

文章编号:1671-1637(2026)04-0015-18

基于多智能体强化学习的多无人机协同 交通监控路径优化

刘晓波^{1,2}, 轩辕婧祎^{1,2}, 谢沅志^{1,2}, 郑芳芳^{*1,2}

(1. 西南交通大学 交通运输与物流学院, 四川 成都 611756;

2. 西南交通大学 综合交通大数据应用技术国家工程实验室, 四川 成都 611756)

摘要:为优化含电池更换站约束的多无人机协同交通监控路径规划问题,构建了基于无人机团队定向问题的混合整数线性规划模型,采用聚类方法确定电池更换站选址,使其分布均匀;提出了基于多智能体Transformer的强化学习(MTRL)算法框架,采用集中式Transformer架构,编码器利用多头注意力机制学习场景图结构全局表示,解码器生成协作式路径决策;设计了基于目标点访问数量的奖励函数,优化无人机访问序列与电池更换策略;引入结构化掩码机制,消除子回路、重复访问及路径冲突,保证解的可行性;在不同目标点数量、电池更换站数量和无人机数量的9类规模场景下开展数值试验。试验结果表明:MTRL在全部9类场景中均能获得高质量可行解,训练收敛稳定;与商业求解器相比,小、中规模场景下平均累计奖励提升9.77%~28.77%,大规模场景下提升9.34%~14.84%,而遗传算法、禁忌搜索在大规模场景下较商业求解器下降28%~41%;推理时间保持毫秒级;在18组跨分布泛化试验中,相对误差均控制在1%以内。所提框架可为无人机集群任务规划、智能交通路径优化及物流配送调度提供高效求解方案,并为多智能体强化学习在复杂约束优化问题中的应用提供方法参考。

关键词:低空交通;无人机;路径优化;深度强化学习;团队定向问题;交通监控

中图分类号:U8 **文献标志码:**A **DOI:**10.19818/j.cnki.1671-1637.2026.161

Cooperative traffic monitoring path optimization for multiple unmanned aerial vehicles based on multi-agent reinforcement learning

LIU Xiao-bo^{1,2}, XUANYUAN Jing-yi^{1,2}, XIE Yuan-zhi^{1,2}, ZHENG Fang-fang^{*1,2}

(1. School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, Sichuan, China; 2. National Engineering Laboratory for Applied Technology of Big Data in Integrated Transportation, Southwest Jiaotong University, Chengdu 611756, Sichuan, China)

Abstract: To optimize the multiple unmanned aerial vehicles (multi-UAVs) cooperative traffic monitoring path planning with battery replacement station constraints, a mixed-integer linear programming model based on the UAV team orienteering problem was constructed, and a clustering

出版历程:2025-09-04 收稿,2025-12-03 修回,2026-01-23 录用

基金项目:国家自然科学基金重点项目(52232011);四川省科技计划项目(2025YFHZ0193,2025HJJP0011)

作者简介:刘晓波(1974-),男,四川成都人,教授,工学博士,E-mail:xiaobo.liu@swjtu.cn。

*通信作者:郑芳芳(1981-),女,浙江浦江人,教授,工学博士,E-mail:fzheng@swjtu.cn。

引用格式:刘晓波,轩辕婧祎,谢沅志,等.基于多智能体强化学习的多无人机协同交通监控路径优化[J].交通运输工程学报,2026,26(4):15-32.

Citation: LIU Xiao-bo, XUANYUAN Jing-yi, XIE Yuan-zhi, et al. Cooperative traffic monitoring path optimization for multiple unmanned aerial vehicles based on multi-agent reinforcement learning[J]. Journal of Traffic and Transportation Engineering, 2026, 26(4): 15-32.

method was adopted to determine the battery replacement stations' locations to achieve uniform distribution. A multi-agent Transformer-based reinforcement learning (MTRL) algorithm framework was proposed, in which a centralized Transformer architecture was adopted. The encoder was used to learn the global graph-structured representation of the scenario via multi-head attention mechanism, and the decoder was used to generate collaborative path planning. A reward function based on the number of visited target nodes was designed to optimize the UAV visiting sequence and battery replacement strategy. A structured masking mechanism was introduced to eliminate subcircuits, repeated visits, and path conflicts, ensuring solution feasibility. Numerical experiments were conducted in scenarios of 9 types of scale with varying numbers of target nodes, battery replacement stations, and UAVs. The results show that MTRL obtains high-quality feasible solutions in all 9 types of scenarios with stable training convergence. Compared with the commercial solver, the average cumulative reward increases by 9.77%-28.77% in small- and medium-scale scenarios and by 9.34%-14.84% in large-scale scenarios, while that of the genetic algorithm and tabu search decreases by 28%-41% in large-scale scenarios. The inference time remains at the millisecond level. In 18 groups of cross-distribution generalization experiments, the relative error is controlled within 1%. The proposed framework provides an efficient solution for UAV swarm mission planning, intelligent transportation path optimization, and logistics distribution scheduling. In addition, it offers a methodological reference for the application of multi-agent reinforcement learning to complex constrained optimization problems.

Keywords: low-altitude traffic; unmanned aerial vehicle; path optimization; deep reinforcement learning; team orienteering problem; traffic monitoring

Publication history: Received 2025-09-04; Received in revised form 2025-12-03; Accepted 2026-01-23

Funding: Key Program of National Natural Science Foundation of China (52232011); Science and Technology Planning Project of Sichuan Province (2025YFHZ0193, 2025HJPJ0011)

* **Corresponding author:** ZHENG Fang-fang, professor, PhD, E-mail: fzheng@swjtu.cn.

0 引 言

随着中国经济的快速发展,汽车保有量持续攀升,交通拥堵问题日益突出,严重制约了城市出行效率并降低了出行体验^[1]。已有研究表明,通过精准监测交通流的动态变化并实施相应的管控措施,是缓解交通拥堵的有效途径^[2]。目前,交通监控技术通常分为固定式与移动式2类。固定式方法包括地磁线圈感应器^[3]、摄像头^[4]和路侧雷达^[5],但各自均存在局限性:如线圈感应器无法识别变道行为,而摄像头和路侧雷达的数据覆盖范围和精度又易受安装位置和角度的影响。移动式方法如浮动车^[6]、探测车^[7]和移动传感器^[8],主要采集样本车辆数据,无法全面反映交通流状况。受制于成本,这类手段通常应用于小规模监控,导致学术界与产业界长期面临监控覆盖不足与视线遮挡频繁等问题^[9-12]。

近年来,无人驾驶航空器(Unmanned Aerial Vehicle, UAV)技术的快速发展为低空交通监控提

供了新的可能^[13]。无人机能够提供实时的空中视角,有效缓解施工、行人及其他车辆造成的遮挡问题^[14]。与固定式路侧设备(如摄像头和雷达)相比,无人机具备灵活部署至目标区域的优势,且成本相对较低,支持在大范围交通网络中大规模应用,从而实现更广域的动态监控^[15-17]。

交通监控任务通常要求在给定时间范围内对关键路段进行周期性巡检,以实时掌握交通流状态并及时发现拥堵、事故等异常情况。然而,在大型交通网络中,需要同时监控数量庞大的目标点。在实际应用场景中,交通监控目标点通常基于历史交通数据和城市规划预先确定,主要包括城市主干道交叉口、快速路出入口以及事故多发路段等关键位置。无人机通过机载传感器采集交通流参数和异常事件信息,并通过无线网络实时回传至交通管理中心。获取的监控信息可用于交通流分析、事故检测、信号控制优化等应用,从而支持交通管理部门及时采取疏导措施,预防或缓解交通拥堵。受限于电池

续航,单架无人机难以完成所有监控任务^[18]。同时,为保证监控的连续性和完整性,以便及时采取有效的管控措施,往往需要在给定的时间范围内完成集中监控。时间限制使得网络内目标点难以实现完全覆盖。因此,在有限续航条件下尽可能扩大覆盖范围,成为一项关键挑战。为此,研究者提出在网络中布设电池更换站,支持无人机按需补能^[19]。通过多架无人机协同作业,并在任务执行过程中多次利用电池更换站补充电池,能够显著提升任务完成度。这一问题可视为经典团队定向问题(Team Orienteering Problem, TOP)的变体。

现有关于无人机TOP的研究主要可分为2类,区别在于对能源约束和补给机制的处理方式。第1类仅考虑能源约束,未引入补给机制。Xu等^[20]针对移动充电器调度场景研究了带成本预算约束的TOP,并提出了具有理论保证的近似算法;Juan等^[21]研究了电池状态影响节点收益的动态定向问题;Ammouriova等^[22]进一步将强化学习应用于带电池约束的动态TOP。此类研究虽然考虑了电池约束,但无人机必须在单次续航时间内完成任务,无法通过中间补给延长作业时间,严重限制了任务覆盖范围。第2类研究引入了能源补给机制,但设施功能单一或位置受限。Lee等^[23]提出了UAV-UGV协同作业方案,由单个UGV沿预定义固定路线移动并为UAV提供无线充电,但充电机会完全依赖于UGV的预设轨迹,缺乏灵活性;秦文龙等^[24]研究了多站点多航次的应急巡检问题,允许无人机在多个站点之间换电,但这些站点同时承担任务起点、终点和换电设施的多重角色,导致无人机需要频繁往返站点,时间成本较高,且路径规划灵活性不足。综上所述,现有针对无人机TOP的研究缺乏无人机续航的补给机制,或将任务起点与换电设施混为一体,尚未引入独立电池更换站的概念,尤其是在电池更换站选址与无人机路径规划的联合优化方面,仍存在研究空白。

TOP的核心目标是在给定时间范围内,最大化访问路径所覆盖节点的总收益。由于其NP难特性,该问题受到广泛关注^[25]。现有TOP研究主要分为3类:精确算法^[26]、启发式算法^[27]和学习算法^[28]。精确算法如分支定界法、动态规划和列生成法,在小规模问题中表现优异,但在大规模实例中计算复杂度急剧增加。启发式算法能在合理时间内获得近似最优解,但解的质量受初始条件和参数设置影响较大,且易陷入局部最优。学习算法通过数据驱动的方式进行路径优化,能够结合在线训练与离线求解,快速

生成近似最优解^[29]。其中,深度强化学习(Deep Reinforcement Learning, DRL)因其通过智能体与环境交互自主生成训练数据而受到广泛关注,不依赖预设数据集,展现出在TOP上获取近似最优解的潜力。然而,针对多无人机协同交通监控管理问题仍然有限,尤其在存在多个电池更换站的场景尚缺乏深入探索。目前Fuertes等^[28]提出的TOP-Former框架采用集中式Transformer架构求解标准TOP,在多智能体路径规划中取得了优异性能,显著优于传统精确算法和启发式方法,然而未考虑飞行途中的电量补给需求。在实际的交通监控应用中,无人机受限于电池容量,需要通过电池更换站进行补能。电池更换站等起降机场是低空空域环境的关键节点^[30],电池更换站等起降站点选址管理工作将直接影响低空航线网络结构和所需的地面基础设施布设^[31]。因此,在TOP中引入电池更换站的补能机制,在保证续航约束的同时最大化监控覆盖范围,成为亟待解决的关键问题。

基于上述研究现状,本文将多无人机协同交通监控管理问题建模为混合整数线性规划(Mixed Integer Linear Programming, MILP)模型,并提出了一种基于多智能体Transformer的强化学习(Multi-agent Transformer-typed Reinforcement Learning, MTRL)算法来求解该问题。本文的创新点如下:①模型层面,提出MILP建模框架,为多无人机协同交通监控管理问题提供精确的数学表达,并通过求解器获得的精确解作为基准,用于评估MTRL方案与精确解之间的差距;②算法层面,设计MTRL算法框架,引入掩码机制确保路径可行性,有效避免电池更换站之间的子循环问题,同时结合聚类方法优化无人机电池更换站的选址,使其在监控区域中分布更加均匀;③试验层面,通过数值试验验证,经过充分训练的MTRL模型在处理不同规模任务时均能获得高质量可行解,且相较基准方法显著缩短求解时间。进一步的分布泛化试验表明,该模型能够有效适应不同节点分布模式生成的实例,无需重新训练,表现出较强的跨分布场景迁移能力。

