基于模型与图强化学习驱动的车辆-无人机协同多目标路径优化方法

杨明园，王 伟†

(哈尔滨工程大学 智能科学与工程学院，哈尔滨 150001)

摘 要: 为了提高物流配送效率, 研究一种考虑工作量均衡的车辆-无人机协同路径问题. 为了求解该问题, 首先以运营成本与车辆-无人机编队工作量均衡建立双目标混合整数线性规划模型. 其次, 提出基于模型与图强化学习驱动的多目标优化方法. 第一, 提出基于混合策略的种群初始化方法和多个局部搜索算子以有效探索解空间; 第二, 提出增强强化学习的帕累托局部搜索算法并将其作为多目标问题的局部搜索算法以进一步提高多目标方法的搜索能力. 其中包括基于图卷积神经网络的特征提取机制和基于长短期记忆网络的策略优化方法. 特征提取机制通过捕捉车辆-无人机路径方案的空间关系, 为智能体决策增加状态表征信息; 策略优化方法通过构建交互环境模型并推演其多步虚拟轨迹提高智能体的训练样本效率. 最后, 通过参数分析和对比实验证实所提模型和算法的有效性以及算法的收敛性和分布性优于精确求解器 CPLEX 和多个先进算法.

关键词: 低空经济；车辆-无人机路径问题；工作量均衡；多目标进化算法；强化学习；图神经网络；长短期记忆网络

中图分类号: TP312 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2025.1120

引用格式: 杨明园, 王伟. 基于模型与图强化学习驱动的车辆-无人机协同多目标路径优化方法 [J]. 控制与决策.

Model and graph-based reinforcement learning driven-multi-objective evolutionary algorithm for vehicles-drone cooperative routing problem

YANG Ming-yuan，*WANG Wei*†

(College of Intelligent Systems Science and Engineering，Harbin Engineering University，Harbin，150001，China)

Abstract: To improve logistics efficiency, we examines a vehicles–drone cooperative routing problem with workload balance. A mixed-integer linear programming model is formulated to minimize both cost and workload imbalance of vehicles–drone. A model and graph-based reinforcement learning-driven multi-objective method is proposed to solve the routing problem. First, the method incorporates a hybrid strategy-based population initialization approach and customizes local search operators to efficiently explore the solution space. Second, a pareto local search algorithm incorporating reinforcement learning is proposed as a local search approach for the problem, thereby enhancing the multi-objective method's local search ability. The feature extraction mechanism captures spatial patterns in routing, and the policy method employs multi-step virtual trajectory to enhance state information and sample efficiency. Finally, through parameter calibration and comparative experiments, the validity of proposals are confirmed, demonstrating that the algorithm outperforms the CPLEX and several state-of-the-art competitors in proposal bi-objective model.

Keywords: low-altitude economy； vehicle-drone routing problem； workload balance； multi-objective evolutionary algorithm； reinforcement learning； graph neural network； long short-term memory

0 引 言

在低空经济快速发展的背景下, 凭借低碳, 低成本, 快速机动等优势, 无人机成为低空智能运输 (LAIT) 的核心组成部分[1]. 然而, 受限于载荷, 航程与抗恶劣天气能力, 无人机难以独立承担大容量和长距离的运输任务[2]. 与之互补的是, 车辆拥有大载量和长续航的优势, 但其灵活性受制于交通状况[3]. 鉴于无人机与车辆在运输能力上呈现互补性, 二者

收稿日期: 2025-10-27；**录用日期:** 2026-02-22.
基金项目: 国家自然科学基金项目 (62271163).
责任编委: 唐加福.
†通信作者. E-mail: wangwei407@hrbeu.edu.cn.

的协同被视为提升物流效率的潜力方案.

在此背景下, Murray 等人开创性地提出车辆-无人机协同路径问题 (VRPD), 将车辆作为无人机的移动起降平台与补给站, 拓展了物流服务的覆盖范围与灵活性[4]. 此后, 众多学者致力于发展 VRPD 变体以应对更复杂的现实需求. 例如, 引入时间窗约束的 VRPDTW 问题[5]; 满足客户同时取货与送货需求的 VRPDPD 问题[6]; 突破单车运力瓶颈的多车-多无人机协同问题 (MVRPD)[7] 等. 尽管现有研究不断推动 VRPD 领域发展, 但一个关键问题被普遍忽视: 不公平的任务分配使得车辆-无人机编队间的工作负荷是非均衡的, 尤其在大规模运输场景中, 不同编队在前往服务区的运输时间和区域内的服务时间上存在显著差异[8]. 公平的工作负荷分配能够提升员工满意度, 改善服务质量等非货币效益[9]. 因此, 编队工作负荷分配是一个亟待优化的重要目标. 同时, 鉴于经济效益是物流优化的首要动机, 运营成本是本文另一个优化目标. 综上所述, 本文研究一种考虑取货, 送货和时间窗口的多车辆-无人机路径问题 (MVRPDPDTW), 旨在构建最小化运营成本与编队间工作量均衡的双目标路线方案.

传统求解路径规划问题的多目标进化方法由于缺乏对搜索状态的感知, 多采用随机或固定序列的局部搜索算子选择策略, 存在收敛缓慢和性能受限问题[10-14]. 基于强化学习 (RL) 的多目标方法通过学习与局部搜索算子交互所获取的奖励反馈, 在后续迭代过程中能倾向选择产生更优后代的搜索算子, 解决上述问题[5, 15-16]. 虽然该方法在求解路径问题中取得了显著成果, 但其多采用宏观特征 (如车辆数目和客户位置等静态信息, 选定操作符以及目标值变化[12-14]) 作为智能体的状态表示内容, 忽略了路径方案的空间拓扑结构. 值得注意的是, 该结构蕴含着服务客户的次序与时间等细节特征, 能反映优化过程的实时进展. 因此, 这些特征的缺失使得智能体尚未得到充分的决策信息, 进而难以有效引导搜索进程. 此外, 现有方法多使用无模型 RL (如 Q-学习, 深度 Q-网络 (DQN), 近似策略优化[11-14, 17]) 指导算子选择. 该类方法通常需要大量试错样本以充分训练智能体. 而每次采样, 多目标方法仅能反馈有限样本, 且采样所需执行的局部搜索操作计算成本昂贵. 这使得训练智能体存在数据匮乏和采样效率低, 进而难以有效指导多目标方法的局部搜索算子选择.

为此, 提出基于模型和图强化学习驱动的多目标优化方法, 通过增加智能体状态信息和提高其训练样本效率进一步提升多目标方法的搜索能力. 模型驱动包含双目标优化模型驱动和基于模型的强化学习驱动两个层面. 算法在优化层面由显式优化机理模型约束搜索空间, 在学习层面由基于模型的强化学习机制共同驱动搜索决策. 首先, 为有效探索解空间, 基于问题特征设计混合随机和贪婪策略的种群初始化方法以及多种邻域结构的局部搜索算子. 其次, 为有效指导搜索算子的选择, 提出增强强化学习的帕累托局部搜索算法 (PLS), 并将其作为多目标方法的局部搜索机制. 其中包括基于图卷积神经网络 (GCN) 的特征提取机制和基于长短期记忆网络 (LSTM) 的 DQN 策略优化方法. 特征提取机制为决策提供丰富的信息状态表征. 具体而言, 将车辆-无人机路径方案转化为图邻接矩阵与特征矩阵后, 运用 GCN 为每个图节点生成高维嵌入向量, 以同时捕捉局部特征与全局结构信息. 随后通过池化技术聚合节点嵌入, 构建路径方案的综合表征. 策略优化方法为智能体提供充分的训练数据. 具体而言, 在采集 DQN 与 PLS 的交互样本后, 通过训练 LSTM 网络构建交互环境模型并推演其多步虚拟轨迹. 随后结合真实交互样本和虚拟推演样本优化智能体的学习策略, 提升样本利用效率.

本文主要内容如下: (1) 建立双目标混合整数线性规划模型以最小化运营成本与车辆-无人机编队工作量失衡. (2) 结合 MVRPDPDTW 的问题特征, 定制基于混合策略的种群初始化方法与多个 PLS 搜索算子以增强解空间的搜索能力. (3) 提出基于 GCN 的特征提取机制, 通过捕捉路径方案中的空间关系, 为智能体决策提供丰富状态表征. (4) 提出基于 LSTM 的 DQN 策略优化方法, 通过推演生成多步虚拟轨迹提高智能体训练样本效率, 进一步有效指导搜索算子选择. (5) 通过参数分析和对比实验证实所提模型和算法的有效性, 算法在优化所提双目标上优于精确求解器 CPLEX 和针对此类问题的多个先进算法.

1 问题描述与建模

本节将 MVRPDPDTW 问题建模为双目标混合整数线性规划模型 (MILP), 同时优化运营成本与编队间工作量失衡. 为便于阅读, 表 1 给出了符号说明.

1.1 问题描述

如图 1 所示, 有 1 个仓库和 19 个客户. 参考配送流程[18-19], 车辆携带无人机从仓库出发, 并前往客户地点执行取件和配送操作. 与无人机不同, 车辆可服务于含重型包裹的客户. 此外, 在仓库 (节点 0 和节点$c+1$) 和客户节点, 车辆需要决定是否发射或

表1　符号说明

符号	说明
	集合和参数
C	客户集合, $C=\{1,...,c\}$.
F	车辆集合, $F=\{1,...,f\}$.
D	车辆携带的无人机集合, $D=\{1,...,d\}$.
N	所有节点集合, $N=C\cup\{0,c+1\}$.
p^D	无人机最大携带包裹重量.
$d^T_{i,j}$	车辆行驶边(i,j)所需的曼哈顿距离.
$d^D_{i,j}$	无人机飞越边(i,j)所需的欧几里得距离.
M	足够大的正值.
$t^{DF}_{i,j}$	无人机飞越边(i,j)所需的飞行时间.
$t^{TF}_{i,j}$	车辆行驶边(i,j)所需的驾驶时间.
$[a_i,b_i]$	客户节点i的服务时间窗口.
s^D_i	无人机在客户节点i的服务时间.
s^T_i	车辆在客户节点i的服务时间.
e^D	无人机的电池能耗率.
w^D_i	客户节点i处无人机携带的包裹重量.
B^D	无人机的最大电池容量.
H	车辆-无人机编队最大工作时长.
	变量
$x_{i,j,f,d}$	若第d架无人机经过边(i,j)并返回第f辆车辆, 则为1, 否则为0.
$y_{i,j,f}$	若第f辆车辆经过边(i,j), 为1, 否则为0.
$z^D_{i,f,d}$	若第f辆车辆的第d架无人机为客户节i提供服务, 则为1, 否则为0.
$z^T_{i,f}$	若第f辆车辆为客户节i提供服务, 则为1, 否则为0.
h^L_i	若节点i是无人机的起飞节点, 则为1, 否则为0.
h^R_i	若节点i是无人机的返回节点, 则为1, 否则为0.
$t^{DA}_{j,f,d}$	第f辆车辆第d架无人机抵达节点j的时间.
$t^{TA}_{j,f}$	第f辆车辆抵达节点j的到达时间.
$t^L_{i,f}$	第f辆车辆在节点i的离开时间.
$t^R_{i,f}$	第f辆车辆抵达返回节点i的时间.
$E^D_{i,d}$	第d架无人机离开节点i时的剩余能量.

