控制与决策
Control and Decision

基于改进 Q学习的电动冷藏车多目标跨区域路径优化

王岩红, 钟颖, 张允华

引用本文:
王岩红, 钟颖, 张允华. 基于改进 Q学习的电动冷藏车多目标跨区域路径优化[J]. 控制与决策, 2026, 41(3): 741-753.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2024.1370

基于改进Q学习的电动冷藏车多目标跨区域路径优化

王岩红[1†]，钟 颖[1]，张允华[2]

(1. 上海工程技术大学 管理学院，上海 201620；2. 同济大学 汽车学院，上海 201804)

摘 要: 面向冷链物流绿色化发展目标和载具电动化趋势, 考虑拥堵路况、充电成本、电量消耗等多目标实施电动冷藏车跨区域路径优化, 提出一种改进的Q学习方法, 设计启发式奖励机制, 引入余弦退火学习率和指数衰减探索率两种动态策略, 提升算法性能并进行仿真实验与对比分析. 实验数据表明, 改进后的强化学习算法能够根据交通运行状态、电动冷藏车的初始电量以及能耗率等, 有效优化跨区域冷链配送路线. 相较于其他 3 种Q学习算法, 在 6 类差异化测试场景下, 其配送方案能够显著降低总里程与电量消耗 ($p < 0.05$, Welch's t-test). 结果表明, 该方法在高速公路、城市道路及充电站投放等环境建模下具备良好的适应性和鲁棒性.

关键词: 电动冷藏车；跨区域路径优化；多目标；动态策略；强化学习；启发式Q学习

中图分类号: TP391；U49 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2024.1370

引用格式: 王岩红, 钟颖, 张允华. 基于改进Q学习的电动冷藏车多目标跨区域路径优化 [J]. 控制与决策, 2026, 41(3): 741-753.

Multi-objective cross-regional path optimization for electric refrigerated vehicles based on improved Q-learning

WANG Yan-hong[1†]，*ZHONG Ying*[1]，*ZHANG Yun-hua*[2]

(1. College of Management，Shanghai University of Engineering Science，Shanghai 201620，China；2. College of Automotive Studies，Tongji University，Shanghai 201804，China)

Abstract: To address the green development goals of cold chain logistics and the trend toward vehicle electrification, this study focuses on optimizing cross-regional routes for electric refrigerated vehicles under multi-objective considerations, including traffic congestion, charging costs, and energy consumption. An improved Q-learning method is proposed, which integrates a heuristic reward mechanism and dynamic strategies combining cosine annealing learning rates and exponential decay exploration rates to enhance algorithm performance. Simulation experiments and comparative analyses are conducted to validate the approach. Experimental data demonstrate that the improved reinforcement learning algorithm effectively optimizes cross-regional cold chain delivery routes by accounting for traffic conditions, the initial battery level of electric refrigerated vehicles, and energy consumption rates. Compared to three other Q-learning algorithms, the proposed method significantly reduces both total travel distance and energy consumption ($p<0.05$, Welch's t-test) across six distinct testing scenarios. The results indicate that the proposed method exhibits strong adaptability and robustness in various environmental modeling scenarios, including highways, urban roads, and charging station deployment.

Keywords: electric refrigerated vehicle； cross-regional path optimization； multi-objective； dynamic strategies； reinforcement learning；heuristic Q-learning

0 引 言

区域协调发展战略为区域物流的转型升级带来重大机遇和挑战: 一是载具电动化和新能源配套设施布局推动的绿色低碳化转型; 二是物联网、大数据、人工智能算法赋能的数字化、智能化转型. 冷链物流, 一方面由于配备冷藏设施, 相较于一般物流载具, 对节能减排要求更为迫切, 冷链产品因其易腐性、高时效性等特性, 对复杂路况下的配送效率和路径优化尤为敏感; 另一方面因关系到社会民生保障与城市安全应急, 对冷链物流设施网络构建、智能化

收稿日期: 2024-11-24；**录用日期:** 2025-05-26.

基金项目: 中国科协“科技智库青年人才计划”项目 (20220615ZZ07110408)；国家自然科学基金青年基金项目 (52206167, 72504174)；上海市“科技创新行动计划”软科学研究领域重点项目 (24692114300).

[†]通信作者. E-mail: yanhong.wang@hotmail.com.

路径优化算法等部署的需求日益提升. 因此, 本文提出基于改进强化学习的电动冷藏车跨区域路径优化方法, 以提高环境适应性、收敛效率, 避免局部最优等问题.

电动冷藏车的路径优化面临“温度衰退”与“电池容量”的双重约束, 当前的研究尚未充分兼顾考虑, 但是对单一约束下的路径优化研究较为丰富. 前者包括多目标优化、动态需求管理以及多温共配等多个方面: Zhang 等[1] 考虑不同类型的车辆协同配送场景, 研究了成本、碳排放量和运输距离三重约束条件下的冷链物流车辆路径问题; Li 等[2] 将客户满意度引入冷链物流网络优化, 实现市场竞争环境的多目标平衡; 谭晓伟等[3] 考虑实时新增订单, 提出一种沿途补货的多配送中心路径规划方法; He 等[4] 开发了一种基于多温区共同配送的电动汽车路径优化系统, 以保证货物新鲜度的同时降低配送成本. 后者在续航短、充电站有限等约束下开发求解方法, 涵盖数学软件法、精确算法、启发式方法等. 其中, 精确算法设计灵活性更好, 启发式算法在大规模算例更具优势. 代表性研究成果包括: Arias-Londoño 等[5] 建立的混合整数线性规划模型, 通过数学软件 GAMS/CPLEX 求解, 为电动物流车规划高质量路线; Wu 等[6] 提出一个两级配送的电动车辆路径问题, 采用分支定价法求解, 并通过与 CPLEX 的对比验证了算法在小规模数据集的有效性; Fan[7] 建立了灵活的能耗估算策略和混合整数规划模型, 结合改进的蚁群算法, 优化了电动车的调度和路线规划; Chen 等[8] 提出一种基于阈值接受的多层搜索算法, 以解决电动物流车的路径规划和充电决策问题.

与上述路径优化方法相比, 以强化学习为代表的 AI 算法表现出更具前途的应用空间, 在面临缺乏先验知识的未知环境时, 基于“规则”的仿生算法和启发式算法难以满足道路拥堵、电量消耗和温度衰退等复杂和贴近现实的优化需求; 而基于“学习”的强化学习方法在路径优化研究方面逐渐开展. 尤其是Q学习算法, 在路径优化领域得到了广泛应用: Wu 等[9] 引入一种适用于无人机搜索与救援任务的Q学习算法, 改善了在未知环境中路径规划效率; Huang 等[10] 采用一种改进的Q学习算法用于自动引导车自主路径规划, 以有效提升作业效率. Zhou 等[11]提出一种优化的Q学习算法, 用以提升移动机器人在复杂环境中的局部路径规划能力与适应性. Zhong 等[12] 基于模拟退火算法和启发式搜索, 设计一种改进的Q学习算法, 以成功应对超出“场地”限制的跨区域场景. 上述研究为基于强化学习的电动冷藏车路径优化提供了新思路, 特别是为未来无人电动车和智能网联车承担跨区域冷链配送奠定了复杂环境适应性与决策优化实时性的基础.

现有相关研究存在一定局限: 一是对于冷链物流路径优化, 多以温度衰退为约束, 对兼顾考虑电池容量、能耗率、拥堵路况等因素的电动冷藏车的关注度不足; 二是对于路径优化场景, 在算法和算力限制下, 多局限于“场地”或区域内规划, 对跨区域交通流动的关注不足, 可以适应无人车的路径优化算法研究更为匮乏. 本文在冷链载具电动化和智能化融合的背景下, 针对电动冷藏车跨区域路径优化问题, 设计一种改进的Q学习方法, 提出多目标启发式奖励机制, 引入余弦退火和指数衰减方法动态调整学习和探索策略, 主要研究工作包括:

1) 以浙江嘉兴至上海嘉定真实道路信息和补能设施为场景, 引入拥堵路况, 构建栅格仿真环境模型;

2) 建立电动冷藏车路径优化的马尔可夫决策过程, 改进强化学习算法;

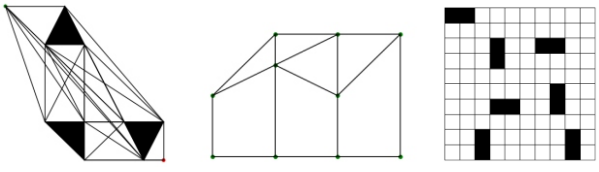
3) 调节路况、初始电量和能量消耗, 实施 3 组对比实验并进行敏感性分析, 验证算法性能和鲁棒性.

