控制与决策
Control and Decision

基于多智能体深度强化学习的船舶协同避碰策略

隋丽蓉, 高曙, 何伟

引用本文:
隋丽蓉,高曙,何伟. 基于多智能体深度强化学习的船舶协同避碰策略[J]. 控制与决策, 2023, 38(5): 1395−1402.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2022.1159

基于多智能体深度强化学习的船舶协同避碰策略

隋丽蓉[1], 高 曙[1†], 何 伟[2]

(1. 武汉理工大学 计算机与人工智能学院，武汉 430063；
2. 闽江学院 物理与电子信息工程学院，福州 350108)

摘 要: 船舶避碰是智能航行中首要解决的问题,多船会遇局面下,只有相互协作,共同规划避碰策略,才能有效降低碰撞风险. 为使船舶智能避碰策略具有协同性、安全性和实用性,提出一种基于多智能体深度强化学习的船舶协同避碰决策方法. 首先,研究船舶会遇局面辨识方法,设计满足《国际海上避碰规则》的多船避碰策略. 其次,研究多船舶智能体合作方式,构建多船舶智能体协同避碰决策模型:利用注意力推理方法提取有助于避碰决策的关键数据;设计记忆驱动的经验学习方法,有效积累交互经验;引入噪音网络和多头注意力机制,增强船舶智能体决策探索能力. 最后,分别在实验地图与真实海图上,对多船会遇场景进行仿真实验. 结果表明,在协同性和安全性方面,相较于多个对比方法,所提出的避碰策略均能获得具有竞争力的结果,且满足实用性要求,从而为提高船舶智能航行水平和保障航行安全提供一种新的解决方案.

关键词: 多智能体深度强化学习；多智能体通信模型；多智能体合作；协同决策；船舶避碰；协同避碰策略

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2022.1159

引用格式: 隋丽蓉, 高曙, 何伟. 基于多智能体深度强化学习的船舶协同避碰策略 [J]. 控制与决策, 2023, 38(5): 1395-1402.

Ship cooperative collision avoidance strategy based on multi-agent deep reinforcement learning

SUI Li-rong[1], *GAO Shu*[1†], *HE Wei*[2]

(1. School of Computer Science and Artificial Intelligence，Wuhan University of Technology，Wuhan 430063，China；
2. College of Physics Electronic Information Engineering，Minjiang University，Fuzhou 350108，China)

Abstract: Ship collision avoidance is the primary issue in intelligent navigation. In multi-ship encounters, only by collaborating and jointly planning collision avoidance strategies, the collision risk can be effectively reduced. In order to make the ship intelligent collision avoidance strategy collaborative, safe and practical, a ship collaborative collision avoidance decision method based on multi-agent deep reinforcement learning is proposed. Firstly, the method of identifying ship encounter situations is studied and a multi-ship collision avoidance strategy that satisfies the "International regulations for preventing collisions at sea" is designed. Secondly, by analysing the cooperation mode of multi-ship agents, a multi-ship agent cooperative collision avoidance decision-making model is constructed. The model uses the attention inference method to extract the key data that is helpful for collision avoidance decisions. And a memory driven experience learning method is designed to effectively accumulate interactive experience. In addition, the noise network and multi-head attention mechanism are introduced into the model to enhance decision-making and exploration capabilities of ship agents. Finally, on the experimental map and the real nautical chart, simulation experiments are carried out on the multi-ship encounter scenarios. The results show that in terms of collaboration and safety, compared with multiple comparison methods, competitive results are obtained and the practical requirements are met using the proposed method, which provides a new solution for improving theintelligent navigation of ships and ensuring navigation safety.

Keywords: multi-agent deep reinforcement learning；multi-agent communication model；multi-agent cooperation；collaborative decision-making；ship collision avoidance；collaborative collision avoidance strategy

收稿日期: 2022-07-01；**录用日期:** 2022-12-20.
基金项目: 绿色智能内河创新国家重大科技专项项目(工信部装函(2019))；国家自然科学基金项目(52172327).
责任编委: 杨涛.
†通讯作者. E-mail: 455125430@qq.com.

0 引 言

在实际航行中,尤其在港口或江海交汇处,多船会遇、交叉等情况十分常见,此时船舶间只有相互协作,在《国际海上避碰规则》(international regulations for preventing collisions at sea,简称"COLREGS规则")的约束下,共同合理规划避碰策略,才能有效应对复杂环境,达到协同避碰的目的,从而降低碰撞风险.近年来,人工智能技术的飞速发展,促进了船舶避碰的智能化研究. 倪生科[1]改进了遗传算法,基于协同和排队理论构建协调避让机制,设计多船避碰策略. 周双林等[2]从单船的角度出发,利用深度Q网络算法控制船舶智能避碰,保障其安全航行. Zhao等[3]利用深度神经网络将遇到船舶的状态映射为本船的转向命令,为多艘船舶制定避碰决策. Chun等[4]提出了一种基于深度强化学习的避碰算法以确定避让时间,并针对最危险的船舶生成符合COLREGS规则的避让路径. Chen等[5]基于深度Q网络强化学习方法,模拟两船间的协同避碰关系. 周怡等[6]利用深度确定性策略梯度算法实现船舶智能避碰.综上,目前现有基于深度强化学习的船舶避碰研究较少考虑船舶间相互影响,避碰策略缺乏协同性,往往使船舶安全避过当前危险局面,却又陷入另一紧迫局面.