回收无人机. 若无人机未被发射, 车辆将其直接运输至下一个节点. 相反, 无人机需要遵守能耗与载荷限制, 在完成客户服务后返回车辆更换电池或包裹, 等待下次发射. 同时, 受交通和天气等不确定因素影响, 车辆与无人机难以同时抵达回收节点. 因此, 率先抵达回收节点的车辆或无人机需等待其他车辆. 最后, 在此问题设定上, 为了减少图 1 中双编队完成运输任务的时间差距, 建立如下双目标模型-运营成本和编队工作量失衡, 确定车辆-无人机协同路线方案.

为了简化问题, 假设如下: (1) 无人机与车辆在回收点处, 允许一方先到并等待另一方以完成汇合; (2) 无人机的发射、回收、维护 (更换电池和包裹) 时间忽略不计; (3) 车辆拥有足够的载货容量, 燃料与电池以完成配送任务; (4) 本文问题为静态规划问题.

1.2　数学模型

本节将 MVRPDPDTW 问题建模为双目标优化模型. 从利润和公平性角度出发, 同时最小化运营成本与车辆-无人机编队工作量失衡具有实际意义. 因此, 本文提出双目标F_1和F_2既反映编队路线支出, 固定部署成本和超时等成本, 又兼顾各编队任务完成时间的差距. 模型具体公式如下:

$$F_1=C_1+C_2+C_3+C_4, \tag{1}$$

$$F_2=\max(t^R_{0,f}-t^R_{0,f'}), f\in F, f'\in F, f\neq f', \tag{2}$$

F_1为编队路径成本C_1, 启动成本C_2, 违反时间窗成本C_3与车辆装卸成本C_4之和. F_2表示不同编队完成时间的最大差值, 用编队完成任务的总作业时间$t^R_{0,f}$衡量.

$$\mu^{f,d}_j=\max(a_j,b_j,t^{DA}_{j,f,d})-\min(a_j,b_j,t^{DA}_{j,f,d})+ a_j-b_j, j\in C, d\in D, f\in F, \tag{3}$$

$$C_1=c_1\sum_{i\in N}\sum_{j\in N}\sum_{f\in F}y_{i,j,f}d^T_{i,j}+ c_2\sum_{i\in N}\sum_{j\in N}\sum_{f\in F}\sum_{d\in D}x_{i,j,f,d}d^D_{i,j}, \tag{4}$$

$$C_2=c_3\sum_{j\in N}\sum_{f\in F}y_{0,j,f}, \tag{5}$$

$$C_3=c_4(\sum_{i\in N}\sum_{j\in N}\sum_{f\in F}y_{i,j,f}\mu^f_j+ \sum_{i\in N}\sum_{j\in N}\sum_{f\in F}\sum_{d\in D}x_{i,j,f,d}\mu^{f,d}_j), \tag{6}$$

$$C_4=c_5\sum_{f\in F}t^R_{0,f}, \tag{7}$$

在式 (3) 中, $\mu^{f,d}_j$是因车辆f的无人机d提前或延迟服务客户j所产生的成本. μ^f_j为车辆f违反客户j时间窗口的成本. c_1, c_2, c_3, c_4和c_5分别对应车辆单位距离行驶成本, 无人机单位距离运行成本, 每编队固定启动成本以及违反时间窗口成本和每小时车辆装卸费. 随后, 模型约束条件被归纳为以下三类:

1.2.1　路径约束

$$\sum_{f\in F}\sum_{d\in D}z^D_{i,f,d}+z^T_{i,f}=1,\ \forall i\in C, \tag{8}$$

$$z^T_{i,f}=\sum_{j\in N}y_{j,i,f},\ \forall i\in C,\ \forall f\in F, \tag{9}$$

$$\sum_{j\in N}y_{j,i,f}=\sum_{j\in N}y_{i,j,f},\ \forall i\in C,\ \forall f\in F, \tag{10}$$

$$\sum_{j\in C}y_{0,j,f}=1,\ \forall f\in F, \tag{11}$$

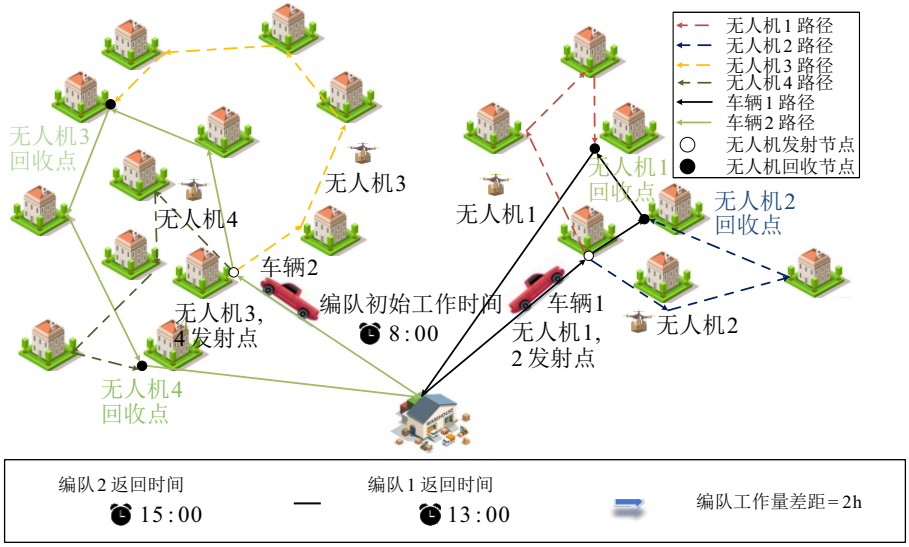


图1 编队工作量失衡的车辆-无人机路径问题示意图

$$\sum_{i\in C} y_{i,c+1,f}=1,\ \forall f\in F, \tag{12}$$

$$z^{D}_{i,f,d}=\sum_{j\in N} x_{j,i,f,d},\forall i\in C,\forall f\in F,\forall d\in D, \tag{13}$$

$$\sum_{i\in N} x_{i,j,f,d}=\sum_{i\in N} x_{j,i,f,d},\ \forall f\in F,\ \forall d\in D,\ \forall j\in C, \tag{14}$$

$$\sum_{i\in C} h^{L}_{i}=\sum_{i\in C} h^{R}_{i}, \tag{15}$$

约束 (8) 规定每个客户被服务一次, 服务方式可为无人机或车辆. 约束 (9) 限制每辆车仅服务其行驶路线中的客户. 约束 (10) 确保每条车辆行驶路径的流量平衡. 约束 (11) 规定每辆车仅从仓库出发一次, 并直接驶往客户. 约束 (12) 规定每辆车仅从客户回到仓库一次. 约束 (13) 确保若无人机d服务路线中的客户i, 则其需要实际飞抵该客户. 约束 (14) 维持每条无人机航线的流量平衡. 约束 (15) 确保无人机起飞节点数与返回节点数相等, 从而维持操作一致性.

1.2.2 时间约束

$$M(1-y_{i,j,f})\geq t^{TA}_{j,f}-t^{TF}_{i,j}-t^{L}_{i,f},\ \forall i\in C,\forall j\in N,\forall f\in F, \tag{16}$$

$$M(2-x_{i,j,f,d}-y_{i,j,f})\geqslant t^{DA}_{j,f,d}-t^{TF}_{i,j}-t^{L}_{i,f},\ \forall i\in C,\forall j\in N,\forall f\in F,\forall d\in D, \tag{17}$$

$$M(1-z^{D}_{i,f,d})\geqslant t^{DA}_{j,f,d}+s^{D}_{i}-t^{L}_{i,f},\ \forall i\in C,\forall f\in F,\forall d\in D, \tag{18}$$

$$M(1-z^{T}_{i,f})\geqslant t^{R}_{i,f,d}+s^{T}_{i}-t^{L}_{i,f},\ \forall i\in C,\forall f\in F,\forall d\in D, \tag{19}$$

$$\max t^{R}_{i,f,d}\leqslant H,\ \forall i\in C,\ \forall f\in F,\ \forall d\in D, \tag{20}$$

$$t^{R}_{i,f}=\max(t^{DA}_{i,f,d},t^{TA}_{i,f}),\ \forall i\in C,\ \forall f\in F, \tag{21}$$

约束 (16) 定义车辆抵达服务客户的时间. 约束 (17) 确保当无人机未发射时, 其抵达节点$i\in N$的时间与运载它的车辆同步. 约束 (18) 定义了无人机从节点$i\in C$的出发时间. 约束 (19) 规定车辆f从客户节点$i\in C$的出发时间等于车辆到达时间$t^{R}_{i,f}$与服务该节点的时间之和. 约束 (20) 表示无人机和车辆司机工作时间的上限H. 约束 (21) 定义车辆f抵达客户节点$i\in C$的到达时间$t^{R}_{i,f}$.

1.2.3 能量和容量约束

$$M(2-x_{i,j,f,d}-h^{R}_{j})\geqslant e^{D}t^{DF}_{i,j}-E^{D}_{i,d},\ \forall i\in C,\forall j\in N,\forall f\in F,\forall d\in D, \tag{22}$$

$$M(1-x_{i,j,f,d}-h^{L}_{i})\geqslant E^{D}_{i,d}-B^{D},\forall d\in D,\ \forall i\in C\cup\{0\},\forall j\in\{C:j\neq i\},\forall f\in F, \tag{23}$$

$$p^{D}\sum_{f\in F}\sum_{d\in D} z^{D}_{i,f,d}\geqslant w^{D}_{i},\forall i\in C, \tag{24}$$

约束 (22) 确保每架无人机拥有足够能量抵达与车辆的会合点. 当无人机从车辆或仓库起飞时, 约束 (23) 保证无人机起飞时的电量不会超过其电池容量B^{D}. 约束 (24) 确保若客户i由无人机服务, 其承载的包裹重量不超过其最大载荷.

