双种群多智能体辅助的周期性车辆-无人机协同路径规划

杨明园，王 伟†，董福王

(哈尔滨工程大学 智能科学与工程学院，哈尔滨 150001)

摘 要: 针对周期性冷链运输中的高碳排放与低服务质量挑战, 提出多周期车辆-无人机协同路径问题, 并以最小化碳排放和最大化客户满意度建立双目标混合整数线性规划模型. 为提高该 NP-hard 问题的求解效率, 提出双种群多智能体引导的多目标优化方法. 首先, 提出双种群演化初始化方法, 通过双目标偏置搜索和种群知识迁移为多智能体引导搜索提供兼具质量与多样性的初始解; 进一步提出多智能体强化学习引导的搜索方法, 设计两个领导者智能体从启发式算子池中序贯地选择扰动和改进算子以协同引导多目标方法的搜索方向, 平衡其收敛性与多样性; 此外, 为缓解多智能体引导过程中独立学习架构的非平稳问题, 提出基于对手建模的策略学习方法, 通过在策略更新中显式刻画对手行为特征, 提高训练的稳定性; 最后, 通过拓展基准测试集. 仿真结果表明, 所提算法在收敛性和多样性上优于精确求解器和先进算法, 可为构建绿色高效的冷链运输系统提供技术支持.

关键词: 低空经济；车辆-无人机协同；周期性运输；多目标进化算法；多智能体强化学习；双种群演化；算子选择

中图分类号: TP312 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2026.0605

引用格式: 杨明园, 王伟, 董福王. 双种群多智能体辅助的周期性车辆-无人机协同路径规划 [J]. 控制与决策, xxxx, x(x): xxxx-xxxx.

Dual-population multi-agent-aided optimization method for periodic vehicle-drone cooperative routing problem

Yang Mingyuan，*Wang Wei*†，*Dong Fuwang*

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

Abstract: To reduce carbon emissions and improve service quality in cold-chain transportation, this study proposes a periodic vehicle routing problem with drones (PVRPD). A bi-objective mixed-integer linear programming (MILP) model is developed to minimize carbon emissions and maximize customer satisfaction. To solve this NP-hard problem efficiently, a dual-population multi-agent-aided multi-objective evolutionary algorithm (DMaMOEA) is proposed. First, a dual-population initialization strategy uses objective-biased search and knowledge transfer to generate high-quality and diverse initial solutions. Second, two agents sequentially select exploration and exploitation operators to balance convergence and diversity. An opponent-model-based learning method is further developed to reduce the non-stationarity caused by independent learning and stabilize policy updates. Experiments on extended benchmark instances show that DMaMOEA achieves competitive results against exact solvers on small-scale instances and outperforms advanced algorithms in convergence and diversity. The proposed method supports the development of greener and more efficient cold-chain transportation systems.

Keywords: low-altitude economy； vehicle-drone coordination； periodic transport； multi-objective evolutionary algorithm；multi-agent reinforcement learning；dual-population evolution；operator selection

0 引 言

冷链运输已成为确保温敏产品 (如药品、生鲜) 完整性与安全性的重要环节. 然而, 相较于传统场景, 冷藏车与温控系统易产生碳排放问题[1]. 其次, 冷链运输多呈现周期性 (如按周、月或固定频次服务), 且考虑时间窗、易腐性和冷藏负荷等因素, 易引发配送延误和产量质量问题, 进而削弱客户满意度[2].

随着低空经济的发展, 集成无人机与车辆的混

收稿日期: 2026-06-21；**录用日期:** 2026-09-20.
基金项目: 国家自然科学基金项目 (62271163).
责任编委: 龙建成.
†通信作者. E-mail: wangwei407@hrbeu.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.

合运输系统为上述挑战提供解决方案. 具体而言, 无人机具备快速响应和低能耗等优势, 而车辆支持高载荷和长续航; 通过将车辆作为无人机的移动起降平台与能量补给站, 无人机作为其配送助手, 可有望降低冷链运输碳排放并提高配送效率[3]. 然而, 车辆-无人机路径问题主要针对单周期场景, 如考虑时变路网[4]、时间窗[5]、多仓库[5] 等因素, 难以充分刻画周期间的路径耦合关系. 此外, 将单周期路径方案移植到多周期场景中可能导致配送效率次优[6]. 由于车辆-无人机路径问题是一个多约束多目标的组合优化问题 (NP-hard), 其难以被精确求解[7]. 因此, 为提高求解效率, 本文进一步研究多目标进化算法.

多目标优化算法的求解效率取决于两个关键阶段: 种群初始化和搜索优化; 且初始种群的分布质量将直接影响优化阶段的算法收敛性与多样性[4]. 然而, 现有初始化方法难以保证初始种群质量: 随机初始化因解集无法均匀覆盖目标空间而收敛性不足[4]; 启发式初始化因偏向单一目标优化, 在多目标冲突关系下降低其他目标的质量. 为此, 本文引入一种新的初始化策略–演化多任务优化方法, 通过并行优化多任务和跨任务知识迁移机制, 为提升初始种群质量与多样性提供新途径[8]. 在搜索优化阶段, 为解决传统算子选择策略自适应搜索能力差的问题, 现有强化学习驱动的进化算法通过学习与搜索算子交互的奖励反馈, 在进化过程中能倾向选择产生更优后代的算子[9]. 虽然该方法提升了搜索能力, 但其多侧重于学习开发算子 (如邻域搜索[10]) 或探索算子 (如遗传交叉[11]、亲本匹配[12]) 的单一偏好, 缺乏在进化过程中维持探索-利用平衡的统一机制. 值得注意的是, 维持探索-利用平衡可以有效避免算法早熟并持续逼近真实帕累托前沿[13]. 因此, 这种单一的学习偏好可能会降低多目标算法的整体优化性能. 同时, 现有分层强化学习方法通过价值估计或层级决策可以缓解探索-利用失衡问题[14], 但是其均建立在单智能体框架下, 难以刻画搜索过程中全局探索与局部开发之间的异质性与交互关系, 限制了多策略协同决策能力. 值得注意的是, 由于多智能体引导过程中其状态转移并非独立, 传统独立学习 (IL) 架构存在非平稳问题[15], 难以有效引导多目标优化算法的搜索进程.

基于上述问题, 本文提出双种群多智能体引导的多目标进化方法, 通过引入双种群和多智能体, 提升多目标算法的搜索效率. 本文贡献如下: (1) 提出多周期车辆-无人机路径问题, 通过引入能耗、制冷、周期性、时间窗和易腐性等冷链特征, 建立最小化碳排放与客户非满意度的双目标混合整数线性规划模型, 提供绿色高效的配送方案. (2) 提出双种群演化初始化方法, 为多智能体引导的搜索提供优质起点. 具体而言, 在双目标偏置搜索生成低碳导向和高满意度导向的主-辅种群后, 引入知识转移机制在主-辅种群内部和之间实现结构重组与信息交互. 进一步结合快速非支配排序准则从进化候选种群中筛选出兼顾质量与多样性的初始解. (3) 提出多智能体强化学习引导的搜索方法, 并将其作为多目标算法的自适应算子选择机制. 具体而言, 通过设计探索和开发智能体, 通过从算子池中序贯地选择扰动和改进算子并执行相应搜索操作, 协同指导多目标优化算法的探索和开发方向, 平衡其进化过程的收敛性与多样性. (4) 提出基于对手建模的策略学习方法, 通过显式建模对手策略分布 (如开发智能体), 将其非平稳的策略演化建模为可观测的增广状态, 提高训练的稳定性. 具体而言, 在蒸馏对手策略的 Q 值函数后, 将其策略估计通过全连接神经网络嵌入到探索智能体的状态中, 进而生成具有对手感知能力的增强状态表征. 进一步学习对手条件动作价值函数, 增强状态的探索智能体实现条件平稳地策略更新.

