

基于改进MADDPG算法的人群疏散仿真研究

杨玉*,王翔,蒋效伟,宋强,陈若彤,戴红伟

(江苏海洋大学 计算机工程学院,江苏 连云港 222005)

摘要:随着社会和经济的快速发展,城市和公共区域的人口密度也逐渐增加,一定程度上增加了紧急情况下人群疏散的难度。针对高密度人群疏散问题提出了一种基于改进多智能体深度确定性策略梯度算法(Improved Multi-Agent Deep Deterministic Policy Gradient, IMADDPG)的人群疏散方法。该方法将注意力机制引入到深度强化学习框架和改进的社会力模型(Social Force Model, SFM)中,通过双层控制策略,宏观层面采用改进的多智能体深度确定性策略梯度算法确定领导者的疏散路径,微观层面采用改进SFM制定避障策略。在多障碍区域与高密度环境中的实验表明,与传统方法相比,该方法的疏散效率平均提高了约16%,疏散时间缩短了约17%。本研究将改进深度确定性策略梯度和改进SFM相结合,为大型空间的人员疏散提供了可靠的解决方案。

关键词:人群疏散;注意力机制;深度强化学习;社会力模型;双层控制

中图分类号:TP391

文献标志码:A

文章编号:0253-2395(2025)01-0130-14

Research on Crowd Evacuation Simulation Based on Improved MADDPG

YANG Yu*, WANG Xiang, JIANG Xiaowei, SONG Qiang, CHEN Ruotong, DAI Hongwei

(School of Computer Engineering, Jiangsu Ocean University, Lianyungang 222005, China)

Abstract: With the rapid development of society and economy, the population density in cities and public areas is gradually increasing, which increases the difficulty of evacuating people in emergency situations. A crowd evacuation method based on the Improved Multi Agent Deep Deterministic Policy Gradient (IMADDPG) algorithm is proposed to solve the problem of high-density crowd evacuation. This method introduces attention mechanism into the deep reinforcement learning-framework and improved Social Force Model (SFM) using a dual layer control strategy. At the macro level, an improved MADDPG is used to determine the leader's evacuation path, while at the micro level, an improved SFM is used to develop obstacle avoidance strategies. The experiments in multi-obstacle areas and high-density environments show that, compared to traditional methods, this method improves evacuation efficiency by an average of 16% and reduces evacuation time by approximately 17%. This study combines the improvement of deep deterministic strategy gradient with the improvement of SFM, providing a reliable solution for personnel evacuation in large spaces.

Key words: crowd evacuation; attention mechanism; deep reinforcement learning; social force model; dual layer control

0 引言

随着社会和经济的快速发展,城市和公共区域的人口密度越来越大,特别是在大型商

场、体育场馆、交通枢纽等相对封闭场所,人口密集现象给城市管理和公共安全带来了一定的挑战。公共场所因为人流量大、空间结构复杂等因素,一旦发生火灾、地震或恐怖袭击等紧

收稿日期:2024-06-30;接受日期:2024-10-24

基金项目:全国高等院校计算机基础教育研究会教学研究项目(2023-AFCEC-307);江苏省计算机学会教学类项目(JSCS2022028);江苏省大学生创新创业项目(202311641003Z)

* 通信作者:杨玉(1979-),女,江苏扬州人,博士,副教授,主要研究方向为智能计算、智慧教育。E-mail:yangyu@jou.edu.cn

引文格式:杨玉,王翔,蒋效伟,等.基于改进MADDPG算法的人群疏散仿真研究[J].山西大学学报(自然科学版),2025,48(1):130-143. DOI:10.13451/j.sxu.ns.2024147.

急情况,由于疏散不及时可能导致拥挤、跌倒甚至踩踏等严重事故^[1]。因此,在突发事故发生时,如何有效对有限空间内的人群进行引导和疏散,对于降低生命和财产的损失,都有其重要的意义。

为了更高效地模拟人群疏散,一部分研究者倾向于通过改进微观模型来实现。例如,Cui等^[2]融合遗传算法^[3]与楼层场元胞自动机模型,改进了行人疏散动力学^[4]研究中的算法,提高了模拟行人特性的准确性,从而优化了疏散效率。Johansson等^[5]基于真实视频数据对社会力模型(Social Force Model, SFM)进行了优化,提出了“局部导航场”概念,使得模型能够更好地模拟行人在拥挤环境中的动态避障行为。李昌华等^[6]通过增加行人运动方向选择、考虑情绪因素如恐慌以及调整吸引力强度对SFM进行了改进,使得模型能够更真实地模拟行人疏散过程中的自组织行为。然而,由于缺乏全局路径规划的指导,这些模型在大规模人群疏散时无法有效引导人群正确选择出口。因而,研究者们将路径规划算法引入到人群疏散模拟中。如Hart等^[7]结合Dijkstra算法^[8]的准确性和启发式搜索的速度提出的A*算法,优先考虑从起点到终点的预计总距离,但仅适用于静态环境。

此外,强化学习(Reinforcement Learning, RL)因其对多变环境的较强适应性,已被广泛应用于解决各种路径规划问题。例如,Li^[9]提出一种多目标强化学习方法,有效增加疏散流量并优化路径带宽,证实其有效性。Yi等^[10]通过将改进后的深度Q网络(Deep Q-network, DQN)算法应用于智能体路径规划,有效增强了智能体在复杂环境中的自主导航以及避障能力。而在紧急疏散过程中,由于需要考虑个体行为、环境变化和人群密度等因素,现有的RL算法在处理大规模人群疏散时存在一定的局限性。这种局限性体现在传统的RL算法在处理人群疏散时,往往因状态空间庞大导致学习效率低下,甚至无法有效学习。在这种情况下,智能体可能会陷入无法有效区分重要和不重要信息的问题,从而无法做出最优的决策。

鉴于此,本文将宏观方法与微观方法进行结合,并在宏观层面采用深度强化学习(Deep

Reinforcement Learning, DRL)算法进行路径规划,在微观层面采用改进的SFM进行避障。本文通过在DRL框架中引入注意力机制^[11],提升智能体识别关键状态信息的能力,从而提高学习和决策的效率。注意力机制的引入使得智能体能够动态地对输入特征进行加权,突出对当前任务更为重要的特征。

本文的主要创新点包括:1)通过合理的人群分组方法和领导者选择机制,将人群划分为若干个分组,并为每个小组选择一个领导者;2)利用引入注意力机制后的多智能体深度确定性策略梯(Multi-Agent Deep Deterministic Policy Gradient, MADDPG)算法对这些领导者智能体进行路径规划;3)结合改进后的SFM进行微观避障,进行人群疏散引导。

1 相关工作

1.1 人群疏散方法

为了能够在紧急情况下进行有效地疏散,研究者们提出了大量的宏观微观结合的人群疏散方法。此类方法采取双层控制,宏观层面采用路径规划算法找出最佳疏散路径,微观层面改进社会力模型以便应用于具体的人群疏散场景。

1.1.1 宏观方法

宏观方法不关注群体中的个体行为,而是将人群作为一个整体进行研究,关注群体的整体流动特性。流体动力学模型^[12]是一种典型的宏观方法。它常常采用流体动力学的原理来模拟人群运动,类似于水流或空气流动的方式,可以有效处理大规模人群的运动和行为。

Farooq等^[13]在高密度人群运动估计中采用了流体力学模型,这一方法从整体层面对行人的运动进行了全面的估算。然而,宏观模型常把人群统一视为同质体,未能反映行人间的个体差异和相互作用,导致仿真与实际情况的偏差较大。

1.1.2 微观方法

与宏观方法相反,微观方法擅长表示疏散过程中的个体行为特征。典型微观方法有Kirchner等^[14]提出的一种元胞自动机人群疏散模拟的代表性方法,该方法以其规则的简洁性和较低的计算需求而受到青睐,特别适合群体

的模拟场景。然而,其简化的规则使得基于该方法的模拟在捕捉个体行为方面表现较差,展示了在描述个体行人特性时的明显局限。Helbing等^[15]提出的SFM是一个用于模拟行人行为和人群动态的微观仿真模型,它被广泛应用于人群管理、建筑设计、紧急疏散计划和虚拟环境设计等领域。Liu等^[16]提出了一种双层控制的人群疏散方法,该方法微观层面采用SFM进行避障,通过将疏散人群划分为两类:领导者和跟随者,并对这两类人群应用不同的模型来实现更有效的疏散过程。Meng等^[17]在他们的研究中对传统的社会力模型进行了扩展,研究在视野受阻条件下的行人疏散问题。研究结果显示,能见度距离在一定范围内越大,疏散过程越快,且不同的能见度条件下,人群密度的波动模式也有所不同。