2 基于模型与图强化学习的车辆-无人机协同多目标路径优化方法

2.1 方法框架

如图 2 所示, 首先基于混合策略的初始化方法, 算法生成多样性良好的初始种群P_0. 其次, 在迭代阶段, 算法先进行选择, 重组等全局搜索操作生成精英解集. 随后, 基于精英解集, GCN 位于底层, 用于从车辆-无人机路径析取图中提取结构化状态特征 (状态s_t). 然后, DQN 位于决策层, 以 GCN 输出的高维

状态嵌入s_t作为输入, 学习路径局部搜索算子的选择策略 (如$N2$), 并执行 PLS 以进一步细致优化解 (生成s_{t+1}). 最后, LSTM 位于环境建模层, 用于刻画 PLS-DQN 交互过程的状态转移, 并生成多步虚拟交互轨迹以辅助策略更新. 这确保 DQN 策略网络和其目标网络的充分训练, 进而提升算法的整体优化效率. 在此阶段, 精英归档策略一直更新并存储全局非支配解. 迭代过程持续至满足预设次数, 确保智能体有效引导路径方案优化.

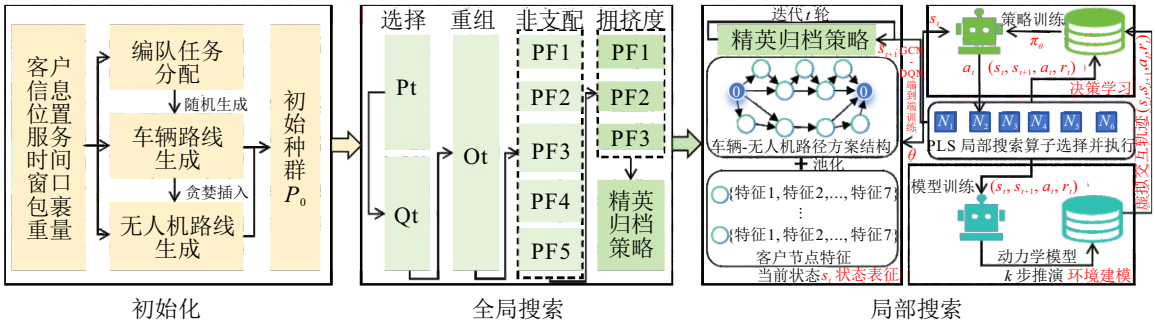


图2　基于模型与图强化学习的车辆-无人机协同多目标路径优化方法框架

2.2　解的编码表示

受启发于文献 [6], MVRPDPDTW 问题的解由五个部分构成. 如图 3 所示, 染色体第一部分的 0 值代表仓库. 第二部分为车辆路线, 由车辆访问客户节点表示 (第二行编码值为 0 值). 第三部分为无人机访问的客户节点, 以客户节点和无人机编号呈现 (第二行编码值为非 0 值). 第四和第五部分是无人机发射与回收节点, 由第二部分的车辆访问客户编号表示. 本文设定无人机拜访单个客户后返回车辆. 因此, 无人机拜访客户的前后节点为其发射与回收节点.

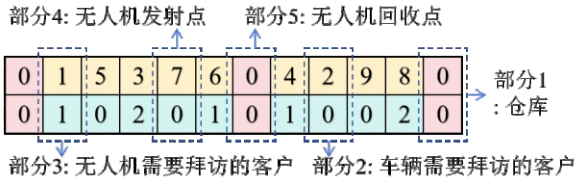


图3　染色体编码示意图

例如, 染色体示意图 3 对应路径方案图 4. 染色体$\langle 0,1,5,3,7,6,0\rangle$表示编队 1 服务的客户编号. 染色体$\langle 0,1,0,2,0,1,0\rangle$表示编队 1 中的无人机 1 服务客户 1 以及无人机 2 服务客户 3. 无人机 1 在仓库起飞, 服务客户 1 后, 被回收于车辆路线的客户 5.

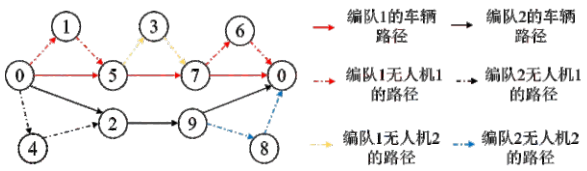


图4　多车辆-多无人机路径示意图

2.3　基于混合策略的初始解生成方法

基于车辆优先, 无人机次要原则[4], 本文设计了一种简单的初始解构造方法. 具体而言, 首先为了满足装载容量限制, 根据客户的包裹重量总和以及车辆最大装载容量计算所需车辆数目. 随后, 通过 K-means++ (仓库为初始聚类中心, 车辆数目为初始聚类簇数) 将邻近客户分配至同一车辆-无人机编[7]. 综上, MVRPDPDTW 被分解为多个独立编队的 VRPDPDTW. 最后, 算法聚焦于如何构建单个车辆-无人机编队的初始路径解.

车辆路径的首尾节点均为仓库, 其余节点为采用随机排序的客户点以增强初始种群多样性. 随后通过贪婪法和如下三个条件将客户c从车辆路线插入至无人机路线. (1) 当节点c插入在车辆路线任意两个相邻节点a和b间, 三个节点组成无人机路线(a,c,b); (2) 节点c可由无人机服务, 即无人机续航与载重能力可以完成服务并返回节点b; (3) 无人机沿路径(a,c,b)成本低于车辆沿路径(a,c,b)成本 (成本见公式 1). 最后, 当插入操作成功时, 将节点c将从车辆路线中移除. 该过程持续直至车辆路线的节点均被遍历. 随后, 算法输出初始路线方案.

2.4　多目标搜索框架

图 2 所示, 在多目标搜索框架中, 每次迭代都会实施一组全局搜索与局部搜索来探索和开发解空间. 在遗传操作后, 帕累托局部搜索通过 DQN 选择 6 个搜索算子中的最优算子以有效提升算法的收敛性.

遗传全局搜索: 为提升种群多样性, 本文设计多点交叉和逆向变异算子. 具体而言, 首先采用锦标赛选择方法从种群中随机选取预定个体, 并根据支配评估结果, 选择最适应的个体进行后续遗传操作. 然后两个亲本染色体在随机选取的多处基因位进行切割与交换, 生成子染色体. 接着对子染色体随机选取两个基因位, 通过水平翻转两个随机点之间的所有

基因实现变异. 如果任一操作改变染色体的可行性, 则该操作将保持染色体不变. 与传统多目标方法类似, 交叉和变异操作受交叉率p_c和变异率p_m控制.

帕累托局部搜索: 传统的 PLS 算法在每次迭代中通过开发所有非支配解迭代更新解集, 效率较低. 值得注意的是, 早期的非支配解可能在下一代被新解所支配. 受文献启发[20], 采用精英归档机制存储搜索过程中的全局帕累托非支配解. 通过 DQN 选择操作算子 (*N1*-*N6*) 并对非支配解执行此操作, 可在寻找新非支配解时降低计算开销.

1) 车辆转无人机操作算子$N1$: 如图 5 所示, $N1$旨在将车辆服务的客户节点转换给无人机. 具体而言, 算法从车辆路线中删除成本最高且符合无人机服务条件的客户, 并采用贪心算法为被删除的客户寻找低成本的无人机分配方案. 同时, 删除操作将减少车辆的服务客户, 因此无人机的服务客户和无人机的发射, 回收节点将根据删除客户相应调整.

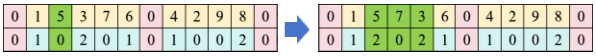


图5 车辆转无人机操作算子 ***N1*** 示意图

2) 无人机转车辆操作算子$N2$: 如图 6 所示, 与$N1$相反, $N2$旨在将由无人机服务的客户移至车辆路线. 具体而言, 算法删除成本最高的无人机路线方案, 并将被删除的客户采用贪心算法插入成本最低车辆路线的相邻节点间. 同时, 由于无人机路线发生变化, 因此调整车辆服务的客户路线.

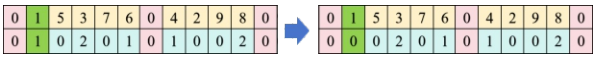


图6 无人机转车辆操作算子 ***N2*** 示意图

3) 交换操作算子$N3$: 如图 7 所示, $N3$通过交换车辆和无人机服务客户节点的位置实现. 随机选取两个客户节点, 且选取的车辆路线节点须为无人机可达节点, 否则该操作后的路径方案是不可行的.

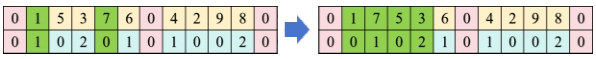


图7 交换操作算子 ***N3*** 示意图

4) 2-*Opt*操作算子$N4$: *2-Opt* 是一种高效的启发式操作, 常用于解决车辆路径问题[21]. 如图 8 所示, *2-Opt* 随机选取客户节点并通过交换路径中任意两条边与其他两条边来优化车辆路线. 同时, 根据置换结果, 车辆或者无人机的服务顺序需要相应调整.

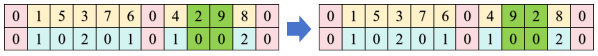


图8 2-*Opt* 操作算子 ***N4*** 示意图

5) 贪婪删除-重新插入操作算子$N5$: 如图 9 所示, 启发式算法$N5$聚焦于无人机服务的客户分配. 首先, 采用与$N2$相同的方法, 识别并删除出成本最高的无人机服务客户节点. 其次, 使用贪婪法将删除的客户节点重新分配给其他可行无人机.

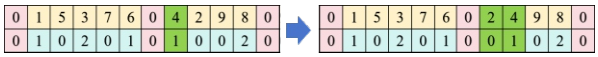


图9 贪婪删除-重新插入操作算子 ***N5*** 示意图

6) 随机删除-重新插入操作算子$N6$: 如图 10 所示, 为了提高种群多样性, 与$N5$类似, $N6$首先删除随机选取的无人机服务客户来优化无人机路线, 随后将该客户节点随机分配给其他可行无人机.