1 问题描述与建模

1.1 问题描述

如图 1, PVRPD 在服务周期内预设客户位置、访问频次和预期访问时间窗. 系统主体由单一车辆构成, 并搭载若干同质无人机, 从仓库出发执行配送. 整个服务周期被划分为多个子周期 (如 7 个), 在每一子周期内, 无人机辅助车辆完成部分配送任务; 车辆负责其余的配送任务, 并在途中发射和回收无人机. 在该子周期结束时, 车辆及其无人机返回仓库并准备执行下一子周期任务, 直至周期循环结束.

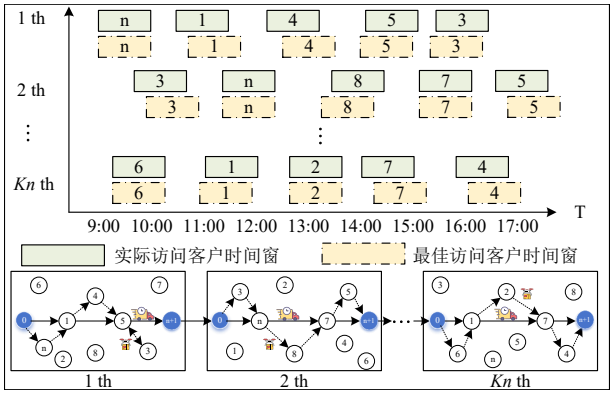

图 1 PVRPD 问题示意图

PVRPD 模型构建如下: 假设图$\mathcal{G}=(V,E)$, 其中V是顶点集, E是边集. 集合V包含与客户相关联的n个顶点, 记作$V_N=\{1,2,\ldots,n-1,n\}$, 和仓库

节点顶点0和$n+1$. 因此, $V=\{0,1,\ldots,n+1\}$, 其中$0\equiv n+1$.

本文考虑一个周期集合$\mathcal{K}=\{1,\ldots,K_n\}$, $K_n\in\mathbb{Z}_{>0}$, $|\mathcal{K}|=K_n$. 此外, 车辆配备一个无人机集合$\mathcal{D}=\{1,\ldots,D_n\}$, $D_n\in\mathbb{Z}_{>0}$, $|\mathcal{D}|=D_n$. 一次无人机作业路径由一个三元组(i,w,j)表述: 无人机在顶点$i\in V_L$从车辆发射, 将包裹送达$w\in V_N$后, 再由同一辆车在顶点$j\in V_R$回收 ($i\neq w, w\neq j$). 同时, 无人机作业路径(i,w,j)小于其最大飞行距离$\mathcal{E}$. 本文参数和变量总结在表 1.

表 1　符号定义

符号	说明
V_D	$V_D\subseteq V_N$称为可由无人机服务的客户子集.
V_L	$V_L=V\backslash\{n+1\}$为V中可发射无人机的顶点子集.
V_R	$V_R=V\backslash\{0\}$为可回收无人机的顶点子集.
d_{ij}, $\bar{d}ij$	车辆和无人机经过边(i,j)的行驶距离. $i,j\in V$.
t_{ij}, $\bar{t}ij$	车辆和无人机经过边(i,j)的行驶时间. $i,j\in V$.
$[g_i^k,b_i^k]$	客户i在周期k的预期时间窗. $i\in V_N, k\in\mathcal{K}$.
$[o_i^k,m_i^k]$	客户i在周期k的可容忍时间窗. $i\in V_N, k\in\mathcal{K}$.
$[e_i^k,l_i^k]$	客户i在周期k的可容忍产品损毁率. $i\in V_N, k\in\mathcal{K}$.
α_1、α_2、α_3	发动机模块系数、速度模块系数和载重模块系数.
v	车辆均匀行驶速度(m/s).
W_0、w_{ij}^k	车辆自重和在周期k时行驶边(i,j)时的负载重量.
η	碳排放因子(kg/L).
a_{n+1}^k	车辆在周期k返回仓库时刻, 即配送全程所花费的时间.
t_s、t_p	车辆在客户点的服务时间和为无人机的装货时间.
β	制冷设备能耗率(L/h).
c_1, c_2	时效性与新鲜度占比客户满意度权重.
ξ_1、ξ_2	产品腐败率和打开车门卸货时产品腐败率.
$tt_s^{j,k}$	在访问客户j之前服务客户的时间之和.
M	一个较大的正整数.
F_j	顾客j的访问频次, $1\leqslant F_j\leqslant K_n, j\in V_N$.
x_{ij}^k	若车辆在周期k经过边(i,j), 为1, 否则为0. $i,j\in V, k\in\mathcal{K}, i\neq j$.
y_{iwj}^{kd}	若无人机d在周期k执行作业(i,w,j), 为1, 否则为0. $i\in V_L, w\in V_D, j\in V, k\in\mathcal{K}, d\in\mathcal{D}, i,j\neq w$.
p_{ij}^k	若在周期k顶点i在车辆的路径中先于顶点j被服务, 为1, 否则为0. $i,j\in V, k\in\mathcal{K}, i\neq j$.
u_i^k	顶点i在周期k时于车辆路径中的位置. $i\in V, k\in\mathcal{K}$.
z_{aiwje}^{kd}	若无人机d在车辆到达i前于顶点a发射, 并在车辆到达j之后于顶点e回收, 为1, 否则为0. $|a,i,w,j,e|=5$表示所有顶点互不相同. 用于指定无人机航线沿车辆方向进行.
a_i^k	车辆在周期k在顶点i处到达时间. $i\in V, k\in\mathcal{K}$.
s_i^k	车辆在周期k从顶点i处出发的最早时间. $i\in V, k\in\mathcal{K}$.
a_i^{kd}	在周期k的无人机d在顶点i处到达时间. $i\in V, k\in\mathcal{K}$, $d\in\mathcal{D}$.
r_i^{kd}	在周期k的无人机d在顶点i处最早被所释放的时间. $i\in V, k\in\mathcal{K}, d\in\mathcal{D}$.
s_i^{kd}	在周期k的无人机d最早从顶点i处的离开时刻. $i\in V, k\in\mathcal{K}, d\in\mathcal{D}$.

1.2　数学模型设计

车辆碳排放f_1由行驶 (ED) 和制冷 (EC) 组成:

$$\min f_1=\eta(ED+EC), \tag{1}$$

$$ED=\sum_{k\in\mathcal{K}}\sum_{i\in V_L}\sum_{j\in V_R}x_{ij}^k\{\alpha_1 t_{ij}+\alpha_2 d_{ij}v^2+\alpha_3(W_0+w_{ij}^k)d_{ij}\}, \tag{2}$$

本文采用综合模式排放模型 (CMEM) 计算车辆的油耗量 (ED), 其与车辆载重和运动状态相关[16].

$$EC=\beta\{\sum_{k\in\mathcal{K}}\sum_{i\in V_L}\sum_{j\in V_R}x_{ij}^k(a_{n+1}^k+t_s)+\sum_{k\in\mathcal{K}}\sum_{d\in\mathcal{D}}\sum_{i\in V_L}\sum_{w\in V_D}\sum_{j\in V_R}y_{iwj}^{kd}t_p\}, \tag{3}$$

假设为无人机装货或为客户卸货, 制冷机组需额外提升功率以补偿温差[16]. 因此, 制冷油耗 (EC) 包含配送全程的正常制冷 (a_{n+1}^k) 和额外制冷 (t_s、t_p).