1 电动冷藏车跨区域路径优化问题

“里程焦虑”是电动载具的痛点问题, 而电动冷藏车的跨区域配送还要应对制冷和保温的能耗需求. 本文旨在解决电量不足以直达目的地时, 电动冷藏车优化选择跨区域路径, 并完成在途充电能源补给, 成功抵达目的地的问题. 具体的优化目标有: 1) 成功到达配送地点; 2) 减少拥堵时间; 3) 缩短路径里程; 4) 降低充电成本. 为应对上述问题, 提出环境建模方法, 并采用改进的Q学习算法进行求解.

1.1 环境建模

首先, 环境建模是基于强化学习路径优化研究的先决步骤, 旨在有效描绘区域道路、建筑物、服务区等环境信息. 常见的环境建模方法有可视图法[13-14]、拓扑法[15]、栅格法[11, 16] 等 (如图 1 所示), 各具特点且涉及不同的变量 (如表 1 所示). 电动冷藏车跨区域路径优化问题覆盖区域面积较广且需要描绘补能设施的精准定位, 栅格法的空间表达力和鲁棒性能够满足建模要求. 然而, 大规模精准的栅格空间建模对



(a) 可视图法 (b) 拓扑法 (c) 栅格法

图1 环境建模方法

表1　环境建模方法对比

建模方法	优点	缺点	主要变量
可视图法	实现简单, 易与算法结合	低效且不灵活	顶点V, 边E
拓扑法	无需详细信息即可确定方位	建立过程复杂	节点N, 边L
栅格法	建模精确, 空间表达力好, 鲁棒性强	对大规模栅格空间的计算资源需求较高	栅格大小s, 栅格值g

算力资源需求高, 限制了仿生算法和启发式算法在栅格地图上求解, 而强化学习具有较好的复杂环境自适应性和不确定性环境的决策能力, 适合应对多维状态和动作空间. 因此, 本文采用栅格法构建环境模型.

其次, 本文选择浙江嘉兴到上海嘉定的跨区域范围作为实验环境 (如图 2 所示), 以提升实验情景的精确性和真实性. 实验区域的选择基于双重考量: 一方面, 嘉兴和嘉定全域已被纳入智能网联汽车测试与示范区, 这为无人电动冷藏车通过城际快速路进行跨区域配送创造了基础条件, 也为跨区域路径优化智能算法的部署提供了落地场景; 另一方面, 作为长三角一体化发展战略的关键节点, 嘉兴国家骨干冷链物流基地不仅服务本地市场, 还辐射邻近的上海嘉定区, 能够有效满足嘉定市场对高效、可靠冷链物流的服务需求. 在道路网络构建过程中, 本文综合考虑冷链运输时效性约束、充电设施空间分布及服务区资源配置等核心要素, 优先选取连接两地的沪昆高速、上海绕城高速等主干高速公路作为跨区域运输动脉, 保障路径网络的拓扑连贯性. 同时, 整合区域内衔接高速路网的城市主干道, 形成覆盖城际运输全流程的多层级路网体系.

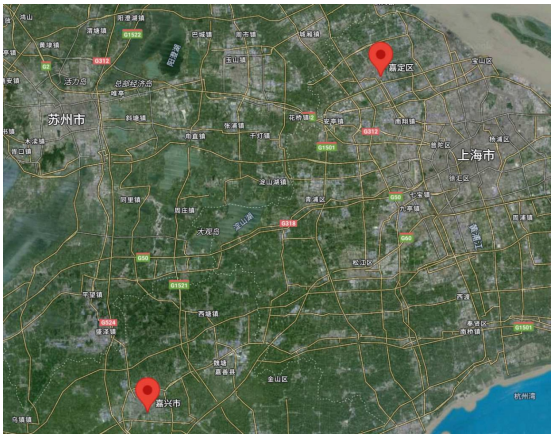


图2　浙江嘉兴与上海嘉定区域地图

最后, 将拥堵延时指数纳入栅格环境模型, 描绘区域内因车流量大、突发交通事故、道路维修等多种因素导致的潜在拥堵状况. 冷链物流的时效性要求较高, 道路拥堵是制约冷链物流效率和服务质量的关键因素. 本文采用高德地图对交通运行状态的 4 等级划分方法 (见表 2), 以拥堵延时指数作为道路拥堵程度的评价指标, 其算法为实际行程时间与自由流状态下行程时间的比值.

表2　拥堵延时指数分类

拥堵延时指数	交通运行状态
$[1.0, 1.5)$	畅通
$[1.5, 2.0)$	缓行
$[2.0, 4.0)$	拥堵
$[4.0, \infty)$	严重拥堵

本研究环境建模的具体步骤如下: 1) 将区域环境抽象为二维空间模型; 2) 将该空间均匀分割为若干栅格单元; 3) 赋予每个栅格单元特定属性, 反映不同的环境信息. 如图 3 所示, 灰色三角形和圆形分别表示起点和目的地, 绿色闪电符号表示电力补给服务区或停车区, 黑色栅格表示不可通行区域, 如街区、建筑物、农田等; 连接嘉兴与嘉定的主要高速公路和城市道路上的各种路况则以不同颜色的栅格表示, 深红色表示严重拥堵, 橙色表示拥堵, 黄色表示缓行, 白色表示畅通; 行进路径以粉色圆点轨迹展示, 轨迹上的粉色闪电符号表示进行了充电操作.

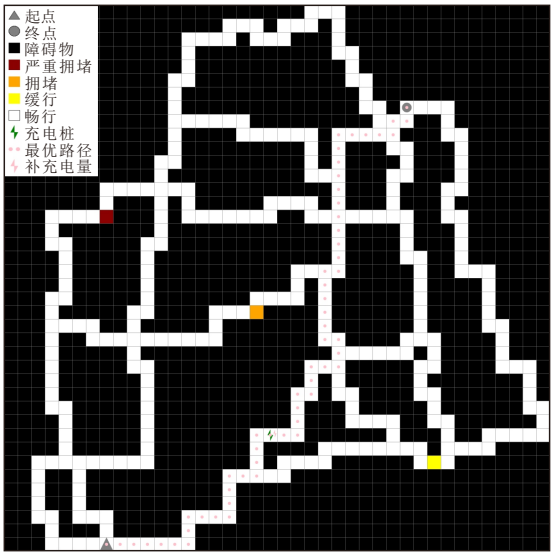


图3　仿真环境栅格地图

1.2　Q学习

Q学习是一种应用广泛的无模型强化学习方法[17]. Q学习基于马尔可夫决策过程 (MDP)[18], MDP 由一个五元组$\{S, A, P, R, \gamma\}$表示[19]. 其中: S表示状态集; A表示动作集; P为状态转移矩阵; R为奖励函数; γ为折扣因子, $\gamma \in (0, 1)$.

Q学习算法的核心思想是建立由状态和动作组成的Q表格, 其中Q值表示在特定状态下执行特定动作所获得的期望收益. 基于该算法, 智能体在每次迭代中, 根据当前状态s_t选择动作a_t, 进入下一状态s_{t+1}并接收环境反馈r_{t+1}, 利用式 (1) 更新Q值. 其中α为学习率, $\alpha \in (0,1)$.

$$Q_{\text{new}}(s_t,a_t)= Q(s_t,a_t)+\alpha[r_{t+1}+\gamma\max_{a\in A}Q(s_{t+1},a)-Q(s_t,a_t)]. \tag{1}$$

在电动冷藏车跨区域路径优化中, 车辆通过与环境交互学习, 逐步优化其路径选择策略. Q表迭代更新完成后, 电动冷藏车基于最优策略进行路径选择. 在每个路径点, 车辆选择Q值最大的动作a, 以完成跨区域冷链配送任务. 动作选择函数定义为

$$\pi=\arg\max_{a\in A}Q(s,a), \tag{2}$$

其中$\arg\max_{a\in A}Q(s,a)$表示当前状态s下, 所有可能动作中Q值最大的动作.

如图 4 所示, 电动冷藏车跨区域路径优化框架[20]采用双层级递进式架构. 在交互学习层, 智能体通过试错机制和反馈循环, 与外部环境持续互动, 捕获交通拥堵状态、车载电池荷电状态及充电服务区分布等关键信息. 随着对环境的不断探索, 智能体积累丰富的决策经验; 路径优化层则利用交互学习的经验对Q表进行迭代更新, 逐步优化动作选择策略, 最终生成最优配送路径. 双层架构通过协同机制实现“感知-决策”闭环, 有效保障电动冷藏车在跨区域多约束条件下的运输效率与系统鲁棒性.