多智能体深度强化学习通过智能体间的协同合作或对抗竞争完成复杂任务,在道路交通协同控制[7]、多无人机协同安全飞行[8]等方面取得了良好效果. 鉴于多智能体深度强化学习可描述为马尔可夫博弈过程,与多船避碰过程具有相似性,均为一系列决策过程,本文将船舶映射为智能体,利用多智能体深度强化学习方法解决多船协同避碰问题. 通过分析多船舶智能体合作方式,构建多船舶智能体协同避碰决策模型,同时在协同避碰决策中引入COLREGS规则,从而在船舶应对复杂会遇局面时,保证避碰决策的协同性、安全性以及实用性,为多船协同避碰提供新的解决方案.

1 问题描述及形式化表示

在船舶避碰研究中,多船会遇局面下避碰决策一直是国内外研究人员关注的焦点. 具体是指会遇局面下,各船舶自主制定避让方案,共同规划避碰策略,以保证避碰决策的安全性、协同性与实用性.

1.1 船舶领域及船舶碰撞风险评估模型

船舶领域是指船舶周围一定范围内的水域,该水域范围内不允许其他船舶侵入. 依据国内外相关研究,选取椭圆形船舶领域构建领域模型,保证船舶避碰的安全性.

为充分利用船舶助航设备获取的数据,本文采用文献[9]中碰撞危险评估方法,利用最终碰撞危险和TCPA (time to closest point of approach)构建船舶碰撞风险评估模型.

1.2 规则约束下多船避碰策略

由于多船会遇局面可分解为多个两船会遇局面,通过辨识两船的会遇局面类别(对遇、交叉和追越)以及船舶角色(直行船或让路船)规划避碰策略,本文采用文献[10]的两船会遇局面辨识方法,融合多因素:限制线(包括船艏限制线和船艉限制线)、船舶航向角及相对方位,辨识船舶会遇局面类别(对遇、交叉和追越),从而提高船舶会遇局面辨识精度. 同时借鉴文献[2]的思想,结合COLREGS规则设计多船避碰策略,具体描述如下(以船舶i为例).

step 1: 计算船舶i与周围船舶j($j\in\{1,\dots,i-1,i+1,\dots,n\}$,$n$为$i$周围船舶数量,包括$i$)的最终碰撞危险和TCPA (见1.1节).

step 2: 若船舶i与j存在最终碰撞危险且TCPA >0,则进行会遇局面辨识.

step 3: 若i周围$n-1$艘船舶未全部完成会遇局面辨识,则转入step 1;否则执行step 4.

step 4: 若辨识船舶i在与任一周围船舶j形成的会遇局面中承担让路船角色(无论是否也承担直行船角色),则船舶i避碰策略均为右转让路,即使为追越局面,也保证右转.

1.3 多船避碰问题的形式化表示

多船避碰过程表现为多艘船舶根据感知信息及其驾驶员积累的航行经验各自规划并执行避碰策略,这与马尔科夫博弈过程相契合. 因此,本文将船舶映射为智能体,基于多智能体深度强化学习,结合船舶航行和避碰特点研究船舶智能体避碰策略,为多船避碰问题提供新的解决思路.

充分考虑船舶的航行目标和实际航行环境中船舶运动的连续性特征,利用航速、航向、目标点与自身的相对位置及周围其他船舶智能体与自身的相对位置构建连续的状态空间o_t^i,基于航速和航向设计连续动作空间$a_t^i=(u_t^i,v_t^i)$. 其中:u_t^i和v_t^i分别表示船舶智能体i在时间步t时的动作V(包括航向和航速)在x轴上的分量和在y轴上的分量.

同时,综合船舶航行目标及避碰安全性设计奖励函数. 以船舶智能体到达目标点的距离作为航行奖励值,且当船舶智能体(包括船舶智能体之间、船舶智

能体与静态障碍物之间)发生碰撞时,给予负面奖励值.其中,船舶智能体间的碰撞以船舶智能体间船舶领域是否重叠为依据(见1.1节).

2　基于多智能体深度强化学习的船舶协同避碰决策

2.1　模型架构

图1为本文构建的多船舶智能体协同避碰决策模型架构,该模型主要由船舶智能体(限于篇幅,以3个船舶智能体为例,且仅画出两个船舶智能体具体结构)、样本缓冲池和先验缓冲池组成.每个船舶智能体包括行动者和评论家两部分(具体结构设计见2.2节),并且拥有独立的样本缓冲池及共享的先验缓冲池.行动者由优化器、现实先验网络、现实消息编码网络、基于噪音网络的现实策略网络、目标网络(包括目标先验网络、目标消息编码网络及目标策略网络,因篇幅所限,图1中没有标识)组成,根据自身的局部观测数据(即本船航行数据)和来自其他船舶智能体的通信数据(即他船航行数据)规划协同避碰策略;评论家由优化器、基于多头注意力机制的现实评价网络、目标评价网络组成,基于样本数据评估策略好坏,指导船舶智能体制定最优的协同避碰决策;缓冲池(包括样本缓冲池和先验缓冲池)存储样本,用于自身训练.

