中国科学技术大学

硕士学位论文

基于人机信任的无人机竞速方法研究

作者姓名：　陈少军

学科专业：　控制科学与工程

导师姓名：　赵云波　教授

完成时间：　二〇二四年五月二十四日

University of Science and Technology of China

A dissertation for master's degree

Research on Drone Racing Methods Based on Human-Machine Trust

Author: Chen Shaojun

Speciality: Control Science and Engineering

Supervisor: Prof. Yun-Bo Zhao

Finished time: May 24, 2024

中国科学技术大学学位论文原创性声明

本人声明所呈交的学位论文，是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。

作者签名：陈少年　　签字日期：2024年5月14日

中国科学技术大学学位论文授权使用声明

作为申请学位的条件之一，学位论文著作权拥有者授权中国科学技术大学拥有学位论文的部分使用权，即：学校有权按有关规定向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文编入《中国学位论文全文数据库》等有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人提交的电子文档的内容和纸质论文的内容相一致。

控阅的学位论文在解密后也遵守此规定。

☑公开　☐控阅（____年）

作者签名：陈少年　　导师签名：

签字日期：2024年5月14日　　签字日期：2024年5月14日

摘　要

无人机竞速技术在军事和民用领域展现出广泛应用潜力。在军事领域，无人机竞速技术可以用于环境极速侦察、实施精准打击等；在民用领域，可以用于灾难响应救援、工业管道检测等。因此，研究无人机的竞速方法具有重要意义。

竞速环境的复杂性给现有无人机竞速方法带来了挑战。一方面无人机在竞速飞行中会遇到诸如湍流、阵风等不确定性因素，这种因素是突发、无征兆的，难以通过概率模型进行建模；另一方面，在多机竞速飞行中，存在着诸如欺骗和故意阻拦等高级博弈行为，这些行为因其复杂性而难以被精准定量描述。这些困难会导致无人机自主决策出错，使得竞速效果显著下降。

人机混合决策在无人机竞速领域的应用备受关注，是解决上述挑战的潜在方向。而人机混合有效决策高度依赖于良好的人机协作关系，否则会出现人类工作负荷大、忽视机器错误决策等问题。考虑到人机信任起着协调人机协作的关键作用，将人机信任融入人机混合决策为解决无人机竞速中的挑战提供了可能。为此，论文提出了适用于无人机竞速场景的人机信任模型，基于该模型，设计了单机场景和多机场景下的竞速策略。论文的主要研究工作分为以下三点：

（1）考虑到现有信任模型可解释性差、未能刻画竞速飞行高动态特点，提出了机器性能驱动的人机信任模型，该模型能够促进人对机器的信任水平和机器实际能力之间的匹配。首先给出了人机信任所具有的特征以及演化规律；其次，引入机器当前表现作为机器性能的评估指标之一，构建机器性能驱动的人机信任模型；最后设计了真实的人机交互实验用于验证所提模型。

（2）针对单机竞速场景中不确定性因素导致无人机竞速效果不佳的问题，提出了基于人机信任的单机竞速共享控制策略，有效地改善了无人机竞速效果，同时提升人类的容错率，降低人类工作量。首先设计用于检测人类失误的无人机轨迹预测模块；其次构建基于人机信任的仲裁机制，实现人机权限的动态分配；最后在实验平台上验证所提方法有效性。

（3）考虑到多机竞速场景中有着对于高级博弈行为的刻画和实时计算的需求，提出了基于人机信任的多机竞速强化学习策略，该方法能够实时给出决策指令，同时提升了机器对于博弈的理解和竞速效果。首先将人机信任作为训练奖励融入到奖励塑造函数中；其次设计对手轨迹预测网络，增加机器对于对手的策略理解；最后设计实验验证所提方法有效性。

关键词：人机信任；共享控制；无人机竞速；强化学习

ABSTRACT

Drone racing technology shows potential for a wide range of applications in both military and civilian fields. In the military field, drone racing technology can be used for environmental reconnaissance, precise strikes, etc; in the civilian field, it can be used for disaster response and rescue, industrial pipeline inspection, etc. Therefore, it is significant to study drone racing methods.

The complexity of racing environment presents challenges for existing drone racing methods. On the one hand, drones may encounter uncertainties such as turbulence and gusts during flights. These uncertainties are sudden and unpredictable, making them difficult to model using probability models. On the other hand, there are advanced game behaviors such as deception and intentional obstruction in multi-drone racing. These behaviors are difficult to precisely quantify due to their complexity. These difficulties can lead to errors in autonomous decision-making for drones, significantly reducing the performance of drone racing.

The application of human-machine hybrid decision-making in the field of drone racing has garnered significant attention as a potential direction for addressing the aforementioned challenges. Effective human-machine hybrid decision-making relies heavily on a strong human-machine collaboration relationship. Otherwise, issues such as high human workload and neglect of machine errors may arise. Due to the critical role of human-machine trust in coordinating human-machine collaboration, integrating human-machine trust into human-machine hybrid decision-making offers the potential to address challenges in drone racing. To this end, this dissertation proposes a human-machine trust model for drone racing. Based on this model, racing strategies are designed for both single-drone and multi-drone racing. The main research work of the dissertation is divided into the following three aspects:

(1) Considering the poor interpretability of existing trust models and their failure to capture the high dynamic characteristics of drone racing, a machine-performance-driven human-machine trust model is proposed. This model can promote the alignment between human trust levels for machines and the actual capabilities of machines. First, we outline the characteristics and evolutionary laws of human-machine trust. Then, current machine performance is introduced as a crucial evaluation metric for constructing trust model. Finally, a real human-machine interaction experiment was designed to validate the proposed model.

(2) For the single-drone racing, where uncertainties lead to poor racing performance, a shared control strategy based on human-machine trust was proposed. This method effectively enhances drone racing performance while improving the human operator's error tolerance and reducing their workload. First, a module for detecting human errors based on drone's trajectory prediction is designed. Then, a trust-based arbitration mechanism is constructed to enable dynamic allocation of human-machine authority. Finally, the effectiveness of the proposed method was validated on the experimental platform.

(3) Considering the demands of sophisticated game behaviors description and real-time computations in multi-drone racing, an reinforcement learning strategy for multi-drone racing based on human-machine trust was proposed. This method can provide real-time decision instructions and improves the machine's understanding of the game and racing performance. First, we incorporate human-machine trust as training rewards into the reward shaping function. Then, we design opponent trajectory prediction networks to enhance the machine's understanding of opponent strategies. Finally, we design experiments to validate the effectiveness of the proposed method.

Key Words: Human-machine trust; Shared control; Drone racing; Reinforcement learning

目　录

插 图 清 单

表 格 清 单

第 1 章　绪　　论

1.1　研究背景与意义

在现实应用背景下，无人机竞速飞行具有显著的实际价值。在军事领域：利用无人机的敏捷性和快速响应能力[1]，可以在短时间内完成对复杂或受限区域的快速侦查任务，比如城市巷战环境、山地丛林等；结合先进的导航定位系统和竞速飞行算法，无人机可在复杂环境下迅速找到并锁定目标[2]，从而实现类似“斩首行动”之类的快速突袭任务。在民用领域：地震、洪水、火灾等突发自然灾害发生后，无人机可以迅速穿越危险区域，进行快速勘查，寻找被困人员，缩短搜救时间；无人机可以在狭小、曲折、高空的管道内部进行高效巡检，及时发现裂纹、腐蚀等问题，预防重大安全事故的发生；在森林防火、野生动物保护等领域，无人机可以在短时间内覆盖大面积区域，尤其在地势陡峭、交通不便的环境中进行实时监控，发现火源、非法伐木或偷猎行为；在娱乐文化产业中，无人机竞速飞行可以拍摄酷炫的打斗场景[3]，降低电影拍摄成本。

无人机竞速飞行涉及面广，具有挑战性，成为了近些年的研究热点。无人机竞速飞行涉及到动力学建模、机载感知、定位和绘图、轨迹生成和最优控制等多个方面，具有较高的技术挑战性与研究价值[4]。学术界和企业界为此举办了许多无人机竞速飞行比赛（如图1.1所示）。2016 年，国际智能机器人与系统大会（IROS）第一次举办无人机自主飞行竞速[5]，之后该活动每年举行一次。2019 年，NeurIPS 举办了涉及多无人机竞速的飞行比赛[6]。洛克希德 · 马丁公司也于 2020 年组织了 Alpha Pilot 比赛[7]，该比赛专注于 AI 算法在无人机竞速飞行中的应用。总的来说，无人机竞速飞行的实际价值和技术挑战，已经引起了学术界和企业界极大的兴趣[8]。

(a) 2016 年 IROS 比赛　**(b) 2019 年 NeurIPS 比赛**　**(c) 2020 年 Alpha Pilot 比赛**

图 1.1　学术界和企业界举办的一些无人机比赛

竞速环境的复杂性给现有无人机自主竞速方法带来了挑战。无人机在竞速飞行过程中，会遇到气流干扰[9]、电磁干扰[10]等不确定因素，这些干扰往往突发、无征兆[11]，因此难以被概率建模，一旦出现机器极可能感知错误导致决策出错。在一些多机竞速对抗环境中，无人机需要揣测对手意图、隐藏自身意图，

具备博弈思维，然而欺骗、故意阻拦等高级博弈行为无法被精准定量描述，因此，这些高级的抽象认知对于机器来说难以做到[12]，这会使得无人机在多机竞速过程中落入下风，导致竞速失败。这些都是机器自主决策所面临的固有挑战，仅依靠机器自身能力难以完全解决[13]。

由于人类具备和机器不同的优势，人机混合决策在无人机竞速领域的应用备受关注，是解决上述挑战的潜在方向。机器的能力优势体现在数值计算、高精度、高存储方面，人类的能力优势体现在应急能力和高级推理方面。在面对突发情况时，人类可以根据其丰富经验进行有效处理[14]。此外，高级推理能力让人类能够理解和把握战略环境，可以为战略决策提供信息的预测和判断[15]。可以看出，人类的这些能力恰好对应了无人机自主竞速中的难点。因此，在无人机竞速研究中，人机混合决策成为了潜在的研究趋势。

为保证良好的人机混合决策，合适的人机信任必不可少，同时，基于合适的人机信任，可以优化人机混合智能系统[16]。人机信任一直是人机领域的重要研究方向[17]，人机信任是人对于机器客观能力的主观认识。过度信任和缺乏信任都会对人机系统产生不良影响[18]：如果人类“过度信任”无人机，人的引入没有起到纠正无人机的作用，无人机决策出错便会导致竞速飞行失败；如果人类对无人机“缺乏信任”，便会浪费机器的能力，增加不必要的人类负担。此外，人机信任是人机之间的“桥梁”，可以体现和调节双方交流的信息，这些信息有助于优化人机混合决策。从具体做法角度来说，合适的人机信任可以在两方面用于人机混合系统的设计：一方面是直接协调人机关系，例如在人机共享控制当中，当人机信任水平较低时，那么此时就应当降低机器的自主权限，提升人类的权限；另一方面是利用人机信任去提升机器的能力，例如通过机器学习的方法，将人机信任作为学习指挥棒，机器可以通过人机信任水平去获悉自身决策的好坏，进而优化决策，提升能力。

然而，尽管人机信任在人机混合决策中具有关键作用，却在无人机竞速实践中未得到应有的重视。因此，针对竞速场景复杂性所带来的挑战，论文将人机信任引入到无人机竞速方法设计中。首先构建适用于竞速场景的人机信任模型，然后分别考虑单机竞速和多机竞速两种不同的场景：在单机竞速中使用人机信任直接调节人机之间的权限，在多机场景中则是利用人机信任去训练机器，让人机优势互补，实现了良好的竞速效果，具有明确的研究意义与价值。

1.2　研究现状

1.2.1　人机信任研究现状

现有关于人机信任的研究主要集中在信任测量和信任演化模型构建两个部分[19]，对此分别介绍。

1. 信任测量

由于信任的潜在性和多面性，无法直接衡量信任。因此，衡量信任取决于捕捉其他因素或基本结构[20]。现有的测量方法主要有三种：自我报告、生理测量、行为测量方法。

自我汇报是当前信任测量适用场景最多的方法[21-22]，也是最早开发出来的方法。该方法使用调查和问卷形式对人（机器操控者）进行调查，最终整合问卷内容，得到人对于机器的信任；此类方法目前没有评估信任的标准，主要是研究人员根据不同的场景设计不同的信任问卷。早期量表一般都是单一量表，例如 Muir 等[23]开发了评分量表用以衡量人对于自动化系统的态度。Jian 等[24]开发了最为常用的信任量表，他们设计了一个三阶段的实验研究，构建多维度的信任测量尺度问卷，设计了用于评估人们对于自动化系统的信任问卷。Madsen 等[25]针对之前量表难以直接推广到其他样本的问题，设计了基于认知和情感的人机信任测量仪器。他们在一项针对运营出租车调度系统用户的现场研究中，对该仪器的结构有效性和量表可靠性进行了测试。Schaefer 等[26]进一步考虑了人机信任动态变化情况下的信任量表设计，开发了一种人对于机器信任变化的有效量表。

生理测量主要是通过分析脑电波、心率等人体生理信号来获得人对于机器的信任，本质上是构建生理信号和信任水平之间的映射关系，是近些年兴起的测量方法。Gupta 等[27]调查了参与者对 VR 环境中搜索任务的听觉辅助系统的信任。他们使用各种传感器收集数据，包括 EEG、HRV 和 GSR 设备等。这项研究发现，生理测量确实可以用来评估人类对机器的信任。Akash 等[28]基于 GSR 和 EEG 的数据提出了一个可以实时测量的经验信任传感器模型。他们在滤除尖峰等噪声干扰的基础之上，通过时域和频域分析方法提取特征，之后把特征放入二次判别分类器进行训练，以此得到最终传感器模型。该模型表明心理生理信号可以用于实时的测量人类的信任水平。Choo 等[29]针对空军多属性任务场景，依靠 EEG 获取人类生理信号，不需要事先假设特征信号与信任之间的映射类型，也不需要假设信任本身的演化，在特征提取之后直接训练深度学习网络模型来搭建脑电波信号和人机信任水平之间的映射关系。这个模型在信任估计和检测方面表现出色，成为人机交互中估计多级信任和检测信任校准的理想选择。此外，它还有助于预防自动化系统的废弃和滥用，通过估计操作员的信任水平并监控

他们对自动化系统的信任校准动作。

行为测量研究较少，这种方法是通过分析人机协作过程中人对外的动作表现来获得人对于机器的信任水平。Hu 等[30]考虑多人同时做实验，每次实验时，记录执行机器决策的人数占据总人数的比例，以此作为群体对于机器的信任水平。计算群体动作比例是绝大多数行为测量的方法，比如 Christensen 等[31-32]。Hudspeth 等[33]方法比较独特，即使用个人的动作作为信任水平。他们巧妙地设计了人和机械臂的协作实验：在某区域内，要求人类放置易碎物品。观察易碎物品的位置，离区域边界越远，说明人对于机器的信任越低。于是就用易碎物品到区域边界距离作为人对于机器信任水平的衡量。

信任测量的三种方法使用场景不尽相同，各有优劣。在无人机竞速场景中，人类是参与决策的，基于生理信号的信任测量需要佩戴专用的生理检测设备，对于人机协作造成了极大的不便，干扰人的决策，因此生理测量并不适合；此外也不存在人类群体动作，基于动作的间接测量也不适用。信任自我汇报使用方便，可操作性较强，而且从原理上说，由于信任本身是主观的，这种测量方式得到的结果是最接近真实的人机信任水平。因此，论文采用信任自我汇报的方式得到人机信任水平。

2. 人机信任演化模型

根据信任演化模型中是否出现过往信任（即历史时刻的信任水平），可将信任模型分成两类：无过往信任模型和有过往信任模型。有过往信任模型根据数学形式又可分为两类：有过往信任时间序列模型和有过往信任概率图模型。

无过往信任模型研究不多，该模型是指信任模型中只有当前时刻信任和机器决策历史中影响信任的相关因素，适用于人机关系简单、影响因素较少的情形。Lee 等[34]给出了信任水平和人机表现、机器故障率之间的线性回归模型，并且利用该模型解释了信任波动方差。但是他们也指出，信任演化是有惯性的，该模型没有任何信任记忆信息，无法反映出信任的动态变化。Wang 等[35]针对人和机械臂协作的场景，考虑机器对于人类的信任，把人类表现作为影响信任的唯一因素，最终将信任建立为所有历史表现的加权和。和之前的研究不同，该模型没有将人类表现用成功率或者故障率来简单表示，而是重点分析了人类表现的计算方法，考虑安全性、机器人奇点、运动的平滑性以及人类的身体和认知表现等因素。基于该模型，可以较好的预测机器的行为动作。Nam 等[36]针对集群机器人搜寻问题，考虑人类无法清楚的感知群体任务的性能的情况下，将信任表述为马尔科夫决策过程，通过逆强化学习得到信任计算模型。他们把集群面积、航向方差、人类干预等信任影响因素作为奖励函数的特征，通过逆强化学习得到奖励函数进而预测信任。该模型可以有效的减少信任预测误差，同时考虑了信任个性化特点，不同操作者会有不同参数的信任模型。

有过往信任时间序列模型[30,34,37-39]是现有研究最多的模型，该模型考虑信任的变化惯性，即当前时刻信任不仅和机器性能、任务环境等因素相关，还和过往时刻的信任有关，最终构建为时序模型。Lee 等[34]指出信任变化是逐渐发生的，不可能瞬间改变，模型需要体现信任变化的因果关系。他们考虑当前时刻信任、上一时刻信任、机器故障率、人机总体表现，构建起自回归移动平均向量信任模型。该模型描述了因果关系所涉及的因素、信任随时间变化的动态、发生故障时信任被侵蚀的速度以及随着系统性能的提高信任建立的速度，是最早的有过往信任时间序列模型。Sadrfaridpour 等[38]针对人与机器合力搬运东西的场景，在考虑过往信任的基础上，加入了人的能力表现，计算人的能力表现和机器能力表现的差值，以此作为信任影响的因素之一。他们认为只要人类和机器人的表现存在相当大的差异，无论哪一个比另一个更大，信任都会降低；相反，如果在一段时间内，性能之间没有显著差异，则信任将会增加。基于该模型设计控制策略，最终提升了人机合作的整体表现。Jonker 等[40]提出信任的变化与机器表现和过去信任的差异成正比。基于这个观点，Hu 等[30]提出了一个针对人机协作场景下的通用信任模型。他们以机器漏报率和错报率作为机器的性能表现，并且把累积信任和人的期望偏置引入到信任模型中，给出了信任演化模型的稳定性条件。此外，该模型还讨论了性别、文化、国度等其他非机器性能因素对于信任模型的影响，他们指出，这些因素会影响信任演化模型中的参数，体现为人对于机器的信任倾向性不同。

有过往信任概率图模型将信任视为服从某个概率分布的随机变量，把信任以及信任的影响因素作为概率图的节点。Xu 等[41]考虑监督协作的人类机器人团队：人类扮演主管角色，负责将任务分配给机器人，同时有权力进行干预和接管控制。他们把机器表现、人的干预、人的反馈、当前时刻信任、上一时刻信任作为概率图节点来构建贝叶斯网络。该模型假定信任水平服从一维高斯分布，其均值和上述影响因素相关，其方差为固定参数，同时利用逻辑斯蒂函数反映人的动作和信任水平之间的联系。该信任模型预测精度较高，其结果表明：当机器出现故障时，人机信任会迅速下降，然而重新上升却很难。Chen 等[42]在考虑人机信任的基础上，把可控措施融入到信任模型中。他们的实验场景是人和机器共同清理桌面，可控措施是：机器可以选择清理高风险或者低风险物品。他们将信任视作隐藏变量，利用 HMM 描述其演化过程进而构建起 POMDP，通过最大化人机团队表现求解 POMDP。和仅有信任预测模型的研究相比，机器可以根据人的信任水平做出最优决策，提升人机协作体验。Akash 等[43]考虑信任自我汇报在序贯决策中难以使用的问题，将人的信任状态作为隐藏状态变量，人的动作作为观测，进而构建起有向概率图信任模型。他们利用大量的人机交互的历史数据，通过贝叶斯推理，学习出模型中的参数，之后利用得到的观测去推理出最合适的信

任值。所提出的模型还考虑了机器透明度、人的工作量，通过求解系统最佳透明度，最终提升了系统性能。

1.2.2　无人机竞速方法研究现状

无人机竞速方法的研究主要涉及感知、决策、执行三个方面[44]。无人机的感知包括检测、识别、跟踪和定位物体以及估计无人机自身状态[45-47]；无人机的决策是根据感知模块获取的信息进行竞速规划，生成顶层指令[48]，如指定位置、速度、角速度等; 执行是底层控制系统根据顶层指令，完成对于电机的控制，例如 PID[49]，LQR[50-51]，SMC[52]等。由于论文研究的是人机决策内容，因此对于无人机的竞速决策研究现状作重点介绍。根据竞速场景不同，分成单机竞速研究和多机竞速研究两部分。

1. 单机竞速控制方法

基于数值优化的规划算法是单机竞速控制的主流研究方向，这类方法是根据竞速任务目标设计优化函数，考虑竞速环境的限制，将竞速问题转化为优化算法实时求解问题。这类方法可解释性强，但是对于实时计算能力要求较高，有些问题无法转化为凸优化问题，只能得到近似解。此外，这些方法对于干扰较为敏感，往往初值条件小幅变动，规划出来的轨迹差异较大。Foehn 等[53]和 Ryou 等[54]将此问题分为规划任务和控制任务，在规划任务中生成全局时间最优轨迹，在控制任务中精确跟踪该轨迹，这也是单机竞速中最常用的方法。Shuli 等[48]针对迭代学习控制 Bristow 等[55]无法用于无人机竞速的问题，提出了时间最优空间迭代学习方法，建立了一个具有路径和障碍物信息的空间虚拟管用以建模比赛场景，考虑无人机通过虚拟管道尽快到达目的地，最终转化为时间优化的求解。Shin 等[56]使用进化算法优化竞速的速度参数，使用在线学习进行轨迹规划，在 2019 年 NeurIPS 举办的比赛中获得了第二级和第三级的第一名。Wang 等[57]针对动态环境中的比赛，考虑速度和避障之间的权衡，提出了一种具有有效多项式轨迹表示的在线重规划框架。MPC（Model Predictive Control）也是解决复杂无人机问题的有效方法[58]，常常用于竞速问题中的最优时间规划。Romero 等[59]针对多项式控制输入不能充分利用致动器的潜力，使控制策略处于次优状态的问题，提出了 MPCC（Model Predictive Contouring Control）竞速方法，该方法考虑了全四旋翼动力学和实际单旋翼推力约束，以实现时间最优四旋翼飞行。Romero 等之后基于 Receding Horizon Approach 方法设计了高效的采样方法，并且扩展了 MPCC 的实现，使其适用于在线竞速重规划[60]。

近些年 AI（Artificial Intelligence）技术大放光彩，AI 赋能的单机竞速控制方法也得到了发展。这其中主要的技术思路有两方面：一方面是利用神经网络替代传统方法中的某些部分，例如无人机的动力学模型等；另一方面是使用端到端的

形式，将感知到的信息通过深度学习直接映射为控制指令。此类方法实时性较好，但是神经网络的训练耗费较高，同时深度学习的不可解释以及鲁棒性差等因素使得该方法同样对于干扰敏感，容易决策失误。Salzmann 等[61]利用神经网络拟合无人机的动力学模型，将其嵌入到轨迹规划当中，实现了良好的穿门竞速效果。Nagami 等[62]考虑 HJB（Hamilton Jacobi Bellman）方法不适合于实时的无人机竞速控制，使用神经网络来模仿 HJB 的控制策略，并且利用强化学习微调控制策略，结合了学习算法与先进控制算法的优势。Song 等[63]针对 MPC 方法中某些难以优化决策的变量，通过神经网络学习经验数据生成高级策略实现优化，实现了控制性能和复杂策略学习的结合。Kaufmann 等[64]开发了基于深度强化学习实现无人机竞速的端到端（Point to Point）框架，感知系统将高维视觉和惯性信息转换为低维表示，决策系统摄取感知系统产生的低维表示并产生控制命令，实现了感知与决策的一体化。

除了上述自主竞速方法，人机混合决策方法在单机竞速中也有应用[65]。这类方法通将人类操作员的专业技能与机器自主决策算法相结合，以提高无人机在复杂、高速竞速环境中的决策效率和性能，在此类方法中，人类操作员通常占据更高级别的决策层级，掌握着较大的控制权限，以便在必要时进行关键决策和干预。Agrawal[66]设计了人机权限交换机制，当检测到人类有干预时，会将全部控制权限移交给人类，同时基于一组探索性问题开发了一个流程，用于引指定无人机的人工干预需求。Hummel 等[67]设计了一个分布式的人与无人机协作架构，该架构充分利用了实时传输的无人机竞速视频影像信息。在这个框架中，人类操作者通过观察和解析实时传送回来的高清图像资料，能够实现对无人机竞速环境中尚未被自动系统充分识别或捕捉到的关键信息进行补充认知和解析，并且同样给予实时干预。

2. 多机竞速控制方法

相较于单机竞速，多机竞速总体上研究较少，正在逐步发展，同时与多机竞速类似的领域（例如多赛车竞速）也取得了一些成果。这些研究集中在基于博弈的优化算法上。技术思路是将单机竞速算法拓展到多机过程中，相当于原有环境中多了一个对手无人机的碰撞问题。Liniger 等[68]考虑双赛车竞速，将竞速场景建模为零和博弈模型，设计博弈代价函数，通过顺序最大化方法来计算纯策略中的斯塔克伯格和纳什均衡。Spica 等[69]针对 2 维的多无人机竞速场景，首次将博弈论引入到无人机竞速当中，利用灵敏度分析改进多机竞速的优化求解，通过迭代最佳响应进行实现近似的纳什均衡。Wang 等[70]针对 3 维场景下的多无人机竞速，考虑对手的意图和反应，通过反复规划自己的轨迹寻求纳什均衡。Schwarting 等[71]利用深度强化学习算法实现多机竞速任务，但是这种算法并没有完整考虑竞速规则，实质上类似单机竞速飞行。Li 等[72]考虑多赛车竞

速问题，设计了最小时间目标下的非线性模型预测控制（NMPC）模型，使用曲线坐标系，将进度时间设置为直接优化目标，基于混合整数二次规划的方法最终生成安全控制策略。He 等[73]考虑多赛车竞速场景，针对周围车辆有无分别设计两种竞速策略，利用障碍函数约束的 MPC 控制器来避障，以保证系统的安全关键性能，实现了无碰撞轨迹的跟踪。Shen 等[74]针对多无人机竞速场景，将 CPC(Complementary Progress Constraint) 扩展到一组无人机，求解碰撞约束优化问题，生成时间最优轨迹，使四旋翼无人机能够以其极端的机动性完全自主地飞行通过赛道，并以最短的飞行时间达到目标。

人机混合决策已在多无人机集群运作中有所实践，例如 Agrawal 等[75]探讨了人与集群无人机间的协同设计，并提出了一个人类操作员与自主无人机群交互元模型。尽管如此，此类研究尚未深入探讨无人机个体间的博弈行为。总体来看，关于人机混合决策在多无人机竞速情景中的应用尚不多见。

1.2.3　研究现状总结

1. 现有信任模型可解释性差、未能刻画竞速飞行高动态特点，不适用于无人机竞速场景

- 现有信任模型对于机器性能的刻画没有体现出无人机竞速的高动态特点，信任和机器性能并不是实时匹配的。现有模型对于机器的性能刻画主要是记录机器所有的决策历史，通过数据处理得到机器的性能评估（例如机器的历史正确率）。这种机器性能评估方法缺少了对于机器当前表现的刻画。而无人机具有高动态特性，机器的当前表现会对人机信任产生重要影响，只考虑机器的历史表现会导致人类的信任水平与机器性能之间存在延时偏差。
- 将信任引入无人机竞速的目的是优化控制设计，提升飞行效果，然而现有信任模型可解释性较差，无法用于优化控制设计。现有研究大多数是考虑某些信任影响因素，将其放入某个数学模型中，进而得出信任演化模型，对于演化模型中参数的意义没有分析，使得信任模型缺乏原理上的解释，难以用于后续的控制优化设计。

2. 单机竞速中，面对不确定性干扰导致竞速效果不佳的问题，自主竞速方法难以解决，而现有的人机混合决策方法中人类工作量大、容错率低

- 无人机自主竞速时，一旦遇到未被建模的不确定性干扰，便极有可能决策出错，这是自主决策方法的固有缺陷，难以解决。由于机器很难对阵风、湍流、电磁干扰这类无征兆的因素进行概率建模，这些因素一旦出现可能会使无人机感知错误，进而决策出错，造成偏离预定轨迹，甚至于竞速失败。
- 现有的人机混合竞速方法普遍赋予人类较高的权限。然而，一旦人类操作

失误，极有可能导致竞速任务失败。因此，人类操作员须持续保持高强度的专注力以杜绝操作失误，这不仅大幅度加重了其工作负担，而且对操作员的错误容忍度也提出了极为苛刻的要求。

3. 多机竞速中，缺少了对于高级博弈行为的刻画，计算实时性达不到要求

- 多无人机竞速中，欺骗、故意阻拦等复杂博弈行为构成了竞争胜负的关键要素。然而，现有的多机竞速算法大多仅侧重于遵循避免碰撞的基本准则，却忽略了对竞争对手复杂博弈行为的辨识与应对，以及自身在复杂博弈环境下的决策制定。
- 多机竞速算法大多使用优化算法，这些算法耗时较长，难以满足多机竞速的高动态要求。多机竞速存在高动态对抗，其计算耗时应当低于单机竞速，然而多机竞速的复杂性使得计算耗时反而高于单机场景，如果对方无人机实时性较好，那么这几乎是降维打击。

1.3　论文工作与结构安排

本节考虑在引入人机信任的无人机竞速场景下，根据上述研究不足，提出本文的研究内容和组织结构。

1.3.1　论文研究内容

针对1.2.3中提到的研究不足，论文从人机协作的角度出发，将人类干预引入到无人机竞速当中，通过人机信任协调人类和机器之间的关系，实现人类和机器的优势互补，最终提升无人机竞速效果。具体包括以下三部分内容：