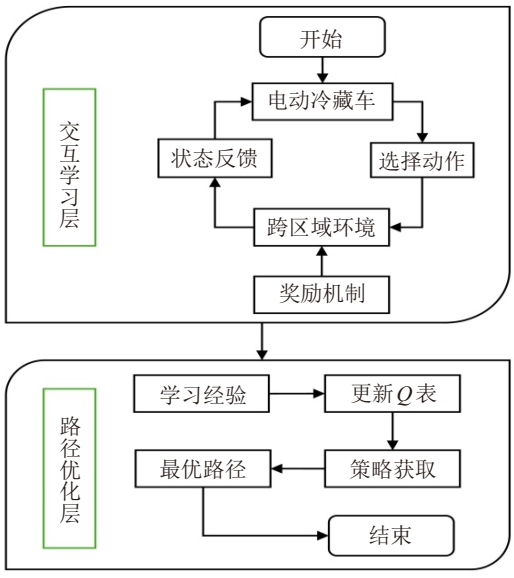


图4 电动冷藏车跨区域路径优化框架

2 强化学习模型与算法设计

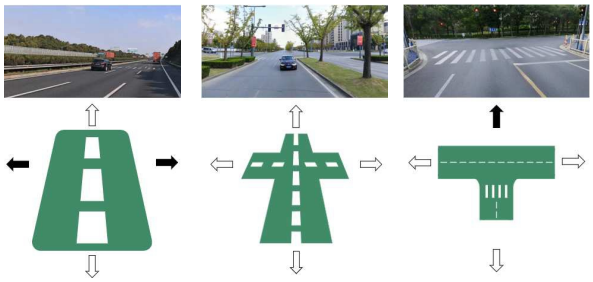
在强化学习模型中, 状态空间和动作空间是两个基本要素. 状态参数影响动作决策, 而动作执行则会导致状态转移, 并触发环境反馈. 此外, 跨区域路径优化具有环境信息不完备、实时路况复杂多变、范围大、距离远等特点, 为了提升计算效率, 需要设计相应的模型和算法. 本文以Q学习为理论框架, 对状态和动作空间、状态转移、奖励函数以及学习和探索策略进行设计, 构建一个适应跨区域冷链路径优化问题的Q学习算法模块.

2.1 状态和动作空间

状态空间S定义为二维变量空间, 即$S=\langle$位置, 电量$\rangle$. 状态变量一是位置信息, 是路径优化的关键决策依据. 鉴于区域环境以$n\times m$的栅格形式呈现, 位置用五元组$\{(x,y),\varphi,C,O,T\}$描述, 其中(x,y)表示智能体在地图上的坐标, 横坐标x的取值范围为$[0,n-1]$, 纵坐标y的取值范围为$[0,m-1]$. 其余 4 个元素φ,C,O,T均与(x,y)关联: φ是拥堵延时指数, 反映交通运行状态, 取值范围为$\{1.0,1.5,2.0,4.0\}$, 分别对应畅行、缓行、拥堵和严重拥堵; C,O,T均为布尔型, 取值为$\{\text{True},\text{False}\}$, 分别表示智能体当前是否位于充电区域、是否处于不可通行区域以及是否已抵达目的地.

状态变量二是电量水平, 智能体作为电动冷藏车的逻辑抽象, 其行动能力受到电池容量制约. 当电量耗尽时, 智能体将无法执行任何动作, 因此电量是影响动作决策的关键因素. 电量用E表示, 定义为电池中的可用能量, 单位为 kWh. 在行驶过程中, 电量会随着行驶距离和供冷时长的增加而逐渐减少. 只有当智能体到达可充电区域时, 即$C=\text{True}$时, 电量才可以得到补充. 在整个行动过程中, 必须确保$E>0$, 否则将被视为行动失败.

在跨区域路径优化中, 电动冷藏车将在状态空间内执行 4 种动作: 前进 (↑)、右转 (↱)、掉头 (↶) 与左转 (↰), 即动作空间$A=\{\text{up},\text{right},\text{down},\text{left}\}$. 如图 5 所示, 智能体主要面临 3 类道路情景, 分别是直行车道、十字路口和丁字路口.



(a) 直行车道 (b) 十字路口 (c) 丁字路口

图5 动作选择

2.2 状态转移机制

在执行动作后, 电动冷藏车会从当前状态转移至下一状态, 此时各个状态变量都将发生相应变化, 并通过数学量化表达. 本文中距离单位统一为 km, 时间单位统一为 h, 电量单位统一为 kWh. 智能体根据动作更新位置状态$\{(x,y),\varphi,C,O,T\}$至$\{(x',y'),\varphi',C',O',T'\}$, 其状态转移函数定义为

$$(x',y')=\begin{cases}(x-1,y),\ x>0\ 且a=\text{up};\\(x,y+1),\ y<m-1\ 且a=\text{right};\\(x+1,y),\ x<n-1\ 且a=\text{down};\\(x,y-1),\ y>0\ 且a=\text{left};\\(x,y),\ \text{otherwise}.\end{cases}\tag{3}$$

$$\varphi'=\begin{cases}1.0,\ (x',y')\ 路况畅通;\\1.5,\ (x',y')\ 路况缓行;\\2.0,\ (x',y')\ 路况拥堵;\\4.0,\ (x',y')\ 路况严重拥堵.\end{cases}\tag{4}$$

$$C'=\begin{cases}\text{True},\ (x',y')\in\text{CA};\\\text{False},\ \text{otherwise}.\end{cases}\tag{5}$$

$$O'=\begin{cases}\text{True},\ (x',y')\in\text{IA};\\\text{False},\ \text{otherwise}.\end{cases}\tag{6}$$

$$T'=\begin{cases}\text{True},\ (x',y')=(x_{\text{target}},y_{\text{target}});\\\text{False},\ \text{otherwise}.\end{cases}\tag{7}$$

其中: 动作$a\in A$, CA 和 IA 分别表示可充电和不可通行区域的位置集合, $(x_{\text{target}},y_{\text{target}})$为目的地坐标.

电动冷藏车的电量变化涉及电量消耗和电量补充两种情形. 电量消耗主要包括两部分: 一是提供车辆动力的能耗, 二是维持低温环境的能耗. 其中, 制冷能耗与行驶过程中的交通拥堵程度密切相关. 补能仅在车辆经过 CA 区域且当前电量低于特定阈值E_{th}时发生. 电量状态E向E'转移函数定义如下:

$$E'=\begin{cases}\min\{E+pt-e_ct,E_{\text{bat}}\},\\\qquad C'=\text{True}\ 且E<E_{\text{th}};\\\max\{E-\Delta E_p-\Delta E_c,0\},\ \text{otherwise}.\end{cases}\tag{8}$$

其中: p 为充电功率, t 为充电时长, e_c 为单位时间制冷耗电, E_{bat} 为电动冷藏车的电池总容量, ΔE_p和ΔE_c分别表示动力系统和制冷系统的能耗.

动力系统能耗ΔE_p和制冷系统能耗ΔE_c可定义为

$$\Delta E_p=e_pL,\tag{9}$$

$$\Delta E_c=\frac{e_cL\varphi'}{v}.\tag{10}$$

其中: e_p 表示单位里程行驶耗电; L 表示从当前位置到下一位置的路径长度, 可通过栅格地图与实际地图的比例计算得出; v 为平均行驶速度.

2.3 多目标奖励函数

强化学习中智能体通过奖励机制提高学习效率和优化策略. 跨区域电动冷藏车, 需在确保电量大于零的前提下, 最小化充电成本和能量消耗, 并尽快抵达目的地. 因此, 奖励函数设计综合考虑电量、距离、时间、经济、能耗等因素, 在奖励函数中引入 5 个目标项 (路径状态项r_{state}, 目标引导项r_{distance}, 拥堵时间项$r_{\text{congestion}}$, 充电成本项r_{cost}和电量消耗项r_{energy}), 得到一个综合的奖励机制, 旨在实现高效经济的路径选择策略, 即

$$r=r_{\text{state}}+r_{\text{distance}}+r_{\text{congestion}}+r_{\text{cost}}+r_{\text{energy}}.\tag{11}$$

2.3.1 路径状态项

路径状态项r_{state}主要由碰撞惩罚r_{obstacle}和终点奖励r_{target}两部分构成. 智能体在到达目标点时会获得正奖励, 碰到障碍物则受到负惩罚, 而其他情况下的奖励值通常设置为 0, 定义为

$$r_{\text{state}}=\begin{cases}r_{\text{obstacle}},\ O=\text{True};\\r_{\text{target}},\ T=\text{True};\\0,\ O=\text{False}.\end{cases}\tag{12}$$

2.3.2 目标引导项

引入目标引导项r_{distance}应对复杂跨区域环境, 避免智能体在可通行路段上做无意义徘徊, 为其提供明确的方向性信息, 提升路径选择效率. 该项涉及智能体在连续两个时刻的位置与目标点之间的欧式距离变化, 当智能体偏离预定目的地时, 环境给予一个负奖励η, 以减少该动作发生的频率. 定义

$$r_{\text{distance}}=\begin{cases}\eta,\ d_{t-1}\leqslant d_t;\\0,\ d_{t-1}>d_t.\end{cases}\tag{13}$$

$$d_{t-1}=\sqrt{(x_{\text{target}}-x_{t-1})^2+(y_{\text{target}}-y_{t-1})^2},\tag{14}$$

$$d_t=\sqrt{(x_{\text{target}}-x_t)^2+(y_{\text{target}}-y_t)^2}.\tag{15}$$

其中(x_{t-1},y_{t-1})和(x_t,y_t)分别是智能体在$t-1$和t时刻的位置坐标.