1 问题描述

本文研究一种多无人机协同交通监控路径优化问题。给定一组监控目标点、若干电池更换站以及单个无人机基地,在有限任务时间和电池续航约束下,通过优化多架无人机的飞行路径,最大化被访问目标点的总收益。具体而言,该问题满足以下约束

条件:①每架无人机从基地出发,完成监控任务后返回基地;②每个目标点最多被一架无人机访问一次;③电池更换站可被多架无人机重复访问,以便无人机进行续航补给;④每架无人机的单次飞行时间不超过其电池续航能力,且总飞行时间不超过给定的任务时间限制。优化目标是在满足上述约束的前提下,最大化所有无人机访问的目标点收益总和。如图 1 所示,无人机基地、目标点和电池更换站分别用蓝色基地、黑色摄像头和绿色电池更换站表示。所有无人机均从同一基地起飞,前往目标点完成监控任务后返回基地。每个目标点仅允许单架无人机访问一次或不被访问;电池更换站则可重复访问或不访问。无人机能耗来源于飞行过程和在被目标点的悬停监测过程。

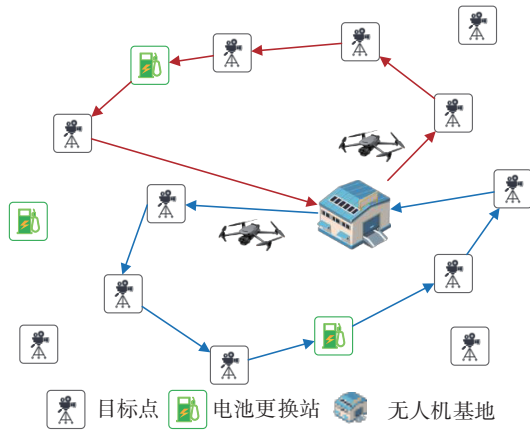


图 1 含电池更换站的无人机协同交通监控路径示例

Fig. 1 Example paths for cooperative UAV traffic monitoring with battery replacement stations

在多无人机协同监控场景中,各无人机需要协调完成以下任务:①合理分配目标点,避免重复访问造成资源浪费;②协调电池更换站的使用,避免多架无人机同时竞争同一换电资源;③在满足各自电池约束的前提下,最大化团队整体的监控覆盖范围。为实现上述协同目标,本文采用集中式决策框架,通过共享包括所有无人机的位置、电量状态和已访问节点信息等在内的全局状态信息,从而使各无人机智能体能够感知彼此状态,在统一的策略网络指导下生成协调一致的路径方案。

电池更换站被假设拥有充足的电池数量,能够为无人机提供即时补能。无人机离开电池更换站后,其续航能力将被完全恢复。电池更换站的空间布局直接影响无人机的补能策略及整体路径优化结果。因此,任务首先需要根据目标点的空间分布合理规划电池更换站位置。随后,综合电池更换站与

目标点的数量和位置,优化无人机访问顺序与电池更换次数,进而生成接近最优的协同飞行路径。在实际应用场景中,交通监控目标点通常基于历史交通数据与城市规划预先确定,包括主要路口、快速路出入口、事故多发路段等关键位置。无人机在监控过程中通过机载传感器如高清摄像头、激光雷达等采集交通流参数和异常事件信息,并通过 5G 无线网络实时回传至交通管理中心。获取的监控信息可支撑多种应用,例如交通流分析(识别拥堵传播模式,预测交通态势演化),事故检测与响应(快速定位事故位置,引导应急车辆),信号控制优化(为自适应信号控制系统提供实时数据输入),交通管理决策(辅助管理部门制定疏导预案,如临时改变信号配时、发布出行诱导信息)等。上述应用能够有效提升交通监控的时空覆盖范围和响应速度。

本文将多无人机协同交通监控管理问题建模为 TOP。该问题可表示为有向图 $\mathcal{G}=(\mathcal{V}, \mathcal{E})$,式中: \mathcal{V} 为节点集; \mathcal{E} 为边集。节点集 \mathcal{V} 包括基地节点 $i=0$ 与 $i=n+1$,目标点集合 $\mathcal{V}'=\{1, \dots, n\}$,电池更换站集合 $\mathcal{S}=\{s_1, s_2, \dots, s_k\}$,其中: i 为节点序号; n 为目标点数量; s_k 为第 k 个电池更换站。边权重定义为 $\mathcal{E}_w=\{t_{ij}\}$, t_{ij} 为无人机从节点 i 到节点 j 的耗时,可通过 i 至 j 的距离除以无人机飞行速度计算得出。目标点 $j \in \mathcal{V}'$ 的悬停监测时间记为 t_j 。为简化建模,假设无人机以恒定速度飞行,起降过程消耗的续航能力忽略不计。可行路线始于基地,在规定时间内访问若干目标点和必要的电池更换站,并最终返回基地。研究目标寻找一组可行路线,使无人机总续航消耗不超过规定时限,同时最大化被监控的目标点数量。

2 多无人机协同监控管理问题建模

2.1 无人机电池更换站选址

针对无人机电池更换站的选址问题,本文采用聚类方法确定其选址。设监控区域内目标点总数为 $|\mathcal{V}'|$,要设置的电池更换站数量为 $|\mathcal{S}|$,将目标点按照其横纵坐标位置聚为 $|\mathcal{S}|$ 类,每类聚类中心即为电池更换站位置。其坐标表达式如下

$$x_k = \frac{1}{|C_s|} \sum_{x_i \in C_s} x_i \quad (1)$$

$$y_k = \frac{1}{|C_s|} \sum_{y_i \in C_s} y_i \quad (2)$$

$$|C_s| = |\mathcal{V}'| / |\mathcal{S}|$$

式中: x_k, y_k 分别为聚类中心的横、纵坐标,即对应的

电池更换站位置坐标; C_s 为第 s 个聚类簇, $s \in \mathcal{S}; |C_s|$ 为簇 s 中目标点个数; x_i, y_i 分别为簇内任意目标点 i 的横、纵坐标。

2.2 多无人机路径规划建模

在确定电池更换站位置后,多无人机协同交通监控管理问题可表示为混合整数线性规划模型

$$\max R = \sum_{j \in \mathcal{V}'} \sum_{i \in \mathcal{V}} r_j \phi_{ij,a} \quad a \in \mathcal{A} \quad (3)$$

式中: R 为所有无人机访问目标点的总奖励; a 为无人机编号; \mathcal{A} 为无人机集合; r_j 为访问目标点 j 的奖励,在实际应用中,交通管理部门根据目标点的历史拥堵频率、事故风险、道路重要性等因素赋予不同目标点不同的奖励值; $\phi_{ij,a}$ 为0-1决策变量,表示无人机 a 是否从节点 i 直接访问节点 j , $\phi_{ij,a} = 1$ 表示直接访问,否则 $\phi_{ij,a} = 0$ 。

当 $j \in \mathcal{V}'$ 时,本文支持3种奖励权重分配模式:

①均等权重,所有目标点奖励相同,适用于需要全面覆盖的场景;②均匀分布权重,目标点奖励在 $[0.01, 1.00]$ 区间均匀分布,用以模拟不同重要度的监控目标;③距离相关权重,目标点奖励与其距离基地的距离成正比,用以体现远距离目标点的战略监控价值。为聚焦验证算法在复杂约束下的协同路径优化性能,本文试验采用均等权重模式(即 $r_j = 1$),在此基础上评估模型的收敛性、求解效率和泛化能力。在实际部署中,可根据具体交通网络特征灵活选择奖励权重分配模式。当 j 为0或 $n+1$ 时,则 $r_j = 0$,即基地无奖励;若 $j \in \mathcal{S}$,则 $r_s = 0$,电池更换站同样没有奖励。约束条件包括以下12点。

(1) 无人机起飞约束

$$\sum_{j=1}^{n+1} \phi_{0j} + \sum_{s \in \mathcal{S}} \phi_{0s} = |\mathcal{A}| \quad (4)$$

式中: ϕ_{0j} 为0-1变量,若无人机从起飞节点行驶到节点 j ,则取值为1,否则取值为0; ϕ_{0s} 为0-1变量,若无人机从起飞节点行驶到电池更换站 s ,则取值为1,否则取值为0; $|\mathcal{A}|$ 为无人机数量。

(2) 无人机返回约束

$$\sum_{i=0}^n \phi_{i(n+1)} + \sum_{s \in \mathcal{S}} \phi_{s(n+1)} = |\mathcal{A}| \quad (5)$$

式中: $\phi_{i(n+1)}$ 为0-1变量,若无人机从节点 i 返回降落节点,则取值为1,否则取值为0; $\phi_{s(n+1)}$ 为0-1变量,若无人机从电池更换站 s 返回降落节点,则取值为1,否则取值为0。

式(4)、(5)确保了无人机的飞行路径数量的正

确性。

(3) 节点流入约束

$$\sum_{i=0}^n \phi_{ij} + \sum_{s \in \mathcal{S}} \phi_{sj} \in [0, 1] \quad j \in \mathcal{V}' \quad (6)$$

式中: ϕ_{sj} 为0-1变量,若无人机从电池更换站 s 行驶至节点 j 则取值为1,否则取值为0。

(4) 节点流出约束

$$\sum_{j=1}^{n+1} \phi_{ij} + \sum_{s \in \mathcal{S}} \phi_{is} \in [0, 1] \quad i \in \mathcal{V}' \quad (7)$$

式中: ϕ_{is} 为0-1变量,若无人机从节点 i 行驶至电池更换站 s ,则取值为1,否则取值为0。

式(6)、(7)限制了目标点和电池更换站之间的访问。由于电池更换站可多次访问,因此对 $j \in \mathcal{S}$ 无访问次数限制。

(5) 节点流平衡约束

$$\sum_{i=0}^n \phi_{ij} + \sum_{s \in \mathcal{S}} \phi_{sj} = \sum_{i'=1}^{n+1} \phi_{ji'} + \sum_{s \in \mathcal{S}} \phi_{js} \quad j \in \mathcal{V}' \cup \mathcal{S} \quad (8)$$

式中: $\phi_{ji'}$ 为0-1变量,若无人机从节点 j 行驶至下一个节点 i' ,则取值为1,否则取值为0。

式(8)为流平衡约束,确保每个节点的流入路径数量等于流出路径数量。

(6) 总飞行时间约束

$$\sum_{i=0}^n \sum_{j=1}^{n+1} t_{ij} \phi_{ij,a} + \sum_{i \in \mathcal{V}} \sum_{s \in \mathcal{S}} t_{is} \phi_{is,a} + \sum_{s \in \mathcal{S}} \sum_{j \in \mathcal{V}} t_{sj} \phi_{sj,a} \leq T \quad (9)$$

式中: t_{is} 为无人机从节点 i 到电池更换站 s 的耗时; t_{sj} 为无人机从电池更换站 s 到节点 j 的耗时; T 为监控时间范围限制。

式(9)表示时间约束,确保形成的路径总续航时间在给定监控时间范围内。

(7) 目标点子回路消除约束

$$u_{i,a} - u_{j,a} + n \phi_{ij,a} \leq n - 1 \quad i, j \in \mathcal{V}' \quad (10)$$