1.2 深度强化学习

DRL由于其对模型结构和超参数较少敏感已经引起了人工智能领域学者的广泛关注。已有诸多学者将多智体系统与DRL结合即多智能体深度强化学习(Multi-Agent Deep Reinforcement Learning, MADRL)进行深入研究。深度强化学习框架主要由以下五个要素组成:智能体、状态、奖励、动作和交互环境,结构如图1所示。

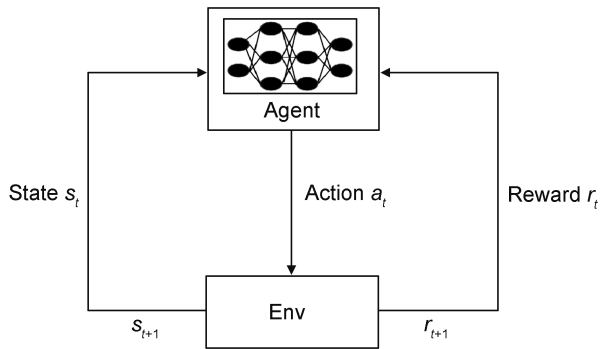


图1 深度强化学习框架

Fig. 1 The framework of DRL

MADRL可以看作是马尔可夫决策过程的一种扩展和具体实现方式,它通过结合深度学习(Deep Learning, DL)和强化学习(Reinforcement Learning, RL)的优势,为解决复杂环境中的决策问题提供了新的途径。马尔可夫决策过程由元组 (S, A, P, r, γ) 构成,其中 S 是状态的集合, A 是动作的集合, γ 是折扣因子,它的取值范围是 $[0,$

$1)$, $r(s, a)$ 为奖励函数, $P(s'|s, a)$ 是状态转移函数,其中 s 为当前状态, s' 为下一状态, a 为采取的动作。优化的目标就是策略函数 $\pi(s, a) = P(A_t = a | S_t = s)$ 通过优化策略 π ,使得智能体在 t 时刻状态 S_t 开始,直到终止状态时,所有奖励的衰减之和的公式(1)最大化,从而使 G_t 有上界,即表示未来奖励对 G_t 的影响将逐步减小。

$$G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \quad (1)$$

Q函数的定义如公式(2)所示:

$$Q^\pi(s, a) = E\pi[G_t | S_t = s, A_t = a] \quad (2)$$

$Q^\pi(s, a)$ 表示在马尔可夫决策过程(Markov Decision Process, MDP)过程中期望回报。而强化学习的目标就是找到最优策略 π^* 使公式(2)的期望最大化,即找到最优策略使公式(3)为真。

$$Q^{\pi^*}(s, a) \geq Q^\pi(s, a) \quad \forall s, a \in S, A. \quad (3)$$

$Q^{\pi^*}(s, a)$ 就是最优动作价值函数,如公式(4)所示:

$$Q^{\pi^*}(s, a) = \max_{\pi} Q^\pi(s, a), \quad \forall s, a \in S, A. \quad (4)$$

1.3 DDPG算法

深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)^[18]主要用于解决连续动作空间的问题。该算法结合了DQN和确定性策略梯度(Deterministic Policy Gradient, DPG)的思想,并引入了经验回放和目标网络来提高稳定性和收敛性。算法采用如图2所示的Actor-Critic框架,网络主要是由策略网络、目标策略网络、价值网络、目标价值网络组成。为了增加环境探索能力,算法在执行动作时加入随机噪声 N_t ,并利用经验回放缓冲区进行离线网络训练,如公式(5)所示:

$$a_t = \mu_{\theta_t}(o_t) + N_t \quad (5)$$

DDPG算法会根据公式(5)选择动作 a_t ,然后执行动作 a_t ,得到奖励 r_t 和状态 s_{t+1} ;接着将 (s_t, a_t, r_t, s_{t+1}) 存入经验回放缓冲区 D 。从 D 中采样 N 个多维数组 (s_t, a_t, r_t, s_{t+1}) ,之后更新critic网络,用公式(6)计算采样得到的策略梯度,然后更新actor网络:

$$\nabla_{\theta} J(\theta) = E_{s, a \sim D} [\nabla_{\theta} \pi(s) \cdot \nabla_a Q^\pi(s, a)] \quad (6)$$

最后通过公式(7)和公式(8)更新target网络参数:

$$\theta' \leftarrow \tau \theta + (1 - \tau) \theta', \quad (7)$$

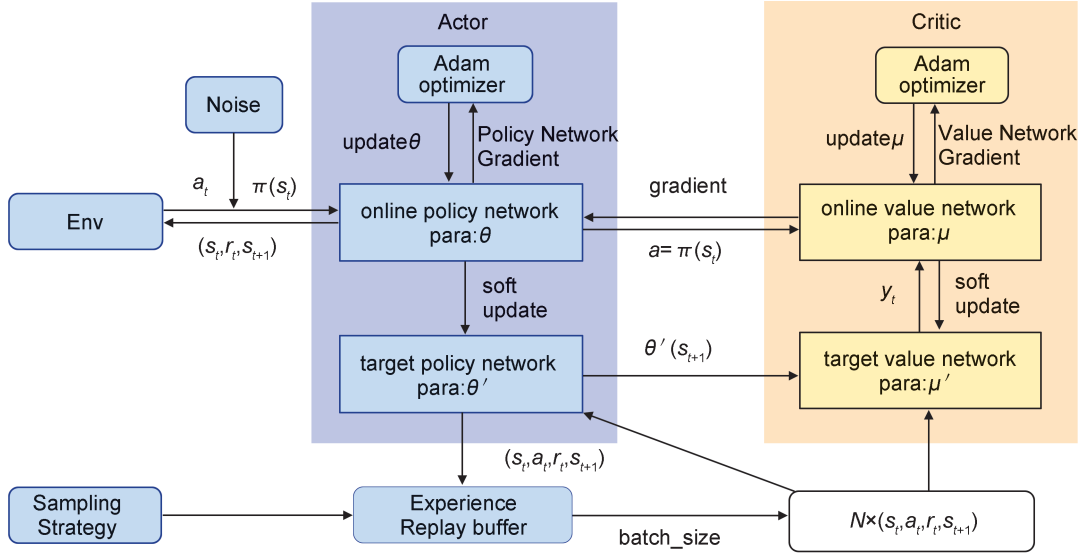


图2 深度确定性策略梯度算法(DDPG)框架

Fig. 2 The framework of deep deterministic policy gradient (DDPG) algorithm

$$\mu' \leftarrow \tau\mu + (1 - \tau)\mu' \quad (8)$$

使用策略梯度方法进行优化,如公式(9)所示:

$$\nabla_{\theta} J(\theta_i) =$$

$$E_{s \sim p^{\theta}, a \sim \pi_i} [\nabla_{\pi_i} \log \pi_i(a_i, o_i) Q_i^{\pi}(x, a_1, \dots, a_N)] \quad (9)$$

其中 $Q_i^{\pi}(x, a_1, \dots, a_N)$ 是中心化的动作价值函数, $x = (o_1, \dots, o_N)$ 包含了所有智能体的观测。

对于确定性策略来说,假设有 N 个连续的策略 μ_{π_i} ,可以得到 DDPG 的梯度公式,如公式(10)所示:

$$\nabla_{\pi_i} J(\mu_i) =$$

$$E_{x \sim D} [\nabla_{\pi_i} \mu_i(o_i) \nabla_{a_i} Q_i^{\mu}(x, a_1, \dots, a_N |_{a_i = \mu_i(o_i)})] \quad (10)$$

1.4 多智能体深度确定性策略梯度算法

多智能体深度确定性策略梯度算法^[19] (Multi-agent deep deterministic policy gradient, MADDPG) 如图 3 所示,每个智能体都独立地应用 DDPG 算法,同时共享一个集中式的 Critic 网络。Critic 网络在训练过程中对各个智能体的 Actor 网络进行评价。在执行阶段,每个智能体的 Actor 网络则自主进行行动,保持动作的去中心化。在智能体博弈中,通过在策略中引入随机性,可以计算每个智能体的期望收益,并

