基于IL+TD3的城市作战场景无人集群协同抗干扰轨迹规划

# 研究综述

城市作战场景下无人机集群轨迹规划协同抗干扰研究主要由多智能体路径规划（multi-agent path finding，MAPF）以及无人机集群城市协同任务两方面研究组成。本节将分别从**MAP**F和**无人机集群城市协同任务**两方面介绍相关的研究发展现状。

## MAPF发展现状

多智能体路径规划（multi-agent path finding，MAPF）是为多个智能体规划路径的问题，其关键约束是多个智能体同时沿着规划路径行进而不会发生冲突，这里的冲突包括和环境的冲突以及智能体之间的冲突。

现有的MAPF算法按照规划方式不同，分为集中式规划算法和分布式规划算法。集中式规划算法是最经典和最常用的MAPF算法，基本由单智能体路径规划算法发展而来，主要分为**基于A\*搜索**、**基于冲突搜索**、**基于代价增长树和启发式算法**四种算法；分布式执行算法主要基于强化学习，智能体根据当前策略不断与环境进行交互，完成多智能体路径规划任务。其目前主要研究方向分为专家演讲型，改进通信型和任务分解型三种。

基于A\*搜索算法在小规模智能体环境中是最佳搜索算法，其由单体A\*算法发展而来。A\*搜索在MAPF中的状态一般被称为k-agent状态空间，其状态数等于不同智能体的联合状态数。通过将所有智能体各自启发式函数之和作为全局启发函数，单智能体A\*搜索能够扩展应用到MAPF 中。在k-agent的搜索空间内，最大的搜索空间为智能体数量的乘积，因为A\*必须沿着最优路径扩展所有的节点，所以A\*算法在解决大规模智能体的MAPF 问题时效率和质量都不高。

**基于冲突搜索算法（CBS）是目前最常用的算法,目前求解MAPF 速度和质量最好的算法大都是在基于冲突搜索算法上进行改进和优化**。其求解过程分低层搜索和高层搜索，底层搜索通过采样为所有智能体单独构建一棵二叉搜索树查找解，高层搜索通过添加限制的方式解决碰撞冲突。具体流程为底层执行带有约束的单智能体规划，使用传统 A\*，RRT\* 算法等进行求解，顶层通过遍历底层的规划路径，检查路径之间是否有冲突，如果有冲突则施加约束重新进行底层单机规划，直到所有底层路径无冲突为止。

基于代价增长树的搜索算法（ICTS）以及其他启发式算法如遗传算法、粒子群算法等也常被用于解决MAPF问题。

使用强化学习方法解决MAPF 问题也有大量研究，其主要利用多智能体强化学习的对应框架如DPDG、TD3、PPO和SAC等进行对应问题的求解。然而其面临的许多挑战，例如环境奖励稀疏、环境动态复杂等。上述任何一种强化学习算法直接应用于MAPF 问题都会出现学习速度慢和求解质量不高的问题。针对以上问题，目前采用了各种组合技术对基于强化学习的MAPF 方法进行改进，使得强化学习的MAPF 方法能够扩展到大规模智能体的环境。按照改进技术的特点，大致将基于RL 的MAPF 方法分为专家演示型、改进通信性和任务分解型三类。

整体来说，传统的算法如基于CBS的多智能体轨迹规划算法能够有效解决复杂场景下的问题并得到优秀的解。但是其场景下一般的优化目标较为简单，主要有最小化所有智能体到达目标位置所需的最大时间，最小化所有智能体到达目标位置所需时间的总和以及最小化所有智能体路径的总长度等三种。传统的算法难以建模智能体之间协同关系，无法在轨迹规划的同时有效地解决智能体之间需要高度协同规划场景下如轨迹协同抗干扰，任务规划等问题。

而基于强化学习的算法主要面临以下三个问题：

①**状态空间维度高**：​随着智能体数量的增加，状态空间呈指数增长；​

②**部分可观察性**：​智能体只能获取有限的环境信息，增加了决策的复杂性；

**③协同策略学习难度大**：​需要智能体之间有效的协作策略，以避免冲突并提高效率。

**基于强化学习的求解方法能够有效地建模智能体之间的协同关系，保障智能体在轨迹规划的主要目标下同时优化求解路径规划过程中的协同任务的执行。**

## 无人机集群协同任务发展现状

近年来，无人机（UAV）已成为未来交付、服务巡逻和作战打击应用的重要组成部分。轨迹规划在多无人机应用中非常重要。作为城市空中应用，自动化多无人机系统能够提供自主规划和执行任务的作用。在很多场景下，多无人机系统需要配合才能完成对应的任务。因此，多无人机系统的组织和管理是更复杂的系统级优化问题，需要高效的轨迹规划和优化策略才能用于实际应用，这个问题涉及轨迹规划和冲突避免两个方面。