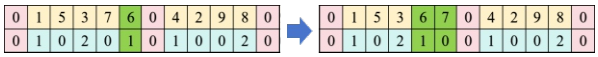


图10 随机删除-重新插入操作算子 ***N6*** 示意图

2.5 基于 GCN 的特征提取机制

如图 11 所示, 为了提供丰富的状态表征信息, 通过采用高效的图卷积网络充分挖掘路线结构, 进而综合表征路径方案[22]. 首先, 将车辆-无人机路径方案转化为图邻接矩阵与特征矩阵后, 采用图卷积神经网络为每个图节点生成高维嵌入向量, 以捕捉路径方案的局部特征与全局结构信息. 随后, 通过池化技术进行节点嵌入聚合, 构建出对整体方案的综合表征, 从而为智能体提供丰富决策支持. 邻接矩阵与特征矩阵是 GCN 的核心输入. 定义如下:

邻接矩阵: 邻接矩阵表示图的结构, 记为$\boldsymbol{A} \in \mathbb{R}^{N\times N}$, 其中$N$为图中节点数, $N=|C|$. C为客户集合. 图由空间边 (E) 构成, 表示客户访问顺序的依赖关系. 其定义如下: 1) 对于任意两个客户a和b, 若a在车辆路线中先于b被服务, 则a和b之间存在空

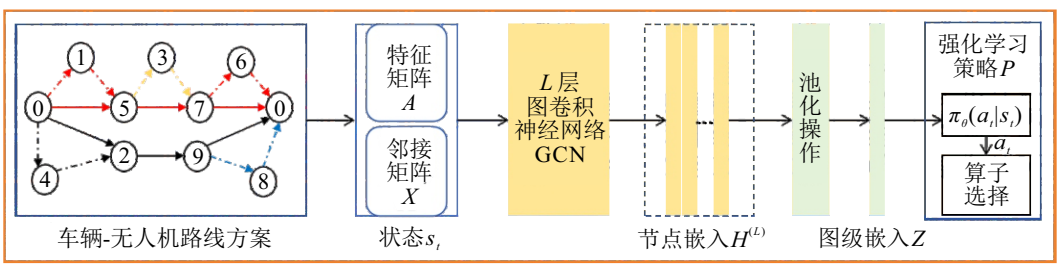


图11 基于 GCN 的特征提取机制示意图

间边. 这些边强制执行同一车辆路线内客户的服务顺序, 确保路线方案的服务优先级. 2) 对于无人机旅行中的任意两个客户a和b, 当节点a在b之前被访问时, a和b之间存在空间边. 该空间边强制执行无人机访问客户的顺序, 确保正确的发射, 旅行和回收任务. 邻接矩阵归一化如下, 其中D为$\boldsymbol{A}$的度矩阵:

$$\widehat{\boldsymbol{A}} = \boldsymbol{D}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}, \tag{25}$$

特征矩阵: 特征矩阵记作$\boldsymbol{X} \in \mathbb{R}^{N\times F}$, 其中$F$为每个节点的特征维度. 每个操作节点的特征$\boldsymbol{x}_{\mathrm{i}} \in \mathbb{R}^{F}$代表第$i$个节点的特征向量. 这些特征精确反映操作的路由状态, 且相互独立以确保无特征间的影响, 进而稳健地表示. 特征的内容如下:

1) 客户索引: 标识所需要服务的客户;

2) 编队索引: 标识为客户服务的编队;

3) 车辆索引: 标识为客户服务的车辆;

4) 无人机索引: 标识为客户服务的无人机;

5) 到达时间: 标识为客户服务的开始时间;

6) 离开时间: 标识为客户服务的完成时间;

7) 服务顺序: 标识为车辆配送路线的服务顺序.

为了有效利用析取图结构与节点特征, GCN 为每个节点及路线方案的表征过程包含两个关键阶段: 首先, 通过 GCN 层处理特征矩阵与图结构, 为单个节点生成高维嵌入, 同时捕捉其局部特征与全局结构上下文. 其次, 运用池化聚合技术将节点嵌入进行整合, 构建整体路径方案的综合表征. 具体如下:

嵌入生成: 通过迭代的 GCN 层更新节点嵌入. 在第l层, 节点嵌入$\boldsymbol{H}^{(l+1)}$计算公式为:

$$\boldsymbol{H}^{(l+1)} = \sigma(\widehat{\boldsymbol{A}}\boldsymbol{H}^{(l)}\boldsymbol{W}^{(l)}), \tag{26}$$

其中, $\boldsymbol{H}^{(l)} \in \mathbb{R}^{N\times F_l}$表示$l$层的节点嵌入向量, F_l为l层的嵌入维度. 对于输入层, $\boldsymbol{H}^{(0)}$为节点特征矩阵$\boldsymbol{X}$. $\widehat{\boldsymbol{A}}$是捕获图结构的归一化邻接矩阵, $\boldsymbol{W}^{(l)}$是第l层 GCN 的训练权重矩阵. $\sigma(\cdot)$为 ReLU 激活函数. 该迭代过程使 GCN 能够聚合邻近节点信息并优化各层嵌入向量, 从而同时捕捉图的局部与全局特性.

嵌入池化: 经过L层图卷积后, 获得节点嵌入向量$\boldsymbol{H}^{(L)} \in \mathbb{R}^{N\times d}$, 其中$d$为每个节点的嵌入维度. 这些嵌入向量通过整合原始特征与邻近节点聚合的结构信息呈现路线空间状态. 为了便于智能体的表示, 对最终的节点嵌入$\boldsymbol{H}^{(L)}$应用全局均值池化操作. 全局均值池化计算图中每个节点嵌入的平均值, 生成固定维度的图级嵌入$\boldsymbol{Z} \in \mathbb{R}^{d}$如下:

$$\boldsymbol{Z} = \frac{1}{|\boldsymbol{N}|}\sum_{i=1}^{N}\boldsymbol{h}_i^{(L)}, \tag{27}$$

其中, $\boldsymbol{h}_i^{(L)} \in \mathbb{R}^{d}$表示第$i$个节点经过$L$层图卷积后的嵌入向量. 聚合操作确保图级嵌入$\boldsymbol{Z}$与图中节点数量无关, 特别适用于处理 VRPD 中不同规模的图. 本文选择$L = 3$. GCN 的输入层规模$F = 7$由特征向量长度决定. 隐藏层规模设为 64, 输出层规模为 32.

2.6　马尔可夫决策过程建模

本节将 DQN 的智能体对局部搜索算子的选取过程建模为马尔可夫决策过程. 在给定状态$\mathcal{S}$下, 智能体依据策略$\mathcal{P}$执行动作$\mathcal{A}$: 从操作池 (*N1*-*N6*) 中选择操作符并开发当前解. 随后智能体基于从环境中获取的两套解计算奖励$\mathcal{R}$, 并通过累积奖励更新策略. MDP 以四元组$\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}\rangle$形式表示:

1) 状态$\mathcal{S}$: 智能体通过观察车辆-无人机路线方案的状态s_t进行决策, 该状态需捕获完整路线方案的所有关键信息. 因此, 为了充分表征路线方案, 状态模型将路线方案转换为邻接矩阵和特征矩阵, 融合了路线结构与客户的数值属性信息. 随后通过图卷积神经网络与池化操作提取高维状态嵌入 (s_t), 为智能体决策框架提供全面的路由方案表征.

2) 动作$\mathcal{A}$: 对应于每次多目标算法迭代的状态s_t, 操作列表中存在 6 种局部搜索操作选择 (*N1*-*N6*). 同时这 6 种操作构成了动作空间$\mathcal{A}$, 并且使智能体能够充分探索解空间, 引导路由方案趋向更优性能.

3) 奖励函数$\mathcal{R}$: 该奖励函数旨在引导智能体优化路由方案, 并鼓励开发高质量解.

$$r_t = \begin{cases} 10, \text{ 若 } s_{t+1} \text{ 支配 } s_t \\ 5, \text{ 若 } s_t \text{ 与 } s_{t+1} \text{ 互不支配} \\ -1, \text{ 若 } s_t \text{ 支配 } s_{t+1} \end{cases}, \tag{28}$$

其中, 在计算奖励r_t时, s_t表示当前解, s_{t+1}表示对s_t应用选定局部操作后得到的新解. 具体而言, 若新解s_{t+1}支配当前解s_t, 则智能体获得 10 点奖励, 以激励种群收敛性改进. 若新解s_{t+1}与当前解s_t互不支配, 则智能体获得 5 点奖励, 以激励种群多样性改进. 在未观察改进情况下, 智能体获得-1 点奖励.

4) 策略$\mathcal{P}$: 策略$\pi_\theta(a_t|s_t)$定义了智能体在t时刻基于观测状态s_t选择动作a_t的策略. 该策略通过参数θ实现, 以神经网络形式输出动作空间$\mathcal{A}$上的概率分布. 策略网络通过 GCN 与池化操作提取的状态嵌入$\boldsymbol{Z}$, 计算每个动作$a_t \in \boldsymbol{A}$的应用概率.

$$\pi_\theta(a_t|s_t) = \frac{\exp(\mathrm{MLP}_\theta(\boldsymbol{Z})_i)}{\sum_{j=1}^{|A|}\exp(\mathrm{MLP}_\theta(\boldsymbol{Z})_j)} \tag{29}$$

$\mathrm{MLP}_\theta(\boldsymbol{Z})_i$表示状态嵌入$\boldsymbol{Z}$下动作索引$i \in \boldsymbol{A}$的分数, $\pi_\theta(a_t|s_t)$为选择动作a_t的概率. 策略网络采用两层多层感知机 (MLP) 结构. MLP 的输入维度与 GCN 的输出维度均为 32. 两层隐藏层维度均为 16.

策略网络的输出层为标量, 代表概率估计下的动作索引.

2.7 基于 LSTM 的 DQN 策略优化方法

如图 12 所示, 本文引入基于模型的策略优化框架 (MBPO)[23], 提出了基于 LSTM 的 DQN 的策略优化方法. 通过训练 LSTM 构建交互环境模型并生成推演虚拟样本为智能体提供双向训练数据, 提高训练样本效率, 进而增强对搜索算子的选择能力.