顾客满意度与访问时间和产品新鲜度相关. 客户满意度和非满意度之和为 1. 为便于双目标优化, 通过 $(1-\mu_w^{kd})$ 将最大化满意度转化最小化非满意度:

$$\min f_2=\sum_{k\in\mathcal{K}}\sum_{d\in\mathcal{D}}\sum_{i\in V_L}\sum_{w\in V_D}\sum_{j\in V_R}y_{iwj}^{kd}[c_1(1-\mu_w^{kd})+c_2(1-\nu_w^{kd})]+\sum_{k\in\mathcal{K}}\sum_{i\in V_L}\sum_{j\in V_R}x_{ij}^k[c_1(1-\mu_j^k)+c_2(1-\nu_j^k)], \tag{4}$$

其中, μ_j^k表示因遵守服务时间窗的客户满意度:

$$\mu_j^k=\begin{cases}\dfrac{a_j^k-o_j^k}{g_j^k-o_j^k}, o_j^k\leqslant a_j^k<g_j^k\\ 1, g_j^k\leqslant a_j^k<b_j^k\\ \dfrac{m_j^k-a_j^k}{m_j^k-b_j^k}, b_j^k\leqslant a_j^k\leqslant m_j^k.\\ 0, \text{其他情况}\end{cases} \tag{5}$$

其中, ν_j^k为考虑产品质量损失下的客户满意度:

$$\nu_j^k=\begin{cases}1, D_j^k\leqslant e_j^k\\ \dfrac{l_j^k-D_j^k}{l_j^k-e_j^k}, e_j^k<D_j^k\leqslant l_j^k,\\ 0, l_j^k<D_j^k\end{cases} \tag{6}$$

其中, D_j^k为考虑送达时间的产品质量腐败函数[16]:

$$D_j^k=1-e^{-\xi_1(a_j^k-tt_s^{j,k})-\xi_2 tt_s^{j,k}}, \tag{7}$$

本文针对 PVRPD 问题设计了车辆与无人机路径流量守恒、空间和时间同步等 3 类, 共 22 条约束, 见附件式 (A1)–式 (A22). 此外, 附件图 A1 采用ε-约束法展示了碳排放和客户非满意度目标之间存在冲

突关系, 即优化一个目标以牺牲另一个目标为代价.

2 双种群多智能体辅助的多目标优化方法

2.1 方法框架

如图 2 所示, 在初始化阶段, 双种群演化初始化方法通过双目标搜索偏置和知识迁移机制, 生成兼具质量和多样性的初始种群P_0. 在进化阶段, 首先从精英解集的车辆-无人机路径解中提取并融合结构特征, 构建智能体状态s_t. 其次, 探索和开发智能体$(e、x)$依次对路径解选择并执行全局搜索算子 (如N_3) 和局部搜索算子 (如N_4), 以协同开发解空间、改进解质量 (新解s_{t+1}). 同时将多智能体与多目标方法的交互样本存入经验回放池. 当回放池收集到最小数目的训练样本后, 基于对手建模的策略学习方法通过状态增强和学习训练, 提高策略更新 (θ_e, θ_x) 的稳定性. 在此阶段, 精英归档策略更新并存储全局最优解. 最后, 迭代过程持续至预设条件, 确保智能体有效引导多目标方法在探索-开发之间保持平衡.

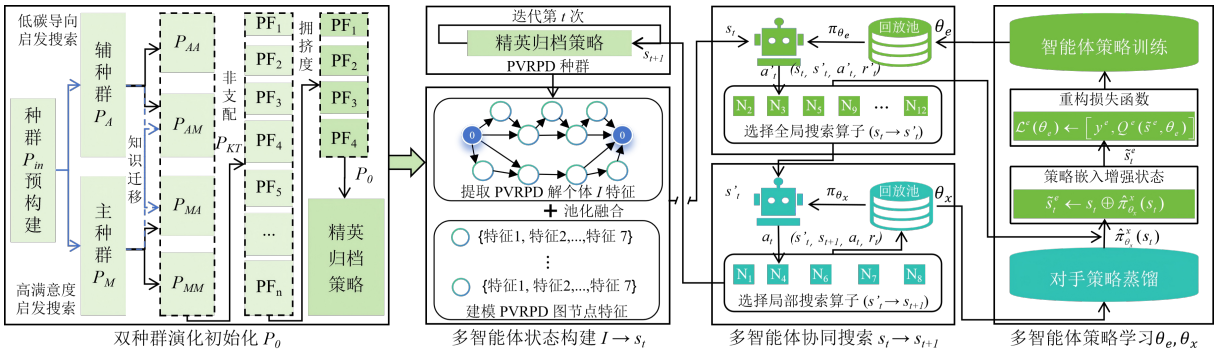

图 2 双种群多智能体辅助的多目标优化算法框架

2.2 解的编码表示

受文献启发, PVRPD 的解由 7 个部分构成[17]. 在图 3 中, 第一部分为车辆路线, 由其访问的客户编号组成. 第二部分的 0 值代表仓库编号. 第三部分为访问周期编号. 第四和第六部分是无人机发射与回收节点, 由第一部分车辆访问的客户编号索引表示. 第五部分为无人机访问的客户, 由客户编号呈现. 第七部分为参与配送的无人机编号.

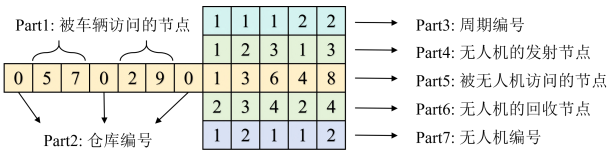

图 3 染色体编码示意图

染色体编码示意图 3 对应路径方案图 4. 染色体$\langle 0,5,7,0,2,9,0\rangle$表示车辆在周期 1 从仓库出发依次访问客户 5 和 7 后返回仓库, 然后在周期 2 访问客户 2 和 9. 染色体$\langle 1,1,1,2,1\rangle$表示在周期 1, 无人机 1 从仓库 (车辆路径的第 1 个节点) 起飞, 访问客户 1 后, 被车辆回收于客户 5 (车辆路径的第 2 个节点).

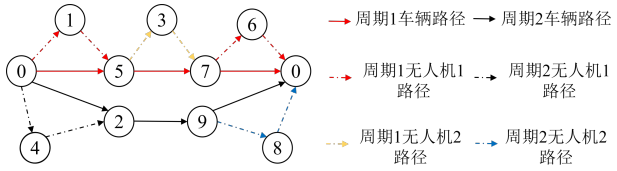

图 4 周期性车辆-无人机协同路径示意图

2.3 双种群演化初始化

受演化多任务优化启发[8], 本文结合 PVRPD 的问题特征, 设计双种群演化初始化方法. 该方法通过初始种群预构建、双目标偏置搜索、知识迁移和个体筛选, 生成兼具质量与多样性的初始种群.

首先, 采用问题感知启发式方法构建预初始化种群P_{in}. 对于客户$c \in V_N$, 根据其服务频率F_c随机分配服务子周期, 形成客户–服务计划对$\{c, (\kappa_1, \kappa_2, \ldots, \kappa_{F_c})\}$及周期调度方案$S = (S_1, S_2, \ldots, S_{|\mathcal{K}|})$, κ_1表示客户首次服务的子周期索引, S_t表示在子周期t接受服务的客户集合. 随后, 根据客户需求重量将S_t划分为车辆服务集合S_t^v和无人机可服务集合S_t^d. 从S_t^d中随机选取子集$\overline{S_t^d}$并将其与S_t^v合并, 通过随机启发式构造车辆路径 $R_t^v = (0, \ell_1, \ell_2, \ldots, \ell_{|S_t^v|+|\overline{S_t^d}|}, n+1)$, 其中节点0和$n+1$表示同一仓库. 然后, 在满足无人机续航约束的条件下, 采用插入策略选择车辆路径中的发射点和回收点, 构造无人机路径 $R_t^d = \{(\chi_1, \psi_1, \omega_1), \ldots, (\chi_q, \psi_q, \omega_q)\}$, $q = |S_t| - |R_t^v| + 2$, $(\chi_i, \omega_i) \in R_t^v$, 且$\psi_i \in S_t \setminus R_t^v$. 重复上述过程, 直至完成所有子周期的路径规划, 得到个体 $I = \{(R_1^v, R_1^d), \ldots, (R_{|\mathcal{K}|}^v, R_{|\mathcal{K}|}^d)\}$.重复生成个体, 直至达到种群规模$ps$, 从而获得 $P_{in} = (I_1, I_2, \ldots, I_{ps})$.

