引用:徐文博,胡春鹤. 基于融合社交模型的多智能体强化学习导航方法[J]. 聊城大学学报(自然科学版),2026,39(4):486-497.

Citation:XU Wenbo,HU Chunhe. An integrated social-model approach to multi-agent navigation[J]. Journal of Liaocheng University (Natural Science Edition),2026,39(4):486-497.

文章编号 1672-6634(2026)04-0486-12 **DOI** 10.19728/j.issn1672-6634.2025120008

基于融合社交模型的多智能体强化学习导航方法

徐文博,胡春鹤

(北京林业大学 工学院,北京 100083)

摘要 密集人群环境中的多智能体协同导航仍面临交互关系复杂、社交规范难以建模等问题。为提升智能体在复杂环境中的避碰效率与社交适应性,构建了一种基于 RVO 与 SFM 建立融合式社会性交互模型的多智能体导航方法(即 SFMRVO-DRL),通过碰撞风险调节两者权重,实现高效避障与自然运动的统一。在此基础上,设计包含图注意力机制的决策结构,用以捕捉智能体间的动态关联,并引入右手通行规则构建多目标奖励函数,以增强策略的社交合规性与稳定性。训练过程中采用 MAPPO 在集中训练、分散执行框架下实现策略优化。实验结果表明,该方法在典型圆形交互场景中,相较 GA3C-CADRL、NH-ORCA 与 HeR-DRL 在成功率、导航时间与速度等指标上均具有显著优势。研究验证了 RVO-SFM 融合交互建模、注意力图社交感知与社交规范驱动的多目标强化学习对提升复杂人群导航性能的有效性。

关键词 深度强化学习;多智能体;社会力模型;社交导航

中图分类号 TP181 **文献标志码** A **开放科学(资源服务)标识码(OSID)**

An integrated social-model approach to multi-agent navigation

XU Wenbo, HU Chunhe

(School of Engineering, Beijing Forestry University, Beijing 100083, China)

Abstract Multi-agent cooperative navigation in dense crowds remains challenging due to complex interaction patterns and the difficulty of modeling social norms. To enhance collision avoidance efficiency and social adaptability in such environments, this paper proposes SFMRVO-DRL, a multi-agent navigation framework that integrates Reciprocal Velocity Obstacles (RVO) with the Social Force Model (SFM) into a unified hybrid social interaction model. The framework adaptively balances the two components based on collision risk, enabling both efficient collision avoidance and naturalistic motion. Building on this hybrid interaction model, we design a decision architecture incorporating a graph attention mechanism to capture dynamic inter-agent relationships, and introduce a multi-objective reward function grounded in right-hand passing conventions to improve social compliance and policy stability. The proposed method employs MAPPO for policy optimization under a centralized-training and decentralized-execution paradigm. Experimental results in a canonical circular crowd-interaction scenario demonstrate that our approach significantly

收稿日期:2025-12-09

基金项目:北京林业大学科技创新计划项目(2024XY-G009)资助

通信作者:胡春鹤,男,汉族,博士,副教授,研究方向:机器人导航制导与控制,E-mail:huchunhe@bjfu.edu.cn。

outperforms GA3C-CADRL, NH-ORCA, and HeR-DRL in terms of success rate, navigation time, and speed. This study highlights the effectiveness of hybrid RVO-SFM interaction modeling, attention-based social perception, and socially normative multi-objective reinforcement learning in advancing multi-agent navigation performance in complex crowd environments.

Keywords　deep reinforcement learning; multi-agent; social force model; socially navigation

0　引言

近年来，多智能体在拥挤空间中的社交导航因其众多的潜在应用而受到广泛关注，与传统避障任务中将障碍物视为被动体不同，社交导航场景中的行人属于具有自主决策能力的主动体，其行为会对智能体的动作产生响应性变化。行人之间及人机之间的交互过程不仅受物理碰撞约束影响，还受到社会规范、舒适距离及让行习惯等因素的制约[1]。单纯基于几何约束的避障模型难以完整刻画此类交互特性，因此，对行人行为的感知对于如何制定优先考虑行人社交安全的智能体社交导航策略至关重要[2]。但预测行人行为是困难的，大多数方法在实验中学习[3]，当人群密集或环境复杂时，智能体社交导航变得更具挑战性[4]。

对于社交机器人的研究通常是将行人视为动态且无法对环境变化作出反应的障碍物，然而，密集人群中的智能体社交导航与传统对于障碍物的避障任务存在本质区别。行人与非生命障碍物（如墙壁、静态物体）具有完全不同的属性。行人不仅具有动态性，更具有社会心理属性和交互反应性[5]。若将行人仅视为通过物理方程描述的动态障碍物，往往会导致机器人产生侵入性行为。并且传统的预测和规划分离的训练方法也会导致“机器人冻结”问题[6]，即一旦环境超过某种复杂度，规划器会认为所有向前的路径都是不安全的，机器人“冻结”在原地或执行不必要的动作以避免碰撞。最近的研究提出了一系列基于深度强化学习的方法来解决耦合模型问题[7-8]。这些方法通过将交互感知机制融入奖励函数的设计，实现了高效机器人策略的训练，从而将复杂耦合模型的处理转移到训练阶段，不同方法的主要区别在于人群交互建模方式的差异。然而这些研究存在局限，一方面完全忽略了行人与行人之间的交互，另一方面仅考虑了有限数量的机器人与行人交互，这种局限性导致这些方法在密集人群或高交互场景中的表现显著降低。一些研究将深度强化学习与注意力图相结合，使用基于图的模型来有效地提取人群特征，并使用注意力机制来推断每个交互的相对重要性以提升机器人与人类之间的交互效果[9]。DS-RNN 模型[10]使用时空图来捕获机器人自身轨迹中的空间交互和时间交互，通过循环神经网络（Recurrent Neural Networks，RNN）表示空间和时间的交互，并为空间边分配注意力权重，以推断最相关的相邻机器人[11]。然而，此类方法每个学习模型的人数是固定的，并且没有考虑机器人间的相互作用。对于单个机器人在密集的人群中的导航，此类方法难以反映真实环境中多智能体之间的交互不能轻易地以当前形式进行扩展以反映真实社交交互并适应多智能体导航。

针对多机器人在复杂环境中的协同导航这一特定场景，多智能体深度强化学习领域取得了重要进展，相关研究成果已在最新综述中得到系统分析[12]。本研究将重点关注协作式解决方案。现有方法主要分为以下几类：基于值函数分解的方法在集中训练分散执行框架下展开研究，如文献[13]提出通过分解个体 Q 值来构建全局 Q 值的方法，QMIX 算法在此基础上引入单调性约束来规范 Q 值合并过程。在策略梯度算法方面，研究表明多智能体 PPO（Multi Agent Proximal Policy Optimization，MAPPO）在协作任务中同样具有良好的效果[14]。另一类研究基于协同图模型展开，文献[15]采用抽象图表示智能体节点，但使用多层感知机（Multilayer Perceptron，MLP）处理节点关系，这种方法存在较大局限性。相比之下，图神经网络能更有效地促进智能体协作，其节点可灵活表示智能体信息。文献[16]采用图卷积网络扩展了参与计算的相邻节点范围。DICG 算法[17]与之类似，但创新性地引入注意力机制动态构建图结构，其图结构使用注意力机制进行动态计算。

在机器人动态避障领域，最优互惠速度障碍法（Reciprocal Velocity Obstacles, RVO）[18]为速度障碍法（Velocity Obstacle, VO）[19]的改进型算法，通过引入互惠性原则，使得交互双方共同承担避障责任，从而产生更加自然和高效的避障轨迹。RVO 算法具有计算效率高、实时性好的特点，适合动态环境场景下的紧急避障。然而，传统 RVO 方法存在对人群社会行为模式考虑不足的缺陷，难以准确反映真实场景中复杂的人

