

容量约束电动汽车路径问题的Q-Learning超启发式优化模型与动态规划充电策略

杨 洋, 刘志威, 李伍奕, 李志莹

重庆科技大学数理科学学院, 重庆

收稿日期: 2026年9月10日; 录用日期: 2026年10月5日; 发布日期: 2026年10月14日

摘 要

本文研究容量约束电动汽车路径问题(CEVRP), 在车辆载重与电池电量双重约束下, 建立以车队总行驶距离最小为目标的0-1混合整数规划模型。令客户集、充电站副本集和扩展节点集分别为 C 、 F' 和 V 。以 x_{ij} 表示弧选择变量, 以 w_i 和 b_i 表示载重与剩余电量状态, 目标函数为 $\min Z = \sum_{(i,j) \in A} d_{ij} x_{ij}$, 并以唯一服务、流守恒、载重递推和电量递推约束定义可行域。针对该NP难组合优化问题, 将低层启发式算子选择建模为有限马尔可夫决策过程 $\mathcal{M} = (S, A, P, R, \gamma)$, 依据Bellman最优方程更新 $Q(s, a)$, 实现自适应算子调度。对候选客户路径, 先插入充电站恢复电量可行性, 再以动态规划递推优化固定客户序列的充电站组合; 初始化阶段依据实例特征确定构造策略的采样权重。在17个IEEEWCCI2020基准实例上进行30次独立运行, 并采用Friedman总体检验及经Holm校正的双侧Wilcoxon符号秩事后检验, 在显著性水平 $\alpha = 0.05$ 下评估算法差异。数值结果表明, 所提算法在Best与Mean指标上具有总体竞争优势, 且所得解满足车辆数、容量与电量约束。

关键词

容量约束电动汽车路径问题, 0-1混合整数规划, 组合优化, 马尔可夫决策过程, Q-Learning超启发式

A Q-Learning Hyper-Heuristic Optimization Model and Dynamic Programming Charging Strategy for the Capacitated Electric Vehicle Routing Problem

Yang Yang, Zhiwei Liu, Wuyi Li, Zhiying Li

School of Mathematical and Physical Sciences, Chongqing University of Science and Technology, Chongqing

文章引用: 杨洋, 刘志威, 李伍奕, 李志莹. 容量约束电动汽车路径问题的 Q-Learning 超启发式优化模型与动态规划充电策略[J]. 运筹与模糊学, 2026, 16(5): 8-21. DOI: 10.12677/orf.2026.165030

Abstract

This paper studies the capacitated electric vehicle routing problem (CEVRP) and formulates a 0-1 mixed-integer programming model that minimizes total fleet travel distance subject to vehicle-capacity and battery-energy constraints. Let C , F' , and V denote the customer set, charging-station-copy set, and extended node set, respectively. The binary arc variable x_{ij} and state variables w_i and b_i describe routing, residual load, and residual energy. Unique-service, flow-balance, load-recursion, and energy-recursion constraints define the feasible region, with objective $\min Z = \sum_{(i,j) \in A} d_{ij} x_{ij}$. For this NP-hard combinatorial optimization problem, low-level heuristic selection is modeled as a finite Markov decision process $\mathcal{M} = (S, A, P, R, \gamma)$, and $Q(s, a)$ is updated by the Bellman optimality equation to adaptively schedule operators. For candidate customer routes, charging-station insertion restores energy feasibility, and a dynamic-programming recurrence then optimizes the charging-station combination for the fixed customer sequence. Instance features determine construction-strategy sampling weights during initialization. Thirty independent runs are conducted on 17 IEEE WCCI 2020 benchmark instances. Differences are evaluated using the Friedman omnibus test and Holm-adjusted two-sided Wilcoxon signed-rank post-hoc tests at $\alpha = 0.05$. The results show that the proposed algorithm is overall competitive in Best and Mean, and every reported solution satisfies the fleet-size, capacity, and energy constraints.

Keywords

Capacitated Electric Vehicle Routing Problem, 0-1 Mixed-Integer Programming, Combinatorial Optimization, Markov Decision Process, Q-Learning Hyper-Heuristic

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

电动汽车配送路径规划同时涉及客户服务顺序与途中补能。相关研究通常通过混合整数规划刻画客户访问、载重传播、电量消耗与充电决策[1]。在此基础上,分支切判定价等精确优化方法可用于求解包含多种充电技术的电动车辆路径问题[2]。

进一步研究围绕充电过程建立改进模型与算法组件,并揭示客户路径结构和充电决策之间的耦合关系[3]。容量约束车辆路径问题的精确求解依赖有效不等式与分支切割框架。Lysgaard 等构造多类容量与组合割,并验证了相应分离算法的求解能力[4]。在电动车辆路径扩展中,Rastani 和 Çatay 将数学规划与大邻域搜索结合,形成面向载重相关能耗和时间窗约束的数学启发式[5]。上述研究分别从精确优化与数学启发式角度说明,模型结构和充电决策的分解能够降低联合求解难度。固定客户顺序下仍存在独立的充电优化空间。Kullman 等将其定义为固定路径车辆充电问题,并提供可嵌入上层路径算法的求解工具[6]。针对非线性充电函数,Bacci 等利用透视割构造精确线性化方法,并通过路径支配预处理缩减候选空间[7]。因此,CEVRP 算法既要生成容量可行的客户路径,也要合理分配充电站插入与固定路径充电优化的计算资源。

搜索控制直接影响邻域算子的使用效率。Tarhini 等将群智能超启发式用于车辆路径问题,表明高层机制可以根据搜索状态组织低层启发式[8]。Fitzpatrick 等将学习机制与组合优化结构结合,以提高容量约束车辆路径问题的大规模求解能力[9]。Pulyassary 等从网络流角度刻画续航、充电策略与站点容量,给出了相关问题的复杂度结果以及精确和近似算法[10]。这些研究表明,CEVRP 的搜索控制既可借助学习机制调度算子,也可依托网络优化结构限定可行决策空间。Karimi-Mamaghan 等进一步将机器学习与元启发式的结合方式归纳为初始化、参数设置、评价与算法选择等类型[11]。Burke 等将超启发式概括为启发式选择与启发式生成两类[12]。Alissa 等从基于特征和无特征两个方向讨论了自动算法选择[13]。Drake 等系统梳理了选择型超启发式的在线决策框架[14],Meignan 等则利用强化学习实现元启发式的自适应调整[15]。据此,本文将 Q-learning 限定为主搜索阶段的算子选择机制。初始化阶段仅以等概率随机方式保持多种构造策略的起点多样性。