式中: $u_{i,a}$ 和 $u_{j,a}$ 分别为节点 i, j 在无人机 a 路径中的位置。

式(10)能够消除无人机在任意2个目标点之间形成的子回路。

(8) 目标点续航约束

$$0 < b_{j,a} \leq B \quad j \in \mathcal{V}' \quad (11)$$

式中: $b_{j,a}$ 为无人机 a 离开节点 j 时的剩余续航时间; B 为无人机的总续航,即满电状态下的续航。

(9) 基地和电池更换站满电约束

$$b_{j,a} = B \quad j \in 0, n+1 \cup \mathcal{S} \quad (12)$$

式(11)确保无人机在任意目标点的续航时间为

正且不超过电池最大容量;式(12)确保离开基地和电池更换站时无人机是满续航。

(10)续航更新约束

$$b_{i,a} - (t_{ij} + t_j)\phi_{ij,a} + B(\phi_{ij,a} - 1) \leq b_{j,a} \quad (13)$$

$$b_{j,a} \leq b_{i,a} - (t_{ij} + t_j)\phi_{ij,a} + B(1 - \phi_{ij,a}) \quad (14)$$

式中: $b_{i,a}$ 为无人机从节点*i*离开时的剩余续航时间; t_j 为节点*j*的监控任务执行时间,即无人机在该节点的停留时长。

式(13)、(14)共同确保无人机*a*从任意目标点*i*到达*j*,并完成监控任务离开时的续航时间为 $b_{j,a} = b_{i,a} - (t_{ij} + t_j)$ 。

(11)电池更换站子回路约束

$$\sum_{i \in \mathcal{S}} \phi_{ij,a} = 0 \quad j \in \mathcal{S} \quad (15)$$

$$\sum_{j \in \mathcal{S}} \phi_{ij,a} = 0 \quad i \in \mathcal{S} \quad (16)$$

(12)节点自环禁止约束

$$\phi_{ii,a} = 0 \quad i \in \mathcal{V} \quad (17)$$

式(15)、(16)确保任意2个电池更换站之间不会形成子回路;式(17)确保无人机在任意节点的访问不会形成自环。

2.3 基于多智能体 Transformer 的强化学习算法

2.3.1 MTRL 模型框架

为解决大规模实例中 MILP 算法求解效率低下的问题,本文提出 MTRL 算法求解多无人机协同交通监控管理问题。该算法将多无人机路径规划建模为马尔可夫决策过程,通过多头注意力机制捕获全局特征,结合掩码策略保证路径可行性,并利用策略梯度方法训练模型,从而高效生成高质量可行解。

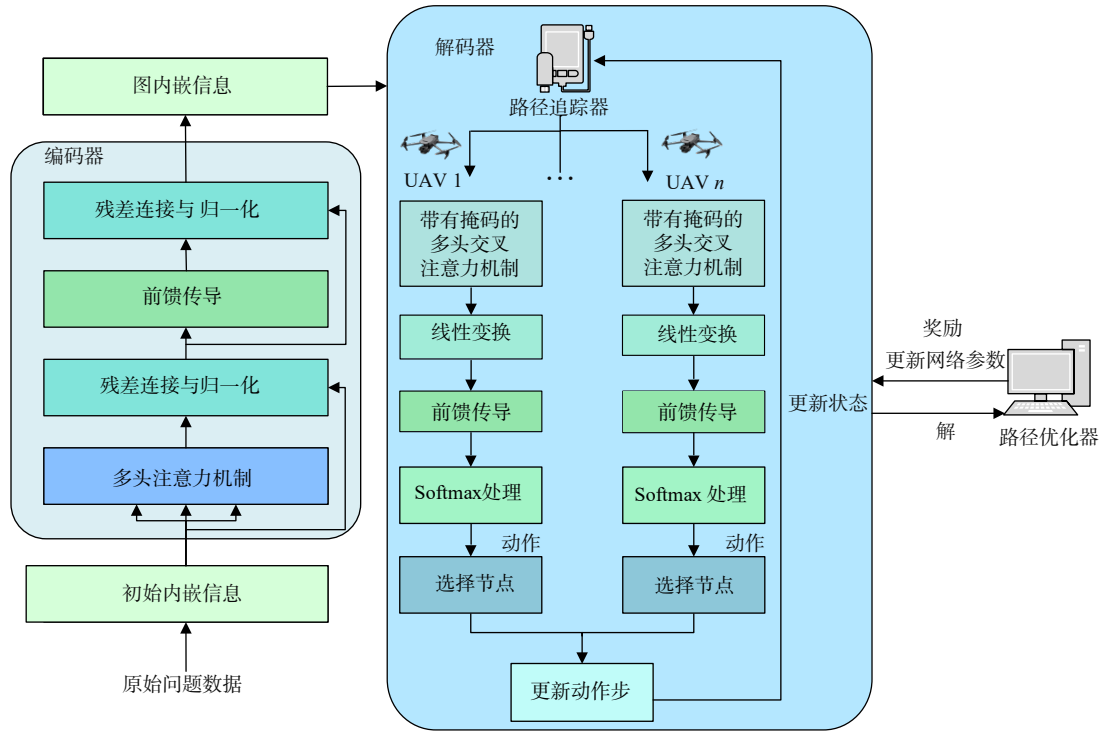


图2 MTRL模型架构

Fig. 2 Architecture of MTRL model

MTRL模型采用集中式架构,其结构如图2所示。首先,原始问题数据通过线性变换层转换为初始内嵌信息,包含电池更换站、目标点和基地的内嵌信息,下文统称节点内嵌信息。随后,初始内嵌信息输入*M*层编码器,如图2编码器部分侧面符号表示,利用多头注意力机制提取全局特征并生成图内嵌信息,图2中采用3个箭头表示从初始内嵌信息到注意力机制的多头数据流向。解码器以图内嵌信息为输入,将每架无人机视为一个独立的智能体*a*,通过掩

码多头交叉注意力机制依次选择下一个节点,形成访问序列。在每个决策步,解码器同时为所有无人机选择下一个访问节点,形成联合动作。各智能体通过共享的图内嵌信息和全局访问掩码感知其他无人机的状态(当前位置、剩余电量)和已访问节点信息,从而实现隐式协同,电池更换站则可被多架无人机重复访问。以上基于共享信息和掩码机制的协同方式,使得各无人机在保持决策独立性的同时,能够生成全局协调的路径方案。掩码机制由路径追踪器

生成,确保所有无人机路径的可行性,且保证当前所有路径的总续航时间消耗不超过给定时间范围。当所有无人机均返回基地后,一次完整的路径规划过程结束。解码器评估这些路线的奖励值并将结果传递至路径优化器。路径优化器通过更新MTRL模型的网络参数提升奖励值,从而启动新一轮路径规划。该过程持续至训练结束,最终,更新后的MTRL模型将作为离线求解器投入使用。

该模型定义了一个随机概率集 P ,用于在当前问题实例 Z 的动作步 δ 选择所有无人机将访问的节点集,规定每架无人机一次只能访问一个节点,其表达式为

$$P = \{p(x_{a,\delta}|Z, \sigma), a \in \mathcal{A}, \delta \in \Delta\} \quad (18)$$

$$p(x_{a,\delta}|Z, \sigma) = p(x_{a,\delta}|Z, \sigma, x_{1,\delta-1}:x_{|\mathcal{A}|,\delta-1}) \quad (19)$$

式中: $p(\cdot)$ 为随机概率; $x_{a,\delta}$ 为无人机 a 在第 δ 步选择的节点; σ 为MTRL的网络策略参数; Δ 为动作步集合; $x_{1,\delta-1}$ 和 $x_{|\mathcal{A}|,\delta-1}$ 分别为首、末无人机在第 $\delta-1$ 步选择的节点。

2.3.2 编码器

MTRL采用标准化的编码器^[32]。原始输入数据包含节点的坐标信息及其消耗的续航时间,通过线性投影层映射为初始内嵌信息 $h_{i,0}, i \in \mathcal{V}$ 。为提取内嵌信息的特征,如图3多头注意力机制所示,多头自注意力层基于输入内嵌信息计算其请求 q_i 、键 k_i 和值 v_i 。通过缩放点积法计算节点 i 的请求 q_i 与节点 j 的键 k_j 之间的兼容性 μ_{ij} 。随后将该结果通过softmax函数处理,以获得节点权重 y_{ij} ,其表达式为

$$\mu_{ij} = \frac{q_i k_j^T}{\sqrt{d}} \quad (20)$$

$$y_{ij} = \frac{e^{\mu_{ij}}}{\sum_j e^{\mu_{ij}}} \quad (21)$$

式中: d 为请求、键和值的维度,三者维度相同; y_{ij} 为节点权重。

将得到的兼容性 μ_{ij} 与节点权重 y_{ij} 相乘并对 j 求和,得到第 m 层编码器的单头注意力得分 $h_{i,m}$ 。多头自注意力机制通过拼接不同注意力头的信息,使模型能够并行捕捉多个子空间的信息^[28]。经过批量归一化(Batch Normalization, BN)层进行归一化。批量归一化的结果 $h_{i,m}$ 将作为线性变换层的输入,形成输出内嵌信息,线性变换层的激活函数统一采用整流线性单元(Rectified Linear Unit, ReLU)激活函数。得到的输出内嵌信息再进行一轮批归一化处理

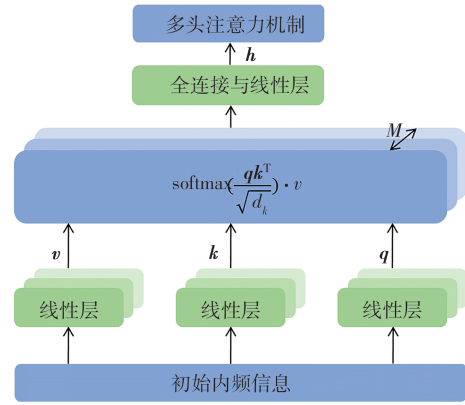


图3 多头注意力机制架构

Fig. 3 Architecture of multi-head attention mechanism

就得到了优势节点内嵌信息 $h_{i,m}$,该过程重复 M 次。最后,编码器将所有节点的内嵌信息聚合,具体为

$$\bar{h} = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} h_{i,M} \quad (22)$$

式中: \bar{h} 为图内嵌信息; $|\mathcal{V}|$ 为节点数量; $h_{i,m}$ 为节点 i 经过第 M 层编码器后的优势节点内嵌信息。

编码层计算的优势节点内嵌信息和图内嵌信息都将作为解码层的输入,用以进行交叉注意力得分计算。

2.3.3 解码器

MTRL解码器由多个深度神经网络构成,每个网络代表一个无人机智能体。这些智能体从路径追踪器接收图信息(如图2所示),并利用该信息确定需依次访问的节点序列。在此过程中,路径追踪器充当环境角色,接收的信息即当前状态,而待访问节点集则代表动作。无人机接收环境状态后通过采取动作改变状态,并在多轮交互中学习最优策略,以实现长期累计奖励的最大化。MTRL解码器的核心要素,包括状态、动作和奖励,具体定义如下。

(1)状态:在动作步 δ 时,状态 s_δ 由静态信息和动态信息两部分构成。

①静态信息:包括节点坐标集 \mathcal{C} 、节点奖励值集 P 和节点续航时间消耗集 \mathcal{D} 。其中,当节点为基地或者电池更换站时,节点奖励值取0。

②动态信息:包括无人机当前位置 $p_{\delta,a} \in \mathcal{V}$ 、剩余电量 $b_{\delta,a} \in [0, B]$ 、从初始步到当前动作步 δ 的累计续航时间消耗 $T_{c,\delta,a}$,以及动作步 δ 时的节点访问掩码 $\mathcal{M}_\delta \in \{0, 1\}$ 。访问掩码用于标记已访问节点(置为1)和未访问节点(置为0)。

无人机 a 在动作步 δ 的剩余电量 $b_{\delta,a}$ 更新为

$$b_{\delta,a} = b_{\delta-1,a} - t_{ij} - t_j \quad (23)$$

累计续航时间消耗为

$$T_{c,\delta,a} = - \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}} (t_{ij} + t_j) \quad (24)$$