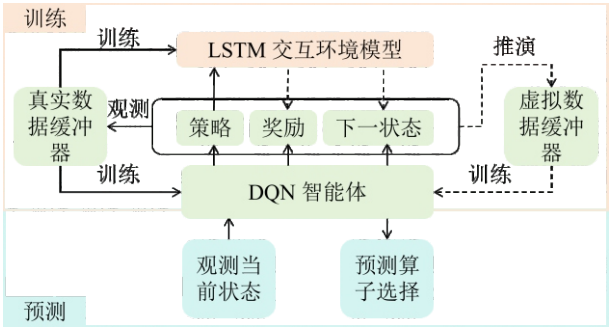


图12 基于 LSTM 的 DQN 策略优化示意图

具体而言, 智能体首先通过与环境交互生成拟合环境模型的真实数据, 随后利用缓冲区中的状态数据进行有限步长的预测. 环境模型基于智能体的动作与当前状态预测路线方案的下一个状态, 从而模拟智能体与局部搜索操作的交互过程. 最后, DQN 智能体通过真实数据与虚拟数据的采样完成训练.

由于预测路线解的下一个状态属于序列预测问题, 即下一个状态由先前状态的组合定义. LSTM 通过隐藏层编码与历史值的长期依赖关系在序列预测问题上比传统神经网络更精准[24]. 因此, 本文采用 LSTM 对经过帕累托局部搜索后的解轨迹建立交互环境模型. 所提交互环境模型的架构采用三层结构设计: 第一层由 50 个 LSTM 单元构成, 其输入状态 s_t 和动作 a_t 的合并 (8 维向量). 第二层与第三层分别包含 128 个和 64 个神经元, 并采用整流线性单元 (ReLU) 激活函数. 同时, LSTM 的输出是预测状态 $\hat{s}_{t+1}^i$ (7 维向量). 整体的策略优化步骤如下:

step 1 采样: DQN 智能体与环境进行交互, 将 M 个样本 $(s_t^i, a_t^i, r_t^i, s_{t+1}^i)$ 存储在真实数据缓存器 $\mathcal{D}_{real}$.

step 2 训练: 基于 $\mathcal{D}_{real}$ 的样本 i, 通过 Adam 优化器最小化 LSTM 损失 $L(\theta') = \dfrac{1}{M}\sum_{i=1}^{M}(r_t^i - \hat{r}_t^i)^2$ 训练 LSTM 参数 θ' 以构建交互环境模型. 其中, 损失函数 $L(\theta')$ 是真实下一状态 s_{t+1}^i 与预测状态 $\hat{s}_{t+1}^i$ 形成的均方奖励误差. $\hat{r}_t^i$ 是 s_t^i 和 $\hat{s}_{t+1}^i$ 之间的奖励值.

step 3 推演: 在 $\mathcal{D}_{real}$ 中随机选择一个样本, 使用 LSTM 网络进行 k 步推演预测并将生成的轨迹集 $B(k) = \{(s, a, r, s')\}$ 添加到虚拟数据缓存器 $\mathcal{D}_{virt}$.

step 4 存储: 从缓存器 $\mathcal{D}_{real}$ 和 $\mathcal{D}_{virt}$ 中抽取 DQN 训练所需最小批次的混合样本存储在其回放池中.

step 5 计算: 对回放池每个样本 j 用目标网络计算 $y_j = r_j + \gamma \cdot \max_{a_j} Q_{\theta^-}(a_j, s_{j+1})$, 其中 γ 为折扣率.

step 6 更新: 最小化目标损失 $(y_j - Q_\theta(s_j, a_j))^2$ 更新 DQN 网络 Q_θ, 同时每 c 步更新一次目标网络 Q_{θ^-}.

2.8 算法收敛性分析

如图 2 所示, 算法的主体结构是全局搜索和局部搜索. 全局搜索采用非支配排序和拥挤度筛选种群规模大小且表现优异的个体进入局部搜索. 局部搜索采用精英存档策略存储搜索过程中的帕累托最优解. 文献证明[25], 具备最优个体保留机制的算法一定能收敛到全局最优解. 此外, 算法没有切割解空间, 不存在遗失最优解的可能. 因此, 算法能够保证在有限迭代次数下收敛至全局最优解.

3 计算实验

为了评估算法性能, 本节开展基准算例仿真. 算法运行在 64 位 Windows 系统, 英特尔酷睿 i7-12700H @2.30 GHz 16GB RAM, NVIDIA GEFORCE RTX 3050 GPU, 6GB 的平台. 软件采用 Python 3.10, PyTorch 1.13 和 CUDA 11.7 编写.

3.1 实验设置

3.1.1 性能指标设定

实验采用超体积 (HV) 与覆盖率 (C-metric) 衡量算法性能[7-8]. $HV(S, R)$ 为解集 S 与参考点 R 构成的立方体体积, 其数值越高, 表明算法收敛性与多样性越优. $C(A, B)$ 表示集合 B 中被集合 A 内任意解所支配的解所占比例, 用于进一步验证算法的收敛性.

$$HV(S, R) = \int_S \prod_{i=1}^{m} \max(0, r_i - f_i)dy, \tag{30}$$

$$C(A, B) = \frac{\{|b \in B| \exists a \in A : a \preceq b|\}}{|B|}, \tag{31}$$

m 表示目标的维度, r_i 表示第 i 个目标值对应的参考点 (参考点为迭代中最差的解). $|B|$ 表示非支配解集 B 中的解的数量. 若 $C(A, B)$ 为 1, 则 B 中每个解至少被 A 中至少一个解所支配或与其相等. 若 $C(A, B)$ 为 0, 则 B 中没有任何解可被 A 中的任意解所支配.

3.1.2 基准测试算例设置

由于所提问题缺乏基准算例, 受启发于文献

[12], 本文扩展了 16 个 Dumas 提出的 VRPPDTW 测试集作为本文算例, 并设定客户规模 (20, 40, 60, 80). 算例名称表示客户数量, 时间窗宽度和测试编号, 如 En20w80_01 表示该算例有 20 个客户, 每个客户的服务时间窗宽度为 80 s. 算例信息包括: 客户索引i, 位置$[x_i, y_i]$ (随机分布在仓库点 [0, 0] 至点 [100, 100] Km 的范围), 需求类型 (40% 和 60% 的客户需要取货, 送货), 时间窗 $[a_i, b_i]$ (与 Dumas 测试集一致) 和包裹重量w_i (65% 的包裹符合无人机载重要求).

根据客户的包裹重量总和以及车辆最大装载容量计算所需编队最小数目. 此外, 针对中小规模场景, 每个编队包含一辆卡车和两架无人机[5-6], 其中车辆和无人机的参数设定如下. 车辆和无人机服务客户时间 (s_i^D, s_i^T) 分别为 60 s 和 0 s. 车辆和无人机行驶速度 (v^D, v^T) 分别为 50 Km / h 和 60 Km / h. 无人机最大载重p^D为 40 Kg. 车辆单位距离行驶成本, 无人机单位距离飞行成本, 编队固定启动成本以及违反时间窗口成本和每小时司机装卸费 (c_1至c_5) 分别为 5 元/Km, 2 元/Km, 10 元, 1 元/h 和 2 元/h. 无人机电池消耗率e^D和容量B^D分别为 29.9 Wh / Km 和 599.4 Wh. 编队最大工作时长H为 8 h.

3.2　参数敏感性实验

为了确定参数的最佳取值, 本节进行田口正交试验[18-19]. 参数候选值如下:

1) 种群大小$PS \in \{60, 80, 100, 120, 140\}$;

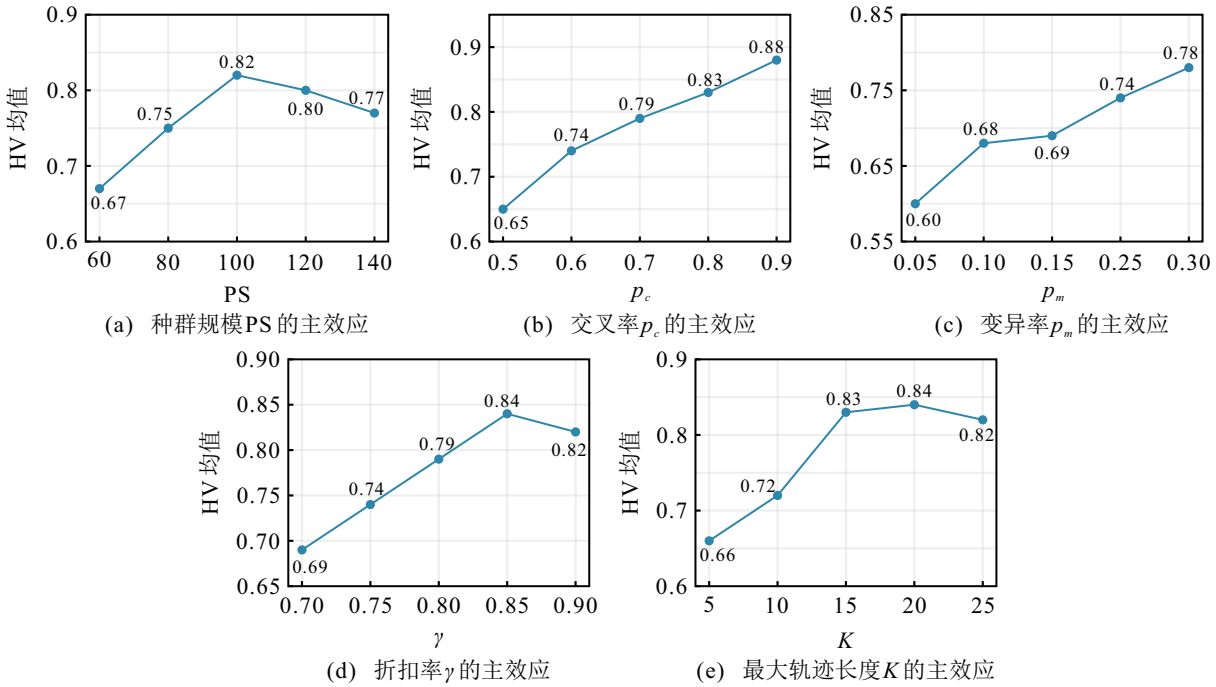
2) 交叉率$p_c \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$;

3) 变异率$p_m \in \{0.05, 0.10, 0.15, 0.20, 0.25\}$;

4) 折扣率$\gamma \in \{0.70, 0.75, 0.80, 0.85, 0.90\}$;

5) 最大轨迹长度$k \in \{5, 10, 15, 20, 25\}$.