随后, 对P_{in}中的每个个体I, 通过双目标偏置搜索构造主种群P_M和辅助种群P_A. P_M将满足无人机载重和续航约束的部分车辆服务客户转由无人机配

送以降低碳排放; P_A通过贪婪交换车辆与无人机服务客户的访问顺序, 减少时间窗违约, 提高客户满意度. 因此, 双种群分别偏向低碳和高满意度区域.

为避免种群同质化, 引入知识迁移机制实现种群内部和种群之间的结构重组. 令$P = P_M, P_A$, 对任一种群$\overline{P}$中的个体I, 以概率$\rho = 0.5$选择进化方式[8]: 当随机数大于ρ时, 从当前种群$\overline{P}$中通过锦标赛选择交配个体I'; 否则从互补种群$P \setminus \overline{P}$中选择I'. 随后执行交叉操作生成子代, 将所有子代存入知识迁移种群P_{KT}. 最后, 依据非支配排序从P_{KT}中筛选ps个个体构成初始种群P_0, 完成种群初始化.

2.4　多智能体引导的自适应算子选择方法

为避免 Q 学习收敛发散, 本文采用深度 Q 网络 (DQN) 完成自适应算子选择, 其过程被建模为马尔可夫决策过程[18], 如下表示$\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$:

1) 状态空间$\mathcal{S}$: 探索和开发智能体共享相同的状态空间. 为了充分感知当前周期性车辆-无人机协同路径方案属性, 状态模型通过多层感知机提取车辆-无人机路径解的 7 个特征 (客户索引、服务周期索引、无人机访问时间、无人机离开时间、服务次序、车辆访问时间、车辆离开时间) 并建模为特征矩阵$X \in \mathbb{R}^{n \times 7}$. 其中, x_i为第i个节点的特征向量. 随后, 通过平均池化生成状态嵌入向量$\mathcal{Z} = \dfrac{\sum_{i=1} x_i}{n}$. 该向量作为智能体的状态包含路径方案的结构和节点信息. 状态空间$\mathcal{S}$由所有路径解构建的嵌入向量$\mathcal{Z}$组成. 多层感知机的核心参数为: 输入层规模为 **7**, 输出层维度为 **32**, 两层隐藏层维度为 **64**.

2) 动作空间$\mathcal{A}$: 表 2 针对 PVRPD 定制 12 个启发式算子以构建动作空间, 见附件图 B1–图 B12, 为搜索操作提供多样化选择证[18]. 由于探索与开发算子具有异质性, 将两类算子置于同一动作空间会因策略不兼容而导致学习效率下降. 针对上述问题, 本文设计双智能体分工架构: 探索智能体e和开发智能体x具有不同的动作空间 ($|\mathcal{A}_e| = 7, |\mathcal{A}_x| = 5$). 动作空间$\mathcal{A}$由算子池构成, 维度$|\mathcal{A}|$由算子数量决定.

表 2　启发式算子池

名称	算子操作内容	开发算子	探索算子
N_1: 2-*Opt*	删除路径中的两条边并重新连接其端点, 以重构访问顺序.	✓	
N_2: 贪婪删除插入	删除代价最高的无人机路径, 并以贪婪策略重新分配无人机配送任务.		✓
N_3: 随机删除插入	随机移除一个无人机子路径, 并重新插入一个可行的无人机配送任务.		✓
N_4: 无人机行程交换	交换无人机航路中的两个节点以调整无人机配送顺序.	✓	
N_5: 车辆转无人机	将车辆路径中的一个客户节点转移至无人机路径中.		✓
N_6: 无人机转车辆	将无人机路径中的一个客户节点移动至车辆路径中.	✓	
N_7: 跨周期交换	在不同车辆-无人机协同配送周期之间交换两个客户节点.	✓	
N_8: 无人机行程重定位	重新分配无人机的起飞与回收节点.	✓	
N_9: 车辆路径跨周期重定位	跨周期将一个客户节点移动至另一车辆路径中.		✓
N_{10}: 车辆路径跨周期交换	在不同车辆路径配送周期之间交换两个客户节点.		✓
N_{11}: 无人机航程跨周期交换	在不同无人机路径配送周期之间交换两个客户节点.		✓
N_{12}: 无人机航程跨周期重定位	跨周期将一个客户节点移动至另一无人机路径中.		✓

3) 状态转移$\mathcal{T}$: 开发智能体x的状态转移定义为: $s_t \xrightarrow{a_t} s_{t+1}$= 开发算子$(s_t, a_t)$. 与其不同, 探索智能体$e$更关注探索算子的选择对下一代或者未来种群质量的影响[12], 则其状态转移定义为: $s_t \xrightarrow{a'_t} s'_t$ =探索算子$(s_t, a'_t) \xrightarrow{a_t} s_{t+1}$= 开发算子$(s'_t, a_t)$.

4) 奖励函数$\mathcal{R}$: 为避免多智能体存在奖励计算复杂度高和奖励稀疏问题, 首先引入基于删除-添加策略的增量式贡献度计算. 具体而言, 通过维护一个动态的支配树, 当新解注入种群时, 仅在该解所支配的子空间区域内进行局部体积积分, 而非扫描整个非支配前沿. 该机制将单步奖励评估的时间复杂度由$\mathcal{O}(n^2)$降低至$\mathcal{O}(k \log n)$, 其中$k = 2$为优化目标数, n为种群规模. 其次, 不同于单一奖励信号, 本文设计了一种混合奖励函数:

$$r(s_{t+1}, s_t) = \Delta HV + \mathbb{I}(\Delta HV = 0) \cdot \Delta Crowd, \quad (8)$$

其中, $\Delta Crowd$为新解引入后对邻域拥挤度的提升量, 即使在 HV 奖励停滞, $\Delta HV = 0$时, 该信号依然保持连续. 具体来说, 引入了前沿位移势能作为内在奖励, 当种群拓展帕累托前沿, 即使没有增加收敛性, 但增加多样性时, 智能体依然获得恒定的正向奖励. 通过这种设计, DQN 的时序差分误差在训练早期被有效激活, 缓解策略网络收敛震荡问题.

$$\Delta HV = \frac{HV(s_{t+1}) - HV(s_t)}{HV(s_t)}, \tag{9}$$

其中, HV 为算法搜寻非支配解x的适应值与参考点之间围成的超体积[17], 通常用于评价多目标优化算法收敛性和多样性的综合表现, 其计算方法如下:

$$HV(\Omega, P) = \bigcup_{x\in\Omega} \nu(x, P), \tag{10}$$

其中, P 是参考点, 为迭代中最差解. Ω为算法搜寻的非支配解集. HV 越大, 则算法收敛性与多样性越优.

2.5 基于对手建模的策略学习方法

由于探索智能体的状态转移依赖于开发智能体, 传统独立学习架构存在非平稳性问题[15]. 为此, 提出基于对手建模的策略学习方法, 相较于传统架构, 探索智能体e采用对手行为的增强状态表示:

$$\tilde{s}_t^e = s_t \oplus \hat{\pi}_{\theta_x}^x(s_t), \tag{11}$$

其中, $\tilde{s}_t^e$表示通过全连接神经网络对s_t和$\hat{\pi}_{\theta_x}^x(s_t)$的拼接操作. s_t表示探索智能体e的原始状态. $\hat{\pi}_{\theta_x}^x(s_t)$表示通过 Softmax 变换将开发智能体$x$的 Q 值函数蒸馏为其在线估计的策略分布[19]:

$$\hat{\pi}_{\theta_x}^x(s_t) = \frac{\exp(Q^x(a_t, s_t)/\tau)}{\sum\limits_{a'\in\mathcal{A}^x} \exp(Q^x(a', s_t)/\tau)}, \tag{12}$$

其中, τ为估计策略分布的平滑度. 在此基础上, 通过建模探索智能体状态与开发智能体策略分布间的映射关系, 其策略更新实现对搜索行为的在线估计, 进而提高探索智能体策略更新的稳定性.

