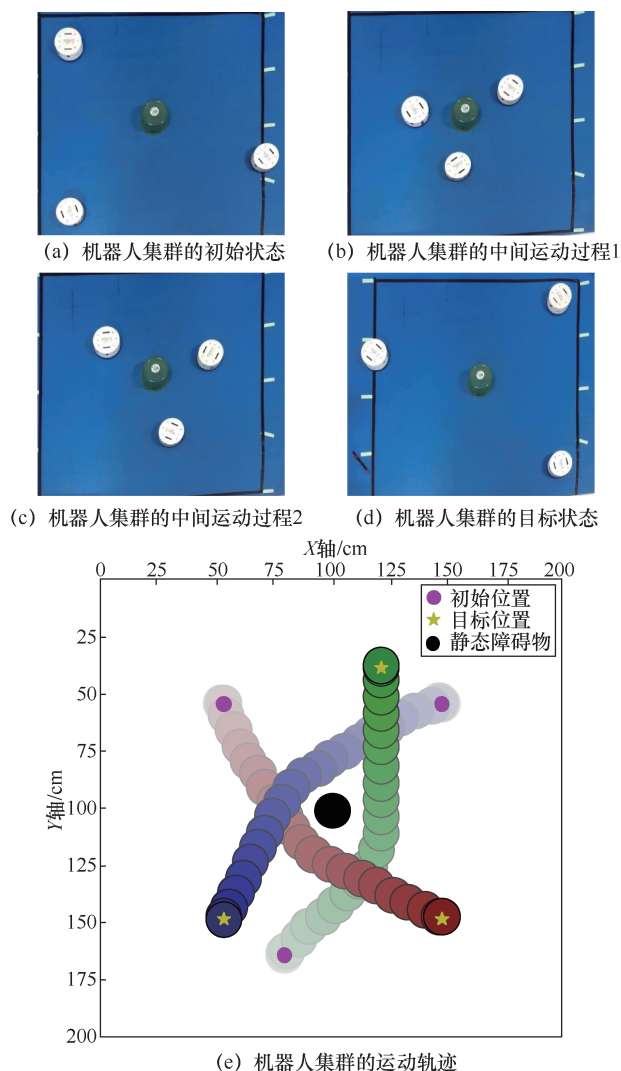


图10 3个Khepera IV机器人在有一个静止障碍物下的协同避障运动过程及运动轨迹

4 结束语

如何设计任务驱动的计算和通信资源高效利用机理是群智系统中亟待解决的关键科学问题。本文借鉴脑注意力机制,成功设计了一种通过群体智能系统合理分配有限资源的方法,可以在执行正常的路径规划与避障任务的同时,将更多资源集中在关键任务,为解决以上科学问题提供了一种新的思路。特别地,本文提出了基于DQN架构的传感器智能选择模块,使得传感器工作状态依照环境所需进行“注意力”动态调整,降低传感器获取信息的冗余度,降低传感器能量资源的消耗。同时,本文利用ORCA算法,将机器人之间近距离通信获得的信息与传感器选择结果结合,对机器人的运动速度进行动态规划,并在双轮式差速驱动的运动模型中实现,完成

整体的路径规划与避障。最终,本文通过仿真以及实际环境验证了上述方法的有效性。

本文虽然假定群智系统智能体间通信距离受限,但仍然假定了理想的通信速率和时延。未来,笔者会在更复杂、更密集、更异构的环境下开展相关研究,并验证其方法的有效性。

参考文献:

- [1] 沈博闻,于宁波,刘景泰. 仓储物流机器人集群的智能调度和路径规划[J]. 智能系统学报, 2014, 9(6): 659-664.
SHEN B W, YU N B, LIU J T. Intelligent scheduling and path planning of warehouse mobile robots[J]. CAAI Transactions on Intelligent Systems, 2014, 9(6): 659-664.
- [2] 张钟元. 室外物流配送机器人的多传感器融合同步定位与建图技术[D]. 上海: 上海交通大学, 2019.
ZHANG Z Y. Multi-sensors fusion for simultaneous localization and mapping of outdoor delivery robots[D]. Shanghai: Shanghai Jiao Tong University, 2019.
- [3] 陶永,高赫,王田苗,等. 移动工业机器人在飞机装配生产线中的应用研究[J]. 航空制造技术, 2021, 64(5): 32-41, 67.
TAO Y, GAO H, WANG T M, et al. Application of mobile industrial robot in aircraft assembly production line[J]. Aeronautical Manufacturing Technology, 2021, 64(5): 32-41, 67.
- [4] 顾宝兴,姬长英,王海青,等. 智能移动水果采摘机器人设计与试验[J]. 农业机械学报, 2012, 43(6): 153-160.
GU B X, JI C Y, WANG H Q, et al. Design and experiment of intelligent mobile fruit picking robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(6): 153-160.
- [5] 王艳新,唐文秀,吴函,等. 基于多传感器融合技术的智能抢险救灾机器人设计[J]. 林业机械与木工设备, 2018, 46(5): 17-20.
WANG Y X, TANG W X, WU H, et al. Design of intelligent emergency rescue robots based on multi-sensor fusion technology[J]. Forestry Machinery & Woodworking Equipment, 2018, 46(5): 17-20.
- [6] 王红君,徐军,赵辉,等. 基于平滑蚁群算法的机器人路径规划[J]. 燕山大学学报, 2017, 41(3): 278-282.
WANG H J, XU J, ZHAO H, et al. Mobile robot path planning based on smoothing ant colony algorithm[J]. Journal of Yanshan University, 2017, 41(3): 278-282.
- [7] 武彬,吴耿锋,马飞,等. 基于免疫算法的移动机器人路径规划系统[J]. 计算机工程, 2004, 30(12): 122-123, 146.
WU B, WU G F, MA F, et al. An immunity algorithm-based motive robot path-planning system[J]. Computer Engineering, 2004, 30(12): 122-123, 146.
- [8] VADAKKEPAT P, TAN K C, WANG M L. Evolutionary artificial potential fields and their application in real time robot path planning[C]//Proceedings of 2000 Congress on Evolutionary Computation. Piscataway: IEEE Press, 2000: 256-263.
- [9] 问泽藤,温淑慧,张迪. 未知环境下移动机器人自主避障算法的研究[J]. 燕山大学学报, 2021, 45(3): 274-282.
WEN Z T, WEN S H, ZHANG D. Research on autonomous obstacle avoidance algorithm for mobile robots in unknown environment[J]. Journal of Yanshan University, 2021, 45(3): 274-282.
- [10] LI Y M, KOŠECKA J. Learning view and target invariant visual servoing for navigation[C]//Proceedings of 2020 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2020: 658-664.

