arXiv:2304.01428v1 [cs.RO] 4 Apr 2023

面向共享空域长时域社交机器人导航的学习型树搜索

Ingrid Navarro[1*]、Jay Patrikar[1*]、Joao P.A.Dantas[2]、Rohan Baijal[3]、Ian Higgins[1]、Sebastian Scherer[1] 和 Jean Oh[1]

摘要—共享空域中对全自主航空作业日益增长的需求，要求开发值得信赖的智能体，使其能够在拥挤、动态的空间中安全、顺畅地导航。在本研究中，我们提出社交机器人树搜索（SoRTS），这是一种用于社交领域移动机器人安全导航的算法。SoRTS旨在为现有的社会感知轨迹预测策略增强蒙特卡洛树搜索规划器，从而改进移动机器人的下游导航。为评估该方法的性能，我们选择通用航空中的社会导航作为应用场景。为支持这一评估，我们还在本研究中推出X-PlaneROS，这是一款高保真航空模拟器，可推动全尺寸航空自主领域的进一步研究。通过一项基于26名美国联邦航空管理局认证飞行员评估的用户研究，我们表明，SoRTS的表现与能力出色的人类飞行员相当，并显著优于我们的基线算法。我们还通过自博弈实验进一步补充这些结果，展示了算法在复杂程度逐步提升的情景中的行为。 [代码][1] | [视频][2]

I. 引言

社交机器人致力于制定决策策略，使其能够与人类顺畅互动，在实现目标点的同时遵循社会规范。尽管社会导航和运动预测 [1]–[3]，已取得显著进展，但要在实现顺畅人群导航的同时平衡社会目标与自身利益，仍然颇具挑战。社会导航可建模为部分可观测随机博弈（POSG） [4], [5]。深度强化学习（DRL）方法 [6] 会明确设定 POSG 的奖励函数，并通过模拟自我博弈实验来推导策略。虽然这类技术在稀疏数据领域颇具前景，因为机器人很容易与人类区分开来，但在同质导航中调整奖励参数并非易事 [7]。DRL 策略也取决于底层模拟器；由于模拟器与真实世界场景缺乏兼容性，仿真到现实的迁移往往会导致不理想的行为。

[1]作者来自美国宾夕法尼亚州匹兹堡市卡内基梅隆大学机器人研究所。 {ingridn, jaypat, ihiggins, basti, jeanoh}@cs.cmu.edu[2]作者来自巴西圣保罗州圣若泽杜斯坎普斯市高等研究院。本文工作是在卡内基梅隆大学访问研究期间完成的。dantasjpad@fab.mil.br[3]作者来自印度北方邦坎普尔市印度坎普尔理工学院。本文工作是在卡内基梅隆大学访问研究期间完成的。 rbaijal@iitk.ac.in

*贡献相同。[1]代码库：https://github.com/cmubig/sorts[2]视频：https://youtu.be/PBE3O4cW2rI

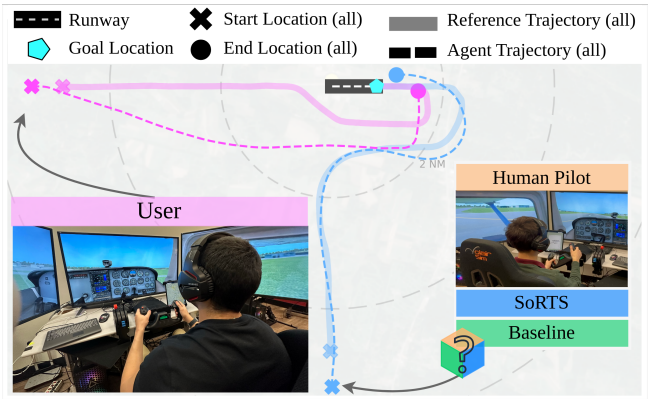


图 1：用户研究中使用的飞行模拟器设置。每位用户按随机顺序与我们的人类飞行员、SoRTS 和基线系统进行交互。图中还展示了用户与SoRTS之间的一次交互。

另一方面，数据驱动方法在社会轨迹预测中很常见。它们旨在直接刻画数据中观察到的人类交互 [2], ，从而无需进行奖励塑形和精确仿真。例如，近期的序列到序列模型在意图预测方面取得了令人鼓舞的成果 [8]–[10]。然而，将这些模型用于下游导航并不容易，因为它们往往会出现预测失败 [11] ，进而损害其泛化能力。这种可能导致不安全行为的风险，凸显了提升部署于真实世界模型鲁棒性的必要性。

在社会导航中，一个智能体的动作会影响另一个智能体的动作，反之亦然 [4], [12]。这种在时间上递归的决策机制已被用于模拟类人游戏玩法 [13], [14]。基于这一思路，我们提出采用基于递归搜索的策略，提升离线训练模型在下游社会导航目标下的鲁棒性。具体而言，我们采用蒙特卡洛树搜索（MCTS） [15] 作为搜索策略，以进行长时域递归仿真、碰撞检测和目标条件化。我们将其与具备社会意识的意图预测模型相结合，以提供短时域智能体间上下文线索并确保运动自然度。我们利用 MCTS 融合这些短时域线索与长时域规划，并纳入全局参考路径来引导树扩展。我们将这一框架称为社交机器人树搜索（*SoRTS*）。

无人飞行器的运营不断增长，推动了与

人类飞行员 [16], [17]同时使用空域的需求。因此，我们选择通用航空（GA）领域来展示我们的方法。近期，GA 被纳入社会导航范式 [9],[10], 其中飞行员 需要 遵循飞行指南，相互协调并尊重彼此的个人空间，以确保飞行安全。这类似于在人群和车辆环境中遵守礼仪。

安全关键领域（如通用航空）要求具备高水平的能力，以确保运行顺畅、安全。这意味着需要开发值得信赖的机器人，使其能够理解并遵循导航规范，同时理解长期动态交互，避免给他人带来危险或不适感。本文将这些方面划分为两个维度：导航性能 和安全性。我们的框架设计与评估均围绕这两个维度展开。我们基于定制的仿真器框架 X-PlaneROS 开展用户研究，邀请了 26 名经验丰富的飞行员，以考察飞行员在真实飞行场景中如何与我们的模型交互。我们进一步分析了他们如何沿这两个维度评估我们的模型，并将其与一名胜任的人类飞行员进行比较。此外，我们还通过复杂度逐步增加的情景开展自我博弈实验，补充了评估结果。

