基于风险损失分析的多无人机巡检路径规划 研究

高 壮,何杏宇

上海理工大学光电信息与计算机工程学院,上海

收稿日期: 2024年4月23日; 录用日期: 2024年5月21日; 发布日期: 2024年5月29日

摘要

当前多无人机系统在执行城市巡检任务时面临巡检点的风险等级冲突性和风险损失动态性问题,为此, 本文提出了一种基于风险损失分析的多无人机巡检路径规划方法。首先,该方法定义了巡检点的风险等 级和损失计算方法;然后,该方法定义了一种巡检点聚类算法,根据巡检点之间的风险等级相异性和地 理位置相近性对巡检点进行聚类,将不同的聚类分配给不同的无人机,避免风险等级相同的巡检点的访 问冲突问题,并优化巡检路径长度;最后,该方法提出了一种基于强化学习的路径规划算法,根据实时 的风险损失和巡检开销来计算奖励函数,以引导无人机的实时巡检路径规划,兼顾了动态风险规避和巡 检效率。实验结果表明,本研究所提出的算法相比于现有算法具有更低的巡检风险损失以及较好的综合 性能表现。

关键词

多无人机系统,城市巡检,路径规划,风险损失,强化学习

Research on Multi-UAV Inspection Path Planning Based on Risk Loss Analysis

Zhuang Gao, Xingyu He

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 23rd, 2024; accepted: May 21st, 2024; published: May 29th, 2024

Abstract

In the execution of urban inspection tasks, multi-Unmanned Aerial Vehicle (UAV) systems are

currently faced with the issues of conflicting risk levels at inspection points and the dynamic nature of risk loss. To address these issues, this paper proposes a multi-UAV inspection path planning method based on risk loss analysis. Initially, this method defines the risk levels of inspection points and the approach for computing the associated losses. Subsequently, it introduces a clustering algorithm for inspection points that considers the heterogeneity of risk levels and the proximity of geographic locations, grouping inspection points into clusters. Different clusters are then allocated to different UAVs to prevent conflicts due to visiting inspection points with the same risk level and to optimize the length of the inspection path. Lastly, the method introduces a path planning algorithm based on reinforcement learning, which employs a reward function calculated from real-time risk loss and inspection costs to guide the UAVs' real-time inspection route determination, thus considering both dynamic risk avoidance and inspection efficiency. Experimental results show that the proposed algorithm outperforms existing methods, offering lower inspection risk loss and better overall performance.

Keywords

Multi-UAV Systems, Urban Inspection, Path Planning, Risk Loss, Reinforcement Learning

Copyright © 2024 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <u>http://creativecommons.org/licenses/by/4.0/</u> Open Access

1. 引言

随着无人机(Unmanned Aerial Vehicle, UAV)技术的迅速发展,其在民用和军事应用中的重要性日益 增加,包括环境监测、无线通信、安全搜救等领域[1]。其中,使用无人机进行城市巡检任务具有成本低 廉、操作简单、机动灵活、安全高效、视野良好等优点,而且无人机更容易与人工巡检相融合,成为了 日常巡检手段[2] [3]。在城市巡检任务中,如何提升无人机路径规划能力,已然成为学界和业界关注的热 点问题[4]。

在无人机巡检路径规划的现有研究中,研究者们提出的算法可以大致归纳为优化算法和智能算法两 大类,以解决效率优化[5]、避障处理[6]、任务调度[7]以及动态适应性[8]等关键问题。其中,优化算法依 据特定的优化策略,搜索最优或者次优路径,如遗传算法[9]、人工蜂群算法[10]和粒子群优化[11]等,这 些方法在处理大规模问题时可能会遇到局部最优和计算复杂性的问题。而智能算法,尤其是深度学习[12] 和强化学习[13] [14] [15],在路径优化中展示了独特优势,其能够通过与环境的交互学习实现优化路径选 择,从而满足巡检无人机对自主决策和适应性的需求。

然而,上述方法都没有涉及巡检点的风险等级冲突性和风险损失动态性的分析:一方面,当多个风 险等级较高(涉及大量的人员安全和经济价值损失)的巡检点同时具有较高的安全事件发生概率时,存在巡 检冲突问题;另一方面,巡检点的人员数量和经济价值随着时间变化(例如办公楼巡检点的人员数量会不 同时段会不同,仓库巡检点的货物储备不同时段也不同),因此具有风险损失动态性,这给实时巡检决策 的风险动态规避带来挑战。