**目前城市场景下无人机集群协同轨迹规划已有大量研究,但是针对无人机集群抗通信干扰的规划研究几乎缺失**。Yao等[1]开发了一种基于扰动流体和轨迹传播的新算法，解决了无人机在静态环境下的三维路径规划问题。Xia等人[2]建立了基于均匀时间间隔的轨迹优化模型，并提出了一种基于梯度的序贯最小优化（GB-SMO）算法来解决无人机规划问题。Gao等人[3]研究了基于战争环境的战场任务轨迹规划，基于Informed-RRT 算法，以最短轨迹和最短规划时间为目标。上述研究对基于不同背景的多无人机轨迹规划进行了有益的探索;**然而，它们都没有涉及协同抗干扰规划领域。**此外，Wang 等[4]提出了一种用于城市路网监控任务的启发式算法，以研究旋翼无人机的多飞机轨迹规划。此外，Zhang 等[5] 基于城市任务背景，针对无人机在城市环境中飞行对行人和公共财产构成较大威胁的问题，建立了无人机轨迹规划的数学模型，并引入了风险评估。此外，他们提出了一种改进的蚁群算法，用于优化轨迹规划，并取得了一定的效果。Na 等人。[6]另一方面，专注于基于城市环境的无人机3D路径优化和实时避障。Muñoz 等[7]反而专注于城市无人机的编队控制。Wang 等[8]还研究了基于城市环境和混合策略的无人机飞行规划，以保证城市飞行的顺利进行。此外，关于基于城市环境的无人机轨迹规划的研究仍然很多[9][10]。**以上研究都对无人机城市领域应用做出了重大贡献。然而，他们的研究目标仅限于最短轨迹和最大安全系数、任务完成率等问题**，并不考虑到无人机之间协同通信能力保持的问题。

**根据上述城市场景无人机集群轨迹规划的相关研究，本研究的创新点在于着重于无人机集群规划过程中的网络通信保持任务，保障无人机集群的通信连通性，也是未来无人机集群智能组网的通信基础。**

## 相关研究不足与本文创新点

现有的关于无人机集群路径规划研究主要有以下两点不足：

①**城市作战场景下无人机集群协同抗干扰规划研究不足：**根据相关研究综述，针对城市作战场景下的无人机路径规划问题，主要存在的研究基本上是**任务分配＋轨迹规划类型**的复合问题，而不考虑无人机集群在运动过程中对通信需求的满足。而城市作战场景通信抗干扰的复杂性主要体现在以下两点：一是建筑物存在导致的视距链路遮挡，使得通信距离大幅度下降；二是干扰信号的存在导致的通信距离下降。现有的研究很少聚焦城市作战场景下无人集群通信需求的满足，而通信需求是实现无人机集群智能化的首要前提，因此本研究主要是为了解决这方面的缺失。

②传**统多智能体轨迹规划算法难以求解对协同性要求高的规划任务：**现有的传统的最常用的多智能体路径规划算法如基于冲突搜索的方式（CBS）都是先进行单智能体路径规划，然后再解决智能体之间路径的冲突并迭代优化全局目标（如最晚到达时间，最短路径总和等），对于无人机运动过程之间的协同需求（如协同抗干扰、协同避障等）难以得到有效解决。针对本文研究场景下无人机集群轨迹规划协同抗干扰任务，任务期间需要保持每一时刻无人机之间网络的完全连通的性质，现有算法难以建模任务规划过程中无人机的高度协同性，导致难以有效解决该问题。

针对在无人机集群城市作战场景下的以上不足，本研究通过场景建模，利用强化学习实现对无人机集群城市作战场景下协同抗干扰轨迹规划任务，主要创新点为：

**①基于TD3实现无人机集群协同抗干扰轨迹规划，补充城市场景下通信协同任务的空白:**本研究基于TD3网络，通过设置对应场景以及信道建模，引入代数连通度作为衡量网络通信能力的指标，设计对应的奖励函数训练网络，实现5节点网络在完成轨迹规划的同时，能够保障无人机集群在飞行过程中，避免建筑物遮挡和干扰信号等带来的通信链路中断，始终保持网络的连通性。

②**基于IL+RL实现大规模无人集群协同抗干扰轨迹规划研究，降低大规模智能规划的实现复杂度：**当无人机集群数目增加时，由于现有的单一多智能体强化学习手段在面对高维度空间时学习效率低，收敛难度高的问题，本研究通过模仿学习加强化学习的架构，首先通过传统的有效的CBS方式生成初步粗粒度的可行解，然后作为强化学习的指导先验知识，训练网络时再引入通信保持作为优化目标，生成最终解。根据这种方式能够极大提高网络学习率并且实现通信优化的目标。