贡献声明：

1) 我们介绍并开源 SoRTS，这是一种基于蒙特卡洛树搜索的长时域导航算法，可增强离线学习得到的具备社会意识的意图预测策略，以用于下游导航。2) 我们推出 X-PlaneROS，这是一个用于共享空域导航的高保真仿真环境；3) 通过对 26 名经美国联邦航空管理局认证的飞行员开展用户研究，并在更复杂的情景中进行自我博弈仿真，我们表明，在导航性能 和安全性 方面，SoRTS 的评价与一名胜任的人类飞行员相当，并且显著优于其基线算法。

II. 相关工作

A. 社会导航算法

社会导航领域已有大量研究，重点关注行人场景和自动驾驶领域 [1]。在行人场景中，人们提出了经典的基于模型的方法 [18], [19]，这些方法至今仍是重要的基线系统。然而，将其扩展到其他领域并非易事。近期采用手工设计的以安全为重点的奖励函数 [22] 的深度强化学习方法 [6], [20], [21] 在这些领域取得了令人鼓舞的成果。然而，仿真器设计[23],和领域特定奖励函数方面的不足限制了其在现实世界中的表现 [22]。实现可扩展性和鲁棒性颇具挑战，通常需要耗费高昂成本进行重新训练。与我们的方法类似，[5] 提出了一种使用蒙特卡洛树搜索训练并部署策略的深度强化学习方法。虽然该方法依赖预先定义的奖励函数和模拟器训练，但我们的工作将这些方法扩展为使用基于离线专家的策略，从而针对社会导航提供领域特定处理。

数据驱动方法侧重于从记录智能体之间交互的数据集中学习策略 [1], [22]。由于这些模型无需显式构造奖励，因此能够捕捉丰富的联合社会动态。然而，由于示范数据存在噪声和缺失，这些方法难以直接部署 [24], [25]。为缓解这一问题，[4] 利用Q值函数的梯度进行模型预测控制，而 [12]则提出使用双重控制进行信念状态传播，将该方法推广。这些方法依赖逆强化学习作为额外步骤来生成Q值函数。也有研究提出直接在优化过程中使用序列模型的梯度 [26], ，但并未考察其收敛特性。我们的方法更直接地使用序列模型，无需计算梯度或Q值。相反，我们将模型的输出转换为动作分布，用于后续规划任务。

B. 社会导航评估

已有研究采用不同指标评估社交机器人导航 [1], [2], [27]。评估的主要分析维度包括：基于参考轨迹的行为自然度，或执行路径的不规则程度 [28], [29], ；性能与效率 [30]–[32], ；以及物理个人空间或不适感相关概念 [33], [34]。用户研究常用于评估更主观的方面，例如机器人在交互过程中引发的不适感和信任 [35]。遵循先前研究，我们重点关注导航性能，以衡量智能体的平滑性及其遵循导航准则的能力；并关注安全性，以评估其尊重他人个人空间的能力。我们还开展了一项用户研究，邀请经验丰富的飞行员在真实的飞行场景中与我们的算法交互，并对机器人的表现、感知安全性和信任度进行评分。

III. 问题表述

我们将社会导航表述为一种近似的部分可观测随机博弈（POSG），这是一个用于去中心化有限时域规划的框架。有关 POSG 及其在社会导航中的应用的更多详情，请参阅 [4], [5], [36]。按照 [9],[10], ，我们假设有 M 个智能体，其中$\mathbf{s}_t^i \in R^3$ 表示智能体 i 在时间步 t 的状态，而$\mathbf{a}_t^i \in \mathcal{A}$ 是离散动作或运动基元的集合，这些动作或运动基元由 $\dot{\mathbf{s}} = f(\mathbf{s}, \mathbf{a})$决定。令$\mathbf{s}_0^i$ 和$\mathbf{s}_g^i$分别表示起始状态和目标状态。系统还可访问一组离线专家 示范数据$\mathcal{D}$ 以及一组全局参考轨迹 τ_R。我们省略上标，以表示所有智能体的联合状态。

因此，给定 M 个智能体的起始**状态**$_0$ 和目标**状态**$_g$ ，目标是找到一系列控制输入$\pi = \{\mathbf{a}_0, \ldots\ldots, \mathbf{a}_g\}$，使智能体沿无碰撞轨迹 $\tau = \{\mathbf{s} \cdots, \mathbf{s}_g\}$行进。生成的轨迹需要满足 $||\mathbf{s}_t^i - \mathbf{s}_t^j|| \geq d \ \forall i, \ j \in \{1, \ldots, M\} \ \forall t$其中 $\mathbf{d}$ 是满足安全目标所需的最小间隔距离。此外，轨迹还需要尽可能贴近参考轨迹 $\min ||\tau - \tau_R||$ ，以满足导航目标，并遵循 $\tau \sim \mathcal{D}$ ，以满足社会目标。不失一般性，我们假设第一个智能体（i=0）是机器人自我智能体，并在每个时间步执行最优动作。