为此,本文设计了一种基于风险损失分析的多无人机路径规划方法,旨在提升多无人机在巡检中的 协同作业效率和动态风险规避能力。具体的,首先,该方法定义了巡检点的风险等级和损失计算方法; 然后,该方法设计了一种巡检点聚类算法,根据巡检点之间的风险等级相异性和地理位置相近性对巡检 点进行聚类,将不同的聚类分配给不同的无人机,避免风险等级相同的巡检点的访问冲突问题,并优化 巡检路径长度;最后,该方法提出了一种基于强化学习的路径规划算法,根据实时的风险损失和巡检开 销来计算奖励函数,以引导无人机的实时巡检路径规划,兼顾了动态风险规避和巡检效率。

2. 无人机巡检系统模型

2.1. 问题描述及假设

本文考虑一种多无人机城市巡检的场景, 巡检范围为 Z, 在巡检范围内有 N 个位置固定的待巡检点, 所有巡检点的集合为 $O = \{O_1, \dots, O_i, \dots, O_N\}$, 其中, 巡检点 O_i 有以下属性: 位置坐标 loc_i 、人员数量 n_i 、经济价值 e_i 以及安全事件发生概率 p_i 。系统中有M台无人机,所有无人机的集合为 $U = \{U_1, \dots, U_m, \dots, U_M\}$ 。多台无人机从同一基地出发, 基地用 O_0 表示, 依次访问所有巡检点, 完成对每个巡检点实际情况的数据采集, 最终返回基地。

在该系统模型中,将一天划分为 *K* 个等长的时段,所有时段的集合为 $T = \{T_1, \dots, T_k, \dots, T_K\}$ 。在本文中,将无人机完成一次完整的巡检任务的总时长控制在一个时段内。在一个时段 T_k 内,巡检点 O_i 的人员数量和经济价值分别为该时段下的人员数量和经济价值的预测值,分别用 $n_i^{T_k}$ 和 $e_i^{T_k}$ 表示,安全事件发生概率为该时段下每个时刻的预测值,用 $p_i^{r_k}$ 表示,其中 $t \in \{t_0, t_1, \dots, t_{\Delta T}\}$ 。

为便于问题分析建模,本文针对研究问题做出以下合理假设:1)假设每台无人机始终在同一高度飞行与悬停,其高度由 H 表示。2)假设每台无人机在飞行过程中的飞行速度为常数 v₀。3)假设整个巡检过程中无人机能量充足,即不考虑无人机充电情况。

2.2. 问题定义

本文采用四维矩阵 $X = \{x_{m,i,j}^t | t \in \{t_0, t_1, \dots, t_{\Delta T}\}, m \in \{1, 2, \dots, M\}, i, j \in \{0, 1, \dots, N\}\}$ 记录无人机在巡检时 段 T_k 内的飞行路径,其中, $x_{m,i,j}^t$ 的值表示 t 时刻 U_m 是否从 O_i 飞到 O_j ,如果该巡检路径存在,则 $x_{m,i,j}^t = 1$, 否则 $x_{m,i,j}^t = 0$,表示为:

$$x_{m,i,j}^{t} = \begin{cases} 1, & t 时刻U_{m} \bigcup O_{i} \ (3) \\ 0, & 其他 \end{cases}$$
(1)

本文对系统中的相关风险损失值作出以下定义,用*V_i*表示*O_i*在巡检时段*T_k*内的风险等级,即巡检综合价值,其计算方式表示为:

$$V_i = \alpha_1 \hat{n}_i^{T_k} + \alpha_2 \hat{e}_i^{T_k} \tag{2}$$

其中, $\alpha_1 \pi \alpha_2$ 为权重参数, $\hat{n}_i^{T_k} \pi \hat{e}_i^{T_k} \beta$ 别为对 $n_i^{T_k} \pi e_i^{T_k}$ 归一化之后的值。