- [11] BERG J V D, GUY S JVD, LIN M, et al. Reciprocal n-body collision avoidance[M]//Springer tracts in advanced robotics. Heidelberg: Springer Berlin Heidelberg, 2011: 3-19.
- [12] LI H, WENG B W, GUPTA A, et al. Reciprocal collision avoidance for general nonlinear agents using reinforcement learning[J]. arXiv preprint, 2019, arXiv:1910.10887.
- [13] 朱杰斌, 秦世引. 无人机编队飞行的分布式控制策略与控制器设计[J]. 智能系统学报, 2010, 5(5): 392-399.
- ZHU J B, QIN S Y. Distributed control strategy and controller design for UAV formation flight[J]. CAAI Transactions on Intelligent Systems, 2010, 5(5): 392-399.
- [14] 刘华罡, 方浩, 毛昱天, 等. 多智能体系统分布式群集运动与避障控制[C]//第二十九届中国控制会议论文集. 皮斯卡特维: 电子与电气工程师协会出版公司, 2010: 4611-4616.
- LIU H G, FANG H, MAO Y T, et al. Distributed flocking control and obstacle avoidance for multi-agent systems[C]//Proceedings of the 29th Chinese Control Conference. Piscataway: IEEE Press, 2010: 4536-4541.
- [15] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint, 2014, arXiv:1409.0473.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the Advances in neural information processing systems. [S.l.:s.n.], 2017: 5998-6008.
- [17] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[J]. IEEE Transactions on Neural Networks, 1998, 9(5): 1054.
- [18] 刘朝阳, 穆朝絮, 孙长银. 深度强化学习算法与应用研究现状综述[J]. 智能科学与技术学报, 2020, 2(4): 314-326.
- LIU Z Y, MU C X, SUN C Y. An overview on algorithms and applications of deep reinforcement learning[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 314-326.
- [19] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [20] XIN J, ZHAO H, LIU D, et al. Application of deep reinforcement learning in mobile robot path planning[C]//Proceedings of 2017 Chinese Automation Congress. Piscataway: IEEE Press, 2017: 7112-7116.
- [21] YUE P Y, XIN J, ZHAO H, et al. Experimental research on deep reinforcement learning in autonomous navigation of mobile robot[C]//Proceedings of 2019 14th IEEE Conference on Industrial Electronics and Applications. Piscataway: IEEE Press, 2019: 1612-1616.
- [22] 罗鹏, 黄珍, 秦易晋, 等. 基于 DQN 的车辆驾驶行为决策方法[J]. 交通信息与安全, 2020, 38(5): 67-77, 112.
- LUO P, HUANG Z, QIN Y J, et al. A method of vehicle driving behavior decision based on DQN algorithm[J]. Journal of Transport Information and Safety, 2020, 38(5): 67-77, 112.
- [23] FIORINI P, SHILLER Z. Motion planning in dynamic environments using velocity obstacles[J]. The International Journal of Robotics Research, 1998, 17(7): 760-772.
- [24] SUNEHAG P, LEVER G, GRUSLYS A, et al. Value-decomposition networks for cooperative multi-agent learning[J]. arXiv preprint, 2017, arXiv:1706.05296.
- [25] RASHID T, SAMVELYAN M, SCHROEDER C, et al. Qmix: monotonic value function factorisation for deep multi-agent reinforcement

learning[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018.

- [26] LUO W H, SYCARA K. Adaptive sampling and online learning in multi-robot sensor coverage with mixture of Gaussian processes[C]//Proceedings of 2018 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2018: 6359-6364.

[作者简介]



项羽铭(1999—), 男, 浙江大学信息与电子工程学院在读, 主要研究方向为群体智能、强化学习。



陈焜(1997—), 男, 浙江大学信息与电子工程学院硕士生, 主要研究方向为群体智能、强化学习。



赵志峰(1975—), 男, 之江实验室总工程师, 主要研究方向为认知无线电、无线网格网络和软件定义网络在无线通信中的应用。



李荣鹏(1989—), 男, 博士, 浙江大学信息与电子工程学院副教授, 主要研究方向为智能通信网络、网络智能、网络切片。



张宏纲(1967—), 男, 博士, 浙江大学电子与信息工程学院教授, 主要研究方向为认知无线电、绿色通信和下一代异构蜂窝网络架构。