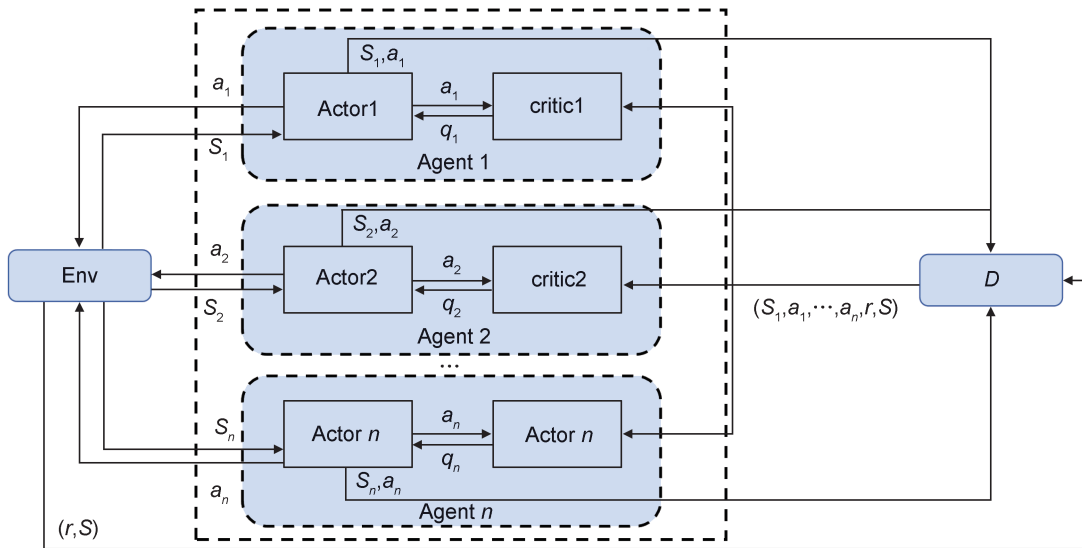


图3 多智能体深度确定性策略梯度算法(MADDPG)框架

Fig. 3 The framework of multi-agent deep deterministic policy gradient (MADDPG)

其中 x 代表所有智能体的观测, μ_i 代表所有智能体的联合动作, 而 Q_i^μ 是在多智能体背景下, 对于智能体 i 而言, 给定状态 x 和所有的智能体动作 (a_1, \dots, a_N) 时的动作价值函数。在多智能体确定性策略梯度的框架下, 每个智能体利用其他智能体的策略来增强自己的策略梯度估计。

智能体的损失函数 $L(\omega_i)$ 通过最小化预测 Q_i^μ 的值和目标 Q_i^* 值之间的差来定义, 如公式 (11) 所示:

$$L(\omega_i) = E_{x,a,r,x'} [Q_i^\mu(x, a_1, \dots, a_N) - y]^2, \quad (11)$$

$$y = r_i + \gamma Q_i^*(x', a'_1, \dots, a'_N) | a'_j = \mu'_j(o_j). \quad (12)$$

其中 y 是目标价值函数, 计算方式如公式 (12) 所示, r_i 是智能体 i 接收到的即时奖励, γ 是用于计算未来奖励现值的折扣因子。 μ' 表示目标策略, 它是基于过去策略 μ 的某种延伸或变种, 用于稳定学习过程。

策略网络的更新是通过梯度上升法来实现的, 目的是最大化每个智能体的期望回报。通过持续地优化这些梯度, 智能体的策略将逐渐趋向于最优。在 MADRL 的环境中, 该方法允许智能体在相互作用和可能存在的非静态环境特征中学习到的复杂的行为。

2 基于改进 MADDPG 算法的人群疏散仿真方法

宏观方法能够有效地规划大规模人群的整体疏散路径, 微观方法可以模拟个体行人的行为动态, 深度强化学习方法如 DDPG 算法则具备自主学习和优化疏散策略的能力。然而, 这些方法也存在一定的不足之处。宏观方法通常忽略个体行为细节, 微观方法在大规模人群场景中计算复杂度高, 而现有的 DDPG 算法在多智能体协同疏散时效果不够理想, 难以应对复杂动态环境下的实时决策需求。

为了解决上述问题, 本文提出了一种基于改进 MADDPG 算法 (Improved Multi Agent Deep Deterministic Policy Gradient, IMADDPG) 的人群疏散仿真方法, 该方法综合了宏观和微观方法的优势, 并在多智能体深度强化学习框架下进行了优化, 如图 4 所示。框架的第一步是初始化疏散环境和人群的规模。然后, 采用

模糊 C 均值聚类的方法对人群进行分组。IMADDPG 算法负责计算领导者的疏散路径, 由领导者引导各自分组内的行人按着规划后的疏散路径运动, 行人采用社会力模型绕过障碍物, 并防范行人之间的碰撞。最终, 该方法顺利完成疏散。

2.1 基于 IMADDPG 算法的路径规划方法

在 MADDPG 算法中, 使用传统的 ϵ -greedy 策略^[20]可能会遇到一系列挑战。首先, ϵ -greedy 中的 ϵ (探索率) 值是固定的, 这意味着在整个训练过程中, 智能体进行随机探索的概率始终保持不变。这种固定探索率可能导致在训练初期探索过多无效或低效的动作, 而在训练后期又因为探索不足而陷入局部最优解。其次, 固定的 ϵ 值无法根据训练的状态和进度动态调整, 缺乏对环境变化的适应性。在某些情况下, 智能体可能需要更多的探索来发现新的策略, 而在其他情况下则应该更多地利用已知信息。最后, 由于 MADDPG 涉及多个智能体之间的交互和相互影响, 平衡探索和利用变得更加复杂和难以掌握。固定的 ϵ 值可能无法在不同的训练阶段和不同的智能体之间找到最佳平衡。为克服上述缺点, 本文根据进度对 ϵ 进行动态调整, 如公式 (13) 所示, 在训练初期保持较高的探索率以发现更多可能性, 而在训练后期逐渐降低探索率以更多地利用已知信息。

$$\epsilon = \max(\epsilon_{\min} - \delta). \quad (13)$$

此外, 为了进一步提高 MADDPG 算法的性能, 任安妮等^[21]通过在策略层引入注意力机制的方式, 使得神经网络能够快速学习有效的策略。本文考虑在策略网络、价值网络、目标策略网络和目标价值网络上引入注意力机制, 如图 5 所示。IMADDPG 动态地对输入特征进行加权, 突出那些对当前任务更为重要的特征。

注意力机制的结构如图 6 所示。

每一个网络的输入特征 x_0 在进入网络前, 首先通过一个全连接层被映射到注意力空间, 目的是将原始特征转换到一个适合注意力加权的, 如公式 (14) 所示:

$$a = \tanh(\omega_{at}x_0 + b_{at}). \quad (14)$$

其中 ω_{at} 和 b_{at} 分别代表全连接层的权重与偏置, a 是映射后的特征。

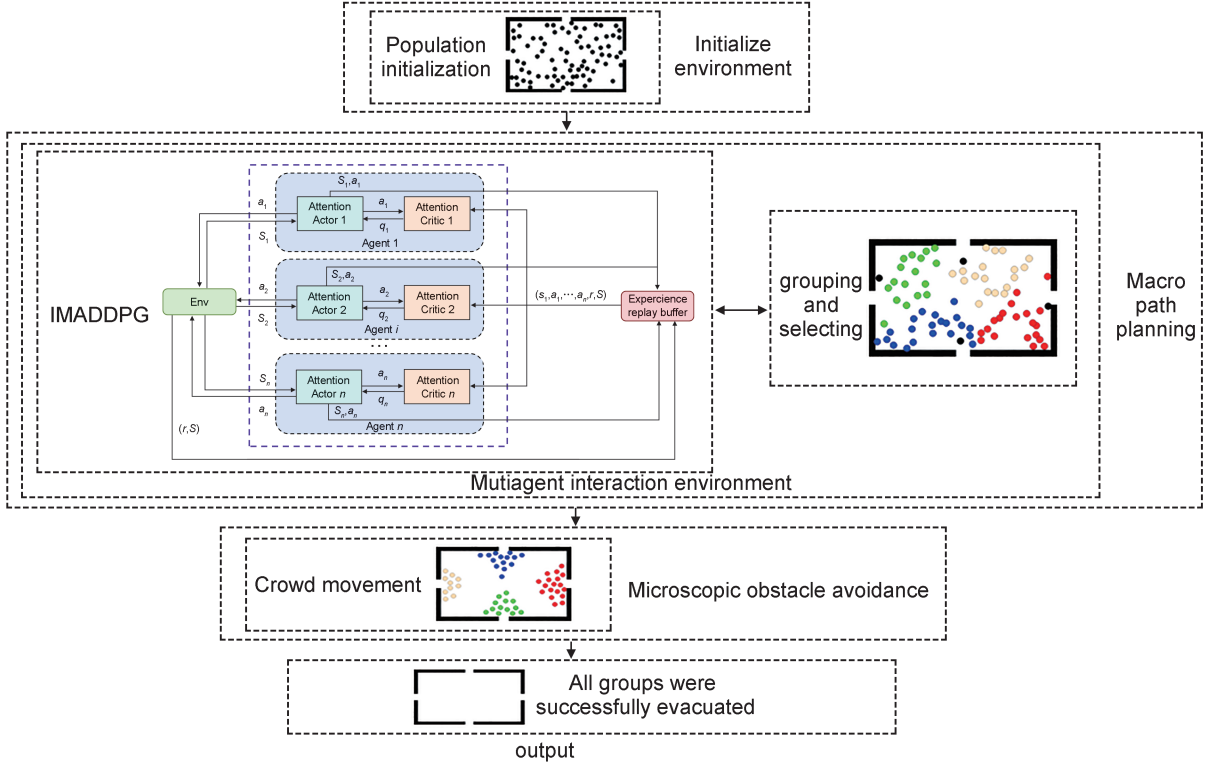


图4 基于改进多智能体深度确定性策略梯度算法(IMADDPG)的人群疏散方法框架

Fig. 4 Framework of crowd evacuation method based on the improved multi agent deep deterministic policy gradient (IMADDPG) algorithm