O,在t时刻的风险损失值用L,表示,其计算方式表示为:

$$L_i^t = V_i \times p_i^t \tag{3}$$

系统在 t 时刻的风险损失值用 L,表示,其定义为当前时刻系统中未被访问的巡检点风险损失值之和, 用 A,表示 t 时刻未被访问的巡检点的下标集合,L,的计算方式表示为:

$$L_t = \sum_{j \in A_t} L_j^t \tag{4}$$

系统的总风险损失值用 L_{total} 表示,其定义为系统在巡检过程中各时刻的风险损失值累计之和,其计算方式表示为:

$$L_{total} = \sum_{t=t_0}^{t_{\Delta T}} L_t \tag{5}$$

由于环境中各巡检点是分开的,因此采用 N + 1行 N + 1列矩阵 $D = \{D_{i,j} | i, j \in \{0,1,\dots,N\}\}$ 来记录每两 个点之间的距离。用 $T_{m,i}$ 表示无人机 U_m 从 O_i 飞到 O_i 的时间成本,其计算方式表示为:

$$T_{m,i,j} = \frac{D_{i,j}}{v_0}$$
(6)

用T(U_m)表示U_m飞往已分配的所有巡检点花费的时间,其计算方式表示为:

$$T(U_m) = \sum_{i=0}^{N} \sum_{j=0}^{N} T_{m,i,j}, m \in \{1, 2, \cdots, M\}$$
(7)

本文的研究目标是为无人机巡检规划出一条能够及时发现安全隐患以最小化系统的总风险损失值的 路径,并且在此过程中使得完成任务的总耗时尽可能少。因此,本文将系统总风险损失与巡检总时长的 加权和作为多无人机执行巡检任务的系统总成本,最小化系统总成本以实现研究目标,用C_{und}表示系统 总成本,其计算方式表示为:

$$C_{total} = \min\left(\delta_1 \hat{L}_{total} + \delta_2 \max_{1 \le m \le M} \hat{T}(U_m)\right)$$
(8)

其中, δ_1 和 δ_2 为权重参数, \hat{L}_{total} 和 $\hat{T}(U_m)$ 分别为对 L_{total} 和 $T(U_m)$ 归一化之后的值。 $\max_{k=n \in M} T(U_m)$ 表示巡检 的总时长,由于多台无人机同时进行巡检,使用系统中无人机完成任务耗时最大的时长作为巡检总时长。

巡检任务的条件约束相关公式描述如下:

1) 巡检点访问冲突约束。多台无人机应避免同时对同一巡检点进行访问,即对于任意给定的巡检点, 同一时刻只有一台无人机可以飞到该点并从该点起飞,表示为:

$$\forall i \in \{0, 1, \cdots, N\}, \sum_{m=1}^{M} \sum_{t=t_0}^{t_{AT}} \sum_{j=0}^{N} x_{m,i,j}^t = 1, \sum_{m=1}^{M} \sum_{t=t_0}^{t_{AT}} \sum_{j=0}^{N} x_{m,j,i}^t = 1$$
(9)

2) 无人机数量约束。选择执行巡检任务的无人机数量不大于无人机总数,其中无人机从同一基地起 飞,表示为:

$$\forall m \in \left\{1, 2, \cdots, M\right\}, \sum_{i=1}^{N} x_{m,0,i}^{t} \le M$$
(10)

3) 访问顺序约束。在任务执行的过程中,每台无人机都应该对巡检点按顺序逐一访问。用数组 $G = \{g_{mi} | m \in \{1, 2, \dots, M\}, i \in \{0, 1, \dots, N\}\}$ 来记录每个点的访问顺序,每个元素的值 g_{mi} 表示 U_m 访问 O_i 顺序 的索引值。由于无人机直接从同一基地起飞,每台无人机的索引顺序 O_0 为1,即 $\forall m \in \{1, 2, \dots, M\}, g_{m,0} = 1$ 。 对于其他巡检点,如果 U_m 从 O_i 飞到 O_i ,则 O_i 的访问顺序比 O_i 多一个,即 $g_{m,i} = g_{m,i} + 1$ 。由于分配给每台 无人机的巡检点数量都不大于总巡检点个数,因此每个巡检点的顺序索引值都小于或等于N+1,表示为:

$$\forall m \in \{1, 2, \cdots, M\}, \forall i \in \{1, 2, \cdots, N\}, g_{m,i} \le N+1$$

$$\tag{11}$$

本文所涉及到的主要变量符号及含义如表1所示。

表 1. 变量符号及含义	
变量符号	含义
0	巡检点集合
U	无人机集合

Table 1. Symbols and meanings

续表	
Т	时段集合
$n_i^{T_k}$	时段T _k 下巡检点i的人员数量预测值
$e_i^{T_k}$	时段T"下巡检点i的经济价值预测值
P_i^t	t时刻巡检点i的安全事件发生概率预测值
$x_{m,i,j}^t$	t 时刻无人机 m 是否从巡检点 i 飞到巡检点 j
V_i	巡检综合价值
L_i^t	t 时刻巡检点 i 的风险损失值
L_{t}	t 时刻系统的风险损失值
A_{r}	t 时刻系统中未被访问的巡检点的下标集合
L_{total}	系统的总风险损失值
$T_{m,i,j}$	无人机 m 从巡检点 i 飞到巡检点 j 的时间成本
$T(U_m)$	无人机 m 的巡检飞行耗时
C_{total}	系统总成本

