

控制与决策

Control and Decision

考虑多换电站的多无人机应急电力巡检路径规划方法

秦文龙, 罗贺, 李晓多, 阴酉龙, 林世忠, 王国强

引用本文:

秦文龙, 罗贺, 李晓多, 等. 考虑多换电站的多无人机应急电力巡检路径规划方法[J]. *控制与决策*, 2025, 40(8): 2391-2399.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2024.0795>

您可能感兴趣的其他文章

Articles you may be interested in

[双重驱动的果蝇优化算法及其在PID控制器中的应用](#)

Double drive fruit fly optimization algorithm and its application in PID controller
控制与决策. 2021, 36(9): 2225-2233 <https://doi.org/10.13195/j.kzyjc.2020.0046>

[面向多目标侦察任务的无人机航线规划](#)

UAV trajectory planning for multi-target reconnaissance missions
控制与决策. 2021, 36(5): 1191-1198 <https://doi.org/10.13195/j.kzyjc.2019.1284>

[基于16方向24邻域改进蚁群算法的移动机器人路径规划](#)

Mobile robots path planning based on 16-directions 24-neighborhoods improved ant colony algorithm
控制与决策. 2021, 36(5): 1137-1146 <https://doi.org/10.13195/j.kzyjc.2019.0600>

[分布式无人机的时变编队非线性控制设计](#)

Time-varying formation nonlinear control of distributed multiple UAVs
控制与决策. 2021, 36(10): 2490-2496 <https://doi.org/10.13195/j.kzyjc.2020.0136>

[车辆与无人机组合配送研究综述](#)

[Review on vehicle-UAV combined delivery problem](#)

控制与决策. 2021, 36(10): 2313-2327 <https://doi.org/10.13195/j.kzyjc.2020.1315>

考虑多换电站的多无人机应急电力巡检路径规划方法

秦文龙^{1,4}, 罗贺^{1,4,5}, 李晓多^{1,4}, 阴酉龙², 林世忠³, 王国强^{1,4,5†}

- (1. 合肥工业大学 管理学院, 合肥 230009;
2. 国网安徽省电力有限公司 无人机巡检作业管理中心, 合肥 230061;
3. 安徽送变电工程有限公司, 合肥 230071; 4. 过程优化与智能决策教育部重点实验室, 合肥 230009;
5. 安徽省空天系统智能管理工程研究中心, 合肥 230009)

摘要: 多无人机应急电力巡检的时间十分有限, 在选择关键巡检目标时需要考虑各目标的故障概率差异, 同时为提升巡检效率, 可以引进换电站降低无人机续航能力不足的影响. 针对上述特点, 将考虑多换电站的多无人机应急电力巡检路径规划问题建模为多站点多航次团队定向问题, 并设计一种融合软演员-评论家模型的遗传算法 (SAC-GA). 首先, 在遗传算法中加入两类局部搜索算子, 以优化多无人机访问目标的选择和缩短无人机飞行路径距离. 其次, 提出一种基于 SAC 模型的参数调优机制, 利用 SAC 模型基于最大熵学习策略的优势, 在遗传算法迭代过程中, 根据历史学习经验和种群的状态动态生成合适的交叉、变异概率以及染色体再插入中的权距比. 实验结果表明, 算法在小规模实验和大规模实验上均具有明显优势, 并通过消融实验验证 SAC-GA 中局部搜索算子的有效性和参数调整方法的优越性. 最后, 通过案例分析验证算法在不同应急场景下的有效性.

关键词: 团队定向问题; 多无人机; 路径规划; 遗传算法; SAC 模型

中图分类号: C93 文献标志码: A

DOI: 10.13195/j.kzyjc.2024.0795

引用格式: 秦文龙, 罗贺, 李晓多, 等. 考虑多换电站的多无人机应急电力巡检路径规划方法 [J]. 控制与决策, 2025, 40(8): 2391-2399.

Multi-UAV emergency power inspection path planning method considering multiple charging stations

QIN Wen-long^{1,4}, LUO He^{1,4,5}, LI Xiao-duo^{1,4}, YIN You-long², LIN Shi-zhong³, WANG Guo-qiang^{1,4,5†}

- (1. School of Management, Hefei University of Technology, Hefei 230009, China; 2. Management Center of UAV Inspection Operation, State Grid Anhui Electric Power Co., Ltd., Hefei 230061, China; 3. Anhui Power Transmission & Transformation Engineering Co., Ltd., Hefei 230071, China; 4. Key Laboratory of Process Optimization & Intelligent Decision-making, Hefei 230009, China; 5. Engineering Research Center for Intelligent Management of Aerospace System, Hefei 230009, China)

Abstract: Limited time for multi-UAV emergency power inspection requires prioritizing targets based on fault probabilities. In order to improve the inspection efficiency, multi-charging stations can be introduced to reduce the impact of insufficient endurance of UAVs. The problem is formulated as a multi-depot multi-visit team orienteering problem and addressed using a genetic algorithm with a soft actor-critic (SAC) model. The algorithm first incorporates two types of local search operators into the evolution process of the traditional genetic algorithm to optimize the selection of multiple UAVs visiting targets and to reduce the flight path distance of the UAVs. Then, a method for dynamically adjusting the parameters of the genetic algorithm using reinforcement learning is proposed. By using the SAC model based on maximum entropy policy learning, during the iteration of the genetic algorithm, dynamically adjusts crossover, mutation rates, and weight distance ratios in chromosome reinsertion based on past learning and population state. Experiments show the algorithm's effectiveness in small and large-scale tests, with ablation

收稿日期: 2024-07-04; 录用日期: 2024-12-18.

基金项目: 国家自然科学基金面上项目 (72271076, 71971075, 71871079); 安徽省自然科学基金项目 (2308085QG 233).

责任编辑: 李登峰.

†通信作者. E-mail: gqwang2017@hfut.edu.cn.

本文附带电子附录文件, 可登录本刊官网该文“资源附件”区自行下载阅览.

experiments validating the local search operator's effectiveness and the superiority of the parameter tuning method. Finally, the algorithm's efficacy in various emergency scenarios is validated via simulations.