- 考虑到现有信任模型可解释性差、未能刻画竞速飞行高动态特点，提出了机器性能驱动的人机信任模型。首先根据人机信任的影响因素分析了人机信任所具有的特征以及演化规律；其次，引入机器当前表现作为机器性能的评估指标之一，进而构建机器性能驱动的人机信任演化模型，给出合理性说明、模型参数的意义和相关性质分析，让模型具备可解释性；最后设计了真实的人机交互实验来验证所提信任模型的有效性。
- 针对单机竞速场景存在的不确定性干扰使得无人机竞速效果不佳的问题，提出了基于人机信任的单机竞速共享控制策略。首先构建人类决策输入的无人机轨迹预测模块，检测人类操作失误的情形，防止人类缺点被引入到竞速飞行中，提升人类的容错率。其次，构建以人机信任作为依据的人机权限动态分配方法，提升无人机竞速飞行效果，同时降低人类工作量；最后，利用实验平台验证所提方法的有效性。
- 考虑到多机竞速场景中有着对于高级博弈行为刻画和实时计算的需求，提

出基于人机信任的多机竞速强化学习策略。首先设计强化学习奖励塑造函数，用人机信任帮助机器强化学习训练，让机器理解博弈，引导机器做出正确决策；其次，构建对手轨迹预测网络，增加机器对于对手的策略理解；最后在实验平台上验证所提方法的有效性。

1.3.2　论文组织结构

论文的组织结构如图1.2所示，共有六章，具体描述如下：

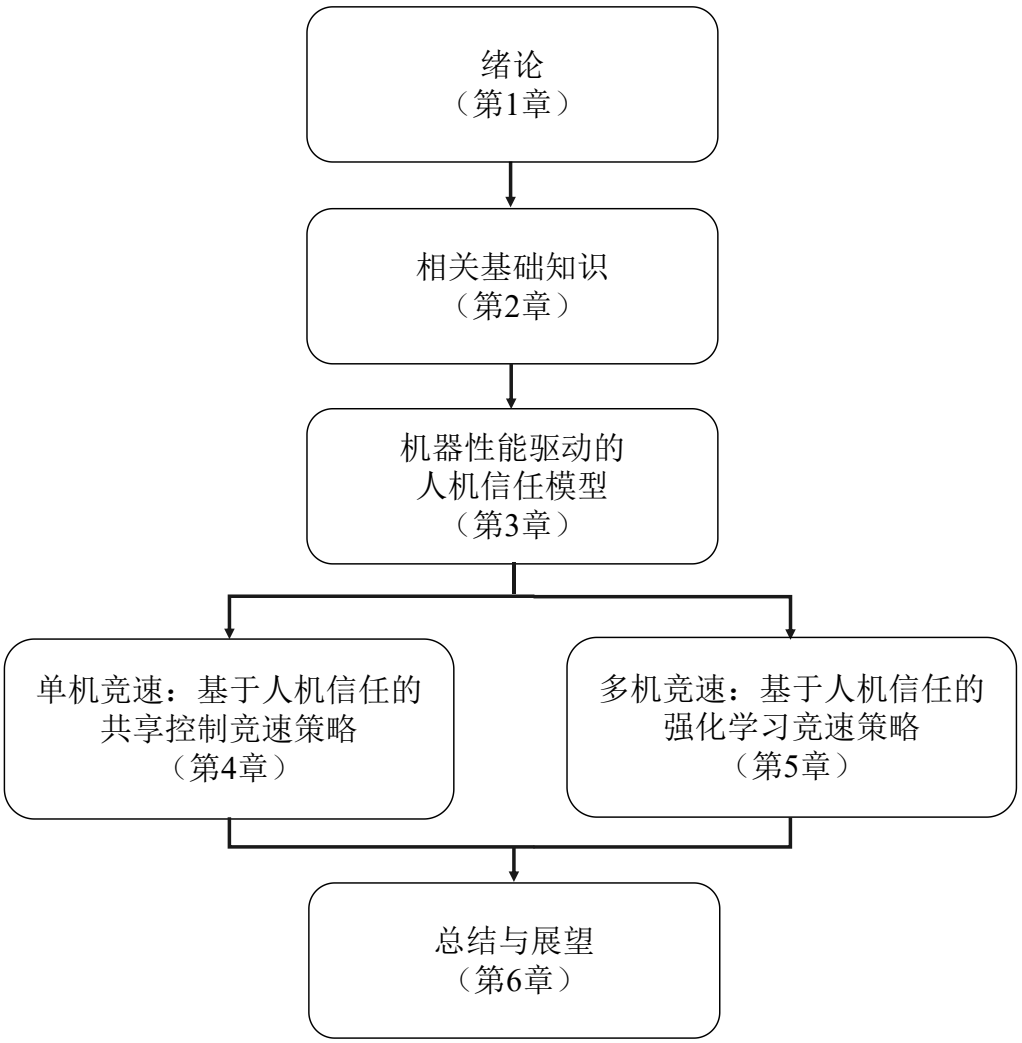


图 1.2　论文的组织结构

第 1 章：绪论。介绍了论文的研究背景，阐明了在无人机竞速中引入人机信任的研究意义。同时对于现有人机信任和无人机竞速的研究现状予以总结，基于现有研究中存在的问题，给出论文的主要研究内容和组织架构。

第 2 章：相关基础知识。介绍了论文涉及到的基础理论以及使用的一些技术方法，包括无人机运动学建模、人机混合智能系统理论，深度学习和强化学习方

法。

第 3 章：机器性能驱动的人机信任模型。研究了在无人机竞速场景中，如何设计信任模型来尽可能的促进信任水平和机器性能之间的匹配。该研究是第 4 章和第 5 章所需要的基础之一，会在 4、5 两章中进一步具体化，用于后续的控制设计。

第 4 章：基于人机信任的单机竞速策略。研究了在单机竞速场景中，将人类操作者引入到控制回路之后，如何避免人类操作失误的影响以及如何利用人机协作让竞速效果得到改善的同时，提升人类的容错率，降低人类的工作量。具体包括人类操作失误检测设计、第 3 章中的信任模型具体化、人机决策仲裁设计、算法仿真与验证等。

第 5 章：基于人机信任的多机竞速策略。研究了在多机竞速场景下，如何利用人机信任提升机器对于复杂博弈和对手策略的理解。具体包括第 3 章中的信任模型具体化、基于人机信任的奖励塑造函数设计、对手轨迹预测设计、强化学习训练框架设计、算法仿真与验证等。

第 6 章：总结与展望。对论文研究内容进行总结，分析了论文研究中存在的不足，以及进一步的改进方向。

第 2 章　相关基础知识

本章集中阐述了研究所需的几大基础知识板块。首先，介绍无人机坐标系及其运动学模型；其次是论文涉及的人机混合智能相关理论；最后是论文中运用到的深度学习与强化学习技术。

2.1　无人机基础知识

2.1.1　两种坐标系

在无人机系统建模中，涉及到两个坐标系，分别称之为世界坐标系和机体坐标系。

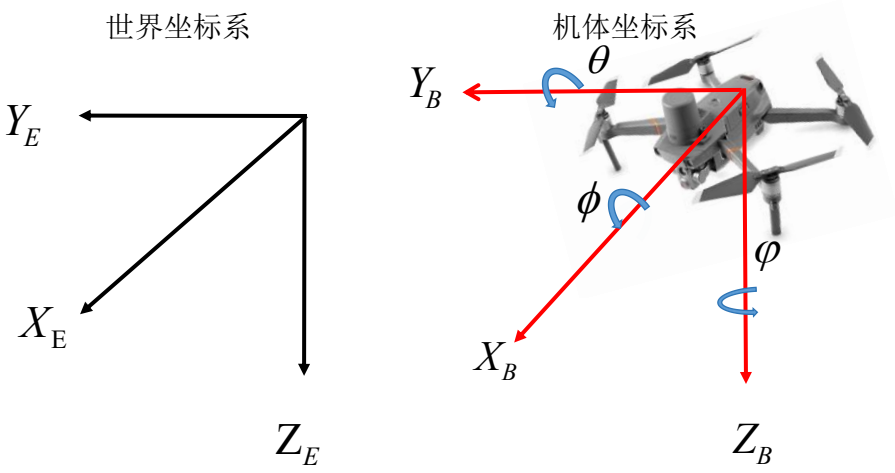


图 2.1　无人机中的世界坐标系和机体坐标系

世界坐标系是一种固定在地面上的坐标系，为无人机提供了一个全局参照系，便于描述无人机在三维空间中的位置、速度和航向，以及进行路径规划和导航计算等。如图2.1所示，世界坐标系采用 NED 形式：正北方向为 X 轴的正方向，正东为 Y 轴正方向，重力方向为 Z 轴正方向，为右手坐标系。

机体坐标系是固定在无人机上的坐标系，主要用于表达无人机自身的姿态（俯仰、滚转、偏航）和内部传感器（如 IMU，惯性测量单元）采集到的数据，这些数据反映了无人机本身的动态变化。如图2.1所示，机体坐标系通常以无人机重心作为坐标原点，将无人机的机头（无人机飞行的前进方向）为 X 轴的正方向，机头右侧为 Y 轴正方向，将旋翼推力反方向定义为 Z 轴，同样也是右手系。

2.1.2　运动学建模

由于无人机基本运动方程中涉及到两个坐标系，因此需要建立二者之间的旋转变换关系。无人机的旋转矩阵 $\boldsymbol{R}$ 通常用来描述二者的旋转关系，$\boldsymbol{R}$ 可以表示为：

$$\boldsymbol{R} = \boldsymbol{R}_Z(\psi)\boldsymbol{R}_Y(\theta)\boldsymbol{R}_X(\phi) \tag{2.1}$$

其中 $\boldsymbol{R}_Z(\psi), \boldsymbol{R}_Y(\theta), \boldsymbol{R}_X(\phi)$ 表达式如下：

$$\begin{cases} \boldsymbol{R}_X(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix}, \\ \boldsymbol{R}_Y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}, \\ \boldsymbol{R}_Z(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{cases} \tag{2.2}$$

其中，偏航角 ψ，俯仰角 θ，滚转角 ϕ 分别是按 Z-Y-X 顺序定义的姿态角（在图2.1中有标注）。式(2.1)表示绕 Z 轴旋转 ψ 角度，然后绕 Y 轴旋转 θ 角度，最后再绕 X 轴旋转 ϕ 角度后的整体旋转效果。

用 $\boldsymbol{p} = [x, y, z]^T$ 表示无人机的位置向量（世界坐标系下），$\boldsymbol{v} = [v_x, v_y, v_z]^T$ 表示无人机的线速度向量（机体坐标系下），$\boldsymbol{\Theta} = [\phi, \theta, \psi]^T$ 表示无人机的姿态角矩阵，$\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^T$ 表示无人机的角速度向量（机体坐标系下），将无人机完整的运动学方程表示如下：

$$\begin{bmatrix} \dot{\boldsymbol{p}} \\ \dot{\boldsymbol{\Theta}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W} \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{\omega} \end{bmatrix} \tag{2.3}$$

其中，$\boldsymbol{W}$ 是姿态角的变化率与机体的旋转角速度之间的变换矩阵：

$$\boldsymbol{W} = \begin{bmatrix} 1 & \tan\phi\sin\psi & \tan\phi\cos\psi \\ 0 & \cos\psi & -\sin\phi \\ 0 & \sin\phi/\cos\theta & \cos\phi/\cos\theta \end{bmatrix} \tag{2.4}$$

在小扰动情况下 $\boldsymbol{W}$ 为单位阵。

2.2　人机混合智能

在人机混合智能系统中，与传统人机协作有所不同，随着决策算法的融入，机器具备了显著的自主行动能力。因此，在这类系统中，人类与机器均展现出了各自的自主性特征，有效分配并协调人与机器的自主权成为了关键问题[16]。共享控制和介入控制是人机混合智能系统的基本控制策略，尽管这二者十分重要，但是现如今学术界对此并没有统一的定义[76]。这里根据文献[16,76-79]，整理出二者的特点，如表2.1所示。

表 2.1　共享控制和介入控制特点对比

角度	介入控制	共享控制
控制目标	主要目标是为了确保人机系统的稳定性和安全性，避免产生不可接受的风险后果	聚焦于系统性能优化，旨在达成显著的效能提升
系统结构	通常仅需单方监控对方状态并判断适宜的介入时机，无需额外增设人机之外的机制	有必要设立一个超越人机层级的统一决策机构，负责执行人机共享的统筹策略
主次地位	在人机交互中，双方地位不等，介入一方较被介入一方地位更高，享有更大决策权限	在系统中，人与机器处于同等地位
相互影响	操控权限可在人与机器之间单向或双向切换，即人可接管机器驾驶，反之亦然，机器也能接手人的操作	人与机器之间的共享表现为深度合作与协同效应，而非简单的单向作用关系
设计要求	尽管执行机制简易，但由于错误介入可能导致整个系统失控，故对接管动作的要求相对严苛	尽管执行机制复杂，但设计上要求一定的灵活性，因其失误结果通常导致优化不足而非系统崩溃

2.3　深度学习

2.3.1　深度学习原理

深度学习是一种基于人工神经网络的机器学习方法，其核心原理在于构建多层非线性变换模型，从而具备了逼近任意复杂函数的能力。这一特性赋予深度学习强大的表征学习能力，使之能够应对诸如图像识别、语音识别、自然语言处理等诸多领域的复杂问题。人工神经元是人工神经网络的组成单元，其结构如图2.2所示，输出 y 计算公式如下：

$$y = f(W^T X + b) \tag{2.5}$$

其中，X 表示输入向量，W 表示权重向量，b 表示偏置量，f 表示激活函数。这里给出论文涉及到的一些激活函数：

- sigmoid 函数:

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{2.6}$$

- tanh 函数:

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2.7}$$

- ReLU 函数:

$$\text{ReLU}(x) = \max(0, x) \tag{2.8}$$

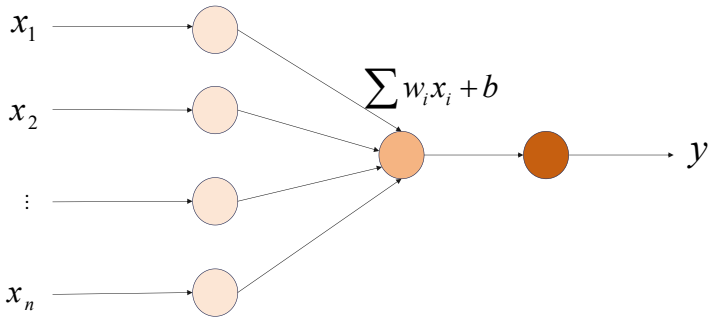


图 2.2　单个人工神经元结构

神经网络的学习流程是以迭代方式进行的，每一次迭代中有三个核心步骤：首先前向传播阶段，计算网络的实际输出；其次是根据实际输出和预期目标计算损失函数值；最后，通过反向传播算法更新网络参数。

深度学习在不同领域有不同的网络架构，例如图像识别的卷积神经网络 CNN（Convolutional Neural Networks），时序数据处理的循环神经网络 RNN(Recurrent Neural Networks)，生成对抗网络 GAN（Generative Adversarial Network）等。论文使用了 RNN 网络，后续仅对 RNN 网络进行介绍。

2.3.2　循环神经网络

在深度学习处理时间序列问题时，RNN 是广泛应用的模型。RNN 能在时序数据分析中表现出色，主要是由于其内在结构特点: 每个时间步 t 的隐藏状态不仅依据当前时刻 t 的输入特征计算，还整合了前一时刻 $t-1$ 的隐藏状态信息（如图2.3所示）。这种机制使得 RNN 能够充分利用历史信息，与仅基于当前时刻输入特征的传统模型形成对比，从而在理解时序动态上有独到优势。

在循环神经网络（RNN）中，每一个时间步 t 的隐藏状态 h_t 的计算通常遵循以下公式：

$$h_t = \tanh(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \tag{2.9}$$

其中：x_t 是当前时间步的输入向量。h_{t-1} 是前一时间步的隐藏状态向量。W_{hx} 是从输入到隐藏层的权重矩阵。W_{hh} 是隐藏层到隐藏层自身的循环权重矩阵。b_h 是隐藏层的偏置向量。tanh 是双曲正切函数，一种常用的激活函数。

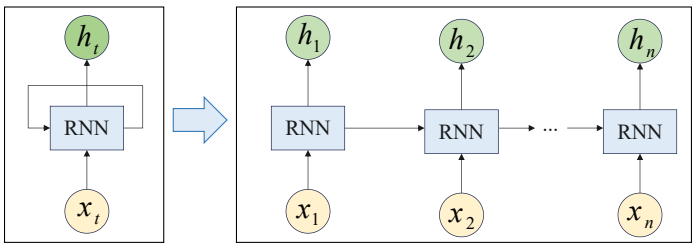


图 2.3　RNN 基本网络结构

此外，输出层的计算公式是：

$$y_t = f_y(W_{yh}h_t + b_y) \tag{2.10}$$

其中：W_{yh} 是从隐藏层到输出层的权重矩阵。b_y 是输出层的偏置向量。f_y 表示输出层的激活函数。

长短期神经网络 LSTM（Long Short-Term Memory）也是一类特殊的 RNN 网络，其结构如图2.4所示。

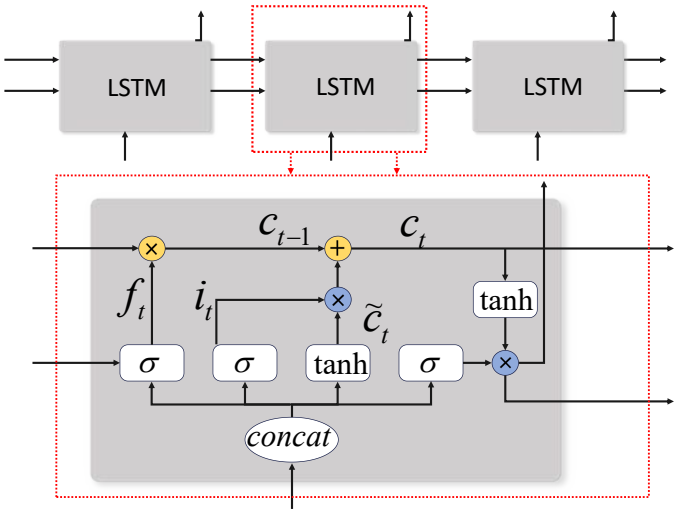


图 2.4　LSTM 基本网络结构

LSTM 利用记忆门、输入门、输出门和遗忘门机制，有效地应对长序列数据挑战。记忆门保留关键信息，输入门控制是否将新输入纳入记忆单元，遗忘门确定何时丢弃现有信息，输出门决定如何将记忆单元内容体现在当前输出中。这些门控机制有助于精准捕获长时序依赖，并能有效缓解梯度消失问题。数据进入到

LSTM 网络中进行的运算如下：

$$
\begin{aligned}
f_t &= \sigma(W_f[h_{t-1}; x_t]) + b_f \\
i_t &= \sigma(W_i[h_{t-1}; x_t]) + b_i \\
o_t &= \sigma(W_o[h_{t-1}; x_t]) + b_o \\
\tilde{c}_t &= f_t \circ c_{t-1} + i_t \circ tanh(W_{\tilde{c}}[h_{t-1}; x_t] + b_{\tilde{c}}) \\
h_t &= o_t \circ tanh(\tilde{c}_t)
\end{aligned}
\tag{2.11}
$$

其中 $[h_{t-1}; x_t]$ 表示 t 时刻的输入 x_t 与上一个单元的隐藏状态 h_{t-1}。σ 是 LSTM 单元中的 sigmoid 激活函数。$\circ$ 表示 Hadamard 运算。$i_t, o_t, f_t, \tilde{c}$ 分别表示每个 LSTM 单元的输入门, 输出门, 遗忘门, 记忆门。$W_f, W_o, W_i, W_{\tilde{c}}$ 与 $b_f, b_o, b_i, b_{\tilde{c}}$ 则为网络参数。

运用 LSTM 模块能有效提炼序列中的周期性特征，进而增强对周期输入序列的预测准确度。此外，鉴于 LSTM 架构的简洁性，相比于繁复的时序预测模型，在保持预测性能相近的前提下，能够提升训练效率，从而节省计算资源和时间消耗。

RNN 网络根据输入数和输出数的不同，有几种常用架构，如图2.5所示：

- 单对多场景：处理非序列输入生成序列输出的任务；
- 多对一场景：当面对的是序列输入但仅要求一个非序列输出时；
- 多对多场景：在此情况下，模型输入和输出均为可变长度的序列数据。

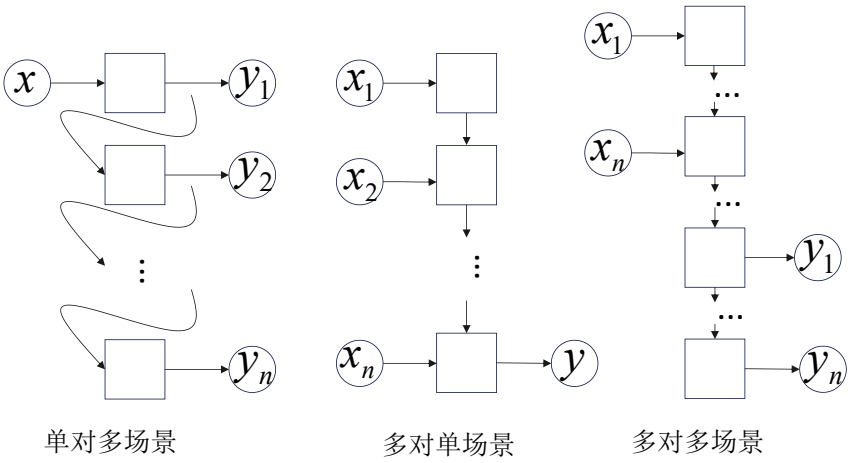


图 2.5　RNN 常见的几种架构

2.4　强化学习

2.4.1　强化学习基本原理

马尔可夫决策过程（Markov Decision Process, MDP）是强化学习中的核心模型，用于描述在一系列离散时间步中，智能体如何通过选择不同动作来最大化累

积奖励的问题。MDP 具有如下基本元素：

- 状态空间 (State Space): 用 $\mathcal{S}$ 表示，智能体可以在其中遍历的所有可能状态集合。
- 行动空间 (Action Space): 用 $\mathcal{A}$ 表示，智能体在每个状态下可以选择执行的所有可能行动集合。
- 转移概率 (Transition Probabilities): 用或 $\mathcal{T}(s'|s,a)$ 表示，在给定状态下采取某个行动后，转移到下一个状态的概率。这里的 s 是当前状态，a 是采取的行动，s' 是下一个可能的状态。
- 奖励函数 (Reward Function): 用 $\mathcal{R}(s,a,s')$ ，当智能体在状态 s 下执行动作 a 并转移到新状态 s' 时得到的即时奖励。
- 折扣因子 (Discount Factor): 通常用 γ 表示，范围在 [0, 1] 之间，用来权衡当前奖励与未来奖励的重要性。未来的奖励会被按照 γ 的指数衰减。

因此，可以用一个五元组 $(\mathcal{S},\mathcal{A},\mathcal{P},\mathcal{R},\gamma)$ 来表示一个 MDP。

强化学习基本的原理就是智能体不断的与环境进行交互，根据环境反馈的奖励去调整自己的策略，定义智能体的策略为：

$$\pi(a_j|s_j) = p\left(\mathcal{A}_j = a_j|\mathcal{S}_j = s_j\right) \tag{2.12}$$

式(2.12)含义是在状态为 s_j 时，采取动作 a_j 的概率。智能体的目标就是最大化获得的累计奖励。强化学习智能体的累计奖励定义为：

$$R = \sum_{t=0}^{\infty}\gamma^t\mathcal{R}_t \quad \mathcal{R}_t : \mathcal{S}_t \times \mathcal{A}_t \times \mathcal{S}_{t+1} \to \mathbb{R} \tag{2.13}$$

式(2.13)本身是一个随机变量，在实际应用过程中，通常考虑最大化期望累计奖励进而得到最优策略 $\pi^*(s)$：

$$\pi^*(s) = \arg\max_{\pi}\mathbb{E}_{\pi}\left[\sum_{t}^{\infty}\gamma^j\mathcal{R}_t \mid \mathcal{S}_0 = s\right] \tag{2.14}$$

进一步的，引入状态值函数 $V(s)$ 与动作值函数 $Q(s,a)$ 用来评估期望累计奖励:

$$V^{\pi}(s) = \mathbb{E}_{a\sim\pi(a,s)}\left[\mathbb{E}_{s'\sim\mathcal{T}(s'|s,a)}\left[\mathcal{R}(s,a,s') + \gamma V(S')\right]\right], \tag{2.15}$$

$$Q(s,a) = \mathbb{E}_{s'\sim\mathcal{T}(s'|s,a)}\left[\mathcal{R}(s,a,s') + \gamma V(S')\right] \tag{2.16}$$

通过这些方程，智能体与环境交互，逐步学习到最优策略 $\pi^*(s)$。

如何学习到最优策略是强化学习领域研究的重点，总的来说，无模型的强化学习算法可以分为三种。第一种是基于值的强化学习算法，模型中只有值评估函数，根据值函数对策略进行改进。第二种是基于策略的算法，通过构建策略函数

$\pi(a|s)$，直接对策略迭代更新参数值，让累积奖励的期望最大；第三种是既有值也有策略的算法，通常称之为 Actor-Critic 框架，Actor 用于计算策略，Critic 用于评估值。论文使用到的强化学习方法是基于 Actor-Critic 框架的，因此后续仅对基于 AC 框架的强化学习方法进行介绍。

2.4.2　基于 AC 框架的强化学习方法

Actor-Critic 总体的思想是：Actor 负责策略更新，Critic 可以对 Actor 更新的方向、快慢等给予指点。下面给出具体做法。

首先定义策略函数 π_θ（一般可以用一个神经网络表示，此时 θ 就是网络的参数），用 τ 表示智能体探索的轨迹，$\mathcal{R}(\tau)$ 表示轨迹 τ 的奖励，$p(\tau;\theta)$ 表示轨迹 τ 出现的概率。因此智能体的目标函数可以表示为：

$$J(\theta) = \mathbb{E}\left[\Sigma_\tau p(\tau;\theta)\mathcal{R}(\tau)\right] \tag{2.17}$$

强化学习目标就是找到最优的 θ 使得 $J(\theta)$ 最大。这就是一个典型的优化问题，通过策略梯度进行优化：

$$\theta_{new} = \theta_{old} + \nabla_\theta J(\theta) \tag{2.18}$$

因此，该问题的关键便是对 $\nabla_\theta J(\theta)$ 的计算，根据 Schulman[80]的博士论文，策略梯度可以写成如下形式：

$$\nabla_\theta J(\theta) = \mathbb{E}\left[\sum_{t=0}^{\infty} \boldsymbol{\Phi}_t \nabla_\theta \ln \pi_\theta(a_t|s_t)\right] \tag{2.19}$$

在式(2.19)中，可以看出有两个部分会影响梯度的变化，一部分是 $\nabla_\theta \ln \pi_\theta(a_t|s_t)$，这是与 π_θ 自身相关的梯度，表示的是参数 θ 如果沿着这个方向走，目标函数变化就最快；另一部分是 $\boldsymbol{\Phi}_t$，$\boldsymbol{\Phi}_t$ 的正负和大小就决定了参数 θ 的变化的正负方向和大小。可以看出 $\boldsymbol{\Phi}_t$ 就起到了对于策略网络 π_θ 更新的指导作用。而 $\boldsymbol{\Phi}_t$ 的计算便需要 Critic（一般也是一个神经网络）参与。一个典型的 AC 算法是利用 TD-error 计算 $\boldsymbol{\Phi}_t$：

$$\boldsymbol{\Phi}_t = \mathcal{R}_t + \gamma V(S_{t+1}) - V(S_t) \tag{2.20}$$

其中，$V(S_t)$ 便是一个 Critic 网络，可以通过 TD-error 进行参数更新。

目前，有许多强化学习算法是基于 AC 框架的，如 SAC[81]（soft Actor-Critic),TRPO[82]（Trust Region Policy Optimization），PPO[83]（Proximal Policy Optimization）等。在这些算法中，PPO 算法的计算与调试的复杂度较低，在许多问题中取得很好的效果，因此论文使用了了 PPO 算法，对其重点介绍。

PPO 与一般的 AC 算法差别在于策略更新方法上。PPO 采用优势函数 $A_t(s_t,a_t) = Q(s_t,a_t) - V(s_t)$ 作为 $\boldsymbol{\Phi}_t$，考虑策略裁剪，PPO 在策略更新时的目

标函数是：

$$J^{CLIP}(\theta) = \hat{\mathbb{E}}\left[\min\left(r_t(\theta)\cdot\hat{A}_t, \text{clip}\left(r_t(\theta), 1-\varepsilon, 1+\varepsilon\right)\cdot\hat{A}_t\right)\right] \tag{2.21}$$

其中：

- $r_t(\theta) = \frac{\pi_\theta(a|s)}{\pi_{\theta_{old}}(a|s)}$ 是基于参数 θ 的新策略与旧策略在状态 s 下选择动作 a 的概率比；
- $\hat{A}_t$ 是在时间步 t 的优势函数的估计值；
- ε 是裁剪窗口大小。

根据式(2.18)和式(2.21)，利用 Adam 优化算法便可以对 PPO 的 Actor 进行参数更新。

PPO 的 Critic 网络更新方法与一般的 Actor-Critic 算法相同。

第 3 章　机器性能驱动的人机信任模型

考虑在无人机竞速场景中，机器的决策环境是高度动态变化的，本章引入机器当前表现作为机器性能的重要评估指标，进而构建机器性能驱动的人机信任模型，促进人对于机器的信任水平和机器能力之间的匹配。首先分析了人机信任的影响因素，基于此给出了人机信任所具有的特征以及演化规律；在此基础上，引入机器当前表现作为机器性能的重要评估指标之一，进而构建机器性能驱动的人机信任模型，并且给出合理性说明以及参数化方法；最后，设计实验验证了本章构建的人机信任模型的有效性。本章所提出的模型是后续两章的无人机控制设计所需要的基础。