3. 基于聚类和强化学习的多无人机路径规划算法

本文的多无人机路径规划算法分为两个阶段,分别为聚类划分阶段和路径规划阶段,算法的整体架 构如图 1 所示。首先,通过聚类算法解决多个高风险巡检点同时访问冲突的问题。然后,通过强化学习 算法解决无人机实时路径规划问题。通过本文算法的设计,实现降低多无人机执行巡检任务的系统总成 本的目的。



3.1. 聚类划分阶段

聚类划分阶段,为了解决多个高风险巡检点被同时访问冲突的问题,本文在设计聚类算法时考虑将

风险等级较高的巡检点划分在不同的群组中,实现同一群组内的各巡检点风险等级差异化。同时,为了降低巡检时间成本,需要均衡每台无人机访问群组的时间,本文在设计聚类算法时考虑将距离相近的巡检点划分在同一群组。

为了兼顾上述两个目标,本文设计了一个综合考虑巡检点之间的风险等级差异和距离相似的巡检相关性函数,基于该函数实现对巡检点的聚类划分。具体地,巡检点 *O_i*和 *O_j*之间的巡检相关性函数 *cor*(*i*, *j*) 表示为:

$$cor(i, j) = \omega_1 \widehat{dif}_{i,j} + \omega_2 \widehat{D}_{i,j}$$
(12)

其中, $\omega_1 和 \omega_2$ 为权重参数, $\widehat{dif}_{i,j}$ 和 $\widehat{D}_{i,j}$ 分别为 $dif_{i,j}$ 和 $D_{i,j}$ 归一化之后的值。具体来说, $dif_{i,j}$ 表示集群 聚类中心点 O_i 和非中心点 O_j 之间的风险等级的差异值, $D_{i,j}$ 表示聚类中心点 O_i 和非中心点 O_j 之间的距 离。 $dif_{i,j}$ 的计算方式表示为:

$$dif_{i,j} = \left| \overline{V}_i - V_j \right| \tag{13}$$

其中,将巡检点 O_i 的巡检综合价值 V_i 作为巡检点的风险等级值, \overline{V}_i 表示聚类中心 O_i 对应集群内所有巡检点综合价值的平均值。 $D_{i,i}$ 的计算公式表示为:

$$D_{i,j} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}$$
(14)

其中, (x_i, y_i) 和 (x_j, y_j) 分别为 O_i 和 O_j 的位置坐标。

本文的聚类算法计算流程如表2所示。

 Table 2. Clustering algorithm based on inspection correlation function

 表 2. 基于巡检相关性的聚类算法

输入 输出	: 聚类群组数 M,集合 O(包含 N个样本点) :聚类结果 clusters (包含 M 个聚类簇)
1:	在集合 O 中任意取 M 个样本作为初始的聚类中心点
2:	循环开始
3:	对每个 O 中的非聚类中心点 O_j ,进行:
4:	利用公式(12)计算相关性,将O _j 分配至相关性最相近的集群下
5:	随机选取一个非聚类中心点 0,
6:	利用公式(12)计算交换聚类中心点 O_i 和 O_j 之后集群的总成本 C_{new}
7:	如果 $C_{new} < C_{old}$:
8:	用 O _i 交换 O _i ,形成一个新的 M 个聚类中心点的集合
9:	循环结束(满足迭代结束要求)

10: 返回集群聚类结果 clusters

3.2. 路径规划阶段

经过聚类划分阶段后,将为每一个聚类群组分配一台无人机进行后续巡检任务,例如,为群组 cluster_m 分配无人机U_m进行后续巡检任务。本文根据环境中动态变化的风险损失值设计了强化学习算法模型的奖 励函数,为无人机规划每个时刻的巡检点访问顺序,实现及时发现安全隐患的效果,同时尽可能降低巡 检时长,以最大限度地降低系统总成本。