为衔接随机初始化、客户路径搜索和充电决策,本文提出一种改进的 Q-learning 超启发式。初始化阶段从最近邻构造、随机客户序列容量构造和极角扫描构造中等概率选择一种策略。客户顺序确定后,启发式充电站插入用于恢复电量可行性,随后采用动态规划优化具有改进潜力的固定客户路径充电方案。主搜索采用多尺度低层启发式,并由 Q-learning 根据修复后完整路径的目标值反馈选择算子。本文在 17 个 IEEE WCCI 2020 实例上评价最终解质量,并通过移除式消融分析主搜索模块的作用。

2. 问题描述

给定一个配送中心,客户集合,充电站集合以及 K 辆同构电动车,每辆车均从配送中心满电出发,完成分配客户的服务后返回配送中心。每个客户具有确定需求且仅能被服务一次,单条路线的累计需求不得超过车辆载重容量。车辆行驶消耗电量,必要时可访问充电站并在完全充电后继续行驶。优化目标是在满足客户服务约束,车辆容量约束和电量约束的前提下最小化车队总行驶距离。

客户分配与车辆内访问顺序共同形成客户路径,但该路径不必然满足电量约束。固定路径充电过程在客户顺序确定后补充充电站访问,因此同一客户路径可能对应多种充电方案。图 1 给出 CEVRP 配送示例,其中圆形节点表示客户,房屋图标表示配送中心,带电池标识的节点表示充电站。

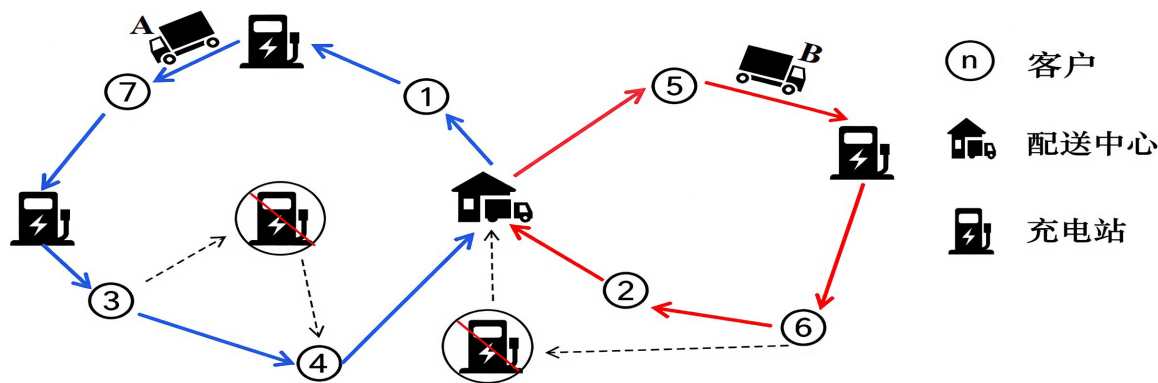


Figure 1. An illustrative CEVRP delivery network

图 1. CEVRP 配送路网示意图

为描述模型,客户节点集合记为 C ,物理充电站集合记为 F 。允许同一物理充电站被多次访问时,以副本节点表示各次充电访问,并构造扩展节点集合 V 。节点 i 至节点 j 的距离记为 d_{ij} ,单位距离能耗系数记为 h 。车辆最大载重容量与电池容量分别记为 W 和 B 。客户 i 的需求量记为 q_i ,车辆完成节点 i 服务后的剩余载重记为 w_i ,到达节点 i 时的剩余电量记为 b_i 。主要符号见表 1。

Table 1. Notations of the CEVRP
表 1. CEVRP 主要符号

符号	含义
x_{ij}	若车辆沿弧 (i,j) 行驶则为 1, 否则为 0
d_{ij}	节点 i 到节点 j 的距离
w_i	车辆完成节点 i 服务后的剩余载重
q_i	客户 i 的需求量
W	车辆最大载重容量
b_i	车辆到达节点 i 时的剩余电量
h	单位距离能耗系数
B	车辆电池容量
C	客户节点集合
F	物理充电站集合
F'	充电站副本集合
V	扩展节点集合
A	弧集合
K	实例先验给定的同构电动车数量

基于上述记号, 采用整数线性规划描述 CEVRP。决策变量 x_{ij} 表示车辆是否沿弧 (i,j) 行驶, w_i 与 b_i 分别表示载重和电量状态。目标函数与约束如下。

$$\min Z = \sum_{(i,j) \in A} d_{ij} x_{ij} \quad (1)$$

$$\begin{aligned} \sum_{i:(i,k) \in A} x_{ik} &= 1, \quad \forall i \in C, \\ \sum_{j:(k,j) \in A} x_{kj} &= 1, \quad \forall j \in C \end{aligned} \quad (2)$$

$$\sum_{i \in V, i \neq k} x_{ik} = \sum_{j \in V, k \neq j} x_{kj}, \quad \forall k \in F' \quad (3)$$

$$w_j \leq W - q_j x_{0j}, \quad \forall j \in C \quad (4)$$

$$w_j \leq w_i - q_j x_{ij} + W(1 - x_{ij}), \quad \forall i \in C \cup F', j \in C \cup F' \cup \{0'\}, (i,j) \in A \quad (5)$$

$$0 \leq w_i \leq W, \quad \forall i \in V \quad (6)$$

$$b_j \leq b_i - h d_{ij} x_{ij} + B(1 - x_{ij}), \quad \forall i \in C, j \in C \cup F' \cup \{0'\}, (i,j) \in A \quad (7)$$

$$b_j \leq B - h d_{ij} x_{ij}, \quad \forall i \in \{0\} \cup F', j \in C \cup F' \cup \{0'\}, (i,j) \in A \quad (8)$$

$$0 \leq b_i \leq B, \quad \forall i \in V \quad (9)$$

式(1)最小化全部车辆的总行驶距离。式(2)保证每个客户仅被访问一次。式(3)规定充电站副本的流平衡。式(4)给出车辆离开配送中心后的初始载重。式(5)描述载重沿路径的传播。式(6)规定载重上界。式(7)描述客户节点间的电量消耗。式(8)描述车辆从配送中心或充电站满电离开后的电量传播。式(9)规定剩余电量范围。后续算法保持该问题定义不变, 并分别处理客户路径搜索与固定路径充电。