3.1　引言

机器性能是影响人类对于机器信任的最重要因素。人类对于机器的信任取决于机器呈现给人类的信息、人类自身以及环境因素。有研究表明[84]机器的外形、交互界面设计、机器的系统透明度、人类性别、种族、任务环境等因素都会影响人类对于机器的信任水平。尽管人机信任受到多种因素影响，机器的性能一直是影响人机信任的最关键因素。这是因为人机系统的目的是为了完成给定任务，那么与完成任务最相关的机器性能自然也就是最重要因素。Hancock et al.[85]运用 Meta 分析方法对现有文献中信任与影响因素的关系进行相关性分析，从数据分析的角度给出了机器的性能对信任的发展贡献最大的结论。

无人机在竞速飞行过程中，所面临的环境是在不断变化的，机器的当前表现会对人机信任产生重要影响。无人机在穿越不同障碍物时，障碍物的大小、形状、方位等是不同的；此外，在不同的地点，所受到的气流干扰、摄像头接收到的图片信息等均不相同。由于环境的不断变化，机器在做决策时所拥有的信息便是不断变化的，也就是说机器当前时刻所做的决策中存在着历史时刻所没有的信息，如果只将机器的历史表现作为机器的性能，那么对于机器性能的刻画就不够完备。此外，人类对于机器的决策历史会产生遗忘，机器当前的状态更容易引起人类的关注，进而改变人类对于机器的信任水平。

对于无人机竞速而言，机器当前表现的刻画是有可能的。因为无人机在竞速过程中，无人机的位置、周围环境等信息都能够被获得，当机器给出飞行指令后（例如飞行速度），可以根据无人机当前环境直接判断飞行指令是否合适，进而给出机器当前表现的评估。

综上所述，为了构建适用于无人机竞速飞行的信任模型，本章引入机器当前

表现的评估，利用机器的当前表现和历史表现共同刻画机器性能，进而构建人机信任演化的模型，促进人对于机器的信任水平和机器能力之间的匹配。

本章结构安排如下：3.2节给出了人机信任模型的构建，包括人机信任的特征以及演化规律、演化方程和合理性说明、参数化方法以及与无人机竞速场景的耦合关系；3.3节给出了信任模型的实验验证，具体有：实验配置、实验结果与分析等；最后3.4节给出了本章研究内容的总结。

3.2　人机信任模型的构建

构建信任模型应当遵循人机信任的独特属性及其固有的发展规律。为此，本节首先提炼出人机信任的核心特征及其动态演化规律，继而在这些理论基础上，构建人机信任动态演化的数学模型，给出模型构建的合理性说明。随后，给出模型的参数化方法，并说明该模型与无人机竞速场景的耦合关系。

3.2.1　人机信任的特征及其演化规律

1. 人机信任的特征

人机信任指的是人类对于机器的能力认知，即：在具有不确定性和风险的环境中，人类认为机器能够完成任务的可能性。在探究人机信任特征时，有必要充分考量决定信任水平的关键因素，从而揭示人机信任的内在规律与特性。依据当前人机信任研究领域的共识[20]，存在两个关键因素，这些因素对人机信任的建立有着实质性的作用：

- 人类偏置因素。鉴于个体间的认知差异，即便是面对同样的机器，人与机器之间的信任程度也可能各异。Hu et al.[30]通过一项涉及 330 余名参与者的实验证明了民族、文化背景、性别等因素对人机信任产生显著影响。
- 机器性能因素。性能因素主要体现在机器在执行任务过程中的实际表现上，如果机器性能越高，那么人类对于机器的信任水平就越高，它是影响信任动态变化的核心维度。

在上述两类影响因素中，人类偏置因素是静态的，不会随着时间发生改变，记为 $\boldsymbol{B}$。而机器性能在人与机器交互过程中会随着时间动态变化的，不仅直接影响到初次建立信任的程度，而且在人机系统的运作中会不断塑造和调整信任关系。为了方便后续的信任演化模型的建立，这里给出机器性能的定义和表示方法。

机器性能定义为人类所能够感知到的机器能力。机器性能有两种理解：一种是机器实际性能，这是机器内部的算法与决策任务共同决定的客观现实；另一种是机器根据任务情况，做出决策之后，人类根据机器决策表现所认识到的机器能

力，也就是人类主观上认识到的机器能力。需要注意的是，机器客观实际能力对于人类来说是难以观察和理解的，例如机器算法中使用到的神经网络具有黑箱特性，人类无法明确其能力边界，因此本文中所讨论的机器性能均为人类主观上认识到的机器能力。

机器性能一般通过机器的实际表现来评估，机器表现可以用一个标量来表示，该标量的计算方法定义如下：

$$\boldsymbol{P} = f_{\boldsymbol{P}}(\boldsymbol{U}) \tag{3.1}$$

其中，$\boldsymbol{P}$ 表示的是机器表现，$\boldsymbol{U}$ 表示机器决策过程中可被人类感知的且与机器性能相关的状态空间，$f_{\boldsymbol{P}}$ 表示由 $\boldsymbol{U}$ 到 $\boldsymbol{P}$ 的映射关系。理论上来说，只要机器的目标任务是明确的，并且数学工具足够，那么便可以衡量机器不同的决策结果与目标任务之间的距离，因此用标量来表示机器的性能是合理可行的。对于简单、指标易量化的任务来说，映射关系 $f_{\boldsymbol{P}}$ 可以显式的表示出来，例如在自动驾驶轨迹跟踪任务中，车辆位置与预定轨迹的偏离程度、轨迹偏航角等就可以很好的衡量机器的性能；对于一些复杂任务，人类可观察的状态维度较多，映射关系 $f_{\boldsymbol{P}}$ 难以显式表达，例如无人机多机博弈过程中，有欺骗、诱导、干扰等行为存在，博弈态势复杂，机器的决策性能无法通过公式化的映射关系直接进行计算，但是映射 $f_{\boldsymbol{P}}$ 依然客观存在，只是难以显式表达。

根据上述人机信任的影响因素，可以得出人机信任具有的特征如下：

- **特征 1**：人机信任由人类和机器共同决定的。人机信任的值取决于两个方面：一个方面是不同人的个性化偏置因素，而另一方面是机器本身在执行任务过程中的具体表现。两者共同决定了人机信任的值。
- **特征 2**：人机信任是可以定量计算的。人机信任考虑到了机器表现的客观因素，由于人的个性化偏置因素并不会随着时间发生改变，在确定人的个性化影响之后，可以通过机器性能计算获得信任的具体数值。
- **特征 3**：人机信任是动态演化的。机器性能是影响人机信任的核心维度，而机器性能在人机协作过程中不是一个固定的值，因此人机信任同样是随着时间动态变化的。

2. 人机信任的演化规律

为了方便叙述人机信任的演化规律，定义三个集合：

- $\boldsymbol{E}$：信任水平（即人机信任的具体的数值）。本文使用一个实数区间 $[0, 100]$ 来表示该集合。数值越高，代表人类越相信机器。0 表示人类完全不信任机器，100 表示人类绝对信任机器。
- $\mathbb{I}_s$：信任影响因素。这类集合有两类，一类是机器的性能表现；另一类是人类偏置因素。与信任类似，采用实数区间表示，这类取值是有序的（例如

数值越高，代表机器性能越好或者人类对于机器信任偏置越大）。

- $\mathbb{T}$：信任演化过程中的时间。

据此，可以定义信任演化函数 f_ϵ：

$$f_\epsilon : \mathbb{I}_s \times \mathbb{T} \to E \tag{3.2}$$

令 $\epsilon \in E$，$I \in \mathbb{I}_s$，$t \in \mathbb{T}$，那么 $\epsilon(t) = f_\epsilon(I,t)$ 表示人类在 t 时刻时对于机器的信任水平。

由于信任具有潜在特性[20]，直接得到人机信任的演化规律是困难的，因此本节在现有研究基础之上，进一步总结提出了几个关于机器性能驱动的信任演化规律的假设：

- **假设 1**：信任水平和机器性能之间存在单调关系。机器性能越高，如果个人偏置因素不变，那么此时信任水平应当更高，即：

$$\frac{\partial f_\epsilon}{\partial I} > 0, \forall t \geqslant 0 \tag{3.3}$$

- **假设 2**：机器性能与人类信任水平差距越大，人类信任水平变化也就越快，即：

$$\frac{\partial^2 f_\epsilon}{\partial t \partial I} > 0, \forall t \geqslant 0 \tag{3.4}$$

  这是因为差距越大，人类接受到的机器性能的刺激会越高，进而更加快速的改变信任水平。
- **假设 3**：未来独立性。当前时刻的信任水平应当只取决于过去以及当下人类所感知到的机器性能，与未来的机器性能无关。
- **假设 4**：机器性能较高导致信任水平上升的速度与机器性能较低导致信任水平下降的速度不同。这是因为不同人的保守程度不同，例如对于相对较保守的人来说，其信任上升的速度是较为缓慢的，而一旦机器性能下降，其信任水平会快速下降。
- **假设 5**：信任最终应当收敛到只和机器性能以及人类偏置相关的稳态数值。

3.2.2　人机信任演化方程及其合理性

1. 信任演化方程

直接获得信任演化函数 f_ϵ 是困难的，考虑从信任更新的角度入手，为此，进行一阶泰勒展开：

$$\epsilon(t + \Delta t) - \epsilon(t) \approx \frac{\partial f_\epsilon}{\partial t} \cdot \Delta t \tag{3.5}$$

根据假设 2，$\frac{\partial f_\epsilon}{\partial t}$ 和 $I-\epsilon(t)$ 应当是单调增长的关系。Jonker 与 Treur[40] 指出人类对于机器的信任变化与影响因素和人类前一时刻信任之差成正比，人机信任领域的学者对此也是达成了共识。因而本文沿用此观点，选取：

$$\frac{\partial f_\epsilon}{\partial t}=\alpha\left(I-\epsilon(t)\right) \tag{3.6}$$

其中，$\alpha>0$ 是待定的系数，代表了信任的更新速率。

根据3.2.1，I 由两部分组成，一部分是机器的表现 P，另一部分是人类的偏置 B。由于机器在一次表现周期中是出于黑箱状态的，人类对其信任值并不会发生变化，因此需要离散化处理：

$$\begin{cases}\epsilon(t)=\epsilon(k), t\in[t_k,t_{k+1}), k\geqslant 0\\ P(t)=P(k), t\in[t_k,t_{k+1}), k\geqslant 0\\ B(t)=B, t\in[0,\infty)\end{cases} \tag{3.7}$$

此外，机器的性能评估既需要考虑历史表现，也需要考虑当前时刻的表现，二者应该分隔开来，这一点在引言中已有说明。为方便叙述，以 P_H 和 P_N 分别表示机器的历史表现和当前表现。根据式(3.5)、式(3.6)和式(3.7)，可以得到：

$$\begin{aligned}&\epsilon(k+1)-\epsilon(k)\\ &=\alpha_H(P_H(k)-\epsilon(k))\\ &+\alpha_N(P_N(k)-\epsilon(k))\\ &+\alpha_B(B-\epsilon(k))\end{aligned} \tag{3.8}$$

其中，α_H,α_N 分别表示机器历史表现和当前表现对于信任变化的影响系数，α_B 表示人类偏置对于信任变化的影响系数，$\alpha_H,\alpha_N,\alpha_B>0$，且 $\alpha_H+\alpha_N+\alpha_B<1$。$\alpha_H$ 的取值如下：

$$\alpha_H=\begin{cases}\alpha_H^+, & P_H(k)-\epsilon(k)>0\\ 0, & P_H(k)=\epsilon(k)\\ \alpha_H^-, & P_H(k)-\epsilon(k)<0\end{cases} \tag{3.9}$$

根据假设 4，$\alpha_H^+\neq\alpha_H^-$。同理，可得 α_N 的取值如下：

$$\alpha_N=\begin{cases}\alpha_N^+, & P_N(k)-\epsilon(k)>0\\ 0, & P_N(k)=\epsilon(k)\\ \alpha_N^-, & P_N(k)-\epsilon(k)<0\end{cases} \tag{3.10}$$

其中，$\alpha_N^+\neq\alpha_N^-$。

至此，完成了信任演化方程的构建，为了方便后文叙述，将该模型简称为 MPTM（Machine Performance Trust Model）。

2. 合理性说明

（1） 考虑人机信任的特征

MPTM 既包含了人类偏置因素 B，也考虑了机器性能因素 P_H 和 P_N，因而符合人机信任的特征 1；同时 MPTM 是定量化的动态模型，因而符合特征 2 和特征 3。因此，从人机信任特征的角度出发，MPTM 具备合理性。

（2） 考虑人机信任的演化规律

针对假设 1，将式(3.6)代入到式(3.5)中，并且离散化处理，得到：

$$\epsilon(k+1) = (1-\alpha)\epsilon(k) + \alpha I(k) \tag{3.11}$$

由于 $\alpha > 0$，信任影响因素 I 数值越高，信任水平 ϵ 越高，因此 MPTM 符合假设 1。

针对假设 3，利用式(3.11)进行递推计算得到：

$$\begin{aligned}\epsilon(k) &= (1-\alpha)^k\epsilon(0) \\ &\quad + \alpha[(1-\alpha)^{k-1}I(0) + (1-\alpha)^{k-2}I(1) + \cdots + (1-\alpha)^0 I(k-1)]\end{aligned} \tag{3.12}$$

可以看出，在任意 k 时刻，信任水平 ϵ 的计算公式中并未出现 k 之后的时刻，因此 MPTM 符合假设 3。

针对假设 5，对于 (3.8)，令 $P_H(k) = P_H(k+1) = P_H$，$P_N(k) = P_N(k+1) = P_N$，可以得到：

$$\epsilon(k+1) = (1-\alpha_H-\alpha_N-\alpha_B)\epsilon(k) + \alpha_H P_H + \alpha_N P_N + \alpha_B B \tag{3.13}$$

迭代计算得到 $T(k)$ 的解析表达式：

$$\begin{aligned}\epsilon(k) &= (1-\alpha_H-\alpha_N-\alpha_B)^k\epsilon(0) \\ &\quad + (1-(1-\alpha_H-\alpha_N-\alpha_B)^k)\frac{\alpha_H P_H + \alpha_N P_N + \alpha_B B}{\alpha_H+\alpha_N+\alpha_B}\end{aligned} \tag{3.14}$$

令 $k \to \infty$，得到 ϵ_{ss} 如下：

$$\epsilon_{ss} = \frac{\alpha_H}{\alpha_H+\alpha_N+\alpha_B}P_H + \frac{\alpha_N}{\alpha_H+\alpha_N+\alpha_B}P_N + \frac{\alpha_B}{\alpha_H+\alpha_N+\alpha_B}B \tag{3.15}$$

可以看出，信任最终收敛到机器性能以及人类偏置加权的定值，说明 MPTM 符合假设 5。

在 MPTM 构建过程中，直接使用了假设 2 和假设 4。至此，MPTM 满足了从假设 1 至假设 5 的各项要求。因此，基于人机信任的演化规律的视角来看，MPTM 模型具有合理性。

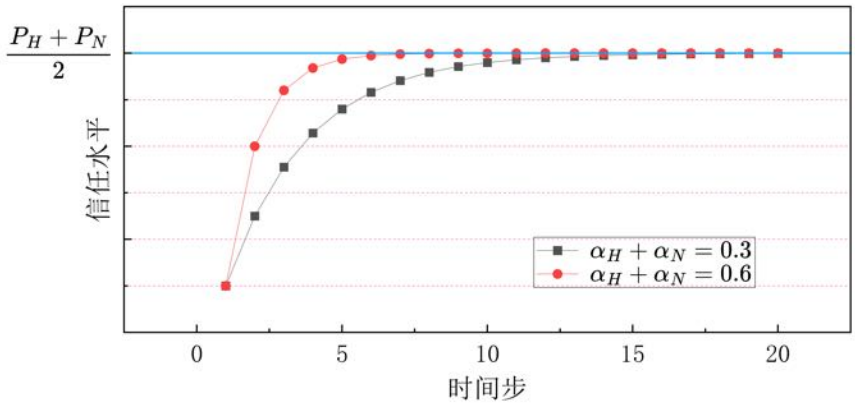


图 3.1　改变 $\alpha_H+\alpha_N$ 之和，信任水平随时间步的变化曲线（$\alpha_H=\alpha_N$，$\alpha_B=0$）

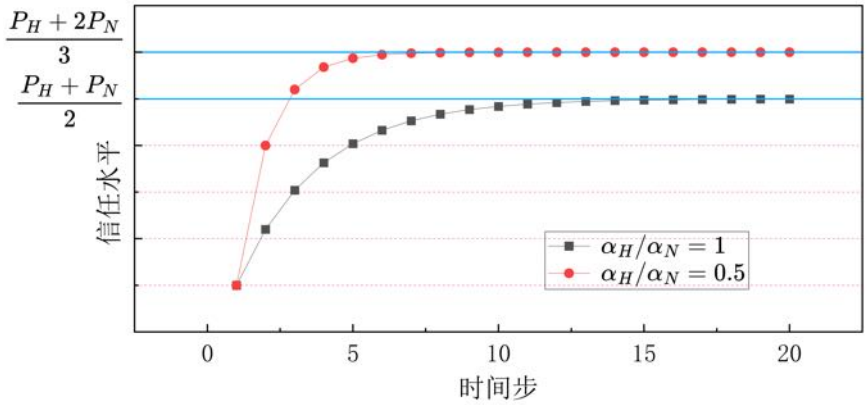


图 3.2　改变 α_H/α_N，信任水平随时间步的变化曲线（$\alpha_B=0$）

（3）　其他说明

针对不同的 α_H、α_N，绘制出相应的信任水平变化曲线，分别如图3.1、图3.2所示。

根据图3.1、图3.2和式(3.15)和可以得出几个重要结论如下：

- 图3.1表明 $\alpha_H+\alpha_N$ 代表了机器性能对于信任水平的影响速度。$\alpha_H+\alpha_N$ 数值越高，影响速度越快，人机信任水平就能更快的收敛到机器性能。
- 图3.2表明 α_H/α_N 代表了不同任务下，人类对于机器历史表现和当前表现的关注程度。当 $\alpha_H>\alpha_N$ 时，表示人类更关注机器的历史表现，当 $\alpha_H<\alpha_N$ 时，表示人类更关注机器的当前表现。
- 由式(3.15)可以看出，α_B 实际上代表了人类的理性程度。当 α_B 较小的时候，说明人类更愿意抛弃掉自己过去对于机器的错误认知，更加注重机器性能因素；如果 α_B 较大，说明人类执着己见，对机器的信任改变较难。当 $\alpha_B>1-\alpha_H-\alpha_N$ 时，说明机器的性能表现已经无法扭转人类的固执偏见，此时人类已经是不理性的了，由式 (3.14)可以看出，信任模型的稳定性也不存在，意味着模型失效。
- 至此，所有讨论对于式(3.1)中的映射 f_P 没有做出额外的假设，因此即使映

射 f_P 无法显式表达，这些结论依然成立。也就是说，无论是容易计算机器性能的简单任务，还是机器性能难以直接计算的复杂任务，该演化模型都是适用的，只是后续的参数化过程会有不同。

3.2.3　人机信任模型参数估计方法

根据式(3.1)中映射 f_P 是否可以显式表达，将参数估计方法分成两种。

方法一：若映射 f_P 可以显式表达，那么使用最小二乘估计得到式(3.8)的参数。利用映射 f_P 计算得到不同时刻的机器性能评估数值，通过信任自我报告得到不同时刻的人类对于机器的信任水平。可以使用 MATLAB 中的非线性最小二乘估计函数得到参数。

方法二：若映射 f_P 无法显式表达，那么使用神经网络来构建由 $\boldsymbol{U}$ 到 ϵ 的映射，该映射即为信任的演化模型，通过梯度下降算法更新演化模型中的参数。映射 f_P 无法显式表达，此时无法直接得到机器性能的评估数值，但是对于具有高级推理和感知能力的人类来说，机器性能是可以被感知的，只是难以量化输出。实际上，机器性能 $\boldsymbol{P}$ 由感知状态空间 $\boldsymbol{U}$ 决定，信任水平 ϵ 由式(3.8)决定，可以看出，由 $\boldsymbol{U}$ 到 ϵ 的映射是存在的。而 $\boldsymbol{U}$ 和 ϵ 都是可以直接得到的，考虑到神经网络强大的拟合功能，进而可以利用神经网络来拟合由 $\boldsymbol{U}$ 到 ϵ 的映射，通过梯度下降算法更新神经网络的参数。需要注意的是，相较于方法一，方法二模型精准度是不足的，成本更高，需要较多的人机交互数据，对于 $\boldsymbol{U}$ 也需要进行筛选挖掘。

3.2.4　人机信任模型与无人机竞速场景的耦合关系

无人机竞速场景对于 MPTM 中的人机信任水平的影响体现在个人偏置和机器性能两个维度上。在个人偏置层面，人类对无人机竞速环境和机器算法的认知，即个人对场景的理解和熟悉程度会间接影响他们对机器的信任水平。在机器性能因素层面，主要考虑无人机在实际飞行竞赛中的具体表现，包括无人机竞速过程中相对于参考轨道的偏离程度、无人机速度等方面的表现，这些实时数据决定了机器的性能表现 $\boldsymbol{P}$，进而影响人机信任水平。更为具体的分析见4.3.2中的单机竞速下的机器性能计算方法和5.3.1中的多机竞速下的人机信任神经网络。

MPTM 对于无人机竞速飞行也会产生影响。这实际上是 MPTM 在竞速飞行中的具体应用，是第 4 章和第 5 章的重点。简单来说，在单机竞速场景中，利用人机信任分配人机控制权限进而影响无人机飞行；在多机竞速场景中，把人机信任作为奖励函数一部分去指导无人机的训练，进而影响无人机的决策。

3.3　人机信任模型验证

3.3.1　场景描述

在现实生活实践中，辅助决策系统（Decision-Aid Systems，DAS）被广泛应用，涵盖了诸如基于疾病检测的个性化健康推荐系统、用于军事战场潜在威胁识别的智能辅助机器人等多种领域。其典型的工作流程，如图3.3所示，表现为如下过程：机器根据具体的任务目标以及实时环境条件，自主产生初步决策结果，

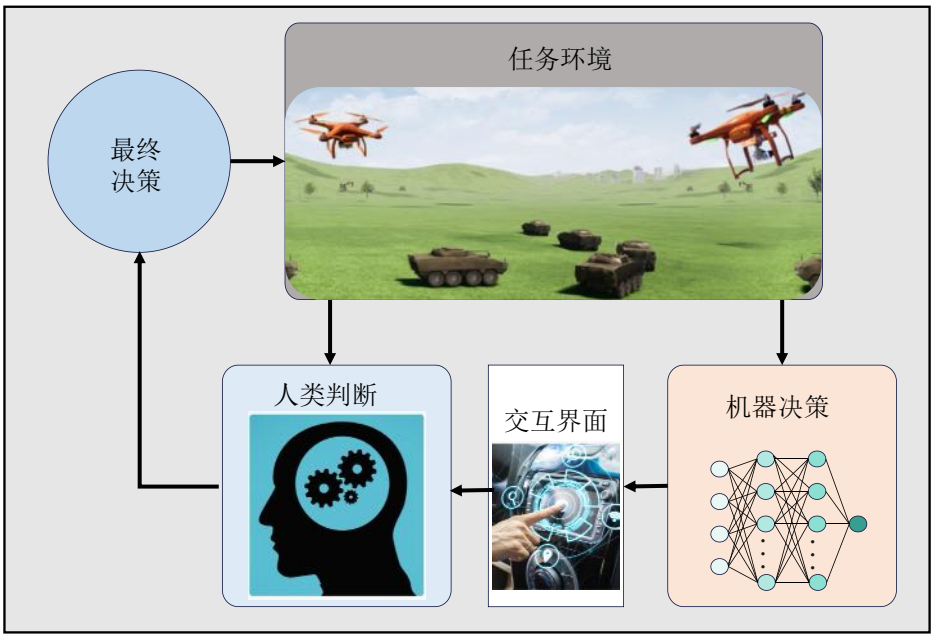


图 3.3　DAS 流程示意图

并通过人机交互界面将此决策传达给人类决策者。在此互动环节中，尽管机器提出了相应的建议，但最终的决策权与执行行为仍然牢牢掌握在人类手中，即人类可以根据机器提供的建议，自主选择采纳或否定。

在本章节中，选取 DAS 作为验证对象，基于三个关键理由。第一，DAS 的核心机制在于运用机器学习技术从一系列备选方案中甄选出最佳项，这一过程实质上可被视为机器学习领域内的一个典型分类问题。鉴于机器学习算法大体上可分为两大类型：分类算法与回归算法，而回归问题经离散化处理后亦可转化为特殊形式的分类问题，因此分类算法具有机器学习算法的广泛代表性。第二，MPTM 针对无人机竞速场景的特点，增加了机器当前表现作为机器性能的评估维度，如果其他场景能够有效刻画机器当前表现，那么 MPTM 也是适用的。第三，现有的 DAS 的信任模型都未曾考虑机器当前表现，使用 DAS 可以更好的体现融入机器当前表现的信任模型在促进信任匹配方面的优势。

3.3.2　机器性能评估计算

在 DAS 中，机器的历史决策正确与否是已知的，因此直接以机器历史决策过程中的成功率作为机器的历史表现，记作 P_H^{DAS}。对于机器当前表现评估，本章使用一种类 KNN（K Nearest Neighbors）方法进行计算。

DAS 实质上是个分类器，因此机器当前表现实际上就是机器当前决策的准确程度，也就是分类器预测效果的精确性。当评价分类器预测效果的精确性时，一个常见做法是参考分类器内建的置信度，例如神经网络模型通过 softmax 函数给出的各类别的概率分布、支持向量机中样本点距离决策边界的空间位置关系等信息。然而，已有相关研究[86-87]指出，直接使用模型产生的这类置信度衡量值可能存在校准不足的问题。即使这些置信度估值已经过有效校准处理，它们的排名也并不可靠。学术文献[88-90]的研究案例揭示，高置信度评分与分类结果的高度可靠性之间不存在必然联系。这意味着分类器自身对于其决策准确性的主观估计可能存在误差和局限性。

为解决上述问题，一种简单有效的方法是比较机器当前的决策状态与其历史正确决策状态。这一方法的基本逻辑在于：若机器当前的决策状态相较于历史正确决策状态更为相近，那么可推断当前机器作出的决策具有更高的正确可能性。为此，需要将机器过往的决策过程进行整理：对于所有的决策历史记录 (x_1^{DAS}, y_1^{DAS})、(x_2^{DAS}, y_2^{DAS}),...,(x_n^{DAS}, y_n^{DAS})，其中 $x_i^{DAS} \in X^{DAS}$ 表示样本，$y_i^{DAS} \in Y^{DAS}$ 代表该样本的正确分类，首先根据样本的分类对所有样本进行划分，为每个 $l \in Y^{DAS}$ 定义集合 H_l:

$$H_l := \{x_i^{DAS} | 1 \leqslant i \leqslant n, y_i^{DAS} = l\} \tag{3.16}$$

之后，对于 Y^{DAS} 中的每一个 l，将 H_l 中的元素组织成一个 KD 树 K_l。换句话说，就是将具有相同分类标签 l 的所有样本放入同一个集合 H_l 中，通过构建一个 KD 树结构 K_l 来高效存储和索引这些数据。

机器当前决策的性能得分的具体计算方法如下：给定当前任务 x_*^{DAS} 以及机器的决策结果 y_*^{DAS}，考虑决策历史中的所有 KD 树 K_l，计算 x_*^{DAS} 与其最近元素之间的距离 d_l。进而得到距离比值 R_*:

$$R_* = \frac{\min\{d_l | l \in Y, l \neq y_*^{DAS}\}}{d_{y_*^{DAS}}} \tag{3.17}$$

其中，$d_{y_*^{DAS}}$ 表示当前样本与机器决策结果 y_*^{DAS} 所对应的 KD 树中的最近元素之间的距离。当 R_* 较小时，意味着此时机器决策结果与历史正确决策结果距离更远，机器犯错的可能性就越高。这里 R_*^{DAS} 只考虑了最近的一个分类，而只要

d_l 比 $d_{y_*^{DAS}}$ 小, 都应当纳入计算范围之内，进一步的，定义集合 Q 如下：

$$Q := \{d_l | d_l < d_{y_*^{DAS}}, l \in Y\} \tag{3.18}$$

最终机器当前表现评估分数 P_N^{DAS}：

$$P_N^{DAS} = \begin{cases} \Sigma_{d_l \in Q} - \text{card}(Q) \cdot d_{y_*^{DAS}} & \text{if card}(Q) \neq 0 \\ R_* \cdot d_{y_*^{DAS}} - d_{y_*^{DAS}} & \text{if card}(Q) = 0 \end{cases} \tag{3.19}$$

其中，card(Q) 表示 Q 集合的元素个数。

至此，可以具体的计算 DAS 中的历史表现 P_H^{DAS} 和当前表现数值 P_N^{DAS}，再通过信任自我汇报得到信任水平，将这些数据带入到式(3.8)中，利用方法一进行参数化就能够得到 DAS 中的 MPTM。

3.3.3　实验设置

为了验证人机信任演化模型的有效性，邀请志愿者参与真实的人机交互实验。实验总体的流程如图3.4所示： 首先，机器针对当前分类任务作出决策并通

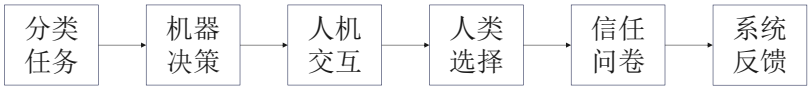


图 3.4　实验总体的流程图

过人机交互界面将结果报告给人类操作员。其次，人类操作员对机器的决策结果进行评估，并作出最终抉择。接着，操作员完成一份信任度问卷（直接通过键盘输入信任程度），以此报告他们对机器针对该特定任务决策的信任程度。最后，系统向人类操作员反馈当前分类任务的正确类别信息。

实验分为两组。第一组信任模型采用本章提出的 MPTM，志愿者可以通过交互界面接收机器的历史表现 P_H^{DAS} 和当前表现 P_N^{DAS}；第二组信任模型采用 Hu et al.[30]提出的信任模型作为对照，该模型与 MPTM 相比只考虑了机器历史表现，没有对机器的当前表现进行评估，志愿者只能通过界面接收机器的历史表现 P_H^{DAS}，将该模型记为 BTM（Baseline Trust Model）。每组均进行 70 次测试，前 50 次试验结果作为拟合数据得到信任模型的参数，后 20 次试验则用于验证该模型的预测效果。