为整合状态信息,在每个动作步 δ 拼接环境内嵌信息 $\mathbf{h}_{\delta,c}$,其由图内嵌信息、优势节点内嵌信息以及无人机的剩余续航时间构成,具体如下

$$\mathbf{h}_{\delta,c} = \text{cat}(\bar{\mathbf{h}}, \mathbf{h}_{1,M}; \mathbf{h}_{|\mathcal{V}|,M}, b_{1,\delta}; b_{|\mathcal{A}|,\delta}) \quad (25)$$

式中: $\text{cat}(\cdot)$ 为拼接操作; $\mathbf{h}_{1,M}$ 和 $\mathbf{h}_{|\mathcal{V}|,M}$ 分别为首、末节点的优势节点内嵌信息; $b_{1,\delta}$ 和 $b_{|\mathcal{A}|,\delta}$ 分别为首、末无人机在动作步 δ 的剩余续航。

(2)动作:在动作步 δ ,无人机智能体 a 的动作即为选择下一个访问节点,记为 $\alpha_\delta = j_a$ 。可行动作集合定义为 $\{\mathcal{A}_{\delta,a} = j \in \mathcal{V} | \mathcal{M}_{\delta,a}(j) = 0\}$,即由未被掩码的节点组成,其中: $\mathcal{A}_{\delta,a}$ 为无人机 a 在动作步 δ 可执行的动作集; $\mathcal{M}_{\delta,a}(j)$ 为无人机 a 在动作步 δ 选择下一个节点 j 时的结构化掩码集。

为计算动作分布,解码器通过多头注意力机制聚合不同节点的内嵌信息,计算出新的多无人机环境内嵌信息 $\mathbf{h}_{\delta,c'}$ 为

$$\mathbf{h}_{\delta,c'} = \text{MHA}(\mathbf{h}_{\delta,c}, \mathbf{W}_{c,q} \mathbf{h}_{\delta,c}, \mathbf{W}_{M,k} \mathbf{h}_M, \mathbf{W}_{M,v} \mathbf{h}_M) \quad (26)$$

$$\mu_{i,u,\delta} = \tau \tanh\left(\frac{\mathbf{q}_\delta \mathbf{k}_{i,\delta}^T}{\sqrt{d}}\right) \quad (27)$$

式中: $\text{MHA}(\cdot)$ 为多头注意力机制层; $\mathbf{W}_{c,q}$ 为候选节点的查询投影矩阵; $\mathbf{W}_{M,k}$ 、 $\mathbf{W}_{M,v}$ 分别为优势节点的键投影矩阵与值投影矩阵,三者均为模型训练过程中自适应学习的参数; \mathbf{h}_M 为编码器计算的优势节点内嵌信息; \mathbf{q}_δ 为注意力机制计算请求; $\mathbf{k}_{i,\delta}$ 为所有节点上的键; τ 为映射范围限制。

为生成协同路径,解码器按如下步骤进行工作。

步骤1:状态编码。在动作步 δ ,解码器整合当前状态 s_δ 与编码器生成的优势节点内嵌信息,通过多头注意力机制计算多无人机环境内嵌信息 $\mathbf{h}_{\delta,c'}$,具体见式(26)。

步骤2:动作概率计算。利用交叉注意力机制,计算每架无人机与各候选节点的兼容性得分,具体见式(27)。

步骤3:可行性约束。引入掩码机制过滤不可行节点。对于被掩码的节点,将其兼容性设为 $-\infty$,确保该节点不会被选择。

步骤4:联合动作选择。将兼容性通过激活函数转换为概率分布,并采用采样或贪婪策略为每架无人机同时选择下一个节点,形成联合动作。

步骤5:状态更新。根据选定的动作更新各无人

机状态和全局访问掩码,并进入下一动作步 $\delta + 1$ 。

上述过程循环执行,直至所有无人机返回基地,生成完整的多无人机协同路径规划方案。

在每一个动作步 δ ,并非所有无人机都能访问 \mathcal{V} 中的所有节点,其可访问性取决于访问掩码 $\mathcal{M}_{\delta,a}$ 。本文设计了几种掩码机制将不可访问的节点在访问掩码中的值设置为1,进而找到所有不可访问节点,通过设置 $\mu_{i,u,\delta} = -\infty$ 来排除不可行动作。该掩码机制描述如下。

(1)重复性约束掩码(\mathcal{M}_1):在动作步 δ 时,已被任何无人机访问的目标点应被掩码;电池更换站节点不受此约束限制。

(2)能量约束掩码(\mathcal{M}_2):若无人机在节点 i 离开时的剩余电量不足以完成 $i \rightarrow j$ 的飞行、完成任务并返回基地,即 $b_{i,\delta-1,a} - (t_{ij} + t_j + t_{j,d}) \leq 0$,则节点 j 应在动作步 δ 被掩码。式中: $b_{i,\delta-1,a}$ 表示无人机 a 在动作步 $\delta - 1$ 离开节点 i 时的剩余电量; $t_{j,d}$ 为从节点 j 返回基地的飞行时间。

(3)累计时间约束掩码(\mathcal{M}_3):若累计续航时间超过最大任务时限 T_{\max} ,即 $T_{c,a} + t_{ij} + t_{j,d} > T_{\max}$ 时,节点 j 应被掩码。式中: $T_{c,a}$ 为无人机 a 的累计任务时间。

(4)电池更换站连续访问约束掩码(\mathcal{M}_4):若无人机在动作步 $\delta - 1$ 刚访问了某个电池更换站,则所有电池更换站节点($s \in \mathcal{S}$)在下一动作步 δ 都应被掩码,以消除电池更换站间的无效跳转。

(5)电量安全约束掩码(\mathcal{M}_5):若访问某目标节点后,剩余电量不足以到达至少一个电池更换站,即 $b_{i,\delta-1,a} - \left[t_{ij} + t_j + \max_{s \in \mathcal{S}} (t_{js}) \right] \leq 0$,则目标节点 j 应被掩码。

(6)基地访问规则掩码(\mathcal{M}_6):起始基地(节点0)在所有动作步中始终被掩码;终点基地(节点 $n + 1$)在任何动作步都可访问,以允许无人机随时结束任务。当无人机已到达终点基地时,所有其他节点都应被掩码。

掩码含义如表1所示,最终掩码通过逻辑或运算整合

$$\mathcal{M}_\delta = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3 \cup \mathcal{M}_4 \cup \mathcal{M}_5 \cup \mathcal{M}_6$$

(3)奖励:MTRL的目标是学习有效策略,使得长期累积奖励最大化。本文将奖励函数定义为所有无人机访问的目标点总数

$$R_t = \sum_{a \in \mathcal{A}} \sum_{\delta=1}^{T_a} r_{\delta,a} \quad (28)$$

表1 掩码含义说明
Table 1 Explanation of masks

掩码符号	掩码名称	掩码含义
\mathcal{M}_1	重复性约束	已被任意无人机访问过的目标点被掩码
\mathcal{M}_2	能量约束	到达下一个目标点后剩余电量不够返回基地的目标点被掩码
\mathcal{M}_3	累计时间	到达下一个目标点并返回基地的累计时间超过最大任务时限的目标点被掩码
\mathcal{M}_4	电池更换站连续访问约束	从电池更换站出发时所有电池更换站被掩码
\mathcal{M}_5	电量安全约束	到达下一个目标点后剩余电量不够到达至少一个电池更换站的目标点被掩码
\mathcal{M}_6	基地访问规则	起始基地始终被掩码

式中: T_a 为无人机 a 的总动作步数; $r_{\delta,a}$ 为无人机 a 在 δ 步的奖励值。

单步奖励值定义为

$$r_{\delta,a} = \begin{cases} 1 & j \in \mathcal{V}' \\ 0 & j \in S \cup \{0, n+1\} \end{cases} \quad (29)$$

无人机访问目标点时奖励为1,否则为0。所有监控任务完成后,计算总奖励 R_t ,并基于该总奖励更新MTRL策略网络的参数。

2.3.4 MTRL训练过程

MTRL编码器和解码器中使用的所有可训练变量均整合为由策略网络 σ 参数化的向量。采用著名的策略梯度方法训练MTRL网络^[33]。MTRL包含2个神经网络:策略网络和基线网络。策略网络根据当前状态计算动作空间上的概率分布,并通过采样选择动作。基线网络采用贪婪策略选择动作以计算基线奖励。基线网络与策略网络采用相同结构。训练过程中利用单边 t 检验验证策略网络是否显著优于基线网络。若单边 t 检验值小于给定阈值,表明策略网络显著优于基线网络,则将基线网络参数替换为策略网络参数。经过充分训练后,策略网络可为生成最大化目标点访问数量的高质量可行解,用于离线部署。

3 数值试验

本节通过数值试验评估所提出的MTRL模型在多无人机协同交通监控管理问题中的性能表现。本文设计了3种试验场景,分别对应2~4架无人机协同执行任务。所有无人机均从同一基地同时出发,在监控区域内协同完成目标点访问任务。每架无人机可在任意电池更换站进行补给,并需在指定的任务时间上限内返回基地。在该设定下,MTRL模型分别管理2、3或4个智能体,通过协调无人机的访问顺序,以最大化总的访问目标点数量。

3.1 试验设置

参照文献[34]~[37],本文设置了试验场景参

数与MTRL模型的超参数。试验中无人机运行参数参照DJI Matrice M300 RTK商用无人机设定,该型号广泛应用于城市交通监控、应急管理等领域,其最大飞行速度 $23 \text{ m} \cdot \text{s}^{-1}$,最大续航时间43 min。考虑搭载高清摄像头、激光雷达等监控传感器后的实际载荷,以及城市交通管控巡检的实际应用场景,本文设定无人机续航时间为40 min,飞行速度为 $12 \text{ m} \cdot \text{s}^{-1}$ 。无人机基地和目标点的位置在边长为1 km的正方形区域内均匀采样生成,每个目标点的监控时间在0~10 min随机分配。在奖励值设置方面,本文采用均等权重模式,将所有目标点的奖励均设为1,目的在于聚焦评估算法在电池约束和电池更换站管理方面的性能表现。MTRL的框架设计支持任意奖励分配模式,试验中的简化假设不影响算法在实际场景中的适用性。起降过程的能耗忽略不计,且所有电池更换站的电池更换时间均设定为1 min,其选址通过聚类方法确定。为全面评估MTRL模型,本文设置了不同规模的试验,采用场景编号对试验场景进行描述,其中T表示目标点数量,B表示电池更换站数量,A表示无人机数量。

(1)小规模试验:20个目标点,2个电池更换站和2架无人机(记为T20B2A2)。

(2)中规模试验:50个目标点,5个电池更换站和2架无人机(记为T50B5A2)。

(3)大规模试验:100个目标点,10个电池更换站和2架无人机(记为T100B10A2)。

在多无人机协同的场景中,若无人机数量为3或4,则分别记为A3和A4。试验步骤如下。

步骤1:对比不同聚类方法在电池更换站选址问题中的性能,包括约束 K 均值聚类(K -means Constrained, KMC)、模糊 C 均值聚类(Fuzzy C -means, FCM)、带噪声的基于密度的空间聚类(Density-based Spatial Clustering of Applications with Noise, DBSCAN)、谱聚类(Spectral Clustering, SC)以及高斯混合模型(Gaussian Mixture Model, GMM)。

步骤2:验证MTRL在不同试验规模和无人机

数量下的训练收敛效果,并与基于指针网络的强化学习(Pointer-network-based Reinforcement Learning, PNRL)算法^[38]及商业求解器 CPLEX 对 MILP 模型的求解结果进行对比,评估计算效率和解的质量。PNRL 基于 RNN 编码器,代表了序列建模的强化学习方法。RNN 按顺序处理节点信息,存在长距离依赖衰减问题,难以有效捕获多无人机、多目标点和多电池更换站之间的复杂全局依赖关系。相比之下, MTRL 采用 Transformer 架构通过多头自注意力机制并行处理所有节点信息,在单次前向传播中建模所有节点间的依赖关系,更适合处理本文的多智能体协同决策和复杂约束问题。因此选择 PNRL 作为对比方法,验证 Transformer 架构在求解多智能体路径规划问题上的优势。