2.3.3 拥堵时间项

考虑冷链物流对高效保质的需求, 引入拥堵时间项$r_{\text{congestion}}$, 该项奖励值基于拥堵延时指数φ设置, 旨在描述因交通拥堵造成的时间损失. $r_{\text{congestion}}$以λ的比例变化, 定义为

$$r_{\text{congestion}}=-\lambda(\varphi-\varphi_0),\tag{16}$$

其中φ_0表示交通畅通状态下的拥堵延时指数.

2.3.4 充电成本项

电动冷藏车充电时, 会产生经济和时间成本, 统称为充电成本, 且随着每一次充电而严格递增. 在奖励机制中考虑充电成本, 激励智能体作出经济性决策. 充电成本项r_{cost}定义为

$$r_{\text{cost}} = -(\theta_1 ptc + \theta_2 t). \tag{17}$$

其中: θ_1和θ_2为比例系数, 分别调节充电经济成本和充电时间成本在奖励值中的权重; c为充电价格, 单位为元/kWh.

2.3.5 电量消耗项

在电池容量约束下, 引入电量消耗项r_{energy}以有效控制能耗, 通过合理规划路径减少能耗, 避免不必要的充电次数和充电时长, 优化充电策略, 提高整体运行效率. r_{energy}的定义基于动力系统能耗 (式 (9)) 和制冷系统能耗 (式 (10)) 两部分, 具体如下:

$$r_{\text{energy}} = -\delta(\Delta E_p + \Delta E_c), \tag{18}$$

其中δ为电量消耗项参数, 用于决定能耗的惩罚程度.

2.4 算法改进设计

针对电动冷藏车跨区域路径优化中续航受限、策略更新不稳定、路径搜索效率低等复合挑战, 本文对Q学习算法进行改进. 通过构建多模块协同优化框架, 在不增加环境信息依赖的前提下提升决策效率: 1) 融合电量衰减预测的启发式奖励机制, 将电量安全阈值纳入策略评估过程; 2) 将余弦退火学习率[21]引入冷链物流路径规划领域, 提高策略更新的稳定性与收敛效率; 3) 结合指数衰减探索策略, 通过预实验标定的差异化衰减速率, 实现场景导向的探索强度调控. 上述创新模块与第 2 节所设计的强化学习架构形成多目标协调系统, 提升整体性能.

2.4.1 启发式奖励机制

基于路径状态的稀疏奖励机制是Q学习中常用的定义方法, 如式 (12) 所示. 对于行驶在可通行道路上的电动冷藏车, 如果电量不为零, 则其奖励值保持恒定. 然而, 稀疏奖励机制可能导致智能体缺乏激励差异, 难以感知电量动态变化对配送风险的影响. 尤其在训练初期, 智能体在跨区域环境中盲目搜索, 无法高效学习电量约束下的最优策略. 针对这一问题, 本文提出融合电量衰减预测的启发式奖励机制, 使智能体能够实时评估电量水平, 并据此调整配送策略: 当电量降低时, 奖励值相应减少, 以引导智能体优先选择更加稳健的路径; 当电量降为 0 时, 则给予更大的负惩罚$r_{\text{depletion}}$, 以避免因电量耗尽而导致配送失败. 该机制可增强智能体对电量管理的敏感性, 提升其前瞻性决策能力, 平衡配送效率与能源安全. 启发式奖励函数r_{state}定义为

$$r_{\text{state}} = \begin{cases} r_{\text{obstacle}}, \ O = \text{True}; \\ r_{\text{target}}, \ T = \text{True}; \\ -\mu_1 \mathrm{e}^{-\mu_2 E}, \ O = \text{False 且} E > 0; \\ r_{\text{depletion}}, \ O = \text{False 且} E = 0. \end{cases} \tag{19}$$

其中: μ_1和μ_2是电量启发项参数, μ_1用于调节电量变化对奖励值的影响, μ_2则控制电量变化时奖励值变化的速率.

2.4.2 余弦退火学习率

学习率α在Q学习算法中决定新信息覆盖旧信息的速度. 较高的α意味着在每次更新Q值时, 新估计值对旧Q值的影响较大, 这在初期有助于加快Q表的收敛. 然而, 较高的α可能导致算法陷入局部最优, 并在后期出现震荡, 甚至可能无法收敛到全局最优路径. 反之, 较低的α虽然在初期收敛速度慢, 但在后期具有更稳定的性能. 针对这一矛盾, 本文引入余弦退火学习率, 对静态学习率α_{static}进行改进.

余弦退火学习策略基于模拟退火算法中“退火”思想, 引入余弦函数实现学习率平滑衰减, 模拟温度逐步冷却的动态过程. 具体而言, 智能体在训练初期以初始学习率α_{initial}迅速适应跨区域环境, 经过N轮迭代后, 学习率逐渐衰减至终值α_{final}, 以确保算法有效收敛至最优路径. 电动冷藏车进行跨区域配送过程中, 面临多变环境和复杂约束, 余弦退火学习率有助于模型初期快速学习有效策略, 后期则通过精细调整提升路径规划的精度与适应性, 从而兼顾计算效率与决策质量. 改进后的第k轮次学习率α_k定义为

$$\alpha_k = \alpha_{\text{final}} + 0.5(\alpha_{\text{initial}} - \alpha_{\text{final}})\left(1 + \cos\left(\frac{\pi k}{N}\right)\right). \tag{20}$$

给定$\alpha_{\text{initial}} = 0.8$, $\alpha_{\text{final}} = 0.01$, $N = 5\,000$, α_k与α_{static}的对比如图 6 所示.

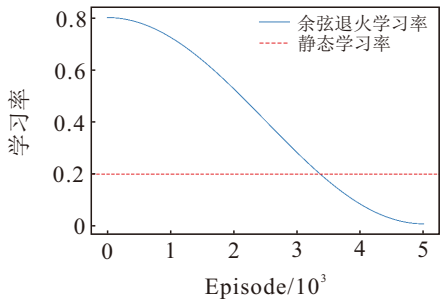


图6 学习率对比

2.4.3 指数衰减探索率

ϵ-贪婪策略常用于Q学习算法, 平衡探索与利用. ϵ代表探索率, 又称贪婪因子, 取值范围为$(0, 1)$.

ϵ表示随机选取动作的概率, 而$1-\epsilon$表示选择现有策略下Q值最大动作的概率. 当ϵ趋近于 1 时, 智能体倾向于探索道路环境, 当ϵ趋近于 0 时, 倾向于利用以往经验选择价值最高的动作. 然而, 固定的ϵ值忽视了智能体在环境试错中的渐进性学习. 较大的ϵ会增加探索性, 导致延迟收敛, 而较小的ϵ则可能使智能体陷入局部最优路径. 为此, 本文引入指数衰减探索率, 对静态探索率ϵ_{static}进行改进.

指数衰减探索策略通过逐步降低探索概率, 使得在学习过程后期, 模型能够更多依赖已有知识进行决策, 从而实现稳定的优化. 针对电动冷藏车的路径优化问题, 路径选择需要兼顾行驶距离、能源消耗、温度控制等多个因素, 同时应对实际运行中的不确定性和复杂性. 在多目标优化场景下, 初始探索率$\epsilon_{\text{initial}}$允许智能体在初期进行大范围路径开发, 以发现多样化的潜在路径. 随着训练轮次递增, 探索因子以衰减率decay_rate逐渐降低至最终探索率ϵ_{final}, 有助于模型在复杂决策空间中逐渐聚焦于最佳路径, 避免无效探索和不必要的计算开销. 改进后的第k轮次的贪婪因子ϵ_k的定义如下:

$$\text{decay\_rate} = \frac{\ln(\epsilon_{\text{final}} \,/\, \epsilon_{\text{initial}})}{N}, \tag{21}$$

$$\epsilon_k = \epsilon_{\text{initial}} \cdot \mathrm{e}^{\text{decay\_rate}\cdot k}. \tag{22}$$

给定$\epsilon_{\text{initial}} = 0.5$, $\epsilon_{\text{final}} = 0.001$, $N = 5\,000$, ϵ_k与ϵ_{static}的对比如图 7 所示.

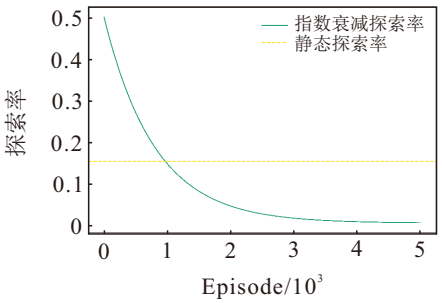


图7　探索率对比

将改进后的Q学习算法用于电动冷藏车跨区域路径优化, 具体训练过程如图 8 所示.