以下是实验内容的具体设置：

- 分类任务：选用的是 scikit-learn 公开数据集中的 “Digits” 分类任务（如图3.5所示）。这是一个用于手写数字识别的经典数据集，共包含 1797 张图像[91]。

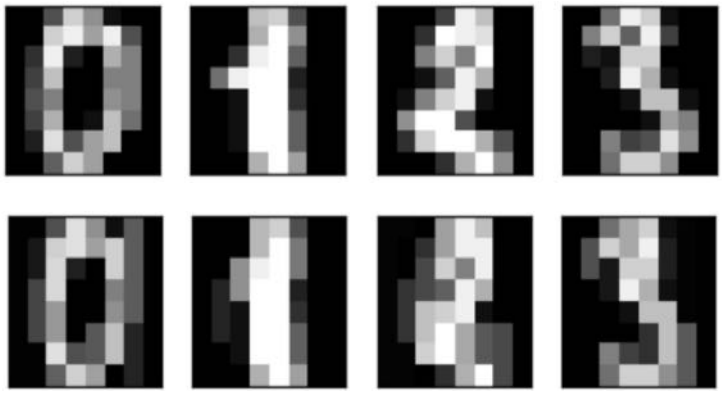


图 3.5　Digits 数据集

- 机器分类算法：朴素贝叶斯算法。在训练阶段，利用数据集前 200 个样本进行训练，使得朴素贝叶斯算法的准确率在 60% 至 85% 之间波动。
- 机器当前表现评估算法初始化的数据集：机器分类算法训练时使用到的 200 个样本及其对应的正确分类标签。
- 人机交互实验中，排除了预先训练使用的 200 个样本，从剩余的数据集中随机抽取 70 个样本进行测试。
- 在人机交互实验开始之前，向志愿者提供了关于所采用的机器决策算法类型、机器决策的准确率、机器性能估计的重要意义以及人机系统力求在最短时间内做出正确决策的目标等相关背景信息。

3.3.4　实验结果与分析

本节从信任模型的预测表现和人机系统的整体表现两个角度给出实验结果与分析。

1. 信任模型的预测表现与分析

MPTM 的信任实际值和模型预测值如图3.6所示，对照组的信任实际值和模型预测值如图3.7所示。

此外，使用平均绝对值误差（Mean Absolute Error，MAE）和均方根误差（Root Mean Square Error,RMSE）用来定量衡量预测效果，具体计算方法如下：

$$\mathrm{MAE} = \frac{1}{n}\sum_{k=1}^{n}|\epsilon(k) - \hat{\epsilon}(k)| \tag{3.20}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\epsilon(k) - \hat{\epsilon}(k))^2} \tag{3.21}$$

其中，$\hat{\epsilon}(k)$ 表示 k 时刻信任的预测值。MPTM 和 BTM 的预测误差指标的结果如表3.1所示。

表 3.1　MPTM 和 BTM 的预测误差指标

信任模型	MAE	RMSE
MPTM	4.47	6.12
BTM	4.75	6.37

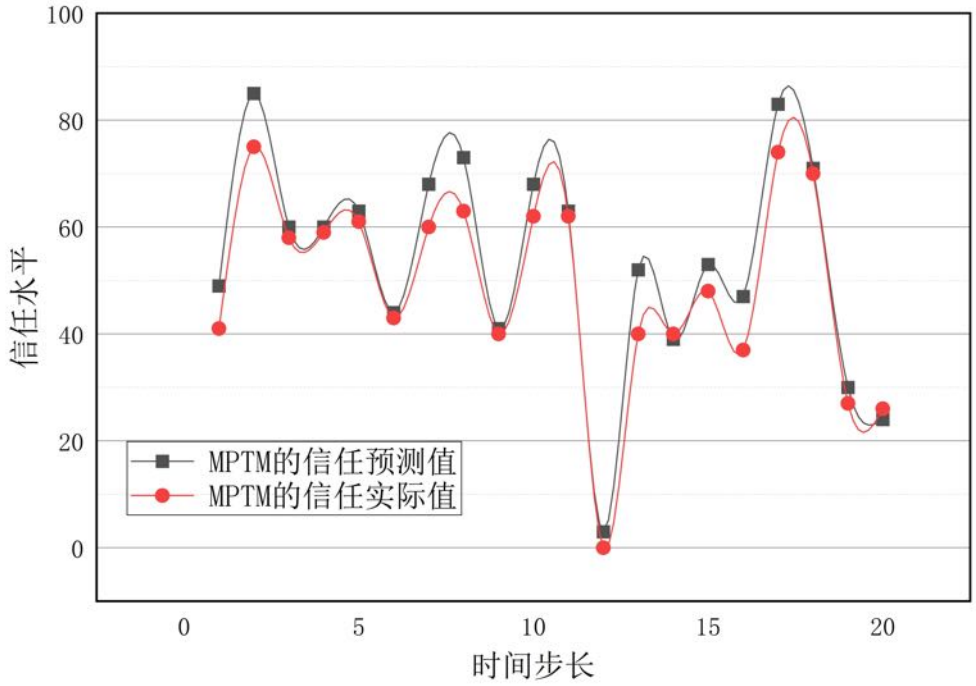


图 3.6　MPTM 的信任实际值和预测值

此外，在 DAS 中，人类决策作为人机交互体系中的决定性环节，信任模型的有效性也需要通过其对人类决策行为的预测效果进行验证与评估。因此构建“信任-动作”的映射模型，这里采用文献[41]中提出的逻辑回归算法来实现。该算法可以将信任值映射到离散动作发生的概率上，并且选择概率最大的动作值作为预测结果。在本文中，我们用数字 1 表示人类接受机器决策，用数字 0 表示人类拒绝机器决策。两组实验获得信任预测值之后，输入到“信任-动作”模型中，结果分别展示在图3.8和图3.9中。

根据图3.6和图3.7，MPTM 的信任实际值在 0 到 85 之间波动，对照组的信任实际值在 50-80 之间波动；MPTM 中的 MAE 与 RMSE 数值均小于对照组 BTM 的数值。相较于对照组，尽管 MPTM 的信任水平波动更强烈，但其预测的均方误差依然能够减小。

根据图3.8和图3.9，在总共 20 次实验中，利用 MPTM 获得的信任水平成功预测人类决策有 19 次，仅有一次误差，且此时 0 和 1 的概率仍处于同一数量级。相比之下，对照组仅正确预测了 16 次。MPTM 模型的预测错误率为 5%，而对照组的错误率为 20%。

可以看出，MPTM 的信任水平的波动更强烈，有两点原因。首先，MPTM 与

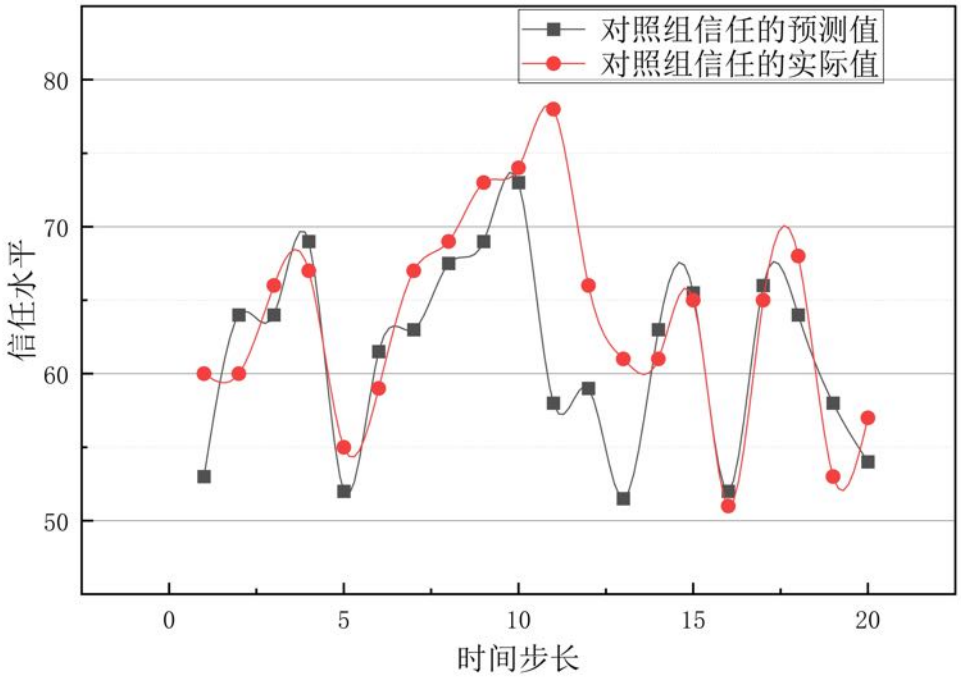


图 3.7　对照组（BTM）的信任实际值和预测值

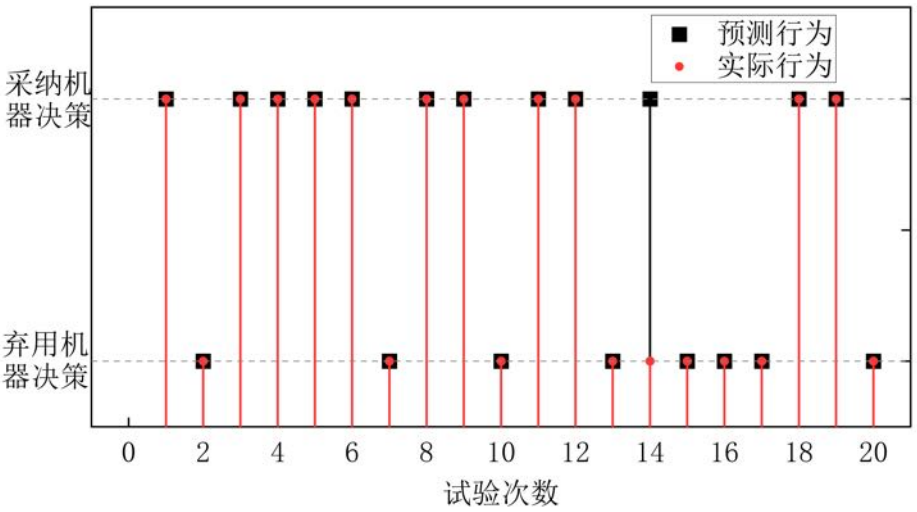


图 3.8　MPTM 的人类行为预测效果

对照组都接受了机器历史表现 P_H^{DAS}，MPTM 中人类操作员则额外感知了机器当前表现 P_N^{DAS}；P_H^{DAS} 是机器总体的准确率，正常情况下较为平稳，波动幅度不大，而 P_N^{DAS} 对于不同的任务图片而言，其数值是不相同的。其次，人类对于机器当前表现更为关注，使得人机信任水平与机器当前表现的联系更为密切。

此外，MPTM 的信任数值预测均方误差更小，并且对于人类动作预测的错误率更低，这是因为 MPTM 减少了人类自我信任评估和人类决策过程中的随机性。P_H^{DAS} 一直较为平稳，给人类的感知刺激的变化较小，人类在输出自身的信任水平时，对于 P_H^{DAS} 的关注度就会不足，进而间接的增加了其他随机因素的干扰程度（比如人类的工作时间、疲劳程度等）；P_N^{DAS} 变化较大，增加了人类对

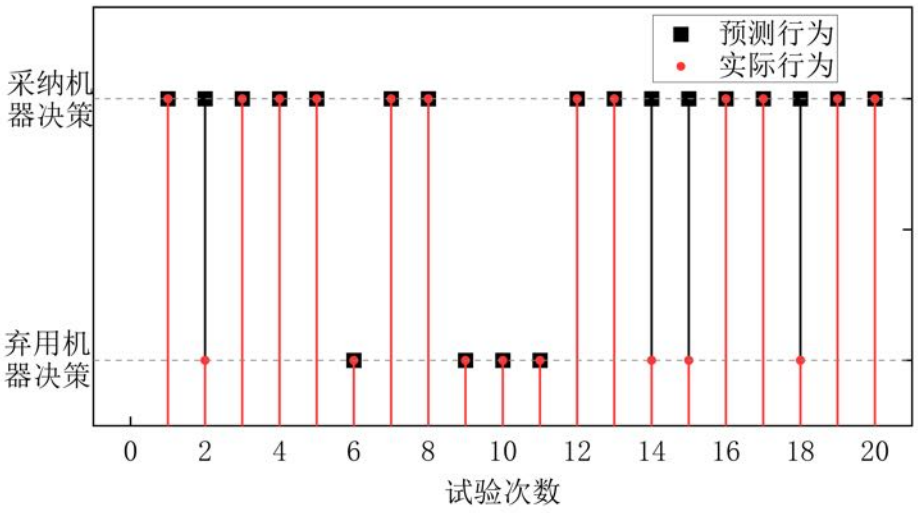


图 3.9　对照组（BTM）的人类行为预测效果

其的关注程度，引导人机信任与之跟随，因此减小了其他随机因素的干扰。人类决策过程也是相同的道理。在这两种信任模型中，随机因素并没有被建模，因此随机因素干扰小的信任模型就会获得更好的预测效果。

上述实验结果和分析有效的验证了 MPTM 的预测性能，证明了引入机器当前表现是十分必要的，能够促进人机信任与机器性能之间的匹配。一方面，从原理上来说，MPTM 的信任水平的波动更强烈，证明 MPTM 可以有效的增加人类对于机器的认知，促进了人机信任与机器性能之间的匹配，使得 MPTM 预测性能更好。人类操作者在与机器合作过程中，信任之所以波动，是因为人类感知到机器可能会出错进而降低信任水平，或者感知到机器可能决策正确进而提高信任水平。而人类感知的信息正是来源于 MPTM 中的 P_N^{DAS} 和 P_H^{DAS}，并且这类信息在后续协作中得到了检验而获得人类认可，否则人类会忽视此类信息。另一方面，从结果上来看，MPTM 信任预测模型实现了均方误差和平均绝对误差都有小幅降低，且人类动作预测错误率下降 15%，这有力地证明了该模型在信任水平预测以及人类动作预测两方面均有提升。

2. 人机系统的整体表现与分析

人机系统的整体表现的计算方法：若机器决策正确且人类选择采纳，或者机器决策错误且人类选择否定，此时记作人机系统正确决策一次；若机器决策正确且人类选择否定，或者机器决策错误且人类选择采纳，此时记作人机系统错误决策一次。

由于对照组和实验组执行的分类任务是随机选择的，因此并非完全相同，这里采用的评估标准是比较人机系统总体性能与机器单独性能之间的差异。依照上述方法，统计前 50 次实验中，MPTM 和对照组的正确率变化曲线，结果如图3.10所示，所有实验完成后最终的准确率如表3.2所示。

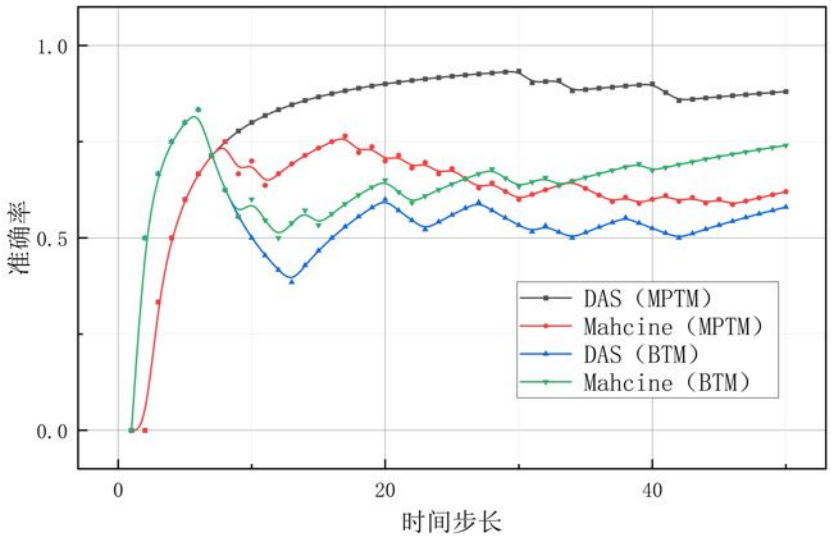


图 3.10　系统整体的准确率比较

表 3.2　MPTM 和 BTM 的系统最终表现（准确率）

信任模型	人机系统的准确率 (%)	只有机器的准确率 (%)	差值 (%)
MPTM	88	62	26
BTM	58	74	-16

根据图3.10和表3.2，在 MPTM 模型中，人机系统的整体性能超过了机器单独运行时的性能，最终准确率提升了 26 个百分点；而在对照组中，人机系统的整体性能比机器单独运行还要差，最终准确率的差值为-16，这说明人类的引入反而使得人机系统整体性能下降了。

上述两种模型中，人机系统整体表现的差异的本质原因在于 BTM 中信任水平和机器水平不能保证实时匹配；而 MPTM 由于考虑了机器当前表现，因而更能促进信任的匹配。为了更为具体的说明，绘制三种曲线：机器决策结果曲线（其中 1 代表机器做出正确决策，0 代表错误决策），人类执行结果曲线（其中 1 表示人类采纳了机器决策，而 0 表示人类否定了机器决策），将机器决策结果与人类决策结果进行同或运算得到人机系统的决策曲线（之所以是同或运算，是因为：人机系统表现好意味着即机器做出正确决策的时候，人类应当选择执行，机器做出错误决策的时候，人类应当选择否决；反之为表现差）。通过这些曲线，可以直观地对比和分析两组实验中人机协作的效果差异。在最初的 25 次实验中，绘制 MPTM 和对照组 BTM 的三种曲线，如图3.11和图3.12所示。

在对照组 BTM 中，人类对于机器表现的认知存在滞后现象，导致人机整体性能显著下降。在图3.11中，红色方框内是决策发生改变的时期，能够明显的观察到，人类决策的改变比机器决策改变慢一个时间步左右，存在滞后。当机器的决策结果发生变化时，人类相应地会倾向于调整自己的决策：例如，当机器从做

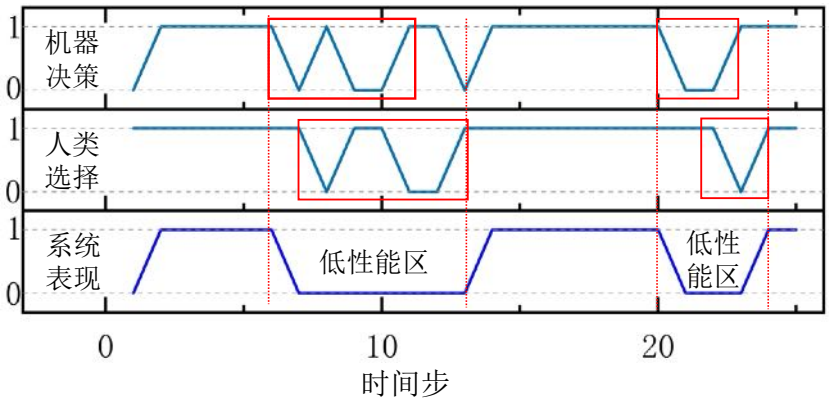


图 3.11　BTM 中的三种曲线：机器决策曲线、人类行为曲线、DAS 性能曲线

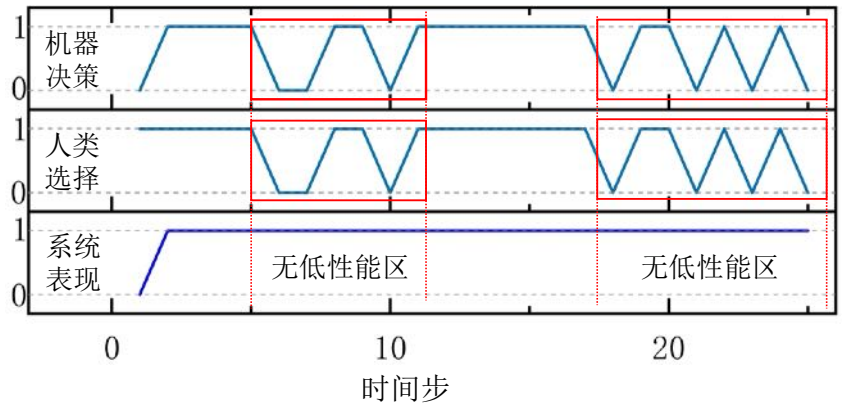


图 3.12　MPTM 中的三种曲线：机器决策曲线、人类行为曲线、DAS 性能曲线

出正确决策转向错误决策时，人类往往会在下一个任务中拒绝执行机器的决策。相反，当机器从做出错误决策转向正确决策时，人类则倾向于在下一个任务中执行机器的决策。然而，问题在于，机器的决策结果变化只存在于历史层面，但人类的决策变化却发生在当下，导致滞后效应。当机器的决策结果周期性地交替（在正确与错误之间震荡），人类对机器的信任程度也随之在过度信任和缺乏信任之间波动，导致人机整体性能显著下降。正如图3.11中的第三条曲线，总体人机低性能时期恰恰对应着机器决策结果震荡的阶段。

滞后现象的根本原因是，在对照组中，缺少机器当前表现的实时反馈。机器的历史准确率变化较小，使人难以准确评估机器当前的真实能力。因此，人类严重依赖于机器先前的决策结果来进行判断，从而产生认知偏差。在 MPTM 模型中，人决策与机器决策结果之间没有观察到滞后现象。因此，尽管 MPTM 中机器的决策结果也可能表现出波动性，但引入了机器当前表现 P_N^{DAS}，人类能够实时跟踪这些波动，进而减少认知偏差。其结果是，人与机器之间的信任水平保持在一个良好匹配的适宜状态。

上述实验结果，说明了 MPTM 通过引入机器当前表现，提升了人对于机器决策的认知水平，保证了人机信任是匹配的，进而提升了人机系统整体的性能。

由此可见在人机信任模型当中，引入当前表现的必要性。

3.4　本章小结

本章考虑到机器当前表现对于无人机竞速中的人机信任的重要影响，针对现有信任模型中缺少对于机器当前表现的刻画的问题，以机器历史表现和机器当前表现作为信任影响因素构建信任演化模型 MPTM。通过辅助决策系统验证了 MPTM 是有效的，能够促进人机信任水平和机器性能之间的匹配。

在后续的第 4 章和第 5 章中，会根据使用场景不同将本章提出的 MPTM 进一步具体化，用于无人机竞速飞行的优化设计。

第 4 章　基于人机信任的单机竞速策略

考虑到在单机竞速场景中存在的阵风等不确定性因素导致无人机竞速效果不佳的问题，本章在第 3 章所提出的信任模型基础之上，提出一种基于人机信任的单机竞速共享控制策略。首先，设计无人机 MPC 竞速算法；其次，考虑人类在介入过程中可能存在误触的问题，设计人类失误检测模块来判断是否误触，提高人类容错率；之后，将第 3 章中的 MPTM 进一步具体化，把人机信任作为仲裁依据来分配人机共享控制权限；最后，在实验平台中进行方法验证，所提方法可以有效增加无人机的竞速飞行效果，同时减轻人类的工作量，提升人类容错率。

4.1　引言

针对竞速场景中存在的阵风等不确定性因素导致无人机竞速效果不佳的问题，人机共享控制是有效的解决方案。人类和机器不同的特性决定了人类可以有效弥补机器本身的不足，而人类操作者干预无人机飞行有两种选择：介入控制和共享控制。介入控制主要为了防止人机系统发生不可接受的后果，其目的是系统的稳定和安全[16]，例如无人机遇到突发问题，机器决策模块直接失效，即将撞击地面，此时人类直接介入，机器没有任何权限。共享控制则是以系统的优化为主，其目的是提升系统的性能[76]，机器和人的地位相同，无优先级之分。本章所考虑的是单机竞速算法中存在的飞行效果不佳的问题，并非类似机器决策模块失效的极端情况，其目的是为了提升无人机的飞行效果。此外，人类尽管具有更高级的观察能力和思维能力，也就是具有更高级的策略，但是其执行能力并不一定超越机器，一个典型的例子便是人类有可能出现误触等行为[79]，这意味着人类的具体操作行为也并非最优，因而机器的决策和人类的决策应当放置在同等地位。综上所述，共享控制显然是在本章所讨论的情境下，最合适的人类干预方式。

在无人机共享控制中，人和机器的控制权限应当是动态分配的，否则人类需要持续关注人机系统并且做出相应的干预行为，这往往需要较大的工作量，同时对于人类的容错率要求极高。当人和机器的控制权限固定时，如果无人机自主飞行效果较好，那么人类应当避免自身的干预；而当无人机自主飞行效果较差时，此时人类需要持续的增加人工干预。如果人类在无人机自主飞行效果较好时仍然施加大量的干预，由于权限固定，此时人类权限并没有减小，人类操作会对无人机产生较大的影响，就可能会导致无人机偏离既有轨道，造成后续飞行效果持

续变差；同理，当无人机自主飞行效果较差时，此时人类权限并没有扩大，因此需要付出更多的干预努力，才能纠正无人机的行为。在这个过程中，可以看出，人类操作员需要时刻关注自主飞行的无人机并且采取适当的行为，不能有差错。

基于人机信任动态分配控制权限是合理的，可以有效的提升人类容错率，降低人类工作量。对于人和机器的控制权限的动态分配，可以确定的是，机器的控制权限应当和机器性能高度关联[78]，即机器权限是机器性能的递增函数。然而，只考虑机器性能依然难以明确性能与权限之间的具体关系，无法得到某一机器性能下的具体权限数值。而上述困难的本质原因在于缺少了对于人类决策的考虑，即权限分配应当同时考虑机器决策和人类决策，理想情况下，当机器性能等于人类的能力时，二者各得一半的权限。人机信任则为二者之间架起了桥梁。从 MPTM 中可以看出，信任不仅仅和性能 P 相关，也和影响系数 α 相关，而 α 便包含了人类自身的因素。比如当人类看到机器性能强于其自身能力，那么就会偏向于提高对该机器的信任水平，换句话说，在人机共享控制过程中，人类对于机器的信任必然会受到人类自身能力的影响。因而使用信任作为权限分配依据比只使用机器的性能更具有合理性。当信任水平较高时，降低人类权限，此时即使人类给予干预，对于系统的影响也比较小，人类不用“小心翼翼”；当信任水平较低时，增加人类权限，此时人类可以耗费较小的努力进而迅速调整无人机的飞行。这样便实现了人类操作者容错率的提升和工作量的下降。

综上所述，针对上述问题，本章提出了基于人机信任的单机竞速共享控制方法。一方面对人类操作是否失误进行检测，避免操作失误对于系统的不利影响；另一方面以人对于机器的信任水平作为权限动态分配依据，提升无人机竞速飞行效果，同时提升人类的容错率，降低人类工作量。

本章结构安排如下：4.2节给出了单机竞速场景的建模以及无人机参考轨迹规划方法，给出问题描述；4.3节给出了基于人机信任的共享控制算法，包括机器决策算法，人类失误检测，人机仲裁机制等；4.4节给出了相关的实验结果和分析，具体有：实验平台、实验配置、消融实验；最后4.5节给出了本章研究内容的总结。

4.2　单机竞速场景描述

4.2.1　竞速场景

竞速场景如图4.1所示：有起点和终点，起点和终点之间具有不同标号的门框，标号代表门的顺序。门具有一定的厚度和朝向，门框是实体的，无人机碰到门框后无法穿越，门的形状是正方形。除了地面和门之外，没有其他实体的赛道边界，无人机可以自由飞行。此外，无人机可以通过 GPS 获得自身的位置以及

门的位置信息。人类操作员可以通过地面端向无人机发送控制指令。

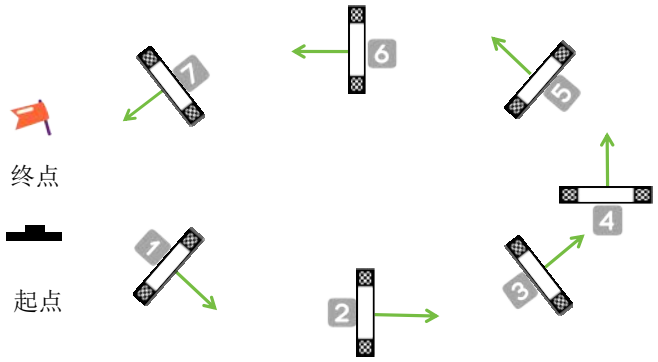
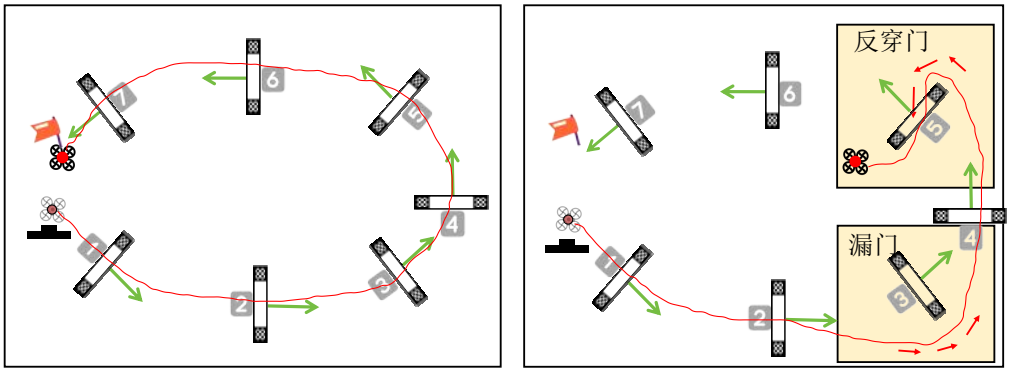


图 4.1　单机竞速环境：箭头表示门的朝向，数字表示门的顺序

无人机的任务是：从初始点起飞，尽可能的花费少的时间，按照门的顺序依次穿过，最终到达终点处。无人机完成任务的成功情形和失败情形分别如图4.2所示。



(a) 成功：按序穿过所有门，并到达终点　　(b) 失败：漏门或反穿门

图 4.2　单机竞速中，竞速失败和成功情形

4.2.2　参考轨迹规划

尽管竞速场景中仅包含门框和地面，没有实体物理障碍和清晰界定的赛道边界，然而，确定一条轨迹作为虚拟的轨道仍然具有重要意义：一方面，无人机在飞行过程中所产生的路径本身就构成了一条空间轨迹；另一方面，确定一条直观且较优的参考轨迹对于后续设计控制方法是必不可少的。