3. 算法设计

本文算法以客户路径为主要搜索对象。低层启发式仅作用于满足载重约束的客户序列，固定路径充电优化再将其转换为电量可行的完整路径。随机初始化负责生成搜索起点，Q-learning 负责主搜索中的算子选择，固定路径充电优化负责候选路径的电量可行性与补能优化。图 2 给出总体流程。

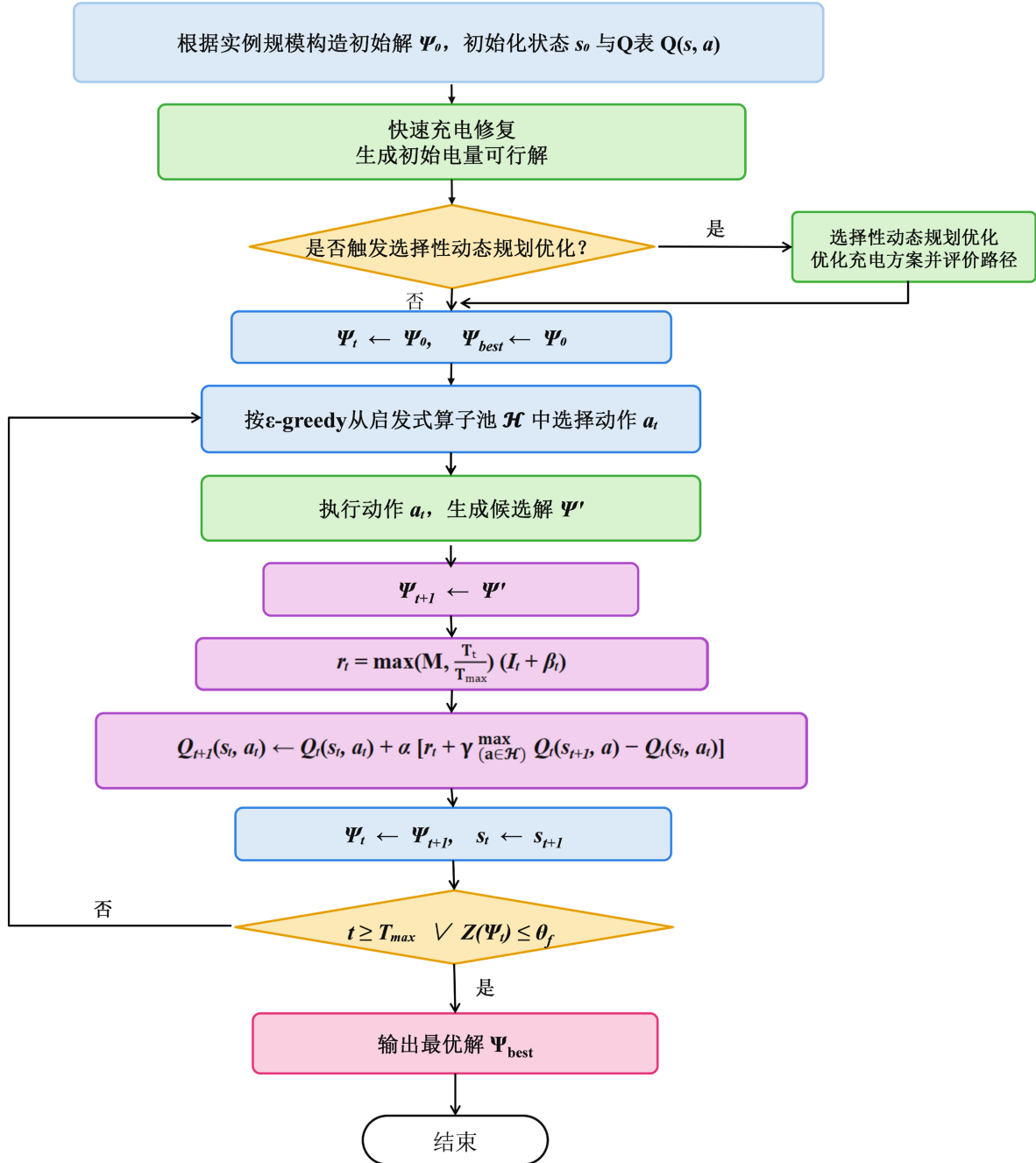


Figure 2. Overall procedure of the proposed algorithm

图 2. 算法总体流程

初始化阶段从最近邻构造、随机客户序列容量构造和极角扫描构造中等概率随机选择一种策略。所

选策略生成满足容量约束的客户路径后，先执行 2-opt 整理，再通过固定路径充电优化形成电量可行解。若路径数量与给定车辆数 K 不一致，则在进入主搜索前进行合并或拆分。重复生成候选解后，选取总距离最短的可行解作为初始解。

客户路径变化可能使原充电方案失效，因此候选解必须在目标值评价前重新完成固定路径充电优化。算法沿客户顺序检查剩余电量，并在无法直接到达下一客户时插入绕行代价较小的可达充电站。形成电量可行路径后，算法删除冗余充电站。启发式修复阶段分别比较原序列与其反向序列的修复结果。确定较优客户序列后，对仍具改进潜力的候选路线，在该客户顺序固定的条件下调用动态规划对充电站组合进行优化。仅当新方案距离更短且满足电量约束时，才替换当前方案。该过程对应固定路径充电优化[6] (图 3)。