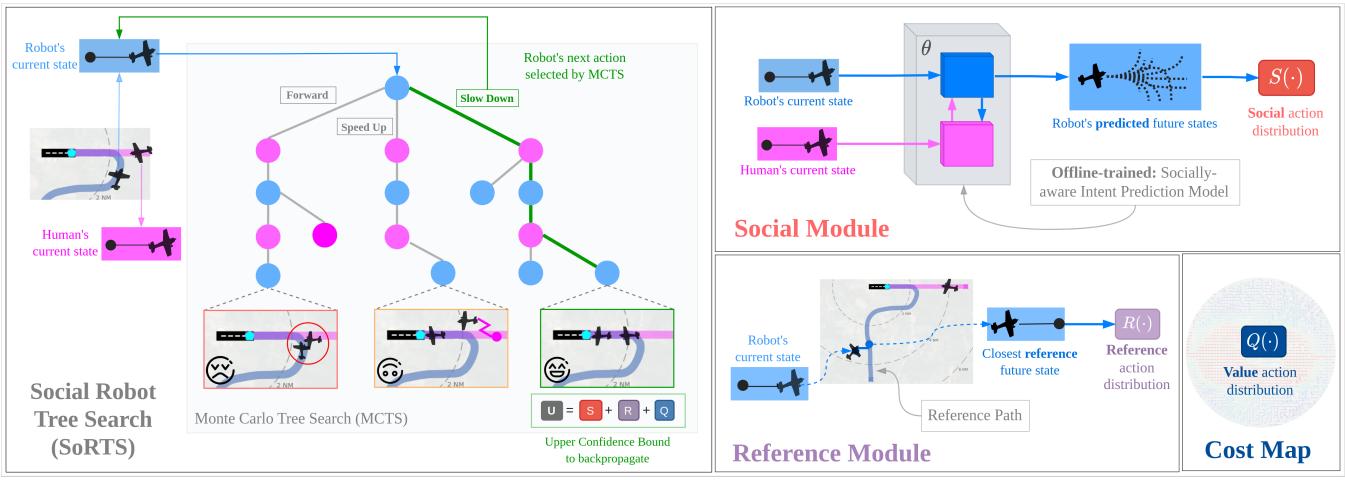


图 2：SoRTS 概览。SoRTS 是一种基于蒙特卡洛树搜索（MCTS）的社交机器人导航规划器，其树搜索由三个组件引导：一个 社交模块、一个 参考模块 和一张 代价地图。社交模块利用离线训练的意图预测模型来刻画智能体间交互，并预测自我智能体的一组可能未来状态。参考模块从全局规划中为智能体提供最近的参考状态；该规划体现了智能体必须遵循的导航准则。代价地图编码全局访问地图，进一步引导导航。MCTS 通过搜索不同决策模式之间的选择，利用这些模块进行在线碰撞检测和长时域仿真。

IV. 方法

我们从以下两个方面设计了 SoRTS： 导航性能 和安全性。SoRTS 的核心思想是利用社交模块和全局参考路径，引导蒙特卡洛树搜索寻找并在各种决策模式之间进行选择。图 2 展示了两架飞机在单一路径上汇合的示例。每架飞机都可以从三种动作中选择，以安全地完成汇合：前进、加速和减速。蒙特卡洛树搜索在前向仿真中不仅会剪除导致未来碰撞的分支，还会利用社交模块在抢行（社交上不理想）和礼让（社交上理想）之间进行选择，从而产生符合社会规范且安全的行为。

A. 模块

SoRTS 是一种基于搜索的规划器，采用蒙特卡洛树搜索（MCTS），其树扩展由三个模块引导。首先，社交模块 负责处理场景中的短时域动态，刻画社会交互及相关线索。其次，参考模块 为智能体提供全局导航准则，例如，机场交通模式。最后，代价地图 编码全局价值地图。MCTS 综合使用这些组件，进行碰撞检测和符合社会规范的长时域仿真。相应的伪代码见算法 1 和2。

1) 社交模块： 我们采用一个使用 θ, $P_\theta(\cdot)$ 参数化的离线训练意图预测算法，以刻画遵循 $\mathcal{D}$ 中专家轨迹的智能体间短期动态。

$$p_\theta(\mathbf{s}_t^i, \mathbf{a}_t^i) \sim P_\theta(\mathbf{a}_t^i \mid \mathbf{s}_{t-t_{obs}:t}, \mathbf{s}_g^i) \tag{1}$$

其中 $p_\theta(\cdot)$ 根据过去的轨迹 $\mathbf{s}_{t-t_{obs}:\,t}$ ， 为智能体 i 提供未来动作的分布$\mathbf{a}_t^i$ 包括所有智能体和目标点 $\mathbf{s}_g^i$ ， 其中 t_{obs} 为观测时域。

此处，我们使用 Social-PatteRNN [10]，这是一种基于学习到的交互来预测多种未来轨迹的算法，能够利用数据中的运动模式信息。

2) 参考路径模块： 给定智能体 i 的起始状态和目标状态，该算法会采样一条合适的参考轨迹 $\tau_r^i \in \tau_R$，第 1 行见算法 1。与第 IV-A.1 节类似，该轨迹用于计算参考动作分布 $p_r(\cdot)$，

$$p_r(\mathbf{s}_t^i, \mathbf{a}_t^i) \sim P_r(\mathbf{a}_t^i | \mathbf{s}_t^i, \tau_r^i) \tag{2}$$

与$\mathbf{s}$ 和 τ_r 在时间步 t的 L2 范数成正比。在算法 2 中，参考动作在第 5 行获得。τ_R 可从专家分布 $\mathcal{D}$、A-Star 等全局路径规划算法以及逻辑规范中得出 [37] ， 也可以手工设计。

3) 代价地图： 该算法还使用表示环境的代价地图，其中表示价值函数 $v(\mathbf{s})$。在我们的场景中，代价地图用于表示状态访问频率，从而引导搜索偏向更理想的区域。算法 2 在第 3 行使用代价地图。该价值函数可以通过自我博弈（例如）学习得到 [5] ， 也可以根据先验分布预先计算，并表示联合状态分布的价值。

B. 社会蒙特卡洛树搜索

蒙特卡洛树搜索是一种基于搜索的算法，它会将树搜索扩展到奖励较高的轨迹。原则上，这是通过在搜索过程中选择使上置信界最大化的节点来实现的 [15]，该上置信界在利用与探索的程度之间进行权衡。SoRTS 利用蒙特卡洛树搜索，并使用上一节介绍的组件来

引导其轨迹推演。形式上，它遵循如下所示的 UCT，

$$U(\mathbf{s},\mathbf{a}) = Q(\mathbf{s},\mathbf{a}) + c_1 \cdot S(\mathbf{s},\mathbf{a}) + c_2 \cdot R(\mathbf{s},\mathbf{a}) \quad (3)$$

其中 $Q(\mathbf{s},\mathbf{a})$ 表示在状态 $\mathbf{s}$ 下采取动作 $\mathbf{a}$的期望值； $S(\mathbf{s},\mathbf{a})$ 是根据社会感知模块进行访问归一化的分量； $R(\mathbf{s},\mathbf{a})$ 是根据参考路径进行访问归一化的分量；c_1 而 c_2 是超参数。为便于表示，我们省略时间 t 下标。在算法14的第 2行，这些值按如下方式更新：

$$Q(\mathbf{s},\mathbf{a}) = \frac{N(\mathbf{s},\mathbf{a}) \cdot Q(\mathbf{s},\mathbf{a}) + v(\mathbf{s})}{N(\mathbf{s},\mathbf{a}) + 1} \quad (4)$$

$$R(\mathbf{s},\mathbf{a}) = \frac{N(\mathbf{s},\mathbf{a}) \cdot R(\mathbf{s},\mathbf{a}) + p_r(\mathbf{s},\mathbf{a})}{N(\mathbf{s},\mathbf{a}) + 1} \quad (5)$$

$$S(\mathbf{s},\mathbf{a}) = \frac{\sqrt{N(\mathbf{s})}}{N(\mathbf{s},\mathbf{a}) + 1} \cdot p_\theta(\mathbf{s},\mathbf{a}) \quad (6)$$