Keywords: team orienteering problem; multi-UAV; path planning; genetic algorithm; soft actor-critic

0 引言

无人机巡检以其高效率、低成本、实时性、灵活性等优势,正在逐步取代人工巡检成为电力巡检的主流方式^[1].在自然灾害等突发事件发生时,输电杆塔等设施可能受到严重损害,需要及时检查电力设施的受损情况并制定相应的抢修和恢复计划.在这种应急场景下,无人机的巡检时间往往是有限的^[2],每个巡检目标发生故障的概率也会存在差异^[3].区别于日常巡检中需要完成所有目标的检查^[4],应急巡检中无人机往往无法在有限时间内访问所有目标,使用多无人机可以在一定程度上提高应急路径规划的灵活性和任务执行效率.针对无人机应急巡检的时间限制、巡检目标之间的差异,如何对巡检区域中巡检目标的选择和访问顺序进行优化是多无人机应急电力巡检路径规划问题的一个重要挑战.

上述问题可以视作团队定向问题(TOP)的一类变体问题,均在路径规划中考虑了时间限制与目标之间的权重差异.TOP最初由Chao等^[5]于1996年提出,多无人机应急巡检路径规划问题与TOP的主要差异是无人机续航性能与车辆相比是不足的,规划的路径必须满足无人机续航能力约束.在现有针对无人机TOP问题的研究中,无人机出发和返回的站点是唯一的^[2,6-7],在应急场景下单站点会限制无人机任务执行范围和路径规划的灵活性.目前为止未发现在应急场景下对多站点无人机TOP问题的研究,部分学者研究了多站点TOP问题^[8-11],这些研究中,站点仅作为车辆的起点或终点.本文针对无人机应急巡检的问题特征,假设所有站点均可以为无人机提供更换电池服务来降低无人机续航时间对巡检效率的影响,研究考虑多换电站的多无人机应急巡检路径规划问题.

在以TOP为基础的路径规划问题中,启发式算法凭借其高效性和实用性,已成为当前广泛应用的求解算法^[12],文献[10-11,13]分别将蚁群优化、遗传算法和粒子群优化运用到TOP类问题的求解当中.近年来,基于强化学习设计启发式算法的参数更新机制受到了密切关注,这种方法能够进一步提高启发式算法的适应性和灵活性.文献[14-15]分别将强化学习应用于遗传算法中交叉算子的选择和交叉与变异概率的确定,取得了良好的效果.Wang等^[16]在武器目标分配问题的算法设计上,利用深度Q学

习网络(DQN)为人工蜂群算法寻找合适的变异算子;Chen等^[17]针对卫星任务规划问题,利用近端策略优化算法(PPO)更新动作选择策略确定进化算法(EA)交叉、突变等参数;Yu等^[18]提出了一种模拟退火强化学习算法解决带时间窗的集合团队定向问题,使用 ϵ -greedy算法确定模拟退火每次迭代中邻域移动的类型.上述研究中,强化学习的基本要素需根据所使用的启发式算法和问题特征进行设置,难以适用于本文研究的问题.

针对考虑多换电站的多无人机应急电力巡检路径规划问题,本文首先将其建模为多站点多航次团队定向问题(MD-MV-TOP),其次提出一种基于SAC(soft actor-critic)模型的遗传算法,最后在实验中对分析本文算法的有效性.

1 问题描述与建模

1.1 问题描述

针对电力设备突发故障,暴风雪、地震等自然灾害后电力故障预警排查等应急电力巡检场景,需要无人机在短时间内获取重要电力设备的图像信息.多架无人机从站点出发,根据应急任务的需求访问若干待巡检目标,每个目标有不同的收益值,收益值越大代表这个目标所代表的设备部位损坏概率越大;在执行任务过程中,当电量不足时,无人机可以选择返回任意站点更换电池,然后以电量充足的状态再次出发执行巡检任务,并在到达应急巡检时间之前返回任意站点.本文旨在设计一组无人机巡检路径,在应急巡检时间内巡检足够多且重要的目标信息,即获得最大收益.

图1展示了该问题的一个实例,在接到应急任务后,两架无人机各自从站点 D_1 、 D_2 出发,以无人机2为例,它从 D_1 起飞,首先巡检了目标6和目标11,当电量不足时,返回 D_1 更换电池,然后再次起飞并依次巡检目标1和目标2,最后在到达应急巡检时

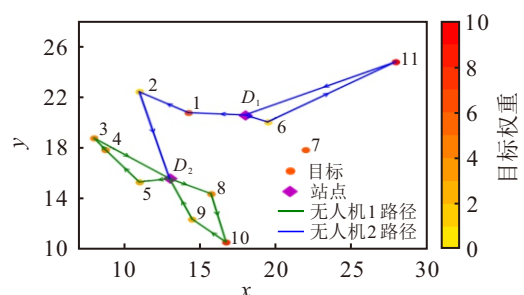


图1 问题实例

间之前返回 D_2 .

1.2 符号说明

模型中所用符号如表1所示.