3　实验与分析

3.1　实验配置

实验配置环境 Python 3.7.3, 操作系统 Windows11 x64, 处理器 Intel(R) Core(TM) i9-13980HX 2.20 GHz. 本文通过 3 组敏感性实验分析, 验证所提出的改进Q学习算法对电动冷藏车路径优化的可行性, 考察其在复杂环境中的适应性与鲁棒性. 实验分别模拟交通运行状态、初始电量和能耗率的变化, 对比 4 种

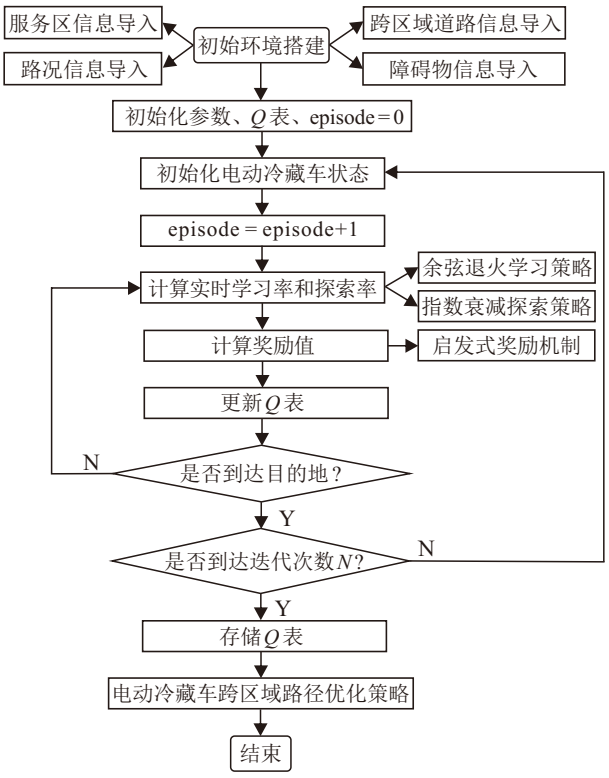


图8　算法流程

Q学习算法的收敛效率与运算结果. 实验对象包括基准算法 (BAS-QL)、文献 [12] 的算法 (SAE-QL)、集成动态学习与探索策略的改进算法 (DDS-QL), 以及本文最终提出的融合启发式奖励机制的双动态策略算法 (HDDS-QL). 为确保公平性, 所有算法均基于统一的多目标奖励函数框架进行测试. 其中: BAS-QL 采用固定参数配置 (α_{static}, ϵ_{static}), DDS-QL 与 HDDS-QL 的唯一区别在于奖励函数的电量启发式约束项. 训练中, 当连续 10 个 episode 内的累积奖励值达到最大值时, 判定算法收敛.

3.2　参数设置

本文的参数定义体系包括 4 类: 系统参数、学习策略参数、探索策略参数和环境反馈参数, 如表 3 所示. 其中: 电池容量等参数参考相关领域文献定义, 平均行驶速度等参数依据实际情景合理取值, 折扣因子、初始学习率等参数通过对比预实验结果设定. 奖励值相关参数的设计充分考虑了各优化目标的数量级差异, 并采用“差异化奖惩-权重平衡”机制, 以确保奖励信号的有效性和任务约束的合理性. 具体而言: 1) 对于避障与电量亏空等任务失败的硬性约束条件, 以及抵达目的地的终极目标, 采用高幅值奖惩, 增强智能体对关键状态的敏感性; 2) 对于充电成本、能耗、拥堵等软约束优化项, 通过调整相关比例因子, 确保不同优化目标之间的权衡与平衡, 从而提升策略的适应性. 具体取值见表 3.

表3 参数设置

类别	参数	符号	值
系统参数	电池容量[22]	E_{bat}	80
	初始电量	E_0	10, 30, 60
	电量阈值	E_{th}	17
	单位里程行驶耗电[1]	e_p	0.28, 0.36, 0.45
	单位时间制冷耗电[23]	e_c	4, 6, 10
	位置转移时路径长度	L	1.82
	平均行驶速度	v	60
	充电功率	p	60
	充电时长	t	0.75
	充电价格	c	1.65
学习策略参数	折扣因子	γ	0.9
	初始学习率	α_{initial}	0.8
	最终学习率	α_{final}	0.01
	第k轮的学习率	α_k	式(20)
	静态学习率	α_{static}	0.2
探索策略参数	初始探索率	$\epsilon_{\text{initial}}$	0.5
	最终探索率	ϵ_{final}	0.001
	第k轮的探索率	ϵ_k	式(22)
	静态探索率	ϵ_{static}	0.15
环境反馈参数	目标点即时奖励	r_{target}	100
	碰撞惩罚	r_{obstacle}	−5 000
	零电量惩罚	$r_{\text{depletion}}$	−100
	反向移动惩罚	η	−1
	到目标点的欧氏距离	d_t	式(15)
	比例因子	λ	1.5
	拥堵延时指数	φ	1.0, 1.5, 2.0, 4.0
	畅通时拥堵延时指数	φ_0	1.0
	比例因子	θ_1	0.001
	比例因子	θ_2	0.3
	比例因子	δ	0.2
	比例因子	μ_1	1
	比例因子	μ_2	0.1
	电量维持奖励	r_{fixed}	0

3.3 敏感性分析

围绕交通运行状态、初始电量以及能耗率实施3组敏感性实验 (Exp 1 ∼ Exp 3), 每组敏感性实验包含 BAS-QL、SAE-QL、DDS-QL 和 HDDS-QL 的子实验 (分别为 Exp 1-1, Exp 1-2; Exp 2-1, Exp 2-2; Exp 3-1, Exp 3-2), 以比较算法性能. 鉴于实验结果的随机性, 4 种算法在指定场景下重复模拟 15 次. 每次模拟中, 均记录算法完成预设训练轮次所需的运行时间, 并收集其最终输出的配送路径规划结果, 包括总里程、拥堵时长、充电费用及耗电量等关键指标. 为检验改进效果是否具有统计学意义, 采用 Welch's t-test 进行分析, 显著性水平设定为 0.05. 如表 4 所示, 实验结果通过三元标注系统进行标识: Y 表示存在统计显著性差异 ($p < 0.05$), N 表示无显著差异, NA 表示不适用.

研究结果表明, 相比其他 3 种Q学习算法, HDDS-QL 在全部测试场景中均能显著降低总里程与电量消耗. 在计算效率方面, 该算法展现出整体性能提升, 但部分场景的运行时间差异未呈现统计显著性. 分析发现, BAS-QL 和 SAH-QL 由于未能找到最优配送路径, 在后期仍需进行大范围的路径探索, 导致车辆误入不可通行区域, 避开了拥堵及充电桩. 因此, 在部分子实验中, 尽管这两种算法在拥堵时长和充电费用指标上与 HDDS-QL 存在量化差异, 但未达到显著水平. 对于采用相同双动态策略的 DDS-QL, 除 Exp 3-1 外, 其路径规划结果与 HDDS-QL 具有一定相似性. 虽然两者在拥堵区域覆盖和充电决策方面表现趋同, 但由于 DDS-QL 的奖励函数未引入电量连续约束机制, 最终生成的路径存在局部冗余. 针对每种情景, 本节从 15 次模拟实验中选取一次进行具体分析.

3.3.1 交通运行状态

将交通运行状态划分为低密度拥堵 (常态) 和高密度拥堵 (高峰时段) 两种典型场景. 基于高德智慧交通平台发布的实时交通数据与历史交通流模式分析, 对关键节点 (含交通枢纽、服务区进出入口等) 的拥堵指数进行人工调整, 以确保仿真环境符合实际路网特征. Exp 1 中, 设置初始电量 $E_0 = 30\,\text{kWh}$、单位里程行驶耗电 $e_p = 0.36\,\text{kWh/km}$ 和单位时间制冷耗电 $e_c = 6\,\text{kWh/h}$ 作为基准参数. 实验结果可视化中, 紫色直线 (BAS-QL)、蓝色虚线 (SAE-QL)、棕色空心圆 (DDS-QL) 与绿色实心圆 (HDDS-QL) 分别表征各算法路径优化结果, 闪电图标则标识执行了充电决策.

图 9 展示出算法性能的显著差异. 低密度拥堵实验 Exp 1-1 中, 经过 3 500 次训练轮次, 除基准算法外, 其他 3 种Q学习算法选择了一条近似配送路径. 其中: HDDS-QL 表现尤为突出, 实现路径长度缩减 25.5 ∼ 92.8 km, 能耗降低 12.6 ∼ 45.9 kWh. 高密度拥堵实验 Exp 1-2 中, HDDS-QL 智能规避两处严重拥堵路段, 规划路径总里程 (111.0 km) 与总能耗 (55.5 kWh) 均达到最优值. 尽管 BAS-QL 与 DDS-QL 两种算法, 相比于 HDDS-QL, 平均路径重合率达 74.9%, 且均在枫泾服务区完成一次充电操作, 但产生了无效路径. 其中: 基准算法冗余里程占比达 31.5%, 导致冷藏车在到达目的地前电量耗尽.