3.2.1. 无人机巡检路径规划强化学习模型

在本文的无人机巡检任务场景中,需要为每台无人机训练路径规划策略,因此无人机可以被视为强 化学习模型中的智能体。在无人机巡检的场景中,由于环境中的风险损失信息随时间变化,很难获取状 态转移概率矩阵,因此本文考虑采用无模型的强化学习。假设无人机 U_m 负责群组 cluster_m 的巡检任务, 群组 cluster_m 中巡检点的数量为 N_m ,满足 $N_m = |cluster_m|$,其中关于状态、动作和奖励函数的定义如下:

1) 状态 *s*_t: 无人机 *U*_m只知道其自身位置信息以及其被分配的群组 *cluster*_m 的信息,包括群组中各巡检点的位置、风险损失值以及访问状态。将这些环境中的动态信息设置在状态中,*t* 时刻的状态 *s*_t 定义为:

$$s_t = \left\{ p_t, d_t, l_t, f_t \right\} \tag{15}$$

其中, p_t 为无人机 t 时刻所在位置的向量, $p_t = (x_t, y_t, z_t)$, 其中无人机的高度为常数 H, 即 $z_t = H$; d_t 为无人机在 t 时刻与集群内各巡检点的距离向量,本文用无人机当前的位置坐标的 x 和 y 分量与巡检点之 间的欧氏距离作为无人机与巡检点的距离,即 $d_t = (d_1^t, \dots, d_i^t, \dots, d_{N_m}^t)$,其中 d_i^t 表示无人机在 t 时刻与巡 检点 O_i 之间的距离; l_t 为 t 时刻集群内每个巡检点的风险损失值的向量,即 $l_t = (l_1^t, \dots, l_N^t, \dots, l_{N_m}^t)$,其中 l_i^t 表示 t 时刻巡检点 O_i 的风险损失值; f_t 为 t 时刻集群内每个巡检点的访问状态的向量,用 0 表示未被访 问过,1 表示已被访问,即 $f_t = (f_1^t, \dots, f_{N_m}^t), f_i^t \in \{0,1\}$ 。

2) 动作 a_t: 无人机的动作为选择集群中的巡检点, 定义为:

$$a_t \in cluster_m$$
 (16)

对于时刻t,无人机根据状态 s_t 执行动作 a_t 表示无人机从当前位置以速度 v_0 向指定的目标巡检点直线飞行。

3) 奖励函数 r_i:为了最小化系统总成本,需要对无人机的动作进行价值评估。基于优化目标,本文 提出了即时奖励函数,表示为:

其中, r,表示t时刻无人机执行完动作之后的奖励, L,表示t时刻系统风险损失值, T,表示t时刻无人机 路径飞行的总耗时, δ₁和δ₂分别表示风险损失和时长的权重参数, R₀和R₁为正整数,分别用于对无人 机完成巡检任务给予奖励和对巡检点重复访问给予惩罚。通过以上奖励函数的设计,引导无人机快速完 成巡检任务,并且在此期间实现最小化系统总成本的目的。

3.2.2. AM-DQN 算法

由于本文设计的无人机巡检路径规划强化学习模型的动作空间为离散动作,并且是一种无模型的强 化学习。因此本文通过对 DQN 进行改进,提出了一种基于 AM-DQN (Action Masking Deep Q-learning Network)的无人机路径规划算法。算法的架构如图 2 所示。

为了使无人机尽快完成一次巡检任务,本文在强化学习模型的奖励函数中设计了对重复访问动作的 惩罚。此设计在一定程度上避免了无人机对刚刚访问巡检点的重复访问,但是在实际操作中,由于无人 机的动作空间为聚类群组中所有的巡检点,导致其在进行动作选择时不可避免地对已经访问的点进行选 择。因此本文将动作掩码机制加入到强化学习的动作选择中[16],过滤掉不可用的动作,避免无人机对环 境采用无效的动作。AM-DQN 通过对当前状态下网络 Q 值进行掩码操作,将无效动作的 Q 值修改为趋 于无穷小的值,在通过 argmax 进行动作选取时,这些被掩码屏蔽的动作将不会被选中。



Figure 2. AM-DQN algorithm architecture diagram 图 2. AM-DQN 算法架构图

算法包含两个神经网络,即估计值网络和目标值网络。用 θ_t^* 表示 t 时刻的目标值网络参数,则 t 时刻目标值 Q_t^* 的计算方式如下:

$$Q_{t}^{*} = r_{t} + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \theta_{t}^{*})$$
(18)