探索智能体损失函数采用随机梯度下降在经验回放池Ξ^e中均匀采样小批量数据, 以均方误差定义:

$$\mathcal{L}^e(\theta_e) = \mathbb{E}_{U(\Xi^e)}[(y^e - Q^e(\tilde{s}^e, a^e, \theta_e))^2], \tag{13}$$

2.6 参数敏感性实验

其中, $Q^e(\tilde{s}_t^e, a_t^e, \theta_e)$为对手策略条件化的动作价值函数; 目标值$y_t^e$则依据贝尔曼最优方程构造:

$$Q^e(\tilde{s}_t^e, a_t^e, \theta_e) = Q^e(s_t, \hat{\pi}_{\theta_x}^x(s_t), a_t^e, \theta_e), \tag{14}$$

$$y_t^e = r_t^e + \gamma_e \max_{a'\in\mathcal{A}_e} Q^e(s_t', \hat{\pi}_{\theta_x}^x(s_t'), a', \theta_e^-), \tag{15}$$

其中, θ_e^-为目标网络参数. 该设计通过目标网络与经验回放机制保证训练稳定性, 并更新探索策略θ_e.

开发智能体的策略训练方法是 DQN[18]. 而探索智能体的策略学习流程如下: 采样: 探索智能体与环境交互, 将交互过程产生的经验样本$(s_t, a_t^e, r_t^e, s_{t+1})$存储至回放池$\Xi^e$. 抽样: 在每个决策点, 从经验回放池$\Xi^e$中均匀随机采样一个最小批次样本集$(s, a^e, r^e, s')$. 蒸馏: 基于采样状态$s_t$, 利用对手策略蒸馏估计开发智能体的策略分布$\hat{\pi}_{\theta_x}^x(s_t)$, 式 (12). 增广: 将对手策略分布嵌入状态信息, 构建探索智能体的增广状态表示, 式 (11). 并据此构造增广批次样本集$(\tilde{s}^e, a^e, r^e, \tilde{s}^{e'})$. 计算: 定义对手条件化动作价值函数, 式 (14). 利用目标网络参数θ_e^-计算贝尔曼更新目标, 式 (15). 更新: 通过随机梯度下降最小化均方贝尔曼误差损失, 式 (13), 以更新探索智能体的Q网络参数θ_e. 同时, 每间隔C步对目标网络参数进行软更新. 此外, 两个智能体网络具有相同的架构: 一个两层全连接神经网络, 第一层有 64 个单元, 第二层有 32 个单元. 重放缓冲区大小为 5 000, 最小批次样本大小为 64, 目标网络更新频率C = 50 步.

3 计算实验

3.1 实验设置

为评估算法性能, 实验基于扩展基准实例进行测试. 算法在英特尔 i7-146500HX (16GB RAM) 处理器、NVIDIA RTX 4 050 及 Windows11 系统的计算机上运行. 代码采用 Python3.10、PyTorch1.13 编写.

3.1.1 评价指标

超体积 (Hypervolum, HV) 和覆盖率 (C-metric) 为评估本文算法的性能指标[17, 20]. HV 计算方法如公式 (10). C-metric 指标计算的公式如下:

$$C(A, B) = \frac{\{|b \in B| \exists a \in A : a \preceq b|\}}{|B|}, \tag{16}$$

其中, $|B|$表示非支配解集B中元素的数目. 关系式 $(a \preceq b)$ 表示解a弱支配解b. 当 $C\ (A, B) = 1$ 时, B 中的每个解至少被 A 中的一个解所支配. 相反, $C\ (A, B) = 0$ 表明 B 中的任意解均未被 A 中的任意解所支配.

3.1.2 基准测试算例设置

由于缺乏统一的基准数据集, 本文在 Luo 构建的 VRPDTW 实例上生成测试集[17], 参数见表 3. 原始实例提供客户索引i、坐标$[x_i, y_i]$、期望访问时间窗$[g_i^k, b_i^k]$、车辆和无人机服务时间t_s与$\bar{t}_s$, 以及无人机装货时间t_p. 最大允许延迟设为时间窗长度的 30%, 并假设 85% 的客户满足无人机配送条件.

无人机速度$\bar{v}$、最大飞行距离$\mathcal{E}$和最大载荷U参考 DJI FlyCart 100[21]; 车辆自重W_0[16] 及行驶速度v参考文献 [17], 且设车辆速度与无人机速度相同. 冷链车辆能耗因子$(\eta, \alpha_1, \alpha_2, \alpha_3, \beta)$、可容忍产品质量损耗区间$[e_i^k, l_i^k]$及质量衰减系数$\xi_1, \xi_2$取自文献 [16, 22]. 根据文献 [22], 考虑冷链企业更关注配送时效, 设置$c_1 = 0.7$和$c_2 = 0.3$. 对于总周期为 7 天的实例, 按照文献 [23] 为每位客户随机生成 1–7 次服

表 3　案例关键参数

符号	描述	值
$t_s, \bar{t}_s$	车辆, 无人机的服务时间	$2 \in [0.5, 10]$ min
$v, \bar{v}$	车辆, 无人机的运输速度	20 m/s (水平飞行)
t_p	车辆为无人机装货时间	$1 \in [0.5, 10]$ min
$\mathcal{E}$	无人机最大飞行距离	12 km (满载, 双电池)
U	无人机最大负载容量	149.9 kg (含空机重量)
W_0	车辆自重	2 500 kg (空机重量)
η	车辆碳排放因子	$2.63 \in [2.3, 2.8]$ kg/L
$\alpha_1, \alpha_2, \alpha_3$	车辆油耗系数	$1.4\times10^{-3}, 4.35\times10^{-7}, 1.04\times10^{-8}$
β	制冷设备能耗率	$0.08 \in [0.05, 0.3]$ L/h
c_1, c_2	时效性和品质性的满意度权重	$0.7, 0.3 \in [0, 1]$
$[e_i^k, l_i^k]$	客户可容忍产品货损范围	$[0.005, 0.02] \in [0, 1]$
ξ_1, ξ_2	产品质量衰减系数	0.002, 0.003 (半衰期= 24 h)

务需求; 参考多无人机配送场景, 每辆车配置 2 架无人机. 最终测试集包含 20 个实例, 客户规模为 20–200. 实例名称 N20W80T7 表示客户数量为 20、时间窗长度为 80 s、规划周期为 7 天.

本文算法有四个关键参数: 种群规模 (ps), 策略估计平滑度 (τ), 探索折扣率 (γ_e) 和利用折扣率 (γ_x). 为确定参数最佳取值, 田口正交试验如下[24]:

1) $ps \in \{50, 100, 150, 200, 250\}$;

2) $\tau \in \{0.03, 0.10, 0.30, 0.60, 0.99\}$;

3) $\gamma_e \in \{0.70, 0.75, 0.80, 0.85, 0.90\}$;

4) $\gamma_x \in \{0.70, 0.75, 0.80, 0.85, 0.90\}$.