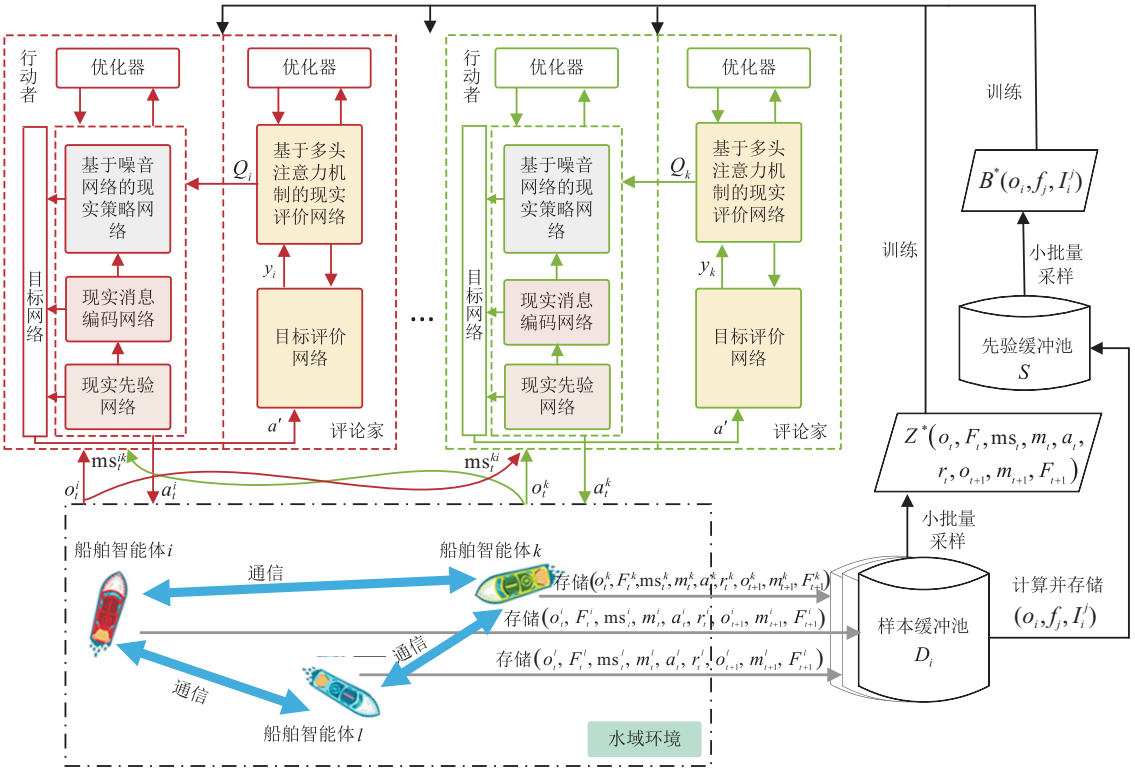


图1　多船舶智能体协同避碰决策模型架构

2.2　面向协同决策的多船舶智能体通信合作方法

多智能体通信研究[11-13]中主要解决智能体通信的时机、通信的对象、通信的内容以及通信内容的处理4个基本问题.独立推理通信框架(individually inferred communication,简称“I2C框架”)[12],通过学习智能体间通信的先验知识确定通信时机和通信对象,利用深度神经网络处理通信内容,是一种简单而有效的通信框架.I2C框架中,智能体仅与可视范围内的其他智能体通信,而船舶受目前通信设备性能所限,仅能采集一定范围内其他船舶的航行数据,因此,选择I2C框架作为船舶智能体协同合作的基础架构.

船舶在会遇局面下需充分考虑相互影响,根据他船对自身的影响程度制定避碰策略.I2C框架虽然通过学习先验知识确定利用周围哪些船舶智能体的通信数据,但对这些被利用的通信数据未进行差异化处理.另外,I2C框架仅根据当前相关数据(即当前时间步下自身局部观测数据和来自周围船舶智能体的通信数据)进行动作决策,没有考虑过往经验的指导作用.因此,本文借鉴重点船避碰[1]思想,利用注意力推理方法,量化其他船舶智能体对当前船舶智能体策略的影响,推理关键通信数据am_t^i,促进船舶智能体的合作避碰;同时受文献[11]和文献[14]基于内存的通信信息表示的启发,基于过往经验m_{t-1}^i进一步探索船舶智能体学习通信数据、累积经验m_t^i的方法.另外,

通过现实先验网络缩减通信数据范围,将来自周围船舶智能体的通信数据(即周围船舶的航行数据)与自身局部观测数据(即本船航行数据)共同作为动作决策依据,促进多船舶智能体合作以完成避碰任务.

2.2.1 注意力推理方法

首先,将船舶智能体i的局部观测数据映射为查询向量q_t^i,将自身局部观测数据及来自周围船舶智能体的通信数据映射为键值向量K_t^i、V_t^i,如下式所示:

$$q_t^i, K_t^i, V_t^i = \varsigma_{\theta_i^{\rm ae}}^{\rm ae}(\mathrm{ms}_t^{i1}, \ldots, \mathrm{ms}_t^{ii-1}, o_t^i, \mathrm{ms}_t^{ii+1}, \ldots, \mathrm{ms}_t^{in}). \tag{1}$$

其中:$\varsigma_{\theta_i^{\rm ae}}^{\rm ae}$表示以$\theta_i^{\rm ae}$为参数的多层神经网络;$o_t^i$为船舶智能体$i$在时间步$t$时的局部观测数据;$\mathrm{ms}_t^{i1}, \ldots, \mathrm{ms}_t^{ii-1}, \mathrm{ms}_t^{ii+1}, \ldots, \mathrm{ms}_t^{in}$为$i$获取的来自周围船舶智能体的通信数据;$n$为与船舶智能体$i$通信的船舶智能体数量(包括$i$).