步骤 3:评估 MTRL 模型的迁移能力:在不同规模和无人机数量的设置中,利用不同分布生成目标点,并加载已训练的模型进行泛化测试。

参照文献[34]~[37],训练样本采用随机生成,每批次包含 256 个样本,每轮训练 200 个批次,总训练轮次数设为 200。编码器与解码器中节点内嵌信息及隐含层维度均设为 128,编码器包含 3 个隐含层。优化器采用 Adam,学习率设为 1.0×10^{-4} 。显著性检验使用单边 t 检验,显著性水平为 5%。所有试验均在搭载 RTX 3090 GPU 的戴尔服务器上运行,试验环境为 Python 3.11.7。

3.2 电池更换站选址方法评估

在多无人机 TOP 中,电池更换站的选址对路径可行性及任务完成率影响显著。为此,本文对比分析了 5 类具有代表性的聚类方法。

(1)KMC 算法:在传统 K -means 的基础上引入聚类规模约束,保证每个聚类目标点数量合理,从而避免电池更换站服务区域过大或过小,提升服务均衡性。

(2)FCM 算法:基于模糊隶属度的“软”聚类方法概念,允许目标点以不同程度隶属于多个聚类中心,能够更好地处理边界模糊的目标点分配问题,适合识别位于多个电池更换站服务边界区域的关键目标点。

(3)DBSCAN 算法:基于密度驱动的聚类方法,能够自动识别任意形状的聚类并有效处理噪声点,适用于目标点分布不规则或存在离群点的复杂环境。

(4)SC 算法:基于图论的谱聚类方法,通过分析目标点空间连接关系识别非凸聚类结构,能够在复杂不规则分布场景下优化电池更换站布局,提升服

务覆盖灵活性。

(5)GMM 算法:基于概率统计的“软”聚类方法,通过 EM 算法估计多个高斯分布的混合参数,不仅提供聚类归属还能量化目标点属于各聚类的概率,适用于目标点呈多峰分布的复杂区域,为电池更换站选址提供概率化的决策依据。

5 种算法分别代表了基于约束优化、模糊隶属度、密度驱动、图论结构和概率统计模型的不同聚类范式,从多个角度评估电池更换站选址问题的解决方案质量,为电池更换站地址选择方法提供全面的性能对比。

为客观评估不同聚类算法在电池更换站选址问题中的性能表现,本文构建了 4 个维度的综合评价指标体系。

(1)Calinski-Harabasz(CH)指数:衡量聚类间分离度和聚类内紧密度,数值越高,聚类效果越好。

(2)最小距离(Minimum Distance, MD):指电池更换站之间的最小欧几里得距离,反映电池更换站的空间分布合理性。

(3)覆盖率(Coverage, CO):即在给定服务半径(0.2 单位)内被电池更换站覆盖的目标点比例,反映对任务目标点的服务覆盖程度,直接影响无人机执行任务时的充电可达性和任务完成率。

(4)均匀度(Uniformity, UN):基于距离变异系数的倒数计算,衡量电池更换站分布的均衡性,直接影响负载均衡与多无人机协同路径优化效果^[39]。

参考相关研究^[34-35,37]的试验设置,本文采用 256 个随机实例的大规模试验对 5 种聚类算法进行验证,其性能表现如图 4 所示。显著性分析表明,5 种方法在所有评价指标上均存在显著差异($p < 0.001$),说明不同聚类算法在电池更换站选址性能方面存在本质区别。图 4(a)为 5 种聚类算法的 CH 指数,可以看出:KMC 和 FCM 的平均 CH 指数分别为 51.646 和 53.705,显著高于 DBSCAN(33.629)、SC(44.899)和 GMM(39.450)。较高的 CH 指数表明 KMC 和 FCM 在目标点聚类结构的紧凑性和可分性方面表现更优,且两者结果接近。

图 4(b)为 5 种聚类算法的最小距离指标结果,可以看出:KMC、FCM、SC 和 GMM 的电池更换站平均间距分别为 0.329、0.328、0.315 和 0.323,均显著大于 DBSCAN 的 0.230,且方差更小。这表明其他 4 种方法得到的电池更换站分布更加均匀,其中 KMC 和 FCM 的结果最为稳定。

图 4(c)为聚类算法的覆盖率指标结果,可以看

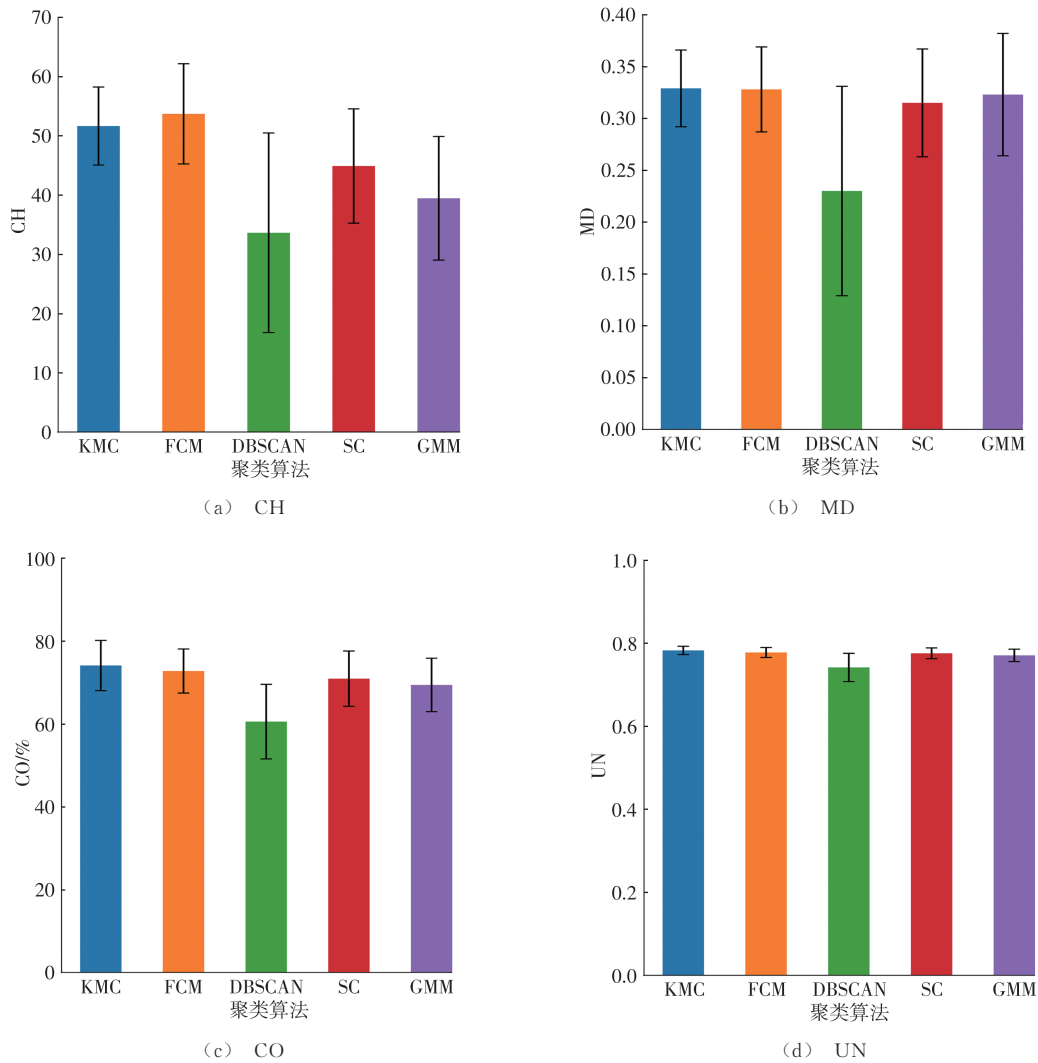


图4 五种聚类算法性能对比

Fig. 4 Comparison of performances among five clustering algorithms

出:KMC和FCM的平均CO分别达到了74.102%和72.766%,显著高于DBSCAN的60.547%,并且优于SC(70.922%)和GMM(69.414%),表明KMC和FCM能够在更广的空间范围内提供更合理的电池更换站布局,为后续的无人机路径优化提供更高的任务覆盖率。

图4(d)为聚类算法的均匀度指标,可以看出:五种方法在电池更换站分布均匀度上(UN)的差异并不显著,但从平均值和标准差来看,KMC的均匀度最佳,其次为FCM、SC、GMM,而DBSCAN最差。

综合CH、MD、CO和UN四个指标的表现,KMC和FCM在平均值和标准差上均显著优于其他算法,SC和GMM的表现次之,DBSCAN效果最差;进一步比较发现,KMC在MD、CO和UN三项指标上整体表现更佳。因此,本文在后续试验中选择KMC作为电池更换站选址方法。图5为电池更

换站选址结果,如图在边长1 km的正方形区域内,图5(a)~(c)分别在20、50、100个目标点的3种场景中通过KMC进行了电池更换站的选址,可以看出,KMC方法在3种场景下均实现合理选址。

3.3 MTRL收敛效果

图6给出了在T20B2A2、T20B2A3、T20B2A4、T50B5A2、T50B5A3、T50B5A4、T100B10A2、T100B10A3和T100B10A4共9个场景下的MTRL模型训练奖励曲线。每一轮训练中,累计奖励根据式(28)计算得出。结果表明,MTRL模型在所有场景中都表现出稳定的收敛趋势,验证了算法训练过程的有效性。

在T20场景中,随着无人机数量的增加,模型收敛速度呈现明显的加快趋势。T20B2A2场景需要75轮训练才能收敛,而T20B2A3和T20B2A4场景分别在23轮和6轮训练后即可达到稳定状态,显示