表1 符号定义

类型	符号	含义
	K	无人机集合
	O	目标集合
	D	站点集合
	S	无人机的航次集合
	N	目标数量
	k	无人机编号
	s	无人机的航次编号
参数	d	站点编号
	h, i, j	目标编号
	r_{ij}	目标 i 与目标 j 之间的路径长度
	r_{dj}	站点 d 与目标 j 之间的路径长度
	r_{id}	目标 i 与站点 d 之间的路径长度
	v	无人机的飞行速度
	e	无人机的最大续航时间
	e_{\max}	应急巡检时间
	w_i	目标 i 的收益值
	决策变量	u_{iks}
x_{ij}^{ks}		无人机 k 第 s 航次从目标 i 飞到目标 j 则为1, 否则为0
y_{dj}^{ks}		无人机 k 第 s 航次从站点 d 飞到目标 i 则为1, 否则为0
	z_{id}^{ks}	无人机 k 第 s 航次从目标 i 飞到站点 d 则为1, 否则为0

1.3 数学模型

在多站点 TOP 问题的基础上, 将其建模为多站点多航次 TOP 问题, 具体如下:

$$\text{Max} \sum_{k \in K} \sum_{s \in S} \sum_{i \in O} \left(\sum_{j \in O} w_i x_{ij}^{ks} + \sum_{d \in D} w_i z_{id}^{ks} \right). \quad (1)$$

s.t. :

$$\sum_{j \in O} \sum_{d \in D} y_{dj}^{ks} = \sum_{i \in O} \sum_{d \in D} z_{id}^{ks} = 1, \forall s \in S, k \in K; \quad (2)$$

$$\sum_{k \in K} \sum_{s \in S} \left(\sum_{d \in D} y_{di}^{ks} + \sum_{h \in O} x_{hi}^{ks} \right) = \sum_{k \in K} \sum_{s \in S} \left(\sum_{j \in O} x_{ij}^{ks} + \sum_{d \in D} z_{id}^{ks} \right) \leq 1, \forall i \in O; \quad (3)$$

$$\left(\sum_{d \in D} y_{di}^{ks} + \sum_{h \in O} x_{hi}^{ks} \right) = \left(\sum_{j \in O} x_{ij}^{ks} + \sum_{d \in D} z_{id}^{ks} \right) \leq 1, \forall i \in O, s \in S, k \in K; \quad (4)$$

$$\sum_{s \in S} \left(\sum_{d \in D} \sum_{j \in O} y_{dj}^{ks} r_{dj} + \sum_{i \in O} \sum_{j \in O} x_{ij}^{ks} r_{ij} + \sum_{d \in D} \sum_{j \in O} z_{jd}^{ks} r_{jd} \right) \leq v e_{\max}, \forall k \in K; \quad (5)$$

$$u_{iks} - u_{jks} + N x_{ij}^{ks} \leq N - 1, \quad 1 < i, j \leq N, i \neq j, \forall k \in K, s \in S; \quad (6)$$

$$\sum_{d \in D} \sum_{j \in O} y_{dj}^{ks} r_{dj} + \sum_{i \in O} \sum_{j \in O} x_{ij}^{ks} r_{ij} + \sum_{d \in D} \sum_{j \in O} z_{jd}^{ks} r_{jd} \leq v e_u, \forall k \in K, s \in S; \quad (7)$$

$$\sum_{i \in O} y_{di}^{ks} = \sum_{i \in O} z_{id}^{ks+1}, \quad \forall k \in K, s, s+1 \in S, d \in D; \quad (8)$$

$$u_{iks} \geq 0, x_{ij}^{ks}, y_{di}^{ks}, z_{id}^{ks} \in \{0, 1\}, \quad \forall i, j \in O, i \neq j, k \in K, s \in S, d \in D. \quad (9)$$

其中: 目标函数 (1) 表示最大化在指定应急巡检时间内无人机检查目标所获得的总收益; 约束 (2) 保证每个无人机每个航次都必须从一个站点出发, 并最终返回到某个站点; 约束 (3) 保证每个目标的入度等于出度, 并且每个目标最多被访问一次; 约束 (4) 保证在任意一个无人机的任意航次中, 每个目标的入度等于出度, 并且每个目标最多被访问一次; 约束 (5) 为应急巡检时间约束, 每架无人机总飞行时间不能超过规定的应急巡检时间, 巡检时间长短和应急场景类型和紧急程度相关; 约束 (6) 为消除子回路约束; 约束 (7) 为无人机的最大续航时间约束; 约束 (8) 为站点次序约束, 无人机降落在某一个站点进行更换电池后, 下一个航次必须从该站点起飞; 约束 (9) 给出了决策变量的取值范围.

2 求解算法

2.1 算法框架

为求解 MD-MV-TOP 问题, 本文提出一种基于 SAC 模型的遗传算法 (SAC-GA), 算法总体框架如图 2 所示.

2.2 遗传算法

2.2.1 染色体编码方式

染色体采用多层编码方式, 如图 3 所示. 染色体被编码为 $|K|$ 层, $|K|$ 为无人机的数量, 每一层代表一架无人机的路径, 红色的数字 $d \in D$ 代表站点, 黑色的数字 $i \in T$ 代表目标.

2.2.2 初始化种群

首先, 从站点集合 D 中选取不同的起始站点生成 $|D|$ 个站点组合, 并依次使用 3 类贪婪策略生成站点间的目标直至达到应急巡检时间限制: 1) 距离最近; 2) 目标奖励值最大; 3) R 最大, R 为未访问的目标收益与当前位置到该目标的距离的比值. 最后, 随机生成种群中剩余的个体, 其中总访问目标数量占原始目标集合 60% ~ 80%.