图 10 给出 4 种算法在不同路况下经过 500 ∼ 3 500 次训练的步数变化. 尽管所有算法在整个训练过程均呈现“L”形趋势 (如图 10(a) 和图 10(b) 右上角), 且在前 500 次迭代中步数从峰值迅速下降, 但仅有 DDS

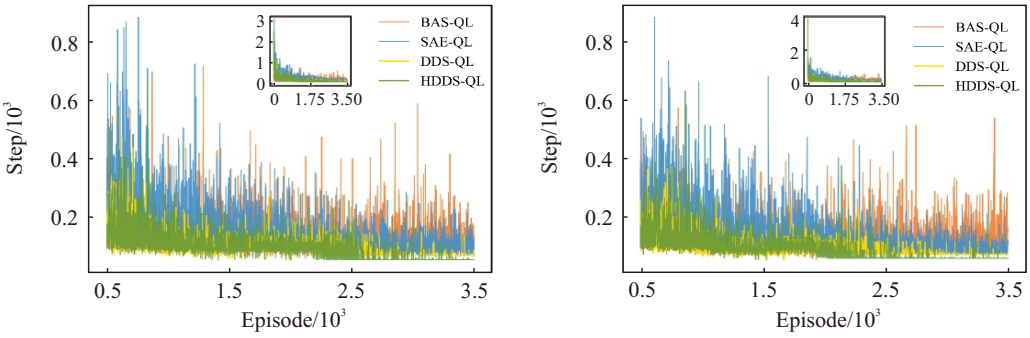
表4　4 种算法的性能对比及统计分析

实验	指标	BAS-QL			SAE-QL			DDS-QL			HDDS-QL		
		Mean	Std.Dev.	*t*-test	Mean	Std.Dev.	*t*-test	Mean	Std.Dev.	*t*-test	Mean	Std.Dev.	*t*-test
Exp 1-1	运行时间 /s	4.1	0.2	N	4.1	0.2	N	4.0	0.2	N	4.0	0.2	NA
	总里程 /km	173.0	17.8	Y	154.2	12.9	Y	129.7	1.3	Y	103.7	0.0	NA
	拥堵时长 /min	10.9	3.9	N	10.3	5.2	N	9.9	0.0	N	9.9	0.0	NA
	充电费用 /元	89.1	30.8	N	84.1	26.1	N	74.2	0.0	N	74.2	0.0	NA
	耗电量 /kWh	86.2	8.7	Y	76.9	6.7	Y	64.8	0.6	Y	52.0	0.0	NA
Exp 1-2	运行时间 /s	4.1	0.3	Y	4.0	0.2	N	3.9	0.2	N	3.9	0.2	NA
	总里程 /km	153.0	19.1	Y	152.8	11.1	Y	136.5	0.0	Y	112.0	1.7	NA
	拥堵时长 /min	10.3	3.6	Y	8.9	1.8	Y	7.2	0.0	N	7.2	0.0	NA
	充电费用 /元	84.1	26.1	N	84.1	26.1	N	74.2	0.0	N	74.2	0.0	NA
	耗电量 /kWh	76.2	9.5	Y	76.1	5.6	Y	68.1	0.0	Y	56.0	0.8	NA
Exp 2-1	运行时间 /s	4.8	0.2	N	4.7	0.2	N	4.7	0.1	N	4.7	0.2	NA
	总里程 /km	163.9	27.8	Y	158.1	23.7	Y	121.9	0.0	Y	119.5	1.8	NA
	拥堵时长 /min	6.3	3.0	Y	5.8	3.7	Y	2.7	0.0	N	2.7	0.0	NA
	充电费用 /元	34.6	38.3	Y	34.6	38.3	Y	0.0	0.0	N	0.0	0.0	NA
	耗电量 /kWh	81.3	13.7	Y	78.4	11.7	Y	60.4	0.0	Y	59.2	0.9	NA
Exp 2-2	运行时间 /s	6.4	0.6	Y	6.3	0.4	Y	5.1	0.2	N	5.0	0.3	NA
	总里程 /km	215.1	43.8	Y	208.3	31.1	Y	124.9	1.5	Y	122.9	1.7	NA
	拥堵时长 /min	5.8	2.9	Y	5.1	2.9	N	3.6	0.0	N	3.6	0.0	NA
	充电费用 /元	193.1	37.7	Y	193.1	37.7	Y	148.5	0.0	N	148.5	0.0	NA
	耗电量 /kWh	106.7	21.8	Y	103.3	15.4	Y	61.9	0.7	Y	61.0	0.8	NA
Exp 3-1	运行时间 /s	6.0	0.2	Y	5.8	0.3	Y	5.8	0.4	Y	5.4	0.2	NA
	总里程 /km	174.2	59.2	Y	155.1	31.7	Y	142.1	29.7	Y	105.2	1.9	NA
	拥堵时长 /min	11.9	1.6	Y	11.1	1.9	Y	10.8	1.3	Y	9.9	0.0	NA
	充电费用 /元	118.8	54.7	Y	113.8	47.5	Y	108.9	38.4	Y	74.2	0.0	NA
	耗电量 /kWh	106.6	35.8	Y	94.9	19.0	Y	87.0	17.9	Y	64.7	1.1	NA
Exp 3-2	运行时间 /s	6.7	0.1	Y	6.6	0.1	Y	6.3	0.3	N	6.2	0.3	NA
	总里程 /km	195.7	49.5	Y	164.9	20.7	Y	145.0	11.1	Y	133.6	11.8	NA
	拥堵时长 /min	4.0	3.0	N	3.9	2.8	N	2.7	0.0	N	2.7	0.0	NA
	充电费用 /元	89.1	30.8	N	74.2	0.0	N	74.2	0.0	N	74.2	0.0	NA
	耗电量 /kWh	64.7	16.3	Y	54.5	6.8	Y	47.9	3.6	Y	44.1	3.9	NA

(a) Exp 1-1　　(b) Exp 1-2

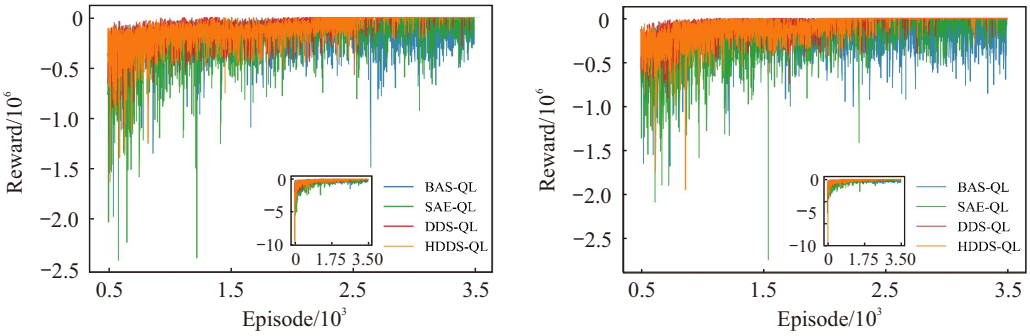
图9　Exp 1 路径优化结果

-QL 与 HDDS-QL 实现稳定收敛. 这得益于动态探索策略与学习机制的协同作用, 有效提升了冷藏车在跨区域配送中的环境适应能力. 然而, 由于 DDS-QL 奖励函数未包含实时电量约束, 其收敛路径偏离全局最优解. Exp 1-1 中, HDDS-QL 在第 2 561 次训练达到收敛, 找到最佳路径; Exp 1-2 中, 该算法从第 2 278 次迭代开始收敛. 这表明, HDDS-QL 在不同交通拥堵情境下保持适应性, 并在复杂路况中展现出



(a) Exp 1-1 (b) Exp 1-2

图10 Exp 1 步数变化



(a) Exp 1-1 (b) Exp 1-2

图11 Exp 1 奖励值变化

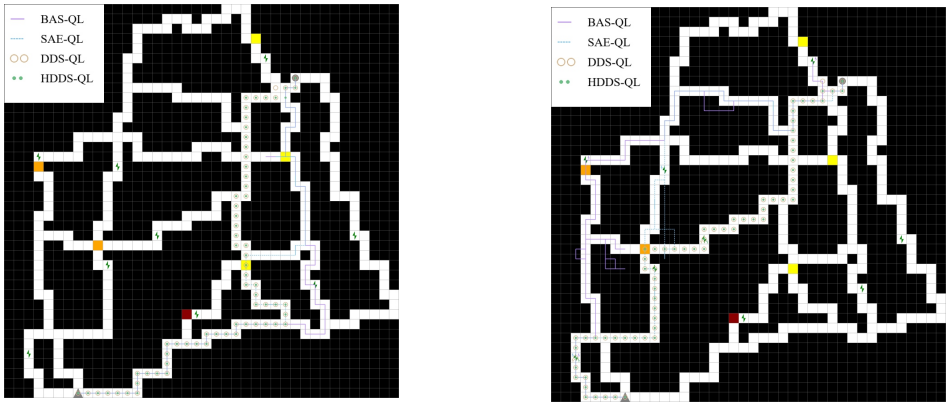
更高的效率. 相反, BAS-QL 和 SAE-QL 在 Exp 1-1 与 Exp 1-2 中均陷入局部最优, 即使在训练后期, 其步数仍在 200 步附近呈现持续发散状态, 反映出策略空间未能形成稳态分布.