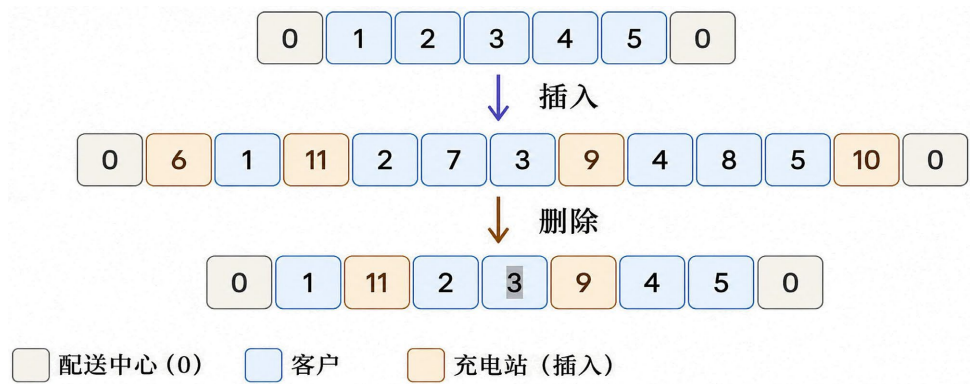


Figure 3. Charging-station insertion and redundant-station removal illustration
图 3. 充电站插入与冗余站点删除示意

主搜索采用 8 种低层启发式。IPS, IPI 和 IO 分别执行路径内客户交换，客户插入与 2-opt。ESS, EPS, EPI 和 Swap-Star 用于跨路径片段交换，客户交换，客户迁移与精细重插入。Ruin-and-Recreate 通过移除并重新插入部分客户形成大尺度扰动。路径间操作先检查容量约束，明显不可行的候选不进入充电处理。动作集合见表 2。

Table 2. Low-level heuristics
表 2. 低层启发式

类别	算子
路径内调整	IPS
路径内调整	IPI
路径内调整	IO
路径间重分配	ESS
路径间重分配	EPS
路径间重分配	EPI
路径间重分配	Swap-Star
破坏重构	Ruin-and-Recreate

算法 1 概括主搜索过程。低层启发式改变客户服务结构，固定路径充电优化恢复并改善电量可行路径，Q-learning 依据完整解反馈调整算子偏好。随机初始化仅用于生成搜索起点，Q-learning 仅作用于搜

索阶段的算子选择，固定路径充电优化仅处理给定客户顺序下的补能决策。

Algorithm 1. Main search loop

算法 1. 主搜索流程

步骤	操作
1	从三种初始化策略中等概率随机选择一种，构造初始客户路径并完成固定路径充电优化。
2	观测当前搜索状态，按 ε -greedy 由 Q-learning 选择一个低层启发式。
3	在客户路径层执行所选算子，并对路径间操作进行容量检查。
4	对容量可行候选执行充电站插入与冗余站点删除，必要时调用动态规划优化固定客户顺序下的充电方案。
5	用修复后的完整路径总距离计算奖励，并更新 $Q(s_t, a_t)$ 。
6	根据接受准则更新当前解，发现更优可行解时更新历史最优解。
7	未达到时间预算时继续搜索，停滞阶段通过破坏重构和重新初始化恢复搜索多样性。

4. 实验结果与分析

4.1. 实验设置

实验采用 IEEE WCCI 2020 CEVRP 基准集中的 17 个公开实例，在表 3 所列对比算法的统一测试环境下开展对比实验，本文算法的主要参数同样汇总于表 3。实例覆盖 E 类与 X 类，客户规模从 22 增至 1001。每个实例给定车辆数 K ，全部结果均要求形成恰好 K 条非空路径，并同时满足客户唯一服务、车辆容量和电量约束。算法使用 C++实现，初始化直接在三类构造策略中等概率随机选择。实验平台配备 Intel Core i7-1165G7 2.80 GHz 处理器。本文算法、各对比算法与搜索模块消融变体在每个实例上均独立运行 30 次。各算法在同一实例上采用相同的时间预算，具体时间设置见表 4。全部实例采用同一组核心参数，不进行实例级定向调参。

Table 3. Main parameters of the proposed algorithm

表 3. 本文算法的主要参数

参数	取值	含义
α	0.90	Q-learning 学习率
γ	0.90	折扣因子
ε	0.10	随机探索率
Q_0	0.01	Q 表初始值
β^+	0.05	正奖励偏置
β^-	-0.10	负奖励偏置
M	0.05	时间因子下界
λ_0	0.05	初始接受阈值
$\lambda_{\min}$	0.001	最小接受阈值
η	0.95	阈值衰减系数
T_d	20	阈值衰减间隔
T_h	150	阈值重热等待步数
T_r	500	重启触发步数
ξ	40	单动作尝试次数
p_2	0.60	2-opt 触发概率

实验报告 Best, Mean 和 Std。Best 为 30 次独立运行中的最小目标值, 用于衡量给定计算预算内可达到的最好解质量。Mean 为 30 次结果的算术平均, 用于评价重复运行的总体表现。Std 为 30 次结果的标准差, 用于描述结果波动。Std 不单独用于判定算法优劣, 而与 Best 和 Mean 结合解释。

4.2. 与已有算法的结果比较

本文将 TAMLs、BACO、CACO-I、CACO-D 和 QHH 作为对比算法。各方法沿用原研究中的参数设置。表 4 报告 Best、Mean 和 Std。从 Best 看, 本文算法在 15 个实例上达到或并列最优, 其中 10 个为严格最优, 5 个为并列最优。X-n214-k11 和 X-n351-k40 的 Best 略高于 QHH。Mean 在 14 个实例上达到或并列最优, 其中 9 个为严格最优, 5 个为并列最优。E 类小规模实例出现较多相同结果, 说明多种算法在有限搜索空间中均能接近同一优质解。规模增加后, 算法差异更易显现。客户数超过 150 的 9 个 X 类实例中, 本文算法在 7 个实例上取得更低 Best, 并在 7 个实例上取得更低 Mean。X-n685-k75、X-n819-k171、X-n916-k207 和 X-n1001-k43 上的优势较为明显。与此同时, X-n214-k11 与 X-n351-k40 的 Best 和 Mean 仍由 QHH 取得更低值, 说明实例规模并非性能差异的唯一来源。