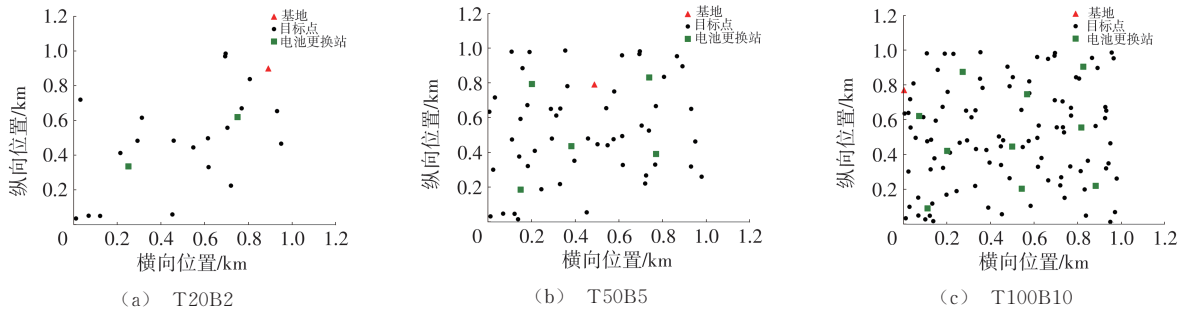


图5 电池更换站选址结果

Fig. 5 Location results of battery replacement stations

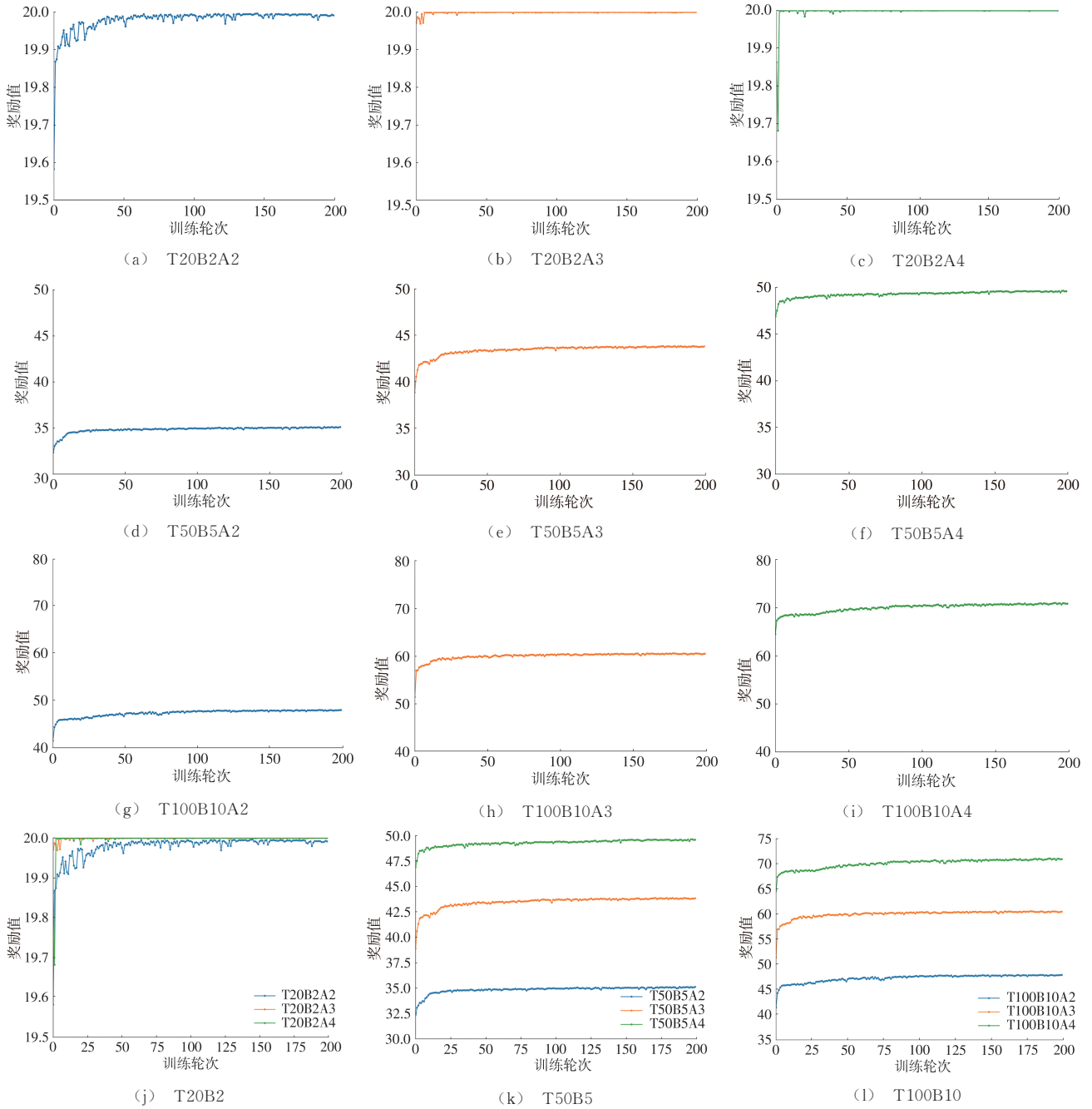


图6 MTRL模型训练奖励曲线

Fig. 6 Training reward curves of MTRL model

出无人机数量增加有助于加快任务分配的收敛速度。尽管无人机数量不同,但由于电量限制、任务时间约束以及电池更换站电池供给能力的制约,3种配置最终都能够实现对全部20个目标点的覆盖。

在 T50 场景中,除了收敛速度随无人机数量增加而加快这一趋势外,最终达到的奖励值也显著提升。T50B5A2、T50B5A3 和 T50B5A4 三种配置的最终奖励值呈递增趋势,依次约为 35、44 和 50。在 T100 场景中, T100B10A2、T100B10A3 和 T100B10A4 三种配置的最终奖励值呈递增趋势,依次约为 46、60 和 71。这一现象反映了在更大规模的任务场景中,增加无人机数量不仅能够提高任务执行效率,还能显著扩大监控覆盖范围。在相同的电量限制和任务时间约束下,更多的无人机通过优化的协同策略能够访问更多目标点,从而获得更高的累积奖励。这种差异表明,对于不同规模的监管任务,无人机数量的配置需要根据目标点密度和任务

复杂度进行合理设计,以实现最优的监控覆盖效果。

3.4 MTRL 模型求解效率和质量评估

为验证所提出 MTRL 模型的计算效率和解质量,将其生成的解与基准方法进行比较。试验选择 T20B2、T50B5 和 T100B10 三个场景,分别在 A2、A3 和 A4 三种无人机配置下进行测试。基准方法为使用 CPLEX 求解多无人机协同交通监控管理问题的整数线性规划模型(设置最大求解时间为 120 s)。在每个场景下,随机抽样 256 个实例进行计算,其中 MTRL 采用贪婪策略生成动作序列。CPLEX 使用相同实例进行对照求解。另外,选择禁忌搜索(Tabu Search, TS)和遗传算法(Genetic Algorithm, GA)2种启发式算法对相同的 256 个案例进行对照求解。比较结果列于表 2,其中“目标值”列为 256 个实例的平均累积奖励,“计算时间”列为平均计算时间。“提升比例”表示相对于使用 CPLEX 所得平均累积奖励的提升比例。

表 2 不同问题规模计算结果比较

Table 2 Comparison of calculation results for problems of different sizes

算法		T20			T50			T100		
		目标值(累计奖励)	提升比例/%	计算时间/s	目标值(累计奖励)	提升比例/%	计算时间/s	目标值(累计奖励)	提升比例/%	计算时间/s
A2	CPLEX	18.215 6	0.00	60.372 4	27.260 0	0.00	107.120 0	41.684 3	0.00	118.763 5
	TS	18.011 7	-1.12	20.271 1	23.179 7	-14.97	27.949 3	24.750 0	-40.63	39.259 9
	GA	18.050 8	-0.90	2.768 0	23.773 4	-12.80	4.030 0	26.238 2	-37.05	5.773 3
	MTRL	19.996 1	9.77	0.003 4	35.101 6	28.77	0.006 7	47.871 8	14.84	0.008 2
	PNRL	19.972 7	9.65	0.006 8	33.738 3	23.76	0.012 3	46.872 3	12.45	0.009 6
A3	CPLEX	16.639 2	0.00	112.956 1	37.726 3	0.00	106.723 9	52.763 2	0.00	117.763 8
	TS	19.996 1	20.17	19.625 0	32.140 6	-14.81	32.057 9	35.543 0	-32.64	44.795 7
	GA	20.000 0	20.20	3.080 1	32.058 6	-15.02	5.224 3	36.878 9	-30.10	7.763 9
	MTRL	20.000 0	20.20	0.006 1	43.750 0	15.97	0.005 7	60.518 1	14.70	0.009 3
	PNRL	19.991 2	20.15	0.007 3	42.339 8	12.23	0.012 3	58.375 9	10.64	0.011 6
A4	CPLEX	16.284 7	0.00	117.485 2	43.681 7	0.00	114.608 3	64.394 6	0.00	118.265 4
	TS	20.000 0	22.81	18.468 8	39.484 4	-9.61	35.205 9	45.058 6	-30.03	51.798 3
	GA	20.000 0	22.81	3.269 0	39.269 5	-10.10	6.441 4	46.183 5	-28.28	9.655 3
	MTRL	20.000 0	22.81	0.004 0	49.539 1	13.41	0.006 1	71.026 7	9.34	0.009 8
	PNRL	19.982 6	22.71	0.008 7	48.668 0	11.42	0.009 5	70.635 7	9.70	0.012 4

表 2 展示了 MTRL 与几种基准算法在不同问题规模下计算结果的指标比较。依据表 2 中目标值和计算时间 2 种指标,可以看出:由于 CPLEX 受到时间限制,未能在大部分实例中找到最优解,导致其解的质量劣于基于强化学习的 MTRL 与 PNRL 算法。对于启发式算法,当问题规模为 T20 时,2 种算法均可快速求出最优解;当问题规模增加到 T50 时,2 种启发式算法的求解效果相较于 CPLEX 限时解具有 9%~15% 的不足,易在迭代早期陷入局部最优解;在更大

规模的问题 T100 案例下,2 种启发式算法的性能出现了更大程度的限制,相比 CPLEX 限时解有 28%~41% 的不足,并大幅度落后于 MTRL。而 MTRL 和 PNRL 在所有测试场景中均表现出优异的实时求解能力,能够在毫秒级时间内生成高质量可行解。进一步对比发现,MTRL 在所有场景下的平均奖励值均高于 PNRL,表明其多层多头注意力机制能够更充分地捕捉多无人机状态与目标点分布的全局特征,从而实现更优的决策效果。此外,MTRL 的计算效率几

乎不受问题规模增加的影响,在小规模、中规模和大规模场景下都能保持毫秒量级。综合分析,MTRL算法在路径优化效果上表现出以下3个显著优势。

(1)解质量优势:在全部9个测试场景中,MTRL相比CPLEX的平均累计奖励提升9.77%~28.77%(中小规模)和9.34%~14.84%(大规模)。

(2)计算效率优势:MTRL求解时间保持在毫秒级,而CPLEX随问题规模增长(至100个目标点),求解单个案例需约120 s。相比之下,MTRL的计算时间随问题规模增长缓慢,展示出显著的效率优势。

(3)规模适应性:随着目标点数量由T20增至T100(增加5倍),MTRL始终能够稳定生成高质量解,且在3种无人机配置(A2、A3、A4)下表现一致,验证了MTRL在不同规模和不同智能体数量场景下的普适性。这些结果表明,所提出的MTRL能够快速找到多无人机协同交通监控管理问题的高质量可行解,且不受问题规模的影响,为实际应用中的多无人机交通监控任务提供了可靠的技术支撑。

MTRL的计算效率优势主要体现在推理阶段。通过一次离线训练,在实际应用的推理阶段,MTRL的时间复杂度主要取决于Transformer模型的前向传播过程。编码器通过多头注意力机制并行处理所有节点特征,其时间复杂度为 p^2 ,其中 p 为节点总数。解码器采用自回归方式生成路径序列,每步通过多头交叉注意力机制从候选节点中选择下一个访问节点,时间复杂度为 pq ,其中 q 为序列长度。MTRL得益于Transformer架构的高效并行计算和GPU加速。相比之下,CPLEX求解整数规划模型的计算时间随问题规模增加指数增长,而MTRL在相同场景下极大缩短了推理时间。这种快速推理的特性使得MTRL非常适合城市交通监控等需要实时响应的应用场景,一次离线训练后,模型可在毫秒级时间内为新的任务实例生成高质量路径规划方案,满足实际应用对计算速度的要求。

3.5 MTRL分布泛化能力评估

为评估MTRL对不同节点分布模式的泛化能力,本文设计了分布泛化试验。在训练阶段,除针对所有场景设置下对使用均匀分布在边长为单位1的正方形区域内生成的节点数据进行训练外,还采用正态分布 $N(0.5, 0.35^2)$ 生成节点位置进行额外训练。正态分布的均值设置为0.5,确保大部分节点仍位于正方形区域内;标准差设为0.35,使得节点呈现明显的聚集分布特征,与训练时的均匀分布形成显