# 基于TD3的无人机集群协同轨迹规划抗干扰研究

## 系统模型

场景示意图如图所示，无人机集群初始位置为地图左下角，其对应目标位置为地图右上角。矩形建筑随机生成在地图范围内，对应的固定位置的干扰生成在地图中央。

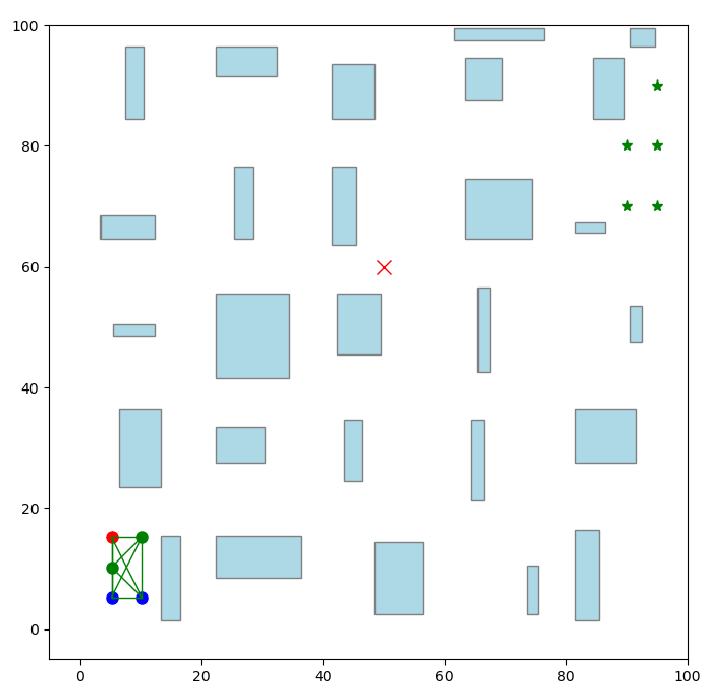


图 ‑1无人机集群协同抗干扰路径规划示意图

每架无人机需要规划好各自的路径到达目的地，并且保证无人机在飞行过程中能够能够完成以下任务：

**①碰撞避免：**无人机需要防止和建筑物碰撞，也需要防止和其他无人机之间的碰撞。

**②通信保持：**无人机需要协调各自位置，保证飞行过程中通信网络的全连通性。

针对通信需求，本研究采取视距链路和城市衰落两种信道模型分别建模有无建筑物遮挡下的情况。其中视距链路采取自由空间衰落模型：

其中建筑物遮挡情况下采取城市衰落模型：

其中rx表示接受功率，tx表示发送功率，σ表示路径损耗因子，反映信号在空间中的衰减速度，典型值范围：2（自由空间）到 6（城市环境），α表示阴影衰落项。

|  |
| --- |
|  |

**为了计算无人机飞行过程中的连通性，引入代数连通度作为表征。**代数连通度是图论中衡量图连接程度的重要指标，是图的拉普拉斯矩阵的第二小特征值。当代数连通度为0时，表示图是不连通的；值越大，说明图的连通性越强。其计算公式为：

其中λ₂是图的拉普拉斯矩阵L = D - A中的第二小特征值，D为度矩阵，A为邻接矩阵。上述无向图的构建由无人机之间的SINR计算得出，不满足双向通信的链路视作不连通。

通过场景信道建模以及代数连通度的引入，能够基本表示城市作战场景下无人机集群协同抗干扰的真实物理需求，后续将在此基础上实现系统问题的求解。

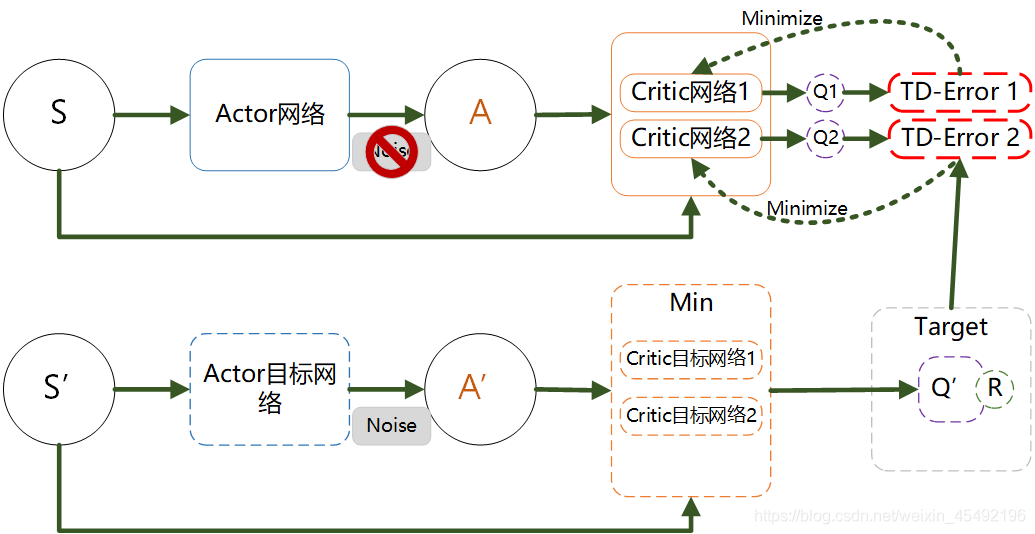
## 算法原理

本研究基于TD3网络实现城市作战场景下的无人机路径规划协同抗干扰任务。下面将分别介绍网络的状态-动作空间、网络模型和奖励函数设置。

**状态空间**描述了环境中所有可能的状态信息。在本研究中状态空间包括无人机的位置、速度、功率等信息。无人机的位置以二维坐标（x, y）表示，反映其在地图中的具体位置；速度以向量（vx, vy）表示，描述其当前的运动状态；功率表示无人机的通信能力，其范围由w\_min\_powers和w\_max\_powers决定。此外，状态空间还可能包含干扰源的位置和功率信息，这些会影响无人机的通信质量，。状态空间的范围整体上受网格大小、最大速度drones\_v\_max等参数的限制。