图 11 展示了 4 种算法在整个训练过程中累积奖励值的变化趋势, 均呈现“Γ”形, 且在前 500 次迭代中奖励值从初始低位迅速攀升. 进入 500 ～ 3 500 次训练阶段后, BAS-QL 和 SAE-QL 的奖励值持续振荡, 而 DDS-QL 和 HDDS-QL 则快速收敛. 特别地, Exp 1-1 中, 即便引入电量衰减惩罚机制 (见式 (19)), HDDS-QL 仍取得 78.9 的最大奖励值, 较 DDS-QL 提升 7.1 个百分点; Exp 1-2 中, 该优势依旧保持在 6.2 个百分点. 对比实验再次验证了本文改进算法在复杂交通环境下的优越性和稳健性.

3.3.2 初始电量

Exp 2 设定两种初始电量水平E_0, 分别为 60 kWh 和 10 kWh. 针对常态低密度拥堵场景, 关键能耗参数 (单位里程行驶耗电e_p、单位时间制冷耗电e_c) 取基准值, 分别为 0.36 kWh/km 与 6 kWh/h.

图 12 呈现 4 种算法在不同初始荷电状态下路径优化结果. Exp 2-1 中, DDS-QL 经过 4 000 次训练迭代, 实现总里程 121.9 km 与能耗 60.4 kWh (较 HDDS-QL 最优值 118.3 km/58.6 kWh, 性能差距约 3.0%), 且全程无需充电. 尽管 BAS-QL 与 SAE-QL 同样未触发充电策略, 但其里程与能耗明显高于 HDDS-QL. Exp 2-2 中, 较低初始电量要求电动冷藏车先补能再

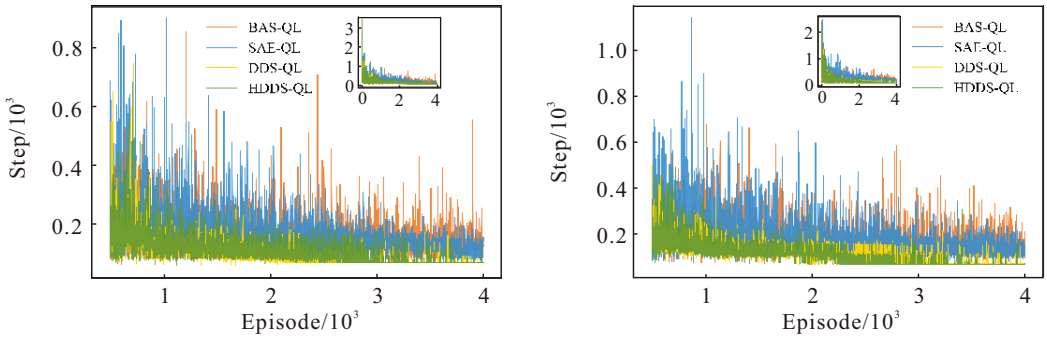


(a) Exp 2-1 (b) Exp 2-2

图12 Exp 2 路径优化结果

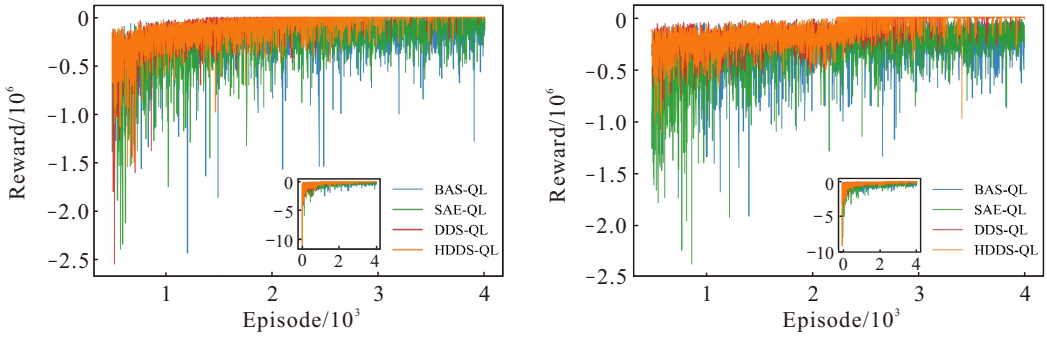
配送. 因此, 4 种算法规划的路径均经过最近可供充电的新塍服务区, 并在后续经历二次充电, 但在路径质量上存在差异. HDDS-QL 通过引入基于荷电状态的启发式奖励函数, 结合动态探索因子和学习率协同调整策略, 实现全局最优路径规划. 然而, DDS-QL 因路径回溯现象产生 7.3 km 的冗余里程, BAS-QL 和 SAE-QL 则因绕行导致能耗分别增加 95.2% 和 50.6%, 显著影响配送效率.

如图 13 和图 14 所示, 随着学习经验积累, 智能体对环境熟悉度提升, 寻找路径所需的步数逐渐减少, 累积奖励值也不断增加. 电动冷藏车的学习目标旨在使累积奖励值收敛至最优值. 经过 4 000 次训练, BAS-QL/SAE-QL 与上述目标仍相去甚远; DDS-QL 通过动态双因子调节实现次优收敛; 唯有 HDDS-QL 率先达成学习目标. Exp 2-1 中, HDDS-QL 经历 2 942 轮训练开始收敛, 并找到最优路径; Exp 2-2 中, 收敛所需训练轮数减少 267 轮, 其可能的原因是, 当初始电量不足时, 电动冷藏车的选择受限, 必须优先经过新塍服务区进行充电, 因而路径选择相对简单; 而当初始电量充足时, 为避免补能产生不必要的经济成本和时间成本, 选择最佳配送路径须进行更广泛的探索.



(a) Exp 2-1　　(b) Exp 2-2

图13　Exp 2 步数变化



(a) Exp 2-1　　(b) Exp 2-2

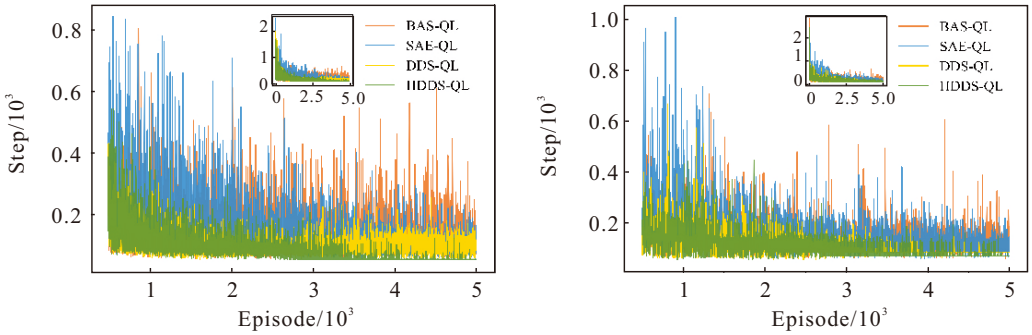
图14　Exp 2 奖励值变化

3.3.3　能耗率

Exp 3 设定高低两种能耗率: e_p=0.45 kWh/km, e_c = 10 kWh/h; e_p = 0.28 kWh/km, e_c = 4 kWh/h. 在常态低密度拥堵下, 初始电量E_0设为 30 kWh.

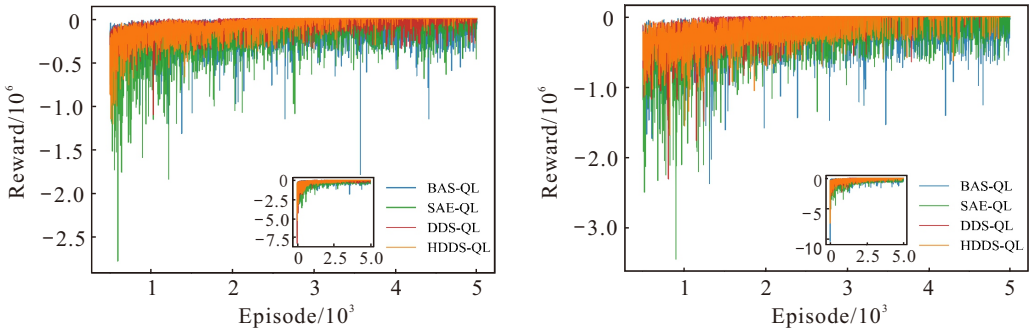
(a) Exp 3-1　　(b) Exp 3-2

图15　Exp 3 路径优化结果



(a) Exp 3-1 (b) Exp 3-2

图16 Exp 3 步数变化



(a) Exp 3-1 (b) Exp 3-2

图17 Exp 3 奖励值变化

图 15 展示出 4 类算法在不同能耗场景下的性能差异. Exp 3 结果表明, 所有算法均需进行一次充电操作, 其中 HDDS-QL 表现最佳. Exp 3-1 中, 为避免电量耗尽, 各算法均倾向选择途经枫泾服务区的节能路径, 而 HDDS-QL 实现最短配送距离 (103.7 km). Exp 3-2 中, 各算法能够有效规避严重拥堵路段, 拥堵时长相比 Exp 3-1 减少超过 57%. 此外, BAS-QL、SAE-QL 与 DDS-QL 三种算法所规划路径在里程及能耗上高度相似, 较 HDDS-QL 产生 23.2% 的路径冗余率与 23.4% 的额外能耗.