著对比。在测试阶段,分别采用均匀分布训练的模型和采用正态分布训练的模型,在对方场景下进行交叉测试。测试场景包括T20B2A2、T20B2A3、T20B2A4、T50B5A2、T50B5A3、T50B5A4、T100B10A2、T100B10A3和T100B10A4,共九类配置,每类配置下生成256个均匀分布和正态分布实例。性能评估指标为相对误差百分比,计算公式为

$$\eta = \frac{|\bar{R}_t - \bar{R}_b|}{\bar{R}_b} \quad (30)$$

式中: η 为相对误差; \bar{R}_t 为交叉测试实例的平均奖励值; \bar{R}_b 为原始方法训练场景的平均奖励值。

表3为不同节点分布模式的性能对比,表中: M_u 为均匀分布模型; M_n 为正态分布模型; T_u 为均匀分布实例集; T_n 为正态分布实例集; η_u 为使用均匀分布模型时,正态分布测试相对于均匀分布测试的误差; η_n 反之。 $M_u \rightarrow T_u$ 表示均匀分布模型在均匀分布实例上的测试, $M_n \rightarrow T_n$ 表示正态分布模型在正态分布实例上的测试, $M_u \rightarrow T_n$ 表示均匀分布模型在正态分布实例上的交叉测试, $M_n \rightarrow T_u$ 表示正态分布模型在均匀分布实例上的交叉测试。可见,MTRL模型在所有12个交叉测试场景下均展现出优异分布泛化能力,相对误差均控制在1%以内。其中,T20场景下的相对误差为0~0.0978%,T50场景的相对误差为0.1563%~0.5567%,T100场景的相对误差为0.1670%~0.5807%。这表明,尽管节点分布模式发生了根本性变化(从均匀分布到正态分布),训练好的MTRL模型仍能保持稳定的求解性能。该试验结果充分验证了MTRL的分布泛化能力,说明其能够适应实际应用中多样化的任务场景,为算法在真实交通监控环境中的部署提供了可能性。

图7分别展示了2~4架无人机协同交通监控管理问题在9个场景下,示例路径可视化结果。可以看出:生成的路径方案均可行,不存在重复访问和子回路现象。在小规模案例下(T20),由于时间限制较为宽松,所有目标点均被访问;在中、大规模案例下(T50、T100)被访问目标点的总数伴随着UAV架数的增加而增加。

从路径可视化结果可观察到以下主要特征与优化效果。

(1)无冲突协同:在所有场景中,不同无人机的路径均未出现对同一目标点的重复访问,验证了掩码机制在避免智能体冲突方面的有效性。

(2)合理的换电策略:无人机根据自身电量和任

表 3 不同节点分布模式的性能对比

Table 3 Performance comparison of different node distribution modes

问题规模	$M_u \rightarrow T_u$	$M_u \rightarrow T_n$	$\eta_u/\%$	$M_n \rightarrow T_u$	$M_n \rightarrow T_n$	$\eta_n/\%$
T20B2A2	19.996 1	19.980 5	0.078 0	19.925 8	19.945 3	0.097 8
T20B2A3	20.000 0	20.000 0	0.000 0	20.000 0	20.000 0	0.000 0
T20B2A4	20.000 0	20.000 0	0.000 0	20.000 0	20.000 0	0.000 0
T50B5A2	35.101 6	34.906 2	0.556 7	34.207 0	34.105 5	0.297 6
T50B5A3	43.750 0	43.664 1	0.196 3	43.363 3	43.183 5	0.416 3
T50B5A4	49.539 1	49.441 4	0.197 2	49.164 1	49.087 4	0.156 3
T100B10A2	47.871 8	47.593 8	0.580 7	47.085 7	46.872 3	0.455 3
T100B10A3	60.518 1	60.257 5	0.430 6	58.473 4	58.375 9	0.167 0
T100B10A4	71.026 7	70.753 5	0.384 6	71.009 8	70.635 7	0.529 6

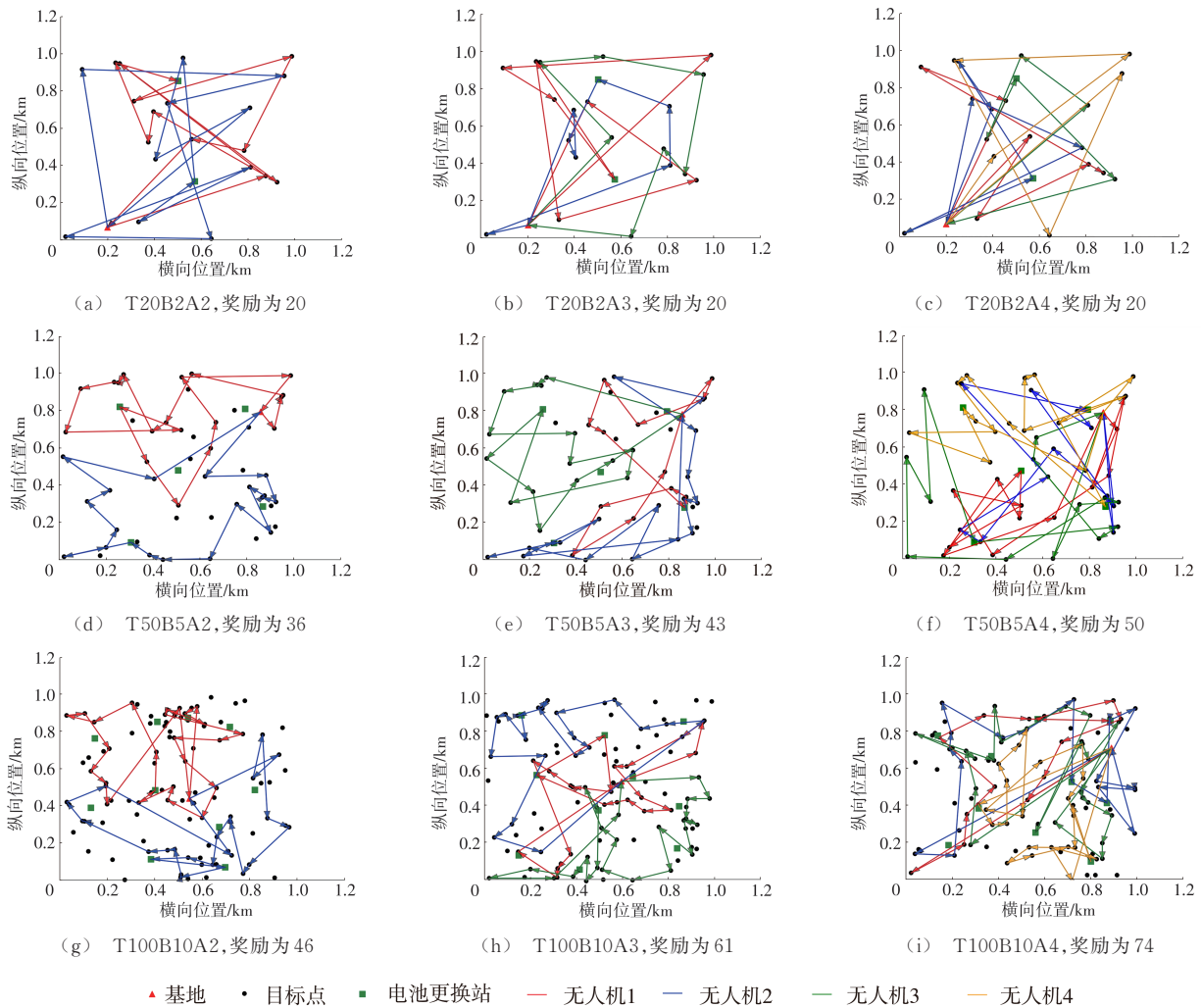


图 7 多无人机路径规划结果可视化示例

Fig. 7 Visualization example of multi-UAV path planning results

务需求,自主选择合适时机访问电池更换站进行补给;在中、大规模场景中,无人机多次访问电池更换站,每次换电后继续执行监控任务,显著扩展了任务覆盖范围;电池更换站的空间分布较为均匀,有效支撑了无人机的长航时作业。

(3)路径可行性验证:所有生成的路径方案均满

足所有约束,未出现违规访问或能量不足情况,表明MTRL算法在约束满足与可行解生成方面具有良好的稳定性。

(4)覆盖效率提升:随着无人机数量的增加,系统总覆盖率稳步提升;相比双机方案,三机与四机配置在中、大规模场景下能够访问更多目标点,体现出

多智能体协同机制在提升监控覆盖效率与空间利用率方面的显著效果。

为进一步验证MTRL算法在真实城市交通监控场景中的适用性,本文在上述泛化试验基础上,测试3种真实场景的复杂分布模式,分别为目标点密集分布(模拟城市核心区)、目标点稀疏分布(模拟城郊高速路网)和差异化奖励分布(采用前文介绍的均匀分布奖励权重、距离相关奖励权重2种差异化奖励区分目标点重要性)。使用在均匀分布下训练得到的MTRL模型在上述3种场景中进行测试,每种场景随机生成256个T50B5A3实例和256个T100B10A4实例。试验结果表明,MTRL算法在3种真实场景下均保持良好性能。目标点密集场景覆盖率达91.77%(T50B5A3)和75.59%(T100B10A4)。目标点稀疏场景覆盖率达82.07%(T50B5A3)和68.38%(T100B10A4)。差异化奖励分布场景中,均匀分布奖励权重覆盖率达84.39%(T50B5A3)和

70.85%(T100B10A4),距离相关奖励权重覆盖率达82.52%(T50B5A3)和69.00%(T100B10A4)。在不同类型的真实场景中,MTRL算法均展现出智能化的目标点优先级处理能力,实现优先访问高奖励目标点,符合实际应用中资源受限时优先保障重点区域的真实需求。此外,所有场景的计算时间均小于1s,满足计算性能的实时性要求。

此外,本文选取真实城市区域对MTRL算法性能进行测试,将研究区域内重点交叉口抽象为目标点,应用KMC聚类方法确定电池更换站的位置,使用训练好的均匀分布MTRL模型对该真实案例进行测试,结果如图8所示。可见,MTRL算法在2、3、4架无人机的3种场景下均求解出了目标点高覆盖率的高质量可行解。本试验进一步验证了MTRL算法的实际应用价值。以上结果证明了MTRL算法从虚拟仿真迁移到实际应用的能力,充分验证了MTRL算法从仿真到真实场景部署的可行性。

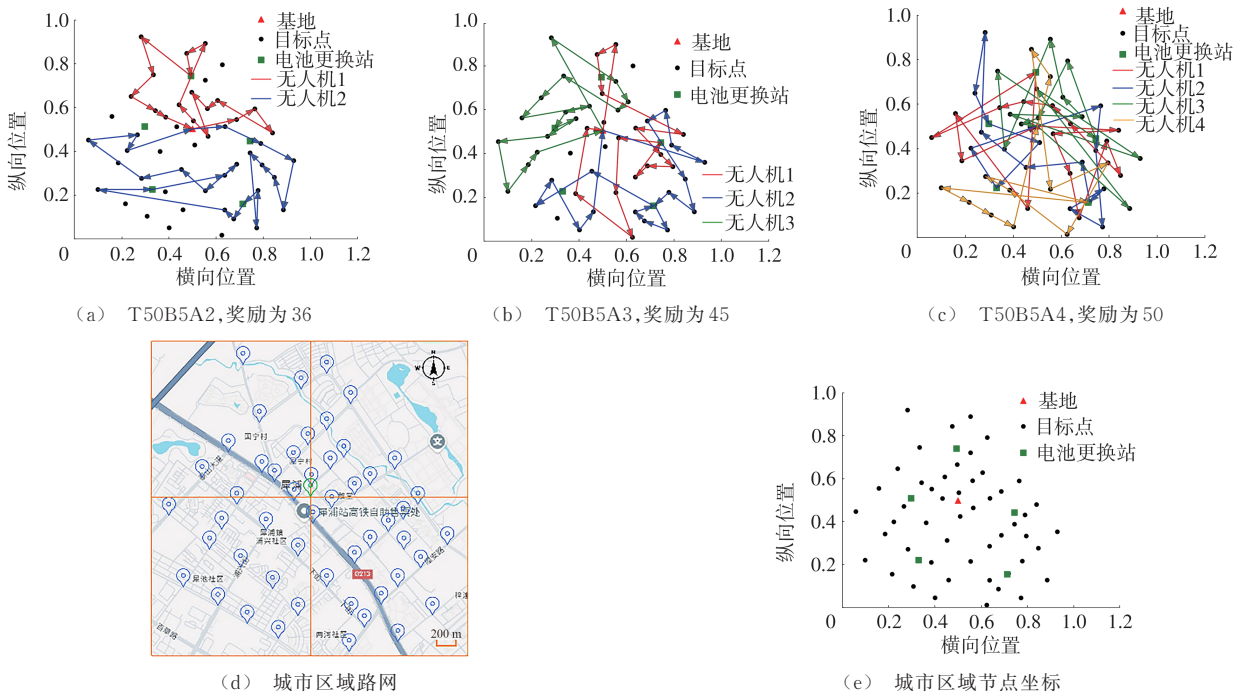


图8 城市道路交通监控结果

Fig. 8 Results of urban road traffic control

4 结语

(1)本文针对带有电池更换站的多无人机协同交通监控管理问题,构建了团队定向问题的混合整数线性规划模型,并提出了基于一种基于多智能体Transformer的深度强化学习算法。该方法采用编码器-解码器架构:编码器负责学习场景的图结构表示,解码器根据各个智能体的状态生成协调的路径