参考轨迹应当过门的中心，且参考轨迹在门中心的切线方向与门的朝向一致。对于依序的第 i 个门，其中心坐标为 $\boldsymbol{p}_i^g = (x_i, y_i, z_i)$，定义门的朝向 $\boldsymbol{G}_i$：门所在平面的单位法向量 (g_x, g_y, g_z)，该法向量需要满足：

$$\boldsymbol{G}_i{}^T(\boldsymbol{p}_{i+1}^g - \boldsymbol{p}_i^g) > 0 \tag{4.1}$$

由于无人机完成任务时必须依序穿过所有的门，那么所有门都应当在参考轨迹附近，由于无人机具有一定的体积，因此当门的中心坐标恰好落在参考轨迹上，

无人机撞门的概率最小；当参考轨迹在门中心的切线方向与门的朝向一致时，此时无人机穿门的空间最大，恰好是门的面积。

依照上述分析，利用分段插值获得参考轨迹。以门的位置 $\boldsymbol{p}_i^g = (x_i, y_i, z_i)$ 作为插值数据点，门的朝向 $\boldsymbol{G}_i = (g_x, g_y, g_z)$ 作为数据点处的切线方向向量。在每个相邻数据点 $\boldsymbol{p}_i^g$ 和 $\boldsymbol{p}_{i+1}^g$ 之间，使用如下形式的三次多项式进行插值：

$$\begin{aligned}\boldsymbol{l}_i(\mu) &= \begin{bmatrix} x_i(\mu) \\ y_i(\mu) \\ z_i(\mu) \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} 1 \\ \mu \\ \mu^2 \\ \mu^3 \end{bmatrix}\end{aligned} \tag{4.2}$$

其中 μ 是插值参数，取值范围为 $[0, 1]$，$x_i(\mu), y_i(\mu), z_i(\mu)$ 是关于参数 μ 的三个分量函数，$a_{11}, a_{12}, ..., a_{34}$ 是待求的插值系数。

为了确定每个插值段的系数，需要利用以下条件：

$$\begin{aligned} &\boldsymbol{l}_i(0) = \boldsymbol{p}_i^g \\ &\boldsymbol{l}_i(1) = \boldsymbol{p}_{i+1}^g \\ &\frac{d}{d\mu}\boldsymbol{l}_i(\mu)\Big|_{\mu=0} = \boldsymbol{G}_i \\ &\frac{d}{d\mu}\boldsymbol{l}_i(\mu)\Big|_{\mu=1} = \boldsymbol{G}_{i+1} \end{aligned} \tag{4.3}$$

求解上述线性方程组，可以得到插值的系数，进一步整理得到插值多项式如下：

$$\boldsymbol{l}_i(\mu) = \begin{bmatrix} \boldsymbol{p}_i^g, \boldsymbol{G}_i, \boldsymbol{p}_{i+1}^g, \boldsymbol{G}_{i+1} \end{bmatrix} \begin{bmatrix} 2\mu^3 - 3\mu^2 + 1 \\ \mu^3 - 2\mu^2 + \mu \\ -2\mu^3 - 3\mu^2 + 1 \\ \mu^3 - \mu^2 \end{bmatrix} \tag{4.4}$$

最后，将所有相邻插值段的插值曲线连接起来，就可以构建整条参考轨迹如图4.3所示。

为了方便后续算法设计，将所有插值曲线形成的轨迹记为 $\boldsymbol{l}$：

$$\boldsymbol{l} : [0, l_s] \to \mathbb{R}^3 \tag{4.5}$$

其中，l_s 等于所有插值数据点的个数减一，$\boldsymbol{l}(0)$ 表示无人机的起点，$\boldsymbol{l}(l_s)$ 表示无人机的目标终点，这是一条光滑可导的曲线。

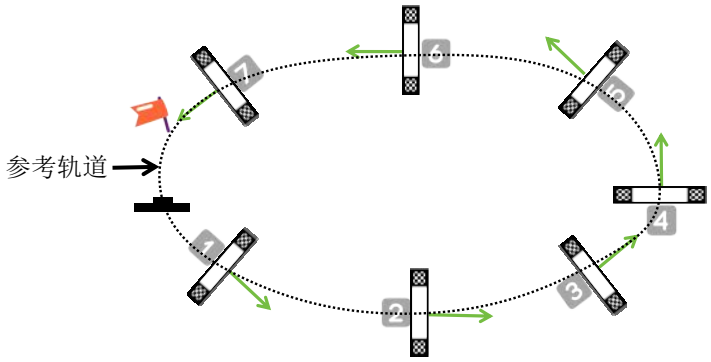


图 4.3　参考轨迹

4.3　基于人机信任的共享控制算法设计

整体的共享控制算法框架如图4.4所示。机器通过模型预测控制（Model Predictive Control，MPC）给出无人机控制指令。人类时刻注意无人机的竞速状态，并且给出自身的控制指令。机器指令和人类指令输入到仲裁模块，仲裁模块会检测人类是否操作失误，若无失误，则根据信任模型获得信任值，并且基于信任值分配机器和人类的控制权重，给出最终的无人机控制指令。无人机的底层控制算法执行控制指令，实现竞速飞行。

需要注意的是：在单机竞速环境下，外部环境对无人机来说是相对静止不变的，导致无人机的行动策略变化不大。当涉及到人机协作时，尽管图像传输等过程中存在的细微时间延迟可能会在操作策略的执行时间与频率上产生微小差异，但这些差异对竞速效果的性能影响甚微。因此，通常没有必要实施特别的干预措施来应对这类延迟问题。

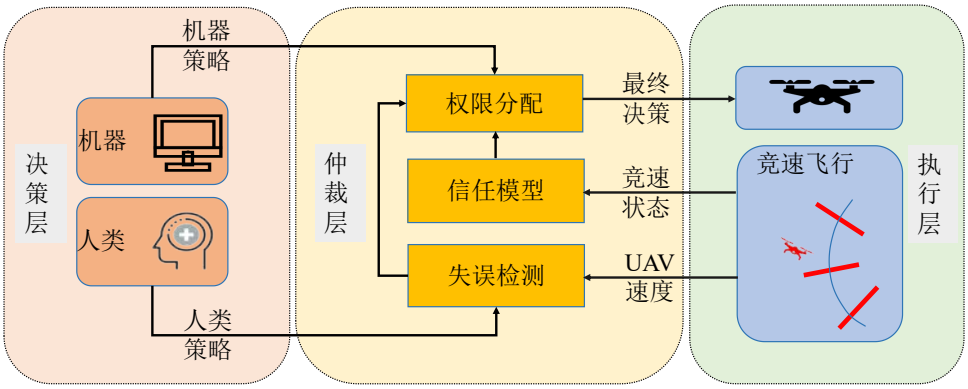


图 4.4　单机竞速共享控制框架

4.3.1　基于 MPC 的自主竞速规划

MPC 是一种用于动态系统的控制策略，具有滚动优化的特性，因而适用于高动态的无人机竞速场景。MPC 通过在一个有限的预测时间内优化控制序列来解决控制问题，其工作过程分为三步：

- 控制对象建模和定义控制目标：用状态空间模型描述系统的动态行为用于预测系统在未来时刻的行为。定义控制器的目标，包括性能指标以及特定约束条件。
- 优化控制序列：观测系统当前状态，使用系统模型和当前状态估计来预测未来一段时间内系统的状态。在预测的时间范围内，通过解决一个优化问题来计算最优的控制输入序列。优化问题的目标是控制器的目标，同时满足系统的动态约束和输入约束。
- 滚动优化：从优化结果中选择第一个控制输入，并将其应用于系统。系统随后前进到下一个时间步，根据系统的测量输出更新状态估计，在下一个时间步骤中重新开始预测和优化过程，不断更新控制输入，实现良好的控制性能。

依照上述 MPC 的工作过程，可以给出无人机中的 MPC 的设计。

首先给出无人机的动态模型。仅考虑高级控制策略，不考虑无人机底层的执行过程，因而采用式(2.3)运动学方程。无人机姿态的变化主要影响无人机的视角，对于通过视觉获得门位置的场景来说，需要进行姿态控制；而这里无人机可以通过 GPS 获得门的位置信息，因此无人机的姿态变化对于穿门影响较小。因此，为了方便后续优化设计，不妨通过姿态控制使无人机姿态角和世界坐标系始终保持一致，此时 $\boldsymbol{R}$ 为单位阵，$\boldsymbol{W}=\boldsymbol{0}$，式(2.3)进一步的简化为质心运动学方程：

$$\dot{\boldsymbol{p}}=\boldsymbol{v} \tag{4.6}$$

接下来给出 MPC 中无人机飞行的目标函数。由于是竞速问题，如果无人机沿着参考轨道 $\boldsymbol{s}$ 走的路程越大，意味着此时无人机距离终点越近。但是无人机不可能完全按照参考轨迹飞行，为此，给出 μ_m 的定义如下：

$$\mu_m(\boldsymbol{p})=\arg\min_{\mu}\frac{1}{2}\|\boldsymbol{l}(\mu)-\boldsymbol{p}\|_2 \tag{4.7}$$

式(4.5)的含义就是从参考轨迹中找出一个离无人机最近的点，该点定义为无人机在参考轨道的位置，记为轨道中心点，其参数就是 μ_m，几何解释如图4.5所示。μ_m 越大，意味着无人机离终点越近。因此，$\mu_m(\boldsymbol{p})$ 就是 MPC 的优化目标。

最后需要考虑无人机在飞行过程中的约束条件。无人机在飞行过程中，需要不断穿门，而门中心在参考轨迹处，因此，无人机任意时刻的位置距离参考轨迹的偏离不应过大，需要满足下式：

$$2\,\|\boldsymbol{l}(\mu_m(\boldsymbol{p}))-\boldsymbol{p}\|_2<w_G \tag{4.8}$$

其中，w_G 表示门内框的宽度。

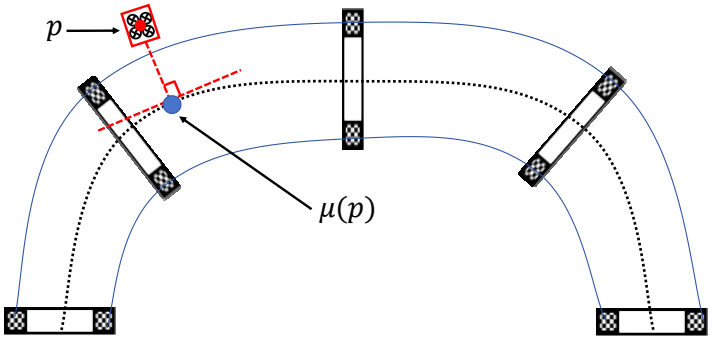


图 4.5　轨道最近点

受限于无人机本身的动力系统和外界环境（比如电池最大输出功率、空气阻力等），无人机本身的速度存在上限：

$$\|\boldsymbol{v}\|_2 \leqslant v_m \in \mathbb{R}^+ \tag{4.9}$$

其中,$\boldsymbol{v}$ 表示无人机的速度，v_m 表示无人机的速度上限。

无人机的电机存在着最高转速限制，且无人机的最大推力过大也可能导致无人机不平衡，因此无人机的加速度也有上限：

$$\|\dot{\boldsymbol{v}}\|_2 \leqslant a_m \in \mathbb{R}^+ \tag{4.10}$$

其中，a_m 表示无人机的加速度上限。

此外，式(4.6)已经构成了一个等式约束。

考虑预测未来 N 个时间步，按照采样时间 Δt 将上述所有等式和不等式离散化。定义序列 $\lambda = (\boldsymbol{p}^1, \boldsymbol{p}^2, ..., \boldsymbol{p}^N, \boldsymbol{v}^1, \boldsymbol{v}^2, ..., \boldsymbol{v}^N)$，可以表示成如下优化问题：

$$\max_{\lambda} \ \mu_m(\boldsymbol{p}^N) \tag{4.11a}$$

$$\text{s.t} \quad \boldsymbol{p}^k = \boldsymbol{p}^{k-1} + \boldsymbol{v}^k \tag{4.11b}$$

$$2\,\|\boldsymbol{l}(\mu_m(\boldsymbol{p}^k)) - \boldsymbol{p}^k\|_2 < w_G \tag{4.11c}$$

$$\|\boldsymbol{v}^k\|_2 \leqslant v_m \tag{4.11d}$$

$$\|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2 \leqslant a_m \tag{4.11e}$$

式 (4.11a)并不是 λ 的显式表达式，无法直接使用优化器求解。为此，设上一次迭代优化得到的终端位置为 $\boldsymbol{p}_m^N$，将 $\mu_m(\boldsymbol{p}^N)$ 在 $\boldsymbol{p}_m^N$ 处一阶泰勒展开，得到：

$$\mu_m(\boldsymbol{p}^N) \approx \mu_m(\boldsymbol{p}_m^N) + \frac{\mathrm{d}\mu_m}{\mathrm{d}\boldsymbol{p}^N}\bigg|_{\boldsymbol{p}_m^N} (\boldsymbol{p}^N - \boldsymbol{p}_m^N) \tag{4.12}$$

舍去其中常值部分，将曲线 $\boldsymbol{l}$ 在 μ_m 处线性化，其方向为 $\boldsymbol{l}$ 在 μ_m 的单位切向量，记为 $\boldsymbol{\tau}$ ，可以看出：

$$\frac{\mathrm{d}\mu_m}{\mathrm{d}\boldsymbol{p}^N} = \boldsymbol{\tau}^T \tag{4.13}$$

将式(4.13)代入到式(4.11)中，得到最终的优化问题如下：

$$\max_{\lambda} \ \boldsymbol{\tau}^T \boldsymbol{p}^N \tag{4.14a}$$

$$\text{s.t.} \quad \boldsymbol{p}^k = \boldsymbol{p}^{k-1} + \boldsymbol{v}^k \tag{4.14b}$$

$$2\,\|\boldsymbol{l}(\mu_m(\boldsymbol{p}^k)) - \boldsymbol{p}^k\|_2 < w_G \tag{4.14c}$$

$$\|\boldsymbol{v}^k\|_2 \leqslant v_m \tag{4.14d}$$

$$\|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2 \leqslant a_m \tag{4.14e}$$

式(4.14)是一个典型的凸优化问题，这里使用 cvxpy 凸优化函数库进行求解。

4.3.2　基于人机信任的仲裁策略

1. 单机竞速下的机器性能计算方法

在第 3 章中，构建了基于机器历史性能和当前性能的信任模型 MPTM，因此需要给出在单机竞速场景下的性能计算方法才能得到 MPTM 的具体参数。

机器的历史性能 P_H^S：考虑 n 个时间步长，计算每个时间步长的轨迹偏移量，将其累加起来作为 P_H^S。具体计算方法如下：

$$P_H^S(k) = \sum_{j=1}^{n} -\|\boldsymbol{l}(\mu_m(\boldsymbol{p}^{k-j})) - \boldsymbol{p}^{k-j}\|_2 \tag{4.15}$$

当无人机偏离轨道较远时，意味着无人机和门框相撞的概率增大，甚至于会漏掉门，导致任务直接失败，即使无人机能够及时返回到轨道中心，也耗费了更多的时间。因此将一段时间内轨道偏移累积量作为历史性能评估是合理的。

对于机器的当前性能 P_N^S 的评估，应该考虑两个角度：一是无人机向轨道中心靠近的速度，靠近的越快，性能越好；二是无人机沿着轨道行进的速度，行进的越快，性能越好。由此，给出 P_N^S 定义如下：

$$\begin{aligned} P_N^S(k) &= P_N^S(k)_1 + P_N^S(k)_2 \\ &= (\boldsymbol{\tau}^T \boldsymbol{v}^k \boldsymbol{\tau} - \boldsymbol{v}^k)^T \frac{(\boldsymbol{l}(\mu_m(\boldsymbol{p}^k)) - \boldsymbol{p}^k)}{\|\boldsymbol{l}(\mu_m(\boldsymbol{p}^k)) - \boldsymbol{p}^k\|_2} + \boldsymbol{\tau}^T \boldsymbol{v}^k \end{aligned} \tag{4.16}$$

式(4.16)有两部分，第一部分中的 $(\boldsymbol{\tau}^T \boldsymbol{v}^k \boldsymbol{\tau} - \boldsymbol{v}^k)^T (\boldsymbol{l}(\mu_m(\boldsymbol{p}^k)) - \boldsymbol{p}^k)$ 表示的是无人机指向轨道中心的速度向量与无人机指向轨道中心位置向量之间的内积。当无人机当前状态是逐渐接近轨道中心时，$P_N S(k)_1$ 为正数，接近的速度越快，$P_N S(k)_1$

的数值越高；当无人机当前状态是逐渐远离轨道中心时，$P_N S(k)_1$ 为负数，远离的速度越快，$P_N S(k)_1$ 的数值越低。第二部分 $P_N^S(k)_2 = \boldsymbol{\tau}^T \boldsymbol{v}^k$ 是无人机速度与轨道中心点的切向量的内积，代表着无人机沿着轨迹飞行的表现。当无人机沿着轨迹飞行越快，$P_N^S(k)_2$ 数值就越高，当无人机出现后退现象时，此时 $P_N^S(k)_2$ 已经为负值。因此，通过式(4.16)计算机器当前性能是合理的。

根据上述讨论，可以看出单机竞速场景下，机器的历史性能 P_H^S 与机器的当前性能 P_N^S 都是标量，并且数值越大，意味着机器表现越好，满足第 3 章中对于机器性能的所有设定要求，因而使用式(4.15)和式(4.16)来构建 MPTM 是可行的。在获得人机信任汇报之后，通过3.2.3中的方法一获得参数估计值。

2. 人类操作失误检测设计

人类操作者在通过交互界面干预无人机飞行的时候，可能存在着操作失误的现象。这里的失误是指人类操作者在执行层的失误，不考虑人类操作者策略层的失误，即人类所希望执行的飞行命令和人类实际干预时执行的命令的不同。例如人类操作者本想让无人机向左飞行，由于紧张或者不熟练等因素，错给成了向右的飞行指令；或者由于分心等原因，人类手指误触按键，导致无人机接受飞行指令。

人类操作失误会使得人类的干预对于人机系统而言是负作用，使得无人机的整体飞行效果大大降低。当机器操控无人机在既定轨道平稳飞行时，如果无人机接收到人类操作者的误触飞行指令，那么便会偏离轨道，耗费整体的竞速时间。此外，如果无人机在轨道转弯时，机身会倾斜，飞行迎风面积变大，外界干扰增加，此时人类的误触指令可能会使得无人机失去稳定性，发生撞门等现象。因此，需要检测人类操作是否失误，进而消除人类给人机系统带来的负面作用，提升整体的飞行表现。

通过预测人类操作员飞行指令所生成的无人机轨迹，并且判断该轨迹是否安全和合理，以此检测人类的失误操作。在辅助驾驶领域，有许多面部识别等方法帮助检测人类驾驶员是否分心，进而判断驾驶员的动作风险。然而，在无人机竞速飞行场景中，无人机的机载计算资源是紧张的，这使得通过面部识别来检测人类是否操作失误是难以实现的。因此，一个更为直接的方法是检测人类控制输入的有效性。具体做法是：考虑人类的控制指令 a_h，将其输入到无人机的运动学模型中，利用式(2.3)预测无人机的未来一个时间步的速度和位置，分别记为 $\boldsymbol{v}'(a_h)$ 和 $\boldsymbol{p}'(a_h)$，定义人类操作失误分数 $\eta(\boldsymbol{v}'(a_h), \boldsymbol{p}'(a_h))$ 和失误判别指标 $\delta(a_h)$ 如下：

$$\begin{aligned}\delta(a_h) &= \operatorname{sgn}(\eta(\boldsymbol{v}'(a_h), \boldsymbol{p}'(a_h))) \\ &= \operatorname{sgn}((\boldsymbol{v}_0 - \boldsymbol{v}'(a_h))^T (\boldsymbol{l}(\mu_m(\boldsymbol{p}'(a_h)) - \boldsymbol{p}'(a_h) - \boldsymbol{\tau})))\end{aligned} \tag{4.17}$$

其中，$\boldsymbol{v}_0$ 表示无人机当前的速度。式(4.17)表示的含义是：当人类的控制输入使

得无人机飞行的预测表现超过当前表现时（由于仅考虑一个时间步长，这里的预测表现通过式(4.16)计算），η 为负值，此时 $\delta = -1$, 表明人类正常干预；反之 $\delta = 1$，说明此时人类操作失误。

3. 基于信任的权限分配策略

在单机竞速过程中，人类和机器都会产生控制决策，此时需要一个仲裁机构来确定最终的控制决策。仲裁机构的输入是人类决策 a_h 和机器决策 a_m，输出是最终决策 a_F：

$$a_F = \rho(a_h, a_m) \tag{4.18}$$

其中，$\rho(\cdot)$ 表示仲裁函数。

如果人类操作员没有干预或者操作失误时，此时不考虑人类控制输入，那么机器应当获得无人机的所有控制权限，即 $a_F = a_m$。

如果人类操作员正常干预，此时 a_F 应当既有人类指令部分，也有机器指令部分。在现有的共享控制研究中，基本上都是使用线性加权的方式得到最终决策，因此这里也同样采用线性加权方式得到最终决策 a_F：

$$\begin{aligned} a_F &= \rho(a_h, a_m) \\ &= \sigma \cdot a_m + (1-\sigma) \cdot a_h \end{aligned} \tag{4.19}$$

其中，σ 表示机器的权重分配系数。由本章引言可知，σ 应当是与人类对于机器的信任水平 ϵ 的标量函数，定义如下：

$$\sigma : \epsilon \to \mathbb{R} \tag{4.20}$$

确定 $\sigma(\epsilon)$ 的具体形式是权限分配设计的关键。由于人机信任水平 ϵ 是连续的，权重分配系数也是连续的，因此，不妨假设 $\sigma(\epsilon)$ 是二阶可导的函数。

人类对于机器的信任水平越高，此时机器应当获得更高的权重系数。无人机在飞行过程中，人类操作员可以观察到无人机飞行的所有表现，因此人类操作者所感知到的机器性能可以近似等于机器的实际性能。如果机器的实际性能越高，那么人对于机器的信任水平就越高。因此，信任水平 ϵ 越高，权重应该越大，于是 $\sigma(\epsilon)$ 应当满足：

$$\frac{\mathrm{d}\sigma(\epsilon)}{\mathrm{d}\epsilon} > 0 \tag{4.21}$$

机器权重最高只能为 1，意味着此时机器拥有全部控制权限；权重最低是 0，意味着此时人类操作员拥有全部控制权限。因此权重分配函数 $\sigma(\epsilon)$ 应当具有上下确界：

$$\begin{aligned} \sup_{\epsilon \in R} \sigma(\epsilon) &= 1 \\ \inf_{\epsilon \in R} \sigma(\epsilon) &= 0 \end{aligned} \tag{4.22}$$

在 $\sigma(\epsilon_0)=0.5$ 附近时，机器和人类的权重相近，处于临界状态，由于边际效用递减规律，此时权重改变值与信任改变值的比值应当最大。而当权重分配系数靠近 1 或 0 时，此时权重改变值与信任改变值的比值应当逐渐减小，否则上下确界将不存在。因此权重分配函数 $\sigma(\epsilon)$ 应当满足：

$$\begin{aligned}\left.\frac{\mathrm{d}^2\sigma(\epsilon)}{\mathrm{d}\epsilon^2}\right|_{\epsilon=\epsilon_0}&=0\\ \frac{\mathrm{d}^2\sigma(\epsilon)}{\mathrm{d}\epsilon^2}\cdot(\epsilon-\epsilon_0)&>0\end{aligned}\tag{4.23}$$

鉴于人的指令和机器的指令地位是相同的，当信任水平 ϵ 与 ϵ_0 距离相同时，权重改变值与信任改变值的比值也应当相同。因此权重分配函数 $\sigma(\epsilon)$ 需要满足：

$$\left.\frac{\mathrm{d}\sigma(\epsilon)}{\mathrm{d}\epsilon}\right|_{\epsilon=\epsilon_0+\Delta}=\left.\frac{\mathrm{d}\sigma(\epsilon)}{\mathrm{d}\epsilon}\right|_{\epsilon=\epsilon_0-\Delta}\tag{4.24}$$

根据式(4.21)-式(4.24)，可以得出 $\sigma(\epsilon)$ 的曲线如图4.6所示。这是一条类双曲正切曲线，选择 $\left.\frac{\mathrm{d}\sigma(\epsilon)}{\mathrm{d}\epsilon}\right|_{\epsilon=\epsilon_0}=\zeta$，最终得到 $\sigma(\epsilon)$ 的函数表达式如下：

$$\sigma(\epsilon)=\frac{e^{\zeta(\epsilon-\epsilon_0)}}{e^{\zeta(\epsilon-\epsilon_0)}+e^{\zeta(\epsilon_0-\epsilon)}}\tag{4.25}$$

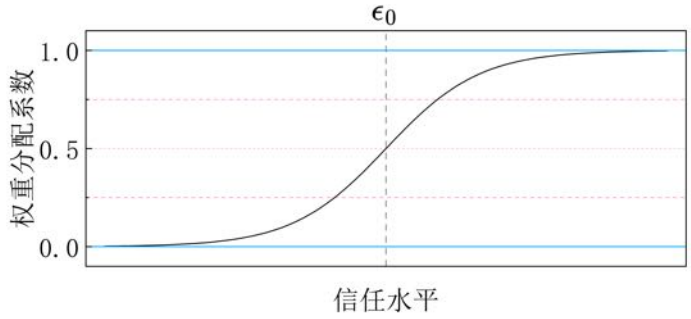


图 4.6　信任分配曲线

至此，得到在整个竞速过程中人机系统的最终决策：

$$a_F=\rho(a_h,a_m)=\begin{cases}a_m & \text{if}\quad \delta(a_h)=1\ \text{ or }\ a_h=0 \quad (4.26a)\\ \sigma(\epsilon)\cdot a_m+(1-\sigma(\epsilon))\cdot a_h & \text{if}\quad \delta(a_h)=-1 \quad (4.26b)\end{cases}$$

4.3.3　算法流程

根据图4.4，以及4.3.1-4.3.2, 可以得到单机竞速场景下的共享控制算法，具体流程如算法4.1所示。

4.4　方法验证

4.4.1　实验平台

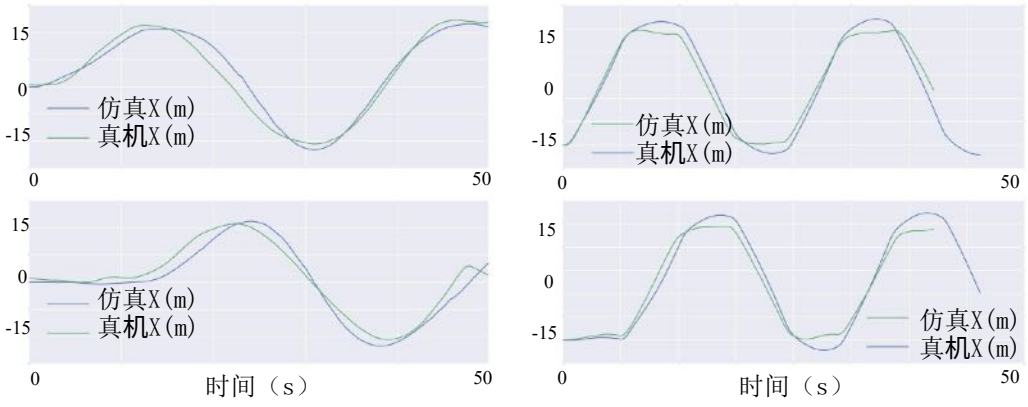
实验平台包括仿真平台和人机交互平台两部分。

算法 4.1　单机竞速场景下的共享控制算法流程

input : 竞速场景：门的参数，起点与终点位置，计算参考轨道 $\boldsymbol{l}$
无人机参数：最大速度 v_m，最大加速度 a_m
算法参数：MPC 预测时域和控制时域长度 N，采样时间 Δt，MPTM 参数，ζ，ϵ_0

while 无人机未到达设定终点 **do**
　获取无人机位置 $\boldsymbol{p}$ 和速度 $\boldsymbol{v}$;
　根据 MPC 算法给出机器决策 a_m，通过交互界面获得人类决策 a_h;
　根据式 (4.17) 计算 $\delta(a_h)$;
　if $\delta(a_h) = -1$ **then**
　　根据式 (4.16) 和式 (4.15) 计算无人机当前性能 P_N^S、无人机历史性能 P_H^S;
　　利用 P_N^S、P_H^S，根据式 (3.8) 计算人类对于机器的信任水平 ϵ;
　　通过 ϵ 和式 (4.25)，得到权限分配系数 $\sigma(\epsilon)$;
　　利用 $\sigma(\epsilon)$ 和式 (4.26b) 得到无人机最终控制指令 a_F;
　end
　else
　　根据式 (4.26a) 得到无人机最终控制指令 a_F;
　end
　将控制指令 a_F 输入到无人机中;
　等待 Δt ;
end

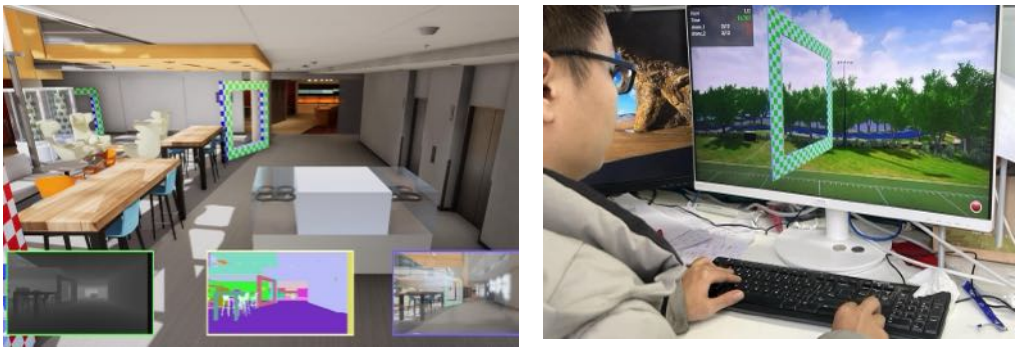
仿真平台是基于 AirSim 仿真器和 airsimneurips 库进行二次开发的人机对抗平台。该平台同时服务于科技部科技创新 2030- “新一代人工智能” 重大项目，安徽省电子产品监督检验所对该平台的有效性进行了系统评测，并且出具了测试证明。仿真平台的仿真截图如图4.8(a) 所示，包含了无人机、门框、物理实体环境、观察视角等。AirSim 是微软公司于 2017 年开发的开源跨平台仿真器，提供多种 API 接口，可实现读取数据、控制车辆、模拟天气等操作，适用于无人机和无人车的物理和视觉仿真。AirSim 基于物理引擎，考虑重力、磁场、空气压力和密度等现实因素，仿真效果与实际效果较为接近，如图4.7所示。airsimneurips 也是微软公司开发的用于无人机竞速的 python 环境，提供了多种无人机的高级控制接口，包括速度控制、转向控制、轨迹跟踪控制等。