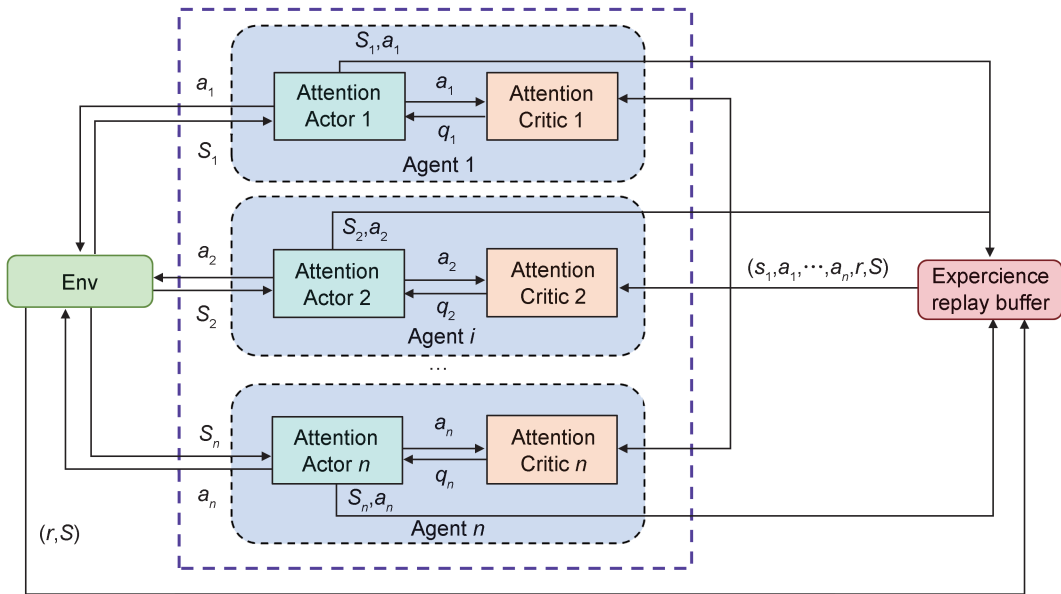


图5 改进多智能体深度确定性策略梯度算法(IMADDPG)框架

Fig. 5 The framework of improved multi agent deep deterministic policy gradient (IMADDPG)

接着,另一个全连接层用于计算每个映射后特征的重要分数 s_0 ,这一步没有使用偏置,强调了权重向量 $V_{ctx_0}^T$ 在映射重要性上的作用,如公式(15)所示:

$$s_0 = V_{ctx_0}^T a_0 \quad (15)$$

然后,使用激活函数对这些重要特征进行归一化,得到特征的注意力权重,如公式(16)所示:

$$\omega = \text{softmax}(s_0) \quad (16)$$

最后,这些注意力权重被用来加权原始输入特征 x_0 ,生成加权特征,如公式(17)所示:

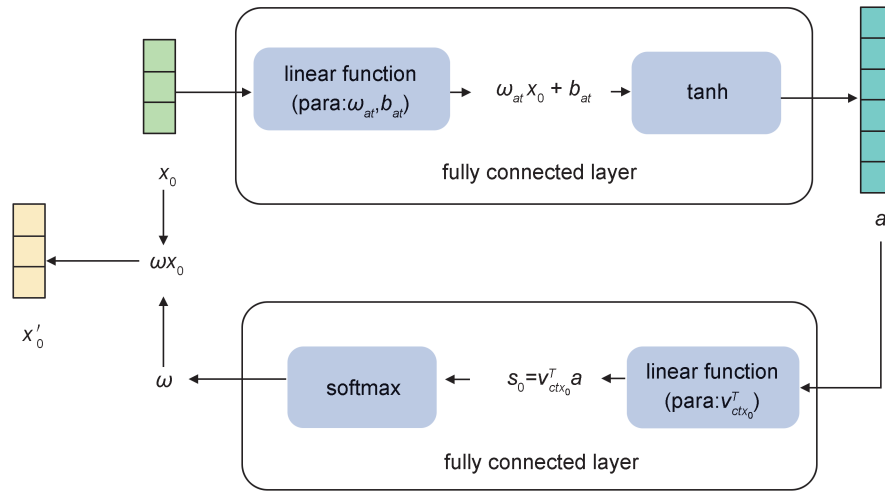


图6 引入的注意力机制的结构

Fig. 6 Structure of the introduced attention mechanism

$$x'_0 = \omega x_0 \quad (17)$$

引入注意力机制可以帮助智能体更好地处理多智能体交互问题,并增强智能体对环境的感知能力。通过这种方式,IMADDPG算法可以提高智能体的协作能力、决策准确性和学习效率,从而实现更高的整体性能。

IMADDPG算法的伪代码如算法1所示。

算法1 IMADDPG算法

- 1 for $episode = 1$ to M do
- 2 初始化一个随机噪声 N , 用于动作探索
- 3 初始化状态 x
- 4 for $t = 1$ to T do
- 5 对于每个智能体 i , 在当前策略和探索下根据公式(5)选择动作 a_i
- 6 执行动作 $a_i = (a_1, \dots, a_n)$ 并得到奖励 r 和新的状态 x' ;
- 7 将 (x, a, r, x') 存放至经验回放池 D
- 8 更新状态 $x \leftarrow x'$
- 9 for 智能体 $i = 1$ to N do
- 10 从 D 中进行随机采样 (x, a, r, x')
- 11 将价值网络的输入特征经过公式(14)—公式(17)后得到新特征
- 12 新输入特征再经过公式(7)和公式(8)中心化训练价值网络
- 13 将策略网络的输入特征经过公式(14)—公式(17)后得到新特征
- 14 新输入特征再经过公式(6)分别训练智能体的策略网络

15 根据公式(10)和公式(11)计算梯度
16 通过公式(7)和公式(8)梯度更新目标网络的参数

17 end for

18 end for

19 end for。

本文采用大O法分析算法的时间复杂度,假设最大迭代次数为 M , 最大时间步长为 T , 智能体数量为 N , 经验回放池中样本数量为 S , 那么 IMADDPG 算法的时间复杂度为 $O(M \times T \times N \times S)$ 。

2.2 人群分组

在紧急疏散过程中,行人的行动显著受到个体间社会关系及其与周围环境交互的影响,导致人群倾向于向空间上接近或存在社会联系的其他个体移动。这一行为模式促使疏散人群在复杂环境中自然而然地形成稳定的小组,并跟随那些因决策能力、经验知识或地理位置优势而自然成为领导者的个体。这些小组一旦形成,其内部结构和成员行为倾向于维持相对稳定状态。本小节的关键在于通过对社会力模型的改进,并评估不同分组策略与领导者选择机制对于提高疏散效率的具体作用。

模糊C均值聚类算法^[22],通过为每个个体赋予隶属于每个群体的概率,而非确定的归属,可以有效弥补多种群文化算法的这些缺点。这种模糊隶属度的引入,允许个体在不同群体之间拥有连续的过渡,更好地反映了人群

之间的自然分布和相似性,尤其适合处理重叠或模糊的数据集。在人群疏散仿真中,这意味着模糊C均值可以更准确地模拟人群的自然聚集行为。此外,模糊C均值聚类的灵活性和自适应性使其在动态变化的环境中表现更佳,能够适应人群疏散过程中的实时变化,如疏散路径的变更或阻塞。总的来说,模糊C均值聚类在人群分组方面提供了一种更灵活、更贴近实际情况的方法,有效地弥补了多种群文化算法在处理复杂、重叠人群数据时的不足。

假设人群样本为 $X=(x_1, x_2, \dots, x_n)$,其中 n 为参与实验的总人数。算法根据每个行人的位置来进行聚类分组,行人位置可用 $X_k=(x_{kx}, x_{ky})$ 表示, x_{kx} 和 x_{ky} 分别对应行人在空间上的横坐标和纵坐标。

本文中采用的是模糊C均值聚类算法(Fuzzy-c Means, FCM)。FCM算法的代价函数见公式(18)。

$$\min J(U, V) = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - v_j\|^2 \quad (18)$$