际交互规律。社会力模型(Social Force Model, SFM)[20]作为一种经典的人群行为建模,描述人群交互的经典方法,通过引入社会力虚拟力场概念来描述行人间的相互作用。该模型将行人运动视为受到多种社会力共同作用的结果,能够较好地模拟人群的宏观流动特征和微观交互行为。但在实际应用中,传统 SFM 存在参数调节困难、计算复杂度高等问题,特别是在处理动态变化的环境时表现欠佳。

综上所述,现有机器人导航算法研究主要集中于单一环境下的避障性能优化,仅针对静态或简单动态环境构建导航模型。然而,在密集动态人群这类复杂场景中,同时考虑避障效率、运动平滑性和人机交互自然度的多目标优化研究仍显不足。传统基于 RVO 的方法虽然计算高效,但难以建模复杂的社会行为规则;SFM 虽能表征人群交互特征,却存在实时性较差的固有局限。近年来,多智能体强化学习算法在复杂决策问题中展现出独特优势,已成为解决动态环境导航问题的重要方法,但由于算法收敛性和策略可解释性等方面的不足,其在真实场景中的应用效果仍有待提升。为此,本研究首先融合 RVO 和 SFM 的建模优势,利用 RVO 处理刚性物理交互以保证安全,利用 SFM 处理柔性社会交互以提升舒适度,构建了同时考虑避障效率、轨迹平滑性和社交合规性的多目标优化框架[21];其次,设计了基于注意力图(Attention Graph)[22]机制的多智能体决策模型,通过图结构捕捉智能体间的动态交互关系,并利用图注意力网络实现对邻近实体影响的自适应加权,从而在复杂人群中实现更精准的社交感知导航;最后,引入基于右手定则的社交规范作为结构化先验知识[23],并结合多目标奖励函数对轨迹合规性、运动平滑性与导航效率进行联合优化,有效引导智能体在密集场景中表现出符合人类预期的通行行为,通过分层奖励设计和渐进式训练方法提升算法在密集动态环境中的适应性和鲁棒性。各模块的数据流与逻辑关系如图 1 所示。

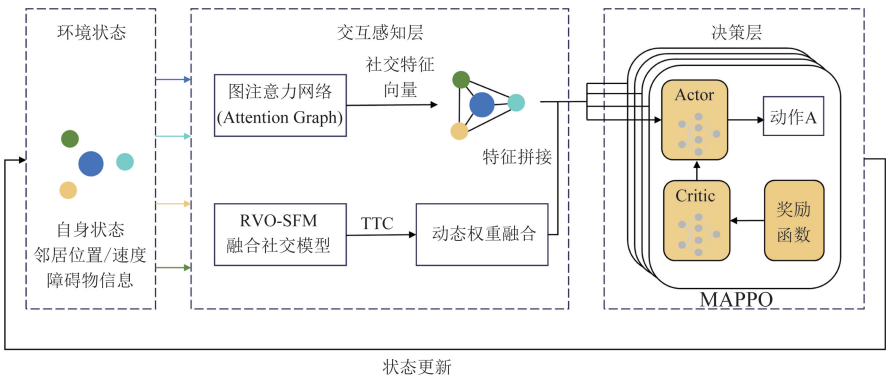


图 1 SFMRVO-DRL 方法的系统架构图

Fig. 1 The overall architecture of the SFMRVO-DRL framework

1 改进型人群交互模型

1.1 传统基于 RVO 的避碰模型

对于半径为 R_A 和 R_B 的圆形机器人 A 和机器人 B,位置和速度可以分别表示为 P_A、P_B、$\boldsymbol{v}_A$、$\boldsymbol{v}_B$。机器人 B 产生的机器人 A 的 VO 区域集合 $S_{AB}(\boldsymbol{v}_B)$ 为

$$S_{AB}(\boldsymbol{v}_B)=\{\boldsymbol{v}_A \mid \lambda(P_A,\boldsymbol{v}_A-\boldsymbol{v}_B)\cap D(P_B,R_A+R_B)\neq\varnothing\},$$

式中 $\lambda(P,\boldsymbol{v})$ 表示起点为 P 且方向为 $\boldsymbol{v}$ 的射线,$D(P,R)$ 表示一个以 B 的位置 P_B 为圆心,半径为 R_A+R_B 的扩充圆的等效构型空间障碍物。如图 2(a)所示,VO 区域可以由一个顶点(其中 $\boldsymbol{v}_P=\boldsymbol{v}_B$)和两个方向向量 $\boldsymbol{v}_l$ 、$\boldsymbol{v}_r$ 来构造。它表示在一段时间内可能引起机器人碰撞的速度集。为了实现与障碍物的碰撞避免,机器人 A 应选择 VO 区域之外的速度,即 $\boldsymbol{v}_A^{'}\notin S_{AB}(\boldsymbol{v}_B)$。同样,静态线障碍物的 VO 区域也是使用相同的定义构建的,如图 2(b)所示。此外,对于静态障碍物阻挡机器人当前位置到目标的路径的情况,需要全局导航策略来产生合理的期望速度。

由于基于 VO 的方法已被广泛用于避让动态障碍物,RVO 更适合一组机器人主动避开对方。如图 2(c)所示,RVO 从 VO 概念的几何延伸而来,$S_{AB}^{\mathrm{R}}(\boldsymbol{v}_B,\boldsymbol{v}_A)$ 代表了机器人 A 在机器人 B 以相似避障方式的速度下的禁止速度区域集合 $S_{AB}^{\mathrm{R}}(\boldsymbol{v}_B,\boldsymbol{v}_A)=\{\boldsymbol{v}_A^{'}\mid 2\boldsymbol{v}_A^{'}-\boldsymbol{v}_A\in S_{AB}(\boldsymbol{v}_B)\}$。

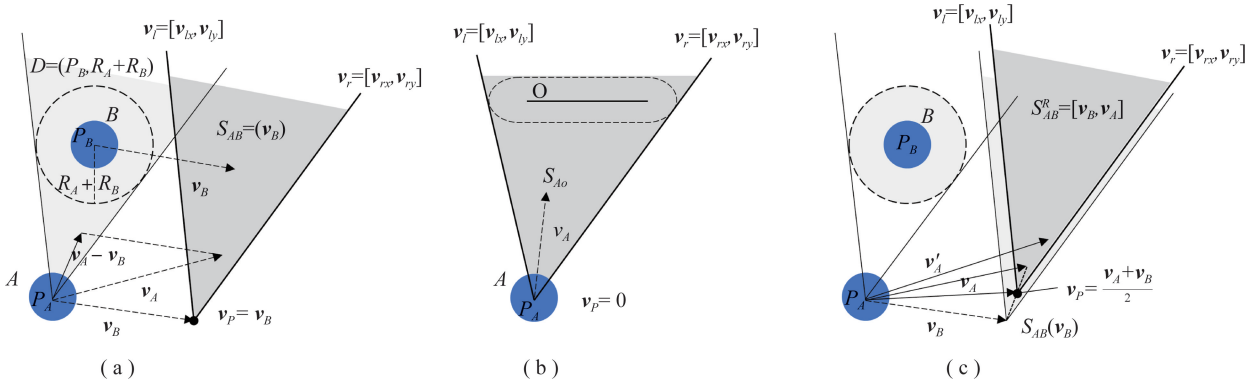


图 2　VO 与 RVO 示意图[24]

Fig. 2　The illustration of VO and RVO[24]