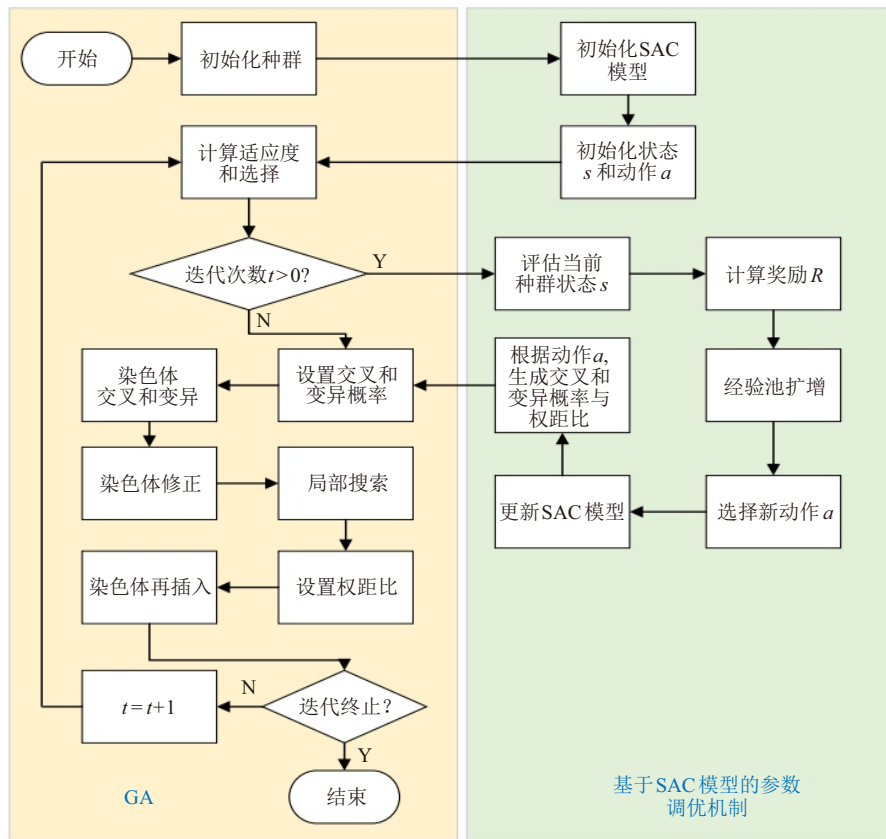


图2 SAC-GA 算法框架

UAV1	1	3	2	2	8	4	1
UAV2	2	7	6	1	5	1	2
	↓		↓				
	站点		目标				

图3 染色体编码方式

2.2.3 计算适应度和选择

染色体适应度评估函数如下:

$$f = \sum_{i=1}^A w_i - e_{sum} + |K|e_{max}, \quad (10)$$

$$e_{sum} = \sum_{k \in K} \sum_{s \in S} \left(\sum_{d \in D} \sum_{j \in O} y_{dj}^{ks} r_{dj} + \sum_{i \in T} \sum_{j \in O} x_{ij}^{ks} r_{ij} + \sum_{d \in D} \sum_{j \in O} z_{jd}^{ks} r_{jd} \right) / v. \quad (11)$$

其中: A 为所有无人机已访问的目标点数量之和, e_{sum} 为染色体对应的无人机飞行时间之和.

选择具体步骤如下: 首先, 使用精英保留策略, 保存原种群适应度值前 20% 精英个体和去除后 20% 的不良个体; 然后, 在剩余的 60% 个体中按照轮盘赌的方法进行选择, 由此完成新种群的生成.

2.2.4 染色体交叉和变异

本文采用单点交叉和 POX 交叉^[19], 每次随机选择出两个染色体上某个无人机的一个航次进行交叉, 站点不参与交叉过程.

采用两点变异和站点变异两种方法优化染色体序列. 其中, 两点变异是在每个无人机的航线中随机更改一个目标, 站点变异是随机更改一个无人机的某个站点编号.

2.2.5 染色体修正

染色体在进行遗传操作后, 可能会变成非可行解. 一方面, 染色体可能会出现重复的目标节点, 此时删除重复目标点; 另一方面, 染色体上无人机的某一个航次飞行时间会超出最大续航时间或者无人机总飞行时间超过应急巡检时间, 此时删除路径中插入性价比 R_{insert} 最低的目标或者随机删除一个目标点直至满足时间要求. R_{insert} 的计算方法为

$$R_{insert} = \delta w_i - c_d. \quad (12)$$

其中: δ 为权距比参数, 用来平衡目标收益 w_i 与加入到现有路径中所需要的距离成本 c_d .

2.2.6 局部搜索

降低无人机访问节点子集所需的旅行成本是优化 TOP 类问题的重要手段^[20], 在 MD-MV-TOP 中, 主要涉及多无人机间的目标分配和无人机目标访问顺序, 合理的目标分配也是提高无人机执行应急任务效率的关键^[21]. 因此, 本文设计两种局部搜索算子分别优化染色体中多无人机目标分配和访问顺序.

1) TAP(task assignment policy) 算子. 首先, 使用

0-1 邻域算子将一个无人机的目标点插入到另一个无人机路径中来减少总飞行时间;其次,使用 1-1

邻域算子交换两个无人机路径中一个目标点来减少总飞行时间. TAP 算子的过程如图 4 所示.

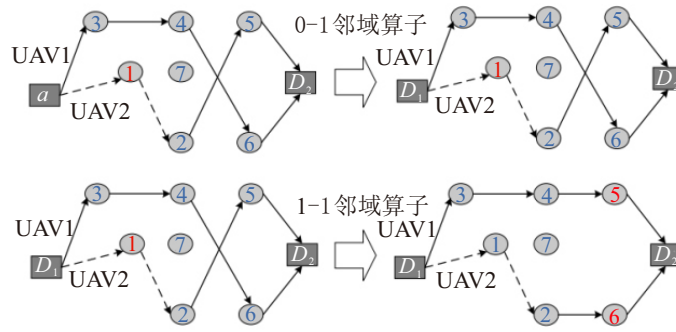


图4 TAP 算子

2) TVO(task visit order) 算子. 首先, 提取出每架无人机访问的目标点以及出发和结束访问的站点;其次, 依据两个目标点间距离从小到大对目标点间的组合进行排序;然后, 基于贪婪方式根据排序结果生成目标点执行顺序, 再通过 2-opt^[22] 优化当前拍摄目标点执行顺序;最后, 基于总访问顺序和续航时间为无人机寻找合适的中间停靠站点.

2.2.7 染色体再插入

在染色体进行修正和局部搜索后, 无人机的每个航次的飞行时间可能会下降, 为在有限的应急巡检时间内获得更大的收益, 需要考虑巡检更多目标的可能性. 本文引入一种插入算法^[20], 即考虑在每个无人机航次中, 在不超过续航时间前提下, 选择插入性价比 R_{insert} 最大的目标点插入.

2.3 基于 SAC 模型的参数调优机制

2.3.1 马尔可夫决策过程基本要素

在 SAC-GA 算法中定义马尔可夫决策过程(MDP)的基本要素: 状态集 S 、动作集 A 、奖励函数 R , 具体内容如下.