其次,通过查询向量和键向量的Hadamard Product计算船舶智能体间的相互影响qi_t^{ij},再应用线性变换生成定义特定船舶智能体权重的标量qs_t^{ij},如下式所示:

$$\mathrm{qi}_t^{ij} = q_t^i \odot k_t^j,\ k_t^j \in K_t^i,\ \mathrm{qi}_t^{ij} \in \boldsymbol{R}^{d_q}; \tag{2}$$

$$\mathrm{qs}_t^{ij} = W_{iq}^{[d_q \times d_1]} \mathrm{qi}_t^{ij},\ \mathrm{qs}_t^{ij} \in \boldsymbol{R}^1. \tag{3}$$

同时,利用softmax操作生成特定于船舶智能体的权重值W_t^i,即

$$W_t^i = \mathrm{softmax}\Big[\frac{\mathrm{qs}_t^{i1}}{\sqrt{d_q}}, \frac{\mathrm{qs}_t^{i2}}{\sqrt{d_q}}, \ldots, \frac{\mathrm{qs}_t^{in}}{\sqrt{d_q}}\Big],\ W_t^i \in \boldsymbol{R}^n. \tag{4}$$

最后,基于注意力向量W_t^i计算从自身局部观测数据和来自周围船舶智能体的通信数据中获取的关键数据am_t^i,即

$$\mathrm{am}_t^i = \sum_{j=1}^{n} w_t^{ij} v_t^{ij},\ w_t^{ij} \in W_t^i;$$

$$\mathrm{am}_t^i \in \boldsymbol{R}^{d_{\rm ms}},\ v_t^{ij} \in V_t^i,\ j \in \{1, 2, \ldots, n\}. \tag{5}$$

其中:w_t^{ij}为船舶智能体j特定于i的权重值;v_t^{ij}为i周围其他船舶智能体j的通信数据的编码值,若$j = i$,则v_t^{ii}为i自身局部观测数据的编码值.

2.2.2 记忆驱动的经验学习方法

船舶驾驶员的航行经验是指导船舶避碰决策的重要依据之一,因此,积累船舶智能体的交互经验,并结合经验信息制定避碰决策,有助于规划更好的避碰策略. 由此,为每个船舶智能体增加一个内存空间,捕获船舶智能体在与环境交互过程中收集、学习的知识,定义为经验信息$m \in \boldsymbol{R}^M$. 它由船舶智能体通过对自身局部观测数据(即本船航行数据)和来自其他船舶智能体的通信数据(即他船航行数据)进行学习获得的,具体方法如下.

1) 编码船舶智能体i的局部观测数据及来自周围船舶智能体的通信数据,将其映射为编码向量E,即

$$E = \{\mathrm{ec}_t^1, \ldots, \mathrm{ec}_t^{i-1}, \mathrm{ec}_t^i, \mathrm{ec}_t^{i+1}, \ldots, \mathrm{ec}_t^n\} = \varsigma_{\theta_i^{\rm ec}}^{\rm enc}(\mathrm{ms}_t^{i1}, \ldots, \mathrm{ms}_t^{ii-1}, o_t^i, \mathrm{ms}_t^{ii+1}, \ldots, \mathrm{ms}_t^{in}), \tag{6}$$

其中$\varsigma_{\theta_i^{\rm ec}}^{\rm enc}$是以$\theta_i^{\rm ec}$为参数的全连接网络.

2) 结合船舶智能体的编码向量E和当前船舶智能体上一时间步的经验信息m_{t-1}^i,在编码向量和上一步的经验信息之间进行相对推理,即

$$\mathrm{mi}_t^{ij} = m_{t-1}^i \odot \mathrm{ec}_t^j,\ \forall \mathrm{ec}_t^j \in E \mathrm{mi}_t^{ij},\ m_{t-1}^i \in \boldsymbol{R}^{d_v}. \tag{7}$$

其中:j为船舶智能体i的通信对象(包括i);mi_t^{ij}为通信船舶智能体i相对于经验信息m_{t-1}^i的影响值;$\odot$为合成算子,采用逐元素乘法的计算方式.

3) 利用前馈神经网络聚合mi_t^{ij}与当前编码向量E,生成经验信息m_t^i,存入内存空间,如下式所示:

$$m_t^i = W_m^{[d_{2v} \times d_v]}[\mathrm{mi}_t^{ij} + \mathrm{ec}_t^j],\ m_t^i \in \boldsymbol{R}^{d_v},\ \mathrm{ec}_t^j \in E, \tag{8}$$

其中$W_m^{[d_{2v} \times d_v]}$表示前馈神经网络.

另外,引入噪音网络,将船舶智能体的现实策略网络的全连接层参数化为噪音网络,其噪音分布采用分解高斯噪音分布,以增强船舶智能体决策探索能力,促进船舶智能体探索更多有助于协同避碰的样本,提高找到最佳协同避碰策略的概率;同时,设计基于多头注意力机制的现实评价网络,量化不同船舶智能体信息的重要性大小,提取来自不同船舶智能体编码向量中的重要信息以计算Q值,促进船舶智能体关注更有利于自己学习更好避碰策略的信息,促进船舶智能体协同避碰,从而提高船舶避碰的安全性.