Table 4. Comparison of values on 17 CEVRP instances

表 4. 17 个 CEVRP 实例上的结果对比

No.	Time/h	Metric	TAMLs	BACO	CACO-I	CACO-D	QHH	Ours
1	0.167	Best	384.67	384.67	384.67	384.67	384.67	384.67
		Mean	385.12	384.67	384.67	384.67	384.67	384.67
		Std	0.45	0.00	0.00	0.00	0.00	0.00
2	0.167	Best	571.94	571.94	571.94	571.94	571.94	571.94
		Mean	571.94	571.94	571.94	571.94	571.94	571.94
		Std	0.00	0.00	0.00	0.00	0.00	0.00
3	0.167	Best	509.47	509.47	509.47	509.47	509.47	509.47
		Mean	509.56	509.47	509.47	509.47	522.50	509.47
		Std	0.09	0.00	0.00	0.00	13.03	0.00
4	0.167	Best	840.14	840.56	840.56	840.56	840.14	840.14
		Mean	840.24	841.95	842.24	840.56	840.14	840.14
		Std	0.10	1.41	1.68	0.00	0.00	0.00
5	0.167	Best	529.90	529.90	529.90	529.90	529.90	529.90
		Mean	542.65	530.41	529.90	529.90	529.90	529.90
		Std	12.75	0.51	0.00	0.00	0.00	0.00
6	0.335	Best	692.64	692.64	692.64	692.64	692.64	692.36
		Mean	714.14	693.03	693.43	692.64	692.89	693.66
		Std	21.50	0.39	0.79	0.00	0.34	2.23
7	0.904	Best	855.87	837.89	840.25	838.93	834.22	833.85
		Mean	879.88	845.24	843.53	840.41	837.81	837.63
		Std	24.01	7.35	3.28	1.48	3.59	2.46

续表

8	1.002	Best	16634.27	16064.74	15917.57	15903.53	15897.24	15864.52
		Mean	17514.75	16164.15	16005.88	15946.53	16036.62	15893.00
		Std	880.48	99.41	88.31	43.00	139.38	11.35
9	1.174	Best	11993.25	11199.52	11114.60	11610.88	11078.27	11083.96
		Mean	11993.25	11329.23	11230.00	11713.91	11162.29	11174.74
		Std	0.00	129.71	115.40	103.03	84.02	35.77
10	1.667	Best	27300.61	26594.68	26469.12	27753.89	26409.67	26478.20
		Mean	27322.69	26862.02	26637.70	27968.96	26570.48	26630.16
		Std	22.08	267.34	168.58	215.07	160.81	62.32
11	1.667	Best	26329.44	24900.99	24824.25	25802.71	24561.65	24454.23
		Mean	26329.44	25049.08	24969.84	25928.03	24785.40	24582.69
		Std	0.00	148.09	145.59	125.32	223.75	81.43
12	1.667	Best	53447.51	52349.36	53979.95	52831.60	51551.36	51465.87
		Mean	53447.51	53178.35	54625.80	52951.81	51692.69	51626.76
		Std	0.00	828.99	645.85	120.21	141.33	117.21
13	1.667	Best	71936.75	71726.16	71966.24	75998.30	70916.83	69412.69
		Mean	71947.72	72206.92	72572.22	76498.85	71342.06	69716.47
		Std	10.97	480.76	605.98	500.55	425.23	129.43
14	1.667	Best	80201.48	80437.08	80410.09	82985.97	79322.23	79169.36
		Mean	80201.48	80830.13	80828.41	83291.54	79608.60	79453.93
		Std	0.00	393.05	418.32	305.67	286.37	165.52
15	2.000	Best	166946.98	164933.60	165074.94	166187.29	161881.70	160012.34
		Mean	166948.77	165581.88	166360.50	166525.12	162419.50	160349.22
		Std	1.79	648.28	1285.56	337.83	537.80	126.97
16	2.000	Best	344217.72	343984.51	345428.58	343724.07	336292.36	332373.35
		Mean	344217.72	345744.33	346616.98	344850.12	336850.15	332786.19
		Std	0.00	1759.82	1188.40	1126.05	557.79	258.72
17	2.000	Best	78184.30	76842.26	78040.30	79412.98	75373.23	74638.66
		Mean	78193.33	78186.81	79029.72	79773.83	75848.78	74983.91
		Std	9.03	1344.55	989.42	360.85	475.55	190.60

4.3. 跨实例统计检验

为避免仅依据胜负实例数量判断整体差异,本文以 17 个实例的 Best 和 Mean 为配对观测进行跨实例非参数检验。首先采用 Friedman 总体检验比较六种算法的平均秩,并以 Kendall 协调系数 W 描述跨算法差异的效应强度。该分析使用逐实例汇总值,反映的是跨实例结果分布,不替代基于原始 30 次运行数据的实例内统计分析(表 5)。

Table 5. Friedman average ranks and omnibus test results**表 5.** Friedman 平均秩与总体检验结果

算法或统计量	Best	Mean
TAMLS	4.529	4.853
BACO	3.765	4.000
CACO-I	4.059	3.971
CACO-D	4.471	3.971
QHH	2.353	2.382
Ours	1.824	1.824
χ^2	44.063	36.657
p	2.25×10^{-8}	7.02×10^{-7}
Kendall W	0.518	0.431

Friedman 总体检验在 Best 与 Mean 上均拒绝六种算法结果分布相同的原假设。本文算法在两项指标上的平均秩均为 1.824。Best 与 Mean 的 Kendall 协调系数 W 分别为 0.518 和 0.431，表明观测到的算法差异具有不可忽略的跨实例效应。

在 Friedman 总体检验达到统计显著后，进一步采用双侧 Wilcoxon 符号秩事后检验比较本文算法与各对比算法。显著性水平设为 0.05。零差值采用 Pratt 零差值处理，并对五组成对比较执行 Holm 校正。

表 6 中的数值均为 Holm 校正后的 p 值。

Table 6. Holm-adjusted p -values for two-sided Wilcoxon signed-rank post-hoc tests**表 6.** 双侧 Wilcoxon 符号秩事后检验的 Holm 校正 p 值

比较	Best	Mean
Ours vs TAMLS	0.00321	0.00160
Ours vs BACO	0.00321	0.00453
Ours vs CACO-I	0.00487	0.00453
Ours vs CACO-D	0.00321	0.00453
Ours vs QHH	0.01419	0.01812