其中 $N(\mathbf{s})$ 表示状态访问次数， $N(\mathbf{s},\mathbf{a})$ 表示给定动作 $\mathbf{a}$时的访问次数。这些更新在时间预算内，或直到发现新状态为止，通过遍历各状态迭代执行。在每个时间步，系统以轮流方式交替扩展各智能体的未来状态，迭代构建一棵新的前向仿真树。会导致碰撞状态的分支将被剪除。[3] 在 $planHorizon$结束时，自车智能体会在算法4第1行选择使访问次数最大的第一个动作。随后重置该树，并持续执行这一过程，直到到达目标点。

算法 1 SoRTS($\mathbf{s}_0$,$\mathbf{s}_g$, θ, τ_R, v)

1:
2: **当** $\mathbf{s}_t^0 \neq \mathbf{s}_g^0$ 或 $timeElapsed \leq planHorizon$ **时执行**
3: $N(\cdot) \leftarrow$ SocialMCTS($\mathbf{s}_t$, $\mathbf{s}_G$, τ_r, θ, v, 0)
4: $\mathbf{a}^0 \leftarrow \arg\max_{\mathbf{a}' \in \mathcal{A}} N(\mathbf{s}_t^0, \mathbf{a}')$
5: $\mathbf{s}_{t+1}^0 \leftarrow f(\mathbf{s}_t^0, \mathbf{a}^0)$ ▷机器人的转移模型。
6: **结束 while**

V. 实验设置

我们的实验重点评估 SoRTS 在我们关注的两个方面的表现：即，导航性能和安全性。因此，本节介绍案例研究评估设置的主要内容及实现细节。

A. 用户研究

我们招募了 26 名经美国联邦航空管理局认证的飞行员（14 名私人飞行员、8 名商业飞行员、3 名飞行学员和 1 名航空运输飞行员），他们平均拥有 986 飞行小时。使用图1所示的驾驶舱配置和模拟器，每名飞行员都被要求将一架飞机降落在指定跑道上。同时，另一名飞行员也在完成同一项任务，因此需要该飞行员的

[3]注意：在实际应用中，对于 $M > 2$，我们在进行树扩展时只使用自我智能体和距离自车智能体最近的智能体。虽然树的显式构建仅针对两个智能体，但 p_θ 为所有智能体提供了高层次的社会背景。这种近似保留了算法的实时性，并且在实际应用中表现良好。

算法 2 SocialMCTS($\mathbf{s}_t$, $\mathbf{s}_G$, τ_r, θ, v, p)

1: **当** p ≤M **时循环**
2: **如果** $\mathbf{s}_t^p \notin S(\cdot)$ **则** ▷新树节点
3: $v_s \leftarrow$ GetValue(v, $\mathbf{s}_t$)
4: $S(\cdot) \leftarrow$ 获取社会行动概率($\mathbf{s}_t$, θ)
5: $p_r(\cdot) \leftarrow$ 获取参考动作概率(τ_r^p, $\mathbf{s}_t^p$)
6: $N(\mathbf{s}_t^p) \leftarrow 1$
7: **返回** v_s
8: **结束 if**
9: $\mathcal{A}' \leftarrow$ 碰撞检测($\mathcal{A}$, d, $\mathbf{s}_t$)
10: $\mathbf{a}^p \leftarrow$ 最大化$_{\mathbf{a}' \in \mathcal{A}'} U(\mathbf{s}_t^p, \mathbf{a}')$ ▷参见公式 3
11: $\mathbf{s}_{t+1}^p \leftarrow f(\mathbf{s}_t^p, \mathbf{a}^p)$
12: p ←p+1 ▷选择要扩展的下一个智能体
13: SocialMCTS($\mathbf{s}$, $\mathbf{s}_g$, τ_r, θ, v, p)
14: 更新(Q, R, S, N) ▷见公式 4 至 6
15: **结束循环**
16: **返回** $N(\cdot)$

协调以安全着陆。这里，第二名飞行员可以是人类、SoRTS 或基线算法[4]。

我们的实验设置遵循下文所述的设计。我们采用被试内 设计，每位用户都会测试每种算法。我们允许用户在测试前熟悉模拟器和控制装置。飞行员从机场半径 10 km 处生成，其进场方向为北（N）、南（S）或西（W），由此定义了六种可能的情景即， $\{(N,S), (S,N), (N,W), (W,N), (S,W), (W,S)\}$。实验开始时，算法顺序和情景随机选定。三次测试期间，情景、初始状态和最终目标点均保持不变。

每次测试后，用户都要填写一份五级李克特问卷，从多个我们认为与高导航性能和安全性相关的因素对第二名飞行员进行评价。由于在安全关键领域开展操作需要具备很强的能力，我们也希望了解用户在判断飞行员是否值得信赖且有能力时优先考虑哪些方面。因此，我们还请用户根据与第二名飞行员的互动以及其在导航和安全性方面的表现，评价他们认为该飞行员有多值得信赖且类人。我们在表 I中总结了用户研究问卷的组成部分。最后，我们还收集了每次实验的轨迹数据，以便使用第 V-C节讨论的相关指标进行进一步分析。

B. 自我博弈

我们通过自我博弈仿真补充用户研究，以评估 SoRTS 和我们的基线系统在更多样的情景中的表现。仿真采用与用户研究类似的设计，但现在允许每个智能体的位置位于 10 千米半径范围内的任意位置，以增加可能情景的数量。我们还考虑了智能体数量从 2 个到 4 个不等的多智能体情景。我们为每种设置随机生成 100 个回合，在每个回合中，如果某个智能体被判定为

[4]我们将第二名飞行员 和算法交替使用。

违反最小间隔距离，或达到允许的最大步数，就视为未成功。

C. 指标

为沿着各分析维度量化我们在用户研究和自博弈实验中收集的轨迹数据，我们考虑以下指标；

1) 相对误差（*RE*）：智能体的参考轨迹与其执行路径之间的欧几里得距离。2) 间隔丧失（*LS*）：两个智能体之间的距离低于最小间隔要求的持续时间。该指标与航空领域相关 [38],但类似于行人场景中社交机器人常用的指标，例如，个人空间 [1], [27]。