(a) 圆形飞行　　(b) 矩形飞行

图 4.7　AirSim 仿真与真机飞行对比[92]

人机交互平台：实验开始前，人类需要通过显示器观测机器控制的无人机的竞速状态以及 P_H^S 和 P_N^S 的数值，通过键盘输出信任水平，进而得到 MPTM 的参数；实验开始后，人类仍需关注无人机的竞速状态，同时使用键盘向无人机发送控制指令。利用 keyboard 库实现发送指令的功能，如图4.8(b) 所示。人类操作员通过按下按键来表达其主要意图，按键按住的持续时间则指示了更具体的命令。例如，如果操作员对当前的左转程度不满意，则可以按住按键，按键持续时间越长，产生的命令值就越大。对应命令的增长率是每秒增加 0.8（经验数值），具体的控制方式如表4.1所示。

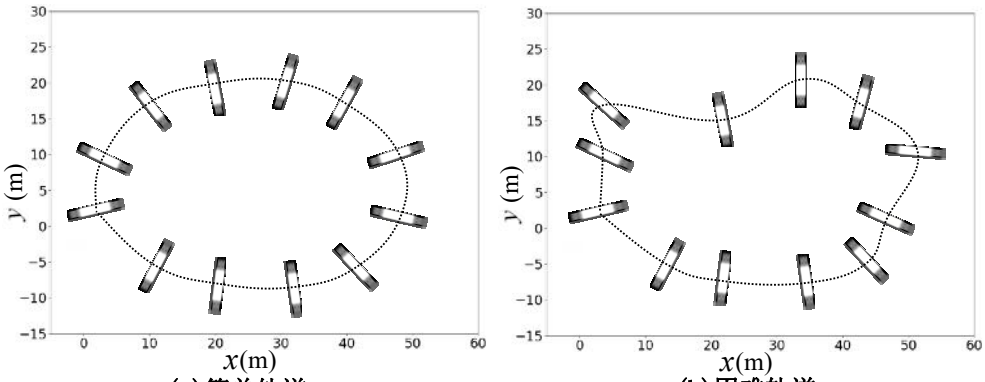


(a)仿真平台　　(b)人机交互平台

图 4.8　实验平台：仿真平台和交互平台

4.4.2　实验设置

门的朝向、位置和参考轨道的参数较多，以可视化的形式给出，如图4.9所示：考虑两种轨道，简单轨道是近似椭圆性质；困难轨道在简单轨道的基础上更改门位置，形成不规则的形状。



(a) 简单轨道　　(b) 困难轨道

图 4.9　单机竞速中门和轨道设置

无人机参数等基本的仿真实验参数的设置如表4.2所示。

为了让人机协作的效果更加明显，设定外界风力干扰：在第 7 个门到第 8 个门之间设定风力干扰，使得无人机在没有人类干预的情况下发生偏航和漏门现

表 4.1　人机交互平台的控制命令和相应的按键

控制命令	按键名称	增长率
向前飞行	W	$0.8\ s^{-1}$
向后飞行	S	
向左飞行	A	
向右飞行	D	
向上飞行	上方向键 (Key up)	
向下飞行	下方向键 (Key down)	

表 4.2　单机竞速实验参数设置

实验参数	
UAV 最大速度	$v_m = 3m/s$
UAV 最大加速度	$a_m = 2m/s^2$
UAV 初始位置	$\boldsymbol{p} = (0, 0, 0)$
UAV 控制方式	速度控制 API
UAV 大小	$0.3m$
门框内部宽度	$w_G = 1m$
门框数量	10 个
预测时域长度	10
控制时域长度	10
采样时间	$\Delta t = 0.05s$
ζ	1
ϵ 的取值范围	$[0, 100]$
ϵ_0	50

象。

人类操作员事先需要获知的信息：信任水平 ϵ 的取值范围是 $[0, 100]$，信任水平越高，机器的权重分配系数越大，在 $\epsilon = \epsilon_0 = 50$ 时，$\delta(\epsilon_0) = 0.5$。

4.4.3　实验结果与分析

本节以纯 MPC 竞速方法作为基准（baseline）方法，对比基准方法和所提共享控制算法的总体表现来说明所提方法的有效性。同时通过消融实验来说明所提共享控制算法中不同模块对于算法性能的影响。

1. 总体表现

所提方法和基准方法分别在简单轨道和困难轨道中单圈飞行 50 次，每次飞行时，门的位置和朝向会随机小幅变动。所提的共享控制方法和基准方法的漏门数量（包含反穿的情形）如表4.3所示，有风干扰时的单圈飞行轨迹和单圈飞行耗时分别如图4.10、图4.11和图4.12所示。可以看出，在有外界未知风力的干扰下，无论是简单轨道还是复杂轨道，所提方法的漏门的数量远低于基准方法；所提方法的单圈耗时也低于于基准方法；在有外界未知风力干扰时，所提方法更能够将无人机保持在门框范围之内。

因此，可以看出，所提方法将人类的自适应特性和抗干扰能力融入到了无人机的飞行控制当中，增强了飞行的鲁棒性能，能够有效的处理未知干扰，明显的

改善单机竞速场景下的飞行效果。

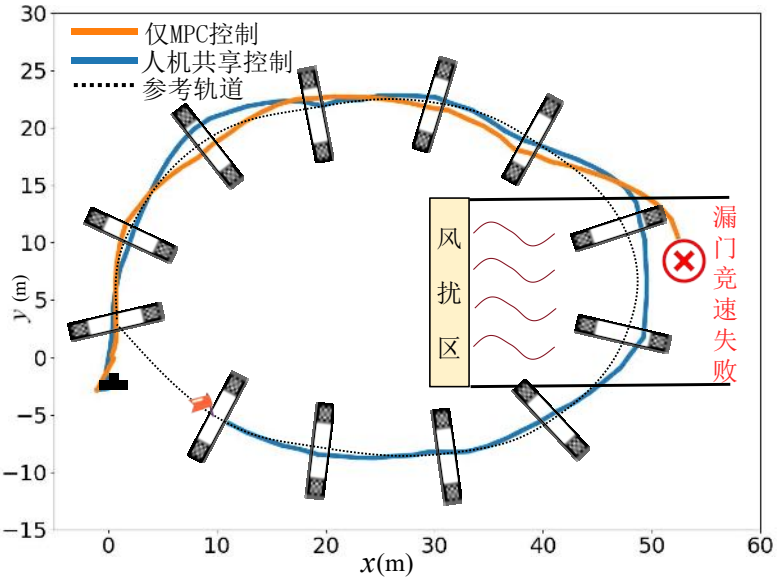


图 4.10　单机竞速简单轨道：所提方法与基准方法飞行轨迹对比

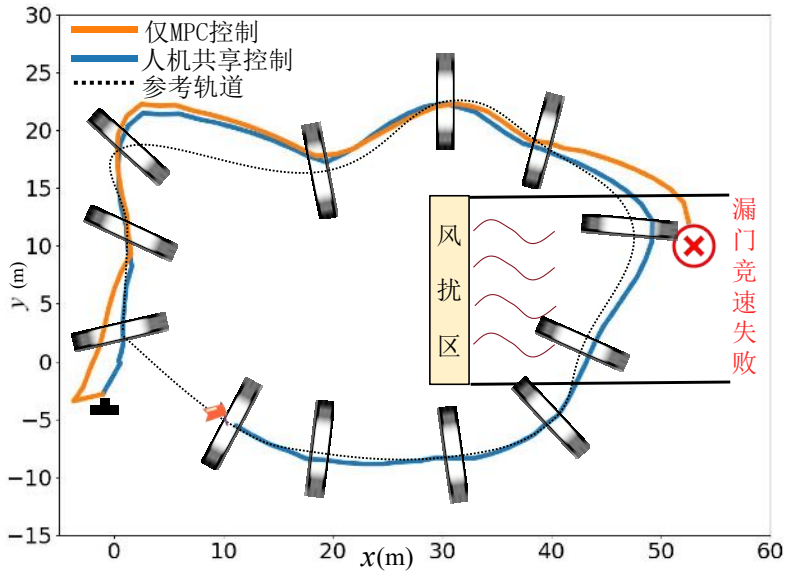


图 4.11　单机竞速困难轨道：所提方法与基准方法飞行轨迹对比

2. 消融实验

人类操作失误检测。 考虑两类失误操作：一种是在无人机飞行效果较好的时候出现失误操作，另一种是无人机飞行效果较差的时候出现失误操作。因此分成两组，有失误检测模块和无失误检测模块在简单轨道和困难轨道各飞 50 次，每次飞行门的位置和朝向小幅变动。单圈飞行轨迹（以简单轨道为例）和单圈耗时分别如图4.13、图4.14所示，漏门数量的对比如表4.4所示。

根据图4.13，有失误检测模块的时候，飞行轨迹并不会因为人类的错误干预出现变化。在没有失误检测模块的时候，如果无人机飞行效果较好，由于信任权限分配的影响，此时人类的权重系数并不多，人类的错误干预并没有让曲线发生

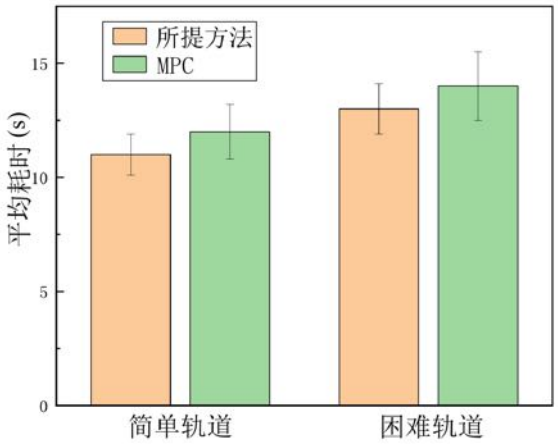


图 4.12　单机竞速：所提方法与基准方法飞行耗时对比

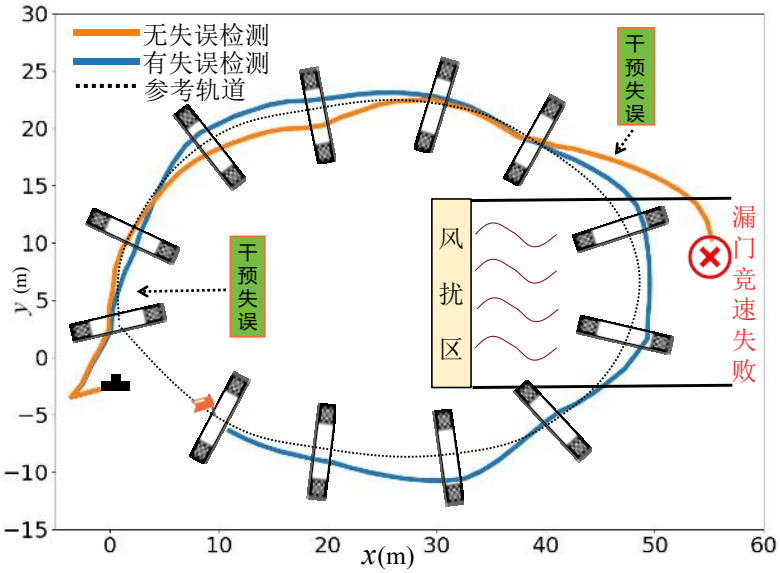


图 4.13　单机竞速的失误检测消融：飞行轨迹对比

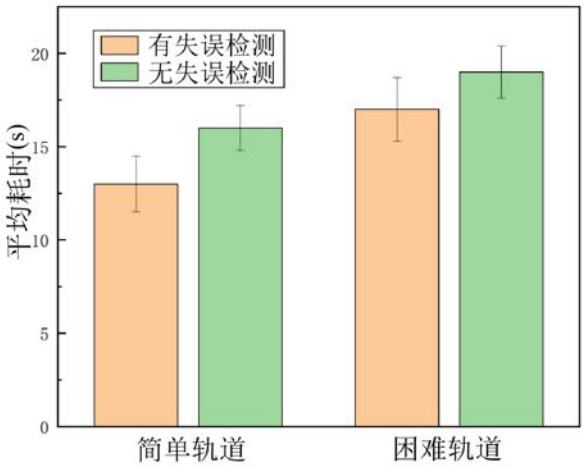


图 4.14　单机竞速的失误检测消融：飞行耗时对比

表 4.3　所提算法与基准算法的漏门数量对比

算法类型	无风情况下单圈漏门数量的均值（个）	有风情况下单圈漏门数量的均值（个）
基准算法（简单轨道）	0.1	1.5
所提算法（简单轨道）	0	0.06
基准算法（困难轨道）	0.2	2
所提算法（困难轨道）	0	0.1

表 4.4　单机竞速的失误检测消融：漏门数量对比

算法类型	无风情况下单圈漏门数量的均值（个）	有风情况下单圈漏门数量的均值（个）
无失误检测（简单轨道）	0.2	1.8
有失误检测（简单轨道）	0	0.08
无失误检测（困难轨道）	0.3	2.6
有失误检测（困难轨道）	0	0.1

过大的变化；而无人机飞行效果较差时，此时人类获得的权重很大，人类的错误干预使得无人机进一步偏离轨道，飞行耗时更多（如图4.14所示），漏掉了更多的门（如表4.4所示），飞行效果极差。

可以看出，人类操作失误检测可以避免将人类的缺点引入到竞速飞行中，防止人机整体的性能下降，因此失误检测模块是有效的且必须引入的。

基于信任的权限动态分配。 无信任动态权限分配的参数设置：机器和人类固定权重，权重比为 1：1。

考虑没有风力干扰的情况，此时无人机自主飞行的效果较好。图4.15是简单轨道下的飞行轨迹，可以看出，在有信任动态权限分配时，人类的决策权重较小，人类可以更大胆的在机器决策基础上进行微调，使得飞行轨迹与参考轨迹较为贴合。没有信任动态权限分配时，人类的权重与机器始终相同，此时人类的干预很容易使得飞行发生“超调”现象，人类需要进一步干预让飞机靠近参考轨道，于是人类干预次数上升（从图4.15可以看出，虚箭头表示一次干预），同时飞行轨迹出现震荡。因此，根据信任动态分配权限能够显著提升人类的容错率。

在有风力干扰情况下，人类单次干预平均时长如图4.16所示，可以看出有信任时，单次干预时长明显下降，说明人类工作量下降。有信任动态权限分配时，由于风力干扰导致机器表现不佳，人类获得了更大的权重，此时人类更容易将无人机调整到参考轨道附近；对于固定权限分配策略，人类权重没有扩大，人类操作员需要付出更多的时间来调整无人机，人类的工作量被迫增加。

基于以上分析，引入信任动态分配权限可以让减少人类的干预次数和干预时长，提升了人类操作的容错率，有效降低了人类的工作量。

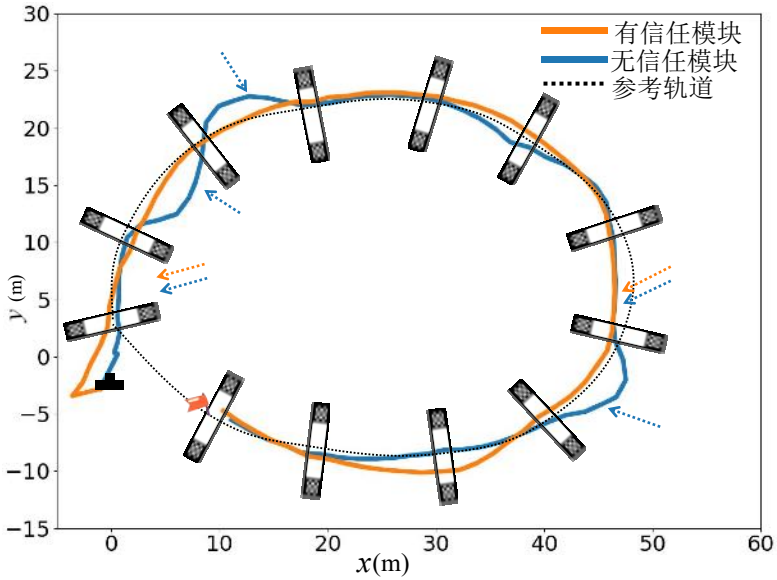


图 4.15　单机竞速信任消融：飞行轨迹对比

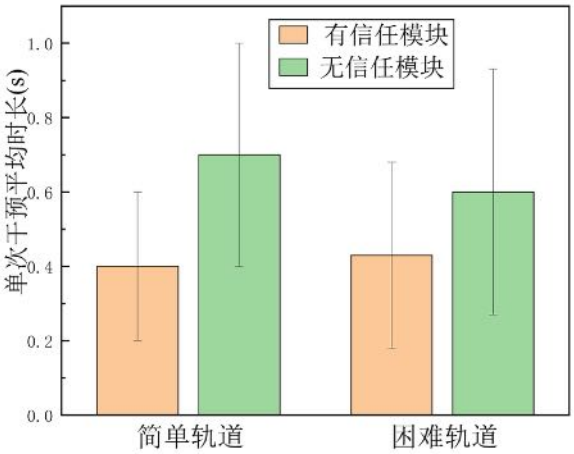


图 4.16　单机竞速信任消融：单次干预耗时对比

4.5　本章小结

本章面向单机竞速场景中存在的不确定性因素导致无人机竞速效果不佳的问题，引入人类操作者辅助无人机飞行，设计了人类操作失误检测模块和基于信任的动态权限分配策略，进而构建人机共享控制算法。本章考虑简单和困难两种单机竞速场景，在实验平台上做了多组对比实验，证明了所提方法可以明显提升无人机飞行效果；消融实验证明了人类操作失误检测可以有效避免人类的负面影响，基于信任的动态权限分配策略可以提升人类的容错率，降低人类的工作量。

第 5 章　基于人机信任的多机竞速策略

考虑到多机竞速场景中有着对于高级博弈行为刻画和实时计算的需求，本章在第 3 章的信任模型和第 4 章的单机竞速的基础之上，重点关注多机博弈竞速场景中对手的动态特性以及博弈态势的刻画，提出一种基于人机信任的多机竞速策略。多机竞速与单机竞速的本质不同之处在于，多机竞速中存在着动态变化的物体，即对手无人机，这使得多机竞速是一个高动态、强对抗、博弈强烈的过程。首先，将多机竞速问题建模为马尔可夫博弈过程；其次，构建人机信任网络作为多机竞速中的 MPTM，基于该网络引导设计奖励塑造函数；之后，利用 LSTM 网络和注意力机制来预测对手轨迹，增加对于对手策略的理解；最后在仿真环境中验证了所提方法的有效性。

5.1　引言

在多机竞速对抗的情境下，对手无人飞行器所展现出的高速动态行为给已方无人机的实时决策带来了困难。这种困难主要体现在计算资源、态势感知等方面。例如在当前时刻观察到敌方无人机的位置，在耗费一段时间优化计算之后得到决策，然而此时敌方无人机的位置早已改变，当前时刻决策的有效性大打折扣。

多机竞速过程中，博弈行为对于机器而言是无法理解的，而且人类也很难通过定量的、可解释的公式去告知机器。博弈本身是一个较为抽象的概念，其中包含了复杂的行为，比如欺骗、故意阻挡等。由于人类具有高级的抽象认知能力，对于人类而言，这些行为可以被有效认知[15]，但是人类却很难设计一个明确的数学公式让机器去理解[13]。这就会使得机器在多机竞速情境下缺少博弈智能，对于博弈态势无法做到准确评估，因而其决策往往显得“幼稚”，并非最优决策。

多机竞速中，无人机需要高速互动和快速决策，人类直接且有效地进行实时操控干预面临显著挑战。人类要想有效决策，必须通过无人机遥感图像传输系统获悉态势，而从无人机采集数据到地面控制中心接收并据此做出判断的过程便会产生不可避免的通信延迟[65]。对于没有强对抗的单机竞速而言，人类决策的有效性并不会因为通信延迟而大幅下降。但是，在多机竞速中，无人机的速度和位置瞬息万变，即使微小的通信延迟所产生的人类控制策略和机器控制策略在时间和频率上的微小差异也可能导致人类决策直接失效。此外，敌方无人机可能采取电磁干扰手段，直接切断了人机的通信，实时的人工操控无法存在。

针对这些问题，本章提出了基于人机信任引导的多机强化学习控制方法。第一，针对对手无人机的动态行为，利用 LSTM 时序网络进行预测，将预测的结果作为参考进行优化决策，增加对于对手策略的理解。第二，由于人类可以有效的感知博弈态势，多机竞速中，已方博弈态势越好，说明此时机器的实际性能越佳，由式(3.15)可知人类的信任水平也就越高，因此人类对于机器的信任可以作为博弈态势的评估，帮助机器间接的刻画了博弈行为。于是，针对机器难以评估博弈态势的问题，构建信任时序网络模型（多机竞速下的 MPTM），把人机信任融入到奖励塑造中，进而引导机器作出正确决策。第三，通过采用强化学习作为机器决策的核心框架，使得机器能直接学习人类智慧。在多机竞速场景下，这允许无人机自主掌握全部操控权限，无需依赖人类操作员的即时介入，从而有效解决了实时人工干预所面临的挑战。第四，所有模型通过神经网络进行构建，在训练完成后，无需实时求解优化问题，解决实时决策的难题。

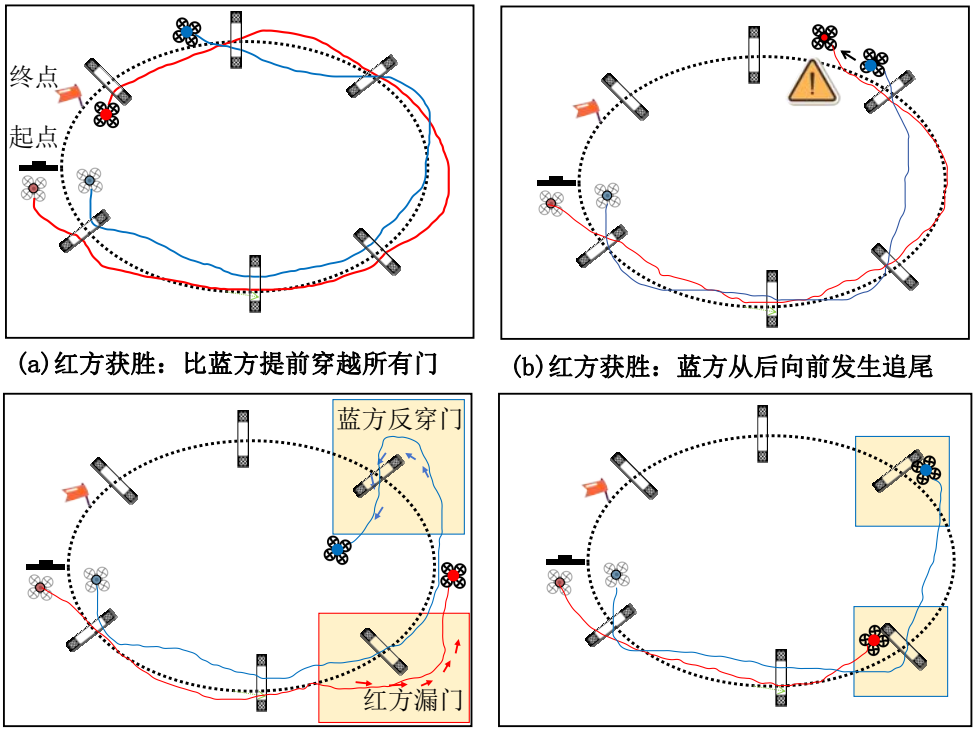
本章的结构如下：5.2节给出了多机竞速场景的建模，给出问题描述；5.3节给出了基于人机信任引导的多机强化学习算法；5.4节给出了相关的实验结果和分析，具体有：实验平台和配置、训练结果、消融实验、算法比较；最后5.5节给出了本章研究内容的总结。

5.2　多机竞速场景描述与建模

5.2.1　场景描述

竞速场景：除了对手无人机之外，其他竞速环境与单机竞速场景相同。无人机可以通过 GPS 获得自身的位置、门的位置、以及对手无人机的位置信息，且不考虑无人机接收人类远程控制指令。

无人机的任务是：从初始点起飞，尽可能的花费少的时间，按照门的顺序依次穿过，且不与对手无人机碰撞坠毁，最终到达终点处。对于碰撞坠毁，这里定义为：当己方无人机由后向前与对手无人机发生追尾，己方无人机坠毁，对方无人机继续飞行。多机竞速中，正是因为有了碰撞坠毁的约束，才有了博弈的存在。一方面，己方无人机需要避免主动碰撞敌方无人机；另一方面，己方无人机可以利用碰撞坠毁约束去阻拦、欺骗敌方无人机，敌对双方因此产生强动态博弈。无人机完成竞速任务的胜利情形（反之为失败情形）和平局情形分别如图5.1所示。



(a) 红方获胜：比蓝方提前穿越所有门　(b) 红方获胜：蓝方从后向前发生追尾

(c) 平局：双方均有漏门或者反穿门　(d) 平局：碰撞门框且停止时间超过规定时长

图 5.1　多机竞速胜利和平局的情形

5.2.2　强化学习建模

1. 部分可观马尔可夫博弈建模

强化学习任务通常使用马尔可夫博弈（Markov Game）来描述。为了方便描述，使用下标 $\cdot_o$ 表示敌方，使用下标 $\cdot_e$ 表示己方。考虑到无人机不可能观察到所有的信息，因此，将多机竞速建模为一个部分可观马尔可夫博弈六元组 $(\mathcal{S},\mathcal{A},\mathcal{O},\mathcal{T},\mathcal{R},\gamma)$。$\mathcal{S}$ 表示包括敌我无人机在内的整体环境的状态空间；$\mathcal{A}$ 表示无人机可以采取的动作空间，包括己方动作空间 $\mathcal{A}_e$ 和敌方动作空间 $\mathcal{A}_o$，也就是无人机三个方向的速度 $\boldsymbol{a}=[\boldsymbol{a}_e,\boldsymbol{a}_o]=[v_e^x,v_e^y,v_e^z,v_o^x,v_o^y,v_o^z]$，其中 $\|\boldsymbol{a}_e\|\leqslant v_m,\|\boldsymbol{a}_o\|\leqslant v_m$；$\mathcal{O}$ 表示无人机可观测空间，包括己方可观空间 $\mathcal{O}_e$ 和敌方可观空间 $\mathcal{O}_o$，可观空间中包含：敌我无人机的位置、敌我无人机速度的夹角、门的位置、我方无人机速度与当前门朝向的夹角、敌方无人机的预测位置；$\mathcal{T}$ 表示状态转移概率；$\mathcal{R}$ 表示奖励函数，这是本章设计的重点之一，在5.3.1会详细说明；γ 表示奖励折扣因子。

2. 训练环境设计

强化学习需要智能体不断的试错，逐步找到最优策略。由于成本和安全因素，无人机不可能通过现实环境直接交互试错，因此需要设计一个多无人机竞速训练环境用于智能体交互。

训练环境应当尽可能的接近现实环境，因此训练环境中应当包括无人机运

动学模型以及门框的相关信息。对于无人机运动学模型，将式(4.6)离散化为：

$$\begin{cases} x^k = x^{k-1} + v_x^k \cdot \Delta t \\ y^k = y^{k-1} + v_y^k \cdot \Delta t \\ z^k = z^{k-1} + v_z^k \cdot \Delta t \end{cases} \tag{5.1}$$

其中，x, y, z 和 v_x, v_y, v_z 分别表示无人机在世界坐标系中的位置和速度。此外，考虑到实际飞行中无人机存在加速度限制，设计速度平滑机制：

$$\begin{cases} v_x^k = \xi_1 \cdot v_x^{k-1} + (1-\xi_1) \cdot \overset{*}{v}_x^k \\ v_y^k = \xi_1 \cdot v_y^{k-1} + (1-\xi_1) \cdot \overset{*}{v}_y^k \\ v_z^k = \xi_2 \cdot v_z^{k-1} + (1-\xi_2) \cdot \overset{*}{v}_z^k \end{cases} \tag{5.2}$$

其中，$\overset{*}{v}^k$ 表示在 k 时刻的速度控制指令，ξ 表示平滑系数，考虑到无人机的动力学特性，z 方向上的平滑系数与 x, y 方向不同。

此外，还需要设置碰撞距离 d_{crash}，两架无人机相距小于 d_{crash} 时，认为无人机会发生碰撞坠毁。

5.3　基于人机信任的多机强化学习算法设计

整体的算法框架如图5.2所示。使用 Actor-Critic 作为强化学习的算法框架。

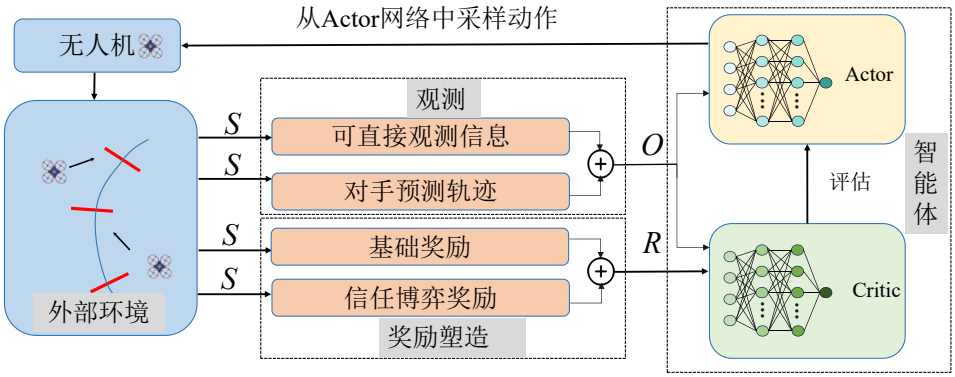


图 5.2　多机竞速强化学习算法框架

智能体的 Actor 网络通过一定的采样策略输出无人机的控制指令，无人机获得指令之后开始竞速飞行。环境状态发生改变，将环境状态分别输入到观测模块以及奖励塑造模块。在观测模块中，一方面从环境状态直接获得可观测的信息，另一方面通过敌机轨迹预测模块获得敌机未来的预测轨迹，两方面信息在一起组成观测状态，观测状态输入到智能体的 Actor 和 Critic 网络。在奖励塑造模块中，