实验采用 $L25(5^4)$ 正交阵列进行校准, 并取 10 次实验结果的平均值. 参数主效应图 5 显示所有参数均对算法性能产生影响. 具体而言, 当种群规模 $ps = 100$ 时, 算法在多样性和过早收敛之间实现了平衡. 当策略估计平滑度$\tau = 0.3$ 时, 算法在保持探索能力的同时, 增强了最优算子选择能力. 探索折扣率 $\gamma_e = 0.85$, 使智能体能够平衡短期与长期收益. 利用折扣率$\gamma_x = 0.80$ 可确保算法得到充分开发, 同时不会显著牺牲多样性. 同时基于望大特性的信噪比 (SNR) 的主效应图 6 展示了各因子最高点所对应的水平: 种群规模 (ps) 的最佳水平为 $L2$; 策略估计平滑度 (τ) 的最佳水平为 $L3$; 探索折扣率 (γ_e) 的最佳水平为 $L4$; 利用折扣率 (γ_x) 的最佳水平为 $L3$. 综上, 参数设定如下: $ps = 100$, $\tau = 0.3$, $\gamma_e = 0.85$, $\gamma_x = 0.80$.

3.2　基于 CPLEX 的对比实验

为验证算法在小规模算例上的精度, 本节分别采用求解器 CPLEX 12.6.3 与本文算法求解 6 个小规模案例. 由于 CPLEX 无法直接处理多目标问题, 本文采用线性加权和法将模型等价重构为单目标形

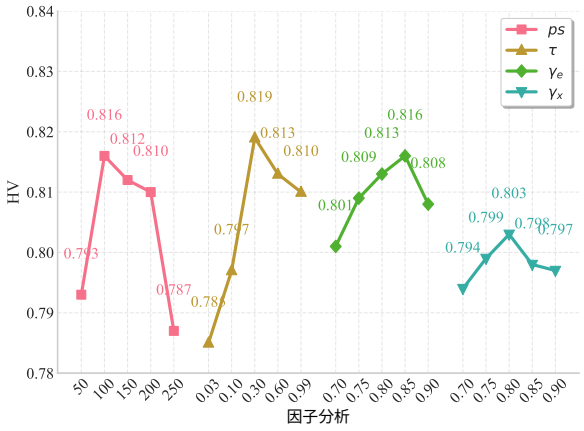

图 5　参数在 HV 指标下的均值主效应图 (最大值)

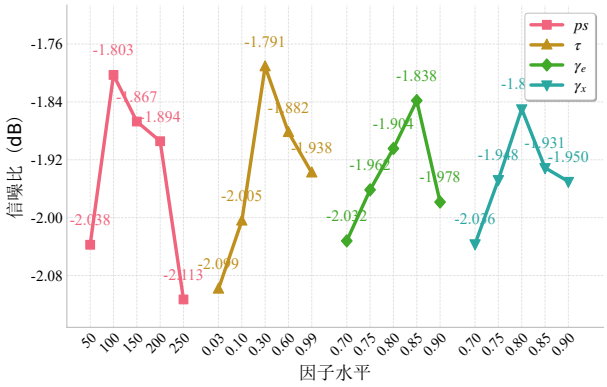

图 6　参数在 HV 指标下的 SNR 均值主效应图 (望大特性)

式以获取参考解[12]. 具体来说, 在加权前对多目标进行了归一化处理, f_i为第i个目标函数:

$$f_i'(x) = \frac{f_i(x) - f_i^{\min}}{f_i^{\max} - f_i^{\min}}, i = 1, 2, \tag{17}$$

其中, $f_i^{\min}$和$f_i^{\max}$分别为目标在可行域内的最优值和最劣值. 通过该极值法, 成本目标和非满意度目标的归一化值均映射至 [0, 1] 区间, 转化后的单目标:

$$F(x) = \sum_{i=1}^{n} \omega_i \cdot f_i'(x), \sum_{i=1}^{n} \omega_i = 1,\ \omega_i \geqslant 0, \tag{18}$$

其中, 为了消除主观偏好对对比结果的影响, 我们采用等权重策略, 即$\omega_1 = \omega_2 = 0.5$. 对比结果如表 4, 其中✓表示 CPLEX 在 3 600 秒内的最优参考解.

表 4　本文算法与求解器 CPLEX 的对比结果

算例	求解器CPLEX				本文算法			
	f_1	f_2	HV	C(CPLEX, –)	f_1	f_2	HV	C(本文算法,–)
N10W80T7	**0.00***	**0.00***	**0.95**	**0.91**	0.001	0.001	0.94	0.10
N15W80T7	**0.00***	**0.00***	**0.95**	**0.88**	0.004	0.001	0.93	0.22
N20W80T7	0.008	**0.00***	**0.94**	**0.76**	**0.00***	0.002	0.93	0.23
N25W80T7	**0.00***	**0.00***	**0.94**	**0.86**	0.001	0.001	0.93	0.14
N30W80T7	0.015	0.006	0.91	0.14	**0.00***	**0.00***	**0.93**	**0.86**
N35W80T7	0.031	0.007	0.91	0.16	**0.00***	**0.00***	**0.92**	**0.84**

注: **0.00***表明求解算法在有限时间预算内获得本案例下归一化后的最小值.

为进一步评估本文算法对最优帕累托前沿的逼近能力, 本文采用ε-约束法生成参考帕累托前沿. 首先将客户非满意度目标转化为不同ε值下的约束条件, 生成$K = 20$个受控帕累托解. 随后, 用 HV 和 C-metric 指标对 CPLEX 与本文算法的解集进行比较, 最优结果以粗体表示. 如表 4 所示, 在 N10W80T7 等小规模算例中, CPLEX 能够获得精确帕累托前沿, 并在 HV 和 C-metric 指标上取得最优结果, 本文算法则能够有效逼近该前沿. 当规模扩展至 N30W80T7 时, CPLEX 在限定时间内仅能获得可行解, 而本文算法在所有算例上的 HV 和 C-metric 表现均与 CPLEX 相当或更优. 结果验证了本文算法在小规模算例上的求解有效性, 以及在较大规模算例中的竞争力.

3.3 消融实验

为评估组件有效性, 变体 1 至变体 6 设计如下:

1) 变体 1 移除种群间知识迁移, 仅种群内进化;

2) 变体 2 移除双种群结构, 仅优化单一种群;

3) 变体 3 采用基于随机策略的算子选择方法;

4) 变体 4 采用基于独立学习的策略学习架构;

5) 变体 5 采用多臂老虎机驱动算子选择机制;

6) 变体 6 采用分层强化学习驱动算子选择策略.

为评估各组件的贡献, 本文基于 30 个随机生成的 N40W80T7 算例, 对 7 种变体算法独立运行 20 次, 并取平均结果作为最终性能. 对于每个算例, 分别按照 HV 和 C-metric 值对 7 种算法进行降序排名, 最优算法记为 1, 依次递增; 随后计算 30 个算例上的平均排名R_j. Friedman 检验的零假设为各算法性能不存在显著差异, 其统计量为

$$\chi_F^2 = \frac{12N'}{k(k+1)}\Big[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\Big], \tag{19}$$

其中, 样本数$N' = 30$, 算法数$k = 7$; 将χ_F^2与自由度$k-1=6$的卡方分布进行比较, 得到相应的p-值. 当$p < 0.05$时, 拒绝零假设, 表明算法间存在显著性能差异. 表 5 给出了显著性水平为 0.05 的 Friedman 检验结果[25], 最优结果以粗体表示.

由表 5 可知, 本文算法在 HV 和 C-metric 指标上均取得最优平均值和排名, 说明各组成模块均具有积极作用. 变体 3 的性能最差, 表明随机算子选择难以识别有效算子; 变体 4 的性能次之, 说明对手建模有助于提高策略训练稳定性; 变体 1 和变体 2 的结果分别验证了种群间知识迁移以及双目标偏置种群对提升初始解质量和多样性的作用; 变体 5 和变体 6 均弱于本文算法, 表明集中式多臂老虎机和分层强化学习难以充分利用不同搜索任务之间的行为关联; 相比之下, 多智能体通过学习差异化选择策略并进行信息交互, 更有效地协调搜索行为, 在收敛性与解集多样性之间取得更好的平衡.