其中 u_{ij} 表示第 i 个样本与第 j 个聚类中心之间的关系程度, m 为模糊度参数, v_j 表示第 j 个聚类中心的位置。在FCM算法中,每个样本可以被分配到多个聚类中心里,因此关系程度 u_{ij} 需要满足公式(19)的条件:

$$\sum_{j=1}^C u_{ij} = 1, i = 1, 2, \dots, N. \quad (19)$$

同时聚类中心需要满足公式(20)的条件:

$$\sum_{i=1}^N u_{ij}^m > 0, j = 1, 2, \dots, C. \quad (20)$$

2.3 领导者智能体

在进行分组后本文将进行领导者智能体的选择,基于适应度函数 $F(x_{ij})$,如公式(21)所示:

$$F(x_{ij}) = \omega_1 \times \max \{u(x_{ij})\} + \omega_2 \times \min \{d(x_{ij})\}, \quad (21)$$

ω_1 和 ω_2 代表的是权重, x_{ij} 代表的是个体 j 在分组 i , $u(x_{ij})$ 代表的是个体距离当前分类中心的距离, $d(x_{ij})$ 代表的是个体到出口的距离。在同一分组中,适应度值最高的被选为领导者。选择领导者可以在一定程度上减少拥堵。

在复杂的大空间环境下进行疏散时,人群的运动受到行人间及行人与障碍区域间相互作

用的影响,导致人们倾向于跟随与自己关系亲近的其他人一同疏散。本研究基于实际场景设定了领导者的各项参数。

每位领导者智能体的位置都会安全出口附近。

疏散开始时,领导者智能体将会引领队伍并位于整个队列的最前方。

领导者的速度如公式(22)所示:

$$\vec{v}_i = v_i^0(t) \vec{b}_i^0(t), \quad (22)$$

$\vec{b}_i^0(t)$ 表示下一个运动方向的向量。对于跟随者的方向设定,他的运动方向不再直接指向出口,而是指向领导者智能体,而当领导者到达出口时。就会在出口附近位置等待组内的人,直到所有人到达出口附近。

行人 j 的运动方向向量计算方式如公式(23)所示:

$$\vec{e}_j^0(t) = ((\vec{p}_i - \vec{p}_j) / d_{ij}) \quad (23)$$

其中 $\vec{e}_j^0(t)$ 为行人 j 的运动方向向量,它的计算方式为行人与领导者的位置向量的差除以行人与领导者的距离, \vec{p}_i 为领导者 i 的位置, \vec{p}_j 为行人 j 的位置, d_{ij} 表示行人 j 与领导者 i 之间的距离。

通过这种方法,模型能够更加真实地模拟出在实际疏散中人们的行为特征,进而优化疏散路径的规划,确保疏散过程的高效和安全。本研究不仅高效地模拟了紧急疏散中的自然人群行为,也优化了疏散过程,确保了仿真的实用性与准确性。

2.4 改进社会力模型

由社会力模型可知,个体会受到来自其周围环境(既包括其他人也包括物体)的社会力和直接作用于其上的物理力的作用。这些力量包括:(1) f_i 驱动力,即内在的动力,驱使个体向目标方向移动;(2) f_{ij} 行人间的相互作用力,用于在身体接触发生之前避免相互之间的碰撞;(3) f_{iw} 个体与障碍物之间的相互作用力,防止与障碍物的碰撞;(4) f_o 干扰力,可能来自环境的其他因素的影响。这些力量的合成作用导致了行人的加速或减速,可通过牛顿第二定律进行描述。社会力模型如公式(24)所示:

$$m_i \frac{dv_i}{dt} = f_i + \sum_{j \neq i} f_{ij} + \sum_w f_{iw} + f_o. \quad (24)$$

原始社会力模型的架构图如图7的(a)所

示,仅向出口移动。

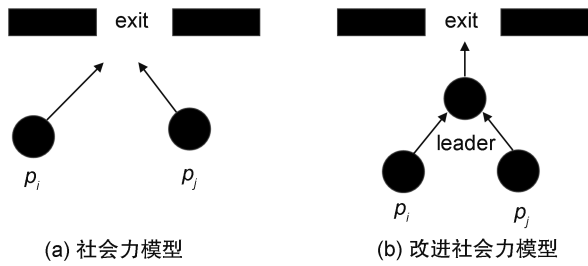


图7 社会力模型与改进社会力模型

Fig. 7 Social force model and improved social force model

经过本文改进的社会力模型,群体内的成员将在受到其他力的同时,不再向出口移动,而是向每个群体的领导者进行移动。改进社会力模型如图7中的(b)所示,公式如(25)所示:

$$m_i \frac{dv_i}{dt} = f_{il} + \sum_{j \neq i} f_{ij} + \sum_w f_{iw} + f_{io}. \quad (25)$$

本文通过公式(18)一公式(20)对人群进行初始化分组,然后选择领导者,选择领导者可以在一定程度上减少拥堵。

为验证改进社会力模型的效果,本文在18 m×18 m的二维空间内进行疏散实验。

图8所展示的本文方法与对比文献[23]的疏散效果对比。其中(a)、(b)、(c)为对比文献[23]的疏散结果,(d)、(e)、(f)为本文方法的结果。由图8可以看出,在改进社会力模型中,行人会跟随各自的领导者运动,从而缓解了堵塞。同时本文的仿真结果还表明,该方法的分组策略及其领导者的选择使得即使在出口资源分布不平衡和受到其他力的影响下仍能快速地完成疏散。

2.5 宏观路径规划的方法

Wang等^[24]所提出的基于双注意力DRL的智能资源分配算法,结合了多维卷积注意力和多头自注意力机制^[25]的优点,可以为分布式节点提供局部和全局特征融合能力,最大限度地提高数据传输的性能。Yu等^[26]设计了混合注意力模块,并通过与MADDPG算法相结合,使智能体更多地关注大量环境信息中的关键信息,提高了经验样本的利用率。借鉴上述方法,本文提出改进后的MADDPG算法进行宏观路径规划,核心架构如图9所示。

在此架构中,领导者作为智能体的实体映射,负责路径规划指导。然后,领导者智能体

将观测状态信息传送至路径规划模块。根据学习到的最佳策略,领导者智能体的神经网络计算路径信息并反馈给领导者执行。这使得领导者成功引导跟随者抵达疏散出口。方法的详细步骤如下:

步骤1:对所有领导者的位置($x_{leader 1}, \dots, x_{leader i}, \dots, x_{leader n}$)进行迭代检查,确定当前状态($o_1, \dots, o_i, \dots, o_n$),并将其传给对应的智能体。

步骤2:进行出口位置的迭代检查。

步骤3:对于每个智能体*i*,依据IMADDPG算法计算路径($P_{path 1}, \dots, P_{path i}, \dots, P_{path n}$)来决定行动 a_i ,每个代理的行动受其输入的状态 o_i 和领导者路径信息 $P_{path i}$ 影响。

步骤4:进入注意力机制(Attention)模块。每个代理的动作 a_i 和观测到的状态 o_i 被传递给对应的注意力Actor和注意力Critic,注意力Critic则给出动作价值 q_i ,注意力Actor则生成动作 a_i 传递给智能体。

步骤5:所有智能体的状态($o_1, \dots, o_i, \dots, o_n$)和动作($a_1, \dots, a_i, \dots, a_n$)被存储到经验回放缓冲池(Experience Replay Buffer Pool)中,以供后续训练更新。

步骤6:每个时间步,领导者智能体依据公式(24)选择动作,并更新自身状态,以引导其路径上的跟随者智能体进行行动。跟随者则根据公式(25)进行移动,确保其保持与领导者的适当距离。

步骤7:监控所有领导者智能体是否逐渐接近出口区域,确认它们是否达到路径终点。

步骤8:当领导者接近出口时,等待最后一个跟随者也抵达出口区域,以保证所有智能体的同步疏散。

步骤9:一旦所有智能体(包括所有领导者和跟随者)均到达出口,宣布疏散过程结束。

3 实验分析

为验证IMADDPG算法的人群疏散效果,参照对比文献[15]对疏散空间的长宽比例、出口位置及障碍区域分布进行设置。具体实验参数设置为:疏散空间大小为50 m×30 m的空间,人为半径0.2 m的圆,出口宽度为1.2 m,人的期望速度大小为1.2 m/s,出口分布在每条边