D. 基线系统

为展示 SoRTS 的优势，我们引入一种基线系统，该系统以朴素的方式按照下式，在状态-动作空间中平衡参考值和社交值，从而选择下一个动作，

$$\mathbf{a} = \arg\max_{\mathbf{a}' \in \mathcal{A}} \left[\lambda \cdot p_r(\mathbf{s}, \mathbf{a}) + (1 - \lambda) \cdot p_\theta(\mathbf{s}, \mathbf{a})\right]$$

其中 $\lambda \in \mathbb{R}$： $[0, 1]$ 用于平衡我们赋予$p_\theta(\cdot)$ 和 $p_r(\cdot)$的重要性。该基线系统等价于将上述算法3中的第4 行替换为上述公式1。

E. 模拟器

为了评估所提出的算法，并推动全尺寸航空自主领域的未来研究，我们推出了 X-PlaneROS。该系统通过 ROS 桥接器将两个主要组件连接起来：X-Plane-11 和 ROS-Plane 自动驾驶仪 [39]，，使用户能够使用高层和低层指令，在真实世界场景中控制通用航空飞机。X-Plane-11 是一款高保真模拟器，可提供逼真的飞机模型和视觉效果。ROS-Plane 是一款广受认可的工具，可提供可靠的自主飞行控制回路。我们的模拟器还支持执行一组选定的运动基元，并提供有助于调整控制回路的可视化工具。相关文档和源代码均已开源[5]。

F. 实现细节

IV-A 节中的模块 使用 TrajAir [9]，，这是一个由非塔台管制终端空域中采集的飞机轨迹数据组成的数据集；社交模块使用 TrajAir 离线训练其意图预测模块，训练细节参照 [10]。我们还使用 TrajAir 构建符合美国联邦航空管理局规定的航路库，供参考模块使用。最后，为构建我们的代价地图，我们根据飞机位姿和风向对 TrajAir 的飞行频率进行了离散化。

规划器的动作空间，$\mathcal{A}$，由用于航空导航的 252 个运动基元组成，这些基元应用于 [37]。根据经验，我们将 $planHorizon$=10s 设为算法2 中的取值，并将 UCT 方程中的探索参数设为 $c_1 = 2$和 $c_2 = 5$，基线规划器则设为 $\lambda = 0.3$ 。

表 I：我们请用户从第二名飞行员 在导航性能 和安全性方面的表现进行评分，并对其可信度和人性化程度进行评分。

导航性能	**安全性**	**第二名飞行员是否.....**
遵循FAA指南	碰撞风险	值得信赖?
总体飞行技能	舒适度	人类?
飞行平稳性	合作性	
	突兀	
	可预测	

VI. 结果

现在我们介绍结果与见解。在以下分析中，有能力的算法指在导航性能和安全性方面表现出色的算法。我们首先考察表I 中的各项因素对于描述有能力算法的可信度有何意义。基于我们的可信度分析结果，我们随后对 SoRTS、基线系统和人类飞行员进行比较分析。最后，我们简要探讨飞行员如何看待问卷中各项因素所反映的人类能力。

A. 关于能力与可信度

为了确定如何更好地评估机器人的能力，例如，以及应优先考虑哪些指标，我们将表 I中的因素与用户感知到的信任进行比较。我们采用重复测量的皮尔逊相关分析，结果汇总于表II。

我们发现，信任与导航性能中的各项组成因素存在较强相关性，其中飞行平稳性的相关性最强。在安全性方面，舒适度的相关性最强，其次是可预测性。尽管人们可能会认为，突兀性和碰撞风险对于信任评级至关重要。我们认为，这一结果部分是因为我们的实验并非专门用于呈现对抗行为。我们也认为，这些结果展现了我们的模型在前所未有的场景中的优势。正如前文所述，数据驱动模型在遇到分布外状态时可能会出现严重异常 [11],SoRTS 是一种在线增强这些模型的方法。由于这项用户研究旨在测试我们的模型在训练数据中未曾出现的情境下的表现，上述组成因素之间较低的相关性可能表明模型对不安全行为具有鲁棒性。