算法的动作选择采用 ε – greedy 策略,用 ε_{min} 表示最小探索因子, ε_{decay} 表示探索因子衰减速率, epoch 表示训练周期,其中探索因子衰减更新公式如下:

$$\varepsilon \leftarrow \varepsilon_{\min} + (\varepsilon - \varepsilon_{\min}) \times e^{-\varepsilon_{decay} \times epoch}$$
(19)

AMDQN 算法训练过程如表 3 所示。

Table 3. AM-DQN based path planning algorithm for UAV inspection 表 3. 基于 AM-DQN 的无人机巡检路径规划算法

输入:无人机和巡检点的状态信息 输出:动作序列、累计奖励和估计值网络 1: 初始化回放缓冲区 D 初始化估计值网络参数 θ ,目标值网络参数 θ^* 2: 初始化探索因子 ε ,最小探索因子 ε_{min} ,探索因子下降率 ε_{decav} ,批次大小B3: 循环每次训练周期: episode = 1, epoch 4: 5: 初始化状态 s, 循环每个时间步: $t = t_0, t_{\Lambda T}$ 6: 用概率 ε 选择一个随机动作 a_r 7: 否则,对当前网络的Q值进行掩码操作,选择Q值最大的动作a, 8: 9: 无人机执行动作 a, , 获得奖励 r, 和下一时刻状态 s,++

续表

- 10: 将 (s_t, a_t, r_t, s_{t+1}) 存储至D中
- 11: 从 D 中均匀随机采样 B 条样本数据 $(s_i, a_i, r_i, s_{i+1}), j \in |B|$
- 12: 根据式(18)计算目标值
- 13: 进行梯度下降计算
- 14: 更新估计值网络参数和目标值网络参数
- 15: 结束时间步循环
- 16: 根据式(19)更新探索因子ε
- 17: 结束训练周期循环

4. 实验与结果分析

4.1. 实验参数设置和目的

4.1.1. 实验参数设置

本文的模拟实验从一个真实的无人机巡检场景迁移,考虑一个 5 km × 5 km 的城市巡检范围,在该城市范围中随机分布若干个位置固定的巡检点,系统由一个无人机基地、多台无人机以及多个巡检点组成。 实验所用服务器的 CPU 为英特尔酷睿 i5-13490F,内存为 32GB,GPU 为 NVIDIA GeForce TRX 4060 Ti,利用 PyTorch 深度学习框架实现算法中的神经网络部分,利用 Python 语言实现强化学习环境部分的编写。详细的实验参数设置如表 4 所示。

名称	值
城市巡检范围 Z	25 平方公里
无人机飞行速度 v _o	10 米/秒
巡检点数量 N	[30,45,60,75,90]
学习率 α	0.002
回放缓冲区 D	2000
折扣因子γ	0.9
探索因子 ε	1
最小探索因子 ε_{\min}	0.01
探索因子下降率 ε_{decay}	0.0000002
批次大小 B	32

Table 4. Experimental parameters 表 4. 实验参数

4.1.2. 实验目的

本文的实验目的包括以下三个方面:

1) 为了验证本文在路径规划阶段所提出的 AM-DQN 算法具有更高的收敛性与学习效率,我们对所 提出的算法和传统的 DQN 算法[13]对无人机执行城市巡检任务的路径规划进行了训练。其中设定当前单 个无人机负责的巡检点数量为 10,模型参数 δ₁和 δ₂分别取 0.7 和 0.3,对于参数选取的实验过程下文将 给出详细过程。具体的,对于这两种算法,我们都进行了一万个训练周期,并且这两种算法都被执行了 十次训练,以便我们能获取奖励值的平均值作为最终效果指标。

2) 为了寻找模型最优参数δ₁和δ₂的值,平衡系统总风险损失和系统总时长之间的关系,以实现本文研究目标。本文设置了不同权重参数下算法的系统总风险损失值和系统总时长情况对比,在当前无人机 任务规模为10的场景下,对不同权重参数进行了实验数据对比并选取最优参数值,后续实验都是基于此 权重参数值进行。

3)为了验证本文所提算法的性能优越性,我们对所提出的算法和聚类 + DQN、MADQN [14]以及 MADDPG [15]对任务规模为 30、无人机数量为 3 的场景下执行任务的不同性能指标数据进行对比,其中 聚类 + DQN 算法中的聚类阶段使用本文所提出的聚类算法。我们在相同的测试环境下,对不同算法进 行了 100 次实验模拟,最终取其结果的平均值进行比较,从而更客观的展示算法的应用效果。