图 16 和图 17 表明 BAS-QL 和 SAE-QL 在整个训练周期内未出现明显收敛趋势, 主要原因在于其探索和学习过程不够充分. 相反, DDS-QL 与 HDDS-QL 采用动态衰减策略 ($\alpha: 0.8 \xrightarrow{\cos} 0.01$, $\epsilon: 0.5 \xrightarrow{\exp} 0.001$), 在初期维持高效探索和学习机制, 激励智能体积极开发潜在路径. 随着训练推进, 参数衰减引发算法收敛行为分化: 低能耗场景下, HDDS-QL 通过引入荷电状态惩罚项成功定位全局最优路径, 而 DDS-QL 虽收敛但易陷入局部最优; 高能耗场景下, 由于单次路径决策导致电量骤减 (最高可达 2.5%/step), DDS-QL 未耦合实时电量反馈机制的缺陷进一步放大, 即使经过 5 000 次迭代仍无法收敛, 而 HDDS-QL 能够维持策略收敛性. 实验数据表明, 尽管能耗率水平不同, HDDS-QL 达到收敛所需的训练轮数无明显差异, Exp 3-1 中, 经过 3 405 轮训练后开始逐渐收敛, Exp 3-2 中, 完成 3 307 次迭代后实现收敛. 这表明本文改进算法具有较强的鲁棒性, 能够适应不同能耗环境的电动冷藏车路径优化.

4 结 论

本文针对电动冷藏车跨区域路径优化问题, 提出了一种基于动态策略的启发式Q学习算法, 提升了冷链配送效率. 首先, 通过建立强化学习模型, 将充电决策与路径规划整合, 实现动作降维, 提高算法运行效率; 然后, 构建以电量为启发式信息的多目标奖励函数, 进一步提升路径优化性能; 最后, 设计动态学习和探索策略, 使得智能体在学习过程中快速探索并稳定收敛至最优路径. 3 组敏感性实验均表明, 本文改进算法在收敛效率与路径质量方面均显著优于对比算法, 不仅有效缩短拥堵时长与行驶总里程, 还可优化能源利用与运营成本. 这表明所做改进提升了算法的学习性能, 可成功为电动冷藏车定制跨区域配送路线. 考虑到Q表储存容量有限, 未来研究将致力于融合深度学习与Q学习框架, 以便更高效地解决跨区域路径优化问题.

参考文献 (References)

[1] Zhang S Y, Guan C L, Qiu Y G, et al. Multi-objective route optimization of urban cold chain distribution using electric and diesel powered vehicles[J]. Research in Transportation Business & Management, 2023, 49: 100969.

[2] Li D, Li K. A multi-objective model for cold chain logistics considering customer satisfaction[J]. Alexandria Engineering Journal, 2023, 67: 513-523.

[3] 谭晓伟, 王雪韵, 胡大伟. 考虑动态需求的多中心沿途补货冷链物流配送路径优化[J]. 四川大学学报: 自然科学版, 2023, 60(2): 70-80.
(Tan X W, Wang X Y, Hu D W. Research on distribution routing optimization of multi-center cold chain logistics for replenishment along the way considering dynamic demand[J]. Journal of Sichuan University: Natural Science Edition, 2023, 60(2): 70-80.)

[4] He M L, Yang M, Fu W Q, et al. Optimization of electric vehicle routes considering multi-temperature co-distribution in cold chain logistics with soft time windows[J]. World Electric Vehicle Journal, 2024, 15(3): 80.

[5] Arias-Londoño A, Gil-González W, Montoya O D. A linearized approach for the electric light commercial vehicle routing problem combined with charging station siting and power distribution network assessment[J]. Applied Sciences, 2021, 11(11): 4870.

[6] Wu Z G, Zhang J L. A branch-and-price algorithm for two-echelon electric vehicle routing problem[J]. Complex & Intelligent Systems, 2023, 9(3): 2475-2490.

[7] Fan L J. A two-stage hybrid ant colony algorithm for multi-depot half-open time-dependent electric vehicle routing problem[J]. Complex & Intelligent Systems, 2024, 10(2): 2107-2128.

[8] Chen Y N, Xue J H, Zhou Y M, et al. An efficient threshold acceptance-based multi-layer search algorithm for capacitated electric vehicle routing problem[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(6): 5867-5879.

[9] Wu J H, Sun Y N, Li D Y, et al. An adaptive conversion speed Q-learning algorithm for search and rescue UAV path planning in unknown environments[J]. IEEE Transactions on Vehicular Technology, 2023, 72(12): 15391-15404.

[10] Huang Y C, Wang C. Improved Q-learning algorithm for AGV path optimization[C]. Advanced Manufacturing and Automation XIII. Singapore: Springer Nature Singapore, 2024: 55-60.

[11] Zhou Q, Lian Y, Wu J Y, et al. An optimized Q-Learning algorithm for mobile robot local path planning[J]. Knowledge-Based Systems, 2024, 286: 111400.

[12] Zhong Y, Wang Y H. Cross-regional path planning based on improved Q-learning with dynamic exploration factor and heuristic reward value[J]. Expert Systems with Applications, 2025, 260: 125388.

[13] Blasi L, D'Amato E, Mattei M, et al. UAV path planning in 3-D constrained environments based on layered essential visibility graphs[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(3): 2359-2375.

[14] 曹路阳, 周乐来, 戴骁蒙, 等. 基于分层边界与可视图的移动机器人自主探索算法研究[J]. 控制与决策, 2025, 40(4): 1207-1216.
(Cao L Y, Zhou L L, Dai X M, et al. An autonomous exploration algorithm of mobile robots based on hierarchical frontier and visibility graph[J]. Control and Decision, 2025, 40(4): 1207-1216.)

[15] Shang Z X, Shen Z G. Topology-based UAV path planning for multi-view stereo 3D reconstruction of complex structures[J]. Complex & Intelligent Systems, 2023, 9(1): 909-926.

[16] 薛阳, 倪大斌, 卢秋红, 等. 基于 PGWO 算法的移动机器人路径规划[J]. 控制与决策, 2025, 40(4): 1395-1401.
(Xue Y, Ni D B, Lu Q H, et al. Path planning of mobile robot based on PGWO algorithm[J]. Control and Decision, 2025, 40(4): 1395-1401.)

[17] Watkins C J C H. Learning from delayed rewards[D]. Cambridge: University of Cambridge, 1989.

[18] Bellman R. A Markovian decision process[J]. Journal of Mathematics and Mechanics, 1957, 6(5): 679-684.

[19] Wang X, Wang S, Liang X, et al. Deep reinforcement learning: A survey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(4): 5064-5078.

[20] 王小康, 冀杰, 刘洋, 等. 基于改进Q学习算法的无人物流配送车路径规划[J]. 系统仿真学报, 2024, 36(5): 1211-1221.
(Wang X K, Ji J, Liu Y, et al. Path planning of unmanned delivery vehicle based on improved Q-learning algorithm [J]. Journal of System Simulation, 2024, 36(5): 1211-1221.)

[21] Loshchilov I, Hutter F. SGDR: Stochastic gradient descent with warm restarts[J/OL]. 2016, arxiv: 1608.03983v5.

[22] 宋丽英, 赵世超, 卞骞, 等. 低碳视角下城乡区域混合车队生鲜配送路径问题研究[J]. 交通运输系统工程与信息, 2023, 23(6): 250-261.
(Song L Y, Zhao S C, Bian Q, et al. Fresh food distribution route optimization of mixed fleets in urban and rural areas under low carbon perspective[J]. Journal of Transportation Systems Engineering and Information Technology, 2023, 23(6): 250-261.)

[23] Leng L L, Wang Z, Zhao Y W, et al. Formulation and heuristic method for urban cold-chain logistics systems with path flexibility — The case of China[J]. Expert Systems with Applications, 2024, 244: 122926.

作者简介

王岩红 (1986–), 女, 副教授, 博士, 主要研究方向为机器学习、决策智能, E-mail: yanhong.wang@hotmail.com;

钟颖 (1995–), 女, 硕士生, 主要研究方向为无人车路径优化、强化学习, E-mail: zhongying0815@outlook.com;

张允华 (1989–), 男, 副教授, 博士, 主要研究方向为移动源节能减排技术, E-mail: zhangyunhua@tongji.edu.cn.