实际上,RVO 区域可以通过从顶点从 $\boldsymbol{v}_B$ 移动到 $(\boldsymbol{v}_A+\boldsymbol{v}_B)/2$ 的 VO 区域平移来获得。因此,VO 和 RVO 都可以用六维向量 $c=[\boldsymbol{v}_P,\boldsymbol{v}_l,\boldsymbol{v}_r]\in\mathbf{R}^6$ 表示,其中 $\boldsymbol{v}_P=[\boldsymbol{v}_x,\boldsymbol{v}_y]$ 表示顶点的坐标,$\boldsymbol{v}_l=[\boldsymbol{vl}_x,\boldsymbol{vl}_y]$ 和 $\boldsymbol{v}_r=[\boldsymbol{vr}_x,\boldsymbol{vr}_y]$ 分别描述左光线和右光线的方向。在一个机器人组中,每个机器人都会选择该 VO 和 RVO 区域之外的速度,以协同完成避碰任务[25]。

1.2　社会力模型

社会力模型通过使用力的合成来建模环境中智能体的交互关系,同时考虑了行人躲避智能体的物理反应以及心理作用。智能体的 SFM 根据牛顿第二运动定律来定义,包括目标吸引力、人-机(人)排斥力、人(机)-物排斥力,通过上述力的合成来表示将智能体与行人的行为的变化(如图 3 所示)。

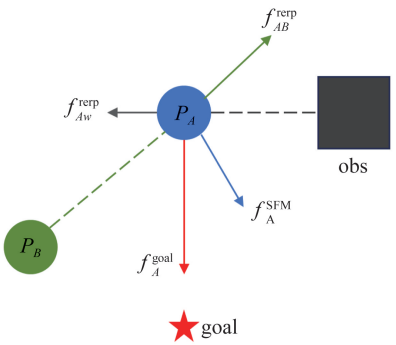


图 3　社会力模型示例

Fig. 3　Examples of social force model

(1) 目标吸引力。智能体 J 以期望的速度 $\boldsymbol{v}_0$ 向期望的方向 $\boldsymbol{e}_J$ 移动。期望速度的方向由智能体从当前位置指向下一个目标位置的矢量给出,速度选择智能体更舒适的数值。τ 为智能体调整其实际速度到期望速度所需的时间,$\boldsymbol{v}_J$ 为智能体 J 的当前速度,则将吸引智能体到达目标的力定义如

$$f_J^{\text{goal}}=\tau^{-1}(\boldsymbol{v}_0\boldsymbol{e}_J-\boldsymbol{v}_J)。\tag{1}$$

(2) 人-机(人)排斥力。智能体 J 因吸引力 f_J^{goal} 的存在,希望保持其朝向目标的期望速度 $\boldsymbol{v}_0$,但它的运动也受到周围其他智能体 K 的影响。这种影响建模为智能体 K 对智能体 J 的排斥力,定义如

$$f_{JK}^{\text{rep}}=\lambda\cdot\exp(\beta^{-1}(r_{JK}-d_{JK}))\cdot\boldsymbol{n}_{JK},\tag{2}$$

式中 λ、β 为人-机(人)间社会力强度、作用范围参数,r_{JK} 为智能体 J、K 之间的社会舒适距离,d_{JK} 为智能体 J、K 之间的欧氏距离,$\boldsymbol{n}_{JK}$ 为智能体 J、K 之间的方向向量。

(3) 人(机)-物排斥力。除了来自其他智能体的排斥力外,障碍物 w 的排斥力也会影响智能体 J 的运动。与其他智能体的排斥力类似,智能体 J 附近所有障碍物的排斥力 f_{Jw}^{obs} 定义如

$$f_{Jw}^{\text{obs}}=\lambda_w\cdot\exp(\beta_w^{-1}(r_{Jw}-d_{Jw}))\cdot\boldsymbol{n}_{Jw},\tag{3}$$

式中 λ_w、β_w 为人(机)-物间社会力强度、作用范围参数,r_{Jw} 为智能体 J 与障碍物 w 之间的社会舒适距离,d_{Jw} 为智能体 J 与障碍物 w 之间的欧氏距离,$\boldsymbol{n}_{Jw}$ 为智能体 J 与障碍物之间的方向向量。

集合 N_J 表示智能体 J 的相邻智能体的集合,即在其感知范围内与其产生社会交互的其他行人,集合 W 表示环境中的障碍物集合,包括墙体及固定障碍物等。最终,智能体 J 的 SFM 由公式(1)~(3)三种力的合成,如

$$f_J^{\text{SFM}}=f_J^{\text{goal}}+\sum_{K\in N_J}f_{JK}^{\text{rep}}+\sum_{w\in W}f_{Jw}^{\text{obs}}。\tag{4}$$

1.3　基于 RVO-SFM 融合社交模型

为兼顾避障效果与行为自然性,本文提出了一种以 RVO 为主、融合 SFM 社会行为机制的行人交互模

型。该融合模型利用 RVO 精确构造的速度障碍机制实现动态避碰,同时引入 SFM 所建模的社会力,增强了行人在非紧急状态下的动作舒适性和自然性。融合策略旨在不同密度与风险条件下,动态地调节 RVO 与 SFM 两种机制对最终运动行为的主导权重。

具体而言,在每个仿真时间步,对于每个行人个体 A,我们首先分别计算其基于 RVO 模块的避碰速度 $\boldsymbol{v}_A^{\mathrm{RVO}}$ 以及基于 SFM 模块的社会力速度 $\boldsymbol{v}_A^{\mathrm{SFM}}$。随后,基于当前场景状态和行人所处的局部环境,采用加权融合的方式生成最终的运动速度 $\boldsymbol{v}_A^{\mathrm{fused}}$,计算方式如

$$\boldsymbol{v}_A^{\mathrm{fused}} = \alpha_A \cdot \boldsymbol{v}_A^{\mathrm{RVO}} + (1-\alpha_A) \cdot \boldsymbol{v}_A^{\mathrm{SFM}}, \tag{5}$$

式中融合系数 $\alpha_A \in [0,1]$ 控制 RVO 部分在最终速度中的占比。当 $\alpha_A = 1$ 时,表示完全依赖 RVO 进行避障决策;当 $\alpha_A = 0$ 时,则完全依赖 SFM 的社会力进行决策。该融合机制为行人提供了在不同情境下的动态适应能力。

基于碰撞风险的融合策略:对于复杂场景或存在显著运动冲突的区域,通过结合碰撞时间(Time-To-Collision, TTC)[26]指标来动态调节融合系数。当预测 TTC 较小时,行人面临更高碰撞风险,优先采用 RVO 控制策略;反之则可增加 SFM 主导行为,表达式为

$$\alpha_A = \min(1, \lambda (t_A^{\mathrm{TTC}} + \varepsilon)^{-1}), \tag{6}$$

式中 λ 为调节参数,t^{TTC} 为碰撞时间,ε 为避免除零的小常数。该策略可进一步提高模型在局部拥堵场景下的鲁棒性。此外,为避免融合后的速度出现突变,本文在生成最终速度时引入了加速度约束机制。具体而言,若当前速度为 $\boldsymbol{v}_A$,则最终输出速度需满足最大加速度限制

$$\| \boldsymbol{v}_A^{\mathrm{fused}} - \boldsymbol{v}_A \| \leqslant a^{\max} \cdot \Delta t。 \tag{7}$$

若超出该范围,则将 $\boldsymbol{v}_A^{\mathrm{fused}}$ 向当前速度方向裁剪至满足限制,确保轨迹平滑、连续并物理可行。

如图 4 所示,该融合模型在避免碰撞的同时保持了行人运动的流畅性和自然性,为密集动态场景下的行人建模提供了一种高精度、高稳定性的解决方法。

1.4 RVO-SFM 融合机理与理论分析

在密集动态人群环境中,单一交互建模方法往往难以同时兼顾避碰安全性、运动平滑性与社交合理性。RVO 与 SFM 分别从几何约束和动力学建模角度刻画智能体间的交互行为,二者在建模假设、作用空间及优化目标上存在显著差异,同时也具有天然的互补性。RVO 方法定义于速度空间,通过构造 VO 集合,将可能导致未来碰撞的速度显式地排除在可行解空间之外。其本质是一种基于几何关系的可行域约束方法,能够在局部范围内提供严格的无碰撞安全保证。因此,RVO 在高动态性和强交互场景中具有响应迅速、计算效率高的优势,尤其适用于紧急避碰情形。相比之下,社会力模型(SFM)定义于力或加速度空间,通过连续的吸引力与排斥力对智能体的运动行为进行调节。该模型强调个体之间的心理和社会影响,其作用结果表现为速度和轨迹的平滑变化,能够自然地刻画人群中常见的社会规范行为,如保持舒适距离、顺势避让等。然而,由于 SFM 缺乏对速度可行域的显式约束,在高密度或高速交互场景中难以单独保证严格的碰撞安全性。因此,从建模空间角度看,RVO 提供了硬约束的速度安全边界,而 SFM 提供了软约束的行为调节机制。二者在速度空间与动力学空间上的差异,使其在融合后能够同时兼顾安全性与行为自然性。