决策。通过引入多头注意力机制和掩码机制,MTRL不仅能够捕获全局特征,还能确保解的可行性,有效避免子回路、重复访问和智能体冲突。与传统数学规划模型的求解方式不同,MTRL利用“状态-动作-奖励”的逐步决策方式,充分考虑全局环境信息,从而构建完整的路径规划方案。

(2)数值试验和对比分析验证了MTRL算法的优越性能。在不同规模场景下,MTRL均能获得高

质量可行解,求解精度和计算效率显著优于传统精确算法。特别是在时间限制条件下,MTRL的解质量明显优于CPLEX。进一步的分布泛化试验表明,MTRL在跨分布测试中依然保持稳定性能:均匀分布训练的模型在正态分布测试案例以及使用正态分布训练的模型在均匀分布测试案例中的相对误差均控制在1%以内,证明了算法对不同节点分布模式的迁移性和适应性,无需额外训练即可迁移应用。尽管MTRL在多无人机协同交通监控路径优化问题上表现出优异性能,但在实际应用中仍需进一步考虑电池更换站容量限制、动态环境变化以及超大规模场景下的计算效率等问题,为后续研究提供了改进方向。

(3)未来的研究将聚焦以下方向:一是面向持续监控任务,针对监控时长超过单架无人机最大续航能力的目标点,探索通过引入更多智能体实现长期覆盖的策略;二是探索动态路径规划问题,考虑天气、空域管制等实时变化因素对弧长的影响,并结合在线微调机制提升模型的动态适应能力;三是引入电池更换站容量约束与排队模型,研究排队等待时间对路径优化的影响;四是提供面向实际部署的系统性指导,包括在目标点密度高的城市核心区优先布设电池更换站等策略;五是大规模协同机制扩展,进一步扩展问题规模,评估集中式方法在处理大量节点和智能体时的性能边界,尝试集中式与分布式决策机制相结合,以突破集中式方法在超大规模场景中的潜在局限性,提升系统的鲁棒性与适应性。

参考文献:

Reference:

- [1] LIU S, BAI Y B. Multiple UAVs collaborative traffic monitoring with intention-based communication[J]. *Computer Communications*, 2023, 210: 116-129.
- [2] 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.
- [3] COIFMAN B. Improved velocity estimation using single loop detectors [J]. *Transportation Research Part A: Policy and Practice*, 2001, 35(10): 863-880.
- [4] KOUTSIA A, SEMERTZIDIS T, DIMITROPOULOS K, et al. Intelligent traffic monitoring and surveillance with multiple cameras[C]//IEEE. 2008 International Workshop on Content-Based Multimedia Indexing. New York: IEEE, 2008: 125-132.
- [5] CAO P, XIONG Z Q, LIU X B. An analytical model for quantifying the efficiency of traffic-data collection using instrumented vehicles [J]. *Transportation Research Part C: Emerging Technologies*, 2022, 136: 103558.
- [6] VANDENBERGHE W, VANHAUWAERT E, VERBRUGGE S, et al. Feasibility of expanding traffic monitoring systems with floating car data technology[J]. *IET Intelligent Transport Systems*, 2012, 6(4): 347-354.
- [7] SEO T, KUSAKABE T, ASAKURA Y. Estimation of flow and density using probe vehicles with spacing measurement equipment [J]. *Transportation Research Part C: Emerging Technologies*, 2015, 53: 134-150.
- [8] LI X, SHU W, LI M L, et al. Performance evaluation of vehicle-based mobile sensor networks for traffic monitoring [J]. *IEEE Transactions on Vehicular Technology*, 2009, 58(4): 1647-1653.
- [9] HUANG P D, CHENG M, CHEN Y P, et al. Traffic sign occlusion detection using mobile laser scanning point clouds [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(9): 2364-2376.
- [10] JIAO J F, WANG H H. Traffic behavior recognition from traffic videos under occlusion condition: A Kalman filter approach[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2022, 2676(7): 55-65.
- [11] SONG X G, PI R D, LV C, et al. Augmented multiple vehicles' trajectories extraction under occlusions with roadside LiDAR data [J]. *IEEE Sensors Journal*, 2021, 21(19): 21921-21930.
- [12] ZHAO J X, XU H, ZHANG Y B, et al. Automatic identification of vehicle partial occlusion in data collected by roadside LiDAR sensors[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2022, 2676(5): 708-718.
- [13] LI M, ZHEN L, WANG S A, et al. Unmanned aerial vehicle scheduling problem for traffic monitoring [J]. *Computers & Industrial Engineering*, 2018, 122: 15-23.
- [14] LI S G, YU H K, ZHANG J R, et al. Video-based traffic data collection system for multiple vehicle types [J]. *IET Intelligent Transport Systems*, 2014, 8(2): 164-174.
- [15] HUANG H L, SAVKIN A V, HUANG C. Decentralized autonomous navigation of a UAV network for road traffic monitoring[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2021, 57(4): 2558-2564.
- [16] 马庆禄,王欣宇,张书,等. 智能网联环境下近邻匝道交通耦合自组织方法[J]. *交通运输工程学报*, 2024, 24(2): 207-220.
MA Qing-lu, WANG Xin-yu, ZHANG Shu, et al. Self-organizing method for traffic coupling between adjacent ramps in intelligent and connected environments[J]. *Journal of Traffic and Transportation Engineering*, 2024, 24(2): 207-220.
- [17] 谢济铭,夏玉兰,钱正富,等. 考虑智能网联近邻车辆信息的交织区换道风险预警[J]. *交通运输工程学报*, 2023, 23(2): 287-300.

- XIE Ji-ming, XIA Yu-lan, QIAN Zheng-fu, et al. Lane-change risk warning in interweaving area considering information from intelligent connected near-neighboring vehicles[J]. *Journal of Traffic and Transportation Engineering*, 2023, 23(2): 287-300.
- [18] YAN H, CHEN Y F, YANG S H. UAV-enabled wireless power transfer with base station charging and UAV power consumption[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(11): 12883-12896.
- [19] COELHO B N, COELHO V N, COELHO I M, et al. A multi-objective green UAV routing problem[J]. *Computers & Operations Research*, 2017, 88: 306-315.
- [20] XU W Z, XU Z C, PENG J, et al. Approximation algorithms for the team orienteering problem [C] //IEEE. *INFOCOM 2020 -IEEE Conference on Computer Communications*. New York: IEEE, 2020: 1389-1398.
- [21] JUAN A A, MARUGAN C A, AHSINI Y, et al. Using reinforcement learning to solve a dynamic orienteering problem with random rewards affected by the battery status [J]. *Batteries*, 2023, 9(8): 416.
- [22] AMMOURIOVA M, GUERRERO A, TSERTSVADZE V, et al. Using reinforcement learning in a dynamic team orienteering problem with electric batteries [J]. *Batteries*, 2024, 10(12): 411.
- [23] LEE J J, RATHINAM S. Team orienteering and scheduling algorithms for collaborative UAV-UGV area coverage with battery constraints[C]//IEEE. *2025 International Conference on Unmanned Aircraft Systems (ICUAS)*. New York: IEEE, 2025: 625-632.
- [24] 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.
- QIN Wen-long, LUO He, LI Xiao-duo, et al. Multi-UAV emergency power inspection path planning method considering multiple charging stations[J]. *Control and Decision*, 2025, 40(8): 2391-2399.
- [25] 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.
- [26] NOVOA C, STORER R. An approximate dynamic programming approach for the vehicle routing problem with stochastic demands [J]. *European Journal of Operational Research*, 2009, 196(2): 509-515.
- [27] KIRÁLY A, ABONYI J. Redesign of the supply of mobile mechanics based on a novel genetic optimization algorithm using Google Maps API [J]. *Engineering Applications of Artificial Intelligence*, 2015, 38: 122-130.
- [28] FUERTES D, DEL-BLANCO C R, JAUREGUIZAR F, et al. TOP-former: A multi-agent transformer approach for the team orienteering problem [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2025, 26(9): 13799-13810.
- [29] BAI L H, ZHENG F F, HOU K N, et al. Longitudinal control of automated vehicles: A novel approach by integrating deep reinforcement learning with intelligent driver model [J]. *IEEE Transactions on Vehicular Technology*, 2024, 73(8): 11014-11028.
- [30] 张洪海, 夷 珈, 李 珊, 等. 低空空域容量评估研究综述[J]. *交通运输工程学报*, 2023, 23(6): 78-93.
- ZHANG Hong-hai, YI Jia, LI Shan, et al. Review on research of low-altitude airspace capacity evaluation [J]. *Journal of Traffic and Transportation Engineering*, 2023, 23(6): 78-93.
- [31] 李诚龙, 屈文秋, 李彦冬, 等. 面向 eVTOL 航空器的城市空中运输交通管理综述[J]. *交通运输工程学报*, 2020, 20(4): 35-54.
- LI Cheng-long, QU Wen-qiu, LI Yan-dong, et al. Overview of traffic management of urban air mobility (UAM) with eVTOL aircraft [J]. *Journal of Traffic and Transportation Engineering*, 2020, 20(4): 35-54.
- [32] 刘 伟, 钟 灿, 曹文明. 基于数据驱动的路网连续交通流短时预测方法综述[J]. *交通运输工程学报*, 2026, 26(2): 24-43.
- LIU Wei, ZHONG Can, CAO Wen-ming. Review of data-driven short-term prediction methods for continuous traffic flow in road networks [J]. *Journal of Traffic and Transportation Engineering*, 2026, 26(2): 24-43.
- [33] LIN B, GHADDAR B, NATHWANI J. Deep reinforcement learning for the electric vehicle routing problem with time windows [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(8): 11528-11538.
- [34] KOOL W, VAN HOOFF H, WELLING M. Attention, learn to solve routing problems! [C]//ICLR. *7th International Conference on Learning Representations*. Washington DC: ICLR, 2019: 39.
- [35] REN L, FAN X Y, CUI J, et al. A multi-agent reinforcement learning method with route recorders for vehicle routing in supply chain management [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(9): 16410-16420.
- [36] FAN M F, WU Y X, LIAO T J, et al. Deep reinforcement learning for UAV routing in the presence of multiple charging stations [J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(5): 5732-5746.
- [37] ZHANG K, HE F, ZHANG Z C, et al. Multi-vehicle routing problems with soft time windows: A multi-agent reinforcement learning approach [J]. *Transportation Research Part C: Emerging Technologies*, 2020, 121: 102861.
- [38] VINYALS O, FORTUNATO M, JAITLY N. Pointer networks [J]. *Advances in Neural Information Processing Systems*, 2015, 28: 2692-2700.
- [39] CALINSKI T, HARABASZ J. A dendrite method for cluster analysis [J]. *Communications in Statistics-Simulation and Computation*, 1974, 3(1): 791519860.