4.2. 结果分析

4.2.1. 强化学习算法收敛性分析

两种算法的奖励收敛情况如图 3 所示,由图 3 可知,所提出的 AM-DQN 算法的奖励值在开始训练后 维持在较高水平,随着训练周期的增加奖励值持续快速增长并最终收敛,而 DQN 算法的奖励值长时间维 持在较低水平,波动较大并且收敛速度较慢。上述比较说明了相比于 DQN 算法,本文所提算法的训练过 程更加稳定和高效。





4.2.2. 不同权重参数的指标数据分析

图 4 为不同权重下训练出的算法模型的系统总风险损失和系统总时长情况,其中模型参数 & 和 &,之

和为 1,将 δ_i 的值范围设置为[0-1],由图 4 可知,随着 δ_i 的值增加,系统总风险损失的占比增加,模型 学习时优先考虑降低系统总风险损失,在 δ_i 取值为 0.7 时,系统总风险损失值趋于稳定,与此同时系统 总时长持续增加,所以本文取 δ_i 为 0.7 作为模型参数。



图 4. 不同权重参数的指标数据

4.2.3. 不同算法的指标数据分析

由于本文算法分为两个阶段,我们首先对 30 个巡检点进行聚类划分为 3 组,然后将聚类群组分别分 配给每台无人机,每台无人机将运行各自的 AM-DQN 算法进行路径规划。聚类分配过程如图 5 所示,其 中巡检的颜色深浅表示巡检点风险等级的高低,颜色越深表示等级越高。



Figure 5. Clustering and allocation process 图 5. 聚类和分配过程

表 5 为四种算法在巡检点规模为 30 时执行任务的指标数据,由表 5 可以看出,本文算法能够保证任务的有效完成,并且拥有更低的系统总成本,在各项指标数据上都拥有较高的性能表现。

Table	5. Indicator	data for task	s performed	by the	four al	gorithms
表 5.	四种算法执	行任务的指	标数据			

指标	本文算法	聚类 + DQN	MADQN	MADDPG
任务完成率	100%	100%	98%	95%
平均系统总成本	55.02	57.24	61.46	63.42
平均系统总风险损失	68.4	71.7	75.5	76.2
平均系统总时长/分	23.8	23.5	28.7	33.6

最后,为了验证本文算法的可扩展性,我们对任务规模分别为[30,45,60,75,90]下的场景进行模拟实 验。图 6 为四种算法在不同巡检点数量下的系统总成本对比,由图 6 可知,随着巡检点数量的增加,四 种算法的系统总成本均在上升,本文算法在不同规模的任务场景下性能表现稳定。其中,在任务规模为 60 以上时,其他三种算法的性能出现了明显的下降,这是因为随着任务规模的增加,算法的动作空间变 大,导致其他三种算法难以完成给定的任务。通过综合分析,在多无人机巡检场景下,本文所提算法拥 有比对比算法更好的表现,是综合性应用的最佳选择。



Figure 6. Comparison of total system cost for four algorithms with different number of inspection points

图 6. 四种算法在不同巡检点数量下的系统总成本对比

5. 总结

本文研究了城市巡检场景下的多无人机路径规划问题,针对巡检场景进行建模,量化环境中的风险

损失,将路径规划的目标设计为最小化系统总风险损失与系统总时长的加权和。本文提出的基于风险损 失分析的多无人机巡检协同路径规划方法,通过聚类划分阶段和路径规划阶段,分别解决了巡检点风险 等级冲突性问题和无人机路径规划风险损失动态适应性问题。该方法兼顾了路径动态风险规避和巡检效 率,能够有效降低巡检的系统总成本。实验结果表明,本文所提出的方法能够有效解决多无人机巡检路 径规划问题,并且相比于现有算法,巡检规划出的路径具有更低的系统总风险损失,在巡检任务中具有 更优的综合性能表现。