1) 状态集 S .

在遗传算法中, 种群平均适应度的提升和最优值的提升是非常关键的, 为平衡 GA 在探索过程中多样性与收敛性之间的关系, 设计 3 个状态 $s_1, s_2,$

$s_3, s = [s_1, s_2, s_3], s \in S$.

s_1 : 种群的多样性. 由种群迭代第 t 次时种群之间的距离评估染色体在解空间中分布的均匀程度, 通过计算初始种群多样性与 s_1 的比值来归一化, 即

$$s_1 = \frac{\max(f(x_u^1)) - \min(f(x_u^1))}{\max(f(x_u^t)) - \min(f(x_u^t))}. \quad (13)$$

其中: $f(x_u^1), f(x_u^t)$ 分别表示第 1 次和第 t 次种群迭代时第 u 个个体的适应度值, $\max(f(x_u^t))$ 表示第 t 次种群迭代时最优染色体的适应度值, $\min(f(x_u^t))$ 表示第 t 次种群迭代时最差染色体的适应度值.

s_2 : 种群的总体水平. 由种群迭代第 t 次时种群适应度均值评估种群总体进化水平, 通过计算初始种群总体水平与 s_2 的比值来归一化, 即

$$s_2 = \frac{\sum_{u=1}^M f(x_u^1)}{\sum_{u=1}^M f(x_u^t)}, \quad (14)$$

其中 M 为种群规模.

s_3 : 种群的近期提升. 由种群迭代第 t 时刻前 10 代的平均种群适应度均值和最优个体适应度提升之和评估种群在某段时间内迭代效果, 通过计算其与总目标收益 R_{max} 与初始种群最优值的差值的比值来归一化, 即

$$s_3 = \frac{0.5 \left\{ \frac{\sum_{u=1}^M f(x_u^t) - \sum_{u=1}^M f(x_u^{t-9})}{M} \right\} + 0.5 \{ \max(f(x_u^t)) - \max(f(x_u^{t-9})) \}}{R_{max} - \max(f(x_u^1))}. \quad (15)$$

2) 动作集 A .

通过选择交叉概率 P_{cross} 和变异概率 P_{muta} 控制种群进化的策略, 并通过调控权距比 δ 的大小改变未访问目标加入染色体的概率, 定义动作 a_1, a_2, a_3 来依次控制 $P_{cross}, P_{muta}, \delta$ 的选择, 动作空间为 $a_1 \in [0.5,$

$0.8], a_2 \in [0.1, 0.4], a_3 \in [0.5, 2], a = [a_1, a_2, a_3], a \in A$.

3) 奖励函数 R .

本文分别基于第 1 次和第 $t-1$ 次种群迭代时染色体最优值与平均适应度的变化, 设计奖励 r_1 和

r_2 来评估第 t 代动作选择的合理性,奖励 R 通过对两类奖励进行加权求和得到,归一化后表达式为

$$r_1 = \frac{\max(f(x_u^t)) - \max(f(x_u^{t-1}))}{\max(f(x_u^{t-1}))}, \quad (16)$$

$$r_2 = \frac{\sum_{u=1}^M f(x_u^t) - \sum_{u=1}^M f(x_u^{t-1})}{\sum_{u=1}^M f(x_u^{t-1})}, \quad (17)$$

$$R = 0.5r_1 + 0.5r_2. \quad (18)$$

2.3.2 参数调优

本文将 SAC^[23]模型嵌入到遗传算法进化环境中,利用 Actor 网络生成的动作调整种群每次迭代所需的进化参数,使用 Critic 网络根据种群提升所获得的奖励去评估每次动作的合理性,根据环境变化和历史样本更新控制种群进化的动作策略.基于 SAC 的遗传算法参数动态调整机制的伪代码如算法 1 所示.

算法 1 基于 SAC 的遗传算法参数动态调整机制

输入: 种群状态 s_t 、动作 a_t 、种群适应度信息;

输出: 新的种群状态 s_{t+1} 、新的动作 a_{t+1} .

- 1) if 迭代次数 $t = 0$
- 2) 初始化 SAC 模型
- 3) 初始化种群状态 s_t 、动作 a_t
- 4) else
- 5) 评估种群当前状态 s_{t+1}
- 6) 计算 $R(s_t, a_t)$
- 7) 经验池扩增 $P \leftarrow P \cup \{s_t, a_t, R(s_t, a_t), s_{t+1}\}$
- 8) 选择新动作 $a_{t+1} \sim \pi_\varphi(a_{t+1} | s_{t+1})$
- 9) 执行动作 a_{t+1} , 生成新的交叉、变异概率和权距比
- 10) $S_t \leftarrow S_{t+1}, a_t \leftarrow a_{t+1}$
- 11) if 经验池 P 达到指定容量 vol
- 12) 从经验池 P 采样 batch 个样本
- 14) 更新 Critic-V 参数 $\varphi \leftarrow \varphi - \lambda_V \widehat{\nabla}_\varphi J_V(\varphi)$
- 15) 更新 Q 网络参数 $\theta_i \leftarrow \theta_i - \lambda_Q \widehat{\nabla}_{\theta_i} J_Q(\theta_i)$ for $i \in \{1, 2\}$
- 16) 更新 Actor 策略网络参数 $\phi \leftarrow \phi - \lambda_\pi \widehat{\nabla}_\phi J_\pi(\phi)$
- 17) 更新策略 π^* 熵的奖励系数 $\alpha \leftarrow \alpha - \lambda_H \widehat{\nabla}_\alpha J_H(\alpha)$
- 18) 更新 Target-V 网络参数 $\bar{\varphi} \leftarrow \tau\varphi + (1 - \tau)\bar{\varphi}$
- 20) end if
- 21) end if

在遗传算法开始迭代前,初始化 SAC 模型并计算种群的初始状态 s_t 和初始动作 a_t , 种群按照动作 a_t 中的交叉概率 P_{cross} 、变异概率 P_{muta} 和权距比 δ 执行第一轮迭代. 在此之后的每一轮迭代中, 计算上一