综上,图1水域环境中船舶智能体i、k、l可以互相通信,收集来自其他船舶智能体的通信数据(即其他船舶的航行数据),通过有效处理通信数据,增强船舶智能体避碰决策的协同性(为了更加清晰直观地呈现船舶智能体间通信合作的过程,图1以船舶智能体i和k为例,ms_t^{ik}和ms_t^{ki}分别为船舶智能体i获取的来自船舶智能体k的通信数据以及船舶智能体k获取的来自船舶智能体i的通信数据):各船舶智能

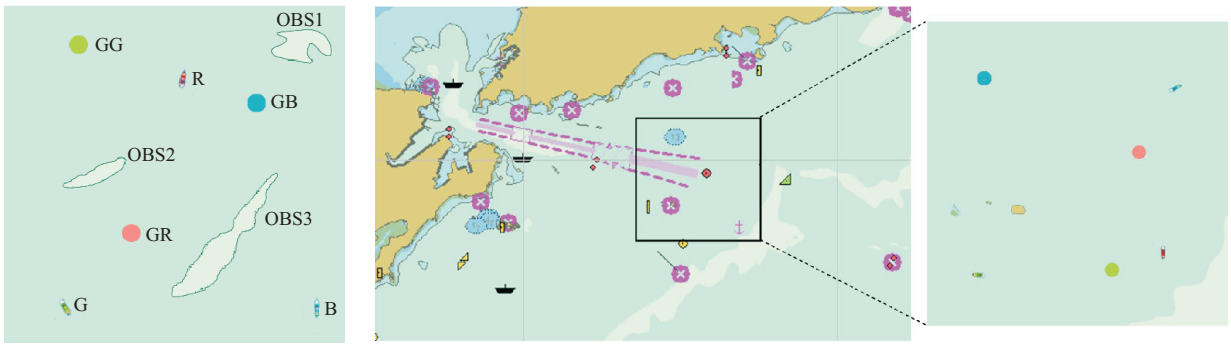
体通过现实先验网络缩减通信数据范围;利用现实消息编码网络提取关键通信数据,积累交互经验,充分利用自身局部观测数据(即本船的航行数据)和来自周围船舶智能体的通信数据(即周围船舶的航行数据),生成协同决策依据;通过基于噪音网络的现实策略网络生成协同避碰策略,利用噪音扰动增强船舶智能体探索能力,提高船舶智能体发现最优协同避碰策略的概率. 各船舶智能体与环境交互组合样本 $(o_t^i, F_t^i, \mathrm{ms}_t^i, m_t^i, a_t^i, r_t^i, o_{t+1}^i, m_{t+1}^i, F_{t+1}^i)$(以船舶智能体 i 为例, o_t^i 和 o_{t+1}^i 分别为 i 在时间步 t 和 $t+1$ 时的局部观测数据, F_t^i 和 F_{t+1}^i 分别为 i 在时间步 t 和 $t+1$ 时抽取的周围船舶智能体的特征, ms_t^i 为 i 在时间步 t 获取的来自其他船舶智能体的通信数据, m_t^i 和 m_{t+1}^i 分别为 i 在时间步 t 和 $t+1$ 时的经验信息, r_t^i 为 i 在时间步 t 时的奖励值),存入各自的样本缓冲池 D_i,每隔固定的时间周期抽取小批量样本,基于多头注意力机制评估策略优劣,有选择地关注不同船舶智能体的策略信息,指导船舶智能体协同合作,保证多船避碰的协同性和安全性.

3　实验结果与分析

3.1　实验环境设置

3.1.1　实验环境

为了验证和评估本文所提出的船舶协同避碰决策模型的性能,保证其既具有普适性,也具有一定的实用性,本文设计两种实验地图:随机实验地图和基于实际电子海图的实验地图,如图2所示. 两种地图皆包含船舶(R、G、B)、船舶航行起点(船舶初始所在位置)、静态障碍物(岛屿、礁石等: OBS1、OBS2和OBS3)和目标点(GR、GG、GB,圆形表示)这4种元素. 其中:随机实验地图静态障碍物数量及位置均随机设置;基于实际电子海图的实验地图取自青岛港附近一定范围内水域.



(a)　随机实验地图　　(b)　基于实际电子海图的实验地图

图2　实验地图示意

3.1.2　参数设置

本实验基于Tensorflow框架,使用Gym、Numpy、Pygame库进行开发,编程语言为Python 3.6.2. 经多次实验调参,算法主要参数设置: buffer_size(缓冲池大小)为1 000 000, batch_size(批量采样大小)为1 024, learning_rate(学习率)为0.001, num_unit(网络单元数)为64, prior_batch_size(先验采样数量)为400, prior_buffer_size(先验缓冲池大小)为60 000, memory_size(内存空间)为10, δ(阈值)为0.5.

3.1.3　评价指标

选取协同性、安全性和实用性3项指标衡量实验结果. 当多智能体深度强化学习应用于合作场景时,智能体以累积最大奖励值为目标,协同合作完成任务[11-12,15]. 因此,统计每1 000个episode的平均累积奖励值、后20 000个episode的平均累积奖励值及其标准差,以衡量算法协同性. 同时统计若干次船舶避碰实验中,避碰决策模型规划出的避碰策略发生船舶碰撞的次数,以衡量避碰方法的安全性. 另外,设计多个船舶会遇局面仿真实验地图,通过记录船舶避碰位置点绘制路径图,可视化避碰效果,判断规划的避碰策略在多船会遇局面下是否满足COLREGS规则,以验证模型的实用性.