表 5 消融实验的 Friedman 秩和检验结果

算法	HV			C（本文算法, –）		
	均值	标准差	排名	均值	标准差	排名
变体1	0.892	0.019	4.06	**0.943**	0.013	2.92
变体2	0.883	0.012	4.94	**0.945**	0.014	3.83
变体3	0.865	0.018	6.77	**0.954**	0.017	5.74
变体4	0.871	0.025	5.82	**0.951**	0.019	5.37
变体5	0.916	0.014	2.18	**0.937**	0.013	1.18
变体6	0.908	0.017	3.13	**0.939**	0.011	2.26
本文算法	**0.928**	0.011	**1.20**	–	–	–
$V_D \subseteq V_N$称为可由无人机服务的客户子集.	$V_L = V\backslash\{n+1\}$为V中可发射无人机的顶点子集.			$V_R = V\backslash\{0\}$为可回收无人机的顶点子集.		

3.4 对比实验

为验证算法有效性, 选取包括经典算法 NSGA-II[26]、MOEA/D[27]、MOPSO[28] 和强化学习机制的 HMOA[17]、NRL[10]、RDMA[12]、QSISRs[29] 和 AMA[11] 进行对比. 为保证公平性, 所有算法均通过田口检验完成参数校准, 校准结果见附录表 C1–C2, 图 C1–C8. 所有算法的最大迭代数为 5 000, 运行 30 次[18]. 本文算法主辅助种群并行演化 100 次; 优化器: Adam; DQN 学习率: 0.000 1; 探索策略: 贪心策略; 采用 Wilcoxon 秩和检验分析结果的统计显著性[25]. 总体而言, 本文算法在大多数测试实例中均取得更优表现. 在图 7 中, C(本文算法, 对比算法) > C(对比算法, 本文算法), 表明本文算法的解集具有更强的支配能力. 在图 8 中, 本文算法具有较高的 HV 值, 说明其解集在收敛性和多样性更具优势.

3.5 讨论

如图 8 所示, 本文算法在算例 N20W80T72 和 N160W80T72 中未取得最优结果, 可能是固定迁移概率难以适应种群状态变化, 导致信息共享不足和局部收敛. 为检验算法差异的显著性, 本文进一步进行了 Friedman 检验, 结果见表 6. 在 HV 和 C 指标上, 本文算法均排名第一, 且显著优于所有对比算法($p < 0.05$),表明其能够有效平衡收敛性与多样性. 如图 9 所示, 本文算法在 N200W80T7 实例上获得了接近参考帕累托前沿且分布均匀的解集. 图 10 表明, 其 HV 收敛值整体高于对比算法, 进一步验证了本文算法在多周期车辆–无人机路径规划中的有效性.

如图 11 所示, 当车载无人机数量m由 1 增至

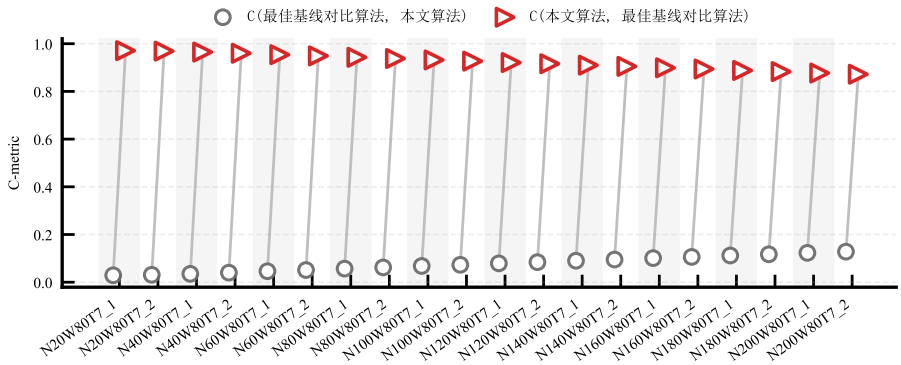

图 7　算法关于 C 指标的对比结果图

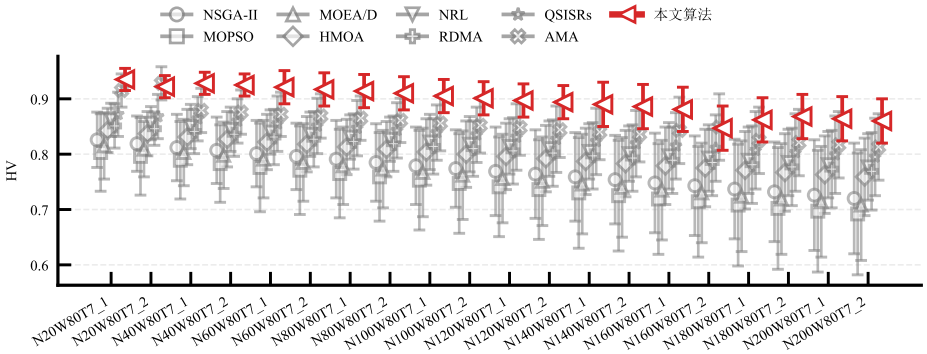

图 8　算法关于 HV 指标的对比结果图

表 6　对比实验的 Friedman 秩和检验结果

算法	HV			C（本文算法,–）		
	均值	标准差	排名	均值	标准差	排名
NSGA-II	0.852	0.028	6.39	**0.920**	0.019	6.11
MOPSO	0.825	0.035	8.93	**0.929**	0.016	8.34
MOEA/D	0.833	0.032	8.49	**0.926**	0.022	7.29
HMOA	0.871	0.025	4.61	**0.914**	0.023	5.08
NRL	0.895	0.020	4.15	**0.902**	0.014	2.85
RDMA	0.943	0.022	4.23	**0.908**	0.011	3.95
QSISRs	0.954	0.018	3.87	**0.896**	0.014	1.72
AMA	0.915	0.015	2.89	**0.929**	0.012	0.66
本文算法	**0.928**	0.011	**1.44**	–	–	–
$V_D \subseteq V_N$称为可由无人机服务的客户子集.	$V_L = V\backslash\{n+1\}$为V中可发射无人机的顶点子集.			$V_R = V\backslash\{0\}$为可回收无人机的顶点子集.		