轮迭代执行动作 a_t 所获得的奖励 $R(s_t, a_t)$ 和种群的新的状态 s_{t+1} , 并储存样本 $\{s_t, a_t, R(s_t, a_t), s_t\}$ 到经验池 P 中, 再由 Actor 网络从动作空间中采样新的动作 a_{t+1} , 种群根据新动作 a_{t+1} 中的参数进行下一轮进化. 在经验池 P 达到指定容量后, SAC 模型进行 Critic-V 网络、 Q 网络的更新, 并基于最大熵学习策略^[24]更新 Actor 网络参数, 通过控制熵的奖励系数 α 使最优策略逐步收敛, 这种策略旨在增强从动作中提取的进化参数的多样性, 使种群在每个状态下所选择的进化参数拥有更多可能性, 以更好地适应遗传算法的随机性探索. 最后, 根据新的 Critic-V 网络参数 φ 柔性更新 Target-V 网络参数 $\bar{\varphi}$.

3 实验结果对比和分析

为测试 SAC-GA 算法性能, 在测试数据集上分别进行小规模实验、大规模实验和消融实验. 最后, 通过案例分析进一步验证本文算法在不同应急场景下的有效性.

3.1 实验数据集设计

基于 Chao 等^[5]提出的经典 TOP 数据集改造得到 48 个 MD-MV-TOP 的标准实例. 数据集实例编号的规则为“TOP 数据集-目标点数量-无人机数量-最大巡检时长”, 如 p1-30-2-40 表示在经典 TOP 数据集 Set1 上改造, 具有 30 个目标点, 2 架无人机, 应急巡检时间限制为 40, 无人机的续航时间固定为 25.

3.2 小规模实验

针对小规模数据集, 将本文算法 SAC-GA 与 CPLEX 求解器的计算结果和时间进行对比. 在 16 个实例中, SAC-GA 均能求解, CPLEX 有 5 个实例无法在设定时间 3 h 内求解. 在 CPLEX 能求解的 11 个实例中, SAC-GA 可以求得 9 个最优解, 其他两个算例求解差距在 1% 以内, 并且随着问题规模增大, SAC-GA 在求解时间上的优势逐渐扩大, 说明在小规模数据集下 SAC-GA 具有良好的求解性能.

3.3 大规模实验

针对大规模数据集, 为验证本文算法 SAC-GA 的优越性, 选用经典遗传算法 (GA)^[25]、蚁群优化算法 (ACOa)^[10] 作为对比算法进行分析, 评价指标为

$$\Delta\text{Best} = \frac{f_{\text{SAC-GA}}^{\text{best}} - f_{\text{Best}}^{\text{best}}}{f_{\text{Best}}^{\text{best}}}, \quad (19)$$

$$\Delta\text{Ave} = \frac{f_{\text{SAC-GA}}^{\text{ave}} - f_{\text{Best}}^{\text{ave}}}{f_{\text{Best}}^{\text{ave}}}. \quad (20)$$

其中: ΔBest 表示本文算法与 3 种算法中表现最好的求解最优值之间的差距, ΔAve 表示本文算法与 3 种算法中表现最好的求解平均值之间的差距,

f_{SAC-GA}^{best} 、 f_{SAC-GA}^{ave} 分别表示 SAC-GA 算法运行 10 次所得目标函数的最优值和平均值, f_{Best}^{best} 、 f_{Best}^{ave} 分别表示

3 种算法中运行 10 次表现最好的所得目标函数的最优值和平均值. 部分实验结果如表 2 所示.

表2 大规模数据集部分实验结果

数据集	GA		SAC-GA		ACoa		求解差距	
	最优值	平均值	最优值	平均值	最优值	平均值	$\Delta Best / \%$	$\Delta Ave / \%$
p5-64-2-25	390	366	440	409.5	390	355.5	0.00	0.00
p5-64-2-40	805	606.5	850	702.5	810	720.5	0.00	-2.56
p5-64-2-45	870	654.5	935	720	890	705.5	0.00	0.00
p5-64-2-50	840	634.5	980	880.5	902	727.4	0.00	0.00
p6-62-2-25	528	509.4	588	575.6	510	498.5	0.00	0.00
p6-62-2-40	792	751.2	834	812.2	698	675.8	0.00	0.00
p6-62-2-45	862	854.6	978	970.6	798	768.9	0.00	0.00
p6-62-2-50	966	960.6	1026	978.8	895	875.5	0.00	0.00
p7-100-3-25	656	623	725	710.9	624	605.5	0.00	0.00
p7-100-3-40	852	805.3	997	979.5	812	802.3	0.00	0.00
p7-100-3-45	935	912.1	1068	1026.3	897	812.1	0.00	0.00
p7-100-3-50	1053	1033	1133	1071.5	1003	973.9	0.00	0.00

由表 2 可见, 随着目标点规模的增加, SAC-GA 与 GA 和 ACoa 的求解差距逐渐扩大. 例如在数据集 p7 上, SAC-GA 较 GA 和 ACoa 在实例 p7-100-2-45 上最优值均提升 30.62%, 平均值分别提升 22.45% 和 25.76%. 在所有 32 个实例中, SAC-GA 的求解最优值和平均值均优于 GA, 仅在两个实例中, SAC-GA 的求解平均值略微落后于 ACoa, 说明 SAC-GA 在解决 MD-MV-TOP 问题上更具优势.

算法 SAC-GA 与算法 GA 在实例 p5-64-2-50 上获得最优解的目标函数值的迭代曲线如图 5 所示.