**动作空间**定义了无人机在每个时间步可以采取的动作，包括加速度、功率调整。加速度以向量（ax, ay）表示，用于控制无人机的运动方向和速度，其范围为[-drones\_a\_max, drones\_a\_max]；功率调整用于控制无人机的通信信号强度，其范围由drones\_min\_power和drones\_max\_power决定；。动作空间是一个连续空间，加速度和功率都是连续变量。

本研究采取TD3网络作为问题求解的研究框架，其具体网络架构如下：



TD3（Twin Delayed Deep Deterministic Policy Gradient）是一种改进的确定性策略强化学习算法，其网络结构主要由一个actor网络和两个独立的 critic网络组成。actor网络负责生成在当前状态下的确定性动作，而两个 critic网络则用于评估该动作的Q值，从而减少由于函数逼近误差导致的过估计问题。TD3在结构上与DDPG相似，但引入了三项关键改进：一是使用两个独立的 critic 网络取其输出最小值来计算目标 Q 值，从而有效缓解过估计偏差；二是引入目标策略平滑，向目标动作中添加微小噪声，使critic网络对输入动作的微小扰动更加鲁棒；三是延迟更新策略网络，即 actor 网络的更新频率低于 critic，从而提升训练稳定性。这些改进使 TD3 在连续动作空间任务中表现出更强的稳定性和性能。

**奖励函数设置**是使得学习任务收敛速度和优化效果的重点，本研究奖励函数由以下三部分组成：**碰撞奖励、连通度奖励和距离奖励。**首先，奖励函数会对与建筑物发生碰撞的无人机和无人机之间的碰撞添加惩罚机制，每次碰撞会减少固定奖励值。其次，根据当前时间步的拓扑连通度与上一时间步的差值来调整连通奖励项，如果当前的拓扑连通度大于零，则奖励值为预定义的固定值；否则，奖励值为零。奖励的增量为当前连通度与上一时间步连通度的差值，同时更新上一时间步的连通度值为当前连通度，最后是距离奖励，根据不同状态距离目标位置的距离差值，计算无人机靠进目标的奖励，远离目标时为负数。

通过这种设计，奖励函数综合考虑了无人机与建筑物的碰撞、无人机之间的碰撞以及拓扑连通度的变化，使得网络输出能够同时实现目标到达、碰撞避免和通信保持三项任务。具体公式如下所示:

其中N为无人机数量，为无人机移动前后距离自身目标点的距离差值，为无人机每发生一次碰撞的奖励值，为无人机是否和建筑物发生碰撞，为当前无人机是否和其他无人机发生碰撞，为无人机网络保持连通时的奖励值，为网络是否连通，由代数连通度计算得到。

## 仿真结果

### 参数设置

参考实际真实场景的参数设置，城市作战场景参数设置如下表：

表 ‑ 仿真场景参数设置

|  |  |
| --- | --- |
| 场景尺寸 | 2km\*2km |
| 无人机数量 | 5 |
| 干扰源数量 | 1 |
| 无人机发射功率 | 1W |
| 干扰源发射功率 | 1W |
| 通信SINR门限 | 10dB |
| 无人机速度范围 | (-10,10)m/s |
| 无人机加速度范围 | (-10,10)m/s^2 |
| 信道频点 | 2.4GHz |
| 带宽 | 1MHz |
| 高斯白噪声功率谱密度 | -174dbm/Hz |
| 信道衰落模型参数 | σ= 5,α= 3.5 |

此外，场景中的建筑物避开对应的干扰源与无人机的位置，随机生成于地图中。建筑物密度可以使用固定参数控制。

### 训练结果

本节主要描述算法收敛性和对应的效果。

**针对5节点场景，**设置actor学习率为1e-3和1e-4，每轮训练最大步数上限为300steps,训练30000个episodes训练结果如图所示，首先是训练过程中每隔30个episodes进行模型测试，获取到的奖励随episodes的变化曲线。然后是使用模型部署后进行轨迹规划测试得出的轨迹示意图。