全部校正后 p 值均低于 0.05。本文与 QHH 的 Mean 比较为 0.01812，是五组 Mean 比较中最大的校正后 p 值。综合平均秩、效应强度与逐实例结果，更稳妥的结论是本文算法在当前基准上具有总体竞争优势，而非在所有实例上持续占优。

4.4. 消融实验分析

主搜索模块采用移除式消融。Full 表示完整算法。w/o RR 关闭破坏重构。w/o DP 关闭固定客户顺序下的动态规划优化。w/o QL 以随机算子选择替代 Q-learning 算子选择。每个变体在每个实例上独立运行 30 次，其余流程与参数保持不变。表 7 以跨实例平均目标值增幅概括模块影响，正值表示移除模块后总体结果变差。

表 7 表明，三个模块对 Best 与 Mean 的影响不同。移除动态规划优化后 Best 的跨实例平均增幅最大，为 0.144%。移除 Q-learning 算子选择后 Mean 的跨实例平均增幅最大，为 0.062%。移除破坏重构后

Best 的平均增幅为 0.118%。Mean 的平均增幅处于 0.048%至 0.062%之间。结果说明不同模块的作用强度随评价指标而变化。

Table 7. Summary of search-module ablation results

表 7. 搜索模块消融结果汇总

变体	Best 平均增幅/%	Mean 平均增幅/%
w/o RR	0.118	0.048
w/o DP	0.144	0.056
w/o QL	0.098	0.062

图 4 根据表 7 中的逐实例 Best 结果绘制。横轴为实例编号。纵轴表示消融变体 Best 相对于 Full Best 的百分比变化。三条曲线分别对应 w/o RR、w/o DP 和 w/o QL。基线 0 表示与 Full 相同。正值表示移除模块后 Best 变差。负值表示该变体在相应实例上获得了略低于 Full 的 Best。

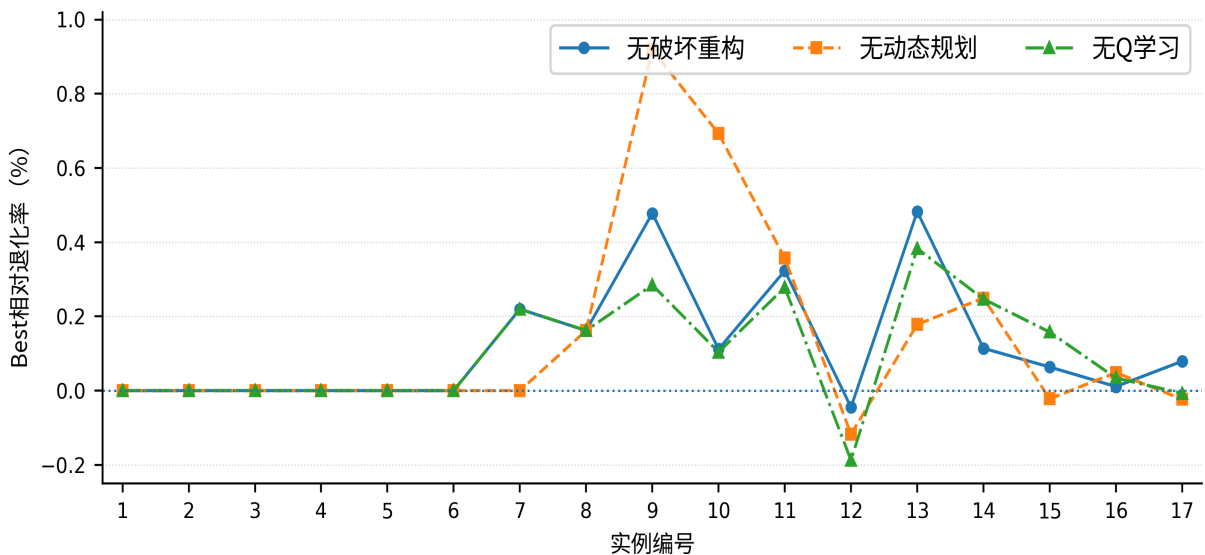


Figure 4. Comparison of per-instance relative degradation across three ablation variants

图 4. 三种消融变体的逐实例相对退化对比

图 4 显示，实例 1 至实例 5 在三个消融设置下均与 Full 一致。差异从实例 6 之后逐步显现。w/o DP 在 X-n214-k11 上的相对退化最大，约为 0.919%。w/o RR 和 w/o QL 在 X-n685-k75 上的相对退化分别约为 0.482%和 0.383%。少数实例出现轻微负值，说明单个实例上移除模块并不必然导致 Best 变差。结合表 7 的跨实例平均结果，三个模块整体上均对 Best 或 Mean 产生正向贡献，但影响程度具有实例差异。

4.5. 约束可行性与固定路径充电优化分析

本文算法在 17 个实例上的全部独立运行均满足客户唯一服务、车辆容量、电量可行性和给定车辆数 K 。为进一步观察固定路径充电优化的作用，选取 9 个代表性实例，比较 30 次运行在初始化阶段与最终解中的平均充电站访问次数 N 和平均充电绕行距离 D 。表 8 中的绝对变化量记为 $|\Delta M|$ 。

9 个实例的充电绕行距离均下降，按表中显示数值计算的改善率介于 30.96%和 80.67%之间。X-n819-k171 与 X-n916-k207 的绕行距离分别减少 1387.23 和 1923.56。充电站访问次数通常随搜索推进而下降，

但 E-n23-k3 由 2.87 次增至 3.00 次，同时绕行距离下降 72.79%。这说明路线改善并不等同于减少访问次数，充电站组合与访问位置同样影响总距离。

Table 8. Changes in number of charging-station visits and charging-detour distance
表 8. 充电站访问次数与充电绕行距离变化