3.1.4　对比模型

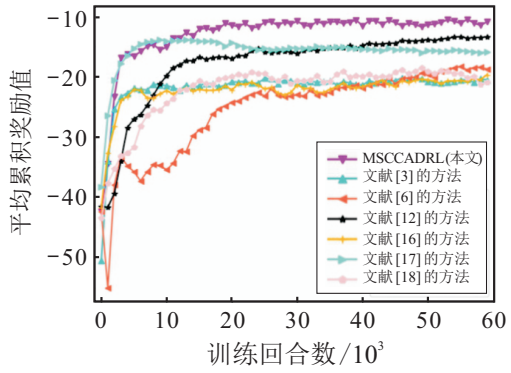
使用文献[3, 6, 12, 16-18]作为对比模型,与本文所提出的多船舶智能体协同避碰决策模型进行比较(为了对比的公平性,这里抽取相应文献中关键算法,设置相同的状态空间、动作空间、奖励函数,并调整相关参数,应用到本文的实验环境中).

3.2　实验及结果分析

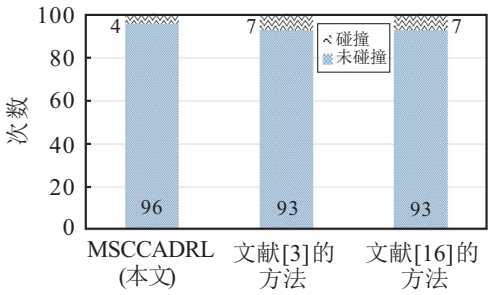
3.2.1　实验1(协同性评估实验)

为评估所提出的多船舶智能体协同避碰决策模型(multi-ship collaborative collision avoidance

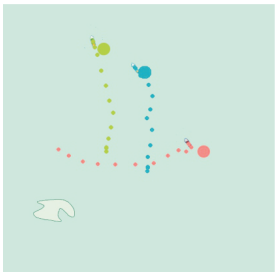
decision-making model based on multi-agent deep reinforcement learning, MSCCADRL) 的协同性, 按照 3.1.1 节的方法搭建随机实验地图, 将本文的 MSCCADRL 与文献 [3, 6, 12, 16-18] 的方法进行对比实验, 统计相同 episode 时的平均累积奖励值. 训练奖励可视化结果如图 3(a) 所示, 后 20 000 个 episode 的平均累积奖励值及其标准差统计结果如表 1 所示. 综合图 3(a) 和表 1 可以看出, MSCCADRL 的平均累积奖励值最高. 实验结果表明, MSCCADRL 进一步增强了船舶避碰的协同性.



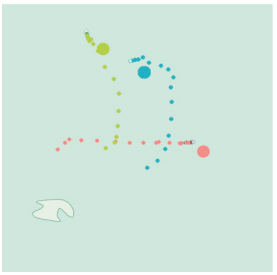
(a) 训练奖励对比曲线



(b) 船舶碰撞次数统计结果

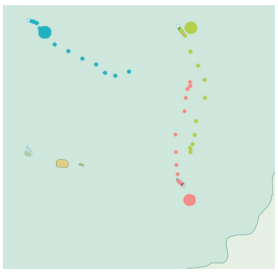


(c1) MSCCADRL(本文) 可视化效果

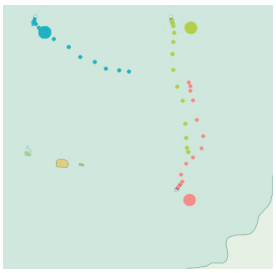


(c2) MSCCADRL/COLREGS 可视化效果

(c) 场景 1(随机实验地图)



(d1) MSCCADRL(本文) 可视化效果



(d2) MSCCADRL/COLREGS 可视化效果

(d) 场景 2(实际电子海图)

图 3 实验结果

表 1 平均累积奖励值统计结果

参数	平均累积奖励值
文献 [3] 的方法	-20.67 ± 0.33
文献 [16] 的方法	-20.82 ± 0.36
文献 [6] 的方法	-19.82 ± 1.12
文献 [17] 的方法	-15.68 ± 0.20
文献 [18] 的方法	-19.67 ± 0.67
文献 [12] 的方法	-13.99 ± 0.45
本文 MSCCADRL	$\mathbf{-11.12 \pm 0.36}$

3.2.2 实验 2 (安全性评估实验)

按照 3.1.1 节的方法搭建随机实验地图, 将本文的 MSCCADRL 与文献 [3] 和文献 [16] 的方法进行对比实验. 生成 100 张随机实验地图, 统计 100 次避碰实验中船舶碰撞次数. 船舶碰撞次数统计结果如图 3(b) 所示, 可以看出, MSCCADRL 碰撞次数小于其他对比算法, 从而验证了 MSCCADRL 的安全性相对较好.

3.2.3 实验 3 (实用性验证实验)

为了验证所提出的多船舶智能体协同避碰决策模型的实用性 (即是否满足 COLREGS 规则), 搭建随机实验地图和基于实际电子海图的实验地图 (详见 3.1 节), 构建船舶会遇局面, 执行 MSCCADRL 和不引入 COLREGS 规则的 MSCCADRL (以下简称 MSCCADRL/COLREGS), 训练船舶智能体避碰模型, 记录船舶智能体每一时间步的位置, 对避碰路径进行可视化分析.