B. 各算法的能力

上一节研究了表 I 中的各项因素如何与值得信赖的飞行员的感知相关。本节将利用前一项分析所得结果，对算法 在用户研究和自我博弈中的表现进行两两比较。

1) 用户研究： 利用表格中可信度的结果 II，我们计算每位用户在各个维度相关性最高的问题上的平均得分，即，对于导航性能，我们计算以下得分的平均值

[5]X-PlaneROS 代码：https://github.com/castacks/xplane_ros

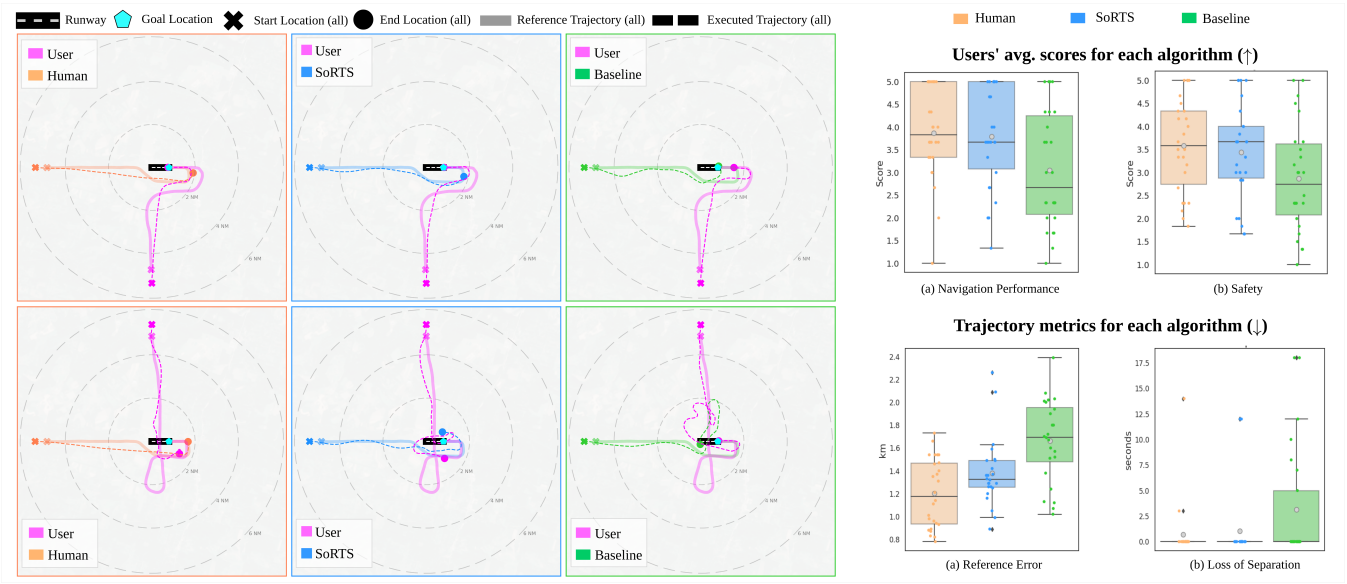


图 3：用户研究结果。**左侧：**两个实验示例。每一行展示了用户与我们的人类飞行员、SoRTS和基线系统交互后产生的轨迹。上方一行展示了一位成功遵循预期引导准则的用户；可以看到，基线系统飞行不如 SoRTS 平稳，并且在接近目标点时突兀地提前结束。下方一行展示了一位未遵循预期引导准则的用户；我们的算法仍能正确导航，而基线系统则表现反常，危险地穿越了跑道。**右侧：**箱线图展示了各算法的用户研究结果分布：上方箱线图展示了用户对导航性能（a）和安全性（b）的平均评分。下方箱线图展示了根据轨迹数据计算得出的平均参考误差（a）和间隔丧失（b）指标。

表 II：通过重复测量，我们展示了各因素的皮尔逊相关系数（R，p值=0.05），见表I与感知到的可信度和人性化程度之间的关系。

Axis	因素	值得信赖?		人类?	
		R	p值	R	p值
导航性能	飞行平稳性	0.81	4.80e-19	0.08	0.47
	遵循FAA指南	0.76	2.99e-15	0.10	0.37
	总体飞行技能	0.71	4.56e-13	0.15	0.20
安全性	舒适	0.92	3.55e-31	0.17	0.14
	可预测行为	0.77	5.19e-16	0.21	0.07
	合作性	0.65	1.61e-10	0.06	0.62
	碰撞风险	-0.56	1.10e-07	-0.11	0.36
	突兀	-0.56	1.13e-07	-0.09	0.42

介于以下美国联邦航空管理局 与 飞行平稳性，对于安全性，我们使用 可预测性和 舒适度。得分见图3，配对统计差异的重复测量方差分析结果见表 III。在这里，我们未发现人类飞行员与 SoRTS 得分存在差异的统计学证据，这表明用户对两者能力的评价相似。结果还表明，我们的基线系统在能力的两个维度上的评分普遍较低，且得分差异更大。

随后，我们考察第 V-C 节中的 RE 和 LS 指标——它们在此分别对应导航性能和安全性——是否反映了用户对各算法的评估。因此，我们在收集到的轨迹上计算这些指标，并在图 3 和表 III 中展示相应的平均分及统计比较。与前文相同，我们在图 3 中展示各算法的平均分，并在

表 III 中展示统计结果。对于 RE 指标，我们发现各算法之间的结果存在显著差异：SoRTS 的误差高于人类飞行员，但明显低于基线系统，且方差小于另外两种算法。类似地，在 0.5 km 处计算的 LS 指标显示，人类飞行员和 SoRTS 通常都不会侵入这一间隔距离。相比之下，基线系统更频繁地侵入他人的个人空间，从而导致更多可能发生碰撞的情形。

表 III：图 3 中的算法配对结果的统计显著性，采用 $t^* \geq 2.060$ 和 $p \leq 0.05$。

算法的 Pair	NP		安全性		RE		LS	
	t值	p值	t值	p值	t值	p值	t值	p值
人类基线	3.121	0.009	3.062	0.016	5.782	0.000	1.321	0.199
SoRTS基线	3.018	0.009	2.626	0.022	2.105	0.011	0.397	0.694
Human-SoRTS	0.415	0.682	0.322	0.322	2.944	0.022	1.211	0.237

导航性能：导航性能，**间隔距离**：间隔丧失，**相对误差**：参考误差。

图 3 展示了我们实验中的一些轨迹示例。每一行代表一位用户与三种对比的算法。我们展示了参考轨迹以及实际执行的轨迹[6]。第一行展示了用户成功完成任务的示例。在这项实验中，我们发现消融版本在接近跑道时意外地提前中断，没有沿交通模式飞行；而 SoRTS 则平稳、安全地完成了任务。第二行展示了一位没有遵循标准起落航线的用户。尽管

[6]人类飞行员看不到参考轨迹，但能看到模拟器中的地图，并被要求遵循相应的交通模式。

如此，SoRTS 仍成功完成了任务，而基线系统表现紊乱，既没有遵循交通模式，还不安全地两次穿越跑道。

根据用户研究，我们得出结论：无论是用户的主观感受，还是通过我们的指标衡量，SoRTS 的表现都与能力出色的人类飞行员相当，并显著优于消融版本。我们的结果进一步凸显了通过蒙特卡洛树搜索进行长时域社会感知仿真的优势。相比之下，仅依赖数据驱动模型在真实环境中运行，并对关注的目标进行简单加权，会导致行为反复无常且不安全。

表 IV：基线系统和 SoRTS 智能体在自我博弈中的任务表现汇总。

# 智能体	算法	成功率（%）	失败率（%）		RE
			LS	超时	
2	基线系统	69.0	30.0	1.0	1.59
	SoRTS	96.5	2.0	1.5	2.01
3	基线系统	48.3	49.3	2.4	1.55
	SoRTS	89.7	8.3	2.0	2.01
4	基线系统	43.5	54.0	2.5	1.56
	SoRTS	71.0	21.0	8.0	1.96

间隔距离：**相对误差**：成功智能体的参考误差。

2) 自我博弈： 现在，我们分析算法在更复杂情景中的表现。表IV 展示了能够降落在跑道上的智能体所占的比例，即成功的智能体，以及因间隔丧失或超过最大允许时间而未成功的智能体。最后，我们展示成功智能体的平均参考误差。

分析该表可见，随着智能体数量增加，任务成功率有所下降。尽管如此，我们可以看到，SoRTS 的表现显著优于消融算法，在 2、3 和 4 个智能体情景中，任务成功率分别提高了 28%、46% 和 38%。我们还发现，SoRTS 的平均参考误差高于消融算法。虽然该误差仅针对成功的智能体计算，但我们推测，误差较高是由于为避免与其他智能体发生碰撞而进行冲突消解所致。

C. 关于能力与人类表现

人们常常争论一个算法能否被视为人类。然而，在航空等以高能力为核心的领域，人们理应会被视为具备这种能力。因此，我们希望算法能够展现出与人类相似的表现。我们通过请用户根据表 I 中的评估结果，判断第二名飞行员是否为人类，来研究这一问题。