基金项目

国家自然科学基金资助项目(61602305, 61802257);上海市自然科学基金资助项目(18ZR1426000, 19ZR1477600)。

参考文献

- Teixeira, K., Miguel, G., Silva, H.S. and Madeiro, F. (2023) A Survey on Applications of Unmanned Aerial Vehicles using Machine Learning. *IEEE Access*, 11, 117582-117621. <u>https://doi.org/10.1109/ACCESS.2023.3326101</u>
- [2] Raja, A., Njilla, L. and Yuan, J. (2022) Adversarial Attacks and Defenses toward AI-Assisted UAV Infrastructure Inspection. *IEEE Internet of Things Journal*, 9, 23379-23389. <u>https://doi.org/10.1109/JIOT.2022.3206276</u>
- [3] Li, Z., Wu, H., Wang, Q., Wang, W., Suzuki, S. and Namiki, A. (2024) Small UAV Urban Overhead Transmission Line Autonomous Correction Inspection System Based on Radar and RGB Camera. *IEEE Sensors Journal*, 24, 5593-5608. <u>https://doi.org/10.1109/JSEN.2023.3317076</u>
- [4] Wu, W., Funabora, Y., Doki, S., Doki, K., Yoshikawa, S., Mitsuda, T. and Xiang, J. (2024) Evaluation and Enhancement of Resolution-Aware Coverage Path Planning Method for Surface Inspection Using Unmanned Aerial Vehicles. *IEEE Access*, 12, 16753-16766. <u>https://doi.org/10.1109/ACCESS.2024.3359056</u>
- [5] Shen, K., Shivgan, R., Medina, J., Dong, Z. and Rojas-Cessa, R. (2022) Multidepot Drone Path Planning with Collision Avoidance. *IEEE Internet of Things Journal*, 9, 16297-16307. https://doi.org/10.1109/JIOT.2022.3151791
- [6] Wang, C., Yang, X. and Li, H. (2022) Improved Q-Learning Applied to Dynamic Obstacle Avoidance and Path Planning. *IEEE Access*, 10, 92879-92888. <u>https://doi.org/10.1109/ACCESS.2022.3203072</u>
- [7] Bayerlein, H., Theile, M., Caccamo, M. and Gesbert, D. (2021) Multi-UAV Path Planning for Wireless Data Harvesting with Deep Reinforcement Learning. *IEEE Open Journal of the Communications Society*, 2, 1171-1187. https://doi.org/10.1109/OJCOMS.2021.3081996
- [8] Yan, C., Xiang, X. and Wang, C. (2020) Towards Real-Time Path Planning through Deep Reinforcement Learning for a UAV in Dynamic Environments. *Journal of Intelligent & Robotic Systems*, 98, 297-309. https://doi.org/10.1007/s10846-019-01073-3
- [9] 李艳, 郭继峰, 罗汝斌. 基于遗传算法与 Dubins 理论的高速无人系统在多障碍环境中的路径规划[J]. 无人系统 技术, 2021, 4(6): 37-45.
- [10] Liu, H.T., Ge, J.Y., Wang, Y., Li, J.C., Ding, K., Zhang, Z.Q., Guo, Z.H., Li, W. and Lan, J.H. (2022) Multi-UAV Optimal Mission Assignment and Path Planning for Disaster Rescue Using Adaptive Genetic Algorithm and Improved Artificial Bee Colony Method. *Actuators*, **11**, Article 4. https://doi.org/10.3390/act11010004
- [11] Phung, M.D. and Ha, Q.P. (2021) Safety-Enhanced UAV Path Planning with Spherical Vector-Based Particle Swarm Optimization. Applied Soft Computing, 107, Article ID: 107376. <u>https://doi.org/10.1016/j.asoc.2021.107376</u>
- [12] Liu, Y., Wang, H., Fan, J., Wu, J. and Wu, T. (2021) Control-Oriented UAV Highly Feasible Trajectory Planning: A Deep Learning Method. *Aerospace Science and Technology*, **110**, Article ID: 106435. https://doi.org/10.1016/j.ast.2020.106435
- [13] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. and Riedmiller, M. (2013) Playing Atari with Deep Reinforcement Learning. arXiv preprint arXiv:1312.5602.
- [14] Hu, W., Yu, Yu, Liu, S., She, C., Guo, L., Vucetic, B. and Li, Y. (2023) Multi-UAV Coverage Path Planning: A Distributed Online Cooperation Method. *IEEE Transactions on Vehicular Technology*, 72, 11727-11740. <u>https://doi.org/10.1109/TVT.2023.3266817</u>
- [15] Zheng, S. and Liu, H. (2019) Improved Multi-Agent Deep Deterministic Policy Gradient for Path Planning-Based

Crowd Simulation. IEEE Access, 7, 147755-147770. https://doi.org/10.1109/ACCESS.2019.2946659

[16] Huang, S. and Ontañón, S. (2020) A Closer Look at Invalid Action Masking in Policy Gradient Algorithms. arXiv preprint arXiv:2006.14171.