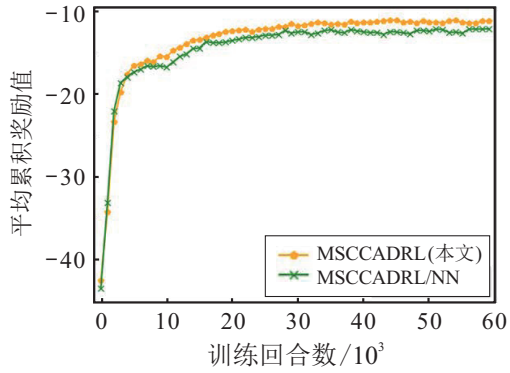
为了有效且清晰地展示结果, 以 3 条船的会遇局面为例, 避碰路径如图 3(c) 和图 3(d) 所示 (限于篇幅, 仅分别给出在随机实验地图和基于实际电子海图的实验地图中任一具有代表性的船舶避碰效果), 图 3(c1)、3(d1) 和图 3(c2)、3(d2) 分别为 MSCCADRL 和 MSCCADRL/COLREGS 的避碰路径可视化效果图. 图 3(c) 场景 1 (随机实验地图) 中, MSCCADRL 指导红色船舶右转, 绿色船舶和蓝色船舶直行, 其避让方案均符合 COLREGS 规则要求; 而 MSCCADRL/COLREGS 方法中, 红色船舶左转, 绿色船舶和蓝色船舶均右转, 其避让策略都不满足 COLREGS 规则. 图 3(d) 场景 2 (实际电子海图) 中, 图 3(d1) 中两船舶 (红色船舶和绿色船舶) 均向右转, 其避碰策略满

足COLREGS规则,同时蓝色船舶也安全靠近其目标点;而图3(d2)中绿色船舶接近直行,红色船舶左转,不符合COLREGS规则要求. 因此,实验结果表明,MSCCADRL能够规划出满足COLREGS规则的避碰策略,从而验证了其在船舶会遇局面中的实用性,为实现船舶智能航行提供了参考.

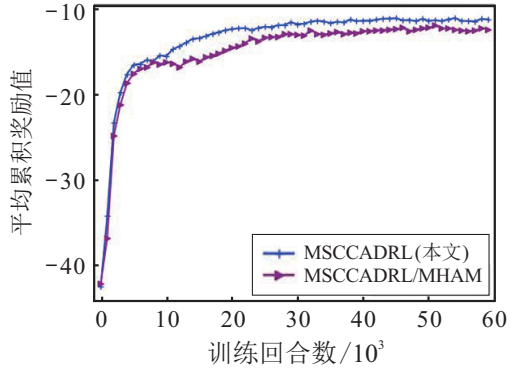
3.2.4　实验4(消融实验)

为验证噪音网络和多头注意力机制的有效性,将MSCCADRL、未融合噪音网络的MSCCADRL(以下简称MSCCADRL/NN)及未融合多头注意力机制的MSCCADRL(以下简称MSCCADRL/MHAM)进行实验. 按照3.1.1节的方法搭建随机实验地图,分别训练船舶智能体避碰模型,统计相同episode时的平均累积奖励值,进行可视化分析.

平均累积奖励值对比结果如图4所示. 对比奖励曲线可以看出,MSCCADRL平均累积奖励值略高于MSCCADRL/NN和MSCCADRL/MHAM,说明噪音网络和多头注意机制能够促进船舶智能体找到更好的避碰策略. 原因可能在于:噪音扰动增大了船舶智能体动作选择的随机性;多头注意力机制促进船舶智能体利用最有效信息进行学习,增强了学习的协同性,因而获得的奖励值更大. 实验结果分别验证了噪音网络和多头注意力机制的有效性.



(a)　噪音网络性能评估结果



(b)　多头注意力机制性能评估结果

图4　性能评估结果曲线

从上述实验可以看出:由于本文的MSCCADRL通过现实消息编码网络提取关键通信数据,学习交互经验,充分利用自身局部观测数据(即本船航行数据)及来自周围船舶智能体的通信数据(即他船航行数据),发挥过往经验的作用,增强了船舶智能体避碰决策的协同性;同时利用基于噪音网络的现实策略网络,引入噪音扰动增强了船舶智能体探索能力,鼓励船舶智能体探索更多的协同避碰样本;基于多头注意力机制设计现实评价网络,学习更有利于自己获取更大回报的策略信息,促进了船舶智能体协同合作,从而进一步增强了多船避碰的协同性与安全性. 另外,引入COLREGS规则,保证了避碰策略的实用性. 结果表明,所提出的MSCCADRL能够训练所有船舶智能体,制定相互协同的、靠近目标点且满足COLREGS规则的安全避碰策略.

4　结　论

本文基于多智能体深度强化学习方法,针对多船协同避碰问题展开了基础研究. 通过充分考虑周围船舶的航行数据协同决策,结合船舶避碰特点和注意力推理方法,设计了记忆驱动的经验学习方法,并引入噪音网络和多头注意力机制,共同构建多船舶智能体协同避碰决策模型,促进了船舶间避碰的协同性,并保证了其安全性. 另外,引入COLREGS规则,增强了避碰策略的实用性,从而为智能化辅助船舶间的协同避碰提供了新的解决思路.