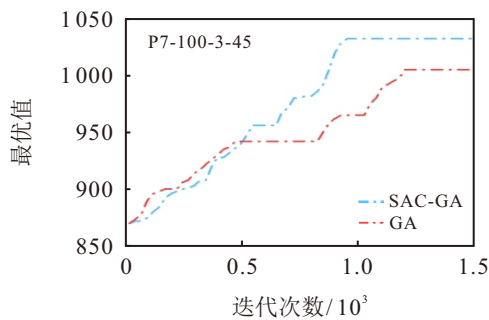


图5 算法迭代曲线对比

从图 5 中可以看出, 场景中, 算法的收敛值差距较小. 在 p7-100-3-45 场景下, SAC-GA 早于 GA 约 250 代完成了收敛, 并且 GA 明显在 500 代时陷入局部最优, 在迭代后期才跳出局部最优找到更优解. 迭代前期, SAC-GA 的收敛速度较差于 GA, 这是由于 SAC 模型回放池中样本数量不足以训练, 作为动作的进化参数输出较为随机, 在收集到足够多的训练样本后, SAC 模型的神经网络的参数训练效果更加优异, 能够进一步指导种群以更加合适的进化参数

进行迭代, 加快了遗传算法的收敛速度.

3.4 消融实验

本节分别分析局部搜索算子 TAP、TVO 和基于 SAC 模型的参数动态调整机制对 SAC-GA 算法的贡献程度. 以未添加改进策略的原始遗传算法 (SGA) 为基准, 将添加局部搜索算子 TAP 的遗传算法 (SGA + TAP)、同时添加局部搜索算子 TAP 和 TVO 以及基于贪婪策略的局部调优算子的遗传算法 (SGA+TAP+TVO)、同时添加两个算子及基于 SAC 模型的参数动态调整机制的本文算法 (SAC-GA) 进行比较. 将最大求解时间限制为 3 min, 分析 4 种算法的求解质量, 其中未加入强化学习调参机制的算法参数仍采用 Taguchi 正交实验的方法设定.

分别使用上述 4 种方法对所有大规模数据集进行测试, 统计算法运行 10 次最优目标函数值平均值, 图 6 是算法在数据集 p5 和 p6 上的结果.

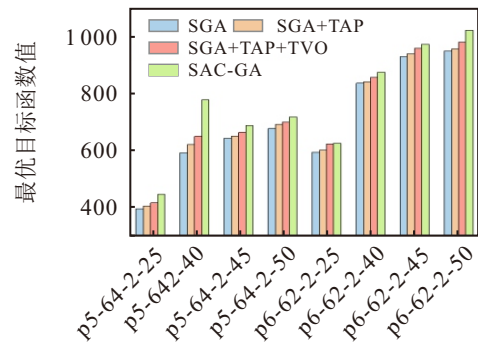


图6 部分消融实验结果

由实验结果可知, 在求解时间相同的情况下, 在数据集 p5 和 p6 中, 局部搜索算子 TAP、TVO 每次的加入都会提升 GA 的求解质量, 验证了局部搜索

策略对于促进遗传算法迭代的有效性;在所有数据集中, SAC-GA 算法均优于 GA + TAP + TVO 算法,说明强化学习动态调参机制均优于 Taguchi 正交实验参数设置方法。

3.5 案例分析

为验证本文算法 SAC-GA 在不同应急场景下的有效性,在安徽省一段输电线路选择 3 座电力杆塔模拟多无人机应急电力巡检,巡检目标为 20 个杆塔上不同类型的零部件。

分别模拟电力杆塔突发故障和暴风雪灾后电力故障预警排查两种应急场景,使用两架无人机执行巡检任务,无人机的续航时间设定为 10 min. 传统巡检方式需要两架无人机遍历所有目标,完成时间约为 25 min,难以满足在应急场景下的巡检需求,本文算法的模拟实验结果如表 3 所示。

表3 模拟实验结果

应急场景	紧急程度	应急时间 / min	目标权重和	巡检收益
突发故障	高	8	40	25
	中	15	40	31
雪灾后 预警排查	高	10	51	30
	中	20	51	40

由表 3 可以看出,本文方法相较于常规巡检策略,能够根据应急巡检时间灵活地生成高效的巡检策略。

4 结论

本文研究了考虑多换电站的多无人机应急电力巡检路径规划问题,建立了以传统 TOP 问题为基础的 MD-MV-TOP 数学模型,并提出了一种基于 SAC 模型的遗传算法. 基于经典 TOP 问题数据集构建了该问题的测试数据集,实验结果表明,本文所提出的 SAC-GA 算法具有良好的求解性能,算法中设计的局部搜索 TAP、TVO 算子以及基于强化学习的参数调优方法对于该问题的求解均可以带来提升. 通过引入强化学习方法,改变了传统依据规则进行参数调优的方式,能够在遗传算法框架和局部搜索算子的基础上实现对算法参数的动态设置,提升了算法的稳定性并加快收敛速度. 本文还进行了应急电力巡检的案例分析,进一步表明,本文方法能为应急电力巡检工作提供有效的巡检方案. 未来的研究工作可以进一步优化算法中 SAC 模型的内部架构,还可以针对 MD-MV-TOP 问题设计精确求解算法。

参考文献 (References)

[1] 赵蒙, 张博, 胡祥培, 等. 电力巡检车载无人机协同作

业优化[J]. *系统管理学报*, 2022, 31(6): 1098-1108.

(Zhao M, Zhang B, Hu X P, et al. Cooperative operation optimization for transmission inspection with vehicle-mounted UAV[J]. *Journal of Systems & Management*, 2022, 31(6): 1098-1108.)

[2] Fuertes D, Del-Blanco C R, Jaureguizar F, et al. Solving routing problems for multiple cooperative unmanned aerial vehicles using transformer networks[J]. *Engineering Applications of Artificial Intelligence*, 2023, 122: 106085.

[3] 杨月, 孙博, 马晓忱, 等. 融合注意力机制的输电线路故障概率预测模型[J]. *电力建设*, 2022, 43(3): 42-49. (Yang Y, Sun B, Ma X C, et al. Prediction model of transmission line fault probability applying attention mechanism[J]. *Electric Power Construction*, 2022, 43(3): 42-49.)

[4] Huang Z, Zhai X F, Wang H X, et al. On the 3D track planning for electric power inspection based on the improved ant colony optimization and a algorithm[J]. *Mathematical Problems in Engineering*, 2020, 2020: 8295362.