每组实验通过 L25 正交表 (5^5) 进行参数校准并取 10 次结果的平均值. 如图 13 所示, 参数在 HV 指标的均值主效应图表明: 所有参数对算法的性能均有影响. 具体而言, 当种群PS = 100 时, 算法在多样性和早熟之间取得平衡[5]. 交叉率p_c = 0.9 有助于维持全局搜索能力[15]. 变异率p_m = 0.25 有效促进种群多样性, 防止早熟收敛[15]. 对于强化学习组件, 折扣率γ = 0.85 使得智能体平衡短期与长期收益[11]; 而最大轨迹长度k = 20 保证智能体训练的充分性且不产生过拟合[23]. 基于此, 算法参数设定为: PS = 100, p_c = 0.9, p_m = 0.25, γ = 0.85, k = 20.



(a) 种群规模PS 的主效应　(b) 交叉率p_c 的主效应　(c) 变异率p_m 的主效应

(d) 折扣率γ 的主效应　(e) 最大轨迹长度K 的主效应

图13　参数在 HV 评价指标下的均值主效应图

为了进一步说明本文算法中 GCN、LSTM 的架构设计以及奖励函数设计的合理性, 对 GCN 网络的层数L (L = 1, 2, 3, 4)、LSTM 模型的推演步骤k值 (k= 1, 3, 5, 7) 以及奖励函数的敏感性 (ω = 0.5, 0.8, 1.0, 1.2, 1.5) 进行敏感性分析, 分析曲线如图 14-16:

结果表明: (i) 当L=1 时, GCN 对复杂空间依赖的刻画能力不足, 整体解质量下降; 当L=4 时，模型出现性能退化和训练震荡, 这是因为过平滑问题的存在; 当L=3 时, 算法在解质量与训练稳定性之间取得了平衡. (ii) 随着k增大, 状态预测误差逐步累积, 导致虚拟样本与真实环境分布的偏差增大. 适中的k (k=3) 能够在保证精度的同时提升样本多样性, 从而

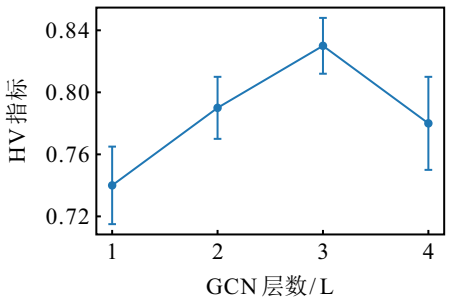


图14 GCN 层数L的主效应图 (L =3)

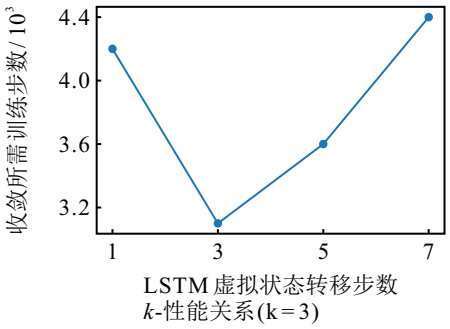


图15 LSTM 虚拟状态转移步数k-性能关系图 (k =3)

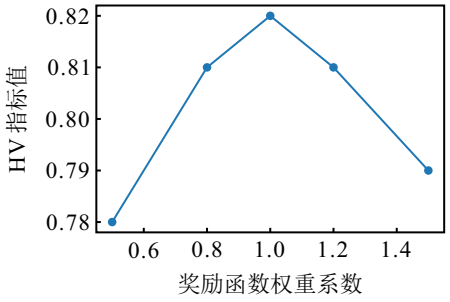


图16 奖励函数的敏感性分析图 (ω=1)

加快策略收敛; 而过大的k则会削弱甚至负面影响训练效果. (iii) 在ω=1 附近区间内, HV 随权重变化平稳且有小幅波动, 说明奖励函数设计具有鲁棒性; ω=1 处于单峰附近, 既能保证解质量也不依赖于精细的调参. 上述结果验证了算法架构参数设计的稳定性和合理性.

3.3 基于 CPLEX 的对比实验

为了验证算法在小规模算例上的收敛性, 本文算法与精准求解器 CPLEX 在 5 个算例上独立运行 10 次并取平均结果[18]. 此外, 本文根据非支配关系和拥挤度选取所提算法和 CPLEX 求解帕累托前沿最优解. 结果如表 2 所示, 其中* 表示 CPLEX 运行 3 600 秒提供的最优参考解, 最优结果已加粗标出.

表 2 对于小规模算例 (如 En10w80_01), CPLEX 能够找到并验证其最优性 (表中以*标出), 而算法能近似匹配该最优解, 证明了其在小规模问题上的收敛能力. 但是随着问题规模扩大至 En15w80_01 及更大, CPLEX 的求解性能急剧下降, 只能在限定时间内提供可行解. 相比之下, 本文算法在所有算例上的性能均等于或优于 CPLEX, 显示其在小规模案例上的有效性及在大规模案例上替代精确求解器的潜力.

表2 本文算法与 CPLEX 的对比结果

算例	CPLEX		本文算法	
	F_1 (元)	$F_2(s)$	F_1 (元)	$F_2(s)$
En10w80_01	**65.4***	**12***	66	13
En15w80_01	71.6	16	**70.3**	**15**
En20w80_01	79.5	18	**77.4**	**18**
En25w80_01	85.2	22	**83.6**	**22**
En30w80_01	90.8	25	**88.2**	**23**

3.4 消融实验

本文算法的核心组件包含混合策略的初始化方法, 基于 GCN 的特征提取机制, 基于 RL 的局部搜索方法和基于 LSTM 的 DQN 策略优化方法. 为了检验不同组件的有效性, 本节开发了如下变体:

1) 算法 1 采用随机方法生成初始种群;

2) 算法 2 采用传统无特征提取的状态表征方法;

3) 算法 3 采用随机策略的搜索算子选择机制;

4) 算法 4 采用无模型的 DQN 策略优化方法.

在算例 En35w80_01 中, 每个算例均独立运行 10 次, 并取平均结果. 表 3 呈现了显著性水平 0.05 条件下的 Friedman 秩和检验结果[17], 最优结果已加粗.

表3 消融实验的 Friedman 秩和检验结果

算法	HV		C-metric	
	排名	p-值	排名	p-值
算法1	2.36		2.55	
算法2	2.86		2.92	
算法3	5.63	1.45E-10	6.21	1.48E-10
算法4	3.25		3.76	
本文算法	**1.25**		**1.92**	

如表 3 所示, 本文算法在 HV 和 C-metric 指标的排名均优于所有变体. 算法 3 的排名最差, 相比于 DQN, 证实随机选择机制难以选择提升搜索质量的运算符. 算法 4 的性能次之, 表明策略优化方法通过持续增加经验有助于获得优质的算子选择策略. 算法 1 和 2 的性能也差于本文算法, 证实混合策略的初始化方法通过丰富初始种群多样性, 为后续搜索提供了优质起始点. 同时, 表明通过捕捉解的空间结构以丰富智能体决策信息可以有效引导搜索过程. 此外, 两种指标的p-值均低于显著性水平 0.05, 证实各算法产出的结果之间存在显著差异.

为了进一步验证 LSTM 虚拟样本生成的有效

性, 本实验将比较使用和不使用 LSTM 虚拟样本时本文算法训练步数的变化曲线.

实验结果如图 17 所示, 引入 LSTM 虚拟样本后, DQN 在早期训练阶段即可获得稳定的策略改进. 相比不使用虚拟样本的配置, 达到相同性能水平所需的训练步数减少. 该结果验证了环境模型的有效性, 其主要贡献在于提升训练效率与样本利用率.

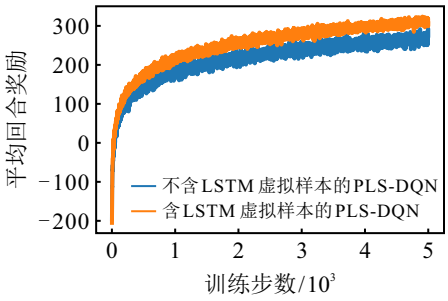


图17　LSTM 生成虚拟样本的训练效率学习曲线

3.5　对比实验

为进一步验证算法的有效性, 本节将对比先进算法 NSGA-II[26], MOEA/D[27], HMOA[5], NRL[12], RDMA[13] 和 AMA[14]. HMOA 结合改进的帕累托局部搜索提升收敛性. NRL 结合 Q-学习为变邻域下降搜索提供合适的搜索算子, 增强其局部搜索能力. RDMA 通过 Q-学习为交叉操作提供合适帕累托前沿区域的父本, 提高其全局搜索能力. AMA 使用多智能体强化学习选择合适的遗传和局部搜索算子, 加强全局和局部搜索的平衡. 值得注意的是, 强化学习均用于策略引导而非监督预测, 其性能评估基于最终获得的 Pareto 解集质量, 而非预测误差. 为了保证公平性, 对比算法也通过田口试验单独运行 10 次以完成参数配置, 结果如表 4 所示.

表4　对比实验的算法参数校准结果

对比算法	参数
NSGA-II[26]	种群大小$PS=140$, 交叉率$p_c=0.6$, 变异率$p_m=0.15$
MOEA/D[27]	种群大小$PS=140$, 交叉率$p_c=0.6$, 变异率$p_m=0.15$, 邻域大小为15
HMOA[5]	种群大小$PS=200$, 交叉率p_c和变异率p_m分别为0.8和0.3, 重启率$\alpha=0.3$
NRL[12]	种群大小$PS=200$, 交叉率p_c和变异率p_m分别为0.7和0.3, 折扣率$\gamma=0.5$, 探索率$\varepsilon=0.8$
RDMA[13]	种群大小$PS=200$, 变异率$p_m=0.5$, 学习率$\beta=0.3$, 折扣率$\gamma=0.5$, 变邻域搜索阈值次数$V=5$
AMA[14]	种群大小$PS=140$, 交叉率p_c和变异率p_m分别为0.6和0.2, 学习率$\beta=0.3$, 折扣率$\gamma=0.5$, 探索率$\varepsilon=0.8$

表 5 和表 6 给出 7 种算法在两种评价指标和相同停止标准 (训练步数上限为 5000 步, 最大训练时间为 60 分钟, 最大运行次数为 200 次) 下对 16 个算例独立运行 20 次的均值结果. 加粗的数值表示最佳值; Wilcoxon 秩和检验用于评价算法结果间的统计差异; +, =, −表示本文算法的表现优于、持平于或逊于其竞争对手. 表 5 和表 6 结果显示, 在大多数算例上, 本文算法在 HV 和 C-metric 指标上优于对比

表5　算法关于 HV 指标的对比结果 (时间单位为秒.)