一方面获得基础飞行奖励，另一方面利用信任模型获得博弈态势奖励，二者奖励之和作为智能体在执行一次动作之后的总奖励，该奖励会输入到 Critic 网络中。Critic 网络根据观测状态和奖励，更新一次网络参数，之后指导 Actor 网络更新一次参数。上述过程不断重复，直至智能体的网络收敛。智能体的网络收敛之后，训练出的策略网络即可部署在无人机中，实现无人机多机竞速飞行。

5.3.1　基于人机信任的奖励塑造

1. 人机信任神经网络

由于博弈态势难以通过定量的公式来表达，也就是式(3.1)中的映射关系 f_P 难以显示表达，根据3.2.3可知，可以使用神经网络来构建多机竞速下的 MPTM。对此需要选择合适的神经网络. 将式(3.8)写成映射形式：

$$\begin{aligned}\epsilon(k) &= f(\epsilon(k-1), \boldsymbol{P}(k); \alpha, \beta, \gamma)) \\ &= f(\epsilon(k-1), f_P(\boldsymbol{U}(k)); \alpha, \beta, \gamma))\end{aligned} \tag{5.3}$$

可以看出，式(5.3)与第二章中 RNN 网络中的式(2.9)输入输出关系相同，因此选用具有时序关系的 RNN 网络作为信任神经网络，记为:

$$\epsilon = F_{trust}(\boldsymbol{U}; \boldsymbol{W}_F) \tag{5.4}$$

其中，F_{trust} 表示神经网络，$\boldsymbol{W}_F$ 是神经网络的参数。

信任神经网络结构如图5.3所示，感知状态 $\boldsymbol{U}$ 输入到 RNN 中，再经过全连接层之后，得到信任水平。

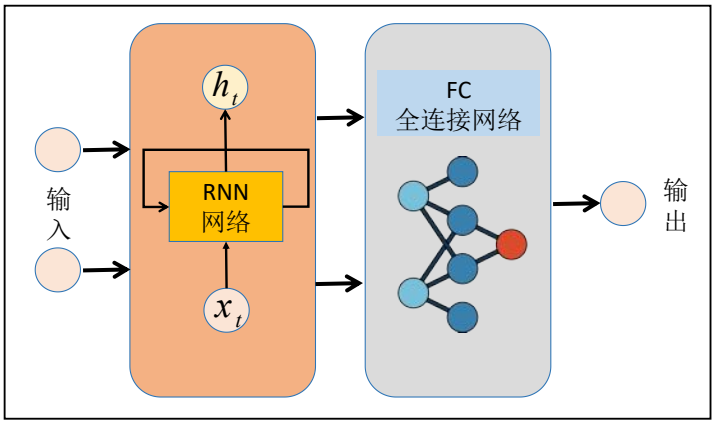


图 5.3　信任神经网络结构

对于神经网络的输入 $\boldsymbol{U}$，应当选择可以影响博弈态势的量。无人机在竞速过程中，存在着阻挡拦截这类博弈行为，因此敌我双方的位置和速度可以影响博弈态势。此外，在靠近门框时，敌我双方需要抢占穿门的有利位置，因此门的位

置和朝向也会影响博弈态势。综上所述，选择 $\boldsymbol{u}$ 作为 RNN 网络中的某时刻的样本输入：

$$\boldsymbol{u} = [\boldsymbol{p}_e, \boldsymbol{p}_o, \boldsymbol{v}_e, \boldsymbol{v}_o, \boldsymbol{p}^g, \boldsymbol{G}] \tag{5.5}$$

此外，人类的信任水平需要考虑机器的历史性能和当前性能，因此选择连续 2 个时间步的 $\boldsymbol{u}$ 作为 $\boldsymbol{U}$：

$$\boldsymbol{U}_k = [\boldsymbol{u}_{k-1}, \boldsymbol{u}_k] \tag{5.6}$$

标签对于训练神经网络 $\boldsymbol{F}$ 而言必不可少，需要人类提供。具体做法是：在仿真平台上，让敌对双方无人机竞速博弈，人类时刻关注博弈态势，并且给出其对于机器的信任水平，进而构建起信任网络的数据集。在这个过程中，需要注意几点：

- 多机竞速是在仿真平台上进行，不是在训练环境中，这是因为训练环境较为简单，而仿真平台是高度相似于现实飞行环境的，这有利于提高智能体的鲁棒性。
- 由于无人机博弈具有高动态特性，人类通过键盘输出信任值需要一定时间，因此需要改变仿真器中的时钟设置并且告知人类，同时每隔一个时间步暂停一次，等待人类评价结束后继续飞行。
- 为了防止数据不平衡导致信任网络出现偏差，需要选择不同的算法、轨道，设置不同的干扰等，让博弈态势有好也有差。由于信任网络训练数据完全来自于无人机飞行数据，如果无人机飞行过程中只出现博弈态势较好或较差的情况（即数据不平衡），在这种情况下，信任网络便失去了对于博弈态势鉴别的功能，导致机器无法正确认知自身的决策性能。
- 通过显示器观察博弈态势时，可能会有飞行视角偏差，为了帮助人类做出信任评价，通过交互界面告知竞速飞行排名。

代价函数选择均方误差代价函数，由于人类不可能做到完全精准的评价，需要避免网络过拟合，对此引入正则项。最终的代价函数如下：

$$\sum_{j=1}^{N} \frac{1}{N} \left\| \epsilon^* - F_{trust}\left(\boldsymbol{U}_j, \boldsymbol{W}_F\right) \right\|_2^2 + \left\| \boldsymbol{W}_F \right\|_2^2 \tag{5.7}$$

其中，ϵ^* 表示的是信任水平的真实值。

2. 奖励塑造

尽管无人机多机竞速的任务目标是明确的，但是状态空间太大且与之相对应的有效奖励信息太少，使得奖励十分稀疏（必须每一回合结束才能知道策略是否正确），导致智能体耗费大量精力探索无意义策略，甚至于智能体会陷入到错误决策中无法寻找到正确决策。为了避免这种情况发生，加快智能体寻找到合适

的策略，需要设计奖励塑造函数（Reward Shaping Function）去引导智能体。本节将奖励塑造分为博弈奖励塑造和基础奖励塑造两部分。

博弈奖励塑造： 由引言可知，机器博弈态势越好，人类对于机器的信任水平越高，因此直接将信任神经网络作为奖励塑造函数 $\mathcal{R}_{game}$：

$$\mathcal{R}_{game}^{k} = F_{trust}(\boldsymbol{U}^{k}; \boldsymbol{W}_{F}) \tag{5.8}$$

基础奖励塑造： 考虑到多机竞速过程中，无人机仍有单机竞速的一些要求（例如不可漏门，尽可能快的到达终点等），因此，可以将第 4 章中单机竞速的目标和约束转化为强化学习基础飞行的奖励塑造。

考虑式(4.14a), 其描述的是无人机应当沿着参考轨道飞行，为了促使智能体沿着轨道飞行，设置沿轨奖励 $\mathcal{R}_{traj}$：

$$\mathcal{R}_{traj}^{k} = \mu_{m}(\boldsymbol{p}^{k}) - \mu_{m}(\boldsymbol{p}^{k-1}) \tag{5.9}$$

$\mathcal{R}_{traj}^{k}$ 越高，说明在第 k 个时间步内，飞机沿着轨道行进越多。

考虑式(4.14c)，其描述的是无人机应当约束在参考轨道的一定范围内，防止无人机出现撞门。然而，与传统规划算法不同的是，强化学习缺少强制约束的能力，因此，将该约束更改为惩罚机制，即负值奖励 $\mathcal{R}_{w_G}$：

$$\mathcal{R}_{w_G}^{k} = -ReLU(\|\boldsymbol{l}(\mu_{m}(\boldsymbol{p}^{k})) - \boldsymbol{p}^{k}\|_{2} - w_{G}) \tag{5.10}$$

其中，$ReLU(\cdot)$ 表示线性整流单元（Rectified Linear Unit）。

等式约束(4.14b)和加速度约束(4.14e)已经体现在竞速训练环境当中，而速度约束(4.14d)可以通过在策略网络中的最后一层添加限幅函数得以实现，对此无需单独设置奖励函数。综上，得到基础奖励函数 $\mathcal{R}_{base}$：

$$\mathcal{R}_{base}^{k} = \mathcal{R}_{traj}^{k} + \mathcal{R}_{w_G}^{k} \tag{5.11}$$

最终，得到强化学习智能体在训练过程中整体的奖励塑造函数如下：

$$\mathcal{R}^{k} = \phi_{1} \cdot \mathcal{R}_{game}^{k} + \phi_{2} \cdot \mathcal{R}_{base}^{k} \tag{5.12}$$

其中 ϕ_1 和 ϕ_2 大小表示对于两种奖励的重视程度。

5.3.2　基于 LSTM 和注意力机制的对手轨迹预测

首先给出敌机轨迹预测的问题描述。选取 10 个连续时间步长作为历史序列 $\boldsymbol{Q}^{k}$：

$$\boldsymbol{Q}^{k} = \{\boldsymbol{q}^{k-9}, \boldsymbol{q}^{k-8}..., \boldsymbol{q}^{k-1}, \boldsymbol{q}^{k}\} \tag{5.13}$$

其中，$\boldsymbol{q}^k = [\boldsymbol{p}_o^k, \boldsymbol{v}_o^k]$ 表示在 k 时刻观测到的对手的运动信息，包括位置 $\boldsymbol{p}_o^k$ 和速度 $\boldsymbol{v}_o^k$。轨迹预测的任务是根据 $\boldsymbol{Q}^k$ 中的历史信息，预测接下来两个时间步长的轨迹 $\hat{\boldsymbol{Q}}^k$：

$$\hat{\boldsymbol{Q}}^k = \{\hat{\boldsymbol{q}}^{k+1}, \hat{\boldsymbol{q}}^{k+2}\} \tag{5.14}$$

其中，$\hat{\boldsymbol{q}}^{k+i} = [\hat{\boldsymbol{p}}_o^{k+i}, \hat{\boldsymbol{v}}_o^{k+i}]$ 表示 k 时刻之后的第 i 个时间步的预测的敌机运动信息。

上述问题是典型的时序预测问题，选用 LSTM 作为轨迹预测的基础网络单元。由于输入序列和输出序列不等长，编码器-解码器（Encoder-Decoder）架构可以有效处理此类问题。此外，将 LSTM 与注意力机制相结合，有助于提取时间序列的相关特征。整体的敌机轨迹预测框架如图5.4所示，将轨迹预测网络记为 Tr-Pred：历史序列进入到 LSTM 编码器中，生成隐藏状态序列，隐藏状态序列进入到注意力模块中，生成带权重的中间上下文向量，该向量输入到 LSTM 解码器模块，经过一个全连接层，输出敌机的预测轨迹。

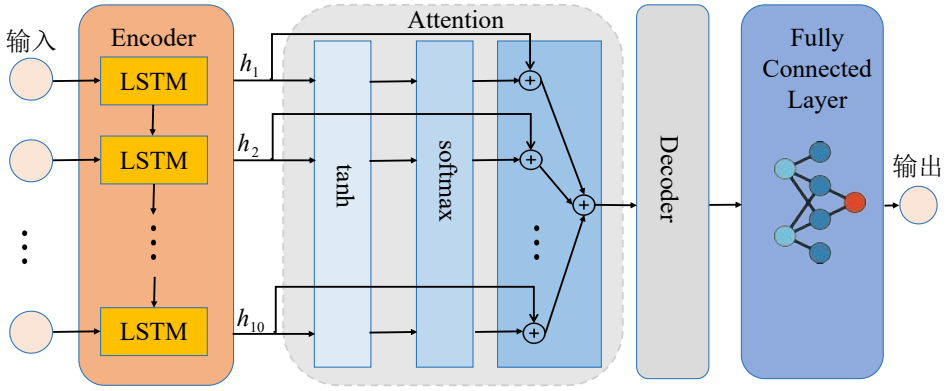


图 5.4　轨迹预测网络结构

对于 LSTM 编码器模块，根据输入序列 $\boldsymbol{Q}^k$ 计算得到隐藏状态变量 $\boldsymbol{h}_{enc}^i$：

$$\boldsymbol{h}_{enc}^i = \mathrm{LSTM}_{enc}\left(\boldsymbol{h}_{enc}^{i-1}, \boldsymbol{q}^i; \boldsymbol{W}_{enc}\right) \tag{5.15}$$

其中，$i \in [k-9, k]$，$\boldsymbol{W}_{enc}$ 表示的是 LSTM 编码器的网络参数。

对于 Attention 模块，选用线性变换用于学习轨迹之间的相互关系：

$$g\left(\boldsymbol{h}_{enc}^i, \boldsymbol{h}_{enc}^k\right) = \tanh\left(\boldsymbol{h}_{enc}^i \boldsymbol{W}_{at} \boldsymbol{h}_{enc}^k\right) \tag{5.16}$$

其中，$i \in [k-9, k-1]$，$\boldsymbol{W}_{at}$ 是待训练的参数。再经过 softmax 层计算权重：

$$\beta_i = \frac{\exp\left(g\left(\boldsymbol{h}_{enc}^i, \boldsymbol{h}_{enc}^k\right)\right)}{\sum_{j=k-9}^{k-1} \exp\left(g\left(\boldsymbol{h}_{enc}^j, \boldsymbol{h}_{enc}^k\right)\right)} \tag{5.17}$$

将权重与对应的隐藏状态相乘求和，得到上下文向量 $\boldsymbol{J}$：

$$\boldsymbol{J} = \boldsymbol{h}_{enc}^k + \sum_{i=k-9}^{k-1} \beta_i \boldsymbol{h}_{enc}^i \tag{5.18}$$

对于 LSTM 解码器模块：

$$\boldsymbol{h}_{dec}^{i} = \mathrm{LSTM}_{dec}\left(\boldsymbol{h}_{dec}^{i-1}, \boldsymbol{J}; \boldsymbol{W}_{dec}\right) \tag{5.19}$$

其中，$i \in [k+1, k+2]$，$\boldsymbol{W}_{dec}$ 表示的是 LSTM 译码器的网络参数。再经过全连接层（FC）得到预测轨迹：

$$\hat{\boldsymbol{q}}^{i} = \mathrm{FC}\left(\boldsymbol{h}_{dec}^{i}; \boldsymbol{W}_{FC}\right) \tag{5.20}$$

其中，$i \in [k+1, k+2]$,$\boldsymbol{W}_{FC}$ 表示的是全连接层的网络参数。

使用均方误差作为网络训练的代价函数，如下所示：

$$\frac{1}{N}\sum_{j=1}^{N}\sum_{i=k+1}^{k+2}\left\|\boldsymbol{q}^{i} - \hat{\boldsymbol{q}}^{i}\right\|_{2}^{2} \tag{5.21}$$

其中，$\boldsymbol{q}^{i}$ 表示 i 时刻的真实值，N 表示训练中的数据量。

5.3.3　强化学习算法训练

本章所考虑的问题中存在敌我双方两个智能体，属于多智能体强化学习问题。由于无人机竞速属于零和博弈，智能体之间不存在通信交流的情况，因此选择完全去中心化的多智能体算法 IPPO（Independent PPO）方法来训练智能体。根据图5.2，以及5.3.1-5.3.2, 可以得到多机竞速场景下的强化学习训练算法。具体流程如算法5.1所示。

算法 5.1　多机竞速的强化学习算法训练流程

input : 训练完成的信任网络 F_{trust} 与轨迹预测网络 Tr-Pred.
多机竞速场景: 门的参数，起点与终点位置，参考轨道 $\boldsymbol{l}_r$.
无人机设置：最大速度 v_m，速度平滑系数 ξ.
智能体数量：2 个（敌方无人机与我方无人机）
output 多机竞速智能体决策模型
:

超参数：折扣因子，学习率，最大训练次数 $\boldsymbol{M}$;
初始化：每个智能体的策略网络 Actor 和价值网络 Critic，随机生成 Actor 网络权重 $\boldsymbol{W}_{ac}$、Critic 网络权重 $\boldsymbol{W}_{cr}$;
for $episode: k = 1, 2, ..., M$ **do**
    智能体在环境中交互，获得轨迹 $\boldsymbol{l}_k$;
    基于轨迹预测网络 $\boldsymbol{Tr\text{-}Pre}$ 获得每个智能体的预测轨迹 $\hat{\boldsymbol{l}}_k$;
    将 $\boldsymbol{l}_k$ 和 $\hat{\boldsymbol{l}}_k$ 整合获得训练轨迹 $\boldsymbol{D}_k$;
    基于信任网络 F_{trust} 和5.12计算奖励数据 $\mathcal{R}_k$;
    对于每个智能体，根据 $\boldsymbol{D}_k$ 和奖励 $\mathcal{R}_k$，基于 PPO 算法更新一次 Critic 网络权重 $\boldsymbol{W}_{cr}$ 以及 Actor 网络权重 $\boldsymbol{W}_{ac}$;
end

5.4　算法仿真与验证

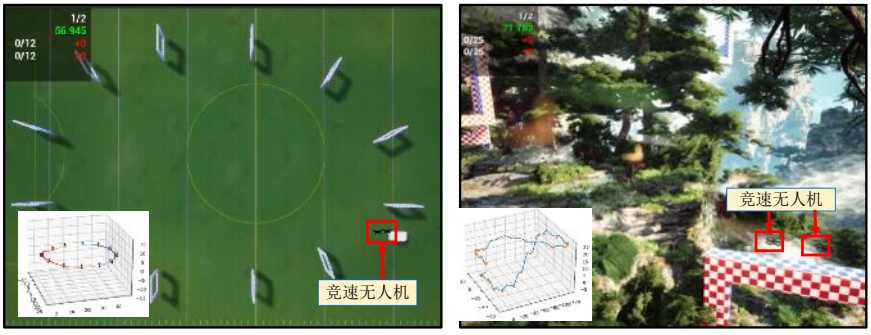
5.4.1　实验平台与实验设置

本章实验中，实验平台、无人机的相关设置以及门框宽度设置与第 4 章相同，不再重复说明。这里主要展示一些核心的实验参数，如表5.1所示。

表 5.1　多机竞速实验参数设置

实验参数		
信任网络参数	RNN 隐藏层神经元数量	hidden-size = 32
	bactch-size	32
	全连接层的神经元数量	FC-size = 16
	学习率	0.001
	网络参数更新方式	Adam
轨迹预测网络参数	LSTM 隐藏层神经元数量	hidden-size = 64
	全连接层的神经元数量	FC-size = 32
	bactch-size	64
	学习率	10^{-3}
	网络参数更新方式	Adam
IPPO 相关参数	总训练步数	1000 万步
	损失函数的熵系数	0.01
	折扣因子	0.9
	网络参数更新方式	Adam
	学习率	10^{-4}
	PPO 网络结构等其余参数	与 OpenAI 开源的基准方法相同[①]
一些训练细节	每次策略迭代前收集数据个数	2048 步
	奖励塑造函数中 ϕ_1 系数	$\phi_1 = 0.6$
	奖励塑造函数中 ϕ_2 系数	$\phi_1 = 0.4$
	单步决策时间	$0.05s$
	最小碰撞距离	$d_{crash} = 0.2m$

门的朝向、位置和参考轨道如图5.5所示：考虑两种轨道，简单轨道是二维的足球场椭圆轨道；困难轨道是三维的张家界山区轨道。同时，在每次竞速飞行时会对门的位置以及朝向进行小幅的随机调整。



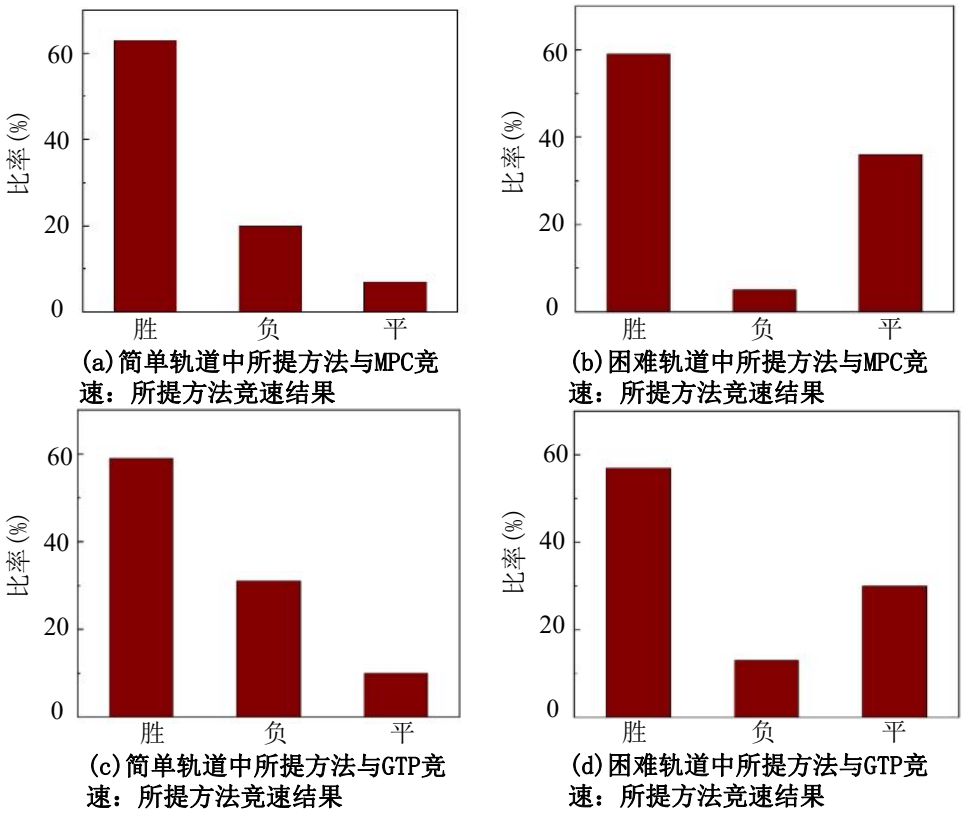
(a) 简单轨道　　(b) 困难轨道

图 5.5　两种多机竞速场景：简单轨道和困难轨道

[①]参考的内容可在托管平台 github 中找到: https://github.com/openai/baselines/tree/master/baselines/ppo2.

5.4.2　实验结果与比较

以 GTP 和 MPC（在式(4.14)中添加碰撞约束即可）作为对比方法。所提方法分别和 GTP、MPC 竞速比较，每组飞行 50 次。所提方法与 GTP、MPC 竞速的胜率情况如图5.6所示，完成一次竞速任务的平均耗时如图5.7所示。可以看出，所提方法无论在二维轨道还是三维轨道，所提方法的胜率均高于 GTP 和 MPC，平均耗时也均小于 GTP 和 MPC，说明所提方法有效。此外，在三维轨道上，所提方法的胜率与其他方法的胜率差值更大，说明所提方法更能有效的应对复杂场景。



(a) 简单轨道中所提方法与MPC竞速：所提方法竞速结果

(b) 困难轨道中所提方法与MPC竞速：所提方法竞速结果

(c) 简单轨道中所提方法与GTP竞速：所提方法竞速结果

(d) 困难轨道中所提方法与GTP竞速：所提方法竞速结果

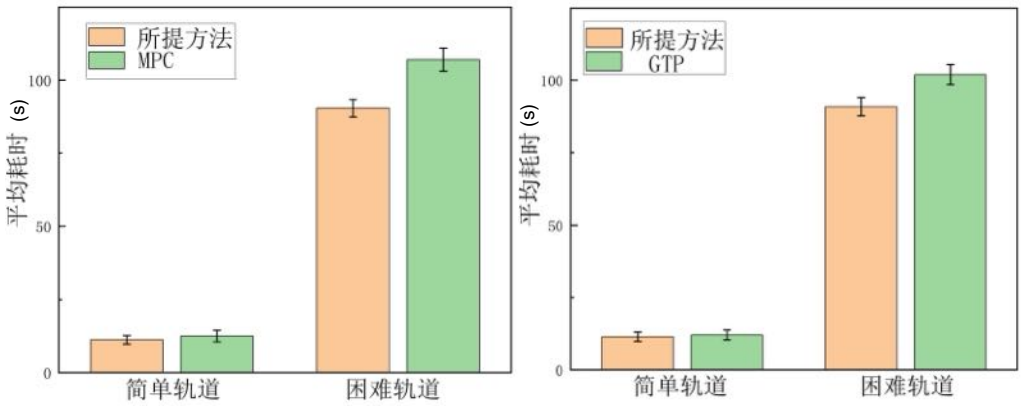
图 5.6　所提方法、GTP 以及 MPC 的胜率情况

5.4.3　消融实验

信任模块消融。取消信任网络 F_{trust} 前后，智能体训练的奖励曲线的对比如图5.8所示。可以看出，在引入信任网络之后，智能体的奖励收敛更快，说明人机信任的引入加快了智能体的训练。

平均耗时和胜率对比如图5.9(a) 图5.12(a) 所示。可以看出，引入人机信任之后，能够有效的减少飞行耗时，提高胜率。

为了深入阐述人机信任机制如何增进机器对博弈的领悟，在一个二维赛道场景中设计了两种情况：第一种是对手无人机大幅落后；第二种则是对手紧随之



(a) 所提方法与MPC比较　　(b) 所提方法与GTP比较

图 5.7　所提方法、GTP 以及 MPC 完成一次任务的平均耗时

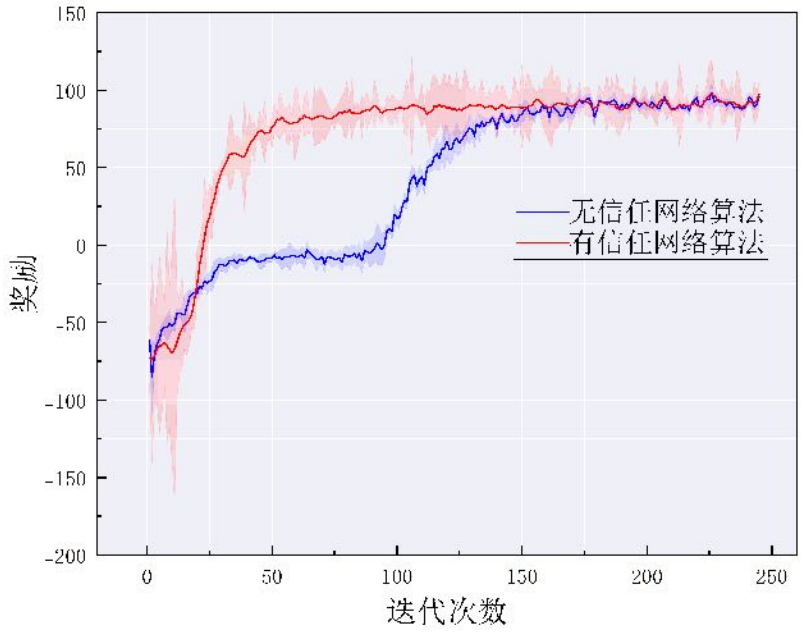
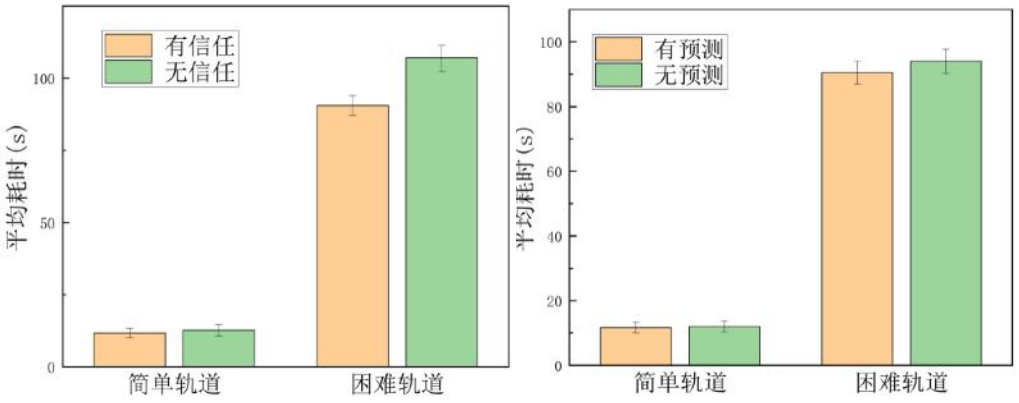
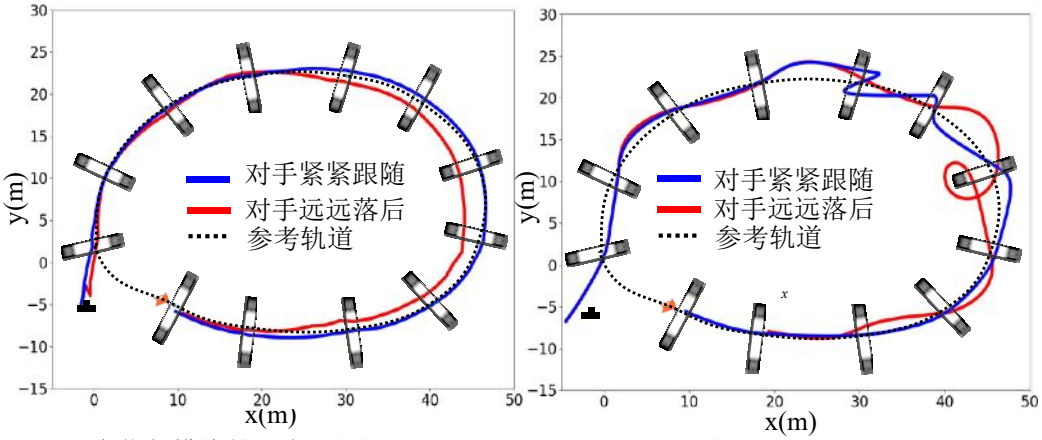