[5] Chao I, Golden B, Wasil E. The team orienteering problem[J]. *European Journal of Operational Research*, 1996, 88(3): 464-474.

[6] Wang K, Wu Q Q, He X T, et al. Optimizing UAV traffic monitoring routes during rush hours considering spatiotemporal variation of monitoring demand[J]. *International Journal of Geographical Information Science*, 2022, 36(10): 2086-2111.

[7] Panadero J, Juan A A, Bayliss C, et al. Maximising reward from a team of surveillance drones: A simheuristic approach to the stochastic team orienteering problem[J]. *European J of Industrial Engineering*, 2020, 14(4): 1-28.

[8] Sankaran P, McConky K, Sudit M, et al. GAMMA: Graph attention model for multiple agents to solve team orienteering problem with multiple depots[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(11): 9412-9423.

[9] Lee D H, Ahn J. Multi-start team orienteering problem for UAS mission re-planning with data-efficient deep reinforcement learning[J]. *Applied Intelligence*, 2024, 54(6): 4467-4489.

[10] Baffo I, Carotenuto P, Rondine S. An orienteering-based approach to manage emergency situation[J]. *Transportation Research Procedia*, 2017, 22: 297-304.

[11] Tlili T, Masri H, Krichen S. Towards an efficient collection and transport of COVID-19 diagnostic specimens using genetic-based algorithms[J]. *Applied Soft Computing*, 2022, 116: 108264.

[12] Gunawan A, Lau H C, Vansteenwegen P. Orienteering problem: A survey of recent variants, solution approaches and applications[J]. *European Journal of Operational Research*, 2016, 255(2): 315-332.

- [13] Dang D C, Guibadj R N, Moukrim A. An effective PSO-inspired algorithm for the team orienteering problem[J]. *European Journal of Operational Research*, 2013, 229(2): 332-344.
- [14] 朱光宇, 张德颂. 基于强化学习的遗传算法求解一种新的钻削路径优化问题[J]. *控制与决策*, 2024, 39(2): 697-704.
(Zhu G Y, Zhang D S. Genetic algorithm based on reinforcement learning for a novel drilling path optimization problem[J]. *Control and Decision*, 2024, 39(2): 697-704.)
- [15] Chen R H, Yang B, Li S, et al. A self-learning genetic algorithm based on reinforcement learning for flexible job-shop scheduling problem[J]. *Computers & Industrial Engineering*, 2020, 149: 106778.
- [16] Wang T, Fu L, Wei Z, et al. Unmanned ground weapon target assignment based on deep Q-learning network with an improved multi-objective artificial bee colony algorithm[J]. *Engineering Applications of Artificial Intelligence*, 2023, 117: 105612.
- [17] Chen Y X, Shen X, Zhang G, et al. Multi-objective multi-satellite imaging mission planning algorithm for regional mapping based on deep reinforcement learning[J]. *Remote Sensing*, 2023, 15(16): 3932.
- [18] Yu V F, Salsabila N Y, Lin S W, et al. Simulated annealing with reinforcement learning for the set team orienteering problem with time windows[J]. *Expert Systems with Applications*, 2024, 238: 121996.
- [19] 李佳磊, 顾幸生. 双种群混合遗传算法求解具有预防性维护的分布式柔性作业车间调度问题[J]. *控制与决策*, 2023, 38(2): 475-482.
(Li J L, Gu X S. Two-population hybrid genetic algorithm for distributed flexible job-shop scheduling problem with preventive maintenance[J]. *Control and Decision*, 2023, 38(2): 475-482.)
- [20] Vansteenwegen P, Gunawan A. Orienteering problems: Models and algorithms for vehicle routing problems with profits[M]. Cham: Springer International Publishing AG, 2019: 58-60.
- [21] 赵畅, 刘允刚, 陈琳, 等. 面向元启发式算法的多无人机路径规划现状与展望[J]. *控制与决策*, 2022, 37(5): 1102-1115.
(Zhao C, Liu Y G, Chen L, et al. Research and development trend of multi-UAV path planning based on metaheuristic algorithm[J]. *Control and Decision*, 2022, 37(5): 1102-1115.)
- [22] Fallahi A, Amani Bani E, Niaki S T A. A constrained multi-item EOQ inventory model for reusable items: Reinforcement learning-based differential evolution and particle swarm optimization[J]. *Expert Systems with Applications*, 2022, 207: 118018.
- [23] Chen S T, Qiu X H, Tan X Y, et al. A model-based hybrid soft actor-critic deep reinforcement learning algorithm for optimal ventilator settings[J]. *Information Sciences*, 2022, 611: 47-64.
- [24] Zheng J H, Kurt M N, Wang X D. Stochastic integrated actor-critic for deep reinforcement learning[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(5): 6654-6666.
- [25] Pop P, Matei O, Pinteau C. A two-level diploid genetic based algorithm for solving the family traveling salesman problem[C]. *Proceedings of the Genetic and Evolutionary Computation Conference*. New York, 2018: 340-346.

作者简介

秦文龙 (2000-), 男, 硕士生, 主要研究方向为无人机路径规划及强化学习, E-mail: 2390662128@qq.com;

罗贺 (1982-), 男, 教授, 博士, 主要研究方向为多主体协同优化与智能决策, E-mail: luohe@hfut.edu.cn;

李晓多 (1996-), 女, 博士生, 主要研究方向为多无人机路径规划, E-mail: lixiaoduo@mail.hfut.edu.cn;

阴酉龙 (1989-), 男, 工程师, 博士生, 主要研究方向为无人机智能巡检技术, E-mail: sgccyinyoulong@foxmail.com;

林世忠 (1982-), 男, 高级工程师, 博士生, 主要研究方向为无人机巡检作业管理, E-mail: shizhong306@126.com;

王国强 (1982-), 男, 副教授, 博士, 主要研究方向为多平台协同优策, E-mail: gqwang2017@hfut.edu.cn.