

同行专家业内评价意见书编号: 20240854167

附件1

浙江工程师学院（浙江大学工程师学院） 同行专家业内评价意见书

姓名: _____ 杨贺磊

学号: _____ 22160074

申报工程师职称专业类别（领域）: _____ 电子信息

浙江工程师学院（浙江大学工程师学院）制

2024年03月20日

一、个人申报

（一）基本情况【围绕《浙江工程师学院（浙江大学工程师学院）工程类专业学位研究生工程师职称评审参考指标》，结合该专业类别(领域)工程师职称评审相关标准，举例说明】

一、对本专业基础理论知识和专业技术知识掌握情况

深入了解并掌握本专业的基础理论知识和专业技术知识。在基础理论知识方面，具备扎实的数学基础，包括线性代数、概率论、微积分和数值分析等，熟练掌握电路分析的基本原理和技术，对信号的时间域和频率域分析有深入的理解，能够熟练应用拉普拉斯变换、傅立叶变换等工具分析和设计控制系统。在专业技术知识方面，深入学习了控制理论的核心概念，包括经典控制理论（如PID控制）和现代控制理论（如状态空间分析、最优控制、鲁棒控制等）。能够应用这些理论设计、分析和优化控制系统。同时，具有将实际控制对象抽象成数学模型的能力，并能使用软件工具（如MATLAB/Simulink）进行系统仿真，以验证控制策略的有效性。在智能控制方面，掌握了机器学习的基本理论和算法，包括监督学习、非监督学习、强化学习等，以及它们在智能控制中的应用。

二、工程实践的经历

工程实践围绕多智能体集群博弈问题展开研究，将带约束的多智能体集群对抗问题抽象为多追逐者多逃避者的集群追逃避障问题，基于多智能体深度强化学习算法对该问题进行深入研究。

在工程实践期间，学习深度强化学习的基本概念和常见算法，了解多智能体强化学习的博弈应用场景，学习移动机器人路径规划算法，编写环境和多智能体强化学习算法实现多追逐者多逃避者的追逐逃避场景下的任务。同时，在实践期间与项目负责人进行沟通和协调，完成整体方案设计、算法设计、仿真实验和机器人实物实验。

三、在实际工作中综合运用所学知识解决复杂工程问题的案例

在专业实践的实际工作中，我参与了一个围绕多智能体深度强化学习的研究项目，旨在探索和解决多移动机器人在执行复杂任务时的控制问题。特别是，我们聚焦于“多追逐者-多逃避者”（pursuer-evader）的追逃策略问题，这是一个经典的多智能体协同和对抗问题，具有广泛的应用背景，如自动驾驶车辆的避障、无人机群的协同控制等。

项目背景与挑战

多移动机器人系统在执行追逃任务时，需要解决两个主要问题：如何有效地协同控制多个追逐者（pursuers）以捕捉逃避者（evaders），同时优化逃避者的逃逸策略以避免被捕捉。这个问题的复杂性在于需要同时考虑多智能体之间的交互、环境不确定性以及策略的实时优化。

解决方案

为了解决这一问题，我们采用了基于多智能体深度强化学习（MADRL）的方法。具体来说，我们设计了一个分布式学习框架，其中每个智能体（无论是追逐者还是逃避者）都能够独立学习并优化其策略，同时通过共享的环境信息实现协同。

1.

环境建模：我们首先建立了一个仿真环境，模拟多追逐者和多逃避者在一个有障碍物的区域内的追逃行为。每个智能体的状态信息包括其位置、速度和周围环境的感知信息。

2.

强化学习算法设计：我们为每个智能体设计了一个深度强化学习模型，使用深度神经网络作为策略网络，来学习如何根据当前状态做出决策。追逐者的目标是最小化捕捉逃避者所需的时间，而逃避者则试图最大化逃逸时间。

3.

协同与对抗学习：通过设计奖励函数促进智能体之间的协同与对抗学习。对追逐者而言，当它们成功捕捉到逃避者时获得正奖励；对逃避者而言，成功逃逸或延长逃逸时间会获得正奖励。

4.