实例	初始 N	最终 N	$ \Delta N $	初始 D	最终 D	ΔD	Gap/%
E-n22-k4	3.17	3.00	0.17	7.59	5.24	2.35	30.96
E-n23-k3	2.87	3.00	0.13	12.42	3.38	9.04	72.79
E-n30-k3	2.07	1.00	1.07	17.19	4.45	12.74	74.11
X-n143-k7	6.47	4.87	1.60	840.43	162.43	678.00	80.67
X-n351-k40	43.80	21.97	21.83	806.78	211.90	594.88	73.74
X-n685-k75	63.60	44.13	19.47	1512.53	428.53	1084.00	71.67
X-n819-k171	234.37	82.70	151.67	1922.66	535.43	1387.23	72.15
X-n916-k207	207.10	119.97	87.13	3870.74	1947.18	1923.56	49.69
X-n1001-k43	30.70	26.10	4.60	1838.11	737.38	1100.73	59.88

4.6. 收敛性分析

为比较不同规模实例上的搜索过程，选取 E-n76-k7，X-n214-k11，X-n749-k98 和 X-n1001-k43 绘制归一化收敛轨迹。纵轴为当前最好目标值相对于各自初始值的比值，因而不同量级实例可在同一图中比较。

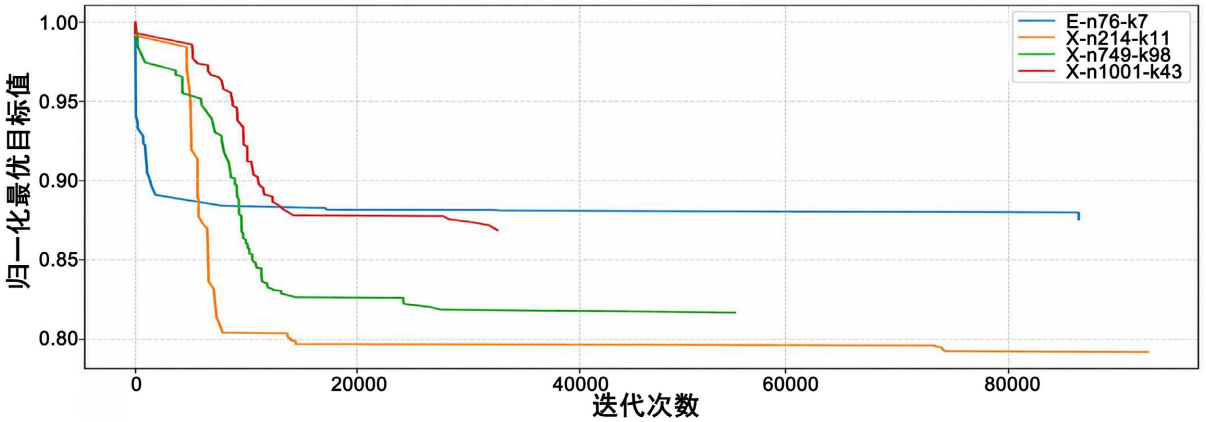


Figure 5. Normalized convergence trajectories on representative instances
图 5. 代表性实例的归一化收敛轨迹

图 5 显示，四个实例在搜索早期均出现较快下降，随后进入幅度更小的阶梯式改进。中大规模实例在后半程仍能产生新的最好解，说明搜索并未在初期改进后立即停滞。该现象与破坏重构和状态相关算子选择所提供的持续搜索能力一致，但收敛曲线本身不能单独区分各模块的因果贡献。

4.7. Q-Learning 机制分析

Q-learning 依据 Bellman 最优关系迭代更新状态动作价值，其有限状态动作空间下的收敛行为可由离散时间切换系统进行分析[16]。Q 表记录不同搜索状态下低层启发式的历史回报。全部状态动作价值初始

为 0.01, 因此搜索开始时不存在预设动作优先级。候选客户路径完成容量检查与固定路径充电优化后, 算法根据完整电量可行路径的目标值变化计算奖励, 并更新本次访问的状态动作对。

机制分析用于解释动作偏好的形成, 不能替代性能验证。表 7 显示, 移除 Q-learning 后 Mean 的跨实例平均增幅为 0.062%, 在三个搜索变体中最大。初始化配置提供搜索起点, 客户路径调整改变服务结构, 固定路径充电优化处理补能方案, 最终结果由各阶段连续作用形成。

状态表示直接影响 Q-learning 的学习效果, 只有能够区分搜索进展的状态才能引导算子选择。搜索状态由连续两次评价的目标值变化刻画。设第 t 次迭代的当前解目标值为 f_t , 上一阶段目标值为 f_{t-1} , 变化指标 P_t 按式(10)计算, 表示当前解相较上一阶段的相对变化幅度。 P_t 大于 0 表示解得到改善, 小于 0 表示解发生退化, 搜索状态按 P_t 的取值划分为 $(-\infty, -3.5)$, $[-3.5, -1)$, $[-1, 1)$, $[1, 3.5)$ 与 $[3.5, +\infty)$ 五个区间。

$$P_t = -\frac{f_t - f_{t-1}}{f_{t-1}} \times 100 \quad (10)$$

奖励决定 Q 表对动作的强化方向, 其计算基准应与最终优化目标一致。奖励以候选解经过充电修复后的实际总距离计算, 而非客户路径距离, 使学习信号与完整解质量保持一致。设动作执行前目标值为 f_{t-1} , 本轮最佳候选解目标值为 $f_{b,t}$, 相对改善率按式(11)计算。搜索后期相同幅度的改进更难获得, 奖励引入时间因子 $\frac{t}{T}$ 保持后期的奖励敏感度, 其中 t 为当前执行时间, T 为最大执行时间。

$$r_t = \max\left(M, \frac{t}{T}\right) \left(\frac{f_{t-1} - f_{b,t}}{f_{t-1}} + \beta \right) \quad (11)$$

5. 讨论

三种构造策略的等概率随机选择能够以较低成本提供不同结构的搜索起点。初始化主要用于保持起点多样性, 最终解质量仍取决于后续客户路径搜索、固定路径充电优化和 Q-learning 算子调度。

完整对比与跨实例检验给出了互补证据。本文算法在 Best 和 Mean 上的平均秩均为 1.824。Friedman 总体检验的 Kendall 协调系数 W 分别为 0.518 和 0.431, 说明算法差异不仅来自少数实例。与五种对比算法的双侧 Wilcoxon 符号秩事后检验经 Holm 校正后均保持在 0.05 以下, 其中与 QHH 的 Mean 比较的 Holm 校正后 p 值为 0.01812。因此, 结果支持本文算法具有跨实例竞争力。本文采用线性能耗与完全充电假设, 尚未考虑时间窗、非线性充电过程和充电站容量限制。