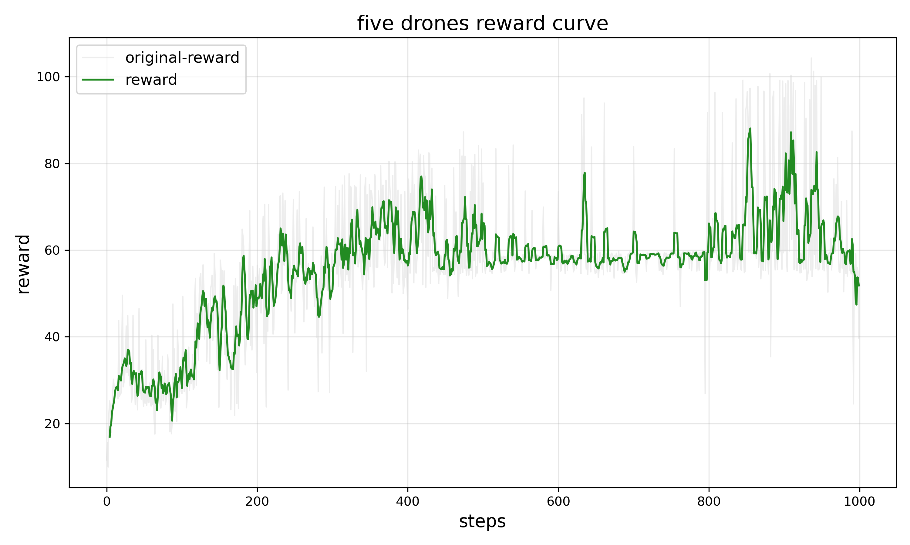


图 ‑2 五节点TD3协同抗干扰路径规划奖励随轮次变化

根据测试中统计的奖励图可以看出随着训练回合数增加，网络决策收到的reward逐渐增加， 代表着节点离目标节点的平均距离越来越接近，最终到达目的地。但是目前还存在**局部最优困境**，测试大量的参数，包括学习率、探索噪声强度、奖励调整后，网络始终趋近于收敛至奖励**60**左右的区间，而目标最优值为**110**左右，分析未经平滑的原始图像可以看出网络在训练过程中的最优结果能够达到**110,**满足规划需求**。**造成这个问题的原因是**干扰信号的存在导致集群在奖励值为60处的训练难度更大，节点针对有效的经验的利用率更低，导致突破该局部最优点难度较大。**

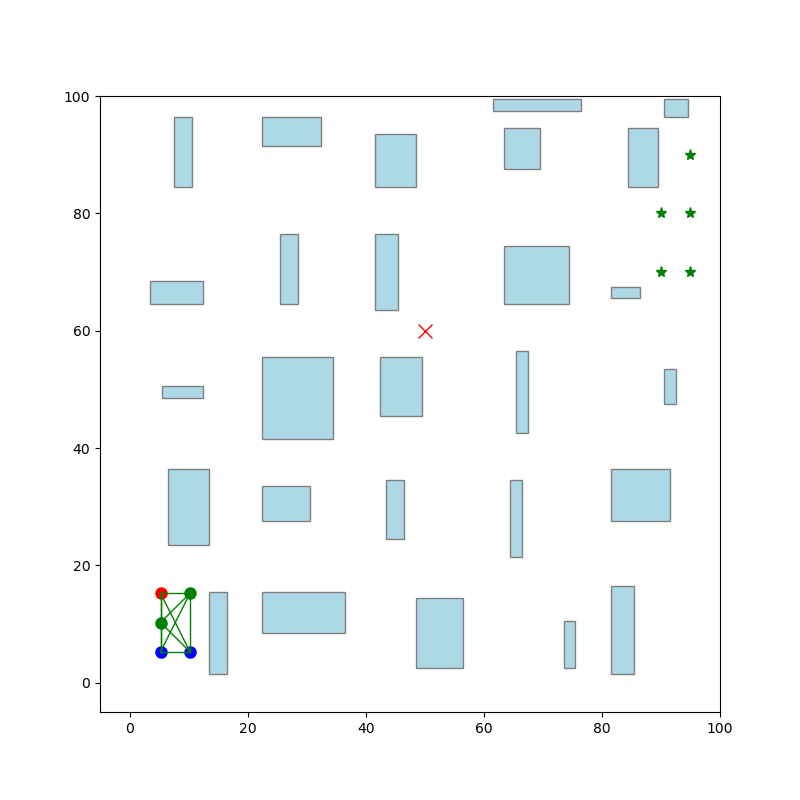


图 ‑3 五节点协同抗干扰路径规划效果图

加载最优效果模型测试可以得出，网络输出的**加速度动作序列**能够保证五架无人机全部到达终点并且在飞行过程中能够**协同规划，避开建筑物以及干扰信号**对链路的影响，**始终保持无人机之间通信的连通性**。

### 结果对比

将基于CBS加RRT\*的传统路径规划算法和本研究提出的基于TD3的协同路径规划算法的规划结果进行比较，统计两种规划算法下，网络通信连通性保持率之比作为对比，其中通信保持率定义为：