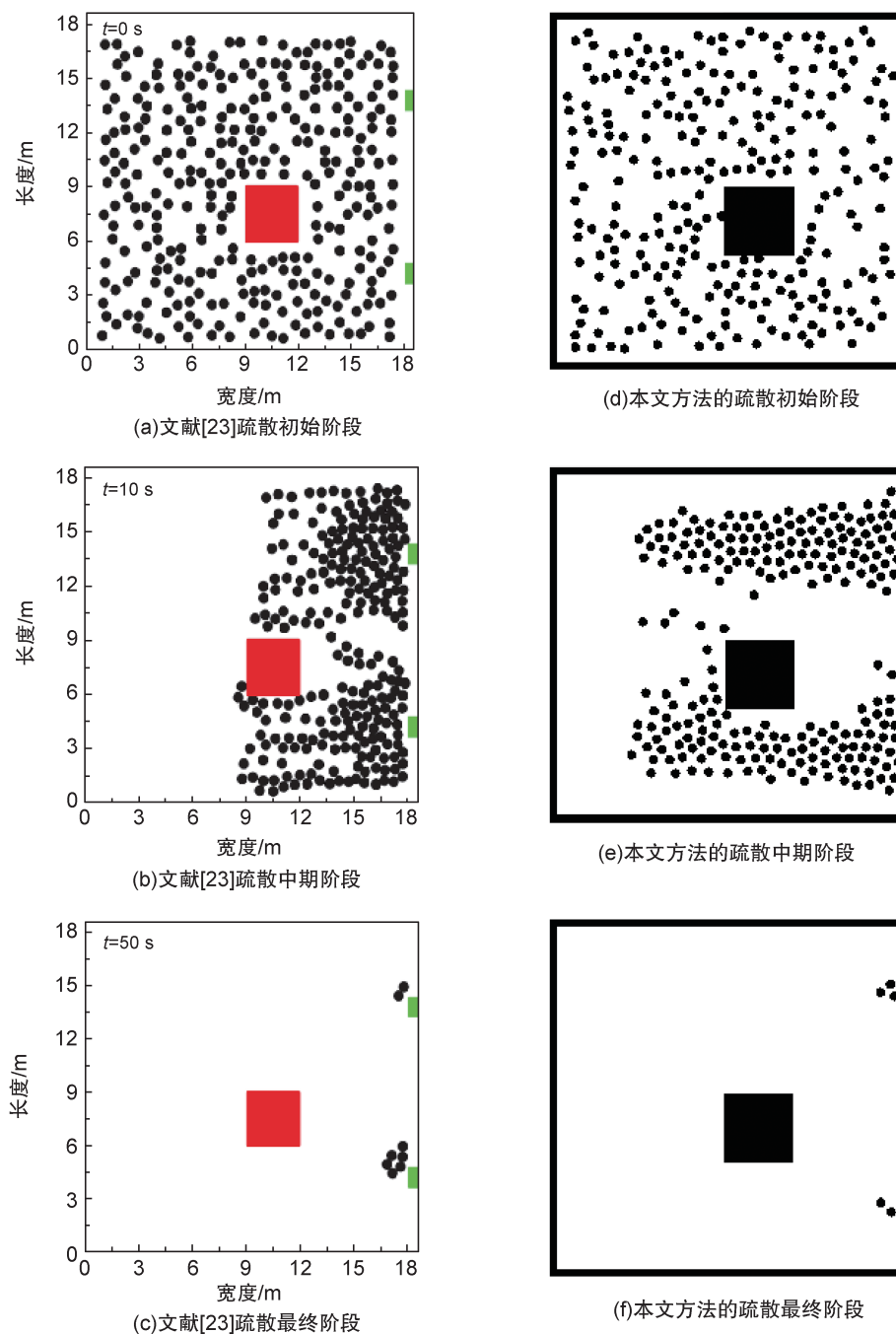


图8 本文方法与文献[23]在18 m×18 m空间的疏散效果对比

Fig. 8 Comparison of evacuation effects between the proposed method and reference [23] in an 18 m×18 m space

的中间位置。为了降低行人分布不均对实验的影响,每次疏散实验都采用了成员分布随机坐标值作为行人疏散开始时的位置。

3.1 宏观路径规划的方法

为了验证本文方法在无阻碍场景中的疏散效果,本文对多种疏散时间和乘客数量进行了比较。对比算法为原始社会力模型方法与MADDPG方法。

首先,本实验验证对比了无障碍、不同人群数量情况下的疏散人员数量效果,对比结果如表1和图10(a)所示。其次,对比了不同方法在相同疏散人员数量情况下的疏散时间效果对比,对比结果如表2和图10(b)所示。

行人疏散的过程中,随着实验人员数量的增长,碰撞的概率就会越来越大。图10展示的是在无障碍的场景中,三种不同方法的疏散

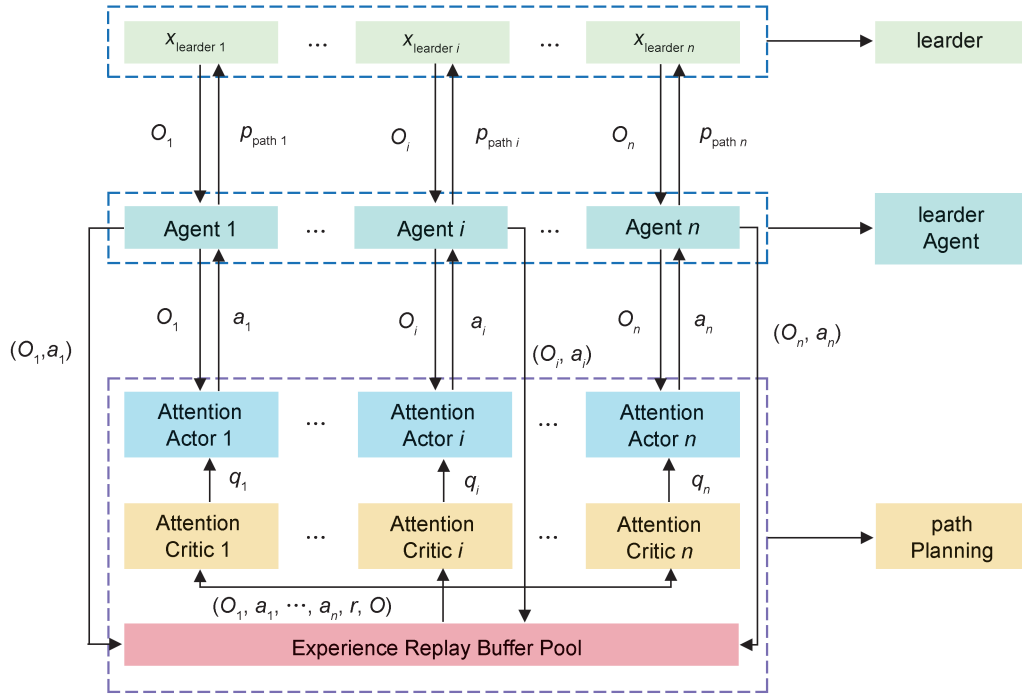


图9 宏观路径规划的结构

Fig. 9 The structure of macro path planning

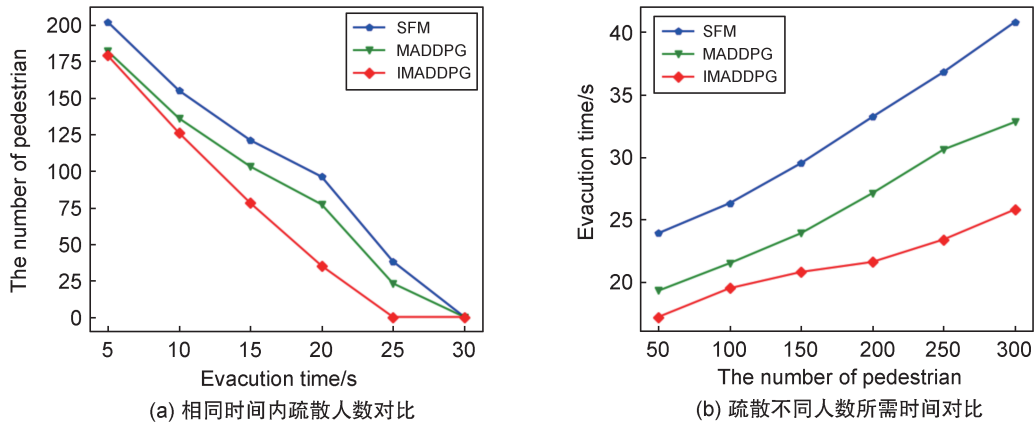


图10 无障碍场景下三种方法的人群疏散效果对比

Fig. 10 Comparison of crowd evacuation effects of three methods without obstacles

性能的比较。图 10(a)和表 1 显示的是同一疏散人员数量(人员设定为 300)下疏散的剩余人员数量随疏散时间的变化,三种方法的人数下降速度差不多。但当时间超过 20 s 后,本文方法开始显示出相对稳定的趋势。本文方法,在 25 s 之前就已经完成疏散,而其他方法则需要更长的疏散时间。从图 10(b)和表 2 可看出,当疏散人员数量的增加,三个方法所需要的时间都在同步增加。当疏散的人数超过 150 人,IMADDPG 算法通过引入注意力机制可以迅速规划路径并疏散行人,合理使用出口空间来避

免拥堵和碰撞,从而进一步提高了人群疏散的速度。

表1 疏散人员数量为 300 时三种方法的疏散人员数量效果对比(剩余人员数量)