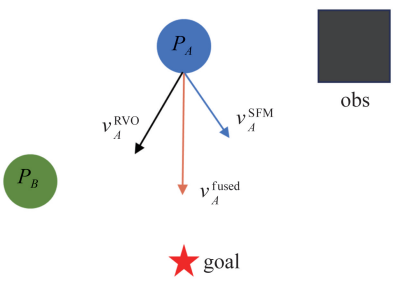


图 4 RVO-SFM 融合模型示例

Fig. 4 Examples of RVO-SFM

RVO 方法可被视为一个带有安全约束的速度选择问题。在给定期望速度 $\boldsymbol{v}^{\mathrm{pref}}$ 的前提下,RVO 通过构造速度障碍区域集合 S^{R},将可能导致未来碰撞的速度从可行解空间中排除,其决策过程可形式化为如下约束优化问题

$$\boldsymbol{v}^{*} = \operatorname{argmin} \| \boldsymbol{v} - \boldsymbol{v}^{\mathrm{pref}} \| \ \text{ s.t. } \boldsymbol{v} \notin S^{\mathrm{R}}, \tag{8}$$

式中,约束条件定义了一个非凸的速度可行域,其几何结构由相对位置、相对速度及智能体半径共同决定。该模型强调安全优先,本质上是一种硬约束优化框架。

相比之下,SFM 并未显式构造可行域约束,而是通过连续力的叠加描述智能体的运动变化,其数学形式

来源于牛顿第二定律

$$m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=f^{\mathrm{goal}}+\sum f^{\mathrm{rep}}+\sum f^{\mathrm{obs}}。\tag{9}$$

在该模型中,交互影响通过连续可导的力场函数体现,智能体的速度变化是一个连续时间动力学系统的演化结果。SFM 更侧重于描述群体行为趋势和社会规范,其约束是“软约束”,并不保证速度解始终满足无碰撞条件。因此,从统一数学视角看,RVO 与 SFM 分别对应为约束优化问题连续动力学系统。

RVO-SFM 融合策略在数学上可被视为一种风险感知的混合控制模型(Hybrid Control)。在每个离散时间步内,系统同时计算基于 RVO 的安全速度 $\boldsymbol{v}_A^{\mathrm{RVO}}$ 与基于 SFM 的行为速度 $\boldsymbol{v}_A^{\mathrm{SFM}}$,并通过连续权重进行加权融合。如公式(6)所示,当 t^{TTC} 趋近于 0 时,融合系数 α_A 增大,RVO 的硬约束机制占据主导,强制切断任何可能导致碰撞的速度分量,解决了 SFM 在紧急状态下反应迟缓或陷入局部极小值的问题;相反,融合系数 α_A 减小,SFM 提供的社会力对 RVO 生成的机械轨迹进行优化,解决 RVO 轨迹生硬的问题。综上,RVO 与 SFM 的融合并非经验性叠加,而是在数学建模层面实现了约束优化与连续动力学的互补统一,从而在复杂多智能体交互场景中同时提升安全性、平滑性与社交合理性。

2　多智能体强化学习

2.1　马尔可夫决策过程

将多机器人的运动场景建模为多智能体马尔可夫决策过程,为系统地描述多智能体在共享环境中的感知、决策与交互过程,本文采用了马尔可夫决策过程(Markov Decision Process,MDP)的多智能体扩展,即部分可观察马尔可夫博弈。N 个智能体的马尔可夫博弈由一组描述所有智能体可能配置的状态 $\boldsymbol{S}$,一组动作 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_N$ 和每个智能体的一组观察值 $\boldsymbol{O}_1,\cdots,\boldsymbol{O}_N$ 定义。考虑到 RVO 与 SFM 均在速度空间进行计算,为了实现精细的运动控制,本文采用连续动作空间,每个智能体 i 在时刻 t 的动作 $\boldsymbol{a}_i^t$ 定义为二维速度矢量 $\boldsymbol{a}_i^t=[\boldsymbol{v}_{ix},\boldsymbol{v}_{iy}]\in\boldsymbol{A}_i$,其中,$\boldsymbol{v}_{ix}$ 和 $\boldsymbol{v}_{iy}$ 分别表示智能体 i 在横向和纵向的期望速度分量。$\boldsymbol{S}$ 表示系统的全局状态空间,由于实际环境中智能体无法获得全局状态信息,本文采用部分可观测建模,$\boldsymbol{O}_i$ 表示智能体 i 的局部观测空间。动作空间受到机器人的最大速度 $\|\boldsymbol{a}_i^t\|^2\leqslant 1.5\ \mathrm{m/s}$ 的约束。为了选择动作,每个智能体 i 使用随机策略 $\boldsymbol{\pi}_{\theta_i}:\boldsymbol{O}_i\times\boldsymbol{A}_i\mapsto[0,1]$,其中 θ_i 表示策略网络参数,该策略刻画了在给定观测条件下,智能体选择不同动作的概率分布,该策略根据状态转换函数 $\boldsymbol{T}:\boldsymbol{S}\times\boldsymbol{A}_1\times\cdots\times\boldsymbol{A}_N\mapsto\boldsymbol{S}'$ 即在当前状态及所有智能体联合动作作用下,系统转移至下一时刻状态,每个智能体 i 获得作为状态和智能体动作的奖励函数 $r_i:\boldsymbol{S}\times\boldsymbol{A}_i\mapsto\mathbf{R}$,并接收与状态相关的观察值 $\boldsymbol{o}_i:\boldsymbol{S}\mapsto\boldsymbol{O}_i$。初始状态由分布 $\rho:\boldsymbol{S}\mapsto[0,1]$ 决定。每个智能体 i 旨在最大化自己的总预期回报 $R_i=\Sigma_{t=0}^{m}\gamma^t r_i^t$,其中 γ 是折扣因子,m 是时间范围。

2.2　注意力图模型

为提升多智能体社交导航中的交互建模能力,我们采用注意力图结构[27]并结合 MAPPO 进行策略优化[28]。首先,利用轨迹预测器估计各智能体的未来位置,从而构建场景图,其中节点表示智能体的连续未来状态,边表示智能体之间的可见性关系。随后,该场景图被输入由两个级联图注意力网络(Graph Attention Network,GAT)组成的注意力模块:第一层主要用于建模环境中其他智能体之间的交互关系,以捕捉局部人群结构特征;第二层则进一步聚焦主智能体与其邻居之间的关联程度,实现对与当前决策最相关交互对象的筛选与强化。经注意力编码后的节点表示最终输入至 RNN 解码器,用于生成智能体的控制动作。注意力图模型的整体结构如图 5 所示。