实验与优化：我们在仿真环境中进行了大量实验，不断调整学习率、奖励函数和策略网络结构，以找到最优的学习策略。通过这些实验，智能体能够学习到复杂的协同和逃避策略，如追逐者学会了分工合作以包围逃避者，而逃避者则学会利用环境障碍物进行躲避。在此基础上，对策略的神经网络结构进行优化设计，能够适应环境中实体数量变化的情景，增强了策略的泛化性、鲁棒性和可迁移性。

成果与应用

通过该项目，我们成功地开发出一套能够处理多智能体协同与对抗问题的深度强化学习框架，并在“多追逐者-

多逃避者”的追逃策略问题上取得了显著的成果。我们的方法不仅提高了追逐者捕捉逃避者的效率，同时也增强了逃避者的逃逸能力，展现了多智能体系统在复杂环境下的自适应与智能行为。

此外，该研究的成果具有广泛的应用前景，如自动驾驶系统中的避障策略、无人机群的协同控制、以及安全防御系统中的入侵检测与反应策略等。通过进一步的研究和开发，我们相信多智能体深度强化学习技术将在未来的智能系统中发挥更大的作用。

(二) 取得的业绩(代表作)【限填3项, 须提交证明原件(包括发表的论文、出版的著作、专利证书、获奖证书、科技项目立项文件或合同、企业证明等)供核实, 并提供复印件一份】

1. 公开成果代表作【论文发表、专利成果、软件著作权、标准规范与行业工法制定、著作编写、科技成果获奖、学位论文等】

成果名称	成果类别 [含论文、授权专利(含发明专利申请)、软件著作权、标准、工法、著作、获奖、学位论文等]	发表时间/授权或申请时间等	刊物名称/专利授权或申请号等	本人排名/总人数	备注
基于强化学习的多机器人围捕多目标的分布式决策方法	发明专利申请	2022年08月15日	申请号: 202210974382.2	2/3	
Large Scale Pursuit-Evasion Under Collision Avoidance Using Deep Reinforcement Learning	会议论文	2023年06月30日	IEEE/RSJ International Conference on Intelligent Robots and Systems	1/5	

2. 其他代表作【主持或参与的课题研究项目、科技成果应用转化推广、企业技术难题解决方案、自主研发设计的产品或样机、技术报告、设计图纸、软课题研究报告、可行性研究报告、规划设计方案、施工或调试报告、工程实验、技术培训教材、推动行业发展中发挥的作用及取得的经济社会效益等】

(三) 在校期间课程、专业实践训练及学位论文相关情况	
课程成绩情况	按课程学分核算的平均成绩： 84 分
专业实践训练时间及考核情况(具有三年及以上工作经历的不作要求)	累计时间： 1.1 年(要求1年及以上) 考核成绩： 86 分(要求80分及以上)
本人承诺	
个人声明：本人上述所填资料均为真实有效，如有虚假，愿承担一切责任，特此声明！	
申报人签名：杨娟	

浙江大学研究生院

攻读硕士学位研究生成绩表

学号: 22160074	姓名: 杨贺磊	性别: 男	学院: 工程师学院	专业: 电子信息	学制: 2.5年						
毕业时最低应获: 24.0学分	已获得: 25.0学分			入学年月: 2021-09	毕业年月: 2024-03						
学位证书号: 1033532024602155	毕业证书号: 103351202402600381			授予学位: 电子信息硕士							
学习时间	课程名称	备注	学分	成绩	课程性质	学习时间	课程名称	备注	学分	成绩	课程性质
2021-2022学年冬季学期	研究生英语		2.0	75	公共学位课	2021-2022学年春季学期	智能移动机器人技术		2.0	65	跨专业课
2021-2022学年秋季学期	中国特色社会主义理论与实践研究		2.0	92	公共学位课	2021-2022学年春季学期	自然辩证法概论		1.0	77	公共学位课
2021-2022学年冬季学期	工程伦理		2.0	93	公共学位课	2021-2022学年夏季学期	制造物联网技术		2.0	95	专业选修课
2021-2022学年秋季学期	数据分析的概率统计基础		3.0	82	专业选修课	2021-2022学年夏季学期	机器人智能控制		3.0	89	专业学位课
2021-2022学年冬季学期	研究生论文写作指导		1.0	94	专业学位课	2021-2022学年春季学期	工程技术发展前沿		2.0	90	专业学位课
2021-2022学年冬季学期	智能工业机器人		2.0	81	专业学位课	2022-2023学年春季学期	研究生英语基础技能		1.0	91	公共学位课
2021-2022学年春季学期	人工智能制造技术		2.0	93	专业学位课						

说明: 1. 研究生课程按三种方法计分: 百分制 (通过、不通过), 两级制 (及格、不及格), 五级制 (优、良、中、及格、不及格)。

2. 备注中“*”表示重修课程。

学院成绩校核章:

成绩校核人: 张梦依

打印日期: 2024-04-02



(12) 发明专利申请

(10) 申请公布号 CN 115220458 A

(43) 申请公布日 2022.10.21

(21) 申请号 202210974382.2

(22) 申请日 2022.08.15

(71) 申请人 浙江大学

地址 310000 浙江省杭州市西湖区余杭塘路866号

(72) 发明人 刘勇 杨贺磊 曹军杰

(74) 专利代理机构 杭州泓呈祥专利代理事务所
(普通合伙) 33350

专利代理师 张婵婵

(51) Int.Cl.