算例	NSGA-II			MOEA/D			HMOA			NRL			RDMA			AMA			本文算法		
	均值	标准差	时间	均值	标准差	时间	均值	标准差	时间	均值	标准差	时间	均值	标准差	时间	均值	标准差	时间	均值	标准差	时间
En20w80_01	0.65	0.05	12.3	0.62	0.06	10.5	0.70	0.04	15.2	0.72	0.03	18.7	0.71	0.04	16.8	0.74	0.03	22.4	**0.82**	0.02	25.6
En20w80_02	0.64	0.05	12.1	0.63	0.05	10.7	0.71	0.04	15.4	0.73	0.03	18.9	0.72	0.03	17.1	0.75	0.03	22.7	**0.83**	0.02	25.9
En20w80_03	0.66	0.04	12.4	0.61	0.06	10.3	0.69	0.05	14.9	0.71	0.04	18.3	0.70	0.04	16.5	0.76	0.03	22.1	**0.81**	0.02	25.3
En20w80_04	0.65	0.05	12.2	0.64	0.05	10.8	0.72	0.03	15.6	0.74	0.03	19.2	0.73	0.03	17.3	0.77	0.02	23.0	**0.84**	0.01	26.2
En40w80_01	0.60	0.06	24.7	0.58	0.07	21.2	0.68	0.05	28.5	0.70	0.04	33.1	0.69	0.05	30.2	0.72	0.04	38.6	**0.80**	0.03	42.3
En40w80_02	0.61	0.06	24.9	0.59	0.06	21.5	0.67	0.05	28.8	0.69	0.04	33.4	0.68	0.05	30.5	**0.75**	0.04	39.1	0.75	0.03	42.8
En40w80_03	0.62	0.05	25.2	0.57	0.07	20.9	0.66	0.05	28.1	0.68	0.04	32.8	0.67	0.05	29.9	0.71	0.04	38.2	**0.78**	0.03	41.9
En40w80_04	0.63	0.05	25.4	0.58	0.06	21.7	0.69	0.04	29.2	0.71	0.03	34.0	0.70	0.04	31.1	0.74	0.03	39.7	**0.81**	0.02	43.5
En60w80_01	0.58	0.06	38.2	0.55	0.07	32.8	0.65	0.05	42.3	0.67	0.04	48.1	0.66	0.05	44.5	**0.77**	0.04	55.4	0.77	0.03	59.2
En60w80_02	0.59	0.06	38.5	0.56	0.07	33.1	0.64	0.05	42.6	0.66	0.04	48.4	0.65	0.05	44.8	0.69	0.04	55.8	**0.76**	0.03	59.6
En60w80_03	0.57	0.06	37.9	0.54	0.07	32.5	0.63	0.05	41.9	0.65	0.04	47.7	0.64	0.05	44.1	**0.75**	0.04	55.0	0.75	0.03	58.8
En60w80_04	0.60	0.06	38.8	0.57	0.06	33.5	0.66	0.04	43.2	0.68	0.03	49.0	0.67	0.04	45.4	0.71	0.03	56.2	**0.78**	0.02	60.1
En80w80_01	0.55	0.07	52.4	0.52	0.07	45.3	0.62	0.05	56.8	0.64	0.04	63.9	0.63	0.05	59.2	**0.74**	0.04	72.8	0.74	0.03	77.1
En80w80_02	0.56	0.06	52.7	0.53	0.07	45.6	0.63	0.05	57.1	0.65	0.04	64.3	0.64	0.05	59.6	0.68	0.04	73.2	**0.75**	0.03	77.5
En80w80_03	0.57	0.06	53.0	0.54	0.07	46.0	0.64	0.05	57.5	0.66	0.04	64.8	0.65	0.05	60.1	0.69	0.04	73.7	**0.76**	0.03	78.0
En80w80_04	0.58	0.06	53.3	0.55	0.06	46.3	0.65	0.04	58.0	0.67	0.03	65.3	0.66	0.04	60.6	0.70	0.03	74.2	**0.77**	0.02	78.5
+/=/−	16/0/0			16/0/0			16/0/0			16/0/0			16/0/0			12/4/0			–		

表6 算法关于 C-metric 指标的对比结果

算例	C(本文算法, -)						C(-, 本文算法)					
	NSGA-II	MOEA/D	HMOA	NRL	RDMA	AMA	NSGA-II	MOEA/D	HMOA	NRL	RDMA	AMA
En20w80_01	**0.95**	**0.97**	**0.88**	**0.86**	**0.89**	**0.84**	0.05	0.03	0.12	0.14	0.11	0.16
En20w80_02	**0.96**	**0.98**	**0.89**	**0.87**	**0.90**	**0.85**	0.04	0.02	0.11	0.13	0.10	0.15
En20w80_03	**0.94**	**0.96**	**0.87**	**0.85**	**0.88**	**0.83**	0.06	0.04	0.13	0.15	0.12	0.17
En20w80_04	**0.97**	**0.99**	**0.90**	**0.88**	**0.91**	**0.86**	0.03	0.01	0.10	0.12	0.09	0.14
En40w80_01	**0.93**	**0.95**	**0.85**	**0.83**	**0.86**	**0.81**	0.07	0.05	0.15	0.17	0.14	0.19
En40w80_02	**0.92**	**0.94**	**0.84**	**0.82**	**0.85**	**0.80**	0.08	0.06	0.16	0.18	0.15	0.20
En40w80_03	**0.91**	**0.93**	**0.83**	**0.81**	**0.84**	**0.79**	0.09	0.07	0.17	0.19	0.16	0.21
En40w80_04	**0.94**	**0.96**	**0.86**	**0.84**	**0.87**	**0.82**	0.06	0.04	0.14	0.16	0.13	0.18
En60w80_01	**0.90**	**0.92**	**0.80**	**0.78**	**0.81**	**0.76**	0.10	0.08	0.20	0.22	0.19	0.24
En60w80_02	**0.89**	**0.91**	**0.79**	**0.77**	**0.80**	**0.75**	0.11	0.09	0.21	0.23	0.20	0.25
En60w80_03	**0.88**	**0.90**	**0.78**	**0.76**	**0.79**	**0.74**	0.12	0.10	0.22	0.24	0.21	0.26
En60w80_04	**0.91**	**0.93**	**0.81**	**0.79**	**0.82**	**0.77**	0.09	0.07	0.19	0.21	0.18	0.23
En80w80_01	**0.85**	**0.87**	**0.75**	**0.73**	**0.76**	**0.71**	0.15	0.13	0.25	0.27	0.24	0.29
En80w80_02	**0.86**	**0.88**	**0.76**	**0.74**	**0.77**	**0.72**	0.14	0.12	0.24	0.26	0.23	0.28
En80w80_03	**0.87**	**0.89**	**0.77**	**0.75**	**0.78**	**0.73**	0.13	0.11	0.23	0.25	0.22	0.27
En80w80_04	**0.88**	**0.90**	**0.78**	**0.76**	**0.79**	**0.74**	0.12	0.10	0.22	0.24	0.21	0.26
+ / = / −	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0						

算法. HV 指标上的优势表明本文算法获得的解集不仅进一步接近真实帕累托前沿, 而且分布广, 多样性佳. C-metric 指标上的优势, 如 C (本文算法, -) 值大于 C (-, 本文算法), 证明所提算法的解集能支配对比算法的解集, 反之则不成立.

表 5 中算法运行时间结果显示, 相较于传统算法 (NSGA-II、MOEA/D), 本文算法的平均运行时间有所增加. 但与其他学习辅助型优化方法 (NRL、RDMA、AMA) 相比, 本文算法的运行时间处于同一数量级, 且未随问题规模扩大呈现指数增长. 此外, 在多数算例中本文算法在相同次数的迭代而取得的收敛值大于对比算法, 体现出算法的收敛速度与收敛效果权衡. 这说明本文算法增加的计算开销被用于提高帕累托解集质量, 而非无效搜索. 此外, 从物流应用可接受角度, 本文研究的车辆-无人机协同路径规划问题属于离线决策场景 (静态规划问题), 通常在配送开始前集中优化, 而非毫秒级实时控制. 因此, 在优化运营成本和编队工作量均衡的前提下, 本文算法所带来的计算代价在物流实践中是可接受的.

然而, 本文算法并未均优于对比算法, 尤其在实例 En40w80_02, En60w80_01, En60w80_03, En80w80_01 中. 这主要因为混合初始化方法中的贪婪策略可能导致初始种群过早收敛到局部最优, 限制了算法探索解空间的能力. 因此, 为进一步验证性能差异的统计显著性, 实验进行了 Friedman 秩和检验与 Mann-Whitney U 检验, 结果如表 7 和表 8 所示.

表7 对比实验的 Friedman 秩和检验结果

算法	HV		C-metric	
	排名	p-值	排名	p-值
NSGA-II	5.25		5.60	
MOEA/D	7.83		6.58	
HMOA	3.98		3.84	
NRL	3.49	1.56E-10	3.56	1.17E-10
RDMA	3.55		3.64	
AMA	2.33		2.58	
本文算法	**1.12**		**1.32**	

表8 对比实验的 Mann-Whitney U 检验结果

算法	p-值(HV)	p-值(C-metric)
(本文算法, NSGA-II)	1.29E-10	1.26E-10
(本文算法, MOEA/D)	1.49E-11	1.58E-11
(本文算法, HMOA)	4.36E-10	4.21E-10
(本文算法, NRL)	2.03E-09	2.32E-09
(本文算法, RDMA)	4.52E-10	4.74E-10
(本文算法, AMA)	5.33E-05	5.57E-05

表 7 和表 8 结果显示, 本文算法在 HV 和 C-metric 指标的排名均为第一, 且与所有对比算法之间的 p-值均小于 0.05. 结果表明, 算法在求解 MVRPDPDTW 时能够兼顾多样性与收敛性. 同时, 本文算法与对比算法性能之间具有显著性差异.

此外, 为进一步可视化算法优势, 图 18 展示了本文算法与先进算法在 En100w80 算例中获得的帕累托前沿分布. 结果显示所提算法的帕累托前沿分布均匀, 同时解更靠近最优帕累托前沿. 结果表明本文算法能够生成多样性和收敛性优越的路径方案.