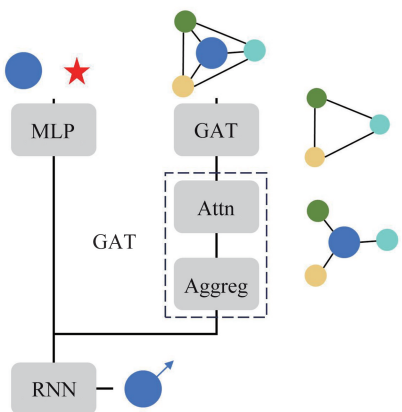


图 5　注意力图模型结构示意图

Fig. 5　Architectures of attention graph

在该注意力图模型中,注意力权重用于刻画不同邻居智能体对主智能体决策过程的影响强度,其大小综合反映了相对位置关系、运动方向差异及潜在交互风险等因素。在迎面交汇、横向穿越等高冲突交互场景中,相关邻居节点通常被分配更高的注意力权重,从而引导决策网络重点关注具有较高碰撞或社交干扰风险

的对象;而对于距离较远、运动方向一致或交互可能性较低的邻居,其注意力权重相对较小,以减少冗余社交信息对决策的干扰。该机制使模型能够在复杂人群环境中自适应地捕捉关键社交交互关系,并赋予注意力权重明确的物理与社交意义。通过使用 MAPPO 算法[29]进行策略训练,实现对注意力图架构的集中训练与分散执行[30],所有智能体共享一套策略网络进行训练。在执行阶段,每个智能体根据其内部状态与视野范围(Field of View,FoV)内的局部观测独立决策。MAPPO 的超参数设置与注意力图模型一致,以提升多智能体场景下训练的稳定性与收敛性。环境根据所有智能体的动作执行状态转移并分别计算奖励,用于驱动策略更新。

2.3 奖励函数

我们在采用注意力图模型及其改进型研究中使用的奖励函数的基础上,通过修改奖励函数配置使得该系统在多智能体与多智能体编队方面取得更好的性能。本方法对智能体在预测路径中的惩罚函数为

$$\begin{aligned} r_{j,i}^{\text{pred}}(s_t) &= \min_{k\in[1,5]}((\boldsymbol{C}_j^{t+k}, i)\,\frac{r_c}{2^k}), \\ r_j^{\text{pred}}(s_t) &= \min_{i\in[1,N]}(r_{j,i}^{\text{pred}}(s_t)), \end{aligned} \tag{10}$$

式中 s_t 是智能体 j 在 t 时刻的状态,r_c 是碰撞的惩罚,$(\boldsymbol{C}_j^{t+k}, i)$ 表示智能体 j 是否与实体 i 的在第 k 个预测位置发生碰撞,r_j^{pred} 为行人处于智能体 j 预测路径上的惩罚。

基于过程的奖励函数引导智能体 j 接近目标奖励函数 $r_j^{\text{pot}} = -d_{j,t}^{\text{goal}} + d_{j,t-1}^{\text{goal}}$。为了使智能体表现出符合人类预期的社交行为,奖励函数的设计至关重要。虽然真实场景中的人类交互包含礼让、协作等复杂的心理博弈,但研究表明,在底层的避碰决策中,人类倾向于遵循简单且直观的启发式规则以快速化解冲突,因此,我们在现有的奖励函数基础上引入了右手通过规则[23,31]。正因如此,采用基于奖励与惩罚的机制来引导机器人学习人类的社会习惯,成为一种有效策略,采用此方法不仅易于实施,也无需依赖精确的数学模型。本文使用以下奖励函数规范诱导机器人遵循右手规则。

$$\begin{aligned} S^{\text{pass}} &= \{s_{j,i}^t \mid d_{j,t}^{\text{goal}} > 3, 1 < \widetilde{p}_x < 4, -2 < \widetilde{p}_y < 0, \mid \widetilde{\varphi} - \varphi \mid > 3\pi/4\}, \\ S^{\text{over}} &= \{s_{j,i}^t \mid d_{j,t}^{\text{goal}} > 3, 0 < \widetilde{p}_x < 3, 0 < \widetilde{p}_y < 1, \mid \widetilde{\varphi} - \varphi \mid < \pi/4\}, \\ S^{\text{cross}} &= \{s_{j,i}^t \mid d_{j,t}^{\text{goal}} > 3, \widetilde{d} < 2, \widetilde{\varphi}_{rot} > 0, -3\pi/4 < \widetilde{\varphi} - \varphi < -\pi/4\}, \end{aligned} \tag{11}$$

$$S^{\text{punish}} = S^{\text{pass}} \cup S^{\text{over}} \cup S^{\text{cross}}, \tag{12}$$

式中 $S^{\text{pass}}, S^{\text{over}}, S^{\text{cross}}$ 分别代表智能体 j 对行人产生的路过、超过和穿越行为情形的集合,S^{punish} 为三种惩罚性交互行为的集合。$\widetilde{d} = \| P - \widetilde{P} \|_2$ 为智能体 j 与行人的距离,$d_{j,t}^{\text{goal}} = \| P^{\text{goal}} - P \|_2$ 为智能体 j 在 t 时所处位置到其目标的距离,$\widetilde{\varphi}^{\text{rot}} = \tan^{-1}((\widetilde{\boldsymbol{v}}_x - \boldsymbol{v}_x / \widetilde{\boldsymbol{v}}_y - \boldsymbol{v}_y))$ 为两个智能体之间的相对旋转角度,角度 $\widetilde{\varphi} - \varphi$ 取值范围为 $[-\pi, \pi]$。综合上述奖励函数,对智能体 j 的完整奖励函数为

$$R_j(s_t, a_t) = \begin{cases} r_c, \text{智能体存在碰撞}, \\ q \cdot I(s_{j,i}^t \in S^{\text{punish}}), \text{智能体存在交互}, \\ r_j^{\text{pot}} + r_j^{\text{pred}}, \text{其他}, \end{cases} \tag{13}$$

式中 q 是标量惩罚。$I(\cdot)$ 是指示函数,当符合惩罚集的定义时函数值为 1,不符合时值为 0。直到所有智能体都发生碰撞或达到目标之前,单一智能体在达到目标时没有任何奖励。因此,每个达到目标的智能体必须等待其他智能体完成。如此设置是为了防止如果达到目标有奖励,一个智能体会获得无限的奖励等待另外的智能体的问题的产生。

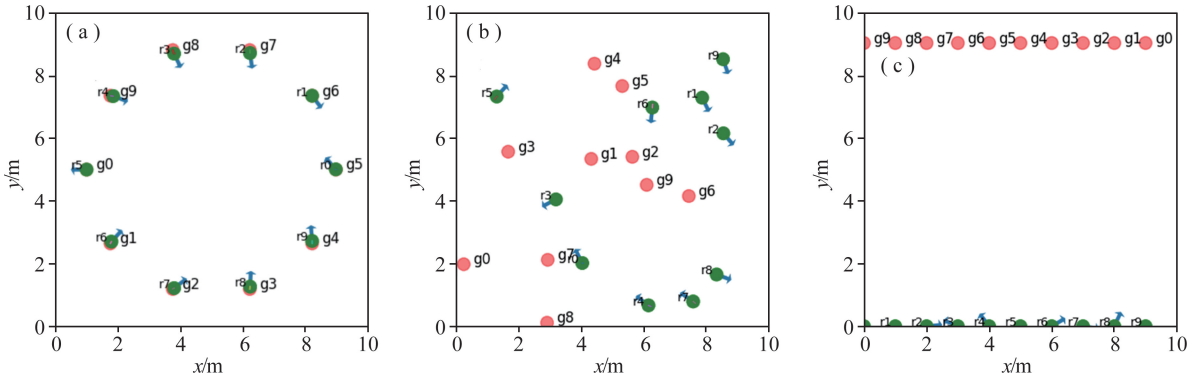
3 仿真实验及分析

我们将 SFMRVO-DRL 算法应用于多智能体导航的实验,并与导航领域广泛应用且具有代表性的 GA3C-CADRL(GA3C-Collision Avoidance with Deep RL)[32],一种经典的通过预测智能体的未来状态辅助决策深度的强化学习的避障算法,NH-ORCA(Non-Holonomic Optimal Reciprocal Collision Avoidance)[33],基于非完整性约束改进的最优互惠速度障碍算法运动规划算法,并与最新的 HeR-DRL(Heterogeneous Relational Deep Reinforcement Learning)[8],利用异构关系图进行复杂环境下决策的深度强化学习