G05D 1/02 (2020.01)

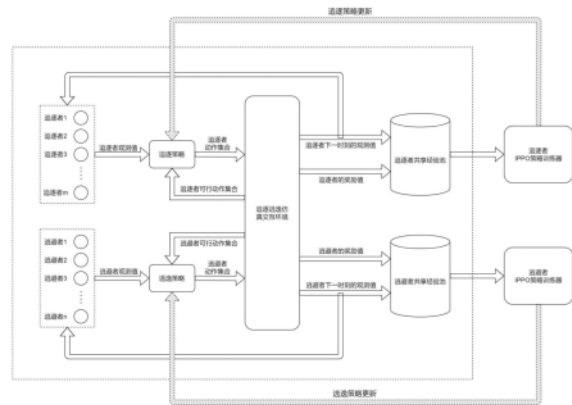
权利要求书2页 说明书10页 附图2页

(54) 发明名称

基于强化学习的多机器人围捕多目标的分布式决策方法

(57) 摘要

基于强化学习的多机器人围捕多目标的分布式决策方法,包括:一、仿真环境初始化,随机生成障碍物、追逐者和逃跑者的状态信息,追逐者和逃跑者都为智能体的机器人;二、获取机器人的观测值信息;三、获取机器人的可行动作,遍历其动作空间得到可行动作集;四、策略神经网络根据观测值信息从当前状态的可行动作集中选择一个动作;五、仿真环境根据选择出的动作对机器人进行更新并计算执行该动作获得的奖励;六、将四-五的决策过程存储到经验收集池中;七、重复二-六直达到单轮最大的仿真时间;八、根据存储器中的各个智能体与仿真环境的交互信息,使用多智能体强化学习算法进行训练;九、重复一-八直达到最大的交互训练轮数。



Large Scale Pursuit-Evasion Under Collision Avoidance Using Deep Reinforcement Learning

Helei Yang¹, Peng Ge¹, Junjie Cao^{1,*}, Yifan Yang¹, and Yong Liu^{1,*}

Abstract—This paper examines a pursuit-evasion game (PEG) involving multiple pursuers and evaders. The decentralized pursuers aim to collaborate to capture the faster evaders while avoiding collisions. The policies of all agents are learning-based and are subjected to kinematic constraints that are specific to unicycles. To address the challenge of high dimensionality encountered in large-scale scenarios, we propose a state processing method named Mix-Attention, which is based on Self-Attention. This method effectively mitigates the curse of dimensionality. The simulation results provided in this study demonstrate that the combination of Mix-Attention and Independent Proximal Policy Optimization (IPPO) surpasses alternative approaches when solving the multi-pursuer multi-evader PEG, particularly as the number of entities increases. Moreover, the trained policies showcase their ability to adapt to scenarios involving varying numbers of agents and obstacles without requiring retraining. This adaptability showcases their transferability and robustness. Finally, our proposed approach has been validated through physical experiments conducted with six robots.

I. INTRODUCTION

The Pursuit-Evasion problem holds significant potential for applications in both civilian [1] and military [2]–[4] domains. A growing body of literature recognizes the multi-agent PEG since the cooperation of autonomous decision-making pursuers can increase the task success rate. While traditional control theories and optimization-based methods have proven effective in scenarios with multiple pursuers and a single evader [5]–[7], they encounter difficulties when modeling complex real-world environments.

Drawing upon the technique of learning from agents' interactions with the environment, Multi-Agent Deep Reinforcement Learning (MADRL) has been successfully applied to the PEG. However, several drawbacks still persist in current works: 1) Single Evader [8]–[11]. Advanced assignment and collaboration strategies of pursuers are crucial for environments that involve multiple evaders. 2) Reliance on rule-based or pre-trained evasion policies. Such policies adopted by the evaders can introduce the risk of overfitting in the pursuers' policy. 3) Insufficient emphasis on collision avoidance [8]–[11]. Insufficient attention has been given in previous studies to the exploration of collision avoidance between robots and obstacles as a necessary constraint, resulting in the game continuing even after a collision occurs. 4) Limited generalization and scalability. Most previous

works require retraining specific policies to accommodate varying numbers of pursuers or evaders due to the fixed dimension of the input.