Table 1 Comparison of the evacuation effect of three methods on the number of evacuees when the number of evacuees is 300 (number of remaining personnel)

方法	疏散时间/s					
	$t=5$	$t=10$	$t=15$	$t=20$	$t=25$	$t=30$
SFM	202	155	121	96	38	0
MADDPG	182	136	103	77	23	0
IMADDPG	179	126	78	35	0	0

表2 三种方法在不同疏散人数情况下的人群疏散时间对比

Table 2 Comparison of crowd evacuation time under different evacuation numbers using three methods

方法	疏散人数					
	50	100	150	200	250	300
SFM	23.9 s	26.3 s	29.5 s	33.2 s	36.8 s	40.8 s
MADDPG	19.3 s	21.5 s	23.9 s	27.1 s	30.6 s	32.8 s
IMADDPG	17.2 s	19.5 s	20.8 s	21.6 s	23.4 s	25.8 s

最后,本文展示了3种方法在4个出口的人员疏散的可视化。疏散的效果可视化如图11所示,MADDPG方法的疏散和IMADDPG方法的疏散可以让行人快速赶往出口,尤其IMADDPG引入注意力机制后可以更快地选择好出口。仿真实验结果表明,本文的方法能够在无障碍区域的疏散大空间场景中更加快速有序地到达安全出口。

3.2 有障碍区域下的人群疏散仿真

为了验证本研究方法在有多个障碍区域的疏散环境下的疏散性能,实验将SFM方法、本方法与MADDPG方法在一个有五处障碍区域的环境,行人只要全部到达出口,就代表疏散成功。首先,通过对比实验验证3种方法在疏散人员数量相同的情况下空间内剩余人数的变化来验证本文的方法在有障碍区域场景中的有效性,如表3、图12(a)所示。然后本实验将改变疏散人员的数量来验证本方法在有障碍区域的环境中进行人员疏散的有效性,如表4、图12(b)所示。为进一步增加实验的真实性,本文采取随机生成栅格坐标的方式初始化乘客位置。

从图12(a)以及表3中可看出,在人群疏散初期,三种疏散策略在减少疏散人数的效率上

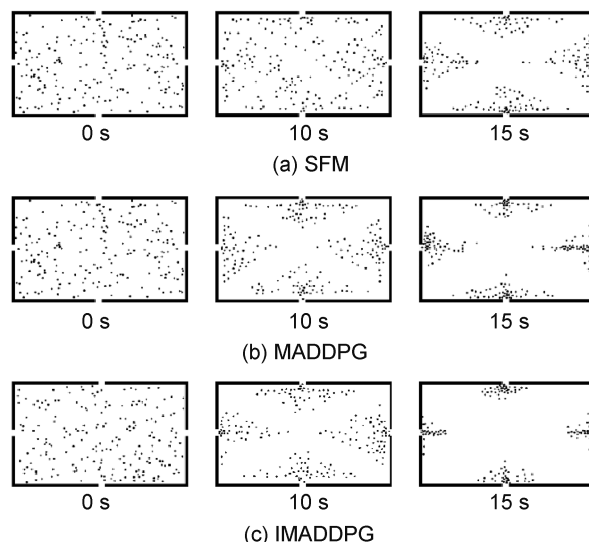


图11 四出口无障碍场景下三种方法的疏散效果对比

Fig. 11 Comparison of evacuation effects with four exits of three methods without obstacles

表3 有障碍场景疏散人数相同时疏散剩余人数统计表

Table 3 Statistical table of remaining evacuation personnel in obstacle scenarios with the same number of evacuations

方法	疏散时间					
	$t=30\text{ s}$	$t=60\text{ s}$	$t=90\text{ s}$	$t=120\text{ s}$	$t=150\text{ s}$	$t=180\text{ s}$
SFM	225	191	165	96	48	0
MADDPG	212	184	139	73	31	0
IMADDPG	205	149	106	46	0	0

并无显著差异,但本文方法相同时间段内疏散的人数都比两个方法的多。从图12(b)以及表4可以看出,在有多障碍区域的场景中,随着人员数量的增加,三种方法的疏散所需时间均呈现上升的一种趋势。每种方法在行人数量较少时都能有效找到疏散路径。然而,当人员数量超过150人时,采用集成注意力机制的确定

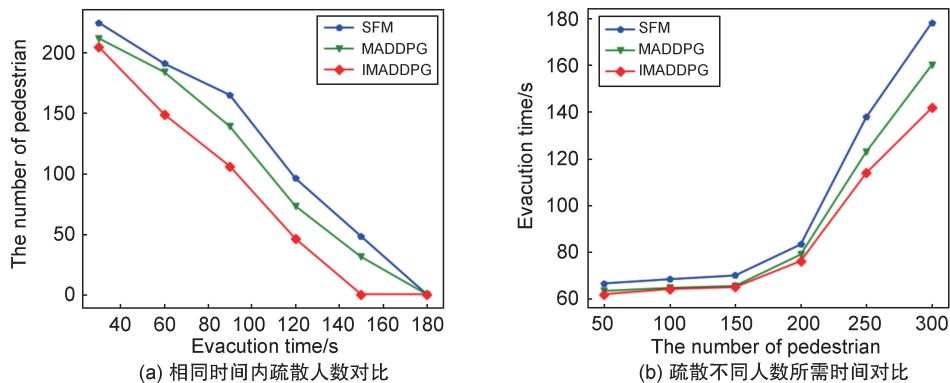


图12 有障碍时三种方法的人群疏散对比

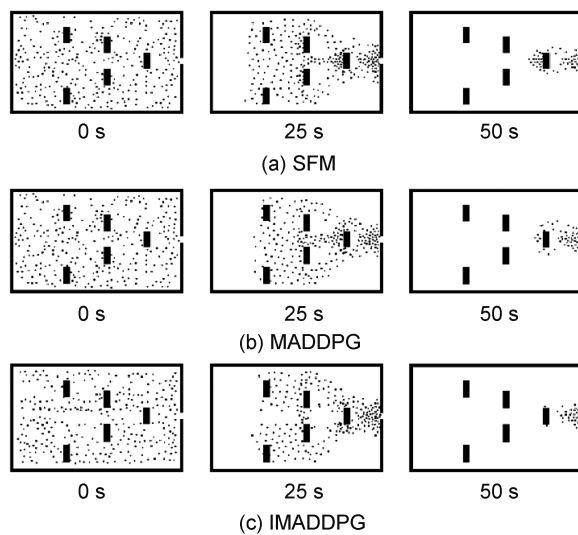
Fig. 12 Comparison of crowd evacuation effects of three methods with obstacles

表4 有障碍场景不同疏散人数时人群疏散时间对比
Table 4 Comparison of crowd evacuation time under different numbers of evacuees with obstacle

方法	疏散人数					
	50	100	150	200	250	300
SFM	66.5	68.3	69.9	83.2	137.7	178.2
MADDPG	63.3	64.6	65.4	78.9	122.9	160.1
IMADDPG	61.9	64.1	64.9	76.1	113.8	141.7

性策略梯度改进算法进行的人群疏散仿真,能够快速计算出最佳疏散路线。与此同时,其他方法因应对人群密集场景的动态不确定性和全局路径规划能力不足,表现出明显的劣势。

为进一步验证本文方法的有效性,设置五个障碍在疏散环境中。图13展示了三种方法对300名人员的疏散效果。实验结果表明,MADDPG算法与传统的SFM在寻找最优疏散路径时遇到困难,尤其是在障碍区域附近容易引发拥堵现象。本文研究提出的新方法有效缓解了拥堵现象,并提升了疏散效率,能够在有障碍物场景中更高效地完成疏散。



注:图中■表示障碍物。

图13 多障碍场景不同方法疏散效果对比

Fig. 13 Evacuation effects of different methods with multiple obstacles

4 结论

为解决大空间高密度人群疏散效率低下问题,提出了一种改进多智能体深度确定性策略梯度(IMADDPG)算法的人群疏散方法。该方法整合了注意力机制和改进的社会力模型

(SFM),实现了宏观层面的路径规划和微观层面的避障策略。实验结果表明,该方法在多障碍和高密度环境中相较于传统方法提升了疏散效率约16%(参照表1和表3),同时减少了约17%的疏散时间(参照表2和表4),验证了其高效性。未来的研究将着眼于深化MADDPG算法,引入更高效的注意力机制模型,并结合环境信息反馈进一步改进社会力模型,以期更符合应用场景并提升疏散效率。

参考文献:

- [1] CUESTA A, BALBOA A, GONZÁLEZ-VILLA J, et al. Evacuation of Vessels in Dockyards: A Model Validation Study[J]. *Fire Technol*, 2023, **59**(5): 2333-2354. DOI: 10.1007/s10694-023-01435-2.
- [2] CUI G, YANAGISAWA D, NISHINARI K. Incorporating Genetic Algorithm to Optimise Initial Condition of Pedestrian Evacuation Based on Agent Aggressiveness[J]. *Phys A Stat Mech Appl*, 2021, **583**: 126277. DOI: 10.1016/j.physa.2021.126277.
- [3] 田雅琴, 胡梦辉, 刘文涛, 等. 基于跳点搜索-遗传算法的自主移动机器人路径规划[J]. *工程设计学报*, 2023, **30**(6): 697-706. DOI: 10.3785/j.issn.1006-754X.2023.03.136. TIAN Y Q, HU M H, LIU W T, et al. Path Planning of Autonomous Mobile Robot Based on Jump Point Search-genetic Algorithm[J]. *Chin J Eng Des*, 2023, **30**(6): 697-706. DOI: 10.3785/j.issn.1006-754X.2023.03.136.
- [4] 张俊, 李红柳, 宋卫国. 基于实验的行人与疏散动力学规律研究[J]. *中国科学技术大学学报*, 2019, **49**(12): 947-956. DOI: 10.3969/j.issn.0253-2778.2019.12.001. ZHANG J, LI H L, SONG W G. Experimental Study on Pedestrian and Evacuation Dynamics[J]. *J Univ Sci Technol China*, 2019, **49**(12): 947-956. DOI: 10.3969/j.issn.0253-2778.2019.12.001.
- [5] JOHANSSON A, HELBING D, SHUKLA P K. Specification of the Social Force Pedestrian Model by Evolutionary Adjustment to Video Tracking Data[J]. *Adv Complex Syst*, 2007, **10**(supp02): 271-288. DOI: 10.1142/s0219525907001355.
- [6] 李昌华, 杨晶, 李智杰. 一种应用于人群仿真的改进社会力模型[J]. *系统仿真学报*, 2021, **33**(3): 521-528. DOI: 10.16182/j.issn1004731x.joss.19-0553. LI C H, YANG J, LI Z J. An Improved Social Force Model for Crowd Simulation[J]. *J Syst Simul*, 2021, **33**(3): 521-528. DOI: 10.16182/j.issn1004731x.joss.19-0553.
- [7] HART P E, NILSSON N J, RAPHAEL B. A Formal Basis for the Heuristic Determination of Minimum Cost

- Paths[J]. *IEEE Trans Syst Sci Cybern*, 1968, **4**(2): 100–107. DOI: 10.1109/TSSC.1968.300136.
- [8] 潘文杰, 李志强, 杨辉. QAOA 最大切割问题的类 Dijkstra 优化及实现[J]. *计算机应用研究*, 2023, **40**(2): 378–382. DOI: 10.19734/j.issn.1001-3695.2022.06.0328.
- PAN W J, LI Z Q, YANG H. QAOA Maximum Cutting Problem Analogous to Dijkstra Optimization and Implementation[J]. *Appl Res Comput*, 2023, **40**(2): 378–382. DOI: 10.19734/j.issn.1001-3695.2022.06.0328.
- [9] LI X L. An Efficient Data Evacuation Strategy Using Multi-objective Reinforcement Learning[J]. *Appl Intell*, 2022, **52**(7): 7498–7512. DOI: 10.1007/s10489-021-02640-8.
- [10] YI C, QI M. Research on Virtual Path Planning Based on Improved DQN[C]//2020 IEEE International Conference on Real-time Computing and Robotics (RCAR). New York: IEEE, 2020: 387–392. DOI: 10.1109/RCAR49640.2020.9303290.
- [11] 申翔翔, 侯新文, 尹传环. 深度强化学习中状态注意力机制的研究[J]. *智能系统学报*, 2020, **15**(2): 317–322. DOI: 10.11992/tis.201809033.
- SHEN X X, HOU X W, YIN C H. State Attention in Deep Reinforcement Learning[J]. *CAAI Trans Intell Syst*, 2020, **15**(2): 317–322. DOI: 10.11992/tis.201809033.
- [12] 苗志宏, 李智慧. 一种基于 SPH 方法的人员疏散混合模型及模拟[J]. *自动化学报*, 2014, **40**(5): 935–941.
- MIAO Z H, LI Z H. A Hybrid Evacuation Model and Simulation Based on SPH Method[J]. *Acta Autom Sin*, 2014, **40**(5): 935–941.
- [13] FAROOQ M U, SAAD M N B M, MALIK A S, *et al.* Motion Estimation of High Density Crowd Using Fluid Dynamics[J]. *Imag Sci J*, 2020, **68**(3): 141–155. DOI: 10.1080/13682199.2020.1767843.
- [14] KIRCHNER A, SCHADSCHNEIDER A. Simulation of Evacuation Processes Using a Bionics-inspired Cellular Automaton Model for Pedestrian Dynamics[J]. *Phys A Stat Mech Appl*, 2002, **312**(1/2): 260–276. DOI: 10.1016/s0378-4371(02)00857-9.
- [15] HELBING D, FARKAS I, VICSEK T. Simulating Dynamical Features of Escape Panic[J]. *Nature*, 2000, **407**(6803): 487–490. DOI: 10.1038/35035023.
- [16] LIU H, LIU B X, ZHANG H, *et al.* Crowd Evacuation Simulation Approach Based on Navigation Knowledge and Two-layer Control Mechanism[J]. *Inf Sci*, 2018, **436/437**: 247–267. DOI: 10.1016/j.ins.2018.01.023.
- [17] MENG Q, ZHOU M, LIU J L, *et al.* Pedestrian Evacuation with Herding Behavior in the View-limited Condition[J]. *IEEE Trans Comput Soc Syst*, 2019, **6**(3): 567–575. DOI: 10.1109/TCSS.2019.2915772.
- [18] DU R Z, WANG J Y, GAO Y. Computing Offloading and Resource Scheduling Based on DDPG in Ultra-dense Edge Computing Networks[J]. *J Supercomput*, 2024, **80**(8): 10275–10300. DOI: 10.1007/s11227-023-05816-w.
- [19] LOWE R, WU Y, TAMAR A, *et al.* Multi-agent Actor-critic for Mixed Cooperative-competitive Environments[J]. *Adv Neural Inf Process Syst*, 2017, **30**: 6379–6390. DOI: 10.5555/3295222.3295385.
- [20] 杨彤, 秦进. 基于平均序列表累计奖赏的自适应 ϵ -greedy 策略[J]. *计算机工程与应用*, 2021, **57**(11): 148–155. DOI: 10.3778/j.issn.1002-8331.2003-0019.
- YANG T, QIN J. Adaptive ϵ -greedy Strategy Based on Average Episodic Cumulative Reward[J]. *Comput Eng Appl*, 2021, **57**(11): 148–155. DOI: 10.3778/j.issn.1002-8331.2003-0019.
- [21] 任安妮, 周大可, 冯锦浩, 等. 基于注意力机制的深度强化学习交通信号控制[J]. *计算机应用研究*, 2023, **40**(2): 430–434. DOI: 10.19734/j.issn.1001-3695.2022.06.0334.
- REN A N, ZHOU D K, FENG J H, *et al.* Attention Mechanism Based Deep Reinforcement Learning for Traffic Signal Control[J]. *Appl Res Comput*, 2023, **40**(2): 430–434. DOI: 10.19734/j.issn.1001-3695.2022.06.0334.
- [22] EFTEKHARI S H, MEMARIANI M, MALEKI Z, *et al.* Hydraulic Flow Unit and Rock Types of the Asmari Formation, an Application of Flow Zone Index and Fuzzy C-means Clustering Methods[J]. *Sci Rep*, 2024, **14**(1): 5003. DOI: 10.1038/s41598-024-55741-y.
- [23] JIANG Y Q, CHEN B K, LI X, *et al.* Dynamic Navigation Field in the Social Force Model for Pedestrian Evacuation[J]. *Appl Math Model*, 2020, **80**: 815–826. DOI: 10.1016/j.apm.2019.10.016.
- [24] WANG Y, SHANG F J, LEI J J. Multi-granularity Fusion Resource Allocation Algorithm Based on Dual-attention Deep Reinforcement Learning and Lifelong Learning Architecture in Heterogeneous IIoT[J]. *Inf Fusion*, 2023, **99**: 101871. DOI: 10.1016/j.inffus.2023.101871.
- [25] YANG H F, YAO L J, CAI J H, *et al.* A New Interest Extraction Method Based on Multi-head Attention Mechanism for CTR Prediction[J]. *Knowl Inf Syst*, 2023, **65**(8): 3337–3352. DOI: 10.1007/s10115-023-01867-w.
- [26] YU L L, HUO S X, WANG Z J, *et al.* Hybrid Attention-oriented Experience Replay for Deep Reinforcement Learning and Its Application to a Multi-robot Cooperative Hunting Problem[J]. *Neurocomputing*, 2023, **523**: 44–57. DOI: 10.1016/j.neucom.2022.12.020.