其中表示无人机集群网络连通的时间长度，表示无人机从出发到全部到达目的地所占的时间。通信保持率越高代表无人机集群在飞行过程中协同通信的能力越强。

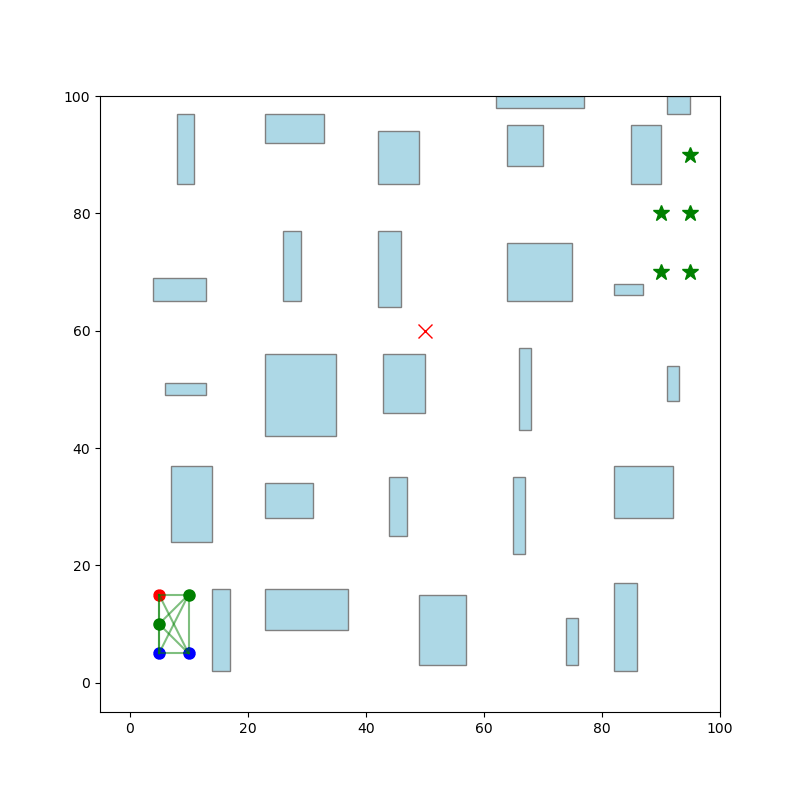


图 ‑4 五节点基于CBS-RRT\*的路径规划效果图

如图展示了基于CBS与RRT\*算法路径规划的效果图，可以看出基于CBS的路径规划方法，难以有效地处理无人机之间的协同关系，无法保证无人机集群在飞行过程中保持网络通信的连接。

表 ‑ 通信保持率对比

|  |  |
| --- | --- |
| TD3-MARL | CBS-RRT\* |
| 100% | 46.7% |

上表展示了在同样的场景参数下，两种算法分别统计的通信保持率数据。可以看出，基于CBS-RRT\*算法所得结果的在统计十次平均规划结果后得到的通信保持率为46.7%，而本研究提出的算法能够保证无人机在飞行过程中保持全程连通，有效保证无人机集群任务过程中的通信连接。

# 基于IL+COMATD3的大规模无人集群协同轨迹规划抗干扰研究

## 算法设计

### 算法框架

由小规模无人机集群迁移到大规模多主体时遇到的最重要的问题是**维度爆炸**：联合搜索方法难以解决该问题，因为状态动作空间组合爆炸，需要不切实际的训练数据量来收敛。 **基于此难点，本文提出的基于模仿学习+强化学习的方案，能够充分利用基于冲突基搜索获得的粗粒度解，降低网络初步的搜索空间，从而解决大规模下网络训练复杂度过高的问题。**如图是本研究提出的算法框架。