This paper investigates a decentralized scenario involving slower pursuers and faster evaders. Both pursuers and evaders adhere to unicycle kinematic constraints. It is assumed that the states of all agents are observable, and obstacles are distributed randomly throughout the environment. When an agent collides with an obstacle, it becomes immobile and loses its ability to interact with the environment. Successful pursuit is defined as the pursuer and evader within a pre-defined capture radius. The proposed scenarios in this study introduce new challenges to the algorithm, including target assignment, obstacle avoidance, and target guidance.

To address the limitation of low generalization and scalability caused by the fixed input dimension, we introduce a state processing technique called Mix-Attention based on Self-Attention. This method provides a compact feature representation for MADRL and is non-parametric in the number of agents and obstacles. We combine Mix-Attention with the widely-used multi-agent reinforcement learning algorithm IPPO [12] to handle the scenario involving multiple pursuers and evaders. Taking inspiration from [13], we train the pursuit and evasion policies synchronously to facilitate the complex co-evolution of policies.

In our simulation experiments, we evaluate the performance of our approach in various scenarios with different numbers of entities. We measure the success rate, travel distance, pursuit cost, and per-evader cost. Additionally, we assess the generalization capability of the trained policy by testing it in environments with different numbers of entities from the training setup. The experimental results demonstrate that our method significantly outperforms other approaches.

The main contributions can be summarized as follows:

- 1) The paper considers collision avoidance and motion constraints within a decentralized multi-pursuer multi-evader pursuit-evasion game.
- 2) A novel state processing method named Mix-Attention is introduced and combined with IPPO. The approach achieves lower training costs, better performance, and improved scalability compared to Bi-RNN and Mean-Embedding.
- 3) The effectiveness of the learned policies is demonstrated through a physical experiment involving four autonomous robots pursuing two targets, showcasing successful policy transfer to the real world.

The remaining sections of the paper are organized as follows: Section II provides a review of relevant works. Our

¹Helei Yang, Peng Ge, Junjie Cao, Yifan Yang and Yong Liu are with the Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou 310027, China.

*Both Yong Liu and Junjie Cao are the corresponding authors, emails: yongliu@iipc.zju.edu.cn, cjunjie@zju.edu.cn

经检索“Engineering Village”，下述论文被《Ei Compendex》收录。（检索时间：2024年3月22日）。

<RECORD 1>

Accession number:20240315412463

Title:Large Scale Pursuit-Evasion Under Collision Avoidance Using Deep Reinforcement Learning

Authors:Yang, Helei (1); Ge, Peng (1); Cao, Junjie (1); Yang, Yifan (1); Liu, Yong (1)

Author affiliation:(1) The Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou; 310027, China

Corresponding authors:Cao, Junjie(yongliu@iipc.zju.edu.cn); Liu, Yong(cjunjie@zju.edu.cn)

Source title:IEEE International Conference on Intelligent Robots and Systems

Abbreviated source title:IEEE Int Conf Intell Rob Syst

Part number:1 of 1

Issue title:2023 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2023

Issue date:2023

Publication year:2023

Pages:2232-2239

Language:English

ISSN:21530858

E-ISSN:21530866

CODEN:85RBAH

ISBN-13:9781665491907

Document type:Conference article (CA)

Conference name:2023 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2023

Conference date:October 1, 2023 - October 5, 2023

Conference location:Detroit, MI, United states

Conference code:195422

Publisher:Institute of Electrical and Electronics Engineers Inc.

Number of references:0

Main heading:Reinforcement learning

Controlled terms:Deep learning

Uncontrolled terms:Collisions avoidance - Curse of dimensionality - Decentralised - High dimensionality - Kinematic constraints - Large-scales - Processing method - Pursuit evasion - Pursuit/evasion games - Reinforcement learnings

Classification code:461.4 Ergonomics and Human Factors Engineering - 723.4 Artificial Intelligence

DOI:10.1109/IROS55552.2023.10341975

Database:Compendex

Compilation and indexing terms, Copyright 2024 Elsevier Inc.

注：

1. 以上检索结果来自 CALIS 查收查引系统。
2. 以上检索结果均得到委托人及被检索作者的确认。