所提算法的优越性源于其整体设计. 首先, 与

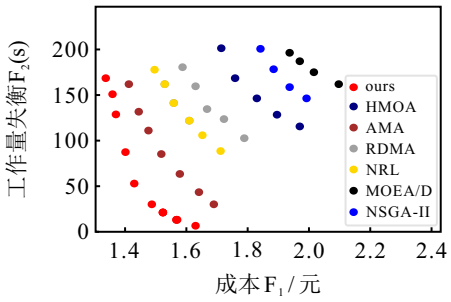


图18　En100w80 算例下对比算法的非支配解集分布图

NSGA-II 和 MOEA/D 相比, 所提算法引入了 PLS, 提高算法的精细化搜索能力. 其次, 与 HMOA 相比, 本文算法引入了基于 RL 的算子选择机制以提高算法的搜索效率. 最后, 与 NRL, RDMA 和 AMA 相比, 所提算法的优势如下三点: (1) 设计基于混合策略的种群初始化方法以保持初始种群的多样性. (2) 利用 GCN 从路线方案的空间拓扑结构中提取关键细节特征, 在现有方法多采用宏观状态表征基础上, 丰富智能体的决策信息. (3) 引入 LSTM 构建交互环境模型, 并结合真实交互数据和虚拟推演数据训练 DQN, 提高现有方法对智能体的训练样本效率. 综上所述, 这些优势使得本文算法能够有效引导局部搜索, 从而获取收敛性和多样性更优的帕累托解集.

4　结　论

本文研究了一种考虑编队工作量均衡的车辆-无人机路径问题, 目标是同时最小化成本和编队工作量失衡. 针对该问题提出了基于模型和图强化学习驱动的多目标优化方法. 首先, 将贪婪策略与随机策略与路径插入法结合, 确保初始种群的多样性. 其次, 通过 GCN 提取路线方案特征并捕捉其复杂结构以增加智能体状态表征信息. 同时, 设计 6 个问题特征的 PLS 操作算子, 实现精细化搜索路径解. 此外, 提出基于 LSTM 的策略优化方法提高 DQN 的训练样本效率, 有效为多目标算法选择最优搜索算子. 最后, 通过参数分析和对比实验证实所提模型和算法的有效性. 同时本文算法在优化运营成本-编队工作量均衡上优于精确求解器 CPLEX 和针对此类问题的多个先进算法, 获得帕累托优的车辆-无人机路径方案.

未来将算法扩展至含有即时物流、动态需求或分布式异构的鲁棒动态车辆-无人机协同路径规划问题[28-30], 是一项重要且具有挑战性的研究方向.

参考文献 (References)

[1] Huang C Q, Fang S F, Wu H, et al. Low-altitude intelligent transportation: System architecture, infrastructure, and key technologies[J]. Journal of Industrial Information Integration, 2024, 42: 100694.

[2] 伍国华, 毛妮, 徐彬杰, 等. 基于自适应大规模邻域搜索算法的多车辆与多无人机协同配送方法[J]. 控制与决策, 2023, 38(1): 201-210.
(Wu G H, Mao N, Xu B J, et al. The cooperative delivery of multiple vehicles and multiple drones based on adaptive large neighborhood search[J]. Control and Decision, 2023, 38(1): 201-210.)

[3] 王俊皓, 李晓玲, 段浩浩, 等. 模因算法求解同时取送货车辆-无人机协同路径优化问题[J]. 控制与决策, 2025, 40(11): 3287-3299.
(Wang J H, Li X L, Duan H H, et al. Memetic algorithm for vehicle-drone collaborative routing problem with simultaneous pickup and delivery[J]. Control and Decision, 2025, 40(11): 3287-3299.)

[4] Murray C C, Chu A G. The flying sidekick traveling salesman problem: Optimization of drone-assisted parcel delivery[J]. Transportation Research — Part C: Emerging Technologies, 2015, 54: 86-109.

[5] Luo Q Z, Wu G H, Ji B, et al. Hybrid multi-objective optimization approach with Pareto local search for collaborative truck-drone routing problems considering flexible time windows[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(8): 13011-13025.

[6] Luo Q Z, Wu G H, Trivedi A, et al. Multi-objective optimization algorithm with adaptive resource allocation for truck-drone collaborative delivery and pick-up services[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(9): 9642-9657.

[7] Mulumba T, Diabat A. Optimization of the drone-assisted pickup and delivery problem[J]. Transportation Research — Part E: Logistics and Transportation Review, 2024, 181: 103377.

[8] Lyu W J, Jin X L, Wang H T, et al. Towards workload-constrained efficient order assignment in last-mile delivery[J]. IEEE Transactions on Mobile Computing, 2025, 24(2): 557-570.

[9] Mancini S, Gansterer M, Hartl R F. The collaborative consistent vehicle routing problem with workload balance[J]. European Journal of Operational Research, 2021, 293(3): 955-965.

[10] Pei J Y, Mei Y, Liu J L, et al. An investigation of adaptive operator selection in solving complex vehicle routing problem[C]. PRICAI 2022: Trends in Artificial Intelligence. Shanghai, 2022: 562-573.

[11] Tian Y, Li X P, Ma H P, et al. Deep reinforcement learning based adaptive operator selection for evolutionary multi-objective optimization[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, 7(4): 1051-1064.

[12] Bektur G. A reinforcement learning-based multiobjective heuristic algorithm for multiple-truck routing problems with heterogeneous drones[J]. Applied Soft Computing, 2024, 167: 112290.

[13] Zhao S C, Zhou H. Learning-driven memetic algorithm

for solving integrated distributed production and transportation scheduling problem[J]. Swarm and Evolutionary Computation, 2025, 96: 101945.

[14] Mara S T W, Sarker R, Essam D, et al. An adaptive memetic algorithm for a cost-optimal electric vehicle-drone routing problem[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(12): 19619-19632.

[15] Wang W Q, Adulyasak Y, Cordeau J F, et al. The heterogeneous-fleet electric vehicle routing problem with nonlinear charging functions[J]. Transportation Research Part C: Emerging Technologies, 2025, 170: 104932.

[16] Jiang Y, Liu M M, Jia X B, et al. The multi-visit vehicle routing problem with multiple heterogeneous drones[J]. Transportation Research — Part C: Emerging Technologies, 2025, 172: 105026.

[17] Zhang Q C, Shao W S, Shao Z S, et al. Graph-based reinforced multi-objective optimization for distributed heterogeneous flexible job shop scheduling problem under nonidentical time-of-use electricity tariffs[J]. Expert Systems with Applications, 2025, 290: 128428.

[18] 贾兆红, 王少贵, 刘闯. 多模式下的车辆和无人机联合配送模型与优化算法[J]. 控制与决策, 2024, 39(7): 2125-2132.
(Jia Z H, Wang S G, Liu C. Vehicle and drones joint distribution model and optimization algorithm in multi-mode[J]. Control and Decision, 2024, 39(7): 2125-2132.)

[19] 罗永琪, 陈彦如, 冉茂亮. 带收益和时间窗的多行程卡车-无人机协同配送问题[J]. 控制与决策, 2025, 40(6): 1817-1826.
(Luo Y Q, Chen Y R, Ran M L. Multi-trip truck-drone routing problem with profits and time windows[J]. Control and Decision, 2025, 40(6): 1817-1826. [自助补缺].)

[20] Ke L J, Zhang Q F, Battiti R. Hybridization of decomposition and local search for multiobjective optimization[J]. IEEE Transactions on Cybernetics, 2014, 44(10): 1808-1820.

[21] Moshref-Javadi M, Hemmati A, Winkenbach M. A truck and drones model for last-mile delivery: A mathematical model and heuristic approach[J]. Applied Mathematical Modelling, 2020, 80: 290-318.

[22] Manessi F, Rozza A, Manzo M. Dynamic graph convolutional networks[J]. Pattern Recognition, 2020, 97: 107000.

[23] Yang Z Y, Fu M S, Qu H, et al. Incremental model-based reinforcement learning with model constraint[J]. Neural Networks, 2025, 185: 107245.

[24] Wang J Q, Liu K, Li H T, et al. Vehicle trajectory prediction using hierarchical LSTM and graph attention network[J]. IEEE Internet of Things Journal, 2025, 12(6): 7010-7025.

[25] Rudolph G. Convergence analysis of canonical genetic algorithms[J]. IEEE Transactions on Neural Networks, 1994, 5(1): 96-101.

[26] Kuo R J, Edbert E, Zulvia F E, et al. Applying NSGA-II to vehicle routing problem with drones considering makespan and carbon emission[J]. Expert Systems with Applications, 2023, 221: 119777.

[27] Guo X, Miao Z H, Pan Q K, et al. Hybrid loading situation vehicle routing problem in the context of agricultural harvesting: A reconstructed MOEA/D with parallel populations[J]. Swarm and Evolutionary Computation, 2024, 91: 101730.

[28] 刘长石, 陈厚吏, 马艺璇, 等. 即时物流系统中骑手+无人机协同配送的路径规划[J]. 控制与决策, DOI: 10.13195/j.kzyjc.2025.1069.
(Liu C S, Chen H L, Ma Y X, et al. Routing for rider-drone collaborative delivery in the real-time logistics system[J]. Control and Decision, DOI: 10.13195/j.kzyjc.2025.1069.)

[29] 马梓元, 冯鹏宇, 龚华军, 等. 基于改进图神经网络算法的异构多智能体动态任务分配[J]. 控制与决策, DOI: 10.13195/j.kzyjc.2025.0887.
(Ma Z Y, Feng P Y, Gong H J, et al. Dynamic task allocation for heterogeneous multi-agent systems based on an improved graph neural network algorithm[J]. Control and Decision, DOI: 10.13195/j.kzyjc.2025.0887.)

[30] 孟祥恒, 郭鹏, 李嘉雯, 等. 基于多智能体深度强化学习的轨道车辆组装分布式异构柔性作业调度[J]. 控制与决策, DOI: 10.13195/j.kzyjc.2025.0857.
(Meng X H, Guo P, Li J W, et al. Distributed heterogeneous flexible job shop scheduling for railway vehicle assembly using multi-agent deep reinforcement learning[J]. Control and Decision, DOI: 10.13195/j.kzyjc.2025.0857.)

作者简介

杨明园 (2001−), 男, 博士生, 主要研究方向为智能优化算法与应用, E-mail: heu_0407_ymy@hrbeu.edu.cn;

王伟 (1979−), 男, 教授, 博士, 博士生导师, 主要研究方向为智能控制理论, E-mail: wangwei407@hrbeu.edu.cn.