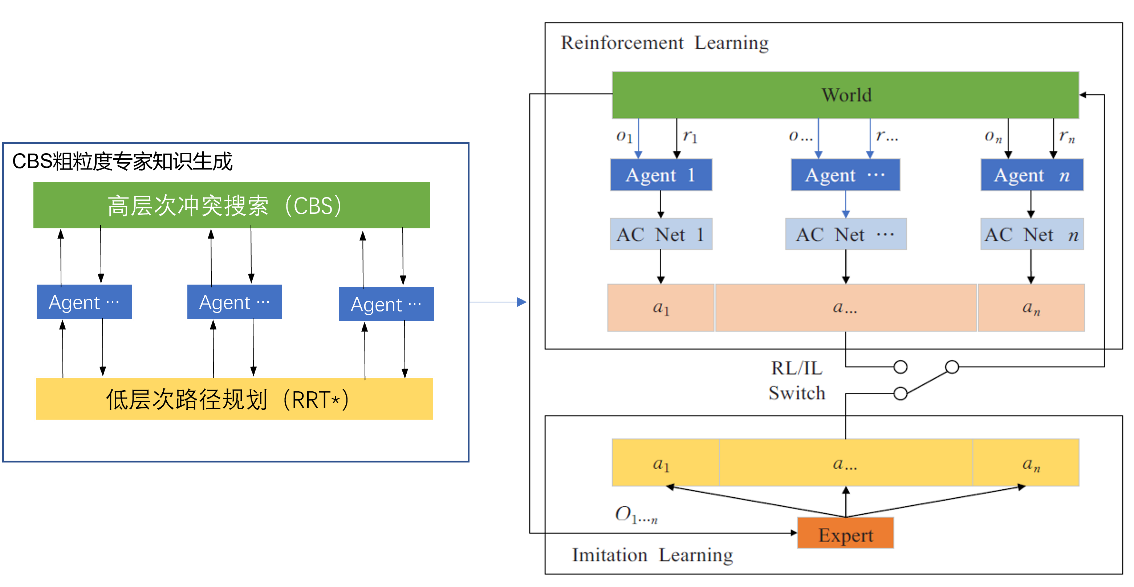


图 ‑ 基于CBS的模仿学习大规模协同路径规划框架

如图该框架由两部分组成：

**第一部分为基于CBS的多智能体路径规划结果**，能够快速搜索出符合无碰撞要求的规划解，作为专家知识输入到模仿学习的网络模型中。

**第二部分为模仿学习模型，**利用基于CBS的多智能体路径规划结果作为专家知识，指导网络训练方向，大幅度降低网络探索的难度，从而加快网络收敛速度，训练满足协同通信的路径规划解。

### 模型设计

和第二节不同，步长固定，输出节点沿x和y移动的距离，都为（-x,x）之间，先用CBS生成初步可行解，step是重要参数，决定一步能走多长，后续网络输出的决策范围也按照这个输出。（-x,x）应该比step大一些，这样绕路了可以赶上，不至于一直延误导致训练不稳定

网络模型主要有两种设计方案：

**①集中式**：state为所有节点的位置 + 下一时刻移动位置的目标（或者step的数量），action为所有节点在两个维度上的移动量，reward为：

1）每个节点根据当前step的目标位置和实际到达位置的距离差值,这种情况下撞墙的话会导致连锁反应，可能永远都无法找到正确的路，如果撞墙主要要随机探索自己找到路

2）撞墙以及撞机会导致对应的飞机回退到上一位置，这样的话需要更新get\_state()返回的状态，每个step应该维护一个数组指示当前决策后，每个无人机的下一个目的状态指针，撞机不更新，不撞机加1，然后get\_state根据这个返回对应的xy以及targetxy。

3）不连通则给于惩罚：软约束，不连通的话给于回退操作并且给于一个非常大的惩罚：硬约束，强制找到一个连通的决策才行继续。

这种集中式的方法容易导致没办法很好对地衡量单个决策的好坏。通过直接采样目标点为移动位置的概率（0.1-0.5之间），其他方向采样为网络输出加上噪声裁剪noise,方差设置为a\_max。

**②分布式：**每个节点训练一个ac：输入为当前节点的位置以及目标点，输出为当前节点的唯一决策。此外另加一个全局critic，输出为全局奖励的计算，用于判断当前决策对全局的影响，也可以选择不启用全局式的，每个节点把其他节点当自然环境，reward函数为自己的靠近奖励，碰撞奖励和全局的连通奖励。这样就可以直接把其他智能体的移动当环境了。

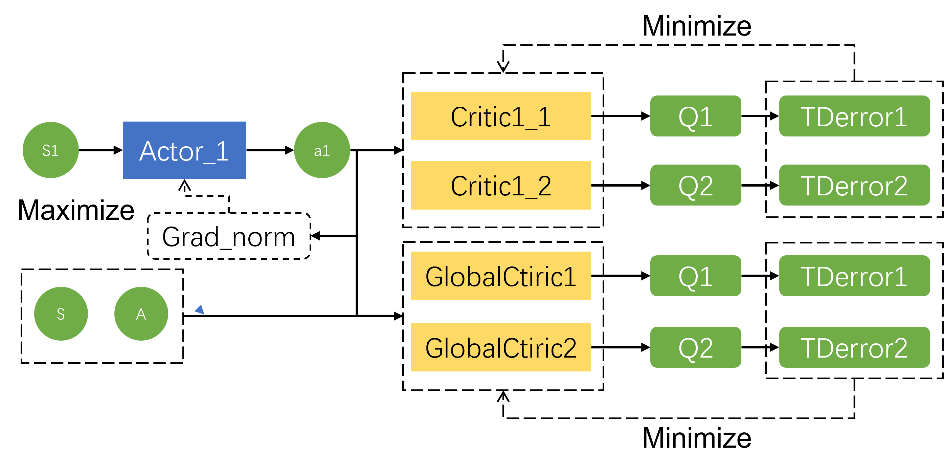


图 ‑2 COMATD3单智能体网络架构

根据以上分析，为了分解全局奖励和单体奖励，降低网络reward贡献分布不均的问题，对于强化学习的架构方面，采取如图基于MATD3的COMATD3网络架构，通过引入全局Critic以及单体critic分别估计单体Q值和全局Q值，再引入Grad\_Norm实现梯度归一化更新，平衡单体Q值和全局Q值的更新幅度。

### 算法细节

1）指向专家目标点的action计算放缩公式。

## 仿真结果

### 基于CBS-RRT\*的算法仿真结果

通过实现基于CBS-RRT\*算法能够进行粗粒度的路径规划结果，此路径规划算法能够有效规划无碰撞路径，但没有有效地实现节点之间的协同通信保持能力，因此作为粗粒度解作为网络初步学习经验。