6. 结论

本文针对 CEVRP 提出一种改进的 Q-learning 超启发式。初始化阶段从三种传统构造策略中等概率随机选择一种, 生成容量可行的客户路径。客户路径变化后, 固定路径充电优化先通过充电站插入恢复电量可行性, 再以动态规划优化固定客户顺序下的充电方案。主搜索采用多尺度低层启发式, 并由 Q-learning 根据完整可行解的目标值反馈选择算子。

在 IEEE WCCI 2020 的 17 个实例上, 本文算法的 Best 在 15 个实例达到或并列最优, Mean 在 14 个实例达到或并列最优。跨实例统计检验进一步支持总体差异。移除式消融显示, 动态规划优化对部分实例的最好结果影响更突出, Q-learning 对重复运行后的平均表现影响更直接, 破坏重构则支持大尺度客户路径调整。本文算法在 17 个实例上的全部独立运行均满足给定车辆数、容量和电量约束。在所选 9 个代表性实例中, 平均充电绕行距离均得到缩短。收敛轨迹显示部分中大规模实例在搜索后期仍能继续产生改进。

基金项目

重庆科技大学研究生创新计划项目(YKJCX2521216)。

参考文献

- [1] Cataldo-Díaz, C., Linfati, R. and Escobar, J.W. (2023) Mathematical Models for the Electric Vehicle Routing Problem with Time Windows Considering Different Aspects of the Charging Process. *Operational Research*, **24**, Article No. 1. <https://doi.org/10.1007/s12351-023-00806-5>
- [2] Ceselli, A., Felipe, Á., Ortuño, M.T., Righini, G. and Tirado, G. (2021) A Branch-and-Cut-and-Price Algorithm for the Electric Vehicle Routing Problem with Multiple Technologies. *Operations Research Forum*, **2**, Article No. 8. <https://doi.org/10.1007/s43069-020-00052-x>
- [3] Froger, A., Mendoza, J.E., Jabali, O. and Laporte, G. (2019) Improved Formulations and Algorithmic Components for the Electric Vehicle Routing Problem with Nonlinear Charging Functions. *Computers & Operations Research*, **104**, 256-294. <https://doi.org/10.1016/j.cor.2018.12.013>
- [4] Lysgaard, J., Letchford, A.N. and Eglese, R.W. (2004) A New Branch-and-Cut Algorithm for the Capacitated Vehicle Routing Problem. *Mathematical Programming*, **100**, 423-445. <https://doi.org/10.1007/s10107-003-0481-8>
- [5] Rastani, S. and Çatay, B. (2021) A Large Neighborhood Search-Based Matheuristic for the Load-Dependent Electric Vehicle Routing Problem with Time Windows. *Annals of Operations Research*, **324**, 761-793. <https://doi.org/10.1007/s10479-021-04320-9>
- [6] Kullman, N.D., Froger, A., Mendoza, J.E. and Goodson, J.C. (2021) Frcvpy: An Open-Source Solver for the Fixed Route Vehicle Charging Problem. *INFORMS Journal on Computing*, **33**, 1277-1283. <https://doi.org/10.1287/ijoc.2020.1035>
- [7] Bacci, T., Gentile, C. and Pizzari, E. (2026) On Tackling Non-Linear Charging Functions in the Electric Vehicle Routing Problem. *Optimization Letters*. <https://doi.org/10.1007/s11590-026-02319-4>
- [8] Tarhini, A., Danach, K. and Harfouche, A. (2020) Swarm Intelligence-Based Hyper-Heuristic for the Vehicle Routing Problem with Prioritized Customers. *Annals of Operations Research*, **308**, 549-570. <https://doi.org/10.1007/s10479-020-03625-5>
- [9] Fitzpatrick, J., Ajwani, D. and Carroll, P. (2024) A Scalable Learning Approach for the Capacitated Vehicle Routing Problem. *Computers & Operations Research*, **171**, Article ID: 106787. <https://doi.org/10.1016/j.cor.2024.106787>
- [10] Pulyassary, H., Kollias, K., Schild, A., Shmoys, D. and Wu, M. (2025) Network Flow Problems with Electric Vehicles. *Mathematical Programming*, **216**, 489-520. <https://doi.org/10.1007/s10107-025-02295-0>
- [11] Karimi-Mamaghan, M., Mohammadi, M., Meyer, P., Karimi-Mamaghan, A.M. and Talbi, E. (2022) Machine Learning at the Service of Meta-Heuristics for Solving Combinatorial Optimization Problems: A State-of-the-Art. *European Journal of Operational Research*, **296**, 393-422. <https://doi.org/10.1016/j.ejor.2021.04.032>
- [12] Burke, E.K., Gendreau, M., Hyde, M., Kendall, G., Ochoa, G., Özcan, E., et al. (2013) Hyper-Heuristics: A Survey of the State of the Art. *Journal of the Operational Research Society*, **64**, 1695-1724. <https://doi.org/10.1057/jors.2013.71>
- [13] Alissa, M., Sim, K. and Hart, E. (2023) Automated Algorithm Selection: From Feature-Based to Feature-Free Approaches. *Journal of Heuristics*, **29**, 1-38. <https://doi.org/10.1007/s10732-022-09505-4>
- [14] Drake, J.H., Kheiri, A., Özcan, E. and Burke, E.K. (2020) Recent Advances in Selection Hyper-Heuristics. *European Journal of Operational Research*, **285**, 405-428. <https://doi.org/10.1016/j.ejor.2019.07.073>
- [15] Meignan, D., Koukam, A. and Créput, J. (2009) Coalition-Based Metaheuristic: A Self-Adaptive Metaheuristic Using Reinforcement Learning and Mimeticism. *Journal of Heuristics*, **16**, 859-879. <https://doi.org/10.1007/s10732-009-9121-7>
- [16] Lee, D., Hu, J. and He, N. (2023) A Discrete-Time Switching System Analysis of Q-Learning. *SIAM Journal on Control and Optimization*, **61**, 1861-1880. <https://doi.org/10.1137/22m1489976>