考虑到船舶避碰过程中不仅受到其他船舶、岛屿、礁石等障碍物的影响,还会受到天气、风、流等因素的影响,因此,未来的研究将考虑更多的船舶避碰决策影响因素;同时将研究基于船舶通信数据的避碰策略预测方法,通过对他船策略的预测更好地规划自身避碰策略,进一步提高船舶避碰决策的安全性和协同性. 另外,应进行实船实验,以进一步验证方法的实用性.

参考文献(References)

[1]　倪生科. 基于规则的船舶智能避碰决策关键技术研究[D]. 大连: 大连海事大学, 2020: 102-117).
(Ni S K. Study on key technologies for ship intelligent decision making for collision avoidance based on rules[D]. Dalian: Dalian Maritime University, 2020: 102-117.)

[2]　周双林, 杨星, 刘克中, 等. 规则约束下基于深度强化学习的船舶避碰方法[J]. 中国航海, 2020, 43(3): 27-32.
(Zhou S L, Yang X, Liu K Z, et al. COLREGs-compliant

method for ship collision avoidance based on deep reinforcement learning[J]. Navigation of China, 2020, 43(3): 27-32.)

[3] Zhao L M, Roh M I. COLREGs-compliant multiship collision avoidance based on deep reinforcement learning[J]. Ocean Engineering, 2019, 191: 106436.

[4] Chun D H, Roh M I, Lee H W, et al. Deep reinforcement learning-based collision avoidance for an autonomous ship[J]. Ocean Engineering, 2021, 234: 109216.

[5] Chen C, Ma F, Xu X B, et al. A novel ship collision avoidance awareness approach for cooperating ships using multi-agent deep reinforcement learning[J]. Journal of Marine Science and Engineering, 2021, 9(10): 1056.

[6] 周怡, 袁传平, 谢海成, 等. 基于DDPG算法的游船航行避碰路径规划[J]. 中国舰船研究, 2021, 16(6): 19-26.
(Zhou Y, Yuan C P, Xie H C, et al. Collision avoidance path planning of tourist ship based on DDPG algorithm[J]. Chinese Journal of Ship Research, 2021, 16(6): 19-26.)

[7] 宋佰霖, 许华, 齐子森, 等. 一种基于深度强化学习的协同通信干扰决策算法[J]. 电子学报, 2022, 50(6): 1301-1309.
(Song B L, Xu H, Qi Z S, et al. A collaborative communication jamming decision algorithm based on deep reinforcement learning[J]. Acta Electronica Sinica, 2022, 50(6): 1301-1309.)

[8] 蒋明智, 吴天昊, 张琳. 基于深度强化学习的无信号交叉口车辆协同控制算法[J]. 交通运输工程与信息学报, 2022, 20(2): 14-24.
(Jiang M Z, Wu T H, Zhang L. Deep reinforcement learning based vehicular cooperative control algorithm at signal-free intersection[J]. Journal of Transportation Engineering and Information, 2022, 20(2): 14-24.)

[9] He Y X, Jin Y, Huang L W, et al. Quantitative analysis of COLREG rules and seamanship for autonomous collision avoidance at open sea[J]. Ocean Engineering, 2017, 140: 281-291.

[10] 沈海青, 郭晨, 李铁山, 等. 考虑航行经验规则的无人船舶智能避碰导航方法[J]. 哈尔滨工程大学学报, 2018, 39(6): 998-1005.
(Shen H Q, Guo C, Li T S, et al. Intelligent collision avoidance navigation method for unmanned ships considering navigation experience rules[J]. Journal of Harbin Engineering University, 2018, 39(6): 998-1005.)

[11] Rangwala M, Williams R. Learning multi-agent communication through structured attentive reasoning[C]. Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 10088-10098.

[12] Ding Z L, Huang T J, Lu Z Q. Learning individually inferred communication for multi-agent cooperation[J/OL]. 2020, arXiv: 2006.06455.

[13] Liu I J, Jain U, Yeh R A, et al. Cooperative exploration for multi-agent deep reinforcement learning[J/OL]. 2021, arXiv: 2107.11444.

[14] Pesce E, Montana G. Improving coordination in small-scale multi-agent deep reinforcement learning through memory-driven communication[J]. Machine Learning, 2020, 109(9): 1727-1747.

[15] Kuba J G, Chen R Q, Wen M N, et al. Trust region policy optimisation in multi-agent reinforcement learning[J/OL]. 2021, arXiv: 2109.11251.

[16] Hu J, Hu S Y, Liao S W. Policy regularization via noisy advantage values for cooperative multi-agent actor-critic methods[J/OL]. 2021, arXiv: 2106.14334.

[17] Iqbal S, Sha F. Actor-attention-critic for multi-agent reinforcement learning[J/OL]. 2018, arXiv: 1810.02912.

[18] Kuba J G, Chen R Q, Wen M N, et al. Trust region policy optimisation in multi-agent reinforcement learning[J/OL]. 2021, arXiv: 2109.11251.

作者简介

隋丽蓉 (1997−), 女, 硕士, 从事深度强化学习的研究, E-mail: suilirong@sina.cn;

高曙 (1967−), 女, 教授, 博士, 从事智能计算、大数据分析及其在智能交通的应用等研究, E-mail: 455125430@qq.com;

何伟 (1982−), 男, 教授, 博士, 从事智能系统与信息融合、船海装备与新能源等研究, E-mail: hewei11@mju.edu.cn.

(责任编辑：李君玲)