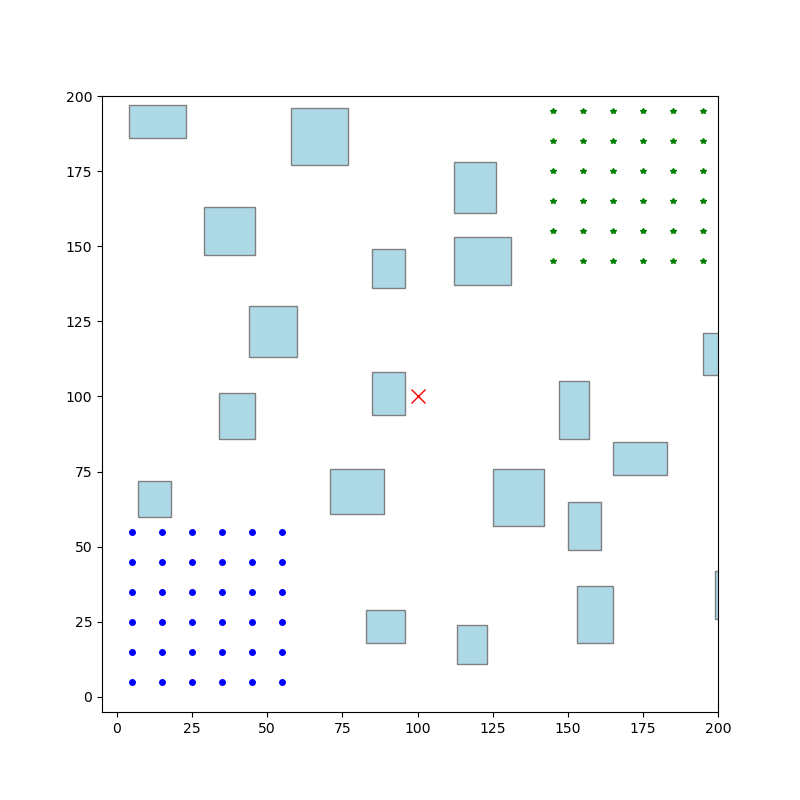


图 ‑3 36节点CBS-RRT\*路径规划效果图

### 基于IL-COTDMA3的算法仿真结果

目前训练效果很差，算法架构比较复杂，还在排查代码方面是否有bug。

# 研究方向与计划

主要是对后续进一步研究方向与计划的思考，分为5节点和36节点两方面进行考虑。

**①5节点场景：**

1)从**干扰信号**方面：调整干扰信号功率、干扰信号运动模型g

2)从**通信维护**方面：连通度从0-1改为连续值进行优化。

②36节点场景：

1)从**编队控制**方面：基于队形变换（formation control）做干扰，在建筑物密度大的情况下比较难。

2)从**建筑密度**方面：可以去除干扰信号，单独考虑建筑密度大的场景下无人机运动的通信保持

1. Yao, P.; Wang, H.; Su, Z. UAV feasible path planning based on disturbed fluid and trajectory propagation. Chin. J. Aeronaut. 2015, 28, 1163–1177.
2. Xia, Q.; Liu, S.; Guo, M.; Wang, H.; Zhou, Q.; Zhang, X. Multi-UAV trajectory planning using gradient-based sequence minimal optimization. *Robot. Auton. Syst.* **2021**,*137*, 103728.
3. Gao, S.; Ai, J.L.; Wang, Z.H. Mixed population RRT algorithm for UAV path planning. *Syst. Eng. Electron.* **2020**, *42*, 101–107.
4. Wang, T.; Huang, P.F.; Dong, G.Q. Cooperation path planning of multi-UAV in road-network continuous monitoring. *Acta Aeronaut. Astronaut. Sin.* **2020**, *41*, 61–70.
5. Zhang, H.H.; Gan, X.H.; Li, S.F.; Feng, Z.; Jin, Y. UAV route planning considering regional risk assessment under complex low altitude environment. *Chin. J. Sci. Instrum.* **2021**, *42*, 257–266.
6. Na, Z.; Zhang, M.C.; Low, K.H. 3D path planning and real-time collision resolution of multirotor drone operations in complex urban low-altitude airspace. *Transp. Res. Part C Emerg. Technol.* **2021**, *129*, 103–123.
7. Muñoz, J.; López, B.; Quevedo, F.; Monje, C.A.; Garrido, S.; Moreno, L.E. Multi UAV Coverage Path Planning in Urban Environments. *Sensors* **2021**, *21*, 7365.
8. Zhang, Q.R.; Wei, R.X.; He, R.K.; Ru, C.J.; Zhou, K. Path planning for unmanned aerial vehicle in urban space crowded with irregular obstacles. *Control Theory Appl.* **2015**, *32*, 1407–1413.
9. Khan, S.I.; Qadir, Z.; Munawar, H.S.; Nayak, S.R.; Budati, A.K.; Verma, K.; Prakash, D. UAVs path planning architecture for effective medical emergency response in future networks. *Phys. Commun.* **2021**, *47*, 101337.
10. Causa, F.; Fasano, G. Multiple UAVs trajectory generation and waypoint assignment in urban environment based on DOP maps. *Aerosp. Sci. Technol.* **2021**, *110*, 106507.