导航算法做为基线方法进行对比实验。还将仅基于 RVO、SFM 的多智能体深度强化学习的 RL-RVO 算法[34]、RL-SFM 算法作为消融实验进行比较。

3.1　实验环境和参数

我们使用多个智能体仅通过圆形场景训练碰撞规避策略。在圆形场景中,所有具有随机方向的机器人都均匀排列,并形成一个圆形,目标位置位于相反侧,这为智能体训练提供了丰富的交互情况。实验场景基于开源框架 CrowdNav[27]进行构建。为了全面评估策略性能,本文设计了三种典型的测试场景:圆形场景,所有机器人均匀排列成圆形,目标点位于直径相对位置。该场景提供了高度对称且密集的正面交互,是评估多智能体避障能力的标准环境;随机场景,机器人的初始位置与目标点在区域内随机生成。此场景具有非结构化特征,用于测试算法在处理不可预测动态干扰时的泛化能力;走廊场景,模拟受限空间内的交互行为,智能体在长方型走廊的一端依次排开,目标点位于对向另一端。该场景限制了智能体的侧向规避空间,能有效验证模型的社交合规性与模型泛化性能。测试场景如图 6 所示。



注:(a) 圆形场景;(b) 为随机场景;(c) 为走廊场景。

图 6　测试场景

Fig. 6　Testing scenario

每个机器人的感知范围为 4 m,最大输入邻居数为 5 个。所有机器人共享相同的策略,并在每个 epoch 后收集观测值以进行参数更新。当发生碰撞或超过最大情节长度时,环境将被重置。为了加快训练收敛速度,我们使用两个阶段来执行训练过程。首先,在 10 m 为直径的圆圈场景中训练策略,使用 4 个智能体进行 200 次 epoch 的训练,以实现基本功能,例如向目标位置移动。然后,在模拟行人环境的仿真场景中继续用 10 个智能体训练策略大约 1 000 次 epoch 的训练。

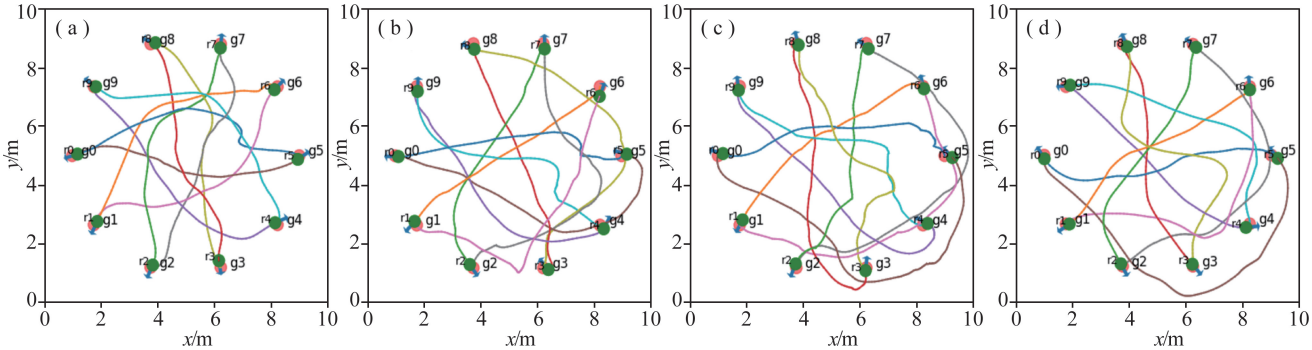
本实验使用三个指标来评估策略效果,包括成功率、平均导航时间和平均导航速度[34]。成功率是在导航过程中没有发生任何碰撞或卡在某处的成功案例的比率,它描述了策略的碰撞规避能力。导航时间是指所有机器人到达目标位置时,仿真中的迭代步数表示的时间量,它反映了策略的运行效率。整个多智能体导航过程的平均速度衡量了策略对最优速度选择的表现。此外,本实验还对单个机器人的控制动作的平均计算成本进行了比较。本实验训练过程通过配备 CPU i7-13650HX 和 GPU Nvidia GTX 4070 的计算机使用 Pytorch(Python 3.8)执行。

3.2　实验结果和讨论

本文在圆形场景、随机场景以及走廊场景三种典型场景下评估了多智能体自主导航性能,其中对比试验圆形场景中的导航轨迹如图 7 所示。通过在上述测试场景中进行对比,本研究系统评估了 SFMRVO-DRL(本文算法)、NH-ORCA、GA3C-CADRL 及 HeR-DRL 的性能,对比指标涵盖成功率、平均导航时间与平均导航速度,旨在全面衡量算法的避障稳健性与规划效率。

如表 1 所示,本文提出的 SFMRVO-DRL 算法在所有对比算法中表现最优,取得了最高的综合成功率,同时平均导航时间与平均导航速度均保持较低水平,这表明该算法在成功规避动态障碍物的前提下,能够规划出相对高效且平滑的移动轨迹。与 HeR-DRL 算法相比,虽然 HeR-DRL 利用异构关系图增强了特征提取能力,但在极端拥堵的对峙点,纯学习驱动的方法偶尔会出现决策震荡;而 SFMRVO-DRL 由于引入了 RVO 与 SFM 的物理先验,为强化学习策略提供了明确的几何安全边界与社交力引导,使其在复杂博弈中

表现出更强的确定性。GA3C-CADRL 算法通过预测动态障碍物的未来状态来辅助决策,在部分场景下展现了良好的前瞻性。然而,实验结果也暴露出其明显的缺陷,在密集、非结构化的动态环境中,其基于预测的动作选择方案容易遇到“冻点问题”,即智能体因过度谨慎或对未来碰撞概率的误判,在决策点附近长时间徘徊而无法做出有效移动,导致平均完成时间显著增加,规划效率下降。NH-ORCA 算法基于优化的速度障碍法,提供了强理论保证的局部避碰,因此取得了较高的综合成功率。但其短板在于缺乏全局或中长程的路径优化机制。结果显示,其平均导航时间与 SFMRVO-DRL 算法相比较长,这是因为 NH-ORCA 更侧重于保证每一步的局部无碰撞,在全局视角下导致智能体选择并非最短的迂回路径,牺牲了路径最优性以换取更高的安全性与成功率。在走廊场景中,HeR-DRL 依靠异构关系图,提高了对其他智能体的特征提取能力,取得了最高的运行成功率。SFMRVO-DRL 则展现了显著的社会合规性与空间适应力。在侧向规避空间受限的狭长区域内,受右手通行奖励函数与 SFM 个人空间感的引导,同向出发的智能体能够自发地形成有序的通行队列,显著降低了冲突以及滞留的时间。实验结果证明,相较于其他纯学习算法或纯几何避障方法,本文方法通过深度融合强化学习的序贯决策能力与改进的物理交互模型,在成功率、路径质量与社交效率三个维度上达成了更佳的综合平衡,有效提升了复杂动态环境下的泛化性能。



注:(a) 为 SFMRVO-DRL 算法;(b) 为 HeR-DRL 算法;(c) 为 GA3C-CADRL 算法;(d) 为 NH-ORCA 算法。

图 7 对比试验圆形场景导航轨迹图

Fig. 7 Trajectory map in the circular scenario for comparative experiments

表 1 对比测试实验的性能指标对比

Tab. 1 Comparison of performance indicators of test experiment

场景	模型	成功率	导航步长	导航速度/(m·s^{-1})
圆形场景	SFMRVO-DRL	**0.99**	**71.47±3.93**	**1.13±0.05**
	GA3C-CADRL	0.82	85.97±8.02	1.11±0.14
	NH-OCRA	0.97	72.96±19.53	1.08±0.10
	HeR-DRL	0.96	72.57±7.34	1.10±0.21
随机场景	SFMRVO-DRL	**0.97**	**73.84±9.71**	**0.79±0.12**
	GA3C-CADRL	0.74	146.52±41.26	0.39±0.34
	NH-OCRA	0.95	75.33±25.64	0.72±0.18
	HeR-DRL	0.92	78.94±10.22	0.70±0.41
走廊场景	SFMRVO-DRL	0.86	**157.84±12.35**	**0.77±0.17**
	GA3C-CADRL	0.58	285.02±44.98	0.49±0.57
	NH-OCRA	0.80	166.74±26.20	0.72±0.26
	HeR-DRL	**0.88**	159.92±14.57	0.73±0.32