令人惊讶的是，我们发现用户对这一问题的回答存在明显分歧，认为人类飞行员 表现得像人类的回答与认为其不像人类的回答几乎各占一半（14：否，12：是）；对于基线系统，回答同样几乎各占一半（14：否，12：是）。而对于*SoRTS*，更多用户认为其表现具有类人特征（8：否，18：是）。此外，表II 显示，用户对以下方面的评估导航表现和安全性与人性化程度预测的相关性较弱。为解释这一相反的结果，我们将针对 SoRTS 的回答与另外两者的回答分开分析，发现可预测性在所有因素中相关性最高，R=0.53，p=0.01。相比之下，人类飞行员和基线系统的数值分别为 R=0.22，p=0.03 和 R=-0.06，p=0.06。我们认为，这一结果可能表明，由于用户普遍认为 SoRTS 的行为更可预测，因此他们更常将其表现评为具有类人特征。

我们认为，这一对比结果不会影响前两节中的分析，因为我们发现，用户对可信度的回答与其对人性化程度的评估也仅呈弱相关，R=0.26，p=0.02。因此，我们得出结论：用户更看重自己对信任的感知，而非所交互智能体的类型。

VII. 结论

我们提出了 SoRTS，这是一种用于长时域社交机器人导航的算法。SoRTS 是一种基于蒙特卡洛树搜索的规划器，它利用离线训练的意图预测模型和体现导航指南的全局路径来引导搜索树的扩展。我们介绍了 X-PlaneROS，这是一款用于全尺寸航空自主研究的高保真模拟器。我们利用该模拟器与经验丰富的飞行员开展用户研究，以考察算法在真实飞行场景中的表现。我们发现，用户认为 SoRTS 的表现可与具备能力的人类飞行员相媲美，且显著优于我们的基线系统。在自我博弈中，我们表明，随着情景复杂度增加，SoRTS 的表现比基线系统高出 28%–46%。据我们所知，这是通用航空社会导航领域的首项研究，尝试将通用航空中的独特问题纳入更广泛的机器人学界的研究视野。

我们指出了未来工作的若干方向。首先，我们假设任务具有同质性，即智能体降落在同一条跑道上；然而，在真实场景中，目标不同的飞行员可能需要相互交互。因此，未来的工作包括研究具有异质任务的智能体之间的交互。我们还假设意图和状态估计是完美的。因此，使预测模型具备不确定性感知和对抗性感知能力，是另一个有前景的方向。最后，提高模型的可扩展性以及探索其他领域，也是潜在的研究方向。

致谢

本研究得到陆军未来司令部人工智能集成中心（AI2C）、产业通商资源部（MOTIE）、韩国技术振兴院（KIAT）和巴西空军的支持。本文所表达的观点不一定代表上述机构的立场。我们感谢 Jasmine Jerry Aloor 提供的支持，也感谢 Condor航空俱乐部（KPJC）和 ABC飞行俱乐部（KAGC）的飞行员参与我们的用户研究。

参考文献

[1] C. Mavrogiannis、F. Baldini、A. Wang、D. Zhao、P. Trautman、A. Steinfeld 和 J. Oh，《社交机器人导航的核心挑战：综述》，arXiv 预印本 arXiv:2103.05668，2021年。[2] A. Rudenko、L. Palmieri、M. Herman、K. M. Kitani、D. M. Gavrila 和 K. O. Arras，《人体运动轨迹预测：综述》，《国际机器人研究杂志》，第39卷第8期，第895–935页，2020年。[3] R. Tian、L. Sun、A. Bajcsy、M. Tomizuka 和 A. D. Dragan，《通过置信度感知的博弈论人类模型为人机交互提供安全保障》，载于2022年国际机器人与自动化会议（ICRA）。IEEE，2022年，第11 229–11 235页。[4] D. Sadigh、N. Landolfi、S. S. Sastry、S. A. Seshia 和 A. D. Dragan，《与人协同行驶的汽车规划：利用对人类行为的影响，围绕人类内部状态开展规划和主动信息收集》，《自主机器人》，第42卷第7期，第1405–1426页，2018年。[5] B. Riviere、W. H¨onig、M. Anderson 和 S.-J. Chung，《面向非合作环境下多机器人规划的神经树扩展》，《IEEE机器人与自动化快报》，第6卷，6875，2021 4，6868期，第–页。[6] S. Matsuzaki 和 Y. Hasegawa，《通过分布式深度强化学习，在具有挑战性的环境中学习具备人群感知能力的机器人导航》，载于2022年国际机器人与自动化会议（ICRA）。IEEE，2022年，第4730–4736页。[7] C. I. Mavrogiannis、V. Blukis 和 R. A. Knepper，《通过学习多智能体路径拓扑实现具备社交能力的导航规划》，载于2017年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，2017年，第6817–6824页。[8] T. Salzmann、B. Ivanovic、P. Chakravarty 和 M. Pavone，《Tra-jectron++：利用异构数据进行动态可行的轨迹预测》，载于欧洲计算机视觉会议。Springer，2020年，第683–700页。[9] J. Patrikar、B. Moon、J. Oh 和 S. Scherer，《像飞行员一样预测：用于预测具备社交感知能力的航空器轨迹的数据集与方法》，载于2022年国际机器人与自动化会议（ICRA）。IEEE，2022年，第2525–2531页。[10] I. Navarro 和 J. Oh，《Social-patternn：以运动模式为引导的社会感知轨迹预测》，载于2022年IEEE/RSJ国际智能机器人与系统会议（IROS）。IEEE，2022年，第9859–9864页。[11] A. Farid、S. Veer、B. Ivanovic、K. Leung 和 M. Pavone，《自动驾驶汽车轨迹预测器的任务相关故障检测》，载于机器人学习会议。PMLR，2023年，第1959–1969页。[12] H. Hu 和 J. F. Fisac，《人机交互中的主动不确定性降低：一种隐式双重控制方法》，arXiv 预印本 arXiv:2202.07720，2022年。[13] D. Silver、T. Hubert、J. Schrittwieser、I. Antonoglou、M. Lai、A. Guez、M. Lanctot、L. Sifre、D. Kumaran、T. Graepel 等，《一种通过自我博弈精通国际象棋、将棋和围棋的通用强化学习算法》，《科学》，第362卷第6419期，第1140–1144页，2018年。[14] A. P. Jacob、D. J. Wu、G. Farina、A. Lerer、H. Hu、A. Bakhtin、J. Andreas 和 N. Brown，《利用KL正则化搜索建模高水平且类人的游戏玩法》，2021年。[15] L. Kocsis 和 C. Szepesv´ari，《基于老虎机的蒙特卡洛规划》，载于《机器学习：第17届欧洲机器学习会议》，德国柏林，2006年9月18日至22日，会议论文集，《计算机科学讲义》系列，J. F¨urnkranz、T. Scheffer 和 M. Spiliopoulou 编，第4212卷。Springer，2006年，第282–293页。[16] J.-P. Aurambout、K. Gkoumas 和 B. Ciuffo，《无人机末端配送：对欧洲各城市可行市场潜力及市民可及性的估算》，《欧洲交通研究评论》，第11卷第1期，第1–21页，2019年。[17] M. Grote、A. Pilko、J. Scanlan、T. Cherrett、J. Dickinson、A. Smith、A. Oakey 和 G. Marsden，《与无人驾驶航空器（UAV）共享空域：通用航空（GA）群体的观点》，《航空运输管理杂志》，第102卷，文章编号102218，2022年。[18] J. v. d. Berg、S. J. Guy、M. Lin 和 D. Manocha，《互惠式N体碰撞规避》，载于《机器人研究》。Springer，2011年，第3–19页。[19] C. Mavrogiannis、P. Alves-Oliveira、W. Thomason 和 R. A. Knepper，《社交动量：具备社交能力的机器人导航框架的设计与评估》，ACM人机交互汇刊（THRI），第11卷第2期，第1–37页，2022年。