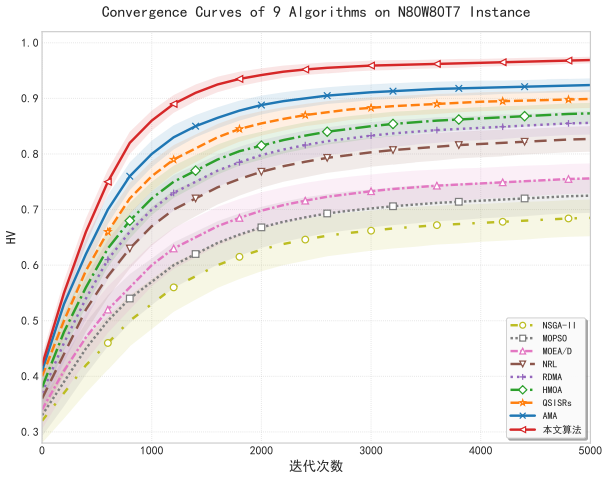

图 10　N200W80T7 算例下算法 HV 指标收敛示意图

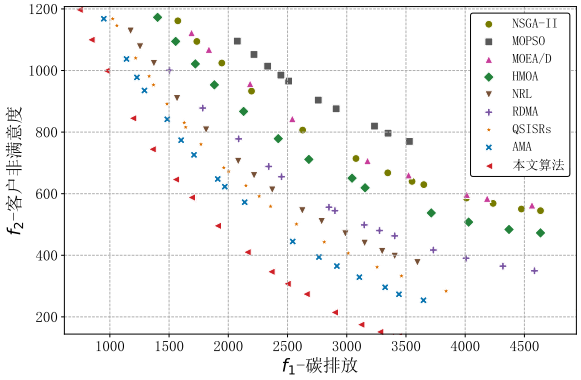

图 9　N200W80T7 算例下算法的帕累托解集示意图

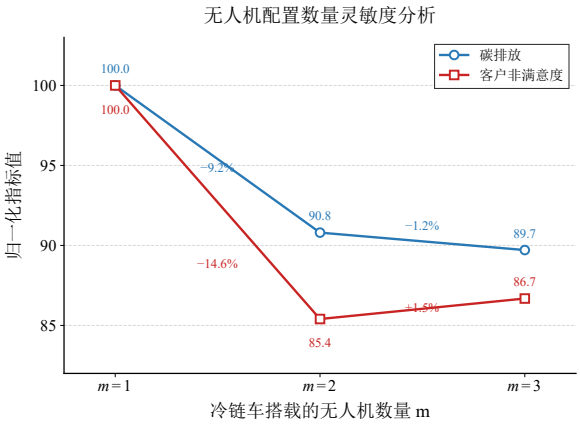

图 11　无人机数量 (m = 1、2、3) 灵敏度分析图

2 时, 并行配送使碳排放和客户非满意度分别降低约 9.2% 和 14.6%. 当m由 2 增至 3 时, 碳排放仅额外降低约 1.2%, 而客户非满意度反而上升约 1.5%. 这是由于多无人机回收增加了车辆等待时间, 并将延误传递至后续客户. 因此, $m = 2$是边际收益拐点, 继续增加无人机可能降低整体效率.

4 结 论

本文通过提出双种群多智能体引导的多目标进化算法解决多周期车辆-无人机协同路径规划问题. 该方法首先通过双种群演化初始化方法提升初始解的质量与多样性. 其次, 通过多智能体引导的搜索方法平衡多目标算法的收敛性和多样性. 然后, 基于对手建模的策略学习方法提高智能体策略训练的稳定性. 实验表明: 本文方法获得了在碳排放与客户满意度帕累托分布优良的周期性车辆-无人机路径方案.

参考文献 (References)

[1] Han J W, Sun C H, Ji Z T, et al. Smart cold chain logistics for fresh agricultural products: Key technologies, challenges, and future trends[J]. Trends in Food Science & Technology, 2026, 167: 105421.

[2] Yuan B F, Tao F M, Qin Y, et al. Can policy achieve the goal of cold chain logistics sustainable development?[J]. Transportation Research Part D: Transport and Environment, 2025, 140: 104607.

[3] 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, 2026, 41(7): 2013-2026.
(Yang M Y, Wang W. Model and graph-based reinforcement learning driven-multi-objective evolutionary algorithm for vehicles-drone cooperative routing problem[J]. Control and Decision, 2026, 41(7): 2013-2026.)

[4] Duan H H, Li X L, Feng Y X, et al. Multi-objective hyper heuristic memetic algorithm for time-dependent vehicle routing problem with drone considering simultaneous pickup and delivery[J]. Expert Systems with Applications, 2026, 297: 129196.

[5] Weng X, She W K, Fan H Q, et al. Multi-depot vehicle routing problem with drones in emergency logistics[J]. Cluster Computing, 2024, 28(1): 64.

[6] Lei D, Jia T. Multi-period time-dependent vehicle routing problem for infectious medical waste collection considering temporal security constraints[J]. Expert Systems with Applications, 2026, 314: 131626.

[7] Wang Y J, Wang G G, Wang L. A tree-based multiobjective evolutionary algorithm for energy-efficient hybrid flow-shop scheduling[J]. IEEE Transactions on Evolutionary Computation, 2025, 29(6): 2313-2327.

[8] Li R, Wang L, Gong W Y, et al. An evolutionary multitasking memetic algorithm for multiobjective distributed heterogeneous welding flow shop scheduling[J]. IEEE Transactions on Evolutionary Computation, 2025, 29(6): 2287-2298.

[9] 王凌, 王锐. 机器学习赋能的智能优化与调度专刊序言[J]. 控制与决策, 2026, 41(4): 865-870.
(Wang L, Wang R. Guest editorial of special issue on machine learning-enabled intelligent optimization and scheduling[J]. Control and Decision, 2026, 41(4): 865-870.)

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

[11] Windras Mara S T, 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.

[12] 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.

[13] Yan Q, Wang H F, Yang S X. A learning-assisted bi-population evolutionary algorithm for distributed flexible job-shop scheduling with maintenance decisions[J]. IEEE Transactions on Evolutionary Computation, 2025, 29(5): 1795-1806.

[14] Zhu N N, Zhao F Q, Yu Y, et al. A hierarchical reinforcement learning-aware hyper-heuristic algorithm with fitness landscape analysis[J]. Swarm and Evolutionary Computation, 2024, 90: 101669.

[15] de Witt C S, Gupta T, Makoviichuk D, et al. Is independent learning all you need in the StarCraft multi-agent challenge? [PP/OL]. 2020, arXiv: 2011.09533. https://arxiv.org/abs/2011.09533.

[16] 周鲜成, 郑梓亮, 杨堃, 等. 城市物流配送中时间-位置依赖型多目标绿色车辆路径问题研究[J]. 控制与决策, 2025, 40(2): 413-422.
(Zhou X C, Zheng Z L, Yang K, et al. Research on time-and-site dependent multi-objective green vehicle routing problem in urban logistics distribution[J]. Control and Decision, 2025, 40(2): 413-422.)

[17] 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.

[18] Guo T, Mei Y, Zhang M J, et al. Learning-aided neighborhood search for vehicle routing problems[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(7): 5930-5944.

[19] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[PP/OL]. 2015, arXiv: 1503.02531.

[20] Van Veldhuizen D A, Lamont G B. On measuring multiobjective evolutionary algorithm performance[C].

Proceedings of the 2000 Congress on Evolutionary Computation. La Jolla, 2002: 204-211.

[21] 大疆官方. DJI FlyCart 100 - DJI 配送无人机[EB/OL]. https://www.dji.com/flycart-100/specs. 2026-04-01.

[22] 高浩然, 张玉林, 张顺顺. 考虑时效性和品质性满意度约束的易腐品冷链物流配送[J]. 控制与决策, 2024, 39(4): 1342-1350.
(Gao H R, Zhang Y L, Zhang S S. Cold chain logistics distribution of perishable products considering timeliness and quality customer satisfaction[J]. Control and Decision, 2024, 39(4): 1342-1350.)

[23] Rodríguez-Martín I, Salazar-González J J, Yaman H. The periodic vehicle routing problem with driver consistency[J]. European Journal of Operational Research, 2019, 273(2): 575-584.

[24] Van Nostrand R C. Design of experiments using the taguchi approach: 16 steps to product and process improvement[J]. Technometrics, 2002, 44(3): 289.

[25] Hollander M, Wolfe D A. Nonparametric Statistical Methods[M]. The 2nd edtion. New York: Wiley, 1999.

[26] Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197.

[27] Zhang Q F, Li H. MOEA/D: A multiobjective evolutionary algorithm based on decomposition[J]. IEEE Transactions on Evolutionary Computation, 2007, 11(6): 712-731.

[28] Coello C A C, Pulido G T, Lechuga M S. Handling multiple objectives with particle swarm optimization[J]. IEEE Transactions on Evolutionary Computation, 2004, 8(3): 256-279.

[29] Chen P, Wang Q L. Learning for multiple purposes: A Q-learning enhanced hybrid metaheuristic for parallel drone scheduling traveling salesman problem[J]. Computers & Industrial Engineering, 2024, 187: 109851.

作者简介

杨明园 (2001−), 男, 博士生, 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, 研究方向为智能控制理论, E-mail: wangwei407@hrbeu.edu.cn;

董福王 (1992−), 男, 副教授, 博士, 博士生导师, 研究方向为智能无人系统, E-mail: dongfuwang@hrbeu.edu.cn.