为了评估算法在实际物理平台上的部署可行性,本文在相同硬件环境下对各算法的单步推理耗时进行了量化对比,对比内容为圆形场景中每个机器人实现单次控制动作的平均计算成本,比较结果如图 8 所示。实验结果显示,虽然由于引入了图注意力机制与融合交互模型,SFMRVO-DRL 的计算成本略高于纯几何

驱动的 NH-ORCA 算法,但其单步平均耗时仅为 0.012 s,远低于常规社交机器人导航所需的实时性阈值(通常为 0.1 s)。相较于 HeR-DRL 算法以及经典的 GA3C-CADRL 算法,本文方法在保持高性能导航的同时,通过优化图结构的稀疏性,实现了更低的计算负载。这证明了 SFMRVO-DRL 在兼顾决策深度与实时性要求方面的优越性,具备在资源受限的移动终端上进行大规模部署的潜力。

为了进一步验证本研究提出的 RVO-SFM 融合交互模型以及各核心组件对导航性能的具体贡献,本文设计并执行了消融实验,其中消融实验中走廊场景的导航轨迹如图 9 所示。实验选取了两种变体算法作为对比基准:RL-RVO 算法,移除社会力模型,仅保留速度障碍引导;RL-SFM 算法,移除速度障碍模型,仅保留社会力引导。通过在相同参数设置的圆形场景、随机场景与走廊场景下进行测试,评估各组件对策略安全性、平滑性及收敛稳定性的影响。

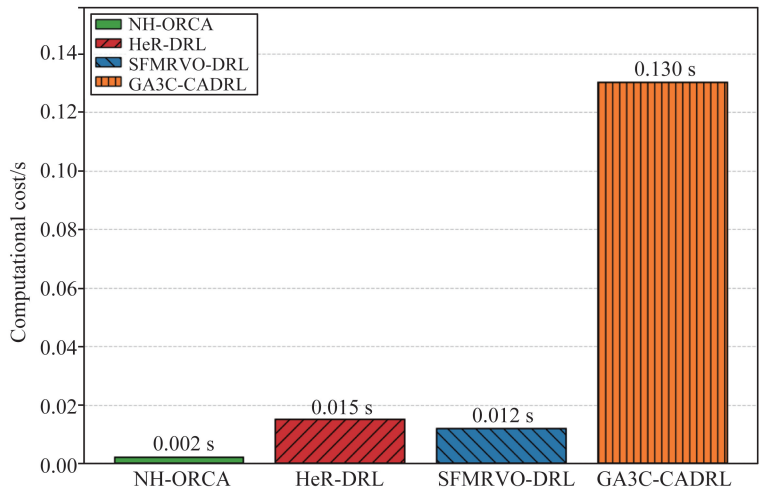
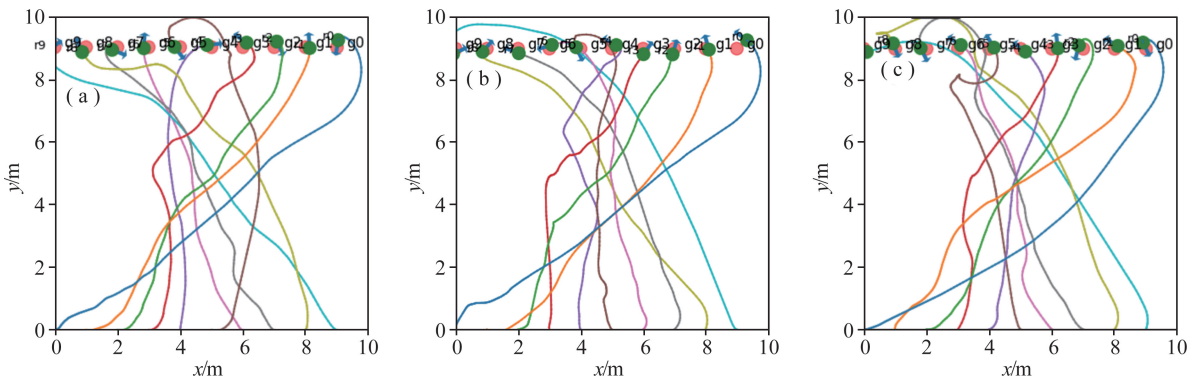


图 8　单步推理耗时对比图

Fig. 8　A comparison of the average computational cost to achieve single control action for each robot



注:(a) 为 SFMRVO-DRL 算法;(b) 为 RL-RVO 算法;(c) 为 RL-SFM 算法。

图 9　消融实验走廊场景导航轨迹图

Fig. 9　Trajectory map in the corridor scenario for ablation studies

消融实验的定量结果如表 2 所示。实验结果表明,完整模型 SFMRVO-DRL 在各项指标上均优于单一模型引导的变体算法。在对比 RL-RVO 时可以发现,虽然 RL-RVO 在避障成功率上表现尚可,但在反映交互自然度的导航速度指标上略逊于本文算法。这是因为 RL-RVO 缺乏对其余智能体个人空间的柔性建模,导致智能体在非紧急状态下的避让动作过于机械,轨迹平滑度不足。通过引入 SFM 模块,完整模型能够利用社会力场实现更早的趋势性避让,从而在物理安全的基础上提升了社交适应性。另一方面,对比 RL-SFM 的结果显示,单纯的社会力引导的方法在密集对流环境下的性能波动较大,尤其是在走廊场景中,其成功率显著下降,这符合了前文所述的 SFM 在处理极近距离冲突时易陷入局部极小值的问题。而 SFMRVO-DRL 通过 TTC 动态调节机制,在碰撞风险升高时及时引入 RVO 的几何硬约束,弥补了 SFM 的失效区间。

表 2 消融实验性能指标对比

Tab. 2 Comparison of performance indicators of test experiment

场景	模型	成功率	导航步长	导航速度/(m·s^{-1})
圆形场景	SFMRVO-DRL	**0.99**	**71.47±3.93**	**1.13±0.05**
	RL-RVO	0.93	74.31±16.77	1.02±0.16
	RL-SFM	0.76	102.01±12.64	1.00±0.13
随机场景	SFMRVO-DRL	**0.97**	**73.84±9.71**	**0.79±0.12**
	RL-RVO	0.82	83.69±19.93	0.74±0.28
	RL-SFM	0.67	86.51±16.84	0.73±0.17
走廊场景	SFMRVO-DRL	**0.88**	**157.84±12.35**	**0.77±0.17**
	RL-RVO	0.64	189.76±22.14	0.61±0.42
	RL-SFM	0.60	192.45±20.58	0.59±0.38

综上所述,SFMRVO-DRL 算法的优越性能得益于其深度融合了强化学习的序贯决策能力与改进的速度障碍法的实时局部避障机制,并通过社会性奖励函数等设计缓解了局部极小与探索不足的问题。相比基础 RL-RVO,显著提升了成功率;相比 GA3C-CADRL,有效避免了“冻点问题”,提升了决策流畅度;相比 NH-ORCA,在保持高成功率的同时,通过全局策略优化生成了更短的路径。实验结果验证了 SFMRVO-DRL 在动态密集环境中,在成功率、导航路径、运行效率三个维度上取得了更好的综合平衡。

4 结论

本文提出了基于课程学习的多智能体导航方法 SFMRVO-DRL,利用社会力模型与 RVO 结合的融合策略解决智能体碰撞避免问题,更好地描述多智能体之间的相互作用。结合社会力模型和 RVO 的避碰策略以及基于社会规则的奖励函数可以实现碰撞互惠行为,并在碰撞风险与旅行时间之间权衡。通过模拟实验评估该策略的导航性能,通过比较包括 NH-ORCA、GA3C-CADRL 和 HeR-DRL 在内的三种策略以及消融实验,在圆形场景、随机场景、走廊场景中的成功率、运行时间和平均速度方面进行了比较。实验结果表明,在拥挤环境中,该方法在碰撞避免能力和时间效率上优于其他方法。

参 考 文 献

[1] 何丽, 张恒, 袁亮,等. 服务机器人社会意识导航方法综述 [J]. 计算机工程与应用, 2022, 58(11): 1-11.