[20] Y. F. Chen、M. Liu、M. Everett 和 J. P. How，“基于深度强化学习的去中心化非通信多智能体碰撞规避”，载于 2017年IEEE国际机器人与自动化会议（ICRA）。IEEE，2017年，第 285–292 页。[21] C. Chen、Y. Liu、S. Kreiss 和 A. Alahi，“人群-机器人交互：基于注意力机制的深度强化学习实现人群感知机器人导航”，载于 2019年国际机器人与自动化会议（ICRA）。IEEE，2019年，第 6015–6022 页。[22] C.-E. Tsai 和 J. Oh，“一种实现社会规范导航的生成式方法”，载于 2020年IEEE国际机器人与自动化会议（ICRA）。IEEE，2020年，第 2160–2166 页。[23] A. Biswas、A. Wang、G. Silvera、A. Steinfeld 和 H. Admoni，“Soc-navbench：用于评估社会导航的基于现实的仿真测试框架”，ACM人机交互汇刊（THRI），第 11 卷，第 3 期，第 1–24 页，2022年。[24] M. A. Bashiri、B. Ziebart 和 X. Zhang，“分布鲁棒模仿学习”，神经信息处理系统进展，第 34 卷，第 24 404–24 417 页，2021年。[25] F. Codevilla、E. Santana、A. M. L´opez 和 A. Gaidon，“探索行为克隆在自动驾驶中的局限性”，载于 IEEE/CVF 国际计算机视觉会议论文集，2019年，第 9329–9338 页。[26] S. Schaefer、K. Leung、B. Ivanovic 和 M. Pavone，“在轨迹优化中利用神经网络梯度实现主动式人机交互”，载于 2021年IEEE国际机器人与自动化会议（ICRA）。IEEE，2021年，第 9673–9679 页。[27] Y. Gao 和 C.-M. Huang，“社会感知机器人导航评估”，《机器人与人工智能前沿》，第 420 页，2021年。[28] C. Mavrogiannis、A. M. Hutchinson、J. Macdonald、P. Alves-Oliveira 和 R. A. Knepper，“不同机器人导航策略对拥挤环境中人类行为的影响”，载于 2019年第14届ACM/IEEE国际人机交互会议（HRI）。IEEE，2019年，第 421–430 页。[29] A. J. Sathyamoorthy、J. Liang、U. Patel、T. Guan、R. Chandra 和 D. Manocha，“Densecavoid：利用预判行为在人群密集环境中实现实时导航”，载于 2020年IEEE国际机器人与自动化会议（ICRA）。IEEE，2020年，第 11 345–11 352 页。[30] S. Liu、P. Chang、W. Liang、N. Chakraborty 和 K. Driggs-Campbell，“基于深度强化学习、用于机器人在人群中导航的去中心化结构化循环神经网络”，载于 2021年IEEE国际机器人与自动化会议（ICRA）。IEEE，2021年，第 3517–3524 页。[31] M. Everett、Y. F. Chen 和 J. P. How，“利用深度强化学习在动态决策智能体之间进行运动规划”，载于 2018年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，2018年，第 3052–3059 页。[32] J. Liang、U. Patel、A. J. Sathyamoorthy 和 D. Manocha，“利用隐式多传感器融合和深度强化学习，在密集人群中为移动机器人实现实时碰撞规避”，arXiv 预印本 arXiv:2004.03089，2020年。[33] E. Torta、R. H. Cuijpers 和 J. F. Juola，“用于机器人社会导航的个人空间参数模型设计”，《国际社会机器人学杂志》，第 5 卷，第 3 期，第 357–365 页，2013年。[34] Y. F. Chen、M. Everett、M. Liu 和 J. P. How，“基于深度强化学习的社会感知运动规划”，载于 2017年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，2017年，第 1343–1350 页。[35] J. Butler 和 A. Agah，“移动个人机器人行为模式的心理影响”，《自主机器人》，第 10 卷，202、03 2001 185页。–。[36] R. Emery-Montemerlo、G. J. Gordon、J. G. Schneider 和 S. Thrun，“具有共同收益的部分可观测随机博弈近似解”，载于第3届自主智能体与多智能体系统国际联合会议（AAMAS 2004），2004年8月19日至23日，美国纽约州纽约市。IEEE计算机学会，2004年，第 136–143 页。[37] J. J. Aloor、J. Patrikar、P. Kapoor、J. Oh 和 S. Scherer，“遵循规则：面向随机领域引导式模仿学习的在线信号时序逻辑树搜索”，arXiv 预印本 arXiv:2209 13737，2022年。[38] T. Glozman、A. Narkawicz、I. Kamon、F. Callari 和 A. Navot，“一种基于视觉的解决方案，用于估算探测与避让中的到达最近接近点的时间”，载于 AIAA Scitech 2021论坛，2021年，第 0450 页。[39] G. Ellingson 和 T. McLain，“Rosplane：面向教育与研究的固定翼自动驾驶仪”，载于 2017年国际无人机系统会议（ICUAS）。IEEE，2017年，第 1503–1507 页。