图 5.8　信任网络消融实验：训练奖励对比



(a) 信任消融：耗时对比　　(b) 轨迹预测消融：耗时对比

图 5.9　两类消融实验中的单圈平均耗时



(a)有信任模块的两类飞行场景对比　　(b)无信任模块的两类飞行场景对比

图 5.10　信任网络消融实验：两类飞行场景对比

后。观察图5.10可以发现，在缺乏信任模块的情况下，无人机在两种场景中的飞行路径呈现出较高的一致性；相反，当融入信任模块后，无人机在两个场景中的飞行策略展现出显著差异——在对手远落后时，它倾向于贴近赛道边缘，利用空间优势加速前进；而当对手紧追不舍时，则采取更为保守的策略，靠近赛道中央飞行，以减小被干扰的风险，这清晰体现了信任模块能够增进机器对于博弈的理解。

此外，可以很明显的看出，引入信任模块的无人机的飞行轨迹更为平滑，缺少信任的飞行轨迹不仅出现抖动，有时甚至存在“后退转圈”的现象。这一现象原因在于智能体的训练环境和实验平台环境并不完全相同（类似于 sim2real）。信任模型数据收集来源于实验平台，同时人类的评价存在一定的不确定性，这些因素都会作用于智能体的训练，使得人机信任引导成为了训练环境和实验环境之间的桥梁，让智能体获得了一定的鲁棒性能。

轨迹预测模块消融。取消轨迹预测网络 Tr-Pred 前后，智能体训练的奖励曲线的对比如图5.11所示。可以看出，在引入 Tr-Pred 网络之后，智能体训练前期的奖励收敛更快，奖励值也更高。

平均耗时和胜率对比如图5.9(b) 图5.12(b) 所示。可以看出，引入轨迹预测网络之后，无人机飞行耗时下降，胜率提高。

基于以上实验结果，可以看出，轨迹预测网络引入，加快了智能体的训练，可以有效提升竞速效果。

5.5　本章小结

本章考虑多机竞速场景中有着对于高级博弈行为刻画和实时计算的需求，提出了基于人机信任的多机竞速强化学习策略。首先将多机竞速问题建模为马尔可夫博弈过程，构建人机信任网络 F_{trust} 作为多机竞速中的 MPTM 并设计奖励

塑造函数，然后基于 LSTM 构建 Tr-Pred 网络用来预测对手轨迹。所提方法在实验平台中和 GTP、MPC 方法进行了对比，验证了所提方法的有效性。消融实验进一步说明，人机信任的引导可以有效的提升机器对于博弈态势的理解，Tr-Pred 网络可以有效的提升竞速飞行效果。

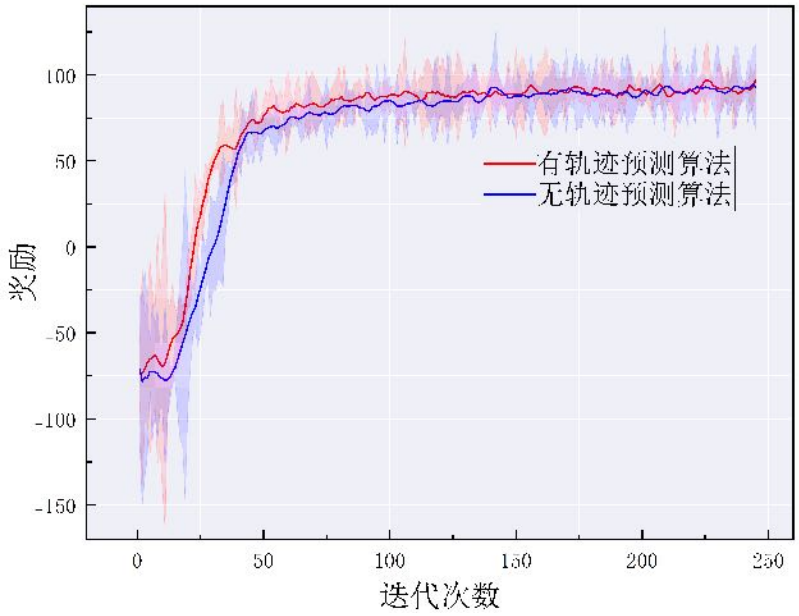
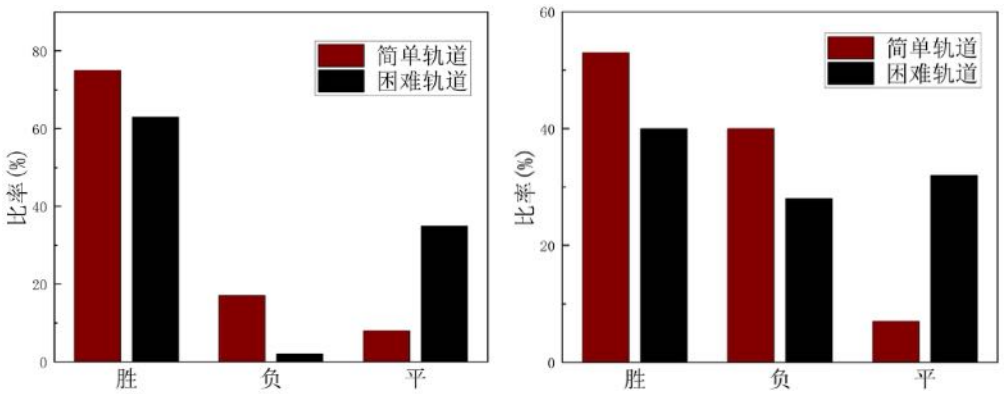


图 5.11　Tr-Pred 网络消融实验：训练奖励变化



(a) 信任消融：有信任模块的竞速结果　(b) 轨迹预测消融：有轨迹预测模块的竞速结果

图 5.12　两类消融实验中的胜率情况

第 6 章　总结与展望

6.1　全文工作总结

无人机竞速飞行在一些灾害响应、军事攻击、娱乐产业等场景具有广阔的应用价值。在无人机竞速中，存在一些难以通过概率模型精确建模的不确定性因素，同时，多机竞速下的复杂博弈行为也难以有效刻画，这些都是现有方法尚未解决的问题。针对这些问题，论文通过人机信任将人类的优势引入到无人机竞速飞行当中，首先解决现有信任模型不适用于无人机竞速场景的问题，然后将竞速场景分为单机竞速和多机竞速两种，根据场景的不同特点分别设计了融入人机信任的竞速方法。

论文的主要研究工作总结如下：

（1）考虑到现有信任模型可解释性差，以及没有刻画无人机竞速飞行中的高动态特点，给出了一种机器性能驱动的人机信任模型，该信任模型能够有效促进人对于机器的信任水平和机器能力之间的匹配，所提模型的信任数值预测均方误差更小，并且对于人类动作预测的错误率更低。首先分析了人机信任的影响因素，基于此给出了人机信任所具有的特征以及演化规律；在此基础上，引入机器当前表现作为机器性能的重要评估指标之一，进而构建机器性能驱动的人机信任演化模型，并且给出合理性说明以及参数化方法；最后设计了真实的人机交互实验用于验证所提出的信任模型。

（2）考虑到在单机竞速场景中存在的不确定性因素导致无人机竞速表现不佳的问题，给出了一种基于人机信任的单机竞速共享控制策略，相较于现有方法，所提方法可以有效增加无人机竞速飞行的效果，单圈耗时明显下降，同时人类的工作量减少，人类容错率也得到提升。首先设计无人机 MPC 竞速算法作为机器的决策算法。在此基础上，引入人类操作者辅助无人机飞行，设计人类失误检测模块以及基于人机信任的决策仲裁机制。最后在实验平台中对所提方法进行验证，设计消融实验证明所提方法中不同模块的有效性。

（3）考虑到多机竞速场景中有着对于高级博弈行为刻画和实时计算的需求，给出了一种基于人机信任的多机竞速强化学习策略，该方法能够实时给出决策指令，同时提升机器对于博弈的理解，竞速胜率超过了现有方法。首先将多机竞速问题建模为马尔可夫博弈过程，设计强化学习训练环境，基于人机信任设计奖励塑造函数，让机器主动学习人类的策略；其次，利用 LSTM 网络和注意力机制设计对手轨迹预测网络，增加智能体对于对手策略的理解。最后在实验平台中对所提方法进行验证，设计消融实验证明所提方法中不同模块的有效性。

论文主要的创新点总结如下：

（1）考虑到现有信任模型可解释性差，以及没有刻画无人机竞速飞行中的高动态特点，提出了一种机器性能驱动的人机信任模型，分析人机信任的特征和演化规律，在此基础上，构建了合理的人机信任演化模型，引入机器当前表现作为机器性能的重要评估指标，进而有效促进人对于机器的信任水平和机器能力之间的匹配。

（2）针对单机竞速场景中存在的不确定性因素导致无人机竞速效果不佳的问题，提出了一种基于人机信任的单机竞速共享控制策略，设计人类失误检测模块，同时设计基于人机信任的决策仲裁机制，实现竞速效果和人类容错率的提升以及人类工作量的减小。

（3）考虑到多机竞速场景中有着对于高级博弈行为刻画和实时计算的需求，提出了一种基于人机信任的多机竞速强化学习策略，将人机信任作为强化学习过程中的一部分奖励，利用 LSTM 网络和注意力机制设计对手轨迹预测网络，提升了竞速效果以及机器对于博弈的理解。

6.2　未来研究展望

综上所述，论文对于无人机竞速飞行所遇到的一些问题开展了相关研究，取得了一些进展和成效。但是，论文的研究仍有一些不足，主要包括以下几个方面：

（1）考虑到现有信任模型可解释性差，以及没有刻画无人机竞速飞行中的高动态特点，论文提出了一种机器性能驱动的人机信任模型，该模型能够促进人对于机器的信任水平和机器性能之间的匹配。但是这种作用是被动的、有限的。因此，可以考虑设计人机信任偏差识别机制，当发现人类对于机器的信任和机器能力不匹配的时候，在交互界面上予以提示，实现实时的信任校准。

（2）针对单机竞速场景中存在的不确定性因素导致无人机竞速效果不佳的问题，论文提出了基于人机信任的单机竞速共享控制策略。但是该方法利用的是简化的无人机运动学模型，和真实的无人机系统模型仍有差距。因此可以考虑在满足计算资源前提下，增加更为复杂和精准的无人机动力学模型，以此增强无人机竞速效果。

（3）考虑到多机竞速场景中有着对于高级博弈行为刻画和实时计算的需求，论文提出了一种基于人机信任的多机竞速强化学习策略。该方法需要人类对于机器竞速表现给出一个具体的信任分数，对于人类而言工作量偏大，而且误差较大。因此，可以考虑采用偏好强化学习技术来简化人类操作，让人们不再需要给出具体信任分数，仅需表明对机器两种竞速表现中更倾向于信任哪一种，从而达到进一步减轻人类任务负荷的目的。

参 考 文 献

[1] DELMERICO J, CIESLEWSKI T, REBECQ H, et al. Are we ready for autonomous drone racing? The UZH-FPV drone racing dataset[C]//2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 6713-6719.

[2] YANG G Z, BELLINGHAM J, DUPONT P E, et al. The grand challenges of science robotics [J]. Science robotics, 2018, 3(14): eaar7650.

[3] 邸健. 竞速无人机自主飞行控制问题研究[D]. 中国科学技术大学, 2023.

[4] LOQUERCIO A, KAUFMANN E, RANFTL R, et al. Deep drone racing: From simulation to reality with domain randomization[J]. IEEE Transactions on Robotics, 2019, 36(1): 1-14.

[5] MOON H, SUN Y, BALTES J, et al. The IROS 2016 competitions[J]. IEEE Robotics and Automation Magazine, 2017, 24(1): 20-29.

[6] KIM D, RYU H, YONCHORHOR J, et al. A deep-learning-aided automatic vision-based control approach for autonomous drone racing in game of drones competition[C]//NeurIPS 2019 Competition and Demonstration Track. PMLR, 2020: 37-46.

[7] FOEHN P, BRESCIANINI D, KAUFMANN E, et al. Alphapilot: Autonomous drone racing [J]. Autonomous Robots, 2022, 46(1): 307-320.

[8] MOON H, MARTINEZ-CARRANZA J, CIESLEWSKI T, et al. Challenges and implemented technologies used in autonomous drone racing[J]. Intelligent Service Robotics, 2019, 12: 137-148.

[9] LUO Y, YU H, ZHANG H, et al. A novel Newton–Euler method-based nonlinear anti-swing control for a quadrotor UAV carrying a slung load[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024, 54(4): 2266-2275.

[10] HE B, JI X, LI G, et al. Key technologies and applications of UAVs in underground space: A review[J]. IEEE Transactions on Cognitive Communications and Networking, 2024: 1-1.

[11] WEI M, ZHENG L, LI H, et al. Adaptive neural network-based model path-following contouring control for quadrotor under diversely uncertain disturbances[J]. IEEE Robotics and Automation Letters, 2024, 9(4): 3751-3758.

[12] MOSTUFA S, CHAKRABARTI K. A genre of cognitive evolutions through artificial superintelligence and robotics technology[M]. Springer Nature Singapore, 2024.

[13] YANG S, MAO X, LIU Z, et al. The accompanying behavior model and implementation architecture of autonomous robot software[C]//2017 24th Asia-Pacific Software Engineering Conference (APSEC). IEEE, 2017: 209-218.

[14] TAO Y, QIU J, LAI S, et al. A human-machine reinforcement learning method for cooperative

energy management[J]. IEEE Transactions on Industrial Informatics, 2021, 18(5): 2974-2985.

[15] JOHNSON J. Automating the OODA loop in the age of intelligent machines: reaffirming the role of humans in command-and-control decision-making in the digital age[J]. Defence Studies, 2023, 23(1): 43-67.

[16] 赵云波, 康宇. 人机混合智能系统自主性理论和方法[M]. 科学出版社, 2021.

[17] GUNIA A. The role of trust in human-machine interaction: Cognitive science perspective[M]. Routledge, 2024.

[18] MAYER M. Trusting machine intelligence: artificial intelligence and human-autonomy teaming in military operations[J]. Defense & Security Analysis, 2023, 39(4): 521-538.

[19] THANGAVEL K, SABATINI R, GARDI A, et al. Artificial intelligence for trusted autonomous satellite operations[J]. Progress in Aerospace Sciences, 2024, 144: 100960.

[20] GEBRU B, ZELEKE L, BLANKSON D, et al. A review on human–machine trust evaluation: Human-centric and machine-centric perspectives[J]. IEEE Transactions on Human-Machine Systems, 2022, 52(5): 952-962.

[21] HAENLEIN M, KAPLAN A. A brief history of artificial intelligence:on the past, present, and future of artificial intelligence[J]. California Management Review, 2019, 61(4): 5-14.

[22] BUCKLEY L, KAYE S A, PRADHAN A K. Psychosocial factors associated with intended use of automated vehicles: A simulated driving study.[J]. Accident; analysis and prevention, 2018, 115: 202-208.

[23] MUIR B M. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems[J]. Ergonomics, 1994, 37(11): 1905-1922.

[24] JIAN J Y, BISANTZ A M, DRURY C G. Foundations for an empirically determined scale of trust in automated systems[J]. International journal of cognitive ergonomics, 2000, 4(1): 53-71.

[25] MADSEN M, GREGOR S. Measuring human-computer trust[C]//11th australasian conference on information systems: volume 53. Citeseer, 2000: 6-8.

[26] SCHAEFER K E. Measuring trust in human robot interactions: Development of the "trust perception scale-HRI" [M]. Springer, 2016.

[27] GUPTA K, HAJIKA R, PAI Y S, et al. Measuring human trust in a virtual assistant using physiological sensing in virtual reality[C]//2020 IEEE Conference on virtual reality and 3D user interfaces (VR). IEEE, 2020: 756-765.

[28] AKASH K, HU W L, JAIN N, et al. A classification model for sensing human trust in machines using EEG and GSR[J]. ACM Transactions on Interactive Intelligent Systems (TiiS), 2018, 8 (4): 1-20.

[29] CHOO S, NAM C S. Detecting human trust calibration in automation: a convolutional neural

network approach[J]. IEEE Transactions on Human-Machine Systems, 2022, 52(4): 774-783.

[30] HU W L, AKASH K, REID T, et al. Computational modeling of the dynamics of human trust during human–machine interactions[J]. IEEE Transactions on Human-Machine Systems, 2018, 49(6): 485-497.

[31] CHRISTENSEN A B, DAM C R, RASLE C, et al. Reducing overtrust in failing robotic systems[C]//2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 2019: 542-543.

[32] XIE Y, BODALA I P, ONG D C, et al. Robot capability and intention in trust-based decisions across tasks[C]//2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 2019: 39-47.

[33] HUDSPETH M, BALALI S, GRIMM C, et al. Effects of interfaces on human-robot trust: Specifying and visualizing physical zones[C]//2022 International Conference on Robotics and Automation (ICRA). IEEE, 2022: 11265-11271.

[34] LEE J, MORAY N. Trust, control strategies and allocation of function in human-machine systems[J]. Ergonomics, 1992, 35(10): 1243-1270.

[35] WANG Q, LIU D, CARMICHAEL M G, et al. Computational model of robot trust in human co-worker for physical human-robot collaboration[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 3146-3153.

[36] NAM C, WALKER P, LI H, et al. Models of trust in human control of swarms with varied levels of autonomy[J]. IEEE Transactions on Human-Machine Systems, 2019, 50(3): 194-204.

[37] SPENCER D A, WANG Y, HUMPHREY L R. Trust-based human-robot interaction for multi-robot symbolic motion planning[C]//2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2016: 1443-1449.

[38] SADRFARIDPOUR B, SAEIDI H, BURKE J, et al. Modeling and control of trust in human-robot collaborative manufacturing[J]. Robust intelligence and trust in autonomous systems, 2016: 115-141.

[39] SADRFARIDPOUR B. Trust-based control of robotic manipulators in collaborative assembly in manufacturing[D]. Clemson University, 2018.

[40] JONKER C M, TREUR J. Formal analysis of models for the dynamics of trust based on experiences[C]//Multi-Agent System Engineering: 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World. Springer, 1999: 221-231.

[41] XU A, DUDEK G. OPTIMO: Online probabilistic trust inference model for asymmetric human-robot collaborations[C]//Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction. 2015: 221-228.

[42] CHEN M, NIKOLAIDIS S, SOH H, et al. Planning with trust for human-robot collaboration

[C]//Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction. 2018: 307-315.

[43] AKASH K, MCMAHON G, REID T, et al. Human trust-based feedback control: Dynamically varying automation transparency to optimize human-machine interactions[J]. IEEE Control Systems Magazine, 2020, 40(6): 98-116.

[44] HANOVER D, LOQUERCIO A, BAUERSFELD L, et al. Autonomous drone racing: A survey [A]. 2023.

[45] ROUHI A, AREZOOMANDAN S, KAPOOR R, et al. An overview of deep learning in UAV perception[C]//2024 IEEE International Conference on Consumer Electronics (ICCE). 2024: 1-6.

[46] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(3): 611-625.

[47] NISAR B, FOEHN P, FALANGA D, et al. Vimo: Simultaneous visual inertial model-based odometry and force estimation[J]. IEEE Robotics and Automation Letters, 2019, 4(3): 2785-2792.

[48] LV S, GAO Y, CHE J, et al. Autonomous drone racing: Time-optimal spatial iterative learning control within a virtual tube[C]//2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023: 3197-3203.

[49] 唐健杰, 王鑫. 基于 PI-PD 控制器的四旋翼姿态控制[J]. 电子技术与软件工程, 2014(16): 175-176.

[50] KAWAI Y, UCHIYAMA K. Design of frequency shaped LQR considering dynamic characteristics of the actuator[C]//2016 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, 2016: 1235-1239.

[51] BOUABDALLAH S, NOTH A, SIEGWART R. PID vs LQ control techniques applied to an indoor micro quadrotor[C]//2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS): volume 3. IEEE, 2004: 2451-2456.

[52] ZHENG E H, XIONG J J, LUO J L. Second order sliding mode control for a quadrotor UAV [J]. ISA transactions, 2014, 53(4): 1350-1356.

[53] FOEHN P, ROMERO A, SCARAMUZZA D. Time-optimal planning for quadrotor waypoint flight[J]. Science robotics, 2021, 6(56): eabh1221.

[54] RYOU G, TAL E, KARAMAN S. Multi-fidelity black-box optimization for time-optimal quadrotor maneuvers[J]. The International Journal of Robotics Research, 2021, 40(12-14): 1352-1369.

[55] BRISTOW D A, THARAYIL M, ALLEYNE A G. A survey of iterative learning control[J]. IEEE control systems magazine, 2006, 26(3): 96-114.

[56] SHIN S, KANG Y, KIM Y G. Evolution algorithm and online learning for racing drone[C]// NeurIPS 2019 Competition and Demonstration Track. PMLR, 2020: 100-109.

[57] WANG Q, WANG D, XU C, et al. Polynomial-based online planning for autonomous drone racing in dynamic environments[C]//2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2023: 1078-1085.

[58] NGUYEN H, KAMEL M, ALEXIS K, et al. Model predictive control for micro aerial vehicles: A survey[C]//2021 European Control Conference (ECC). IEEE, 2021: 1556-1563.

[59] ROMERO A, SUN S, FOEHN P, et al. Model predictive contouring control for time-optimal quadrotor flight[J]. IEEE Transactions on Robotics, 2022, 38(6): 3340-3356.

[60] ROMERO A, PENICKA R, SCARAMUZZA D. Time-optimal online replanning for agile quadrotor flight[J]. IEEE Robotics and Automation Letters, 2022, 7(3): 7730-7737.

[61] SALZMANN T, KAUFMANN E, ARRIZABALAGA J, et al. Real-time neural MPC: Deep learning model predictive control for quadrotors and agile robotic platforms[J]. IEEE Robotics and Automation Letters, 2023, 8(4): 2397-2404.

[62] NAGAMI K, SCHWAGER M. HJB-RL: Initializing reinforcement learning with optimal control policies applied to autonomous drone racing.[C]//Robotics: science and systems. 2021: 1-9.

[63] SONG Y, SCARAMUZZA D. Policy search for model predictive control with application to agile drone flight[J]. IEEE Transactions on Robotics, 2022, 38(4): 2114-2130.

[64] KAUFMANN E, BAUERSFELD L, LOQUERCIO A, et al. Champion-level drone racing using deep reinforcement learning[J]. Nature, 2023, 620(7976): 982-987.

[65] MIRRI S, PRANDI C, SALOMONI P. Human-drone interaction: state of the art, open issues and challenges[C]//Proceedings of the ACM SIGCOMM 2019 Workshop on Mobile Air-Ground Edge Computing, Systems, Networks, and Applications. 2019: 43-48.

[66] AGRAWAL A. Human-drone collaborations in human-on-the-loop emergency response systems[M]. University of Notre Dame, 2022.

[67] HUMMEL K A, POLLAK M, KRAHOFER J. A distributed architecture for human-drone teaming: Timing challenges and interaction opportunities[J]. Sensors, 2019, 19(6): 1379.

[68] LINIGER A, LYGEROS J. A noncooperative game approach to autonomous racing[J]. IEEE Transactions on Control Systems Technology, 2019, 28(3): 884-897.

[69] SPICA R, CRISTOFALO E, WANG Z, et al. A real-time game theoretic planner for autonomous two-player drone racing[J]. IEEE Transactions on Robotics, 2020, 36(5): 1389-1403.

[70] WANG Z, TAUBNER T, SCHWAGER M. Multi-agent sensitivity enhanced iterative best response: A real-time game theoretic planner for drone racing in 3D environments[J]. Robotics

and Autonomous Systems, 2020, 125: 103410.

[71] SCHWARTING W, SEYDE T, GILITSCHENSKI I, et al. Deep latent competition: Learning to race using visual control policies in latent space[A]. 2021.

[72] LI N, GOUBAULT E, PAUTET L, et al. Autonomous racecar control in head-to-head competition using mixed-integer quadratic programming[C]//Opportunities and challenges with autonomous racing, 2021 ICRA workshop. 2021.

[73] HE S, ZENG J, SREENATH K. Autonomous racing with multiple vehicles using a parallelized optimization with safety guarantee using control barrier functions[C]//2022 International conference on robotics and automation (ICRA). IEEE, 2022: 3444-3451.

[74] SHEN Y, ZHOU J, XU D, et al. Aggressive trajectory generation for a swarm of autonomous racing drones[C]//2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2023: 7436-7441.

[75] AGRAWAL A, CLELAND-HUANG J, STEGHÖFER J P. Model-driven requirements for humans-on-the-loop multi-uav missions[C]//2020 IEEE Tenth International Model-Driven Requirements Engineering (MoDRE). IEEE, 2020: 1-10.

[76] ABBINK D A, CARLSON T, MULDER M, et al. A topology of shared control systems—finding common ground in diversity[J]. IEEE Transactions on Human-Machine Systems, 2018, 48(5): 509-525.

[77] REDDY S, DRAGAN A D, LEVINE S. Shared autonomy via deep reinforcement learning [A]. 2018.

[78] FLEMISCH F, ABBINK D A, ITOH M, et al. Joining the blunt and the pointy end of the spear: towards a common framework of joint action, human–machine cooperation, cooperative guidance and control, shared, traded and supervisory control[J]. Cognition, Technology & Work, 2019, 21: 555-568.

[79] FLEMISCH F, ABBINK D, ITOH M, et al. Shared control is the sharp end of cooperation: Towards a common framework of joint action, shared control and human machine cooperation [J]. IFAC-PapersOnLine, 2016, 49(19): 72-77.

[80] SCHULMAN J. Deep reinforcement learning through policy optimization[D]. University of California, Berkeley, 2016.

[81] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]//International conference on machine learning. PMLR, 2018: 1861-1870.

[82] SCHULMAN J, LEVINE S, ABBEEL P, et al. Trust region policy optimization[C]// International conference on machine learning. PMLR, 2015: 1889-1897.

[83] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms

[A]. 2017.

[84] MADHAVAN P, WIEGMANN D A. Similarities and differences between human–human and human–automation trust: An integrative review[J]. Theoretical Issues in Ergonomics Science, 2007, 8(4): 277-301.

[85] HANCOCK P A, BILLINGS D R, SCHAEFER K E, et al. A meta-analysis of factors affecting trust in human-robot interaction[J]. Human factors, 2011, 53(5): 517-527.

[86] GUO C, PLEISS G, SUN Y, et al. On calibration of modern neural networks[C]//International conference on machine learning. PMLR, 2017: 1321-1330.

[87] KRIZHEVSKY A, HINTON G, et al. Learning multiple layers of features from tiny images [M]. Toronto, ON, Canada, 2009.

[88] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and harnessing adversarial examples[A]. 2014.

[89] NGUYEN A, YOSINSKI J, CLUNE J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 427-436.

[90] PROVOST F J, FAWCETT T, KOHAVI R, et al. The case against accuracy estimation for comparing induction algorithms.[C]//ICML: volume 98. 1998: 445-453.

[91] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: Machine learning in python[J]. the Journal of machine Learning research, 2011, 12: 2825-2830.

[92] SHAH S, DEY D, LOVETT C, et al. Airsim: High-fidelity visual and physical simulation for autonomous vehicles[C]//Field and Service Robotics: Results of the 11th International Conference. Springer, 2018: 621-635.

致　谢

三年的硕士时光一晃而过，二十年的求学生涯即将画上句号。这一路走来，遇到过一些坎坷，但更多的是收获受益终身的知识、道理和本领，在此向所有关心和帮助过我的人们致以诚挚的谢意。

感谢我的导师赵云波教授。作为赵老师在科大的第一届硕士研究生，我深感荣幸并由衷感激当初夏令营时赵老师给我求学科大的宝贵机会。在我的科研旅程中，无论是思维模式的塑造还是科研方法论的建立，每一步都凝聚着赵老师细致入微的指导与关怀。面对科研路上的种种挑战与疑惑，赵老师总是能够迅速察觉并给予我及时的沟通与支持，帮助我克服难关。更重要的是，赵老师对我的影响远远超出了学术范畴。当我站在职业生涯抉择和个人发展规划的十字路口，感到彷徨无措时，赵老师主动与我沟通，给予我宝贵的引导。这些人生指引，使我在心智与人格上得到了全面的成长与发展。

感谢邸健师兄、刘斌琨师兄和李鹏飞师兄。邸健师兄从研二的时候便开始指导我，教会了我科研论文查找、代码编写、数据绘图等各方面基础技能；在我研三之后，邸健师兄毕业了，工作很忙，但是仍然会从本就不多的空余时间中抽出一部分帮我看毕业论文，给出了很多有用的意见。在一些细节问题不懂的时候，刘斌琨师兄总是不厌其烦的花时间教我，遇到困难泄气的时候，斌琨师兄会鼓励我，帮我寻求解决方法。在一些科研问题上，李鹏飞师兄花了很多时间与我讨论，对于我的科研思维有很大的提升，在完成毕业论文时，李鹏飞师兄耐心地帮我逐句更改摘要、绪论等内容。

感谢我所有的朋友们。无论是在挫折面前相互倾诉的慰藉时光，还是在庆祝成功时共享的欢声笑语，你们的存在如同调色盘一般，为我的生活添上了斑斓的色彩。随着毕业的钟声渐近，我们即将踏上各自的征途，前往不同的工作岗位，探索各异的城市，遇见形形色色的人。在此临别之际，我由衷地希望我们每个人都能保持那份纯真的快乐与对生活的热爱，无论未来道路如何延展，愿我们在各自的生活舞台上继续绽放。

我要感谢我的父母。你们的开明思想和远见卓识为我提供了坚实的教育平台，用无条件的爱与鼓励浇灌了我的成长之路。正是有了你们作为坚强后盾，我才能在求学的征途中不断进步。如今，随着我即将完成学业，踏入职场，角色虽变，但我对你们的感激与责任愈加深厚。正如你们一直守护我那样，未来我将成为你们的依靠。

最后，愿在今后的生活中，自己能永保初心，心怀感恩，一往无前。

在读期间发表的学术论文与取得的研究成果

已发表论文

1. **Chen Shaojun**, Yun-Bo Zhao, Yang Wang and Junsen Lu, “A Human-Machine Trust Model Integrating Machine Estimated Performance”, 6th International Symposium on Autonomous Systems (ISAS), Nanjing, China, 2023, pp. 383-388.

已投稿论文

1. Yun-Bo Zhao, Junsen Lu,Yang Wang, **Chen Shaojun**, Ruiyu Xia, Pengfei Li, Yu Kang “A Machine-Performance-Driven Human Trust Model for AI-Powered Human-Machine Systems”, IEEE Transactions on Human-Machine Systems,under review.