[2] SINGAMANENI P T, BACHILLER-BURGOS P, MANSO L J, et al. A survey on socially aware robot navigation: taxonomy and future challenges[J]. The International Journal of Robotics Research, 2024, 43(10): 1533-72.

[3] HUANG Z, LI R H, SHIN K, et al. Learning sparse interaction graphs of partially detected pedestrians for trajectory prediction[J]. IEEE Robotics and Automation Letters, 2021, 7(2): 1198-205.

[4] MIRSKY R, XIAO X S, HART J, et al. Conflict avoidance in social navigation-a survey[J]. ACM Transactions on Human-Robot Interaction, 2024, 13(1): 1-36.

[5] 姚陈鹏, 石文博, 刘成菊,等. 移动机器人导航技术综述[J]. 中国科学:信息科学, 2023, 53(12): 2303-2324.

[6] 姜杨,赵天祥,孙若怀,等.风险导向的深度强化学习人群导航策略[J].东北大学学报(自然科学版), 2025, 46(12):1-8.

[7] JIA Z T, GUO S, WANG K, et al. Autonomous navigation for robots in crowd with graph representation and active learning[C]//Proceedings of 4th 2024 international Conference on Autonomous Unmanned Systems,2025,1379:570-583.

[8] ZHOU X Y, PIAO S H, CHI W Z, et al. HeR-DRL: heterogeneous relational deep reinforcement learning for single-robot and multi-robot crowd navigation[J]. IEEE Robotics and Automation Letters, 2025,10(5):4524-4531.

[9] YANG S H, LI B, XU Y M, et al. Robot crowd navigation incorporating spatio-temporal information based on deep reinforcement learning[C]//Proceedings of 36th Chinese Control and Decision Conference,2024:5773-5780.

[10] ZHOU Y Y, GARCKE J. Learning crowd behaviors in navigation with attention-based spatial-temporal graphs[C]//Proceedings of IEEE International Conference on Robotics and Automation, 2024:5485-5491.

[11] MA T, LIU Z, LIU T, et al. A spatiotemporal graphical attention navigation algorithm based on limited state information[J]. IEEE Transactions on Computational Social Systems, 2024, 11(5): 6407-6421.

[12] CHEN W N, CHI W Z, JI S H, et al. A survey of autonomous robots and multi-robot navigation: perception, planning and collaboration [J]. Biomimetic Intelligence and Robotics, 2025, 5(2): 100203.

[13] HUANG A Q, WANG Y L, SANG J H, et al. DVF: multi-agent Q-learning with difference value factorization[J]. Knowledge-Based Systems, 2024, 286: 111422.

[14] 方宝富，余婷婷，王浩，等. 稀疏奖励场景下基于适应性状态近似的多智能体强化学习[J]. 机器人，2024，46(6)：663-671.

[15] GUPTA N, HARE J Z, KANNAN R, et al. Deep Meta Coordination Graphs for Multi-agent Reinforcement Learning[EB/OL]. arXiv preprint arXiv:2502.04028, 2025.

[16] WU H M, TIAN L, TANG H J, et al. Graph convolutional reinforcement learning-guided joint trajectory optimization and task offloading for aerial edge computing[J]. IEEE Transactions on Intelligent Transportation Systems, 2024,26(10): 17487-17498.

[17] LI S, GUPTA J K, MORALES P, et al. Deep implicit coordination graphs for multi-agent reinforcement learning[EB/OL]. arXiv preprint arXiv:2006.11438, 2021.

[18] 陈应霞，卫明明，汪勇，等. 融合 RVO 和路径规划的塔机动态避障算法[J]. 安全与环境学报，2025，25(7)：2643-2655.

[19] 陈奕梅，康雪晶，徐鹏. 复杂环境下多移动机器人控制算法研究[J]. 电光与控制，2021，28(4)：48-52.

[20] HELBING D, MOLNÁR P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282-4286.

[21] FLÖGEL D, FISCHER L, RUDOLF T, et al. Socially integrated navigation: a social acting robot with deep reinforcement learning [C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2024:4785-4792.

[22] 景荣荣，吴兰，张坤鹏. 基于 Transformer 的自动驾驶交互感知轨迹预测[J]. 科学技术与工程，2023，23(26)：11414-1123.

[23] 顾金浩，况立群，韩慧妍，等. 动态环境下共融机器人深度强化学习导航算法[J]. 计算机工程与应用，2025，61(4)：90-98.

[24] CHEN L, WANG Y N, MIAO Z Q, et al. Reciprocal velocity obstacle spatial-temporal network for distributed multirobot navigation [J]. IEEE Transactions on Industrial Electronics, 2024, 71(11): 14470-14480.

[25] LE H T, NGUYEN D T, TRUONG X T. Socially aware robot navigation framework in crowded and dynamic environments: a comparison of motion planning techniques[C]//Proceedings of 8th NAFOSTED Conference on Information and Computer Science,2021:95-101.

[26] JAFARI A, LIU Y C. Time-to-collision based social force model for intelligent agents on shared public spaces[J]. International Journal of Social Robotics, 2024, 16(9): 1953-1968.

[27] LIU S J, CHANG P X, HUANG Z, et al. Intention aware robot crowd navigation with attention-based interaction graph[C]//Proceedings of IEEE International Conference on Robotics and Automation, 2023: 12015-12021.

[28] GAO Y, YANG F K, FRISK M, et al. Learning socially appropriate robot approaching behavior toward groups using deep reinforcement learning[C]//Proceedings of 28th IEEE International Conference on Robot and Human Interactive Communication,2019:1-8.

[29] ZHOU Z Y, LIU G J, TANG Y. Multiagent reinforcement learning: methods, trustworthiness, applications in intelligent vehicles, and challenges[J]. IEEE Transactions on Intelligent Vehicles, 2024,9(12):8190-8211.

[30] LOGOTHETIS M, KARRAS G, ALEVIZOS K, et al. Efficient cooperation of heterogeneous robotic agents: a decentralized framework [J]. IEEE Robotics & Automation Magazine, 2021, 28(2): 74-87.

[31] MOUSSAÏD M, HELBING D, THERAULAZ G. How simple rules determine pedestrian behavior and crowd disasters[J]. Proceedings of the National Academy of Sciences, 2011, 108(17): 6884-6888.

[32] EVERETT M, CHEN Y F, HOW J P. Motion planning among dynamic, decision-making agents with deep reinforcement learning[C]// Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems,2019:3052-3059.

[33] ZHANG W, LIU Z, GAO J. Decentralized navigation of multiple robots based on event-triggered NMPC and NH-ORCA[C]//Proceedings of 14th Asian Control Conference,2024:2277-2282.

[34] HAN R, CHEN S, WANG S, et al. Reinforcement learned distributed multi-robot navigation with reciprocal velocity obstacle shaped rewards [J]. IEEE Robotics and Automation Letters, 2022, 7: 5896-903.

（责任编辑：刘庆松）